Multicriteria Competency Clustering Framework

Exploiting employees' competencies for Strategic Workforce Planning

by
R.D. van der Velden
(SNR: 1271672)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Business Analytics and Operations Research

Tilburg School of Economics and Management
Tilburg University

Supervised by:
dr. Emiel Caron
MSc. Jasper Kunst AAG (PwC NL)

Second reader:
dr. Juan Vera Lizcano

May 22, 2022

Acknowledgements

This thesis is the final element to the completion of my academic career, which started 7 years ago. From the moment I started the bachelor's program in Econometrics and Operation Research at Tilburg University, I knew I was in the right place. Countless lessons were learned and memories and friendships were made over the years. I have also had the privilege to engage in many extracurricular activities connected to the study programs and TiSEM faculty. Starting right at the first month of my studies, by joining the program's Sounding Board; whilst doing an exchange and some board years and jobs in between; this grew into participating in the main participation body within the faculty, the Faculty Council during my master's.

All of this would not have been possible without a number of people. Firstly, I would like to thank Emiel Caron for our first, unpresumptuous and spontaneous sparring session about HR analytics. Which evolved into Emiel supervising me through the academic field for the last 8 months, of which this thesis is the final result. Secondly, I would like to thank Jasper Kunst for being my external supervisor at PwC for his time, our weekly meetings, his persistent questions about my planning and for challenging my ideas. I also want to thank the Data & Analytics team at PwC Consulting for making me feel welcome right from the start. In particular Charlotte Bech, for introducing me to the topic of Strategic Workforce Planning and sharing her knowledge and experience in the field, both in academics and in practice. I am looking forwards to starting my professional career in this team.

I also would like to express my gratitude towards my friends, not only during the process of writing this thesis, but mostly for their support, friendship, laughter and tears throughout the past 7 years. Lastly, I would like to thank my family and boyfriend Guus for their unconditional love and support; even when life isn't always fair.

Rachel van der Velden

- Utrecht, May 17, 2022

Abstract

An organization's workforce is subject to many uncertainties and changes; engaging in Strategic Workforce Planning (SWP) is vital for an organization's health. Technological developments cause the current approach to SWP to not be future-proof, since job titles may no longer be representative for long term strategic plans. To provide a more robust description of the workforce, we proposes to describe employees by their competencies. This research extends the current approach to SWP by exploiting employees' competencies.

To this end, we propose to group employees into employee profiles based on their competencies. We find that it is valid to describe and differentiate employees by their competencies, which, at the moment, is not done in a quantitative Strategic Workforce Planning setting. Employees' competency data exhibits two features: multi-dimensionality and high dimensional and low sample size. In the literature, there are no directly applicable methods. Therefore, we propose a framework for employee profile design: the Multicriteria Competency clustering framework; which deals with the specific features of the competency data. Next, we present methods for integrating employee profiles into the SWP approach; which allows for more sophisticated internal mobility plans, resulting in a more efficient use of the available resources. Finally, a case study illustrates the use of the new methods. Results from both a mathematical and an HR perspective prove its interpretability, explainability and applicability.

Keywords: Strategic Workforce Planning, Competency Management, Competency data, clustering, multi-criteria data, high dimensionality & low sample size data

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ASWP Analytical Strategic Workforce Planning.

CQM Cluster-quality measure.

DA Deductive Algorithm.

FTE Full-time equivalent.

HDLSS High Dimensionality and Low Sample Size.

HR Human Resources.

HRA Human Resources Analytics.

HRIS Human Resources Information Systems.

HRM Human Resources Management.

LP Linear Programming.

MCC Multicriteria Competency.

MCDA Multicriteria decision aid.

MDP Markov Decision Process.

O*NET Occupational Information Network.

OO Ordinal Optimization.

OR Operations Research.

PCA Principal Component Analysis.

SDG Sustainable Development Goals.

SOC Standard Occupational Classification.

SWP Strategic Workforce Planning.

WPP Workforce Planning Process.

1 — Introduction

1.1 The Future of Work is changing

A company's workforce is subject to many uncertainties and external- and internal changes ([13]). Think of technological developments, legal standards, societal demands, - needs and - changes and the company's strategic intent; below, these topics and their influence on a company's workforce are presented.

Technological developments change the demand for employees. Take for example a supermarket in the Netherlands: where there used to be 5 cashiers, there now are 2 cashiers and 6 self checkout systems. Also, the type of employee that companies are looking for has changed, as well as their required skills: where before a marketeer had to implement commercials through local TV and -newspaper, marketeers nowadays need to navigate their way through many different (social) media channels, using Marketing Analytics and Big Data. This increasing demand for digital skills (even accelerated by the changes in work life due to Covid [52]) and the automation of jobs, highly impact the composition of a company's workforce: generally, the portion of "administrative" employees goes down, where the amount of "executive" employees goes up.

Legal Standards also influence a company's workforce. New legislation, such as the recently approved women's quota for supervisory boards of Dutch stock market listed companies or laws around the retirement age, should be taken into account.

Societal demands, needs and changes too are an external change that influences a companies (future) workforce. There is an increasing societal need for more diversity in the workforce, not only regarding gender, but also regarding race, culture, religion, age, sexual orientation and physical abilities . Also, recent graduates have a different mindset than the current workforce. They give and expect more attention and priority to a healthy work-life balance, flexible working hours and a company's sustainability goals ([23]). Another societal influence is the changes of a country's entire labor force over time. Demographic shifts such as aging and fluctuations in the growth or shrinkage of the labor force impact the labor market.

Additionally, the company's own *strategic intent* requires the company's workforce to be changed over time. When the company's goal is to increase production of some product, more resources are needed, which includes labor. Also, when aiming to become more innovative, a company needs to invest in adequate financial resources, as well as attracting new talented employees such that their research and development department is able to scaled up.

Engaging and using these transformations to one's advantage and implementing the suitable Human Resources strategy is vital for an organization's health ([54]). Without the right employees, an organization is unlikely to achieve its strategic goals or sustain their current enterprise. Especially in the current "War for Talent" it is crucial that companies attract, hire and retain employees ([39]). Nonetheless, a survey conducted by PwC has shown that 68% of organisations do not consistently take a strategic scenario-based approach to workforce plans ([55]).

This strategic scenario-based approach, the method that aims to provide actionable insights in the workforce development in order to enable realisation of the organization's strategy, is *Strategic Workforce Planning* (SWP). By engaging in Strategic Workforce Planning, a company commits to a time-consuming and organization-changing process. It focuses on long term strategic decision making in order to ensure the right people, at the right time and at the right costs, are part of the organization.

At this moment, approaches and algorithms used for Strategic Workforce Planning are often based on the different job titles within a company. However, given the uncertainties and changes described above, can we depend on these jobs titles? Research has shown that digitalization and evolving skills needed in the workplace are putting nearly 1.6 million jobs in the Netherlands at risk of becoming obsolete in the medium to long term ([53]). These "zombie jobs", such as retail sales associates, waiters and call center agents, are highly susceptible to automation ([53]). As a consequence of constantly evolving jobs and their requirements, these job titles may no longer be representative for long term strategic plans.

1.2 Purpose of this research

This research aims at formulating a more future-proof and robust approach to Strategic Workforce Planning by extending a SWP approach. As introduced in the section above, with the rising technological developments, job titles may no longer be adequate descriptors of a company's workforce. Instead, we want to use other characteristics to describe the workforce. From this, the main research question follows:

Main RQ: *How can we extend the approach for Strategic Workforce Planning, by exploiting employees' competencies?*

This overarching research question can be split up in the research questions presented below. Each of these questions tackles an aspect of the required research in order to adequately assess the overall research objective. The order and sequentially of these questions also give rise to the structure of this thesis.

RQ

1: In relation to Human Resources Management and - Analytics, what exactly is *Strategic Workforce Planning* and how does it add value to companies?

2: What is the current approach, and which *HR tools or principles* should be used to extend it?

3: What are *competencies* and *competency data*, and do appropriate quantitative methods exists for grouping employees based on their competencies?

4: Are we able to create *new clustering methods* to design employee profiles?

5: How can we *integrate* these employee profiles into the Strategic Workforce Planning approach?

6: Can we apply these new methods in practice by means of a *case study*?

1.3 Research approach

By answering the research questions, we get closer to extending the approach to Strategic Workforce Planning. In this research, we plan to do so as follows. First, we need to find out why exactly a company's workforce is important and what value it brings. In order to do this, we dive into Human Resources Management literature to review the role and added value of HR departments within companies. Obtaining this foundation on HR, allows us to learn about Strategic Workforce Planning and why it is important for companies to engage in. Next, we go more into the analytical steps of the SWP approach and see where these methods belong within the field of Operations Research and why there is a need for extending them. This allows us to propose an extension to the current approach. As said before, job titles may no longer be adequate measures for a company's long term workforce. Instead, we find that the workforce can be better described by the actual employees performing those jobs. Consequently, we find that those employees are best described by their *competencies*, i.e. the set of knowledge, skills and abilities they bring to the company. Once we describe a workforce in terms of their competencies, we are able to look past these job definitions. In order to find which employees are similar to each other, we want to group the employees based on their competencies to create employee profiles, in a quantitative manner. To do so, we first explore competency data, from which we find that we need to account for several features of the data when grouping employees. Later, we create our own methodology, that deals with these unique data features. Next, we purpose methods to integrate this into the existing SWP approach. Finally, we perform these new methods in a case study on a company within the airline industry.

1.4 Research Relevance

This section discusses the relevance of this research, from an academic-, business-, and societal point of view.

Academic relevance

Firstly, this research is relevant to Human Resources literature; as we bridge the gap between Competency Management and quantitative approaches to SWP. We validate and motivate the use of competencies to describe employees in a strategic setting and propose to extend the current approach to SWP by using employee profiles. Thereafter, this research adds to the field of Operations Research by creating a new framework for designing employee profiles based on employees' competency data, dealing with specific data features. Additionally, we propose method on how these employee profile can be integrated into the SWP approach.

Business relevance

From a business perspective, this research creates new methods and ways to look at a workforce. Evaluating individuals based on their competencies allows companies to think beyond jobs, functions or departments. It gives rise to a new level of insights; instead of knowing "how many FTEs are spend on which function?", one could answer the question: "what are the capabilities of my workforce?". As such, one makes use of the available resources more efficiently and thus reduces costs. This is desirable, since it is much harder and more expensive to fire current- and hire new staff, than it is to retain and retrain a current employee. In fact, research shows that new recruits from outside the company, take up to three years to do as well as internal candidates ([21]). Also, this has a positive effect on Employee Engagement, resulting in a happier workforce.

Societal relevance

This research is also relevant from a societal perspective. If all companies would engage more and better in their workforce and its development, this would have an impact on the labor force in general. Companies knowing and planning which competencies they want and need their employees to possess, ensures in time up- and reskilling within the company itself. Then, more employees will have adequate skills, and a company is less likely to fire employees, and hire new ones. As a result, employees will have more continuous career development paths, resulting in a more stable labor force.

In addition, this contributes to the Sustainable Development Goals (SDGs) of the United Nations[1]. Namely, SDG 8, which is about "promoting sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all". More specificially, this research adds to target

[1] https://sdgs.un.org/goals

8.2: “Achieve higher levels of economic productivity through diversification, technological upgrading and innovation, including through a focus on high-value added and labour-intensive sectors”. Also it is inline with target 4.4, part of SDG 4 on Quality Education: “By 2030, increase by x% the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship”.

1.5 Research Outline

The remainder of this research is structured as follows. Chapter 2 provides the foundation, relevant literature and definitions on Human Resources Management, HR Analytics and Strategic Workforce Planning. Chapter 3 goes into more analytical detail on the existing Strategic Workforce Planning approach. Subsequently, we propose an extension to the approach in order for it to be more robust to (technological) developments. Appendix A refers to both Chapter 2 and 3. Next, Chapter 4 explores the definitions and use cases of this HR principle, together with the corresponding data and quantitative methods for grouping employees. New methodologies are proposed in Chapter 5 and 6, to design and integrte employee profiles into the SWP approach. Appendix B refers to both Chapter 4 and 5. In Chapter 7, a case study is performed where these new methodologies are tested and results are presented. Appendix C refers to the case study in Chapter 7. Following the results, the research questions are concluded upon in Chapter 8; along with recommendations for business application, contribution to literature, limitations of the research and avenues for future research.

2 — Human Resources Management & Strategic Workforce Planning

This chapter provides an introduction to the field of Human Resources Management (HRM). Many readers of this research will have profound knowledge of the analytical part of this research, but have no knowledge of the field of HRM. For that reason, we also go into the reasons why and how HRM is relevant for organizations. Next, the field of Human Resources- or People Analytics and its history is explored in more detail. Finally, the sub field of "Strategic Workforce Planning" is elaborated upon.

2.1 Introduction to Human Resources Management

The role of the Human Resources function is widely studied and is constantly changing. When thinking of your own interactions with "HR", it is very likely that mostly administrative tasks come to your mind. Tasks such as recruiting new candidates, processing payrolls and managing your remaining days off, are likely to be on the list. However, HR is much more than that. Think of more strategic tasks: improving and maintaining employee engagement and productivity, aligning the company culture and values and managing employees' performance and aligning this with training, upskilling and promotion trajectories.

These different aspects of the added value of HR have been contained in a framework by Ulrich ([65]). The framework is comprised of four key roles or 'result domains' that HR professionals have to fulfil; of which an overview is presented in Figure 2.1 below. According to Ulrich, HRM has to deliver results in each of these domains, since each are equally important. The framework says that HR professionals must learn to be both strategic and operational, focusing on the long and short term. The horizontal axis, in it's turn, represents that the activities of HR professionals should range from managing processes (HR tools and systems) to managing people.

	Future/Strategic Focus		
Processes	Strategic Partner: Management of Strategic Human Resources	Change Agent: Management of Transformation and Change	**People**
	Administrative Expert: Management of Firm Infrastructure	Employee Champion: Management of Employee Contribution	
	Day-to-day/Operational Focus		

Figure 2.1: Ulrich's multiple-role framework on the added value of the HR function: four result domains

To better understand these roles, Ulrich describes these roles in terms of three dimensions: the *deliverables* that constitute the outcome of the role; the *metaphor* or visual image that accompanies the role and the *activities* the HR professional must perform to fulfill the role. The descriptions of the different roles is presented below.

Role	Deliverable	Metaphor	Activity
Management of Strategic Human Resources	Executing strategy	Strategic Partner	Aligning HR and business strategy: "Organizational diagnosis"
Management of Transformation and Change	Creating a renewed organization	Change Agent	Managing transformation and change: "Ensuring capacity for change"
Management of Firm Infrastructure	Building an efficient infrastructure	Administrative Expert	Reengineering Organization Processes: "Shared services"
Management of Employee Contribution	Increasing employee commitment and capability	Employee Champion	Listening and responding to Employee: "Providing resources to employees"

Table 2.1: Ulrich's multiple-role framework on the added value of the HR function: definitions of HR roles

Hailey, Farndale and Truss ([32]) assessed the role of the HR department on organizational performance. They identified and discussed the inherent conflict in Ulrich's framework between the process-oriented and people-oriented roles, already noticed by other research. Their study supported this by finding how the HR department may become more important strategically, but the human factor of people's everyday work experience may deteriorate. Subsequently, they find that HR professionals' predominant focus on being a strategic partner is related to longer-term damage to the financial performance of organizations.

Buyens and De Vos ([10]) researched the perceptions of the value of the HR function and find an inconsistency between the literature and practice. In the literature it is often claimed that HRM would only add value to the company if it was a full strategic partner. However, in their own research, they conclude that the HR professionals are working on highly diverse tasks, some of them being purely administrative and others being very strategic. They extend Ulrich's framework on the added value of the HR function, by proposing a framework on the perceived value of the HR function.
They do this by first distinguishing four different stages of involvement in the decision-making processes: problem definition (very early), development of a solution (early), implementation (late) and control (very late). Buyens and De Vos claim that HR departments deliver value at each stage of the decision-making process; only different capabilities are needed. Below in Table 2.2 the different stages and their characteristics are depicted.

Stage of decision-making process	Involvement of HR function	Characteristics
Very early: problem definition	Value-driven HRM	Anticipative; Recognise and determine; Give meaning
Early: development of a solution	Timely involvement of HRM	Active adaption; Conceptual understanding; Instrumental
Late: implementation	Executive HRM	Passive adaptation; Executing; Here-and-now problem solving
Very late: control	Reactive HRM	Reactive; Glue; Resolve misfits

Table 2.2: Buyens and De Vos' framework on the perceived added value of the HR function: involvement of HRM in decision-making processes

Secondly, they apply this framework to the four result domains designed by Ulrich. For each role, they study the impact of the HR professionals, whether they were involved in discussions and from what

point in time they were involved in a certain process; by researching the perceptions of HR managers, top managers and line managers. Their research concludes that the HR function delivers value within different areas of the company, ranging from administrative to strategy formulation, confirming the multiple-role model designed by Ulrich.
Their newly designed framework on the perceived value of the HR function integrates the four domains in which HRM delivers value with the four stages of involvement in decision-making processes. Their framework is presented below, in Figure 2.2. It presents the employees in the center, as a pivot on which the HR policies have to be based. Ulrich's "four domains in which the HR function offers added value" are centered on this core. The exterior circle is not static, but moves around the four HR roles. This indicates that each HR role is more or less involved in a decision-making process.

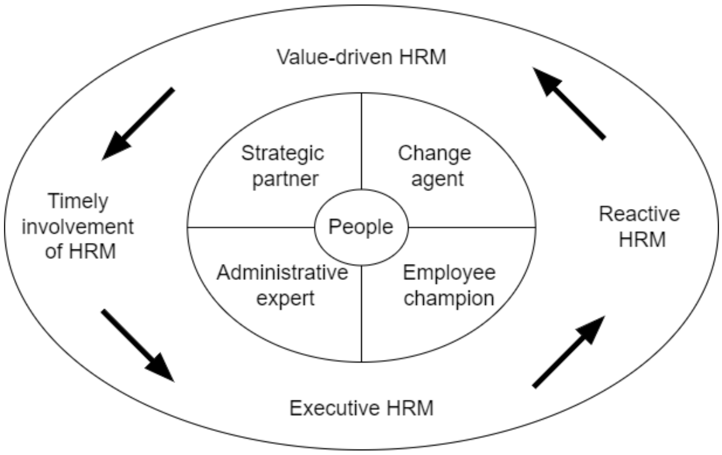

Figure 2.2: Buyens and De Vos' framework on the perceived added value of the HR function: Integrated involvement of the HR function

All in all, it is clear that HR departments add values to companies in more than one way and that it is vital for an organization's health.

2.2 Human Resources Analytics

Traditionally, HR is a field dominated by administrative tasks, soft skills and "gut feeling". It is only recent that data and analytics have come into play. According to an evidence-based review of "HR Analytics" (HRA) literature by Marler and Boudreau in 2017 ([45]), the term "HR Analytics" first appeared in the HR published literature only in 2003-2004. They found that just four articles were deemed to be of sufficient quantitative level and concluded that there is still much room for academic researchers to add to the HR Analytics literature and conversation.

Also, there seems to be an ongoing debate on the appropriate name for the field, as the terms "HR Analytics", "Talent Analytics", "People Analytics", "Workforce Analytics" and "Human Capital Analytics" are all used alongside each other. In business, the fields are perceived quite differently. In 2015, Josh Bersin (one of the leading thinkers in HRA) claimed in a Forbes blog ([8]) that "HR Analytics solves HR problems that seem interesting to HR managers, typically business people don't care; People Analytics solves real-world business problems that help run the company better." and even said that he does not believe People Analytics should belong in the field of HR or with an organization's HR department.

In 2021, Margherita ([44]) advanced the review by Marler and Boudreau ([45]). They considered a larger population of articles by dropping some of the constraints. Multiple key words were used to search articles, amongst others: HR analytics, Workforce analytics, Talent analytics and Human capital analytics. Firstly, they studied the different definitions of HR analytics over the years. Table A.1 in the Appendix A.1 presents some of the most structured definitions found in the literature; along with references and (in italic) some peculiar aspects that are identified in the definition, gathered by Margherita (2021, [44]). For this research, we use the most recent definition by Kryscynski et al., 2017 ([41]):

> *Data, metrics, statistics and scientific methods*, with the help of technology, to gauge the impact of human capital management practices on business goals.

Secondly, they derived a structured inventory of 106 specific concepts framed within a purposeful classification framework. They built a more comprehensive identification of "enablers", "applications" and "value creation drivers" associated to HR analytics. These concepts and sources are presented in Table A.2 in Appendix A.1. Many of these concepts are actually optimization problems, such as "job scheduling", "expertise recommendation and allocation" and "Voluntary turnover prediction".

Notice how all four of Ulrich's roles are represented in these concepts and sources of HR analytics. For example, "Data-driven decision making" (#95) and "Workforce forecasting modelling" (#67) are easily related to the "Strategic Partner" role; "Employee sentiment analysis" (#45) and "Support to organizational change management" (#106) to the "Change Agent" role; "Workplace attendance, accidents, injuries tracking" (#68) and "Employee performance evaluation and rewards" to the "Administrative Expert" and "Proactive predicting decision on people matters" (#61) and "Improved employee experience" (#82) to the "Employee Champion". This suggests that the concepts and sources related to applications and values in HR analytics, are contained within Ulrich's multiple-role framework on the added value of the HR function. As a result, we see that HR analytics is able to add value in all aspects of the HR function; making the term "HR analytics" the most all-encompassing, hence it is used for the remainder of this research.

The rise of HR analytics is preceded by the increasing utilization of IT within the HR department (Gardner, 2003 [28]). HR departments had an increasing demand to adopt computer technology in order to more effectively and efficiently process for example employee information (Kavanagh & Johnson, 2017 [38]). To this end, *Human Resources Information Systems (HRIS)* are designed. Chauhan et al., (2011, [12]) define a HRIS as "the integration of software, hardware, support functions and system policies and procedures into an automate process designed to support the strategy and operational activities of the HR department and managers throughout the organization". Also, they identify three types of HRIS and their corresponding applications. Firstly, *operational* HRIS provide data to support routine and repetitive HR decisions. Secondly, *tactical* HRIS provide support for decisions that emphasize the allocation of resources. Finally, *strategic* HRIS provides significant value increased by different alignment of business processes and product lines with the strategic objectives of the organization (Agiu et al., 2014 [3]). Again, we see the different roles of the HR function represented.

Now the reader has a general overview of the literature, as well as the different tasks and values of the field of Human Resources- Management and Analytics, we explore the field of Strategic Workforce Planning.

2.3 Strategic Workforce Planning

Strategic Workforce Planning (SWP) focuses on long term HR decision making, for instance 6, 12, 36, 60 or 120 months. In this research, the following definition[1] is used:

> *Strategic Workforce Planning* provides actionable insights in the workforce development to enable realisation of the organization's strategy. It does so by developing an aligned set of HRM policies and practices that ensure the appropriate workforce, at the right costs, is available when needed.

It is important for companies to invest in and prioritize SWP. Reasons include, but are not limited to, achievement of business goals and objectives, financial benefits and improvement of the employee experience. For example, SWP factors in internal and external changes and allows for finding, hiring and retaining[2] the right people, with the right skills, at the right time, at the right costs. A lack of a strategic succession planning could affect productivity and retention of top performers, while the lack of well-being, health, and safety policies could undermine engagement, performance, and productivity of the workforce and increase operating costs (Isson and Harriott, 2016 [34]).

An appropriate approach to SWP differs for each organization, as each has its own (future) challenges. Phillips and Gully (2015, [51]) present a generic approach to the "Workforce Planning Process" (WPP), shown below in Figure 2.3. They provide a general framework that is applicable to many types of organizations.

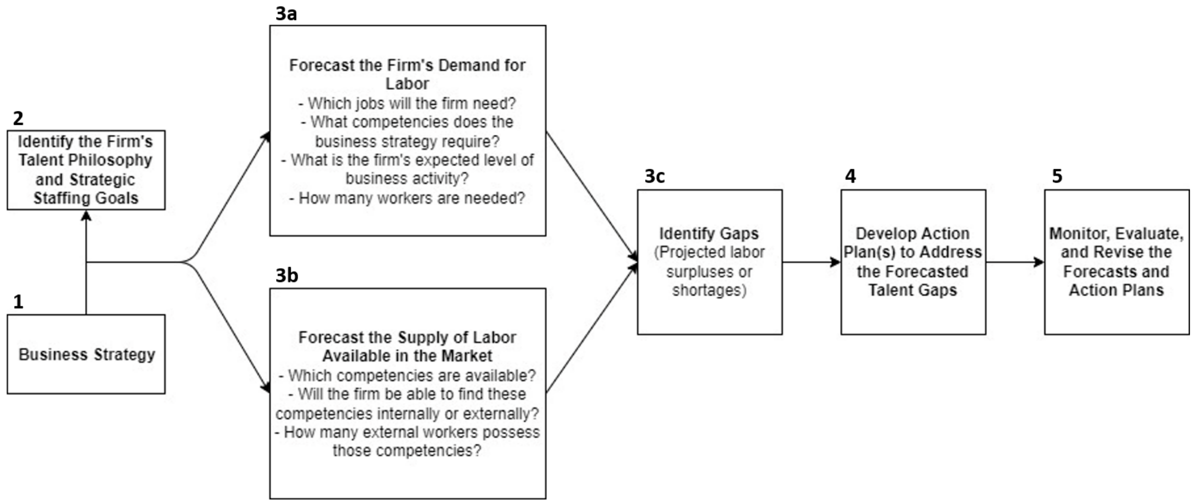

Figure 2.3: Phillips and Gully's (2012, [51]) approach to the "Workforce Planning Process"

The first two steps focus on the organization itself: what is the strategic vision, mission and business strategy, but also, what are their staffing preferences, i.e. how do they deal with promotions, demotions and other staffing decisions. The third step is to conduct a workforce analysis, where both the labor demand and -supply is forecasted and gaps between those two are identified. For forecasting an organization's labor demand, they discuss five of the most commonly used types of information to evaluate general business trends in the economy: seasonal factors, interest rates, currency exchange rates, competitive changes, and industry and economic forecasts. For forecasting an organization's labor supply, they describe how this is influenced by the organization's internal and external labor markets. The

[1] Based on: Isson and Harriott (2016, [34]); Taylor (2005, [60]); De Bruecker (2015, [17]); Emmerichs et al. (2004, [22]); Phillips and Gully (2012, [51]); Kunc (2008, [42])

[2] Retaining an employee is when that employee keeps working at the organization

fourth step is to develop action plans to address these gaps, consistent with their talent philosophy. The fifth step is to evaluate how effective the organization's workforce plan has been in terms of meeting the company's recruiting and hiring goals.

From a managerial perspective, Willis, Cave and Kunc (2018, [69]) provide extensive arguments as to why the steps 1, 2 and 5 are crucial for a sustainable engagement to SWP. They find that methods and studies do not only need to capture the complexity of the workforce mathematically, but also need to properly design policies for managing the workforce, as well as aligning all stakeholders involved for a shared view of future challenges, appropriate implementation of methods and in agreeing political decisions.

Engaging in SWP is a time-consuming, iterative and organization-changing process, and does not only involve the HR departments within an organization. Business leader input is as important as HR input, and thus workforce planning is an organizational initiative and not something that is solely done by HR (Phillips and Gully, 2012, [51]). Within the HR department, we actually see all four of Ulrich's HR roles represented when engaging with SWP: the *Strategic Partner* aligning business and HR strategy; the *Change Agent* managing transformations initiated by SWP; the *Administrative Expert* facilitating the necessary processes and data; and the *Employee Champion* aligning the new workforce planning with employees.

Strategic Workforce Planning is considered a part of HR analytics, as it usually entails data, dashboards, KPI's[3] and analyses, but mostly for descriptive purposes. For example, Isson and Harriott ([34]) define Strategic Workforce Planning Analytics as *"...the process of injecting advanced analytics into workforce planning in order to optimize outcomes and ensure human capital planning success"* (p. 101). Also, they say that workforce planning analytics helps organizations to create economic value from their human capital data and how it creates high business impacts. Where their book provides plenty of advice and best practices on big data in SWP; the technical aspects of SWP analytics are minimal. Except for listing the types of data and workforce characteristics organizations would need, no actual statistical- or forecasting model or -technology is put forward.

2.3.1 Analytical Strategic Workforce Planning approach

The Analytical Strategic Workforce Planning (ASWP) approach used in this research is an application of the WPP approach by Phillips and Gully ([51]). It is an application, because all the steps can be mapped onto the process created by Phillips and Gully. Simultaneously, the approach is much more tangible and concrete, with descriptive, predictive and prescriptive analytics incorporated in the approach. This ASWP approach is described in Figure 2.4. The approach starts from a high-level, organizational perspective, which allows for quantitative analyses through steps 1 to 5. The organizational perspective is where we consider all jobs within the company, but also the organization's overall goals. Then, those steps lead to actionable insight which can be used on a low-level, individual perspective. This individual perspective is more qualitative and is about implementing career planning, upskilling and transitioning for each employee individually, which is not something considered in this research.

[3]KPI: Key Performance Indicators. A quantifiable measure of performance over time for a specific objective

Figure 2.4: The Analytical Strategic Workforce Planning approach

The first step is important, because the organization needs to have a clear view of its strategic intent, as this is the foundation of Strategic Workforce Planning. This first step can be mapped to both steps 1 and 2 of Phillip and Gully's WPP approach in Figure 2.3. In the second step, the current and historical workforce is analysed in order to get more insight in the development of the workforce; this is related to step 3b in the WPP. The organization's strategy is translated into future workforce requirements in the third step (see step 3a in Figure 2.3). For example, when the strategy is to grow a certain department, this should be reflected in these requirements. We call this the *desired workforce* and its composition is quantified by the number of employees in jobs, costs of employees and the ratio's between jobs. Now that step 2 and 3 discussed the current and desired future workforces, we get to making a planning in the fourth step (see step 3c and 4 in Figure 2.3). Firstly, the insights into the workforce development are used to forecast the expected future workforce. Secondly, with both the expected- and desired future workforce known, we are able to optimize *interventions*. We use interventions to guide the expected future workforce to the desired workforce. Examples of such intervention are HR policies, recruitment plans, retention plans and reskilling schemes. Finally, the fifth step (see step 5 in Figure 2.3) shows how the findings, interventions and their impact on the company have to be *monitored and iterated* thoroughly with all stakeholders on their feasibility and implementation.

This chapter aims at providing the reader with an introduction to the field of HRM and its subfields HRA and SWP. That is, their history, most influential literature, definitions, ongoing debates, applications and how they bring value to organizations. In the next chapter, we go into more detail about the analytical and quantitative aspects of the Analytical Strategic Workforce Planning approach.

3 — Approaches for Strategic Workforce Planning

In this chapter, we describe the analytical steps of Strategic Workforce Planning in detail, as well as their place in the field of Operations Research. We also argue where they fall short and why there is a need for a new approach. To that end, we propose an extension of the SWP approach. Lastly, we refine the scope of this research accordingly.

3.1 Analytical Strategic Workforce Planning approach

As mentioned in the previous chapter, much of the literature and methods for SWP are devoted to the qualitative and managerial aspects of SWP (Isson and Harriott, 2016 [34]; Tursunbayeva, Di Lauro and Pagliari, 2018 [61]). Usually, the steps do involve analytics, however these are usually descriptive in nature. The ASWP approach depicted in Figure 2.4 provides room for not only descriptive, but also predictive and prescriptive analytics in three of the steps. This is different from other approaches, because it uses *optimization methods* to determine the appropriate HR interventions, at its core. This section elaborates on step 2, 3 and 4, which are illustrated more detailed in Figure 3.1.

STEP 2 The current and historical workforce is analysed and descriptive analytics provide insights regarding the size, structure, diversity, age and financial value of the workforce. Most importantly, historical trends of inflow, throughflow and outflow are gathered. These trends will form the basis for predicting the expected future workforce.

STEP 3 The organization's *desired workforce* is designed; this is a translation of the client's strategy into future workforce requirements. The organization's strategy takes into account technological trends and innovations, as well as for example, changing client demand and sustainability goals. These factors need to be translated to future workforce requirements; for example growing or shrinking departments and committing to entirely new business themes, requires a different workforce.

STEP 4 Predicting the *expected future workforce* of the current situation is done using the historical trends obtained in step 2. Then, *Optimization* is used to find the best combination of HR interventions to be put in place to reach the organizations desired workforce as close as possible. In this context, the optimization is usually a *minimization problem*, as the goal is to close the gap between the expected future- and desired workforce. These HR interventions form a concrete action plan that the organization should follow to reach its desired workforce. For example, the plans could say that employees in job 1 should be offered early retirement plans; or that x and y new employees should be hired for job 4 in 2 and 3 years for now, respectively; or that z employees in job 5 should be promoted to job 6 in year 3. In Section 6, we describe the optimization problem and algorithm in more detail.

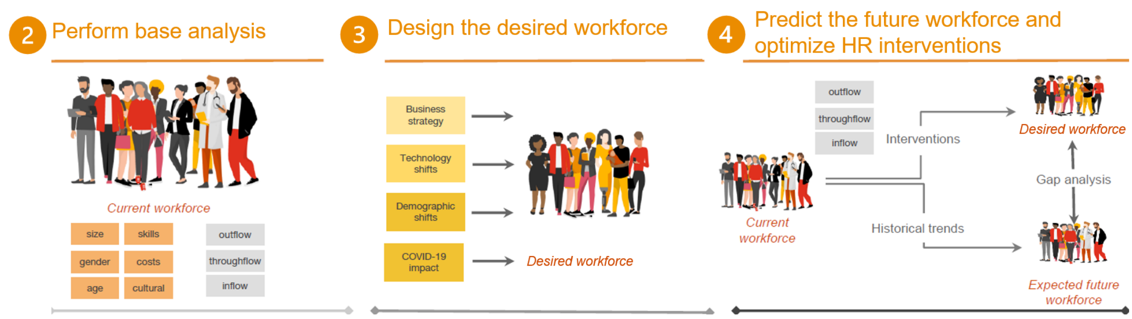

Figure 3.1: The ASWP approach: step 2, 3 and 4 depicted in more detail

3.2 SWP in the field of Operations Research

This section explores the position of Strategic Workforce Planning in the field of Operations Research (OR). Kunc's (2008, [42]) paper on the dynamics of HRM, found that there are three types of workforce planning methods: judgemental, mathematical and a mix of both. The mathematical methods fall into the field of Operations Research.

De Bruecker et al. (2015, [17]) reviewed literature regarding workforce planning problems incorporating skills. Here a mismatch within the field was also found. The mathematical methods were found to be simplifications of complex workforce situations, without considering actual implementation; whereas the judgemental methods provided descriptive explanations of complex workforce modes, but without the support of mathematical Operations Research. Finally, most of the OR mathematical models were not applied on a strategic level, but on a more operational or process level, such as staff scheduling.

As discussed in Section 2.3, Wilis, Cave and Kunc (2018, [69] also found that that methods and studies do not only need to capture the complexity of the workforce mathematically, but also need to properly design policies for managing the workforce, as well as aligning all stakeholders involved for a shared view of future challenges, appropriate implementation of methods and in agreeing political decisions. In 2018, they describe the development and implementation of a framework for SWP for the English healthcare system. The presented multi-methodology approach is based on System Dynamics. System Dynamics is argued to be the best OR methodology, as their application on the healthcare system involved systematic delays[1] and combinatorial- and dynamic complexity[2]. Therefore, their framework was specific for the healthcare system in the UK, making it not fit for general application.

Optimization of HR interventions

In 2013, Bech ([6]) investigated algorithms to find an optimal recruitment strategy in a more generic setting. An overview of the papers analysed can be found in Table A.3 in Appendix A.2. Their research was mostly focused on workforce structures shaped like a pyramid: a hierarchical organizational structure where the higher the job level, the lower the number of employees in that level is; this structure is applicable in, for example, consultancy firms, law firms, banks and hospitals. They concluded that algorithms proposed in the literature were not adequate in practice, as those were only fit for theoretical purposes and not in practice with larger company sizes. They describe the problem in mathematical terms, resulting in a Linear Programming (LP) problem.

Three algorithms were researched to solve this LP problem:

[1]For example development delays in the workforce caused by education and training times of doctors and nurses.
[2]For example, many different care paths and branches existing in medical training, different models in care delivery, diverse training schemes, movement between training paths and the feedback processes between availability of treatments and patient behaviour.

- Algorithm based on a Markov Decision Process (MDP);
- Deductive algorithm based on simulation of the evolution of the workforce (DA);
- Algorithm based on Ordinal Optimization (OO).

They conclude that the size of an organization is an important factor to determine which algorithm should be applied. The empirical results from the case studies showed that the MDP algorithm is theoretically the most pure algorithm, and gives the most reliable optimal recruitment strategy over the entire time horizon. Also, they find that the Ordinal Optimization algorithm, which aims at finding a "good" recruitment strategy, does indeed find a sufficient one. However, these two algorithms have a major drawback on their scalability; as the MDP algorithm already struggled with the complexity of a company consisting of 5 employees and the OO algorithm works for medium sized companies. Finally, they find that the deductive algorithm is the best algorithm for finding an appropriate recruitment strategy in practice. It is seen that the algorithm, gives an optimal recruitment strategy per time period and a suboptimal recruitment strategy over the entire time horizon, even for large companies.

3.3 Proposed extension

Now that we have a better view on the current approach to ASWP, we also see where we could possible extend it. As described in Section 1.1, the *Future of Work* is changing. Due to external- and internal changes outside and within the company, it is no longer evident that jobs continue to exist as they are now, the workforce's current skills remain relevant and companies need to start accounting for these transformations.

3.3.1 Exploiting employees' competencies

In this research we aim at extending the ASWP approach. In Section 1.1, we found that, in the lights of technological developments, job titles may no longer be adequate descriptors of a company's workforce. Rather than using job titles to describe employees, we propose to use the **employees' competencies** instead[3]. By describing employees based on their set of knowledge, skills and abilities, instead of their job title, we provide a more robust description of the workforce. Having this information could change the way a business views its workforce, and hence allows for better application of HR interventions.

We want to do this by grouping employees based on their competencies, into *employee profiles*. Then, employees with similar competencies are grouped into the same profile. For example, one can imagine how the competency requirements for an Aerospace Engineer are quite similar to the competency requirements for an Operational Research Analyst, since both have a predominant mathematical focus. Also a relatedness is seen between for example a Sales Manager and a Marketing Manager; even though they are likely to work in different departments within an organization, both have to show leadership, supervise others, run day-to-day operations, oversee strategy, set goals and track performances of their team. By grouping employees based on their competencies, we want to find these employee profiles within organizations in a quantitative manner, using *competency data*.
Once we have found that some employees are part of the same employee profile, we are able to change the way companies transform their expected future workforce to their desired workforce. For example, it may become clear that some employees are part of a similar employee profile, however, one of the jobs is in danger of becoming obsolete, whereas the other is expected be remain relevant. This allows for more sophisticated up- and reskilling schemes within an organization and increases the employees' internal mobility. As such, one makes use of the available resources more efficiently.

[3]In Chapter 4, we will provide the formal definition of *competency*.

3.4 Refinement of scope

Having more information on HR practices, Strategic Workforce Planning approaches and algorithms, we refine the scope of this research.

This research focuses on altering the fourth step of the ASWP approach, described in Section 3.1 and illustrated in Figure 3.1. In this step, the expected future workforce is forecasted and interventions to reach the desired workforce are optimized. However, we do not research how a company should translate organizational strategy into future workforce requirements in order to attain this desired workforce. That is a major task in itself, considering that, among others, labor market-, macroeconomic and technological- developments should be taken into account.

In the next chapter, it becomes clear that we are able to collect data on employees' competencies based on their job titles. Ideally, we would want a company to have specific data on the competencies of each employee working there, as this would allow for the creation of individual career paths and up- and reskilling schemes. However, from experience we know that this is not something that most companies have available. So instead, we use job related competency data as a proxy for individual competency data.

4 — Competency Management & Clustering

Now that we have developed a more clear view on HR, Strategic Workforce Planning approaches and the scope of this research; we take a closer look at these *competencies.* We first review what "competency" is from a HR perspective, and why they adequately describe employees. Next, we look at what exactly *competency data* is, the background of the database used in this research and particular features of the data. Finally, we look at methods for grouping data that exhibits these features.

4.1 Introduction to Competency Management

The term "competency" started to emerge in the late 1960s and early 1970s. A lot of research has been done on the definition of "competency"[1]. In 1982, Zemke ([71]) concluded that *"...the word "competencies" is a term that has no meaning apart from the particular definition with whom one is speaking".* The term occurs in many contexts, such as legal, clinical, psychology, vocational, educational and industrial psychology. In all these contexts, the term defines "successful" performance of certain tasks or activity or adequate knowledge of a certain domain of knowledge or skill (Schippmann et al., 2000, [58]).

Schippmann et al. (2000, [58]) found that the term "competency" was adopted in the vocational[2] counseling profession to define broad areas of knowledge, skills and abilities linked to specific occupations. It caused a shift in thinking within HR, which started the growth in competency-based frameworks. Instead of formal qualifications and past experiences, more emphasis was placed on the behavior of the employee (Gigliotti, 2019 [29]). The roots of this shift lie in the work of psychologist David McClelland (1973, [46]). McClelland claimed that employers should test for personality or competencies of life outcomes (e.g. communication skills, patience, moderate goal-setting, and taking initiative) and that the outcomes of such test would be a stronger indicator of one's abilities than the results of traditional test of intelligence. For this research, we define an individual's competencies as follows:

> The combination of knowledge, skills and abilities, which are required in order for an individual to successfully perform a specific occupation; any individual characteristics that can be observed, measured, learned, acquired and enhanced.

Later on in this chapter, we focus on the corresponding data, which we acquire through the *O*NET database.* For that reason, we explain the more explicit definitions and a selection of examples of these "knowledge", "skills" and "abilities" based on their definition (Fleisher et al., 2018 [26]). "Knowledge" are *organized sets of principles and facts applying in general domains.* For example, Medicine & Dentistry and Therapy & Counseling are in the "Health Services" domain and Computer & Electronics and Building & Construction are in the "Engineering & Technology" domain. A "skill" is a *proficiency that is developed through training or expertise.* Basic skills facilitate the acquisition of new knowledge, for

[1] In Table B.1 in Appendix B.1, a summary of definitions of "çompetencies" from noted scholars, federal agencies and subject matter experts is presented, based on Schippmann et al., [58] and Gigliotti, 2019 [29]

[2] Vocational: relating to an occupation or employment

example: Reading, Writing, Mathematics, Science and Learning Strategies. On the other hand, cross-functional skills extend across several domains of activities, examples are Negotiation, Programming, Repairing, Time Management and Complex Problem Solving. An "ability" is defined as an *enduring attribute of the individual that influence performance.* Some examples of different types of abilities are: Oral Expression, Written Comprehension, Explosive Strength, Originality, Inductive Reasoning, Mathematical Reasoning, Memorization and Hearing Sensitivity.

4.1.1 Competency modelling

With the rising of the term "competency", came the practice of *competency modelling*, which has exploded in the 1990s. Through competency modelling, job positions today are written based on the needs of the organization and are directly aligned with a core set of competencies required for the position (Schippmann et al., 2000 [58]). Competency modelling is when organizations create a competency-based framework that "involves identifying the varied knowledge, values, abilities and behaviors that people need to possess and exercise to achieve the strategic objectives, goals and performance expectations of the organization" (Croft & Seemiller, 2017 [14]).

In 2011, Campion et al. ([11]) compared competency models to the more traditional "job analysis" framework and concluded that competency modelling has an impact far outstanding that of traditional job analysis. They presented a set of best practices for competency modelling, the many roles competency models play in HR systems, how to use this competency information and examples of the impact in organization. Most importantly, Campion et al. find that competencies are used to develop and align Human Resources Systems across an organization; some of the uses they present are:

- *Hiring* new employees based on procedures that measure competency; the models distinguish the top performing employees from average employees.
- *Providing* courses specific to the development of certain competencies to train employees.
- *Evaluating* employees' performance by creating assessment tools around competency and their levels of mastery.
- *Defining* and describing what effective performance and successfulness is for employees.
- *Promoting* employees by making use of competency based promotion criteria.
- *Developing* and aligning employee careers by guiding the allocation of job assignments using competency models.
- *Managing* employee information by gathering employee skill, training, and job experience information.
- *Compensating* employees by connecting business objectives and performance levels to assess pay differences or to evaluate employees for pay increases.
- *Supervising* the acquisition and retention of critical competencies related to current and future strategic intent.

Another important use of competency information, directly linking to the goal of this research, is that competencies support organizational change efforts. It helps organizations through a transition by giving HR the ability to train, assess, select, promote, and reward employees in alignment to a desired future state (Campion et al., 2011, [11]). This links with the goal of this research, by having a more future, or in other words, *strategic* view.

4.2 Competencies in Strategic Workforce Planning

Now that we have a better view of what *competencies* exactly are, we need to validate whether they fit the purpose of this research. Also, we review how competencies are currently used for the purpose of Strategic Workforce Planning.

In 2020, Gonsalvez et al. ([30]) clustered competencies of psychology practitioners and revealed that the competencies had a very hierarchical structure. An important result from their research is that *competency typology* appeared to play a significant role in determining empirical clustering outcomes. In other words, the differentiation between knowledge, skills, relationship and attitudes provide a better insight into the anatomy of competencies of psychology practitioners, than differentiation along functional or thematic lines. Also, they found at macro-level that there appear to be two main-branches in competencies: knowledge and technical skills versus relationship, attitude and value competencies.
Their findings are in line with our discussion on competencies above: employees are better described and clustered based on their competencies, than by the tasks or activities associated with their specific occupation. So, in order for an individual to successfully perform a specific occupation, it is not that they have to be good at certain tasks, instead certain competencies (a combination of knowledge, skills and abilities) are required.

So indeed, it is valid to describe employees using competencies; and this is not new in Strategic Workforce Planning. Competencies in fact are used within companies and SWP, however they are used in a more qualitative manner. Recall the literature on competency modelling, in the previous section, where competencies are used for hiring, training and managing (new) employees, which also links to SWP. Also recall that big parts of SWP are formulating the company's strategic intent and analysing the current workforce, which is often done through competencies (Phillips and Gully, 2012, [51]). In this research, on the contrary, we want to ensure that competencies are an integral part of the quantitative algorithms used for Strategic Workforce Planning. Also, we want to do this in a quantitative manner and for that, we need data; which is discussed in the next section.

4.3 Competency data: the O*NET database

As discussed before in Section 3.4, most companies do not have competency information on their employees. For that reason, we will make use of the O*NET (Occupational Information Network) database, which is set up by the U.S. Department of Labor, Employment and Training Administration and is continuously maintained and updated through their data collection program. The database is constructed based on *occupations*; an occupation is the same as a job or profession. For this research, we use the data from version 26.1 (November 2021).

4.3.1 Occupational taxonomy

The O*NET-SOC (Standard Occupational Classification) occupational taxonomy ([63]) includes 1016 occupational titles. All SOC occupations are assigned a six-digit code. These digits represent the different levels of aggregation within the structure. For example:

- 17-0000 Architecture and Engineering Occupations (SOC major group)
  - 17-2000 Engineers (SOC minor group)
    - 17-2110 Industrial Engineers, Including Health and Safety (SOC broad occupation)
      - 17-2112 Industrial Engineers (SOC detailed occupation)

This six-digit SOC code, is referred to as an SOC-level occupation and is also assiged with a ".00" extension. In cases where an occupation is more detailed than the SOC-level occupation, it is assigned this SOC-level code, along with a two-digit extension, depending on the number of detailed SOC occupations linked the the SOC-level occupation. For example:

17-2112.00 Industrial Engineers (SOC-level)
17-2112.01 Human Factors Engineers and Ergonomists (detailed O*NET-SOC occupation)
17-2112.02 Validation Engineers (detailed O*NET-SOC occupation)
17-2112.03 Manufacturing Engineers (detailed O*NET-SOC occupation)

The O*NET database collects data on the detailed O*NET-SOC occupation level and the SOC-level. In total, 923 out of 1016 occupations are on the O*NET data-level and up until now[3], the data on 873 (detailed) occupations is collected.

4.3.2 Conceptual foundation

The conceptual foundation of O*NET is the *Content Model*. The Content Model (see Figure 4.1) provides a framework that identifies the most important types of information about work and integrates them into a theoretically and empirically sound system (O*NET, [62]). It makes a distinction between *worker-oriented* and *job-oriented* descriptors. The worker-oriented descriptors reflects the character of people, whereas the job-oriented descriptors reflects the character of specific occupations.

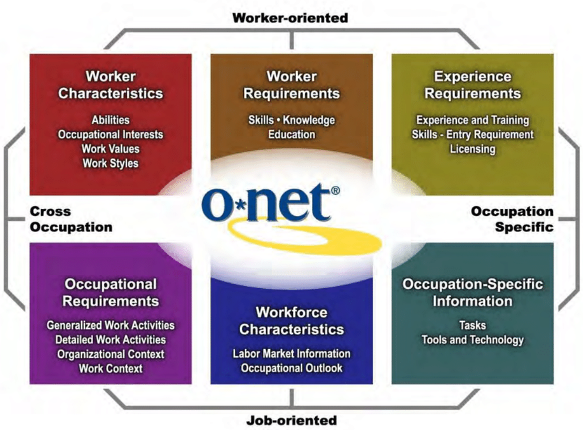

Figure 4.1: O*NET Content Model

We found that employees are better described by their competencies, than by the tasks or activities associated with their specific occupation. This implies using the *worker-oriented* descriptors, which are again divided into three major domains: “Worker Characteristics”, “Worker Requirements” and “Experience Requirements”. In Section 4.1, we defined an individual’s competency as something that can be observed, measures, learned, acquired and enhanced. This definition helps us in selecting the right descriptors for this research.

We are not as much interested in “Experience Requirements”, as we are more looking at an individual, rather than to previous work experience or licensing. For the same reason, the *Educational* of the “Working Requirements” domain is not relevant. Since recently, organizational literature supports the inclusion of interests, values, and work styles as descriptors of “Worker Characteristics”. This is reasonable, considering people follow their interest and prefer to work at a company that matches their values and work style. These analyses however, do not represent why a person would, or would not, be successful in performing a specific job. Also, interests, values and work styles are not characteristics that an individual is able to learn or enhance. Considering all available descriptors, the attributes of *Knowledge, Skills* and *Abilities* are the ones that best represent the competencies needed for individuals to be successful at performing a specific job.

[3] O*NET 26.1 Database (November 2021)

4.3.3 Competency data

This section elaborates on the actual information from the O*NET database that is used for this research: the *Knowledge, Skills* and *Abilities* databases[4]. The Content Model distinguishes 33 types of knowledge, 35 types of skills and 52 types of abilities, which together add up to 120 different types of competencies. These three databases itself are again structured into multiple layers of information.

The 33 types of knowledge, are divided into 10 categories or domains, such as Engineering & Technology, Communications and Health Services. The 35 types of skills are divided into two main categories: basic- and cross-functional skills. Basic skills are skills that facilitate learning or the acquisition of knowledge, such as *Reading*, *Writing* and *Critical Thinking*. Cross-functional skills are skills that facilitate performance across jobs, such as *Negotiation*, *Programming* and *Management of Financial Resources*. The 52 types of abilities are structured into four main categories: cognitive-, psychomotor-, physical- and sensory abilities, which are again divided into sub-categories. The complete overview of all types of knowledge, skills and abilities and their structure is presented in Table B.2, B.3 and B.4 in Appendix B.2, respectively.

For each type of competencies, the database provides two scores for every (detailed) SOC occupation: the importance- and level score. The *importance score* measures how important that type of competency is with respect to the specific occupation on a scale of 1 to 5; with the values: "not important" (1), "somewhat important" (2), "important" (3), "very important" (4) and "extremely important"(5). The *level score* measures the level of mastery of the people in that specific occupation regarding that type of competency, on a scale from 0 to 7; with 0 being the lowest and 7 the highest possible attainable level. In order to keep the results objective, O*NET provides *level scale anchors* for each type of competency that provide an additional source of clarity. These anchors indicate the degree, or point along a continuum, to which a particular descriptor is required or needed to perform a specific job. Each level scale includes examples near the lower end, midpoint, and higher end of the scale to provide additional context. Take for example the task to determine the level of knowledge about *Administration and Management* that is needed to perform a job. Then on the scale of 0 to 7, three anchors are provided: "Approve a reimbursement request" at the value 2; "Monitor progress of a project to ensure timely completion" at level 4; and "Manage a multi-million dollar company" at value 6.

A detailed overview of the structure of the databases is displayed in Table B.5 in Appendix B.2. Amongst others, the files contain statistical information on the sample size, standard error, lower- and upper confidence bounds. For the *knowledge* attribute, for some of the occupations, the statistical information and precision indicator are not provided. Also, a "low precision" indicator and "relevancy" indicator are provided:

- The low precision indicator ("Recommend Suppress") has the value "Yes" if the data values is considered to have low precision[5]. O*NET encourages users to use estimates exhibiting "low precision" with caution and for many applications users are advised to consider suppressing these estimates.

- The relevancy indicator ("Not Relevant") has the value "Yes" if the level score for that type of competency is identified as "not relevant". This is the case when 75% of the respondents rated importance for that type of competency as "not important" (score 1 on the scale of 1 to 5). O*NET encourages users to provide their end-users with an indication that the item level rating is "not relevant" rather than displaying the level value or displaying no level information.

[4]These files are available for downloading on the O*NET website: `https://www.onetcenter.org/database.html#individual-files`

[5]A value has "low precision" when the standard error is greater than 0, 51, the sample size is less then 10, the variance is 0 and the sample size is less than 15 or when the relative standard error (RSE) is greater then 0, 5.

4.3.4 Data features

There are two features of the data that have to be taken into account when looking at grouping employees based on competency data.

Firstly, the competency data is three dimensional. Usually data is two dimensional; a common example is where n individuals are described by p characteristics or variables, like their age, gender and education level. In our case, the data is three dimensional, because for n jobs or occupations, we have the information on m types of competencies, for which we each have 2 scores: the importance- and the level score. The first score measures how important that type of competency is with respect to the specific occupation; the second score measures the level of mastery that the people with that specific occupation have regarding that type of competency. To illustrate: for the occupation of "Flight Attendant", the skill of "Speaking" has an importance score of 4.12 and a level score of 3.75; the knowledge of "Building & Constructing" has an importance score of 1.41 and a level score 0.75 [6].

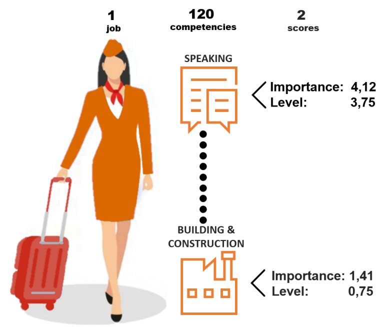

Figure 4.2: Illustration of three-dimensional competency data

Secondly, we see that an organization's competency data experiences High Dimensionality and Low Sample Size (HDLSS). The data has *high dimensionality*, since we have 120 type of competencies, for which we each have 2 scores, resulting in many variables. This will be generally true for competency data, as it is not possible to adequately capture an individual's competencies using just a few attributes. There is a *low sample size* because for example 2000 employees is not equivalent to 2000 data points; instead these 2000 employees comprise 50 different occupations, leaving us with merely 50 data points. This is a feature that is generally true for most companies. This is because for overall data analysis a "normal" amount of samples would be at least 1000; which is an unrealistic number of jobs for a company.

Other researches using the O*NET database, often neglect the level score of the competency data in order not to deal with the multi-dimensionality feature. However, when describing employees based on their competencies, we believe the combination of the importance- and level score give the most valuable information. For that reason, we take both of these data features into consideration when grouping employees.

4.4 Methods for designing employee profiles

This section explores methods for grouping employees. As we have seen in Section 4.3.1, there are existing occupational taxonomies, classifying and ranking jobs in many different categories. However, as we have seen in Section 3.3 and 4.2, we should not rely on these classifications of job titles in the long term and describe employees' by their competencies.

In Section 4.1 we established that competency modelling is used to developed and align HR systems and policies within companies. So HR practitioners might already have information on their employees, as well as some guidelines or commitments to specific core competencies the company wishes to strengthen. For that reason, HR practitioners already have a lot of information that could be used to group employees. However, we want to take all types of competencies into account, and design employee profiles based on similarities between jobs that may not be obvious from an HR perspective. For that reason we turn to competency data as described in Section 4.3.

[6]Importance scores are between 1 and 5; level scores are between 0 and 7.

As we have quantitative data and try to extract information out of this, without prior knowledge, we resort to *unsupervised learning techniques*. The method deemed most fit is *Clustering*, or *Cluster Analysis*, where the task is to group objects. Clustering is *unsupervised*, because we do not have classifications of the data beforehand from which we can learn, and also, we do not know the number of clusters (Shalev-Schwartz and Ben-David, 2014 [59]).

This section is structured as follows: first we go into more detail about clustering algorithms for multi-dimensional and high dimensional, low sample size competency data in Section 4.4.1. Then, we go into detail on how we can assess whether these clusters are meaningful, from a statistical, as well as an HR perspective in Section 4.4.2.

4.4.1 Clustering methods

The main objective of clustering is to group data in such a way that there is a large similarity within a cluster and large dissimilarity between different clusters. There are different types of clustering, and in this research, we consider *strict partitioning clustering*, meaning that each objects belongs to exactly one cluster. In clustering literature, clustering multi-dimensional data is referred to as "multi-criteria clustering". The remainder of this section elaborates on several types of clustering methods, including relevant extensions that deal with the features of the data.

Centroid-based clustering: k-means and relevant extensions

One of the most well-known clustering methods is the *k-means algorithm*, introduced by Lloyd in 1957 (published in 1982, [43]). The k-means clustering problem aims to partition n observations into k clusters, such that the within-cluster variances are minimized; each observation belongs to the cluster with the nearest cluster center. Because this problem is NP-hard (Aloise et al., 2009 [4]), heuristics are used. Lloyd's heuristic or algorithm is often referred to as the "naive k-means algorithm". However, k-means is not designed for multi-criteria clustering.

In 2018, Wasid and Ali ([68]) incorporate multi-criteria ratings in a k-means algorithm by altering the distance formula. They are considering a recommender system for movies, in the case of multi-criteria ratings. Recommender systems for movies usually have user-movie ratings, where each user (n) rates each movie (p) with one score. They improved the recommender system by using multi-criteria ratings that represent the preferences of uses on several aspects of movies; so each user (n) rates one movie (p) on several items (s), such as story, visual, acting and directing. This is in line with the dimensions of our data, where each job (n) is described by m types of competencies using s scores. They incorporate this multi-criteria aspect in a k-means algorithm by altering the distance formula; their algorithm is displayed in Figure B.1 in Appendix B.3. We also translate Wasid and Ali's Euclidean distance formula to match our competency framework. When taking a closer look at their distance formula [7], we find that it does not actually take into account the multi-criteria aspect of the data. Instead, they simply treat the multi-criteria scores as independent variables.

In 2003, De Smet and Guzmán ([19]) proposed another extension of the k-means algorithm to the multi-criteria framework. They propose a completely different definition of a multi-criteria distance, which is based on a preference structure defined by the decision maker. With this multi-criteria distance, they are able to partition "alternatives" into classes that are meaningful from a multi-criteria perspective. Alternatives could be objects, actions or characteristics of customers. They do this by introducing the notion of "profile" to each alternative, using *preference modelling*. Preference modelling usually considers the following relations: Preference (P), Indifference (I), and Incomparability (J). These are matrices that indicate that for example alternative a_i is preferred over alternative a_j, or that the decision maker is indifferent between alternatives a_i and a_j. Then, each alternative has a certain profile, which is a 4-tuple of sets, indicating which other alternatives are considered better, worse, incomparable

[7] Formula B.2 in Appendix B.3

and indifferent. Then, they define a distance metric between alternatives that takes into account the multi-criteria nature: basically, two alternatives are as close as their profiles are alike.

The k-means algorithm also has drawbacks, which are especially prominent considering our HDLSS data. Firstly, the k in the k-means algorithm has to be specified in advance; if chosen poorly, this could lead to severe degradation of the quality of the clustering results (Raykov et al., [56]). Secondly, k-means is highly dependent on the initialization of the algorithm. As the algorithm uses a heuristic, we only find a local optimum which changes with every random initialization. That's why in most methods, the algorithm is performed multiple times and the best solution is used. Also, our HDLSS data causes problems. High dimensionality imposes a fundamental problem: the contrast between points that are near and far away no longer exist. The distance measure becomes ill-defined and even the concept of proximity may not be meaningful from a qualitative perspective (Aggarwal et al., [2]). In order for a clustering to be meaningful, *stability* is viewed as a necessary condition. Regarding the required sample size for segmentation analysis, a data analyst is left guessing, as Dolnicar et al. ([20]) found in 2016. They concluded that insufficient sample size has serious negative consequences for the quality of the clustering.

All in all, a k-means approach might lead to unstable or unreliable results.

Hierarchical clustering: agglomerative clustering and relevant extensions

Another type of clustering is *hierarchical clustering*, which builds a hierarchy of clusters. *Agglomerative clustering* is where clusters are created from a *bottom-up approach*: All observations start in their own cluster and the pairwise dissimilarity is measured between the clusters. The two clusters that are most similar to each other, are fused into on cluster. This iterates, until all observations are in one cluster. In order to measure the distance between two clusters, the notion of *linkage* is developed.

There are some benefits to hierarchical clustering method, compared to the k-means algorithm. Firstly, there is no need to specify the number of clusters k beforehand, making the method more robust. Secondly, the greedy method generates one local solution, and is not dependent on a random initialization and hence, the low sample size is not such a big problem anymore. Also, specifically in our case, competencies have a hierarchical structure, as we found in Section 4.3, which might be beneficial for the results of the clustering. However, the problem of an ill-defined distance measure due to the high-dimensionality data might still a problem.

Ferligoj and Batagelj (1992, [24]) presented two types of modified agglomerative algorithms to deal with the multicriteria issue. They assumes that there are k "criterions" instead of 1 and try to find the best hierarchical solution which satisfies all k dissimilarity matrices as much as possible. In their first method, they derive the dissimilarity matrix D as a function of all possible k dissimilarity matrices: $D = f(D^t, t = 1, \cdots, k)$. Where the function f acts almost like a linkage criterion, determining the distance between two clusters, based on k dissimilarity matrices. Their second approach is to select which clusters to merge by searching for the *Pareto nearest* pair of clusters. A pair of clusters is Pareto nearest if there is no other pair of clusters, where for at least one dissimilarity matrix the distance between the clusters is smaller, while the other distances are remain the same. Now, it is possible that at each step more than one Pareto nearest pair of clusters exists. Hence the procedure might give several (Pareto) hierarchical solutions, which can be avoided by including additional decision rules. With this second approach, it is not as easy to make a visual representation of the hierarchical solution, as there is no unique level at each step of the merges. In their paper, their method using the Pareto efficiency criterion works very well and fast. They say it is possible to analyze 100 data points for 10 criteria in reasonable time. However, as their procedure yield several hierarchical solutions, a procedure for an efficient review of the obtained solutions should be found.

4.4.2 Cluster evaluation

Besides exploring methods for designing clusters, we also need methods for evaluating these clusters, as this is not evident. For supervised learning techniques, there are labels telling you whether your method predicted the output right or not, resulting in some kind of *prediction accuracy.* In contrast, with unsupervised clustering techniques, we try to organize the data in some meaningful way and to learn about its structure. As a result, there is no clear success evaluation procedure for clustering (Shalev-Schwartz and Ben-David, 2014 [59]).

In fact, we still do not really know what a clustering *is*; there are fundamental challenges associated with clustering, highlighted by Jain and Dubes (1998, [36]), which cause an inherent vagueness in the definition of a cluster, even in the-present-day (Jain, 2010 [35]). Generally, there are two approaches to tackling this issue:

- The first approach is to attempt to define a *clustering function.* In 2003, Kleinberg ([40]) advocates the development of a theory of clustering that will be "independent of any articular algorithm, objective function, or generative data model". Aiming to define what a clustering function is, Kleinberg addresses this via an axiomatic approach, and failed. He suggested seemingly natural and plausible axioms, but shows that these lead to a contradiction. From this his *impossibility theorem* arises: there exists no clustering function that satisfies all axioms. Often, these results are interpreted as the impossibility of defining what clustering is, or even of developing a general theory of clustering. ([1], [40], [59]).
- The second approach is to acknowledge that one such function does not exist. This type of "absolute" approaches to defining clustering has been discredited by multiple researches. Users of these methods often reject the notion that clustering is a domain-independent subject (von Luxburg, Williamson and Guyon, 2012 [67]). Von Luxburg et al. argue that *"clustering should not be treated as an application independent mathematical problem, but should always be studied in the context of its end-use"* (p. 65). They discuss the shortcomings in the current clustering evaluation methods and conclude that they *"...are very problematic and do not serve their purpose"* (p. 68) and that *"these methods cannot be used to evaluate the usefulness of the clustering: usefulness cannot be evaluated without a particular purpose in mind"* (p.68). Jain ([35]) argues that the representation of the data must go hand in hand with the the purpose of grouping. Furthermore, they claim it is up to the user to carefully choose his representation to obtain a desired clustering.

In the following two sections, both approaches are applied to find ways to evaluate employee profiles.

Cluster-quality measures

Just like Kleinberg, Ackerman and Ben-David (2008, [1]) aim towards the development of a general clustering theory. They disagree with the interpretation that the impossibility theorem shows there is no possibility of developing a general theory of clustering. Instead, they show that the impossibility was due to the specific formalism used by Kleinberg and not because of an inherent feature of clustering. Additionally, they claim that many techniques for evaluating *cluster validity*, do not satisfy the need for a general theory of clustering. Instead of attempting to define a clustering function, they tackle the closely related issue of evaluating the *quality of a given data clustering.* They do so by developing a formalism and a consistent axiomatization of *clustering-quality measures.*

Cluster-quality measures are defined as follows:

> A *clustering-quality measure* (CQM) is a function that maps pairs of the form *(data set, clustering)* to some ordered (for example a non-negative real numbers), so that these values reflect how "good" or conclusive the clustering is.

CQMs measure the quality of a given data clustering by quantifying how good any specific clustering is, as well as being able to help clustering model-selection by comparing different clusterings over the same data set. Particularly, Ackerman and Ben-David (2008, [1]) introduce quality-measures that reflect the underlying intuition of center-based and linkage-based clustering; which are also consistent for a set of axioms[8]. Ackerman and Ben-David introduce two new CQMs that can be distinguished between two types of CQMs: measures that reflect the underlying intuition of center-based- and linkage-based clustering. These new measures are presented below, as we use these as a tool to evaluate the designed employee profiles.

Centroid-based clustering The quality measure reflecting the center-based clustering is the *Relative Margin*. For each point in the data set, we consider the ratio of the distance from the point to its closest center to the distance from the point to its second closest center. Intuitively, we want the point to be close to its own cluster center, and far away from the second closest center. The average of this ratio for all points is the Relative Margin quality measure. The smaller the ratios are, the more confident points are about their cluster membership, thus smaller values of Relative Margin indicate better clustering quality.

Definition 4.4.1 (Relative Point Margin) *The K-Relative Point Margin of* $x \in X$ *is* $K\text{-}RM_{X,d}(x) = \frac{d(x,c_x)}{d(x,c_{x'})}$, *where* $c_x \in K$ *is the closest center to* x, $c_{x'} \in K$ *is a second closest center to* x, *and* $K \subseteq X$.

A set K is a *representative set* of a clustering C if it consists of exactly one point from each cluster of C; this is formalized in Definition 4.4.2.

Definition 4.4.2 (Representative Set) *A set* K *is a representative set of clustering* $C = \{C_1, C_2, \cdots, C_k\}$ *if* $|K| = k$ *and for all* $i, K \cap C_i \neq \emptyset$

Definition 4.4.3 (Relative Margin) *The Relative Margin of a clustering* C *over* (X, d) *is*

$$RM_{X,d}(C) = \min_{K is\ a\ representative\ set\ of\ C} avg_{x \in X \setminus K} K\text{-}RM_{X,d}(x)$$

Using Relative Margin, it takes $O(n^{(k+1)})$ operations to compute the clustering quality of a data set, exponential in k. If a set of centers is given, the Relative Margin can be computed in $O(nk)$ operations. Relative Margin satisfies all axioms of evaluating the quality of a given data clustering, as provided by Ackerman and Ben-David.

Hierarchical clustering To assess the quality of a linkage-based clustering, the *weakest link* quality measure is proposed. Whenever a pair of points share the same cluster they are connected via a tight chain of points in that cluster. The weakest link quality measure focuses on the longest link in such a chain.

Definition 4.4.4 (Weakest Link Between Points) *The Weakest Link Between Points* x *and* y *is:* $C\text{-}WL_{X,d}(x,y) = \min_{x_1,x_2,\cdots,x_l \in C_i}(\max(d(x,x_1), d(x_1,x_2), \cdots, d(x_l,y)))$, *where* C *is a clustering over* (X,d) *and* C_i *is a cluster in* C

The weakest link of C is the maximal value of $C - WL_{X,d}(x,y)$ over all pairs of points belonging to the same cluster, divided by the shortest between-cluster distance.

[8] Axioms for clustering-quality measures:

· Scale invariance: output of a clustering is invariant to uniform scaling of the input.

· Consistency: if within-cluster distances are decreased, and between-cluster distances are increased, then the output of a cluster does not change.

· Richness: by modifying the distance function, any partition of the underlying data set can be obtained.

· Isomorphism Invariance: clustering should be indifferent to the individual identity of clustered elements.

Definition 4.4.5 (Weakest Link of C) *The Weakest Link of a clustering C over (X,d) is:*

$$WL(C) = \frac{\max_{x \sim_C y} C\text{-}WL_{X,d}(x, y)}{\min_{x \not\sim_C y} d(x, y)}$$

The range of values of weakest link is $(0, \infty)$.

Evaluation from an HR perspective

As discussed by von Luxburg et al. ([67]) and Jain ([36]), a clustering can not be evaluated without the context of its end-use. In the case of clustering employees based on their competencies to design employee profiles, we also need clusters to be *interpretable* and *explainable* to be meaningful from an HR perspective.

Firstly, clusters need to be meaningful in the sense that each cluster actually contains relevant information that an HR practitioner is able to interpret and utilize. For example, if jobs are clustered into two equal sized groups, these groups do not really provide much insights that can be used for decision making. Also, when a method is applied and the resulting cluster membership is presented, this only shows the quantitative perspective. However, the HR practitioner is more interested in why certain jobs are in one cluster, and why others not. They want insights into the employees, their competencies and the (dis)similarities between them.

Secondly, in order for any methodology to be meaningful to HR practitioners, it needs to be *explainable.* HR practitioners and companies will not adapt to new methods or base important decision on black box methods for which they do are not able to explain its inner workings. Results might actually impact people's lives, therefore methodologies and their results need to be justified and explainable.

When evaluating methods and results from an HR perspective, these two aspects need to be taken into account.

This chapter was focused on Competency Management and clustering. First, literature on competency and competency modelling was reviewed. Also, we researched how competencies are currently used in SWP and we validated that employees can be describe and differentiate by their competencies. Next, we explored what competency data looks like, using the O*NET database. We found that multi-dimensionality and high dimensionality and low sample size are generic features of the data. Finally, we reviewed existing methods for grouping data that exhibits these features. We concluded that there are no adequate methods. In the next chapter, we propose a new methodology with the goal of designing employee profiles. Then, in Chapter 6, we propose new ways as to how these employee profiles can be integrated into the Analytical Strategic Workforce Planning approach.

5 — Proposal for a Framework for Employee Profile Design

In the previous chapter, we introduced the notions of *competency management*, the corresponding *competency data* and we researched *clustering methods* for HDLSS and multidimensional data. In this chapter, we use all before mentioned information and propose a new methodology for designing employee profiles based on employees' competencies. First, a preliminary section will summarize the obtained information and describe a clustering algorithm that our new method is based on. In the second section, the elements of the original algorithm are translated to the competency framework. Using these translations, the third and fourth section introduce new algorithms.

5.1 Preliminary work

In Section 3.3, the idea to group employees based on their competencies, was proposed. For this, we want to find groups for which we have differences between groups and similarities within one group. This led us to research clustering techniques that could be used to this end in Section 4.4. More specifically, we want a *strict partitioning*, which means that each employee is grouped into one cluster precisely one time. Also, we want to make sure that we keep the multidimensional character of the data, as well as make sure we deal with the high dimensionality and low sample size characteristic of our data.

The new method is an alteration of the multicriteria decision aid (MCDA) clustering algorithm introduced by De Smet and Guzmán (2003, [19]), which was already mentioned in Section 4.4. This method suits our needs, as it was the method that actually integrated the multi-dimensional nature of the problem into the algorithm the most. Also, the method (unintentionally) takes care of the HDLSS feature of our data, while maintaining a very intuitive and tractable character.

In this section, we explain the original multicriteria decision aid (MCDA) clustering algorithm by De Smet and Guzmán (2003, [19]). We do this by going through the different elements of their approach,after that, we translate those elements to our competency framework in Section 5.2.

5.1.1 The original multicriteria decision aid (MCDA) clustering algorithm

De Smet and Guzmán aim to group "alternatives" into homogeneous classes, in order to aid the decision making processes where the decision maker has to decide between these "alternatives". They do this by extending the k-means algorithm to the multi-criteria framework. The clustering model is based on the idea that all the alternatives inside the same cluster are similar in the sense that they are preferred, indifferent and incomparable to more or less the same alternatives. Therefore, they introduce the notion of a "profile" to each alternative, using *preference modelling*. Then, they define a distance metric between alternatives that takes into account the multi-criteria nature: basically, two alternatives are as close as their profiles are alike. With this new multi-criteria distance, they are able to partition "alternatives" into classes that are meaningful from a multi-criteria perspective. Below, the different elements are elaborated upon.

Aim

The aim of their algorithm is to group a set of alternatives $A = \{a_1, a_2, \cdots, a_n\}$, that are evaluated on m criteria $\{g_1, g_2, \cdots, g_m\}$, into categories that remain as homogeneous as possible.

To make the method more tangible, we consider the case where a HR professional is the decision maker, choosing between alternatives. For example, a recruiter has to decide which candidate to hire for an open vacancy. Then, we have a set of candidates (alternatives), whom are each evaluated on different criteria, such as their previous working experience, educational background and personality traits. The recruiter wants to group the candidates based on these criteria, such that he has more insights into the (dis)similarities of the candidates, in order to make a more informed decision.

Preference modelling and structure

Preference modelling usually considers the following relations: Preference (P), Indifference (I), and Incomparability (J). These are displayed as square matrices, indicating the relationship between two alternatives a_i and $a_j \in A$: alternative a_i is *preferred* over alternative a_j if $a_i P a_j$ holds true; or, the decision maker is *indifferent* between alternatives a_i and a_j if $a_i I a_j$ holds true; or, the decision maker thinks that a_i and a_j are *incomparable*, then $a_i J a_j$ holds true.

The three relations $\{P, I, J\}$ make up a *preference structure* on A if they satisfy certain conditions[1] and if, given any two elements a_i, a_j of A, one and only one of the following properties is true: a_i is preferred over a_j ($a_i P a_j$); or, a_j is preferred over a_i ($a_j P a_i$); or, indifferent between a_i and a_j ($a_i I a_j$); or, a_i and a_j are incomparable ($a_i J a_j$).

In the example introduced earlier, the recruiter (decision maker) would decide for each candidate (alternative), whether this candidate is preferred, indifferent or incomparable to each of the other candidates; these comparisons of candidates results in a preference structure.

Profiles

This preference structure $\{P, I, J\}$ is used to design profiles for each alternative. The *profile* $P(a_i)$ of alternative $a_i \in A$ is defined as being a 4-tuple $\langle J(a_i), P^-(a_i), I(a_i), P^+(a_i)\rangle$ where:

- $J(a_i) = \{a_j \in A | a_i J a_j\} = P_1(a_i)$ all alternatives a_j, for which a_j incomparable to a_i
- $P^-(a_i) = \{a_j \in A | a_j P a_i\} = P_2(a_i)$ all alternatives a_j, for which a_j is preferred over a_i
- $I(a_i) = \{a_j \in A | a_i I a_j\} = P_3(a_i)$ all alternatives a_j, for which indifferent between a_i and a_j
- $P^+(a_i) = \{a_j \in A | a_i P a_j\} = P_4(a_i)$ all alternatives a_j, for which a_i is preferred over a_j

Using the created preference structure, the recruiter is now able to represent each candidate by means of a profile.

[1]For every $a_i, a_j \in A$, we have:

· P is asymmetric:	$a_i P a_j \Rightarrow a_j \neg P a_i$	if a_i is preferred over a_j, a_j is not preferred over a_i
· I is reflexive:	$a_i I a_i$	all diagonal element are 1, decision maker is indifferent between a_i and a_i
· I symmetric:	$a_i I a_j \Rightarrow a_j I a_i$	if indifferent between a_i and a_j, then also indifferent between a_j and a_i
· J is irreflexive:	$a_i \neg J a_i$	all diagonal elements are 0: alternatives a_i and a_i are not incomparable
· J is symmetric:	$a_i J a_j \Rightarrow a_j J a_i$	if a_i and a_j are incomparable, then also a_j and a_i incomparable

Distance metric

In order to account for the multicriteria nature of the problem, De Smet and Guzmán create a new distance metric based on the just defined profiles. The distance metric is defined as [2]:

> *Let $P(a_i)$ be the profile of alternative a_i, the distance between two alternatives $a_i, a_j \in A$ is defined as follows:*
>
> $$d(a_i, a_j) = 1 - \frac{\sum_{k=1}^{4} \left| P_k(a_i) \cap P_k(a_j) \right|}{n} \tag{5.1}$$

The intuition behind this concept is that two alternatives are as close as their profiles are alike. When calculating the distance between a_i and a_j, for each element of the profiles, the amount of alternatives in common are counted. When the profiles are more like each other, the fraction becomes bigger and the distance between the alternatives becomes smaller. By using this distance metric, De Smet and Guzmán are able to extend the k-means algorithm to the multicriteria framework.

Now, the recruiter is able to use the defined profiles and distance metric, to find which of the candidates have similar profiles.

Construction of cluster centers

With this new distance metric, determining the central element of a cluster of alternatives is not as straightforward as it is with more traditional distance metrics, like the Euclidean distance formula. The construction of the central elements is based on a voting procedure:

> *Let $\{a_{i_1}, \cdots, a_{i_p}\}$ be the p alternatives in the i^{th} cluster. The profile of the i^{th} central element, noted $P(c_i)$, is determined by a voting procedure:*
>
> $$a_j \in P_k(c_i) \iff k = \text{argmax} \sum_{a_{i_l}} \mathbb{1}_{\{a_j \in P_k(a_{i_l})\}} \tag{5.2}$$

To determine the profile of cluster center c_i, each alternative is evaluated and assigned to the profile separately. Each alternative is assigned to the element (k) of the profile, for which it occurs most in the profiles of the alternatives in the cluster. For example, suppose that cluster i contains 10 alternatives: $\{a_{i_1}, \cdots, a_{i_{10}}\}$; then c_i is the central element of cluster i and we aim to define the profile of c_i: $P(c_i)$. Take for example alternative a_{i_5}, if alternative a_{i_5} is preferred (P_2) over 5 alternatives, not preferred (P_4) over 1 alternative, incomparable (P_1) to 3 alternatives and indifferent (P_3) with respect to itself; then alternative a_{i_5} will be assigned to $P_2(c_i)$.

Once there is a cluster of candidates, this voting procedure allows to find the "average" of the profiles of all candidates within that cluster.

MCDA clustering algorithm

Armed with the (new) definitions of the preference structure, profiles, distance metric and a method to construct cluster centers, De Smet and Guzmán present their extension of the k-means algorithm to the multicriteria framework: the multicriteria decision aid (MCDA) clustering algorithm.

[2] For the proof that this formula is in fact a valid distance metric, we refer the reader to the original publication (De Smet and Guzmán, 2003 [19]).

5.2 The Multicriteria Competency clustering framework

In this section, we alter the MCDA algorithm, such that it fits our clustering setting and present the Multicriteria Competency (MCC) clustering framework. In order to do so, we make changes to each of the elements of the MCDA algorithm: aim, preference structure, distance metric and construction of cluster centers. In short, we create our own *competency structure*, by assigning competencies to different categories, taking both the level- and importance scores into account. With this competency structure, we define the notion of profile to each job. Which in its turn is used in a new distance metric formula and for the construction of cluster centers.

Aim

The aim of the new algorithm is to group a set of jobs or functions $J = \{J_1, J_2, \cdots, J_j, \cdots J_n\}$, that are described by a set of competencies $X = \{x_1, x_2, \cdots, x_i, \cdots, x_m\}$, each for two scores $S = \{s_1, s_2\} = \{\text{level, importance}\}$.

Competency modelling and structure

De Smet and Guzmán created a model for grouping alternatives, using the decision maker's preference structure. In our application, such a preference structure is not relevant. Instead, we define a *competency structure*.

We create this competency structure for each job, by categorizing all m types of competencies, based on their scores. In order to make relevant categories, we look at the available information from the competency data. Recall that the importance scale ranges from 1 ("not important") to 5 ("extremely important") and that the relevancy indicator is defined as: a type of competency is "Not Relevant" when 75% of the importance ratings are rated as "not important". Note that it is possible for a type of competency to be relevant for that specific occupation, but not important. Also recall, the level scores are in the range 0 (lowest) to 7 (highest). For both scores, we create different categories, displayed in Table 5.1 and 5.2:

Importance category	Criterion
irrelevant	relevancy indicator has value "Yes"
low	importance score < 2.5
average	$2.5 \leq$importance score < 3.5
high	importance score ≥ 3.5

Table 5.1: Importance score categories

Level category	Criterion
low	level score < 2.5
average	$2.5 \leq$ level score < 4
high	$4 \leq$level score < 5.5
exceptionally high	level score ≥ 5.5

Table 5.2: Level score categories

In order to grasp the multi-dimensional aspect of the data, we create *overall categories*. In this way, we aim to preserve the relationship between the importance- and level scores. We do so by intersecting the importance- and level categories, resulting in 8 overall categories displayed in Table 5.3. They are first ordered by the importance category and then by the level category, from the lowest to the highest.

Overall category	Category name	Importance category	Level category
O_1	Irrelevant	irrelevant	—
O_2	Unimportant	low	—
O_3	Applicable	average	low
O_4	Favorable	average	average
O_5	Useful	average	high
O_6	Significant	high	average
O_7	Essential	high	high
O_8	Crucial	high	exceptionally high

Table 5.3: Overall competency categories

Then, the categories $\{O_1, O_2, O_3, O_4, O_5, O_6, O_7, O_8\}$ make up a *competency structure* if they satisfy the following conditions, $\forall J_j \in J$:

- $\sum_{k=1}^{8} \mathbb{1}_{\{x_i \in O_k\}} = 1 \;\; \forall x_i \in X$ — Each type of competency occurs exactly once in all categories
- $\sum_{k=1}^{8} |O_k| = m$ — A total of m types of competencies are categorized

In essence, for each job we need all types of competencies to be placed into different categories, based on both the importance- and level score, thus maintaining the multicriteria nature of the problem.

For the O*NET database, we found that with these categories, generally all combinations of importance- and level scores are contained within the categories for all available O*NET data. As there is data on 873 jobs, for which there is information on 120 types of competencies, there are $873 \times 120 = 104760$ combinations of jobs and types of competencies. Out of all these combinations, only 7 result in an importance- and level score that is outside of the defined categories. In case this occurs when categorizing the types of competencies, we propose that one manually adds this instance to one of the existing categories at their own discretion.

Profiles

Now, using the competency structure, we define the *profile* $P(J_j)$ of job $J_j \in J$ as being an 8-tuple: $< O_1(J_j), O_2(J_j), O_3(J_j), O_4(J_j), O_5(J_j), O_6(J_j), O_7(J_j), O_8(J_j) >$ where

- $O_1(J_j) = \{x_i \in X | x_i \in O_1\} = P_1(J_j)$ all types of competencies that are *irrelevant* to job J_j
- $O_2(J_j) = \{x_i \in X | x_i \in O_2\} = P_2(J_j)$ all types of competencies that are *unimportant* to job J_j
- $O_3(J_j) = \{x_i \in X | x_i \in O_3\} = P_3(J_j)$ all types of competencies that are *applicable* to job J_j
- $O_4(J_j) = \{x_i \in X | x_i \in O_4\} = P_4(J_j)$ all types of competencies that are *favorable* to job J_j
- $O_5(J_j) = \{x_i \in X | x_i \in O_5\} = P_5(J_j)$ all types of competencies that are *useful* to job J_j
- $O_6(J_j) = \{x_i \in X | x_i \in O_6\} = P_6(J_j)$ all types of competencies that are *significant* to job J_j
- $O_7(J_j) = \{x_i \in X | x_i \in O_7\} = P_7(J_j)$ all types of competencies that are *essential* to job J_j
- $O_8(J_j) = \{x_i \in X | x_i \in O_8\} = P_8(J_j)$ all types of competencies that are *crucial* to job J_j

Distance metric

Now that we have defined the competency structure and the notion of a profile for each job, we define a distance metric. Again, the intuition behind this concept is that two jobs are as close as their profiles are alike.

> *Let $P(J_j)$ be the profile of job J_j, the distance between two jobs $J_j, J_l \in J$ is defined as:*
>
> $$d(J_j, J_l) = 1 - \frac{\sum_{k=1}^{8} \left| P_k(J_j) \cap P_k(J_l) \right|}{m} \tag{5.3}$$

The proof by De Smet and Guzmán showing that Formula 5.1 is indeed an appropriate distance metric, is easily extended to Formula 5.3 in Section B.4. Since this is an unconventional distance metric, we no longer have the problem that the distances between points are ill-defined; as is the case with high dimensional data.

Construction of cluster centers

Also the way in which cluster centers are constructed, is slightly altered to fit the MCC clustering framework.

> *Let $\{J_{c_1}, \cdots, J_{c_p}\}$ be the p jobs in the c'th cluster. The profile of the c'th central element, noted $P(C_c)$, is determined by a voting procedure:*
>
> $$x_i \in P_k(C_c) \iff k = \operatorname{argmax} \sum_{J_{c_l}} \mathbb{1}_{\{x_i \in P_k(J_{c_l})\}} \tag{5.4}$$

In this case, a type of competency is assigned to the element (k) of the profile of the cluster center, for which it occurs most in the profiles of the jobs in the cluster.

The definitions of *aim*, *competency structure*, *profiles*, *distance metric* and *cluster center construction method* together make up the Multicriteria Competency clustering framework. This framework is very flexible and can be altered to match other use cases. For instance, the framework is not dependent on the O*NET database and can also be used on other types of competency data. Only the categorizations in Table 5.1, 5.2 and 5.3 might need some altering in order for them to match the corresponding scale ranges. Also, the amount of (overall) categories can be changed, along with the requirements, as long as there are at least two overall categories.

The MCC clustering framework can be used in multiple clustering algorithms. The following sections explain how the framework can be applied to a k-means clustering algorithm, as well as a hierarchical clustering algorithm.

5.3 Multicriteria Competency k-means clustering algorithm

Using the above defined competency structure, notion of profiles, distance metric and procedure to construct cluster centers, the k-means method is extended to the Multicriteria Competency k-means clustering algorithm, presented below.

Algorithm 1: The MCC k-means clustering algorithm

Input: $J = \{J_1, J_2, \cdots, J_j, \cdots J_n\}$, set of jobs
Input: k, the desired number of clusters
Input: Definition of $P(J_j)\ \forall J_j$ such that we have the 8-tuple:
$< O_1(J_j), O_2(J_j), O_3(J_j), O_4(J_j), O_5(J_j), O_6(J_j), O_7(J_j), O_8(J_j) >$
Input: Distance metric 5.3
Output: $C = \{C_1, \cdots, C_k\}$, set of clusters
Randomly initialize k cluster centers $(B_1, \cdots, B_k)$ such that $\forall c, \exists j\ B_c = J_j$;
Each cluster C_c is associated with cluster center B_c ;
repeat
    for *jobs* J_j, $j \in \{1, \cdots, n\}$ **do**
        Assign J_j to the cluster C_c with the nearest cluster center B_c i.e., $d(J_j, B_c) \leq d(J_j, B_l)$ where $c, l \in \{1, \cdots, k\}$
    end
    for *clusters* C_c, $c = \{1, \cdots, k\}$ **do**
        Update the cluster center B_c to be the centroid of C_c
    end
until *the cluster membership no longer changes*;

This algorithm is an application of the heuristic algorithm that is often referred to as the naive-, or Lloyd's k-means algorithm (Kanungo et al. [37]). The only difference is the individual elements that are now translated to the MCC clustering framework. In the iterative algorithm, first the cluster centers are randomly initialized to be one of the jobs. Then, the jobs are assigned to the nearest cluster, after which the cluster center is recalculated. This process iterates until the algorithm converges and the cluster memberships no longer change.

Metaprocedure

The algorithm is dependent on the random initialization, hence the algorithm converges to a local minimum. For that reason, a *metaprocedure* is necessary. In the metaprocedure, the MCC k-means clustering algorithm is executed many times, in order to keep on improving the local minimum solution found thus far. We assess the quality of the clusters using the Relative Margin metric, defined in Section 4.4.2.

Choice of k

As we have an unsupervised clustering problem, we do not know the number of clusters k. Therefore we perform the algorithm for different values of k, after which we can find the best number of clusters.

In order to do this, we resort to the *Relative Margin* clustering-quality measure, defined in Section 4.4.2. For multiple values of k, we run the algorithm and save the Relative Margins. Recall that we want the Relative Margin to be as small as possible, then the points are more confident about their cluster membership. However, when the number of clusters (k) grows, it automatically decreases the Relative Margin: if there are more cluster centers, the cluster centers will be closer to each other, decreasing the ratios. This introduces a trade-off between the number of clusters and the Relative Margin.

In order to choose the best k, we use the *Kneedle algorithm*, introduced by Satopää, Albrecht, Irwin and Raghavan (2011, [57]), to deal with this trade-off. They define a *knee point* as: *"...a point at which the relative costs to increase some tunable parameter is no longer worth the corresponding performance benefit. These knee points represent beneficial points that system designers have long selected to best balance inherent trade-offs."* (p. 1). Their generic approach to find these knee points in discrete data is the *Kneedle algorithm.* The algorithm uses the mathematical definition of curvature for a continuous function as the basis for their knee definition. Knees occur when a curve becomes more "flat", indicating a decrease in curvature; they summarize. Looking for a knee points in a graph portraying the trade-off between k and the Relative Margin, will help us select the best k.

Drawbacks of k-means

The proposed k-means algorithm still experiences the drawbacks that are known to affect the algorithm, on which we elaborated in Section 4.4.1.

5.4 Multicriteria Competency hierarchical clustering algorithm

The MCC clustering framework can also be applied to an agglomerative hierarchical clustering algorithm, simply by incorporating the distance metric introduced in Section 5.2. Hierarchical clustering algorithms have some advantages over the k-means algorithm. Firstly, the algorithm is not dependent on random initialization and an initial choice of k. Because of this, the algorithm is deterministic and no metaprocedure is needed. Secondly, clusters made with an hierarchical clustering algorithm can be visualized in a *dendrogram*, showing the hierarchical relationship between the clustered items.

Algorithm 2: The MCC hierarchical clustering algorithm

Input: $J = \{J_1, J_2, \cdots, J_j, \cdots J_n\}$, set of jobs
Input: Definition of $P(J_j)\ \forall J_j$ such that we have the 8-tuple:
$< O_1(J_j), O_2(J_j), O_3(J_j), O_4(J_j), O_5(J_j), O_6(J_j), O_7(J_j), O_8(J_j) >$
Input: Distance metric 5.3
Consider each item as its own cluster: initialize n clusters $(C_1, \cdots, C_n)$ such that $\forall c\ C_c = J_c$;
Output: A hierarchy structure; dendrogram
repeat
  Calculate the distances between all clusters, using *complete linkage*:
  $\forall i, j \in \{1, \cdots, c\}\ \ D(C_i, C_j) = \max_{x_1 \in C_i, x_2 \in C_j} d(x_1, x_2)$;
  Merge the two clusters with the closest distance;
  $c = c - 1$
until *there is one cluster left*;

Linkage methods

Whereas it is evident how to calculate the distance between two points, it is not evident how one should calculate the distance between a point and a cluster, or between two clusters. In order to decide which clusters to merge, we use *linkage methods*; some examples are: single linkage, complete linkage, centroid linkage, Ward's linkage and average linkage (Murtagh, 1983, [49]).

In our MCC hierarchical clustering algorithm, we propose the use of *complete linkage*. Then, two clusters with the closest maximum distance are merged. We propose to use the complete linkage, because it is able to handle our unusual distance metric and it has a simple and tractable character. The single linkage is also able to handle the distance metric, however the disadvantage is that the method exhibit the "chaining" effect, is undesirable in practice (Murtagh, 1983, [49]). Other linkage methods may be less suited, as they rely on averages or other transformation of the distances, which will not work for our distance metric.

Choice of k

Still, the choice on the numbers of clusters is ambiguous. For this, the *Weakest Link* cluster-quality measure introduced in Section 4.4.2 could be used.

6 — Employee Profile Integration in ASWP

In the past chapters, we laid a clear path to extending the Analytical Strategic Workforce Planning approach. Literature on the Analytical Strategic Workforce Planning approach, Competency Management, Competencies in SWP, Competency data and on methods for designing employee profiles was reviewed. This chapter uses all gathered information to extend the ASWP approach, and making it more robust to future (technological) developments.

The first section goes into more detail on the existing algorithm for optimization of HR interventions. In the second section, we propose how one could integrate employee profiles into the algorithm and the implications for the overall ASWP approach.

6.1 Optimization of HR interventions

In Section 3.3, some literature on SWP algorithms and methods was discussed. Bech (2013, [6]) concluded that there was in fact no good method that worked in practice. They defined mathematical definitions for targets, staff flow and assumptions, which we used to define a Linear Programming (LP) problem accordingly. After looking into multiple algorithms for solving this LP problem, Bech (2013, [6]) concludes that the *deductive algorithm* based on simulation of the evolution of the workforce, was best for finding an appropriate recruitment strategy in practice.

The algorithm has multiple variants, for this research, we will consider the algorithm that works on a macroscopic level. Meaning that the algorithm considers all employees of one specific job as one group, and not as individuals with their own transition probabilities and costs. So for one job, the employees have the same behaviour and costs. The deductive algorithm is able to deal with as many employees and jobs as necessary, as it relies on a deterministic approach. Below the algorithm is explained in more detail. For the specific mathematical definitions and assumptions, we refer to the original paper by Bech (2013, [6]).

6.1.1 Bech's Deductive Algorithm

The current- and desired workforce are both represented by three components: the total number of employees in the organization; the ratios or proportions by which the jobs occur in the workforce; and the financial value of the particular composition of employees. From this, three targets for the optimization model are defined, that aim to transform the current workforce into the desired workforce. The model minimizes:

- The headcount gap: the difference between the desired number of employees and the realized number of employees.
- The ratio gap: the quadratic sum of the gaps between the desired ratios and the realized ratios of all jobs.
- The financial gap: the gap between the desired financial value and the realized financial value; this could be in terms of either revenues, costs or profits.

The algorithm has two degrees of freedom, meaning that it is possible for two out of three target gaps to be put to zero; if firing employees is allowed. However, this option is not usually considered in related literature and Bech also demonstrates how firing leads to very drastic and unrealistic actions. Instead, the deductive algorithm is able to only set one of the three gaps to zero, while another gap can be set very close to zero. If desired, a company could find the optimal recruitment strategy for each of the three targets. Subsequently, they could take the weighted sum of the recruitment strategies found for each of the solutions. This however, is not the same as minimizing a weighted combination of gaps, which Bech proposes as an avenue for further research.

The shape of a company's workforce changes over time, this is what is called the *staff flow*. This flow is divided into three stages: outflow, throughput and inflow. *Outflow* is when employees leave the organization, which could be due to employees resigning, or by age dependent outflow like retirements. *Throughput* is when there are transitions of employees within the company, such as promotions and demotions to different function levels, but also transfer between teams or departments. *Inflow* is the actual hiring of new employees from outside the company. Now, the *goal* of the algorithm is to give a forecast on the inflow for each time period, resulting in an optimal recruitment strategy for the entire time horizon.

Every time period, the outflow and throughput are simulated. After that, the company knows how many employees are left in which function, thus the costs of the workforce can be calculated. The inflow for each function level is determined based on the number of vacancies in each function level and the chosen target to minimize. Once this is done for all time periods, the result is the optimal recruitment strategy per time period, which is suboptimal over the entire time horizon.

Now that we have seen the routine of the algorithm, we take a closer look how and where we can integrate the use of employee profiles in the Analytical Strategic Workforce Planning approach.

6.2 Employee profile integration in ASWP approach

This section explores ways to integrate employee profiles in the Analytical Strategic Workforce Planning approach, explained in Section 2.3.1, 3.1 and 6.1.1.

6.2.1 Designing the desired workforce

In their highly influential paper, Frey and Osborne (2017, [27]) examine how susceptible jobs are to computerisation. They implement a methodology to estimate the probability of computerisation for 702 six-digit SOC-level[1] occupations from the O*NET database. They used the O*NET database because of two important features. Firstly, *"...it defines the key features of an occupation as a standardised and measurable set of variables"* (p.263). Secondly, because it *"...provides open-ended descriptions of specific tasks to each occupation"* (p. 263). They aimed to determine which problems engineers need to solve for specific occupations to be automated from a technological capabilities point of view. Then, they looked for the similarities between the characteristics of these problems and characteristics belonging to specific occupations. In this way, they were able to research the way in which technological changes impact the occupational composition of the labour market, but also the number of jobs at risk if these technological changes actually happen. Their findings suggest that *"...recent developments will put a substantial share of employment, across a wide range of occupations, at risk in the near future."* (p. 266). According to their estimates around 47% of total US employment is in the high risk category. These jobs at risk are jobs that are expected to be automated relatively soon, perhaps over the next decade or two.

These automation probabilities can be used in determining the desired future workforce, in combination with the designed employee profiles. Once we have employee profiles, we can check whether one of these

[1] As explained in Section 4.3.1

jobs is at a great risk of becoming automated. In that way, companies can think ahead about up- and reskilling employees in those jobs, to jobs within the employee profile, that are less susceptible to automation. This can be implemented in the actual algorithm by means of the transition methods that calculate the company's throughput.

6.2.2 Transition methods: push and pull

The algorithm has multiple ways to calculate the company's throughput; both push- or pull, as well as hybrid push-pull method can be used. The core idea of the *push method* is that an employee gets promotion when s/he has developed enough skills and knowledge and is performing well in her/his current function, regardless if there is enough space or money available for this transition. With the *pull method*, the company's desired workforce is taken into account, as transitions only occur when they are desired. Essentially, employees are pulled into different jobs once there are open vacancies.

Push method

The push method is considered a Markov Process, where the internal transitions take place according to a discrete Markov chain. The Markov chain is characterized by its state space and transition probability matrix. Let P denote a transition probability matrix consisting of the transition probabilities p_{ij}, with p_{ij} the probability that an employee in function level i makes a transition to function level j in time interval $[t, t+1]$. The transition probabilities can be determined based on historical information or an expert opinion.

The transition probability matrix for a hierarchical organizational structure, as used in [6], typically looks like the matrix presented in Table 6.1. In this hierarchical organizational strucuture, job A is on the lowest level and job G on the highest level. This P-matrix also includes some transition rules, restricting the transitions to lower levels and to other levels than the one directly "above" the current level.

Job i/j	**A**	**B**	**C**	**D**	**E**	**F**	**G**
A	0.6	0.4	0	0	0	0	0
B	0	0.6	0.4	0	0	0	0
C	0	0	0.5	0.5	0	0	0
D	0	0	0	0.65	0.35	0	0
E	0	0	0	0	0.85	0.15	0
F	0	0	0	0	0	0.9	0.1
G	0	0	0	0	0	0	1

Table 6.1: Usual push transition probability matrix for a hierarchical organizational structure

Pull method

In the pull method, transitions only occur when they are desired; employees are "pulled up" once there is a vacany in other function levels. Similar to the push method, a transition probability matrix can be defined for the pull method. Let S be the $(k \times k)$ transition probability matrix with element s_{ij} representing the probability that a vacancy in function level j will be filled by an employee in function level i. The transition probabilities can be determined based on historical information or an expert opinion.

The transition probability matrix typically looks like the matrix presented in Table 6.2. This S-matrix also includes some transition rules, restricting the transitions to lower levels and to other levels than the one directly "above" the current level.

Job i/j	**A**	**B**	**C**	**D**	**E**	**F**	**G**
A	0	1	0.3	0.1	0.1	0	0
B	0	0	0.7	0.25	0.1	0	0
C	0	0	0	0.65	0.2	0	0
D	0	0	0	0	0.7	0.2	0
E	0	0	0	0	0	0.8	0.1
F	0	0	0	0	0	0	0.9
G	0	0	0	0	0	0	0

Table 6.2: Usual pull transition probability matrix for a hierarchical organizational structure

Integrating employee profiles

These push and pull methods very heavily depend on the hierarchical structure of the company, and have a very restricting transition rules. Now that HR practitioners have employee profiles, they are able to take another approach to constructing these matrices. Transitions that were not deemed possible before, now are possible.

Engaging in Strategic Workforce Planning for the first time, will create an optimal recruitment strategy and will provide information as to in which positions new employees should be hired. This information can be used to create up- and reskilling schemes for its employees within certain employee profiles. Then, these schemes can be integrated into the ASWP approach itself. This can be done by reflecting the schemes into the transition probability matrices. For example, when a certain upskilling scheme is designed for employees, the transition probabilities between those jobs might change. Also HR policies can affect the transition probabilities. If there is a clear goal of decreasing the amount of employees in one function, the transition probabilities could reflect this by having a small probability of staying in the same job and a higher probability of going to another job. Obviously, these systems should also work together and involve prior knowledge on the company; as it should not be possible for anyone to become Chief Executive, just because they are in the same employee profile. In that way, companies can get more in control of their workforce development. This is a continuous process that has to be monitored and evaluated regularly.

An example of such a new transition probability matrix (for the push) method, could look like the one in Table 6.3. Note that there will still be some hierarchy within the different jobs. In this example, we could say that jobs C, F, G and H are in the same employee profile; with the employees in job C being upskilled to other jobs within the employee profile.

Job i/j	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
A	0.6	0.4	0	0	0	0	0	0	0	0
B	0	0.6	0.4	0	0	0	0	0	0	0
C	0	0	0.2	0	0	0.6	0	0.4	0	0
D	0	0	0	0.7	0.3	0	0	0	0	0
E	0	0	0	0	0.85	0.15	0	0	0	0
F	0	0	0	0	0	0.9	0.1	0	0	0
G	0	0	0	0	0	0.3	0.6	0	0	0
H	0	0	0	0	0	0.3	0	0.7	0	0
I	0	0	0	0	0	0	0	0	0.9	0.1
J	0	0	0	0	0	0	0	0	0	1

Table 6.3: Possible new push transition probability matrix, using employee profiles

Before, the transition matrices were stationary, meaning that for each time period, the transitions probabilities are similar. However, with this new information, the company could resort to more complex transition matrices. For example in the first year, not much will change; but maybe in the second year, the up- and reskilling program is beginning to yield results and the probabilities of changing to another job might become higher/lower for some jobs. In this way, the up- and reskilling programs within the company, can be reflected in the transition probability matrices.

7 — Case Study on Designing & Integrating Employee Profiles

In this chapter, the information and algorithms introduced in the chapters before, are applied in a case study on a company within the airline industry. First, we look at the company's data and how we should prepare this for the analysis and methods. Next, we obtain more insight into the data, amongst others by using Principal Component Analysis (PCA). Then, we apply the Multicriteria Competency clustering framework and algorithms to the company's data and show their results. Lastly, we look at how these results can be integrated into the ASWP approach.

7.1 Employee data

For this research, employee data from a company within the airline industry is used. This is a dummy dataset, inspired by previous clients of PwC and amended in order to make it permissible for usage in this research. In order to get some insights into the company and its workforce, this section contains some descriptive information.

To get a better idea of the workforce, we look at some of the workforce's demographics and characteristics. In total, the data contains information on 1301 unique employees over the years of 2017 and 2018; which work in 35 different occupations within the company. Table C.1 in Appendix C.1 contains these occupations and the corresponding number of employees. Information like the amount of FTE worked and the distribution between female and male employees are shown in Table 7.1 below. Also, the employees' ages in 2018 are depicted in Figure 7.1, showing the age diversity within the company.

(#)	**2017**	**2018**
Employees	1200	1216
average FTE	0,803	0,804
FTE	963,6	977,8
Unique jobs	35	35
Females	577 (48,1%)	590 (48,5%)
Males	623 (51,9%)	626 (51,5%)

Table 7.1: Employee demographics and characteristics

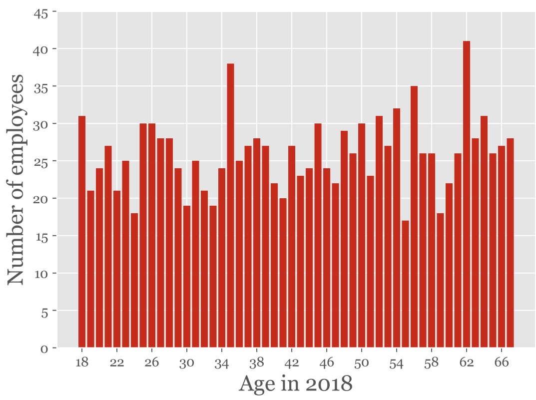

Figure 7.1: Employees' ages in 2018

To gain insights into the development of the workforce, we look at transformations that took place in the last few years in Table 7.2. It is possible for employees to either join, leave or change jobs within the company. Figure 7.2 shows the length of employment within the company for current employees.

	Hired (#)	Stopped (#)	Transitions (#)
2017	0	85	4
2018	101	—	—

Table 7.2: Workforce transformations

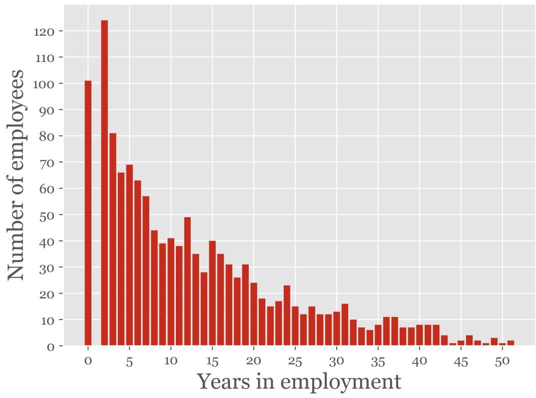

Figure 7.2: Length of employment of current employees in 2018

7.2 Data preparations

This section explains the steps required to prepare the data such that it is ready to use for descriptive analyses and the algorithms explained in Section 5.2.

Firstly, the employee data and the competency data have to be matched. For each job within the company, the corresponding O*NET occupation and unique SOC-code should be found. For our data, the combinations of these jobs and SOC-codes are displayed in Table C.1 in Appendix C.1. Also, this table contains the *probabilities of automation* for each job, as calculated by Frey and Osborne (2017, [27]). The only occupation that is missing, is the "'Human Resources Specialists" ($13-1071.00$); instead, the probability of automation for "Human Resources; Training; and Labor Relations Specialists; All Other" (13-1078) is used.

Secondly, we only need the competency data corresponding to the (future) jobs of the company; all competency data on other jobs is disregarded.

Thirdly, the most recent version of the data does not contain competency data on two of the company's occupations: "Public Relations and Fundraising Managers" and "Business Operations Specialists, All Other".

The occupation "Public Relations and Fundraising Managers" (SOC Code 11-2031.00) is recently discontinued[1] and divided into two new occupations: "Public Relations Managers" (11-2032.00) and "Fundraising Managers" (11-2033.00). However, since they only separated recently, there is no competency data available yet. For that reason, we use the data from the O*NET 25.0 database, published in August 2020, as an alternative for this occupation's competency data. In that version of the data however, the type of knowledge *Administrative* was named *Clerical* ; this should be adjusted.

The occupation "Business Operations Specialists, All Other" (SOC Code 13-1199.00) has no competency data in the O*NET database, as it is not included in the data collection plan. The "All Other" in the title represent occupation with a wide range of characteristics which do not fit into one of the detailed O*NET-SOC occupations. The occupations listed under this title are: "Business Continuity

[1] Report August 2019: `https://www.onetcenter.org/reports/Taxonomy2019.html`

Planners" (13-1199.04) and "Sustainability Specialists" (13-1199.05), "Online Merchants" (13-1199.06) and "Security Management Specialists" (13-1199.07)[2]. Hence, the average scores of these underlying occupations are used as a proxy for the scores of the occupation "Business Operations Specialists, All Other".

Lastly, both the types of *skills* and *knowledge* descriptors contain *Mathematics*. Hence, these should be renamed, such that they are distinct, for example: "mathematics_skills" and "mathematics_knowledge".

As a result of these preparatory steps, the competency data on 35 jobs is presented in one big overview. This overview contains 8400 rows (= 35 jobs × 120 types of competencies × 2 scores) and 15 columns, as described in Table B.5 in Appendix B.2.

7.3 Preliminary data insights

In this section, we perform some descriptive analyses which provide insight into the data available for this research.

7.3.1 Data (multi-)dimensionality

After the data preparations of the previous section, the big overview consist of 8400 records of data: each line contains an occupation ($J = 35$), a type of competency ($n = 120$) and its corresponding score ($S = 2$). Additionally, the overview contains statistical information and the indicators mentioned before.

This overview is transformed to only include the actual data values and not all complementary information. This results in a dataframe that contains 35 rows (occupations) and a two-level hierarchical column indexing, with 120 types of competencies, which each have two types of scores. This feature of our data is referred to as *multi-dimensionality*. We do not simply have 2D data, with observations and variables; we have 3D data, as illustrated in Figure 7.3.

Now that the most relevant data is in one dataframe, we clearly see the data exhibits *High Dimensionality and Low Sample Size (HDLSS)*: with 35 observations and 120 (or 240) dimensions, we have $J >> n$.

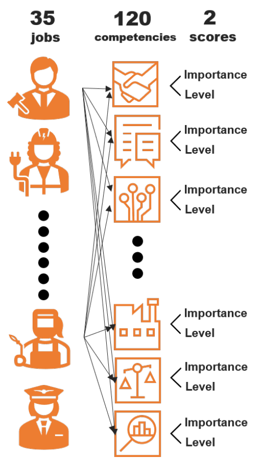

Figure 7.3: Illustration of company's three-dimensional competency data

[2]Before August 2019, also "Energy Auditors" (13-1199.01, now: 47-4011.01), "Security Management Specialists" (13-1199.02, now: 13-1199.07) and "Customs Brokers" (13-1199.03, now:13-1041.08) were listed under this title. From August 2019 onward, they were assigned new categories.

7.3.2 Competency scores

Table C.2 in Appendix C.1 displays statistics on the level- and importance scores for all of the types of competencies. For both scores, the mean, median, minimum and maximum values are displayed.

Figure 7.4 shows 4200 data points: for 35 jobs with 120 types of competencies, their importance- and level scores are depicted on the axes. It is easily seen that the importance- and level scores are positively correlated. When a type of competency becomes more important to an occupation, it makes sense that the individuals performing that occupations have a higher level in that type of competency.

Figure 7.4: Plot of 4200 (35×120) points displaying the importance- and level scores of each job, for each type of competency

Table C.1 in Appendix C.1 also shows the percentage of the data that is considered to have low precision (as described in Section 4.3) or when the precision is not known, for each O*NET occupation. Before, it was already explained that some of the occupations in the *knowledge* attribute, did not contain the statistical information and thus also no indicator of precision. In Table C.1, these are recognized by the value of $27,5\%$, because for a total of 66 out of 240 data values, the statistics are unavailable[3].

7.3.3 Data relevancy

The data contains information on the relevancy of the types of competencies for each occupation. The column "Not Relevant" attains the value "Y" for the level scores, if that level score is deemed irrelevant. This is the case if 75% of the respondents rated importance for that type of competency as "not important" (score 1 on the scale of 1 to 5). Otherwise, the value is equal to "N".

In an effort to reduce the dimensionality of the data, we looked at the relevance of all types of competencies. In the case that a certain competency is not relevant for any of the occupations, we could consider disregarding that type of competency. Table C.1 in Appendix C.1 displays the amount of irrelevant competencies per occupation. We see that on average, $25,5$ out of 120 competencies are irrelevant, with the minimum and maximum at 7 and 43 types of irrelevant competencies per occupation, respectively.

Also, Table C.2 displays all the types of competencies, together with the amount of occupations for which the type of competency is relevant. These numbers are also displayed in Figure 7.5 below. It is seen that on average, each type of competency is relevant to $27,5$ out of 35 jobs. In total, for 63 of the 120 types of competencies, the particular competency is relevant to all occupations. Also, only one type of competency (Dynamic Flexibility) is relevant to just one job. In Figure 7.5 we again see the positive correlation: as a type of competency becomes more relevant (thus more important) to more jobs, the mean level scores of this competency go up.

[3] $\frac{66 \text{ (2 scores} \times \text{33 types of knowledge)}}{240 \text{ (2 scores} \times \text{120 types of competencies)}} = 0,275$

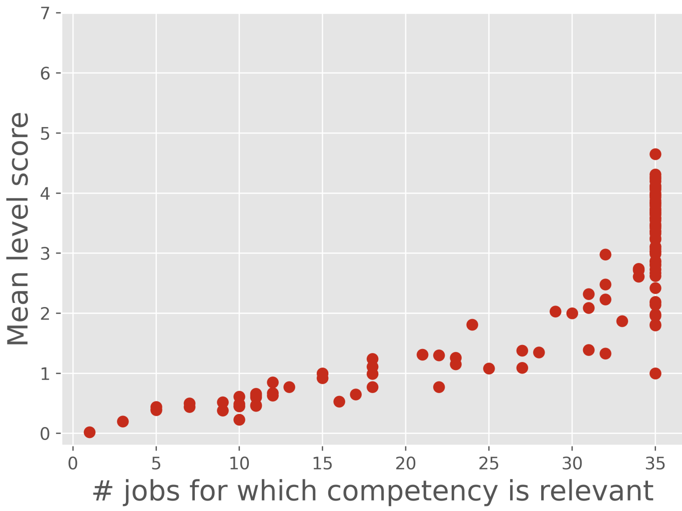

Figure 7.5: Mean level scores of all types of competencies, with a count on how many times this type of competency is relevant for the company's occupations

All in all, we found that each type of competency, is relevant for at least one occupation. Hence, leaving out types of competencies and thus reducing dimensionality, is not desirable as this leads to losing possibly relevant information when representing the company's workforce by their competencies.

7.3.4 Data visualization using PCA

In order to get a better idea of the occupations and their competencies and to get more insights into the similarities between occupations, we use Principal Component Analysis (PCA). PCA is a method that summarizes (correlated) variables into a few representative variables and is often used as a dimension reduction method, or to visualize high dimensional data.

PCA does not allow for multi-dimensional data, as the underlying optimization problem depends on two-dimensional matrices. In order to use PCA, we change the distance measure to Formula B.1, introduced by Wasid and Ali (2018, [68]). In this way we transform our 3D data into 2D data, simply by aggregating the importance- and level scores. Now, our data consists of 35 occupations, with 240 types of competencies and -scores. When performing PCA, the data should be standardized (mean 0 and unit variance). Also, performing PCA by using the NIPALS method is advised, as this method deals better with high dimensional data.

Figure 7.6 below shows 2-dimensional points that best represent the 240-dimensional data, corresponding to the 35 occupations, together with an indicator of their corresponding "SOC major groups". From this very basic, exploratory analysis, we already see how occupations and their competency scores, relate towards each other. Some occupations are much alike, with almost overlapping points. Other occupations stand more on their own, showing some *distance* between that occupation and others. What stands out is that we see that even though occupations belong to the same SOC major group, in this analysis, they are still relatively far apart. On the other hand, we also see occupations with a different SOC major groups being close to each other. This implies that we should not simply use the occupational taxonomy in order to see which employees should be grouped together. Instead, the information on employees' competencies really have additional value.

Figure 7.6: Principal Component Analysis with NIPALS algorithm: First two components for 35 jobs from 9 "SOC major groups"

However, using this Principal Component Analysis to cluster employees based on their competencies, it not desirable.

Firstly, reducing dimensionality by using PCA leads to losing quite a lot of information. Figure 7.7 shows how much of the variance is explained by the number of components used. Using more components, leads to having more information, which then represent the true data better. In Figure 7.6, we used only two principal components which only accounts for about 60% of the information contained in the true data. This means that this two dimensional representation does not account for 40% of the information, which could be valuable. Also, since we aggregated the importance- and level scores, we possibly lost the information embedded in the multi-dimensionality of the data.

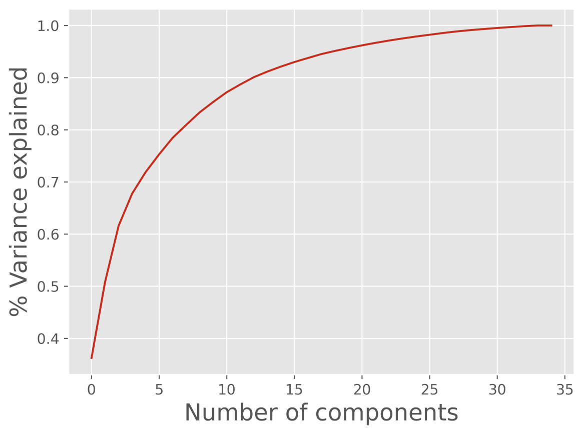

Figure 7.7: Principal Component Analysis with NIPALS algorithm: cumulative percentage of variance explained by components

Additionally, our data still has a low sample size, which causes problems. Because the NIPALS algorithm uses regression estimates which are dependent on sample size, this PCA analysis is probably inconsistent. Meaning that these principal components do not actually represent the data as best as possible. Hence it would be inaccurate to use PCA.

Despite the arguments for not using PCA, this analyses still give us a better insight into the data. All in all, these insights illustrate and motivate the need for more complex clustering methodologies to make meaningful clusters of employees; taking into consideration our *multi-dimensional* and *High Dimensionality and Low Sample Size* data. In the next section, we apply the framework and algorithms presented in Chapter 5.

7.4 Applying the Multicriteria Competency clustering algorithms

In this section, we apply the Multicriteria Competency clustering algorithms to the case of the airline company. First, the MCC clustering framework is created and requirements are checked. The MCC k-means and - hierarchical algorithm are applied and results are provided, in the second and third section respectively.

7.4.1 The Multicriteria Competency clustering framework

The aim of applying the algorithm is to group the set of 35 jobs $J = \{J_1, J_2, \cdots, J_j, \cdots J_{35}\}$, based on 120 types of competencies $X = \{x_1, x_2, \cdots, x_i, \cdots, x_120\}$, that each have two scores $S = \{s_1, s_2\} = \{\text{level, importance}\}$.

For each job, we define the competency structure by categorizing all 120 types of competencies, based on their importance- and level scores. In order for the competency structure to be consistent, we need to satisfy the following requirements, $\forall J_j \in J$:

- $\sum_{k=1}^{8} \mathbb{1}_{\{x_i \in O_k\}} = 1 \;\; \forall x_i \in X$ Each type of competency occurs exactly once in all categories
- $\sum_{k=1}^{8} |O_k| = m$ A total of m types of competencies are categorized

These requirements ensure that all competencies are categorized correctly and completely. For this case study, the categorization of the scores are displayed in Figure 7.8; we see that all combinations of importance- and level scores fall within the defined categories, thus satisfying the requirements.

Next, using these competency structures, each job can be represented by a profile $P(J_j)$ for job $J_j \in J$. Finally, distances between jobs are defined by distances metric 5.3 and cluster centers are constructed as explained in Section 5.2.

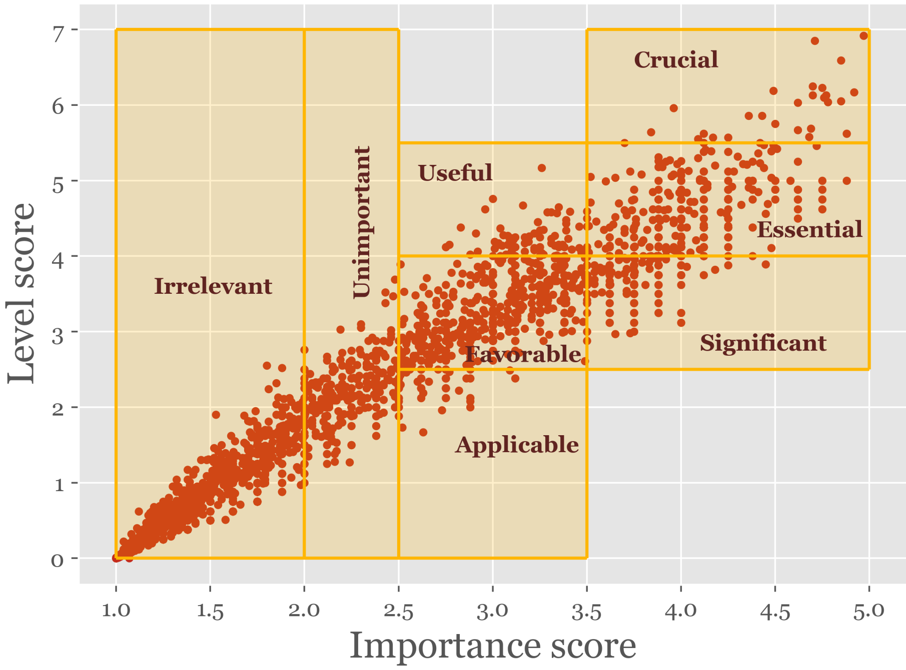

Figure 7.8: Competency structures of 35 jobs

7.4.2 Results MCC k-means clustering algorithm

Ready to apply the MCC k-means clustering algorithm explained in Section 5.3, we first need to search for the best number of clusters k.

Choice of k

When choosing the value of k, we want to make sure that we find the number of clusters, that best represents the data and that organizes the data in some meaningful way. It is always important to include intuition and prior knowledge in the decision, as we have seen that we should always study a clustering in the context of its end-use, in Section 4.4.2.

As explained in Section 5.3, we run the metaprocedure for the values of $k = 2$, up to $k = 18$; each time, the algorithm is run 5000 times. Then, we will use the *Kneedle algorithm* to find the value of k that has the best trade-off.

In Figure 7.9, boxplots of the Relative Margins of each run are displayed for the various values of k. Recall that the quality of the clustering is better if the Relative Margin is lower. The box extends from the lower (Q_{25}) to upper (Q_{75}) quartile values of the data, with a line at the median; the whiskers extend from the box to show the range of the data. Also, the red line plots the median values. We use the median, as this value is more robust to outliers, compared to the average of the Relative Margins. Figure 7.10 shows the size of the box, $Q_{75} - Q_{25}$, for each of the different number of clusters k.

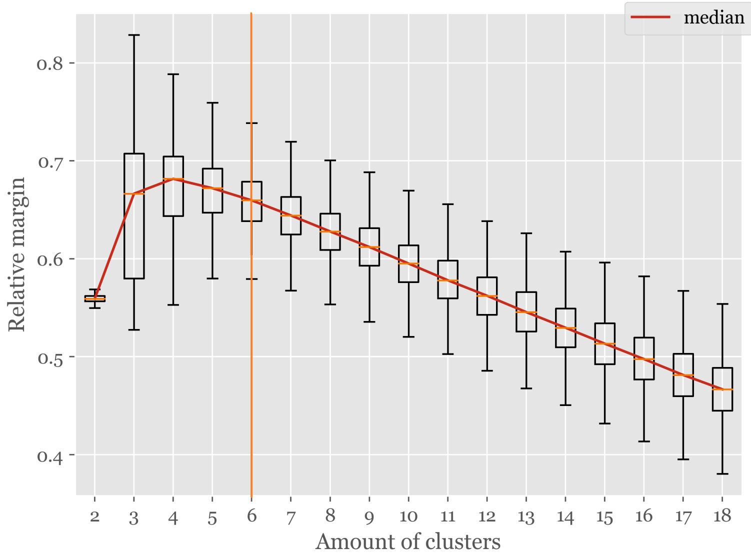

Figure 7.9: Relative Margins for different number of k, each for 5000 runs

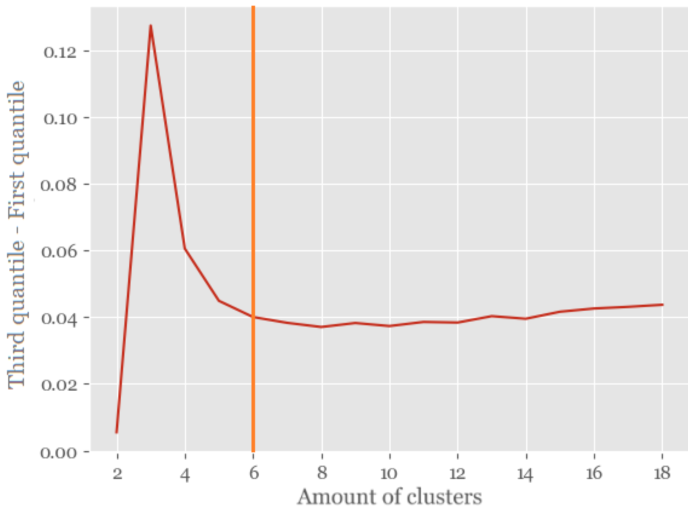

Figure 7.10: Size of box or Relative Margins: $Q_{75} - Q_{25}$

Applying the Kneedle algorithm to the median Relative Margin, results in selecting the number of cluster to be 6. Meaning that when going from 5 to 6 clusters, we still gain knowledge; however increasing the number of clusters from 6 to 7 or more, no longer gives new information about the structure of the jobs. Also, when looking at Figure 7.10, we see that increasing the number of clusters from 5 to 6, really decreases the size of the box. However, increasing the number of clusters more, does not results in less variable clusterings. From these analyses we conclude that we cluster the data into 6 groups.

Results for $k = 6$

Having determined the number of clusters, we run the metaprocedure again, but this time for 10.000 runs. Table 7.3 shows the best result of clustering the 35 jobs into 6 clusters using the MCC k-means algorithm; with a Relative Margin of 0.48574.

Employee clusters
(Relative Margin = 0.48574)

Cluster #	Occupations
A	General and Operations Managers; Compliance Officers; Administrative Services Managers; Industrial Production Managers; Airfield Operations Specialists
B	Flight Attendants
C	Aircraft Mechanics and Service Technicians; Aircraft Structure, Surfaces, Rigging, and Systems Assemblers; Aerospace Engineering and Operations Technologists and Technicians
D	Airline Pilots, Copilots, and Flight Engineers; Commercial Pilots
E	Air Traffic Controllers
F	Business Operations Specialists, All Other; Management Analysts; Lawyers; Market Research Analysts and Marketing Specialists; Financial Managers; Human Resources Specialists; Sales Managers; Computer and Information Systems Managers; Training and Development Specialists; Marketing Managers; Chief Executives; Logisticians; Human Resources Managers; Transportation, Storage, and Distribution Managers; Operations Research Analysts; Compensation, Benefits, and Job Analysis Specialists; Purchasing Managers; Travel Agents; Public Relations and Fundraising Managers; Aerospace Engineers; Training and Development Managers; Advertising and Promotions Managers; Compensation and Benefits Managers

Table 7.3: Result MCC k-means clustering algorithm with $k = 6$ and Relative Margin = 0.48574

In this clustering, we see that there are 5 well distinguished clusters: A, B, C, D amd E. These cluster contain 1, up to 5 jobs that are distinct from the other jobs. However, we also see that the sixth cluster contains 23 jobs. This is more than half of the total jobs, essentially making the sixth cluster "everything else", and hence not as clearly defined as the other clusters.

Besides looking at the clusters of the jobs themselves, we also take a deeper look into each of the clusters. The centers of the clusters contain a lot of valuable information and we already represent the centers by means of a profile. Table 7.4 shows the profile of the center of cluster C, from Table 7.3. In Appendix C.2, the profiles of clusters A, B, D, E and F are also depicted, in Table C.3, C.4, C.5, C.6 and C.7, respectively. Cluster B and E only comprise of one job, hence their cluster center is similar to the job itself, as well as their profiles.

Employee Profile for cluster C:
Aircraft Mechanics and Service Technicians; Aircraft Structure, Surfaces, Rigging, and Systems Assemblers; Aerospace Engineering and Operations Technologists and Technicians.
(Relative Margin = $0,48574$)

Category	**Types of competencies**
Crucial	Mechanical
Essential	Written Comprehension; Problem Sensitivity; Inductive Reasoning; Critical Thinking
Significant	Oral Expression; Deductive Reasoning; Near Vision; Reading Comprehension; Complex Problem Solving; Quality Control Analysis
Useful	—
Favorable	Administration and Management; Administrative; Customer and Personal Service; Production and Processing; Computers and Electronics; Engineering and Technology; Design; Mathematics (knowledge); Education and Training; English Language; Public Safety and Security; Oral Comprehension; Written Expression; Fluency of Ideas; Originality; Information Ordering; Category Flexibility; Mathematical Reasoning; Memorization; Speed of Closure; Flexibility of Closure; Perceptual Speed; Visualization; Selective Attention; Time Sharing; Arm-Hand Steadiness; Manual Dexterity; Finger Dexterity; Control Precision; Multilimb Coordination; Reaction Time; Static Strength; Trunk Strength; Extent Flexibility; Far Vision; Visual Color Discrimination; Depth Perception; Hearing Sensitivity; Auditory Attention; Speech Recognition; Speech Clarity; Active Listening; Writing; Speaking; Mathematics (skills); Science; Active Learning; Learning Strategies; Monitoring; Social Perceptiveness; Coordination; Persuasion; Instructing; Service Orientation; Equipment Selection; Operations Monitoring; Operation and Control; Equipment Maintenance; Troubleshooting; Repairing; Judgment and Decision Making; Systems Analysis; Systems Evaluation; Time Management
Applicable	Rate Control
Unimportant	Economics and Accounting; Sales and Marketing; Personnel and Human Resources; Building and Construction; Physics; Chemistry; Psychology; Geography; Therapy and Counseling; Law and Government; Telecommunications; Communications and Media; Transportation; Number Facility; Spatial Orientation; Response Orientation; Wrist-Finger Speed; Speed of Limb Movement; Explosive Strength; Dynamic Strength; Gross Body Coordination; Gross Body Equilibrium; Night Vision; Peripheral Vision; Glare Sensitivity; Sound Localization; Negotiation; Operations Analysis; Technology Design; Installation; Programming; Management of Financial Resources; Management of Material Resources; Management of Personnel Resources
Irrelevant	Food Production; Biology; Sociology and Anthropology; Medicine and Dentistry; Foreign Language; Fine Arts; History and Archaeology; Philosophy and Theology; Stamina; Dynamic Flexibility

Table 7.4: Employee profile of cluster center C

This representation of the cluster center really shows why the three jobs of “Aircraft Mechanics and Service Technicians”, “Aircraft Structure, Surfaces, Rigging, and Systems Assemblers” and “Aerospace Engineering and Operations Technologists and Technicians” are put into one cluster by the algorithm. Cluster C is the only cluster that has the types of competencies of “Mechanical” and “Quality Control Analysis” in one of the higher categories. For all other clusters profiles, these types of competencies fall in much lower categories.

For cluster A, in Table C.3, we see that there is no “crucial comeptency” connecting all the jobs. Also the types of competencies in the “Essential” and “Significant” categories are not as original, compared to the other profiles. What does stand out are the “Coordination” and “Computers and Electronics” types of competencies, which are categorized higher compared to the other profiles.

Cluster B, in Table C.4, only contains the job of “Flight Attendants”. For that reason, the profile of the cluster center is similar tot he profile of that job. Generally, we see that there are not many types of competencies in the three highest categories. But we do notice that there are some types of competencies that are categorized much higher than they are for other employee profiles, for example: “Foreign Language” , “Medicine and Dentistry”, “Stamina” and “Static Strength”. These are types of competencies that are unimportant or irrelevant to all other employee profiles, ensuring that this job distinguishes itself from other jobs.

For cluster D, in Table C.5, which is the cluster of the two different types of pilots, we see many types of competencies in the top three categories. Also, many of these types of competencies are unimportant or irrelevant for all other employee profiles, for example: “Spatial Orientation”, “Geogrpaphy”, “Multilimb Coordination” and “Response Orientation”. This shows that the high scores on the level and importance of those types of competencies really ensures that those two jobs are put in the same category.

Cluster E, in Table C.6, only contains the job of “Air Traffic Controllers”. Thus again, the profile of the cluster center is similar tot he profile of that job. Here, we see that there is the type of competency of “Time Sharing” is crucial. Also, we see many types of competencies that are essential, that are not common for the other employee profiles. This shows that an Air Traffic Controller has a very specific set of competencies, differentiating the job from other jobs.

For cluster F, in Table C.7, we see quite some types of competencies in the higher categories. However, they are not very original and other employee profiles also categorized these type of competencies quite high. We see that the top 4 categories mostly contain types of competencies that are related to competencies generally related to business and general operations of the company. What does stand out is the fact that there are very many irrelevant types of competencies. Whereas the other employee profiles have around 10 irrelevant types of competencies, the employee profile of cluster F has 33. Overall, cluster F contains many jobs that do not embody a unique aspect or outstanding type of competency as to why these jobs are in one cluster. That makes cluster F a collection of “all other” jobs, resulting in a very all-round profile.

Results for $k = 10$

Not all clusters in Table 7.3 are equally meaning full, partly because the amount of jobs in the clusters is unequally divided. For that reason, we perform the algorithm again, but this time with with 10 clusters, in an effort to split up this big cluster. We choose $k = 10$, because in Figure 7.10 we see a small drop in the size of the box, compared to the sizes of the boxes at 9 and 11 clusters.

We run the metaprocedure again with 10.000 runs and $k = 10$. Table 7.5 shows the best clustering found, with a Relative Margin at 0.43137.

Employee clusters
(Relative Margin = 0.43137)

Cluster #	Occupations
A	Air Traffic Controllers
B	Airline Pilots, Copilots, and Flight Engineers; Commercial Pilots
C	Compliance Officers
D	General and Operations Managers; Administrative Services Managers
E	Industrial Production Managers
F	Airfield Operations Specialists
G	Aerospace Engineering and Operations Technologists and Technicians
H	Aircraft Mechanics and Service Technicians; Aircraft Structure, Surfaces, Rigging, and Systems Assemblers
I	Flight Attendants
J	Business Operations Specialists, All Other; Management Analysts; Lawyers; Market Research Analysts and Marketing Specialists; Financial Managers; Human Resources Specialists; Sales Managers; Computer and Information Systems Managers; Training and Development Specialists; Marketing Managers; Chief Executives; Logisticians; Human Resources Managers; Transportation, Storage, and Distribution Managers; Operations Research Analysts; Compensation, Benefits, and Job Analysis Specialists; Purchasing Managers; Travel Agents; Public Relations and Fundraising Managers; Aerospace Engineers; Training and Development Managers; Advertising and Promotions Managers; Compensation and Benefits Managers

Table 7.5: Results MCC k-means clustering algorithm with $k = 10$, Relative Margin = 0.43137

Here, we actually see that the already existing clusters have become even smaller, or more specialized. Whereas the big cluster has remained exactly the same. This actually does not tell us more about the jobs in cluster J, but it does say more about the structure of the other clusters. This finding might indicate a hierarchical structure within the clusters.

7.4.3 Exploration MCC hierarchical clustering algorithm

This section explores the results of the MCC hierarchical clustering algorithm. Applying the MCC hierarchical algorithm presented in Section 5.4, does not require an initial choice of k and the result is the dendrogram presented in Figure 7.11.

Choice of k

The resulting cluster membership does of course dependent on the choice of k and that choice remains ambiguous. Due to limited time, we have not calculated the Weakest Link cluster-quality measure introduced in Section 4.4.2 for different clusterings. For that reason, we will provide the results for $k = 6$ and $k = 10$, similar to the previous section. This also allows us to compare the results of the two different algorithms.

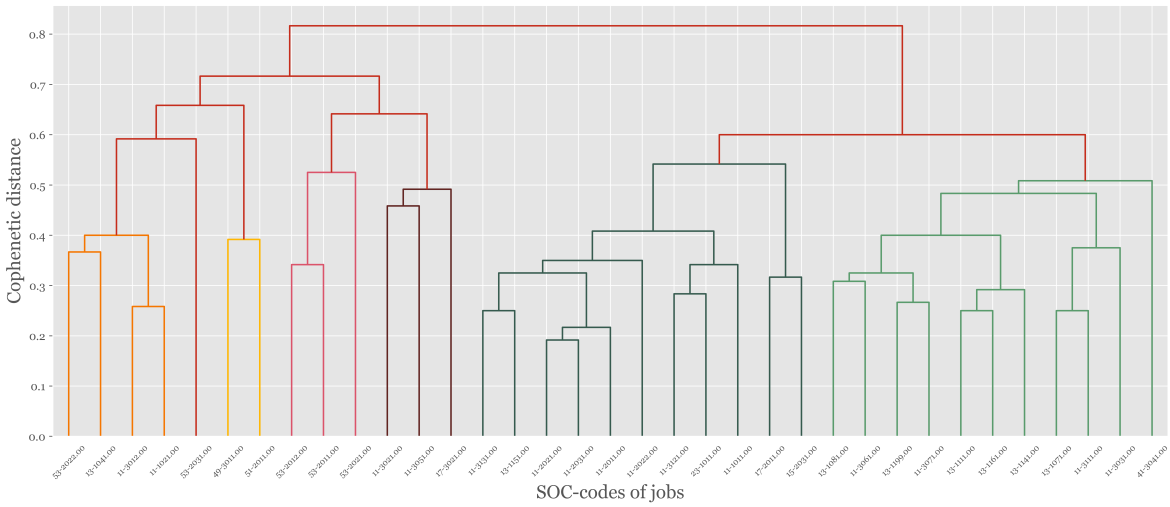

Figure 7.11: Dendrogram for MCC hierarchical clustering algorithm

Results for $k = 6$

In Table 7.6, the resulting cluster membership is presented for $k = 6$.

Employee clusters

Cluster #	**Occupations**
A	Aircraft Mechanics and Service Technicians; Aircraft Structure, Surfaces, Rigging, and Systems Assemblers
B	General and Operations Managers; Compliance Officers; Administrative Services Managers; Airfield Operations Specialists; Flight Attendants
C	Air Traffic Controllers; Airline Pilots, Copilots, and Flight Engineers; Commercial Pilots
D	Computer and Information Systems Managers; Industrial Production Managers; Aerospace Engineering and Operations Technologists and Technicians
E	Lawyers; Sales Managers; Training and Development Specialists; Marketing Managers; Chief Executives; Human Resources Managers; Operations Research Analysts; Public Relations and Fundraising Managers; Aerospace Engineers; Training and Development Managers; Advertising and Promotions Managers
F	Business Operations Specialists, All Other; Management Analysts; Market Research Analysts and Marketing Specialists; Financial Managers; Human Resources Specialists; Logisticians; Transportation, Storage, and Distribution Managers; Compensation, Benefits, and Job Analysis Specialists; Purchasing Managers; Travel Agents; Compensation and Benefits Managers

Table 7.6: Results MCC hierarchical clustering algorithm with $k = 6$

We see that the number of jobs per cluster is more equally divided, as there is no longer one big cluster containing “everything else”. More specifically, we see that the large “all-round” cluster in Table 7.3 has been roughly divided between two clusters: E and F. Despite the similarities that can be seen between the two clustering, there are some differences that make the clusters of Table 7.6 less specific; for example the fact that Air Traffic Controllers and Flight Attendants are now included in other clusters.

Results for $k = 10$

In Table 7.6, the resulting cluster membership is presented for $k = 10$.

Employee clusters

Cluster #	Occupations
A	Aircraft Mechanics and Service Technicians; Aircraft Structure, Surfaces, Rigging, and Systems Assemblers
B	Travel Agents
C	Flight Attendants
D	Operations Research Analysts; Aerospace Engineers
E	Airline Pilots, Copilots, and Flight Engineers; Commercial Pilots
F	Air Traffic Controllers
G	Computer and Information Systems Managers; Industrial Production Managers; Aerospace Engineering and Operations Technologists and Technicians
H	General and Operations Managers; Compliance Officers; Administrative Services Managers; Airfield Operations Specialists
I	Lawyers; Sales Managers; Training and Development Specialists; Marketing Managers; Chief Executives; Human Resources Managers; Public Relations and Fundraising Managers; Training and Development Managers; Advertising and Promotions Managers
J	Business Operations Specialists, All Other; Management Analysts; Market Research Analysts and Marketing Specialists; Financial Managers; Human Resources Specialists; Logisticians; Transportation, Storage, and Distribution Managers; Compensation, Benefits, and Job Analysis Specialists; Purchasing Managers; Compensation and Benefits Managers

Table 7.7: Results MCC hierarchical clustering algorithm with $k = 10$

Because of the inner workings of hierarchical clustering, the clusters of Table 7.7 will always be subsets of clusters of Table 7.6. We see that the cluster membership is starting to look more like the membership in Table 7.5; Air Traffic Controllers, Pilots, Flight Attendants and Aircraft Mechanics and Technicians are each clustered into their own cluster again. We also see some new relationships, that were not there in the results of the MCC k-means algorithm. Namely the grouping of “Operations Research Analysts” and “Aerospace Engineers”, for the k-means results, these were part of the sixth, less defined cluster before.

Employee profile design

The employee profiles corresponding to the clusters memberships of Table 7.6 and 7.7 can be created by using Formula 5.4; similar as to how cluster centers were constructed.

7.4.4 Cluster evaluation

In Section 4.4.2, two approaches were presented to evaluate employee profiles.

The first approach was focused on the statistical quality of a clustering, measured by clustering-quality measures. This approach was used in the metaprocedure of the k-means algorithm. Here, a clustering is deemed of better quality than another clustering if the Relative Margin has a lower value. For the hierarchical algorithm, the corresponding CQM was not implemented.

The second approach is to evaluate clusters in the context of their end-use. In our case, we need the clusterings and clustering functions to be *interpretable* and *explainable*. The MCC clustering framework ensures that both of these needs are met. Because we categorize the types of competencies for each job, we no longer describe the jobs by 240 numbers, but by their profile. This ensures that the employee profiles are very interpretable and explainable, as is also illustrated by the discussion on the results in the previous sections. From the results of both algorithms, we could say the the MCC hierarchical clustering algorithms provides results that are "better" from an HR perspective, as the number of jobs in the clusters is more evenly distributed and thus contain more insights into the grouping of employees.

As we will see in the following section, the MCC clustering framework also ensures that relevant information is extracted from the employee profiles and is able to be used in the ASWP approach.

7.5 Employee profiles integration in ASWP approach

In Chapter 6.2 some methods for integrating employee profiles into the ASWP approach were discussed. Now that we have obtained employee profiles, we can see how this could be done for this case study. Table C.8 in Appendix C.2 shows the 35 jobs, their cluster membership for the MCC hierarchical algorithm for $k = 6$ and the probabilities of automation.

For cluster A, we actually see that both jobs actually have a high probability of automation: 71% and 79%. So even reskilling within the employee profile on the short term, but not be beneficial.

For cluster B, we see that two out of out of five jobs have a probability of automation of 73% and 0.71%; the other three have probabilities of automation of 8%, 16% and 35%. This indicates that maybe the company should create an up- and reskilling scheme within this employee profile, to make better use of its resources in the long term.

For cluster F, we see a wide variety of probabilities of automations, ranging from 1% to 96%. Since this cluster contains so many jobs, we can not assume that there is the possibility of reskilling for each of the jobs. An HR practitioner from the company has to look into the possibilities.

This analysis is very basic, but it does demonstrate the applicability of the employee profiles.

8 — Conclusions & Discussion

In this final chapter, we conclude by answering the research questions introduced in Section 1.2. Following the conclusions of this research, we present recommendations for organizations for applying the proposed methods in business. Next, this research's contribution to the HRM- and OR literature are presented. Finally, the limitations and avenues for further research are presented.

8.1 Conclusions

In Section 2.3, we found that the labor market is changing rapidly and it is important for companies to engage in Strategic Workforce Planning. Engaging in SWP is vital for an organization's health, reasons include achievement of business goals, financial benefits and improvement of the employee experience. A lack of SWP could affect performance and productivity of the workforce, increase operating costs and sustainability of the enterprise in the long run.

Strategic Workforce Planning provides actionable insights in the workforce development, in order to enable realisation of the organization's strategy. It does so by developing an aligned set of HRM policies and plans that ensure the right people, at the right time and at the right costs, are part of the organization.

We found that approaches to SWP are often based on different job titles within an organization. However, considering the changes in the labor market, especially regarding technological developments, these job titles may no longer be representative for long term strategic plans. This research extends the current approach to Analytical Strategic Workforce Planning, to account for these developments. In the following sections, the conclusions of this research are presented by answering the research questions; as a preparation for answering the main research question: *How can we extend the approach for Strategic Workforce Planning, by exploiting employees' competencies?*. The answers to those research questions can be divided into two categories. The first three questions can be considered as much needed background and context; the answers to these questions motivate and validate the proposed methods. The final three research questions are regarding the proposed methods for designing and integrating employee profiles into the ASWP approach and their results.

Validation & motivation

The first research question on the definitions and added value of Human Resource Management, HR Analytics and Strategic Workforce Planning was answered by means of a literature review. We presented a generic, well-used approach to the "Workforce Planning Process" (Phillips and Gully, [51]). Change management and implementation were found to be crucial elements for a sustainable engagement to SWP. Also, we found that, even though SWP is considered a part of HR analytics, generic approaches rarely rely on quantitative methods.

The second research question was answered by presenting the Analytical Strategic Workforce Planning (ASWP) approach, which is an application of the WPP approach but is much more tangible and concrete, with descriptive, predictive and prescriptive analytics profoundly included in the approach.

The ASWP approach belongs to the field of Operations Research, as it uses an optimization problem at its core. Additionally, we found a mismatch in the OR field regarding workforce planning. Mathematical methods were found to be simplifications of complex workforce situations, with limited consideration of actual implementation; whereas the judgemental methods provided descriptive explanations of complex workforce modes, but without the support of mathematical Operations Research. Furthermore, most of the OR mathematical models are not applied on a strategic level, but on a more operational or process level, such as staff scheduling.

Bech (2013, [6]) concluded that the mathematical algorithms proposed in the literature are not adequate in practice and investigated algorithms to find an optimal recruitment strategy in a more generic setting. They conclude the deductive algorithm is the best algorithm. This algorithm is still very much dependent on job titles, that may no longer be adequate descriptors of an organization's workforce, as we found in Section 1.1. To provide a more robust description of the workforce, we propose to describe employees based on their set of knowledge, skills and abilities. We do this by grouping employees into employee profiles based on their competencies. This allows for more sophisticated up- and reskilling schemes within an organization and increases the employees' internal mobility, resulting in a more efficient use of the available resources.

In preparation for designing methods for our proposed extension, we turned to the next research question: What are *competencies* and *competency data*, and do appropriate quantitative methods exists for grouping employees based on their competencies? An individual's competencies are defined as the combination of knowledge, skills and abilities, which are required in order for an individual to successfully perform a specific occupation; and that these individual characteristics that can be observed, measured, learned, acquired and enhanced. Competencies are used in competency modelling, where they are used to develop and align Human Resource Systems across an organization. Examples are: hiring new employees based on procedures that measure competencies, or by developing employee career paths. Next, we found that it is valid to describe and differentiate employees by their competencies, however this is not done in a quantitative Strategic Workforce Planning setting.

When looking at competency data, we found two features that are generally true and are not dependent on the specific database used. Firstly, competency data is multi-dimensional; because for an occupation, a type of competency is rated on the importance of that competency to the job, as well as the level of mastery of an individual in that job. Secondly, an organization's data experiences high dimensionality and low sample size. As we are describing employees by their competencies, it is inevitable that this will results in many features, as we can not describe an employee's competency by merely 6 features, for example. Low sample size is caused by the fact that the data is on job level and generally, companies do not have 1000 different jobs. When grouping employees based on competency data, these two features of the data should be taken into account. Regular clustering methods do not account for these features and we found that in the literature, there are no suitable methods that could be used directly.

Employee profiles design & integration

The fourth research question is answered by presenting the Multicriteria Competency clustering framework to design employee profiles, which deals with the specific features of the competency data. The framework is based on a framework by De Smet and Guzmán (2003, [19]). Five elements of the framework are altered to fit the competency framework: the aim, competency structure, profiles, distance metric and the construction of cluster centers. Using these elements, the Multicriteria Competency k-means clustering algorithm and MCC hierarchical clustering algorithm are presented.

Next, we presented methods for integrating employee profiles into the ASWP approach; thereby answering the fifth research question. In order to see where the current approach can be altered, we took a closer look at Bech's (2013, [6]) deductive algorithm for optimization of HR interventions. We propose to integrate the employee profiles in the transition methods that are used to calculate the expected future workforce in step 4. This is done by reflecting the insights from the employee profiles into the transition probability matrices used in the deductive algorithm. This also leads to the possibility to represent the new up- and reskilling programs within the company in the transition probability matrices.

Finally, the proposed methods for designing and integrating employee profiles were applied to a case study, allowing us to answer the final research question: Can we apply these new methods in practice by means of a *case study*? The case study's preliminary data insights illustrate and motivate the need for more complex cluster methods, compared to the exploratory PCA analysis. The Multicriteria Competency clustering framework is successfully applied to the company's employee- and employee competency data. Following, the Multicriteria Competency k-means clustering algorithm is performed and results are provided. We found that for the company that has 35 jobs, 6 was the best number of clusters. Five out of the six clusters were well-defined and contained insights on the jobs in the cluster. This could be seen through the distinctive features of the clusters employee profile, for example by an outstanding combination of types of competencies in certain categories. The sixth cluster, however, contained all other jobs, leading to a very all-round employee profile which did not embody a unique aspect of the employees in that profile. The results of the MCC hierarchical clustering algorithm showed a more even distribution or job within the clusters and provided some insights that were not available for the k-means algorithm. Generally, the framework was also well evaluated through the HR perspective. A very nice feature of the generic and explainable MCC clustering framework, is that the resulting employee profiles are very interpretable. From an employee profile, it can be seen what certain jobs have in common to ensure they are clustered together. This gives the framework a very intuitive character, which can be easily assessed and validated by a HR practitioner. Also, the employee profiles and their interpretable character ensure that the insights can also be used in other parts of HR decision making.

In conclusion, through the design and integration of employee profiles, we were able to exploit the employees' competencies in an approach for Strategic Workforce Planning.

8.2 Recommendations for business application

In Chapter 6, we discussed ways in which organizations can extract value out of employee profiles by integrating them in the Analytical Strategic Workforce Planning approach. However, the employee profiles can also be used in other HR decision making processes. Some possible applications are presented below.

- *Recruitment process:* employee profiles could be used during the recruitment process. Before, companies might really focus on specific (study or work) backgrounds of possible candidates when recruiting for specific vacancies. Now, insights on employee profiles give indications as to what type of competencies the organization wants and needs in employees, which allows for more robust and versatile recruitment strategies.
- *Hiring process:* employee profiles can be used during the hiring process of a candidate, for example to find out which job within the organization matches the candidate's competencies best. This could be done quantitatively, by actually gathering the competency data of the candidate and calculating the distances between the candidate's profile and the employee profiles. Or, in a more qualitative manner, where the HR practitioner assesses the candidate's competencies and tries to find the best matching employee profile.
- *Creating a new job:* as organizations grow, they might want to create new roles within the organization to provide room for growth. A (potentially) new job can be mapped onto the employee profiles to see which current employees might be able to take the new role upon them. This also provides employees room for growth within the company, increasing *employee retention.*

Before an organization is able to design and apply employee profiles to HR decision making, there are requirements that need to be met. Regarding Strategic Workforce Planning, organizations need to be willing to fully engage in the process and invest in change management, as they will have to make big, strategic decisions that change the organization and its workforce. Next to that, the most important requirement is that the organization's data management needs to be up to par. Without accurate and reliable information on the current state of the workforce, long term workforce planning is not pointless. Also competency data needs to be managed; larger companies often do already engage in competency modelling. However, collecting level- and importance scores on many different types of competencies, is not something that is generally done. As an alternative, an organization could map their jobs onto the jobs provided in the O*NET database, or any other preferred competency database. This does give the limitations that the data might not accurately represent the specific jobs within the organization. Still, it can lead to useful insights into the organization's workforce.

8.3 Contribution to literature

This research contributes to both Human Resources- and Operational Research literature in multiple ways, presented in this section.

Human Resources literature

Firstly, this research adds to the HR literature by *introducing* a more concrete, tangible and analytics focused approach to Strategic Workforce Planning; of which the steps were able to be mapped onto a more general approach often used in HR literature. Secondly, by *identifying* that the notion of competency is widely used within HR, but is not used as an integral part of the quantitative algorithms used for Strategic Workforce Planning. Thirdly, by *creating* a new, interpretable, explainable and generic clustering framework that can be used by HR practitioners to gain insights into an organization's workforce and its development based on employees' competency data. And lastly, *integrating* employee profiles into the ASWP approach, allows HR practitioners to have a better understanding of the workforce development.

Operations Research literature

Additionally, this research adds to the OR literature by *exploring* different clustering methods for multi-dimensional, high dimensionality and low sample size data. Secondly, by *creating* a new clustering framework to deal with multi-dimensional, high dimensionality and low sample size data; whilst maintaining a interpretable and explainable character. Within the framework, a centroid-based, as well as an hierarchical algorithm were presented. Thirdly, by *expanding* the field of mathematical SWP methods by introducing a new approach. In this new approach we proposed methods for integrating the framework's results into the ASWP approach. Lastly, this research *continues* the work of De Smet en Guzmán ([19]). The originality in their approach comes from the definition of a distance metric that takes into account the multicriteria nature of the problem. In their closing remarks, they propose that: "Future studies will be based on the application of the multicriteria distance introduced in this framework to other clustering or classification algorithms" and "...the application of the presented method to other real problems will allow us to confirm its coherence and to further analyze its distinctive features.". The contribution and originality of this research can be found in both of these aspects. Firstly, we explored the application of the modified framework also to a hierarchical clustering algorithm, whereas De Smet and Guzmán only applied it only to a k-means algorithm. Secondly, we applied the modified framework to a different, real world problem. In doing so, we found that the method indeed is able to take the multi-criteria feature of the data into account. Also, the method's distinctive features (construction of some type of structure, distance metric and the construction of cluster centers) were proven to enhance the interpretability, explainability and applicability of the results.

8.4 Limitations & Avenues for Future Research

(Analytical) Strategic Workforce Planning approach

A limitation of this research is that we used competency data on a job level as a proxy for individual competency data. Competency data on a job level is a simplification of the workforce, because of course, each employee has their own unique set of competencies. Because of this, the approaches and algorithms in this research considers all employees in one job to be equal, with the same costs, behaviour and competency scores. Despite the fact that organizations themselves also do not have this data on an individual level, this is still a limitation to the framework. A possible avenue for future research is to extend the approach to also include individual competencies. Actually, this would not change much to the MCC clustering framework. The difference would be that we have more data points, as each employee will be represented by one data point. This would actually work in favor of the framework, as we found that the low sample feature of the data caused negative consequences for the quality of a clustering. However, the integration of the employee profiles into the ASWP approach would become different. We considered Bech's deductive algorithm at the macroscopic level, where all employees in one job level are considered as one group. Bech also proposed a deductive algorithm at a *microscopic level*, where each employee is treated as an individual. In that stochastic approach, personal costs and career paths are forecasted by throwing a dice. The disadvantage of this microscopic level is that a lot of information on individuals is required, which also leads to a higher complexity of the algorithm; as well as a higher complexity for HR practitioners. Also, since this is a stochastic approach, each simulation run will find a different optimal recruitment strategy, which allows for risk analysis.

An avenue for future research, which would improve the overall methodology, regards the improvement of the third step of the Analytical Strategic Workforce Planning approach: the design of the desired workforce. This topic is not within the scope of this research, but, elements of this research could be used.

MCC clustering framework

The MCC clustering framework also experiences some limitations. Firstly, similar to the original framework by De Smet and Guzmán (2003, [19]), the framework is based on aggregated data and thus may loose some information. Secondly, the k-means algorithm suffers from the same bottlenecks as the original k-means algorithm. We still have that it is depends on the choice of k and that choosing k remains to be somewhat arbitrary; also the non-uniqueness of a solution is a limitation to the algorithm. Future research could dedicate to incorporating more sophisticated k-means algorithms. Also, future research could look at implementing the Weakes Link cluster-quality measure for the evaluation of the clusters created by the hierarchical algorithm.

Future research could commit to more complex clustering algorithms in general. For example, the notion of "fuzzy clustering" could be researched. This is a form of clustering where a data point can belong to more than one cluster, to a certain degree. This might be beneficial from an application perspective, especially when there is individual competency data, as it might better apprehend the complexity of employees' individual competencies. It might allow for more customization of the application for an organization. Another possible extension is to include weights in the distance metric. In the current distance metric, no distinction is made between the categories; either the type of competency is in the same cluster, or it is not. It does not take into account whether the type of competency is in a similar of totally different category. Also, for example more weight could be put onto the higher categories, as they are more descriptive to a job or employee.

Case study

The case study is performed using competency data from the O*NET database and employee data from a dummy dataset; which introduces limitations to this research. Firstly, the ASWP approach and MCC clustering framework might not work for all types of organizations. In our case study, the organization has employees with very different backgrounds, ensuring different types of employee profiles. However, when an organization is more specified to one sector, this might not be the case. As an example, take a marketing agency. Then, many employees within the organization will have similar scores on certain types of competencies, like the knowledge of "Sales and Marketing", the skill of " Social Perceptiveness" and the ability of "fluency of ideas". This increase in similarities between employees has the possibility of distorting the clustering algorithms. Secondly, the framework and its results are very dependent on the quality of the competency data. This is relevant, when using the O*NET, as well as when using an organization's own data. The data is collected through interviews with employees and despite efforts to mitigate biases, this is still something to look out for. More general limitations (and strengths) of the O*NET database are described in Handel (2016, [33]).

References

1. Ackerman, M. & Ben-David, S. Measures of Clustering Quality: A Working Set of Axioms for Clustering. *Advances in neural information processing systems* **21.** `https://proceedings.neurips.cc/paper/2008/file/beed13602b9b0e6ecb5b568ff5058f07-Paper.pdf` (2008).
2. Aggarwal, C. C., Hinneburg, A. & Keim, D. A. *On the surprising behavior of distance metrics in high dimensional space* in *International conference on database theory* (2001), 420–434.
3. Agiu, D., Mateescu, V. & Muntean, I. Business intelligence overview. *Database Systems Journal* **5,** 23–36 (2014).
4. Aloise, D., Deshpande, A., Hansen, P. & Popat, P. NP-hardness of Euclidean sum-of-squares clustering. *Machine learning* **75,** 245–248 (2009).
5. Bassi, L. J., Carpenter, R. & McMurrer, D. *HR analytics handbook* (Reed Business, 2010).
6. Bech, C. *New practical strategic workforce planning algorithms and their theoretical foundations.* MA thesis (VU University Amsterdam, July 2013).
7. Bersin, J., Collins, L., Mallon, D., Moir, J. & Straub, R. People analytics: Gaining speed. *Global Human Capital Trends–the new organization: Different by design,* 87–95 (2016).
8. Bersin, J. *The geeks arrive in HR: people analytics is here.* Feb. 2015. `https://www.forbes.com/sites/joshbersin/2015/02/01/geeks-arrive-in-hr-people-analytics-is-here/?sh=7f5f187673b4`.
9. Boudreau, J. W. & Jesuthasan, R. *Transformative HR: how great companies use evidence-based change for sustainable advantage* (John Wiley & Sons, 2011).
10. Buyens, D. & De Vos, A. Perceptions of the value of the HR function. *Human resource management journal* **11,** 70–89 (2001).
11. Campion, M. A. *et al.* Doing competencies well: Best practices in competency modeling. *Personnel psychology* **64,** 225–262 (2011).
12. Chauhan, A., Sharma, S. K. & Tyagi, T. Role of HRIS in improving modern HR operations. *Review of Management* **1,** 58–70 (2011).
13. Cheremond, G. R. *HR leaders take note: social developments, digital business, consumer behaviors, emerging technologies and more will change how people will work in 2028.* Aug. 2019. `https://www.gartner.com/smarterwithgartner/6-ways-the-workplace-will-change-in-the-next-10-years`.
14. Croft, L. & Seemiller, C. Developing Leadership Competencies. *New directions for student leadership* **156,** 7–18 (2017).
15. Davenport, T. & Harris, J. *Competing on analytics: The new science of winning. Boston: Harvard Business School Press, Boston Massachusetts. A short description is retrieved at March 20, 2016* 2007.
16. Davenport, T. H., Harris, J. & Shapiro, J. Competing on talent analytics. *Harvard business review* **88,** 52–58 (2010).

17. De Bruecker, P., Van den Bergh, J., Beliën, J. & Demeulemeester, E. Workforce planning incorporating skills: State of the art. *European Journal of Operational Research* **243,** 1–16 (2015).

18. De Feyter, T. Modeling mixed push and pull promotion flows in manpower planning. *Annals of Operations Research* **155,** 25–39 (2007).

19. De Smet, Y. & Guzmán, L. M. Towards multicriteria clustering: An extension of the k-means algorithm. *European Journal of Operational Research* **158,** 390–398 (2004).

20. Dolnicar, S., Grün, B. & Leisch, F. Increasing sample size compensates for data problems in segmentation studies. *Journal of Business Research* **69,** 992–999 (2016).

21. Economist, T. *Why companies are so bad at hiring.* May 2019. `https://www.economist.com/business/2019/05/09/why-companies-are-so-bad-at-hiring`.

22. Emmerichs, R. M., Marcum, C. Y. & Robbert, A. A. *An operational process for workforce planning* (Rand Corporation, 2004).

23. Executive, H. R. *How to develop a winning strategy in the war for talent.* Jan. 2022. `https://hrexecutive.com/how-to-develop-a-winning-strategy-in-the-war-for-talent`.

24. Ferligoj, A. & Batagelj, V. Direct multicriteria clustering algorithms. *Journal of classification* **9,** 43–61 (1992).

25. Fitz-Enz, J. *The new HR analytics* (American Management Association, 2010).

26. Fleisher, M. S. & Tsacoumis, S. 'O*NET analyst occupational skills ratings: Procedures update. *Raleigh, NC: National Center for O*NET Development* (2012).

27. Frey, C. B. & Osborne, M. A. The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change* **114,** 254–280 (2017).

28. Gardner, S. D., Lepak, D. P. & Bartol, K. M. Virtual HR: The impact of information technology on the human resource professional. *Journal of vocational behavior* **63,** 159–179 (2003).

29. Gigliotti, R. A. An Introduction to Competencies and Competency-Based Leadership. *Competencies for Effective Leadership* (2019).

30. Gonsalvez, C. J., Deane, F. P., Terry, J. & Hoong Gooi, C. Anatomy of psychology practitioner competence: Clustering approach reveals the hierarchical organization of competencies. *Clinical Psychology: Science and Practice* **27,** e12347 (2020).

31. Guenole, N., Feinzig, S., Ferrar, J. & Allden, J. Starting the workforce analytics journey (2015).

32. Hailey, V. H., Farndale, E. & Truss, C. The HR department's role in organisational performance. *Human resource management journal* **15,** 49–66 (2005).

33. Handel, M. J. The O* NET content model: strengths and limitations. *Journal for Labour Market Research* **49,** 157–176 (2016).

34. Isson, J. P. & Harriott, J. S. *People analytics in the era of big data: Changing the way you attract, acquire, develop, and retain talent* (John Wiley & Sons, 2016).

35. Jain, A. K. Data clustering: 50 years beyond K-means. *Pattern recognition letters* **31,** 651–666 (2010).

36. Jain, A. K. & Dubes, R. C. *Algorithms for clustering data* (Prentice-Hall, Inc., 1988).

37. Kanungo, T. *et al.* An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence* **24,** 881–892 (2002).

38. Kavanagh, M. J. & Johnson, R. D. *Human resource information systems: Basics, applications, and future directions* (Sage Publications, 2017).

39. Kelly, F. J. *A War For Talent Is Starting—Spoiler Alert: Workers Will Win* Apr. 2021. `https://www.gartner.com/smarterwithgartner/6-ways-the-workplace-will-change-in-the-next-10-years`.

40. Kleinberg, J. An impossibility theorem for clustering. *Advances in neural information processing systems* **15** (2002).

41. Kryscynski, D., Reeves, C., Stice-Lusvardi, R., Ulrich, M. & Russell, G. Analytical abilities and the performance of HR professionals. *Human Resource Management* **57,** 715–738 (2018).

42. Kunc, M. Achieving a balanced organizational structure in professional services firms: some lessons from a modeling project. *System Dynamics Review: The Journal of the System Dynamics Society* **24,** 119–143 (2008).

43. Lloyd, S. Least squares quantization in PCM. *IEEE transactions on information theory* **28,** 129–137 (1982).

44. Margherita, A. Human resources analytics: A systematization of research topics and directions for future research. *Human Resource Management Review,* 100795 (2021).

45. Marler, J. H. & Boudreau, J. W. An evidence-based review of HR Analytics. *The International Journal of Human Resource Management* **28,** 3–26 (2017).

46. McClelland, D. C. Testing for competence rather than for" intelligence." *American psychologist* **28,** 1 (1973).

47. Mehlmann, A. An approach to optimal recruitment and transition strategies for manpower systems using dynamic programming. *Journal of the Operational Research Society* **31,** 1009–1015 (1980).

48. Mishra, S. N., Lama, D. R., Pal, Y., *et al.* Human Resource Predictive Analytics (HRPA) for HR management in organizations. *International Journal of Scientific & Technology Research* **5,** 33–35 (2016).

49. Murtagh, F. A survey of recent advances in hierarchical clustering algorithms. *The computer journal* **26,** 354–359 (1983).

50. Nilakantan, K. & Raghavendra, B. Control aspects in proportionality Markov manpower systems. *Applied Mathematical Modelling* **29,** 85–116 (2005).

51. Phillips, J. & Gully, S. M. *Strategic staffing* 3rd ed. (Pearson, Boston, 2015).

52. PwC. *De verschuiving van vaardigheden: een analyse van de evolutie van onze competenties.* 2019. `https://www.pwc.nl/nl/themas/de-toekomst-van-werk/de-verschuiving-van-vaardigheden.html`.

53. PwC. *How to make vulnerable jobs ready for the future: Realigning people between industries or regions.* 2021. `https://www.pwc.nl/en/insights-and-publications/services-and-industries/people-and-organisation/from-zombie-to-angel-job-through-re-learning-and-re-alignment.html`.

54. PwC. *Preparing for tomorrow's workforce, today: insights from a global survey of business and HR leaders.* 2018. `https://www.pwc.com/gx/en/people-organisation/pdf/pwc-preparing-for-tomorrows-workforce-today.pdf`.

55. PwC. *PwC's global workforce strategy diagnostic* `https://www.pwc.com/gx/en/services/people-organisation/workforce-strategy-diagnostic.html`.

56. Raykov, Y. P., Boukouvalas, A., Baig, F. & Little, M. A. What to do when K-means clustering fails: a simple yet principled alternative algorithm. *PloS one* **11,** e0162259 (2016).

57. Satopää, V., Albrecht, J., Irwin, D. & Raghavan, B. *Finding a" kneedle" in a haystack: Detecting knee points in system behavior* in *2011 31st international conference on distributed computing systems workshops* (2011), 166–171.

58. Schippmann, J. S. *et al.* The practice of competency modeling. *Personnel psychology* **53,** 703–740 (2000).

59. Shalev-Shwartz, S. & Ben-David, S. *Understanding machine learning: From theory to algorithms* (Cambridge university press, 2014).

60. Taylor, S. *People resourcing* (CIPD Publishing, 2005).

61. Tursunbayeva, A., Di Lauro, S. & Pagliari, C. People analytics—A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management* **43,** 224–247 (2018).

62. U.S. Department of Labor, E. & Administration, T. *O*NET database* 2019. `https://www.onetcenter.org/`.

63. U.S. Department of Labor, E. & Administration, T. *O*NET Taxonomy* 2019. `https://www.onetcenter.org/taxonomy.html`.

64. Ulrich, D. & Dulebohn, J. H. Are we there yet? What's next for HR? *Human Resource Management Review* **25,** 188–204 (2015).

65. Ulrich, D. *Human resource champions: The next agenda for adding value and delivering results* (Harvard Business Press, 1996).

66. Van den Heuvel, S. & Bondarouk, T. The rise (and fall?) of HR analytics: A study into the future application, value, structure, and system support. *Journal of Organizational Effectiveness: People and Performance* (2017).

67. Von Luxburg, U., Williamson, R. C. & Guyon, I. *Clustering: Science or art?* in *Proceedings of ICML workshop on unsupervised and transfer learning* (2012), 65–79.

68. Wasid, M. & Ali, R. An improved recommender system based on multi-criteria clustering approach. *Procedia Computer Science* **131,** 93–101 (2018).

69. Willis, G., Cave, S. & Kunc, M. Strategic workforce planning in healthcare: A multi-methodology approach. *European Journal of Operational Research* **267,** 250–263 (2018).

70. Zanakis, S. H. & Maret, M. W. A Markov chain application to manpower supply planning. *Journal of the Operational Research Society* **31,** 1095–1102 (1980).

71. Zemke, R. Job competencies: can they help you design better training? *Training* (1982).

A — Human Resources Management & Strategic Workforce Planning

This appendix presents additional information on HRM and SWP. More specifically, Section A.1 refers to Section 2.2 and Section A.2 refers to Section 3.2.

A.1 HR analytics

Table A.1: Definitions of HR analytics adapted from cited literature in chronological order (Margherita, 2021 [44])

Adapted definition	**Source**
Extensive use of data, statistical and quantitative analysis, *explanatory and predictive models*, and fact-based management to drive decisions and actions involving personnel	Davenport & Harris, 2007 ([15])
A set of *six kinds of analytics* in terms of human-capital facts, analytical HR, human-capital investment analysis, workforce forecasts, talent value model, and talent supply chain	Davenport et al., 2010 ([16])
Logical analysis that uses objective business data as a basis for reasoning, discussion, or calculation	Fitz-enz, 2010 ([25])
Evidence-based approach to managing people and people processes within organizations	Bassi et al., 2010 ([5])
Evidence-based HR driving *strategic impact* based on logic-driven analytics, segmentation, risk leverage, synergy and integration and optimization	Boudreau & Jesuthasan, 2011 ([9])
Approaches for uncovering unique *insights* about people that enable faster and more accurate *decision-making* to executives	Guenole et al., 2015 ([31])
Rigorously *tracking HR investments and outcomes*	Ulrich & Dulebohn, 2015 ([64])
Multidisciplinary approach to integrate methodologies for improving the quality of people-related decisions	Mishra et al., 2016 ([48])
Bringing *together HR and business data* for analyzing people-related risks, performance characteristics, engagement and culture, and identifying career paths	Bersin et al., 2016 ([7])
A HR practice enabled by information technology that uses *descriptive, visual, and statistical analyses* of data related to HR capital and organizational performance to establish business impact and enable data-driven decision-making	Marler & Boudreau, 2017 ([45])
HR analytics is the *systematic identification and quantification* of the people drivers of business outcomes	van den Heuvel & Bondarouk, 2017 ([66])
Data, metrics, statistics and scientific methods, with the help of technology, to gauge the impact of human capital management practices on business goals	Kryscynski et al., 2017 ([41])

Table A.2: 106 key concepts associated to HR analytics, divided into 3 main categories and 6 sub-categories (Margherita, 2021 [44])

Concepts and sources related to enablers/factors in HR analytics	
technology enablers	organizational factors
1. Artificial Intelligence	19. Academic and practitioner integration
2. Chatbots	20. Agile workforce analytics
3. Cloud-based systems	21. Analytics function centralization
4. Data clustering tools	22. Analytics skills of HR professionals
5. Employee information systems	23. Analytics team creation
6. HR big data	24. Awareness of analytics opportunities
7. HR databases	25. Awareness of challenges and criticisms
8. HR information systems	26. Data governance and ethics
9. HR platforms	27. Degree of individual adoption
10. HR software and applications	28. Employees' perceived accuracy and fairness
11. HR statistic tools and algorithms	29. Ethics issues in HR data analysis and use
12. Internet of things devices and sensors	30. Focus on actionable insights
13. Job search engines	31. HRM team preparation and expertise
14. Machine learning applications	32. Knowledge and competence hubs
15. Multi-cue semantic information	33. Organization and industry implementation barriers
16. Natural language processing	34. Organizational complementarities
17. Neural fuzzy networks	35. Organizational readiness
18. Social media and professional networks	36. Outside-in approach with focus on actionable metrics
	37. People specialist team creation
	38. Performance pay policy
	39. Privacy issues in HR data analysis and use
	40. Six thinking hats approach
	41. Virtue ethics approach

Concepts and sources related to applications in HR analytics	
descriptive applications	predictive/prescriptive applications
42. Adaptive scoring algorithm	57. Dynamic talent flow analysis
43. Competence analytics	58. Expertise recommendation and allocation
44. Employee engagement	59. Prediction human resources modelling
45. Employee sentiment analysis	60. Predictive data profiling
46. Expertise estimation and competence assessment	61. Proactive predictive decision on people matters
47. HR information retrieval, fusion and completion	62. Probabilistic learning framework
48. Intelligence applicants shortlisting	63. Propensity modeling
49. Job scheduling	64. Sentiment analysis
50. Latent ability modelling	65. Turnover costs and recruitment decision
51. Occupational skills normalization	66. Voluntary turnover prediction
52. Online recruitment	67. Workforce forecasting modelling
53. Real-time data collection	68. Workplace attendance, accidents, injuries tracking
54. Semantic web human resource résumés	
55. Skill assessment, identification and normlization	
56. Talent hiring, engagement and retention	

Concepts and sources related to values in HR analytics	
employee-related value	organizational value
69. Appropriate recruitment profile selection	89. Automated decision-making
70. Employee attrition and loyalty analysis	90. Automated management style
71. Employee attrition prediction	91. Business and organizational performance
72. Employee churn prediction	92. Business value creation and business model innovation
73. Employee engagement and commitment	93. Competitive advantage and enterprise analytics
74. Employee fraud risk management	94. Customer satisfaction
75. Employee performance evaluation and rewards	95. Data-driven decision making
76. Employee profiling	96. Data-oriented leadership
77. Employee reskilling and competence update	97. Evidence-based predictive decision making
78. Employee sentiment analysis	98. Managerial efficiency
79. Forecasting of HR capacity and recruitment needs	99. Organizational agility
80. Global recruitment optimization	100. Organizational effectiveness
81. HR external and internal marketing	101. Organizational resilience
82. Improved employee experience	102. People-driven competitive advantage
83. Job turnover and transition networks	103. Strategic change
84. Leadership development	104. Strategic execution of organizational plans
85. Real-time workforce performance awareness	105. Support to agile project management
86. Skill-job fit, customized training/pay and loyalty	106. Support to organizational change management
87. Sustainable talent acquisition	
88. Wage transparency	

A.2 SWP in the field of Operations Research

Author(s)	Year	Keywords
Nilakantan and Raghavendra [50]	2005	A hierarchical organization under the influence of "proportional" policies.
De Feyter [18]	2007	Mixed push and pull method and its application on the maintainability and attainability of an organization.
Mehlmann [47]	1980	The influence of inflow and transition behavior on the size and structure of an organization using dynamic programming
Zanakis and Maret [70]	1980	Markov chain application

Table A.3: Literature referred to by Bech (2013, [6])

B — Competency Management & Clustering

This appendix contains additional information on Chapter 4 and 5.

B.1 Introduction to Competency Management

Table B.1: Summary of definitions of “competencies” from noted scholars, federal agencies and subject matter experts is presented, based on Schippmann et al., [58] and Gigliotti, 2019 [29]:

Definition	**Source**
A mixture of knowledge, skills, abilities, motivation, beliefs, values, and interests	(Fleishman, Wetrogan, Uhlman, & Marshall-Mies, 1995)
A combination of motives, traits, self-concepts, attitudes or values, content knowledge or cognitive behavior skills; any individual characteristic that can be reliably measured or counted and that can be shown to differentiate superior from average performers	(Spencer, McClelland, & Spencer, 1994)
An underlying characteristic of an individual that is causally related to effective or superior performance in a job	Boyatzis, 1982)
A measurable pattern of knowledge, skills, abilities, behaviors, and other characteristics that an individual needs in order to successfully perform work roles or occupational functions. Competencies specify the “how” of performing job tasks, or what the person needs to do the job successfully	(U.S. Office of Personnel Management, 2018).
The skills, knowledge, and behaviors that lead to successful performance	(U.K. Civil Service, 2018)
A knowledge, skill, ability, or characteristics associated with high performance on a job	Mirabile, 1997
A written description of measurable work habits and personal skills used to achieve work objectives	Green, 1999
Observable, behavioral capabilities that are important for performing key responsibilities of a role or job	Subject Expert Matter Schippmann et al., 2000.
Knowledge, skills, abilities, traits or motives directly resembling or related to the job or job performance or some other important life outcome	McClelland, 1973

B.2 Competency data

All types of competencies

Table B.2, B.3 and B.4 present all types of knowledge, skills and abilities and their structures as defined by the O*NET Content Model.

Knowledge

DOMAIN				
Business and Management	**Engineering and Technology**	**Mathematics and Science**	**Arts and Humanities**	**Transportation**
Administration and Management Administrative Economics and Accounting Sales and Marketing Customer and Personal Service Personnel and Human Resources	Computers and Electronics Engineering and Technology Design Building and Construction Mechanical	Mathematics Physics Chemistry Biology Psychology Sociology and Anthropology Geography	English Language Foreign Language Fine Arts History and Archaeology Philosophy and Theology	Transportation
Communications	**Law and Public Safety**	**Manufacturing and Production**	**Health Services**	**Education and Training**
Telecommunications Communications and Media	Public Safety and Security Law and Government	Production and Processing Food Production	Medicine and Dentistry Therapy and Counseling	Education and Training

Table B.2: Types of *knowledge* per category

Skills

BASIC		**CROSS-FUNCTIONAL**				
Content	**Process**	**Social**	**Complex Problem Solving**	**Technical**	**Systems**	**Resource Management**
Reading Comprehension	Critical Thinking	Social Perceptiveness	Complex Problem Solving	Operations Analysis	Judgment and Decision Making	Time Management
Active Listening	Active Learning	Coordination		Technology Design	Systems Analysis	Management of Financial Resources
Writing	Learning Strategies	Persuasion		Equipment Selection	Systems Evaluation	Management of Material Resources
Speaking	Monitoring	Negotiation		Installation		Management of Personnel Resources
Mathematics		Instructing		Programming		
Science		Service Orientation		Operations Monitoring		
				Operation and Control		
				Equipment Maintenance		
				Troubleshooting		
				Repairing		
				Quality Control Analysis		

Table B.3: Types of *skills* per (sub-)categories

Abilities

COGNITIVE						
Verbal	**Idea Generation and Reasoning**	**Quantitative**	**Memory**	**Perceptual**	**Spatial**	**Attentiveness**
Oral Comprehension	Fluency of Ideas	Mathematical Reasoning	Memorization	Speed of Closure	Spatial Orientation	Selective Attention
Written Comprehension	Originality	Number Facility		Flexibility of Closure	Visualization	Time Sharing
Oral Expression	Problem Sensitivity			Perceptual Speed		
Written Expression	Deductive Reasoning					
	Inductive Reasoning					
	Information Ordering					
	Category Flexibility					

PSYCHOMOTOR			PHYSICAL			SENSORY	
Fine Manipulative	**Control Movement**	**Reaction Time and Speed**	**Endurance**	**Physical Strength**	**Flexibility, Balance and Coordination**	**Auditory and Speech**	**Visual**
Arm-Hand Steadiness	Control Precision	Reaction Time	Stamina	Static Strength	Extent Flexibility	Hearing Sensitivity	Near Vision
Manual Dexterity	Multilimb Coordination	Wrist-Finger Speed		Explosive Strength	Dynamic Flexibility	Auditory Attention	Far Vision
Finger Dexterity	Response Orientation	Speed of Limb Movement		Dynamic Strength	Gross Body Coordination	Sound Localization	Visual Color Discrimination
	Rate Control			Trunk Strength	Gross Body Equilibrium	Speech Recognition	Night Vision
						Speech Clarity	Peripheral Vision
						Depth Perception	
						Glare Sensitivity	

Table B.4: Types of *abilities* per (sub-)categories

Structure of O*NET databases

The knowledge, skills and abilities O*NET database have the structure and information as presented in Table B.5.

Column	Type	Column Content
O*NET-SOC Code	Character(10)	O*NET-SOC Code
Title	Character Varying(150)	O*NET-SOC Title
Element ID	Character Varying(20)	Content Model Outline Position
Element Name	Character Varying(150)	Content Model Element Name
Scale ID	Character Varying(3)	Scale ID
Scale Name	Character Varying(50)	Scale Name
Data Value	Float(5,2)	Rating associated with the O*NET-SOC occupation
N	Integer(4)	Sample size
Standard Error	Float(5,2)	Standard Error
Lower CI Bound	Float(5,2)	Lower 95% confidence interval bound
Upper CI Bound	Float(5,2)	Upper 95% confidence interval bound
Recommend Suppress	Character(1)	Low precision indicator (Y=yes, N=no)
Not Relevant	Character(1)	Not relevant for the occupation (Y=yes, N=no)
Date	Character(7)	Date when data was updated
Domain Source	Character Varying(30)	Source of the data

Table B.5: O*NET database: structure and description

B.3 Clustering methods

Wasid and Ali's algorithm for multi-criteria k-means

Wasid and Ali (2018, [68]) incorporate multi-criteria ratings in a k-means algorithm by altering the distance formula; their algorithm can be found in Figure B.1.

Input: Dataset Multi-criteria ratings, number of clusters K
Output: Set of user clusters

Step 1: [*Initialization*]
Initialize randomly K clustering centers $\{c_1, c_2, c_3, .., c_k\}$ from the dataset
Step 2: [*Assignment*]
a. Find the closest cluster center c for each user a using the following Euclidean distance formula.

$$Edis_{c,a} = \sqrt{(M_{c,1} - M_{a,1})^2 + (M_{c,2} - M_{a,2})^2 + ... + (M_{c,n} - M_{a,n})^2},$$

where: $Edis_{c,a}$ – distance between the cluster center c and the user a
n – Number of user rated items
$M_{c,n}$ - Multi-criteria ratings of cluster center c for item n
$M_{a,n}$ - Multi-criteria ratings of user a for item n

b. Arrange user a by distance from each clustering centers $\{c_1, c_2, c_3, .., c_k\}$
c. Assign the user a with the nearest distance cluster center c
Step 3: [*Convergence*]
a. Iteratively process cluster reassignment until convergence criteria is reached.
b. if convergence criteria met, then the algorithm terminates and returns a set of user clusters $\{c_1, c_2, c_3, .., c_k\}$, otherwise go to step 2.

Figure B.1: Wasid and Ali's algorithm for multi-criteria k-means ([68])

We translate Wasid's Euclidean distance formula, such that it matches our problem:

$$Edis_{c,j} = \sqrt{(S_{c,1} - S_{j,1})^2 + (S_{c,2} - S_{j,2})^2 + \cdots + (S_{c,i} - S_{j,i})^2 + \cdots + (S_{c,n} - S_{j,n})^2} \tag{B.1}$$

where,

$$\begin{aligned}
J = \{J_1, J_2, \cdots, J_j, \cdots J_n\} &\text{ - Set of jobs} \\
C = \{C_1, C_2, \cdots, C_c, \cdots C_k\} &\text{ - Clusters} \\
X = \{x_1, x_2, \cdots, x_i, \cdots, x_m\} &\text{ - Set of types of competencies} \\
Edis_{c,j} &\text{ - distance between the cluster center } C_c \text{ and job } J_j, \\
& \quad\; c = 1, \cdots, k;\ j = 1, \cdots, n \\
S_{c,i} = \begin{pmatrix} S_{c,i,level} \\ S_{c,i,importance} \end{pmatrix} &\text{ - Multi-criteria scores of cluster center } c \text{ for type of competency } i \\
S_{j,i} = \begin{pmatrix} S_{j,i,level} \\ S_{j,i,importance} \end{pmatrix} &\text{ - Multi-criteria scores of job } j \text{ for type of competency } i
\end{aligned}$$

Writing out one of the terms of Formula B.1 gives:

$$\begin{aligned}
(S_{c,i} - S_{j,i})^2 &= \left(\begin{bmatrix} S_{c,i,level} \\ S_{c,i,importance} \end{bmatrix} - \begin{bmatrix} S_{j,i,level} \\ S_{j,i,importance} \end{bmatrix} \right)^2 \\
&= \left(\begin{bmatrix} S_{c,i,level} - S_{j,i,level} \\ S_{c,i,importance} - S_{j,i,importance} \end{bmatrix} \right)^2 \\
&= [S_{c,i,level} - S_{j,i,level}, S_{c,i,importance} - S_{j,i,importance}] \cdot \begin{bmatrix} S_{c,i,level} - S_{j,i,level} \\ S_{c,i,importance} - S_{j,i,importance} \end{bmatrix} \\
&= (S_{c,i,level} - S_{j,i,level})^2 + (S_{c,i,importance} - S_{j,i,importance})^2
\end{aligned}$$

Then, using this equality in all terms of Formula B.1:

$$
\begin{aligned}
Edis_{c,j} = \Big[& (S_{c,1,level} - S_{j,1,level})^2 + (S_{c,1,importance} - S_{j,1,importance})^2 \\
& + (S_{c,2,level} - S_{j,2,level})^2 + (S_{c,2,importance} - S_{j,2,importance})^2 \\
& \vdots \\
& + (S_{c,i,level} - S_{j,i,level})^2 + (S_{c,i,importance} - S_{j,i,importance})^2 \\
& \vdots \\
& + (S_{c,n,level} - S_{j,n,level})^2 + (S_{c,n,importance} - S_{j,n,importance})^2 \Big]^{\frac{1}{2}}
\end{aligned}
\tag{B.2}
$$

We find that Wasid's distance measure in fact is the regular Euclidean distance formula. The multi-criteria issue is solved by simply putting all scores across all types of competencies in one vector, treating them as independent variables.

B.4 The MCC clustering framework

Proof of distance metric 5.3

The proof below shows that Formula 5.3 is indeed an appropriate distance metric. This proof is an extension of proof of Theorem 2 by De Smet and Guzmán (2003, [19], p.397).

Notations:

- $A = \{a_1, \cdots, a_n\}$
- $ab = \{a_1 \cap b_1, a_2 \cap b_2, \cdots, a_8 \cap b_8\}$
- $|a_i|$ = cardinality of a_i

Theorem 1. The application

$$d : P \otimes P \to \{\frac{h}{n} | h \in \{1, \cdots, n\}\} : (P(a), P(b)) \to d(a, b) = 1 - \frac{\sum_{i=1}^{8} |(ab)_i|}{n}$$

is a distance.

Proof

- $d(a, b) = d(b, a)$ by definition
- $d(a, b) \geq 0$

  $d(a, b) \geq 0 \iff \sum_i |(ab)_i| \leq n$
  $\sum_i |(ab)_i| \leq \sum_i |(a)_i| = n$
- $d(a, b) = 0 \iff a = b$

  $\Leftarrow$ If $a = b$
  It follows from the definition of d that $d(a, b) = 0$
  $\Rightarrow$ If $d(a, b) = 0$

then $\sum_i |(ab)_i| = n$

As $|(ab)_i| \leq |a_i|$ and $|(ab)_i| \leq |b_i|$, it follows that $\sum_i |(ab)_i| = n = \sum_i |a_i| = \sum_i |b_i|$

And so $(ab)_i = a_i = b_i, \forall i$

- $d(a,c) + d(c,b) \geq d(a,b)$ (triangle inequality)

  Then we have $1 - \frac{\sum_i |(ac)_i|}{n} + 1 - \frac{\sum_i |(cb)_i|}{n} \geq 1 - \frac{\sum_i |(ab)_i|}{n}$

  Rewriting gives $1 + \frac{\sum_i |(ab)_i|}{n} \geq \frac{\sum_i |(ac)_i|}{n} + \frac{\sum_i |(cb)_i|}{n}$

  If $c = b$ or $c = a$ the bound is reached. Let us take $c = b$ and swap k elements in the profile of c. We will always have $\sum_i |(bc)_i| = n - k$

  Then $1 + \frac{\sum_i |(ab)_i|}{n} \geq \frac{\sum_i |(ac)_i|}{n} + \frac{n-k}{n}$

  rewriting leads to $\sum_i |(ac)_i| \leq \sum_i |(ab)_i| + k$

$\square$

C — Case Study on Designing & Integrating Employee Profiles

This appendix contains additional information on the case study performed in Chapter 7: descriptive information on the employee- and competency data in the first section and results from the MCC clustering framework in the second section.

C.1 Employee- and competency data

SOC-code	Occupation	2017 (#)	2018 (#)	Irrelevant competencies (#)	Recommend Suppress (%)	Probability of automation
11-3012.00	Administrative Services Managers	20	21	11	0,42	0.73
11-2011.00	Advertising and Promotions Managers	2	2	31	1,67	0.04
17-3021.00	Aerospace Engineering and Operations Technologists and Technicians	1	1	18	4,17	0.48
17-2011.00	Aerospace Engineers	6	6	35	3,33	0.02
53-2021.00	Air Traffic Controllers	2	2	25	0,42	0.11
49-3011.00	Aircraft Mechanics and Service Technicians	13	12	9	2,92	0.71
51-2011.00	Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	3	4	8	2,92	0.79
53-2022.00	Airfield Operations Specialists	1	1	9	1,67	0.71
53-2011.00	Airline Pilots, Copilots, and Flight Engineers	89	93	14	3,33	0.18
13-1199.00	Business Operations Specialists, All Other	80	78	25	0,00	0.23
11-1011.00	Chief Executives	17	17	30	2,92	0.02
53-2012.00	Commercial Pilots	40	42	10	1,25	0.55
11-3111.00	Compensation and Benefits Managers	1	1	40	27,50	0.96
13-1141.00	Compensation, Benefits, and Job Analysis Specialists	7	6	28	27,50	0.47
13-1041.00	Compliance Officers	23	24	13	4,58	0.08
11-3021.00	Computer and Information Systems Managers	30	26	28	4,58	0.04
11-3031.00	Financial Managers	45	44	43	4,58	0.07
53-2031.00	Flight Attendants	306	308	12	1,25	0.35
11-1021.00	General and Operations Managers	174	184	9	1,25	0.16
11-3121.00	Human Resources Managers	11	11	31	27,50	0.01
13-1071.00	Human Resources Specialists	45	46	42	3,75	0.31
11-3051.00	Industrial Production Managers	13	16	7	0,83	0.03
23-1011.00	Lawyers	50	51	30	1,67	0.04
13-1081.00	Logisticians	13	14	28	27,50	0.01
13-1111.00	Management Analysts	54	57	33	27,50	0.13
13-1161.00	Market Research Analysts and Marketing Specialists	48	48	37	27,50	0.61
11-2021.00	Marketing Managers	18	18	34	0,83	0.01
15-2031.00	Operations Research Analysts	8	8	34	27,50	0.04
11-2031.00	Public Relations and Fundraising Managers	5	6	34	2,50	0.02
11-3061.00	Purchasing Managers	6	4	32	27,92	0.03
11-2022.00	Sales Managers	29	28	24	27,50	0.01
11-3131.00	Training and Development Managers	3	3	33	27,50	0.01
13-1151.00	Training and Development Specialists	23	22	36	27,92	0.01
11-3071.00	Transportation, Storage, and Distribution Managers	9	9	26	0,42	0.59
41-3041.00	Travel Agents	5	3	34	4,17	0.10

Table C.1: All 35 unique occupations within the company: employee- and competency data

Table C.2: Competency data: relevancy of types of competencies and the mean, median, minimum and maximum value for both level- and importance scores.

Element ID	Element Name	Jobs	Level (range: 0 to 7)				Importance (range: 1 to 5)			
		#	mean	median	min	max	mean	median	min	max
Knowledge										
2.C.1.a	Administration and Mgmt.	35	4,2	4,41	2,6	6,23	3,68	3,83	2,54	4,75
2.C.1.b	Administrative	35	3,68	3,92	1,21	5,27	2,87	2,94	1,69	3,94
2.C.1.c	Economics and Accounting	34	2,61	2,94	0,6	4,48	2,62	2,72	1,3	3,96
2.C.1.d	Sales and Marketing	32	2,98	2,94	0,28	6,13	2,65	2,38	1,13	4,85
2.C.1.e	Customer and Personal Service	35	4,65	4,85	2,4	5,86	3,79	3,88	2,2	4,51
2.C.1.f	Personnel and Human Res.	35	3,58	3,74	1,01	6,17	3,05	3,04	1,57	4,92
2.C.10	Transportation	31	2,32	2,1	0,38	6,04	2,46	2,2	1,14	4,78
2.C.2.a	Production and Processing	32	2,48	2,41	0,61	5,36	2,41	2,34	1,31	4,41
2.C.2.b	Food Production	5	0,4	0,18	0	1,98	1,21	1,09	1	2,18
2.C.3.a	Computers and Electronics	35	3,84	3,86	2,54	6,13	3,17	3,21	2,23	4,77
2.C.3.b	Engineering and Technology	32	2,23	2,08	0,14	6,1	2,31	2,13	1,05	4,76
2.C.3.c	Design	29	2,03	2,03	0,2	5,41	2,1	2,03	1,17	4,13
2.C.3.d	Building and Construction	15	1	0,75	0	2,62	1,54	1,37	1	2,51
2.C.3.e	Mechanical	24	1,81	1,11	0,05	6,85	2,03	1,64	1,02	4,71
2.C.4.a	Mathematics_knowledge	35	3,88	3,74	1,92	6,25	3,23	3,26	2,06	4,7
2.C.4.b	Physics	21	1,31	1,14	0	4,84	1,76	1,58	1	4,03
2.C.4.c	Chemistry	18	1,11	0,96	0	3,64	1,55	1,46	1	2,83
2.C.4.d	Biology	10	0,61	0,56	0,01	1,95	1,3	1,28	1,01	2,18
2.C.4.e	Psychology	34	2,74	2,74	0,71	4,38	2,54	2,52	1,32	3,75
2.C.4.f	Sociology and Anthropology	30	2	1,81	0,2	4,58	2,04	1,91	1,19	3,31
2.C.4.g	Geography	31	2,09	1,81	0,39	4,61	2,1	1,81	1,11	3,96
2.C.5.a	Medicine and Dentistry	17	0,65	0,55	0	2,22	1,4	1,35	1	2,82
2.C.5.b	Therapy and Counseling	27	1,38	1,25	0,04	3,71	1,7	1,68	1,02	2,88
2.C.6	Education and Training	35	4,1	4,02	2,66	6,92	3,13	3,03	2,33	4,97
2.C.7.a	English Language	35	4,27	4,17	2,98	5,69	3,98	3,97	3,28	4,69
2.C.7.b	Foreign Language	27	1,09	1,08	0,27	3,04	1,59	1,53	1,12	2,96
2.C.7.c	Fine Arts	7	0,5	0,38	0	2,3	1,27	1,19	1	2,39
2.C.7.d	History and Archaeology	18	0,77	0,6	0	1,82	1,38	1,38	1	1,79
2.C.7.e	Philosophy and Theology	22	1,3	1,25	0,21	2,67	1,56	1,54	1,08	2,12
2.C.8.a	Public Safety and Security	35	2,62	2,56	1,09	4,56	2,68	2,5	1,61	4,48
2.C.8.b	Law and Government	35	2,86	2,84	1,64	5,46	2,86	2,86	1,72	4,72
2.C.9.a	Telecommunications	35	1,81	1,65	0,43	3,46	2,31	2,24	1,33	3,65
2.C.9.b	Communications and Media	35	2,86	2,68	0,99	5,21	2,71	2,54	1,78	4,35
Skills										
2.A.1.a	Reading Comprehension	35	4,08	4	3,12	5	3,87	3,88	3	4,38
2.A.1.b	Active Listening	35	4	4	3,12	4,88	3,95	4	3,12	4,62
2.A.1.c	Writing	35	3,79	3,88	2,88	4,88	3,56	3,75	2,75	4
2.A.1.d	Speaking	35	3,96	4	3	4,88	3,92	3,92	3	4,62
2.A.1.e	Mathematics_skills	35	2,99	3	1,88	4,75	2,85	2,88	2	4,5
2.A.1.f	Science	28	1,35	1	0	5,12	1,86	1,75	1	4
2.A.2.a	Critical Thinking	35	4,06	4	3,12	4,88	3,87	3,88	3,12	4,38
2.A.2.b	Active Learning	35	3,71	3,75	3	4,75	3,47	3,5	2,88	4
2.A.2.c	Learning Strategies	35	3,37	3,25	2,25	5,25	3,11	3	2,5	4,38
2.A.2.d	Monitoring	35	3,93	4	2,88	5,5	3,61	3,75	2,88	4,12
2.B.1.a	Social Perceptiveness	35	3,59	3,75	2,25	4,25	3,52	3,62	2,75	4,25
2.B.1.b	Coordination	35	3,64	3,75	2,88	5,12	3,51	3,62	2,88	4,25
2.B.1.c	Persuasion	35	3,45	3,5	2,25	5	3,24	3,12	2,25	4,12
2.B.1.d	Negotiation	35	3,23	3,25	2,12	4,62	3,11	3,08	2,25	4,12
2.B.1.e	Instructing	35	3,36	3,25	2,62	4,75	3,11	3,12	2,62	4,25
2.B.1.f	Service Orientation	35	3,25	3,25	2,38	4	3,18	3,12	2,38	4,12
2.B.2.i	Complex Problem Solving	35	3,66	3,75	2,62	5	3,59	3,62	2,88	4,38
2.B.3.a	Operations Analysis	34	2,72	2,88	0,25	5	2,63	2,75	1,25	3,88
2.B.3.b	Technology Design	35	1	0,79	0,38	3,75	1,8	1,75	1,38	3,25
2.B.3.c	Equipment Selection	5	0,44	0,12	0	3	1,28	1,12	1	3,25
2.B.3.d	Installation	3	0,2	0	0	3	1,13	1	1	2,88
2.B.3.e	Programming	35	1	0,75	0,38	2,75	1,72	1,62	1,38	2,75
2.B.3.g	Operations Monitoring	35	1,96	1,75	0,5	4,88	2,35	2	1,5	4,62
2.B.3.h	Operation and Control	18	1,24	0,62	0	5,62	1,8	1,38	1	4,88

Continued on next page

Table C.2 – continued from previous page

Element ID	Element Name	Jobs #	Level (range: 0 to 7) mean	median	min	max	Importance (range: 1 to 5) mean	median	min	max
2.B.3.j	Equipment Maintenance	7	0,44	0	0	4,75	1,29	1	1	4,25
2.B.3.k	Troubleshooting	18	0,99	0,38	0	4	1,66	1,38	1	4,12
2.B.3.l	Repairing	5	0,39	0	0	4,25	1,27	1	1	4,25
2.B.3.m	Quality Control Analysis	33	1,87	1,88	0,25	3,88	2,2	2,12	1,12	3,88
2.B.4.e	Judgment and Decision Making	35	3,83	3,88	2,88	5,75	3,68	3,75	3	4,5
2.B.4.g	Systems Analysis	35	3,42	3,46	1,88	5,38	3,24	3,25	2	4,12
2.B.4.h	Systems Evaluation	35	3,46	3,5	1,88	5,12	3,17	3,12	2	4,12
2.B.5.a	Time Management	35	3,55	3,75	2,88	4,75	3,4	3,5	3	4
2.B.5.b	Management of Financial Res.	35	2,42	2,25	0,62	5,5	2,46	2,25	1,5	4,12
2.B.5.c	Management of Material Res.	35	2,19	1,88	0,75	4,75	2,36	2,12	1,75	3,88
2.B.5.d	Management of Personnel Res.	35	3,33	3,25	1,62	5,38	3,1	3	1,88	4,25
Abilities										
1.A.1.a.1	Oral Comprehension	35	4,31	4,25	3,75	5	4,01	4	3,12	4,62
1.A.1.a.2	Written Comprehension	35	4,12	4	3	5,12	3,96	4	3	4,5
1.A.1.a.3	Oral Expression	35	4,24	4,12	3,12	5	4,04	4	3	4,88
1.A.1.a.4	Written Expression	35	3,96	4	2,75	5	3,72	3,88	2,88	4,12
1.A.1.b.1	Fluency of Ideas	35	3,58	3,63	2,38	4,62	3,36	3,38	2,75	4
1.A.1.b.2	Originality	35	3,46	3,5	2,12	4,25	3,25	3,25	2,38	3,88
1.A.1.b.3	Problem Sensitivity	35	3,98	4	2,88	5	3,88	3,88	3,5	4,75
1.A.1.b.4	Deductive Reasoning	35	4,12	4,12	3,25	5	3,85	3,88	3,12	4,38
1.A.1.b.5	Inductive Reasoning	35	3,85	3,88	3	5	3,72	3,88	3	4,25
1.A.1.b.6	Information Ordering	35	3,68	3,75	3	4,12	3,59	3,62	3	4
1.A.1.b.7	Category Flexibility	35	3,48	3,5	3	4,12	3,24	3,25	3	3,75
1.A.1.c.1	Mathematical Reasoning	35	3,11	3,12	1,88	4,75	2,96	3	2	4,75
1.A.1.c.2	Number Facility	35	3,11	3,12	2	4,5	2,89	3	2,12	4,25
1.A.1.d.1	Memorization	35	2,67	2,75	2	3,38	2,62	2,62	2	3,12
1.A.1.e.1	Speed of Closure	35	2,8	2,75	2	4,88	2,68	2,62	2	4,25
1.A.1.e.2	Flexibility of Closure	35	3,03	3	2,25	4,5	3,01	2,88	2,5	4,25
1.A.1.e.3	Perceptual Speed	35	2,82	2,75	2	4,38	2,85	2,75	2	4,12
1.A.1.f.1	Spatial Orientation	12	0,63	0,12	0	4,62	1,42	1,12	1	4
1.A.1.f.2	Visualization	35	2,87	2,75	1,75	4,5	2,72	2,75	2	3,5
1.A.1.g.1	Selective Attention	35	3,02	2,88	2,38	4,88	3,06	3	2,62	4,5
1.A.1.g.2	Time Sharing	35	2,73	2,62	2	5,62	2,79	2,75	2,12	4,12
1.A.2.a.1	Arm-Hand Steadiness	23	1,15	0,5	0	3,75	1,83	1,5	1	4
1.A.2.a.2	Manual Dexterity	25	1,08	0,5	0	4,12	1,81	1,5	1	3,88
1.A.2.a.3	Finger Dexterity	35	1,98	1,88	1	4,12	2,25	2,12	1,5	4
1.A.2.b.1	Control Precision	23	1,26	0,75	0	4,5	1,87	1,62	1	4,62
1.A.2.b.2	Multilimb Coordination	15	0,92	0,12	0	4,25	1,63	1,12	1	4
1.A.2.b.3	Response Orientation	13	0,77	0	0	5,62	1,51	1	1	4,88
1.A.2.b.4	Rate Control	12	0,67	0	0	5	1,46	1	1	4,12
1.A.2.c.1	Reaction Time	12	0,85	0	0	5	1,52	1	1	4,25
1.A.2.c.2	Wrist-Finger Speed	22	0,77	0,5	0	2,75	1,53	1,38	1	2,75
1.A.2.c.3	Speed of Limb Movement	9	0,38	0	0	2,75	1,24	1	1	2,75
1.A.3.a.1	Static Strength	11	0,63	0	0	3	1,4	1	1	2,88
1.A.3.a.2	Explosive Strength	10	0,23	0	0	1,12	1,18	1	1	1,88
1.A.3.a.3	Dynamic Strength	16	0,53	0,38	0	2,12	1,41	1,25	1	2,75
1.A.3.a.4	Trunk Strength	32	1,33	1,12	0	3	1,85	1,75	1	3
1.A.3.b.1	Stamina	10	0,45	0	0	2,62	1,31	1	1	2,75
1.A.3.c.1	Extent Flexibility	11	0,66	0	0	3,5	1,42	1	1	3,12
1.A.3.c.2	Dynamic Flexibility	1	0,02	0	0	0,38	1,02	1	1	1,38
1.A.3.c.3	Gross Body Coordination	11	0,46	0	0	2,75	1,31	1	1	2,88
1.A.3.c.4	Gross Body Equilibrium	11	0,47	0	0	3,12	1,31	1	1	3
1.A.4.a.1	Near Vision	35	3,82	3,75	3,25	4,62	3,65	3,62	3,12	4,25
1.A.4.a.2	Far Vision	35	3,07	2,88	1,88	5,38	2,9	2,75	2,12	4,25
1.A.4.a.3	Visual Color Discrimination	35	2,18	2	0,75	3,88	2,32	2,08	1,75	3,62
1.A.4.a.4	Night Vision	9	0,52	0	0	3,88	1,33	1	1	3,38
1.A.4.a.5	Peripheral Vision	10	0,49	0	0	4	1,33	1	1	3,88
1.A.4.a.6	Depth Perception	31	1,39	1	0	4,38	1,97	1,75	1	4,12
1.A.4.a.7	Glare Sensitivity	11	0,6	0	0	4,25	1,38	1	1	3,62
1.A.4.b.1	Hearing Sensitivity	35	1,8	1,62	0,5	4	2,14	1,88	1,5	3,75
1.A.4.b.2	Auditory Attention	35	2,14	1,88	0,62	4	2,33	2,12	1,62	3,5
1.A.4.b.3	Sound Localization	10	0,45	0	0	2,88	1,29	1	1	2,88

Continued on next page

Table C.2 – continued from previous page

Element ID	Element Name	Jobs	Level (range: 0 to 7)				Importance (range: 1 to 5)			
		#	mean	median	min	max	mean	median	min	max
1.A.4.b.4	Speech Recognition	35	3,72	3,75	3	4,62	3,66	3,75	3,12	4,12
1.A.4.b.5	Speech Clarity	35	3,74	3,62	2,88	5	3,78	3,88	3	4,25

C.2 Results MCC clustering algorithms

Employee Profile for cluster A:
General and Operations Managers; Compliance Officers; Administrative Services Managers; Industrial Production Managers; Airfield Operations Specialists.
(Relative Margin = $0,48574$)

Category	Types of competencies
Crucial	—
Essential	Administration and Management; Customer and Personal Service; Oral Comprehension; Written Comprehension; Oral Expression; Written Expression; Reading Comprehension; Active Listening; Speaking; Monitoring; Coordination
Significant	English Language; Problem Sensitivity; Deductive Reasoning; Inductive Reasoning; Near Vision; Speech Recognition; Speech Clarity; Critical Thinking; Social Perceptiveness; Judgment and Decision Making
Useful	Computers and Electronics
Favorable	Administrative; Personnel and Human Resources; Mechanical; Mathematics (knowledge); Psychology; Education and Training; Public Safety and Security; Law and Government; Fluency of Ideas; Originality; Information Ordering; Category Flexibility; Mathematical Reasoning; Number Facility; Memorization; Speed of Closure; Flexibility of Closure; Perceptual Speed; Visualization; Selective Attention; Time Sharing; Far Vision; Visual Color Discrimination; Auditory Attention; Writing; Active Learning; Learning Strategies; Persuasion; Negotiation; Instructing; Service Orientation; Complex Problem Solving; Operations Analysis; Operations Monitoring; Systems Analysis; Systems Evaluation; Time Management; Management of Financial Resources; Management of Personnel Resources
Applicable	Telecommunications
Unimportant	Economics and Accounting; Sales and Marketing; Production and Processing; Engineering and Technology; Design; Building and Construction; Physics; Chemistry; Biology; Sociology and Anthropology; Geography; Medicine and Dentistry; Therapy and Counseling; Foreign Language; History and Archaeology; Philosophy and Theology; Communications and Media; Transportation; Spatial Orientation; Arm-Hand Steadiness; Manual Dexterity; Finger Dexterity; Control Precision; Multilimb Coordination; Response Orientation; Rate Control; Reaction Time; Wrist-Finger Speed; Speed of Limb Movement; Static Strength; Explosive Strength; Dynamic Strength; Trunk Strength; Stamina; Extent Flexibility; Gross Body Coordination; Gross Body Equilibrium; Night Vision; Peripheral Vision; Depth Perception; Glare Sensitivity; Hearing Sensitivity; Sound Localization; Mathematics (skills); Science; Technology Design; Programming; Operation and Control; Troubleshooting; Quality Control Analysis; Management of Material Resources
Irrelevant	Food Production; Fine Arts; Dynamic Flexibility; Equipment Selection; Installation; Equipment Maintenance; Repairing

Table C.3: Employee profile of cluster center A

Employee Profile for cluster B:
Flight Attendants.
(Relative Margin = $0,48574$)

Category	Types of competencies
Crucial	—
Essential	Customer and Personal Service; Public Safety and Security; Oral Comprehension; Oral Expression; Speech Clarity
Significant	English Language; Problem Sensitivity; Deductive Reasoning; Near Vision; Speech Recognition; Active Listening; Speaking; Monitoring; Social Perceptiveness; Coordination; Service Orientation
Useful	—
Favorable	Administration and Management; Sales and Marketing; Personnel and Human Resources; Computers and Electronics; Psychology; Sociology and Anthropology; Geography; Education and Training; Foreign Language; Law and Government; Communications and Media; Transportation; Written Comprehension; Written Expression; Inductive Reasoning; Information Ordering; Category Flexibility; Speed of Closure; Flexibility of Closure; Perceptual Speed; Selective Attention; Time Sharing; Arm-Hand Steadiness; Manual Dexterity; Finger Dexterity; Multilimb Coordination; Static Strength; Trunk Strength; Stamina; Extent Flexibility; Gross Body Coordination; Gross Body Equilibrium; Far Vision; Visual Color Discrimination; Auditory Attention; Reading Comprehension; Writing; Critical Thinking; Active Learning; Persuasion; Negotiation; Instructing; Complex Problem Solving; Judgment and Decision Making; Time Management
Applicable	Medicine and Dentistry; Therapy and Counseling; Fluency of Ideas; Originality; Hearing Sensitivity; Learning Strategies; Management of Personnel Resources
Unimportant	Administrative; Economics and Accounting; Production and Processing; Food Production; Engineering and Technology; Design; Mechanical; Mathematics (knowledge); Physics; Chemistry; Biology; Philosophy and Theology; Telecommunications; Mathematical Reasoning; Number Facility; Memorization; Visualization; Control Precision; Response Orientation; Rate Control; Reaction Time; Wrist-Finger Speed; Speed of Limb Movement; Explosive Strength; Dynamic Strength; Depth Perception; Glare Sensitivity; Mathematics (skills); Science; Operations Analysis; Technology Design; Programming; Operations Monitoring; Operation and Control; Troubleshooting; Quality Control Analysis; Systems Analysis; Systems Evaluation; Management of Financial Resources; Management of Material Resources
Irrelevant	Building and Construction; Fine Arts; History and Archaeology; Spatial Orientation; Dynamic Flexibility; Night Vision; Peripheral Vision; Sound Localization; Equipment Selection; Installation; Equipment Maintenance; Repairing

Table C.4: Employee profile of cluster center B

Employee Profile for cluster D:
Airline Pilots, Copilots, and Flight Engineers; Commercial Pilots.
(Relative Margin $= 0,48574$)

Category	Types of competencies
Crucial	Transportation; Response Orientation; Operation and Control
Essential	Customer and Personal Service; Geography; English Language; Oral Comprehension; Written Comprehension; Oral Expression; Problem Sensitivity; Deductive Reasoning; Inductive Reasoning; Information Ordering; Perceptual Speed; Spatial Orientation; Selective Attention; Control Precision; Multilimb Coordination; Rate Control; Reaction Time; Near Vision; Far Vision; Depth Perception; Hearing Sensitivity; Speech Recognition; Reading Comprehension; Critical Thinking; Active Learning; Monitoring; Operations Monitoring
Significant	Public Safety and Security; Flexibility of Closure; Time Sharing; Manual Dexterity; Visual Color Discrimination; Speech Clarity; Active Listening; Speaking; Coordination; Complex Problem Solving; Judgment and Decision Making
Useful	Visualization; Instructing
Favorable	Administration and Management; Personnel and Human Resources; Computers and Electronics; Mechanical; Mathematics (knowledge); Physics; Psychology; Education and Training; Law and Government; Written Expression; Fluency of Ideas; Originality; Category Flexibility; Mathematical Reasoning; Number Facility; Memorization; Speed of Closure; Arm-Hand Steadiness; Finger Dexterity; Wrist-Finger Speed; Extent Flexibility; Night Vision; Peripheral Vision; Glare Sensitivity; Auditory Attention; Writing; Mathematics (skills); Science; Learning Strategies; Social Perceptiveness; Persuasion; Negotiation; Service Orientation; Troubleshooting; Quality Control Analysis; Systems Analysis; Systems Evaluation; Time Management; Management of Personnel Resources
Applicable	Engineering and Technology
Unimportant	Administrative; Economics and Accounting; Production and Processing; Design; Chemistry; Sociology and Anthropology; Medicine and Dentistry; Therapy and Counseling; Foreign Language; Telecommunications; Communications and Media; Speed of Limb Movement; Static Strength; Explosive Strength; Dynamic Strength; Trunk Strength; Stamina; Gross Body Coordination; Gross Body Equilibrium; Sound Localization; Operations Analysis; Technology Design; Programming; Equipment Maintenance; Management of Financial Resources; Management of Material Resources
Irrelevant	Sales and Marketing; Food Production; Building and Construction; Biology; Fine Arts; History and Archaeology; Philosophy and Theology; Dynamic Flexibility; Equipment Selection; Installation; Repairing

Table C.5: Employee profile of cluster center D

Employee Profile for cluster E:
Air Traffic Controllers .
(Relative Margin = $0,48574$)

Category	Types of competencies
Crucial	Time Sharing
Essential	Customer and Personal Service; Education and Training; English Language; Public Safety and Security; Transportation; Oral Comprehension; Oral Expression; Problem Sensitivity; Deductive Reasoning; Inductive Reasoning; Information Ordering; Category Flexibility; Speed of Closure; Flexibility of Closure; Perceptual Speed; Visualization; Selective Attention; Near Vision; Far Vision; Speech Recognition; Speech Clarity; Reading Comprehension; Active Listening; Critical Thinking; Monitoring; Coordination; Complex Problem Solving; Judgment and Decision Making
Significant	Written Comprehension; Auditory Attention; Speaking; Active Learning; Time Management
Useful	Geography
Favorable	Administration and Management; Computers and Electronics; Mathematics (knowledge); Psychology; Law and Government; Telecommunications; Written Expression; Fluency of Ideas; Originality; Mathematical Reasoning; Number Facility; Memorization; Finger Dexterity; Control Precision; Visual Color Discrimination; Depth Perception; Hearing Sensitivity; Writing; Mathematics (skills); Learning Strategies; Social Perceptiveness; Persuasion; Instructing; Service Orientation; Operations Analysis; Operations Monitoring; Quality Control Analysis; Systems Analysis; Systems Evaluation; Management of Personnel Resources
Applicable	Arm-Hand Steadiness; Negotiation; Operation and Control
Unimportant	Administrative; Economics and Accounting; Personnel and Human Resources; Production and Processing; Engineering and Technology; Design; Mechanical; Physics; Therapy and Counseling; Communications and Media; Spatial Orientation; Manual Dexterity; Multilimb Coordination; Response Orientation; Rate Control; Reaction Time; Trunk Strength; Night Vision; Peripheral Vision; Glare Sensitivity; Sound Localization; Science; Technology Design; Programming; Troubleshooting; Management of Financial Resources; Management of Material Resources
Irrelevant	Sales and Marketing; Food Production; Building and Construction; Chemistry; Biology; Sociology and Anthropology; Medicine and Dentistry; Foreign Language; Fine Arts; History and Archaeology; Philosophy and Theology; Wrist-Finger Speed; Speed of Limb Movement; Static Strength; Explosive Strength; Dynamic Strength; Stamina; Extent Flexibility; Dynamic Flexibility; Gross Body Coordination; Gross Body Equilibrium; Equipment Selection; Installation; Equipment Maintenance; Repairing

Table C.6: Employee profile of cluster center E

Employee Profile for cluster F:

Business Operations Specialists, All Other; Management Analysts; Lawyers; Market Research Analysts and Marketing Specialists; Financial Managers; Human Resources Specialists; Sales Managers; Computer and Information Systems Managers; Training and Development Specialists; Marketing Managers; Chief Executives; Logisticians; Human Resources Managers; Transportation, Storage, and Distribution Managers; Operations Research Analysts; Compensation, Benefits, and Job Analysis Specialists; Purchasing Managers; Travel Agents; Public Relations and Fundraising Managers; Aerospace Engineers; Training and Development Managers; Advertising and Promotions Managers; Compensation and Benefits Managers.

(Relative Margin $= 0,48574$)

Category	**Types of competencies**
Crucial	—
Essential	Administration and Management; Customer and Personal Service; Mathematics (knowledge); English Language; Oral Comprehension; Written Comprehension; Oral Expression; Written Expression; Fluency of Ideas; Problem Sensitivity; Deductive Reasoning; Inductive Reasoning; Reading Comprehension; Active Listening; Writing; Speaking; Critical Thinking; Monitoring; Social Perceptiveness; Judgment and Decision Making
Significant	Information Ordering; Near Vision; Speech Recognition; Speech Clarity; Complex Problem Solving
Useful	Administrative; Education and Training
Favorable	Economics and Accounting; Personnel and Human Resources; Computers and Electronics; Law and Government; Communications and Media; Originality; Category Flexibility; Mathematical Reasoning; Number Facility; Memorization; Speed of Closure; Flexibility of Closure; Perceptual Speed; Visualization; Selective Attention; Time Sharing; Far Vision; Mathematics (skills); Active Learning; Learning Strategies; Coordination; Persuasion; Negotiation; Instructing; Service Orientation; Operations Analysis; Systems Analysis; Systems Evaluation; Time Management; Management of Financial Resources; Management of Personnel Resources
Applicable	—
Unimportant	Sales and Marketing; Production and Processing; Engineering and Technology; Design; Psychology; Sociology and Anthropology; Geography; Therapy and Counseling; Foreign Language; History and Archaeology; Philosophy and Theology; Public Safety and Security; Telecommunications; Transportation; Manual Dexterity; Finger Dexterity; Wrist-Finger Speed; Trunk Strength; Visual Color Discrimination; Depth Perception; Hearing Sensitivity; Auditory Attention; Science; Technology Design; Programming; Operations Monitoring; Quality Control Analysis; Management of Material Resources
Irrelevant	Food Production; Building and Construction; Mechanical; Physics; Chemistry; Biology; Medicine and Dentistry; Fine Arts; Spatial Orientation; Arm-Hand Steadiness; Control Precision; Multilimb Coordination; Response Orientation; Rate Control; Reaction Time; Speed of Limb Movement; Static Strength; Explosive Strength; Dynamic Strength; Stamina; Extent Flexibility; Dynamic Flexibility; Gross Body Coordination; Gross Body Equilibrium; Night Vision; Peripheral Vision; Glare Sensitivity; Sound Localization; Equipment Selection; Installation; Operation and Control; Equipment Maintenance; Troubleshooting; Repairing

Table C.7: Employee profile of cluster center F

Cluster membership and probabilities of automation for result MCC hierarchical clustering algorithm with $k = 6$

SOC-code	Occupation	Probability of automation	Cluster membership
11-3011.00	Administrative Services Managers	0.73	B
11-2011.00	Advertising and Promotions Managers	0.039	E
17-3021.00	Aerospace Engineering and Operations Technologists and Technicians	0.48	D
17-2011.00	Aerospace Engineers	0.02	E
53-2021.00	Air Traffic Controllers	0.11	C
49-3011.00	Aircraft Mechanics and Service Technicians	0.71	A
51-2011.00	Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	0.79	A
53-2022.00	Airfield Operations Specialists	0.71	B
53-2011.00	Airline Pilots, Copilots, and Flight Engineers	0.18	C
13-1199.00	Business Operations Specialists, All Other	0.23	F
11-1011.00	Chief Executives	0.02	E
53-2012.00	Commercial Pilots	0.55	C
11-3111.00	Compensation and Benefits Managers	0.96	F
13-1141.00	Compensation, Benefits, and Job Analysis Specialists	0.47	F
13-1041.00	Compliance Officers	0.08	B
11-3021.00	Computer and Information Systems Managers	0.03	D
11-3031.00	Financial Managers	0.07	F
53-2031.00	Flight Attendants	0.35	B
11-1021.00	General and Operations Managers	0.16	B
11-3121.00	Human Resources Managers	0.01	E
13-1071.00	Human Resources Specialists	0.31	F
11-3051.00	Industrial Production Managers	0.03	D
23-1011.00	Lawyers	0.03	E
13-1081.00	Logisticians	0.01	F
13-1111.00	Management Analysts	0.13	F
13-1161.00	Market Research Analysts and Marketing Specialists	0.61	F
11-2021.00	Marketing Managers	0.01	E
15-2031.00	Operations Research Analysts	0.04	E
11-2031.00	Public Relations and Fundraising Managers	0.02	E
11-3061.00	Purchasing Managers	0.03	F
11-2022.00	Sales Managers	0.01	E
11-3131.00	Training and Development Managers	0.01	E
13-1151.00	Training and Development Specialists	0.01	E
11-3071.00	Transportation, Storage, and Distribution Managers	0.59	F
41-3041.00	Travel Agents	0.10	F

Table C.8: Cluster membership and probabilities of automation