A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

DOES THE BEHAVIOUR ON PROFESSIONAL NETWORKING SITES DIFFER BETWEEN FEMALE AND MALE STARTUP FOUNDERS?

LUISA MARIA KRAUT

Work project carried out under the supervision of:

Francisco Pizarro Beleza Rodrigues Queiró

20/05/2022

Abstract

Although the gender gap in entrepreneurship has been discussed extensively in research, factors originating from the internet have barely been considered to date. As professional networking sites become increasingly popular in the emerging remote working society, a better understanding of founders behaviour on these sites is crucial. This work therefore provides insights into possible differences between female and male founders profiles and actions on LinkedIn. A sample of 1284 European early-stage founders was tested to validate the proposed hypothesis. The results confirm differences in the activity and willingness to self-report information, differences in founders' networking behaviour, however, proved insignificant.

Keywords: Entrepreneurship, Venture Capital, Social Networks, Gender Gap, Gender Stereotypes

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1 Introduction

Shifts towards a more gender equal future are gradually transforming the workforce with female entrepreneurs becoming an integral part in many industries (Elam et al. 2019). Women led companies tend to outperform comparable male entrepreneurs and money invested in women was found to provide higher returns (Abouzahr et al. 2018). Furthermore, women are said to help avoid inefficient decision making and groupthink (Ewens and Townsend 2020). The importance of female founders has also been understood by investors, with newly established programs specifically catered at, and venture fonds exclusively investing in women being on the rise (Hodgson 2021; Lewin 2021; APX 2020). Although this shift is gaining momentum, the gender gap in entrepreneurship is still prevalent and men are undoubtedly dominating the field (Flèche, Lepinteur, and Powdthavee 2021; Guzman and Kacperczyk 2019; Elam et al. 2019; OECD and European Commission 2019). This topic has been extensively discussed in research for decades, but still only little advancements have been made (Hodgson 2021). Besides the discussed views in literature on the causes of these inequalities in the offline world, the emerging digital opportunities open a further field of research. Especially in light of the recent COVID-19 pandemic the work life has rapidly shifted to working remotely, changing the way people collaborate. Interactions online differ significantly from behaviour offline due to the invisibility of facial emotions and expressions (Bai et al. 2019). However, the online world offers a more scalable approach to connecting with people and building a network far beyond the size that networking offline would allow. This is an important factor for founders whose network is crucial for their ventures success (Stam, Arzlanian, and Elfring 2014). But not only the number of connections matter, also the interaction and maintenance of these (Smith and Smith 2021). Professional networking sites facilitate this, while also offering the possibility to display an all-time accessible, up-to-date CV, as well as a personal newsfeed featured by others in the network (Smith and Smith

2021). Hence, this is a source of rich information on founders actions and behaviour online and is worth analysing, as any differences found might be a further factor impacting gender imbalances in entrepreneurship.

After introducing the gender gap in entrepreneurship and funding, scientific insights into the origins of these are given. This is followed by the hypothesis development and data analysis. Using statistical tests and applying cluster analysis the results are interpreted and implications are given in the discussion section. Finally, the conclusion marks the last section of this work. The table of contents is displayed in ***Appendix 1***.

2 Literature Review

Entrepreneurship refers to an individual's or team of individuals attempt of creating a new business or venture, "such as self-employment, a new business organization, or the expansion of an existing business" (Reynolds, Hay, and Camp 1999, 3). Individuals creating ventures are often referred to as startup founders or entrepreneurs. While both terms describe a person creating a new business, they are used to distinguish the business itself. Startup founders intend their venture to be disruptive and scalable which must not be the case for entrepreneurs (Riani 2021). In the following, the gender gap in entrepreneurship as well as in startup funding is examined and theories for the existence of these will be described in more detail.

2.1 Gender gap in Entrepreneurship

The difference in the share of men and women that engage in entrepreneurial activity is referred to as gender gap in entrepreneurship (Vossenberg 2013). Women are substantially less likely to be self-employed than men (OECD and European Commission 2019) and men were found to be significantly more likely to succeed as entrepreneurs (Flèche, Lepinteur, and Powdthavee 2021; Guzman and Kacperczyk 2019). In high-income countries the total entrepreneurial activity (TEA) rate of women is 8.1% which corresponds to approximately

66% of men's TEA rate (Elam et al. 2019, 8). The disparity of male and female entrepreneurs is at odds with the fact that the contribution of female entrepreneurs to economic growth, development and poverty alleviation is significant (Meunier, Krylova, and Ramalho 2017). Looking specifically at startups, the gender gap is also very apparent. Around 80% of all startups globally in 2019 were led by men (Crunchbase 2019, 6). Especially for high-growth ventures, there are large disparities between the share of female founders (FF) and male founders (MF) (Guzman and Kacperczyk 2019). Nonetheless, an increase over the last decade can be observed. The global number of startups with at least one FF, that raised a first funding round, increased from 10% of startups being female co-led in 2009 to every fifth startup in 2019 (Crunchbase 2019, 5). Furthermore, In 2019, 21 unicorns[1] had at least one FF, compared to only four unicorns in 2013 (Crunchbase 2019, 17). However, this number dropped to 10 out of 120 unicorns in 2020 (Teare 2020). The share of unicorns with at least one FF, out of all existing unicorns is 12%, from a total of 1,107 unicorns, as of December 2021 (Crunchbase 2021a). In the European union, this share is only half, with only 12 out of a total of 200 unicorns that have a FF (European Startups 2021). To understand the origin of these imbalances, it is necessary to examine different factors such as gender stereotypes, human capital and social capital.

2.1.1 Gender stereotypes

Literature suggests that entrepreneurship is highly gendered and that those gender stereotypes play a role in women not pursuing an entrepreneurial career path (Ladge, Eddleston, and Sugiyama 2019; Guzman and Kacperczyk 2019; Thomas et al. 2021). While masculine traits like aggressiveness, ambition or risk-taking are commonly related with entrepreneurship, cultural beliefs about female characteristics, such as being caring and warm, having children

[1] Startup whose first equity round was raised at a post money valuation of more than one billion dollars.

or being attractive, do not match this view (Thébaud 2010; Ladge, Eddleston, and Sugiyama 2019). As a result, the perceived incongruity of entrepreneurship and women puts a disadvantage on women that have personal attributes confirming the female stereotypes (Tinkler et al. 2015). Although egalitarian gender attitudes are commonly spread and household as well as childcare duties are more of men's responsibilities nowadays than in the past, the essential beliefs that family caregiving and household chores are mostly women's responsibility, continue to be a dominant cultural ideology. Thus, a work-family conflict arises, especially for mothers who are often more intensively involved in their children's activities (Kacperczyk and Younkin 2021; Cha and Weeden 2014; Guzman and Kacperczyk 2019). In order to balance this conflict, risk-bearing entrepreneurial paths are often not an option for women. The fact that men are almost twice as likely as women to start a new business supports this (OECD and European Commission 2019). Furthermore, the recent COVID-19 pandemic shows that this gendered view of the distribution of responsibilities is still prevalent. The increased demand for caretaking is mostly covered by women, which have to re-evaluate their work-family situation, leading to more and more women considering to stop working at all (Thomas et al. 2021).

2.1.2 Human capital

Goldin (2016) defines human capital as the skillset that enhances a working individual's productive capacity, it is regarded as an asset resource and develops through education, training and work experiences. Entrepreneurial activity is positively correlated with the human capital for both men and women. Generally, when comparing post-secondary or higher degrees by gender and national income level, it can be seen that women in low-income countries are less likely to start a businesses than men with the same education level, while the likeliness is similar in high-income countries (Elam et al. 2019). Hence, the lack of female founders in high growth ventures, that are typically in technical and scientific fields, must be explained by low representation of females in those field. This can be traced back to negative gender stereotypes and few girls in science, technology, engineering and mathematics (STEM) subjects in schools and colleges and both reasons reinforcing each other (Guzman and Kacperczyk 2019; Murciano-Goroff 2021). In his research, Murciano-Goroff (2021) was able to identify that even when making it into male-dominated fields, women were found to be less confident when it comes to self-reporting their technological skills and experiences compared to men with similar skills and job experiences. He further stated that self-promoting females have on average more experience in the reported skill than comparable males. This finding can be explained through the work of Ladge and colleagues (2019) that covers the topic of confidence regarding success and skills in the context of women in entrepreneurship. They show that women face difficulties identifying themselves and their role in the masculine domain of entrepreneurship that lacks other female role models. Resulting in women being more likely to suffer from imposter fears than men, meaning they tend to credit their success to luck or other external rationales, rather than their own competences. These findings are in line with women being reported to have a lower confidence level than men when it comes to their capability in starting a business (Elam et al. 2019). The imposter phenomenon among women is enhanced by people discounting the female entrepreneurs experience or intent as being less meaningful than that of similar male entrepreneurs (Elam et al. 2019; Kacperczyk and Younkin 2021). Microaggressions like interruptions or comments, not concerning women's professional behaviour, further support this (Thomas et al. 2021).

2.1.3 Social capital

Besides the entrepreneurs human capital, a ventures success heavily depends on their social capital (Stam, Arzlanian, and Elfring 2014; Mckeever, Anderson, and Jack 2014; Bernstein, Korteweg, and Laws 2017). Social capital is defined as an individual's ability to make use of its network of relationships to facilitate the achievement of desired outcomes like realising opportunities, acquiring resources or gaining legitimacy. It is understood to comprise both the network of relationships and the actual and potential resources accessible through that network and is a key result of networking (Nahapiet and Ghoshal 1998; Adler and Kwon 2002). Scholars also suggests that the structure and composition of social networks differs between male and female entrepreneurs. Females are said to have strong homogenous networks with close ties (Uzzi 2019). As a result of a less diverse network and missing endorsement benefits, women face more challenges or have limited access to and or exclusion from valuable resources and opportunities, leading their social capital to be less useful for developing and validating new business ideas (Upton, Broming, and Upton 2018; Guzman and Kacperczyk 2019; Tinkler et al. 2015). Men however, show to be better able to make use of their experience, position and existing contacts by forming a centralized, more diffuse network (Guzman and Kacperczyk 2019; Uzzi 2019; Tinkler et al. 2015).

2.2 Gender Gap in Funding

Another approach to explain the gender disparity in entrepreneurship is the unequal access to funding. Women, especially in early-stage ventures, are less likely to receive external funding then men. Male-only founded companies received 91% of total Venture Capital funding in 2020. With a deal volume of $5 billion (bn), female-only founded companies received just 2.3% in the same year, a decrease from 3% in 2019 (Bittner and Lau 2021). The numbers for European FF are even lower, out of the total capital raised between 2017 and 2021, companies with at least one FF received only 1.3% of a total funding volume of €187.1 bn (Hodgson 2021). Investments require decisions from two sides, investors offering it and entrepreneurs accepting it. Opinions in research about the reasons for the funding gap in entrepreneurship are therefore divided in demand side and supply side (Guzman and Kacperczyk 2019; Kanze et al. 2017).

2.2.1 Demand-side

The demand-side or entrepreneur-driven rationale argues that the variance in funding outcomes is due to FF seeking less and therefore receiving less funding than MF. The lower demand from FF can be explained by several reasons, most prominently, that women are active in less capital-intensive businesses (Coleman and Robb 2009; Flèche, Lepinteur, and Powdthavee 2021; Henrekson and Du Rietz 1999). This is often explained by gender stereotypes of familial responsibility that play a role in women not pursuing businesses with high-growth attributes, that involve financial risks and long working hours. Entrepreneurship in the form of lifestyle-businesses and home-based ventures, however, can enable women to have more control about their own schedule to make work compatible with family demands. A report by the OECD and European Commission (2019) supports this with their findings that not only do women typically engage in smaller businesses than men, i.e. having less employees, they also operate in different industries, especially in the areas of personal and household services. This complies with findings that female founded companies are less profitable and show less likeliness to have growth ambitions (Cha and Weeden 2014; Guzman and Kacperczyk 2019; Kanze et al. 2017; Flèche, Lepinteur, and Powdthavee 2021; Ladge, Eddleston, and Sugiyama 2019). Besides that, there is a difference in the initial motivation for starting a business between women and men. While MF motivation is related to material success, FF are said to value career satisfaction from socioemotional sources over financial success (Manolova, Brush, and Edelman 2008).

2.2.2 Supply-side

Some scholars are of the opinion that the disparity in venture capital funding between FF and MF cannot be entirely explained by the above mentioned entrepreneur-driven factors (Tinkler et al. 2015; Gompers and Wang 2017). Since almost all early stage investors are male, e.g.

only 8% of US Venture Capital fonds (VCF) (Gompers and Wang 2017, 10) and only 10% of VCF decision makers in Europe are female (Hodgson 2021), the funding gap must also be a result of male investors decision making. Recent research found that the response to FF differed between male and female investors. For comparable companies, male investors demonstrate a preference for MF, while FF were more successful with female investors (West and Sundaramurthy 2020; Ewens and Townsend 2020; Kanze et al. 2017). This suggests gender bias and homophilic preferences by either female or male investors (McPherson, Smith-Lovin, and Cook 2001).

The gender stereotypes explained in section 2.1.1 are also found in an investor context. This is demonstrated by the fact that women are facing more critical questions and pushback during presentations to investors compared to MF. Women are asked to demonstrate basic technological knowledge to show that they understand the underlying topic, which is not the case with MF (Abouzahr et al. 2018; Ewens and Townsend 2020). In addition to that, Kanze and colleagues (2017) point out that MF were mostly being asked promotion-focused questions including for example the startups growth potential, while FF were asked prevention-focused questions, concerning the potential for losses in order to prove their ability to generate safe returns. The study also found that in general, entrepreneurs that were asked mostly prevention question raised on average about seven times less than the entrepreneurs that received promotion questions.

Another study by Malmstrom and colleagues (2017), analysed Swedish venture capitalists (VCs) language when discussing final funding decisions. They were able to show that the investors “rhetorically produce stereotypical images of women as having qualities opposite to those considered important to being an entrepreneur, with VCs questioning their credibility, trustworthiness, experience, and knowledge” while MF entrepreneurial potential was reinforced (Malmstrom, Johansson, and Wincent 2017, para. 7). For example, the entrepreneurs age was perceived differently depending on their gender. While young FF were considered as inexperienced, MF were viewed as promising. Unsurprisingly, they found that on average, FF were only awarded one fourth of the applied-for-amount, while MF received more than half of what they had asked for. According to Ewens and Townsend (2020) signs of traction, like customer numbers, sales or in taking part in an incubator program, help MF garner interest from male investors. FF, however, benefit significantly less from this, proving a further sign of discounting credentials and competence of FF. The aforementioned recent studies support earlier research such as Foschi (2000, 28), which focused on double standards for competence, i.e. "the application of different criteria for competence to actors of different status". She proved that in the case of gender, when men and women perform at the same level, women face stricter standards than men. Women have to put in more effort and are allowed fewer mistakes, to be attributed the same level of competences as men. Unlike for men, women's failure confirms expectations, while success is due to luck.

In addition to gender stereotypes in entrepreneurship, research suggests that homophilic tendencies can explain the lower amount of FF being funded. "Tech and VC are industries built on two things: relationships and insider knowledge" (Teare 2020, para. 29). This demonstrates, that a lack of either indirect or direct connections to a VCF, places significant hurdles for fundraising (Banerji and Reimer 2019; Tinkler et al. 2015). FF experience difficulties breaking into the 'boys club', especially when decision makers are facing uncertainty, the tendency for similarity-attraction increases (Bittner and Lau 2021; Guzman and Kacperczyk 2019). Hence, gendered expectations and homophily are significantly disadvantageous for female founders when trying to access funding.

3 Methodology

In this section the research approach is briefly described, including data collection, sample size and data preparation as well as the derivation and application of the measurement items.

3.1 Research approach

To achieve the research goal, the deductive approach, deducing hypotheses from literature and testing these with the collected data, was applied (Saunders, Lewis, and Tornhill 2009). A lack of scientific findings of the behaviour on professional networking sites (PNS) in the context of entrepreneurship, in combination with gender inequalities was identified as the research gap. The few to date existing references on founders' activities online focus on venture funding outcomes (Jin, Hitt, and Wu 2017; Banerji and Reimer 2019), the development of social connections (Smith, Smith, and Shaw 2017; Smith and Smith 2021) or the founders personalities (Obschonka, Fisch, and Boyd 2017). On a company level, the effect of social media presence has been studied with regards to uncertainty reduction regarding firm quality and differentiation from competitors (Fischer and Reuber 2014).

However, to date there is no research that analysed the founders actions on PNSs and whether this is another factor that might contribute to the gender gap in entrepreneurship and funding. Therefore, the goal of this work is to examine, whether the behaviour of FF and MF differs online and if so, whether it is congruent with findings in research regarding gender differences offline. The literature review enabled the deduction of three hypothesis that were tested to answer the research question.

Based on the findings on human capital in section 2.1.2, that women are more reluctant to self-promotion than men, the first hypothesis is as follows:

H1: The amount of information displayed on PNS profiles differs between FF and MF.

Section 2.1.3, on the importance and differences in social capital allowed the derivation of the second hypothesis:

H2: There is a difference in networking on PNS between FF and MF.

Finally, combining the insights from section 2.2.2 regarding investor bias and translating Fischer and Reuber's (2014) findings, concerning the ability to signal quality through active and frequent social media use, to an entrepreneurial individual, leads to the third hypothesis:

H3: Women are more active on PNS compared to men, attempting to prove their credibility and intent.

For the scope of this work, 'online behaviour' is therefore measured along the three developed hypothesis, and referred to as the founders activity, self-reporting and networking.

3.2 Data collection

In order to test the hypothesis, Crunchbase (CB) and LinkedIn (LI) data are used to identify European early stage FF and MF to analyse and compare their actions online. CB is an online database that provides crowdsourced data on private companies, founders, VCFs and other investors. It has 70 million annual visitors and claims to have the "best-in-class live data powered by [their] unique community of contributors, partners, and in-house data experts" (Crunchbase 2022). CB developed its own metric, called Crunchbase rank (CBR) that dynamically ranks all entities (People, companies, investors) in the dataset and measures the 'prominence' of the entity. It takes many different factors into account and is fluid, i.e. falling or rising over time as significant changes like fundraising or news affect it. Since CB does not provide in-depth information about the identified startup founders, LI is chosen to systematically record their profile and activity information. LI is a PNS that was launched in 2003 and currently consists of 830 million members worldwide (LinkedIn 2022). Individual's user profiles are structured like an online resume, providing information about users professional experience and educational background. Furthermore, users can provide information about their skills and accomplishments. Users can demonstrate their interests by following influencers, companies, schools and organisations. Moreover, users have the possibility to give or receive recommendations about or from other users and endorse their skills. Information about a user's network size can also be inferred as LI provides the number of followers (subscribers) a user has. The number of followers does not equal the number of connections, however, as two users can be connected but not follow each other and vice versa. The exact number of connections, if exceeding 500 is not displayed. Besides the users profile, LI also provides the users activities. These include their own posts as well as the user's reaction (Like [Like, Celebrate, Support, Love, Insightful, Curious], Comment, Share) to other users posts, comments or updates. LI therefore is not only a resource for objective network data but also provides rich information about a user's experience and its behaviour on the site.

For this work, data was extracted from CB in October 2021. Only continental European founders were selected to ensure comparability in terms of markets and industry groups. Only early stage founders (Angel – Series A funding) were selected based on the assumption that founders' companies at a funding stage beyond Series B investments have already proven to have a useful social network, close ties to investors and in case of FF already overcome gender bias. Moreover, only companies founded in or after 2010 were selected from CB. For FF, this query gave 3,111 results, with CBRs from 851 to 1,257,979. The same query for MF gave around nine times more results, 28,197 MF, with CBRs from 24 to 1,262,542. To get early stage company data, an additional query on European companies was run and returned 24,809 results. Finally, the founder dataset and company dataset were merged, using Python 3.8. The merged dataset consisted of 8170 matching records, of which 89% were MF and 11% FF.

In order to find the corresponding founders on LI, the CB data had to be transformed. First, companies needed to be comparable for the industries they were operating in. As the labels that were assigned to the 'Industry' and 'Industry Groups' columns by CB for each company resulted in 4847 different combinations for 'Industry' and 2911 for 'Industry Groups' (Crunchbase 2021b), new industry labels were created to reduce the variety and can be seen in ***Appendix 2***. In order to do so, two different Neutral Language Processing (NLP) algorithms, NLTK (Bird, Loper, and Klein 2006) and Yake (Ricardo Campos et al. 2020) were applied on CB's 'Full Description' column, which states the companies description in full sentences. Results of each algorithm were ranked by performance and compared to the original industry group tags to ensure an accurate assignment. Next, founders whose LI profiles and activities could be retrieved were identified. Hence, founders whose LI URL was not provided on CB were excluded from the dataset, resulting it to decrease in size to 6696 records. The gender split remained unchanged, however. Since the dataset consisted of nine times more MF than FF, the number of MF was reduced by only selecting comparable MF to the FF in the sample. Comparability was based on the founders CBR (ranges of 50,000 ranks), the industry label and the last funding type. This resulted in a dataset consisting of 642 URLs for each gender. The identified URLs were used to extract information from the founders LI profile page and the 200 most recent activities. Finally, the profile and activity data from LI was merged onto the CB data to obtain the final dataset.

3.3 Data preparation

For numeric data, the method of winsorization was applied to handle outliers. In comparison to more radical methods like keeping or completely removing the outliers and their corresponding records, winsorization recodes the outliers which allows to keep as much data as possible, while not deferring measures by taking extreme points into account (Leys et al. 2019). Outliers, in this work regarded as data points beyond three standard deviations from the mean (outer fence), were recoded to the value of the k^{th} and $1\text{-}k^{th}$ percentile for the observation. The value for k was chosen based on the closest proximity outside the outer fence. In this dataset, missing values demonstrate the founders will to not provide certain information and hence do not need imputation, except for two cases as explained in the data dictionary in ***Appendix 3***. Besides the information retrieved from the founders LI pages, additional variables were created. In case the founders information was not in English, texts were translated using the google translate API (SuHun Han 2020). Moreover python NLP library TextBlob (Loria 2020) was used for sentiment analysis of the founders posts.

3.4 Measurement Items

In line with the hypothesis formed in section 3.1, three scores were created to measure founders behaviour online, consisting of the individual variables displayed in ***Appendix 4.*** In order to measure how much information founders were revealing about their experiences, skills and accomplishments, the *self-reporting score* (SRS) was calculated (*H1*). The *networking score* (NS), helps to understand the competences to build and engage with the network (*H2*). Finally, the *activity score* (AS) was created to measure the activeness or passiveness of a founder on LI (*H3*). The variables used within each score were scaled to a range from zero to ten. The final score value was calculated by taking the mean of the respective variables.

4 Results

In the following section, the final dataset is explored and statistically evaluated, testing for significant differences between gender, applying t-test and Chi-Squared test. Moreover, different user segments were identified using the k-means clustering algorithm. The attained clusters were tested for significant differences using Kruskal-Wallis and Chi-Squared test.

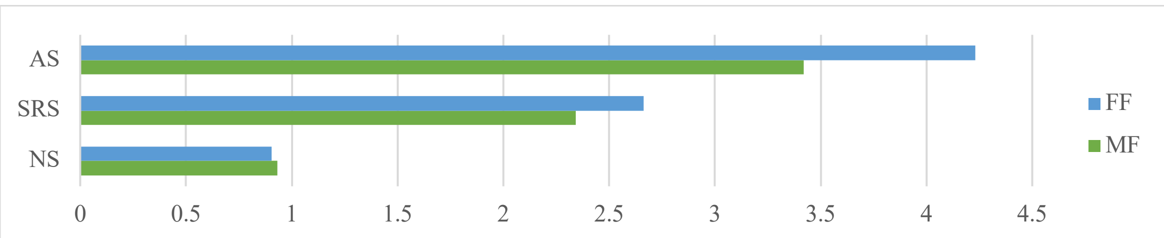
4.1 General Data Exploration

The final dataset consisted of 1284 founders, equally split between FF and MF. ***Appendix 5*** provides an overview of demographic variables. The three industries more than half of all founders operate in were (social) media entertainment (22%), health (16%) and enterprise services (16%). Most founders were based in the UK (30%), Germany (14%) and France (8%). The top cities were the respective capitals. The country with the highest share of FF (73%) was Sweden. The top three schools that founders attended are the University of Oxford, University of Cambridge and Imperial College London and 20% of all founders were holding an executive or doctoral degree. Half of all founders reported between five and ten years of work experience. For 71% of the analysed startups, the last funding type was a Seed round. More than half of all founders employed 1-10 employees and 88% of companies had a founding team size of 1-3 people, with two founders being the most common team size. Most companies were between three and five years old. Half of all founders had a personal CBR below 300,000. FF received a total funding amount of \$1,6 bn compared to \$1,4 bn for MF. The mean funding amount however, was slightly lower for FF than for MF with \$3.4m and \$3.6m respectively. The number of most recent funding rounds for FF was higher for the years 2019 to 2021 than for MF, however, for both the number decreased over the last three (FF) and four (MF) years, as depicted in ***Appendix 6.***

Comparing the founders actions, skills and interests it was found that the most frequent action was ‘like’. FF made more use of the different reaction types than MF. The most frequently used ones were ‘celebrate’ and ‘love’, which were used only one third and two fifth as much by men. The top skill reported by almost 40% of all founders was ‘management’, followed by ‘business strategy’ (35%). The top most followed influencers for all founders were Richard Branson and Bill Gates. While the top ten influencers of MF only consisted of men, the FF top ten included two females, Arianna Huffington and Melinda Gates. For both FF and MF LI post's, around three quarters of all posts contained an image and 10% of all posts only contained text. An overview can be found in ***Appendix*** *7*. Examining the content of the posts, the analysed FF incorporated more emojis, with an average of one emoji per post compared to 0.8 by MF. The most used emoji by both FF and MF was the 'rocket' emoji, which appeared 304 times in posts by FF and 260 times by MF. Differences could also be seen in heart emoji use, FF used 20 different types with a frequency of 352 times compared to only 14 types and a total of 149 times by MF. The top two hashtags for all founders were 'innovation' and 'startup'. 81% of FF have used at least one hashtag per post compared to 71% of MF. Both FF and MF mentioned their company on average in every fourth post. Examining the founders posts NLP results, it was found that FF have slightly higher scores for average polarity and subjectivity than men. For polarity, both, female and male content is on average rather positive (score > 0 < 0.5) and for subjectivity slightly objective (score > 0.25 < 0.5). The Exact values are represented in ***Appendix 8*** and **9**.

Figure 1 below shows that the mean values for FF and MF were similar for NS. For SRS and AS however, FF showed higher means than MF, also shown in ***Appendix 10***. Comparing the individual variables used in the creation of the scores, the biggest disparity could be found for *lenDescription* (M_{FF} = 530, M_{MM}= 431) from SRS. Regarding AS, the average *actionDateDays* differed most (M_{FF} = 692, M_{MF}= 922), translating to 61% of FF being active on LI on a daily basis, compared to 44% of MF. Moreover, on average FF have a higher value for *shareEndorsedSkills* than MF (M_{FF} = 0.84, M_{MF}= 0.78) as included in NS.

Figure 1: *mean scores by gender*



4.2 Correlation Analysis

Through conducting a pearson correlation analysis it was found that *avgLikeCountFP* was positively correlated with *avgCommentCountFP* (r_{FF}=.80, r_{MF}=.81, $p < .001$). Moreover, *numGivenLikes* was strongly positively correlated with *numActions* (r_{FF}=.96, r_{MF}=.96, $p < .001$) and negatively correlated with *actionDateDays* (r_{FF}=-.61, r_{MF}= -.55, $p < .001$). *NumHashtags* was positively correlated with *numFoundersPosts* (r_{FF}=.62, r_{MF}=.56 $p < .001$). *AvgPolarity* was positively correlated with the *avgSubjectivity,* indicating the more positive a post, the more subjective it was (r_{FF}=.73, r_{MF}=.72, $p < .001$). Moreover, the *avgEndorsement* was positively correlated with the number of *subscribers* (r_{FF}=.56, r_{MF}=.56, $p < .001$). Besides that, no significant relations between independent variables were found.

4.3 Statistical Analysis

Independent samples t-tests were applied to the three scores to investigate the significance of the difference between the FF and MF. The underlying assumptions for the use of the t-test, normally distributed data as well as homogeneity of variances, can be tested with the Shapiro-Wilk test and Levene's test respectively (Saunders, Lewis, and Tornhill 2009). The Shapiro-Wilk test finds $p < .05$ for each score. Conducting Levene's test, only SRS tested true for equal variance ($p > .05$). Research found however, that t-test is performing well regardless of normality when comparing two groups of the same size (Poncet et al. 2016; Saunders, Lewis, and Tornhill 2009). For scores with unequal variances (AS & NS, $p < .05$), Welch's test (unequal variances t-test) was conducted as suggested by (Ruxton 2006; Kasuya 2001). While the means of the three scores let undertake that females are on average more active (M_{AS-FF} = 4.23, M_{AS-MF}= 3.42) and more likely to self-report (M_{SRS-FF} = 2.66, M_{SRS-MF}= 2.34), they were found to have weaker network engagement (M_{NS-FF} = .91, M_{NS-MF}= .93). Considering the t-test results, it was possible to validate that the descriptively examined mean values for AS and

SRS were indeed statistically significant ($p < .05$). The null hypothesis for *H1* and *H3* could therefore be rejected. The expectation for *H1* was a lower SRS score for FF than for MF, but the opposite was found. For NS however, it was determined that there was no significant difference between FF and MF ($p = .384$), hence, *H2* had to be rejected. T-test results for the numeric variables in the dataset were conducted under the same procedure and can be found in ***Appendix 9***. Chi-squared test was conducted for the categorical variables. Only the variables *country*, *schoolNames*, *highestDegree*, *schoolYears* and *cbRank* were found significantly different, $p < .05$, as shown in ***Appendix 6***. Regarding actions, skills and interests (***Appendix 7***) as well as posts (***Appendix 8***), all variables differed significantly.

4.4 K-Means Cluster Analysis

In an attempt to determine if different LinkedIn user segments exist and moreover, gender imbalances within these, the k-means clustering algorithm, an unsupervised machine learning algorithm commonly used for customer and market segmentation, was chosen (Dabbura 2018). It requires a predetermined number of clusters (n). Two common methods to identify the optimal value for n are the silhouette analysis[2] and the elbow method[3] (VanderPlas 2016; Navlani, Fandango, and Idris 2021). To identify the best set of features for the algorithm, different sets were tested and the best performing one chosen based on the results of silhouette analysis as depicted in ***Appendix 11***. The highest silhouette coefficient was .62 for n = 2 and found in both, set 5 and set 6. Usually, the highest silhouette coefficient indicates

---

[2] The silhouette analysis provides a measure to compare tightness and separation between clusters. The resulting coefficient is on a range of [-1;1]. A silhouette coefficient near -1 indicates that samples have possibly been assigned to the wrong cluster, while a value close to +1 shows that clusters are well separated

[3] It is based on a plot of the Within-Cluster-Sum-of-Squares. The optimal number for n is chosen where adding an additional cluster does not cause a huge change in variance, i.e. where the line begins to flatten.

the optimal value for n (Bhattacharyya and Dutta 2012). However, besides the coefficient value, one must also consider the thickness and size of the silhouette plot representing the clusters (Pedregosa et al. 2011). In both cases for n = 2 no meaningful practical information could be derived, as it would just separate barely active users from the other users. Therefore the silhouette plots for a range of n = 2-6 for both sets were evaluated. As set 5 with n = 4 shows the most similar plot sizes and second highest coefficient compared to set 6 n=3 ***(Appendix 12)***, it was identified as best option. This decision was further supported by the results of the elbow method. The graph in ***Appendix 13*** shows that the line begins to flatten at n = 4, indicating the optimal number of clusters (Navlani, Fandango, and Idris 2021).

As a result, four LinkedIn user segments could be identified, as shown in ***Appendix 14.*** Segment 1, the *passive users* represent 29% of all founders. They were barely active, reluctant to self-promote and grow their network. In comparison, segment 2, the *enthusiastic users* (17%) were both active and willing to provide content, however, not able to fully leverage their network. The third segment, the *reserved users* (44%), even though active, were providing little content and possess a small network. Finally, the *fully engaged users*, the smallest segment, representing 10% of all founders, were very active and frequent content providers and highly engaged within their large network.

In the next step, the same variables used for gender comparison were chosen to characterize the cluster structure and identify significantly different dimensions. To compare the segments Chi-Squared test for categorical and Kruskal-Wallis test for numeric variables were applied. The results from the demographic profile analysis can be seen in ***Appendix 15.*** Opposed to the results from gender separation, *gender*, *industry*, *numJobs* and *numEmployees* differed significantly between the four segments, *country* and *schoolYears* in contrast did not. FF make up roughly one third of both *the passive users* and the *fully engaged users*. While enthusiastic users are almost balanced between FF and MF, FF are more prominent in the *reserved users* segment. The highest and lowest share of founders with an executive degree can be found in the two extremes, *passive* and *fully engaged* segments. Moreover, *fully engaged users* show the highest share of having more than 10 jobs, indicating a very experienced segment. Most interestingly, *enthusiastic* as well as *fully engaged users* have more users with personal CBRs below 150,000, compared to the other two segments. *Fully engaged* users, even though the smallest segment, show the highest numbers in *emoji* use with an average of two *emojis* per post as well as the highest average number of reported *skills*. Contrary to gender as a groups separator, the four clusters were found to be significantly different from each other for each of the three scores, as proved by the Kruskal-Wallis test ($p < .05$), as shown in ***Appendix 16***. Looking at the scores per segment separated by gender in ***Appendix 17*** it can be seen that females do have higher average scores for each of the three scores in all four segments, except for SRS, NS of the passive users. Looking at the mean value of *subscribers* (NS), *enthusiastic users* made up roughly 57% of *fully engaged users* (7377 subscribers), compared to 22% of *reserved users* and 11% of *passive users*. The fully engaged users also dominated the *lenDescription* variable (SRS) with a mean value of 1740, with 49%, 84%, and 70% of that value for Segment 1-3. As expected, the passive users had the highest amount of *actionDateDays* with an average of 1639 days. For *fully engaged users*, one year has passed for *reserved users* 1.4 years and *enthusiastic users* 1.3 years. Along these variables the segments also differed significantly ($p < .05$).

5 Discussion

This work aimed to examine if differences in the behaviour on PNS between FF and MF exist. In particular, differences in activity, self-reporting and networking have been investigated. This section focuses on the main findings and explains how they relate to the initial literature review and the research question.

In line with research on the gender gap in entrepreneurship, it was found that also for European early stage founders reported on CB, the share of FF is substantially low (11%). However, in comparison to the global number of founders that have raised at least one funding round (Crunchbase 2019, 5), the European numbers make up only half. This raises the question if there is a reluctancy of European FF to publish their company on CB, out of fear of failure in public (Smith and Smith 2021) or if this reflects the true number. Supporting Elam and colleagues findings (2019) that women and men in high income countries are equally likely to start a business for the same education level, the results of the gender comparison for education (*highestDegree, numberUniqueDegrees*) were found to be indifferent, except for *schoolYears* for which FF reported on average more years than men. Moreover the total founding amount, was found to be indifferent between FF and MF, which contradicts research on the gender gap in funding. This could either be true or accredited to the sample selection of MF. If it were true, this could be an indication of a reduction of investors gender bias and homophilic tendencies and a step towards more equal treatment. While the skills reported differed significantly between FF and MF, no factual comparison of the actual skillset is possible. However, either findings on fewer women in STEM can be confirmed (Guzman and Kacperczyk 2019), as programming skills like software development, Java Script and Java were only reported by MF, or the finding of females not reporting their technical skills holds true (Murciano-Goroff 2021). The most common skills that only FF reported were Marketing Communication and Social Media Marketing, confirming business fields with a high share of women (Guzman and Kacperczyk 2019). Comparing the top ten list of influencers followed, similarity-attraction bias for men as well as a lack of missing entrepreneurial role models for women could be confirmed. Moreover, gender stereotypes, like females being considered more caring and warm, can be further validated with this work (Thébaud 2010; Ladge, Eddleston, and Sugiyama 2019). This can be

seen in the differential and more frequent use of the ‘celebrate’ and ‘love’ reactions by FF compared to MF. Moreover, FF posts were on average more positive and subjective and they used heart emojis more often than MF. In general, FF used emojis more often than MF, which supports research findings on this topic (Bai et al. 2019; Oleszkiewicz et al. 2017).
Evaluating the results of the score comparison, it was found that the extend of self-reporting was higher for FF than for MF, which is contrary to the expectations and research findings that women are less likely to do so (Ladge, Eddleston, and Sugiyama 2019; Murciano-Goroff 2021). The reason for higher scores for FF could be due to females being aware of gender bias and actively trying to offset it or by FF simply being more open about providing information online. The second hypothesis had to be rejected, indicating no significant differences in networking between FF and MF which is also contradictory to the findings from literature. However, this might be due to the fact that the developed score did not have the power to measure the quality of the connections, nor if the connections were leading to beneficial support and resources. Taking t*faUsd* into account, this would support however, that both male and females have benefitted from their network equally. However, the correlation analysis could not validate any significant relation. Finally, the hypothesis, that FF were more active on LI than MF could be confirmed. However, based on this dataset it is not possible to validate, if the assumption that this is due to FF being aware of investor bias and therefore try to signal competence, is correct or whether this is due to other reasons.
Through the cluster analysis four different LI user segments were identified. Female founders were less present in the two extreme segments, *passive* and *fully engaged users*, which could be due to an unwillingness of being perceived as either lazy or uncommitted or on the other end too stubborn and exaggerating. In line with research on social capital and its essence for a ventures success, it was found that founders with the highest average NS, *enthusiastic users,* did also show the highest share of low CBRs, the relation was found insignificant however.

5.1 Limitation and future research

Despite the findings and insights gained, this study as part of a master's thesis in founders behaviour on PNS was limited in time and resources. Thus, some limitations must be considered when interpreting the results and conclusions as pointed out in this section. First of all, the collected data must be critically evaluated. While CB claims to be the most up to date source in the startup ecosystem it also states that there are gaps in the dataset. This can be a result of companies not listing each or any founder at all, or due to other missing information, like no gender associated to the founder. Hence the sample data extracted might not be representative for the entire sample, indicating the possibility of Type I errors. This also applies to the selection of MF, which was based on the CBR, 'Industry' and 'Funding stage' variables and not on randomness, which on one side ensures comparability between FF and MF but might also inflate Type I Error. Moreover, the CBR is highly fluid and might only be consistent for a couple of days or even hours, which could result in different findings if the matching would be repeated CBRs on any other date.

In addition, the data retrieved from LinkedIn must be critically evaluated. Differences in the information provided might not only be due to differences in experience or skills but also due to the founders personal style and preference of displaying information online. Furthermore, different founders share their personal information in a different degree of detail. For example, some report all skills they have, others report only those, they are truly competent in. Therefore, it must be considered, that the self-reported data is not fact-checked and can only be assumed to be true. Moreover, due to the time and size specifications of this thesis, only the 200 most recent activities on LI were analysed and only CB and LI were chosen as data sources, while others, e.g. AngelList, PitchBook or twitter were not taken into consideration.

The recommendation for future studies would therefore be, to find means to compare actual experience and skills with the reported information online. Conducting a long-term study, that is able to identify and prove a relationship between startup founders behaviour on PNS with venture success would also be beneficial. It would moreover be advisable, to conduct interviews with FF post-analysis of their online behaviour, to determine if they have been aware of the gender bias, and whether this impacted their behaviour. Finally, it will be interesting to see, how and if the increasing share female founder role models impacts other women's decisions to start a venture.

6 Conclusion

This work has outlined the different causes for the gender gap in entrepreneurship and in startup funding based on previous literature on this topic. Factors identified include negative gender stereotypes for women and differences in human and social capital. The lower rate of investor backed female founders was explained by structural factors and self-imposed limitations as well as biased investors with homophilic tendencies. Beyond that, a research gap regarding effects of the online world on the gender gap in entrepreneurship was identified. This work therefore analysed the online behaviour of European early stage founders on the PNS LinkedIn and aimed to identify differences in the activity, likeliness to self-report and networking between female and male founders. While female founders were found to be significantly more active on the site, they also tended to self-report more information than male founders. Differences regarding networking were found to be insignificant. Beyond that, some gender stereotypes identified in the literature review could be confirmed while others were neglected based on the sample at hand. Behaviour on PNS is an upcoming research area that has a great future perspective and will gain increasing importance. Therefore, some investigation gaps will need to be addressed in future research on this topic.

Bibliography

Abouzahr, Katie, Matt Krentz, John Harthorne, and Frances Brooks Taplett. 2018. 'Why Women-Owned Startups Are a Better Bet'. BCG.Com. 2018. https://www.bcg.com/publications/2018/why-women-owned-startups-are-better-bet.

Adler, Paul S., and Seok-Woo Kwon. 2002. 'SOCIAL CAPITAL: PROSPECTS FOR A NEW CONCEPT'. *Academy of Management Review* 27 (1): 17–40.

APX. 2020. 'Introducing APX Female Founder Office Hours and AMA's'. *APX Voices* (blog). 2020. https://medium.com/apxberlin/why-are-only-15-percent-of-european-founders-female-9fdcaef35962.

Bai, Qiyu, Qi Dan, Zhe Mu, and Maokun Yang. 2019. 'A Systematic Review of Emoji: Current Research and Future Perspectives'. *Frontiers in Psychology* 10.

Banerji, Devika, and Torsten Reimer. 2019. 'Startup Founders and Their LinkedIn Connections: Are Well-Connected Entrepreneurs More Successful?' *Computers in Human Behavior* 90: 46–52.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws. 2017. 'Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment'. *The Journal of Finance* 72 (2): 509–38.

Bhattacharyya, Siddhartha, and Paramartha Dutta. 2012. *Handbook of Research on Computational Intelligence for Engineering, Science, and Business*. Vol. 1. 2 vols. Hershey PA: IGI GLobal.

Bird, Steven, Edwardd Loper, and Ewan Klein. 2006. *Natural Language Processing with Python. O'Reilly Media Inc*. O'Reilly Media Inc.

Bittner, Ashley, and Brigette Lau. 2021. 'Women-Led Startups Received Just 2.3% of VC Funding in 2020'. Harvard Business Review. 2021. https://hbr.org/2021/02/women-led-startups-received-just-2-3-of-vc-funding-in-2020.

Cha, Youngjoo, and Kim A. Weeden. 2014. 'Overwork and the Slow Convergence in the Gender Gap in Wages'. *American Sociological Review* 79 (3): 457–84.

Coleman, Susan, and Alicia Robb. 2009. 'A Comparison of New Firm Financing by Gender: Evidence from the Kauffman Firm Survey Data'. *Small Business Economics* 33 (4): 397.

Crunchbase. 2019. 'Funding to Female Founders'.

———. 2021a. 'Query Builder | Companies'. Crunchbase. 2021. https://www.crunchbase.com/search/organization.companies/520374bbb1827e3bda9d1db59d08418a.

———. 2021b. 'What Industries Are Included in Crunchbase?' Crunchbase | Knowledge Center. 2021. https://support.crunchbase.com/hc/en-us/articles/360043146954-What-Industries-are-included-in-Crunchbase-.

———. 2022. 'About Crunchbase'. Crunchbase. 2022. https://about.crunchbase.com/about-us/.

Dabbura, Imad. 2018. ‘K-Means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks’. *Towards Data Science* (blog). 2018. https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a.

Elam, Aamanda B., Candida G. Brush, Patricia G. Greene, Benjamin Baumer, Monica Dean, and René Heavlow. 2019. ‘GEM 2018/2019 Women’s Entrepreneurship Report’. 978-1-9160178-1–8. Global Entrepreneurship Monitor. GEM, Babson College, Smith College, KOREA ENTREPRENEURSHIP FOUNDATION.

European Startups. 2021. ‘The European Unicorn Club’. European Startups. 2021. https://app.europeanstartups.co/unicorns/f/regions/allof_European%20Union/tags/allof_verified%20unicorns%20and%20%241b%20exits.

Ewens, Michael, and Richard R. Townsend. 2020. ‘Are Early Stage Investors Biased against Women? | Elsevier Enhanced Reader’. *Journal of Financial Economics* 135: 653–77.

Fischer, Eileen, and A. Rebecca Reuber. 2014. ‘Online Entrepreneurial Communication: Mitigating Uncertainty and Increasing Differentiation via Twitter’. *Journal of Business Venturing* 29: 565–83.

Flèche, Sarah, Anthony Lepinteur, and Nattavudh Powdthavee. 2021. ‘The Importance of Capital in Closing the Entrepreneurial Gender Gap: A Longitudinal Study of Lottery Wins’. *Journal of Economic Behavior & Organization* 188: 591–607.

Foschi, Martha. 2000. ‘Double Standards for Competence: Theory and Research’. *Annual Review of Sociology* 26 (1): 21–42.

Goldin, Claudia. 2016. ‘Human Capital’. In *Handbook of Cliometrics*, edited by Claude Diebolt and Michael Haupert, 55–86. Heidelberg: Springer-Verlag.

Gompers, Paul, and Sophie Wang. 2017. ‘Diversity in Innovation’. w23082. Cambridge, MA: National Bureau of Economic Research.

Guzman, Jorge, and Aleksandra (Olenka) Kacperczyk. 2019. ‘Gender Gap in Entrepreneurship’. *Research Policy* 48 (7): 1666–80.

Henrekson, Magnus, and Anita Du Rietz. 1999. ‘Testing the Female Underperformance Hypothesis’. *Small Business Economics* 14 (1).

Hodgson, Leah. 2021. ‘Senior Female VCs Call out Major Funding Disparity in Europe’. PitchBook. 2021. https://pitchbook.com/news/articles/senior-female-vcs-call-out-major-funding-disparity-in-europe.

Jin, Fujie, Lorin Hitt, and Andy Wu. 2017. ‘Social Is the New Financial: How Startup Social Media Activity Influences Funding Outcomes’. In *Proceedings*. Vol. 1. Mack Institute for Innovation Management. https://mackinstitute.wharton.upenn.edu/2017/social-is-the-new-financial-how-startup-social-media-activity-influences-funding-outcomes/.

Kacperczyk, Olenka, and Peter Younkin. 2021. ‘A Founding Penalty: Evidence from an Audit Study on Gender, Entrepreneurship, and Future Employment’. *Organization Science*, April. https://doi.org/10.1287/orsc.2021.1456.

Kanze, Dana, Laura Huang, Mark A. Conley, and E. Tory Higgins. 2017. ‘We Ask Men to Win and Women Not to Lose: Closing the Gender Gap in Startup Funding’. *Academy of Management Journal* 61 (2): 586–614.

Kasuya, Eiiti. 2001. ‘Mann–Whitney U Test When Variances Are Unequal’. *Animal Behaviour*, no. 61: 1247–49.

Ladge, Jamie, Kimberly A. Eddleston, and Keimei Sugiyama. 2019. ‘Am I an Entrepreneur? How Imposter Fears Hinder Women Entrepreneurs’ Business Growth’. *Business Horizons* 62: 615–24.

Lewin, Amy. 2021. ‘Germany’s First VC Fund for Female Founders Launches’. *Sifted.Eu* (blog). 2021. https://sifted.eu/articles/vc-fund-female-founders-auxxo/.

Leys, Christophe, Marie Delacre, Youri L. Mora, Daniël Lakens, and Christophe Ley. 2019. ‘How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration’. *International Review of Social Psychology* 32 (1): 5.

LinkedIn. 2022. ‘About LinkedIn’. About LinkedIn. 2022. https://about.linkedin.com/.

Loria, Steven. 2020. ‘TextBlob: Simplified Text Processing’. TextBlob. 2020. https://textblob.readthedocs.io/en/dev/.

Malmstrom, Malin, Jeaneth Johansson, and Joakim Wincent. 2017. ‘We Recorded VCs’ Conversations and Analyzed How Differently They Talk About Female Entrepreneurs’. *Harvard Business Review*, 17 May 2017. https://hbr.org/2017/05/we-recorded-vcs-conversations-and-analyzed-how-differently-they-talk-about-female-entrepreneurs.

Manolova, Tatiana S., Candida G. Brush, and Linda F. Edelman. 2008. ‘What Do Women Entrepreneurs Want?’ *Strategic Change* 17 (3–4): 69–82. https://doi.org/10.1002/jsc.817.

Mckeever, Edward, Alistair Anderson, and Sarah Jack. 2014. ‘Social Embeddedness in Entrepreneurship Research: The Importance of Context and Community’. In , 222–36.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. ‘Birds of a Feather: Homophily in Social Networks’. *Annual Review of Sociology* 27 (1): 415–44.

Meunier, Frédéric, Yulia Krylova, and Rita Ramalho. 2017. ‘Women’s Entrepreneurship: How to Measure the Gap between New Female and Male Entrepreneurs?’ Working Paper 8242. Policy Research Working Paper. World Bank Group.

Murciano-Goroff, Raviv. 2021. ‘Missing Women in Tech: The Labor Market for Highly Skilled Software Engineers’. *Management Science*.

Nahapiet, Janine, and S. Ghoshal. 1998. ‘Social Capital, Intellectual Capital, and the Organizational Advantage’. *Academy of Management Review* 23 (2): 242–66.

Navlani, Avinash, (Armando Fandango, and Ivan Idris. 2021. *Python Data Analysis*. 3rd ed. Livery Place 35 Livery Street Birmingham B3 2PB, UK.: Packt Publishing Ltd.

Obschonka, Martin, Christian Fisch, and Ryan Boyd. 2017. ‘Using Digital Footprints in Entrepreneurship Research: A Twitter-Based Personality Analysis of Superstar Entrepreneurs and Managers’. *Journal of Business Venturing Insights* 8: 13–23.

OECD and European Commission. 2019. ‘The Missing Entrepreneurs 2019: Policies for Inclusive Entrepreneurship’. Paris: OECD Publishing.

Oleszkiewicz, Anna, Maciej Karwowski, Katarzyna Pisanski, Piotr Sorokowski, Boaz Sobrado, and Agnieszka Sorokowska. 2017. ‘Who Uses Emoticons? Data from 86 702 Facebook Users’. *Personality and Individual Differences* 119: 289–95.

Pedregosa, F., G. Varoquaux, A Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. ‘Scikit-Learn: Machine Learning in {P}ython’. *Journal of Machine Learning Research* 12: 2825–30.

Poncet, Antoine, Delphine S. Courvoisier, Christophe Combescure, and Thomas V. Perneger. 2016. ‘Normality and Sample Size Do Not Matter for the Selection of an Appropriate Statistical Test for Two-Group Comparisons’. *European Journal of Research Methods for the Behavioural and Social Sciences* 12 (2).

Reynolds, Paul D., Michael Hay, and S. Michael Camp. 1999. ‘GEM 1999 Executive Report’. Executive Report. Global Entrepreneurship Monitor. Kauffman Center for Entrepreneurial Leadership at the Ewing Marion Kauffman Foundation. https://www.gemconsortium.org/report/gem-1999-global-report.

Riani, Abdo. 2021. ‘The Difference Between An Entrepreneur And A Startup Founder’. *Forbes*, 22 April 2021. https://www.forbes.com/sites/abdoriani/2021/04/22/the-difference-between-an-entrepreneur-and-a-startup-founder/.

Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, and Adam Jatowt. 2020. ‘YAKE! Keyword Extraction from Single Documents Using Multiple Local Features’. *Information Sciences* 509: 257–89. https://doi.org/10.1016/j.ins.2019.09.013.

Ruxton, Graeme D. 2006. ‘The Unequal Variance T-Test Is an Underused Alternative to Student’s t-Test and the Mann–Whitney U Test’. *Behavioural Ecology* 17 (4): 688–90.

Saunders, Mark, Philip Lewis, and Adrian Tornhill. 2009. *Research Methods for Business Students*. 5th ed. Edinburgh Gate Harlow Essex CM20 2JE England: Pearson Education Limited.

Smith, Claudia G., and J. Brock Smith. 2021. ‘Founders’ Uses of Digital Networks for Resource Acquisition: Extending Network Theory Online’. *Journal of Business Research* 125: 466–82.

Smith, Claudia, J. Brock Smith, and Eleanor Shaw. 2017. ‘Embracing Digital Networks: Entrepreneurs’ Social Capital Online’. *Journal of Business Venturing* 32 (1): 18–34.

Stam, Wouter, Souren Arzlanian, and Tom Elfring. 2014. ‘Social Capital of Entrepreneurs and Small Firm Performance: A Meta-Analysis of Contextual and Methodological Moderators’. *Journal of Business Venturing* 29 (1): 152–73. https://doi.org/10.1016/j.jbusvent.2013.01.002.

SuHun Han. 2020. ‘Googletrans: Free and Unlimited Google Translate API for Python — Googletrans 3.0.0 Documentation’. Googletrans 3.0.0 Documentation. 2020. https://py-googletrans.readthedocs.io/en/latest/.

Teare, Gené. 2020. ‘Global VC Funding To Female Founders Dropped Dramatically This Year’. Crunchbase News. 21 December 2020. c.

Thébaud, Sarah. 2010. ‘Gender and Entrepreneurship as a Career Choice: Do Self-Assessments of Ability Matter?’ *Social Psychology Quarterly* 73 (3): 288–304.

Thomas, Rachel, Marianne Cooper, Kate Mcshane Urban, Gina Cardazone, Ali Bohrer, Sonia Mahajan, Lareina Yee, et al. 2021. ‘Women in the Workplace’. McKinsey & Company, LeanIn.Org.

Tinkler, Justine E., Kjersten Bunker Whittington, Manwai C. Ku, and Andrea Rees Davies. 2015. ‘Gender and Venture Capital Decision-Making: The Effects of Technical Background and Social Capital on Entrepreneurial Evaluations’. *Social Science Research* 51 (May): 1–16.

Upton, Lee O., Emma J. Broming, and Dr. Rebecca L. Upton. 2018. ‘Research on Women Entrepreneurs’ Social Networks’. National Women’s Business Council.

Uzzi, Brian. 2019. ‘Research: Men and Women Need Different Kinds of Networks to Succeed’. *Harvard Business Review*, 2019. https://hbr.org/2019/02/research-men-and-women-need-different-kinds-of-networks-to-succeed.

VanderPlas, Jake. 2016. *Python Data Science Handbook*. 1005 Gravenstein Highway North, Sebastopol, CA 95472: O’Reilly Media, Inc.

Vossenberg, Saskia. 2013. ‘Women Entrepreneurship Promotion in Developing Countries: What Explains the Gender Gap in Entrepreneurship and How to Close It?’ Working Paper 2013/08. Maastricht School of Management.

West, Collin, and Gopinath Sundaramurthy. 2020. ‘Women VCs Invest in Up to 2x More Female Founders’. Kauffman Fellows. 2020. https://www.kauffmanfellows.org/journal_posts/women-vcs-invest-in-up-to-2x-more-female-founders.

Appendix

Appendix 1: Table of Contents

Appendix 2: New industry labels

New Industry Labels	CB Industry group
Consumer products	community and lifestyle, sports, clothing and apparel, consumer goods, events, consumer electronics
Food & beverage	food and beverage
Financial services	financial services, payments, lending & investments
Education	education
Health	health care
IT	information technology, privacy & security
Mobility	transportation, travel & tourism, navigation & mapping
(social) media & entertainment	media & entertainment, gaming, music & audio, platforms, video, sports, content & publishing, commerce & shopping
Enterprise services	messaging & telco, professional services, sales & marketing, administrative services, advertising
Manufacturing	manufacturing
Agriculture farming	agriculture and farming
Real estate	Real Estate
Science & engineering	science and engineering, biotechnology
Privacy &security	government and military, privacy and security
Resources	energy, natural resources, sustainability

Appendix 3: Data dictionary

Name	Type	Definition	Calculation/ creation	Example	Origin
gender	str	The founders gender		female	CB
cbRank	int	A founders Crunchbase rank		75000	CB
country	str	The country of the founders companies HQ		UK	CB
city	str	The city of the founders companies HQ		London	CB
industryLabel	str	The industry the company operates in	Derived from CB Industry labels (see ***Appendix*** 1)	health	CB
companyAge	int	The number of years from founded date until 2021	2021- founded date	8	CB
numFounders	int	The founding team size		2	CB
lastFunding Stage	str	The last type of funding series		Seed	C
lastFunding Date	str	The date of the last funding round		10-2021	CB
tfaUsd	int	Total founding amount in USD		2018	CB
numJobs	int	Total number of jobs		7	LI
yearsJob Experience	int	Years of job experiences reported	Number of years from start date of the first job until the end of 2021, assuming there have been no gap years in-between. For the founders that did not add date ranges to their job experiences, the job years	10	LI

			were calculated by multiplying the number of jobs by the mean yob years (1.97 for women and 2.0 for men).		
numSchools	int	Total number of schools		3	LI
schoolYears	int	The number of school years (excluding secondary education and other)	For founders that did not provide date ranges, the average number of other founders total school years with the same degree of education is imputed. The average number of years for each degree for women were: 9.3 years for doctoral 7.3 for executive programs, 5.5 for postgraduate, and 3.4 for undergraduate. The average number of school years is 5.23. For men: 8.7 years for doctoral, 6.6 for executive programs, 5.7 for postgraduate, and 3.8 for undergraduate	6	LI
schoolNames	str	Name of the schools attended as reported on LinkedIn		University of Amsterdam	LI
highestDegree	str	The highest degree reported	1449 different Degree terms existed, to reduce the variety, new, unique Degree names were created and ranked: secondary education (1), undergraduate (2), postgraduate (3), executive (4), doctoral (5).	postgraduate	LI
numUnique Degrees	int	The number of unique degrees	excluding high school and others like summer school. i.e. if two Bachelors degrees and one Masters degree were reported, the count would be 2	3	LI
typeFP	str	List of the different types of posts ((Image, video (both, from LinkedIn or external source), text, article, poll, live video) used by founder and their respective occurrence count		[('Image', 10), ('text', 4), ('acticle', 1)]	LI
avgLike CountFP	int	The average number of likes on a founders posts	Total number of likes of all posts/ number of posts	30	LI
avgComment CountFP	int	Average number of comments on a founders posts	Total number of comments of all posts / number of posts	4	LI
emojis	str	The emojis used in all posts stored in a list		[🚀 , 👉 , 👇 , 💪 , 🙌 , 🤩]	LI
numEmojis	int	The total number of emojis used		10	LI

avgNum Emojis	int	The average number of emojis per post	Total number of emojis of all posts/ number of posts	1	LI
hashtags	str	All hashstags of all posts stored in list		['startups', 'ai', 'sustainability', 'fintech']	LI
avgNum Hashtags	int	The average number of hashtags per post	Total number of hashtags of all posts / number of posts	2	LI
avgNum Company Mention	int	The average number of company mentions per post	Total number of company mentions of all posts / number of posts	0.8	LI
avgSubject-ivity	int	The average subjectivity score of all posts *	Sum of subjectivity scores for all posts / number of posts	0.3	LI
avgPolarity	int	The average polarity score of all posts**	Sum of polarity scores for all posts / number of posts	0.15	LI
actionType	str	List of the different action types (like, like_comment, post, comment, celebrate, love, reply_comment, support, insightful curious)		['like_c', 'like', 'comment', 'post', …]	LI
numGiven Likes	int	The number of likes given by a founder		130	LI
allSkills	str	List of all skills resported by a founder		['Management', 'Business Strategy', 'Strategy', 'Marketing']	LI
avgEndorse-ment	int	The average number of endorsements per skill reported	Sum of all skills/ sum of all endorsements	8	LI
numInfluen-cers	int	Number of influencers followed		7	LI
Influencer Name	str	Name of the followed influencers		['Bill Gates', 'Richard Branson']	LI
lenHeadline	int	Number of characters in headline text		51	LI
lenDescription	int	Number of characters in description text		480	LI
avgLenJob Desc	int	Average number of characters of job description texts		608	LI
numSkills	int	Number of skills reported		22	LI
avgPost LenFP	int	Average number of characters of a founders post	Sum of all post length/ number of posts	300	LI
sumAllAcc	int	The sum of all accomplishments (Publications, Patents, Courses, Projects, Honors & awards, Test scores, Languages,		4	LI

		Organizations, Causes)			
subscribers	int	Number of followers on LinkedIn		3640	LI
Like Comment Ratio	int	The share of likes to comments on a founders posts	Likes/(likes+comments)	0.95	LI
Share Endorsed Skills	int	The share of the endorsed skills out of all skills	Number of endorsed skills / all skills	0.8	LI
Avg Engagement Rate	int	The average engagement rate over all posts	((sum of all likes /number of posts)+ (sum of all comments /number of posts)/subscribers)	0.3	LI
Received RecRate	int	The share of the received recommendations in relation to the number of subscribers	Number of received recommendations / subscribers	0.05	LI
Endorsement Rate	int	The number of endorsements as share of subscribers	Number of endorsements/ subscribers	0.13	LI
avgNum Hashtags	int	The average number of hashtags used per post		26	LI
numHashtags	int	The total number of hashtags used		26	LI
avgNum Company Mention	int	The avg number of company mentions per post	Number of company mentions / number of posts	0.8	LI
Num Founders Posts	int	Number of a founders' posts		13	LI
numActions	int	Number of all actions (including posts, maximum 200)		185	LI
Action DateDays	int	Number of days from first until most recent action (the more days, the less active the founder)	LinkedIn provides the time delta between the date of the action and the current time and date of observation in the form of minutes, hours, days, weeks, months or years. For all actions per founder, these date formats were converted into days, e.g. 18 hours are 0.75 days, 2 months are 60 days etc.	720	LI
action Frequency	int	The frequency in days in which actions occur	Total number of days/ total number of actions	7	LI
sumAll Interests	int	The sum of interests (influencers, groups, schools, companies)		50	LI
givenRec Rate	int	The share of the given recommendations in relation to the subscribers	Nmber of given recommendations / subscribers	0.07	LI

likeAction Ratio	int	The number of likes given as share of all actions (including all types of likes).	Number of likes given / number of actions	0.8	LI

* The subjectivity score is a result of NLP Sentiment Analysis. The algorithm returns a score on a scale from 0 to 1 for the analysed text. If score < 0.25 text is considered 'Objective', if score < 0.5, itis considered 'Slightly objective', if score < 0.75 it is considered 'Slightly subjective', if score < 1 it is considered 'Subjective'

** The polarity score is a result of NLP Sentiment Analysis. The algorithm returns a score on a scale from -1 to 1 for the analysed text. If the score is < -0.5 the sentiment of the text is considered 'Negative', if score < 0, it is considered 'Rather negative', if score == 0 it is considered 'Neutral'. If the score < 0.5 it is considered 'Rather positive' and if the score is < 1 it is considered 'Positive'

Appendix 4: Individual variables used for score creation

Score	**Description**	**Variables**
Self-Reporting Score (SRS)	Measures how much information founders are revealing about themselves	lenHeadline lenDescription avgLenJobDesc numSkills avgPostLenFP sumAllAcc
Activity Score (AS)	Measures how active a founder is	numFoundersPosts numActions actionDateDays[1] givenRecRate likeActionRatio[1] sumAllInterests
Networking Score (NS)	how well a founder is engaging his network into his actions	subscribers LikeCommentRatio[2] shareEndorsedSkills AvgEngagementRate receivedRecRate endorsementRate avgNumHashtags

[1] As the emphasis is put on activeness, and in these cases a high value would suggest passiveness, the value of the ratio is subtracted from 1.

[2] Assuming that more comments (the lower the ratio) mean higher network engagement, hence the value of the ratio is subtracted from 1.

[3] Values are assigned to each degree: undergraduate: 1, postgraduate: 2, executive education: 3, PhD: 4

Appendix 5: Demographic characteristics

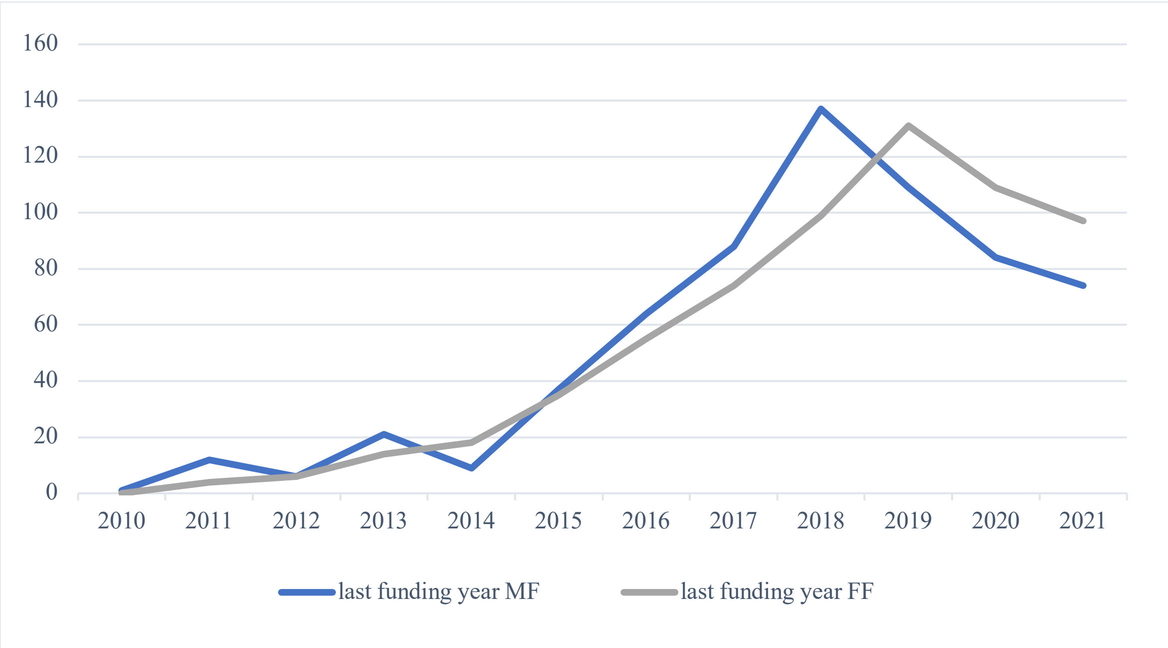
Variable	Subcategory	FF %	MF %	Chi-Squared test p-value
gender	female	50%	0%	1
	Male	0%	50%	
industry	Social media & entertainment	11%	11%	1
	Health	8%	8%	
	Enterprise services	8%	8%	
	Consumer products	6%	6%	
	Financial services	4%	4%	
	Science & engineering	3%	3%	
	Mobility	2%	2%	
	IT	2%	2%	
	Food & beverage	2%	2%	
	Education	1%	1%	
	Resources	1%	1%	
	Real estate	1%	1%	
	Agriculture & farming	1%	1%	
	Manufacturing	1%	1%	
	Privacy & security	< 1%	< 1%	
top 10 countries*	United Kingdom	17%	12%	< 0.05
	Germany	6%	8%	
	France	4%	4%	
	Sweden	3%	1%	
	Spain	3%	3%	
	The Netherlands	2%	2%	
	Finland	2%	2%	
	Switzerland	2%	3%	
	Ireland	2%	1%	
	Denmark	2%	2%	
top 10 cities*	London	13%	9%	0.056
	Berlin	3%	3%	
	Paris	2%	3%	
	Stockholm	3%	1%	
	Barcelona	1%	1%	
	Dublin	1%	1%	
	Amsterdam	1%	1%	
	Copenhagen	1%	1%	
	Helsinki	1%	1%	
	Madrid	1%	1%	
top 3 schools*	University of Oxford	1%	1%	< 0.05

	University of Cambridge	1%	1%	
	Imperial College London	1%	1%	
highestDegree	postgraduate	23%	20%	< 0.05
	undergraduate	12%	10%	
	other	6%	9%	
	doctoral	5%	6%	
	executive	4%	5%	
	secondary education	0%	0%	
numJobs	< 5	11%	12%	0.56
	< 10	26%	26%	
	< 15	11%	10%	
	> 15	2%	2%	
lastFundingStage	Seed	36%	36%	0.80
	Series A	8%	9%	
	Pre-Seed	4%	4%	
	Angel	2%	1%	
numEmployees	1-10	26%	29%	0.98
	11-50	21%	19%	
	51-100	2%	2%	
	101-250	< 1%	< 1%	
	251-500	< 1%	< 1%	
numFounders	1	12%	11%	0.38
	2	21%	20%	
	3	12%	13%	
	4	4%	5%	
	5	1%	1%	
	6	< 1%	< 1%	
	7	< 1%	0%	
cbRank	< 150,000	12%	12%	< 0.05
	< 300,000	12%	12%	
	< 600,000	16%	18%	
	< 1,200,000	10%	8%	
numUniqueDegrees	0	2%	3%	< 0.05
	1	13%	17%	
	2	21%	18%	
	3	12%	10%	
	4	2%	2%	
	5	< 1%	< 1%	
SchoolYears	< 3	10%	15%	< 0.05
	< 6	22%	15%	
	<=9	13%	14%	
	> 9	5%	6%	

companyAge	< 3	4%	4%	< 0.05
	< 6	24%	26%	
	<= 9	20%	18%	
	> 9	2%	2%	

*As the original amount of categories was too large to produce reliable results for the Chi-Squared test, only the top 10 categories were used for testing.

Appendix 6: Years of the last funding rounds



Appendix 7: founders actions, skills and interests

Variable	Subcategory	Female N	Male N	Chi-Squared test p-value
actionType	like	63207	59289	< 0.05
	like comment	10415	6626	
	post	8334	8333	
	comment	6608	5010	
	celebrate	5272	3299	
	love	4418	1749	
	reply comment	3848	2383	
	support	575	385	
	insightful	351	343	
	curious	117	106	
top 5 skills*	Management	272	231	< 0.05

	Business Strategy	230	223	
	Strategy	227	164	
	Entrepreneurship	216	204	
	Marketing	207	125	
top 5 influencers*	Bill Gates	171	132	< 0.05
	Richard Branson	151	140	
	Simon Sinek	98	72	
	Jeff Weiner	92	60	
	Arianna Huffington	110	30	

*As the original amount of categories was too large to produce reliable results for the Chi-Squared test, only the top 10 categories were used for testing.

Appendix 8: Founders posts

Variable	**Subcategory**	**FF N**	**MF N**	**Chi-Squared test p-value**
typeFP	Image	6301	6024	< 0.05
	Text	811	884	
	Video (LinkedIn Source)	541	584	
	Article	386	442	
	Video (External Source)	255	323	
	Poll	23	29	
	Live Video	5	3	
	Overlay Image	2	3	
top 10 emojis*	🚀	304	260	<0.05
	👇	216	155	
	👉	216	191	
	👏	143	63	
	🙏	130	84	
	❤️	129	50	
	💪	114	85	
	🙌	113	56	
	🤩	108	45	
	🎉	103	79	
heart emojis	❤️	129	50	< 0.05
	💜	40	0	
	😍	38	31	
	💚	30	17	
	🥰	26	8	
	💙	21	6	
	💛	17	14	

	🧡	14	2	
	♥	11	6	
	💖	4	3	
	💕	4	3	
	💗	3	0	
	💘	3	1	
	🤍	3	1	
	💓	3	0	
	🖤	2	1	
	❣️	1	6	
	💟	1	0	
	😻	1	0	
	🤎	1	0	
top 10 hashtags*	innovation	247	184	< 0.05
	startup	139	184	
	ai	142	168	
	hiring	137	125	
	startups	82	86	
	sustainability	100	57	
	fintech	86	67	
	technology	86	54	
	digitalhealth	44	96	
	healthcare	56	55	

*As the original amount of categories was too large to produce reliable results for the Chi-Squared test, only the top 10 categories were used for testing.

Appendix 9: Statistics for numeric variables

Variable	N	FF M	FF SD	MF M	MF SD	T-test p-value
cbRank	1284	373791	257099	353457	243162.3	0.15
numFounders	1284	2.26	1.06	2.29	0.99	0.65
tfaUsd	1284	3388699	5967852	3634378	7395597	0.22
numJobs	1284	7.59	4.02	7.64	4.24	0.81
yearsJobExperience	1284	15	6	16	7	0.11
numSchools	1284	2.78	1.56	2.41	1.49	< 0.05
schoolYears	1284	5.23	3.5	4.94	3.69	0.15
numUniqueDegrees	1284	1.99	0.91	1.78	0.96	< 0.05
avgLikeCountFP	1284	33	52	25	47	< 0.05
avgCommentCountFP	1284	3.14	6.94	2.14	5.26	< 0.05
numEmojis	1284	11	20	8	20	< 0.05

avgNumEmojis	1284	1	1.48	0.84	1.58	< 0.05
numHashtags	1284	28	50	24	50	0.17
avgNumCompanyMention	1284	0.24	0.4	0.22	0.35	0.14
avgSubjectivity	1284	0.44	0.14	0.41	0.16	< 0.05
avgPolarity	1284	0.25	0.12	0.22	0.13	< 0.05
numGivenLikes	1284	142	61.62	119.7	73.05	< 0.05
sumEndorsements	1284	209	217	187	204	0.06
avgEndorsement	1284	8.5	9.38	7.66	8.78	0.10
numInfluencers	1284	5.18	7.05	3.43	5.03	< 0.05
actionFrequency	1284	16	46	22	35	< 0.05
lenHeadline	1284	55	38	48	33	< 0.05
lenDescription	1284	530	610	431	535	< 0.05
avgLenJobDesc	1284	656	788	573	755	0.05
numSkills	1284	~~23~~	~~12~~	23	13	0.71
avgPostLenFP	1284	347	211	318	226	< 0.05
sumAllAcc	1284	4.02	3.58	3.13	2.66	< 0.05
subscribers	1284	2558	2335	2231	2008	< 0.05
LikeCommentRatio	1284	0.93	0.07	0.93	0.07	< 0.05
shareEndorsedSkills	1284	0.84	0.3	0.78	0.32	< 0.05
AvgEngagementRate	1284	0.02	0.04	0.02	0.02	< 0.05
receivedRecRate	1284	0.03	0.06	0.03	0.08	0.34
endorsementRate	1284	0.1	0.08	0.1	0.1	0.19
avgNumHashtags	1284	2.05	2.25	2	2.51	0.22
numFoundersPosts	1284	13	14	12	15	0.34
numActions	1284	161	65	136	78	< 0.05
actionDateDays	1284	692	780	922	867	< 0.05
givenRecRate	1284	0.04	0.07	0.05	0.09	0.29
like_action_ratio	1284	0.87	0.13	0.86	0.15	0.48
sumAllInterests	1284	80	76	62	66	< 0.05

Appendix 10: Score statistics by gender

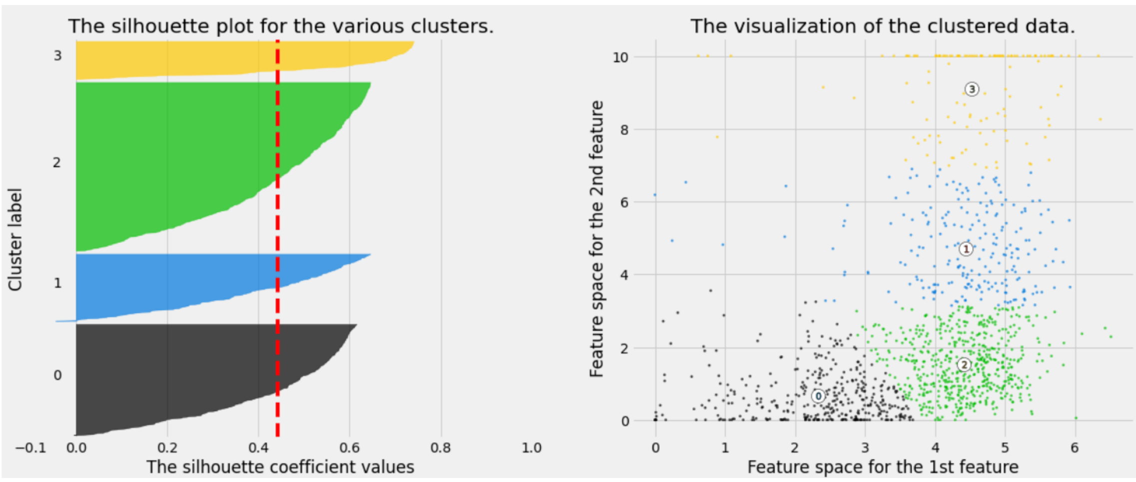
Scores	**Gender**	**N**	**M**	**SD**	**Percentiles**			**T-test**	**Welch's test**
					25	**50**	**75**	**p-value**	**p-value**
SRS	FF	642	2.66	1.25	1.75	2.50	3.40	< 0.05	
	MF	642	2.34	1.18	1.48	2.16	3.03		
NS	FF	642	0.91	0.50	0.57	0.81	1.15		0.38
	MF	642	0.93	0.57	0.51	0.83	1.26		
AS	FF	642	4.23	1.05	3.62	4.53	4.96	< 0.05	
	MF	642	3.42	1.26	2.61	3.73	4.25		

Appendix 11: Results of Silhouette Analysis for the different sets of variables

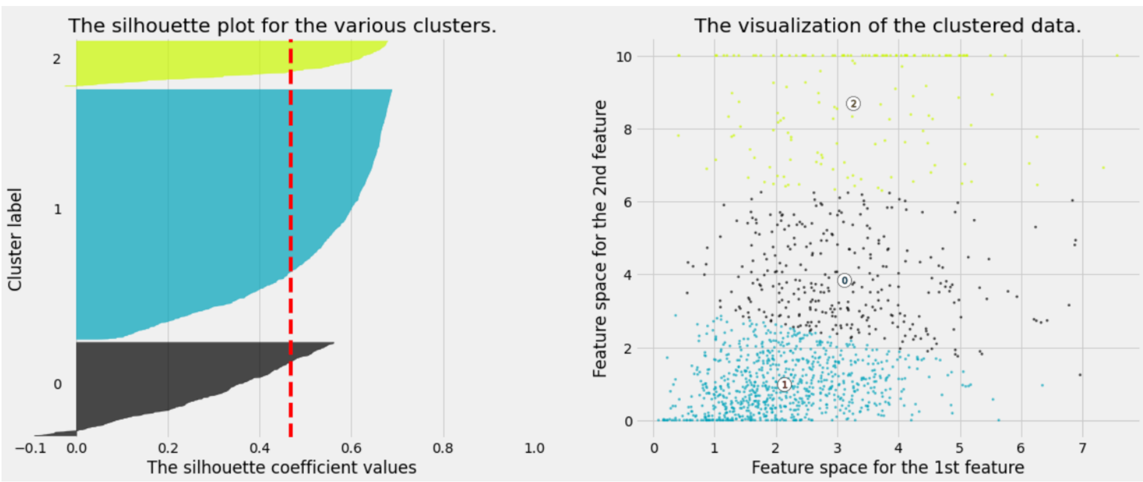
Set	Variable combination	Values for n				
		2	3	4	5	6
1	AS, SRS, NS	0.2773	0.2351	0.2387	0.2286	0.2205
2	AS, SRS	0.3836	0.3898	0.327	0.3524	0.3531
3	AS, SRS, NS, subscribers	0.5023	0.2737	0.2512	0.2226	0.2106
4	AS, SRS, subscribers	0.5533	0.3285	0.3266	0.3027	0.2912
5	AS, subscribers	0.6174	0.4215	0.4429	0.4215	0.4027
6	SRS, subscribers	0.6231	0.4693	0.3901	0.4092	0.4063
7	NS, AS	0.5423	0.3848	0.3809	0.3796	0.3705
8	NS, SRS	0.4999	0.4225	0.3789	0.3924	0.3701

Appendix 12: Silhouette analysis for k-means clustering plots for set 5 and set 6

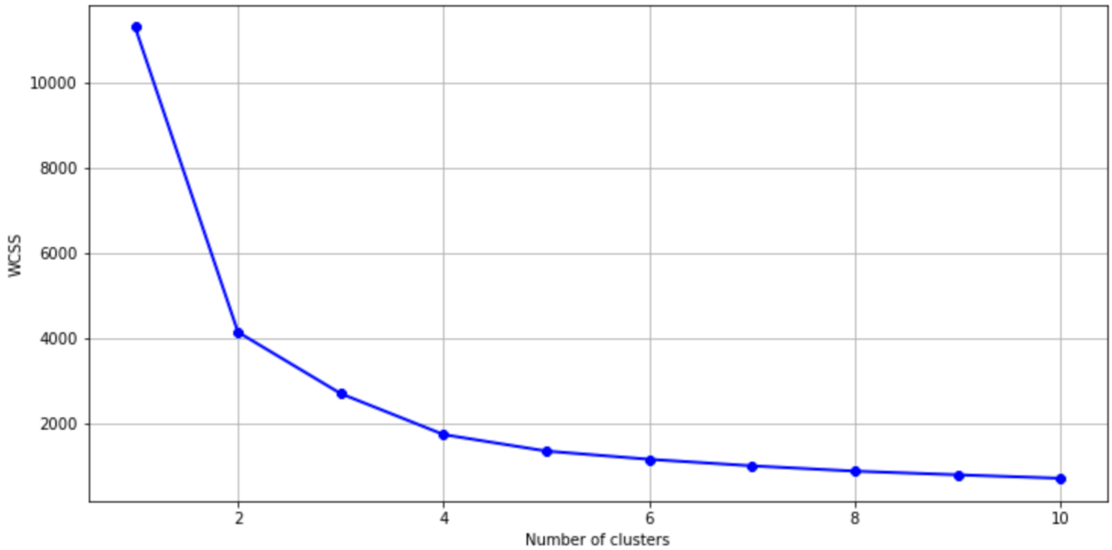
Set 5, n=4



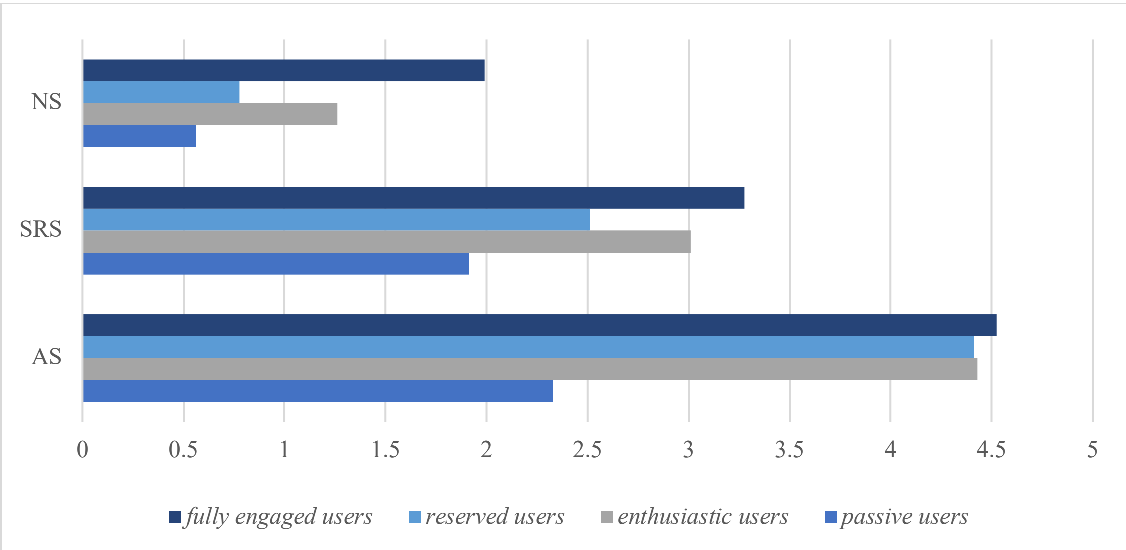
Set 6, n = 3



Appendix 13: Plot of Elbow Method



Appendix 14: Scores per segment identified through k-means clustering



Appendix 15: Demographic variables for the four LI user segments

User segment		**passive**	**enthusiastic**	**reserved**	**fully engaged**	**Chi-Squared test p-value**
Total	N	371	224	562	127	< 0.05
	% from total	29%	17%	44%	10%	
gender	female	38%	51%	61%	34%	
	male	62%	49%	39%	66%	
industry	Social media & entertainment	27%	17%	21%	21%	< 0.05
	Agriculture & farming	2%	1%	2%	1%	
	Consumer products	13%	8%	12%	8%	
	Education	3%	4%	3%	4%	
	Enterprise services	10%	25%	15%	20%	
	Financial services	4%	11%	7%	13%	
	Food & beverage	3%	4%	3%	5%	
	Health	18%	12%	17%	12%	
	IT	3%	3%	4%	4%	
	Manufacturing	2%	1%	2%	0%	
	Mobility	3%	6%	6%	5%	
	Privacy & security	0%	0%	0%	1%	
	Real estate	1%	2%	2%	4%	
	Resources	2%	2%	2%	2%	
	Science & engineering	9%	4%	4%	2%	
top 10 countries*	United Kingdom	28%	33%	28%	34%	< 0.05
	Germany	16%	12%	13%	16%	
	France	5%	9%	9%	12%	
	Sweden	3%	2%	7%	0%	
	The Netherlands	3%	4%	6%	2%	
	Switzerland	4%	6%	5%	2%	
	Spain	5%	9%	4%	10%	
	Denmark	2%	3%	4%	3%	
	Finland	4%	2%	4%	2%	
	Ireland	3%	3%	3%	3%	
top 10 cities*	London	21%	24%	20%	25%	0.08
	Berlin	5%	5%	7%	11%	
	Paris	3%	6%	6%	9%	
	Stockholm	3%	1%	6%	0%	
	Barcelona	2%	3%	2%	4%	
	Dublin	2%	3%	2%	2%	

	Amsterdam	2%	3%	2%	1%	
	Copenhagen	1%	3%	2%	2%	
	Helsinki	2%	1%	2%	2%	
	Madrid	1%	4%	1%	5%	
top 3 schools*	University of Oxford	2%	2%	2%	5%	< 0.05
	University of Cambridge	1%	6%	1%	1%	
	Imperial College London	2%	1%	2%	1%	
highestDegree	executive	3%	11%	12%	14%	< 0.05
	other	17%	14%	14%	19%	
	doctoral	12%	12%	12%	10%	
	postgraduate	39%	42%	40%	38%	
	Secondary education	5%	0%	1%	0%	
	undergraduate	23%	20%	21%	19%	
numJobs	< 10	51%	56%	52%	47%	< 0.05
	< 15	12%	22%	23%	24%	
	< 5	36%	15%	20%	14%	
	> 15	2%	7%	5%	14%	
lastFundingStage	Seed	77%	67%	69%	72%	0.21
	Series A	14%	20%	18%	17%	
	Pre-Seed	6%	9%	9%	9%	
	Angel	3%	4%	4%	2%	
numEmployees	1-10	60%	42%	59%	44%	< 0.05
	11-50	36%	50%	37%	48%	
	51-100	2%	7%	2%	7%	
	101-250	1%	1%	1%	1%	
	251-500	< 1%	0%	0%	0%	
numFounders	1	20%	22%	25%	25%	< 0.05
	2	44%	41%	39%	36%	
	3	23%	24%	23%	27%	
	4	10%	10%	9%	10%	
	5	3%	2%	2%	2%	
	6	0%	0%	0%	1%	
	7	0%	0%	< 1%	0%	
cbRank	< 150,000	13%	32%	23%	42%	< 0.05
	< 300,000	23%	28%	24%	17%	
	< 600,000	40%	28%	35%	27%	
	< 1,200,000	25%	13%	18%	15%	
	0	10%	4%	3%	3%	< 0.05

numUnique Degrees	1	34%	25%	29%	35%	
	2	40%	41%	38%	38%	
	3	15%	25%	24%	18%	
	4	2%	5%	4%	6%	
	5	0%	0%	0%	0%	
schoolYears	< 3	30%	23%	22%	31%	0.41
	< 6	34%	38%	39%	34%	
	< = 9	26%	29%	28%	24%	
	> 9	9%	10%	11%	10%	
companyAge	< 3	4%	6%	9%	7%	0.22
	< 6	49%	55%	47%	55%	
	<= 9	41%	35%	38%	32%	
	> 9	6%	4%	5%	6%	

*As the original amount of categories was too large to produce reliable results for the Chi-Squared test, only the top 10 categories were used for testing.

Appendix 16: statistics for the three scores by segment

Score	**user segment**	**N**	**M**	**SD**	**Percentiles**			**Kruskal-Wallis test p-value**
					25	**50**	**75**	
SRS	passive	371	1.91	1.07	1.14	1.73	2.47	< 0.05
	enthusiastic	224	3.01	1.25	2.1	2.83	3.69	
	reserved	562	2.51	1.06	1.73	2.39	3.16	
	fully engaged	127	3.28	1.43	2.19	3.22	4.26	
NS	passive	371	0.56	0.37	0.32	0.51	0.72	< 0.05
	enthusiastic	224	1.26	0.31	1.05	1.23	1.43	
	reserved	562	0.78	0.3	0.57	0.75	0.95	
	fully engaged	127	1.99	0.33	1.81	1.95	2.18	
AS	passive	371	2.33	0.86	1.88	2.49	2.95	< 0.05
	enthusiastic	224	4.43	0.91	4.03	4.55	5.02	
	reserved	562	4.41	0.58	4.02	4.43	4.8	
	fully engaged	127	4.52	0.98	4.07	4.59	5.08	

Appendix 17: Score results for the three segments separated by gender