UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Data-Driven Solutions in VC Investments

A cross-sectional study on best-in-class data strategies within deal origination

School of Business, Economics and Law at Gothenburg University
MSc in Knowledge-Based Entrepreneurship
Master Thesis, Graduate School, Spring 2022
Authors: Robin Gerdemann and Carlos Heredia Alcaraz
Supervisor: Marouane Bousfiha

Data-driven Solutions in VC Investments
A cross-sectional study on best-in-class data strategies within deal origination

Written by Robin Gerdemann and Carlos Heredia Alcaraz

School of Business, Economics and Law, University of Gothenburg, Vasagatan 1, P.O. Box 600,
SE 405 50, Gothenburg, Sweden
Institute of Innovation and Entrepreneurship

Contact: Gerdemann@gmx.de and carlosheredialcaraz@gmail.com

ABSTRACT

The development of new technologies and the innovative trends that have emerged over the last years are transforming and disrupting several traditional industries. In a world that is becoming increasingly connected and digitized, data-driven strategies have captured the interest of venture capital investors. Actors within the VC industry have identified the potential of data-driven solutions to improve the operational efficiency within different investment stages - particularly to strengthen and enhance the deal sourcing phase, which is relatively prone to suboptimal resource allocation.

The primary aim of this thesis was to investigate, together with Volvo Group Venture Capital, which are the best data-driven strategies currently used by top-performing VC firms to adapt to this observable transformation. In order to achieve that, the authors conducted a cross-sectional study primarily based on interviews with top-performing VC firms as well as with experienced professors within the entrepreneurial finance and venture capital ecosystem.

The findings provide relevant insights on how data-driven VCs are currently performing their deal sourcing strategies while simultaneously highlighting the role of data-driven tools to support investment firms in retrieving, organizing, and presenting the data. By studying and combining these best practices, the authors were able to provide a selection of data dimensions and databases together with relevant tools for the organization and structuring of the data to transform it into useful information for strategic financial decision-making. Furthermore, the results of the study can also be considered as support and a starting point for VC firms currently redefining their data strategies.

Keywords*: Venture Capital, Deal Sourcing, Corporate Venture Capital, Independent Venture Capital, Data, Data-Driven solutions, Database, Proprietary Algorithms, CRM, Unbiased Decision-Making, Volvo Group Venture Capital*

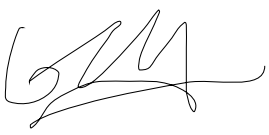
ACKNOWLEDGEMENT

We would like to express our gratitude to all those who supported us throughout this master thesis project. The cooperation of the sampled interviewees has led to deep insights into the VC industry and contributed to the overall quality of this paper.

We also want to thank Martin Witt and the entire team at Volvo Group Venture Capital for providing us with the relevant guidance in the context of this research and for letting us be part of the team over the last six months. In particular, we want to show our appreciation for Erik Johansson and Adam Albertsson for our weekly meetings. Your knowledge and expertise built the foundation of this report.

Lastly, we would like to thank Marouane Bousfiha who has been an excellent supervisor for this project and continuously supported us along this exciting journey.

Gothenburg, June, 2022



Robin Gerdemann *Carlos Heredia Alcaraz*

Table of Contents

TERMINOLOGY

BDA	Big data and related analytics
CRM	Customer Relationship Management
CVC	Corporate Venture Capital
GPs	General Partners
IPO	Initial Public Offering
IVC	Independent Venture Capital
LPs	Limited Partners
M&A	Mergers and Acquisitions
PA	Proprietary Algorithms
PE	Private Equity
VC	Venture Capital
VGVC	Volvo Group Venture Capital

1. INTRODUCTION

1.1. Background

The ongoing process of globalization promotes international competition and exposes companies to highly dynamic, interconnected, and thus complex business environments. This economic and societal development is evident in financial markets and investment projects, which are already characterized by a high degree of uncertainty and risk in their very nature. This is especially noticeable within the Venture Capital (VC) industry. The internal and external conditions of VC investments are shaped by opacity, lack of transparency and are therefore difficult to predict (Cumming, Fleming and Schwienbacher, 2005). The VC industry, however, provides several investor types with early-staged and high yield investment opportunities and thus syndicates various interest groups. In general, the business environments of potential target companies in this industry are characterized by a high degree of uncertainty, which makes the identification of profitable projects even more difficult. It has been proven that most VC investments can be considered unsuccessful as they fall short of investor-specific expectations which is also due to the optimizable risk premium ratio in the entrepreneurial sector (Vereshchagina and Hopenhayn, 2009; Bhakdi, 2013).

One explanation for this particular circumstance could be either the complete lack of relevant information or suboptimal quality of the underlying data (Povel and Sertsios, 2014). However, the availability of necessary information about the financial and operative performance of respective targets plays a critical role in VC investments as it forms the basis for an initial filtering followed by a target evaluation. Relying on the strength of the public capital market's information efficiency and one's own ability to extract the relevant information for individual target valuation can lead to significant deviations from the actual target value. In addition, VC investors cannot rely on a substantial amount of historical data due to the short operating performance history of the companies under investigation (Fama, 1970; Arroyo et al., 2019). The initial step of most investment procedures - the deal origination process - requires a significant amount of information in particular as investors try to identify the main economic drivers of a certain industry including key market participants. The compilation of a long list with potential targets, however, is a process that is relatively prone to suboptimal resource allocation. VC firms spend significant amounts of their time networking and structuring deal terms, yet an essential part of success comes from selecting the right companies right from the beginning. On average, VC firms need to screen between 80 and 100 potential targets for their respective financial and operational performance in order to close a deal (Teten and Farmer, 2010).

However, data-driven solutions could potentially help to clean up information asymmetries and search costs that are very likely to occur within the origination phase (Capron and Shen, 2007; Teten and Farmer, 2010; Gompers et al., 2020). A more data-driven approach can support the deal sourcing phase by gathering more relevant and high-quality information

already from the origin of the project - especially about the target companies (Harvard Business Review, 2021). An improved compilation of the underlying long-list taking into consideration the respective investor-specific criteria could result in a higher probability of success of the project at hand.

1.2. Research Question

The present research addresses the question of whether an increasing and widespread application of data-driven solutions within the VC industry can improve the deal origination process in VC investments. Additionally, the authors of this paper explored how VC firms store and transform the data collected into helpful and supportive information for potential investment projects. It is also of great importance to investigate how well VC funds can qualify target companies based on data-driven approaches. The focus in this thesis is primarily on identifying, analyzing and comparing pre-existing patterns within the VC ecosystem to provide both economic potential and learning opportunities. The underlying research question therefore is *how can the deal making process within venture capital investments be improved by identifying best practices in applying data-driven solutions?*

The research is conducted in collaboration with Volvo Group Venture Capital (VGVC) - the venture arm of the global commercial vehicle company Volvo Group. The Group is a multinational company in the mobility and infrastructure sector and was founded in 1927 in Sweden. The company manufactures trucks, buses, boat engines and construction equipment, and also offers service solutions related to these products. The company currently has over 100 000 employees and is operating in 190 countries all over the world. The function of VGVC is to identify young and innovative companies in order to develop and support new business areas and opportunities for the Volvo Group as well as increase profitability for existing customers. The aim is to drive business growth by establishing strategic partnerships with innovative companies on a global scale primarily within logistic services, electric infrastructure and site solutions. Founded in 1997, the company has already participated in 80 investments in Europe, Israel and the United States. Recent investments such as the shareholding in the Israeli Upstream Security Ltd. underline the strategic focus on technological and sustainable solutions and transport systems of the Swedish manufacturer (Volvo Group, 2021). The vision of VGVC is to be a proactive driver of a jointly brought about transformation of transportation and infrastructure by empowering innovative companies (Volvo Group, 2021).

1.3. Research Scope and Delimitations

Having a sector-focused perspective on potential data-driven solutions during the study, incorporating the respective macro- and microeconomic variables that are characterized by a high degree of diversity, would be a major undertaking. Therefore, the authors of this paper have attempted to limit the scope of the study to the data analysis tools and techniques currently in use by VC players who either market their investment activities with a high degree of data integration and automation or have a strong reputation in the marketplace. Regardless of the industry, geography, or investment stage focus of investors, this paper seeks to review

current data-driven solutions across the global VC ecosystem and ultimately aims to provide readers with valuable findings in the form of recommendations for potential adoption of the tools.

Given the time constraints of this project, the authors of this paper conducted a sample of 20 interviews with VC investors within and outside the VGVC network. The interviews were complemented with knowledge and input from VGVC in the context of weekly discussions. However, the findings and recommendations presented at the end of this paper are based only on the primary data collected in the interviews which is then supplemented by findings from previous academic contributions. In addition, the authors of the paper focused on the investor perspective of a corporate investment arm that has resources similar to VGVC. Therefore, the scope and objective of the study were primarily tailored towards the strategic needs and vision of the VGVC, making the empirical findings and recommendations generalizable only to similarly structured and operating CVC investors. Furthermore, alternatives for startups to implement more data-driven elements to fundraising are not considered in this paper, although a collaborative approach between ventures, investors and third-party data providers is critical to data mining.

1.4. Disposition of the Paper

The present paper includes seven chapters and is structured as follows. The first chapter is the introduction and aims to provide a motivation for the investigation of the topic at hand, the objectives of the study, the research question, and potential delimitations. The second chapter presents the literature review and builds the theoretical framework as a foundation for the overall research. The literature review highlights relevant concepts observable in the VC industry and within the investment processes of VC firms. Simultaneously, the focus here is on the deal sourcing process and the existing literature on data-driven solutions within the VC ecosystem is introduced. The subsequent third chapter presents the research methods chosen by the authors to conduct the thesis. That includes a description of the research strategy, research design, data collection process, data analysis, and relevant aspects of the research quality. The fourth chapter outlines the research findings identified through the primary data, covering the main themes and illustrating the highlights of the data collected throughout the interviews. Based on the empirical findings, the following analysis focuses on interpreting and integrating the main findings of the study while trying to establish connections with the previously existing literature and the current conditions of the VC ecosystem. The sixth chapter discusses and provides insights on how the development of data-driven solutions in deal origination practices can potentially transform the VC industry in the coming years. Lastly, the seventh chapter concludes the paper by answering the research question while summarizing the most important findings of the research and highlighting how this research contributes to the existing literature.

2. LITERATURE REVIEW

This chapter explains the emergence and development of Venture Capital to be used as a financing instrument for startups. This is followed by a presentation of financing processes with the help of venture capital, including its premises and individual stages with a particular focus on the deal origination phase. Also, the role that venture capitalists themselves take within the whole economic system - especially for young companies - will be considered and elaborated on. However, the investor's perspective on VC investment will be the particular focal issue in the context of this paper.

2.1. The Venture Capital Ecosystem

2.1.1. What is Venture Capital

Venture Capital (VC) as it is understood today, was first originated in 1946 when the first VC firm was established in the United States (Gompers, 1994). That institution was the American Research and Development, and had the aim to finance innovations and technologies developed in the U.S. after World War II. Since then, VC investments have been a relevant funding source for entrepreneurial firms, particularly for innovative fast-growing startups. VC is categorized as a subtype of Private Equity (PE) investment, and consequently, VC firms take an equity stake in the firms they invest in (Metrick and Yasuda, 2011).

Given the characteristics of entrepreneurial firms and the uncertain business environment they are confronted with within the early stages, venture capitalists must assume a high degree of risk and consequently expect high economic rewards from their investments (Gompers, 1994). VC investors expect around ten times the return on investment over five years, however, the return on activity and risk premium ratio in the entrepreneurial sector is relatively low given the high level of uncertainty compared to more mature industries (Zider, 1998; Vereshchagina and Hopenhayn, 2009). In order to achieve high returns and compensate for the risks involved, VCs try to identify and invest in those startups that have the strongest growth potential and become successful in a relatively short amount of time (Metrick and Yasuda, 2011). Inevitably, a significant number of VC-backed startups are technological-based young entrepreneurial firms that can penetrate large and attractive markets, and ultimately scale fast (Metrick and Yasuda, 2011).

VC firms also aim to reduce the involved transactional risks by co-investing with other VC firms. Lintner (1972) emphasized that risk-averse behavior among investors can be increasingly observed in the market, which points to the high relevance of diversification. The advantage of diversification became known through Markowitz's portfolio theory and can also be applied to syndicated VC-investments (Markowitz, 1952; Lintner, 1972). According to Zider (1998), venture firms therefore prefer to have two or three groups involved in most stages of financing series. Typically, there will be a leading investor and several so-called followers.

2.1.2. The role of Venture Capital in society

To better understand VC and its role in society, it is worth bearing in mind the different funding sources for entrepreneurs as well as acknowledging the challenges that small and young firms encounter when obtaining financial resources. A general classification to categorize these funding sources is the distinction between debt capital and equity capital (Isaksson, 2006). Through the provision of traditional loans, banks have played a crucial role in financing new ventures with debt capital over the last years. Nonetheless, there is still a significant number of small and new firms that cannot obtain debt capital from banks as they lack certain tangible assets that can be used as collateral or guarantees for banks (Gompers, 1994). Furthermore, most entrepreneurial ventures lack a track record and positive financial performance results during the first years of development which hinders their access to debt capital (Isaksson, 2006). Also the credit crunch in the late 2010's and the recession caused by Covid-19 impact with its consequences being perceptible on a global scale led to higher levels of risk-aversion of banks when financing young ventures (Kleijn, 2011; Doleschel & Manu, 2021). It is under these circumstances that venture capitalists play a crucial role in financing entrepreneurial firms to support their development in the early stages as they enable startups to continue their operations and thereby ultimately drive innovation by stimulating high-technology development (Florida and Smith, 1990).

The creation, growth and promotion of new and small firms is also a relevant element to enhance economic growth (Davidsson, Achtenhagen and Naldi, 2010). As demonstrated by Kortum and Lerner (1998), VC activity positively influences and increases the level of innovation. Based on their empirical findings, VC investments account for a significant percentage of the industrial innovations developed in the U.S. As stated in the economic literature, the high levels of VC activity in the U.S. are one of the key underlying reasons explaining the North American leadership role in developing and commercializing technological innovations globally. Elan and Goodrich (2010) further explain this high VC investment dynamic with the low risk-aversion among U.S.-American investors. When looking at the VC industry in Europe, the VC activity and its relevant impact on the European economic development has significantly increased over the last years (Bottazzi and Da Rin, 2002). VC investments have played a crucial role in the economic growth of European countries through the commercialization of innovations (Faria and Barbosa, 2014). In this vein, European governments have identified the VC industry as a relevant element to attain job creation, innovation developments, and economic growth (Bottazzi and Da Rin, 2002).

When studying and analyzing the VC industry, it is also relevant to highlight the value-added that VC firms provide to the companies in which they invest. Clearly, the capital obtained by the portfolio companies is one of the most relevant contributions to the target company (Metrick and Yasuda, 2011). However, it is not the only value that VCs deliver. According to the definition of Gompers (1994), VC funds are active investors due to their strategic involvement and positioning within the startup company, which differentiates them from other investor types. For instance, VCs closely monitor the progress of their portfolio companies,

provide industry and managerial advice to founders, provide access to professional networks, and, in some occasions, can also sit on the board of directors (Gompers, 1994).

2.2. Types of Venture Capital

When studying the different types of VCs, one of the most commonly used categorizations differentiates between Independent Venture Capital (IVC) and Corporate Venture Capital (CVC). One of the essential differences between these two types of VC investors are their underlying investment objectives. Hellmann (2002) suggests that IVCs tend to focus more on financial gains as their ultimate investment goal, while CVCs also consider the strategic value of the financial participation within the target company. Nonetheless, it is worth noting that this classification is rather broad and in practice, both types of VC investors can pursue these two investment objectives to a similar extent. The investment objectives, however, are not the only element differentiating between IVCs and CVCs. For instance, the origin of the funds is another relevant difference between these two types of VC investors. The following subsections present a detailed description of the two different types of VCs, highlighting their particular characteristics and differences.

2.2.1. Independent Venture Capital

Independent Venture Capital (IVC) is the predominant type of VC investors (Bertoni, Colombo and Grilli, 2013). IVCs are financial intermediaries by nature, as they raise funds from investors and make equity investments in startup companies (Metrick and Yasuda, 2011). The typical investors in IVC funds are institutional investors such as university endowments, public pension funds, corporate pension funds, or insurance companies (Zider, 1998). These large institutions allocate fairly small percentages of their total investments into high-risk investments through VC funds. From a legal and economic perspective, IVC firms establish Limited Partnerships agreements with investors, in this case also known as Limited Partners (LPs) while venture capitalists assume the role of General Partners (GPs) (NVCA, n.d). Therefore, based on this legal and economic relationship, the VC firm acts as the fund manager, while the LPs share the ownership of the fund with the VC firm. Figure 1 illustrates the typical structure of an IVC fund, showing the relationship between GPs, LPs, and portfolio companies. As previously mentioned, an essential characteristic of IVCs is that their objective is to maximize financial gains, and thus, maximize the return to the investors (Metrick and Yasuda, 2011). In order to obtain the highest financial returns from their investments, IVCs aim to exit the portfolio companies through an Initial Public Offering (IPO) or via a merger or takeover transaction (M&A).

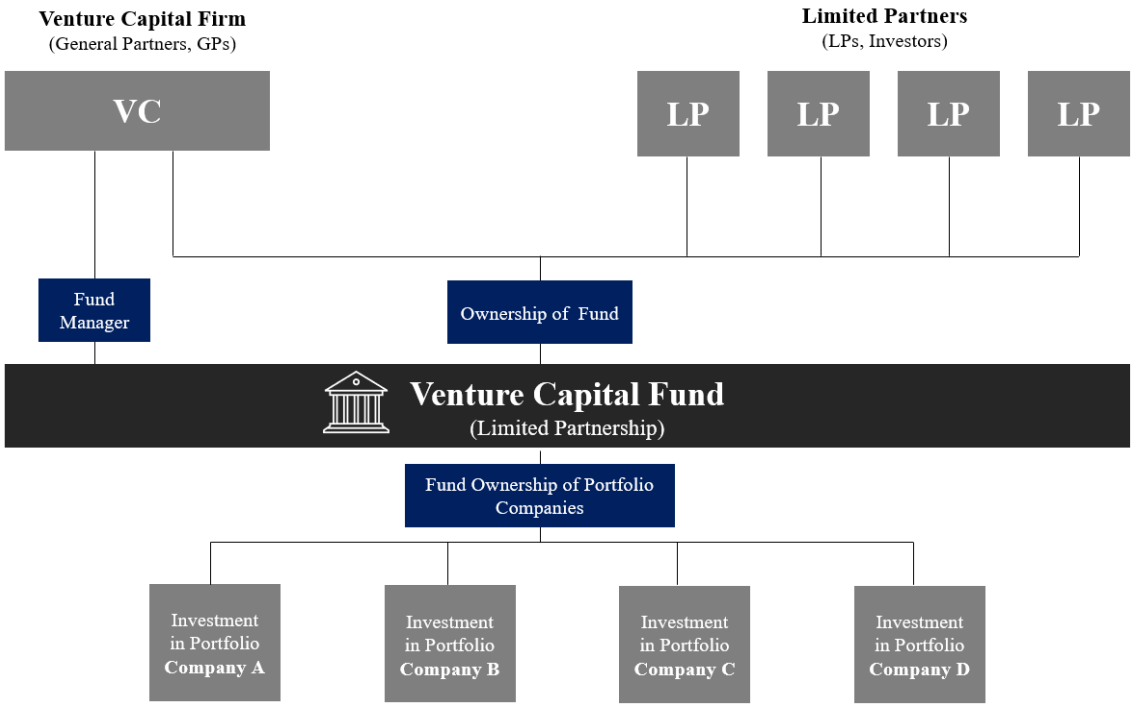


Figure 1. Structure of a VC fund. Adapted from NVCA (n.d.)

2.2.2. Corporate Venture Capital

Corporate Venture Capital (CVC), also known as corporate venturing, describes a variant of venture funding that is of particular importance in the framework of this research. CVCs, like traditional VC investments, serve to support young, innovative, and legally independent companies that need financing to carry out the business activities they aspire to. In this context, however, the funding sources are limited to incumbent firms that possess sufficient capital resources. Similar to individual angel investors and conventional VC funds, these corporate divisions are willing to invest in startups in all growth phases, but it is very likely that their investment strategy defines a specific stage that is of particular interest to them. The tactical orientation of the fund can also specify the geographical focus for investments. Most VC investors focus their investment activity primarily on entrepreneurial hubs of their operational interest because they are very likely to enable tacit and explicit knowledge transmission, promote networking, and have greater appeal to skilled employees (Galope, 2014). This investment strategy, however, is questioned within the economic literature. Teten and Farmer (2010) were able to find evidence for strong investment performance outside of entrepreneurial hubs.

CVCs also differ from other institutional forms in their organizational structure as well as their legal and economic affiliation. Figure 2 illustrates the typical structure of a CVC, showing the relationship between the parent company, the CVC fund, and portfolio companies. With few exceptions, the corporate venturing departments are subsidiaries of their respective parent companies, which makes them distinctly visionary in line with the group's strategic orientation. Unlike IVC firms, CVCs therefore do not have full authority of the allocated capital to them and need the approval of the group’s investment committee (Rajan,

Servaes and Zingales, 2000; Chemmanur, Loutskina and Tian, 2014; Titus and Anderson, 2016).

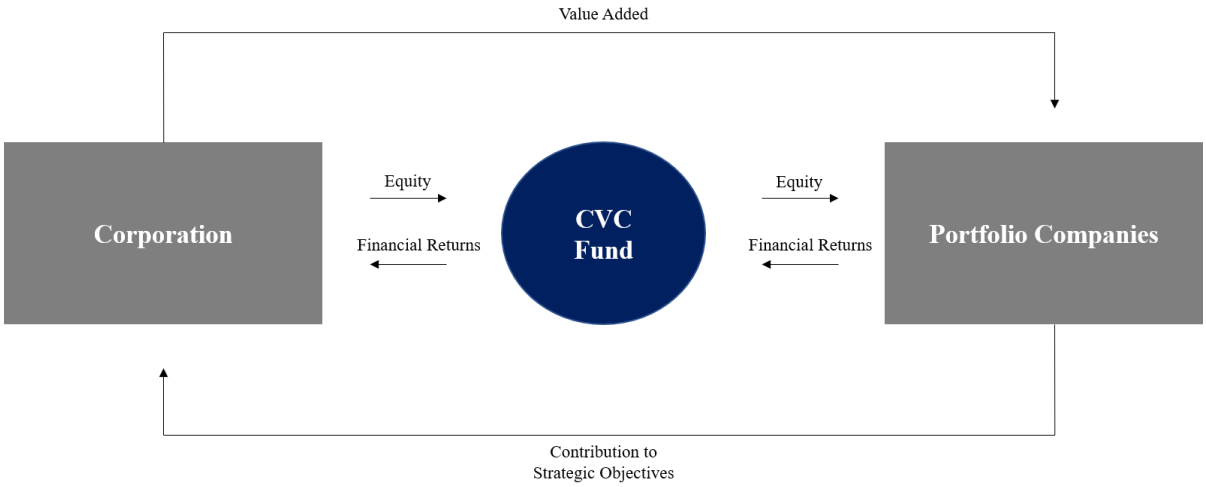


Figure 2. Structure of a CVC fund. Adapted from Ernst, Witt and Brachtendorf (2005)

The distinction from traditional and independent venture capital investment behavior is then made primarily based on goals pursued as well as the extent of management support as a result of a shareholding within the company. CVCs are very likely to have a two-bottom-line approach that incorporates strategic expectations towards the investment in addition to the desired financial returns for profitability (Van de Vrande, Lemmens and Vanhaverbeke, 2006; Benson and Ziedonis, 2009). While the strategic element is aiming at accessing new technologies developed by startups, incumbents primarily want to generate synergies and intent to create both value and sector insights for the corporate - which is also referred to as so-called strategic investing (Woolridge and Snow, 1990). On the one hand, the anticipated organic growth of the target company already provides CVC investors with high-yield opportunities (Engel, 2002). With the help of post-investment participation in innovative companies, however, additional synergy-driven increases in enterprise value can be achieved through joint R&D activities, market expansion, and the forming of long-term alliances - which are all of high strategic value (Metrick and Yasuda, 2011; Napp and Minshall, 2011; Dushnitsky and Lenox, 2006).

Therefore, a strategic investor such as a CVC division evaluates investment projects differently from traditional VC funds. While independent VCs have to employ mechanisms that ensure financial returns to their limited partners, corporate divisions can not only step out of the role as a public financial intermediary and apply longer investment horizons but also expand their specific investment requirements from being solely return-oriented (Metrick and Yasuda, 2011). The lack of a purely performance-based system then allows CVCs to integrate a more sensitive approach towards the respective portfolio companies, which ultimately has an effect on their positioning within the VC ecosystem. The market signals of their investments are less aggressive and - due to the extensive industry knowledge the firm possesses compared to individual VCs - indicate a high potential of the target company

(Chemmanur, Loutskina and Tian, 2014; Zhu, 2009). Ravid and Spiegel (1999) point out that the acquisition of a stake in a company can increase the perceived enterprise value of the target company through the signaling effect, as the market expects the realization of synergy-related value increases after an economic participation.

The strategic importance of corporate venturing as a profitable investment instrument is becoming more and more apparent. Global CVC-backed investments have accelerated recently, reaching a new high in the first half of 2021 alone. The industry recorded an all-time high with 2.099 disclosed deals and an accumulated investment volume of 69 billion EUR dollars (CB Insights, 2021). Recent transactions that gained public attention include Censy's 30.8 million EUR Serie B investment round which was led by Intel Capital - an established subsidiary of Intel Corporation to manage corporate venture capital and global investments. The financial participation of previous investors in Censy such as Google Ventures and Greylock Partners demonstrates not only CVCs' competitiveness but also the incumbents' awareness of the attractiveness of corporate venturing when they want to operate on the frontier of emerging technologies and markets (Intel Capital Stats, 2022; Intel Capital, 2022).

2.3. The Venture Capital Investment Process

The overall VC transaction is a detailed process in which many stakeholders have to be taken into account. Thus, it is difficult to isolate a salient motive for single transactions as many perspectives are simultaneously involved in the investment decision process. In addition to that, the motives are highly investor-specific and differ within the network of capital providers. Unlike the underlying motives of such investments among the different types of VC, however, the process and sequence of VC investments are structured and organized in the same way most of the time (Bender, 2011). Tyebjee and Bruno (1984) suggested a descriptive investment process model that is generally accepted and in wide application within the VC industry until today. The authors identified and conceptualized a five-step sequence that attempts to model critical investment activities and their respective milestones. The model differentiates between deal origination, deal screening, deal evaluation, deal structuring, and lastly post-investment activities.

Deal origination marks the first activity of a VC investment and describes the phase in which venture capitalists seek to identify potential investment opportunities in their area of activity or interest. The goal of this sourcing procedure is to compile a long-list of potential targets. This step is followed by an initial ***screening*** where VC funds apply certain criteria that match with the characteristics of the target companies that made it on the long-list. Without an extensive review, the aim is to reject unsuitable companies in order to allocate resources more efficiently (Fried and Hisrich, 1994). The following step, ***target evaluation***, tries to assess the startup on the basis of a predetermined multidimensional set of criteria having in mind the risk premium ratio within the entrepreneurial sector. The venture capital investment agreement is developed in the ***deal structuring*** phase. Here, among other terms, the investment volume and the corresponding equity share for the investors are defined and the deal is ready to be closed. Finally, ***post-investment activities*** include the shift of the investor's role to a cooperator of the

portfolio company, who very likely has formal representation on the board. This will facilitate monitoring activities and also allows for the application of cash-out mechanisms such as IPO or M&A transactions (Tyebjee and Bruno, 1984; Fried and Hisrich, 1994; Bender, 2011).

Fried and Hisrich (1994), however, criticized that the initially developed model did not highlight the significant necessity of information-gathering activities within the overall VC investment. Sandberg, Schweiger and Hofer (1989) supported this argument by emphasizing that the analysis of purely financial performance of a target company is not sufficient for a holistic decision-making process as perceptual, emotional, and cognitive elements possessed by the founding team of the target company must also be taken into account.[1] The researchers thus contest the traditional theory of neoclassical economics, which assumes complete investor rationality with primary focus on financial cash flow (Weintraub, 2002). Built upon the finding of Tyebjee and Bruno (1984), therefore, Fried and Hisrich (1984) split the screening phase into VC firm-specific screening on the one hand, and a generic screening process on the other. While the firm-specific screening phase includes an examination of investor-specific requirements regarding the target characteristics, the latter focuses on an analysis of the business model coupled with the pre-existing information the investor might have about the startup.

This paper tries to complement the strength of both models and is based on the following sequence depicted in Figure 3: Deal origination, deal screening, deal due diligence, deal structuring, investment development and investment exit.

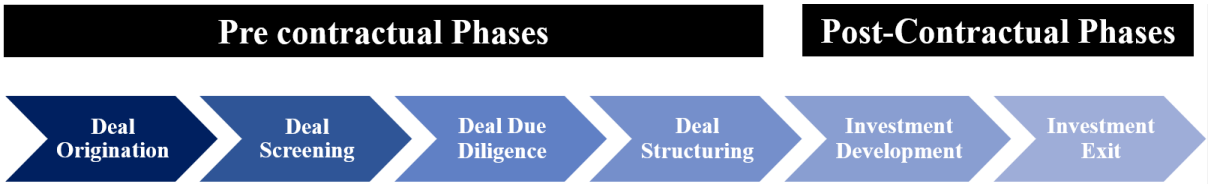


Figure 3. Investment decision process in VC investments. Adapted from Bender (2011)

The following subsection provides a detailed explanation of the deal origination process, as it is the most relevant step of the different phases previously highlighted for the scope of the present paper.

2.3.1. The Deal Origination Process

The deal origination process, as shown above, represents the first phase of the overall VC investment, and thereby, builds the foundation for all further activities. The strategic importance of deal origination for the overall deal-flow is not only a matter venture funds are aware of, but is also highlighted within the VC-related literature. Teten and Farmer (2010) were able to show with their empirical findings that funds can significantly influence and

[1] This is also shown by Fowle (2019) and Ullah (2015). According to them, characteristics of the founding team can play a critical role in the sourcing stage of an investment. In this context, Agarwal et al. (2022) mention the importance of standardized assessment of elements that are more qualitative in nature to enable equitable allocation of resources.

improve their investment performance if they identify and apply industry best practices as part of the mapping process. Due to the fast-changing environment of the VC ecosystem, however, best practices are changing rapidly and market participants need to remain agile in order to identify and adapt to them (Oxford University, 2022).

The sourcing stage is primarily focusing on identifying potential investment opportunities and aims at ensuring a desired and viable deal-flow for the investor. It is crucial for VC funds to have access to an adequate number of deals that can add both strategic and economic value. The sufficient number of investment opportunities then allows the investor to be more selective within the screening and due diligence phases, which ultimately affects the average quality of the deals (Silva, 2004; Bender, 2011). The literature distinguishes between two different types of origination strategies that can be employed by the investor, namely *direct* and *indirect* (Dealroom, 2022a). The differentiation, however, does not indicate that either one or another approach has to be employed. The strategies are by no means mutually exclusive and can complement each other.

The direct approach includes, among others, the predominant technique of cold contacting, where the venture capitalist and the entrepreneurial team that is subject of the potential transaction do not know each other beforehand. The visibility and public presentation of the VC are crucial in this step, as the target will most likely collect and verify information about the VC on its own once the contact is initiated. Cold contacting also includes active search activities with the help of technology that supports an investigation of the business environment. However, also previous contact with the respective target in the form of e.g. previous investments could facilitate the contact initiation process and represent another type of a direct approach (Bender, 2011).

Indirect deal origination, on the other hand, includes different types of third parties into the mapping process that initiate the communication between potential investors and targets. Syndicated VC-Investments - the co-investing of two or more investor parties - are often a form of indirect deal origination. In this case, it is very likely that at least one investment participant has a previous syndication relationship with the target company (Sharifzadeh and Walz, 2012). Another source of deal flow can also be entrepreneurial teams that are listed in the VC's investment portfolio whose already existing network can be exploited. In addition to that, financial intermediaries such as banks, consultancies, or public institutions can establish the initial contact between investors and target firms (Bender, 2011). Figure 4 illustrates and summarizes the different forms of contact initiation.

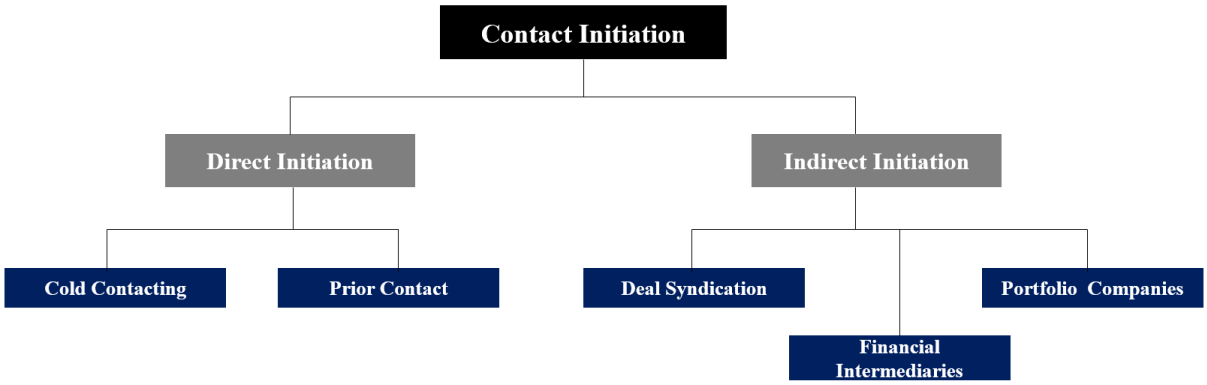


Figure 4. Forms of Contact Initiation within the deal origination process. Adapted from Bender (2011)

2.4. Data-Driven Solutions in the VC Industry

Given the fast pace of technological development, multiple industries adopt innovative practices supported by new technologies, thus, facing significant structural transformations. One of these technological trends is the utilization of data through the adoption of data-driven solutions. These practices have become increasingly important as more data has been generated and made accessible to companies - especially to and by VC firms (Weibl and Hess, 2019). The understanding of data-driven solutions in this paper, however, is rather broad and includes heterogeneous data tools. The authors analyze the added value of the efficient usage of publicly available search engines and databases up to complexly designed AI algorithms for the identification process of industry best practices.

The importance of relevant data for the successful closing of financial transactions has already been highlighted within the economic literature (Ravid and Spiegel, 1999; Povel and Sertsios, 2014). However, the ongoing process of globalization has caused a sophistication of the VC market structures which ultimately resulted in a significant increase of available market data. Market participants are confronted with a rapidly changing technological development in which it is becoming increasingly difficult to track the underlying dynamics, capture the volume of available data that is constantly generated and process it in a way that provides useful information for strategic decision-making (Koratana, McLetchie and Van Ingelgem, 2019). Minor information asymmetries can already result in either significant evaluation errors or decisive competitive advantages (Cerra, Easterwood and Power, 2012). Other industries, such as the M&A sector, have already tried to enhance the potential of data-driven approaches and aim at excluding subjective elements in decision-making as much as possible by using and primarily relying on big-data solutions and supportive data-analytic tools. However, conditions in the M&A market are somewhat facilitated for the application of data-driven strategies, because, unlike in the VC ecosystem, the relevant investors are more likely to interact with companies that do not have limited financial information (Koratana, McLetchie and Van Ingelgem, 2019; Fairview Capital, 2018). Data-analytic tools such as Deloitte's iDeal, for example, are supposed to support every investment phase, however are

tailored towards target companies with a certain performance history (Fanning and Drogt, 2014; Deloitte 2022).

While the value of data-driven solutions in financial transactions is evident, actors in the VC ecosystem still face challenges in identifying, implementing and deploying them adequately as their potential remains largely untapped (Loeb, 2012). As mentioned above, the analysis of solely financials of a target is not sufficient for a comprehensive investment decision as perceptual, emotional, and cognitive elements possessed by the target management team have to be considered as well (Sandberg, Schweiger and Hofer, 1989). Big data and related analytics (BDA) aim at supporting the identification process of unknown patterns and thereby enhance the information quality about the target. BDA applications, however, are yet not able to incorporate these elements extensively into the decision making process (Zetsche et al., 2017).

2.4.1. Data-driven solutions to improve the deal origination process

As previously highlighted, the deal origination process is one of the most relevant stages within the VC investment process having a significant influence on the final success of the investments made by VC funds (Teten and Farmer, 2010). Up to date, this first investment stage primarily relies on the interaction of VC investors with their environment in the form of human relationships - i.e. leveraging their existing network and connections (Tyebjee and Bruno, 1984). Relying on the venture capitalists' network has been an efficient strategy to map and screen promising startups which are also in line with the investment strategy of the VC firm (Weibl and Hess, 2019). The predominant way to source for deals within the VC industry, as of today, is therefore to combine the analysis of the few target company information which is stored in public databases and other external web-based services together with a personal meeting of the founding team. Popular databases that are in widespread usage within the entrepreneurial sector are for example CB Insights, Pitchbook and Dealroom (Splenda and Barnhart, 2017; Retterath and Braun, 2020; CB Insights, 2022; Pitchbook, 2022; Dealroom, 2022b). This data can be also complemented with information found on various professional and social networks such as LinkedIn and Angellist (Weibl and Hess, 2019). Also events such as conferences can connect companies in need of funding with potential investors. GU Ventures' "Connect2Capital", for example, brings together international investors with innovative entrepreneurs, however does not necessarily represent a purely data-driven approach. Another relevant European event is the 4YFN which is a startup event bringing together VC investors and entrepreneurs.

In the case of a CVC, sometimes additional internal sources from previous analyst work of the respective business units are additionally being exploited in order to provide the investment team with deal leads (Rauser, 2003; Retterath and Braun, 2020). In order to stand out from competition, however, it is required to identify and deploy alternatives to the usage of public databases and internal deal leads that create additional value within the deal origination. To allow for a structured overview of market developments including key potential targets, certain VC funds have developed algorithms that allow them to extract actual

company information. Proprietary algorithms (PA), for instance, provide an automated means of improving target selection efficiency. These algorithms represent a form of machine learning tailored towards the strategic needs of a VC fund operations to enrich respective ad-hoc analysis (Holland, 1992). According to Kurlandski and Bloodgood (2022), the value of these algorithms becomes particularly apparent when investors are interested in specific target company information which builds the foundation for further filtering and is highly needed in order to meet respective investor-specific criteria for successful investments.[2]

One component that has to be considered as well is the way how respective information is organized and processed by the investment team within VC firms. Customer relationship management systems (CRMs) can support the structuring as well as the evaluation of data and in this regard, and therefore, CRMs can also be classified as data-driven solutions within the deal origination process. Widely used market products are e.g. Jira, Pipedrive and Trello (Atlassian, 2022; Trello, 2022; Pipedrive, 2022). Tanner et al. (2005), however, emphasize the importance of using an appropriate CRM system in order to ultimately benefit from it. Reinartz, Krafft and Hoyer (2004) also argue that the simple implementation of such tools is not sufficient and the VC investment department should carefully evaluate the strategic contribution of CRM processes to the overall success of the transaction. According to the authors, the wrong instrument can have a significant negative impact on investment performance.

Nonetheless, given the development and maturation process that startups and VC ecosystems have experienced over the last years, more potential opportunities are available for global VC firms (Zhong et al., 2018). In this vein, the CVC industry is not an exception showing a significant increase in investment activity over the last decades (CB Insights, 2021). Consequently, CVC firms encounter more challenges to identify and map the most promising investment opportunities in terms of strategic value and financial importance (Xu, Chen and Zhao, 2017). For instance, the number of German startups in the logistics sector has substantially grown, making it increasingly difficult to keep an accurate overview of the industry's microeconomic drivers (Zielske and Held, 2020). This economic development illustrates the importance of having both a holistic and actual understanding of macroeconomic dynamics as they can significantly change within a short amount of time (Osterwalder and Pigneur, 2010; Ernst, Witt and Brachtendorf, 2005). Market intelligence tools, such as Mattermark and Statista, can help to provide investment units with a comprehensive overview of their respective sectors including the business models of relevant target companies (Blank and Dorf, 2012). Guarda, Augusto and Lopes (2019) identify a competitive advantage in the application of market intelligence tools - especially in fast changing business environments.

In front of this dynamic and competitive landscape, data-driven tools are presented as a potential solution to support VC firms in identifying the best investment opportunities (e.g.,

[2] Business Sale (2022) is further supporting the value of proprietary algorithms for investment purposes.

Weibl and Hess, 2019; Jain, 2018; Xu et al., 2017; Kirchhoff, Schiereck, and Mentz, 2006; Bhabra and Hossain, 2018). The potential of data-driven approaches to enhance the deal origination process in the VC ecosystem is still a relatively novel and underexplored topic and only little research has been done to date. Nevertheless, within this research scope, Weibl and Hess (2019) conducted interviews with several German VC firms and found that the application of data-driven solutions has a significant impact on the overall investment process of VCs, and particularly, on the average quality of deal origination activities and general screening. As suggested by Weibl and Hess (2019), the utilization of data-based strategies leads VC investors to improve the identification of successful investment opportunities through the automation of the deal origination process. From an economical perspective, that alone can significantly reduce the operational costs of the VC fund. Consequently, data-driven tools can not only increase the profit margin of VCs but also improve the identification of investment opportunities in a relatively short amount of time and ultimately improve the success rates of VC investments (Weibl and Hess, 2019). The most relevant classes of data-driven solutions can be found in Table 1 below.

It is worth noting that not all the existing data-driven practices are optimal as stand-alone solutions. In this vein, Retterath and Braun (2020) investigated the quality of the data provided in some of the most widely spread databases among VC investors. The authors evaluated the quality and coverage of the information provided by these third-party data providers by taking three dimensions as reference: general company data, founding team, and funding information. Overall, the size of the funding rounds determines how much data is available about that specific company, and therefore, larger funding rounds will have more probability to be reported. Given the current scenario, the authors suggested the usage and combination of different databases to achieve the most optimal coverage of relevant data. Based on their investigation, the ideal combination of databases to accurately cover data about the company, the founding team, and funding information would include VentureSource, Crunchbase, and Pitchbook (Retterath and Braun, 2020).

	Data Sources	Description	Data extracted	Main Investment Stage	Examples
Internal Systems	**Leads from internal Business Units**	Deal leads and industry/company information provided by other departments.	Current industry trends, strong performing competitors and potential targets.	Deal Origination, Screening	Internal Reports and general communication
	Proprietary Algorithms	Internally developed algorithms in order to extract relevant and investor specific data for filtering process	Includes data such as firm's momentum, financial performance, market size and competitors	Deal Origination, Screening	Balderton Capital, NGP, Connetic Ventures
	CRM Systems	Data from internal enterprise systems will be processed with standard workflow tools.	Data system to support VCs in storing and working with data collected from other sources.	Screening	Jira, Pipedrive, Trello, Affinity
External Web-based Services	**Social Media Platforms**	Gather public information about individuals as well as general information about the company.	Professional/ Academic Information of Funding team, daily target information.	Screening	LinkedIn, Xing, Angellist
	Market Intelligence Platforms	Run an analysis of respective sectors of macroeconomic nature.	Key macroeconomic drivers.	Deal Origination, Screening	Mattermark, Statista, Delphai
	Databases	Company Investigation with a microeconomic focus.	Key microeconomic drivers including their financial and operational information.	Deal Origination	CB Insights, Pitchbook, Crunchbase, Dealroom

Table 1. Overview of different data sources used by VCs and their respective descriptions.Adapted from Weibl and Hess (2019)

3. METHODOLOGY

This chapter provides an overview and explanation of the methodology implemented by the authors to conduct the study and ultimately find an answer to the research question. Therefore, the research strategy and design are explained as well as the motivation and justifications for choosing these approaches.

3.1. Research Strategy

Since the deal origination process and its integration of data-driven solutions are not widely researched in the VC industry - yet play a crucial role in the successful structuring of the overall VC investment - an inductive research strategy is considered the adequate approach to study the topic. As proposed by Bell, Harley and Bryman (2019), inductive research builds on empirical observations to propose new explanations and theories, and in this case, to ultimately provide an unbiased answer to the research question using primary data. Therefore, based on this relationship between theory and observations made by the executors of the study, a qualitative research approach is commonly suggested to be applied as the overarching research strategy (Bell, Harley and Bryman, 2019). Considering the exploratory nature of the research question, a thematic analysis resulting out of the qualitative approach did not only allow for an appropriate discussion and interpretation of the underlying data but also for categorizing and classifying the findings in a reasonable way. This supports the study as its focus is on the perspectives, attitudes, and opinions of the respective interviewees to serve as empirical observations. For this research, the authors conducted 20 interviews with experienced and knowledgeable individuals from the VC industry to serve as the target sample for primary data collection. The sample is primarily composed by VC actors with practical experience, but also includes researchers who approach the area under investigation from a more theoretical and academic perspective. In order to be defined as knowledgeable, the individuals need to have either a proven track record of previous VC investments, several years of experience within the VC industry, cross-sectoral expertise, or experience with the application of data-based operations within the deal-origination process - preferably at the international level. In addition, VGVC has provided valuable industry insights to the executors of the study in weekly meetings throughout the project, as the investment company and its investment managers combine all of the above criteria.

In order to enhance the analytical power of the study of the underlying data, the authors also included a ranking and mapping of the interviewees based on the primary data being collected, which will serve as a complement to the research strategy described above and help to develop a deeper understanding of the sample at hand. The ranking allows the analysis of the relation between firm characteristics and sourcing behavior within the VC industry and to correlate investment variables such as the level of automation with the level of data integration within their origination strategy (Sandelowski, 2000). By doing that, the authors of this paper tried to expand the scope of the research - which ultimately enriches the investigation - as they seek to capture several deal mapping methods within the VC industry. Throughout the analysis

section, the empirical results are also either compared, critically discussed, or validated with industry figures and reports.

3.2. Research Design

Among the most commonly used research designs, cross-sectional design seems to be the most suitable for the present study, given the exploratory nature and presented characteristics of the research at hand. A cross-sectional study comprises the collection of data from several cases at a single point in time (Bell, Harley and Bryman, 2019). In this case, the authors studied the best practices in data-driven deal sourcing that are currently being used by VC investors, capturing the ongoing data-driven trends in deal origination. It is worth noting that the capability of this study to provide findings related to long-term trends can be limited, given that the study focuses on a single moment in time. And that is one of the implicit drawbacks when choosing a cross-sectional study as the research design. By adopting a cross-sectional design, the authors are able to understand the main variations in the chosen sample, in this case, the variations of different VC firms when adopting and implementing data-driven solutions to improve their deal sourcing practices. That variation is highlighted by Bell, Harley and Bryman (2019) as one of the core advantages of cross-sectional design. These differences between the VC firms of the sample allowed the authors of the present research to generate a benchmark to classify the firms according to their data-driven practices in deal sourcing. Furthermore, given the time boundaries of the project, the cross-sectional design allows the authors to compare these diverse sets of VCs in a relatively short amount of time, which would not be possible with other similar research designs like longitudinal research. Consequently, after considering these dimensions, a cross-sectional design was chosen to be applied in this research as it allows for an adequate and comprehensive analysis of the topic and ultimately draw a generalized conclusion.

The overall structure of the research is then organized as follows and is depicted in Figure 5. First of all, the literature review builds the theoretical baseline for the paper. By examining existing literature for previous findings, discussions, and general research streams on the VC industry, the reader of this paper is provided with a deeper understanding of the underlying structures and dynamics of not only the VC ecosystem, but also the data-driven elements currently in use (Easterby-Smith and Junshan, 2018). This theoretical framework is very important in order to follow the analysis and thus understand the formulated recommendations at the end of this paper. To ensure a high level of quality throughout the literature review, the study and analysis of the literature is divided into two levels. The first stage involves reviewing the literature for an individual and holistic understanding of the theoretical background and the main concepts. The second level then includes the presentation of the literature that examines and juxtaposes the relationship between the main identified dimensions in the analysis (Knopf, 2006).

Based on the theoretical understanding, the authors of this paper could focus more on the gathering of empirical data that can be used for finding an answer to the research question. Sources for this information were the internal discussions with VGVC, conducted interviews with other VC investors and researchers and the initially used secondary literature. The

respondents stood out for their in-depth knowledge of the VC industry, technological expertise or networking skills. To validate the results, industry reports and figures were additionally being used in order to illustrate the relevance of the findings in today's market. It should be noted that the cross-sectional research design used was not linear, and integrated feedback loops after the interviews and after early findings in the empirical results section helped the researchers iterate the model and make it more responsive to the research question.

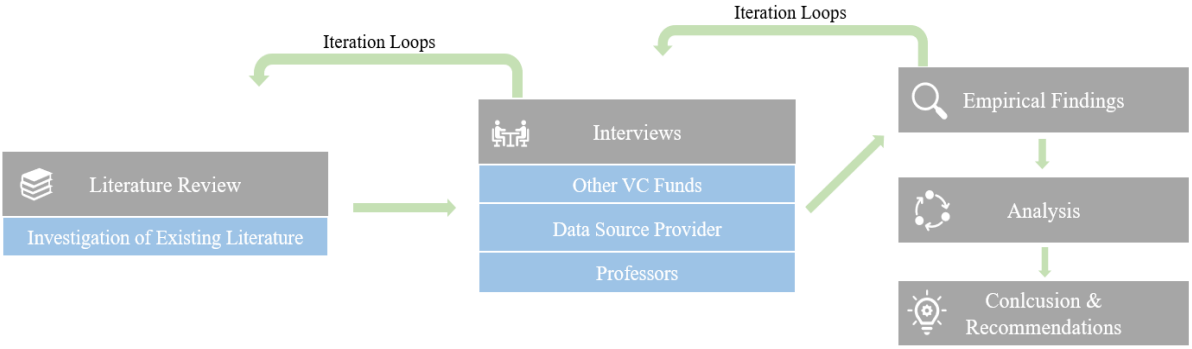


Figure 5. Overview of the cross sectional research method used in this paper.

3.3. Data Collection

To collect the primary data that is required to conduct the study and provide an answer to the research question, the authors carried out semi-structured interviews as it allows the interviewee to balance between structured elements and a certain level of flexibility (Bell, Harley and Bryman, 2019). As this research aims to get an in-depth understanding of the complex and dynamic business environments within the VC industry, semi-structured interviews were chosen as an appropriate technique. The flexibility given by this method supported the interviewers to engage in richer discussions with the interviewees, gathering more insightful data in respective dimensions than just having pre-formulated questions that are to be answered by the respondents. By following this strategy, the authors of this paper were able to identify i.a. new strategic sourcing activities that have not been previously observed. On the other hand, having a certain degree of structure was needed to perform an appropriate comparison of the data collected. The interviews were planned and designed to last around 30 minutes. However, given the semi-structured nature of the interviews, some of them lasted between 35 and 40 minutes. After the interviews, the authors of the paper transcribed the respective discussions to support and conduct the qualitative data analysis based on the notes.[3] Furthermore, the executors of the research designed and employed two different interview guides: One guideline was created for VC investors and another one for professors - which helped to support and serve as a more tailored directive for the interviews. The objective of these interview guides is to guarantee a certain degree of structure, and simultaneously, ensure that the discussions provide the relevant data needed for the research. The interview guideline can be found in Appendix 1 and 2.

[3] Once the interviewee agreed to record the meeting, the authors of this paper were able to create transcripts based on the recordings.

To reach out to knowledgeable individuals within the VC industry and set up the interviews, the authors leveraged their international network and connections. By studying and working in different European countries, the authors have developed an international network that facilitated the access to some relevant international VC actors. Furthermore, other sources were employed to establish new connections and ultimately set up interviews with other actors within the VC space that were considered relevant for this study. Finally, by conducting this research in collaboration with VGVC, the authors were given access to a large professional network and received support from Volvo Group in the form of contact initiation within the VC area. Table 2 provides an overview of all the interviewees with practical experience in the VC industry that are currently operating in either CVCs or IVCs. Robin Tech represents the only exception as he is part of a public database team. Table 3, on the other hand, depicts all interview partners within the academic field specialized in VC and Entrepreneurial Finance.

Name	**VC firm**	**Position**	**Interview Date**
Erik Johansson	Volvo Group VC	Investment Director	Weekly discussions[4]
Dennis Levien	Porsche Ventures	Investment Manager	31-01-2022
Philipp Rollwage	BMWi Ventures	Investment Analyst	31-01-2022
Johanna Holmström	Connect Sverige[5]	Head of Investment	02-02-2022
Alessio Fannelli	645 Ventures	Principal	04-03-2022
Kim Banham	Connetic Ventures	Partner	10-03-2022
Carles Illa	Nauta Capital	Head of Engineering	22-03-2022
Jacob Fellman	NGP Capital	Vice President	23-03-2022
Martin Eriksen	Viking Venture	Investment Manager	24-03-2022
Moreno Bonaventura	InReach Ventures	Lead ML Engineer	24-03-2022
Daniel Blomquist	Creandum	Operating Partner	25-03-2022
Sami Niemi	Spintop Ventures	Partner	29-03-2022
Kerstin Cooley	Brightly Ventures	Managing Partner	30-03-2022
Francesco Corea	Balderton Capital	Research Lead	01-04-2022
Rocio Pillado	Adara Ventures	Partner	19-04-2022

[4] The weekly meetings with Erik Johansson and other team members at VGVC helped the authors to complement and contrast the findings from the rest of the interviews, providing insights on the topic at hand and guidance for the study.
[5] Johanna Holmström left Connect Sverige to co-found Nordic Node, a consultancy that helps young ventures pursue global ambitions with international venture capital.

Jason Brenier	Georgian Partners	Head of Innovation	20-04-2022
Dr. Robin Tech	Delphai (Search Engine)	Managing director & Co-founder	16-03-2022

Table 2. List of interviewees including VC investors.

Name	**Institution**	**Position**	**Interview Date**
Roberto Ragozzino	Nova SBE	Professor	16-03-2022
Pedro Carvalho	Nova SBE	Assistant Professor	17-03-2022
Francisco Queiro	Nova SBE	Assistant Professor	21-03-2022
Peter Bryant	IE Business School	Adjunct Professor	25-03-2022

Table 3. List of interviewees including academics and professors.

3.4. Data Analysis

After carrying out the interviews with knowledgeable individuals within the VC industry, the authors performed the corresponding qualitative data analysis. One of the main objectives of qualitative data analysis is to provide the researcher with a better understanding of the overall topic, specifically, in regards to the research question at hand. The qualitative data analysis should support the researchers in the phase of processing and summarizing the data by identifying underlying relations within it. In order to identify relevant patterns, the thematic analysis strategy can be applied for a structured analysis of the collected data (Bell, Harley and Bryman, 2019). This technique relies on searching for themes by identifying different codes within the raw data in a relatively short amount of time. When conducting a thematic analysis, coding the collected data and forming themes makes large amounts of data more accessible for the following analysis and is therefore an appropriate approach for dealing with semi-structured interview transcripts such as those integrated in this research project. Moreover, thematic analysis is an efficient approach to capture the essence of respondents' perceptions and opinions on a particular topic, which is a crucial factor in this study. Considering the aforementioned advantages and the strategic suitability for the project at hand, the authors decided to apply the thematic analysis in this work (Bell, Harley and Bryman, 2019).

In order to ensure the conduction of an appropriate data analysis, the authors followed a four step process, which is in line with the suggested approach by Bell, Harley and Bryman (2019). First, the researchers reviewed the transcripts in order to familiarize themselves with the data collected. Second, the authors proceeded with the data coding process, which included two separate stages: An initial coding as well as a following review to ensure that the coding was appropriate and no relevant data had been omitted. Third, after identifying the codes, the main

themes were formed and formulated. Equivalent to the second step, the authors reviewed and refined the results. Finally, the researchers related the results of the thematic analysis with the research question at hand, aiming to provide an in-depth and insightful answer.

3.5. Data Needs

A detailed description of the systematic approach the students employed in order to collect, structure, and analyze the data and come up with reasonable recommendations was presented above. To answer the given research question at hand, however, appropriate and reliable data providing the executors of the investigation with deep industry insights had to be collected. First of all, the authors collected data about the characteristics of the VC. This section includes firm size, investment ticket size, headcount of the investment team, and other company features. This information helped not only to examine possible relationships between organizational elements and investment behavior but also to relate investment variables such as the level of data integration to the size or available resources of the firm. This data also supported the next investigation area in which the students ask for the current level of automation within the respective VC deal mapping activities.[6] The authors needed to develop a deeper understanding of internal deal mapping and how the deal origination process is currently structured within both VGVC and the respective interviewed market players of the VC industry. Since one of the main objectives of this paper also is to improve VGVC's mapping practices, the authors aimed at checking for potential partial adoptions and integrations from competitors' solutions and industry best practices. The identification of these two elements is thus critical in order to structure this process successfully. However, also the attitude towards the application of data-driven solutions within the deal initiation stage is of particular interest as it provides an overview of the industry's perception and acceptability of corresponding tools in the market. Given that several interviewed market players have recently undergone a transition away from the traditional human-capital intensive approach to more data-driven solutions, the authors are also interested in information about how to organize and structure a potential shift to respective solutions and what potential challenges might have to be encountered along the strategic transformation.

The primary data collected was then supplemented with secondary data in the form of literature, such as related academic journals and their conclusions as well as recommendations. Literature and secondary data being presented in the literature review that explicitly deals with the deal origination process was used in order to critically discuss the student's output and findings and allowed to see beyond the conducted research. By doing this, the expressiveness, validity and reliability of the thesis can be improved. The inclusion of academic literature is of particular importance when the collected data is processed using an inductive research strategy where the student's conclusions are drawn from the input data without having predetermined expectations of the analytical process which significantly would affect theory building and recommendations (Bell, Harley and Bryman, 2019). This provides the reader of this paper with both a rich and comprehensive analysis and discussion

[6] Deal mapping, unless not explicitly indicated otherwise, will be used as a synonym for deal origination and deal sourcing throughout this paper.

of the application of data-driven solutions within the deal origination process in the VC industry as this approach aims at incorporating both diverse and differing investor-specific perspectives into the work.

3.6. Research Quality

In this section, the validity and reliability of the thesis are evaluated as these are essential components of the research quality (Bell, Harley and Bryman, 2019). Furthermore, the ethical considerations contemplated by the authors while conducting the research are also reviewed in the present section.

3.6.1. Validity

According to Bell, Harley and Bryman (2019), validity can be defined as the most relevant criterion of quality research. Validity relates to the level of integrity reached by the authors when drawing the conclusions derived from the findings and results of the study (Bell, Harley and Bryman, 2019). Within the context of this research, the authors conducted the research by following various principles and methods to ensure the validity of the present study.

To collect the primary data through semi-structured interviews, the authors selected individuals with extensive knowledge of the research topic, both at the industry level and the academic level. The partner companies who have collaborated in the research through different discussions and interviews are some of the most relevant and prominent VC investors. They have an extensive track record and have backed some of the most innovative companies, including Revolut, Spotify, Klarna and others. Thus, the findings of the research include the perspectives and the best practices in deal sourcing currently used by distinguished players in the VC ecosystem. On the other hand, by selecting interviewees with different perspectives who have addressed the topic from different angles, the authors could triangulate the results and ultimately increase the validity of the findings. Furthermore, the data analysis has been meticulously performed through thematic analysis, examining the transcripts and notes from the interviews exhaustively to identify codes from the most frequent words and expressions highlighted by the interviewees. Later on, the themes were formed after reviewing the coding analysis and having discussions between the supervisors and the authors. During the different phases of the study, the authors additionally maintained frequent discussions both with the academic supervisor and the contacts at VGVC as the collaboration partner. The assessment throughout these discussions helped the authors to contrast the knowledge gained and the perspectives developed over the research period. The fact of receiving academic and industry orientation from knowledgeable supervisors lead the research to ensure a greater level of validity. As stated by Bell, Harley and Bryman (2019), it is worth noting that qualitative research can implicitly be subjective. Therefore, the authors have continuously considered this aspect throughout the different phases of the study, in order to approach the research from an objective and non-discriminatory angle.

3.6.2. Reliability

As defined by LeCompte and Goetz (1982), reliability refers to the degree to which the meaning and interpretation of a given result reach a minimum level of congruency when analyzed by various interpreters. Reliability can be classified into two different types: external and internal reliability (Leung, 2015).

On the one hand, external reliability is defined by Bell, Harley and Bryman (2019) as the degree to which the results of the research are repeatable. External reliability can be challenging to achieve when conducting qualitative research (Bell, Harley and Bryman, 2019). To tackle this issue, the authors clearly described and documented the most relevant steps that took place during the study. Furthermore, the decision-making process considered and applied by the authors throughout the research is also comprehensively described in the different sections of the paper. Additionally, the transparency applied by the authors can simultaneously increase the level of replicability of the study. On the other hand, internal reliability refers to how data is understood and interpreted by the members of the research team, and most importantly, the level of agreement on what they hear and see (Bell, Harley and Bryman, 2019). Within the context of the present study, two researchers have designed and conducted the different sections of the research which can enhance the level of reliability of a study (Bell, Harley and Bryman, 2019). In this vein, both authors have been equally involved during the research, reviewing the literature, participating in the interviews, analyzing the results, and continuously maintaining discussions with the supervisors. Furthermore, in each of the different steps and decision-making actions, the authors have discussed and exchanged their points of view, aiming to reach a significant level of internal reliability for the outcome of the study.

3.6.3. Ethical considerations

In terms of ethical considerations, the most relevant issue in the context of the present study is the ethical considerations between the researchers and the research participants such as the collaborator companies (Bell, Harley and Bryman, 2019). To satisfactorily tackle potential ethical issues, the authors have addressed certain commonly recurrent issues that were defined by Diener and Crandall (1978). These ethical principles are harm to participants, lack of informed consent, invasion of privacy, and deception.

Regarding the harm to participants, one of the concerns highlighted by Diener and Crandall (1978) is the origination of stress in the research participants. To tackle that issue, the authors were flexible and adaptive to the schedules and time availability of the interviewees. Furthermore, the authors did not pressure or lead the interviewees in any direction or toward questions that the interviewee would not feel comfortable with or able to respond to. That also helped the authors to ensure that there was no invasion of privacy in this regard. Moreover, the authors explicitly asked the participants if their names and roles could be shared in the study. However, the possibility of not sharing their data, information, and responses was offered to the participants. Regarding lack of informed consent, the interviewees were informed before their participation about the aim of the research, the context of the study, the

purpose of the interviewees, and how the data would be handled. To prevent deception, the authors transparently communicated the research purpose and context to the participants. Furthermore, the authors kept this clear and transparent communication, before, during, and after the interviews.

4. EMPIRICAL FINDINGS

The empirical findings presented in this chapter are structured as follows. The first section of the chapter will present the different data sources that were mentioned by the interviewees. Subsequently, the dimensions that VC investors are monitoring to identify and qualify deals in their deal sourcing practices are described. This section includes different subsections to elaborate on each of the relevant dimensions highlighted by the interviewees. In the third section, the organization and automation of data by VC firms during the deal sourcing process are described, where the relevancy of proprietary algorithms and CRM software are emphasized. The fourth section of the chapter introduces the theme of unbiased decision-making processes in VC investments and its relationship with the application of data-driven solutions. Lastly, the last section elaborates on the outlook and future perspectives of data-driven deal sourcing practices within the VC industry, contrasting the different points of view of the interviewees. Each section includes quotations from the interviews to elaborate on the themes and illustrate the essence of the empirical findings. Moreover, an overview of all the respective themes and codes that were identified through the analysis of the empirical findings is displayed in Table 4. The themes identified throughout the thematic analysis serve as the foundation of the main sections of the present chapter.

4.1. The Different Data Sources in VC Deal Sourcing

VC actors have to find certain ways in order to extract the relevant data and information that can support their data-driven deal sourcing practices. The sampled VC investors rely on a combination of several data sources - some of them being internal and some others more of external nature. Based on the results from the interviews, one of the most relevant data sources is the ones offered by third party providers in the form of public databases. The interviewees differentiate themselves in terms of what kind of database they subscribed to, the number of subscriptions and the way they structure, organize and make use of the data for financial decision making. The latter one will be the focal issue in the following section. However, most of the VC funds of the given sample have a significant amount of data sources in common.

To begin with, Illa highlighted the importance of a diverse selection of data sources that build the foundation for their investment decisions. Nauta Capital is aware of the fact that public databases are in widespread use among VC investors. Therefore, through both additional data sources such as events and news websites, and their way to process the data, Nauta Capital tries to differentiate themselves and stand out from competition.

> *"The most relevant data sources for our investment decision-making process are Crunchbase, Pitchbook, Delaroom and other public databases. We have seen that the database of Crunchbase is improving every year, but it is also used by all other VCs. Another good source for information are events, because many companies that want to raise money participate in these events. We try to get as many sources as we can."* - Carles Illa, Head of Engineering at Nauta Capital

NGP Capital tries to utilize as many data sources as possible to extract diverse data and relevant information. Jacob Fellman noted that data plays a significant role for their investment strategy and therefore they spend a considerable amount of capital in data resources. Fellman also mentioned that they are continuously trying to improve their sourcing tool by identifying other data sources that can be of strategic value.

> "*At NGP we have a mix of different data sources, including traditional databases like Pitchbook, Tracxn and Dealroom as we have subscriptions to all of these databases. We actually have a significant investment in data resources."* - Jacob Fellman, Vice President at NGP Capital

Eriksen from Viking Venture stated that the fund is using a public database focussing on the Nordic countries as a primary source for the target list compilation. In addition to that, Viking Venture uses Pitchbook, Crunchbase and Vainu as complementary data providers. Eriksen mentioned that in Viking Venture's case, the geographical focus influences their data source choice as they focus their investments in the Nordic countries where companies are obligated to disclose a significant amount of company information already due to the legal framework.

> "*Our team goes through the list of companies that are extracted from all available domains from Proff*[7] *which is one of the main data sources. In addition to that, we use Pitchbook as an alternative data provider. However, in the Nordics it is different because you have access to all company information. [...] We use Crunchbase as well. And also a finnish data provider called Vainu."* - Martin Eriksen, Investment Manager at Viking Venture

Creandum has a fragmented and diverse strategy in order to extract relevant target information. In addition to checking the public databases, the Swedish VC also receives not only company data and deal leads from their personal network but also information such as employee fluctuation from previous and current unicorns.

> *"Creandum has subscriptions with Crunchbase, Pitchbook and Dealroom as most VCs also have subscriptions. [...] For us at Creandum, founders and their backgrounds are an important factor when it comes to finding and evaluating deals. For instance, someone leaving Spotify or Revolut - that usually is a key signal for Creandum and something we get informed about through our strategic network. You can also use LinkedIn, targeted searches on people and if something changes in their resumes - which of course can be automated somehow."* - Daniel Blomquist, Operating Partner at Creandum

BMWi Ventures strategy, in addition to the usage of public databases, tries to leverage the availability of and close operational connection to other internal business units of the BMW

[7] All domains of Proff refer to proff.se, proff.no, proff.fi and proff.dk

group as they provide them with possible deal leads. The internal communication represents a strong and reliable data source for the corporate investment arm of the German group.

> *"We seek to leverage the strategic relationships with the various business units of the whole BMW group and adopt the initial industry screening that they have conducted in their respective market of interest. Thereby we can identify this connection as another source for potential deal leads."* - Philipp Rollwage, Investment Analyst at BMWi Ventures.

Johansson from the Volvo Group Venture Capital notes that the relationship between the corporate investment arm and the parent company should lead to mutual synergies. In addition to the previous argument by Rollwage, the Swedish group's investment director highlights the sector intelligence that can be gained by interacting with the relevant markets - which are then ultimately unlocked for the entire group.

> "*When we are investing in our selected business models and entrepreneurs, we automatically receive industry insights from the sectors of our interest. That can be a result of the interaction with the venture, but is also a consequence of investigating the market in more detail before placing the respective investments. Just as we receive information from other Volvo Group business areas, this newly acquired market knowledge is then made available to other departments to support their strategy and thus increase the Group's overall competitiveness.*" - Erik Johansson, Investment Director at Volvo Group Venture Capital

Balderton Capital also relies on a very diverse range of data sources. The venture fund has subscriptions to several public databases and additionally tries to enrich the decision making process with product reviews, employee fluctuation and announcements.

> "*Our data sources include different things. For instance, more general information about the startups is retrieved from public data sources such as Pitchbook, Crunchbase, CB Insights, Preqin, Capital IQ and Dealroom. This covers basic general and financial information about the company as a starting point like who the investors are, where the company is located, amount of funding, among others."* - Francesco Corea, Research Lead at Balderton Capital.

4.1.1. The availability of information and reliability of existing data

Even though most data-driven VCs utilize a diverse range of data sources, it is worth noting that reliable information and data about young ventures are not always publicly available for VC investors. While there are several data providers offering services in the form of public databases, that does not mean that the reliability of the data is always flawless. Some data might be outdated, leading VC investors to miss relevant data points from a venture. For instance, the latest funding round raised by the company might not be updated, or an important announcement of the venture might not be covered by the database. Even in some cases, the

company profile of a specific venture might exist in the database, but without relevant available data about the company in question. Consequently, although VC investors would like to utilize more data, they are very likely to face the challenge of unreliable information or even lack of data in the case of young ventures that lack market traction.

In contrast to the widespread use of public databases highlighted by the various VC stakeholders above, Robin Tech, co-founder of Delphai, questioned the overall quality of the underlying data for financial decision making. For him, accurate information cannot be taken for granted when employing public databases in sourcing strategies.

> *"Significant data points provided by public databases have to be questioned because of unknown sources, redundant data or manually extracted information that is prone to errors. Considering the strategic and financial value of investments - and we are talking about a highly uncertain business environment - the quality and reliability of the little data available for decision-making must be ensured at all times."* - Robin Tech, Managing Director and Co-Founder at Delphai.

This lack of data and specified information about the startup companies and the respective degree of opacity was also pointed out by other interviewees.

> *"Since startups in their early stages are not public companies, quantifiable data about those companies can be very difficult to access which can be an important challenge for VCs when becoming more data-driven. This can be even more difficult if the company is not part of the VC portfolio or within their network."* Francisco Queiro, Assistant Professor at Nova Business School.

Pillado from Adara Ventures also mentioned that for some companies in their very early stages, there is no product or service launched to the market yet, and therefore, there is no traction to generate data or indicators.

> *"In early-stage investments, sometimes it is just an entrepreneur with an idea, in some cases there is not even a product or service launched yet, and consequently, there is not too much public data about these very young companies."* Rocio Pillado, Partner at Adara Ventures.

This previous challenge regarding the unavailable information partly determines which are the dimensions that VCs monitor and observe based on availability of data and relevancy of that data to identify investment opportunities. These dimensions are covered in the next subsection, showcasing the most highlighted dimensions by the VCs in the sample.

Codes	Themes
Signals, combination of signals, investment criteria, data points, dimensions about the founders, dimensions about the company. Traction, momentum, reputation, events, news coverage, accelerators, angel investors, seed investors, startup stage, headcount, web traffic, social media traction, job ads, early stage, late stage, background of the founder, stamp of approval, previous investors, number of employees, patents, product reviews.	**Dimensions that VCs are using to source deals.** Most highlighted dimensions by the VC firms in the sample.
Data sources, data providers, subscriptions, investment strategy, network, available data Pitchbook, Dealroom, Crunchbase, CB Insights, Tracxn, LinkedIn, Capital IQ, Twitter, AngelList	**Data sources used by VC firms.** Most highlighted data sources by the VC firms in the sample.
Automation of deal sourcing process. Development and implementation of proprietary proprietary systems/platforms/algorithms. Proprietary system based on AI and ML models. In-house data engineering team working and focusing on improving the proprietary data-driven solutions. Automatic gathering of data and filtering processes. Automated integrations into CRM software. Proprietary systems help screening opportunities by filtering data and presenting the most relevant investment opportunities. Feedback loops to improve the ML models and therefore, get smarter suggestions from the system. Proprietary systems can provide a scoring system to support the investment team in making decisions. CRM, automation of processes, integrations, flexibility of the CRM software, customizable, features, CRMs to support the investment team in making decisions, CRM software used as a system of record. Affinity, Pipedrive, Zendesk	**How VCs are organizing data and automating the deal sourcing process.** The important role of CRM tools. Most highlighted CRMs by the VC firms in the sample
Equal opportunities, avoid biases, remove biases, personality traits, reinforcement biases, minority groups, unbiased decisions, diverse team, awareness	**Unbiased decision-making process in VC investments.**
Industry changes, utilization of data, implementation of data-driven solutions, VC industry, data-driven trends, human-driven business, competitive advantage, human interaction, networking, future	**Outlook and future perspectives of the VC industry.**

Table 4. Codes and themes identified from the primary data.

4.2. Dimensions applied by VC Investors to Source Deals

Throughout the conversations with the different interviewees, another main theme was related to the type of data that VCs are using to improve their deal sourcing process and financial decision-making. Therefore, the VC investors highlighted various dimensions that are relevant in their current data-driven deal origination practices to identify and qualify potential deals. It is worth noting that monitoring these dimensions partly depends on the availability of data which in turn, depends on the startup stage in which the VC firm is investing. Hence, in early-stage VC investments, there are fewer dimensions that investors can leverage to support their deal origination practices. On the other hand, when moving to later stages, startups have already developed and refined their products and services. That usually takes place after raising pre-seed or seed funding rounds, where early-stage investors, including angel investors and early-stage VCs, have already analyzed, validated, and invested in these companies. On many occasions, at this point, the products and services developed by those startups have been launched in their target markets and potential clients have been able to interact, in one way or another, with the company and its business model. Therefore, as an outcome of this traction, more publicly available information can be found about these startups, and VCs can utilize and complement a variety of dimensions. That can become a highly creative process, where the dimensions and signals are endless. Consequently, VCs should choose what they believe are the most relevant dimensions that could help them to screen, map, and identify the most promising investment opportunities.

Each of the following subsections presents the relevant dimensions and the reason why VCs monitor these dimensions to identify and qualify investment opportunities.

4.2.1. The founding team and their career movements

The importance of the founding team and their background was also highlighted by various interviewees as one of the relevant dimensions, especially when there exists a lack of other data about the target company at hand. This was mentioned by Blomquist from Creandum:

> *"In early stages like pre-seed, there is usually very little information on the company, but maybe more available data about the founders. Founders and their professional and educational backgrounds are an important factor when it comes to screening and assessing potential deals."* - Daniel Blomquist, Operating Partner at Creandum.

Thus, the available data about the founding team plays a critical role from the VC investors' perspective and can support the VC firm not only to evaluate existing startups but also to proactively identify the creation of potential new ventures. In this vein, various interviews mentioned how they have been able to proactively anticipate the creation of new ventures by tracking career movements and decisions of talented individuals within recognized organizations. In other words, when knowledgeable individuals with entrepreneurial or management experience decide to leave an organization that can be seen as a signal, as these individuals might start their own venture. This type of dimension is considered by global VCs - including Creandum:

> *"When C-level executives or management employees decide to leave one of the large technology companies can be considered a signal, as these individuals could start their own venture."* - Daniel Blomquist, Operating Partner at Creandum.

Blomquist provided a clear example within the FinTech space, as it is one of the sectors the firm has a significant interest to invest in.

> *"For example, if a C-level employee from Revolut leaves the company, we check the next movements of that person, because Creandum is interested in Fintech and that former manager from Revolut potentially could be the next fintech founder."* - Daniel Blomquist, Operating Partner at Creandum.

From a slightly different perspective, other interviewees highlighted the importance of tracking M&A operations within the industries of interest of the VC firm. The reason for tracking potential acquisitions is because some talented individuals with industry experience leave the organization after being acquired and decide to capitalize on their industry experience by starting up new ventures. This is a dimension that is being utilized by the North-American VC firm 645 Ventures:

> *"At 645 Ventures, we track M&A operations and especially if there are employees in senior positions departing the company that is being acquired. Some M&A operations involve a change in the corporate culture and bring new dynamics to the organization, consequently, some employees decide to leave the organization to start their own ventures."* - Alessio Fanelli, Principal at 645 Ventures.

4.2.2. Online presence

Other dimensions remarked by the interviewees were those related to internet buzz like news coverage, social media activity, and company website traffic. These are relevant dimensions that could be tracked in several stages of startup growth. For instance, a startup raising a seed round or being selected for one of the most prestigious accelerator programs could be covered by media sources. Similarly, a company raising a funding round in later stages could also generate internet buzz on social media and consequently increase their company website traffic. For those companies in later stages which have certain market traction, these types of dimensions could provide indicators to VC investors on how customers value the products and services offered by the startup. Furthermore, by monitoring these types of dimensions, VCs can not only identify new investment opportunities but also monitor market trends and the development of new technologies that could be of interest to the VC firm. Nonetheless, given the scope of the present thesis, the focus will be on the relevancy of these dimensions for the deal origination stage.

The online presence was highlighted by Corea as a dimension commonly used by Balderton Capital to identify and qualify deals. Within these dimensions, Corea mentioned a variety of data points that can be gathered by looking at the internet presence of the company.

> *"We gather and screen many dimensions that can include, among others, data from mobile apps, website data - several data points can be observed here, for instance, number of visitors or from which other website you are landing from-, product reviews, announcements - new office, funding rounds, and so on.-, headcount, or patents."* - Francesco Corea, Research Lead at Balderton Capital.

Similarly, Blomquist from Creandum indicated the relevancy of observing the social media traction of the company to complement other dimensions and ultimately qualify potential investment opportunities.

> *"When moving towards later stages of the venture like Series A and onwards, there exists more information about the startups. At this point, it is still relevant to have the focus on the funding team as an important element, but it is needed to analyze a combination of data points and different signals. Then, the available data points can be very broad and different, from LinkedIn data, for instance, if a company is growing in terms of the number of employees to social media traction like Twitter followers, or in some cases activity of Github repos."* - Daniel Blomquist, Operating Partner at Creandum.

At Nauta Capital, they try to scrape as much relevant data as possible, and a significant part of that data comes from the internet presence of the company. For instance, identifying if interesting target startups are participating in events or are mentioned in the news coverage.

> *"In terms of data points and data sources, we try to get as much relevant data as possible. Part of that data comes from scraping the Internet activity by looking at events, directories, news, and various websites with relevant startup and VC information."* - Carles Illa, Head of Engineering at Nauta Capital.

4.2.3. Growth in employment, a positive signal for VCs

A dimension that several interviewees highlighted is the recruitment activity in startup companies. In other words, when startups are hiring new talent, it is understood as a relevant positive signal that can be interpreted as an indicator of company growth. Furthermore, a startup hiring new employees signals to the market that they have the economic resources to invest in talented individuals that will help the company become more competitive to achieve its corporate goals. This perspective is shared both by the academic interviewees and by VC investors.

> *"Start-ups recruitment activity is an important indicator of growth - and growth is a key factor for VCs. In this sense, AngelList has become one of the most relevant*

platforms connecting startups that are hiring with talented individuals. The data that AngelList possesses is highly valuable because it shows which are the startups that are growing." - Francisco Queiro, Assistant Professor at Nova Business School.

Other interviewees from well-recognized VC firms also emphasized the importance of keeping track of the hiring activity of relevant companies that they could have under the screening radar.

"We had previously partnered with an external software development firm to help us develop and implement a web crawler that collects all those Norwegian companies that had posted more than five recruitment ads in the previous month on LinkedIn. The reason to invest in this data-driven tool is that the hiring activity of a startup is a relevant early indicator of growth." - Martin Eriksen, Investment Manager at Viking Venture.

"Some of the data sources that we are currently using show you how the number of employees of a company is changing over time. If the company is currently hiring, that is considered an interesting data point for Balderton." - Francesco Corea, Research Lead at Balderton Capital.

4.2.4. Previous investors as a quality stamp for VCs

Three interviewees highlighted the importance of screening those companies that have been backed by well-recognized and prestigious investors, including those startups that were part of reputable accelerator programs. According to what VCs responded during the interviews, being selected for a well-recognized accelerator program like Y Combinator or having a distinguished investor in your cap table is interpreted as a quality stamp by VCs.

"We have a list of angel investors and pre-seed funds that we believe have good criteria. The main idea behind following the investment activity of these investors is because they are sort of a "stamp of approval". This is also something common in the entrepreneurial and VC ecosystem, if a founder is part of Y Combinator, he or she is probably a good founder." - Daniel Blomquist, Operating Partner at Creandum.

The investment activity of certain investors is also a dimension that Nauta Capital is tracking, not only to qualify potential deals but also as a way to identify new investment opportunities that otherwise would take longer time to identify through other mechanisms.

"Among the different dimensions that we gather to feed the algorithm, one important data point is the previous investor of a specific startup. From another perspective, we also look at the recent activity of some interesting early-stage VCs and accelerators" - Carles Illa, Head of Engineering at Nauta Capital.

Spintop Ventures do also rely on this type of dimension, particularly as it helps the investment team to qualify deals. Therefore, for the Nordic VC firm, a startup who is backed by a relevant investor within the Spintop Ventures network is seen as a relevant distinctive signal that could influence moving that company forward to later stages of the investment funnel.

> *"When identifying and evaluating investment opportunities a crucial aspect is to see who the previous investors are in that company. We like to see that other VCs, and particularly VCs that we know, have invested in that company. It is a quality stamp because investors that we trust have already done some analysis on the startup at hand."* - Sami Niemi, Partner at Spintop Ventures.

This dimension can support the deal sourcing process of VCs in two different ways. First, monitoring and tracking recent investment rounds made by relevant early-stage investors in a specific industry could help VCs recognize potential investment opportunities. Second, if those startups have been already identified by VCs, being backed by well-recognized investors can be a relevant dimension showing the potential of the startup. Therefore, it positively influences the decision-making process within the deal sourcing stage.

4.3. How VCs are organizing data and automating the deal sourcing process

The first two sections of the empirical findings explained which are the most relevant data sources and data dimensions that VCs are utilizing to identify and qualify potential investment opportunities. However, becoming more data-driven does not only include the usage of more data and sources during the deal origination process but also how the fund is organizing, handling, and working with that data in a strategically and competitive way. Therefore, a crucial component to effectively identifying and qualifying the most relevant deals is the level of automation of the deal sourcing process. In this vein, one of the most important priorities of data-driven firms is to efficiently retrieve and sort out the data from a variety of data sources, and present the data in a user-friendly manner to the investment team. To perform this process effectively, several VC firms have developed their internal data-driven tools, including smart algorithms and machine learning models to help the firm automate the process. Furthermore, these tools support the investment team in identifying opportunities and highlighting which of those deals could be more relevant based on the VC investment criteria. In this way, the investment team directly focuses on the most promising deals, which helps the team to be more time-efficient by reducing manual work during the compilation and analysis of data.

Nauta Capital, is one of the pioneer firms in data-driven deal sourcing practices, and they have developed an internal solution that helps the firm to source and qualify deals in an automatic manner with the implementation of algorithms and data models.

> *"Through our internally developed system based on the application of algorithms, we automate the gathering and filtering of data. Within the system, the scrapers are*

constantly scraping data from the internet, and the curated information with potential companies that could be interesting to analyze is provided to our investment teams periodically." - Carles Illa, Head of Engineering at Nauta Capital

In a similar manner, NGP Capital is another interviewed VC firm which has also developed an internal data tool that helps the firm to automate the deal sourcing process. Their internally developed algorithms, called Q, are additionally fed with feedback from the investment team so it can work more efficiently over time. Furthermore, they have a data and analytics team working with the platform on a daily basis.

"Our machine learning models are retrieving data from many different data sources to track more than two million companies in real time. The model considers stage, momentum and sector fit of the venture. Our platform automatically filters those companies that do not fall under our investment scope. Q is also providing a score for each company which makes it easier to prioritize. We have integrated both the scoring system and the platform into our CRM, so feedback is provided based on the progress of those companies through the investment funnel. The feedback loop is very important for NGP´s system to improve over time." Jacob Fellman, Vice President at NGP Capital.

Even though Blomquist from Creandum stated they have not developed highly automated systems with advanced algorithms to source deals, they still have programmed some integrations to help them automate part of the process.

"We have a tool that brings new deals based on certain parameters. We get pinged when there are relevant signals in the market like the creation of a new company or the announcement of a new funding round coming soon. This approach provides the team with a weekly list of companies. Then we manually review this list in order to check if there are actually any strategic matches with the identified startups. If there is a fit then that company is exported to our CRM." Daniel Blomquist, Operating Partner at Creandum.

Other VC firms utilizing internal data-driven tools to organize and qualify data about potential investments are working slightly differently. For instance, Balderton Capital's internally developed platform is providing the investment team with relevant flags in their system reporting announcements of companies that fall under the investment criteria of the VC firm.

"Our internally developed software is then processing all this data. The final outcome of the system is, however, not a list of companies, but rather a specific flag in our CRM that suggests the investment team to review a specific company. For example, these flags could contain something like hiring ads from a startup, or if a company has reported an increase in revenue." - Francesco Corea, Research Lead at Balderton Capital.

4.3.1. The role of customer relationship management tools (CRMs)

Customer Relationship Management (CRM) software plays a relevant role in VC investments by supporting VC firms in managing the leads and the communication between the investment team and the leads while storing and organizing, in one place, data that could be relevant to making investment decisions. In other words, CRMs allow VC investors to efficiently manage their leads and keep track of the progress of each investment opportunity along the different stages of the pipeline. Therefore, CRMs organize and store relevant data about startups or entrepreneurs that could be interesting for the VC firm. Since CRMs contain data and are present from the very early stages of the investment funnel, including deal sourcing, they should be considered when studying the data-driven deal sourcing of VCs and addressing the research question of this thesis.

Within the CRM space, several providers are designing and delivering CRM solutions in different forms, which can be adapted to the users' traits and needs. According to the results of the interviews, there seems to be a variety of CRM software that are currently used by the different VC firms. On one side, some VC firms are using generic CRMs, for example, Zendesk, Pipedrive, or Coda, which could also be used by other types of industries, not limited to investment firms. On the other hand, the empirical findings show how other VC investors are working with CRMs that are specifically designed and tailored to the VC investment activity. For instance, the second group of CRMs would include tools like Affinity.

> *"In terms of CRM, we have been working with Affinity for five years. Some other VCs are working with Asana or Salesforce, however, Affinity is well known and used in the VC space. We believe that Affinity has some useful features for instance looking at relationships and networking. Furthermore, the tool allows the automation of certain processes, including the upload of new deals."* - Daniel Blomquist, Operating Partner at Creandum.

Pillado from Adara Ventures highlighted how helpful certain features provided by the CRM can be in order to leverage existing connections and networks to connect with potential companies and get referrals.

> *"One of the main reasons why we are using Affinity is because it has a functionality that helps our investment team to easily track the email exchanged or other types of interactions with other companies in an automatic way. We do not have to manually input that data, it is automatically monitored and updated by the system. Furthermore, we can see how the investment team is connected through their network to external companies or entrepreneurs."* - Rocio Pillado, Partner at Adara Ventures.

Based on the conversations with various interviewees, some VCs are also building integrations between their CRM software and other data-driven solutions supporting the deal origination process. Therefore, the flexibility of the CRM system to provide integrations with other tools used by the VCs seems to play an important role when deciding on the CRM provider. By

developing and implementing these integrations, VCs can tailor the CRM features to the VC needs while automating part of the process and reducing manual tasks that otherwise would be time-consuming for the investment team. Most importantly, the integrations built by some VCs help them to support their data-driven solutions, and therefore, their CRM becomes a relevant element in helping the firm become more data-driven.

> *"We believe that Affinity is a good CRM system. We have an API that helps us to link Affinity with the datasets we collect through our internal data-driven tool. Moreover, in Affinity there are already some great integrations built by the provider which helps to connect the CRM with data from external data sources like Pitchbook or Crunchbase. VCs can also build more complex integrations to automate certain processes."* - Francesco Corea, Research Lead at Balderton Capital

Other interviewees also remarked the importance of selecting flexible CRMs to utilize the existing integrations provided by the CRM company or building your own integrations. That flexibility in terms of potential integrations is one of the reasons why NGP Capital is working with Coda, as stated by Fellman.

> *"In terms of CRM we are using Coda. It is very flexible and allows you to build many different things in a customized way and without much coding. It is also relatively more affordable than other CRMs. So, at NGP Capital we have built our own CRM on top of Coda. I believe Coda is one of the most flexible tools in this sense."* - Jacob Fellman, Vice President at NGP Capital.

In a similar fashion, Spintop Ventures is working with Pipedrive as their CRM tool, and as previously mentioned by other interviewees, they have also built integrations on top of the standard CRM solution.

> *"At Spintop Ventures we are working with Pipedrive. It works quite well for us, and we have built some internal integrations to establish certain automatic connections to other systems."* - Sami Niemi, Partner at Spintop Ventures.

4.4. Unbiased decision-making process in VC investments

The importance and prominence of unbiased decision-making was also highlighted by the sampled interviewees. It seems that impartiality clearly plays a strategic role in VC investments for them, and according to the interviewees, it is influenced by the gradual integration of data-driven solutions in deal origination. However, the respective interviewees also mentioned the different levels and approaches to address this challenge. Each of the following VC funds consider biases in their decision-making process from a different perspective.

For example, Kim Banham from Connetic Ventures noted that the traditional pitch presentation setting provides the opportunity for a high number of implicit biases. As some founders might be exceptional performers in presenting their business idea, it says little about the market fit and operational eligibility of the business model at hand. According to Banham, certain background information, such as previous employment, visited universities and age of the founders can already result in implicit biases within the investment decisions.

> "*The target company pitch is broken and biased, because extroverts are expected to present in a more persuasive manner - meanwhile this ability has little meaning for a successful business development in the future. How an entrepreneur is wired cannot be identified only through a traditional pitch presentation. We believe that founders are the most important piece of the investing puzzle - so analyzing founders and team members is the most heavily weighted part of our algorithm and diligence process. However, we do not believe that school, experience, background, or demographic data is an effective way - and certainly not an unbiased way - to analyze startup founders.*"
> - Kim Banham, Investment Partner at Connetic Ventures

NGP Capital also realizes the existence of possible biases in VC investments, noting that they are working on solving this challenge on a continuous basis. The corporate investment arm of Nokia tries to overcome misconceptions in various ways. By increasing the range of target companies to consider for their long list, focussing more on the business model at hand and the strategic fit with their parent organization, NGP Capital works towards more unbiased decision-making.

> "*We believe that unbiasedness is an important aspect and there are a lot of layers to consider when adding personal data from the founders in the system. We are constantly working towards a solution in order to have a more unbiased decision-making process. In one way, our system is not very biased in the sense that we are weighting other factors such as momentum, stage fit and thematic fit stronger than personal traits of the founding team. The founding team does not have significant influence in determining a score for the target company. We are also reducing biases by looking at a wide range of companies."* - Jacob Fellman, Vice President at NGP Capital

Creandum is another VC firm recognising the possibility of impartial decision-making in VC investments. However, Creandum's sourcing strategy is primarily based on personal interaction with the founding team - without the support of data-driven tools to reduce irrationality in their respective filtering decisions. The VC fund believes that it is very important to assemble a diverse team to get to the root of the challenge. This is also because of the wide range of signals that draw attention to certain target companies in the fund's deal origination process. However, not only historical data of the target company but also the characteristics of the founding team are included in Creandum's evaluation of the target company. According to Daniel Blomquist, being aware of potential biases is already the initial step to avoid partial elements in decision-making.

"*Creandum prefers to meet and personally talk to the founders, rather than just letting data analytics tell us which founders could be worth investing in or not. But of course by doing so, one cannot avoid unbiased decisions. We believe that, even though you are using an algorithm, decisions can still be biased. We are trying to build a diverse team at Creandum to help us make less biased decisions - or at least be aware of potential biases. In Series A it may be easier to avoid them, because you have other signals or data points to support decisions, while in pre-seed and seed stages it is a lot about the founders*" - Daniel Blomquist, Operating Partner at Creandum

4.5. Outlook and Future Development of Data-Driven Solutions in VC Deal Sourcing

The future application and integration of data-driven solutions is also considered by the interviewees in this research. However, perspectives differ from investor to investor as well as from actors with practical VC experience to researchers who approach the topic from an academic angle. Opinions of the interviewees reach from an expected stagnation of the current level of data-driven solution integration to entire automated deal sourcing procedures in the future. In addition to that, several risks and shortcomings are being identified by the interviewees. Spintop Ventures, for example, highlights the difficulty to use data in early-stage investments. The investment focus in terms of stage and company maturity play a critical role for the fund. The venture's strong market reputation helps to primarily rely on their networking abilities and recommendations from strategic partners.

"We *believe that in early-stage investments it will be difficult to apply more data-driven solutions. Pre-seed VC investments usually involve little quantitative data and we do not believe that this will change in the near future. However, at Spintop Ventures we want to reach higher levels of automation for certain processes of the deal sourcing such as the better mapping of companies.*" Sami Niemi, Investment Partner at Spintop Ventures.

Fanelli from 645 Ventures also mentioned the relevance of stage focus. He expects the data challenge to become even bigger in the future as seed investments tend to accrue even earlier than nowadays. Therefore, Fanelli awaits an increase in data-driven solutions - especially those that are focusing on improving network-related activities as they are highly relevant in early-stage investments.

"*Over the last 2 years, the competition in the VC market has become pretty tight and seed rounds were made even earlier and earlier. In the past, you might seed a beta product out or similar, while nowadays the founders come out of known companies like Uber, Airbnb, etc and are able to raise money very quickly with little work done. At the same time, the "Growth Seed" asset class has disappeared, as a result of early Series As. We therefore think seed funds might spend more time building network-related software.*" - Alessio Fanelli, Principal at 645 Ventures.

For Illa at Nauta Capital, however, the VC sector will see a general increase in the usage of data-driven solutions. The current status quo in the industry, with its already widespread use of data analytics, will be subject to change towards an even higher intensity of automation tools. Especially to identify market trends in a time-saving manner, the role of automated market intelligence will become even more prominent. However, background checks of the founding team and competitor analysis do not have to be conducted manually.

> "*We strongly believe that the VC industry is becoming more data-driven and more VCs are implementing data-driven solutions. What we see today is just the starting point, where data is applied to identify startups. However, the aim is to analyze and make decisions in an automated way. In other words, that would be to add more filters to reduce the manual work while you move through the investment stages more automatically. Another interesting point would be to identify market trends and see if the startup could add value there with algorithms.*" - Carles Illa, Head of Engineering at Nauta Capital.

In line with the previous statement, Fellman at NGP Capital also expects an increase of overall data usage in the deal origination process. He identified the possession of relevant and up-to-date data as a level playing field. VC investments, however, still involve social, interactive elements between potential investors and fund recipients. Fellman therefore expects a balance between data-driven solutions and human relationships, but with a significant increase in data integration.

> "*I believe that the importance of leveraging data is definitely increasing all the time. It becomes more of a competitive factor rather than a competitive advantage to possess data. The issue here is that from there it is still a very human driven business based on relationships. Even if you find the best companies: How do you get access to them? How do you convince the team to collaborate with you? From that perspective, we believe that the human element is never gonna disappear - at least in the coming years. The qualitative aspect including human interaction and assessment in parallel are also needed*." - Jacob Fellman, Vice President at NGP Capital.

Eriksen at Viking Ventures observed an increase in data analytics as well. For him, however, the geographical investment focus plays a critical role in the way data-driven solutions will be incorporated into the deal origination strategy of a certain fund. Nordic countries and their transparent disclosure requirements allow for easily accessible target company information, while other regions are more opaque in nature due to their legal framework.

> "*Data-driven solutions will definitely become not only very important in the VC industry but more and more distributed in the next coming years. However, something to bear in mind is that VCs invest in small and rather new companies, where less information is available. Unlike PE investments where there is more public information for instance through Capital IQ or Bloomberg. In this vein, the geographic focus of the VC firm*

could also have a certain influence on this aspect. However, it is not only about data, but also how VCs work and handle that data." - Martin Eriksen, Investment Manager at Viking Ventures.

Johansson does not question the fact that data-driven solutions will be in widespread application in the future, however, he highlighted the importance to align data analytics with the strategic orientation and internal capabilities of the VC firms.

"*I believe that the VC industry is going to be more data-driven in the future. But there are different approaches on how the firm should implement that. If the firm wants to automate, it will also depend on the stage the VC is investing in - and also about the VCs capabilities. The key to success is to find the right balance between who you are as a VC, what capabilities does the firm have and what is the advantage in your focus market. In general, everyone will go more data-driven in the future, but there is no blueprint of how this should look like.*" - Erik Johansson, Investment Director at Volvo Group Venture Capital.

Corea also stated that the right data strategy is highly important when it comes to their implementation and application. In order to remain competitive, however, it is inevitable to engage with automated data analytics as it provides a competitive pace advantage. For him, however, human relations will always be a part of the VC investment procedure.

"*We believe that there are more VCs trying to implement data-driven solutions to improve the deal sourcing process. The way in which VCs are tackling this issue is different, and it depends on their respective strategies. I have the feeling that being data-driven is the only path that VCs can follow. Eventually, it is going to be very hard to not have any of these things internally. Some of these tools are quickly becoming the baseline. In 10 years you can be left out already, because you will not be able to see all the deals that you need and want to see. Of course, there will always be this networking and referrals playing an important role.*" - Francesco Corea, Research Lead at Balderton Capital.

Queiro identified the increasing usage of data-driven solutions in investment decision-making. However, he believes that a high level of automation will take some time to achieve and will not play a decisive role in the near future. In addition to that, given the fund's willingness to engage directly with the target company from an early point, Queiro expects increased ambition to improve its own fund's networking capabilities.

"*I believe that the complementarity aspect of data in order to support and complement VC investments does already exist. It is a continuum from more basic usage of data to fully automated investment decision processes. We are somewhere in between those two extremes, and we are moving more towards automation. However, I think that we are quite far from a world where automation plays a major role in the investment decision*

process. VCs, on the other hand, are trying to get close to founders early on and that is all about networking. This trend is then going in the opposite direction." - Francisco Queiro, Professor in Finance and Venture Capital at Nova SBE.

5. ANALYSIS

Based on the empirical results and the presented secondary literature from the previous chapters, the following analysis aims to critically discuss the results with the real-life conditions of the VC ecosystem. The chapter will start by mapping the surveyed VC actors regarding their level of data integration and automation in order to provide the reader with a better understanding of the used sample in this research. The mapping is important to illustrate the individual strengths of the sampled VC firms and how one has to interpret and classify the statements of the respective investors. Secondly, the industry best practices for extracting data sources are juxtapositioned. Based on the analysis of the data sources, the authors examined the different dimensions VC actors can be looking for within the sources. Having extracted all the relevant information, the organization and structuring of the information play a critical role for investors and are subsequently discussed. Finally, the way investment teams are extracting and handling the data has a significant influence on the level of unbiasedness in decision making and is therefore investigated as well.

5.1. Ranking and Mapping of the Sampled VC Investors

The following ranking and mapping can give the reader an indication of not only which VC funds are implementing industry best practices, but also in which way. The transcripts, interview notes and internal discussions with VGVC build the underlying explanatory basis for the ranking. Since VGVC has interacted with a significant amount of the VC players of the sample in the form of syndicated investments, the information on respective networking is partly based on their experience. Simultaneously, the conducted interviews provided the authors of this paper with insights about the level of data integration of the investors.[8] A ranking scale from zero to five for the two given characteristics then allows for an appropriate and sufficient differentiation between the firm and a following expressive way of visualization.

First of all, a fair amount of VC actors characterize themselves with a strategic focus on interpersonal communication with the potential target companies, other VC investors as well as accelerators. Even current portfolio companies and their respective professional network are sometimes exploited in order to identify deal leads. Spintop Ventures, Viking Venture, Adara Ventures and Brightly Ventures, for example, primarily rely on their networking capabilities and therefore can be grouped in the same investor class. In addition to that, they try to leverage their strong reputation and market previous investments that work as strong and positive market signals. This is also the case for more strategic investors like the corporate investments arms of the BMW Group and Porsche. VGVC, while using data-driven elements in their sourcing process, is also still focusing more on interpersonal connections in the industry to identify potential investment opportunities. Networking, therefore, represents the

[8] The conducted interviews had an average duration of 30 minutes and therefore the insights are limited in their expressiveness.

most important source of their deal origination strategy. Based on these findings, the authors of this paper allocated very few points to the level of data integration and rewarded the networking activities of the respective VC funds with an above-average score.[9]

Secondly, a group of technology specialists could be identified. Investors like Connetic Ventures and InReach Ventures differentiate themselves with a high level of automation and internally developed sourcing tools in the form of proprietary tech which builds the primary or even only source for initial deal mapping. At the same time, interpersonal interactions with the target company enjoys less focus in their early sourcing strategy. Another element that distinguishes them from other market participants is their technological competence and amount of resources they spend on data-driven solutions. While Connetic Ventures has developed an AI-algorithm to analyze teams and business models, InReach Ventures has automated their investment decision entirely. In addition, the British investment fund only employs experts with a technical or programming background. Therefore, these two VCs were given only a low score on networking, but relatively high points for data integration.[10]

Lastly, VC firms with a diverse and balanced sourcing strategy are trying to incorporate both elements in their operations. On the one hand, they identify themselves with a high level of data integration. The usage of respective target data plays a critical role in their initial deal mapping activities and is heavily used in the first step of the investment process. Data-pioneers like Balderton Capital, NGP Capital, Georgian and Nauta Capital have additionally employed concepts to automate the processing of relevant target data that is needed for financial decision making - which further differentiate themselves from competition in this regard. On the other hand, networking activities are equally weighted in their sourcing strategy as they try to leverage and continuously expand their strong and large professional network within the VC ecosystem. A significant share of their investments therefore also originates from external deal leads. VC actors like Creandum and 645 Ventures can thus also be classified as trade-off strategists who have a versatile portfolio of sources they equally seek to exploit in order to identify deals. Consequently, both characteristics can be assigned an above-average score.[11]

The corresponding ranking is shown in Table 5. The final score is the result of the addition of the points for networking efforts and usage of data.

[9] This group is labeled “A” in Figure 6.
[10] This group is labeled “B” in Figure 6.
[11] This group is labeled “C” in Figure 6.

VC	Type of VC	Location	Stage	Networking [0-5]	Usage of Data [0-5]	Score
Balderton Capital	IVC	U.K.	Seed and onwards	5	4.5	9.5
NGP Capital	CVC	U.S.A	Serie B	4.5	4.5	9
Nauta Capital	IVC	Spain	Seed - Series A	4.5	4.5	9
Creandum	IVC	Sweden	Pre-seed and onwards	5	3.5	8.5
Georgian Partners	IVC	Canada	Growth Stage	4	4.5	8.5
Spintop Ventures	IVC	Sweden	Early stage	5	2.5	7.5
Brightly Ventures	IVC	Sweden	Pre-seed and Seed	5	2.5	7.5
Volvo Group VC	CVC	Sweden	Series A and onwards	4	3	7
Viking Venture	IVC	Norway	3M EUR in Revenue	4	3	7
BMWi Ventures	CVC	U.S.A	Seed and onwards	4	3	7
Porsche Ventures	CVC	Germany	Seed and onwards	4	3	7
645 Ventures	IVC	U.S.A	Seed - Series A	3	3.5	6.5
Connetic Ventures	IVC	U.S.A	Pre-seed and Seed	2	4	6
InReach Ventures	IVC	U.K.	Pre-seed	0	5	5
Adara Ventures	IVC	Spain	Seed and onwards	4	1	5

Table 5. The ranking of all interviewed VC firms regarding their networking efforts and usage of data in deal origination.

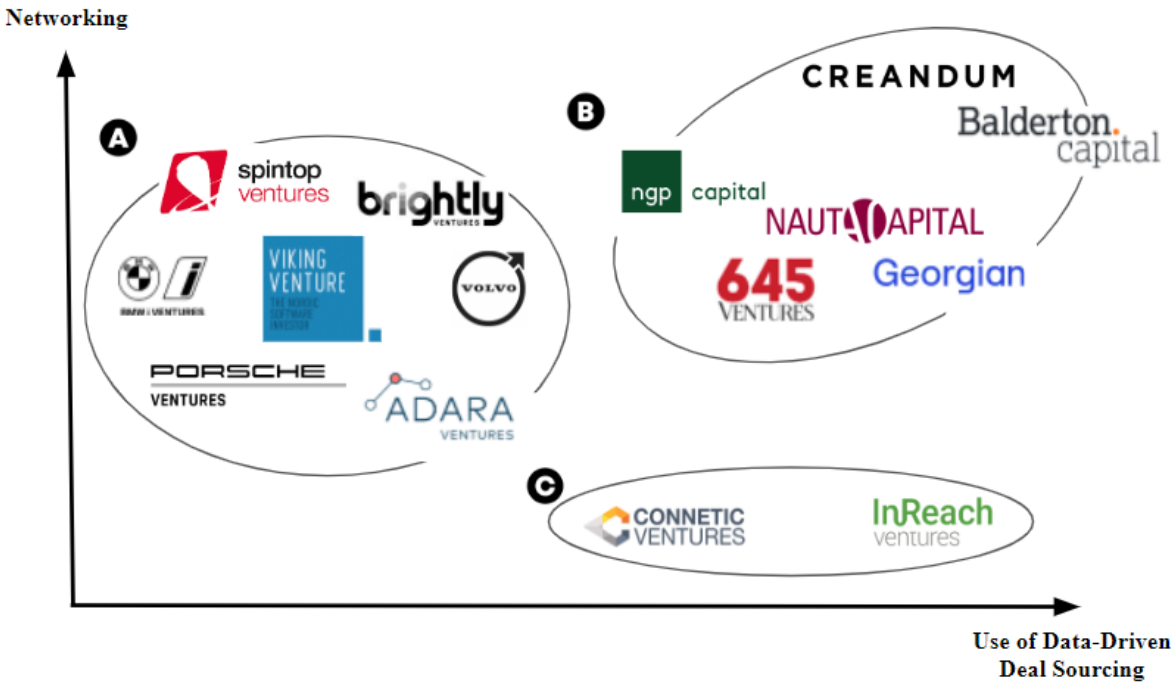


Figure 6. Graphical illustration of the ranking according to their networking efforts and level of data integration for deal sourcing activities

The ranking presented above then made it possible to group several VC players in the sample and present the funds in a graphical representation according to their strategies pursued. It must be noted that the authors of the paper cannot claim that one strategy is superior to another. Even within the proposed classes, actors can differ significantly in their operations, as VC funds perform differently successfully having in mind their resources and investor-specific expectation towards individual investments.

In addition to that, it is also worth mentioning that the delimitation of the proposed classes might not be as strict as depicted in Figure 6. Overlaps are very likely to occur and continuous reorientations and strategic iterations have to be expected in a highly dynamic and uncertain VC industry.

5.2. Analysis of Data Sources that VCs are using to source Deals

According to Loeb (2012), the potential that data-driven solutions can unlock in deal origination could be significant, as the application can result in a more efficient and resource-saving process that provides VC funds with competitive advantages over other market participants. Connetic Ventures highlighted the pursuit of funds to invest in disruptive technologies, however, the measures in order to identify those deals have not been changing over the last decades. Particularly in uncertain environments that by nature provide little information for decision-making due to limited disclosure requirements - such as in the VC industry - data analytics offer great advantages to extract target information. In line with this argument, Francesco Corea emphasized the real practical value-added of proactive and data-driven deal sourcing in helping Balderton Capital to have a more diverse range of target companies to consider when compiling the longlist. In the practice of VC operations, attempts are made to extract data nodes from various sources and link them to the respective company profile under investigation. It is possible to classify the different data sources into three segments that offer diverse benefits to its applicant - as can be seen in Figure 7.

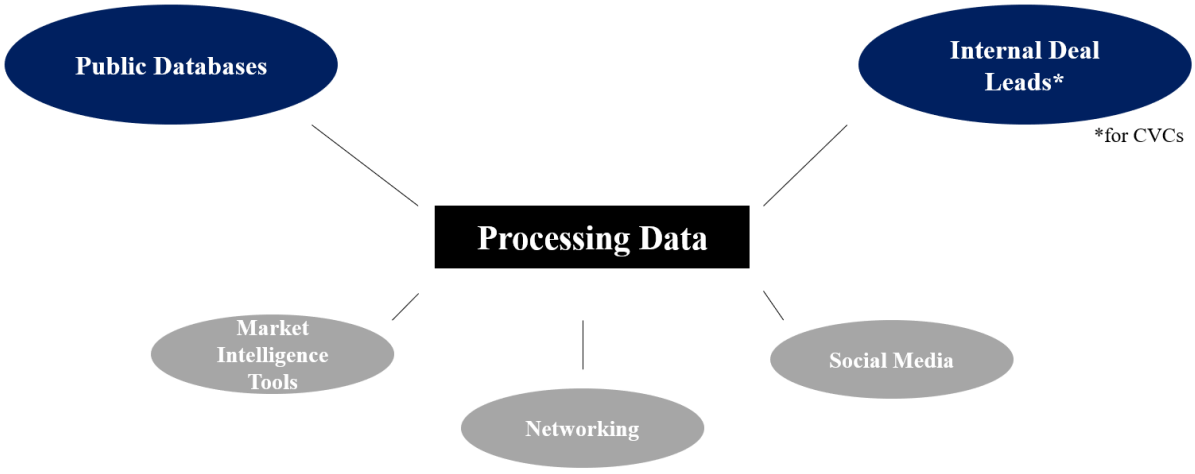


Figure 7: Illustration of different data sources within the deal origination of VC investments

5.2.1. Mining Data from Public Databases as a standard Data-Driven Tool

Public databases, for instance, provide a relatively fast way to private equity and venture capital information. Popular sources in this section include, among others, CB Insights, Pitchbook, and Dealroom (Splenda and Barnhart, 2017; Retterath and Braun, 2020). Creandum referred to the daily interaction with these sources having subscriptions not only to one single database but a broad selection of them. The advantages of having a diverse portfolio of data sources are also highlighted in academic research, emphasizing the wide scope of external industry information that can be extracted from public databases (Kirchhoff, Schiereck, and Mentz, 2006; Bhabra and Hossain, 2018; Retterath and Braun, 2020). It becomes apparent that public databases are in widespread use within the VC ecosystem, which was i.a. supported by NGP Capital. The global operating corporate venture arm of Nokia, therefore, noted that using public databases can, or even must, be considered a level playing field and thus does not provide VC players with a competitive advantage nowadays.

Although public databases are relatively popular, they can have significant shortcomings as stated by Robin Tech. According to the CEO and founder of Delphai, a fair amount of data from public databases has to be questioned because of untraceable sources, outdated information, or manual data extraction that can easily result in inaccuracy and errors. This is also supported by Retterath and Braun (2020), who explored the quality of data in some of the most popular databases among VC investors. Considering the strategic and financial value of CVC or M&A investments, Tech stressed the importance of high quality and reliable data for decision-making. In addition to that, Tech criticized the short and manually created company descriptions provided by some public databases which in most cases build the only foundation for the filtering process conducted by investors during the deal origination stage. Especially in the entrepreneurial sector, the operative orientation of respective firms can be subject to a multitude of strategic iterations and reorientations which would require an update of the company profile (Osterwalder and Pigneur, 2010). In addition, business models that emerge in today's interconnected markets can be able to satisfy a diverse catalog of customer requirements and thus are not only limited to a singular need as may be stated in the characterization provided by the database (Blank and Dorf 2012). Given these dynamics, company descriptions of young companies used and presented in public databases may be outdated and need to be at least both questioned and double-checked. When sourcing for strategic fits in deal origination, however, databases are usually filtered for long list compilations based on the specified portrayal of the venture. This systematic filtering process is particularly detrimental to VC funds that seek to identify and create strategic fits and synergies through their investments, as the respective company descriptions may not cover the VCs' operational focus and thus a significant number of investment opportunities may be missed (Ernst, Witt and Brachtendorf, 2005).

A list of all the public databases and market intelligence tools that were mentioned throughout the conduction of the interviews can be found in Table 6.

Data Source	**Mentioned by**	**Description**
Crunchbase	Most VCs	Crunchbase is a source of company information. The database depicts performance history of companies worldwide and is built on a data extracting approach that leverages information from a community of contributors, external network, and internal content teams.
Dealroom	Most VCs	Dealroom is a provider of data on young ventures, VCs and corporations with a strong focus on theEuropean space.
Pitchbook	Most VCs	PitchBook is a resource that delivers data, research, and technology. PitchBook covers the entire sector of the private capital market with its sub-categories VC, PE, and M&A transactions.
CB Insights	Most VCs	CB Insights provides market data, insights, future predictions, and workflow tools to investors that can be used for financial decision making.
Tracxn	NGP Capital	Tracxn provides comprehensive data about startups and young private companies.
Vainu	Viking Ventures	Vainu is a sales intelligence, prospecting and insights platform with a database of companies.
Capital IQ	Balderton Capital	Capital IQ allows access to comprehensive industry data, models, and workflow solutions in order to evaluate and manage counterparty credit risk, address operational and enterprise risk, perform entity due diligence and meet regulatory requirements.
Preqin	Balderton Capital	Preqin is a market intelligence tool to support investors at every stage of the investment cycle with information about i.a. previous investors, fundraising and research reports.
LinkedIn	Most VCs	LinkedIn is a social media platform allowing users to build, connect and maintain a professional network.
Y Combinator	Creandum, 645 Ventures	Y Combinator is a seed-stage venture firm and specializes in funding early stage startups, primarily in the software and web services arena.
Proff	Viking Ventures	Overview of i.a. key figures, official positions, addresses for all companies that deliver annual reports to the Swedish Companies Registration Office

Table 6. List of data sources used and mentioned by the interviewees.

5.2.2. Exploiting Data Sources from Internal Business Units

In some cases, the information from public databases is further complemented with leads and target information from internal sources such as other business units. This is particularly the case for CVC investment arms with strategic support from their parent organization (Rauser, 2003; Retterath and Braun, 2020). BMWi Ventures indicate that they seek to leverage the strategic relationships with the various business units of the BMW group and adopt the initial industry screening that they have conducted in their respective market of interest. Thereby the corporate venture arm can identify this connection as a source for potential deal leads. While VGVC also benefits from the in-depth industry knowledge of its parent company, it emphasizes the importance of creating synergies in the form of a mutual exchange of information as they provide other internal departments with newly acquired market insights. Titus and Anderson (2016) support this argument by suggesting a close collaboration between the CVC department and the parent organization to allow for maximum value creation for the parent company resulting from the risk-bearing investments. Furthermore, Dushnitsky and Lenox (2006) find empirical evidence for a higher value increase of the parent firm through more strategic-driven investments done by its CVC department. Both findings not only highlight the necessity of operational fits between CVCs and their parent organization but also point to the importance of strategic parental influences on their corporate investment arms by i.a. providing them with deal leads.

5.3. Analysis of the Dimensions that VCs are using to source Deals

Based on the empirical findings, it becomes apparent that VC firms utilize and monitor a variety of dimensions throughout the deal sourcing process. A diverse set of dimensions provides investors with information about investment opportunities within a particular geographic market or industry sector. By tracking and monitoring these KPIs, VCs can identify potential new deals and qualify them more accurately against their investor-specific criteria. Therefore, VCs should try to determine the dimensions that are most beneficial for them and ultimately provide the most relevant insights to anticipate the creation of new ventures, identify existing companies, and qualify leads.

As remarked by the interviewees, the unavailability of data or the potential low reliability of the existing target information are challenges that VC investors face when they want to integrate data sets into automated processes on a larger scale. For instance, VCs with an investment focus on early-stage ventures might be limited in their attempt to retrieve data on startups automatically due to the companies' short operational performance history, and ultimately the restricted existence of target information. This aspect has been highlighted by various interviewees and is also supported by industry reports (Fairview Capital, 2018). Given underlying databases that are characterized by low reliability and limited expressiveness, the transition to a more data-driven model would not provide any strategic benefits to VC firms. Retterath and Braun (2020) found that not all available data has the same level of accuracy and some information about companies can be outdated. The authors are thereby emphasizing the importance of selecting the most relevant combination of data sources depending on the investor-specific criteria.

5.3.1. Mapping of the most relevant Dimensions used in Data-Driven Deal Sourcing

The most highlighted dimensions by the interviewees can be classified into two different categories. On the one hand, one type of dimension refers to the characteristics of the entrepreneur or founding team which can be examined at an early stage. They do, however, require some form of interaction between the entrepreneur and the investor as the interpretation of the data can differ significantly among VCs - which ultimately makes it challenging to store the data in a standardized way. On the other hand, the dimensions refer to the venture and its performance. For this second type to be available for VC investors, some of the target companies should have had some previous market traction and product development. Thus, VCs will track these dimensions if they are VC funds investing in later stages. The decision of grouping the dimensions into two different types was based on the study and framework developed by Retterath and Braun (2020) who investigated the availability of startup-data in VC databases based on three dimensions: the company, the founders, and funding information. Table 7 summarizes these dimensions based on their type and the stage in which VCs tend to look at these dimensions to support their data-driven deal sourcing practices. Furthermore, there is a brief description of why VC investors consider this data relevant elements to identify and qualify deals.

It is worth noting that the table does not claim to incorporate all dimensions that VCs utilize to identify and qualify deals, but these are the most highlighted ones by the interviewees in this research. Furthermore, the assigned stage investment focus in Table 7 does not necessarily apply to every investor.[12] As previously mentioned, each sector and type of business model has to be considered unique and potentially requires a different deal sourcing approach. However, the above dimensions may be more or less relevant depending on the stage the fund is focussing on with its strategy. Based on the empirical findings, a successful data-driven deal origination strategy will include the combination and variety of dimensions, just looking at one dimension is not enough to identify and qualify the most promising investment opportunities.

[12] The investment focus in table 7 is indicated by the arrow on the left-hand side.

Early-stage

Dimensions	Type of Dimension	Purpose
Founders background	Founding team	Founders with specific experience in certain roles or academic backgrounds can demonstrate the ability of a founder to execute and develop business ideas.
Employment changes	Founding team	Talented employees from specific industries and companies leaving their former organizations can help VCs identify future entrepreneurs.
Internet buzz	Founding team / Company	News coverage, social media posts, and other announcements can help VCs identify new deals or potential entrepreneurs.
Previous investors	Company	Previous investors are seen as quality stamps by VC investors, helping them identify and qualify deals. These can sometimes include accelerators or angel investors, among others.
Growth in employees	Company	Companies hiring is understood as a signal of growth for VC investors.
Social Media traction	Company	Similar to internet buzz, when a company starts gaining followers and traction in their social media profiles can be understood as a signal of market acceptance and interaction with potential buyers.
Website and mobile data	Company	Shows if the company is being successful in their digital platforms, for instance, in terms of website visitors or sales made through their website.
Unit economics and other key metrics	Company	Key performance indicators can help VCs to identify and qualify deals based on how the company is developing their business operations, these can also be used as benchmarks.
Customer feedback and reviews	Company	Signalize the acceptance of the product or service by the target market in which the company operates.

Late-stage

Table 7. Summary and classification of dimensions used by VCs.

5.4. How VCs are organizing Data and Automating the Deal Sourcing Process

Another relevant element when studying how VC firms are utilizing data to improve their deal origination process is the level of automation implemented in their practices. Various interviewees highlighted how they are becoming more data-driven, not only by utilizing more data but also by automating the process. Based on the data collected through the interviews, it can be observed how the level of automation can be classified on a scale with different approaches. That means that there are different degrees to which VC firms are automating the processes of retrieving, structuring, organizing, and screening the available data.

As mentioned in previous sections, some of the most data-driven VC firms have internally developed data-driven solutions that support the firm in identifying and qualifying the most relevant startups within their markets of interest. Various of these data-driven VCs have developed machine learning algorithms, which, in combination with other tools, can

automatically retrieve, organize and present the data to the respective investment teams. Some other VC firms have also internally developed data-driven solutions that, without including machine learning models, still allow the firm to automate many tasks of their data-driven deal sourcing strategies. In line with these findings, it is worth noting that some VC firms were born as data-driven VCs and they have leveraged the power of AI and other data-driven solutions since the beginning of their operations. As can be seen in figure 8, firms with those characteristics from this sample are Connetic Ventures and InReach Ventures. On the other hand, there are highly data-driven VC firms that have a long performance history and built important networks around them - such as Balderton Capital and NGP Capital. These VCs have relied upon, and still do, more traditional approaches of deal sourcing when they started their investment activities. However, in the last years, they have realized which value data-driven solutions can provide and decided to start working on an increasing implementation as a crucial element of their deal origination process. Furthermore, these firms enjoy a strong track record and robust reputation within the VC space and the entrepreneurial ecosystem, which support them in keeping a highly dynamic deal flow.

5.4.1. Leveraging the Power of Proprietary Algorithms

In order to outperform competitors, further value-creating alternatives to enhance the data flow in deal origination are in use. The so-called automated proprietary algorithms (PAs), for instance, are highly beneficial to investors as they represent a form of machine learning that is tailored towards the respective strategic focus of investors and their investment activity - especially when the investors have defined unique investment criteria for their filtering process. Such algorithms are able to provide investors with a comprehensive overview of the operational health of young companies and identify and predict future success (Holland, 1992; Kurlandski and Bloodgood, 2022). One of the main advantages stated by the interviewees is the ability of internal PAs to scrape data from hundreds to thousands of data sources and present the retrieved information in a structured and user-friendly manner to the investment team. The outputs of the systems received by the investment team can differ. For instance, some PAs provide the investment team with a long list of relevant companies according to the VC firm's investment criteria, while others supply the investment team with warnings or notifications about relevant news regarding companies they have previously identified. VC firms also have incorporated scoring models in their internal data-driven systems, which provide the investment team with a score for each of the companies identified through the tool based on certain predetermined parameters. That helps investment teams to set priorities and focus on the most relevant opportunities. In other cases, some VCs are implementing machine learning models that are continuously fed with feedback loops to help the algorithms become more tailored towards the investor-specific criteria and ultimately provide more accurate investment opportunities based on the available data. In line with the empirical findings, Weibl and Hess (2019) found that VC firms utilizing data-driven solutions in their deal origination and screening stages experienced both short and long-term benefits. Therefore, the study conducted by Weibl and Hess (2019) supports the findings of the present paper by also stating

that data-driven approaches provide competitive advantages and value-added to investment firms.[13]

Both Balderton Capital and NGP Capital employ internally developed PAs for data extraction to validate their investor-specific investment criteria. In these cases, internal programming and data analytic teams are working on the maintenance and functionality of the algorithms. NGP Capital noted that they can thereby screen for key signals such as the target company's current traction, the attractiveness of the market under investigation, relevant financial performance, and information about the founding team. Viking Venture, even though they do not deploy such tools, mentioned another value-adding element of PAs. The Norwegian VC stated that relevant companies are occasionally not listed in traditional public databases as i.a. Pitchbook as they mainly focus on companies with recent funding rounds. The same VC, however, defines their preferred target company as a venture that generates around 3 million SEK of revenues only through internal bootstrapping excluding external investors. While the PA often extracts company information from public databases as well, it also connects data from other sources such as press releases, company websites, and conferences and thereby allows for more investor-specific sourcing strategies (CB Insights, 2021). That benefits investors in that they can reduce the time spent extracting data from multiple sources themselves as well as get exactly the target information they need for decision making. Simultaneously, this approach ensures that more relevant investment opportunities are identified, as the automation of processes can escalate, reaching and handling larger data sets than a human performing these tasks. Thus, by using PAs, VC investors can differentiate themselves from competition not only in terms of time savings but also regarding the qualification of potential deals, and thereby overcome the problem of competitive parity in the use of data sources.

When PAs are somewhat laid out more publicly to serve a wide audience of customers, general dimensions are being extracted, collected and visualized in a comprehensive way. Delphai, for example, has developed several PAs that collect updated data from over 15 000 sources worldwide including not only financial statements and investor portfolios but also inconspicuous signals such as recent job openings. By doing this in an automated way and offering full-text search, the Berlin-based company enables a significant increase in eligible target companies to consider within deal sourcing and thereby reduces the probability of missing profitable deals.

[13] In the conducted study, Weibl and Hess (2019) investigated the deal sourcing strategy and application of data-driven solutions of 13 VC investors in the German VC market.

5.4.2 Implementing CRM Software to support Data-Driven Deal Sourcing.

Based on the primary data collected through the interviews, various interviewees described the role that CRMs are playing in their data-driven deal origination practices. Therefore, even though they are usually not considered data-driven tools per sé, CRMs are closely related to the main data-driven tools that VC firms have in place. Similar to what Weibl and Hess (2019) found in their research with German VCs, the interviewees of the present research also highlighted that CRMs are relevant elements in supporting VC investors in storing and organizing the data to make more efficient investments decisions. As described by Weibl and Hess (2019), CRMs are the backbone of VC IT systems. The typical flow of information during an investment decision can be seen in Figure 8.

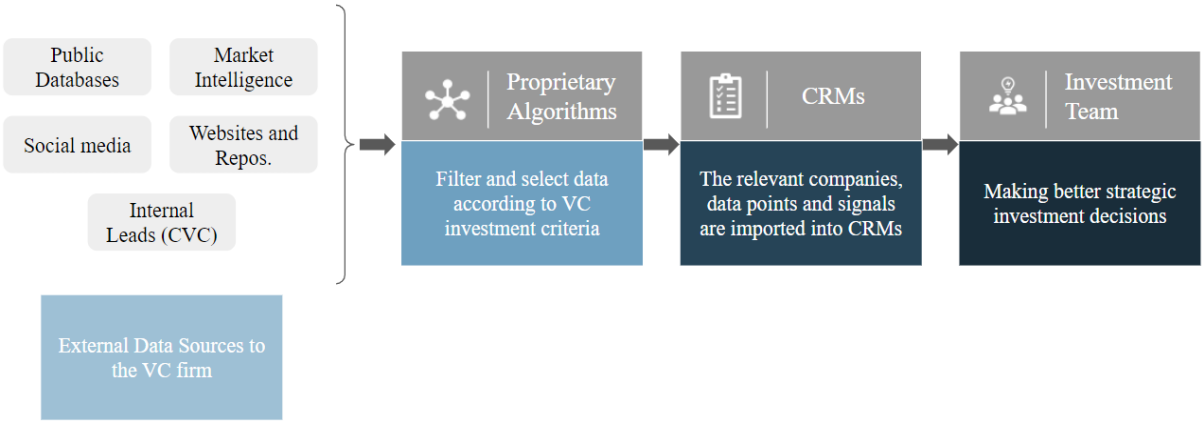


Figure 8. Overview of the flow of information within data-driven deal sourcing

The VCs interviewed in the present investigation remarked that one of the most relevant attributes of CRMs is choosing a CRM software that allows the firm to build integrations with their existing systems. By doing that, VC investors can also automate part of their processes within the deal origination and screening stages. Therefore, in line with Tanner et al. (2005), the empirical findings of this paper also highlight the importance of choosing a CRM that adds value to the VC firm and their existing operational systems to ultimately work more efficiently and data-driven. That is further supported by Reinartz, Krafft and Hoyer (2004), who found that selecting the right CRM tool can significantly contribute to the long-term results of the investment firm.

To summarize the investment process and show how data-driven solutions support the deal origination and screening phases, the authors of the research have condensed the different steps in figure 8, based on the empirical findings previously described.

5.5. How Data can contribute to making more Unbiased Decisions

The importance of unbiased decision-making - especially in the early stages of VC investments - is becoming increasingly prominent. Weintraub's (2002) traditional theory of neoclassical economics and its assumption about complete rationality of investment behavior is difficult to sustain in today's complex market environments. Several characteristics of the target company and its founding team can significantly influence financial decision-making, regardless of the respective business model at hand. Ullah (2015) argues that only with a deep understanding of the potential biases and awareness of the underlying dynamics of decision making higher levels of rationality can be maintained. Therefore, financial investors should understand the different layers that are to consider in and linked to venture capital investment decisions. The original interview guideline, however, was not necessarily designed to explore this development. Nevertheless, the magnitude of this phenomenon within the interview process was significant as many data-driven pioneers used this term to market their respective investment procedures. Therefore the authors of this research were able to identify unbiased decision making as a separate theme in their analysis

5.5.1 Identifying attractive Business Model outside of Entrepreneurial Hubs

For instance, the geographical location of the entrepreneur or startup plays a critical role in the financing process for a startup. Most VC investors focus their investment and sourcing activities primarily on entrepreneurial hubs, as they expect them to generate knowledge transfers, allow for stronger networking and depict greater attractiveness for qualified employees across startups (Galope, 2014). Bloomberg Data (2022) is supporting this argument by stating that the Silicon Valley region remains the area for highest VC investor dynamic in terms of money raised and closed deals. Connetic Ventures, however, questions this approach by noting that it is important to focus more on the value creation through the created business model presented by the target company rather than placing too much weight on the firm’s current location. Empirical evidence is additionally indicating that sourcing activities outside of geographic VC centers can result in top investment performance (Teten and Farmer, 2010). The Kentucky based VC identifies the entire north-american VC ecosystem as financially strong - particularly in the Silicon Valley area and New York. However, they further observe that startups located outside of the entrepreneurial hubs might have promising business models with a diverse team, yet diminished access to venture capital sources. Their machine learning application “Wendal” allows the firm to consider young ventures in a more unbiased way regardless of their location. In line with the findings of proprietary methodological research, Connetic Ventures deployed a standardized and automated process to assess the business model isolated from the current location of the target company. Agarwal et al. (2022) proposed that this technique ensures lower irrationality in financial decision making and are even able to show statistical significance for this argument. Furthermore, the authors highlighted the possibility to automate the procedure in the form of a computerized investment decision making system.

5.5.2. Coping with Implicit Biases in VC investments

Another element to consider are the unconscious biases of the decision-maker within the deal origination process.[14] The context of a business model pitch is in most cases the first time that the potential investor and the founding team meet face to face. It takes place relatively early in a possible collaboration and is followed by a filtering decision (Fried and Hisrich, 1994). The setting of a pitch therefore opens up a multitude of opportunities for prejudices inferences based on race, gender, age or sexual orientation of the individuals in the founding team. Investment decisions based on individual instinct exacerbate the problem of diversity in the VC ecosystem. In this context, the application of data-driven solutions can help to significantly reduce implicit biases in venture capital decisions, as a standardized and automated assessment design allows for a stronger focus on the behavior of the founding team and provides every applicant with an equal chance of receiving funding (Fairview Capital, 2018). Connetic Ventures does link their operational strategy to this insight by weighting the founders as the most important element in their review process. In doing so, they assign members of the target team a critical role in the overall investment decision. However, Connetic Ventures does not believe that school, experience, background, or demographics are unbiased criteria to effectively analyze startup founders.

While data analytics can help make the investment process more impartial, Blomquist emphasized the fact that even solely data-driven decisions are not entirely unsusceptible against biases. The operating partner at Creandum highlighted the importance of the underlying sample, training data and the feedback given to the respectives algorithms in order to allow for effective and impartial analysis of the data set that is processed by automated procedures. The ubiquity of social identifiers and their correlations to concrete behaviors can feed the system with certain information and therefore train the algorithm for actual discrimination. Adara Ventures is also aware of this concern and further claims that the origin of bias can also be rooted in the team responsible for decision-making. Blomquist noted that moving from qualitative to quantitative data in the decision-making process could indeed ensure a higher degree of impartiality. For him, however, it cannot be considered a generally applicable and final solution as it is very important to have a diverse investment team that is highly aware of the biases they face and are susceptible to. Thus, the importance of a diverse decision making unit as well as the awareness of their own respective biases are critical elements to the selection process.

[14] The so-called implicit biases are automatically generated associations about certain persons or groups of people (Pritlove et al., 2019). The authors also stated that human beings feel more comfortable with groups of people that are similar to one's own characteristics and behaviors which could significantly influence the decision making process of whether to invest or not.

6. DISCUSSION

As discussed above, the intensity with which data-driven tools are being used by VC actors in the future will be partly influenced by the focus investment stage as well as the current state of the venture. As indicated by Tyebjee and Bruno (1984), investors can only rely on little historical performance data when evaluating early-stage ventures. Having the practical industry expertise, NGP Capital supports this argument by stating that one of the key differences between early- and later-stage investors is the availability of data about the targets - which ultimately makes it more challenging to automate certain data extraction activities in deal sourcing. It becomes apparent that whenever traditional KPIs such as financial performance and previous funding rounds are not existing or publicly available, early-stage VC investors tend to consider other criteria and signals that allow for inferences for respective success predictions. Balderton Capital highlighted the diversity of dimensions they consider - including social media traction, growth in employees and employment changes of respective entrepreneurs - in order to forecast desired developments of the early-stage targets.

However, the potential impact of data-driven solutions in deal origination is viewed differently by academics and those actively involved in VC deals and industry proximity. While the interviewees with an academic background can identify a growing trend towards a more frequent use of automated applications in the initial investment phases, they question whether the data-driven approach can take an equally relevant part within deal sourcing. According to Queiro, the information required for VCs' decision-making is difficult to express in quantitative measures which makes it difficult to feed algorithms properly. Qualitative elements such as passion, resilience, and determination of the entrepreneur are also crucial and necessary to consider, however, these are difficult to integrate into automated solutions (Fowle, 2019).

The generic usage of data-driven solutions in future deal sourcing activities, on the other hand, is what every interviewee and previous studies can agree on and therefore almost undeniable. Corea from Balderton Capital believes that the industry will be even more data-driven and automated in the future and identifies today's approaches including the level of automation and data integration as a starting point only. Illa from Nauta Capital further stated that he expects data analytic tools to take over market intelligence functions which allow for an overall more resource-saving process. Another important aspect that influences the deal strategy of VCs is the funds current recognition on the market which is highly shaped by previous activities. Creandum, for instance, is highlighting its market position within the entire VC ecosystem. Due to their previous participation in e.g. Spotify, Klarna, and TradeRepublic, the Stockholm-based venture fund enjoys a strong market reputation and therefore has a relatively high share of incoming funding requests. These conditions can significantly change the intentions and ambitions within the deal sourcing. Well-known and high-performing VC funds could thus take a more passive role instead of proactively seeking deals themselves (Weibl and Hess, 2019). Therefore, the market position and brand reputation

of VC firms are very likely to influence the level of data integration and automation within deal mapping activities.

Simultaneously, Fellman from NGP Capital mentioned the challenge of not only initiating contact with the target company but also convincing them to participate in a strategic partnership that can result in investments. Consistent with this argument, Eriksen from Viking Venture noted greater bargaining power for promising companies due not only to a wide range of possible funding sources, but also to the ease of accessing them in today's globalized and interconnected market structures. The increasing negotiating power of auspicious targets therefore can be identified as another trend. In addition to that, Fanelli sees a declining seed asset class and argues that network-related software is very likely to be a focal issue of VC funds on the tactical and strategic horizon.

On the other hand, VC investments are also affected by elements that do not have a well-explored research stream as of today. For instance, the increasing prominence of unbiased decision making, can play a critical role in order to enable an equitable distribution of a fund's financial resources. Simultaneously, higher levels of investor rationality can help to extend the current geographical focus of VCs and thereby result in the identification of more attractive and profitable business models. Data-driven solutions can support the process of becoming more unbiased through a standardized evaluation model for founding teams and the respective business models at hand.

7. CONCLUSION

This research intended to examine the application, motives and benefits of data-driven solutions within deal origination activities of VC investments. A significant amount of VC investments fail to realize the synergy-related value expectations and consequently do not generate the desired increase in company value. In this context, the increasing value of information in a dynamic and digitized VC ecosystem has been highlighted by the authors. The deal origination phase plays a critical role in this process as it is relatively prone to suboptimal resource allocation and only a few investments placed by VC funds generate the desired positive financial returns. To identify more deals that actually meet investor-specific criteria, VCs can profit from adopting data-driven tools. By conducting 20 interviews with knowledgeable actors within the VC ecosystem and receiving industry insights from VGVC, the authors of this paper identified and analyzed currently applied industry-best practices in data-driven deal sourcing. A comprehensive overview of used data strategies deployed by top-performing VCs presented in this paper can help to successfully structure an in-house transformation towards more data-driven solutions. The application of data analytics is then not only limited to data mining activities but also to structure and organize data sets more effectively which helps to transform single data points into more tangible information for financial decision making. Lastly, this paper also outlines the already existing and anticipated increasing relevance of data-driven solutions in future deal origination activities.

It is worth noting that there is no monocausal answer to the right application of data-driven solutions within deal origination. The wide range of data analytics, due to today's technological capabilities, allows investors to consider significantly higher quantities of target companies for long-list compilation. Information stocks from public databases should be complemented with deal leads from the professional network or - in the case of CVCs - from other internal business units. Corporate investment arms like VGVC are already exploiting this data source to enrich their information pool for decision making. In addition, investors can rely on automation and standardization to place investments with a higher degree of rationality so that they are more focused on the business model at hand and do not fall for implicit biases. In this vein, PAs support investment teams in qualifying the investment opportunities in a time-saving manner by automatically checking them against investor-specific criteria. It is required and highly important, however, to tailor the technology being used to the investors' own preferences. With the availability of more data points and a certain degree of automation, VCs have the ability to reduce the initial long list of target companies more efficiently.

Another key finding of this research is that the human component anchored in the traditional sourcing approach is unlikely to disappear entirely as the VC ecosystem is still a very people-intensive business. VC firms need to be aware of their capabilities and market position to choose the right balance between networking and data-drivenness. In order to formulate a successful future strategic direction, this study shows that by identifying an adequate trade-off between the two approaches presented, VC funds are able to leverage the advantages of

both concepts while avoiding potential pitfalls. The right balance will be key to either gaining or maintaining a robust market position in the industry and ensuring that one is operating on the cutting edge of market developments.

The paper contributes to the existing literature by highlighting how VC investors develop and apply data-driven approaches to improve deal sourcing practices. Furthermore, the study includes the perspective of VC actors regarding future trends that are shaping, or could eventually transform, the deal sourcing process of VC investments by the application of data-driven tools. Another relevant contribution of the paper is the impact of investors' rationality and the degree of unbiasedness in investment decision-making, which plays a crucial role in the intersection of data-driven solutions and the VC industry. Nonetheless, this topic remains understudied in the literature, and even though this study provides novel knowledge and findings to the field, an investigation of the actual influence of implicit bias and impartiality in investments in the form of a qualitative study could help to reach a scenario in which each entrepreneur has an equal chance of obtaining capital resources. It is also recommended to conduct a quantitative study that follows the entire transaction cycle in order to evaluate the influence of data-driven solutions on the overall success rate of VC investments. Therefore, future research is required to develop a deeper understanding of the underlying dynamics and consequences of using data-driven solutions within VC deal sourcing processes.

REFERENCES

Agarwal, A., Agarwal, H. and Agarwal, N. (2022). Fairness Score and Process Standardization: Framework for Fairness Certification in Artificial Intelligence Systems. *arXiv preprint arXiv:2201.06952*.

Arroyo, J. et al. (2019). Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments. *IEEE access*, 7, pp.124233–124243.

Atlassian. (2022). The #1 software development tool used by agile teams. https://www.atlassian.com/software/jira. [Retrieved: 28.02.2022]

Bell, E., Bryman, A. & Harley, B. (2019). *Business research methods* Fifth edition Emma Bell, Alan Bryman, Bill Harley.

Bender, M. (2011). *Spatial proximity in venture capital financing: A theoretical and empirical analysis of Germany.* Springer Science & Business Media.

Benson, D., & Ziedonis, R. H. (2009). Corporate Venture Capital as a Window on New Technologies: Implications for the Performance of Corporate Investors When Acquiring Startups. *Organization Science*, *20*(2), 329–351.

Bertoni, F., Colombo, M.G. and Grilli, L. (2013). Venture capital investor type and the growth mode of new technology-based firms. *Small Business Economics*, 40(3), pp.527-552.

Bhabra, H.S. and Hossain, A.T. (2018). Did the Sarbanes-Oxley Act result in a strategic shift in M&A motives?. *Managerial Finance*.

Bhakdi, J. (2013). Quantitative VC: A New Way to Growth. *The journal of private equity*, 17(1), pp.14–28.

Blank, S. and Dorf, B. (2012). The startup owner's manual: The step-by-step guide for building a great company. K & S Ranch. *Inc, California. PubMed Abstract OpenURL*.

Bloomberg. (2022). Startups Raked In $621 Billion in 2021, Shattering Funding Records. https://www.bloomberg.com/news/articles/2022-01-12/startups-raked-in-621-billion-in-2021-shattering-funding-records. [Retrieved: 30-05-2022]

Bottazzi, L. and Da Rin, M. (2002). Venture capital in Europe and the financing of innovative companies. *Economic policy*, *17*(34), pp.229-270.

Capron, L. & Shen, J.-C. (2007). Acquisitions of private vs. public firms: Private information, target selection, and acquirer returns. *Strategic management journal*, 28(9), pp.891–911.

CB Insights. (2021). The 2021 Mid-Year Global CVC Report. https://www.cbinsights.com/research/report/corporate-venture-capital-trends-h1-2021/. [Retrieved: 03.02.2022]

CB Insights. (2022). Without data, you're just another person with an opinion. https://www.cbinsights.com. [Retrieved: 08.02.2022]

Cerra, A., Easterwood, K., & Power, J. (2012). Transforming business: Big data, mobility, and globalization. John Wiley & Sons.

Chemmanur, T. J., Loutskina, E., & Tian, X. (2014). Corporate Venture Capital, Value Creation, and Innovation. The Review of Financial Studies, 27(8), 2434–2473. http://www.jstor.org/stable/24465662

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188.

Cumming, D., Fleming, G. & Schwienbacher, A. (2005). Liquidity Risk and Venture Capital Finance. *Financial management*, 34(4), pp.77–105.

Davidsson, P., Achtenhagen, L. and Naldi, L. (2010). *Small firm growth*. Now Publishers Inc.

Dealroom. (2022a). A Beginner's Guide to M&A Deal Origination. https://dealroom.net/blog/a-beginners-guide-to-m-a-deal-origination. [Retrieved: 04.02.2022]

Dealroom. (2022b). Discover the world's most promising companies and tech ecosystems. https://dealroom.co. [Retrieved: 08.02.2022]

Deloitte. (2022). iDeal – defining M&A analytics. https://www2.deloitte.com/us/en/pages/mergers-and-acquisitions/solutions/mergers-and-acquisitions-data-analytics-ideal.html. [Retrieved: 07.02.2022]

Diener, E. and Crandall, R. (1978). *Ethics in social and behavioral research*. U Chicago Press.

Doleschel, J., Manu A. (2021). *Scarring effects of the COVID-19 pandemic on the global economy – reviewing recent evidence.* . https://www.ecb.europa.eu/pub/economic-bulletin/focus/2021/html/ecb.ebbox202107_01~f8314090a4.en.html [Retrieved: 28.05.2022]

Dushnitsky, G. and Lenox, M.J. (2006). When does corporate venture capital investment create firm value?. *Journal of business venturing*, *21*(6), pp.753-772.

Easterby-Smith, M., & Junshan, G. (2018). Vision, mechanism and logic: understanding the strategic investment decision making process. In *Management Issues in China: Volume I* (pp. 106-126). Routledge.

Elan, S.L. and Goodrich, M. (2010). Behavioral patterns and pitfalls of US investors. In *Federal Research Division, Library of Congress* (p. 01).

Engel, D. (2002). The impact of venture capital on firm growth: an empirical investigation. https://madoc.bib.uni-mannheim.de/442/1/dp0202.pdf. [Retrieved: 03.02.2022]

Ernst, H., Witt, P., & Brachtendorf, G. (2005). Corporate venture capital as a strategy for external innovation: an exploratory empirical study. r&d Management, 35(3), 233-242.69

Fairview Capital. (2018). The future of data-driven venture firms. Fairview Capital Partners Inc.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417.

Fanning, K., & Drogt, E. (2014). Big data: New opportunities for M&A. *Journal of Corporate Accounting & Finance*, 25(2), 27-34.

Faria, A.P. and Barbosa, N. (2014). Does venture capital really foster innovation?. *Economics Letters*, *122*(2), pp.129-131.

Florida, R. & Smith, Jr., Donald F. (1990). Venture capital, innovation, and economic development. *Economic development quarterly*, 4(4), pp.345–360.

Fowle, M. (2019). The entrepreneurial dream: Happiness, depression and freedom. *International Review of Entrepreneurship*, *17*(3), pp.305-322.

Fried, V. H., & Hisrich, R. D. (1994). Toward a Model of Venture Capital Investment Decision Making. *Financial Management*, *23*(3), 28–37.

Galope, R.V. (2014). What Types of Start-ups Receive Funding from the Small Business Innovation Research (SBIR) Program?: Evidence from the Kauffman Firm Survey. *Journal of technology management & innovation*, *9*(2), pp.17-28.

Gompers, P.A. (1994). The rise and fall of venture capital. *Business and Economic History*, pp.1-26.

Gompers, P.A. et al. (2020). How do venture capitalists make decisions? *Journal of financial economics*, 135(1), pp.169–190.

Guarda, T., Augusto, M. F., & Lopes, I. (2019, June). Geographic market intelligence as a competitive advantage. In 2019 14th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-5). IEEE.

Harvard Business Review Analytic Services. (2021). The Evolution of Decision Making: How Leading Organizations Are Adopting a Data-Driven Culture. https://hbr.org/resources/pdfs/tools/17568_HBR_SAS%20Report_webview.pdf. [Retrieved: 09.01.2022]

Hellmann, T. (2002). A theory of strategic venture investing. *Journal of financial economics*, 64(2), pp.285-314.

Holland, C., Lockett, G., & Blackman, I. (1992). Planning for electronic data interchange. *Strategic Management Journal*, 13(7), 539-550.

Intel Capital. (2022). Censys Completes $35 Million Series B Funding Round Led by Intel Capital - Intel Capital. https://www.intelcapital.com/censys-completes-35-million-series-b-funding-round-led-by-intel-capital/. [Retrieved: 03.02.2022]

Intel Capital Stats. (2022). A force multiplier for early-stage start-ups. https://www.intelcapital.com/#stats. [Retrieved: 03.02.2022]

Isaksson, A. (2006). *Studies on the venture capital process* (Doctoral dissertation, Handelshögskolan vid Umeå universitet).

Jain, C. (2018). *Artificial intelligence in venture capital industry: opportunities and risks* (Doctoral dissertation, Massachusetts Institute of Technology).

Kirchhoff, M., Schiereck, D. and Mentz, M. (2006). Market valuation of real estate finance mergers: a note. *Journal of Property Investment & Finance*.

Kleijn, M., Advisory Council for Science and Technology Policy (2011). Capital opportunities: Smart money for ambitious entrepreneurs. Derived from http://www.awt.nl/publicaties/kapitalekansen/item79.

Knopf, J.W. (2006). Doing a literature review. *PS: Political Science & Politics*, *39*(1), pp.127-132.

Koratana, A., McLetchie, J., and Van Ingelgem, K. (2019). M&A success, powered by advanced analytics. https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/m-and-a-success-powered-by-advanced-analytics. [Retrieved: 08.02.2022]

Kortum, S. & Lerner, J. (2001). Does venture capital spur innovation? In *Entrepreneurial inputs and outcomes: New studies of entrepreneurship in the United States*. Emerald Group Publishing Limited, pp. 1–44.

Kurlandski, L. and Bloodgood, M. (2022). Impact of Stop Sets on Stopping Active Learning for Text Classification. *arXiv preprint arXiv:2201.05460*.

LeCompte, M.D. and Goetz, J.P. (1982). Problems of reliability and validity in ethnographic research. *Review of educational research*, *52*(1), pp.31-60.

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care*, *4*(3), p.324.

Lintner, J. (1972): The Market Price of Risk, Size of Market and Investor's Risk Aversion: A Reply, in: *The Review of Economics and Statistics*, Vol. 54, No.2, S. 206-208

Loeb, S. (2012). In light of the limitations of data-driven decision making. *Education Finance and Policy*, *7*(1), pp.1-7.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91.

Metrick, A. & Yasuda, A. (2021). *Venture capital & the finance of innovation* Third Edition, Wiley

Napp, J.J. & Minshall, T. (2011). Corporate Venture Capital Investments for Enhancing Innovation: Challenges and Solutions. *Research technology management,* 54(2), pp.27–36.

NVCA. n.d. What is Venture Capital?. https://nvca.org/about-us/what-is-vc/#toggle-id-2. [Retrieved: 30.05.2022]

Osterwalder, A. and Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers* (Vol. 1). John Wiley & Sons.

Oxford University. (2022). Risk and Repeat: How the Venture Capital Ecosystem Drives Entrepreneurship and Innovation. https://www.law.ox.ac.uk/business-law-blog/blog/2022/01/risk-and-repeat-how-venture-capital-ecosystem-drives-entrepreneurship. [Retrieved: 07.02.2022]

Povel, P., & Sertsios, G. (2014). Getting to know each other: The role of toeholds in acquisitions. *Journal of Corporate Finance*, 26, S. 201-224.

Pritlove, C., Juando-Prats, C., Ala-Leppilampi, K., & Parsons, J. A. (2019). The good, the bad, and the ugly of implicit bias. The Lancet, 393(10171), 502-504.

Pipedrive. (2022). The all-in-one sales platform for growing revenue https://www.pipedrive.com/sv. [Retrieved: 28.02.2022]

Pitchbook. (2022). VC, PE and M&A—done better. https://pitchbook.com. [Retrieved: 08.02.2022]

Rauser, I. (2003). Value Added of Corporate Venture Capital: How do CVC Units Benefit from their Organizational Core?-Case Studies of the Relationship Between CVC-Unit and Business Units.

Rajan, R., Servaes, H., & Zingales, L. (2000). The Cost of Diversity: The Diversification Discount and Inefficient Investment. *The Journal of Finance*, *55*(1), 35–80.

Ravid, S. A., & Spiegel, M. (1999): Toehold strategies, takeover laws and rival bidders. *Journal of Banking & Finance*, 23(8), S. 1219-1242.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, *41*(3), 293–305.

Retterath, A. and Braun, R. (2020). Benchmarking venture capital databases. *Available at SSRN 3706108*.

Sandberg, W. R., Schweiger, D. M., & Hofer, C. W. (1989). The use of verbal protocols in determining venture capitalists' decision processes. Entrepreneurship theory and practice, 13(2), 8-20.

Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed- method studies. Research in nursing & health, 23(3), 246-255.

Sharifzadeh, A., & Walz, U. (2012). Syndication of VC Investments, Governance, and Contract Design. In *The Oxford handbook of venture capital.*

Silva, J. (2004). Venture capitalists' decision-making in small equity markets: a case study using participant observation. Venture capital, 6(2-3), 125-145.

Splenda, R. and Barnhart, M.E. (2017). Venturing forth: Venture capital and private equity information in academic business libraries. *Journal of Business & Finance Librarianship*, *22*(3-4), pp.208-221.

Tanner, J. E., Ahearne, M., Leigh, T. W., Mason, C. H., & Moncrief, W. C. (2005). CRM in Sales-Intensive Organizations: A Review and Future Directions. *The Journal of Personal Selling and Sales Management*, *25*(2), 169–180.

Teten, D., & Farmer, C. (2010). Where Are the Deals? Private Equity and Venture Capital Funds' Best Practices in Sourcing New Investments. *The Journal of Private Equity*, *14*(1), 32–52.

Titus, V., & Anderson, B.S. (2018). Firm Structure and Environment as Contingencies to the Corporate Venture Capital-Parent Firm Value Relationship. *Entrepreneurship Theory and Practice, 42*, 498 - 522.

Trello. (2022). Trello helps teams move work forward. https://trello.com/en/home. [Retrieved: 28.02.2022]

Tyebjee, T. T., & Bruno, A. V. (1984). A Model of Venture Capitalist Investment Activity. *Management Science*, *30*(9), 1051–1066.

Ullah, S. (2015). An empirical study of illusion of control and self-serving attribution bias, impact on investor's decision making: moderating role of financial literacy. *Research Journal of Finance and Accounting*, *6*(19), pp.109-118.

Van de Vrande, V., Lemmens, C., & Vanhaverbeke, W. (2006). Choosing governance modes for external technology sourcing. *R&D Management*, 36(3), 347-363.

Vereshchagina, G. and Hopenhayn, H.A. (2009). Risk taking by entrepreneurs. *The American Economic Review*, *99*(5), pp.1808-30.

Volvo Group. (2021). Volvo and Upstream Security collaborate for safer connected vehicles. https://www.volvogroup.com/en/news-and-media/news/2021/nov/volvo-and-upstream-security-collaborate-for-safer-connected-vehicles.html?sc_cid=soco-8hgtLGsa. [Retrieved: 09.01.2022]

Volvo Group. (2022). Volvo Group Venture Capital. https://www.volvogroup.com/en/about-us/organization/other-entities/volvo-group-venture-capital.html#:~:text=Volvo%20Group%20Venture%20Capital%20AB%20was%20founded%20in,companies%20participating%20in%20the%20transformation%20of%20our%20industry. [Retrieved: 10.01.2022]

Weibl, J. and Hess, T. (2019). January. Finding the Next Unicorn: When Big Data Meets Venture Capital. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Weintraub, E. R. (2002). Neoclassical economics. The concise encyclopedia of economics, 1, 1.

Woolridge, J. R., & Snow, C. C. (1990). Stock market reaction to strategic investment decisions. *Strategic management journal*, 11(5), 353-363.

Xu, R., Chen, H., and Zhao, L.J., *Predicting Corporate Venture Capital Investment*, Proceedings of the 38th International Conference on Information Systems, 2017, pp.1-10.

Zetsche, D. A., Buckley, R. P., Arner, D. W., & Barberis, J. N. (2017). From FinTech to TechFin: the regulatory challenges of data-driven finance. *NYUJL & Bus.*, 14, 393.

Zhong, H., Liu, C., Zhong, J. and Xiong, H. (2018). Which startup to invest in: a personalized portfolio strategy. *Annals of Operations Research*, *263*(1), pp.339-360.

Zhu, Y. (2009). Market Signals about Investment Decisions: Do Managers Listen? *Academy of Management Perspectives*, 23(1), S. 84-85.

Zider, B. (1998). How venture capital works. *Harvard business review*, *76*(6), pp.131-139.

Zielske, M. and Held, T. (2020). The Use of Agile Methods in Logistics Start-ups: An Explorative Multiple Case Study. *International Journal of Innovation and Technology Management*, *17*(06), p.2050042.

APPENDIX

Appendix 1. Interview Guide for VCs

Section 1: Introduction to the interview: Provide background to the interviewee.
Short introduction and presentation of the students and the research project.

Section 2: General questions about the VC fund and their investment criteria.
This section is only relevant for some VCs, since most of the funds share this information publicly.

- How would you define (Name of the VC Firm)? What is your vision and mission as a (Corporate/Independent) VC?
- What falls under your investment criteria?
    - Sectors/industries
    - Stage growth
    - Geographic focus
    - Average investment ticket size
- What are your current Assets under Management?
- How many investments has your VC firm made over the last 3 financial years?
- Portfolio companies and exits

Section 3: Specific questions about the research topic and interviewees opinion.

- How is the deal sourcing process currently conducted at your firm? Do you rely mainly on human relationships, such as networking and referrals? Or maybe other sources to identify potential investment opportunities?
- Are you currently applying data-driven tools to support/improve the deal sourcing process?
  By data-driven solutions we mean everything from public databases (e.g. CB Insights, Pitchbook, etc.) to more advanced tools like AI or proprietary algorithms.
- If so, how are you applying these data-driven solutions?
    - Mapping investment opportunities: how to ensure the firm is not missing any promising investment opportunities.
    - Storing the data: how the firm is storing the information that is found regarding startups and investors. What CRMs and other complementary tools?
    - Are these data-driven tools also helping/supporting the partners to make decisions and move on to the next step of the funnel?
- Has your VC firm increased the usage of data-driven tools within the deal origination process over the last 3 years?
- Do you think your firm has plans to increase the adoption of data-driven tools to improve the deal sourcing process? Can you identify the possible competitive advantage?

- Challenges within deal-origination and the application of data-driven solutions? Have you identified any industry trends?
- How do you see the future development of data-driven solutions within the VC industry, particularly as an approach to improve the deal sourcing process?

Appendix 2. Interview Guide for Professors

Section 1: Introduction to the interview: Provide background to the interviewee.
Short introduction and presentation of the students and the research project.

Section 2: Specific questions about the research topic and interviewees opinion.
It is often discussed that VCs invest in new technologies and innovative solutions, but VC firms, up to date, have not been innovating to improve their operations. That means that most of the VC firms still rely on more traditional approaches, such as networking, to make investment decisions.

From an academic/research perspective:
- What is your research scope?
- Is there an active research stream about the application of data-driven solutions within deal sourcing? How important do you perceive the deal origination?
- How do you perceive the relevance of data-driven solutions within deal origination? What do you think is the role of data-driven solutions as a potential approach to improve the decision making process in VC investments?
  → *Define what we mean by data-driven solutions.*

From an industry/practical perspective:
- Based on your experience, would you say that the VC industry has been applying innovative tools in recent years?
- Is the deal sourcing process currently changing in its execution?
- Which are those innovative solutions and how are they improving the VC investment process?
  - To map investment opportunities.
  - To collect more data from target startups.
  - To store the data from target startups. (Increase usage and improvement of CRM tools)
  - To support the partners in the decision making process
- How do you think it is going to develop over time? Are VCs going to rely more on data-driven tools rather than human relationships and networking practices in the future? Have you identified any trends?
- Which are the challenges that the VC industry could face in becoming more data-driven?