Haaga-Helia
University of Applied Sciences

Data mining and Data Analysis using KNIME: Job satisfaction of Foreigners in Finland.

Tsedenia Adnew

Bachelor's Thesis
Degree Programme in BITe…
2021

Haaga-Helia
University of Applied Sciences

Abstract

Date

Author(s)
Tsedenia Adnew

Degree programme
Business Information Technology

Report/thesis title
Data Mining and Data Analysis using KNIME: Job satisfaction of Foreigners in Finland**.**

Number of pages and appendix pages
58

The topic of job satisfaction is becoming a major concern for companies in recent years. The well-being of employee's not only physically but mentally as well is crucial for a company's better performance and public image. The rapid increase in globalization has affected Finland positively, as the country has a more diversified workforce, thus needing to adopt to more cultures and values.

This thesis aims to understand the general level of satisfactions foreigners in Finland have in their respective jobs. KNIME analytical platform has been used primarily to analyse the data. And the analysed data has been written in an easy-to-understand format for readers.

This thesis covers the definition of job satisfaction, importance and how we can use data mining and analysis to determine how foreign employees in Finland are coping with working there.

The results of the findings explain and answer the research´s main and sub research question. The research also provides the current rating of 3 factors of job satisfaction, physical and mental health, personal growth and cultural fit.

This research aims to be useful for future studies in providing insight of the current job satisfaction and provide basis for comparing different factors and establish corelation to improve the job satisfaction of the current and future workforce.

Keywords
Data Mining, Data Analysis, KNIME, Job Satisfaction, Machine learning

Haaga-Helia
University of Applied Sciences

Abstract

Date

Table of contents

Acknowledgement

I would like to express my gratitude to my thesis advisor Dr. Amir Dirin for suggesting the thesis topic and helping me through out. Guiding me with all the extra materials to get a better understanding of the aim of my thesis. And giving me the encouragement to explore the data mining and analysis world.

I would also like to thank all the participants that took part in the survey and helped me gather the data I needed. Last but not least I would like to thank God and my family for being there for me during my studies and supported me through out especially my sister in giving me encouragement to keep being focused.

Tsedenia Adnew
20 June 2021
Helsinki

1 Introduction

Job satisfaction is the combination of environmental, psychological and physiological events that will determine if a person can truthfully say they are satisfied with a job (Happock, 1935). Although written almost a century ago this aptly named book, is still effective in defining the meaning of job satisfaction because, pleasant or unpleasant circumstances an employee faces at their work will influence their view on their work.

Economic and technological changes all around have brought about a radical change in working nature. The perception of people with regards to their work has also changed, although we would need to determine if it is for the better or for worse. Good knowledge on how employees feel about their workplace is one of the most crucial things a company needs to determine.

As the business world is growing more and more each day, businesses are expanding and with it the demand of employees is at an all-time high to meet their goals for the day. A company or a business is as good as the service or product they give and, the high competition between industries and businesses demands them to have high performance in-order-to establish competitive advantage. In relation, having high performance for a company means having employees that perform well. Job satisfaction and performance is a hand and glove concept. One is very dependent on the other. Mathematically, they are directly related. In simpler terms the increase of one will increase the other and vise versa.

According to data provided by Statistics Finland there are a total of 423,494 people with non-Finnish origin (Statistics Finland 2021). That is approximately 8% of the entire population of Finland, which can not be considered slight. Of that population a little over 200,000 are of working age and are currently working. Having this many workers in the economy will significantly affect the way the economy performs with regards to productivity.

As mentioned in the above paragraph, employee satisfaction and performance are dependent on each other. Thus, companies should not focus only on their employee's performance but also their satisfaction.

This thesis aims to show what the level of the current foreigners job satisfaction is, and make a general summary of where the lowest satisfaction lies. Which industry is the most prevalent in dissatisfaction and why?

This research will heavily rely on Data Mining and Analysis to get the most accurate reading of the collected data. Nowadays, companies can gain a substantial amount of useful information by analysing a plethora of data and extract patterns or find a connection (SOJECT, 2020). Data mining is one of the core competencies of many companies these days be it large or small. There is tremendous important information to be gained by analysing data such as, make better decisions, find leads, predict future trends and understand their own company better.

KNIME Analytics Platform is a platform that has broad prediction abilities and as well as many analytical possibilities by taking huge amounts of data. By analysing the data that is collected using a questionnaire, I will use KNIME to group similar categories(cluster), find trends and why certain industries have an over all low rating satisfaction level.

In conclusion, I will be using machine learning and data mining and analysis with the help of KNIME to collect, analyze, implement and make general conclusion based on the data collected through a survey that has been sent out to about 120 people and the interviews of four people with varying professions. By utilizing KNIME I will be answering all the research questions and provide a step by step development of the project on the platform.

This thesis will give companies or readers an understanding of their current employee satisfaction status and adjust their values and processes accordingly so it can be all inclusive for a better performance.

2 Research Question and Research Methods

The main research question of this thesis is investigating how to analyse the Job satisfaction and performance of foreigner employees in finland through KNIME analytics Platform.

2.1 Sub-Questions

In-order-to grasp the full impact of this thesis' main question we need to go through the sub questions that will clarify it further:

1. How we can use KNIME to analyse the current level of job satisfaction of foreigners in Finland. The use of KNIME not only ensures the accuracy of the analysed data result but it makes for an easy understanding for the reader what the results will be.
2. What is the lowest score category and what the reason is?
3. What common factors affect the satisfaction level of employees?

Job satisfaction is one of the core concepts businesses and companies should give emphasis on. But surprisingly job satisfaction surveys are done once or twice a year with a long gap in between, with no way of measuring if the changes implemented have been helpful or not. Data mining along with KNIME can be helpful to answer these sub questions and better explain the data set, the process and the final result in a comprehensible manner so that readers can easily understand it.

3 Methods

In this chapter, we can get a clear picture of the methodology that is followed during writing the thesis which is mixed methods and workflow process using CRISP-DM.

3.1 Methodology

Data mining, data analytics and machine learning are implementations that are complicated, as a result the implementation process should be organized very well. This thesis is organized in to three iterations:

The first iteration will be collecting the data through interviews and questionnaire. The second iteration focuses on developing the project in KNIME Analytical Platform to create the concept. And the third iteration is the interpretation of the KNIME results along with their discussion and possible recommendation for companies.

3.2 Mixed Methods

This thesis uses mixed methodology to conduct its research. Mixed method is a methodology that involves the combination of both qualitative and quantitative method to answer research questions (Renwick, 2021). Separately quantitative and qualitative methods have their own strength and weaknesses but by combining them we can get a better and stronger outcome and validation with little to no weaknesses.

Combining different methods can be difficult if the data set is of different nature. This thesis has conducted both qualitative and quantitative method using a questionnaire and semi-structured interviews.

The data collected quantitatively is the questionnaire that has been answered by 102 people deciding their emotions regarding their current job. The qualitative data is collected through a semi-structured interview where the interviewees were asked to rate their satisfaction and other open ended questions that can be quantifiable when analyzed.

3.3 Workflow Process

As mentioned in previous chapters it is quite important to understand and organize the implementation process of a complicated subject such as data mining and analysis. As a result the meticulous planning of the whole process will provide a structured and easily decipherable project. This thesis will use CRISP-DM method to present the data science life cycle.

CRoss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (CRISP-DM) is a model that catagorizes the data science life cycle in to six stages (Vorhies, 2016). By breaking down stages in to six categories and basing each stage's around data mining using structured approach, this methodology has become essential in predictive analytics.

The structured and organized nature of this methodology allows data miners to stay focused on the goal. When there is huge amounts of data coming, often it is common to lose sight of the purpose of the analysis and the process will keep repeating itself or not have and end. The break down of these stages and iterative nature of CRISP-DM will allow the project to evaluate each stages and make changes accordingly. This will minimize risks of not reaching the ultimate goal.

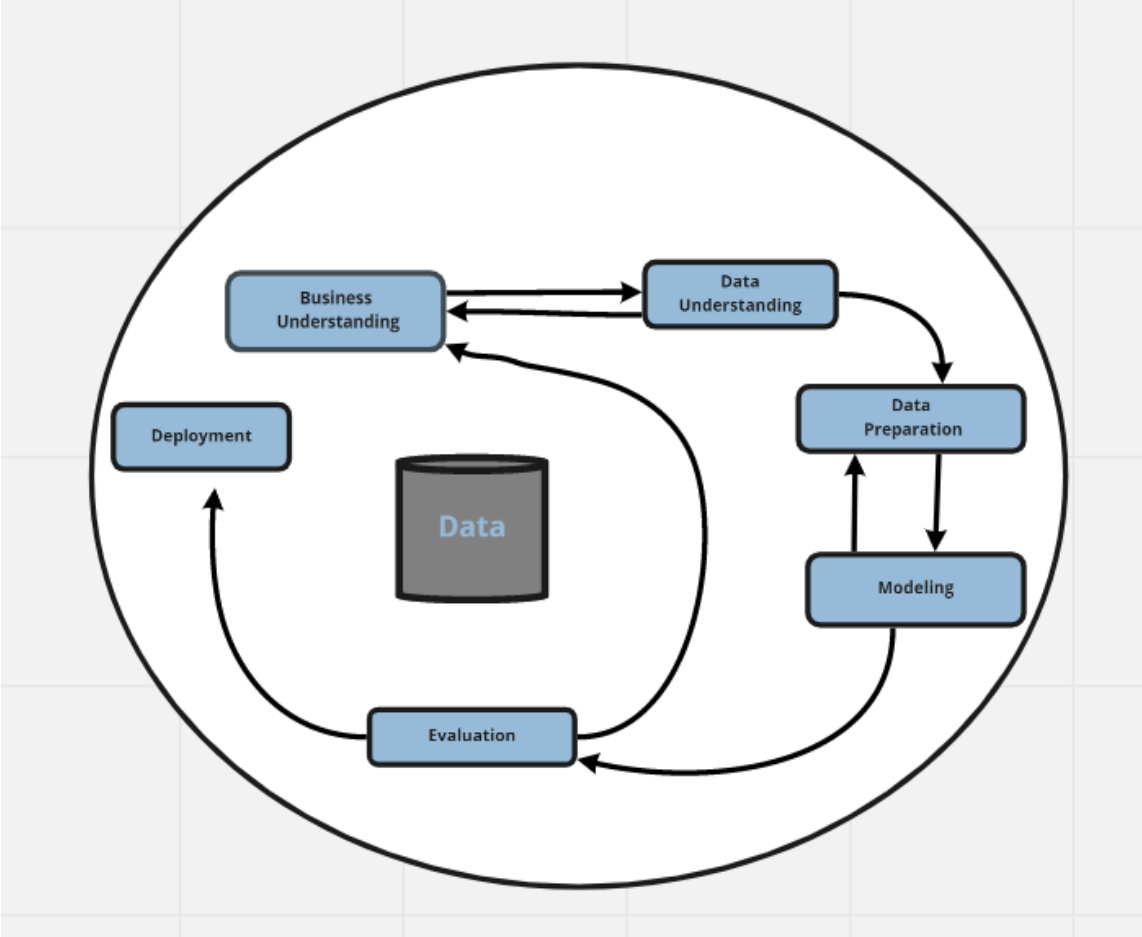


Figure1. CRISP-DM

A. Business Understanding

The business understanding phase is the most fundamental part, where we focus on understanding the objectives and requirements of the project. This phase is essential in that it could lay a strong foundation for the success of the project (Vorhies, 2016).

Business understanding determines the business objectives having this phase in first place will help steer the project the right way and determine the focus areas for the project. We can easily determine what the problem is that needs to be solved.

Each business and project are vastly different from each other, one size does not fit all, so tailoring the way the project will be implemented with the business in focus might reduce risk of failure.

B. Data Understanding

After determining the focus of the business, we add more foundation to the understanding by identifying, collecting, discover insights to the data and analysing the data set that will assist in achieving the project goal. While collecting data, we will only have raw data, if not understood accurately it can lead to misinterpretations that could possibly jeopardize the project or business. This phase has four tasks:

- **Collect initial data**: understanding the source of the data and determining the relevance of the data.
- **Describe data**: examining the raw data collected and determining its properties.
- **Explore data:** taking a closer look in to the collected data to explore the relationships amongst the data.
- **Verify data quality**: determining the quality and relevance of the data.

Understanding the data set we will be working with is a crucial step in solving the initial business problem we are facing. It is imperative we determine the importance of the data we have collected and how accurate it is regarding the business problem because it is an important tool we will be using to come up with resources and methods to start the project. In addition, the inability to understand the data set can mean that we would need to stop the project to re-evaluate the scope and problem of the project.

C. Data Preparation

In most projects data preparation constitutes about 80% of the project (Data Science Project Management, 2019). Commonly known as data munging, will organize the final data that will be used in the modelling phase. In Data preparation phase we will convert the raw data we collected to fit the final dataset we will be preparing. This phase has five tasks:

- **Select data**: filtering out data that is relevant enough to be included or excluded.
- **Clean data**: avoiding garbage-in, garbage-out by removing erraneous values that can affect the dataset.
- **Construct data**: exploring all the collected data and determining new attributes that will be advantageous for the next phase.
- **Integrate data:** by combining data with common traits we create new data set.
- **Format data:** re-formatting data that will fit the project or business.

D. Modelling

By taking into consideration the data set we have and the tools we have used we will build and assess different models that bet suit the business or project. Designing the best model for the project will increase the accuracy of the prediction and provide a solid model that can be relied on heavily and that is not susceptible to risk of failure.

The ideal approach for modelling for a project is having more than one model which you can compare for the best possible outcome. The copious amount of modelling techniques makes the task of creating various data models easier. This phase has four tasks:

- **Selecting modelling techniques**: deciding which methods to use neutral, cluster, regression…
- **Generate test design**: breaking down the modelling approach in to training, testing and validation.
- **Build model:** creating a model that fits the business requirement or objective.
- **Assess model:** by comparing the various models that have been build, deciding which is the best fit.

E. Evaluation

In this phase, the models that have been created will be evaluated to determine if they have met the objectives that were set in previous phases or not. Based on the evaluation's success we can then move forward to the next phase, if not however, we must go back and alter the model to make it a better fit.

The deeper the evaluation is the better it is to secure success, reviewing the steps involved and the creation of the model and assessing it thoroughly can be beneficial in the long run to avoid mistakes and stay focused on the objectives that were set. This phase has three tasks:

- **Evaluate results**: determining if the created models meet the objectives or not.
- **Review process**: thorough evaluation of all the steps taken in-order- to determine if any steps were overlooked or not given much emphasis.
- **Determine next steps:** following the above two tasks we will then decide if we will go ahead with the project or if we would need to change or adjust further.

F. Deployment

Based on the requirements set initially, this phase can either be simple as possible or very difficult (CRISP-DM, 2013). The accessibility of a model is as crucial as the model itself. This phase has four tasks:

- **Plan deployment:** map out the deployment process and model
- **Plan monitoring and maintenance**: detecting and avoiding issues while deploying is very important, having a plan to effectively carry out this task can reduce risks.
- **Produce the final report**: the project team will create a summart of the data mining results so it can be inspected further.

- **Review project:** breaking down the project process to learn from the mistakes, successes, and improvement areas.

4 Related Research

In this chapter the topics about job satisfaction, Data mining and Machine learning will be discussed in depth so readers can get better understanding of each method used during this thesis. Furthermore, it will determine the importance of each topic pertaining to the thesis.

4.1 What is Job Satisfaction

Job satisfaction is a phrase we seem to hear often, but as much as it is being talked about it does not have an exact definition. How a person defines their satisfaction is very dependent on their own views and values thus, it makes it hard to generalize the definition of job satisfaction. But the definition that comes close to describing this term is, the positive emotional feeling people feel in response to the experience they have in doing their job (BasuMallick, 2020).

The favourable or unfavourable conditions employees are met with daily in their workspace will lead them to determine whether they are satisfied or not. There are many things to take into consideration for a person to say that they are satisfied with the job they do. Because of that job satisfaction is an unquantifiable metric that is hard to measure, it varies from employee to employee even when said employees are working under the same conditions the result is often different.

It is important for companies to measure and monitor their employee's satisfaction frequently. Research shows that satisfied employees have higher motivation to work, very engaged and perform better (Bourne J., 2020). The benefit of this has not gone unnoticed by companies, hence the reason for the bi-yearly or quarterly satisfaction survey every company conducts to gauge the overall mind set their employees are in.

The satisfaction surveys contain a list of question that are tailored to measure the feeling of employee's at their job, be it positive, negative or neutral (De La Haye, 2020). These questions are specific to the industry the company is in and takes in to consideration different factors, characteristics and level of their employees like age, gender, level of education… Having these factors as a basis of the survey can help determine the trends in the organization that was not otherwise detectible if a survey was not used.

In the old days there were not surveys but performance level was the determinant of the importance of an employee. Regardless of the working conditions, an employee was meant to produce "x" amount of product, sell "x" amount of product or have "x" amount of clients and if they fail to do so they were deemed unimportant and thus expendable. Fortunately, that has changed. Companies now not only take the working conditions in to considerations but other economical, political and social factors when creating these surveys. They make sure that each employee is valued and that on its own can boosts the satisfaction level each employee will have.

Since satisfaction is not a quantifiable element, the measurement can be tricky and open to interpretation. But most surveys take on a form of answer being "strongly dissatisfied" and "strongly satisfied", scoring in this way will simplify the data analysis and is more-quicker for employees to finish the survey.

The satisfaction survey conducted for this thesis took similar form where each employee answers ranged from very dissatisfied being a score of "1" increasing all the way to very satisfied "5". The questions on the conducted survey are very general as this survey was not for a specific industry but is effective in giving accurate and general satisfaction level idea to the researcher.

4.1.1 Importance of Job Satisfaction

Over the decades, the world has moved from being a product and manufacturing market to service oriented market(Bathena Z, 2018). Most of the growth in the market has been towards soft powers and services economies, and companies realize that they must shift their focus on the human resources they possess. This focus is about giving their employees the necessary attention to the issues they face, and keep the employees happy. One

of the main goals of businesses or organizations is ensuring the satisfaction of employees with regards to how they feel about their daily job.

Employee retention is a key goal for organization, not just from reducing cost perspective, in terms of training a new employee and building trust. But creating loyal employees that will treat the firm as if it were their own and give their best performance. A retained and satisfied employee will not only be an ambassador for their brand with in the company but also praise the company externally also.

Employee satisfaction needs to be viewed in both short term and long term (Bathena Z. 2018). The short term is linked to the way the employees see the company, if they perceive it in a positive light they can see themselves working there in the future. If not, they will look for change as soon as possible, and because they have not built a sense of trust and loyalty to the company it will be easy for them to do so. Resulting in high employee turnover.

In the long-term vision is important to determine which of the employees are dissatisfied, because having to work at a place they are not satisfied with will spur them to look for other reasons to dislike the company even more. A dissatisfied employee might associate their dislike for the company with the values it has and derive a skewed assumption, for example favoritism, gender inequality, wage gap… Assumptions like this will hurt the company's reputation and cast a bad light and can be projected extrinsically affecting its equity.

Moreover, in due time it comes to a point where the employee satisfaction will affect profitability because of sub-par performance of employees in treatment of customers. For this additional reason companies should invest in their employees well-being. Research shows that service organizations that have great performance and customer satisfaction have better return on investment (Deeprose, 2006).

When a company is praised for its great working conditions and opportunities, prospects will line up in hopes of getting a job there. And the opposite is true if it is criticized. Therefore, companies need to identify dissatisfied employees and work towards fixing their grievances and leaving a positive impression.

Companies have adopted the “Customer is always right” mantra especially in the service sector. This has greatly affected the employees and, employee turn-over had sky rocketed (Bourne J, 2020). So, through time this mantra has since been modified to include the employees as well and be fair in deciding who was wrong in difficult situations. The idea these days service giving companies are adopting is that if the company takes care of the employees then the employees will take care of the customers to the best of their abilities. This has brought balance to both entities and employees feel like they have a voice in standing up for themselves in cases of disputes with customers.

The questionnaire in this research has included questions about the feelings employees have towards the values of the company they are employed in. Whether or not the way the company enforces its values and is inclusive of their cultural differences. Having questions like this will demonstrate the notion they have towards their companies and if they feel recognized or not.

4.1.2 Factors and out-comes of Job Satisfaction and Performance

Work is one of the most relevant aspects in a person’s life. Understanding the effects several conditions have on job satisfaction is crucial to enhance employee’s productivity and performance.

The most essential outcomes of job satisfaction is job performance (Robbins and Judge, 2019). Satisfaction of employees in their work will guarantee great customer service and increase customer satisfaction. Job satisfaction correlates with performance strongly, and the effect it has can be indicative of the quality of work that is done.

When referring to working conditions to make employees more satisfied, there are external, internal, environmental, and personal things to consider. In each aspect, there is elements we can change to increase the satisfaction of employees. At the same time there are countless aspects that are out of the control of the company such as, weather, travel, personal relationships… These things occur outside of the work and thus cannot be

changed or controlled so they cannot be considered as attributes if an employee is dissatisfied.

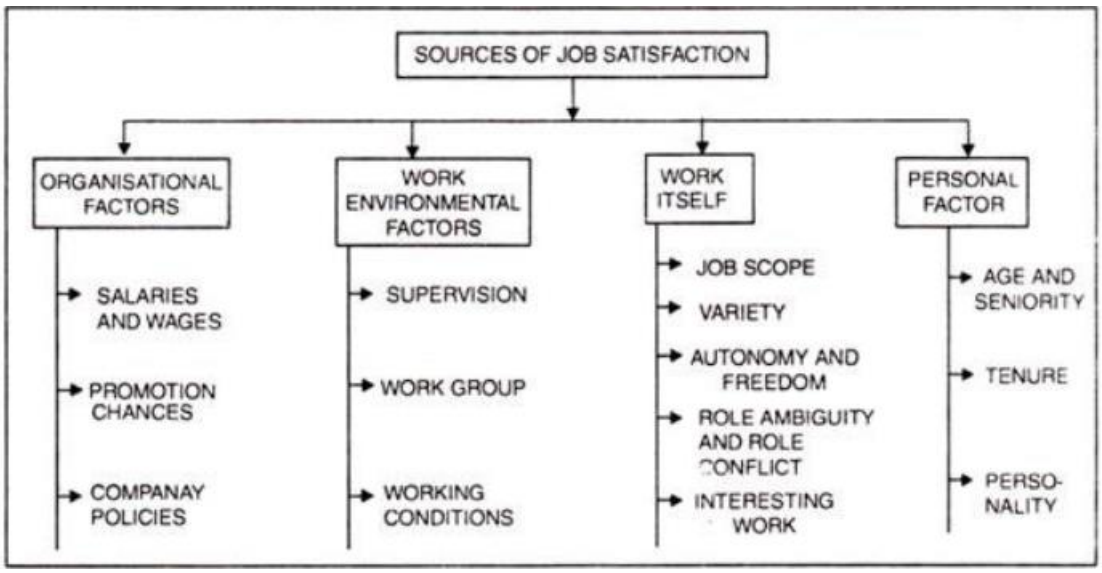


Figure 2. Sources of Job satisfaction (Ventakesh, 2019)

Organizational factors play a significant role in influencing the satisfaction of an employee. The list in the above picture is according to the priorities given by most employees, salaries and wages is in the top. According to Maslow's model of satisfaction, money is viewed as an important instrument in making sure that a persona's needs are fulfilled. Furthermore, it is indicative of the worth of the employee in terms of their work being important. This attributes to the employees feeling of accomplishment and achievement (Ventakesh, 2019).

Similarly, work environmental factors influence the job satisfaction in a sense that how the environment is portrayed by the employees. To some extent supervision is necessary to make sure that employees are doing their jobs in a timely manner, if too much supervision is demonstrated the workplace will quickly be resented by employees. The support and co-cooperativeness employees show each other when working on team project is also a very essential, it serves a sort of camaraderie and assistance and creates a smoother and productive working environment.

The questionnaire and interview conducted for this thesis had, multiple questions related to the salary, promotion opportunities, and the freedom to use their skill to execute a task. These questions will shed light on the organizational and work itself, and how it influences them daily.

4.1.3 Measuring Job Satisfaction

Job satisfaction is a concept that is difficult to measure. It is a feeling; therefore, there is no exact metric we can use to measure it. The feelings are all subjective, what one employee views as an asset of the company can be a nuisance to another so the balance to find the exact metric is very challenging.

For many years it was believed that high performing employees are very satisfied with their job. While that might be true when thought about rationally, studies have shown that it is not necessarily always true (iEduNote, 2019). The lack of corelation between the two has caused companies to rethink the path they take to determine the satisfaction of their employees.

Traditionally, companies used annual surveys and reviews to gauge their employee satisfaction, but as of late the frequency has changed to bi-annually or even more depending on the size of the company.

Figure 3. Measuring employee satisfaction (Totah Z. 2020)

The first step an organization needs to determine is the importance of what employee satisfaction means. Its only when a company knows this importance they will give the necessary focus to help their employees. Understanding the need for employee satisfaction will ultimately lead to listing the benefits it will bring. And the approach to accomplish this should be viewed from all angles that can affect the employee's satisfaction and hence their performance. And finally based on the data gathered from the survey the company needs to implement the necessary alterations to better fit their employees.

The correlation between employee satisfaction and engagement is very straight forward. When an employee is satisfied with their work and working conditions, they are more likely to be engaged and dedicated to help their company achieve its mission (Nevogt D, 2020). Furthermore, there is higher possibility that a satisfied person is not going to leave their job.

Now it is completely possible for employees to have positive feelings towards their job without being highly engaged or motivated. For example, a person in their late fifties who

is comfortable with a secure and steady job, who is fine with the routine job they would need to do daily. So, job satisfaction does not necessarily mean there is high engagement. But the reverse is not true. An employee cannot have low satisfaction with being highly engaged. So routinely measuring satisfaction level is beneficial for companies to weed out the employees that have exhibited low levels of engagement and satisfaction and try to fix it or move on to a better solution.

This thesis emphasizes the importance of the frequency of these measures among other things, for a company regardless of its size. Finding out problems occurring in the business is important to find out early on instead of letting it fester for a while where it gets to a point it will affect the company negatively.

4.2 Data Analysis

Data analysis is the process of collecting, modelling and analyzing data set to elicit insights that will facilitate decision making (Calzon B, 2021). It is examining a set of data that we have collected to draw reliable conclusions about the data collected to expand our knowledge on the topics.

Industries nowadays utilize data analysis to help them in making an informed decision and actions. Data analysis is not subject to a specific industry it can be applicable in any time of industry, this makes it a very valuable.

Data analysis helps us to induce the underlying knowledge from the data by removing irrelevant chaos that is created by the raw data. Data collection is a continual process, this makes data analysis an iterative process where one must repeatedly collect new data and make necessary adjustments to the final result (Arora S. 2021).

Qualitative analysis

This approach answers the question ‘what, ‘how’ or ‘why’. This analysis is used for data that cannot be measured or counted or given numerical values instead expressed in language(.Stevens E, 2020)

This thesis mainly uses qualitative data to gather the second part of the survey's question. Where respondents need to answer each question in terms of being satisfied or dissatisfied.

Quantitative Analysis

This approach is measured in numbers, scales and can be further manipulated for a more statistical approach. Anything that can be counted and measured with numerical value is called quantitative data (Stevens E., 2020). Thus it can be measured using quantitative approach.

Text analysis

Test analysis is a technique that can be used to analyse machine readable facts, to create structured data out of unstructured content. By slicing and dicing filed and data to easily to read and interpret data one can extract information.

This thesis has used all three techniques to achieve the intended goal. The qualitative and quantitative approaches are mostly evident in the interview and questionnaire parts, where the employees participating were asked to answer the questions that covered the what, how and why to deciding how they feel about certain aspects of their job by rating them in the numbers ranging from 1- 5. Furthermore, this thesis has used text analysis or text mining to cluster, predict and associate data results.

4.2.1 Descriptive data analytics

The descriptive data analytics model is initial process to any analytical process, it describes or summarize data points for better understanding of the data in a way patterns can be identified (Rawat, 2021).

Although it is the initial phase in analysing data it can be used to draw conclusions and predict future trends, but the accuracy of the model will not be as full proof as predictive data analysis.

Descriptive data analysis can be used as a basis for understanding the different business aspects of an organization and be a foundation to determine the organizations' overall structure.

4.2.2 Predictive Data analytics

Predictive data analysis uses previous data from a company and with the help of machine learning model determines patterns and predicts trends. The model created is then applied to the present data in-order-to predict what will happen in the future (Arora S. 2021).

Almost all businesses apply predictive data analysis to predict the future trends to help make decision for a successful future. Furthermore, it is preferred by organizations because of the type of data and the volume, cost effective computers and tighter economic conditions(Arora S. 2021).

The technological improvement on the tools used for predictive analysis has generated a more sophisticated model. As these tools become more advanced the use of predictive analysis has evolved from just predicting the future to also identifying the “unknown” that were not able to be addressed (DiFransa A, 2021). Identifying the unknown has helped organizations plan for a model that can best align with other made up scenarios that might happen in the future.

This thesis has utilized predictive analysis to determine the future trends in defining whether or not certain aspects of the satisfaction scale will increase or decrease. Moreover, there is recommendations on how to avoid these predictions (if viewed as negative) from happening or if they do happen how they can be resolved.

4.3 Data Mining

The process of extracting information from a collection of data be it from a data warehouse or collection of other data sets is known as data mining. In computer science determining patterns and association of in large volumes is referred to knowledge discovery in database which is another term for data mining (Clifton C, 2020). This process combines tools from statistics and AI with database management to analyse data sets.
Data mining is used in the business sector dominantly but it’s not limited to that sector only.

Linear regression

Linear regression is a graphical representation of all the known points with a straight line that passes through the centre of if these points(DiFrenza, 2021). The line is the representation of the smallest distance between the points. Linear regression mathematical modelling can be used to predict non-existent data, based on the relationship that exists between the line and the other data points.

Text Mining

As linear regression is limited to numeric data, text mining is used for making predictions about non-numeric factors. Text mining is the process of taking an unstructured text and organizing it in a structured format that can be useful in identifying patterns and insights (IBM Cloud education, 2020).

About 80% of data in the world is made up in an unstructured format and text mining can be an asset to be utilized in an organization. Transformation of unstructured data to a structured one that has more meaning enhances the improvement of decision making that is based on these informative analysed data leads to better results for the organization.

Cluster Models

Clustering models as the name suggests is this method is concerned with finding different groups with similar attributes in a data. When entering data in to the machine it has to familiarize itself with the features and patterns without any external mapping of input-output (Data Flair,2020).

4.4 Data Mining with KNIME

KNIME, the Konstanz Information Miner is a free open-source data analytics tool that can be used to integrate machine learning and data mining through the many features it offers. As data science is becoming more and more prominent in our lives, data analysis tools are also becoming more advanced. KNIME analytics platform has features like machine learning with functions like regression and prediction which are important for creating a visual representation of data that is easy to understand (Breaker, 2019).

KNIME is a platform that is built for GUI based workflow, this enables everyone to be able to use this platform without needing to code. This makes it easier to perform tasks such as data manipulations, transformations and data mining to create a single workflow that consolidates the functions in to a single workflow (Goyal S, 2020). Because of these features and more, the researcher found KNIME to be the ideal platform to analyze the gathered data efficiently.

This thesis' other main focus is the detailed use of KNIME in setting up the data to be analyzed and the process of analyzing it. The result of analyzed data is then presented in different models and to predict the future trends of job satisfaction and correlation of different elements within the questionnaire and finally answer each of the sub-questions presented in the research question.

4.5 Minnesota Satisfaction Questionnaire

The Minnesota satisfaction questionnaire (MSQ), is a questionnaire that is filled with questions designed to measure an employee's satisfaction with their job. The fist questionnaire was designed in 1963, and adjusted as time went on. This questionnaire provides an in depth understanding of what an employee values in their job with regards to what they find rewarding.

The format of this questionnaire is designed to be administered to groups or individuals with consideration of employees who can read at a fifth-grade level or higher. The format of this questionnaire is divided in to the long form or short form, each has the same purpose but slightly different intensity. The MSQ forms are gender neutral and measures job satisfaction on 20 five-item scales that are important elements to determine what a person values the most in a job.

The MSQ has an option to use either a long form that takes 20-15 minutes to complete and a short form that that takes 5 minutes to complete. In this thesis the 1977 short version was used as it is just as effective as the long form.

5 Empirical study/Designing the Research

The questionnaire was suggested by the thesis advisor as the MSQ was created by consultation with psychologists and has been used for over four decades with improvements over the years. It contains questions that are key to determine the feeling of employees towards their job. This questionnaire is applicable for any job industry thus why it has been used for so many years.

In this thesis the 1977 short form was used to measure the satisfaction of employees. This form consists of 20 items that best represent scales that are important in the work environment, such as activity, authority, advancement, company policy, creativity. All these factors are covered in each question and their responses are imperative to accurately determine the level of satisfaction each employee has.

The questionnaire was created in a way that respondents can rate each question 1-5 based on the matrix described at the beginning of the questionnaire. For example, 1 signifies ‘very dissatisfied’, 2 ‘Dissatisfied’, 3 ’Neutral’, 4’Satisfied’ and 5’very satisfied’. All 20 questions were rated using this matrix, having the same response options for all simplifies the conversion of responses to numeric form or string form while using KNIME.

Further information of the employee was also included in the questionnaire, like level of education, industry of work and the relation of their education to their job as it can be a crucial factor to affect the employees feeling toward their job. Since this research is not based on a specific industry, like tech or medical, it was essential to include the question of what line of work the respondents are in. The use this information is further explained in the next chapter as we make comparisons and correlations.

Following the compilation of questions, a Webropol was created and the link was sent to respondents by email, shared on Instagram, Facebook, LinkedIn and WhatsApp groups for maximum responses. Whatsapp groups of Haaga-Helia students and LinkedIn, is a very diversified place in terms of nationality, age and field of work utilizing these advantages played an imperative role in getting as much responses as possible. Instagram and Facebook were partially good platforms to get responses but not nearly as LinkedIn and Whatsapp.

The MSQ was filled by 116 people, with various nationalities. After collection of responses the data was converted in an excel file that was then plugged in to the KNIME analytical tool to achieve the goal of analysing the providing the result.

The second method of gathering data was using interviews. The interview method used was a combination of structured and unstructured, so that the researcher could gather more information for the reasoning behind each of the MSQ answers. The interview used the MSQ as a basis for the structured aspects of the interview and the unstructured part as a follow up to previous questions that can be helpful reason out the responses to questions.

Fifteen interviews were conducted with similar questions as the MSQ but open ended. The interviewees were kept anonymous, but for the presentation of results they will be given names by combining their nationality and a number. For example, Phil1 would be the first Philipino interviewee. There were 2 people from the Phillipines, 3 people from India, 2 people from Ethiopia, 2 people from China, 2 people from German, 2 people from Greece and 2 people from Sweden.

The findings from the interview are analysed using the thematic approach with underlying support from the inductive and latent approaches. In this method the researcher searches for any emerging theme's based on the responses from the interviewees. Thematic analysis is one of the common methods in qualitative analysis, it is the study of patterns and deriving meaning from it (Corsley J. 2021).

The inductive approach involves creating themes without any preconception of what the emerging theme will be, allowing the data to determine the theme. Latent approach focuses on underlying meanings and these meanings can be theorized (Crosley J. 2021). By combining these approaches the researcher can form a reliable data analysis result.

6 Data Analysis

This chapter gives a detailed overview of how the analysis was implemented using KNIME. It Follows the KNIME workflow mentioned in Chapter 3. And the clearly defined CRISP- DM method will make understanding the entire process easier, as it is very structured and straight forward.

6.1 KNIME Workflow Development

This thesis aims to give a general understanding of employee satisfaction of foreigners in Finland thus, it is not specific to a group of people or industry. As a result, the findings presented in this thesis are a generalization of the data analysis collected from the questionnsire. The aim is to provide a comprehensible and interactive view on the general feelings of employees towards their jobs currently.

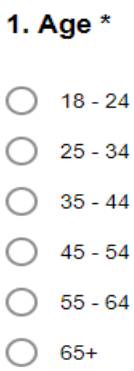
The detailed description of how we can create a KNIME workflow will be explained step by step.

6.2 Data Understanding

In this step we define clearly what questions the researcher is trying to answer, and what the collected information is. Defining these elements initially will help in shaping the analysis process to achieve the goal. With relation to the case, this section will explain the data set provided.

As mentioned in Chapter 4 the MSQ provided the basis for collecting the data set. The first 6 questions covered basic information about the respondents.

1) Age

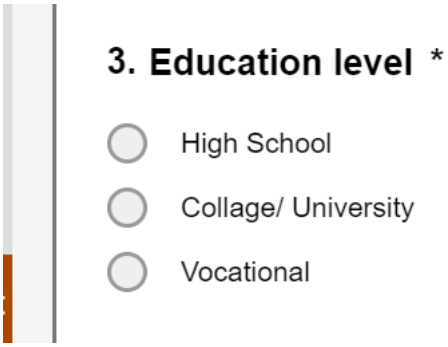


2) Nationality

There was an empty box so that respondents can fill out their nationality.

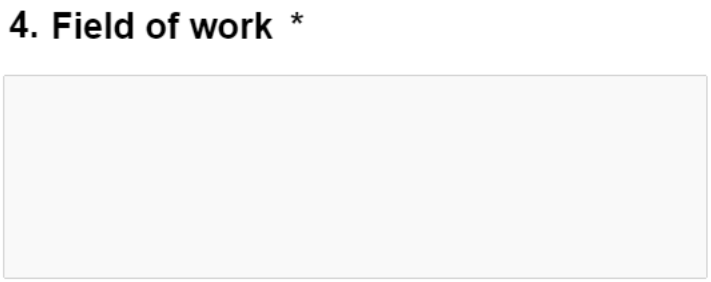
3) Education level

Data gathered from this question when combined with the fourth and fifth question of the questionnaire can give an insight on what the factor is when people score a lower on their satisfaction. The Asterix next to the question signifies that it is a mandatory question to fill out, because data gathered from this question can be used as a basis for comparison with other aspects of the questionnaire.



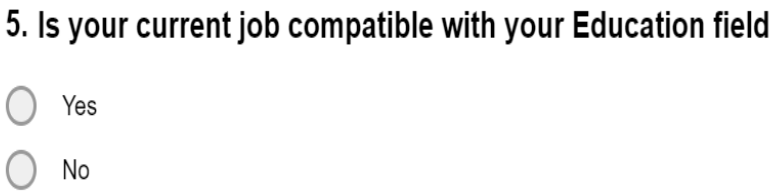
4) Field of work

Similar to question number three this question is also mandatory to decide the industry that has the lowest satisfaction level upon further analysis.

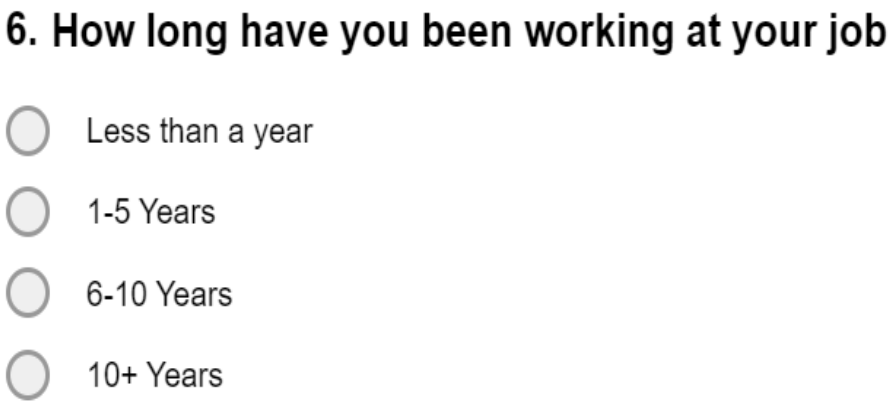


5) Job compatibility with Education

This question plays an important role in presenting the results, it can be used as rational argument point as to why the level of satisfaction is low.



6) How long have you been working at your job



Each question in the MSQ cover the three categories that are prominent factors in influencing job satisfaction and performance of employees. Cultural fit, personal growth and employee mental health and creativity. Basing the questions around these three factors will ensure that the employees are being asked questions that can give a concrete understanding of their satisfaction state. It will also have an impact on the industry in determining the areas in which they are not focusing on.

6.3 Data Preparations

After the collection of data through the MSQ, the researcher transformed the collected data in excel sheet. KNIME mainly uses CSV or Excel to read data as a result the collected data can then easily be plugged to KNIME.

During the data analysis phase excel reader node and string nodes were used among other features. Each features used are discussed below.

Excel Reader node

The excel reader node does exactly as its name suggests and reads Excel Files provided by the user. It can read a single file or multiple files at the same time depending on the user's configurations. Furthermore, it can read live data from the data source (KNIME Hub). Data that is being currently gathered for example, the responses we get from people filling out a Google form can be read by KNIME as the form is being filled.

The initial step to our data analysis was the excel reader node and using this phase as the basis researcher then carried out the rest of the functions.

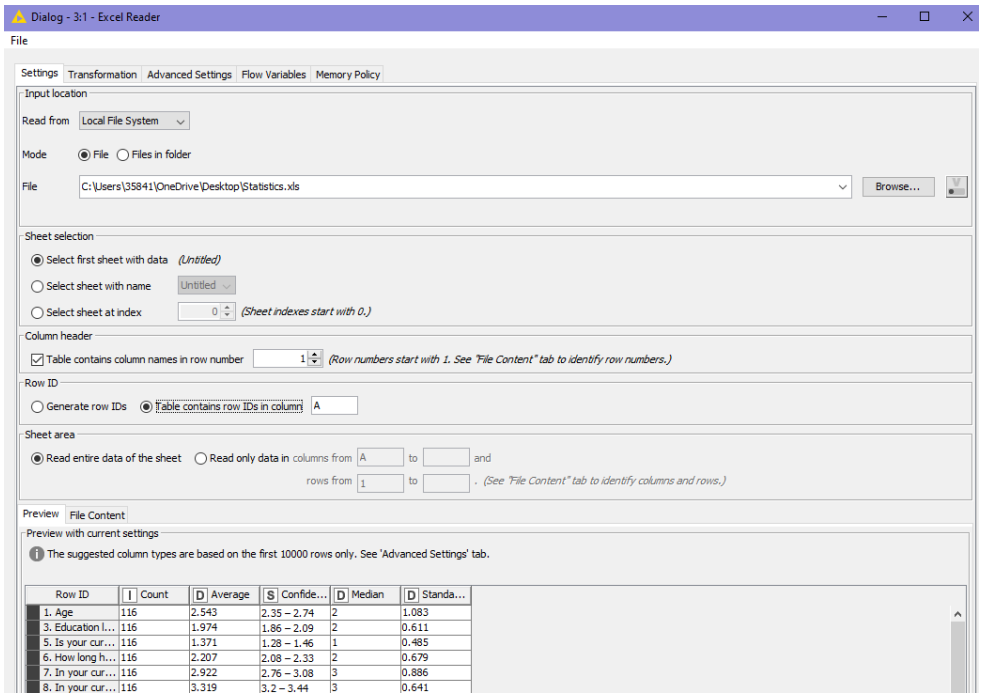


Figure 4. Excel reader node and the Configuration

In the above figure the excel reader node is displayed along with the configurations. After choosing the correct excel file one needs, it is easy to alter the displayed file by changing the configurations shown on the screen. It is possible to switch the row ID, sheet area and sheet selection.

Row and Column Filter

As the name suggests these filters are used to exclude or include rows or columns from the table. By using the include and exclude list users can move the columns and rows to their liking.

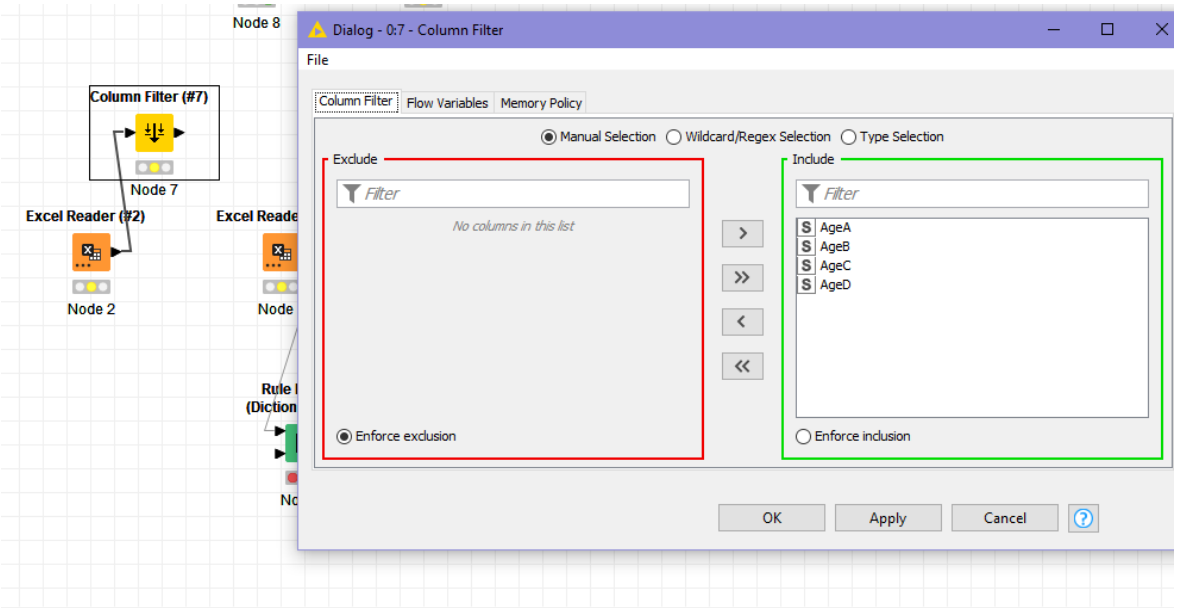


Figure 5: Column Filter (Age or respondents)

In figure 6 the researcher has decided to exclude the column named ‘AgeD’. This column represents the percentage of how many respondents are in that age range in the questionnaire, which means that excluding that column will only show the age and amount of people included in that age range.

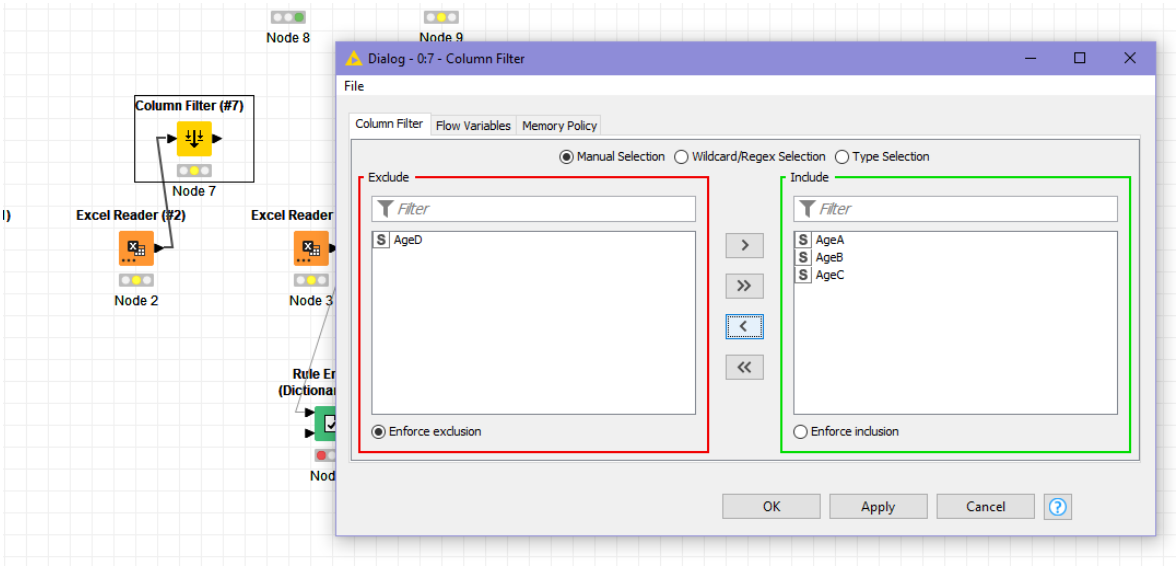
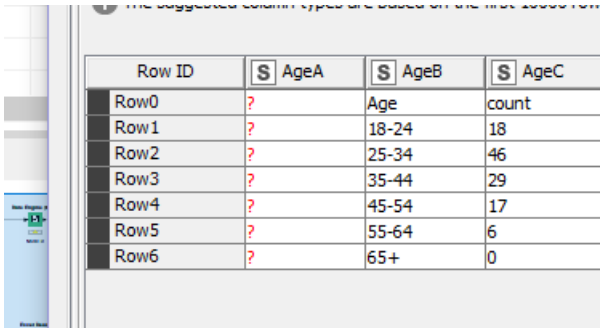


Figure 6: Column Filter (Excluding Column D)



Row ID	AgeA	AgeB	AgeC
Row0	?	Age	count
Row1	?	18-24	18
Row2	?	25-34	46
Row3	?	35-44	29
Row4	?	45-54	17
Row5	?	55-64	6
Row6	?	65+	0

Figure 7: Column Filter of Age and Count

Similarly, the row filter can also be used to exclude or include specific data the same as the column filter as shown in figure 5. By clicking on the single or double arrow located between the two boxes(Figure 5), users can weed out the column or row they do not want. This method can be used to get comparisons from different factors of the data gathered.

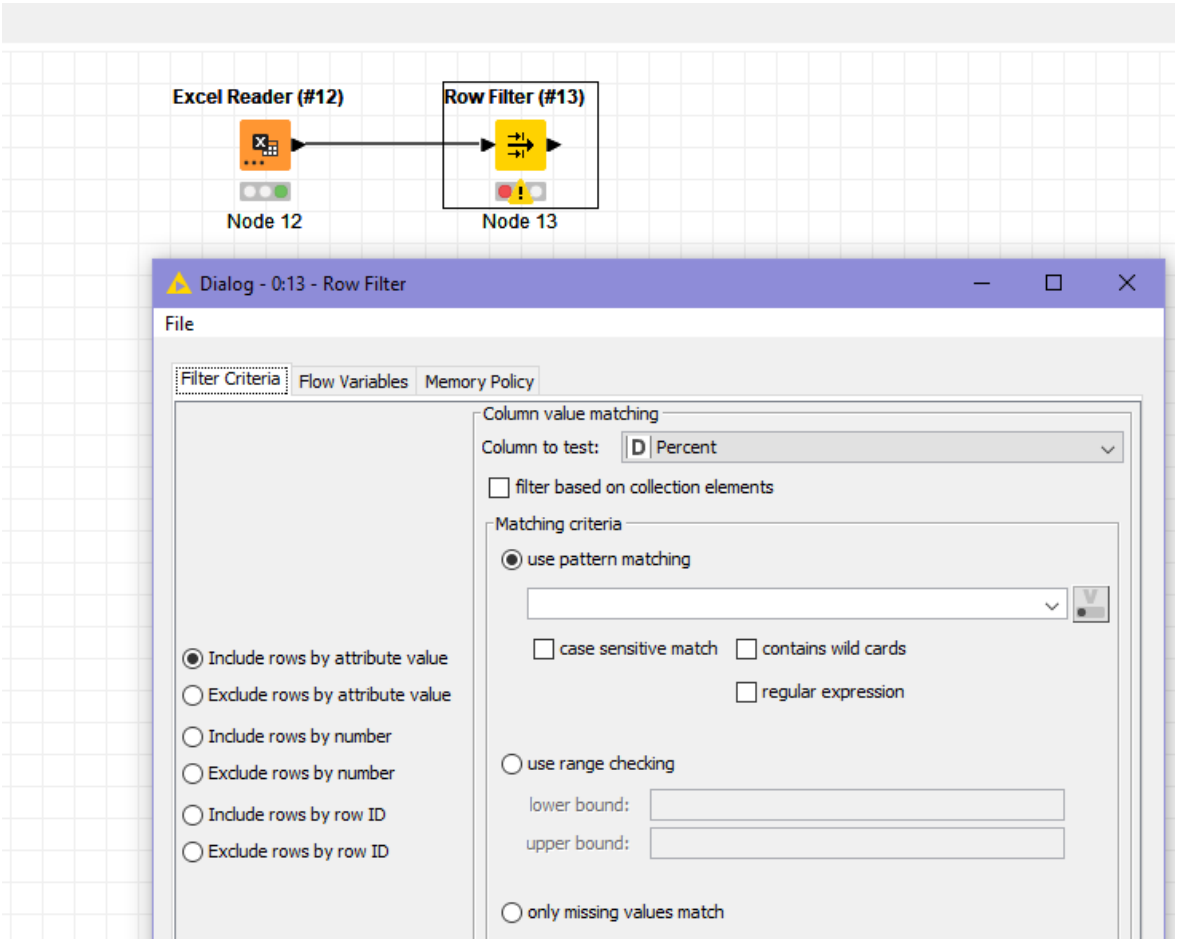


Figure 8: Row Filter

By taking the data gathered from the question 'Is your current job compatible with your education we can analyse if the cause for employee dissatisfaction is because of incompatibility between the job they are given and their education level. In figure 9 below, we can decide to exclude 'Row1', and focus on 'Row0' to see if there are respondents that have a low satisfaction level despite being in the field of work they are educated for.

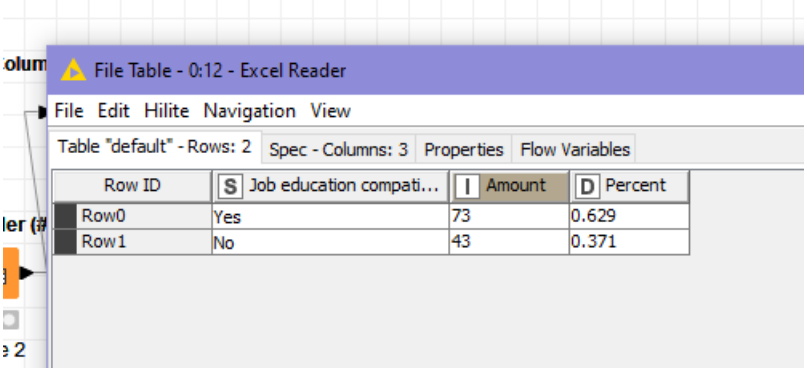
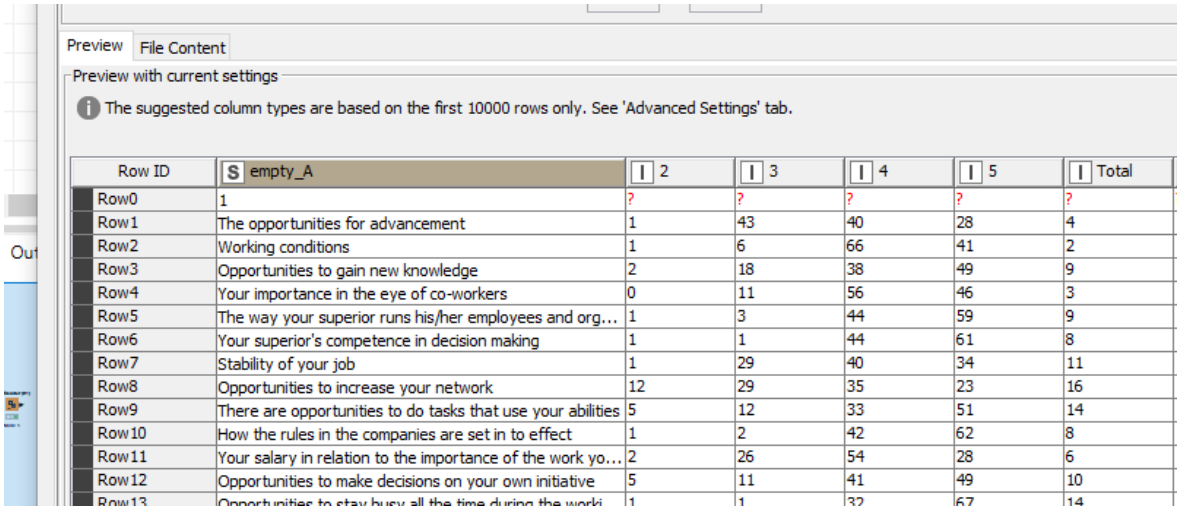


Figure 9 Row filter



Row ID	empty_A	2	3	4	5	Total
Row0	1	?	?	?	?	?
Row1	The opportunities for advancement	1	43	40	28	4
Row2	Working conditions	1	6	66	41	2
Row3	Opportunities to gain new knowledge	2	18	38	49	9
Row4	Your importance in the eye of co-workers	0	11	56	46	3
Row5	The way your superior runs his/her employees and org...	1	3	44	59	9
Row6	Your superior's competence in decision making	1	1	44	61	8
Row7	Stability of your job	1	29	40	34	11
Row8	Opportunities to increase your network	12	29	35	23	16
Row9	There are opportunities to do tasks that use your abilities	5	12	33	51	14
Row10	How the rules in the companies are set in to effect	1	2	42	62	8
Row11	Your salary in relation to the importance of the work yo...	2	26	54	28	6
Row12	Opportunities to make decisions on your own initiative	5	11	41	49	10
Row13	Opportunities to stay busy all the time during the worki	1	1	32	67	14

Figure 10: Respondents satisfaction Metrix

Figure 10 depicts the list of questions on the column ‘empty A’ and the rate of their satisfaction or dissatisfaction with the numbers ‘1,2,3,4,5’. When working with many rows and columns KNIME makes it easier to work with so that users can filter the data they would not need or is irrelevant to the analysis. In the above figure (Figure 9), companies may be more keen to figure out the amount of people that have given a response of about ‘3’ or lower thus they might choose to remove columns ‘4’ and ‘5’. This will only leave the file that has respondents with medium to lower satisfaction. And help shift the focus to them.

Rule Engine

The Rule engine node takes a list of rules or predefined rules from a second input as shown in the figure below and tries to match the row to the table. If a rule is fulfilled then the result is added to a new column.

In this data analysis, rule engine node was instructed to form a new column that says ‘Match’ if the ‘Job education compatibility’ was a ‘Yes’.

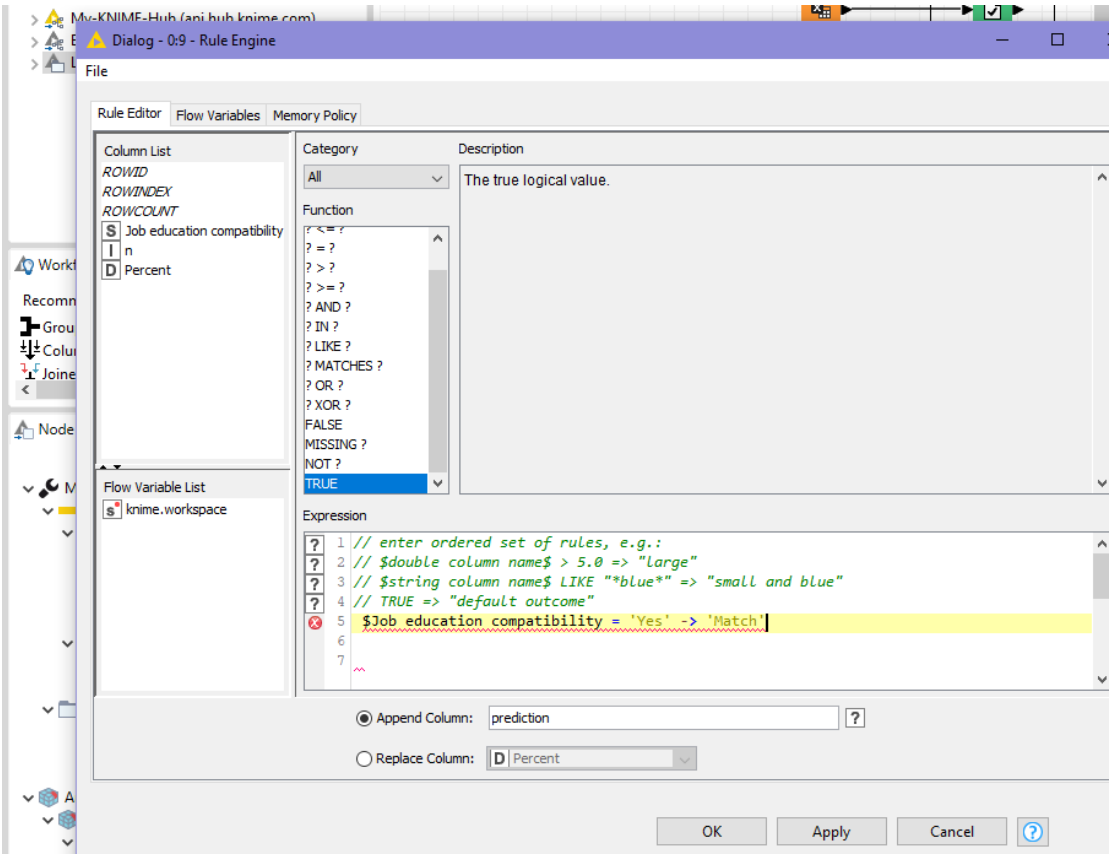


Figure 11: Rule engine condition

In the above figure, the condition the researcher has set is based on the answers the respondents gave as to whether or not the job they are currently in matches their education. After this condition is applied a new column appears with ‘Match’ in the list.

Random Forest Learner

This feature in KNIME learns a random forest, that is composed of chosen decision trees and each tree is analyzed on a different set of rows and columns.

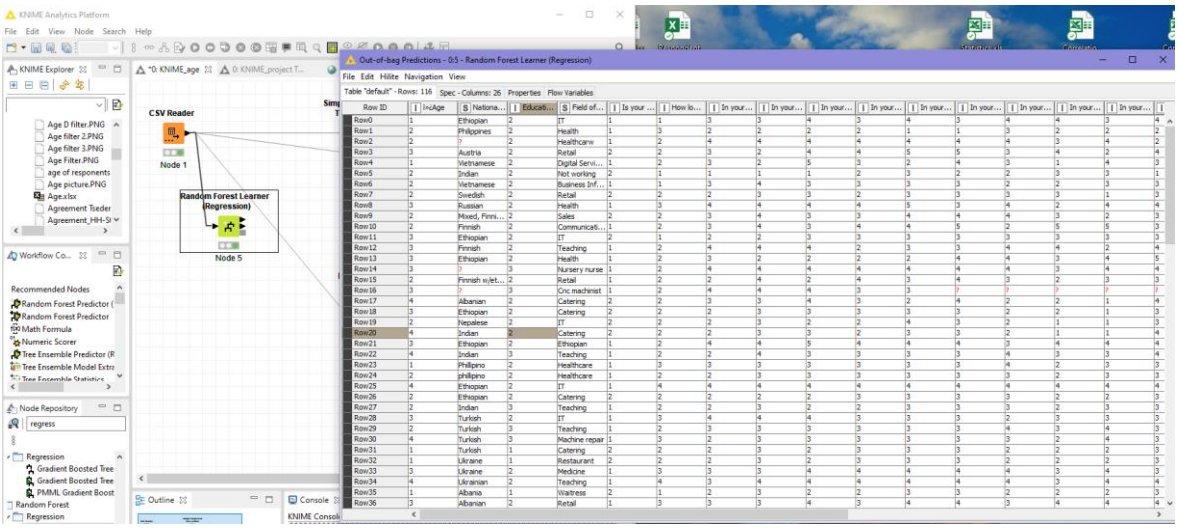


Figure 12: Random Forest Learner

7 Result

In this chapter, the results obtained from analysing the collected data is presented. The results are classified into Participant's profile, comparisons of average low score points and comparison of factors in the questionnaire that are valuable for example, comparison of Job vs Education compatibility, correlation of statements and an over view of respondents matrix.

7.1 Descriptive Data Analysis

Initial data exploration was done to get preliminary insights on getting a summarization of the data set and determining its main features. Figure 11 below shows a table that presents the information about the lowest rating registered, highest rating registered, the standard deviation, variance(measure of the spread of numbers in a data set), skewness(measures the deviation of variables given distribution from the normal distribution and entropy.

Statistics

Question	Count	Average	Confidence interval	Median	Standard deviation	Skewness	Kurtosis	Entropy
1. Age	116	2.54	2.35 – 2.74	2	1.08	0.49	-0.4	2.07
3. Education level	116	1.97	1.86 – 2.09	2	0.61	0.01	-0.26	1.32
5. Is your current job compatible with your Education field	116	1.37	1.28 – 1.46	1	0.49	0.54	-1.74	0.95
6. How long have you been working at your job	116	2.21	2.08 – 2.33	2	0.68	0.74	1	1.41
7. In your current job are you satisfied with the following: The opportunities for advancement	116	2.92	2.76 – 3.08	3	0.89	0.38	-0.75	1.78
8. In your current job are you satisfied with the following: Working conditions	116	3.32	3.2 – 3.44	3	0.64	-0.2	0.87	1.37
9. In your current job are you satisfied with the following: Opportunities to gain new knowledge	116	3.39	3.22 – 3.55	3.5	0.9	-0.35	-0.31	1.86
10. In your current job are you satisfied with the following: Your importance in the eye of co-workers	116	3.35	3.23 – 3.48	3	0.69	-0.11	-0.31	1.5
11. In your current job are you satisfied with the following: The way your superior runs his/her employees and organization	116	3.62	3.49 – 3.75	4	0.71	-0.36	0.91	1.51
12. In your current job are you satisfied with the following: Your superior's competence in decision making	115	3.64	3.52 – 3.76	4	0.66	-0.36	1.26	1.4
13. In your current job are you satisfied with the following: Stability of your job	115	3.22	3.04 – 3.39	3	0.96	0.15	-0.8	1.93
14. In your current job are you satisfied with the following: Opportunities to increase your network	115	3.02	2.8 – 3.24	3	1.2	0.09	-0.85	2.22
15. In your current job are you satisfied with the following: There are opportunities to do tasks that use your abilities	115	3.5	3.32 – 3.68	4	0.99	-0.63	0.16	1.94
16. In your current job are you satisfied with the following: How the rules in the companies are set in to effect	115	3.64	3.52 – 3.77	4	0.68	-0.45	1.24	1.44
17. In your current job are you satisfied with the following: Your salary in relation to the importance of the work you do	116	3.09	2.93 – 3.24	3	0.86	0.17	-0.17	1.81
18. In your current job are you satisfied with the following: Opportunities to make decisions on your own initiative	116	3.41	3.24 – 3.58	4	0.93	-0.59	0.35	1.88
19. In your current job are you satisfied with the following: Opportunities to stay busy all the time during the working day	115	3.8	3.67 – 3.93	4	0.69	-0.53	1.56	1.46
20. In your current job are you satisfied with the following: Opportunities to try your own methods to get the job done	115	3.69	3.53 – 3.84	4	0.84	-0.7	0.85	1.73
21. In your current job are you satisfied with the following: Opportunities to work alone in your job	116	3.7	3.54 – 3.86	4	0.9	-0.84	0.91	1.79
22. In your current job are you satisfied with the following: Compliments you receive for a job well done	115	2.65	2.5 – 2.81	3	0.86	0.32	0.3	1.8
23. In your current job are you satisfied with the following: The sense of accomplishment you get after completing your work	116	3.66	3.53 – 3.8	4	0.73	-0.32	-0.03	1.57

Figure 12: Descriptive Statistics of Numerical Features

7.2 Participant profile

The respondents to the MSQ were as many as 116 people, although the survey was opened by around 287 people. All the participants were not Finnish as the thesis is focused on foreigners' emotions towards their job.

Nationality

The respondent country of nationality is very diversified covering about 21 countries. And each country was represented by two or more respondents, this strengthens the data towards creating a pattern.

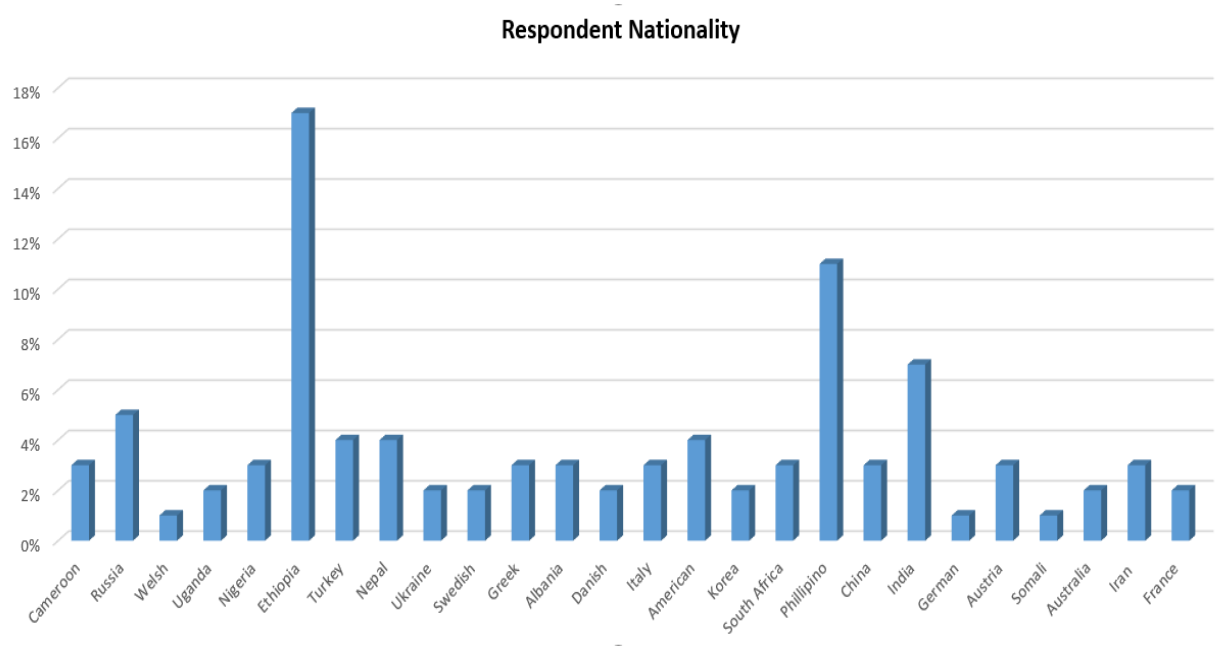


Figure 12: Bar Graph: Respondent's country of origin

As shown in the figure above the majority of the participants were Ethiopian and Phillipino with 17% and 11% respectively. It was followed by India 7 %, Russia with 5%, Nepal, America and turkey are 4% of the respondents and Albania, Ukraine, France, welsh, Australia, Greek, Iran, china, Korea, Sweden, Denmark, German, Somali and Nigeria all below 4%.

Although the respondent distribution has a clear gap and does not correlate to the number of foreigners represented with that nationality living in Finland. It gives a base for understanding and readers can use it to gauge the satisfaction levels of these foreigners.

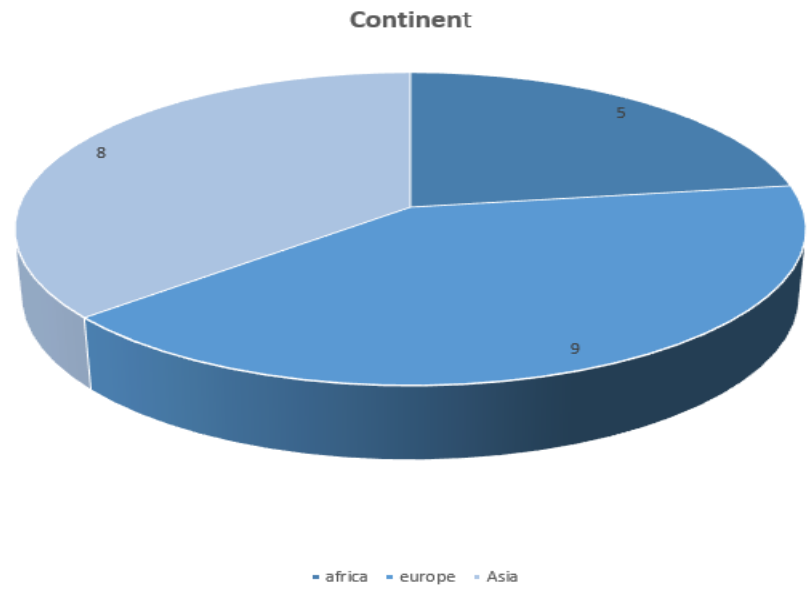


Figure 13: Continental distribution of respondents

As shown in the figure above when classified according to continent the most respondents are from different parts of Europe. Where there were as many as 9 nationalities from the respondents. The second highest participant is Asia with 8 nationalities taking part and finally Africa with 5 countries participating.

Age

Another question included in the MSQ to further gather basic information about the respondents was their age. The importance of age will be further discussed in the explanation parts in the coming pages. Up on analysing the collected data, it revealed that the majority of foreign employees are ages 25- 34 at 39.60%, 35-44 coming close at 25%, in third place 18-24 at 15.50%, 45-54 at 14.70% and lastly 55-64 at 5.20%. 65 and above was included in the questionnaire but there were no respondents by that age.

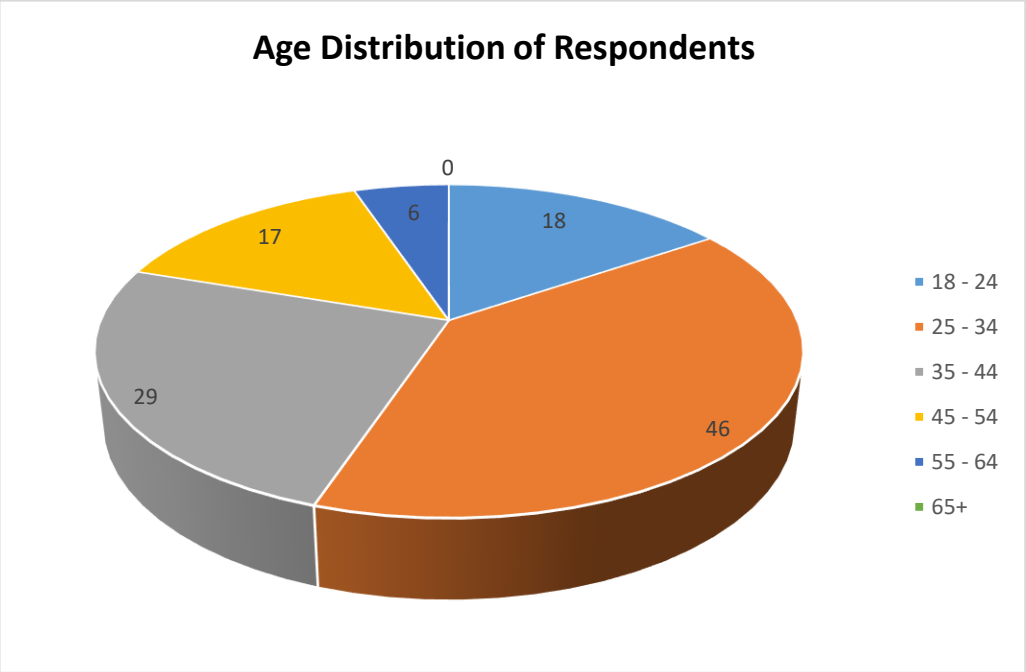


Figure 14: Age distribution of respondents

Education Level

In the first part of the questionnaire where respondents were asked basic information about themselves that would help the researcher further, such as their level of education showed that almost 63 percent of the working foreign population in Finland is at a university/collage level of educated. 19.80% is high school graduated and 17.30% of the respondent population has vocational education.

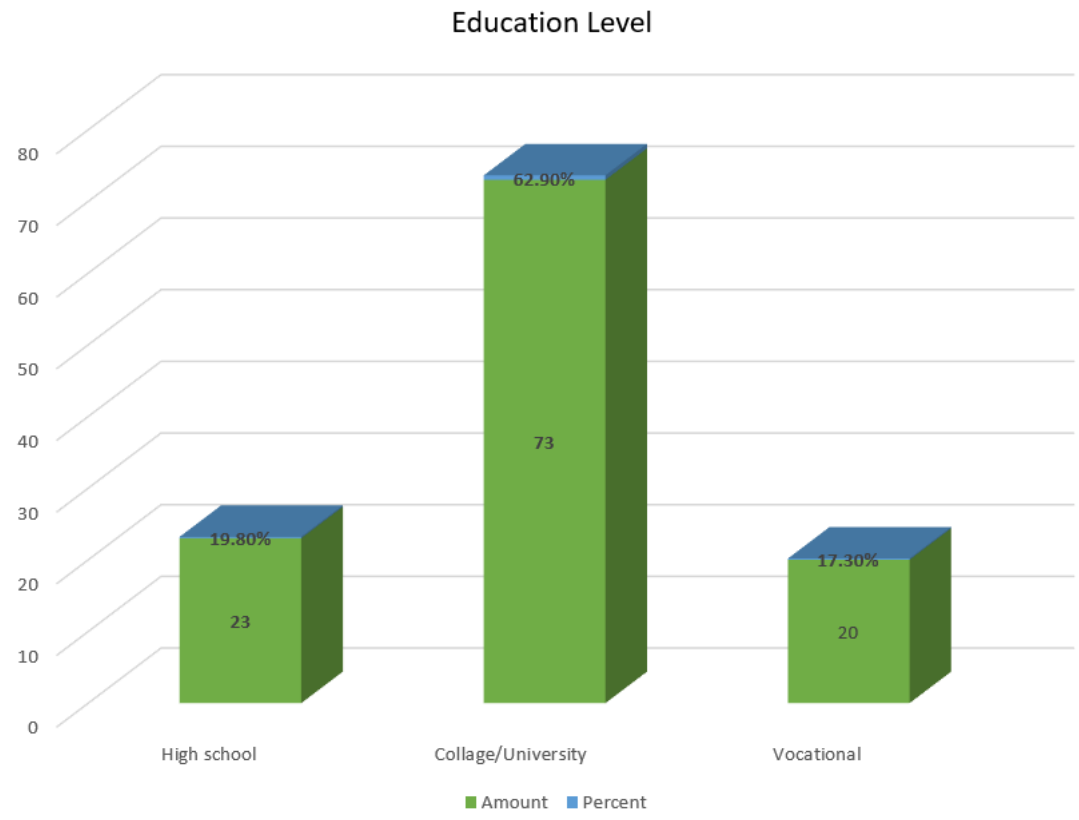


Figure 15: Bar Graph: Education level

The job education compatibility was another major question to determine the source of satisfaction or dissatisfaction employees might have, which its importance be discussed in the coming pages as well. The figure below shows that the 63% of the respondents have job to education compatibility and 37% determined that there is no compatibility.

Job to Education Compatibility

The Job to education compatibility is analysed in this thesis to determine how many foreigners are working in a field that does not corelate with their education. The analysis determined that 62.90% (73 people) responded yes and 37.10% (43 People) answered no.

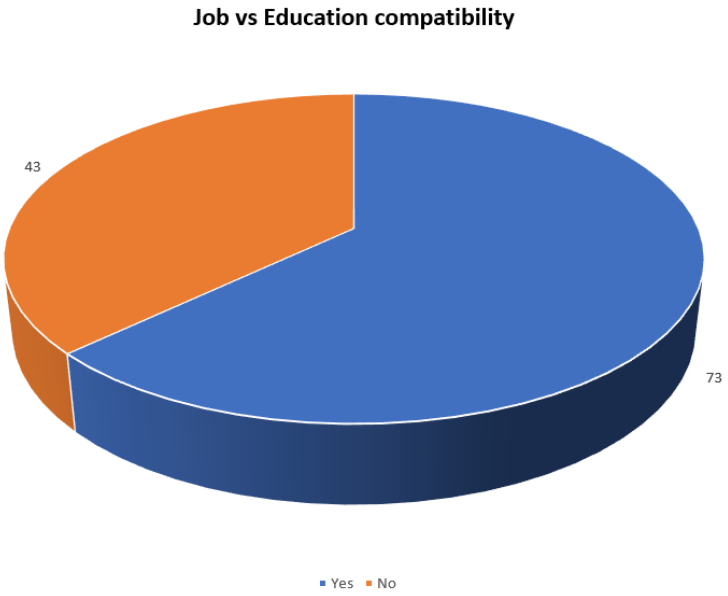


Figure 16: Pie chart: Job vs Education Compatibility

7.3 Comparison

Closer analysis is made based on Age, Duration of stay in current job, Education level and Field of work. An in depth analysis of each of these elements will define how many foreigners are using their education or training efficiently in the working environment.

Age

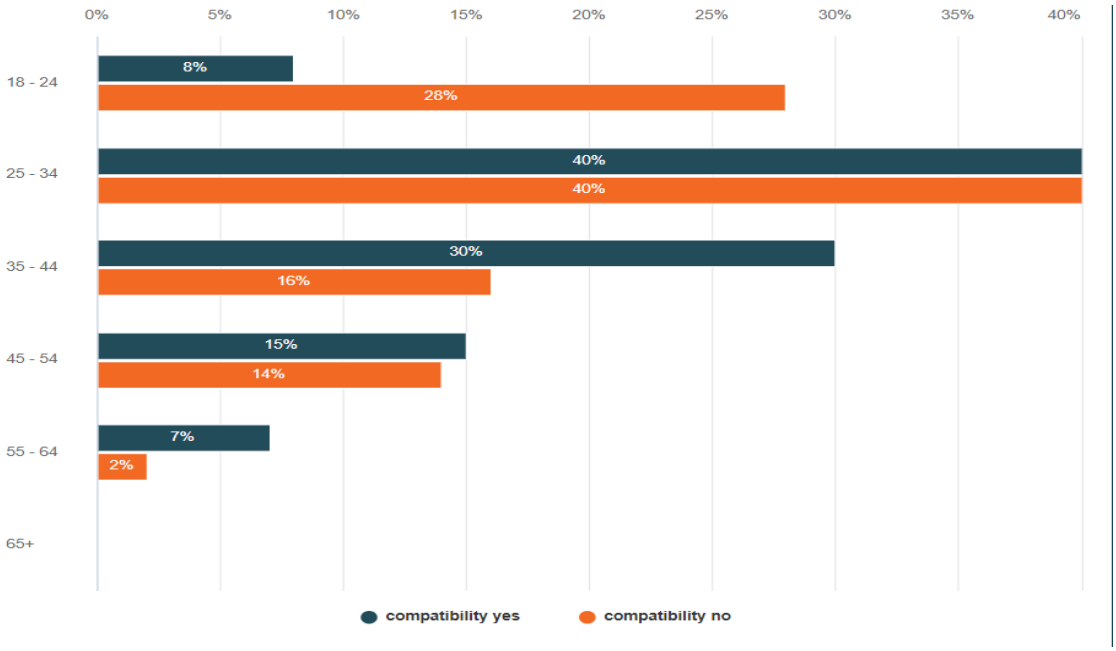


Figure 17: Job vs Education compatibility (Age)

Figure 17 supports previous claim that most of the work force of the foreigner population is in their prime working age (24-34 and 35-44). Furthermore, the figure above shows that from the 116 respondents, 70% are working in the field of what they are trained for or educated on. And foreigners between the ages 45- 54 are almost equal with 15% of the respondents having job to education compatibility and 14% responding no compatibility.

The highest level of incompatibility is recorded in the respondent population between the ages of 25-34. Although the same age group had the highest compatibility rate the opposite was the same amount. The second highest incompatibility was the age group between 18-24. And the least incompatibility is recorded between the ages of 55-64 which was only 2%.

Education Level

	compatibility yes		compatibility no		
	n	Percent	n	Percent	Total
High School	2	2.7%	21	48.8%	23
Collage/ University	54	74.0%	19	44.2%	73
Vocational	17	23.3%	3	7.0%	20
Total	**73**		**43**		**116**

Figure 18: Job vs Education compatibility(Education level)

The table in figure 18 shows that the biggest percentage of respondents that replied ‘Yes’ to ‘job and education compatibility’ have a Collage/University level education, vocational is the second highest with 23.3% (17 respondents) and the lowest which is 2% for High school level education.

On the contrary, majority of the respondents who answered ‘No’ to the compatibility question are also at the High school education level with 48.8% (21 people), closely followed by Collage/University level education with 44.2% (19 people) and 7% (3 people) with Vocational education.

Duration of stay in current job

Figure 19 shows a table that has calculated the compatibility of job vs education with regards to how much respondents have been working at their job. Out of the 73 respondents that have job vs education compatibility 44 of them have been working at their current job for 1-5 years. 19 respondents 6-10 years, 6 respondents more than 10 years and 4 respondents less than a year.

On the other hand, respondents that do not have job vs education compatibility majority of the respondants (32 respondents) have stayed at their current job for 1- 5 years, 4 respondents 6-10 years, 7 respondents less than a year and there were no respondents that have been at a job for more than 10 years

	compatibility yes		compatibility no		
	n	Percent	n	Percent	Total
Less than a year	4	5.5%	7	16.3%	11
1-5 Years	44	60.3%	32	74.4%	76
6-10 Years	19	26.0%	4	9.3%	23
10+ Years	6	8.2%	0	0.0%	6
Total	**73**		**43**		**116**

Figure 19: Job vs Education compatibility (Duration of stay in current job)

Field of work

compatibility yes	compatibility no
IT	Retail
Health	Not working
Healthcarw	Retail
Digital Service Design	Sales
Business Information technology	IT
Health	Catering
Communications	Catering
Teaching	IT
Health	Catering
Nursery nurse	Catering
Retail	Catering
Cnc machinist	Restaurant
Ethiopian	Waitress
Teaching	Catering
Healthcare	Restaurant
Healthcare	Catering
IT	Retail
Teaching	Catering
IT	Retail
Teaching	Catering
Machine repair	Cleaning/ janitorial services
Medicine	Retail
Teaching	Retail
Retail	DJ
Healthcare	Retail
IT	Machine technician
Teaching	Waitress
Teaching	car Sales
Sales	Consruction
Healthcare	Restaurant
IT	Janitorial Services
Teaching	Waiter
Teaching	Photography/video making
IT	Sales
Teaching	Sales
Sales	Sales
IT	Retail
child care	car sales
IT	Retail
Sales	Cleaning
IT	Retail
IT	catering
Child care	Retail
Teaching	
Medical	
translator	
Teaching	
Car Sales	
IT	
Translator	
Retail	
Restaurant	
IT	
Child care	
Retail	
Catering	
IT	
IT	
IT	
Art	
Music	
Teaching	
Translator	
IT	
teaching	
Art	
IT	
Translator	

Figure 20: Job vs Education compatibility (Field of Work)

The diversity of jobs shown in figure 20 further supports and gives more detail to the claims shown in figure 16 ‘ Job vs Education compatibility’ where it showed that 73 respondents are in the working field that is compatible with their education and 43 respondents working field was incompatible.

Respondent Satisfaction Matrix

The final part of the questionnaire was the list of questions that will be able to determine the satisfaction level of employees. Respondents were given the matrix to rate their feeling towards each question. This is the most crucial part of data gathering as it contains the main information for the basis of the thesis.

As mentioned in data preparations section of this thesis, the questionnaire was built around three factors that affect job satisfaction level: cultural fit, personal growth and Employee mental health and creativity. The first figure shows the result for the cultural fit of factor.

	1	2	3	4	5	Total	Average	Median	Mode
The opportunities for advancement	1	43	40	28	4	116	2.9	3.0	2
Working conditions	1	6	66	41	2	116	3.3	3.0	3
Opportunities to gain new knowledge	2	18	38	49	9	116	3.4	3.5	4
Your importance in the eye of co-workers	0	11	56	46	3	116	3.4	3.0	3
The way your superior runs his/her employees and organization	1	3	44	59	9	116	3.6	4.0	4
Your superior's competence in decision making	1	1	44	61	8	115	3.6	4.0	4

Figure 17: Respondent satisfaction Matrix (First Factor)

Out of the above six questions depicted in figure 17, the lowest score belongs to the question ‘opportunities for advancement’. These questions are related to an employee’s personal growth. 40 of the 116 participants were neutral to that question with regards to their current jobs and 43 participants are dissatisfied with their job’s inability to give opportunities for advancement and 1 participant is very dissatisfied.

The second wave of questions covered the personal growth factor. As figure 18 presents The average of each of these questions is 3 and above. The highest score is for the question ‘Opportunities to stay busy all the time during the working day’ with an average score of 3.8 and 67 of the participants are ‘satisfied’ and 15 are ‘very satisfied’.

Referring again to figure 18, ‘Opportunities to work alone in your job’ and ‘opportunities to try your own methods to get the job done’ is the second highest average both being an average of 3.7. Over half of the respondents rated these questions ‘satisfied’ and ‘very satisfied’.

Stability of your job	1	29	40	34	11	115	3.2	3.0	3
Opportunities to increase your network	12	29	35	23	16	115	3.0	3.0	3
There are opportunities to do tasks that use your abilities	5	12	33	51	14	115	3.5	4.0	4
How the rules in the companies are set in to effect	1	2	42	62	8	115	3.6	4.0	4
Your salary in relation to the importance of the work you do	2	26	54	28	6	116	3.1	3.0	3
Opportunities to make decisions on your own initiative	5	11	41	49	10	116	3.4	4.0	4
Opportunities to stay busy all the time during the working day	1	1	32	67	14	115	3.8	4.0	4
Opportunities to try your own methods to get the job done	2	7	31	60	15	115	3.7	4.0	4
Opportunities to work alone in your job	3	8	27	61	17	116	3.7	4.0	4

Figure 18: Respondent Satisfaction matrix (Second factor)

Figure 19 shows that the last factor is employee mental health and creativity. The question ‘compliments you receive for a job well done’ scored and average of 2.7 which is one of the lowest score from all the questions. But on the contrary the last question ‘The sense of accomplishment you get after completing your work’ had an average score of 3.7 with 73 people rating it ‘satisfied’ or ‘very satisfied’.

Compliments you receive for a job well done	8	42	50	12	3	115	2.7	3.0	3
The sense of accomplishment you get after completing your work	0	7	36	62	11	116	3.7	4.0	4

Figure 19: Respondent Satisfaction matrix (Third Factor)

7.4 Text mining

As mentioned in Chapter 4, text mining was used to find recurring words from the respondents answers. Text mining was utilized in parts of the questionnaire where respondents had to write something instead of checking a box that already had a predetermined answer. The ‘Nationality’ and ‘Field of work’ were two of the questions that needed text mining to determine a theme with in the replies. And the result is shown as follows.

Field of work

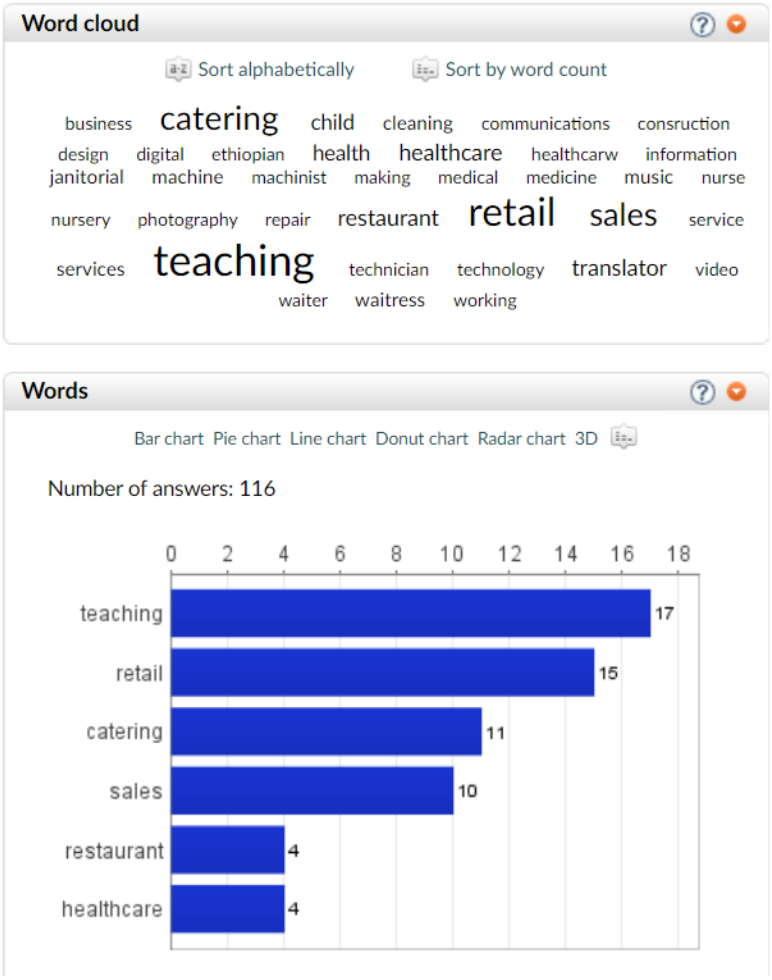


Figure 21: Text mining (Field of work)

In figure 21, the word cloud shows the random answers of the respondents in different fonts their size implying the vague amount of respondents who answered using the same word. Larger fonts means that there were more people in that working field and smaller ones signify small amount of people. The word cloud is supported by a more solid figure by the bar chart below it where it shows the amount of respondents in numbers that has the same wording.

Majority of the respondents are working in the teaching field (17 respondents), retail (15 respondents, catering (11 respondents), sales (respondents) and health care along with restaurant each have 4 respondents.

Although in Figure 21 it was shown that there were more working fields text mining focuses on words that have the most count. As a result the bar graph only shows the words used to describe the working field that has used the same word.

Nationality

Similarly in figure 22, text mining shows the nationalities that have the most frequency in bigger fonts and gives detail on how many people have answered the same way. The dominant nationality in this set of respondents is Ethiopian which consisted of 33% of the respondents, 16% Indian, 13% both American, Vietnam and Russian and 11% Pilipino.

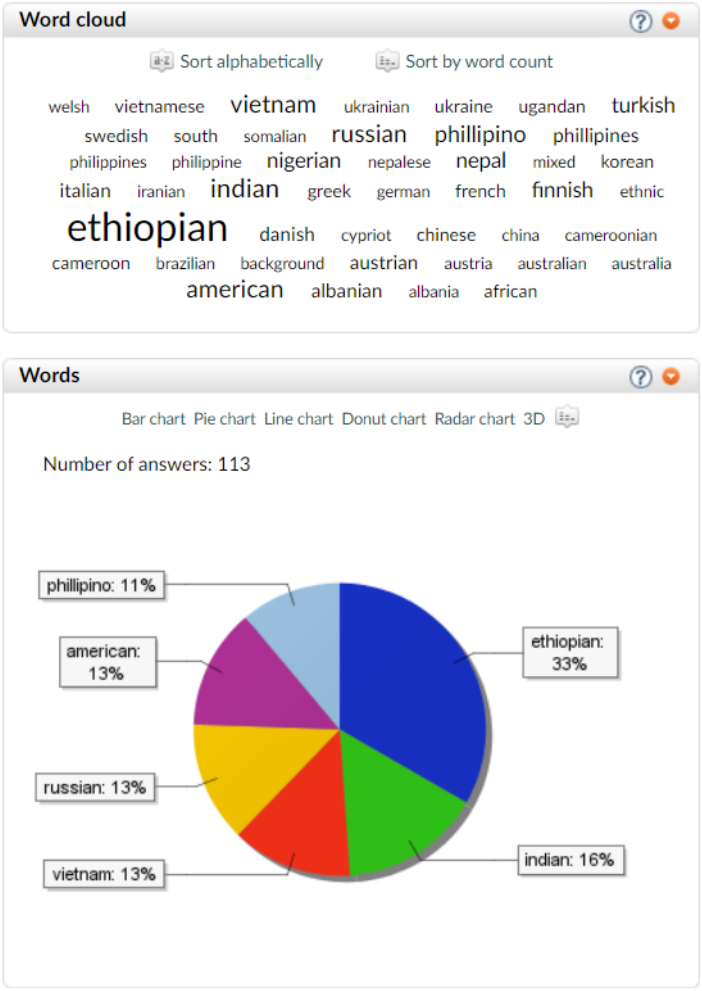


Figure 22: Text Mining (Nationality)

7.5 Correlation

The correlation matrix in figure 24 was used to determine the collinearity of independent variables. The questions 'Is your current job compatible with your education' and 'education level' had strong correlations thus either one of the statements should be removed from the selection of data to avoid a skewed result because redundancy of information will affect the regression model.

Correlations

	1	3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	1	0.34	-0.22	0.61	0.22	0.24	0.21	0.17	0.14	0.2	0.29	0.18	0.2	0.24	0.25	0.18	0.14	0.11	-0.03	0.1
	0.34	1	-0.5	0.2	0.01	0.04	0.02	-0.04	-0	0.03	0.4	0.07	0.19	-0.03	0.05	0.06	-0.04	0.01	-0.01	0.02
	-0.22	-0.5	1	-0.31	-0.52	-0.36	-0.49	-0.4	-0.27	-0.32	-0.61	-0.42	-0.59	-0.31	-0.22	-0.4	-0.25	-0.38	-0.28	-0.18
	0.61	0.2	-0.31	1	0.29	0.33	0.31	0.29	0.24	0.24	0.33	0.3	0.31	0.24	0.28	0.23	0.07	0.08	-0.03	0.02
	0.22	0.01	-0.52	0.29	1	0.63	0.75	0.69	0.52	0.53	0.31	0.54	0.5	0.45	0.41	0.47	0.52	0.49	0.38	0.32
	0.24	0.04	-0.36	0.33	0.63	1	0.54	0.61	0.5	0.47	0.26	0.42	0.41	0.4	0.42	0.45	0.32	0.4	0.24	0.31
	0.21	0.02	-0.49	0.31	0.75	0.54	1	0.72	0.55	0.58	0.37	0.59	0.6	0.54	0.42	0.47	0.49	0.47	0.35	0.31
0	0.17	-0.04	-0.4	0.29	0.69	0.61	0.72	1	0.69	0.66	0.28	0.59	0.68	0.46	0.43	0.65	0.43	0.56	0.3	0.41
1	0.14	-0	-0.27	0.24	0.52	0.5	0.55	0.69	1	0.78	0.31	0.57	0.55	0.54	0.47	0.62	0.47	0.57	0.44	0.46
2	0.2	0.03	-0.32	0.24	0.53	0.47	0.58	0.66	0.78	1	0.31	0.55	0.58	0.65	0.44	0.51	0.53	0.57	0.45	0.43
3	0.29	0.4	-0.61	0.33	0.31	0.26	0.37	0.28	0.31	0.31	1	0.37	0.51	0.4	0.36	0.35	0.29	0.33	0.2	0.24
4	0.18	0.07	-0.42	0.3	0.54	0.42	0.59	0.59	0.57	0.55	0.37	1	0.66	0.55	0.49	0.54	0.43	0.58	0.39	0.45
5	0.2	0.19	-0.59	0.31	0.5	0.41	0.6	0.68	0.55	0.58	0.51	0.66	1	0.54	0.45	0.71	0.44	0.72	0.6	0.44
6	0.24	-0.03	-0.31	0.24	0.45	0.4	0.54	0.46	0.54	0.65	0.4	0.55	0.54	1	0.55	0.44	0.43	0.54	0.44	0.3
7	0.25	0.05	-0.22	0.28	0.41	0.42	0.42	0.43	0.47	0.44	0.36	0.49	0.45	0.55	1	0.42	0.38	0.36	0.36	0.32
8	0.18	0.06	-0.4	0.23	0.47	0.45	0.47	0.65	0.62	0.51	0.35	0.54	0.71	0.44	0.42	1	0.39	0.73	0.45	0.48
9	0.14	-0.04	-0.25	0.07	0.52	0.32	0.49	0.43	0.47	0.53	0.29	0.43	0.44	0.43	0.38	0.39	1	0.62	0.49	0.43
0	0.11	0.01	-0.38	0.08	0.49	0.4	0.47	0.56	0.57	0.57	0.33	0.58	0.72	0.54	0.36	0.73	0.62	1	0.64	0.56
1	-0.03	-0.01	-0.28	-0.03	0.38	0.24	0.35	0.3	0.44	0.45	0.2	0.39	0.6	0.44	0.36	0.45	0.49	0.64	1	0.43
2	0.1	0.02	-0.18	0.02	0.32	0.31	0.31	0.41	0.46	0.43	0.24	0.45	0.44	0.3	0.32	0.48	0.43	0.56	0.43	1
3	0.17	0.02	-0.4	0.21	0.48	0.47	0.5	0.55	0.51	0.45	0.39	0.59	0.75	0.56	0.5	0.64	0.51	0.72	0.56	0.55

Figure 24: Correlation matrix

7.6 Interview Results

The interview questions as mentioned previously, were very similar to the MSQ but instead of rating each questions respondents had to describe their opinion. The interview was conducted after collecting the MSQ data. The reason for this is because the researcher will be able to ask the MSQ that had either the lowest score or highest score so as to get a detailed and quantifiable answer.

Based on figure 17, 18 and 19, the researcher determined that the following questions were the most crucial topics to cover in order to get a response that can help companies make quantifiable changes to better their organizations for the improvement of working environment for employees.

- Are you satisfied with the opportunities for advancement in your workplace? Why?

  The answer for this question was diverse. Of the 15 interviewees, 7 had professional jobs(IT, Healthcare, Teaching), and 8 had non-professional jobs(Catering, translator, janitorial services and waitress). All 8 interviewees with non-professional jobs replied they were not satisfied with the inability to have advancement in the workplace. While 4 of the interviewees with professional jobs replied they were satisfied. On the contrary 3 of the interviewees with professional jobs(teaching) said needs to be improvement in their line of work as there are little to no opportunities for advancement.

8 Discussion

In this section the research sub questions will be answered along with recommendations for specific work industries on how to better interact with their foreign employees.

Based on the interview results and the analysed questionnaire the following answers were determined.

8.1 Answers to research questions

What are the major satisfaction factors according to the respondents?
Referring to figure 18, the major satisfaction factor from the results was for the question 'opportunities to stay busy all day'. This question had an average or 3.8, this question scored mostly neutral, satisfied and very satisfied amongst respondents this raised the average to be higher. The majority of the respondents rating this question very well can mean that employees value staying busy and accomplishing tasks.

Companies can conduct their own survey to get a more in-depth analysis of how these employees can stay busy all day without burning out. The questions presented in this survey is a great way to analyse the current state of employees because as mentioned it covers the three factors that affect job satisfaction. Emphasizing only one question and leaning only towards the positive scale is not efficient for companies. As a result, each question should be given proper emphasis and not just the lowest or highest scored question.

How we can use KNIME to analyse the current level of job satisfaction of foreigners in Finland.?

As mentioned in the thesis KNIME is an easily understandable platform to use by anyone regardless of knowing how to code or not. Each excel can be converted to a node that can then be connected based on the type of feature we are looking for be it association, corelation, including or excluding elements based on ones preference. For this thesis these features were the only ones used but KNIME offers so much more features that can be

utilized to get the best result. Furthermore the interactive and user friendly graphs, charts and visuals help the reader understand the result of the data set.

Researcher has used the column and row filter to filter out irrelevant factors when analyzing the job to education compatibility. The Column row filter was used to delete the row that had the answer 'No' so that the study can be focused on employees that are in the same field as their education but are still not satisfied. Because it is easily arguable that the cause for dissatisfaction for employees that have scored questions low could be because there is no job and education compatibility.

This thesis based its result on the response of 116 employees, if more research was done and more participants were included the result could have an astounding effect on the foreign workforce for the better.

What is the lowest score category and what the reason is?

Although the result is different in every aspect of the respondents answer and is affected by the age, education level and compatibility of job to education. The lowest level of dissatisfaction was for the question 'Compliments for a job well done' which scored an average of 2.7 (Figure 19).

Psychologically humans respond to positive reinforcement and incentives rather than the pressure of being fired. And according to the results(Figure 19) analysed the average of this particular question was much lower than all the other questions. The lack of acknowledgement or encouragement of the company towards its employees has created a major dissatisfaction among them as figure 19 shows. Further research is recommended to be done to understand just how much this factor might have affected employees.

What common factors affect the satisfaction level of employees?

Based in Figure 24 of the correlation matrix, the factors that were redundant or employees had the same feelings towards were, ‘opportunities for advancement’ and ‘job vs education compatibility’. These two questions were highly correlated and although they are not the only factors they are the most common factors.

Employees feel like they are not being given the opportunity for workplace advancement or there is very little effort towards that they feel like it has affected their feelings towards their job negatively.

8.2 Out of scope

This thesis was based on answers received from respondents that are currently employed. It has given great insights to the current level of satisfaction the work force has, but it is also beneficial to conduct this same research with employers or managers.

The reversal of this research will help strengthen the results that were received and give better insight to the reasons of low or high score with in the questionnaire and what can be done to improve these factors.

Furthermore, the result of the research conducted with managers and employers can be imposed with the result gathered from employees. This can determine correlations that have significant impact on the workforce, in all three factors(personal growth, cultural fit and mental and physical health of employees) that would need to be fulfilled to ensure satisfaction in the work place.

8.3 Validity and reliability

This research has utilized MSQ as a main source of data gathering. The MSQ has questions that are complied with the help of psychologist and has been used for decades and is still being used by several companies today. The efficiency is still not being proved as companies use most or all of the question to conduct their satisfaction surveys.

Furthermore the questionnaire uses the satisfaction matrix to cover each negative, positive and neutral feelings employees might have b giving them an extreme choice from very dissatisfied to very satisfied with variations in between of satisfied and dissatisfied and a middle ground for neutral. This further ensures that all the grounds with respect to feelings have been covered.

The thesis questionnaire was suggested by the thesis advisor himself Dr.Amir Dirin. Thesis topic was also suggested by the advisor as there were several thesis done prior to this one.

Based on the above we can say that the thesis holds reliability and validity in terms of study and results.

9 Conclusion

Analysing job satisfaction level of foreigner in Finland was the main goal of this thesis by using KNIME analytical platform. Each data that was collected was reliable in that it was answered by currently employed people.

The job industry is a fickle industry, there are many things affecting the satisfaction levels of the employees, be it personal or job related the decrease of satisfaction level will impact the company negatively in both performance and image. It is advisable that companies conduct job satisfaction survey frequently and keep track of their employees' feelings and determine if it is negative or positive. And based on the result make the necessary changes to better accommodate the employees.

Employee satisfaction is a growing concern for companies these days and it is important that companies focus on it. There are many benefits and downfalls of focusing or neglecting this topic. Great benefits if handled properly and very costly downfalls if not given the necessary emphasis. So every company regardless of its size should have the topic of job satisfaction of employees at its core.

Furthermore, KNIME is a great tool to utilize in a cost-efficient manner to analyse not just the job satisfaction but any line of industry can use it to analyse any kind of data they will need. With the business world digitalizing more and more the inflow of data is growing at a rapid pace, being able to adopt and handle these changes and use them to an advantage is crucial for companies.

In conclusion, this research-based thesis has analysed the gathered data accurately and it can be used to aid other similar topics in the future.

References

Bourne J. 2020. *What is Job Satisfaction and Why is it important?* URL: https://positivepsychology.com/job-satisfaction/ Accessed: 14.05.2020

De La Haye, 2019. *Employee satisfaction Survey: The Complete Guide* URL: https://www.pinpointhq.com/insights/employee-satisfaction-surveys/#what-is-an-employee-satisfaction-survey Accessed: 01.05.2021

Ventakesh, 2019. *Factors Influencing Job satisfaction*. URL: https://www.yourarticlelibrary.com/hrm/jobs/factors-influencing-job-satisfaction-with-diagram/53312 Accessed: 28.04.2021

iEduNote, 20 Factors affecting Job Satisfaction. URL: https://www.iedunote.com/factors-affecting-job-satisfaction Accessed: 17.05.2021

Manikslay S. 2017 Job Satisfaction Theory. URL: https://www.economicsdiscussion.net/human-resource-management/job-satisfaction-theory/31947 Accessed: 23.04.201

iEduNote, 2019. Ways of Measuring Job Satisfaction. URL: https://www.iedunote.com/measuring-job-satisfaction Accessed: 10.04.2021

Nevogt D. 2020. How to Measure and Improve Employee Satisfaction URL: https://blog.hubstaff.com/measuring-employee-satisfaction-with-survey/ Accessed: 06.04.2021

Totah Z. 2019. How to effectively measure employee satisfaction URL: https://www.selecthub.com/hris/measure-employee-satisfaction/#:~:text=Measuring%20employee%20satisfaction%20also%20helps%20determine%20their%20needs,employee%20Net%20Promoter%20Score%20%E2%80%94%20known%20as%20eNPS. Accessed: 17.05.2021

Arora S. 2021. What is Data Analysis, Methods, Techniques and Tools. URL: https://hackr.io/blog/what-is-data-analysis-methods-techniques-tools Accessed: 10.04.2021

Calzon B, 2021. Your Methods Business Guide to Data Analysis Methods and Techniques. URL: https://www.datapine.com/blog/data-analysis-methods-and-techniques/ Accessed: 26.05.2021

DiFranza A, 2021 Predictive Analytics: What it is & Why it's Important URL: https://www.northeastern.edu/graduate/blog/predictive-analytics/ Accessed: 10.04.2021

Statistics Finland 2019. Persons with foreign background URL: Persons with foreign background | Statistics Finland Accessed: 17.04.2021

Vorhies, W. 2016. CRISP-DM- a standard methodology to ensure a good outcome

URL: CRISP-DM – a Standard Methodology to Ensure a Good Outcome - Data Science Central Accessed: 10,05.2021

Crosley J. 2021. What (Exactly) is Thematic Analysis? A plain- Language Explanation & Definition (with examples). URL: https://gradcoach.com/what-is-thematic-analysis/ Accessed: 18.06.2021.