Czech University of Life Sciences Prague

Faculty of Economics and Management

Bachelor Thesis

Assessment of the Human Resources database with the special focus on data science positions

Alla Sumkina

Declaration

I declare that I have worked on my bachelor thesis titled "Assessment of the Human Resources database with the special focus on data science positions" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on _____

Acknowledgement

I would like to thank my supervisor Tomáš Hlavsa for support and advice during my work on this thesis.

Assessment of the Human Resources database with the special focus on data science positions

Abstract

This thesis determines the extent to which factors in the careers of data scientists affect their desire to change a job position. The analysis was conducted using an HR dataset, which is allowed for open usage. The dataset enables us to analyze the relationship between factors, which can be informative for the future research of human resource managers. The thesis demonstrates the potential of using a database of candidates for management research and its usefulness in forming strategic decisions. In the thesis, the appropriate use of the database will be demonstrated, which will contribute to the most focused work of managers. For clarity, the database of data scientists was chosen as the most promising and popular area of the 20th century. Based on the main factor in the database-target, which indicates whether the specialist is currently looking for a job or not, the remaining related factors will be determined by the degree of influence on this main factor. For greater accuracy, statistical methods are used using SAS software. It was determined that the city development factor, relevant experience, university, level of education, major study or studied discipline, work experience, size of the company in which the candidate works, and last new job factor have a relationship with the target.

Such factors like gender, the type of company in which the candidate works, and training hours have no relationship with the target. These are factors, that HR managers should not consider crucial during managerial processes. Determined factors influencing the desire of specialists to change jobs allow managers to make targeted proposals for open vacancies and thus find the most perspective candidates according to the company's goals.

Keywords: Human Resource Management, Chi-square test, relative risk, odds ratio

Table of content

List of pictures

List of formulas

List of tables

Chapter 1

Introduction

Information is accumulating in every sphere in which we interact. Over the past 20 years, information technologies have started to develop and accumulate on the electronic devices that turn them into motionless masses. Almost everywhere, the industry assumes that the database is consuming resources, and the human resource management industry is no exception. Data is a big resource that can be used for the different manipulations that are necessary for development. Information provided by specialists and collected in the database can be beneficial to the company's goals.

The statistical analysis of data is considered from the side of Human Resource Management. Based of Human Resourse Management the importance of data analysis is presented. As it was mentioned before, the nessesarity of data usage is essentially that will be proved in the practical part. We focus at the challenge of using auxiliary data in the estimation and design stages in this thesis.

It is recommended to the readers who would like to understant the main principles of Human Resourse Management functioning to read Literature review in which main principles, meanings and work proccesses are desciped based on theoretical resourses. The information presented in these chapter can be preparable for the whole understanding of the thesis. Moreover, the information about data scientists and the role of HR in data science the reader can read in article „HR management in data science sphere".

This thesis is based on database analysis. The data in the set is related to data science candidates. To test the factors, one main factor was chosen, which is called "target." Other factors, which indicate or do not indicate a dependence in relation to the main factor, were also chosen. The analysis is considered based on the human resource management point of view to present how statistical analysis is necessary for managers and the company as well.

Dealing with categorical data, the relevant statistical tests will be applied. Moreover, all data is going to visually presented for better visual understanding and evaluation.

With the help of statistical tests, the factors that influence the desire of specialists to make changes in their career will be indicated. In the thesis, the importance of data analysis for HR management will be shown, which proves the necessity of data application and analysis for every company.

1 Objectives and Methodology

1.1 Objectives

The thesis has one main goal: to determine what makes data scientists want to change jobs. Based on the data obtained as a result of the analysis, human resources specialists will make decisions about inviting certain candidates who are open to vacancies. Each characteristic will be considered in relation to the main factor (target). It is assumed that each factor in the

dataset is associated with a target, with a factor that represents information about whether the data scientist is currently looking for a job or not. Found consequences are the essential resource for the successful being of the organization. Hence, the database of data science positions will be considered and valued from the human source managers' point of view. In the project, the information based on the dataset will be assessed and consequences will be made that will be useful for work of HR specialists. Thus, the importance of data analysis for managers will be presented.

1.2 Methodology

This thesis is primarily based on factors analysis. The data for this research is based on a database of human resource managers, which includes a high number of variables and personnel. Human capital measurement has been defined by IDS (2004) as being 'about finding links, correlations and, ideally, causation, between different sets of (HR) data, using statistical techniques' [2]. That is why, statistical approaches are able to find patterns in the data and solve particular tasks of Human Resource department. Dealing with categorical variables, the Chi-sqaure test will be applied. This method allows to identify which of the factors affect the factors that candidates are ready to change jobs. Thus, we will be able to achieve the goal with regard to assessing the database in terms of target. This will allow for the most detailed analysis of the HR database of candidates in the data science field. Furthermore, SAS studio software will be applied for analysis, allowing for massive datasets to be managed. Data visualization will be used for increased clarity, allowing data relations to be presented in the most obvious way possible.

The null hypothesis assumes that variables have no relationship with each other. The alternative hypothesis is that it is assumed that variables have a relationship. Alternative hypotheses should be accepted in case the null hypothesis is rejected. The decision rule is:
If the P-value is less than alpha, the null hypothesis is rejected. In this case, the alternative hypothesis is accepted.
If the P-value is greater than alpha, the null hypothesis is accepted.

As it was mentioned, the Chi-square test was chosen for categorical data analysis in this thesis. Let O_i be the observed number of observations in category i. Also, let E_i be the expected number of observations in category i if the population were normal with the same mean and standard deviation as in the sample.
The chi-square test for independence is based on the counts in a contingency (or cross-tabs) table. It tests whether the row variable is probabilistically independent of the column variable. The idea of the test is to compare the actual counts in the table with what would be expected under independence. If the actual counts are sufficiently far from the expected counts, the null hypothesis of independence can be rejected.
Here, O_{ij} is the actual count in cell i,j (row i, column j), E_{ij} is the expected count for this cell assuming independence, and the sum is over all cells in the table. [15]
Then, the test statistic is:

$$\text{chi-square test statistic} = \sum_{ij}(O_{ij} - E_{ij})^2/E_{ij}$$

(1)

Formula 1. Chi-square test statistic

[15]
Chi-square test statistic for every factor, as well as other values and test results, will be presented in the practical part. To determine the result of the tests, we will need to compare the alpha and P-value for each factor. It should be noted that the results will be obtained using SAS Studio and interpreted as well.

For a more detailed assessment of factors, relative risk and odds ratio will be applied to see if the presence of a risk factor or intervention changes the risk of an outcome compared to the absence of a risk factor or intervention.

"Odds" and "Risk" are the most frequent terminology which are used as measures of relationship between variables. "Risk" refers to the probability of occurrence of an event or outcome. Statistically, risk is equal to chance of the outcome of interest to all possible outcomes, while "odds" refers to the probability of occurrence of an event to probability of the event not occurring.

The relative risk compares the frequency of an outcome between groups of factors. To expain the formulas assume, n_{11} is a value in the first row and the first column, n_{10} is a value in the first row and the second column, n_{01} is a value in the second row and the first column and n_{00} is a value in the second row and the second column. The risk of the outcome occurring in the group placed in first row is

$n_{11}/n_{11}+n_{10}$ (2.1)

and the risk in the group placed in the second row is

$n_{01}/n_{01}+n_{00}$ (2.2)

The relative risk is the ratio of the probability of the outcome occurring in the first row group versus a second row group is:

$n_{11}/n_{11}+n_{10} / n_{01}/n_{01}+n_{00}$ (2.3)

[16]
Formula 2. Relative risk

It can be easily shown that the odds ratio is a good approximation to the relative risk when the incidence or risk rate is low, for instance, in rare diseases, and can largely overestimate the relative risk when the outcome is common in the study population.

The odds ratio is the ratio of the odds of outcome occurring in the first row group to the odds of it occurring in the second row group. The odds of outcome in the first row group is n_{11} n_{10} and the odds of outcome in the second row group is n_{01} n_{00}. The odds ratio thus becomes

$\theta(n_{01}, n_{11}) = n_{11} n_{00}/n_{10} n_{01}$ (3)

[16]
Formula 3. Odds ratio

These values will allow a clearer understanding and evaluation of groups of variables and their comparison with each other, which will allow HR managers to better understand the risks and probabilities regarding the desire of candidates to be open to new career offers.

Chapter 2

Literature Review

2.1 The Meaning of Human Resources Management

Companies are managed by people and through people. It is common knowledge that without people, organizations can never exist. Indeed, people who make up the "human resources" of a company are able to influence the performance of an organization, depending on their level of commitment, professional skills, contribution and cooperation. It is quite relevant to know that the human resources managers are cooperating with are a significant part of the success of the company. For instance, Leon C. Megginson discusses about human resource management following: "From the national point of view, human resources are knowledge, skills and attitude obtained in the population; while for individual enterprises, they represent the total of the abilities, knowledge and skills of their employees" [3]. The 'human resource' is composed of all the efforts, skills and capabilities of all the people who work for that organization [4]. First of all, talking about human resource management, we should consider the "human resources" as material the managers work with. HRM must also develop mechanisms that will help multicultural individuals work together [5]. The term 'HRM' can be divided into three parts. The first sub-field of micro-HRM focuses on the functional aspects of HRM (e.g. the employment lifecycle such as on HR practices of HR planning, recruitment and selection, induction and socialisation, training and development, performance management, rewards and remuneration, managing employment relationships). International HRM sub-field deals with the management of people in global and multinational enterprises. Issues of local responsiveness or integration of the parent firm's global best-practices at its subsidiary locations, better coordination and control are among the key areas of emphasis in managing people. Finally, the last sub-field of HRM research adopts a strategic approach to HRM (or SHRM for short) [6].
The terms 'human resource management' (HRM) and 'human resources' (HR) have largely replaced the term 'personnel management' as a description of the processes involved in managing people in organizations[2].

2.2 Roots of HR Management

History of HRM dates back to BC. Entrepreneurs have recognized the importance of physically relocating managers to foreign locations where business operations are based since approximately 1900 BC. Even at this stage, locals were viewed as inferior and restricted to lower level jobs while parent country nationals (PCNs) were charged with

running foreign operations and afforded superior conditions, similar to modern day expatriates [1].

The field further developed with the arrival of the Industrial Revolution in the latter part of the 18 th century, which laid the basis for a new and complex industrial society. In simple terms, the Industrial Revolution began with the substitution of steam power and machinery for time-consuming hand labor. Working conditions, social patterns, and the division of labor were significantly altered. A new kind of employee—a boss, who wasn't necessarily the owner, as had usually been the case in the past—became a power broker in the new factory system [7].

The field further developed with the arrival of the Industrial Revolution in the latter part of the 18 th century, which laid the basis for a new and complex industrial society. In simple terms, the Industrial Revolution began with the substitution of steam power and machinery for time-consuming hand labor [7].

There is no specific date assigned to the appearance of the first personnel department, but around the 1920s more and more organizations seemed to take note of and do something about the conflict between employees and management [7].

Change in organizations focused on the group-level rather than the individual by ensuring equitable recruitment, retention and basic support structures for minority individuals. The 'discrimination-and-fairness' paradigm emerged from the Civil Rights movements of the 1960s in the United States. This led to the introduction of legislation which made it unlawful for organizations to discriminate on the basis of race, colour, religion, sex and origin [1] This was based on moral and ethical values of fair treatment and compensatory fairness for historically disadvantaged groups in business and the society.

Later on, various approaches focused on enhancing the working environment started to appear. The studies pointed out the importance of the social interaction and work group on output and satisfaction. The human relations movement eventually, around the mid-1960s, became a branch of and a contributor to the field of organizational behavior [7].

But beginning with the turbulence in society and organizations starting in the 1970s, we have seen an increase in "new" employment relationships, and a decline in "traditional" or long-term employment relationships. The current employment scene has elements of both the "traditional" and the "new" (or changed) relationship [8]. Since the development of human resource management (HRM) as a field of scientific research in the 1980s, many changes have taken place in this area. An important development has been the integration of HRM into the strategic management process [1]. It emphasized the new role of strategic planning in HRM. In the 1990s, there was a primary emphasis on deriving HR strategies from the master strategic plan of the organization. After 2000, there has been more interest in integrating closely into the primary strategy, logic dictating that the informational input from HR could signifi cantly improve or rather change the overall approach [1].

2.3 Modern understanding of HRM

In the last decade, there has been a realization amongst managers that good plans, effective organizational structures, impressive plants and modern technologies do not guarantee the

success of the organization unless they hire the right specialists, train them suitably, appraise them

properly, and constantly motivate them to get the desired results. The advent of the information age represents both a challenge and an opportunity for knowledge and competence management.

New forms of knowledge extraction and expert location are deeply impacting companies around the world [9].

In fact, technological progress has had a significant impact on the nature of HR work. Traditional kinds of jobs have become technologically challenging because of rapid technological progress. The impetus for the HRM change comes from recognition of recent developments in the HRM profession and a realization that current practices do not reflect those changes, especially those concerning IT strategies. The problem often results in policies, practices, and strategies that may be outdated [10]. This progress requires the labor of more educated and more skilled employees.

The emergence of new technologies, competitive pressures from an everchanging work environment, globalization of the world economy, and organizational changes, such as workforce diversity, delayering, outsourcing, offshoring, teamwork, networks, and the entry of Generation Xers in the workplace, are factors that call for a redefinition of the concept of a career [10].

Formalized HR policies can be used in induction, team leader and management training to help participants understand the philosophies and values of the organization, and how they are expected to behave within that context [2]

To be successful HR specialists must play a role of analytic, so it leads to changings in work of HR manager. By the way, HR specialist must be able to operate different programs to work with huge databases, because the forecasts cannot be made with absolute certainty, reasonable estimates for the future requirements can be made [4].

Moreover, network information helps management to identify critical and central groups for negotiation and consensus building [10].

This modern understanding of HR objectives is focused on the fact that the company should provide equal opportunities to all its employees and all qualified applicants for employment. The equal opportunity policy should spell out the organization's determination to give equal opportunities to all, irrespective of sex, race, creed, disability, age or marital status. The policy should also deal with the extent to which the organization wants to take 'affirmative action' to redress imbalances between numbers employed according to sex or race, or to differences in the levels of qualifications and skills they have achieved [2]. Their primary goal is to ensure that policies and procedures are enforced properly, and to permit a wronged employee a forum to obtain relief. As part of this role, too, comes the disciplinary process. These representatives see that appropriate disciplinary sanctions are used consistently throughout the organization [5].

2.4 Features of Human Resource Management

HRM performs different characteristics which is neccessary for the company. The defining characteristic of HCM is this use of metrics to guide an approach to managing people that regards them as assets and emphasizes that competitive advantage is achieved by strategic investments in those assets through employee engagement and retention, talent management and learning and development programmes [2].

To achieve its objective, management typically requires the coordination of several vital components that we call functions [5].

The Human Resources Effectiveness Diagram (HRED) is a model created by Aileen G. Zaballero. It was designed to provide HR leaders with an objective, comprehensive, yet easily applied assessment of both the effectiveness and relevance of their HR goals [11].

Figure 1. HRED Model

[11]

The HRED is an approach of modern human resource management. It can be used in multiple scenarios utilizing a strategic, operational, or transactional mindset [11]. HRED consists of five part which are performed in a form of frames on the Figure 1.:

1. Recruitment and Staffing,
2. Talent Management and Development,
3. Work/Performance Analysis and Rewards
4. Employee Relations and Communications
5. Measuring Organization Effectiveness

[11].

Each part of the HRED model contains papers and forms. The purpose of each white paper

is to provide an authoritative document that helps frame a specific concept. The goal is to educate and support HR professionals and practitioners to make better decisions.

The purpose of each form, the intended audience, and suggested uses are presented. Each form addresses a critical business process for an HR function. Forms can be used as is or modified. While each form reflects a good working model that can be used immediately, we encourage readers to modify forms as needed or desired [11].

Through the job analysis process, HRM identifies the essential qualifications for a particular job and includes them in the job description [5]. HRM aims to support the development of firm-specific knowledge and skills that are the result of organizational learning processes [2].

Indeed, HRM is aimed at several functions. As part of the strategic HR responsibility, HR assists decision makers in evaluating where the organization currently stands, deciding where the organization wants to be in the future, developing a plan to achieve those goals, implementing the plan, and checking progress toward those goals [5].

2.5 Scope of HRM

Nowadays, the scope of HRM is widespread. An understanding of human resource management is essential for anyone whose work is related to the organization. HR issues become important wherever there is a group of workers. Staffing is performed by all the managers, either directly or indirectly through the HR department. So, all managers are, in this way, HR managers since they get involved in HR activities like choosing, training, inducting, compensating and motivating the employees along with industrial relations activities [3]. Thus, they must understand the scope and application of the personnel policies of their organization in order to ensure that their everyday personnel actions are consistent with those policies as any violation of such policies may get them into confrontations with their subordinates. Further, knowledge of the basics of HRM is important even to non-managerial employees as they may be keen to know the impact of the personnel policies of their organization on their own compensation, training and career growth aspects. Thus, the personnel aspects of management run through the entire organization. HR professionals must assist in managing the multiple change processes involved and also help those involved in the process to gather whatever learning is available as effectively as possible.

The focus is on the *relationship builder* and *knowledge facilitator* roles of HR professionals in developing a climate of trust, innovation, and learning [10].

2.5.1 Job Analysis Process

Job analysis is the process of breaking down a specific job into parts and scrutinizing each of them to gather the necessary information. It requires a systematic and efficient examination of the tasks, duties, responsibilities and accountability of a job [3]. All Job Analysis steps are visually presented in the Figure 2.
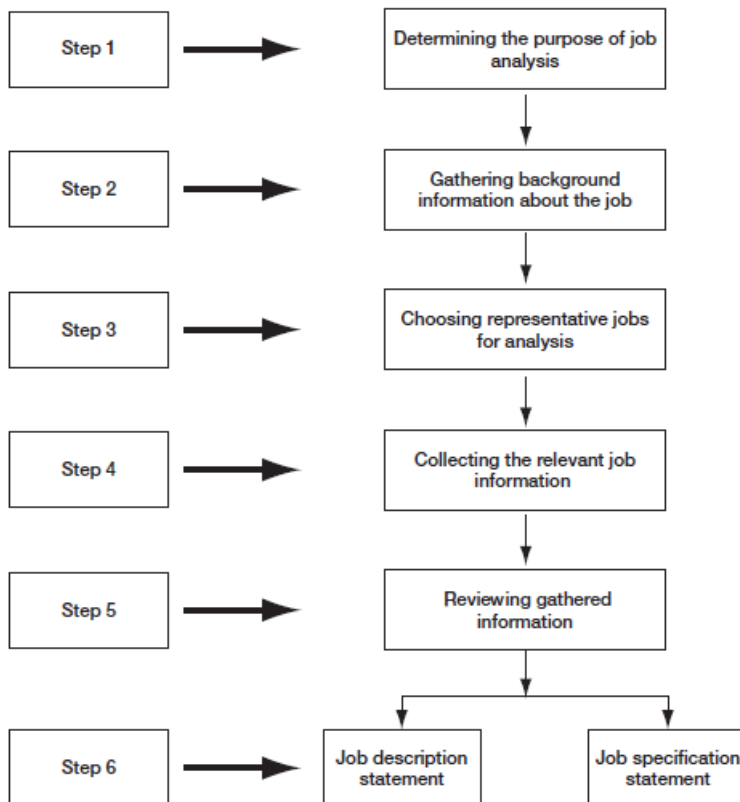
Figure 2. Steps in Job Analysis process

[3]

The first step of the job analysis process is to define the purpose of the future job analysis. True, job analysis has relevance for almost all HR activities of an organization. Depending upon the purpose of the analysis, the organization may focus specifically on certain aspects of the job that are considered important [3].

More than that, it is necessary for an organization to gain the background information about the job to determine its relative importance for the organization. While doing so, the job analysis should focus on identifying information relating to all the important elements of the job [3].

Often, organizations choose only a few jobs for analysis instead of analysing all of them [3] because many vacancies have similar characteristics.

Information on various aspects of the job is gathered in the fourth stage. Specifically, the information about the duties, level of responsibility, authority, accountability, content and context of the job, desired employee behaviour, critical knowledge, knowledge skills and abilities (KSA), and training requirements are gathered [3]. These details may be obtained by contacting current and former employees of the job, as well as supervisors and managers.

At the next stage, the collected data is needed to carefully analyzed and then reported. Generally, the specific job holders and their immediate supervisors are included in such an analysis process. The opinion of the employee performing the job is ascertained to verify the correctness and completeness of the job analysis report. In case the employee or supervisor points out incompleteness or discrepancies in the report, supplementary information could be collected [3].

This is the final stage consists of two parts: job description statement and job specification statement. The two essential documents of the firm, namely, the job description and job **specification**, are prepared on the basis of the job analysis report. A job description is a

written statement containing information about the duties, responsibilities, accountabilities, working conditions and risks associated with the job [3].

2.6 Techniques of Data Collection for Job Analysis

When you are planning the data collection for an evaluation it is usual to
try to obtain a range of different types of data [12]. It does not matter what the methodology applied for data collection and future analysis, the focus should be on gathering information about the tasks associated with the job and the desired human behavior at work [3]. However, the characteristics of the job, the purpose of the job analysis, and the quality of information are the factors that usually influence the decisions for analysis. Choosing a specific method for gathering data, the right care should be taken to ensure that the chosen method generates sufficient information about the major job requirements and the skills, knowledge, and ability necessary for performance. It might sometimes become difficult for a job analyst (an expert in job analysis) to observe job-related behaviors that are abstract in nature.
Collection of data for analysis is usually done at one or more steps. These levels are: individual, group, organizational, and community. A specific level depends on the nature of job information required and the characteristics of the job position. Besides, selection of a specific technique is based on specific factors such as: type of data for collection, the methods of data collection, the source used for data collection, and the process of data analysis. Methods of data collection are various.

2.6.1 Method of Questionnaire

In this method, a worker is suggested to fill out a questionnaire, because employees "are part of the job analysis process [3]. Some organizations apply structured questionnaires with direct and predetermined questions. Others use a structured checklist method. This method notes the responses in a predetermined statement form. A few other companies use open questionnaires to collect more detailed answers for each job studied. The main advantage of a questionnaire method is that it is the easiest and most efficient method to collect information from a large number of employees within a short period. The information collected in this manner will be used to determine the relative contribution on all jobs in the company on a consistent basis [11]. However, its main disadvantage is the time consuming and high cost of its preparation.

2.6.2 The Job Performance Method

In the job performance method, the job is completed by the HR himself to find out the different requirements of working positions. For example, there is considered information like the nature of the job, the level of difficulty, the level of concentration needed. High-performance working involves the development of a number of interrelated processes which together make an impact on the performance of the firm through its people in such areas as productivity, quality, levels of customer service, growth, profits and, ultimately, the delivery

of increased shareholder value [2]. Nevertheless, the main disadvantage of this method is that it may not be adopted for certain types of jobs.

2.6.3 Observation Method

In this method, the HR analyst observes the performance of the employee in the job. Observation means studying role holders at work, noting what they do, how they do it, and how much time it takes. This method is most appropriate for routine administrative or manual roles, but it is seldom used because of the time it takes [2]. The manager records the various tasks performed by the employee as part of the required job. In fact, this method is ideal for gathering first-hand information relating to working conditions, the physical activities involved in the job, the normal time required to complete the job, and so on. It is ideal for manual jobs completed within a short period. It is usually free from the information [3]. However, the observation method can increase the effectiveness of the job analysis if it is combined with other methods like interviews or questionnaires. In the observation method, HR must take in account the following:

• The observation must be made during the whole work cycle.
• The characteristics and needs of the job alone should be recorded and not the personal traits or behavior of the employees not related to the work position.
• The job analyst should just observe the working process with any intervention.
• The observation must be done during normal conditions; the behavior of employees and their activities should be recorded under normal conditions.

2.6.4 The Critical Incident Method

A critical episode is a substantial occasion that make an employee react in the form of performance or specific behavior. The critical-incident technique is a means of eliciting data about effective or less effective behaviour that is related to examples of actual events – critical incidents [2]. The aim of that method is to gain the relevant description of the job influenced by behavior. The episode and the response are valued based on the actions, effects, processes and conclusions resulting from the incident. The behavioral feedback response to the episode can be or not become part of the job, depending on performance effectiveness. The main drawback of this method is the difficulty in converting the behavioral feedback into practical information for making conclusions. Hearing and handling constructive feedback can be a powerful development tool, and the results and analysis of exit interviews can provide relevant and useful data directly into the organization development processes [11].

2.6.5 Method of Interview

In this method, the employee is asked questions about different aspects of the job. For data collection, the different types of interview are applied. The aim of the interview is to obtain

all the relevant facts about the role to provide the information required for a role profile. It is helpful to use a checklist when conducting the interview [2]. They can be in the form of either structured or unstructured interviews. A structured type of interview has a definite format and questions, while an unstructured interview has no predetermined questions. However, a structured interview format allows for comparing the gathered information. But, the open questions in an interview can be helpful in gathering the relevant information. It is also believed that the interview offers the greatest value as a selection device in determining an applicant's organizational fit, level of motivation, and interpersonal skills [5]. When preparing the interview questions, HR must ensure that the questions are not difficult to understand by the employees. Moreover, the HR must present these questions in a simple, clear manner to the employees. HR should also provide clarification about the questions whenever the employees need it. The main disadvantage of this method is its cost and time-consuming. It can be very time-consuming to analyse data from interviews and observations, but these approaches often collect very relevant data [12]. And, finally, the interviewer may require sufficient training and experience for successfully conducting a job analysis interview [3].

2.7 Recruitment and selection processes

One of the most important parts is hiring new employees is known as the recruitment. It is known that the term "recruitment" refers to the process of filling of vacancies in an organization [4]. According to Konopaske, recruitment is directly related to both human resource planning and selection. In addition, recruiting often represents the first contact between organizations and prospective employees. As such, care should be taken to create a positive first impression with these job applications. Consider recruitment and selection practices and the crucial role played by HR specialists [7].

Recruitment and staffing are responsible for identifying and providing candidates for an organization's open positions [11]. The recruitment process includes job analysis which is necessary for the determination of employment conditions. The HR prepares the clear job description that gives the detailed information about the job and performs the worker's duties with complete understanding. The HR department is able to decide whether the candidate is fit: physically, mentally, skills and experience for the vacancy he applied for. Recruitment is also refed to the first stages in the process of filling vacancies in an organization [4].

For some positions, certain tests might be used before the interview. The HR department with existing specialists can prepare work tests to check candidates' skills and aptitude tests to show their skills in doing simple tasks. Besides, the candidates can be asked to do intelligence tests for measurement of reasoning ability or personality tests to find out "the possession or lack of certain character traits [3].

It is quite common for new people to be offered to work a "trial period" before getting the position. If an employee is offered and accepts suitable alternative work within the company, it will take effect without a break from the previous employment and will be confirmed in writing. If the offer is refused, the employee may forfeit his or her redundancy payment. Employees will receive appropriate training and will be entitled to a four-week trial period

to see if the work is suitable. This trial period may be extended by mutual agreement to provide additional training. During this period, employees are free to terminate their employment and if they do, would be treated as if they had been made redundant on the day the old job ended. They would then receive any redundancy pay to which they are entitled [2].

At the end of the probationary period, the new employee can be called to the HR department where his progress can be discussed. If both sides are agreed, the employee is "confirmed" in the post, otherwise he/she should terminate the position in the company. The final stage is introductions to the work team. People and teams are always crucial as they can make the project succeed or fail [12]. Induction is such a stressful process for new comer, so HR manager should arrange this procedure with more responsibility to reduce stressful situations for all parties.

Successful candidates who get the positions can face adaptation or "induction". It is established that induction is meant to induct a new employee into the new social settings of his work [4]. This process is a duty of the HR manager and should be organized productively for the effective interaction among all employees. In some companies, induction includes any number of stages. In the first stage, the HR department gives general information about the organization. In the second stage, the supervisor of the department where the employee is going to work informs them about duties, responsibilities, and communication processes with colleagues.

2.8 The Scope of Training

Nowadays, the scope of training programs is wide. Training is the use of systematic and planned instruction activities to promote learning. The approach can be summarized in the phrase 'learner-based training'. It involves the use of formal processes to impart knowledge and help people to acquire the skills necessary for them to perform their jobs satisfactorily. It is described as one of several responses an organization can undertake to promote learning [12]. All training programs are produced by the organizations. Training is usually focused to ensure that each individual has the skills and knowledge necessary to enable him or her to perform effectively in his or her job [12]. One more important objective of HRM department is to find a trainer. Selecting a trainer that has a limited understanding of learning, training delivery, individual motivation, and training techniques is doomed to fail. As a cautionary suggestion, a review of the trainer's understanding of theory, methods, and delivery should precede placing anyone in a trainer position [7].

To avoid any misunderstanding or unproductive work, the promotional courses are established. The amount of training that can be offered in a project depends on the length of the project and the amount of training that an individual needs to be able to complete the tasks required. Training is not the answer to everything but is often important in bringing performance up to the required level [12]. A lot of companies have higher requirements for certain positions, so employees who would like to get a promotion should complete prepared promotional training to be promoted.

The trainee would then be provided with feedback that demonstrates their strengths and weaknesses on the subject matter, and be provided with references that would allow them to review areas where they are deficient [10].

Training imparts a job-related knowledge to the participants and enables them to understand what they must do in their job and how they should [3]. Individual differences among trainees affect the learning process. First of all, different people learn at different rates, a fact that should be considered in designing training programs. Second, people with different characteristics (such as intelligence levels or various learning styles) may learn best using different training approaches. And third, contrary to many stereotypes, older adults can learn as well as younger adults, but they do learn differently [8]. Indeed, organizations should accumulate its knowledge to employ them for future impovements. In the short term, they need the ability to bring as much organizational knowledge, wisdom, and experience to bear on business challenges as possible, and faster than ever. In the long term, and this is where corporate universities come in, organizations need to determine gaps in their knowledge competencies and work diligently to close those gaps [10].

2.8.1 Knowledge

Knowledge workers are defined as workers whose skills or knowledge are inextricably linked with the product or service of their employing organizations [2]. Knowledge of training design issues—in particular, the conditions of practice—should also be used to maximize learning. These conditions include active practice, massed versus spaced practice sessions, whole versus part learning, overlearning, knowledge of results, and task sequencing [8]. The main purpose of any training program is to provide the participants with the knowledge necessary to achieve the goals of the job. Knowledge is generally developed through the processes of perception, learning and thinking. Training includes knowledge related to a job and provides the concepts for understanding what they must do in their job and what methods they can apply to do their work successfully.

2.8.2 Skills Acquisition

A skill is the ability to find solutions to some problems. New management practices must focus on combining understanding, knowledge, skills and attitudes when assembling work teams; analyzing requirements for performing work and establishing new work settings are a huge component of tacit knowledge, or knowledge that is mainly embedded in people [9]. It is developed by persistent training and application of other practices. Employees may require one or more skills such as social skills, technical, intellectual skills, management skills, motor skills, mental skills or may be perceptual skills. In general training, employees learn those skills that are necessary to do the jobs effectively and to use their knowledge productively. Requests and contributions are combined to determine the competence created in terms of personal, relational, technical and professional skills [9].

Ethics commonly refers to a set of rules or principles that define right and wrong conduct [5]. Ethics are the principles of right and wrong that are accepted by an individual or a social group [3]. Ethics have to do with what we believe is right or wrong. Taken together, values and ethics encompass the dimension of morality, which is sometimes sadly lacking in modern management [11]. The management of human resources often involves ethics for decision-making like fair treatment and justice. Organizations should provide familiarity with ethical policies as well as the necessary training in ethical practices. can help employees understand the presence of ethical dilemmas in situations which require decision-making and learn the techniques for using ethical codes to settle problems involving ethical dilemmas in a fair and just manner. Moreover, it must ensure the adoption of ethical values in every aspect of cooperation with the people.

2.9 Human Resources Planning

The first step in the process of recruitment is the forming of human resources plans by the HR department. Panning in human resource management is concerned with forecasting and estimating the future demand for labor by an organization [4]. HR-planning can be divided into quantitative and qualitative HR-planning, followed by the planning of personal actions. The whole process is known as the planning part of human resource management (HRM). Quantitative HR-planning deals with the number of employees or headcount (HDCNT) and the full time equivalent (FTE) in relation to time and space. Qualitative HR-planning in ERP-systems planning is about the employees' occupational aptitude, depending on time and space as well [9]. So, the first step is the estimation of the future HR requirements of the company and the decision to achieve the goals of the company through recruitment.

Analytical thinking refers to the systematic way of thinking to realize problems, develop alternative plans of action, choose the best way of action, and implement the selected plan to successfully solve the problems. It is possible that the organization needs more skilled line managers who are able to carry out training needs analysis before any decisions are made about exactly what sort of training is needed [12]. Employees benefit from continuous training programs that provide them with the necessary knowledge for analytical thinking as well as the skills to apply methods for problem solving.

2.10 Measurement Tools to Evaluate HR Programs

The initial weakness of HRM in many organizations is the absence of correct and effective tools to measure the performance of HR programs. In most cases, top managers expect HR managers to express the results of their work in measurable formats, so HR managers' focus in the future will be on the development and introduction of reliable and accurate HR tools to measure the efficiency of HR functions [3]. Organizational effectiveness includes all performance measurement at any level and the overall corporate culture of the organization. The performance measurement and human resource metrics in this part of the book will enable an HR department to analyze its own level of effectiveness against agreed-on human resource metrics [11]. Obviously, HR managers should be focused on the fact that organizations will require the introduction of reliable and accurate HR tools and

knowledgeable HR specialists who will be able to perform the relevant analytics which are necessary for company effectiveness.

2.11 The Future Role of HRM

Obviously, time of technologies plays the main role in HRM development. The data managers accumulate, will be essential resource of company evolution. But the data would be very useful, as it can be fully integrated into the data model without great efforts, because pre-dated notifications of personal actions are quite similar to existing personal actions. On the contrary the information about pre-dated notifications is used in reports and planning processes [9].

Effective HRM is concerned with helping an organization achieve its objectives in the future by providing for competent, well-motivated employees. Thus, human resources need to be incorporated into an organization's long-term strategic plans [7]. Thus, HR managers have to think not only about the traditional activities like planning, welfare measures and internal relations. Such forecasting often relies on a combination of quantitative and/or qualitative techniques. From an operational perspective, HR planners consider a number of useful data points: determine past, current and future trends in workforce movements in an organization [6].

With employee self-services and delegated HR-planning a new approach has been started, which will lead into a decrease of personnel maintaining personal actions and master data. These administrative tasks will be taken over by the employees themselves or the mid-level management [9].

2.12 HR Management in the Data Science Sphere

Data scientists are passionate about data and finding creative ways to solve problems and portray information [13]. Data Science provides ways to deal with and benefit from Big Data: to see patterns, to discover relationships, and to make sense of stunningly varied images and information [13]. Data Science projects tend to use many types of data sources, including large or unconventional datasets [13]. Data science encompasses a variety of disciplines and areas that enable us to provide an accurate, objective, and rational examination of raw data. Thus, companies looking for people who can work with complex data have had good luck recruiting among those with educational and work backgrounds in the physical or social sciences [14]. Considering the difficulty of finding and keeping data scientists, one would think that a good strategy would involve hiring them as consultants. Most consulting firms have yet to assemble many of them [14].

Data scientists generally have five main sets of essential skills which are shown in Figure 3:

- Quantitative skill: such as mathematics or statistics
- Technical aptitude: namely, software engineering, machine learning, and programming skills
- Skeptical mind-set and critical thinking: It is important that data scientists can examine their work critically rather than in a one-sided way.

- Curious and creative: Data scientists are passionate about data and finding creative ways to solve
- problems and portray information.
- Communicative and collaborative: Data scientists must be able to articulate the business value
- in a clear way and collaboratively work with other groups, including project sponsors and key stakeholders.

[13].

Figure 3. Profile of a Data Scientist

[13]

Data science is relatively new field, so responsibilities and roles are not always set clear, nowadays. Data scientists tend to be more motivated, too, when more is expected of them. The challenges of accessing and structuring big data sometimes leave little time or energy for sophisticated analytics involving prediction or optimization. Yet if executives make it clear that simple reports are not enough, data scientists will devote more effort to advanced analytics. Big data shouldn't equal "small math." [14]

Working in the Data Science sphere, an HR manager must know the whole process of working with those specialists. Even as the ranks of data scientists swell, competition for top talent will remain fierce. Expect candidates to size up employment opportunities on the basis of how interesting the big data challenges are [14].

Given that today's most qualified prospects come from nonbusiness backgrounds, hiring managers may need to figure out how to paint an exciting picture of the potential for breakthroughs that their problems offer [14].

If companies sit out this trend's early days for lack of talent, they risk falling behind as competitors and channel partners gain nearly unassailable advantages. Think of big data as an epic wave gathering now, starting to crest. If you want to catch it, you need people who can surf [14]. HR managers' focus in the future will be on the development and introduction of reliable and accurate HR tools to measure the efficiency of HR functions [3].

Chapter 3

Practical Part

This research used hypothesis testing to collect data that signified a relationship between the dependent variable, which is called the target, and the set of independent variables. The data set consists of 11 independent variables, which will be evaluated in order to check their relations according to the target variable.

The rest of this chapter is organized as follows. First, we will explain the data and methods that will be applied. Second, we present our method to visually understand the correlation and interaction between dependent and independent variables. Some interesting interactions between review attributes that arise in this step suggest the potential for trivial as well as non-trivial correlations. Finally, we present our results, which are able to identify the factors which can influence wish of the candidates to change jobs.

3.1 Data description

For the data analysis, the HR dataset, which consists of 8956 candidates, was chosen from the open data source Kaggle.com. Our goal, as was mentioned before, is based on an analysis of the factors that make data scientists desire to change their job position. The variable, which is called "target," is determined as a binary variable, which is explained as 0, which means that the candidate is not looking for a job now, and 1, which means that the candidate is looking for a job change. The target variable is presented more clearly in the following table:

Main variable	Description of main variable
Target	0 – *Not looking for job change*
	1 – *Looking for a job change*

Table 1. Main variable description

For the clear analysis of the main variable, the set of factors was chosen, and each of these factors will be statistically analysed to determine the most statistically significant related to the wish of the specialists to make the changes in their career. All dependances will be resolved through independent variables chosen from the dataset. A description associated with every factor is presented in the Table 2:

| Variables for analysis	Description of variables	Unit
City development index	Development index of the city (scaled)	
Gender	Gender of candidate	
Relevant experience	Relevant experience of candidate	
Enrolled university	University course enrolled or no enrollment	
Education level	Education level of candidate	
Major discipline	Education major discipline of candidate	
Experience	Candidate total experience	years
Company size	Number of employees in current company of a candidate	employees
Company type	Type of current employer's company	
Last new job	Difference in years between previous job and current job	years
Training hours	Training hours completed	hours

Table 2. Description of variables
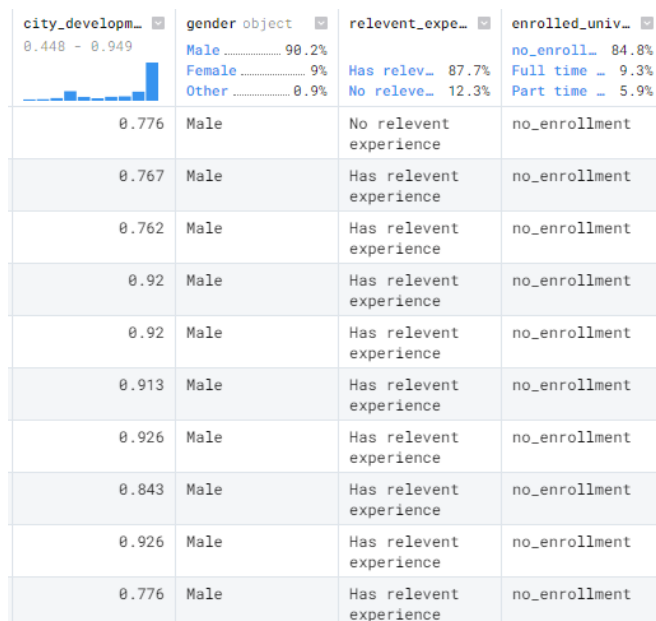
Figure 4. Deepnote- factors presentation

The factors will be observed separately. Figure 4 shows the first four factors visually. The city development index is defined as a range from zero to one. The most developed cities have a value close to one, while the least developed cities have a value close to zero. Also in Figure 4, we can see factors such as gender and the relative experience of the candidates. In addition, Figure 4 shows the enrollment factor, which indicates whether the candidate is studying at the university or not. We can observe that university students are divided into full-time and part-time groups. These factors will then be combined as "enrolled" to apply to the test. Another important factor is the level of education of the candidates, as shown in Figure 5. The definition of a degree indicates the bachelor's education of the candidates, which, as we can see, is the main share. In addition, the data in Figure 5 makes it possible to consider what discipline data scientists are studying or have studied. And as you can see, the majority of applicants, which is 89.2%, are studying or have studied STEM (Science, Technology, Engineering, and Mathematics). In addition, we can evaluate the experience of candidates. The experience of candidates over the years can be considered, as well as the size of the company in terms of the number of employees where a certain candidate works. We can also consider the type of company the candidate works for, in order to then consider the impact of this factor on the target. In Figure 5, we also see the last new job factor. This factor is the gap in years between the end of the previous work and the beginning of a new one. Moreover, we can then look at the number of training hours of candidates in the direction of data science. And finally, the relative target, to which all previous variables will be considered for the presence of relationships. A more detailed explanation of the target factor is given in Table 1.
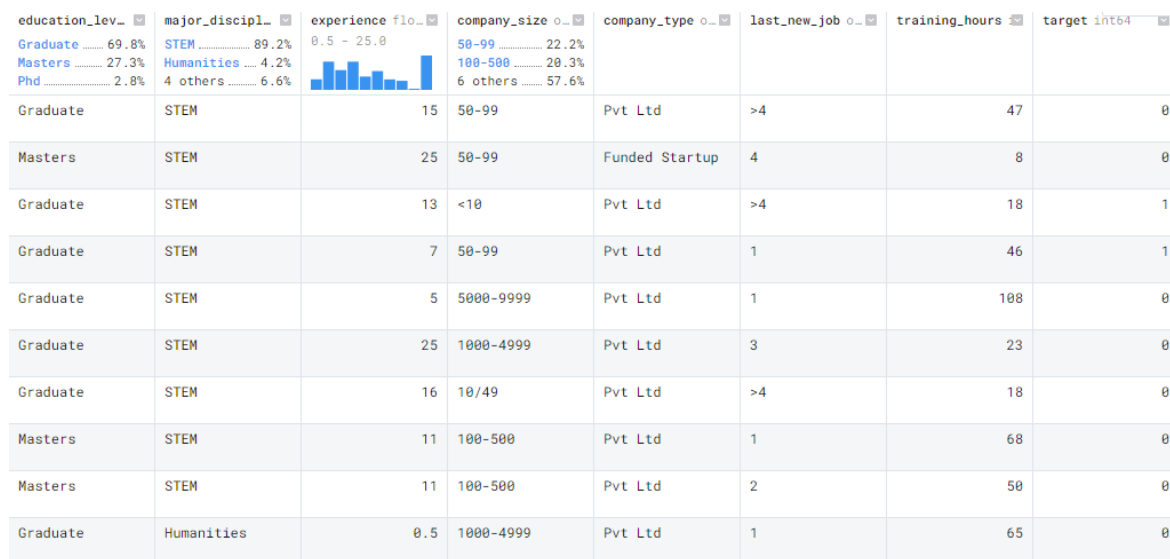
education_lev…	major_discipl…	experience flo…	company_size o…	company_type o…	last_new_job o…	training_hours	target int64
Graduate …… 69.8% Masters …… 27.3% Phd …… 2.8%	STEM …… 89.2% Humanities …… 4.2% 4 others …… 6.6%	0.5 - 25.0	50-99 …… 22.2% 100-500 …… 20.3% 6 others …… 57.6%				
Graduate	STEM	15	50-99	Pvt Ltd	>4	47	0
Masters	STEM	25	50-99	Funded Startup	4	8	0
Graduate	STEM	13	<10	Pvt Ltd	>4	18	1
Graduate	STEM	7	50-99	Pvt Ltd	1	46	1
Graduate	STEM	5	5000-9999	Pvt Ltd	1	108	0
Graduate	STEM	25	1000-4999	Pvt Ltd	3	23	0
Graduate	STEM	16	10/49	Pvt Ltd	>4	18	0
Masters	STEM	11	100-500	Pvt Ltd	1	68	0
Masters	STEM	11	100-500	Pvt Ltd	2	50	0
Graduate	Humanities	0.5	1000-4999	Pvt Ltd	1	65	0

Figure 5. Deepnote- factors presentation

From the following page, the tests will be given their interpretation and also, with the help of SAS, the variables in relation to the target will be clearly demonstrated.

3.2 Data analysis

Section 4.2 will apply tests to determine the relationship between each of the 11 factors and the target. The chi-squared test, relative risk, and odds ratio were chosen as the methodologies, the results of which will also be interpreted.
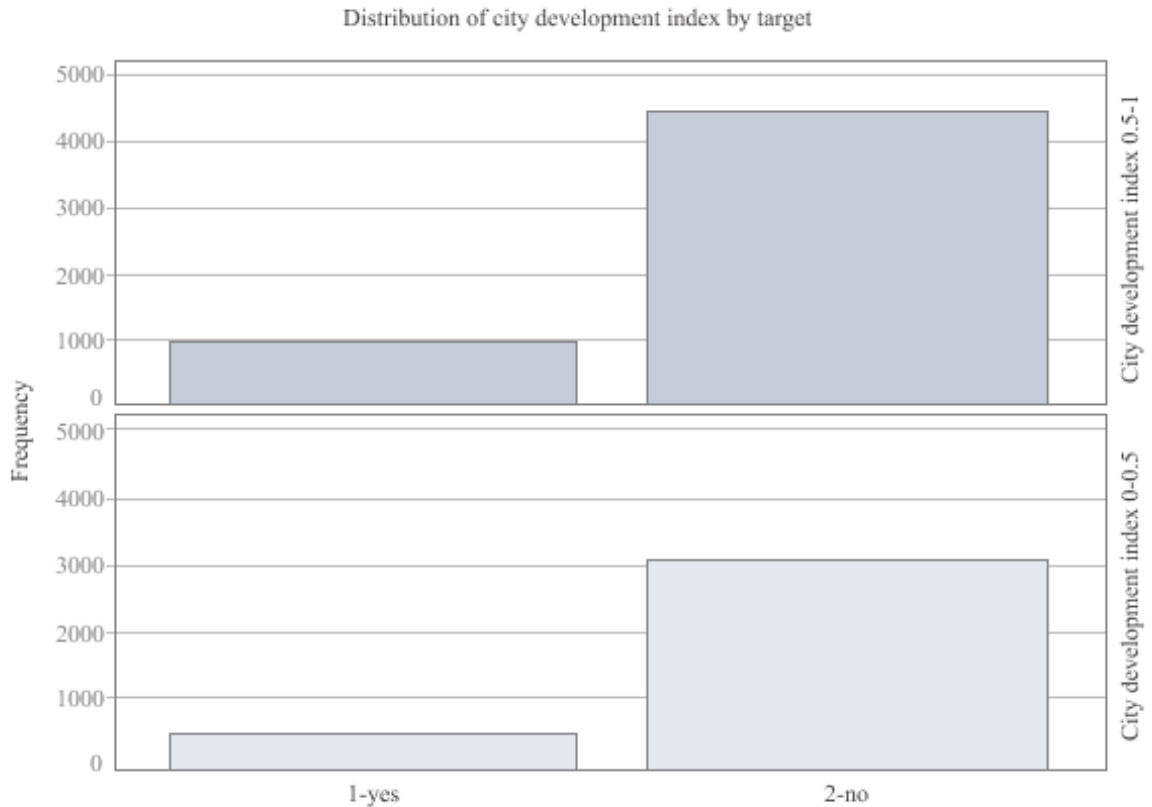

Figure 6. SAS- city development index

3.2.1 Analysis of city development index factor

In the Figure 6, the distribution of the city development index related to the target is presented. The city development index is determined as an interval from 0 to 1, where indexes closer to 1 indicate the most highly developed cities. The "city development index" factor was divided into two groups: candidates with an index interval of 0 to 0.5 and candidates with an index interval of 0.5 to 1. For the relationship between variables determination, the Chi-square test is applied.

A Chi-square test is used in order to find out if there is any relationship between the variables. For making the table, the "city development index" factor was divided into two groups: candidates with an index interval of 0 to 0.5 and candidates with an index interval of 0.5 to 1. The null hypothesis assumes that the city development index of specialists and the target are not related to each other, while the alternative hypothesis is based on the assumption that there is a relationship between the city development index of specialists and the target. The gained Chi-square value is equal to 30.0640 and the P-value is equal to 0.0001. 0.05. Comparing this value with the gained P-value value, we see that the P-value is less than

alpha. So, our null hypothesis is rejected. That means that there is a relationship between the city development index of specialists and their wish to make changes in career. Relative risk for the first column shows that there is a 1.3156 times greater risk that data scientists from more developed cities are looking for a job change, while relative risk for the second column represents a 0.9489 times less risk that data science specialists from more developed cities are not looking for a new job. As for odds ratio, gained value determines that there is a 1.3857 times greater chance that specialists from developed cities will be more interested in career changes than specialists from less developed cities. Thus, the "city development index" is a factor that should be accepted during recruiting processes in the HR department.
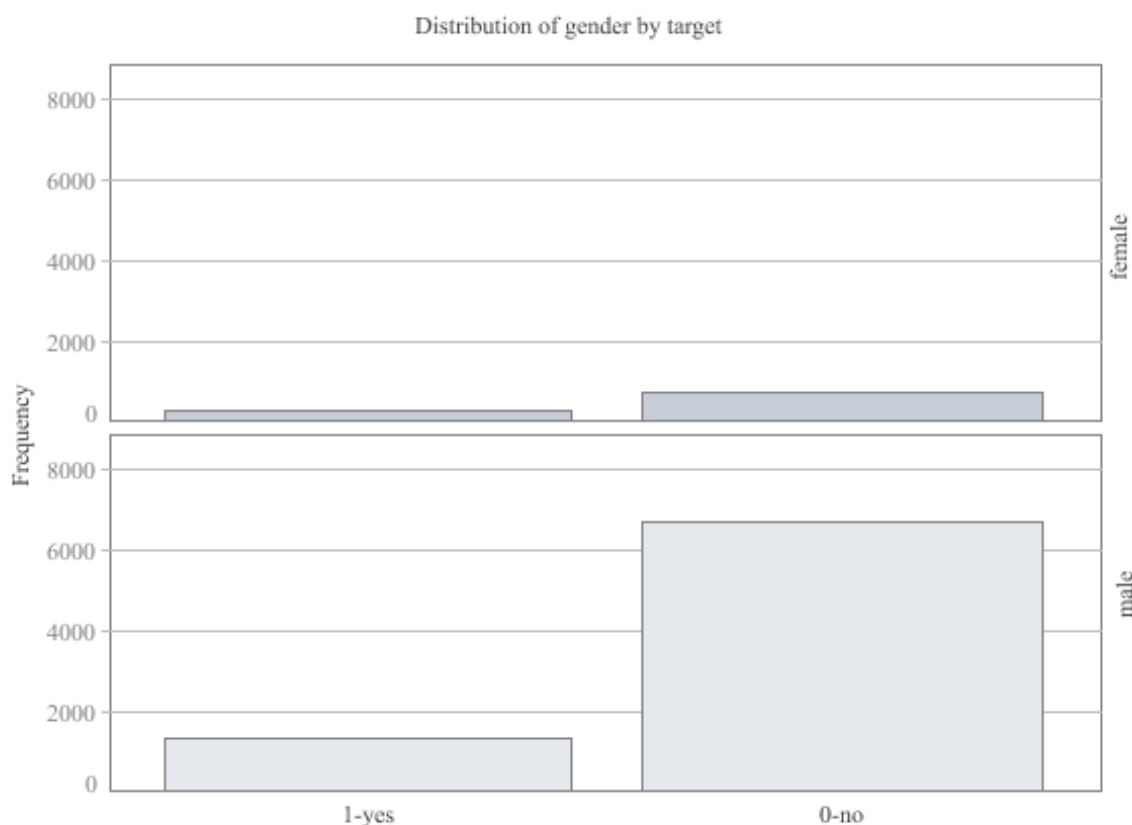

Figure 7. SAS- gender

3.2.2 Analysis of gender factor

In the Figure 7 proportions based on gender and related to target are performed. The majority of candidates in the both groups who is not looking for a job and who is looking for a job. According the data on the pie chart the minority of females in the data sector is presented.

A chi-square test will be applied to determine whether there is any relationship between the target and the gender of the candidates. The total number is 8955, which means that the Chi-square test can be applied. The zero hypothesis assumes that there is no relationship between target and the gender of data scientists. According to the results of the test, the Chi-square value is equal to 0.0492. The Alpha number that was assumed was 0.05. The P-value is 0.8245. We can compare those values: $0.8245 > 0.05$, so alpha is less than the P-value, which

means that zero hypothesis is accepted. So, the target and the gender of the candidates have no relationship. The odds ratio is 0.9780, which means that the risk of the female is looking for a job is 0.9780 times less than male. Relative risk for the first column is 0.9816, so it is 0.9816 times less risk that a female is looking for a job. For the second column, the relative risk is 1.0037, so there is a higher risk that a female is not looking for a job. Thus, the HR department should pay attention to male candidates during the selection process.
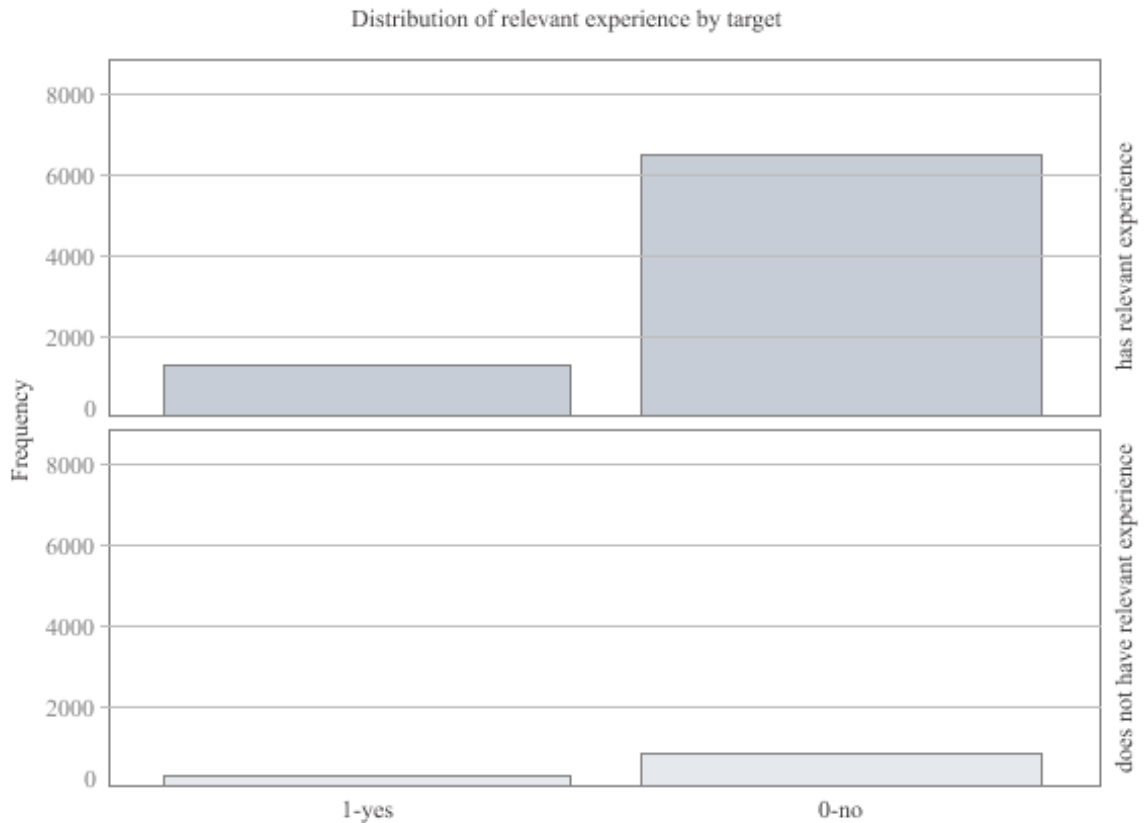

Figure 8. SAS- relevant experience

3.2.3 Analysis of relevant experience factor

The following diagram shows the distribution of the target according to the relevant experience of specialists. As it can be seen, the proportions of candidates who are and those who are not looking for a job in both groups are not similar. We can see unequal distribution among those who has relevant experience but does not want to change their career, to prove it statistically, turn to tests and values.

A Chi-square test is applied to determine whether there is any relationship between the target and the relevant experience of the candidates. The total number is 8877, which means that the Chi-square test can be applied. The Hull hypothesis that there is no relationship between the target and the relevant experience of the data scientists. In terms of the alternative hypothesis, it was assumed that there was no relationship between the target and the relevant experience of data scientists. According to the results of the test, the Chi-square value is equal to 12.0657. The alpha value that was assumed was 0.05 and the P-value is equal to

31

0.0005. We can compare those values. 0.0005 is less than 0.05, which means that zero hypothesis cannot be accepted as accepted. So, the target and the relevant experience of the candidates have a relationship with each other. Considering relative risk for the first column, we can say that those candidates who have relevant experience are 0.7945 times less likely to be looking for a job. As for the relative risk for the second column, it is 1.0520 times more candidates with relevant experience are not looking for a job. Then, the determined odds ratio shows that 0.7552 times less candidates with relevant experience will be uninterested in job changes than data scientists who lack relevant experience.
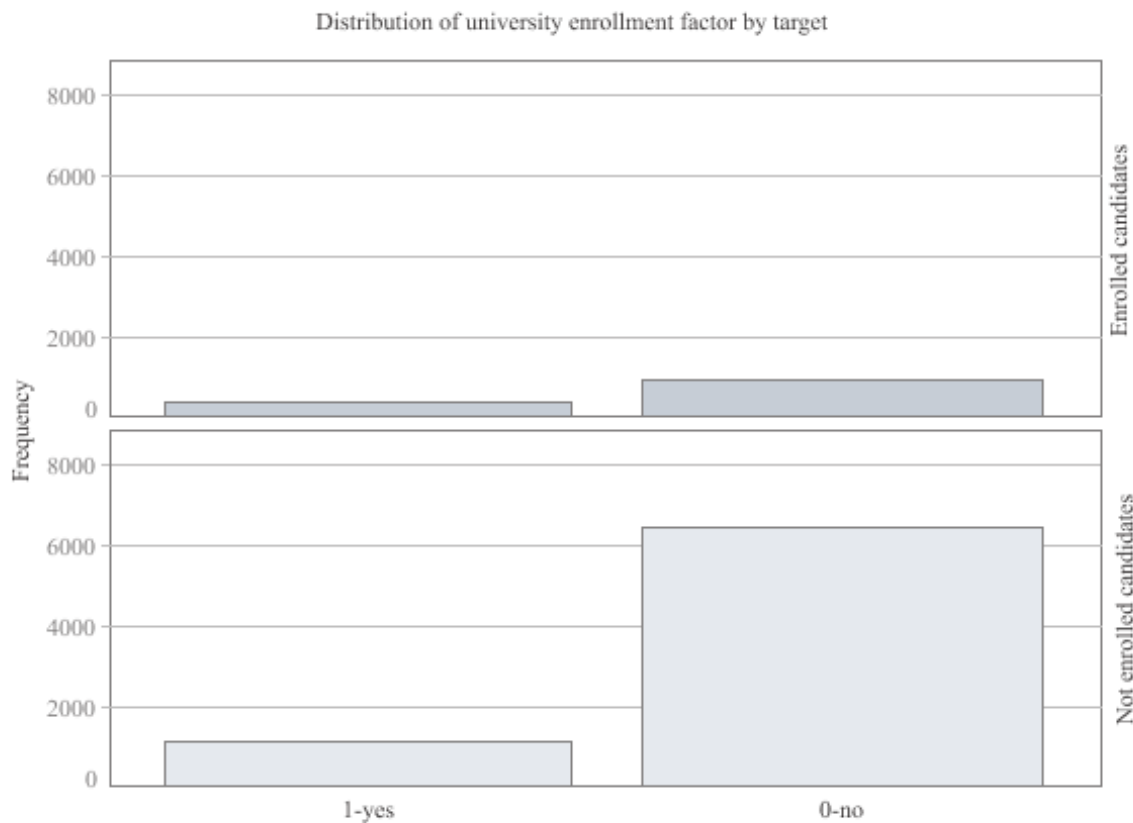

Figure 9. SAS- enrolled university

3.2.4 Analysis of enrolled university factor

The Figure 9 presents the distribution of the university enrollment status of candidates according to target characteristic. The university enrollment factor indicates whether the candidate is studying at a university or not. Those who study at the university in Figure 9 are designated as enrolled candidates. Those who are not studying, that is, already have a degree, are determined as not-enrolled candidates. Nevertheless, to find out if there is any relationship, statistical tests must be provided.

In the form of a null hypothesis, it was supposed that there was no relationship between the enrolled university factor and the target factor. In contrast, it was supposed in the form of the alternative hypothesis that enrolled university factor and target factor have a relationship. The Chi-square value that was gained in the Chi-square value provided in SAS is equal to 49.2335, and our accepted significance level is 0.05. The P-value is equal to 0.001. While

comparing those values, we found that the alpha value is less than the P-value, which means that we cannot accept the null hypothesis. So, the enrolled university factor and the target factor have a relationship.

In terms of odds ratio, candidates who are enrolled in university have a 1.6483 times greater chance of looking for a job change. Relative risk for the first column represents the 1.4987 times higher risk that enrolled in the university, candidates are more likely to want to change their job. Relative risk for the second column shows that unenrolled candidates are 0.9093 times less likely to be looking for a new job. That means that HR managers should select new employees among the group of students.

Figure 10. SAS- education level

3.2.5 Analysis of education level factor

The following diagram shows the distribution of the education level factor among candidates to target. Based on the data in the diagram, we can notice that many specialists who would like to change their position have graduated. Besides, we can notice that there is an almost equal distribution of each category in every target group.

A chi-square test is provided to determine whether there is any relationship between the target and the education level of the candidates. The sample size is greater than 50, so we have a reason to apply the Chi-square test. To make a two-by-two table for providing the Chi-square test, the "master" and "PhD" sectors were combined into "nongraduated". The Hull hypothesis assumes that there is no relationship between target and education level of

data scientists, while the alternative hypothesis assumes that there is a relationship between target and education level of data scientists. According to the results of the test, the Chi-square value is equal to 12.5359. The alpha number that was assumed to be 0.05 has a P-value of 0.0019. We can compare those values: 0.0019<0.05, so alpha is greater than the P-value, which means that zero hypothesis is rejected. So, the target and education level of the candidates have a relationship with each other. Relative risk for the first column is 1.2014, so it is relatively more risk that bachelor candidates are looking for a job 1.2014 times more than "master" and "PhD" candidates. For the second column, the relative risk is 0.9649, so there is a lower risk that master and doctoral candidates are less interested in changing jobs than bachelor and college candidates. The odds ratio value, which is 1.2514, means that a group of graduated candidates are 1.2514 times more likely to like to change jobs than "master" and "PhD" specialists.
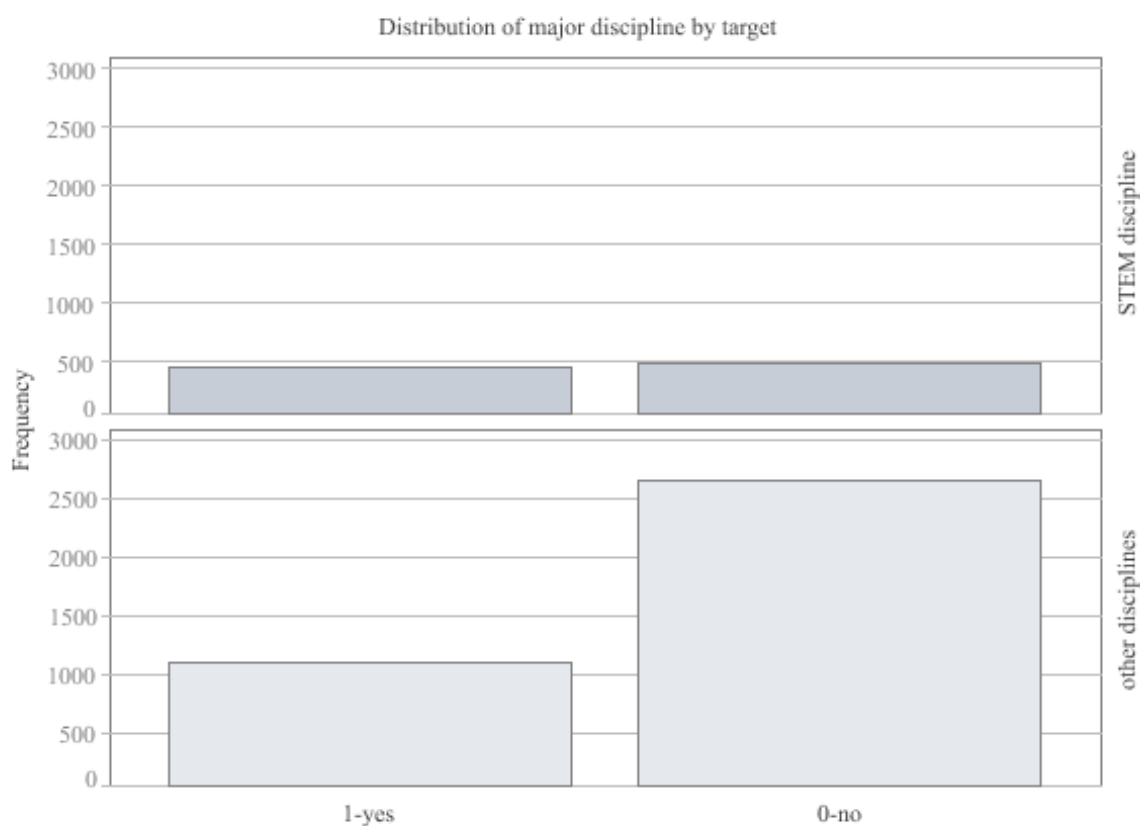

Figure 11. SAS- major discipline

3.2.6 Analysis of major discipline factor

According to the diagram above, the tendency of STEM (Science, Technology, Engineering, and Mathematics) specialists in data science is clearly reflected. The STEM candidates who are looking for a job are major group of the total candidates in that group.

A chi-square test is provided for the determination of the relationship between the target and the major discipline of the candidates. The sample size is greater than 50, so we have a reason to apply the Chi-square test. To make a two-by-two table for providing a Chi-square test, we

divided this factor into two groups: STEM discipline and other kinds of major discipline. The Hull hypothesis assumes that there is no relationship between the target and the major study discipline of data scientists, while the alternative hypothesis assumes that there is a relationship between the target and the major study discipline of data scientists. According to the results of the test, the Chi square value is equal to 123.3022. The Alpha number that was assumed was 0.05. The P-value gained is 0.0001. We can compare those values: $0.0001 < 0.05$, so alpha is greater than the P-value, which means that null hypothesis cannot be accepted. So, the target and major study discipline of data scientists have a relationship with each other. Relative risk for the first column value shows that 1.6033 times more STEM candidates will be interested in changes in their career. Relative risk for the second column represents 0.9272 times less risk that STEM candidates are not looking for a job. In terms of odds ratio, candidates with major disciplines in Science, Technology, Engineering, and Mathematics are 1.7292 times more likely to be interested in career changes than candidates with other major disciplines.
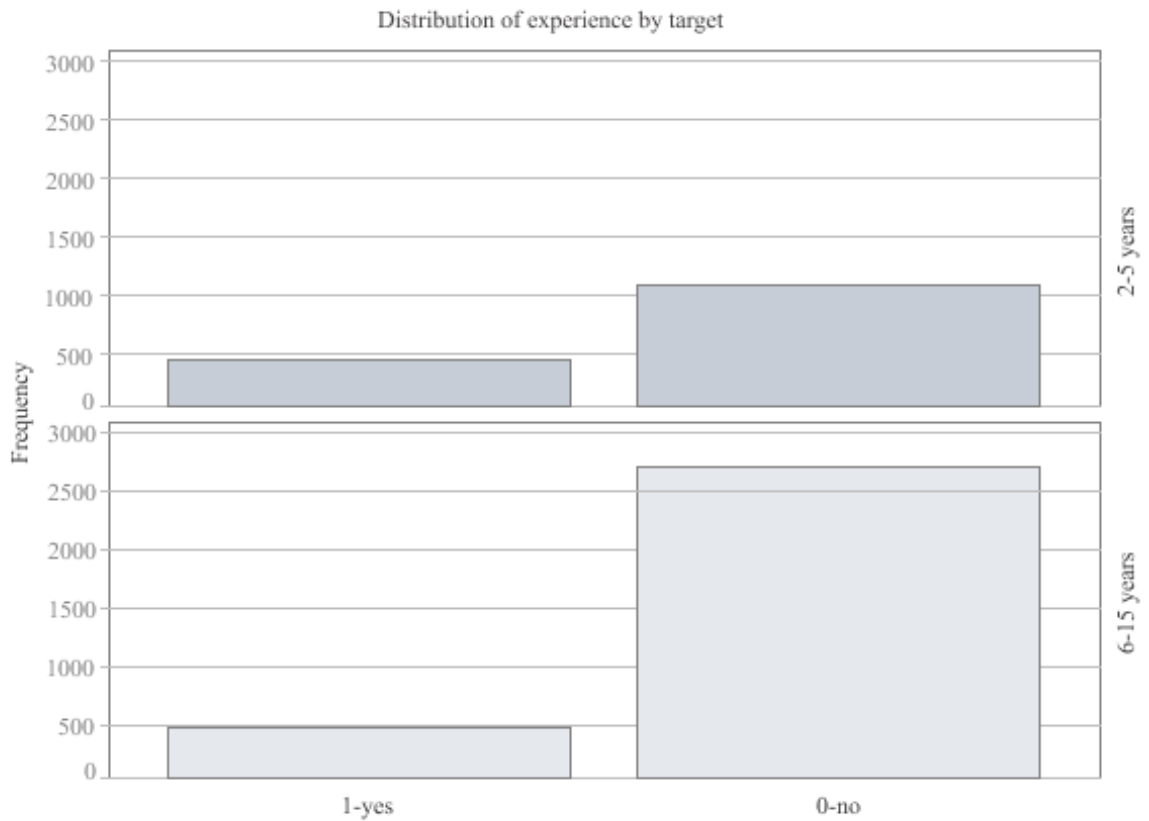

Figure 12. SAS- experience

3.2.7 Analysis of experience factor

The diagram above represents the distribution of experience (in years) by target. The diagram is divided into two groups: 2-5 and 6-15 years of experience. The distribution of this factor according to the target we can see visually. Obviously, more experienced specialists are uninterested in changes in their career.

In the form of a null hypothesis, it was supposed that there was no relationship between the experience of candidates and the target factor. In contrast, it was supposed in the form of the alternative hypothesis that experience factor and target factor have a relationship. The chi-square value that was gained in the Chi-square test provided in SAS is equal to 123.3022, and our accepted significance level is 0.05. The P-value is 0.0001, which means the alpha value is greater than the P-value , so we need to reject the null hypothesis. So, experience and the target factor have a relationship. That means that HR specialists should pay attention to the experience of candidates and select candidates according to the experience the company needs. According to the relative risk for the first column, it is proven that there is a 1.6535 times positive risk that candidates with 2–5 years of experience are more likely to change jobs. Relative risk for the second column shows that candidates with 6–15 years of experience are 0.7313 less likely to change jobs. The odds ratio represents that there is a 2.2612 times greater chance that candidates with 2–5 years of experience will be more interested in career changes than candidates with 6–15 years of experience. So, experience of the candidates is an important factor which HR managers should take in account during selection processes.
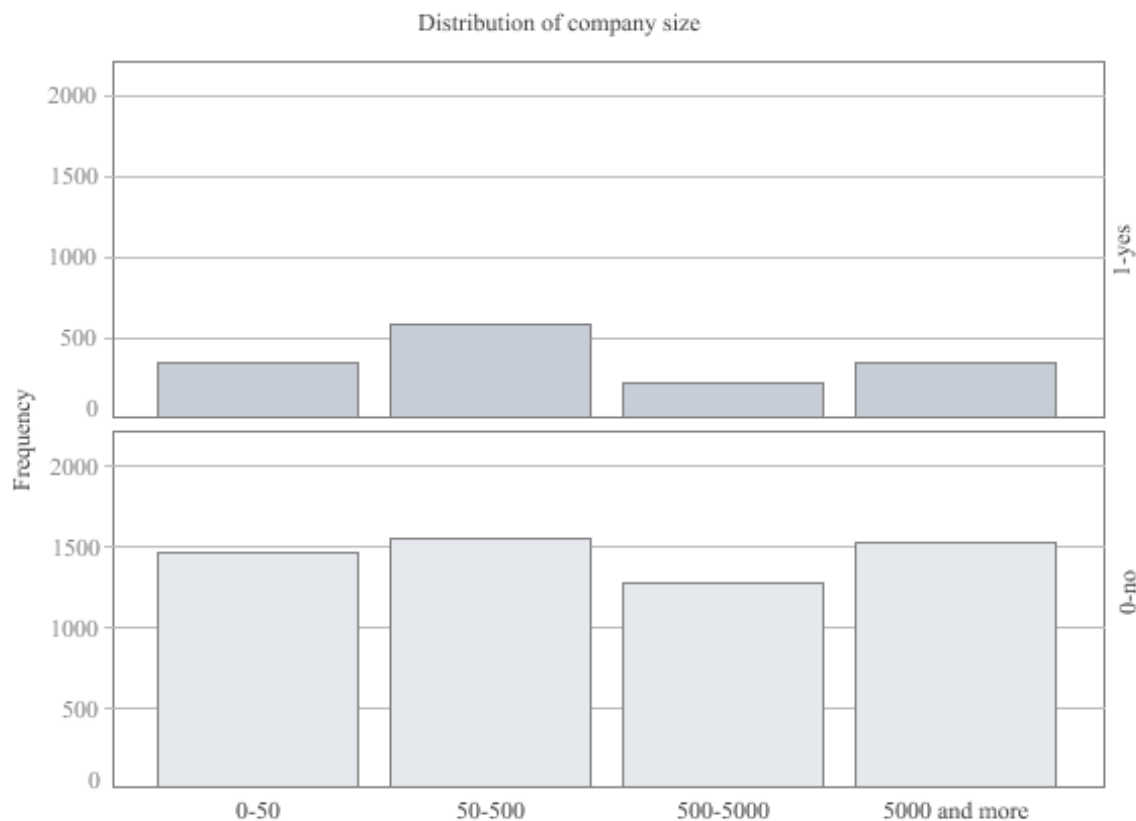

Figure 13. SAS- company size

3.2.8 Analysis of company size factor

The diagram above represents the distribution of company size the specialists are working now by target. The top part of diagram shows the company size of candidates who are looking for a job, while bottom part of the diagram shows the company size of candidates who are not looking for a job. The columns are divided into four part by quantity of the

employees in the companies: 0-50 employees, 50-500 employees, 500-5000 employees, 5000 employees and more.

A chi-square test is provided for the determination of the relationship between the target and the company size within which the candidates are working. The sample size is greater than 50, which means that the chi-square test can be applied. To make a contigency table for providing the chi-square test, we divided this factor into four groups according to employee quantity: 0–50, 50–500, 500–5000, and more than 5000 employees group. The Hull hypothesis, which is based on the assumption that there is no relationship between target and company size where the data scientists are working, while the alternative hypothesis is based on the assumption that there is a relationship between target and company size when the data scientists are working. According to the results of the test, the chi square value is equal to 130.3861. The alpha number that was assumed was 0.05 and the p-value was 0.001. We can compare those values, so alpha is greater than the P-value, which means that the zero hypothesis is rejected. So, the target and company size factors have a relationship with each other, and company size is one more factor that HR managers should take in account.

Figure 14. SAS- company type

3.2.9 Analysis of company type factor

On the diagram above, we can see the distribution of candidates from Pvt. Ltd. (Proprietary Limited Company) companies, which is on the left side of the diagram, and candidates who are working in other types of companies, which is on the right side of the diagram. A

statistical test should be provided in order to find out the relationship between the company type of the data scientists and their wish to change careers.

In the form of a null hypothesis, it was supposed that there was no relationship between company type, the candidates' work, and the target factor. In contrast, it was supposed in the form of the alternative hypothesis that company type and target factor have a relationship. The Chi-square value that was gained in the Chi-square test provided in SAS is equal to 0.2737, and our accepted significance level is 0.05. The gained P-value is 0.6009. While comparing those values, we found that the alpha value is less than the P-value, which means that we need to accept the null hypothesis. So, company type and target factor have no relationship. Relative risk for the first column shows that there is 1.0297 times more risk that candidates from proprietary limited companies are looking for a job. Relative risk for the second column shows that there is 0.9943 less risk that candidates from proprietary limited companies are not looking for a job. And finally, the odd ration represents that candidates from proprietary limited companies are 1.0356 times more likely to change their positions than candidates from other types of companies.
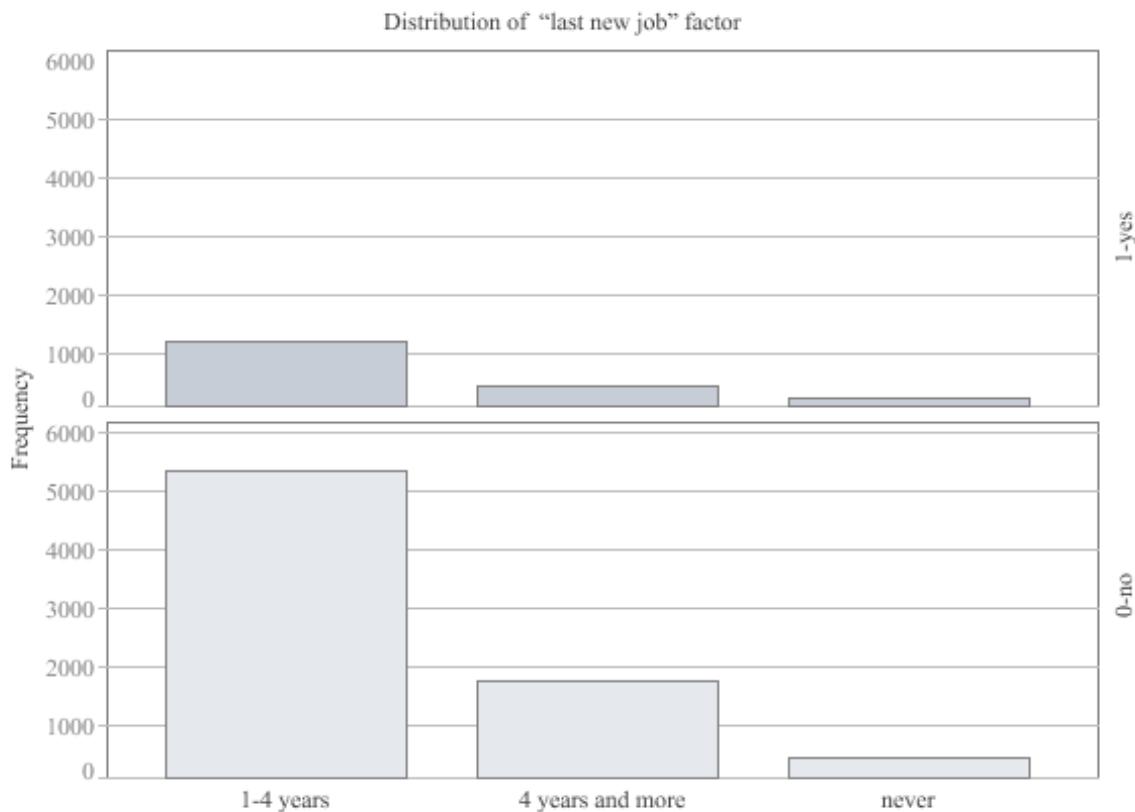

Figure 15. SAS- last new job

3.2.10 Analysis of last new job factor

"Last new job" criteria shows how many years are left from finishing the last position and starting a new one. "Last new job" criteria is divided into three groups such as 1-4 years gap, 4 years or more gap and "never", which means that the candidates have never had a gap between their previous and current position.

A chi-square test can be used for the determination of the relationship between the target and the "last new job" factor. The sample size is more than 50, which means that the Chi-square test can be applied. To make a tree by two table for applying the Chi-square test, we divided this factor into two groups according to employee quantity: 1–4-year gap, 4–6-year gap, and group "never", which means that the candidates have never had a gap between their previous and current position. The Hull hypothesis assumes that there is no relationship between the target and the "last new job" factor, while the alternative hypothesis states that there is a relationship between the target and the "last new job" factor. According to the results of the test, the chi-square value is equal to 66.4464. The Alpha number that was assumed was 0.05. The P-value is equal to 0.0001. Those values can be compared, so if alpha is less than the P-value, that means that zero hypothesis is rejected. So, the target and the "last new job" factor have a relationship with each other. That means that a gap between the previous and current position is an important factor for HR management processes.
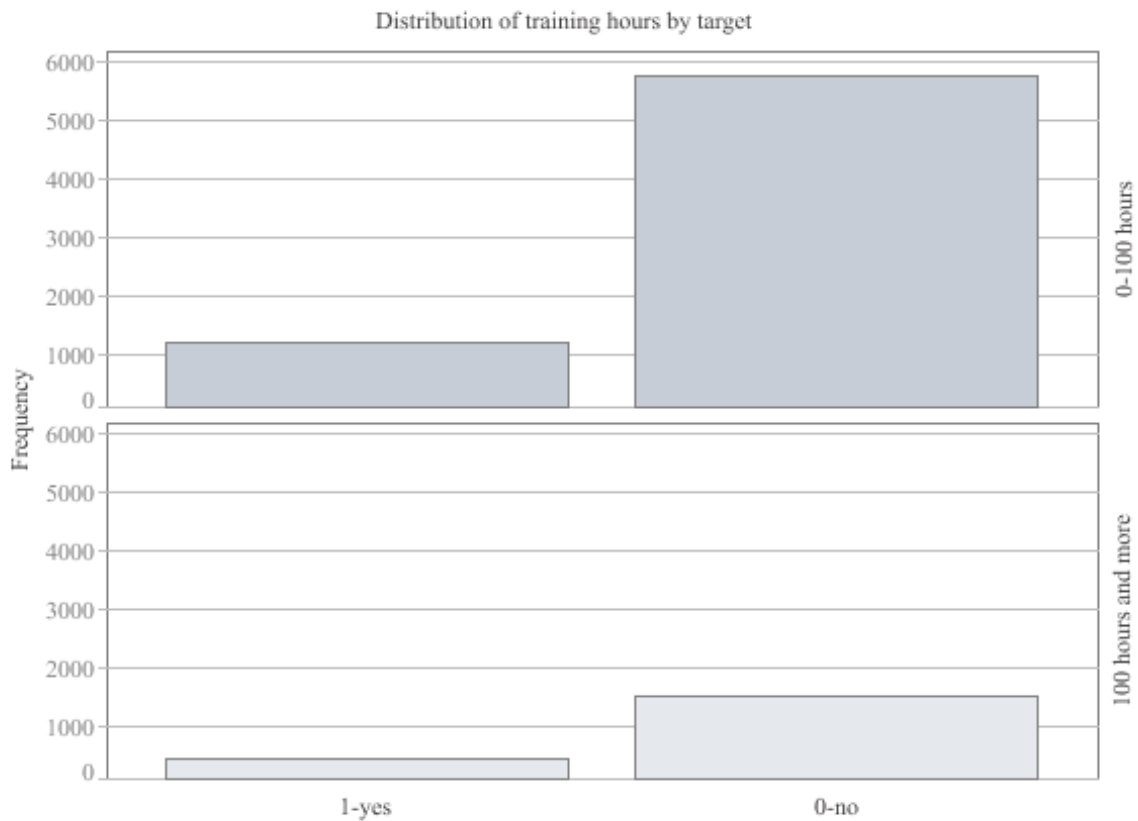

Figure 16. SAS- training hours

3.2.11 Analysis of training hours factor

The diagram above reflects distribution of training hours of specialist according to target. Quantity of training hours are divided into two intervals. One of them represents training hours from zero to 100, another one is determined as 100 hours and more. It can be determined visually that there are almost equal proportions among both intervals in the group of those who are looking for a job and those who are not looking for a job.

In the form of a null hypothesis, it was supposed that there was no relationship between number of training hours of the candidates and the target factor. In contrast, it was supposed in the form of the alternative hypothesis that training hours of the candidates and target factor have a relationship. The Chi-square value that was gained in the Chi-square test provided in SAS is equal to 0.7373, and our accepted significance level is 0.05. The gained P-value is 0.3905. While comparing those values, we found that the alpha value is less than the P-value, which means that we need to accept the null hypothesis. So, training hours of the candidates and target factor have no relationship. Relative risk for the first column shows that there is 1.0525 times more risk that candidates who have training hours from zero to 100, are looking for a job. Relative risk for the second column shows that there is 0.9901 less risk that candidates who have training hours from zero to 100 are not looking for a job. And finally, the odd ration represents that candidates who have training hours from zero to 100 are 1.063 times more likely to change their positions than candidates who have training hours in total from 100 and more.

3.2.12 Results of Chi-square test

It was determined that the city development factor, relevant experience, university, level of education, major studying or studied discipline, work experience, size of the company in which the candidate works and last new job factor have a relationship with the target. Factors such as gender, the type of company in which the candidate works and training hours factors have no relationship with the target. These are factors that HR managers should not look at when looking for data scientists who are ready to change jobs. More detailed analysis is presented in the chapter Results and Dicussion. The results of the Chi-square test are presented in the Table 3.

Analysis by target	Results of Chi-square test
City development index	Have a relationship
Gender	No relationship
Relevant experience	Have a relationship
Enrolled university	Have a relationship
Education level	Have a relationship
Major discipline	Have a relationship
Experience	Have a relationship
Company size	Have a relationship
Company type	No relationship
Last new job	Have a relationship
Training hours	No relationship

Table 3. Results of Chi-square test

Chapter 4

Results and Discussion

According to statistical tests, several relations can be identified which are essential for making decisions by the HR department. The obtained results can be used to identify candidates who are ready to change jobs. Finally, the implication of each dependent variable in the prediction of the independent variable (target) is interpreted.

According to the Chi-square test for the city development index variable, the development index of the candidates' cities and target factor are related with each other. It can be assumed that overall economic tendencies that specialists who are from small cities try to find new job position. It indicates the tendency of specialists to relocate in more developed cites receive more experience and career opportunities in big companies , which are likely in big cities.

The next step was to examine the relationship between the gender of the candidates and their desire to find a job. It was determined that the gender of the scientists and their desire to change jobs were not related to each other. However, we are seeing a huge gap in the number of women and men in data science. It was visually demonstrated that men occupy the majority of this area.

Then an analysis was made of the relationship between the relevant experience of data scientists and their desire to find a job. It was determined that the experience of candidates and their desire to find a job are related to each other, which clearly shows the desire of candidates with less experience to look for a new job position. It also shows the tendency of employees to seek career opportunities and gain experience.

We analyzed if there was a link between their university enrollment and their desire to find work. It was found that the availability of university enrollment and the desire of candidates to find a job are interconnected. Also, due to the relative risk values, it has been determined that it is better for HR professionals to look for candidates who are looking for work among students, since data scientists who do not have university enrollment are more likely to already have a job. This result presents opportunities for the company to invite candidates for internships and bachelor practice.

The next step was to analyze the level of education of the candidates and their willingness to make career changes. As a result of the test, it was found that the level of education of candidates is related to their desire to look for a job. Data scientists with a bachelor degree will be more likely to seek employment than data scientists with a master and doctoral degree. It was also noticed that there are more students with a bachelor level of education both in the group of those who are looking for a job and in the group of those who are not looking for a job.

After that, an analysis of the desire to find a job and the main studied or studied discipline was carried out. STEM is a set of disciplines such as science, technology, engineering, and mathematics. It is found out that there is a relationship between the desire to find a job and the main studying or studied discipline. In addition, candidates with STEM as their primary

discipline are most likely to be looking for jobs in data science. This confirmed fact corresponds to reality. Indeed, subjects such as science, technology, engineering, and mathematics are the main ones to study in the career of data science. Thus, HR managers should look for candidates among those who have studied STEM.

Then, the dependence of the criteria on experience and the desire of workers to look for work were considered. As a result of the test, it was revealed that the amount of experience and the desire of workers to look for a job have a relationship with each other. We see that workers with less experience (two to five years) will be more interested in finding a new job than candidates with six to 15 years of experience. This trend is associated with the desire of the candidate to change positions and possibility to have more career opportunities in a new company.

In the next step, we examined the relationship between the size of the company in which the data scientists currently work and their desire to find a job. The interrelation of these factors was found. Obviously, specialists from small companies will want to change jobs, which again is associated with the search for career opportunities and higher earnings.

When analyzing the types of companies in which data scientists currently work, it was determined that these two factors do not have a relationship with each other. Thus, HR managers should not rely on what type of company a particular candidate is from, as interested individuals may come from different types of companies.

Then an analysis was made of whether there is a relationship between the finishing of the previous and the beginning of the new job of the candidate and their desire to make changes in their careers. It has been established that there is a relationship between the finishing of a candidate's job and the start of a new job and their desire to make career changes. Thus, managers during the recruiting process should pay attention to the gap between the end of the previous and the beginning of the candidate's new job and their desire to find a job. Then an analysis was made of the relationship between the training hours of the candidates and their desire to find a job. It was determined that the training and hours of candidates and their desire to find a job are not related to each other. It turns out that the number of training hours does not influence the desire of data scientists to look for work.

The research presents the objective valuation of the application of statistics in HR management based on the data science field. Due to research, it can be shown how data is valuable and what necessary information it is able to give. The implications that were gained during the analysis will allow the HR managers to offer the open vacancies to specialists who are able to provide real value to the company as well as perform more focused actions, which will also affect the expected productivity of the HR department.

Conclusion

In this thesis, we have focused on data analysis of human resource professionals, which are data science specialists. The majority of the work consisted of analyzing the relationship between one main variable and the rest of the factors. The study began with determining whether there was a relationship between the target and any one factor.

A Chi-square test was applied to determine if there was a relationship between the target and other factors. This method will not be the most appropriate since we are dealing with categorical data and the target factor is a binary value where those who are looking for a job are represented by the number 1 and those who are not looking are represented by the number 0. It was found that the city development factor, relevant experience, university, level of education, major studying or studied discipline, work experience, size of the company in which the candidate works and last new job factor have a relationship with the target. Such factors like gender, the type of company in which the candidate works and training hours factors have no relationship with the target. These are factors that HR managers should not look at when looking for data scientists who are ready to change jobs.

Statistical analysis allows managers to monitor candidates' willingness to change occupations, providing them to objectively evaluate potentials for vacancies and offer open vacancies to the candidates who are able to make any changes in their careers.
So, HR managers can make accurate and informed decisions during the recruiting, selection, performance analysis, and talent management processes by using information received from the analysis. Data is undoubtedly a significant resource, and managers should make best of the situation in their job. The significance of using statistical analysis in areas like human resource management in the field of data science was demonstrated in this thesis. Additionally, the thesis provides the advantages of employing software, which is an essential component of successful performance of HR managers.

References

1. Szamosi LT: **Human Resource Management: A Critical Approach**. 2018.
2. Armstrong M: **A handbook of human resource management practice**: Kogan Page Publishers; 2006.
3. Durai P: **Human resource management**: Pearson Education India; 1990.
4. Dhiraj A, Joshi M, Kumar A: **Human resource management and character building: route to business sustainability**. *International Journal of Business and Globalisation* 2020, **25**(3):332-348.
5. DeCenzo DA, Robbins SP, Verhulst SL: **Fundamentals of human resource management**: John Wiley & Sons; 2016.
6. Malik A: **Strategic human resource management and employment relations**: Springer; 2018.
7. Ivancevich JM, Konopaske R: **Human Resource Management Twelfth Edition**. *Training and Development* 2013:395-397.
8. Werner JM, DeSimone RLRL: **Human resource development**: Cengage Learning; 2011.
9. Association IRM: **Human resources management: concepts, methodologies, tools, and applications**: IGI Global; 2012.
10. Torres-Coronas T, Arias-Oliva M: **E-human resources management: Managing knowledge people**: IGI Global; 2005.
11. Rothwell WJ, George M: **Encyclopedia of human resource management, critical and emerging issues in human resources**: John Wiley & Sons; 2012.
12. Martin V: **Managing projects in human resources training and development**: Kogan Page Publishers; 2006.
13. Dietrich D, Heller B, Yang B: **Data Science & Big Data Analytics Discovering, Analyzing, Visualizing and Presenting Data (pp. 420)**. In.; 2015.
14. Davenport T, Patil D: **Data Scientists: The Sexiest Job of the 21st Century–Meet the people who can coax treasure out of messy, unstructured data**. *Harvard Business Review* 2012:2-8.
15. Albright SC, Winston WL: **Business analytics: Data analysis & decision making**: Cengage Learning; 2014.
16. Robbins AS, Chao SY, Fonseca VP: **What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes**. *Annals of epidemiology* 2002, **12**(7):452-454.