The Formula For Fortune

Analysis of Start-up Success

Team:

Amit Peled
Anurag Agarwal
Chakita Muttaraju
Manasa Shivappa Ronur
Shivam Agarwal

Abstract

In this study, we investigate the factors influencing startup success using data from Crunchbase, focusing on investments and associated attributes including the profiles of the key personnels involved in a startup. For analysis purposes, success is defined as either company's acquisition or operating for over 5 years. We explore several research questions, including the relationship between industry type and success rate. Our findings reveal a significant association, with manufacturing demonstrating the highest success proportion. We also examine the impact of location on success, noting the associations at both country and city levels, with the U.S. leading in success proportion. Surprisingly, we find no relationship between founders holding STEM degrees and success rate of the startups. Furthermore, our analysis uncovers that equity crowdfunding tends to decrease the success probability of a startup while the secondary markets are more associated with success. Notably, recession years show marked differences in startup success metrics. This report provides entrepreneurs, investors, and policymakers with invaluable guidance by synthesising insights gleaned from these multifaceted analyses. It emphasises how crucial it is to take into account regional variables, funding sources, industry dynamics, and prevailing economic conditions when attempting to maximise strategic choices and increase the probability of startup success.

Table of Contents

1. Introduction

The startup ecosystem has drawn a lot of attention recently as entrepreneurs from all over the world set out to realise their dreams. The desire to start something from scratch and the possibility of quick expansion and influence has led to an increase in entrepreneurial activities. This analysis explores the various aspects that go into making a startup successful, looking at the essential components that distinguish successful endeavours from the inevitable obstacles they must overcome. This project aims to discern patterns, identify influential factors, and gain a deeper understanding of what contributes to startup success or challenges.

The questions addressed in this study are,

1) Some industries, such as healthcare and banking, are more regulated and capital-intensive than others, making it challenging for startups. So, do few industries see more success than others?
2) Location can influence the logistics cost, wages, and networking opportunities, which can further affect the startup operation. So, is there a correlation between the startup's success & operating location?
3) Does the involvement of specific personnel and background (Founders, Co-Founders, Board Members, Investors) affect the success of a company?
4) Does the recession phase affect the company's success timeline?
5) Does the type of funding (venture capital or debt financing) influence the success rate?

2. Methods

2.1. Data Description

2.1.1. Data Source

Startup Investments, performance and associated attributes data is available on the Crunchbase platform. It was downloaded as a mysqldump from the data.crunchbase.com and processed into csv files using MySQL. The processed csv files are available on Kaggle.

It is a comprehensive dataset with csv files containing information about the startup ecosystem: organisations, individuals, office locations, funding rounds, initial public offerings and acquisitions. It includes data of startups from 1902 to 2014 globally. Each file has a unique id associated with each record that enables us to join each file to one another for analysis purposes. The table below gives the list of all csv files along with the brief description of the contents.

File Name	Description of the contents
acquisitions.csv	Contains information about startups that have been acquired.
ipos.csv	Contains data on initial public offerings.
degrees.csv	Detail for people's education background
investments.csv	All investments made by investors
people.csv	Contains information about individual and profiles associated with the startup
funds.csv	Details for investors' investment funds
funding_rounds.csv	Details for each funding round in the dataset
objects.csv	Organization profiles available on Crunchbase platform
milestones.csv	Contains significant events within the startup ecosystem.
offices.csv	Contains information about startup company offices.
relationships.csv	Contains relationship data that links companies to individuals and their positions.

Table 2.1.1.1 *- Data files used for analysis and their description.*

The data dictionary with details about fields in each csv is available here.

The dataset related to Global Economic Recession Timeline is available on this website - https://fred.stlouisfed.org/series/USREC.

2.1.2. Study Design

This was an observation study conducted over 32,765 observations of startups collected and maintained by crunchbase platform. One of the main limitations of the dataset is that the observational data did not have a specific column to indicate the success or failure of a startup directly. To overcome this, a custom calculated field "success_metric_updated" was introduced to indicate if the startup was successful or not. To do so, we added a new field age calculated using the field founded_year. Next a startup was labelled as "Successful" if it is acquired or it is in operating status for more than 5 years, "Potentially Successful if it is operating for less than 5 years", and everything else as "Unsuccessful".

The below pie chart shows the distribution of each status in the newly created column "success_metric_updated" in the dataset.

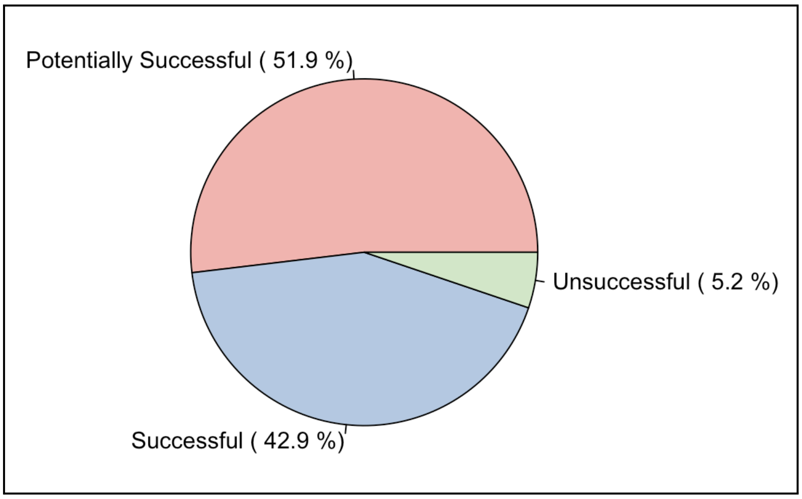

Figure 2.1.2.1 *- Spread of success status across the dataset.*

2.2. Statistical Methods

2.2.1. Do startups in few industries see more success than others? - By Manasa

To analyse the effect of industry type on the success status, we used the success_metric_updated column that was created as described in the study design section, and the custom column Industry. The industry column was created based on the market field which has multiple tags associated with each startup indicating the market it belongs to. For the purpose of this study, only the primary tag was used to assign the start up to a particular industry. Each observation was assigned to one of the seven major industry types - Technology, Manufacturing, Healthcare, Finance and Banking, Services, Retail and Transportation. The ones that do not fall under any of these seven industry types were labelled as "Others". The initial EDA as seen in ***Figure 2.2.1.1*** showed notable variations in success proportions across different industries.

For statistical analysis, we consider a startup as "***Successful***" if the field success_metric_updated is successful and ***"Unsuccessful"*** if the field success_metric_updated is either Potentially successful or Unsuccessful. The null and alternative hypothesis was defined as below.

<u>Null Hypothesis:</u> There is no association between the type of industry and proportion of startups that are successful..

<u>Alternative Hypothesis:</u> There is an association between the type of industry and proportion of startups that are successful.

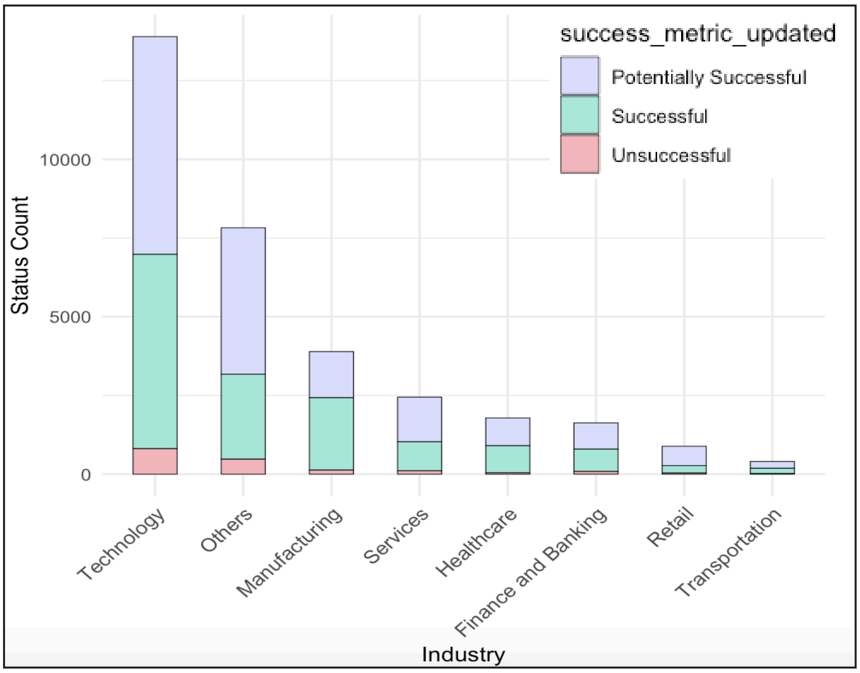

Figure 2.2.1.1: *Success status counts by Industry*

Chi-squared test of independence was chosen as the test method because both the variables involved are categorical and given the large sample size, EDA revealed the expected cell count is greater than 5 for each combination of the variables. So this method of test seemed a perfect fit. In the dataset each observation pertains to one startup and for analysis each observation is assumed to be independent of other observations. To conduct the chi square test , a contingency table containing industry type and the status (success or unsuccessful) was created.

Pearson's Chi-Squared test statistic and corresponding p-value was calculated using the R functions. We are using the significance level 0.05 for all our tests to determine whether to accept or fail to reject the null hypothesis. Additionally 95% wald based confidence interval was calculated to provide a range of plausible values for a population proportion of success status based on the sample proportion for each industry.

2.2.2. Does location of origin of startup influence success status

For this analysis we will be looking mainly at country level and city level data. For the city level analysis, we will focus on cities based in the US, since the dataset consists of mostly US based startups. In this analysis, the success status for the top 5 countries (by number of startups) are examined. The rest of the 101 countries are grouped together under the "Others" category. Similarly for cities, the top 5 cities in the US (by number of startups) were picked with the remaining cities being grouped into the "Others" category. From the EDA it can be observed that there are differences between countries and cities in their success status. Table 2.2.2.1 shows the success status proportion for countries and cities respectively. The "Potentially Successful" status was merged along with the "Unsuccessful" status group.

To determine if the difference in success status between countries and cities in the US is statistically significant, a chi-square test is performed for countries and cities(since both variables are categorical). We assume that the success statuses of the different locations are independent. The cell counts for each cell in the table are greater than 5. Considering these factors, it is appropriate to use a Chi-square test to test for association between the location and success status. The null and alternative hypothesis for the analysis are as given below.

H_0: The location(country/city) of origin of a startup has no statistically significant association with its success status

H_a: The location(country/city) of origin has a statistically significant association with its success status.

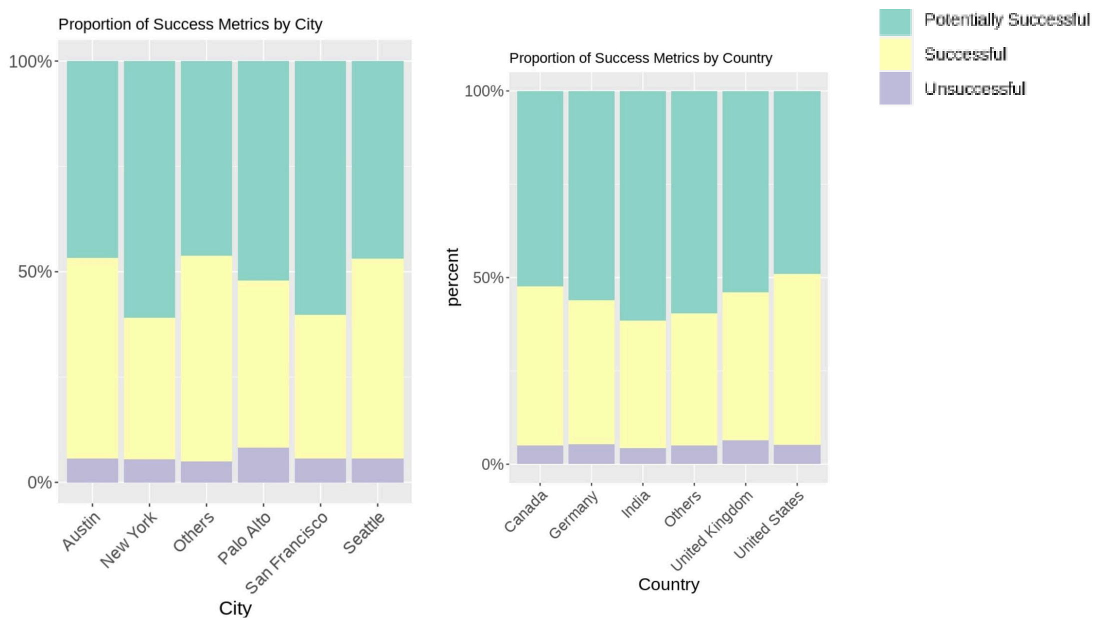

***Figure 2**: Success proportion(%) by country and city*

Country	Successful	Unsuccessful	City	Successful	Unsuccessful
USA	45.78 %	54.2%	***Seattle***	47.54%	52.4%
UK	39.64%	60.36%	***Austin***	47.6%	52.4%
Canada	42.6%	57.4%	***NYC***	33.57%	66.4%
Germany	38.6%	61.4%	***SFO***	34.06%	65.9%
India	34.1%	65.84%	***Palo Alto***	39.6%	60.3%
Others	35.35%	64.6%	***Others***	48.8%	51.1%

***Table 2.2.2.1:** Success proportions for cities and countries*

2.2.3. Does the involvement of specific personnel and backgrounds (Founders, Co-Founders, Board Members, Investors) affect the success of a company? - By Shivam

To analyze the association between specific personnel and backgrounds (Founders, Co-Founders, Board Members, Investors) with **STEM** backgrounds and startup success, we employed a **Chi-squared test of independence**. This test was chosen for its ability to assess the independence between two categorical variables, which in this case are the **personnel/background involvement** (STEM background or not) and **startup success** categories. We first formulated the null hypothesis, suggesting no significant association between these personnel /backgrounds with STEM backgrounds and startup success, with the alternative hypothesis proposing a significant association.

During the data preparation, we grouped all **STEM** subject degree types, whether they possessed a ‘**Bachelor's’**, ‘**Master's’**, or ‘**Ph.D.’** degree. Conversely, all other degree types were classified as **'Other'**. This categorization was essential to maintain consistency and clarity throughout our analysis. ***Table 2.2.3.1*** shows the success status proportions for STEM and Non-STEM degree types.

Also in our analysis, we focused solely on companies labeled as either "**Successful**" or "**Unsuccessful**". We decided to include all instances labeled as "**Potentially Successful**" under the "**Unsuccessful**" category. This decision was motivated by the inherent uncertainty surrounding the success of startups categorized as "**Potentially Successful**". Since their ultimate success is not guaranteed, categorizing them as "**Successful**" could introduce a misleading representation of their status. By aligning them with the "**Unsuccessful**" category, we ensure a more accurate reflection of the uncertainty surrounding their outcomes.

Following established methodologies, we verified our assumptions, ensuring the independence of observations and maintaining sufficient cell counts in the contingency table. Considering a significance level of **0.05**, we proceeded to assess the statistical significance of the relationship between personnel /background involvement with STEM backgrounds and startup success. This assessment relied on Pearson’s Chi-Squared test statistic and its corresponding p-value, providing a robust framework for our analysis.The null and alternative hypotheses for the analysis are given below:

Null Hypothesis (H0): There is no significant association between having a STEM degree among specific personnel and backgrounds (Founders, Co-Founders, Board Members, Investors) and startup success.

Alternate Hypothesis (Ha): There is a significant association between having a STEM degree among specific personnel and backgrounds (Founders, Co-Founders, Board Members, Investors) and startup success.

success_metric_updated	Successful	Unsuccessful
degree_category		
Bachelors in STEM	446	341
Masters in STEM	145	96
Other	690	563
PhD	40	22

Table 2.2.3.1: *Success proportions for STEM and Non-STEM degree types*

2.2.4. Does the type of funding influence the success rate? - By Amit

After exploratory data analysis revealed that there seemed to be a discrepancy in the allocation of funding amounts for start ups depending on the type of funding that was employed to support a company (Figure 2.2.4.1), the question of whether the type of funding influences the success rate was formulated.

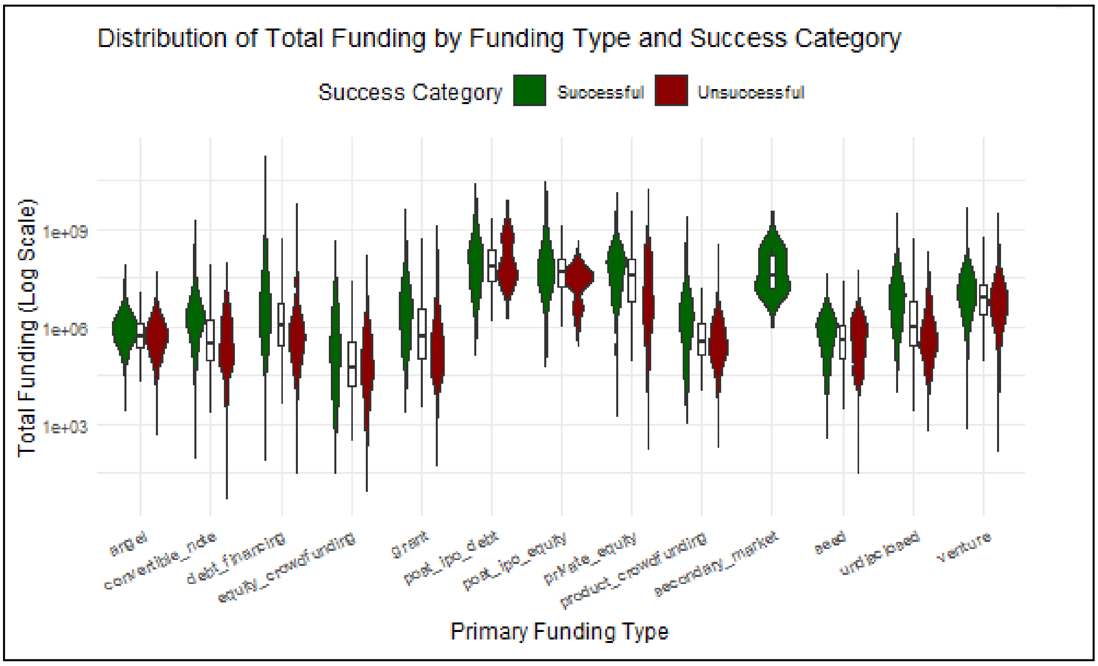

Figure 2.2.4.1*: Distribution of Total Funding by Funding Type and Success Category. Shows how some funding types result in differing total funding than others.*

Two primary statistical methods were utilized to investigate the influence of funding type on startup success rates: the Chi-Square Test of Independence and Logistic Regression.

Chi-Square Test of Independence

The Chi-Square Test was applied to determine whether a statistically significant relationship exists between the categorical variables: the type of funding and startup success rate.

Null Hypothesis (H_0) - There is no association between the funding type and startups' success rate.
Alternate Hypothesis(H_a) - There is a significant association between the type of funding a startup receives and the success rate of startups.

The dataset categorises startups as either successful or unsuccessful, pooling those labelled “potentially successful” into the unsuccessful group. This test was chosen for its ability to evaluate the association between categorical variables in a contingency table format.
To implement this analysis, one would construct a contingency table summarising the frequencies of observed successes and failures across various funding types. The test compares the observed frequencies against the expected frequencies, which are calculated under the assumption that the variables are independent. For the Chi-Square Test to be valid, an important assumption is that the expected frequencies in each cell of the contingency table should be greater than 5. This ensures that the test has enough power to detect an association if one exists.
A significant Chi-Square statistic may suggest that the type of funding does, in fact, influence a startup success rate, indicating that certain funding types are associated with higher or lower success rates. This may guide investors and entrepreneurs in making informed decisions about the funding strategies they want to pursue for their investments.

Logistic Regression
Logistic Regression was utilised to model the probability of startup success as a function of funding type. As the success metric is binary in our dataset, this method is appropriate for estimating the odds ratios of success across different funding types.
To implement this analysis, the categorical variable indicating the type of funding would be transformed into a series of dummy variables, allowing for them to be used as predictors in the regression model. The dependent variable, startup success, would also be coded in binary terms. The logistic regression model is then defined and fitted to the dataset, allowing the estimation of success probabilities conditional on funding types.
Interpreting the coefficients from the model would provide a quantitative measure of the association between funding type and success odds. This method offers an in-depth perspective of how different types of funding might affect the chances of a startup's success beyond the potentially discovered existence of an association indicated by the Chi-Square Test.

Together, these methods form a robust approach to understanding the dynamics between funding type and startup success rates, critical for entrepreneurs and investors in strategic financial decision-making.

2.2.5. Does the recession phase affect the company's success timeline? - By Anurag

The objective of this inquiry was to ascertain whether the occurrence of a recession in a given year influences the success of startups. To conduct this analysis, I examined the counts of various metrics, including IPO listings, acquired startups, closed startups, and startups receiving funding within a year. The goal was to determine if there is a difference in the mean number of these metrics between recession years and non-recession years.

I utilised data from the USREC.csv file, which contained information spanning 170 years. After incorporating data from other CSV files and merging them, I obtained a cleaned dataset with the counts arranged in the desired format.

These are my null and alternate hypothesis:

Null Hypothesis (Ho): There is no significant difference in mean number of IPO listings, mean number of Acquired Startups, Mean number of Startups getting funds and Mean number of Startups closing for year being a recession year or not.

Alternate Hypothesis (H1): There is a significant difference in mean number of IPO listings, mean number of Acquired Startups, Mean number of Startups getting funds and Mean number of Startups closing for year being a recession year or not.

Here are the initial exploratory data analysis (EDA) findings and preliminary observations:

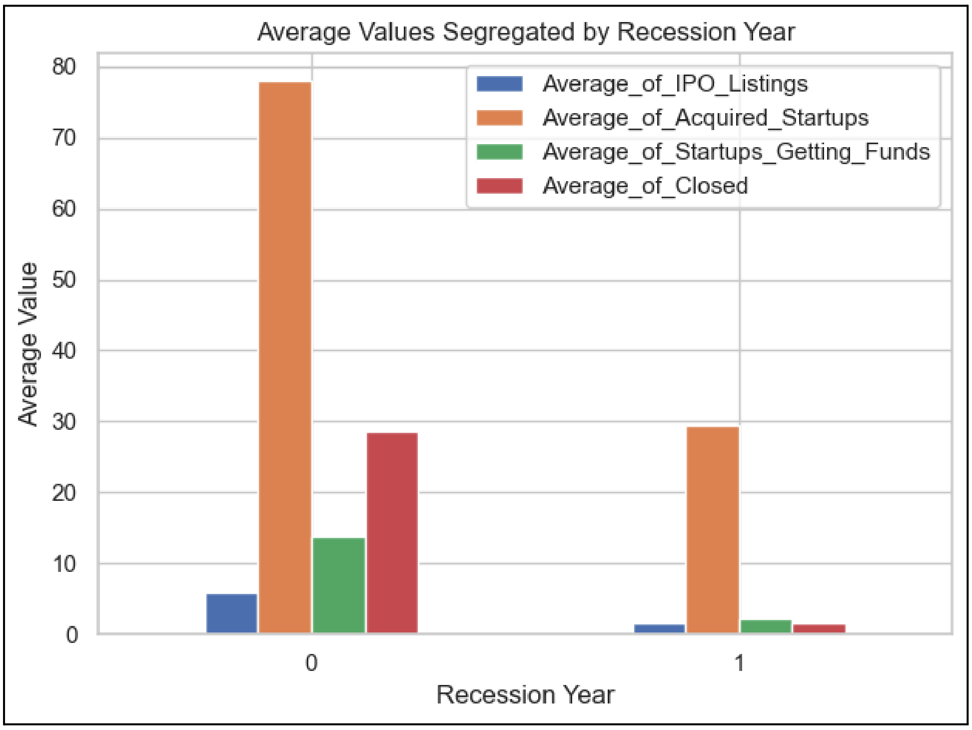

***Figure 2.2.5.1**: Mean count of startups for each metric selected*

Before proceeding with the t-test to compare sample mean counts, I conducted checks to ensure that the assumptions of normality and equal variances were met between the two groups. Shapiro's test was employed to assess normality, while Levine's test was utilised to evaluate the equality of variances between the groups.

Shapiro's test evaluated the distribution of data within each group, determining if it followed a normal distribution. Meanwhile, Levine's test compared the variances between the two groups, potentially indicating unequal variances, which could violate the equal variances assumption necessary for the t-test.

By conducting these tests, I ensured that the subsequent t-test analysis would be valid and reliable, providing accurate comparisons between the means of the two groups.

Metric	Shapiro's test passed	Levine's test passed
IPO listings	No	No
Acquired Startups	No	Yes
Funded Startups	No	Yes
Closed Startups	No	Yes

Table 2.2.5.1: *Results for checking normality and homogeneity of variances*

Since Shapiro's test for normality failed for all metrics of interest, indicating that our data does not follow a normal distribution, I believe the conservative approach of opting for a non-parametric test like the Mann-Whitney U test is appropriate. This test enables the comparison of means (or distributions) between two independent groups, such as recession years and non-recession years, without relying on the assumption of normality.

3. Results

3.1. Do startups in few industries see more success than others? - By Manasa

The results of the Pearson's Chi-squared test and Cramer's V value are as shown in the table below. At the 0.05 significance level, we reject the null hypothesis. There is statistically significant evidence to conclude that there is an association between the proportion of success status and the type of the industry. Cramer's V value further supports the association, by value there seems to be a weak association.

X^2 Statistic	Degree of freedom	p-value	Cramer's V
799.37	7	< 2.2e-16	0.156

Table 3.1.1: *Results of Chi-squared test and Cramer's V value*

Even though the results are statistically significant, one might worry about the practical significance given the small cramer's V value. However we believe that this small value could be significant in the practical world. As we know few industries are more capital intensive, more regulated, have a slow evolving landscape compared to others which plays a crucial role in the success of the startup.

Below figure gives the 95% wald based confidence interval estimate for the true population proportion of success for each industry. Some industries such as manufacturing have more success than others as seen in the figure. Retail industry seems to have the lowest success rate compared to all the other industries. We are able to make more precise estimates for industries such as technology, manufacturing and services where we have a bigger sample size, as opposed to transportation where the confidence interval is very wide due to small sample size.

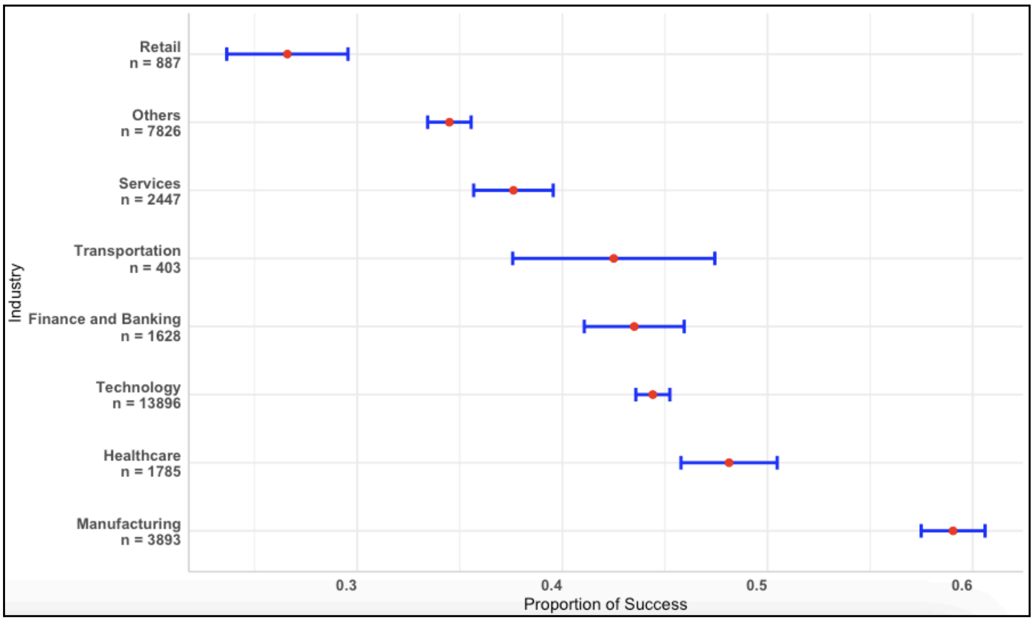

Figure 3.1.1: *95% Wald based Confidence Interval for each industry*

3.2. Does location influence startup success status? - By Chakita

The chi-squared statistic for the country level analysis was 257.39 with degrees of freedom 5. For the city level analysis the test statistic value was 311.33 also with degrees of freedom 5. A p-value of less than the order of 2.06e-16 was obtained. At a significance level of 0.05, this result is statistically significant and we reject the null hypothesis for both the county and city level analyses as the test result indicates a presence of association between location and success status. As a post hoc test, a cramer's V test was performed to check the effect size of the association. For the country level analysis the cramer's V test yielded a value of 0.088 and for the city level analysis a value of 0.118. This implies that the association between location and success status as observed in the dataset is weak.

For further insights on the association, a test was performed to analyse if the success proportion for the US is statistically significantly different from other countries. A two sample z-test for proportions (with equal variance) was performed at a 0.05 confidence level. A similar test was conducted to test the difference in success proportion between Seattle and other US cities. For this analysis all countries except the US were grouped into the “Others category”. Similarly for cities, all cities except Seattle were grouped into the “Others category”. Note that the “Others” category for cities still only contains cities within the US. The results obtained are tabulated in the table 3.2.1. The z-test results suggest that at a significance level of 0.05, there is a difference in success proportion between the US and other countries as well as between Seattle and other US cities. However, on examining the 95% confidence interval for this difference in success proportion, we see that the difference is practically insignificant, for cities. This corroborates with the effect size results obtained from the post hoc analysis of the chi-square test. The 95% confidence interval for difference in success proportion between the US and other countries is about 7 to 10%. At a country scale, we could think of this difference to be practically significant, although a subject matter expert would have to examine various factors to ascertain this. To conclude,the association between location and success status as observed in the data, is too weak to be of practical significance, especially for cities.

	z-score	p-value	95% wald C.I
USA vs “Others”	15.13249	9.887838e-52	(0.077 , 0.0999)
Seattle vs “Others”	1.992275	0.04634092	(0.00076 , 0.0930)

Table 3.2.1: *Results of the two sample z-test for proportions*

3.3. Does the involvement of specific personnel and backgrounds (Founders, Co-Founders, Board Members, Investors) affect the success of a company? - By Shivam

The analysis conducted to investigate the influence of specific personnel(Founders, Co-Founders, Board Members, Investors) with **STEM** backgrounds and the success of a company produced the following results:

Test Statistic	Degree of freedom	p-value	Interpretation
3.98	3	0.26	Not Significant

Table 3.3.1: *Results of Chi-squared test*

At a significance level of **0.05**, the chi-squared test of independence **did not yield a statistically significant result**. This suggests that there is no significant association between the involvement of founders, co-founders, board members, or investors with STEM backgrounds and the success of a company.

However, it's important to note that this finding does not discount the potential influence of other factors on company success. Additional factors such as market conditions, product innovation, marketing strategies, financial management, and organizational culture may play crucial roles in determining the success of a company. Future research could explore these factors in more depth to gain a comprehensive understanding of the complex dynamics that contribute to company success.

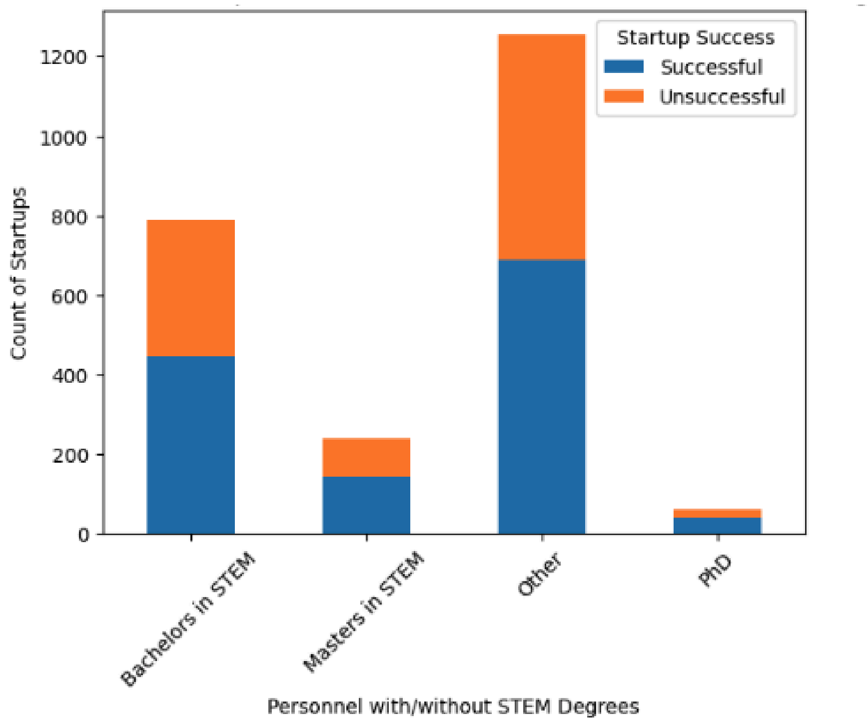

Figure 3.3.1: *Degree type for STEM and non-STEM proportion to success of startup*

3.4. Does the type of funding influence the success rate? - By Amit

In the analysis of the impact of funding type on the success rate of startups, the following results were obtained from both descriptive and inferential statistical methods.

Chi-Square Test for Independence

The Chi-Square test returned a test statistic of 4971.9 with 12 degrees of freedom and a p-value less than 2.2e-16 (Table 3.4.1), leading us to reject the null hypothesis at the 0.05 significance level. This outcome suggests a significant association between the type of funding and the success rate of startups.

Test Statistic	Degree of freedom	p-value	Interpretation
4971.9	12	< 2.2e-16	reject at 0.05 sig level

Table 3.4.1: *Results of Chi-squared test for Funding Type*

A normalized heatmap was created to better visualize the distribution of success across different funding types, as illustrated in Figure 3.4.2. The visual evidence from the heatmap indicates that venture and grant funding are more prevalently associated with successful startups. It should be noted that our success labels are biased towards companies that have gone public due to the time constraints in our logic for determining success. Conversely, funding types such as equity crowdfunding and product crowdfunding are less frequently associated with successful startups. This pattern may suggest that larger and more formal rounds of funding, such as venture capital, are better predictors of startup success compared to smaller, individual investments like angel investing.

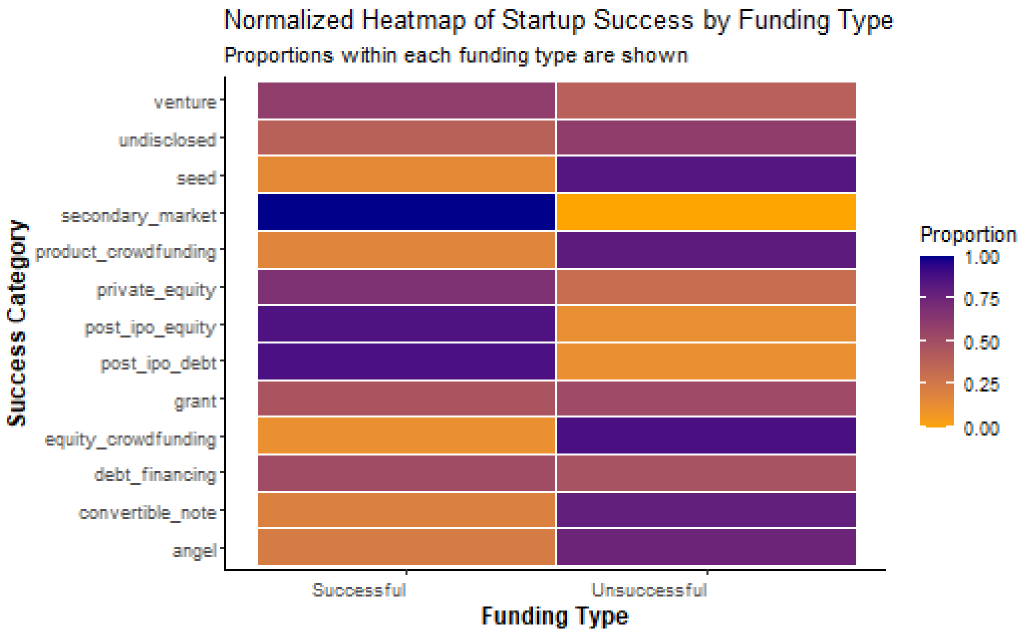

Figure 3.4.2: *Normalized Heatmap to Visualize Effect of Funding Type on Success Rate*

Logistic Regression Analysis

Following the logistic regression analysis, the model allowed for the extraction of probability predictions for each startup success category based on the funding type. All predictors in the model were found to be statistically significant, with the highest p-value for any term being 0.0461. The results from the logistic regression are visually presented in Figure 3.4.3, where the point estimates with their corresponding 95% confidence intervals indicate the influence of different funding types on startup success. Types of funding that are plotted to the right of the zero line (colored in blue) exercise a positive influence, implying that, when other factors are held constant, they are associated with increased odds of startup success. On the contrary, funding types situated to the left of the zero line (colored in red), such as \`product_crowdfunding\` and \`equity_crowdfunding\`, exhibit a negative relationship with the likelihood of startup success.

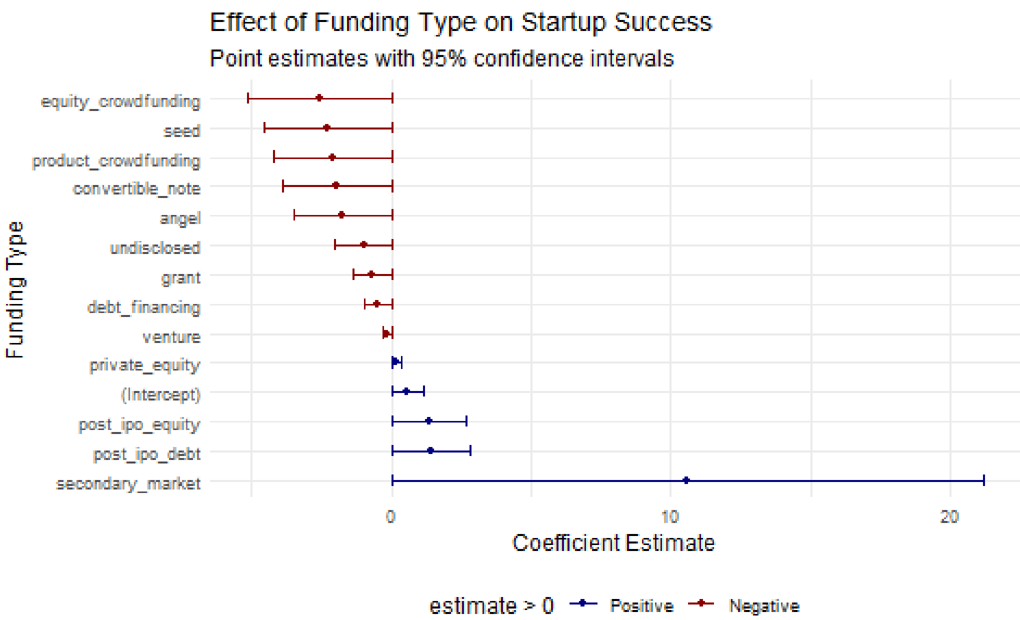

Table 3.4.3: *Visualising the Effect of Funding Type on Success Rate through Point Estimates and their confidence intervals.*

3.5. Does the recession phase affect the company's success timeline? - By Anurag

The Mann-Whitney test results indicate a significant difference in means for all metrics of interest, except for the count of closed startups, between recession and non-recession years. This means that for IPO Listings, Acquired Startups and Startups getting funds, there is a significant difference in the mean counts of these metrics for the year being a recession year or not. The same is not true when comparing the mean number of startups getting a closed status.

Considering the majority of these results, we reject the null hypothesis and accept the alternative hypothesis, suggesting a significant difference in the mean number of success metrics between recession and non-recession years. Thus we can come to the conclusion that the recession period does affect a startups' success.

Variable	Test statistic	p-value	Statistically significant
Number of IPO Listings	2785.0	0.000291	Yes
Number of Acquired Startups	2882.5	0.000931	Yes
Number of Startups Getting Funds	3076.5	0.007193	Yes
Count of Closed Startups	3560.0	0.517655	No

Table 3.5.1: *Results of Mann-Whitney test for interest metrics. (For 0.05 significance level)*

4. Discussion

The statistical analysis conducted on various factors influencing startup success reveals fascinating perspectives on the workings of the entrepreneurial environment.To summarise,

Effect of Industry on Startup Success: Upon analysing the relationship between industries and startup success, our analysis, supported by Pearson’s Chi-squared test and Cramer’s V value, shows an association, which we concluded are significant statistically and practically. Notably, certain sectors, such as manufacturing, exhibit a higher propensity for success compared to others. Conversely, the retail industry emerges with a comparatively lower success rate. The limitation of this analysis is the assumption of independence. In the real world a company might operate in more than one industry sector, so to know more on the effects of industry type on success rate, we might have to drop our assumption of independence and perform tests like correlated data analysis.

Geographical Influences on Startup Triumph: The analysis explored the connection between location (country and city) and success. While a statistical association was found (highly significant chi-square test), it doesn't necessarily translate to a strong real-world link. This is because further tests revealed the association's weakness. The effect size (Cramer's V) was low, and although there was a non-negligible difference in success rates between the US and other countries (7-10%), its practical importance depends on the specific field. Similarly, the statistically significant difference between Seattle and other US cities was minimal in practical terms (confidence interval close to zero).

Therefore, while a location-based influence might exist, this data suggests it is weak and may not be practically significant, especially for cities. It's crucial to remember that other factors likely play a significant role in determining success, and further research is needed to understand the complex interplay between location and success.

Founder Backgrounds and Success Rates: Contrary to prevailing assumptions, the statistical analysis yields no statistically significant association between the backgrounds of founders, co-founders, board members, or investors with STEM (Science, Technology, Engineering, and Mathematics) backgrounds and the attainment of startup success. This notable finding underscores the limited influence of individual backgrounds on the overarching success trajectory of entrepreneurial ventures, challenging conventional wisdom and highlighting the importance of holistic assessments in gauging startup viability.Building upon the existing findings concerning founder backgrounds and their correlation with success rates in startups, there is a promising avenue for future research to explore non-traditional factors that significantly impact the outcomes of these ventures. Diving deeper into these factors could involve investigating aspects such as team dynamics, leadership styles, cultural alignment within the industry, and the influence of external market conditions. By expanding the analytical lens beyond individual founder backgrounds, researchers have the opportunity to develop a more nuanced and comprehensive understanding of the diverse factors that shape entrepreneurial success or failure.

Funding Modalities and Success: The analysis unearths a compelling association between the types of funding procured and the likelihood of startup success. Notably, ventures buoyed by venture and grant funding exhibit heightened prospects of triumph, while those reliant on equity crowdfunding and product crowdfunding face comparatively diminished success rates. These findings underscore the pivotal role of funding modalities in shaping the success paradigms of nascent ventures, prompting stakeholders to strategically navigate the funding landscape to optimize growth trajectories and bolster resilience against market vagaries.

Economic Cycles and Success: The statistical examination elucidates the profound impact of economic cycles on the success trajectories of startups, underscoring significant disparities in success metrics between recessionary and non-recessionary periods. These findings highlight the dynamic interplay between economic contexts and entrepreneurial outcomes, prompting stakeholders to adopt adaptive strategies that reconcile with the prevailing economic milieu and fortify organizational resilience against the exigencies of fluctuating market conditions.

In conclusion, the success of the startup depends on the combination of various factors such as industry type, location, funding and global economic situation. This statistical analysis report acts as an informational tool for stakeholders navigating the challenging landscape of entrepreneurship in addition to offering an in-depth knowledge of the factors that influence startup success. These findings provide practical insights that enable decision-makers to anticipate market dynamics, develop long-term competitive advantages, and create well-informed strategies despite the inherent volatility of the entrepreneurial landscape. Equipped with this empirical comprehension, stakeholders can more adeptly maneuver through obstacles, seize chances, and guide their endeavours toward enduring expansion and wealth in a constantly changing commercial landscape.

5. References

- https://www.statstest.com/cramers-v-2/
- https://data.crunchbase.com/docs/data-dictionary
- https://statisticsbyjim.com/hypothesis-testing/practical-statistical-significance/
- Shapiro test for normality: https://www.sciencedirect.com/topics/mathematics/wilk-test
- Mann Whitney Non-Parametric test: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/mann-whitney-u-test/

6. Code

The code used for the entire study can be found in the git repository here.