universität wien

MASTERARBEIT / MASTER´S THESIS

Titel der Masterarbeit / Title of the Master´s Thesis

The influence of different determinants on employee turnover in the context of start-ups

verfasst von / submitted by

Tobias Quebe, B.A.

Angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet	UA 066 915
Studienrichtung lt. Studienblatt / Degree programme as it appears on the student record sheet:	Masterstudium Betriebswirtschaft
Betreut von / Supervisor:	Assoz. Prof. Steffen Keck, PhD

Abstract

The population on earth continues to grow steadily, but on average, it is getting older and older. Moreover, companies face, besides an older workforce, the problem of a shortage of skilled workers. Also, employees tend to switch jobs more frequently compared to the past. Therefore understanding employee turnover is of high interest to companies. Research on this topic has been going on for several decades, defining different types of turnover, such as voluntary turnover or avoidable turnover, theories on how it comes to turnover have been developed, and determinants that can be used to predict employee turnover have been investigated. However, start-ups were left out of this research.

The present study wants to investigate whether already tested determinants in established companies show similar significances and correlations as in start-ups. Due to the special characteristics of start-ups, such as the enormous growth, further determinants, such as the funding volume and the number of patents and trademarks, are examined. For this purpose, two different data sets were used and merged. One dataset contained company information on start-ups founded since 2017 and the other public LinkedIn profiles. These datasets were processed and linked using Python, resulting in a data set with 1296 start-ups. This sample was analyzed using multiple regression analyses. The results of the analyses show that the number of patents and trademarks have no significant influence on the employee turnover of a startup ($r = -.41$ $p > .1$), whereas the funding volume has a significant influence ($r = -.065$, $p < .01$). A possible reason for the influence of the funding volume could be that it offers employees security in the sense of punctual payment. Nevertheless, only a small part of the variance in the variable employee turnover is explained by the model ($r^2 = .138$). Finally, this study provides initial insights into the topic of employee turnover in start-ups and opens up many further research opportunities. Thus, the present model can be expanded by including additional determinants, which have already been proven to be good predictors for established companies, such as job satisfaction or organizational commitment.

Abstrakt

Die Bevölkerung der Erde wächst stetig, wird aber im Durchschnitt immer älter. Außerdem stehen die Unternehmen neben einer älteren Belegschaft vor dem Problem des Fachkräftemangels. Darüber hinaus neigen die Arbeitnehmer im Vergleich zu früher häufiger dazu, den Arbeitsplatz zu wechseln. Daher ist das Verständnis der Mitarbeiterfluktuation für die Unternehmen von großem Interesse. Seit mehreren Jahrzehnten wird zu diesem Thema geforscht, wobei verschiedene Arten der Fluktuation wie freiwillige Fluktuation oder vermeidbare Fluktuation definiert, Theorien zur Entstehung von Fluktuation entwickelt und Determinanten zur Vorhersage der Mitarbeiterfluktuation untersucht wurden. Start-ups wurden bei diesen Untersuchungen jedoch nicht berücksichtigt.

In der vorliegenden Studie soll untersucht werden, ob bereits getestete Determinanten in Start-ups ähnliche Signifikanzen und Korrelationen aufweisen wie in etablierten Unternehmen. Aufgrund der besonderen Charakteristika von Start-ups, wie dem enormen Wachstum, werden weitere Determinanten, wie das Finanzierungsvolumen und die Anzahl der Patente und Schutzmarken, untersucht. Zu diesem Zweck wurden zwei verschiedene Datensätze verwendet und zusammengeführt. Ein Datensatz enthielt Unternehmensinformationen zu Start-ups, die seit 2017 gegründet wurden, der andere öffentliche LinkedIn-Profile. Diese Datensätze wurden mit Python verarbeitet und verknüpft, so dass für die statistische Analyse ein Datensatz mit 1296 Start-ups vorlag. Diese Stichprobe wurde mithilfe multipler Regressionsanalysen analysiert. Die Ergebnisse der Analysen zeigen, dass die Anzahl der Patente und Marken keinen signifikanten Einfluss auf die Mitarbeiterfluktuation eines Start-ups hat ($r = -.41$, $p > .1$), während das Finanzierungsvolumen einen signifikanten Einfluss hat ($r = -.065$, $p < .01$). Ein möglicher Grund für den Einfluss des Fördervolumens könnte sein, dass es den Mitarbeitern Sicherheit im Sinne einer pünktlichen Bezahlung bietet. Dennoch wird nur ein kleiner Teil der Varianz in der Variable Mitarbeiter Fluktuation durch das Modell erklärt ($r^2 = .138$). Abschließend lässt sich festhalten, dass diese Studie erste Einblicke in das Thema Fluktuation in Start-ups liefert und viele weitere Forschungsmöglichkeiten eröffnet. So kann das vorliegende Modell um weitere Determinanten erweitert werden, die sich bereits für etablierte Unternehmen als gute Prädiktoren erwiesen haben, wie z.B. Arbeitszufriedenheit oder organisationales Engagement.

Table of content

Table of figures

List of abbreviations

Abbreviation	Meaning
CLM	Classical Linear Model
CLT	Central Limit Theorem
CLT	Central Limit Theorem
DV	Dependent Variable
ICT	Information and Communication Technology
IP	Intelligent property
IQR	Interquartile range
IV	Independent Variable
NACE	General Industrial Classification of Economic Activities in the European Communities
Obs	Observations
OCQ	Organizational Commitment Questionnaire
R&D	Research and Development
SD	Standard Deviation
US	United States of America

1. Introduction

The world's population is growing unprecedentedly and will reach 8 billion people for the first time in 2022. Moreover, the global population is predicted to grow to 8.5 billion in 2030 and 9.7 billion in 2050. This growth is mainly due to declining mortality levels and increased global life expectancy (United Nations Department of Economic and Social Affairs, Population Division, 2022). Especially developed countries are seeing major demographic shifts in their population as baby boomers get older and subsequent generations become smaller and grow slower. Consequently, in OECD countries, the share of people above 50 is predicted to be at 45% in 2050. Hence the percentage of the working population is declining, and talent shortages are rising, which is shown by an increase in the old-age dependency ratio from 30% in 2020 to 50% in 2050. (OECD, 2020). Because of this prediction, companies face changes in the workforce size and structure they can access (Clark & Ritter, 2020; Kulik et al., 2014; E. S. W. Ng & Burke, 2005; Taylor & Bisson, 2020).

Companies have to face another significant challenge: the skill shortage in the workforce (Brunello & Wruuck, 2021; Kulik et al., 2014). Skills in human capital are a considerable element for companies to gain a competitive advantage over others. Hence skill shortage can affect organizational performance in a negative way (Malik et al., 2019).

Besides the two aspects mentioned above of demographic change and skill shortage in the workforce, employees are turning more frequently than in the past. In the past, employers did not worry about turnover as much as today because new joiners could easily absorb a company's culture. As companies now need to be international, work from home policies, etc., it is more challenging to replace a leaving employee. For example, in September 2021, more than 3% of the US workforce quit their jobs Feld (The Economist, 2021). This massive turning and resignation trend is called the "Great Resignation." as this is not a trend for a particular industry or country, every business must face the challenge of turning employees. Therefore, fluctuation is an essential issue for every company to deal with daily.

At first glance, turnover is associated negatively with companies' performance (Laulié & Morgeson, 2021; Meier & Hicklin, 2008; Park & Shaw, 2013; Shaw, Gupta et al., 2005; Wynen et al., 2019)(De Winne et al., 2019; Glebbeek & Bax, 2004). Nevertheless, studies also showed that turnover could benefit companies (Simón et al., 2022). However, most of the literature shows the negative effect of turnover (Hancock et al., 2013, 2017; Hausknecht & Trevor, 2011; Heavey et al., 2013; Park & Shaw, 2013; Skelton et al., 2019). For example, the costs associated with turnover for hiring and training a new employee can be over 150% of their annual payment (Hansen, 1997). Additionally, Allen et al. (2010) mention that the cost of turnover can range from 90% to 200% of the employee's annual payment. These are costs for recruiting, selecting an appropriate candidate, and providing training for the successor (Allen et al., 2010). Moreover, turnover also has indirect costs when looking at skill depreciation. David and Brachet (2011) found that turnover is twice as much related to organizational forgetting than skill depreciation (David & Brachet, 2011). Other researchers also mention that turnover can be costly in terms of a decrease in efficiency (Kacmar et al., 2006), in terms of knowledge and social capital (Dess & Shaw, 2001), or even in a reduction of customer satisfaction (McElroy et al., 2001).

Nevertheless, also beneficial aspects of turnover were found in research, as there are justified doubts that there is not only a negative linear relationship but also a curvilinear relationship (Park & Shaw, 2013; Wynen et al., 2019). Researchers also mention that there might be some curvilinearity in the relationship between turnover and workforce performance (Meier & Hicklin, 2008; Shaw, Gupta, et al., 2005). Furthermore, other researchers mention that the relationship between turnover and organizational performance differs between the types of turnover, as there is also a linear and positive relationship regarding involuntary turnover (An, 2019; S. Lee, 2018). Another positive aspect is motivation since the motivation of the remaining employees can rise after a turnover and may outweigh its associated cost (Laulié & Morgeson, 2021).

Hence, "it is too early to conclude that the relationship between turnover and firm performance is straightforwardly negative” (Hancock et al., 2013, p. 574). On the one hand, it negatively affects organizational performance, generating costs for companies, decreasing efficiency, and possible customer dissatisfaction. On the other hand, it has beneficial aspects for the workforce performance, the performance of shops, and motivational benefits for stayers. Subsequently, turnover has been studied for decades in the literature, and several theories (Heavey et al., 2013; P. W. Hom et al., 1984; T. W. Lee & Mitchell, 1994; March & Simon, 1958; Mobley, 1977; Mobley et al., 1979; Porter & Steers,

1973) about it have been developed, and it is still of high interest in the literature (P. W. Hom et al., 2017).

1.1 Objective of this work

While the world's population is aging and skill shortage is rising all over the globe, it shows that companies must deal even more with turnover in their daily business. Several theories about turnover have been developed in the literature, and various determinants to predict turnover have been investigated. That research shows that companies are affected in a negative as well as a positive manner by turnover.

As turnover happens in every company, whether established or recently founded, research about turnover in start-ups should be addressed. However, while the literature investigates turnover by determining predictors of it and further developing created theories, it is mainly researched for already established companies or governmental facilities rather than start-ups. Furthermore, predictors and variables fluctuate in their significance Feld (Kammeyer-Mueller et al., 2013), which underlines the necessity to research turnover for recently established firms.

Moreover, companies must deal with the "war for talents" (Beechler & Woodward, 2009; Löffler & Giebe, 2021). The "war for talents" refers to the competition for the best talents in the labor market so that they join the company. We are not talking about talents who have a company car or a prestigious office as a criterion for choosing a company, but about people who are characterized by enormous creativity and willingness to perform and who pay more attention to freedoms such as time with the family. This challenges companies to retain employees in the long term (Busold, 2019). Hence, the "war for talents" affects a company's employee turnover. Besides the "war for talent" deals with people who also focus on aspects like freedom, payment is still essential in negotiating with them (Berwick & Raval, 2022). Since some talents consider prestige in the form of payment or a prestigious name to be essential factors in choosing an employer, start-ups, which usually have fewer monetary resources available compared to established companies, are at a disadvantage. This represents another aspect of why the turnover topic at start-ups should be investigated (Buschow et al., 2015).

Hence this master's thesis aims to analyze the significance of several turnover determinants for start-ups and provide the first input to the literature.

1.2 Motivation

The motivation for this work is to analyze specific determinants to predict turnover. Since turnover is a challenge for companies as they face skill shortages and demographic changes in the population, it is necessary to understand the topic and how specific determinants can affect turnover. Moreover, an extension for the turnover literature will be provided by analyzing the different determinants in the context of start-ups and not established ones.

Another critical factor is that the significance of variables can fluctuate over time (Kammeyer-Mueller et al., 2013). For example, turnover among women is higher than for men (P. Hom et al., 2008; Stroh et al., 1996; Weisberg & Kirschenbaum, 1993), whereas other studies found that men have a higher turnover compared to women (Light & Ureta, 1992). In the case of actively searching for a new job, no support was found that a gender difference exists (T. H. Lee, 2012). Hence the correlation and the significance of already researched determinants might change, have a similar or different correlation for newly established firms, and if new determinants show a significant correlation for turnover.

2. Literature Review

A literature review is conducted first to understand the subject of the master's thesis. This review should give the reader an overview of the current state of research. In this context, different constructs and theories regarding turnover will be presented, and the most critical determinants of turnover will be explained.

2.1 Employee turnover

Turnover, or more precisely, employee turnover, is a topic of such high interest in the literature that it is already investigated for decades and is still under investigation (P. W. Hom et al., 2017). Hence, knowing what is understood under the term turnover itself is mandatory. Price (1977) defined turnover “as the degree of individual movement across the membership boundary of a social system” (Price, 1977, p. 4). Hence turnover is the movement of an individual from one organization to another and between associations. Therefore it is the process of an individual leaving a job (Brotherton, 2010). Furthermore, turnover refers to the movement of workers in the labor market. This process suggests that workers can move between enterprises or stages of employment, unemployment, and inactivity (Abbasi & Hollman, 2000; Burgess, 1998).

Besides, movement turnover can be classified into several types, which the cause of movement can distinguish. For example, turnover can be voluntary or involuntary, i.e., an employee leaves the company at his request due to a better job offer. Alternatively, the organization dismisses an employee due to poor performance, which would be involuntary turnover (P. W. Hom et al., 2019).

Moreover, managers refer to the term turnover as the entire process of filling a vacant position in a company. Hence it includes steps like searching for a replacement, interviewing possible new employees, negotiating with them, and training the new employee on the job. This means that a vacant position in a company needs to be filled with a new employee who will be trained for that position Feld (Ongori, 2007).

Hence turnover is a movement of employees in the labor market from one company to another or from employment to unemployment or inactivity/retirement. On the managerial side, turnover includes several associated process steps, like training the replacing employee. Moreover, it can be categorized into various types of turnover, like voluntary or involuntary. Several types of turnover will be described in the following capital.

2.2. Types of employee turnover

So far, the literature has primarily examined self-initiated turnover, i.e., voluntary turnover (P. W. Hom et al., 2019). However, other types of turnover should be addressed, so this section describes different kinds of turnover.

2.2.1 Voluntary vs. Involuntary turnover

As described before, the literature differentiates between voluntary and involuntary turnover. The literature focuses more on the voluntary type because it is easier to validate turnover theories regarding their choice behavior (P. W. Hom et al., 2019). Most theories assume that employees make their own decisions about turnover because of better job opportunities or an uncomfortable job situation (Griffeth & Hom, 1995). Hence, the turnover is self-initiated by the employees and out of the organization's control. Nevertheless, as this definition deals with voluntary organizational quits, it does not consider job transfers within a company or promotions (P. W. Hom et al., 2019).

Involuntary turnover is initiated by the organization of an employee, meaning that the employee has to leave the company involuntarily (Hausknecht & Trevor, 2011; Shaw et al., 1998). Hence the organization decides when an employee must leave the company. These involuntary dismissals are, for example, done due to bad performance of the employee (P. W. Hom et al., 2019),

Therefore it can be stated that voluntary and involuntary turnover differ mainly by the initiator, either the employee or the organization (An, 2019). For voluntary turnover, the employee initiated the turnover (Shaw et al., 1998), whereas the organization began the process for involuntary turnover.

2.2.2 Avoidable and unavoidable turnover

Another differentiation of turnover is avoidable and unavoidable turnover, which Abelson first described in 1987. The study expanded the taxonomy of employee turnover, which provides a more precise indication between variables that tend to predict turnover and employee turnover itself (Abelson, 1987). In this context, avoidable and unavoidable turnover refers to a company's perspective. This distinguishes it from voluntary and involuntary turnover, which refers to the employee's perspective. For this reason, there are different possible combinations for avoidable and unavoidable fluctuation. One possible combination would be that the turnover could have been avoided from the organization's point of view. However, from the employee's point of view, it was a voluntary turnover, for example, because he or she received a better job offer at another company. The combination of involuntary from the employee perspective but in the avoidable control area of the company is dismissals or layoffs.

Moreover, unavoidable but voluntary turnover may be career changes of the employee or a movement to another state or country done by the employee. Involuntary turnover for the employee, but unavoidable for the company, could be a severe medical issue of the employee or the death of one (Abelson, 1987). Figure 1 provides an overview of the possible combinations mentioned above.

		Employee control: Yes	Employee control: No
Organizational control	Avoidable	• Better payment • Better working conditions • Problems with leadership	• Dismissals • Layoffs
	Unavoidable	• Home staying • Career change • Movement of location	• Death • Medical reasons

Figure 1 own illustration of avoidable vs. unavoidable taxonomy based on Abelson, 1987

Hence avoidable and unavoidable turnover is an expansion of the avoidability taxonomy about employee turnover by considering the organization's viewpoint. Based on this classification, researchers attempt to predict avoidable and unavoidable employee turnover by investigating determinants such as work attitudes. (P. W. Hom et al., 2019)

2.2.3 Collective turnover

Collective turnover can be described as the aggregated departure of several employees within an organization or a team (Hausknecht & Trevor, 2011; Heavey et al., 2013). Thus, collective turnover does not deal with the departure of an individual employee but with the departure of several employees from a team or an organization (Hausknecht, 2017).

Moreover, Nyberg and Ployhart (2013) describe collective turnover as "the quantity and quality of depletion of employee knowledge, skills, abilities, and other characteristics" (Nyberg & Ployhart, 2013, p. 109). Therefore it can also be described as the collective loss of employees and their capabilities (Nyberg & Ployhart, 2013).

Collective turnover is typically measured in rates, with the numerator calculated as the sum of departures during a given period, e.g., annually, divided by the denominator, which represents the total number of employees in the organization, team, or collective either at the beginning of the period or on average during that period (P. W. Hom et al., 2019).

As this measurement provides a reasonably straightforward calculation, it has also been criticized, e.g., by Hausknecht and Holwerda (2013), who noted that the measure focuses only on the number and not the quality of departures (Hausknecht & Holwerda, 2013; P. W. Hom et al., 2019).

All in all, collective turnover represents the loss of employees and their skills, measured in rates (departures in a period divided by the number of employees at the beginning of that period). Besides the straightforward calculation of the turnover rate, the quality of the departures is not included.

2.3 Determinants of turnover

As described above, several theories have been developed to predict employee turnover. However, several determinants to predict turnover have also been developed and investigated. Some of these determinants show higher correlations than others, and their significance also differs. Also, different significance levels for the same determinant can be found. Hence, several determinants will be described in the following chapter, and their significance will be evaluated.

2.3.1 Job satisfaction

Employees' satisfaction with a job is crucial for business as they ensure an organization's production (Ali, 2016). Job satisfaction is one of the most used determinants for employee turnover, as shown in a meta-analysis done by Griffeth et al. (2000). Job satisfaction was used in over 60 studies to predict turnover (Griffeth et al., 2000). Moreover, Rubenstein et al. (2018) did a more recent meta-analysis showing that job satisfaction was used in over 170 studies to predict turnover (Rubenstein et al., 2018).

However, to understand this determinant's significance and predictability power, it is essential to have a general understanding of the term job satisfaction. Job satisfaction is generally described as an employee's contentment with their job. It represents the emotion that results from the perception of a workplace. This perception includes the worker's physical, social, and psychological needs and expectations and whether the workplace meets them. (Azeez et al., 2016). Hence employees will be satisfied if their expectations and needs are met or exceeded, or they become unsatisfied.

Questionnaires or surveys are commonly used in the literature to measure employees' job satisfaction. Andrews and Withey (1976) developed a questionnaire containing five questions where each question is answered by selecting one option from a seven-point Likert scale (Andrews & Withey, 1976). Other researchers used different questionnaire templates like the Minnesota Job Satisfaction Questionnaire (Arvey et al., 1989). Skelton (2019) used the questionnaire of Andrews and Withey (1976) to measure employees' job satisfaction in the manufacturing sector (Skelton et al., 2019). Tarigan and Ariani used the shortened Minnesota Job Satisfaction Questionnaire provided by Arvey et al. (1989) to measure job satisfaction (Tarigan & Ariani, 2015). Besides these multi-item questionnaires, researchers also use single-item questionnaires to determine job satisfaction by asking, "Overall, how satisfied are you with your current job?" (Scanlan & Still, 2019). Answers to this question are again entered into a Likert scale. Hence job satisfaction can be measured using either single-item surveys or multi-item ones (Sessa & Bowling, 2020).

However, the correlation between job satisfaction and turnover varies across studies. Mossholder et al. (2005) investigated employee turnover in the healthcare sector. Their study showed a correlation between job satisfaction and turnover of -.13 (Mossholder et al., 2005). Moreover, Griffeth et al. (2000)

meta-analysis showed that job satisfaction is one of the best predictors of employee turnover, with a corrected average correlation of -.22 (Griffeth et al., 2000). Furthermore, Rubenstein et al. (2018) conducted another meta-analysis and found significant results for job satisfaction on turnover. Their meta-analysis showed an even higher correlation between job satisfaction and turnover of -.28 on average (Rubenstein et al., 2018).

Job satisfaction describes the perceived satisfaction with a job, meaning whether the employee's social, psychological, or physical needs and expectations are met. It is one of the most critical company assets and shows high correlations with turnover, which is why it is one of the most used determinants. Satisfaction is often measured through various surveys with one or more questions. Furthermore, several studies have shown a significant correlation between job satisfaction and turnover, showing a significant average correlation of r = -.28 in over 170 studies.

2.3.2 Organizational commitment

As committed employees are likelier to stand with an employer (Mowday et al., 1979), the employee's organizational commitment is another used determinant. Much research has also been done on this determinant, which is why it has similar values to job satisfaction in the number of studies that use it as a determinant in various meta-analyses (Griffeth et al., 2000; Rubenstein et al., 2018). The meta-analysis by Griffeth et al. (2000) cites 67 studies with over 27.000 employees in the samples, and Rubenstein et al. (2018) mention 129 studies with more than 71.000 employees (Griffeth et al., 2000; Rubenstein et al., 2018) considering an organizational commitment to predict turnover.

Meyer and Allen (1991) developed a three-component conceptualization of organizational commitment. The three mentioned components are the desire (affective commitment), the need (continuance commitment), and the obligation (normative commitment) to stay with an organization (Meyer & Allen, 1991). It can be derived that organizational commitment is the degree to which an employee desires to stay and is loyal to an organization (Rubenstein et al., 2018). The literature primarily uses this model to determine employees' affiliation with a firm (Solinger et al., 2008).

Like job satisfaction, organizational commitment is also measured by questionnaires and surveys. An example is the Organizational Commitment Questionnaire (OCQ) developed by Mowday Steers and Porter in 1979 (Mowday et al., 1979). This Questionnaire consists of 15 questions like "I am proud to tell others that I am part of this organization" (Mowday et al., 1979, p. 228), and each response is collected on a 7-point scale (Mowday et al., 1979). Nevertheless, researchers have criticized the questions in this questionnaire as there is an overlap between the questions and turnover itself, which should be measured by the questions (Cohen, 1993; O'Reilly & Chatman, 1986; Reichers, 1985). Therefore, some researchers use a shortened version of the OCQ, excluding the problematic question terms (Cohen, 1993). Moreover, the shortened OCQ provides slightly better reliability than the original OCQ, α = 0.90 compared to α = 0.88, respectively (Bartlett & McKinney, 2004). However, some researchers still stick to the more extended version; for example, Vance (2017) used the OCQ with 15 statements to measure employees' organizational commitment to a company in a manufacturing plant (Vance, 2015).

Organizational commitment correlates to turnover in several studies, as shown in the meta-analysis done by Griffeth et al. (2000) and Rubenstein et al. (2018). The analysis by Griffeth et al. (2000) mentions that in the investigated studies, the correlation of organizational commitment and turnover showed a corrected mean r of -.27 (Griffeth et al., 2000). Rubenstein et al. (2018) found a slightly higher corrected mean r of -.29 (Rubenstein et al., 2018).

All in all, organizational commitment is a determinant for turnover, which describes the affiliation of an employee to their organization. This belonging can be an affective commitment, a continuance commitment, or a normative commitment. Several studies have investigated this determinant, showing a significant average correlation of -.27 to -.29 Feld (Griffeth et al., 2000; Rubenstein et al., 2018) to turnover. Questionaries like the OCQ, consisting of 9 or 15 items, are used to measure it.

2.3.3 Company size

Another determinant that is used to predict turnover is company size. In most studies, the determinant is used as a control variable. Therefore, it is mainly controlled for the effect of company size as studies are aware of its effect on employee turnover. Nevertheless, controlling the company size is essential, as larger companies are more likely to offer better development opportunities for their employees (P. G. Benson et al., 1987).

Alterman et al. (2021) examined the effects of wage nondisclosure on voluntary turnover rates in firms while also controlling for the effects of firm size. They controlled for the effect of organizational size because earlier studies showed that larger companies are more likely to offer better career development opportunities (G. S. Benson et al., 2004) and higher wages (Kalleberg & Van Buren, 1996). At a significance level of $p < 0.05$, they found a correlation between the control variable and the independent variable pay secrecy practice of 0.26 and a positive correlation of $r = .1$ to the voluntary turnover rate of the company (Alterman et al., 2021). From this, it can be deduced that both the exercise of pay secrecy and the size of a company positively influence the turnover rate of a company.

A study by Shaw, Gupta, et al. (2005) investigated the relationship between voluntary turnover and workforce performance (Shaw, Gupta, et al., 2005). The study comprised over 140 American Concrete Pipe Association facilities in the United States and Canada. They controlled each facility's size by the number of production workers. They found a correlation of $r = .16$ at a significance level of $p < .01$ (Shaw, Duffy, et al., 2005). Hence the company size showed an even higher positive correlation to voluntary turnover.

A study by Hancock et al. (2013) suggested a moderating effect of the company size on the relationship between collective employee turnover and performance (Hancock et al., 2013). They mention other reasons why it is crucial to control company size because a company's size can impact how performance is affected. It can also impact the company's resources to manage departures and new hires. Based on that, they infer that larger firms could better compensate for the negative aspects of turnover than small- and medium-sized organizations. This leads them to hypothesize that the size moderates the relationship more negatively for smaller and medium-sized organizations than larger ones. The study showed that company size was, at least partially, a significant moderator with a chi-squared value of 14.27 at a significance value of $p < .05$. Thereby the medium-sized and large-sized organizations showed a correlation of $r = -.07$ and $r = -.06$ and, therefore, a stronger negative relationship than small-sized organizations with a correlation of $r = -.01$. Nevertheless, only the correlation of large-sized companies was significant at $p < .05$ (Hancock et al., 2013). It implies that the company's size if it is large, weakens the relationship between collective turnover and performance.

Also, the meta-analysis by Rubenstein et al. (2018) showed some weak effects of the organization size ($r = .03$) by considering 15 studies and, in total, over 30000 employees. However, they do not state if the variable was used as a control, moderating, or independent variable in the studies (Rubenstein et al., 2018).

Company size is mainly used as a control variable or moderator for employee turnover. Besides the fact that the correlation differs across studies, company size correlates positively or negatively with turnover. Reasons for a negative relationship might be that larger companies can offer better career development opportunities (G. S. Benson et al., 2004) and higher wages (Kalleberg & Van Buren, 1996).

2.3.4 Tenure

Employee tenure is another determinant used in the literature to predict turnover. Tenure is when an employee has worked for their employer in a job.

Yanadori and Kato (2007) investigated the relationship between the voluntary turnover ratio and firm labor productivity in Japanese companies. They wanted to expand the literature by extending the generalizability of the relationship. Based on human capital theory, which suggests that firm-specific human capital is developed with the length of an employee's tenure and therefore affects a firm's effectiveness, the firm average employee tenure mediates the negative relationship between the workforce's productivity and voluntary employee turnover. This is also shown in their regression analysis as the average employee tenure shows a correlation of $r = -.38$ at a significance level of $p < .01$ on the turnover ratio (Yanadori & Kato, 2007). Their findings also show a negative relationship similar to studies done with US-based companies (Kacmar et al., 2006; Shaw, Duffy, et al., 2005; Shaw, Gupta, et al., 2005).

Benson et al. (2004) investigated if the investment in general skill development would reduce employee turnover. In their study about how general skill development and promotion relate to voluntary turnover, they argue that employees with lower tenure levels are likelier to quit. In their analysis, tenure is negatively associated with turnover, with a correlation of -.27 at a significance level of $p < .01$ (G. S. Benson et al., 2004). This suggests that with each additional year of tenure, the likelihood of employee turnover decreases and supports the assumption that employees with longer terms are less likely to quit compared to colleagues with less tenure.

Also, another study that investigated the relationship between an employee's age and voluntary turnover used the determinant tenure in moderator tests (T. W. H. Ng & Feldman, 2009). Although age and tenure are positively correlated, they are not identical because age and tenure are often divergent when job mobility increases. Moreover, long-tenured employees are also less likely to quit because of extrinsic motivations like financial benefits provided by the organization. Therefore, tenure was used as a moderating variable to test whether it attenuates the negative relationship between voluntary turnover and age. Their regression analysis confirmed the assumption that organizational tenure would moderate the relationship by providing a correlation of -.18 and -.09. The two different correlations result from the two used subgroups in sample one for tenure lower than six years and one for equal or above six years. The group with longer organizational tenure shows a higher negative correlation than the others. Hence the moderating effect is more substantial for employees with longer organizational tenure, and tenure should be considered when predicting employee turnover (T. W. H. Ng & Feldman, 2009). Based on the findings of the previous study of Ng and Feldman, Peltokorpi et al. (2015) controlled for tenure. They also find a negative relationship to voluntary employee turnover with a correlation of -.08 at a significance level of $p < .05$ (Peltokorpi et al., 2015).

The meta-analysis of Rubenstein et al. (2018) shows the widespread use of the determinant seniority. More than 110 studies were considered in this analysis, which showed an average correlation to voluntary turnover of -.27. Thereby tenure is shown as one of the strongest individual predictors in the analysis (Rubenstein et al., 2018).

Regardless of the type of variable, control variable, or moderator, tenure is often used in predicting employee turnover. One reason for this is that longer tenure makes turnover less likely. Various studies have confirmed this by finding a correlation of -.38 to -.09 (G. S. Benson et al., 2004; T. W. H. Ng & Feldman, 2009; Peltokorpi et al., 2015; Yanadori & Kato, 2007). Several studies found a significant average correlation of -.27 across over 110 studies (Rubenstein et al., 2018). Hence, tenure should be considered when predicting employee turnover.

2.3.5 Gender

The gender of an employee is another determinant used in the literature to predict turnover. Often, the literature provides findings that women tend to have higher turnover rates than men (CEDEFOP, 2020; Cotton & Tuttle, 1986; P. Hom et al., 2008; G. B. Lewis & Park, 1989; Maurer & Qureshi, 2021; Weisberg & Kirschenbaum, 1993). However, some researchers found contradicting results where women had a lower turnover rate than men (Lyness & Judiesch, 2001). There are several reasons why women tend to leave a job more frequently than men. Often they take over more responsibilities for home and family (Hewlett, 2006; Maurer & Qureshi, 2021), less payment, fewer development opportunities, and they have to face more career obstacles than men (P. Hom et al., 2008).

Lewis and Park (1989) analyzed the effect of gender on turnover in the public sector. Their study shows that women's higher turnover is not due to gender but the average age, pay, or tenure. Instead, it shows that comparing women with men, who have similar average age, pay, etc., the turnover is quite similar (G. B. Lewis & Park, 1989).

Grissom et al. (2012) examined whether supervisor gender impacts employee turnover or job satisfaction in the public sector. Although there are mixed results in the literature as to which gender has a higher impact on turnover, the consensus in the literature is that the gender of the supervisor should be included to predict turnover (Grissom et al., 2012).

All in all, gender has been a determinant of turnover for a long time; the meta-analysis done by Hom and Griffeth (1995) took 15 studies for their analysis which showed a correlation of r = -.07 (Griffeth & Hom, 1995). The following meta-analysis by Griffeth et al. (2000) already considered 45 studies, which provided an average corrected correlation of r = -.11 (Griffeth et al., 2000). The most recent analysis by Rubenstein et al. (2018) used 89 studies in their research and found an average correlation of .00 (Rubenstein et al., 2018).

2.3.6 Diversity

Another determinant is diversity in an organization. Diversity can, besides gender diversity, also be found in race or age (Leonard & Levine, 2006).

Leonard and Levine (2006) analyzed the effect of diversity on turnover in over 800 workplaces (Leonard & Levine, 2006). Based on social science theory, they hypothesized that turnover would be lower for workplaces with a lower diversity by considering diversity for gender, race, and age. The analysis showed only minor and no significant effects of diversity, either gender, race, or age. However, the results also showed that men´s rate was unaffected by gender diversity. However, women left diverse workplaces more often than with high shares of men or high percentages of women (Leonard & Levine, 2006).

Maurer and Qureshi (2019) investigated the impact of the representation of women on collective turnover (Maurer & Qureshi, 2021). They suggested that higher gender diversity would provide spillover effects to women, who tend to leave a company earlier than men, and men, resulting in a general reduction of collective turnover. Therefore, increasing diversity would reduce collective employee turnover (Maurer & Qureshi, 2021). The analysis found proof for their suggestion.

As diverse as diversity can be in the form of age, gender, and ethnic origin, different results can also be found in the literature. For gender diversity, there are both significant and non-significant results. Accordingly, further research is needed to determine whether diversity impacts employee turnover.

2.3.7 Industry

The industry in which a company operates is also a determinant used to predict employee turnover. Dependent on the industry, the turnover rate can vary drastically. In the US, for example, the state and local education sector had the lowest turnover, with 16%. In contrast, the accommodation and food service industry had the highest turnover, of over 86%, in 2021 (U.S. BUREAU OF LABOR STATISTICS, 2022). These different turnover rates can also be seen on a global scale. For example, the tech and media sector had a turnover rate of almost 13%, and the governmental administration sector provided the lowest turnover rate, with around 8.4 % (G. Lewis & Soroñgon, 2022).

Nevertheless, several researchers focus only on a specific industry like hospitality (Dipietro & Condly, 2007; Okae, 2018; Tews et al., 2013), retail (Olubiyi et al., 2019), healthcare (Antwi & Bowblis, 2018; Collini et al., 2015), or manufacturing (Chin, 2018) rather than comparing different industries at once.

The meta-analysis by Hancock et al. (2013) investigated the moderating effect on the relationship between employee turnover and performance. Depending on the industry, turnover in a company can have problematic consequences. An industry where standardized systems and technologies are standard might have it easier to find replacements than industries with high knowledge requirements, where it might be more challenging to find replacements. Furthermore, human capital represents a different competitive advantage in each industry (Hancock et al., 2013). They clustered the investigated companies into four categories, the first consisting of financial, service, health, and tech, the second one of manufacturing and transport, the third of retail and food, and lastly, multiple. The different industries showed various moderating effects on performance. The most substantial effect was found in the manufacturing and transportation industries, with a correlation of -.07. In contrast, the weakest effect with a correlation of -.02 was shown in the category of various industries (Hancock et al., 2013). Hence the moderating effect of the industry varies depending on the industry the organization operates.

A reason for different turnover rates in various industries is that human capital is more interchangeable in some sectors than others (Shaw, Duffy, et al., 2005). For example, the restaurant industry experiences high turnover rates as the human capital are more alike than in other industries.

Overall, the industry can have a significant moderating role in the turnover performance relationship. Nevertheless, different industries have varying degrees of impact on this relationship (P. W. Hom et al., 2019).

2.3.8 Country

The country has not been used in the literature as an explicit determinant of employee turnover. Only a few studies have been conducted in individual countries (De Winne et al., 2019; Khera & Divya, 2019; S. Lee, 2018; Sender et al., 2018). Whether the country influences employee turnover in a company has yet to be researched. In most cases, researchers only examine turnover in one country and do not compare it to other countries. Nevertheless, some studies also make a comparison between the two countries. Sender et al. (2018) analyzed the effect of job embeddedness on turnover in China and Switzerland. Their analysis showed that higher job embeddedness would result in a lower likelihood of turnover in Switzerland compared to China (Sender et al., 2018).

However, a lot of the research about turnover has taken place by investigating employee turnover in the US or having a data set about US companies or people (Collins et al., 2015; S. Lee, 2018; Skelton et al., 2019; Stamolampros et al., 2019). However, there are also studies concentrating on other countries besides the US. Kera and Divya (2019) used a machine-learning model to predict employee turnover in the Indian IT Industry (Khera & Divya, 2019). In addition, De Winne et al. (2019) made a longitudinal analysis of Belgian companies about the impact of employee turnover on labor productivity. Their results show that organizational productivity increases when facing low levels of turnover, and after a peak is reached, it harms the performance of the Belgian companies (De Winne et al., 2019).

Examining whether the respective country correlates with employee turnover does not occur at all or only very sporadically. The country itself does not play a significant role in the research; instead, data from a specific country are primarily collected and analyzed. Furthermore, the research focuses primarily on companies and individuals from the USA.

2.4 Startup

Start-ups have become an integral part of the corporate landscape in recent years and are driving economic growth (Jurgens, 2022). Start-ups are known for their innovative ideas and unique approaches to solving problems. To understand what a startup means in the later course of this thesis, this literature review will discuss various definitions to gain an understanding of the term.

A company must be characterized by three characteristics to be called a startup. These three characteristics are age, innovation, and growth. To be considered a startup, the company must be at most ten years, it should be innovative with the products, services, or technologies it offers, and it should be able to show remarkable growth in terms of revenue or the number of employees. Here, age is assigned as the most important because these characteristics must be fulfilled, whereas, of the other two, only one more must be fulfilled (Haag, 2021).

Also, a startup is a newly established company that focuses on creating and introducing innovative products or services to the market to make them appealing and indispensable to customers. They aim to improve upon existing products or introduce new ones and often disrupt traditional industries (Baldridge & Curry, 2022).

It has been suggested that a company's age is a crucial indicator in determining whether it is classified as a startup (Zaech & Baldegger, 2017). As such, this paper will also consider a company's age as a factor in determining whether it is considered a startup. However, there is no universally accepted definition of a startup (Haag, 2021), and different researchers use different ranges of years to classify a company as a startup, for example, 5-12 years (Bruneel et al., 2010; Pellegrino et al., 2012; Zaech & Baldegger, 2017). Therefore, the age of a company is one of the most important but not the only indicator to classify a company as a startup.

3. Current state of research

As described earlier, turnover is a significant challenge every organization faces, mainly due to the demographic change in the population of the world, skill shortage in the labor workforce as well as the trend of faster churning and resignation (Clark & Ritter, 2020; Kulik et al., 2014; E. S. W. Ng & Burke, 2005; Taylor & Bisson, 2020). In addition, employee turnover is already a topic of research that has been studied for several decades (P. W. Hom et al., 2017). In this context, different types of turnovers were elaborated, as well as theories on how turnovers occur. Additionally, various determinants were investigated to predict employee turnover.

These determinants include job satisfaction, organizational commitment, company size, tenure, gender, industry, and the country, among many others. While much research has been done on job satisfaction and organizational commitment, determinants such as company size, tenure, gender, industry, and country have been less studied. These have often only been examined as control variables, or a moderating effect has been analyzed.

Research on the relationship between firm size and turnover has focused primarily on using firm size as a control variable. Studies have found that larger companies are more likely to offer better development opportunities and higher wages, which can affect the turnover rate (P. G. Benson et al., 1987). Several studies found a positive relationship between firm size and turnover, with correlations of r = .26 and .1, respectively. Another study by Hancock et al. (2013) found that firm size moderates the relationship between collective turnover and performance, with larger firms having a stronger negative relationship than smaller firms. A meta-analysis by Rubenstein et al. (2018) found a weak effect of firm size (r = .03) on turnover. Overall, while company size is mainly used as a control variable, it is positively or negatively correlated with turnover, depending on the study.

Tenure, i.e., the length of time an employee has been with the employer, is a determinant frequently studied in the literature. Studies have consistently found a negative relationship between tenure and turnover, with longer tenure making turnover less likely. This relationship can be seen in companies and countries like the U.S. and Japan. Moreover, the human capital theory states that firm-specific human capital develops with an employee's tenure, affecting a company's effectiveness. Also, a negative moderating effect is found between workforce productivity and voluntary turnover (Yanadori & Kato, 2007). The meta-analysis by Rubenstein et al. (2018) examined over 110 studies, providing an average correlation of -.27 between tenure and employee turnover.

For the gender of employees, the literature generally suggests that women tend to have higher turnover rates than men (CEDEFOP, 2020; Cotton & Tuttle, 1986; P. Hom et al., 2008; Maurer & Qureshi, 2021; Weisberg & Kirschenbaum, 1993), but there are also a few studies that have found the opposite (Lyness

& Judiesch, 2001). Reasons for this discrepancy may include differences in home and work responsibilities and obstacles women face in their careers (P. Hom et al., 2008). However, some researchers have found that these differences in turnover rates are due to factors such as age, salary, or tenure rather than gender (G. B. Lewis & Park, 1989). Several meta-analyses have been conducted for employee turnover, including gender as a determinant. Nevertheless, the correlation between those two differs across the studies from -.11 (Griffeth et al., 2000) to .00 (Rubenstein et al., 2018). Hence there is no final. Therefore, no clear result shows the correlation between gender and employee turnover.

Diversity is a determinant that has shown mixed results in research. Leonard and Levine (2006) studied more than 800 workplaces. They found that diversity had little or no significant impact on employee turnover when diversity was considered in terms of gender, race, and age. However, their study also found that women were more likely to leave jobs with high diversity than jobs with a high percentage of men or women (Maurer & Qureshi, 2021). In addition, diversity can have spillover effects that reduce collective employee turnover.

Furthermore, the industry is used to predict turnover, and turnover rates can vary drastically across industries. For example, the state and local education sector had the lowest turnover rate in the U.S. at 16%. In contrast, the accommodation and food services industry had the highest turnover rate at over 86% in 2021 (U.S. BUREAU OF LABOR STATISTICS, 2022). Nevertheless, these variations can also be seen on a global scale. The technology and media sector has a turnover rate of nearly 13%, while the government sector has the lowest turnover rate at around 8.4% (G. Lewis & Soroñgon, 2022). These vast differences in 2021 are probably influenced mainly due to the Corona crisis in 2021. However, many studies focus on a specific industry rather than comparing different industries simultaneously. A meta-analysis by Hancock et al. (2013) found that the industry in which a company operates has a significant moderating effect on the relationship between employee turnover and performance, with different industries being affected to different degrees. A possible reason might be the interchangeability of knowledge in some industries (Shaw, Duffy, et al., 2005).

The country is a determinant used less frequently in the literature. Only a few studies have been conducted considering the country in their analysis. Most research has focused on turnover in the U.S. (Collins et al., 2015; S. Lee, 2018; Skelton et al., 2019; Stamolampros et al., 2019). Nevertheless, a few studies also compare countries. For example, Sender et al. (2018) compared turnover between Switzerland and China. Hence turnover is less investigated between several countries than investigating turnover in one single country, especially the U.S.

A company is considered a startup when it is younger than ten years, innovative with the products, services, or technologies it offers, and able to show tremendous growth in revenue or the number of employees. However, there is no universally accepted definition of a startup, and different researchers use different ranges of years to classify a company as a startup (Haag, 2021). Age is considered the most critical indicator for a startup, but it is not the only one (Zaech & Baldegger, 2017). Nevertheless, research on employee turnover in the context of start-ups is not present yet.

Several determinants have been studied concerning employee turnover, including job satisfaction, organizational commitment, company size, tenure, gender, industry, and country. Research has found mixed results for these determinants, with some studies finding a positive correlation and others finding no correlation. Studies on company size and tenure have found a positive correlation with turnover, while studies on gender and diversity have mixed results. The industry in which a company operates has also been found to have varying turnover rates across different industries. Further research is needed to fully understand the relationship between these determinants and startup employee turnover.

3.1 Research gap and research question

Many studies have been conducted on employee turnover and have shown several correlations between individual determinants, like job satisfaction, tenure, or company size. By doing so, the research is concentrated on investigating these determinants in established companies rather than start-ups. Hence, already investigated determinants need to be investigated in the context of start-ups. In addition, determinants are not necessarily stable and can fluctuate over time (Kammeyer-Mueller et al., 2013). In the context of start-ups, it is also reasonable to consider other determinates than the already established ones. The research gap of this thesis is to investigate the effects of various established and new determinants on turnover in startup companies. Hence, the research questions are:

1. *Do establish determinants of turnover have a similar effect on start-ups as on already established firms?*
2. *Can newly developed determinants significantly predict turnover in start-ups?*

The findings of this paper are expected to extend the literature around the explanatory power of determinants in start-ups and provide new determinants in the context of newly established firms to predict turnover.

3.2 Hypothesis

Based on the literature review and the current state of research in the literature about employee turnover in start-ups, a model for the thesis can be developed. The literature has already investigated several determinants for turnover for already established firms like job satisfaction, company size, tenure, or gender (Alterman et al., 2021; Rubenstein et al., 2018; Shaw, Gupta, et al., 2005; Skelton et al., 2019; Yanadori & Kato, 2007). One characteristic of start-ups is that they are characterized by great uncertainty, unlike already established companies. This is illustrated by the fact that nine out of ten start-ups fail for various reasons (Patel, 2015). As a result of this uncertainty, previously studied determinants of employee turnover in start-ups show a different effect than in already established companies. Furthermore, they are characterized by excessive growth rates (Jurgens, 2022), which conversely means they need to gain employees rather than lose the existing ones. Additionally, other determinants might be useful to consider in predicting turnover in the unique context of start-ups. For example, the funding volume a startup has gathered. Funding volume is the total amount of money a startup has received. There are several types of funding, e.g., crowdfunding, angel investment, loans, or venture capital (Calopa et al., 2014; Paschen, 2017). Regardless of the type of financing, it is crucial for a startup because it enables the company to operate. A startup needs employees to provide the service or product to ensure business operation. Funding is needed for a startup to pay employees, which is why funding can also be seen as a security aspect for the employees. The funding ensures payment and, thus, the business operation of the startup. As described, early payment motivates employees to stay with an employer and provide their labor to the employer (P. W. Hom et al., 2019). In the case where a startup can only raise a small amount of funding, this puts employees at risk of not being able to get paid. As explained before, payment is a motivating factor for employees, but if employees are not paid, this leads to dissatisfaction. Dissatisfaction among employees in turn leads to them leaving the startup (P. W. Hom et al., 2019; Porter & Steers, 1973).

Conversely, this also eventually means the cessation of business operations. Hence, the funding volume of a startup might have an impact on the turnover of the startup. Based on these considerations in conjunction with other studies, the first hypothesis is:

H1: *A small funding volume for a startup will result in a higher employee turnover rate, whereas a high funding volume will result in a lower employee turnover rate for a startup*

Nevertheless, another factor that might influence the turnover of a startup is the construct of innovativeness. Start-ups tend to enter the market with a disruptive or innovative business model/idea (Spender et al., 2017). By doing so, they want to revolutionize the market or a market segment by offering a new and innovative service or product. Moreover, many people, especially younger generations, prefer innovative companies over traditional ones (Slaski, 2018). Therefore, employees are less likely to leave an innovative company than a traditional one. Hence, it can be assumed that an innovative startup has a lower turnover rate than a less innovative one. Various factors can measure a company's innovativeness, like the amount spent on R&D, the intellectual property a company owns, or the provided culture in the company (Anzola-Román et al., 2018; Taques et al., 2021). Intellectual property can be measured by the number of trademarks or patents a company owns. The more trademarks or patents a company owns, the higher the intelligent property, at least, which is protected.

Nevertheless, it must be stated that owning a trademark or patent does not imply anything about the quality of intellectual property. However, as people tend to stay longer at an innovative company, it can be assumed that companies with more intellectual property have lower turnover rates. This implies that patents and trademarks might influence the employee turnover rate of a company. Based on these considerations, the second hypothesis is:

H2: *The more trademarks and patents a startup owns, the lower the employee turnover rate will be, and the lower the number of trademarks and patents a startup owns, the higher the employee turnover rate will be*

After developing the two above-mentioned hypotheses, the following conceptual model can be developed (see figure 2).

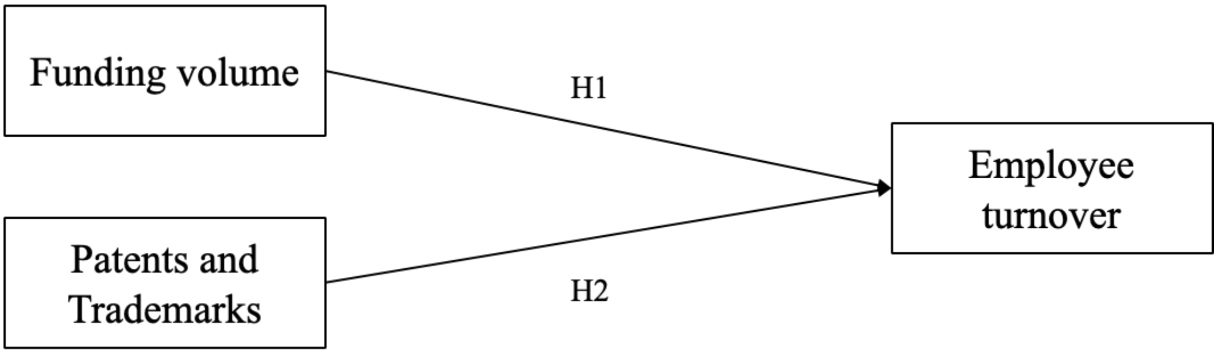


Figure 2 conceptual model

4. Methodology

In order to analyze the previous hypotheses, two data sets are merged into one. After that, this data set will be analyzed statistically. This chapter deals with the approach for the evaluation. For this purpose, a description of the basic data sets will be carried out first, followed by an explanation of the further processing of the data sets. Subsequently, a statistical regression analysis is performed, and the Gauss-Markov assumptions are checked.

4.1 Data collection

The data analysis was based on two different data sets, the first consisting of company information gathered from Crunchbase and the second one including people profiles from LinkedIn. Crunchbase is a commercially US-based database company that has provided company information about start-ups since 2007. It has increased significantly in size over the last couple of years, gathering information about more and more companies. Especially in startup academia, the Crunchbase database gained popularity among researchers (Dalle et al., 2017). Crunchbase provides various information about companies from different areas. Starting with general information about the company, like size, headquarter, region, and funding date. Besides this general information, several more insights are provided, including financial data like funding volume, number of investors and their type (business angel, venture capital, etc.), number of investment rounds, as well as acquisitions, or technical data, like monthly visits and number of patents and trademarks. Also, information about people related to the company is provided, like names of founders, members of the board, or contact possibilities (Dalle et al., 2017). For this master's thesis, a Crunchbase data set consisting of over 59.000 records was provided by Bright Data. Bright Data is a web data platform that provides ready-to-use data sets for download (Bright Data, 2022). The gathered data set consists of up to 86 attributes for each company, out of which eleven were used for further processing. From the data set, the company's name, a unique id, the funding date, the region (EU), the country code, as well as the operating status and IPO status, were gathered. In addition, the funding volume of each company, the number of patents and trademarks, the industry, and the number of employees were extracted. A description of how the data set was processed can be found in more detail in the next section.

Data sets from Crunchbase are often linked to other data sources for research purposes (Dalle et al., 2017). Kaminski et al. (2019) linked a Crunchbase data set with KickStarter to investigate if there is a long-run relationship between crowdfunding and venture capital investments (Kaminski et al., 2019). Other researchers linked the database with Twitter to analyze the temporal and self-focus and how it affects the rise of their funding (Tata et al., 2016). Banerji and Reimer (2019) linked a Crunchbase data set with data from LinkedIn to investigate if well-connected founders are more successful than less-connected ones (Banerji & Reimer, 2019).

In this master thesis, the Crunchbase data set is linked to a LinkedIn data set. LinkedIn is a social network that is used for business purposes, like connecting with business partners and enabling new business opportunities. In the network, people create profiles about themselves and provide various information like job positions, certifications, experience, qualifications, and more. LinkedIn is used as a data source in various fields in the literature. For example, Johnson and Leo (2020) use LinkedIn data to analyze the effectiveness of LinkedIn in the job search of people (Johnson & Leo, 2020). Moreno-Delgado et al. (2020) used LinkedIn data to determine the feasibility of ranking universities based on the number of graduates who work in top companies (Moreno-Delgado et al., 2020). The LinkedIn data set used in this master thesis consists of over 150 million publicly available profiles of people provided by Bright Data.

The previously mentioned data set was filtered before downloading and further processing. Only profile information of people were downloaded that have their current job at one of the companies from the Crunchbase data set or worked for one of them in the past. To match profiles and companies the company id from the Crunchbase data set was used. For each profile, the data set provides up to 115 attributes, but only the attributes person's name, the current company and the working experience are used and processed further. From the person’s name the gender is estimated, and from the current job position as well as the previous positions it is checked if and how long the person had worked for a certain company. A more detailed explanation of this will be provided in the next section.

All in all, two different data sets, one from Crunchbase and one from LinkedIn, both provided by Bright Data, are merged based on the company id out of the Crunchbase data set. The data set is processed further by using Python, which will be explained in more detail in the next section.

4.2 Data processing

The two previously described data sets were combined and further processed using Python. The two data sets were processed by cleaning, filtering, and transforming the data set using Python`s data manipulation libraries to ensure a final data set that consisted of accurate and relevant data. This chapter briefly describes the approach to data preparation. The processing was done in three different Python scripts to ensure faster processing times.

The first script processes the raw Crunchbase data set provided by BrightData, with around 59.000 records. Firstly, the script reads in the raw data set as a data frame. From this data frame, only specific columns are selected and assigned to a new data frame (“dfData”). Moreover, some columns are added to the “dfData” data frame with an initial value of zero. Then the script opens the raw Crunchbase data set again to read each row as a dictionary. After that, it is iterated through each row, and several pieces of information, if they exist, about the funding volume, the number of trademarks, and the number of patents, are extracted. Thereafter the extracted information is updated in the new data frame. After all rows of the raw data set had been processed, in the new data frame, all rows where no funding volume and either no trademarks or patents were found were dropped. The data frame consists of 3043 rows, where each row stands for a single company, and 11 columns, where each column represents a particular attribute, like the company name or number of trademarks. Figure 7 in the appendix provides a detailed view of the script.

Since processing the new data frame with 3043 rows is much faster than the original raw data set, a second Python script was developed and used, which processes the new data frame further. The second script was used to prepare the provided data frame for the statistical analysis. Firstly, the preprocessed data frame of 3043 rows was read in and saved as a data frame containing the same data. After that, the preprocessed data frame was opened again, and it was iterated through each row of it. In each row, some editing was done regarding the format. A categorical variable (1-10, 11-50, etc.) was created from the column for the number of employees. In the column for the industries, all values were extracted that contained an industry value, and from the foundation date, the foundation year was extracted. Later on, each company was assigned to an industry sector based on the NACE Revision 2 code classification (Eurostat, 2008). All values were saved in a new column. All columns that were no longer needed and contained the original format were removed from the data frame. Finally, the order of the columns was adjusted, and the data frame was saved. Figure 8 in the appendix provides a detailed view of the script.

In the third and final Python script, the merger between the processed Crunchbase data frame and the raw data set containing the LinkedIn profiles is done. Hence the preprocessed Crunchbase data set is read in and saved in a data frame variable. Several new columns, which will contain data later or are only dummy columns to enable several calculations, are added to the data frame. The raw LinkedIn data set is loaded and opened in Python. Afterward, it is iterated through each row of the data set, several data are extracted, and several functions are executed.

The first function is done to extract data for the turnover calculation later in the script. The function is extracting the person's start date and end date at one of the companies from the processed Crunchbase data frame. The extracted start and end date are then used to update the regarding column by one, which is the value of how many people have started or ended their job in a particular year by that company. This will later be used to calculate the turnover rate of a particular year. Figure 9 in the appendix provides a detailed view of the script.

The second function, executed when the previous function was successful, calculates the number of employees and the average tenure at the company. For that, the person's duration at the company is extracted and transformed into a monthly format. These months are then added to the total months for that company. Moreover, a counter value is created by always adding one if a duration is found. The average tenure is calculated later in the code since all rows of the LinkedIn data set need to be processed before further calculations can be done. The last function is used whenever the previous two functions work successfully and will gather the gender of the employee by using a Python library called gender-detector. Based on a person's first name, which is transferred to the function, it will guess the gender of that person. The guessed gender is used to update the regarding column (male, female, or unknown) by one to get the number of employees for each gender. Based on the number of employees of a particular gender, the diversity is calculated later. Figure 10 in the appendix provides a detailed view of the script.

After all rows of the LinkedIn data set have been processed, further calculations can be done. These calculations include the average tenure at the company, the diversity based on the proportion of female employees, the total number of trademarks and patents, and the annual turnover rates. Moreover, more dummy columns are created to calculate the company's average turnover. The dummy columns include the value one if the annual turnover rate exceeds zero. After that, the sum of all dummy columns is calculated to provide the correct denominator for calculating the average turnover rate. The final steps of the script are dropping all the dummy and unnecessary columns and saving the data frame. Figure 11 in the appendix provides a detailed view of the script.

All in all, the final data frame consists of 3043 companies, and for each company, 28 attributes, from the company name to the average turnover, are provided. Nevertheless, not all companies can be used in the statistical analysis due to missing values in the data basis For a detailed view of the full third script, see figure 12 in the appendix.

5. Data analysis

In the following chapter, the statistical analysis of the preprocessed data will be described. This is done to check the previously provided hypothesis and answer the research questions of this master thesis. The analysis of the processed data will be done by using the software STATA. The analysis will start with a descriptive analysis of the sample, followed by checking several assumptions before performing four regression analyses

5.1 Descriptive analysis

In the first step of the statistical analysis, some irrelevant variables, such as the company name or the company id, were removed for the evaluation. The remaining variables were labeled with an explanation description of themselves. The data of the processed data set showed that for some companies, no average employee turnover could be calculated. For this reason, the companies for which no employee turnover could be calculated were removed from the sample. This resulted in a reduction of the original sample n = 3042 to n = 1375. Moreover, it was checked if other variables had any missing values (see table 1, below).

Table 1 Overview of missing observations in the sample

Variable	Missing	Total	Percent Missing
Founding_year	0	1,375	0.000
Country_code	0	1,375	0.000
Size	0	1,375	0.000
Industry	0	1,375	0.000
Funding_volumee	0	1,375	0.000
No_trademaarks	0	1,375	0.000
No_patents	0	1,375	0.000
Avg_tenure	0	1,375	0.000
Total_IP	0	1,375	0.000
Average_Turnover	0	1,375	0.000
Diversity	32	1,375	2.330
Nace	74	1,375	5.380

As shown in table 1, there were some missing observations for the variables Diversity and Nace. Hence these missing observations were also removed from the sample, which reduced the sample size from n = 1375 to n = 1296 observations for each variable as the final sample size.

The descriptive statistics for each variable are in table 2 below. As shown in the table, the variable "Nace" is missing. This was done as it is a non-numeric variable and was only used to convert the "Industry" variable into a numeric, categorical variable in STATA. For example, if the NACE code was equal to "A", this resulted in category 0 in the variable "Industry".

The table shows, that the average turnover ranges from 0% to 350%, which reflects a broad spectrum. The mean over all observations is about 32%, suggesting that some extreme values might be present. The standard deviation is about 29.5, which in contrast to the funding value, represents a very low dispersion to the mean value. The funding value shows a minimum value of only 736 US dollars up to a maximum value of over 2.4 billion US dollars. The spread (SD = 1.502e+08) around the mean value of over 34 million dollars is correspondingly large.

On the other hand, the number of trademarks and patents show a maximum value of 27 and 31, respectively. Furthermore, it is understandable that both variables show a minimum value of zero since, as described above, only companies that must show either a trademark or a patent were considered. This also explains the minimum of one for the variable Total_IP since it represents the sum of both variables. Total_IP has a standard deviation of about three and a mean of about 3.7. In addition, the control variables move in a comprehensible range. Country_code, Size, and Industry are categorical variables, where for example, a one for Industry would stand for the industry sector manufacturing. For this reason, it can be deduced for the company size that an average startup in the sample belongs to category one, which corresponds to a size of approximately 11-50 employees. An average startup from the sample was founded in 2018 (Founding year M = 2018,204) and operates in the industry "Information and communication" (Industry M = 8). An explanation of each value for the categorical variables founding year, country code, size, and industry can be found in table 6 of the appendix.

The ordinal variable, Avg_tenure, describes the average duration an employee has worked for a company and has values between 1 and 64 months with an SD of 8.7. Since these companies were founded from 2017 up to and including 2022, the maximum duration an employee could have spent there would be around 72 months. Nevertheless, on average, employees spend just under 20 months at a startup from the sample, corresponding to approximately 1.5 years. The diversity shows a minimum value of 0 and a maximum value of 100. This suggests that there are companies that have only women employed (Diversity maximum = 100), while other start-ups have no women in their workforce (minimum = 0). On average, the proportion of women in the start-ups is around 30%, with an SD of around 23.8. When calculating the proportion of women, the assumption was made that the proportion between men and women would be the same for people who were classified as unknown gender. Hence the number of unknown genders classified employees were deducted from the employee count, and then the proportion of women was calculated.

Table 2 descriptive statistics of the sample

Variable	Obs	Mean	Std. Dev.	Min	Max
Average Turnover	1296	32.086	29.499	0	350
Funding volume	1296	34479177	1.502e+08	736	2.429e+09
No trademarks	1296	2.265	2.686	0	27
No patents	1296	.755	2.34	0	31
Total IP	1296	3.021	3.707	1	42
Founding year	1296	2018.204	1.185	2017	2022
Country code	1296	4.284	4.238	0	12
Size	1296	1.008	.948	0	4
Avg tenure	1296	19.792	8.772	1	64
Diversity	1296	30.272	23.795	0	100
Industry	1296	8.61	3.578	0	16

Next the assumptions of normal distribution and outliers will be checked. A normal distribution is a symmetrical distribution of values around the mean, which results in a bell-curve appearance (Weinberg

& Abramowitz, 2020). Considering the Central Limit Theorem (CLT), a normal distribution can be assumed if the sample size is sufficiently large. In alignment with other researchers, a sufficiently large sample size is present when n >= 100 observations (Weinberg & Abramowitz, 2020). Hence, a normal distribution is expected with a total sample size of over 1200 observations.

Nevertheless, there are different opportunities to check for a normal distribution, like statistical testing with the Shapiro-Wilk test or data visualization by creating histograms of the variables. Besides assuming a normal distribution will be present, it will be checked for the dependent variable (Average Turnover) and the independent variables (Funding value and Total IP). Therefore, a histogram for each of the variables is done. These histograms can be seen in figure 3.

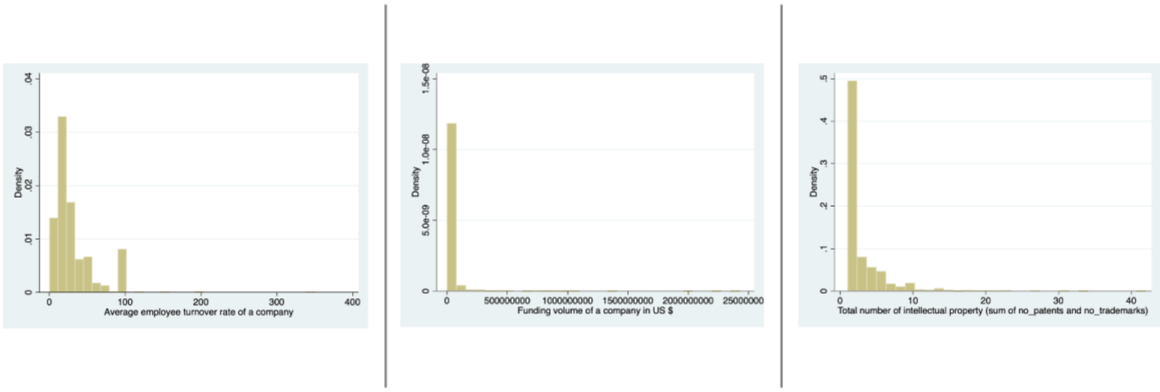


Figure 3 histogram of DV and IVs

As shown in the figure above, all histograms show that all variables are rightskewed. This shows that besides the assumption of a normal distribution, it is not given in this sample, and there is no symmetrical distribution. Therefore, the variables will be logarithmized to generate a better normal distribution. After the logarithmization of the variables, the distribution looks as shown in figure 4.

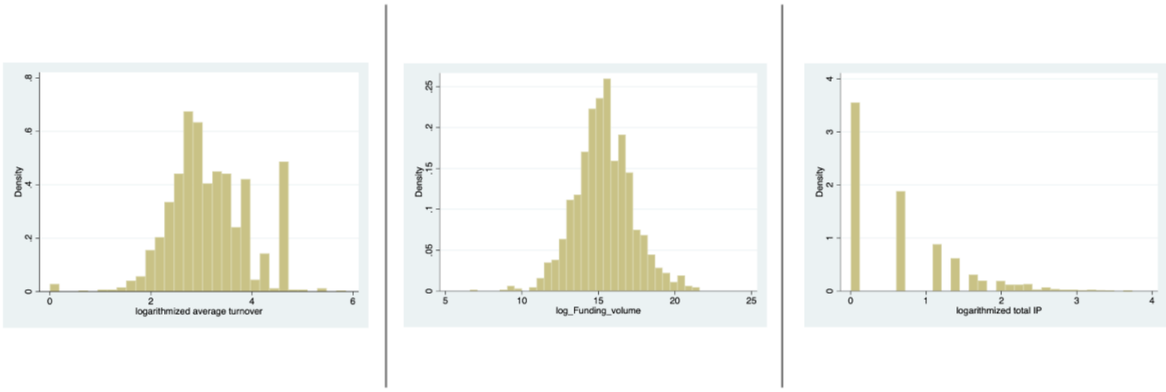


Figure 4 histogram of logarithmized DV and IVs

After checking for a normal distribution, it is also necessary to check for outliers. Outliers differ heavily from the rest of the data and, therefore might influence the mean value of the standard deviation (Weinberg & Abramowitz, 2020). STATA defines a value as an outlier in a boxplot graph when it falls more than 1.5 interquartile ranges beyond the 75th or 25th percentile. Another way to detect outliers is by using the z-criterion. The z-criterion shows how many of standard deviations the estimation is away from the mean (Weinberg & Abramowitz, 2020). As shown in figure 5, there are several outliers for the DV and IVs.

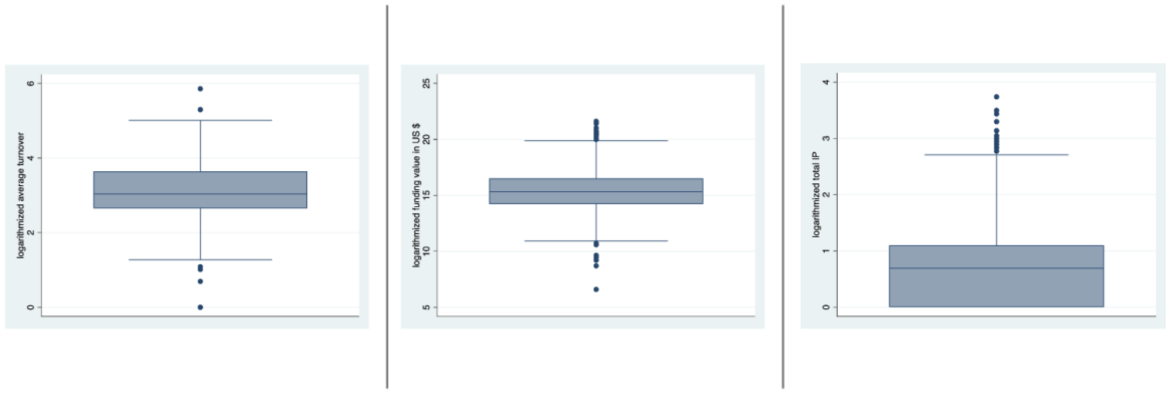


Figure 5 box-whisker plots of DV and IVS

This is also shown in the calculated z-score of the variables. The threshold for the z-score was set to three (Nevil, 2022). This threshold was chosen because there is a broad spectrum in the investment volumes of start-ups. Some start-ups receive very little capital, while a few have a very large investment volume. To generate better information, this spectrum should be reduced by removing the extreme values from a threshold value of three. Hence only values that show a very large z >= 3 are identified as an outlier and will be dropped from the sample. Considering the cutoff threshold value of three for the average turnover, nine values were identified as outliers, for the funding value eleven, and the total intelligent property ten observations were identified as outliers. The table showing the z-values of each variable can be found in the appendix (table 7-9). As these observations impact the mean value and standard deviation, the observations which fulfill the z-criterion will be dropped. This reduced the sample size to n = 1263.

Moreover, the correlations between the different variables were investigated by generating a correlation table (see table 3, below). As shown in the table, the logarithmized average turnover has no high correlations with either the independent or control variables. The average employee turnover shows the highest correlation with the independent variable funding value ($r = -.165$, $p < .01$). The second highest correlation is shown with the control variable average tenure ($r = -.135$ $p < .01$). Moreover, all variables show a negative correlation to employee turnover besides diversity. However, all variables show a significant correlation (at least $p < .1$) to the average employee turnover besides the industry ($p = .427$) and the country code ($p = .361$) (see table 3, below).

The independent variable funding value shows a quite high correlation of $r = .638$ with the startup size ($p < .01$). This is quite reasonable because larger start-ups probably need more money to provide payment for their employees and enable the operation of the business. Moreover, the funding value shows a significant correlation ($r = .290$, $p < .01$ with the other independent variable total intelligent property. The variables founding year, country code, average tenure, diversity, and industry show a negative correlation to the funding value, and all show a significant p-value besides the industry. The industry shows a non-significant correlation to the funding value ($r = -.022$, $p = .427$) (see table 3, below).

The total intelligent property variable shows a quite high correlation with the number of trademarks and the number of patents ($r = .749$ and $r = .54$, $p < .01$, respectively). This does not come as a surprise, as this variable is the sum of both variables, as mentioned earlier. Nevertheless, the correlation table shows that there is no multicollinearity between the IVs. All control variables show a significant positive correlation with the real intelligent property. Only diversity and the industry show a positive correlation but on an insignificant level (see table 3, below).

Within the control variabels, the highest correlation exists between the company's founding year and the average tenure at a company. They show a significance level of $p < .01$ and a correlation of $r = -.308$. However, within the control variables, there are no more quite high ($r >= .4$) correlations (see table 3, below).

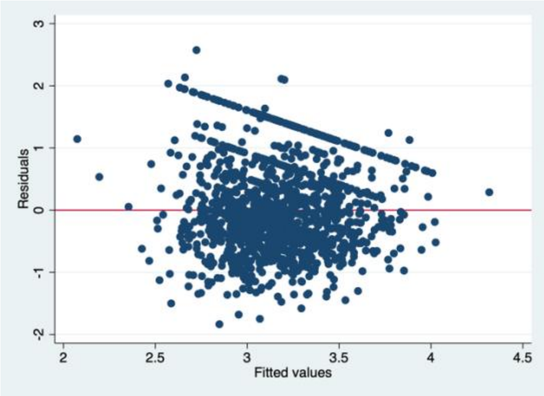
Table 3 Pearson's correlation table of variables with significance levels

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) log_average_turover	1.000										
(2) log_Funding_volume	-0.165	1.000									
	(0.000)										
(3) log_Total_IP	-0.096	0.290	1.000								
	(0.001)	(0.000)									
(4) No_trademarks	-0.070	0.278	0.749	1.000							
	(0.013)	(0.000)	(0.000)								
(5) No_patents	-0.071	0.096	0.540	0.040	1.000						
	(0.011)	(0.001)	(0.000)	(0.154)							
(6) Founding_year	-0.075	-0.077	-0.192	-0.119	-0.127	1.000					
	(0.008)	(0.006)	(0.000)	(0.000)	(0.000)						
(7) Country_code	-0.026	-0.187	-0.074	-0.066	-0.067	-0.073	1.000				
	(0.361)	(0.000)	(0.009)	(0.019)	(0.017)	(0.009)					
(8) Size	-0.128	0.638	0.178	0.256	-0.017	-0.056	-0.094	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.553)	(0.047)	(0.001)				
(9) Avg_tenure	-0.135	-0.072	0.103	0.008	0.138	-0.308	0.092	-0.091	1.000		
	(0.000)	(0.011)	(0.000)	(0.771)	(0.000)	(0.000)	(0.001)	(0.001)			
(10) Diversity	0.058	-0.053	0.014	0.047	-0.026	-0.089	-0.016	-0.012	-0.114	1.000	
	(0.040)	(0.059)	(0.621)	(0.094)	(0.355)	(0.002)	(0.569)	(0.679)	(0.000)		
(11) Industry	-0.022	-0.002	0.021	0.011	-0.003	-0.039	0.002	-0.038	0.063	0.069	1.000
	(0.427)	(0.941)	(0.456)	(0.690)	(0.907)	(0.167)	(0.937)	(0.177)	(0.025)	(0.015)	

5.2 Regression analyses

Four regression analyses were performed to test the hypotheses and further interpret the data, which will now be compared. As for the models the DV and the IVs were entered in a logarithmic form. Therefore, a log-log model was developed and the necessary interpretation is needed (see table 4, below). The control model consits of all control variables an the DV average employee turnover. The “Model 1” in the table below included only the independent variable funding value and all of the controls, whereas “Model 2” included only the IV of total intelligent property and controls. The last model is the complete model containing all controls and independent variables (see table 4, below).

But before the models can be compared, the two assumptions need to be checked. First, the assumption of no multicollinearity and the second assumption of homoscedasticity needs to be checked to enable an unbiased interpretation of the results. No multicollinearity assumes that no regressor is a linear function of another regressor (Weinberg & Abramowitz, 2020). Partially this can already be seen in the pairwise correlation table above (see table 4, below). Nevertheless, the variance inflation factor is measured to check if some variables influence each other by their interaction. The test showed a mean-variance inflation factor of 2.2. Hence it can be assumed that the models are not driven by multicollinearity, and the assumption of no multicollinearity is fulfilled. The second assumption which needs to be checked is homoscedasticity. It assumes that the variance of the errors or residuals in a regression model is constant across all independent variable levels (Weinberg & Abramowitz, 2020). Hence the spread of the data points around the regression line is consistent across the range of the independent variable. To test for homoscedasticity, the residuals can be plotted or the Breush-Pagan/Cook-Weisberg test for hetersokedasticity can be (see figure 6, below).




```
Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of log_average_turnover

H0: Constant variance

    chi2(1) =  20.78
Prob > chi2 = 0.0000
```

Figure 6 plotted residuals to fitted values and BP test

As the fitted values plot shows, some heteroscedasticity can be assumed. This is confirmed by the Breusch-Pagan/Cook-Weisberg test, which shows that the null hypotheses of constant variance can be rejected as the p-value is smaller than $p < .01$.

The control model shows a general r-squared of .123, which means that around 12% of the variance in the DV of average employee turnover is explained by this model. Moreover, the categorical control variables founding year, country code, size, and industry are included in the model. An overview of the meaning of the individual values can be found in the appendix (see appendix table 6). The four categories (2018, 2019, 2020, and 2021) founding year negatively influences the average employee turnover compared to the base category of 2017 ($r = -.254, -.352, -.282, -.452$, $p < .01$). Only the category 2022 does not have a statistically significant influence on turnover ($r = -.104$). The following control variable is the country code representing the country where a company was founded. The base category is Great Britain (see appendix table 6), and only value 5 (Switzerland) shows an influence on the DV ($r = -.196$, $p < 0.1$). All other countries do not show a significant influence compared to the base category.

Furthermore, the control variable size shows a significant influence on the average employee turnover of a company. The influence ranges from $r = -.544$ to $-.297$ at a significance level of at least 5%. This means that the bigger a startup is, the lower the employee turnover rate will be. For the categorical variable industry, only category 6 showed a significant influence on employee turnover ($r = .502$, $p < .01$. When a startup operates in the transportation and storage sector, it will have a higher turnover rate than the agriculture, forestry, and fishing base category. The average tenure has a significant ($p < .01$) influence of $r = -.0156$ on the turnover. The diversity (proportion of females in the startup) does not show a significant influence. The intercept of the control model is 3.777, meaning that every startup has at least an average turnover of 43%.

Comparing the second model (Model 1), it also includes the first independent variable funding value. The funding value shows a significant influence ($p < .01$) on a company's average employee turnover rate with -.0703. This shows that there is a negative relation between the two variables. Moreover, there is a slight increase in the explained variance of the DV ($r^2 = .137$). Hence there is a slight improvement of 1% between the model with only controls and the model with the funding value. The control variables in the second model show quite similar values and significance levels as in the model with controls only. The only difference is that country code 12 (Others) also became significant, whereas the size 4 (more than 250 people) in a company became insignificant. Also, the intercept of the model raised to 4.816 and remains significant at $p < .01$.

The third model ("Model 2") shows similar values as the control model and the second model (model 1). Here the IV total intelligent property shows a statistically significant ($p < .05$) influence on the average employee turnover rate. Nevertheless, the influence with $r = -.069$ is relatively small. The r^2 is only 0.4% higher than the control model but 1% lower than model 1. Therefore, less variance in the DV is explained by only considering the total intelligent property compared to only considering the funding value and controls. Besides that, the control variables show no significant change only category four of the control variable size became significant ($r = -.256$, $p < .1$).

Lastly, the entire model is compared to the other three models. The entire model shows the highest r^2 value with an explained variance in the DV of 13.8%, which is only slightly higher than model 1 (13.7%) and one percent higher than the control model and model 2. Hence the entire model explains the most variance in the DV between all the models. Moreover, the independent variable funding value decreases a bit but remains significant ($r = -.065$, $p < .01$). The other independent variable became insignificant in the entire model compared to model 2 ($r = -.041$ $p > 0.1$). The control variable's founding year shows similar coefficients and significance levels, and the funding year 2022 remains insignificant. The entire model shows, equal to model 1, that country codes 5 and 12 are significant ($p < .05$ and $p < .1$, respectively) compared to the base category.

In contrast, the other two models showed only country code 5 as significant. For the control variable size, it is the other way around. The control model and model 3 show for size 4 (250 people and more) a significant influence ($p < .05$ and $p < .1$, respectively), while the entire model show and model 1 show category four as insignificant. All other sizes show a significant negative significant impact. The entire model shows the same results for the industry sector, where only category 6 shows a significant influence ($r = .481$, $p < .01$). Moreover, the intercept shows a quite similar value to model 1 with only the funding value and the control, but slightly decreased from 4.816 (model 1) to 4.781 (entire model).

Table 4 regression analysis for all models

VARIABLES	(1) Controls DV: log Avg. Turnover	(2) Model 1 DV: log Avg. Turnover	(3) Model 2 DV: log Avg. Turnover	(4) Entire Model DV: log Avg. Turnover
log_Funding_volume		-0.0703***		-0.0653***
		(0.0155)		(0.0159)
2018.Founding_year	-0.254***	-0.264***	-0.257***	-0.265***
	(0.0519)	(0.0515)	(0.0518)	(0.0515)
2019.Founding_year	-0.352***	-0.373***	-0.360***	-0.376***
	(0.0600)	(0.0597)	(0.0599)	(0.0597)
2020.Founding_year	-0.282***	-0.302***	-0.307***	-0.316***
	(0.0709)	(0.0705)	(0.0716)	(0.0711)
2021.Founding_year	-0.423***	-0.453***	-0.452***	-0.468***
	(0.117)	(0.116)	(0.117)	(0.116)
2022.Founding_year	-0.104	-0.0910	-0.0779	-0.0763
	(0.300)	(0.297)	(0.299)	(0.298)
1.Country_code	0.0702	0.0639	0.0691	0.0637
	(0.0679)	(0.0673)	(0.0677)	(0.0673)
2.Country_code	-0.119	-0.110	-0.125	-0.114
	(0.0828)	(0.0822)	(0.0827)	(0.0823)
3.Country_code	-0.0816	-0.140	-0.0947	-0.144
	(0.0885)	(0.0887)	(0.0885)	(0.0887)
4.Country_code	-0.0497	-0.0764	-0.0450	-0.0717
	(0.0905)	(0.0899)	(0.0903)	(0.0900)
5.Country_code	-0.196*	-0.199**	-0.197**	-0.199**
	(0.100)	(0.0994)	(0.1000)	(0.0993)
6.Country_code	0.0182	-0.00117	0.00905	-0.00525
	(0.102)	(0.102)	(0.102)	(0.102)
7.Country_code	0.00768	-0.0396	0.00821	-0.0359
	(0.104)	(0.104)	(0.104)	(0.104)
8.Country_code	-0.174	-0.196	-0.161	-0.187
	(0.133)	(0.132)	(0.133)	(0.132)
9.Country_code	0.00306	0.00588	-0.00425	0.00131
	(0.134)	(0.133)	(0.134)	(0.133)
10.Country_code	0.0921	0.0105	0.0776	0.00757
	(0.145)	(0.145)	(0.144)	(0.145)
11.Country_code	-0.203	-0.216	-0.192	-0.208
	(0.149)	(0.148)	(0.149)	(0.148)
12.Country_code	-0.0766	-0.126*	-0.0893	-0.130*

	(0.0705)	(0.0708)	(0.0706)	(0.0708)
1.Size	-0.355***	-0.273***	-0.339***	-0.269***
	(0.0476)	(0.0506)	(0.0480)	(0.0507)
2.Size	-0.544***	-0.362***	-0.519***	-0.360***
	(0.0803)	(0.0892)	(0.0808)	(0.0892)
3.Size	-0.421***	-0.179*	-0.386***	-0.175*
	(0.0876)	(0.102)	(0.0887)	(0.102)
4.Size	-0.297**	0.0377	-0.256*	0.0388
	(0.148)	(0.164)	(0.149)	(0.164)
1.Industry	0.0975	0.100	0.0980	0.100
	(0.185)	(0.184)	(0.185)	(0.184)
2.Industry	0.259	0.271	0.245	0.262
	(0.187)	(0.186)	(0.187)	(0.186)
3.Industry	-0.00192	-0.264	0.0160	-0.235
	(0.534)	(0.533)	(0.533)	(0.533)
4.Industry	0.104	0.103	0.0697	0.0831
	(0.221)	(0.219)	(0.221)	(0.220)
5.Industry	0.143	0.132	0.119	0.118
	(0.169)	(0.168)	(0.169)	(0.168)
6.Industry	0.502***	0.496***	0.477***	0.481***
	(0.175)	(0.174)	(0.175)	(0.174)
7.Industry	0.243	0.222	0.214	0.206
	(0.173)	(0.172)	(0.174)	(0.172)
8.Industry	0.175	0.179	0.148	0.162
	(0.150)	(0.148)	(0.150)	(0.149)
9.Industry	0.196	0.229	0.161	0.205
	(0.155)	(0.154)	(0.156)	(0.155)
10.Industry	0.236	0.229	0.194	0.204
	(0.189)	(0.187)	(0.189)	(0.188)
11.Industry	0.163	0.136	0.129	0.117
	(0.172)	(0.171)	(0.172)	(0.171)
12.Industry	0.352	0.360	0.304	0.331
	(0.331)	(0.328)	(0.331)	(0.329)
13.Industry	0.229	0.198	0.200	0.183
	(0.201)	(0.199)	(0.201)	(0.199)
14.Industry	0.0935	0.114	0.0952	0.114
	(0.156)	(0.155)	(0.155)	(0.155)
15.Industry	0.165	0.151	0.144	0.140
	(0.180)	(0.178)	(0.179)	(0.178)
16.Industry	-0.237	-0.244	-0.287	-0.274
	(0.355)	(0.352)	(0.355)	(0.353)
Avg_tenure	-0.0156***	-0.0161***	-0.0153***	-0.0159***
	(0.00255)	(0.00253)	(0.00255)	(0.00254)
Diversity	0.000432	0.000181	0.000395	0.000177
	(0.000916)	(0.000910)	(0.000914)	(0.000910)
log_Total_IP			-0.0686**	-0.0410
			(0.0291)	(0.0297)
Constant	3.777***	4.816***	3.840***	4.781***
	(0.171)	(0.285)	(0.172)	(0.286)
Observations	1,263	1,263	1,263	1,263
R-squared	0.123	0.137	0.127	0.138

Standard errors in parentheses | *** p < 0.01, ** p < 0.05, * p < 0.1

6. Results

This thesis aims to check if already established determinants for employee turnover would also be relevant in the context of start-ups. As start-ups show other characteristics than established firms, like insecurity or high growth rates (Jurgens, 2022), other determinates like the funding volume or the total intelligent property a company owns were also considered. After the provided statistical analysis of the gathered data in the previous section, the findings will be interpreted in the context of the previously provided research questions (see chapter 3.1) in this chapter. The presented hypotheses (see chapter 3.2) will be checked, and the findings will be compared to the literature.

6.1 Interpretation

In the context of the interpretation, the first step is to check if support was found for the provided hypotheses in chapter 3.2. The first hypothesis deals with the funding volume's effect on a startup's average employee turnover. To operate the business, a startup needs employees and a certain amount of funding to provide payment for them. In addition, payment is a motivating factor for employees (P. W. Hom et al., 2019), which could prevent them from switching to another employer. To enable payment for their employees, the startup relies on its funding volume. Hence it was assumed that the funding volume might provide security and a motivational factor for the employees. Based on this, the first hypothesis suggested that a higher funding volume would decrease a startup's average employee turnover rate and vice versa. The analysis in chapter 5.1 shows that this hypothesis can be accepted. The analysis shows that a higher funding value statistically significantly decreases a startup's average employee turnover. An increase in the funding volume by 10% would result in a decrease of 0.65%. Hence an increase of 10% results in a relatively low decrease in the employee turnover rate. Moreover, an overview of the effects a change in the funding volume might have on the turnover rate can be seen in table 5, below.

Table 5 Overview of result in percentage change of the funding volume

Coefficient Funding Volume	Percentage increase of the funding	Reduction in average employee turnover
-0,0653	5	-0,3265
	10	-0,653
	15	-0,9795
	20	-1,306
	25	-1,6325
	30	-1,959
	35	-2,2855
	40	-2,612
	45	-2,9385
	50	-3,265
	55	-3,5915
	60	-3,918
	65	-4,2445
	70	-4,571
	75	-4,8975
	80	-5,224
	85	-5,5505
	90	-5,877
	95	-6,2035
	100	-6,53

As shown in table 5, an increase in the funding volume only slightly impacts the turnover rate. For example, an increase of 80% in the funding volume would result in a decrease of turnover of around 5% for the startup.

Start-ups tend to enter the market with a disruptive or innovative business model/idea (Spender et al., 2017). More and more people want to work in open and innovative companies compared to traditional ones (Slaski, 2018). Based on this, the second hypothesis was developed. The innovativeness of a company can be measured by the construct of intelligent property, which in turn can be measured by the number of patents and trademarks a company owns (Taques et al., 2021). Therefore, the second hypothesis suggested that more patents and trademarks would decrease the average employee turnover as employees tend to stay longer in more innovative companies. Considering the analysis in section 5.2, no statistically significant effect of an increase in the number of trademarks or patents could be found (see table 4 in section 5.2). Hence no support for hypothesis two was found, and it has to be rejected.

Comparing the findings of the control variables with the literature shows similar results for established firms as for start-ups. The size showed negative and positive correlations to employee turnover in the literature (see section 2.3.3). The analysis found support that the size can have a decreasing effect on employee turnover for start-ups up to 250 employees (category 3). When a startup has over 250 employees, size had no significant influence on the employee turnover rate (see section 5.2). Hence the company's size does not have an influence when the workforce has reached over 250 employees.

Moreover, the literature showed that an employee's tenure would reduce the turnover rate (Yanadori & Kato, 2007). The analysis also showed a similar result in the context of start-ups. However, the effect is relatively low in start-ups (see section 5.2). In contrast to the literature, it was not found that a company's diversity (proportion of females in the workforce) significantly influenced the turnover rate. This analysis could not confirm these findings, as the diversity did not show a statistically significant effect. A possible reason for that might be that only 14% of the total share of the sample had a diversity of 50% and higher. This means that most start-ups in this sample consist of a workforce dominated by men. Hence further research has to be done with a sample where the proportion of more diverse workforces is higher. The analysis confirmed the findings in the literature for the industry sector only partially, as only for one sector the data showed a significant impact on employee turnover. This sector is the transportation and storage industry sector. It is shown, that this sector increases the turnover rate of the startup ($r = .481$, $p <.01$). As the literature has shown that the turnover ratio is very different across industries, this might assume that it is not that different in the context of start-ups. As this is only a sample of around 1300 start-ups from Europe it needs to be checked with further research to enable generalizability. The last two variables which were controlled in the sample were the funding year and the country. All different founding years (2017-2022) significantly impacted the turnover rate besides the year 2022. A reason for that might be that only six countries in the sample were founded in 2022. Interesting to see is that all the founding years show a negative impact, besides economically bad situations like the corona crisis, which would assume that founding a company in those years would increase the turnover rate as companies needed to do savings to survive the overall bad economic situation. Lastly, the country of origin of the company is considered. The analysis showed that only one country compared to the base category, significantly influenced the average turnover rate. This is also in alignment with previous research about the impact of countries on employee turnover (see section 2.3.8). The data shows that only one country, Switzerland (category 5), had a significant influence on employee turnover ($r = -.199$, $p < .05$). This would mean that a startup from Switzerland will probably have a slightly lower employee turnover rate compared to start-ups from other countries. As mentioned earlier, only few research has been done about the country's influence on employee turnover, even for established companies. Therefore further research about this variable and its influence should be done to check if the current results are also valid in other research.

Hence the research questions from chapter 3.1 can be answered as follows. Some of the already established determinants show similar results in the context of start-ups, for example, the size, whereas others, like diversity, did not show similar results. Moreover, the newly introduced determinants to predict employee turnover in the context of start-ups, funding volume, and total intelligent property provided only partially significant results. The analysis showed that the funding volume could be used to predict turnover, but it has only slight economic effects for start-ups. As shown in table 5, even a very high increase in the funding volume will result in only slight decreases in the employee turnover rate. However, the second independent variable of the total amount of intelligent property is no valuable predictor for employee turnover as no significant effect was found. All in all, the whole model in section 5.2 explained around 14% ($r^2 = .138$) of the variance in the dependent variable.

7. Limitations

The analysis also shows some limitations, which will be explained in this chapter. First, it is necessary to mention that the analysis was done for only European start-ups with a founding year later than 2017, which brings up the limitation of generalizability. Hence the findings need to be further checked with start-ups from other regions, for example, with start-ups from Asia or America.

Moreover, the average turnover was calculated from LinkedIn data, bringing up two different limitations. First, no actual turnover data from the start-ups were present. The calculation was based on the findings in the previously mentioned data sets. Hence the calculated average turnover estimates the actual turnover rate one of the start-ups has. Furthermore, the data set provides the second limitation as there are probably people who work for the startup but are not included in the data set and therefore excluded from the calculations. Moreover, there are probably employees of the startup who do not have a public LinkedIn profile or no profile at all and are excluded for this reason.

The provided analysis showed a relatively low explanation of the variance in the dependent variable of employee turnover. To get better predictive power for the whole model, further variables which have previously shown significant imfleunce in other analyes should be included. For example, job satisfaction or organizational commitment would be further valuable predictors. These variables can be added to the analysis by gathering data from job portals like Glassdoor or Indeed.

8. Conclusion and outlook

All in all, the purpose of this master's thesis was to investigate employee turnover in the context of start-ups. Due to the unique characteristics of a startup, the literature should be extended by checking if already established determinants show similar correlations and significances in the context of start-ups. Furthermore, new determinants, funding volume, and the total amount of intelligent property should be analyzed to check if they could be valuable determinants in predicting the turnover rate and what effect those determinants would have on the turnover rate.

Two data sets from Bright Data were gathered, one with company information from Crunchbase, and the other containing publicly available LinkedIn profiles. Both data sets were processed using several Python scripts and merged. The statistical analysis was done by using STATA. Several assumptions, like normal distribution or multicollinearity, were checked for the variables and were fulfilled. Only the assumption of heteroskedasticity was not fulfilled, and therefore, the predictors are biased. The analysis showed that funding volume predicts turnover. Nevertheless, the effect of funding volume on the employee turnover rate can be interpreted as rather low ($r = -.0653$). Moreover, the analysis showed that no significant effect of the total intelligent property was found. Therefore, the first hypothesis could be accepted, whereas the second one needs to be rejected. The construct of the intelligent property might not be the best to measure the company's innovativeness. Further research could extend the model by other operationalization options, for example, the amount spent for research and development. The analysis showed the investigated control variables similar but also contradicting results compared to the literature. Hence the controls should be investigated further in the context of start-ups.

The research questions from section 3.1 can be answered, as some determinants of employee turnover show similar results in the context of start-ups, as in previous research about employee turnover in established companies. The newly developed determinants (funding volume and intelligent property) can only partially significantly predict employee turnover, as mentioned in section 5.3.

Moreover, the overall explained variance was relatively low in the model ($r^2 = .138$). The model could be extended in the future by already established determinants like job satisfaction or organizational commitment. This could be done by extending the data with job ratings from job platforms like Indeed or Glassdoor. In addition, the analysis was conducted on an estimated average employee turnover rate. Further research could investigate the determinant for start-ups by gathering collaboration with several start-ups to analyze the effects on actual data.

Bibliography

Abbasi, S. M., & Hollman, K. W. (2000). Turnover: The real bottom line. *Public Personnel Management*, *29*(3), 333–342.

Abelson, M. A. (1987). Examination of avoidable and unavoidable turnover. *Journal of Applied Psychology*, *72*(3), 382–386. https://doi.org/10.1037/0021-9010.72.3.382

Ali, W. (2016). Understanding the Concept of Job Satisfaction, Measurements, Theories and its Significance in the Recent Organizational Environment: A Theoretical Framework. *Archives of Business Research*, *4*(1), Article 1. https://doi.org/10.14738/abr.41.1735

Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining Talent: Replacing Misconceptions With Evidence-Based Strategies. *Academy of Management Perspectives*, *24*(2), 48–64. https://doi.org/10.5465/amp.24.2.48

Alterman, V., Bamberger, P. A., Wang, M., Koopmann, J., Belogolovsky, E., & Shi, J. (2021). Best Not to Know: Pay Secrecy, Employee Voluntary Turnover, and the Conditioning Effect of Distributive Justice. *Academy of Management Journal*, *64*(2), 482–508. https://doi.org/10.5465/amj.2019.0231

An, S.-H. (2019). Employee Voluntary and Involuntary Turnover and Organizational Performance: Revisiting the Hypothesis from Classical Public Administration. *International Public Management Journal*, *22*(3), 444–469. https://doi.org/10.1080/10967494.2018.1549629

Andrews, F., & Withey, S. (1976). *Social Indicators of Well-Being: American's Perceptions of Life Quality*. Plenum Press.

Antwi, Y. A., & Bowblis, J. R. (2018). The Impact of Nurse Turnover on Quality of Care and Mortality in Nursing Homes: Evidence from the Great Recession. *American Journal of Health Economics*, *4*(2), 131–163. https://doi.org/10.1162/ajhe_a_00096

Anzola-Román, P., Bayona-Sáez, C., & García-Marco, T. (2018). Organizational innovation, internal R&D and externally sourced innovation practices: Effects on technological innovation outcomes. *Journal of Business Research*, *91*, 233–247. https://doi.org/10.1016/j.jbusres.2018.06.014

Arvey, R. D., Bouchard, T. J. Jr., Segal, N. L., & Abraham, L. M. (1989). Job satisfaction: Environmental and genetic components. *Journal of Applied Psychology*, *74*(2), 187–192. https://doi.org/10.1037/0021-9010.74.2.187

Azeez, R., Foluso, J., & Adeoye, A. (2016). Job Satisfaction, Turnover Intention and Organizational Commitment. *BVIMSR Journal of Management Research, 8*, 102–114.

Baldridge, R., & Curry, B. (2022, October 16). *What Is A Startup? The Ultimate Guide*. Forbes Advisor. https://www.forbes.com/advisor/business/what-is-a-startup/

Banerji, D., & Reimer, T. (2019). Startup founders and their LinkedIn connections: Are well-connected entrepreneurs more successful? *Computers in Human Behavior*, *90*, 46–52. https://doi.org/10.1016/j.chb.2018.08.033

Bartlett, K. R., & McKinney, W. R. (2004). A study of the role of professional development, job attitudes, and turnover among public park and recreation employees. *Journal of Park and Recreation Administration*.

Beechler, S., & Woodward, I. C. (2009). The global "war for talent." *Journal of International Management*, *15*(3), 273–285. https://doi.org/10.1016/j.intman.2009.01.002

Benson, G. S., Finegold, D., & Mohrman, S. A. (2004). You Paid for the Skills, Now Keep Them: Tuition Reimbursement and Voluntary Turnover. *The Academy of Management Journal*, *47*(3), 315–331. https://doi.org/10.2307/20159584

Benson, P. G., Dickinson, T. L., & Neidt, C. O. (1987). The Relationship Between Organizational Size and Turnover: A Longitudinal Investigation. *Human Relations*, *40*(1), 15–29. https://doi.org/10.1177/001872678704000102

Berwick, I., & Raval, A. (2022, November 16). How to win the war for talent. *Financial Times*. https://www.ft.com/content/72c2debe-1fa0-41db-af6c-ef1b2bc7fba1

Bright Data. (2022). *Web data collection for Businesses. Bright Data—Crunchbase*. https://brightdata.comhttps://brightdata.comhttps://brightdata.comhttps://brightdata.com/cp/datasets

Brotherton, P. (2010). Warning signs of turnover waiting to happen: Employees stressed out by major changes made during Great Recession are more likely to leave. *T+D*, *64*(12), 24–25.

Bruneel, J., Yli-Renko, H., & Clarysse, B. (2010). Learning from experience and learning from others: How congenital and interorganizational learning substitute for experiential learning in young firm internationalization. *Strategic Entrepreneurship Journal*, *4*(2), 164–182. https://doi.org/10.1002/sej.89

Brunello, G., & Wruuck, P. (2021). Skill shortages and skill mismatch: A review of the literature. *Journal of Economic Surveys*, *35*(4), 1145–1167. https://doi.org/10.1111/joes.12424

Burgess, S. (1998). Analyzing firms, jobs, and turnover. *Monthly Labor Review: MLR*, *121*(7), 55–57.

Buschow, C., Laugemann, R., & Schneider, B. (2015). Wer gewinnt den ‚War for Talent' in der Medienwirtschaft: Konzerne oder Start-ups? [Who wins the ‚War for Talent' in media economy: Corporations or startup companies?]. *MedienWirtschaft - Zeitschrift Für Medienmanagement Und Medienökonomie*, *12*, 24–31. https://doi.org/10.15358/1613-0669-2015-4-24

Busold, M. (Ed.). (2019). *War for Talents: Erfolgsfaktoren im Kampf um die Besten*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-57481-2

Calopa, M. K., Horvat, J., & Lalic, M. (2014). Analysis of Financing Sources for Start-up Companies. *Management : Journal of Contemporary Management Issues*, *19*(2), 19–43.

CEDEFOP. (2020, October 16). *Job Turnover*. CEDEFOP. https://www.cedefop.europa.eu/en/tools/skills-intelligence/job-turnover

Chin, C. L. (2018). The Influence of Job Satisfaction on Employee Turnover Intention in the Manufacturing Industry of Malaysia. *Social Sciences*, *1*(2).

Clark, R. L., & Ritter, B. M. (2020). How Are Employers Responding to an Aging Workforce? *The Gerontologist*, *60*(8), 1403–1410. https://doi.org/10.1093/geront/gnaa031

Cohen, A. (1993). Organizational Commitment and Turnover: A Meta-Analysis. *The Academy of Management Journal*, *36*(5), 1140–1157. https://doi.org/10.2307/256650

Collini, S. A., Guidroz, A. M., & Perez, L. M. (2015). Turnover in health care: The mediating effects of employee engagement. *Journal of Nursing Management*, *23*(2), 169–178. https://doi.org/10.1111/jonm.12109

Collins, S. K., McKinnies, R. C., Matthews, E. P., & Collins, K. S. (2015). A Ministudy of Employee Turnover in US Hospitals. *The Health Care Manager*, *34*(1), 23. https://doi.org/10.1097/HCM.0000000000000038

Cotton, J. L., & Tuttle, J. M. (1986). Employee Turnover: A Meta-Analysis and Review with Implications for Research. *The Academy of Management Review*, *11*(1), 55–70. https://doi.org/10.2307/258331

Dalle, J.-M., den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research. *OECD Science, Technology and Industry Working Papers*, *2017*(8), 0_1,3,5-29. https://doi.org/10.1787/6c418d60-en

David, G., & Brachet, T. (2011). On the Determinants of Organizational Forgetting. *American Economic Journal: Microeconomics*, *3*(3), 100–123. https://doi.org/10.1257/mic.3.3.100

De Winne, S., Marescaux, E., Sels, L., Van Beveren, I., & Vanormelingen, S. (2019). The impact of employee turnover and turnover volatility on labor productivity: A flexible non-linear approach. *The International Journal of Human Resource Management*, *30*(21), 3049–3079. https://doi.org/10.1080/09585192.2018.1449129

Dess, G. G., & Shaw, J. D. (2001). Voluntary Turnover, Social Capital, and Organizational Performance. *The Academy of Management Review*, *26*(3), 446–456. https://doi.org/10.2307/259187

Dipietro, R. B., & Condly, S. J. (2007). Employee Turnover in the Hospitality Industry. *Journal of Human Resources in Hospitality & Tourism*, *6*(1), 1–22. https://doi.org/10.1300/J171v06n01_01

Eurostat. (2008). *NACE Rev. 2*. European Commission. https://www.bundesbank.de/resource/blob/619164/ccae9c8d5f91f785cb1693e60c77f3a9/mL/nace-rev-2-data.pdf

Glebbeek, A. C., & Bax, E. H. (2004). Is High Employee Turnover Really Harmful? An Empirical Test Using Company Records. *Academy of Management Journal*, *47*(2), 277–286. https://doi.org/10.5465/20159578

Griffeth, R. W., & Hom, P. W. (1995). *Employee Turnover*. South-Western College Pub.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium. *Journal of Management*, *26*(3), 463–488. https://doi.org/10.1177/014920630002600305

Grissom, J. A., Nicholson-Crotty, J., & Keiser, L. (2012). Does My Boss's Gender Matter? Explaining Job Satisfaction and Employee Turnover in the Public Sector. *Journal of Public Administration Research and Theory*, *22*(4), 649–673. https://doi.org/10.1093/jopart/mus004

Haag, P. (2021). Einführung: Management von Start-ups und KMU. In P. Haag (Ed.), *KMU- und Start-up-Management: Strategische Aspekte, operative Umsetzung und Best-Practice* (pp. 1–13). Springer Fachmedien. https://doi.org/10.1007/978-3-658-34700-0_1

Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. (2013). Meta-Analytic Review of Employee Turnover as a Predictor of Firm Performance. *Journal of Management*, *39*(3), 573–603. https://doi.org/10.1177/0149206311424943

Hancock, J. I., Allen, D. G., & Soelberg, C. (2017). Collective turnover: An expanded meta-analytic exploration and comparison. *Human Resource Management Review*, *27*(1), 61–86. https://doi.org/10.1016/j.hrmr.2016.06.003

Hansen, F. (1997). Currents in Compensation and Benefits. *Compensation & Benefits Review*, *29*(5), 6–18. https://doi.org/10.1177/088636879702900502

Hausknecht, J. P. (2017). Collective Turnover. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(1), 527–544. https://doi.org/10.1146/annurev-orgpsych-032516-113139

Hausknecht, J. P., & Trevor, C. O. (2011). Collective Turnover at the Group, Unit, and Organizational Levels: Evidence, Issues, and Implications. *Journal of Management*, *37*(1), 352–388. https://doi.org/10.1177/0149206310383910

Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. (2013). Causes and consequences of collective turnover: A meta-analytic review. *Journal of Applied Psychology*, *98*(3), 412–453. https://doi.org/10.1037/a0032380

Hewlett, S. A. (2006). Leadership in your midst: Tapping the hidden strengths of minority executives. *Development and Learning in Organizations: An International Journal*, *20*(3). https://doi.org/10.1108/dlo.2006.08120caf.007

Hom, P., Roberson, L., & Ellis, A. (2008). Challenging Conventional Wisdom About Who Quits: Revelations From Corporate America. *The Journal of Applied Psychology*, *93*, 1–34. https://doi.org/10.1037/0021-9010.93.1.1

Hom, P. W., Allen, D. G., & Griffeth, R. W. (2019). *Employee Retention and Turnover: Why Employees Stay or Leave*. Routledge. https://doi.org/10.4324/9781315145587

Hom, P. W., Griffeth, R. W., & Sellaro, C. L. (1984). The validity of mobley's (1977) model of employee turnover. *Organizational Behavior and Human Performance*, *34*(2), 141–174. https://doi.org/10.1016/0030-5073(84)90001-1

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, *102*(3), 530–545. https://doi.org/10.1037/apl0000103

Johnson, M. A., & Leo, C. (2020). The inefficacy of LinkedIn? A latent change model and experimental test of using LinkedIn for job search. *Journal of Applied Psychology*, *105*, 1262–1280. https://doi.org/10.1037/apl0000491

Jurgens, J. (2022, May 12). *How startups drive economic recovery while growing responsibly*. World Economic Forum. https://www.weforum.org/agenda/2022/05/how-startups-help-drive-economic-recovery-and-growth/

Kacmar, K. M., Andrews, M. C., Van Rooy, D. L., Steilberg, R. C., & Cerrone, S. (2006). Sure Everyone Can Be Replaced... but at What Cost? Turnover as a Predictor of Unit-Level Performance. *The Academy of Management Journal*, *49*(1), 133–144.

Kalleberg, A. L., & Van Buren, M. E. (1996). Is Bigger Better? Explaining the Relationship Between Organization Size and Job Rewards. *American Sociological Review*, *61*(1), 47–66. https://doi.org/10.2307/2096406

Kaminski, J., Hopp, C., & Tykvová, T. (2019). New technology assessment in entrepreneurial financing – Does crowdfunding predict venture capital investments? *Technological Forecasting and Social Change*, *139*, 287–302. https://doi.org/10.1016/j.techfore.2018.11.015

Kammeyer-Mueller, J. D., Rubenstein, A. L., Long, D. M., Odio, M. A., Buckman, B. R., Zhang, Y., & Halvorsen-Ganepola, M. D. K. (2013). A Meta-Analytic Structural Model of Dispositonal Affectivity and Emotional Labor. *Personnel Psychology*, *66*(1), 47–90. https://doi.org/10.1111/peps.12009

Khera, S. N., & Divya. (2019). Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques. *Vision*, *23*(1), 12–21. https://doi.org/10.1177/0972262918821221

Kulik, C. T., Ryan, S., Harper, S., & George, G. (2014). Aging Populations and Management. *The Academy of Management Journal*, *57*(4), 929–935.

Laulié, L., & Morgeson, F. P. (2021). The end is just the beginning: Turnover events and their impact on those who remain. *Personnel Psychology*, *74*(3), 387–409. https://doi.org/10.1111/peps.12422

Lee, S. (2018). Employee Turnover and Organizational Performance in U.S. Federal Agencies. *The American Review of Public Administration*, *48*(6), 522–534. https://doi.org/10.1177/0275074017715322

Lee, T. H. (2012). Gender Differences in Voluntary Turnover: Still a Paradox? *International Business Research*, *5*(10), Article 10. https://doi.org/10.5539/ibr.v5n10p19

Lee, T. W., & Mitchell, T. R. (1994). An Alternative Approach: The Unfolding Model of Voluntary Employee Turnover. *The Academy of Management Review*, *19*(1), 51–89. https://doi.org/10.2307/258835

Leonard, J. S., & Levine, D. I. (2006). The Effect of Diversity on Turnover: A Large Case Study. *Industrial and Labor Relations Review*, *59*(4), 547–572.

Lewis, G. B., & Park, K. (1989). Turnover rates in federal white-collar employment: Are women more likely to quit than men? *American Review of Public Administration, 19*(1), 13–29.

Lewis, G., & Soroñgon, J. (2022, November 8). *Industries with the Highest (and Lowest) Turnover Rates*. LinkedIn Talent Blog. https://www.linkedin.com/business/talent/blog/talent-strategy/industries-with-the-highest-turnover-rates

Light, A., & Ureta, M. (1992). Panel Estimates of Male and Female Job Turnover Behavior: Can Female Nonquitters be Identified? *Journal of Labor Economics*, *10*(2), 156–181.

Löffler, L., & Giebe, C. (2021). *Generation Z and the War of Talents in the German banking sector*. *4*, 1–18. https://doi.org/10.35409/IJBMER.2021.3319

Lyness, K. S., & Judiesch, M. K. (2001). Are female managers quitters? The relationships of gender, promotions, and family leaves of absence to voluntary turnover. *Journal of Applied Psychology*, *86*(6), 1167–1178. https://doi.org/10.1037/0021-9010.86.6.1167

Malik, A. R., Manroop, L., & Patel, P. C. (2019). An empirical examination of the relationship between skills shortage and firm performance: The role of high-performance work systems. *Journal of Management and Organization*, *25*(5), 695–710. https://doi.org/10.1017/jmo.2017.30

March, J. G., & Simon, H. A. (1958). *Organizations* (Vol. 2). Wiley.

Maurer, C. C., & Qureshi, I. (2021). Not just good for her: A temporal analysis of the dynamic relationship between representation of women and collective employee turnover. *Organization Studies*, *42*(1), 85–107. https://doi.org/10.1177/0170840619875480

McElroy, J. C., Morrow, P. C., & Rude, S. N. (2001). Turnover and organizational performance: A comparative analysis of the effects of voluntary, involuntary, and reduction-in-force turnover. *Journal of Applied Psychology*, *86*(6), 1294–1299. https://doi.org/10.1037/0021-9010.86.6.1294

Meier, K. J., & Hicklin, A. (2008). Employee Turnover and Organizational Performance: Testing a Hypothesis from Classical Public Administration. *Journal of Public Administration Research and Theory*, *18*(4), 573–590. https://doi.org/10.1093/jopart/mum028

Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. *Human Resource Management Review*, *1*(1), 61–89. https://doi.org/10.1016/1053-4822(91)90011-Z

Mobley, W. H. (1977). *Intermediate linkages in the relationship between job satisfaction and employee turnover*. https://oce-ovid-com.uaccess.univie.ac.at/article/00004565-197704000-00020/PDF

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin*, *86*(3), 493–522. https://doi.org/10.1037/0033-2909.86.3.493

Moreno-Delgado, A., Orduña-Malea, E., & Repiso, R. (2020). LinkedIn as a data source to rank universities according to graduate's employability in top companies. *Transinformação*, *32*, e200023. https://doi.org/10.1590/2318-0889202032e200023

Mossholder, K. W., Settoon, R. P., & Henagan, S. C. (2005). A Relational Perspective on Turnover: Examining Structural, Attitudinal, and Behavioral Predictors. *The Academy of Management Journal*, *48*(4), 607–618.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, *14*(2), 224–247. https://doi.org/10.1016/0001-8791(79)90072-1

Nevil, S. (2022, December 21). *How to Calculate Z-Score and Its Meaning*. Investopedia. https://www.investopedia.com/terms/z/zscore.asp

Ng, E. S. W., & Burke, R. J. (2005). Person–organization fit and the war for talent: Does diversity management make a difference? *The International Journal of Human Resource Management*, *16*(7), 1195–1210. https://doi.org/10.1080/09585190500144038

Ng, T. W. H., & Feldman, D. C. (2009). Re-examining the relationship between age and voluntary turnover. *Journal of Vocational Behavior*, *74*(3), 283–294. https://doi.org/10.1016/j.jvb.2009.01.004

Nyberg, A. J., & Ployhart, R. E. (2013). Context-Emergent Turnover (cet) Theory: A Theory of Collective Turnover. *The Academy of Management Review*, *38*(1), 109–131.

OECD. (2020). *Promoting an Age-Inclusive Workforce: Living, Learning and Earning Longer*. OECD. https://doi.org/10.1787/59752153-en

Okae, B. H. (2018). The Effect of Employee Turnover in the Hospitality Industry: Quantitative Correlational Study. *International Journal of Learning and Development, 8*(1), 156–183.

Olubiyi, O., Smiley, G., Luckel, H., & Melaragno, R. (2019). A qualitative case study of employee turnover in retail business. *Heliyon*, *5*(6), e01796. https://doi.org/10.1016/j.heliyon.2019.e01796

O’Reilly, C. A., & Chatman, J. (1986). Organizational commitment and psychological attachment: The effects of compliance, identification, and internalization on prosocial behavior. *Journal of Applied Psychology*, *71*(3), 492–499. https://doi.org/10.1037/0021-9010.71.3.492

Park, T.-Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology*, *98*(2), 268–309. https://doi.org/10.1037/a0030723

Paschen, J. (2017). Choose wisely: Crowdfunding through the stages of the startup life cycle. *Business Horizons*, *60*(2), 179–188. https://doi.org/10.1016/j.bushor.2016.11.003

Patel, N. (2015, January 16). *90% Of Startups Fail: Here's What You Need To Know About The 10%*. Forbes. https://www.forbes.com/sites/neilpatel/2015/01/16/90-of-startups-will-fail-heres-what-you-need-to-know-about-the-10/

Pellegrino, G., Piva, M., & Vivarelli, M. (2012). Young firms and innovation: A microeconometric analysis. *Structural Change and Economic Dynamics*, *23*(4), 329–340. https://doi.org/10.1016/j.strueco.2011.10.003

Peltokorpi, V., Allen, D. G., & Froese, F. (2015). Organizational embeddedness, turnover intentions, and voluntary turnover: The moderating effects of employee demographic characteristics and value orientations. *Journal of Organizational Behavior*, *36*(2), 292–312. https://doi.org/10.1002/job.1981

Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin*, *80*(2), 151–176. https://doi.org/10.1037/h0034829

Price, J. L. (1977). *The Study of Turnover* (First editioin). Iowa State University Press.

Reichers, A. E. (1985). A Review and Reconceptualization of Organizational Commitment. *The Academy of Management Review*, *10*(3), 465–476. https://doi.org/10.2307/258128

Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology*, *71*(1), 23–65. https://doi.org/10.1111/peps.12226

Scanlan, J. N., & Still, M. (2019). Relationships between burnout, turnover intention, job satisfaction, job demands and job resources for mental health personnel in an Australian mental health service. *BMC Health Services Research*, *19*(1), 62. https://doi.org/10.1186/s12913-018-3841-z

Sender, A., Rutishauser, L., & Staffelbach, B. (2018). Embeddedness across contexts: A two-country study on the additive and buffering effects of job embeddedness on employee turnover. *Human Resource Management Journal*, *28*(2), 340–356. https://doi.org/10.1111/1748-8583.12183

Sessa, V. I., & Bowling, N. A. (Eds.). (2020). *Essentials of Job Attitudes and Other Workplace Psychological Constructs*. Routledge. https://doi.org/10.4324/9780429325755

Shaw, J. D., Delery, J. E., Jenkins, G. D., & Gupta, N. (1998). An organization-level analysis of voluntary and involuntary turnover. *Academy of Management Journal*, *41*(5), 511–525.

Shaw, J. D., Duffy, M. K., Johnson, J. L., & Lockhart, D. E. (2005). Turnover, Social Capital Losses, and Performance. *The Academy of Management Journal*, *48*(4), 594–606.

Shaw, J. D., Gupta, N., & Delery, J. E. (2005). Alternative Conceptualizations of the Relationship between Voluntary Turnover and Organizational Performance. *The Academy of Management Journal*, *48*(1), 50–68.

Simón, C., Avgerinos, E., & Revilla, E. (2022). Quits Versus Discharges Across Job Levels: Revisiting the Positive Side of Turnover. *Journal of Management*, 01492063221108921. https://doi.org/10.1177/01492063221108921

Skelton, A. R., Nattress, D., & Dwyer, R. J. (2019). Predicting manufacturing employee turnover intentions. *Journal of Economics, Finance and Administrative Science*, *25*(49), 101–117. https://doi.org/10.1108/JEFAS-07-2018-0069

Slaski, C. (2018, April 9). *Innovation Can Make Or Break Employee Retention*. Forbes. https://www.forbes.com/sites/forbeshumanresourcescouncil/2018/09/04/innovation-can-make-or-break-employee-retention/

Spender, J.-C., Corvello, V., Grimaldi, M., & Rippa, P. (2017). Startups and open innovation: A review of the literature. *European Journal of Innovation Management*, *20*(1), 4–30. https://doi.org/10.1108/EJIM-12-2015-0131

Stamolampros, P., Korfiatis, N., Chalvatzis, K., & Buhalis, D. (2019). Job satisfaction and employee turnover determinants in high contact services: Insights from Employees'Online reviews. *Tourism Management*, *75*, 130–147. https://doi.org/10.1016/j.tourman.2019.04.030

Stroh, L. K., Brett, J. M., & Reilly, A. H. (1996). Family Structure, Glass Ceiling, and Traditional Explanations for the Differential Rate of Turnover of Female and Male Managers. *Journal of Vocational Behavior*, *49*(1), 99–118. https://doi.org/10.1006/jvbe.1996.0036

Taques, F. H., López, M. G., Basso, L. F., & Areal, N. (2021). Indicators used to measure service innovation and manufacturing innovation. *Journal of Innovation & Knowledge*, *6*(1), 11–26. https://doi.org/10.1016/j.jik.2019.12.001

Tarigan, V., & Ariani, D. W. (2015). Empirical Study Relations Job Satisfaction, Organizational Commitment, and Turnover Intention. *Advances in Management and Applied Economics*, *5*(2), 21–42.

Tata, A., Laureiro-Martínez, D., & Brusoni, S. (2016). Don't Look Back? The Effect of Attention to Time and Self on Startup Funding. *Academy of Management Proceedings*, *2016*(1), 13926. https://doi.org/10.5465/ambpp.2016.13926abstract

Taylor, M. A., & Bisson, J. B. (2020). Changes in cognitive function: Practical and theoretical considerations for training the aging workforce. *Human Resource Management Review*, *30*(2), 100684. https://doi.org/10.1016/j.hrmr.2019.02.001

Tews, M. J., Michel, J. W., & Ellingson, J. E. (2013). The Impact of Coworker Support on Employee Turnover in the Hospitality Industry. *Group & Organization Management*, *38*(5), 630–653. https://doi.org/10.1177/1059601113503039

United Nations Department of Economic and Social Affairs, Population Division. (2022). *World Population Prospects 2022: Summary of Results*. 52.

U.S. BUREAU OF LABOR STATISTICS. (2022, October 3). *Table 16. Annual total separations rates by industry and region, not seasonally adjusted—2022 M01 Results*. https://www.bls.gov/news.release/jolts.t16.htm

Vance, T. J. (2015). An Analysis of Employee Turnover in a Manufacturing Plant [D.B.A., Northcentral University]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/docview/1669973499/abstract/C3B94FA4BBC641F7PQ/1

Weinberg, S. L., & Abramowitz, S. K. (2020, February 27). *Statistics Using Stata: An Integrative Approach*. Higher Education from Cambridge University Press; Cambridge University Press. https://doi.org/10.1017/9781108770163

Weisberg, J., & Kirschenbaum, A. (1993). Gender and turnover: A re-examination of the impact of sex on intent and actual job changes. *Human Relations*, *46*(8), 987.

Wynen, J., Van Dooren, W., Mattijs, J., & Deschamps, C. (2019). Linking turnover to organizational performance: The role of process conformance. *Public Management Review*, *21*(5), 669–685. https://doi.org/10.1080/14719037.2018.1503704

Yanadori, Y., & Kato, T. (2007). Average employee tenure, voluntary turnover ratio, and labour productivity: Evidence from Japanese firms. *The International Journal of Human Resource Management*, *18*(10), 1841–1857. https://doi.org/10.1080/09585190701570981

Zaech, S., & Baldegger, U. (2017). Leadership in start-ups. *International Small Business Journal*, *35*(2), 157–177. https://doi.org/10.1177/0266242616676883

Appendix

```python
"""
@author: Tobias Quebe
"""

import pandas as pd
import json
import csv


# Crunchbase dataset (saved in same directory)
dfcrunch = pd.read_csv("bd_crunchbase_dataset.csv", sep = ',', encoding = 'utf-8-sig')

# selecting only the necessary columns from the raw dataset
dfData = dfcrunch[['name','region', 'country_code', 'operating_status', 'founded_date', 'id', 'industries', 'num_employees']]

# adding new columns to the new data frame dfData
dfData['funding_value'] = 0
dfData['no_trademarks'] = 0
dfData['no_patents'] = 0


# open the raw dataset and iterate through each row, extract neede information and update the values in the new dataframe dfData in the regarding columns
with open('bd_crunchbase_dataset.csv', encoding = 'utf-8-sig')  as crunchbasedata:
    crunchbaseReader = csv.DictReader(crunchbasedata)
    for row2 in crunchbaseReader:
        # Funding
        tmpFund = json.loads(row2['financials_highlights'])
        # search for funding amount in USD
        if 'funding_total' in tmpFund:
            fundingvalue = tmpFund["funding_total"]["value_usd"]
            #update value in final dataframe
            dfData.loc[dfData["name"] == row2['name'], 'funding_value'] = fundingvalue
        # trademarks and patents
        tmpTrade = json.loads(row2['ipqwery'])
        # search for trademarks
        if 'ipqwery_num_trademark_registered' in tmpTrade:
            noTrade = (tmpTrade["ipqwery_num_trademark_registered"])
            #update value in final dataframe
            dfData.loc[dfData["name"] == row2['name'], 'no_trademarks'] = noTrade
        # search for patents
        if 'ipqwery_num_patent_granted' in tmpTrade:
            noPatent = (tmpTrade["ipqwery_num_patent_granted"])
            #update value in final dataframe
            dfData.loc[dfData["name"] == row2['name'], 'no_patents'] = noPatent

# search process is completed
print("Finished searching")


# rows which do not contain a funding volume and either no number of patents and no number of trademarks are droppen/deleted
dfData = dfData[ (dfData['funding_value'] != 0) & ((dfData['no_patents'] != 0) | (dfData['no_trademarks'] != 0))]
print("Finished dropping")
```

Figure 7 script for processing of the raw Crunchbase data set

```python
"""
@author: Tobias Quebe
"""

import pandas as pd
import json
import csv

#Crunchbase dataset (saved in same directory)

dfData = pd.read_csv("Aufbereitung_Crunch.csv", sep = ',', encoding = 'utf-8-sig')

# open the preprocessed Crunchbase dataset, iterate through each row and process the data of some columns
with open("Aufbereitung_Crunch.csv", encoding = 'utf-8-sig')  as crunchbasedata:
    crunchbaseReader = csv.DictReader(crunchbasedata)
    for row in crunchbaseReader:
        # get a better format for the number of employees
        if row['num_employees']!= "":
            size = row['num_employees']
            size = size.split("_")
            size.pop(0)
            size[0] = size[0].lstrip('0')
            size[1] = size[1].lstrip('0')
            num_empl = '-'.join(size)
            # udate the values in the data frame
            dfData.loc[dfData["name"] == row['name'], 'size'] = num_empl
        # get all the industrie values
        if  row['industries'] != "":
            tmpindustries = json.loads(row['industries'])
            listOfInudstries = []
            for a in tmpindustries:
                indus = a["value"]
                listOfInudstries.append(indus)
            # convert to a string to add later in the data frame
            stIndus = ", ".join(listOfInudstries)
            # update the values in teh data frame
            dfData.loc[dfData["name"] == row['name'], 'Industries'] = stIndus
        # extract the year from the founded_date column and save in a new column
        if row['founded_date']!= "":
            year = row['founded_date']
            year = year.split("-")
            year = year[0]
            # enter the extracted value in the new column "Funding_Year"
            dfData.loc[dfData["name"] == row['name'], 'Founding_Year'] = year


#dropping not needed columns
dfData.drop(['Unnamed: 0','num_employees','industries'], axis = 1, inplace = True)

#rearrange columns to a new order
dfData = dfData.loc[:, ["name","id", "Founding_Year", "founded_date", "region", "country_code", "operating_status", "size", "Industries", "funding_value", "no_tradem

print("Done with calculations and rearranging")
```

Figure 8 script for further processing of the preprocessed Crunchbase data set

```python
# function to get the start_date and end_date of a person at the respective company. Moreover counting the number of employees who started and ended their carrer a the respective company in a certain year.
def countTurnover(companyName, experience):
    for exp in experience:
        startdate = 0
        enddate = 0
        success = 0
        #check if the start date exists and is not empty:
        if 'start_date' in exp:
            if exp['start_date'] is not None:
                startdate = exp['start_date']
                # check the length of the list start date and extract the start date if it is a digit.
                if len(startdate) >= 2:
                    if exp['start_date'].split(' ')[-1].isdigit():
                        startdate = int(exp['start_date'].split(' ')[-1])
                else:
                    if startdate.isdigit():
                        startdate = int(startdate)
                    else:
                        startdate = 0
        # check if the end_date exists and is not empty
        if'end_date' in exp:
            if exp['end_date'] is not None:
                enddate = exp['end_date'].split(' ')
                enddate = enddate[-1]
                if enddate.isdigit():
                    enddate = int(enddate)
                # enddate is always zero if the LinkedIn profile contains a string not a number, as than somenthing like "present" is entered which is treated like no end date --> no turnover
                else:
                    enddate = 0

        #checks if somebody started at a certain company in the specific year,counts and assigns the number of new joiners in the respective year
        if startdate == 2017:
            initEmplCnt2017 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2017"]
            newCount = initEmplCnt2017 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2017"] = newCount
            success = 1
        if startdate == 2018:
            initEmplCnt2018 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2018"]
            newCount = initEmplCnt2018 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2018"] = newCount
            success = 1
        if startdate == 2019:
            initEmplCnt2019 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2019"]
            newCount = initEmplCnt2019 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2019"] = newCount
            success = 1
        if startdate == 2019:
            initEmplCnt2019 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2019"]
            newCount = initEmplCnt2019 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2019"] = newCount
            success = 1
        if startdate == 2020:
            initEmplCnt2020 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2020"]
            newCount = initEmplCnt2020 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2020"] = newCount
            success = 1
        if startdate == 2021:
            initEmplCnt2021 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2021"]
            newCount = initEmplCnt2021 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2021"] = newCount
            success = 1
        if startdate == 2022:
            initEmplCnt2022 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2022"]
            newCount = initEmplCnt2022 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2022"] = newCount
            success = 1

         #get the number of employees who turned in a certain year and assign them into the respective column in the data frame
        if enddate == 2017:
            initEmplEndCnt2017 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2017"]
            newCount= initEmplEndCnt2017 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2017"] = newCount
        if enddate == 2018:
            initEmplEndCnt2018 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2018"]
            newCount= initEmplEndCnt2018 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2018"] = newCount
        if enddate == 2019:
            initEmplEndCnt2019 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2019"]
            newCount = initEmplEndCnt2019 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2019"] = newCount
        if enddate == 2020:
            initEmplEndCnt2020 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2020"]
            newCount = initEmplEndCnt2020 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2020"] = newCount
        if enddate == 2021:
            initEmplEndCnt2021 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2021"]
            newCount = initEmplEndCnt2021 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2021"] = newCount
        if enddate == 2022:
            initEmplEndCnt2022 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2022"]
            newCount = initEmplEndCnt2022 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2022"] = newCount
        # return the success variable for checking later if the function was successfull
        return success
```

Figure 9 script for extracting the start and end date of a person

```python
# function to add the months and employee count in the final dataframe
def addTimeAndCount(companyName, positions):
    durationInMonths = 0
    # checks all postions the person has/had
    for pos in positions:
        duration = pos["duration_short"]
        complete = 0
        # check if the duration is not empty
        if duration is not None:
            duration = duration.split()
            years = 0
            months = 0
            # check if the length of the list is either 4 elements or 2 elements long
            if len(duration) == 4:
                if duration[0].isdigit() and duration[2].isdigit():
                    years = int(duration[0])
                    months = int(duration[2])
                    complete = 1
            if len(duration) == 2:
                if duration[0].isdigit():
                    #mit dieser Variante werden nur Jahre erfasst welche wir folgt eingetragen werden, alle andere werden als Monate erfasst
                    if any(x in duration[1] for x in ['year', 'years', 'Jahr', 'Jahre','año', 'años', 'ani', 'سنين']):
                        years = int(duration[0])
                        complete = 1
                    else:
                        months = int(duration[0])
                        complete = 1

            # update durationInMonths so that it contains a monthly sum
            durationInMonths += years * 12 + months

    # takes duration already saved and adds the current duration
    initDuration = dfData.loc[dfData["name"] == companyName]["duration"]
    newDuration = initDuration + durationInMonths

    # takes employee count already saved and adds the current employee count
    initCount = dfData.loc[dfData["name"] == companyName]["emplCnt"]
    newCount = initCount + 1

    # update of values
    dfData.loc[dfData["name"] == companyName, "duration"] = newDuration
    dfData.loc[dfData["name"] == companyName, "emplCnt"] = newCount

    # return the complete variable to check if the function was able to find a value and worked correctly for further processing
    return complete
```

Figure 10 Python script to calculate the duration at a company and no. of employees

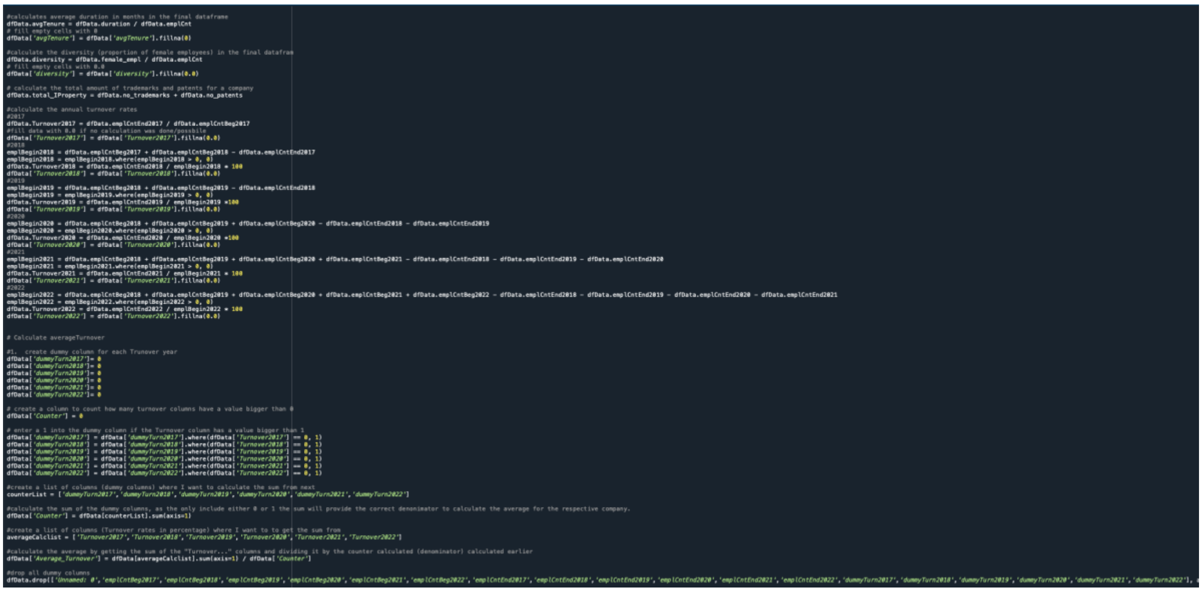
```
# function to guess the gender and update into the final data frame
def getGender(companyName, firstname):
    g = guesser.Detector()
    gender = g.get_gender(firstname)
    #check if male, female or unknown and adding 1 to the respective column
    if gender == "female" or gender == "mostly_female" :
        initfemale = dfData.loc[dfData["name"] == companyName]["female_empl"]
        num_females = initfemale + 1
        dfData.loc[dfData["name"] == companyName, "female_empl"] = num_females
    if gender == "male" or gender == "mostly_male":
        initmale = dfData.loc[dfData["name"] == companyName]["male_empl"]
        num_males = initmale + 1
        dfData.loc[dfData["name"] == companyName, "male_empl"] = num_males
    if gender == "unknown":
         initunkown = dfData.loc[dfData["name"] == companyName]["unkown_gender"]
         num_unknown = initunkown + 1
         dfData.loc[dfData["name"] == companyName, "unkown_gender"] = num_unknown
```

Figure 11 script for the third function to guess the gender of a person

```python
"""
@author: Tobias Quebe
"""

import pandas as pd
import json
import csv
import gender_guesser.detector as guesser

# save preprocessed Crunchbase dataset (saved in same directory) in a new dataframe
dfData = pd.read_csv("Aufbereitung_Crunch_Final.csv", sep = ',', encoding = 'utf-8-sig')

#adding columns with initial value of 0
dfData['duration'] = 0
dfData['emplCnt'] = 0
dfData['avgTenure'] = 0
dfData['total_IProperty'] = 0
dfData['male_empl'] = 0
dfData['female_empl'] = 0
dfData['unkown_gender'] = 0
dfData['diversity'] = 0

#Turnover columns with initial value of 0.0
dfData['Turnover2017'] = 0.0
dfData['Turnover2018'] = 0.0
dfData['Turnover2019'] = 0.0
dfData['Turnover2020'] = 0.0
dfData['Turnover2021'] = 0.0
dfData['Turnover2022'] = 0.0

#adding temporary columns which will be deleted later on:
dfData['emplCntBeg2017']= 0
dfData['emplCntBeg2018']= 0
dfData['emplCntBeg2019']= 0
dfData['emplCntBeg2020']= 0
dfData['emplCntBeg2021']= 0
dfData['emplCntBeg2022']= 0
dfData['emplCntEnd2017']= 0
dfData['emplCntEnd2018']= 0
dfData['emplCntEnd2019']= 0
dfData['emplCntEnd2020']= 0
dfData['emplCntEnd2021']= 0
dfData['emplCntEnd2022']= 0


# function to add the months and employee count in the final dataframe
def addTimeAndCount(companyName, positions):
    durationInMonths = 0
    # checks all postions the person has/had
    for pos in positions:
        duration = pos["duration_short"]
        complete = 0
        # check if the duration is not empty
        if duration is not None:
            duration = duration.split()
            years = 0
            months = 0
            # check if the length of the list is either 4 elements or 2 elements long
            if len(duration) == 4:
                if duration[0].isdigit() and duration[2].isdigit():
                    years = int(duration[0])
                    months = int(duration[2])
                    complete = 1
            if len(duration) == 2:
                if duration[0].isdigit():
                    #mit dieser Variante werden nur Jahre erfasst welche wir folgt eingetragen werden, alle andere werden als Monate erfasst
                    if any(x in duration[1] for x in ['year', 'years', 'Jahr', 'Jahre','año', 'años', 'ani', 'سنين']):
                        years = int(duration[0])
                        complete = 1
                    else:
                        months = int(duration[0])
                        complete = 1

            # update durationInMonths so that it contains a monthly sum
            durationInMonths += years * 12 + months

    # takes duration already saved and adds the current duration
    initDuration = dfData.loc[dfData["name"] == companyName]["duration"]
    newDuration = initDuration + durationInMonths

    # takes employee count already saved and adds the current employee count
    initCount = dfData.loc[dfData["name"] == companyName]["emplCnt"]
    newCount = initCount + 1

    # update of values
    dfData.loc[dfData["name"] == companyName, "duration"] = newDuration
    dfData.loc[dfData["name"] == companyName, "emplCnt"] = newCount

    # return the complete variable to check if the function was able to find a value and worked correctly for further processing
    return complete
```

```python
# function to get the start_date and end_date of a person at the respective company. Moreover counting the number of employees who started and ended their carrer a the respective company in a certain year.
def countTurnover(companyName, experience):
    for exp in experience:
        startdate = 0
        enddate = 0
        success = 0
        #check if the start date exists and is not empty:
        if 'start_date' in exp:
            if exp['start_date'] is not None:
                startdate = exp['start_date']
                # check the length of the list start date and extract the start date if it is a digit.
                if len(startdate) >= 2:
                    if exp['start_date'].split(' ')[-1].isdigit():
                        startdate = int(exp['start_date'].split(' ')[-1])
                else:
                    if startdate.isdigit():
                        startdate = int(startdate)
                    else:
                        startdate = 0
        # check if the end_date exists and is not empty
        if'end_date' in exp:
            if exp['end_date'] is not None:
                enddate = exp['end_date'].split(' ')
                enddate = enddate[-1]
                if enddate.isdigit():
                    enddate = int(enddate)
                # enddate is always zero if the LinkedIn profile contains a string not a number, as than somenthing like "present" is entered which is treated like no end date --> no turnover
                else:
                    enddate = 0

       #checks if somebody started at a certain company in the specific year,counts and assigns the number of new joiners in the respective year
        if startdate == 2017:
            initEmplCnt2017 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2017"]
            newCount = initEmplCnt2017 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2017"] = newCount
            success = 1
        if startdate == 2018:
            initEmplCnt2018 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2018"]
            newCount = initEmplCnt2018 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2018"] = newCount
            success = 1
        if startdate == 2019:
            initEmplCnt2019 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2019"]
            newCount = initEmplCnt2019 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2019"] = newCount
            success = 1
        if startdate == 2019:
            initEmplCnt2019 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2019"]
            newCount = initEmplCnt2019 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2019"] = newCount
            success = 1
        if startdate == 2020:
            initEmplCnt2020 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2020"]
            newCount = initEmplCnt2020 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2020"] = newCount
            success = 1
        if startdate == 2021:
            initEmplCnt2021 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2021"]
            newCount = initEmplCnt2021 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2021"] = newCount
            success = 1
        if startdate == 2022:
            initEmplCnt2022 = dfData.loc[dfData["name"] == companyName]["emplCntBeg2022"]
            newCount = initEmplCnt2022 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntBeg2022"] = newCount
            success = 1

         #get the number of employees who turned in a certain year and assign them into the respective column in the data frame
        if enddate == 2017:
            initEmplEndCnt2017 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2017"]
            newCount= initEmplEndCnt2017 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2017"] = newCount
        if enddate == 2018:
            initEmplEndCnt2018 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2018"]
            newCount= initEmplEndCnt2018 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2018"] = newCount
        if enddate == 2019:
            initEmplEndCnt2019 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2019"]
            newCount = initEmplEndCnt2019 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2019"] = newCount
        if enddate == 2020:
            initEmplEndCnt2020 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2020"]
            newCount = initEmplEndCnt2020 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2020"] = newCount
        if enddate == 2021:
            initEmplEndCnt2021 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2021"]
            newCount = initEmplEndCnt2021 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2021"] = newCount
        if enddate == 2022:
            initEmplEndCnt2022 = dfData.loc[dfData["name"] == companyName]["emplCntEnd2022"]
            newCount = initEmplEndCnt2022 + 1
            dfData.loc[dfData["name"] == companyName, "emplCntEnd2022"] = newCount
        # return the success variable for checking later if the function was successfull
        return success
```

```python
# function to guess the gender and update into the final data frame
def getGender(companyName, firstname):
    g = guesser.Detector()
    gender = g.get_gender(firstname)
    #check if male, female or unknown and adding 1 to the respective column
    if gender == "female" or gender == "mostly_female" :
        initfemale = dfData.loc[dfData["name"] == companyName]["female_empl"]
        num_females = initfemale + 1
        dfData.loc[dfData["name"] == companyName, "female_empl"] = num_females
    if gender == "male" or gender == "mostly_male":
        initmale = dfData.loc[dfData["name"] == companyName]["male_empl"]
        num_males = initmale + 1
        dfData.loc[dfData["name"] == companyName, "male_empl"] = num_males
    if gender == "unknown":
         initunkown = dfData.loc[dfData["name"] == companyName]["unkown_gender"]
         num_unknown = initunkown + 1
         dfData.loc[dfData["name"] == companyName, "unkown_gender"] = num_unknown


#loads the linkedIn data CSV file and iterates through ebery row
with open('bd_LinkedIn_Data.csv', encoding = 'utf-8-sig') as linkedInData:
    linkedInDataReader = csv.DictReader(linkedInData)
    #variable to chekc which column is currently iterated through
    a = 0
    for row in linkedInDataReader:
        # extract the experience and get the name of the persons profile
        tmpExp = json.loads(row['experience'])
        tmpGen = row['name'].split(' ')[0]
        #goes through every work exp of the person
        for exp in tmpExp:
            #checks if the value in exp['company'] is not empty
            if exp['company'] is not None:
                # any of the characters in brackets is in the name the company will be skiped as otherwise compiler error appears
                if any(x in exp['company'] for x in ['*', '+', '(', ')','\\', '?', '!', '[',']','{','}']):
                   continue
                # if none of the above mentioned characters is included this will be executed
                else:
                    # checks if the company is also in the dataframe
                    if dfData["name"].str.contains(exp['company']).sum() > 0:
                        success = countTurnover(exp['company'], exp['positions'])
                        # if the countTurnover funciton was succesfull (success = 1) proceed otherwise skip
                        if success == 1:
                            completed = addTimeAndCount(exp['company'], exp['positions'])
                            # if the addTimeAndCount funciton was succesfull (completed = 1) proceed otherwise skip
                            if completed == 1:
                                getGender(exp['company'],tmpGen)



        # adding one as the next iteration is done
        a = a+1
        print("iteration: ", a)
```



```
#calculates average duration in months in the final dataframe
dfData.avgTenure = dfData.duration / dfData.emplCnt
# fill empty cells with 0
dfData['avgTenure'] = dfData['avgTenure'].fillna(0)

#calculate the diversity (proportion of female employees) in the final datafram
dfData.diversity = dfData.female_empl / dfData.emplCnt
# fill empty cells with 0.0
dfData['diversity'] = dfData['diversity'].fillna(0.0)

# calculate the total amount of trademarks and patents for a company
dfData.total_IProperty = dfData.no_trademarks + dfData.no_patents

#calculate the annual turnover rates
#2017
dfData.Turnover2017 = dfData.emplCntEnd2017 / dfData.emplCntBeg2017
#fill data with 0.0 if no calculation was done/possbile
dfData['Turnover2017'] = dfData['Turnover2017'].fillna(0.0)
#2018
emplBegin2018 = dfData.emplCntBeg2017 + dfData.emplCntBeg2018 - dfData.emplCntEnd2017
emplBegin2018 = emplBegin2018.where(emplBegin2018 > 0, 0)
dfData.Turnover2018 = dfData.emplCntEnd2018 / emplBegin2018 * 100
dfData['Turnover2018'] = dfData['Turnover2018'].fillna(0.0)
#2019
emplBegin2019 = dfData.emplCntBeg2018 + dfData.emplCntBeg2019 - dfData.emplCntEnd2018
emplBegin2019 = emplBegin2019.where(emplBegin2019 > 0, 0)
dfData.Turnover2019 = dfData.emplCntEnd2019 / emplBegin2019 *100
dfData['Turnover2019'] = dfData['Turnover2019'].fillna(0.0)
#2020
emplBegin2020 = dfData.emplCntBeg2018 + dfData.emplCntBeg2019 + dfData.emplCntBeg2020 - dfData.emplCntEnd2018 - dfData.emplCntEnd2019
emplBegin2020 = emplBegin2020.where(emplBegin2020 > 0, 0)
dfData.Turnover2020 = dfData.emplCntEnd2020 / emplBegin2020 *100
dfData['Turnover2020'] = dfData['Turnover2020'].fillna(0.0)
#2021
emplBegin2021 = dfData.emplCntBeg2018 + dfData.emplCntBeg2019 + dfData.emplCntBeg2020 + dfData.emplCntBeg2021 - dfData.emplCntEnd2018 - dfData.emplCntEnd2019 - dfData.emplCntEnd2020
emplBegin2021 = emplBegin2021.where(emplBegin2021 > 0, 0)
dfData.Turnover2021 = dfData.emplCntEnd2021 / emplBegin2021 * 100
dfData['Turnover2021'] = dfData['Turnover2021'].fillna(0.0)
#2022
emplBegin2022 = dfData.emplCntBeg2018 + dfData.emplCntBeg2019 + dfData.emplCntBeg2020 + dfData.emplCntBeg2021 + dfData.emplCntBeg2022 - dfData.emplCntEnd2018 - dfData.emplCntEnd2019 - dfData.emplCntEnd2020 - dfData.emplCntEnd2021
emplBegin2022 = emplBegin2022.where(emplBegin2022 > 0, 0)
dfData.Turnover2022 = dfData.emplCntEnd2022 / emplBegin2022 * 100
dfData['Turnover2022'] = dfData['Turnover2022'].fillna(0.0)


# Calculate averageTurnover

#1.  create dummy column for each Trunover year
dfData['dummyTurn2017']= 0
dfData['dummyTurn2018']= 0
dfData['dummyTurn2019']= 0
dfData['dummyTurn2020']= 0
dfData['dummyTurn2021']= 0
dfData['dummyTurn2022']= 0

# create a column to count how many turnover columns have a value bigger than 0
dfData['Counter'] = 0

# enter a 1 into the dummy column if the Turnover column has a value bigger than 1
dfData['dummyTurn2017'] = dfData['dummyTurn2017'].where(dfData['Turnover2017'] == 0, 1)
dfData['dummyTurn2018'] = dfData['dummyTurn2018'].where(dfData['Turnover2018'] == 0, 1)
dfData['dummyTurn2019'] = dfData['dummyTurn2019'].where(dfData['Turnover2019'] == 0, 1)
dfData['dummyTurn2020'] = dfData['dummyTurn2020'].where(dfData['Turnover2020'] == 0, 1)
dfData['dummyTurn2021'] = dfData['dummyTurn2021'].where(dfData['Turnover2021'] == 0, 1)
dfData['dummyTurn2022'] = dfData['dummyTurn2022'].where(dfData['Turnover2022'] == 0, 1)

#create a list of columns (dummy columns) where I want to calculate the sum from next
counterList = ['dummyTurn2017','dummyTurn2018','dummyTurn2019','dummyTurn2020','dummyTurn2021','dummyTurn2022']

#calculate the sum of the dummy columns, as the only include either 0 or 1 the sum will provide the correct denominator to calculate the average for the respective company.
dfData['Counter'] = dfData[counterList].sum(axis=1)

#create a list of columns (Turnover rates in percentage) where I want to to get the sum from
averageCalclist = ['Turnover2017','Turnover2018','Turnover2019','Turnover2020','Turnover2021','Turnover2022']

#calculate the average by getting the sum of the "Turnover..." columns and dividing it by the counter calculated (denominator) calculated earlier
dfData['Average_Turnover'] = dfData[averageCalclist].sum(axis=1) / dfData['Counter']

#drop all dummy columns
dfData.drop(['Unnamed: 0','emplCntBeg2017','emplCntBeg2018','emplCntBeg2019','emplCntBeg2020','emplCntBeg2021','emplCntBeg2022','emplCntEnd2017','emplCntEnd2018','emplCntEnd2019','emplCntEnd2020','emplCntEnd2021','emplCntEnd2022','dummyTurn2017','dummyTurn2018','dummyTurn2019','dummyTurn2020','dummyTurn2021','dummyTurn2022'],
```

Figure 12 complete third Python script for data processing

Table 6 Explanaition of value of categorical variables

Variable	Value	Meaning
Founding year	base category	2017
	2018	2018
	2019	2019
	2020	2020
	2021	2021
	2022	2022
Country Code	base category	Great Britan
	1	Germany
	2	France
	3	Spain
	4	Sweden
	5	Switzerland
	6	Netherlands
	7	Italy
	8	Finland
	9	Belgium
	10	Poland
	11	Denmark
	12	Others
Size	base category	1-10 people

	1	11-50 people
	2	51-100 people
	3	101-250 people
	4	more than 250 people
Industry	base category	Agriculture, Forestry and Fishing
	1	Manufacturing
	2	Electricity, Gas, Steam and Air Conditioning Supply
	3	Water Supply; Sewerage, Waste Management and Remediation Activities
	4	Construction
	5	Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
	6	Transportation and Storage
	7	Accommodation and Food Service Activities
	8	Information and Communication
	9	Financial and Insurance Activities
	10	Real Estate Activities
	11	Professional, Scientific and Technical Activities
	12	Administrative and Support Service Activities
	13	Education
	14	Human Health and Social Work Activities
	15	Arts, Entertainment and Recreation
	16	Other Service Activities

Table 7 z-values average turnover

Standardized values of log_average_turnover	**Freq.**	**Percent**	**Cum.**
-3.975893	7	77.78	77.78
-3.10169	1	11.11	88.89
3.412178	1	11.11	100.00
Total	9	100.00	

Table 8 z-values funding volume

Standardized values of log_Funding_volume	Freq.	Percent	Cum.
-4.632481	1	9.09	9.09
-3.518274	1	9.09	18.18
-3.25356	2	18.18	36.36
-3.250819	1	9.09	45.45
-3.116936	1	9.09	54.55
-3.022726	2	18.18	72.73
3.193925	1	9.09	81.82
3.259228	1	9.09	90.91
3.300036	1	9.09	100.00
Total	11	100.00	

Table 9 z-values total intelligent property

Standardized values of log_Total_IP	Freq.	Percent	Cum.
3.054099	4	40.00	40.00
3.257308	3	30.00	70.00
3.432392	1	10.00	80.00
3.511627	1	10.00	90.00
3.817262	1	10.00	100.00
Total	10	100.00	