INVESTIGATING ASSOCIATIONS BETWEEN FOUNDER CHARACTERISTICS AND STARTUP SUCCESS

by

Dmytro Rohozhynskyi

A thesis submitted in partial fulfillment of the
requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2025

Thesis Supervisor: _____ Professor Olha Halytsia

Approved by _____
Head of the KSE Defense Committee, Professor [Type surname, name]

_____

_____

_____

Date _____

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

VC Venture Capital

AI Artificial Intelligence

Bn billion

LLM Large Language Model

FOALED Fighters, Operators, Accomplishers, Leaders, Engineers and Developers

R&D Research and Development

P&L Profit and Loss

IPO Initial Public Offering

CEO Chief Executive Officer

SaaS Software as a Service

CHAPTER 1. INTRODUCTION

Venture Capital is a vibrant niche. Some people compare it to gambling, and not without reason. Suppose in roulette you choose between black and red and have approximately 50/50 chances of payoff. In that case, VC is interested in startups that will raise the second round of investment with a higher valuation or will be bought by a bigger corporation. The probability of such an event is a dozen times smaller, with 90% of the startups closing after 1 year and 70% more in years 2-5. Essentially, VC job consists of constant betting on zero. But unlike gamblers, VCs don't just watch the wheel spin. They step into the game: they mentor founders, introduce them to clients, push them toward better business models, even recruit missing team members. The capital they bring is only half of the value - the other half is the network, the credibility, and the know-how that makes the next round possible. Smart VCs don't wait for luck. They rig the game.

Venture capital runs on time. Long time. A fund you read about today won't know for a decade if it won or lost. Returns come in cycles – launches, crashes, hype, collapse. Patience disguised as urgency. The irony? It's fueled by investors who want instant fireworks. So the VC stands between two clocks: fast money and slow progress.

And here's the twist – the best deals never look good. They look risky, strange, or wrong. Until they aren't. But that's exactly where the magic tends to hide. Facebook in 2004 was just a college project; Airbnb in 2008 looked like a bad joke about strangers sleeping on your couch. The nature of Venture Capital is to embrace the absurd early, precisely because the obvious opportunities are already priced in. By the time the mainstream believes, it's too late.

If the percentage of success is so small, why play in the first place? (you might ask). Payoff. There is no other field of investment that can bring you that much money back. Unicorns

(companies that have reached $1bn valuation) are the holy grail that everyone is trying to find. One good investment can cover losses from dozens of bad ones.

From such disproportional spread in losses and gains arises an interesting observation. If a hypothetical company has a 5% success rate of its investment that covers 95% of its losses, it means that a small change of 1% in the success rate can bring tremendous profits. And this is exactly the reason why this research exists. Imagine if there are factors that affect the success of a startup and can be easily determined in the earlier stages. Such patterns need to have the smallest effect to become useful. They don`t have to be the absolute solution to the question of whether to invest or not, but they should work as a guiding beam and additional filter.

So, what exactly are we looking for?

In plain words - early, cheap, and observable signals that tilt the odds ever so slightly away from zero. There is nothing more available than a founder's LinkedIn page. So why not check whether their previous founding track record, gender, a certain set of skills, their level of education and the ranking of their alma mater, the number of languages that they know, or just the overall number of founders might affect startup success? Each of these datapoints on its own is hardly prophetic, but, layered together, they may nudge that 5 % hit-rate to 6 % or 7 %. And if the math we just walked through holds, that incremental lift is enough to turn a merely good fund into a top-quartile performer.

Obviously, many more reasonable factors might affect startup success, but these ones have an incredible advantage. They are on the surface. There is almost no paywall, no need to get private investigators or pay thousands of dollars for some ultra-niche datasets. Open source intelligence at its finest. Sure, it doesn`t come without any caveats. The info might not be spot-on because we're stuck with whatever founders chose to share about themselves. LinkedIn does some moderating, yet that mainly catches outright scams and skips minor errors. Of course, such signals are not absolute. Many of them reflect biases

rather than qualities of a founder. A prestigious degree may ease fundraising but say little about execution. Gender differences in outcomes may highlight investor prejudice more than business fundamentals. Distinguishing between correlation and causation is essential.

The point isn't guessing the next big hit for sure, but dodging clearly bad options and putting money toward trends worth noticing. This is the practical view: spot simple, visible traits of founders that don't cost much to check, use them widely, mix in tests over different places and years, then stack them into choices - but never let them stand alone. In an area where tiny shifts flip results, just this much makes a difference.

CHAPTER 2. INDUSTRY OVERVIEW AND RELATED STUDIES

Venture capital keeps going back and forth, yet the figures still shock - 2024 wrapped with about US $314 billion in worldwide VC cash, just a slight 3% jump from 2023, but nowhere near 2021's high; AI monster deals led the charge, grabbing close to a third of all money tossed in.

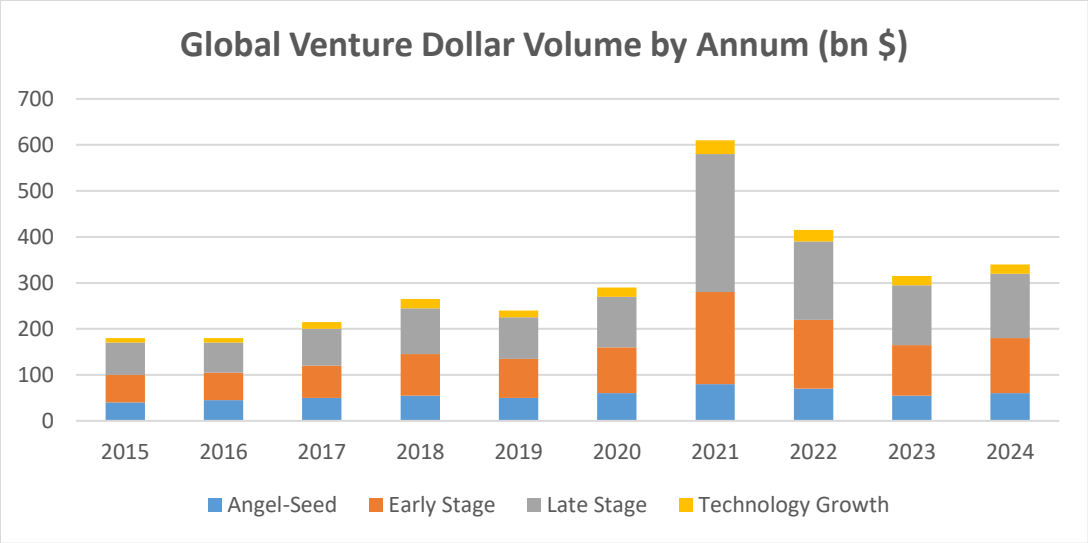

Figure 1: Global Venture Dollar Volume By Annum
Source: Crunchbase

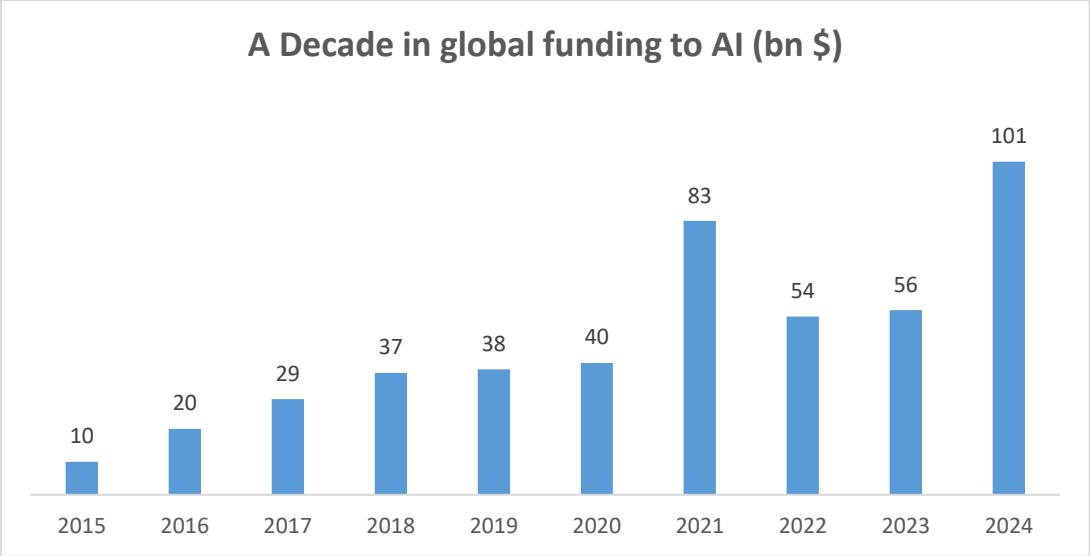

Figure 2: Global Funding To AI
Source: Crunchbase

Geographically, the United States captured 57 % of all capital - half of it concentrated in the Bay Area - while Asia–Pacific fell to a nine-year low and Europe slipped to an 11 % share in Q1 2025 after a stronger 2024. The rebound has been powered by late-stage deals and a handful of outsized Series C+ rounds; seed- and Series A-volume, however, continues to sag, reflecting investors' lingering risk aversion.

Against this backdrop, failure remains the industry's default outcome. Recent surveys put the overall startup mortality rate at ≈ 90 %, with only 30-35 % (out of the 10% left) surviving beyond year five.

	Quote / Statistic	Source
1	The highest ten-year survival rate for new businesses is agriculture, forestry, fishing, and hunting, at 50.5%.	U.S. Bureau of Labor Statistics
2	Founders of a previously successful business have a 30% chance of success with their next venture.	Skill vs. Luck in Entrepreneurship and VC
3	The main challenge to the success of a startup is product market fit.	Emerline
4	82% of successful business owners admit they have the right qualifications and backed up experience to run a company, even with limited cash flow.	Small Business Trends
5	Paying attention to your customers is important since 14% of startups fail due to not regarding customers' needs.	Fundera
6	Founders who have failed previously have a 20% chance of success while first-time founders have an 18% chance of success.	Skill vs. Luck in Entrepreneurship and VC
7	In 2022, U.S. healthcare startups were the strongest industry, bringing in $12.6 billion in revenue.	Fierce Healthcare

Table 1: Startup Success Rates (Source: Embroker)

Exits are equally scarce: global IPO counts have yet to recover to pre-2022 levels, and M&A appetite is restrained outside of AI infrastructure and climate-tech niches. In other words, the funnel that takes a seed-stage company to a meaningful liquidity event keeps getting narrower - and costlier.

One of the largest contributions to the work on startup performance is the 2023 Scientific Reports paper The impact of founder personalities on startup success. Using 21187 Crunchbase-listed companies, McCarthy shows that:

-Big-Five facets such as Adventurousness, low Modesty, and high Activity Level are over-represented among successful founders;

-Founders cluster into six archetypes - Fighters, Operators, Accomplishers, Leaders, Engineers and Developers (FOALED).

-Personality-diverse teams with three or more founders double the odds of a successful exit, relative to solo founders.

Complementary evidence comes from Freiberg & Matz (2023), who evaluate 10,541 founder–startup dyads and find that Openness helps at fundraising, while Emotional Stability matters all the way through exit. Their research, combined with findings from the FOALED project, shifts focus away from old sayings like "back the jockey" and looks instead at actual personal qualities and how team members fit together.

A different paper that stood out was Hanák's 2020 study titled 'Entrepreneur's Experience as Business-Angel and VC Decision Criteria'. Starting with a broad review of 54 real-world studies on how angels and VCs evaluate startups, Hanák lists each personal trait investors consider during assessment. Instead of statistical models, he uses a clear side-by-side comparison to track which career backgrounds get mentioned most - and where existing research disagrees or overlaps oddly. Experience ranks highest in the "entrepreneur/team" category for both angel investors and venture capitalists. It gets broken down into groups like these:

- Industry / domain-specific - Years in same sector, patent count, prior industry network

- Start-up / serial - Number of ventures founded, "track record" exits

- Managerial / leadership - Years managing people or P&L responsibility

- Functional - Depth in marketing, finance, R&D, production

- International - 1 year or more abroad in a business role

Even though such segmentation nicely covers the field, it has some issues. For example, due to overlap, the same founder often ticks multiple boxes simultaneously, making it hard to isolate causal weight.

This study nicely suggests where to look. Based on its results, it is obvious that industry experience matters a lot (it shows up in 11 of 13 angel studies and all VC studies that list experience explicitly). Across 65,390 firms, performance dips on a founder's second venture and rebounds on the third and fourth, which indicates that a track record matters.

Here, I`ll be moving a bit towards some thoughts on approaching the evaluation of founder profiles. I would like to mention Özince & Ihlamur (2024) "Automating Venture Capital: Founder assessment using LLM-powered segmentation, feature engineering and automated labeling techniques". They had a natural question of whether a large language model can take over the grunt work of first-pass founder due diligence. (To add some context, it is exactly what most relatively small-sized and progressive venture firms are trying to implement). They began with a pool of 3,716 U.S. LinkedIn profiles (1,022 "successful", 2,694 "unsuccessful") and then drew a balanced working sample of 300 founders - 150 who steered companies to an IPO, >US $500 m exit, or >US $500 m fund-raise, and 150 who did not.

The clever bit is what happens next. Using a chain-of-thought prompt, GPT-4o first generates a structured prose "founder summary" from each raw LinkedIn JSON. The same model then auto-assigns three layers of labels:

- A ten-rung "Level" ladder (L1–L10) that proxies entrepreneurial stature - from "nascent founder" to "serial unicorn builder";

- Twenty "Persona" archetypes that capture style and background (e.g., "military-service founder", "PhD founder");

- Twenty-three Boolean flags, such as big-tech alum or top-tier consultant.

Success rates climb monotonically up the Level ladder, from ≈35 % in L1–L4 to 100 % in L9 and L10 (founders who have already run unicorns or multiple nine-figure exits). Among personas, award-winning founders (Q) and those with military backgrounds (R) exceed 90

% success, whereas nonprofit veterans (P) bottom out at zero. Feeding all labels into a simple random forest delivers 70 % precision on a balanced test set - respectable lift for a zero-touch screen.

Nowadays, the new models have way better performance than GPT-4o, and I sure will be implementing this approach in my term paper.

CHAPTER 3. METHODOLOGY

Methodology starts with determining the word "success" in the case of the research. Previously, before the data gathering process, I wanted to mark it through the time that it took a company to gather their round, the amount of money they raised, and at what valuation, to further understand the trajectory, since valuation is one of the direct indicators of progress/stagnation. But after the data issues, which I will be touching on later in Chapter 4, strategy took its turn. Due to limitations, I was forced to adapt and took 4 factors as determinants of success: the last funding round stage, the overall number of rounds that were gathered, total equity funding gathered, and the last funding date. So each outcome is modeled separately to respect its statistical nature.

1. Stage Model (Latest Funding Stage)

The latest_stage outcome is treated as an ordinal categorical variable reflecting a natural progression of funding stages: Pre-Seed, Seed, Series A, Series B, Series C, Series D, and Series E. It is analyzed using a proportional-odds ordered logit model, where the log-odds of being at or above a given stage are modeled as a linear function of CEO characteristics and team composition. The model is checked for heteroskedasticity.

The model assumes the proportional odds assumption, which implies that the effect of each predictor is constant across all thresholds of the outcome. The closed firms are excluded.

$$Pr(Stage \leq j) = logit - 1(\theta j - X\beta)$$

Outcome: latest_stage (ordinal factor)

Model: Ordered logistic regression (polr)

Predictors: Number.of.Founders, ceo_prev_startups, ceo_edu, ceo_top_uni, skill_tech, skill_business, skill_sales, languages_n, gender

2. Round Count Model

The round_count outcome is treated as an ordinal count variable representing the number of distinct funding rounds a startup has raised. It is analyzed using a proportional-odds ordered logit model, following the same structure as the stage model.

$$Pr(Rounds \leq j) = logit - 1(\theta j - X\beta)$$

Outcome: round_count (ordinal factor or integer count)

Models: Ordered logistic regression (polr)

Predictors: Number.of.Founders, ceo_prev_startups, ceo_edu, ceo_top_uni, skill_tech, skill_business, skill_sales, languages_n, gender

3. Total Equity Funding Model

The Total.Equity.Funding.Amount..in.USD. outcome is treated as a continuous numeric variable reflecting the total funding a startup has raised to date. It is modeled using ordinary least squares (OLS) linear regression.

$$Funding\ i = Xi\beta + \varepsilon i$$

Outcome: Total.Equity.Funding.Amount..in.USD. (numeric)

Model: OLS linear regression

Predictors: Number.of.Founders, ceo_prev_startups, ceo_edu, ceo_top_uni, skill_tech, skill_business, skill_sales, languages_n, gender

4. Last Funding Date Model

The Last.Funding.Date is treated as a numeric variable, derived from the date format and expressed as the number of days since a baseline date (or optionally converted to a year-based scale). This variable is interpreted as a proxy for how recently a company has received funding - a recency-based measure of success. It is analyzed using OLS linear regression, treating the date numerically.

$$Date\ i = Xi\beta + \varepsilon i$$

Outcome: Last.Funding.Date (converted to numeric)

Model: OLS linear regression

Predictors: Number.of.Founders, ceo_prev_startups, ceo_edu, ceo_top_uni, skill_tech, skill_business, skill_sales, languages_n, gender

Predictors. The predictor set includes number of founders in the company, CEO experience, education, university prestige, skills, languages, and gender. Education is modeled as an ordinal variable (None, BA, MA/MSc, MBA, PhD), reflecting the hierarchical structure of degrees. University prestige is categorized into bins (Top 50, Top 200, Top 1000, Not ranked). Skills are represented both as grouped binary indicators (technical, business/finance, sales/marketing). Continuous variables such as prior startups and languages are standardized to z-scores for comparability of effect sizes.

Diagnostics and Robustness

Models got worked out using only full rows - any with gaps or endless numbers got left out. In descriptive statistics for Total Equity Funding Model there was a high skewnes which indicated outliers, yet I kept them in because the original data set was tiny enough. For ordered results like Stage or Round Count, I used proportional odds models - assuming each predictor behaves the same across levels. For each model I ran a Studentized Breusch-Pagan test just to see if variance changes with predictions and whether I'd need robust standard errors.

Limitations

Independent variables describe founders more than firms. Company factors such as business model, revenue, or team size are missing. Some variables, like education or skills, were simplified into categories, flattening important differences. The data is cross-sectional, showing one moment in time, not change or cause and effect. Biases are also likely: underrepresented groups and failed startups are less visible in public data, limiting how far the results can generalize.

Also, one of the biggest issues is that due to the specifics of the early-stage startups and their mortality rate, Crunchbase data has a huge dip in coverage of closed companies, which, taking into consideration the literature coverage, can not be representative in the closed-to-active company ratio and, as a result, heavily distorts the possible model based on it.

Result Interpretation

1. Stage Model

The results are being interpreted through Average Marginal Effect and their p-values. The stage variable moves through funding steps. A positive coefficient means the predictor increases the odds of being at a later stage. Significance shows whether that shift is meaningful. The ordered logit assumes effects are roughly the same at each jump upward.

2. Round Count Model

The results are being interpreted through Average Marginal Effect and their p-values. Similar idea but focused on the number of rounds. A higher coefficient points to a greater chance of more rounds being completed.

3. Total Equity Funding Model

Linear regression interprets coefficients as the expected change in total funding for each one-unit change in the predictor, usually in millions. Importance comes from effect size, significance, and overall model fit. Outliers can pull results hard, so robustness matters when reading large coefficients.

4. Last Funding Date Model

The dependent variable measures time since a fixed reference date. A positive coefficient suggests more recent funding. Later dates can reflect both success and simple survival, so results show association, not causation. Duration and censoring effects aren't captured.

CHAPTER 4. DATA

As the research progressed, the question of data gathering and the overall dataset structure appeared to be one of the most difficult ones. I happened to have access to the paid Crunchbase account, which somewhat eased the pain, but wasn't the absolute remedy. Before the start of the research, on the idea level, I was sure that I could find data on all of the investment rounds for each company with precise dates, sizes of the rounds, and the companies' evaluations at the moment. This would have allowed me to build a complex timeseries dataset that would be usable for establishing a robust company success indicator, and later in the stage would have allowed me to use the Cox proportional hazards model. Frankly speaking, I overestimated the Crunchbase capabilities.

The actual dataset was formed by the next filtration:
Headquarters Location – United States.
Industry – SaaS, Consumer Applications, Mobile Apps, Apps, Web Apps
IPO Status - Private
Founded – 01/01/2018 – 01/01/2019

The reason behind the filtration choice was rather pragmatic and led by the desire to minimize any possible variation in the starting positions (fair rules for everyone). Industry options are chosen in a similar range of scalability. Apps are an easy-to-scale a very popular niche that guarantees us a large dataset with companies that are having lots of movement in a chosen period of time. The founding dates are in the range of a year and are almost equally distant from the Covid-19 crisis, which reinforces the "fair rules for everyone" policy and somewhat excludes the risk of an anomaly triggered by the unusual global shift in the market dynamics. IPO status is also one of the equaling factors due to the

unpredictable influence of the publicly traded companies on the data (There were not so many publicly traded companies, so they may be considered extremes)

The total number of companies within this range ended up being 4296. Out of them:
Active – 4041
Active & was acquired – 135
Closed – 255
Closed & was acquired – 54

The proportions of Active to Closed companies do not reflect reality. The true percentage of companies that have gone bankrupt, ceased their operations, or resorted to temporarily stopping their activity – can not be determined due to technical limitations. What we know from relevant sources/literature (previously covered) it should be closer to 90-95%. Such an issue is the cause of the inability to compare these factors/variables, preserving the true situation on the market.

It also reveals an uncomfortable truth about startup research in general – we mostly study those who survived long enough to leave a trace. The quiet failures, the ones that shut down before anyone noticed, are almost completely invisible in the data. That alone creates a massive survivorship bias, one that distorts not only the numbers but also our collective perception of what "success" looks like. Without access to those missing stories, any model – no matter how advanced – will only show a part of the picture.

The total dataset is minimized down to 162 companies (1:20 scale), excluding all the ones that were acquired, and taking only the ones that are active. The sole purpose of limiting the dataset down to a smaller scale was the issue with Crunchbase limitations, which did not allow me to export all of the LinkedIn accounts of CEOs of all companies, to later scrape all the necessary data, and as a result, forced this task on me in a manual labour form. So to save time and my overall mental stability, I had to resort to the actions that may have an effect on the overall quality of the research.

Coming up to the manual labour part, I ended up limiting the founder's data only to the CEOs. Throughout factors that were gathered are: the number of previously founded companies, the CEO's education (BA, MA, MBA, Ph.D.), the CEO`s university place in top 1000 (top 50, top 200, top 1000, not top), technical/business/sales skills (where 1 is present and 0 is absent), the number of languages CEO knows and their gender.

It's not a perfect setup – far from it – but it's a realistic one. Working with what you have instead of what you wish you had is a kind of discipline on its own. Each factor here represents an attempt to capture something intangible, something about a person's professional DNA. Education, experience, skill – all of these are only fragments of a much larger story. Still, even through this narrow lens, you can sense the outlines of a pattern.

Variable	Frequencies (% of valid)	Valid (%)	Missing (%)
Number.of.Founders	1: 25 (15.4%) · 2: 72 (44.4%) · 3: 46 (28.4%) · 4–8: 19 (11.7%)	100	0
round_count	1: 15 (9.3%) · 3: 44 (27.2%) · 4: 36 (22.2%) · others: 67 (41.3%)	100	0
latest_stage	Pre-Seed: 6 (3.7%) · Seed: 41 (25.3%) · A: 48 (29.6%) · B: 41 (25.3%) · C: 20 (12.3%) · D–E: 6 (3.8%)	100	0
ceo_prev_startups	0: 76 (46.9%) · 1: 47 (29.1%) · 2: 27 (16.7%) · 3–5: 12 (7.4%)	100	0
ceo_edu	Bachelors: 65 (42.2%) · Masters: 47 (30.5%) · MBA: 29 (18.8%) · PhD: 13 (8.4%)	95.1	4.9
ceo_top_uni	not top: 40 (25.8%) · top 1000: 26 (16.8%) · top 200: 33 (21.3%) · top 50: 56 (36.1%)	95.7	4.3
skill_tech	TRUE: 83 (51.2%) · FALSE: 79 (48.8%)	100	0
skill_business	TRUE: 113 (69.8%) · FALSE: 49 (30.2%)	100	0
skill_sales	TRUE: 50 (30.9%) · FALSE: 112 (69.1%)	100	0
languages_n	1: 53 (32.7%) · 2: 67 (41.4%) · 3: 27 (16.7%) · 4–5: 15 (9.2%)	100	0
gender	female: 43 (26.5%) · male: 119 (73.5%)	100	0

Table 2: Descriptive Statistics

Variable	Type
Number.of.Founders	int
round_count	int
latest_stage	ordered factor (7 levels)
Last.Equity.Funding.Amount..in.USD.	int
ceo_prev_startups	numeric
ceo_edu	ordered factor (4 levels)
ceo_top_uni	ordered factor (4 levels)
skill_tech	logical
skill_business	logical
skill_sales	logical
languages_n	int
gender	factor (2 levels)

Table 3: Data Types (STR)

Caption:

1 - (0.05 ≤ p < 0.1)

2 - p < 0.05

3 - p < 0.01

MODEL 1

LATEST STAGE OF INVESTMENT

factor	AME	SE	z	p	lower	upper
ceo_eduMasters	0.0152	0.0145	1.0494	0.2940	-0.0132	0.0436
ceo_eduMBA	-0.0084	0.0136	-0.6211	0.5345	-0.0350	0.0181
ceo_eduPh.D.	0.0414	0.0375	1.1050	0.2692	-0.0321	0.1149
ceo_prev_startups	-0.0140	0.0094	-1.4980	0.1341	-0.0324	0.0043
ceo_top_unitop 1000	-0.0645	0.0378	-1.7073	0.0878	-0.1384	0.0095
ceo_top_unitop 200	-0.0549	0.0344	-1.5960	0.1105	-0.1224	0.0125
ceo_top_unitop 50	-0.0512	0.0320	-1.5987	0.1099	-0.1140	0.0116
gendermale	-0.0499	0.0266	-1.8741	0.0609	-0.1020	0.0023
languages_n	0.0022	0.0064	0.3401	0.7338	-0.0104	0.0147
Number.of.Founders	-0.0009	0.0050	-0.1856	0.8527	-0.0106	0.0088
skill_business	0.0166	0.0167	0.9933	0.3206	-0.0162	0.0494
skill_sales	-0.0033	0.0123	-0.2700	0.7871	-0.0274	0.0207
skill_tech	-0.0203	0.0141	-1.4342	0.1515	-0.0479	0.0074

Table 4: Latest Stage of Investment

Studentized Breusch-Pagan test

BP = 13.141, df = 13, p-value = 0.437

No proofs of Heteroskedasticity

$$\begin{aligned}\Delta P(Y) = {} & + (0.0152) \cdot ceo\_eduMasters - (0.0084) \cdot ceo\_eduMBA + (0.0414) \\ & \cdot ceo\_eduPhD - (0.0140) \cdot ceo\_prev\_startups - (0.0645) \\ & \cdot ceo\_top\_unitop1000 - (0.0549) \cdot ceo\_top\_unitop200 \\ & - (0.0512) \cdot ceo\_top\_unitop50 - (0.0499) \cdot gender\_male \\ & + (0.0022) \cdot languages\_n - (0.0009) \cdot Number\_of\_Founders \\ & + (0.0166) \cdot skill\_business - (0.0033) \cdot skill\_sales \\ & - (0.0203) \cdot skill\_tech\end{aligned}$$

None of the AMEs are strongly significant at 5%, but two variables show marginal significance (p between 0.05 - 0.1):

- ceo_top_unitop 1000
- gender male

Both effects are negative, meaning they correlate with a slight reduction in the likelihood of being in a higher stage. The magnitude of AME (around −0.05 to −0.06) means approximately a 5–6% change in probability associated with a one-unit change in the predictor.

MODEL 2

THE TOTAL NUMBER OF INVESTMENT ROUNDS

factor	AME	SE	z	p	lower	upper
ceo_eduMasters	0.0034	0.0094	0.3603	0.7186	-0.0151	0.0219
ceo_eduMBA	0.0317	0.0225	1.4060	0.1597	-0.0125	0.0759
ceo_eduPh.D.	0.0048	0.0167	0.2866	0.7744	-0.0279	0.0375
ceo_prev_startups	-0.0043	0.0053	-0.8133	0.4160	-0.0148	0.0061
ceo_top_unitop 1000	-0.0021	0.0180	-0.1162	0.9075	-0.0374	0.0332
ceo_top_unitop 200	-0.0112	0.0164	-0.6844	0.4938	-0.0434	0.0210
ceo_top_unitop 50	-0.0157	0.0154	-1.0184	0.3085	-0.0458	0.0144
gendermale	0.0196	z0.0145	1.3553	0.1753	-0.0088	0.0480
languages_n	0.0014	0.0051	0.2738	0.7843	-0.0087	0.0115
Number.of.Founders	-0.0073	0.0042	-1.7249	0.0845	-0.0156	0.0010
skill_business	0.0116	0.0127	0.9138	0.3608	-0.0133	0.0364
skill_sales	-0.0230	0.0131	-1.7578	0.0788	-0.0486	0.0026
skill_tech	-0.0033	0.0102	-0.3235	0.7463	-0.0233	0.0167

Table 5: The Total Number of Investment Rounds

Studentized Breusch-Pagan test

BP = 17.891, df = 13, p-value = 0.1617

No proofs of Heteroskedasticity

$$\begin{aligned}\Delta P(Y) \;=\; &+ (0.0034) \cdot ceo\_eduMasters \;+\; (0.0317) \cdot ceo\_eduMBA \\ &+ \;(0.0048 \cdot ceo\_eduPhD \;-\; (0.0043) \cdot ceo\_prev\_startups \\ &-\; (0.0021) \cdot ceo\_top\_unitop1000 \;-\; (0.0112) \\ &\cdot ceo\_top\_unitop200 \;-\; (0.0157) \cdot ceo\_top\_unitop50 \\ &+\; (0.0196) \cdot gender\_male \;+\; (0.0014) \cdot languages\_n \\ &-\; (0.0073) \cdot Number\_of\_Founders \;+\; (0.0116) \\ &\cdot skill\_business \;-\; (0.0230) \cdot skill\_sales \;-\; (0.0033) \\ &\cdot skill\_tech\end{aligned}$$

Only two variables show weak (marginal) significance:

- Number.of.Founders
- skill_sales

Both have negative AMEs, suggesting that - all else equal - more founders and having sales skills correlate with a slight reduction in the likelihood of being in a higher outcome stage.

MODEL 3

TOTAL EQUITY FUNDING AMOUNT IN USD

Variable	Coefficient	Std. Error	Significance
Number.of.Founders	0.075	-0.105	–
ceo_prev_startups	0.212	-0.116	*
ceo_edu).L	-0.27	-0.306	–
ceo_edu).Q	-0.267	-0.275	–
ceo_edu).C	-0.375	-0.252	–
ceo_top_uni).L	0.505	-0.221	**
ceo_top_uni).Q	-0.503	-0.242	**
ceo_top_uni).C	0.333	-0.266	–
skill_tech	0.456	-0.252	*
skill_business	-0.298	-0.279	–
skill_sales	0.067	-0.271	–
languages_n	-0.186	-0.126	–
gendermale	0.817	-0.275	***
Constant	16.28	-0.45	***

Table 6: Total Equity Funding Amount (in USD)

Studentized Breusch-Pagan test

BP = 9.3784, df = 13, p-value = 0.7438

No proofs of Heteroskedasticity

$$\begin{aligned}\Delta P(Y) \;=\; &+ (0.075) \cdot Number.of.Founders \;+\; (0.212) \cdot ceo\_prev\_startups \\ &- \;(0.270) \cdot ceo\_edu.L \;-\; (0.267) \cdot ceo\_edu.Q \;-\; (0.375) \\ &\cdot ceo\_edu.C \;+\; (0.505) \cdot ceo\_top\_uni.L \;-\; (0.503) \\ &\cdot ceo\_top\_uni.Q \;+\; (0.333) \cdot ceo\_top\_uni.C \;+\; (0.456) \\ &\cdot skill\_tech \;-\; (0.298) \cdot skill\_business \;+\; (0.067) \cdot skill\_sales \\ &- \;(0.186) \cdot languages\_n \;+\; (0.817) \cdot gendermale \;+\; (16.280)\end{aligned}$$

Statistically significant predictors are "Number.of.Founders" and "ceo_prev_startups".

Marginally significant predictors ($p < 0.1$) are "ceo_top_uni.L ", "ceo_top_uni.Q" and "skill_techTRUE".

MODEL 4

LAST FUNDING DATE

Variable	Coefficient	Std. Error	Significance
Number.of.Founders	-0.007	-0.053	–
ceo_prev_startups	-0.057	-0.059	–
ceo_edu).L	-0.088	-0.156	–
ceo_edu).Q	0.244	-0.14	*
ceo_edu).C	0.389	-0.127	***
ceo_top_uni).L	0.017	-0.113	–
ceo_top_uni).Q	-0.013	-0.123	–
ceo_top_uni).C	0.093	-0.136	–
skill_tech	0.016	-0.128	–
skill_business	0.017	-0.142	–
skill_sales	0.317	-0.138	**
languages_n	0.13	-0.063	**
gendermale	-0.058	-0.14	–
Constant	6.351	-0.232	***

Table 7: Last Funding Date

Studentized Breusch-Pagan test

BP = 12.352, df = 13, p-value = 0.499

No proofs of Heteroskedasticity

$$\begin{aligned}\Delta P(Y) = &-(0.007) \cdot Number.of.Founders - (0.057) \cdot ceo\_prev\_startups \\ &- (0.088) \cdot ceo\_edu.L + (0.244) \cdot ceo\_edu.Q + (0.389) \\ &\cdot ceo\_edu.C + (0.017) \cdot ceo\_top\_uni.L - (0.013) \\ &\cdot ceo\_top\_uni.Q + (0.093) \cdot ceo\_top\_uni.C + (0.016) \\ &\cdot skill\_tech + (0.017) \cdot skill\_business + (0.317) \cdot skill\_sales \\ &+ (0.130) \cdot languages\_n - (0.058) \cdot gendermale + (6.351)\end{aligned}$$

Only four variables demonstrate meaningful or marginal significance:

- ceo_edu.Q
- ceo_edu.C
- skill_salesTRUE
- languages_n

All four coefficients are positive, suggesting that - all else equal - higher education, sales skills, and proficiency in multiple languages are associated with a greater likelihood of reaching a higher investment stage.

TOTAL RESULTS

Independent Variable	Latest Stage	Round Count	Total Equity Funding	Last Funding Date
Number.of.Founders		*	***	
ceo_prev_startups				
ceo_edu.L				
ceo_edu.Q				*
ceo_edu.C				***
ceo_top_uni.L	*		**	
ceo_top_uni.Q	*		**	
ceo_top_uni.C	*			
skill_techTRUE			*	
skill_businessTRUE				
skill_salesTRUE		*		**
languages_n				**
gendermale	*		***	

The results indicate a partial correlation between the number of founders, the university background of the CEO, sales skills and gender with the overall success of the startup. These factors were the only ones that demonstrated at least some level of statistical significance across two of the four main models.

However, these findings do not imply causality. Given the limitations discussed in earlier chapters and the fact that none of the variables reached significance in three or more models, the observed relationships should be interpreted as associative rather than causal.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

To put it short, the results are undermining to say the least. Even though some factors do show a significant influence and might be indicators of success, none of them managed to show these dependencies on the scale of at least 3 tests (I am not even talking about all 4 of them). The total number of founders shows great results in Total Equity Funding, and a marginal significance in terms of the Total Round Count, but fails to satisfy the Latest Stage and Last Funding assumptions. CEOs' education seems to be relevant only in the Last Funding Date, which, on paper, still is one of the weakest models as an indicator of success. The university had some effect in Latest Stage and Total Funding. In terms of skills, business ones are thrown aboard, technical – were proven marginally significant in Total Funding, and sales seem to be the most valuable one, appearing in both Round Count and Last Funding Date. Languages had a mediocre significance in the case of the Last Funding Date.

What actually impressed me (and not in a good way) are the results of the number of startups (companies) previously funded by the CEO and the gender statistics. The second one is at least understandable since the niche is historically a male-dominated industry, and the distortion on a larger scale should be visible. What I totally did not expect is that the most widely known as the absolute factor (there are several studies that conclude that approximately after 3+ previously founded companies, CEOs' chances of success increase drastically), the factor that is beign used as one of the crucial determinants for many in the investment field, would end up absolutely non-existent in the results. After that, it is easy to conclude that the results of this research are to be taken skeptically. Such a yield might have been caused by the low unit count in the dataset, or by the niches I have chosen, but we will never be certain.

Another thing I wanted to mention is regarding the available data. I already covered it shortly in the Data Description part, but nonetheless. One of the biggest issues ended up being the pure mechanical inability of Crunchbase to gather information on all of the

startups that "did not make it". And I am not talking about those who worked for a decent amount of time and went bankrupt. I am talking about those who started as bootstraps and quickly needed to shut down due to various other factors. I am talking about the ones that stopped their activity for a while and are awaiting better times. I am talking about those who did not even pop on their radars due to the smaller scale. The result of such a thing is that if we were to determine "success" as an active company, and a closed company will be the opposite, the results will not be significant whatsoever.

When talking about just the CEO, this is pretty much what you would expect. Even though he/she is the most important person in the company, they are far from being the sole determining factor. But it doesn't necessarily mean that the approach itself is wrong, or that I was digging in the wrong direction. I already went through this several times in the limitations part, but I think it is essential to understand; the idea was to take the open source data and try to squeeze something useful from it. This research is extremely easy to replicate and I encourage you to try gathering your own dataset, because this is the type of research I believe to be the most useful one. The type that uses scrap, sees patterns, and transforms it into something inherently mind-opening.

Obviously, it would have been great to tell you how significant my discoveries are and how influential they will be to the whole Venture industry, but my recommendation would be to refrain from the use of the results in any way that might be responsible for your financial decisions and the overall financial wellness. This material is to be considered an interesting experiment and a modest carcass for future works. Nothing more, nothing less.

But even with the limitations mentioned, I still believe the results point in some interesting directions. The fact that things like team size, sales talent, or schooling pop up more than once across different setups might hint that there's some underlying link the data just didn't pick up.

To conclude, I believe that research did what it was supposed to do – it tested assumptions and challenged expectations. I hope others will replicate it, expand it, or even prove it wrong. That's the whole point.

REFERENCES

Crunchbase News. "Startup Funding Regained Its Footing In 2024 As AI Became The Star Of The Show." Crunchbase News. Published on January 7, 2025.
https://news.crunchbase.com/venture/global-funding-data-analysis-ai-eoy-2024

Crunchbase News. "These 11 Charts Show The State Of Startup Investing At The Beginning Of 2025." Crunchbase News. Published on April 21, 2025.
https://news.crunchbase.com/venture/startup-investment-charts-q1-2025/

Embroker. "106 Must-Know Startup Statistics for 2024." Embroker Blog. Published on April 25, 2025.
https://www.embroker.com/blog/startup-statistics/

Freiberg Brandon, and Sandra C. Matz. "Founder Personality and Entrepreneurial Outcomes: A Large-Scale Field Study of Technology Startups." Proceedings of the National Academy of Sciences (PNAS). Published on May 1, 2023.
https://www.pnas.org/doi/10.1073/pnas.2215829120

Hanak Robert. "Entrepreneur's Experience as Business Angels' and Venture Capitalists' Decision Criteria." ResearchGate. Published on January 2020.
https://www.researchgate.net/publication/346562184_Entrepreneur's_Experience_as_Business_angels'_and_Venture_Capitalists'_Decision_criteria

McCarthy Paul X., Xian Gong, Fabian Braesemann, and Fabian Stephany. "The Impact of Founder Personalities on Startup Success." ResearchGate. Published on October 2023.
https://www.researchgate.net/publication/374781588_The_impact_of_founder_personalities_on_startup_success

McDonald Rory, Dan Wang, and Emily Cox Pahnke. "The Past Is Prologue? Venture-Capital Syndicates' Collaborative Experience and Start-Up Exits." Harvard Business School Working Paper.
https://www.hbs.edu/ris/Publication%20Files/McDonald_Rory_J04_The%20Past%20is%20Prologue_91cc16d6-0a6a-4be3-8ee1-1026fde10922.pdf

Özince, and Ihlamur. "Automating Venture Capital: Founder Assessment Using LLM-Powered Segmentation, Feature Engineering and Automated Labeling Techniques." ResearchGate. Published on July 2024.
https://www.researchgate.net/publication/382080195_Automating_Venture_Capital_Founder_assessment_using_LLM-powered_segmentation_feature_engineering_and_automated_labeling_techniques

SeedBlink Blog. "State of Fundraising in Q1 2025: Key Findings from Market Reports." SeedBlink. Published on April 14, 2025.
https://seedblink.com/2025-04-14-state-of-fundraising-in-q1-2025-key-findings-from-market-reports