

Empowering the Crowd: Equity Crowdfunding's Contribution to Startup Funding and Retail Engagement

MSc In Fintech

Habil Shihoriwala
Student ID: X21197962

School of Computing
National College of Ireland

Supervisor: Theo Mendonca

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Habil Firozbbhai Shihoriwala
.....
Student ID: X21197962
.....
Programme: MSc In Fintech **Year:** 2022-2023
.....
Module: MSc Research Project
.....
Supervisor: Theo Mendonca
.....
Submission Due Date: 14th August 2023
.....
Project Title: Empowering the Crowd: Equity Crowdfunding's Contribution to Startup Funding and Retail Engagement
.....
6651 27
Word Count: **Page Count:**.....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Habil Firozbbhai Shihoriwala
.....
Date: 14th August '23
.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only
Signature:

Date:	
Penalty Applied (if applicable):	

Empowering the Crowd: Equity Crowdfunding's Contribution to Startup Funding and Retail Engagement

Habil Shihoriwala
X21197962

Abstract

Equity crowdfunding serves as a gateway for retail investors to partake in early-stage startup investments, thereby democratizing access to such opportunities. This study consists of five prominent US-based equity crowdfunding platforms, namely SEEDINVEST, STARTENGINE, WEFUNDER, NETCAPITAL and REPUBLIC. By analysing data spanning 2018 to 2023, the author investigates factors catalysing retail engagement and funding success. Focusing on pivotal attributes like funding amount, investor count, and funding status, the researcher employs a statistical approach encompassing data visualization and machine learning models, including SVM, Random Forest, and PCA.

Incorporating a dataset of 4144 campaigns, author establish a significant linkage between attributes such as number of investors, annual revenue, analyst report, amount raised and average daily raise. The survey acts as an additional guidance in helping identify factors that the crowd thinks are relevant in making an investment decision. This supplementary analysis corroborates the robust relationship between variables and identified independent factors, thereby underscoring their pivotal role in bolstering retail investor participation and steering startup funding dynamics.

In summation, the study augments the understanding of the intricate dynamics underpinning equity crowdfunding. By illuminating the substantial impact of variables, the researcher accentuates their significance in fostering investor engagement and shaping the trajectory of startup funding.

Key words: Equity crowdfunding, retail participation, funding, success, platform, campaign.

1 Introduction

Entrepreneurs turned to equity crowdfunding for their funding needs due to factors such as ease of raising capital, lack of private equities that understand the domain and difficulty in meeting capital raising requirements of the traditional channels. It is a lot more convenient for investors to invest through a platform as well as founders to raise capital online. Now through equity crowdfunding entrepreneurs are able to get funding from general public. As startups boom, people have come to realization that it is the angel investor or private equity

that is able to take advantage of early growth of startups and the general public or retail investor crowd is only able to get advantage when these investors are exiting their stakes at the IPO stage. Through this medium even retail is able to make substantial investments at the pre-funding stage. Although the risk appetite of those angel investors is high and involves huge stakes, it was primarily dominated by the High-net-worth investors. Equity funding has solved the problem, making it easier for retail to participate for as low as 100\$ dollars. It is important to see if this is really happening or it is just a new fad and what are the factors that are allowing investors to participate as well as entrepreneurs to raise funds. Also, what parameters are leading to a campaign's success ultimately proving the former two points. Similar to the research on parameters leading to retail participation, some reports have used signals that can be used to determine the success even before it is listed (Kleinert et al., 2021).

1.1 Motivation

The researcher seeks to examine the equity crowdfunding space to analyze the impact that this will have on retail participation and entrepreneurs that want to raise capital at an early stage. This research will help investors and founders to know what parameters influence successful equity crowdfunding. It will help on the below listed factors:

- To verify the relationship between variables that motivate in determining the crowd engagement and ultimately the success of the platform
- To understand the perspective of potential investors
- To gauge the awareness of the investors about the platform and if they are willing to invest in it.
- To ascertain the parameters associated with a successful campaign that help determine whether to invest or not.

1.2 Objective

The research focuses on understanding the relationships of the variables that lead to retail participation. Researcher focuses on broadly three things. Understanding the drivers that are able to explain the changes in retail participation and the funding status of the campaign. Using the data of five platforms to understand the relationship of total valuation and number of raises have on investors. The campaign data set is to be used to determine the relationship the variables have with each other as well as understand their part in the determination of a successful campaign(Estrin et al., 2021).The campaign data set has headers like number of number of raises, total valuation, annual revenue, total amount raised in a month that will help in understanding if they have a trend, ultimately answering our main research question revolving around retail participation, startup funding requirements and parameters (Lo & San-Yih Hwang, 2017) leading to a successful crowdfunding campaign The data taken is from 2016 till 2023.It couldn't happen earlier like other countries due to regulatory challenges and JOBS act (Jegelevičiūtė & Valančienė, 2015) in the US passing late.

As a part of the project, the researcher has considered the data of five US based equity crowdfunding companies namely Seedinvest, StartEngine, Wefunder, Net capital and

Republic and studying the relationship of the campaigns done by them using Logistic regression (Reichenbach & Walther, 2021), descriptive statistics, data visualization, machine learning tools like random forest and Support vector mechanism and principal component analysis. The purpose of taking these 5 platforms is because they are the top five platforms for equity crowdfunding in United states. With two sets of data, one consisting of 4144 campaigns having details of annual revenue, valuation, number of investors, amount raised, revenue multiple and their funding status. The other set is the monthly data of five platforms considered with their total valuation, number of raises and the total amount raised. The paper is more inclined to entrepreneurial finance to ascertain if crowdfunding is propelling retail participation and enabling founders to fund startups using those platforms. It is imperative to find whether these platforms have resulted in funded rather than not funded. Author uses logistic regression as well as regress them independently with the funding status an increasing trend would mean it does answer our research question. The goal is to show through the research that this is forming a new ecosystem of raising capital as well as allowing a diverse and global range of investors to contribute towards it. Some TV shows like “Shark Tank “show us how entrepreneurs come to that platform and raise capital from elite panel of investors. Someone that has no expertise in Artificial intelligence can never make a substantial investment but equity crowdfunding connects a wide plethora of founders and general public that might have domain expertise and connect with the synergy and give funding. The effectiveness and success of the crowdfunding platform will be known from how many successful fundings it has obtained and whether that number is trending upwards with time, that also means retail participation increasing.

A survey in addition to the data is to prove these points asking relevant questions to the mass. Here, the response is taken from the general public whether they are aware of equity crowdfunding and would opt for it as an alternative investment. Followed by asking the impact of equity crowdfunding. Most of the retail investor lack of awareness as the key factor but if they had a chance to diversify and invest in early startups, they would. Most of them considered that this trend of equity crowdfunding is going to increase in the future. For this I have floated my survey on several crowdfunding groups on social networking site like LinkedIn as well as propagated it to those involved in the financial domain.

The use data visualization tools to show if there is a relationship between analyst ratings, funding status, annual revenue and Investors. The researcher establishes a relationship of variables with the help of scatter plots. A regression analysis of the platforms with total valuation and amount raised every month is conducted, the campaign analysis is done with the help of logistic regression as our data is binary is nature. In order to determine whether parameters affect the funding status. This research is useful to entrepreneurs that would gain an insight as to what parameters lead to funding success of their campaign.

1.3 Research Question:

To verify whether a relationship exists between the equity crowdfunding attributes and the funding status, contributing to the determination of retail engagement.

2 Related Work

In the report “Equity crowdfunding: First or Last resort?”, (Walthoff-Borm et al., 2018) the author takes into account the data of 277 firms of UK on leading platform Crowd cube to establish when the entrepreneurs use equity crowdfunding. He uses the pecking order theory to determine that external financing is usually the last source of funding used for early-stage startups. This is in coherence with the report “Working the crowd: Improvisational entrepreneurship and equity crowdfunding in nascent entrepreneurial ventures (Ross Brown, 2017). This research talks about how higher debt levels are responsible for founders opting equity crowdfunding. In this quest, the author finds that except asset tangibility, none of the profitability parameters govern success of equity crowdfunding platforms. (Xavier Walthoff-Borm, 2018) The research answers to the question as to why entrepreneurs come to equity crowdfunding using pecking order theory.

The literature titled “Which updates increase crowd participation” outlines what updates increase crowd participation in equity crowdfunding campaigns. (Jorn Block, 2017) The research length and complexity of updates as the main variables to determine the success of the campaign. The author uses logistic regression and correlation to establish relationship. However, the updates have a positive relationship on the number of investments. Updates in simple language increase participation. The research moves on similar lines but it does not focus on updates, rather proven using data visualization of platform data set by showing increase in total valuation and number of raises over time. A hypothesis test is used determine the p value which gives us a result that it is significant in showing that it is increasing with time and thereby leading more participation of investors. As valuation/capitalization and number of raises increases month on month, the funding as well as general public participation increases.

The author in “Equity Crowdfunding a new Phenomenon”, (Nir Vulkan, 2016) analyzes 636 campaigns ,17188 investors and 64831 investments between 2012 to 2015 from a leading European equity crowdfunding platform determines the success of the crowdfunding platform by taking into account variables like successful campaigns, pre money valuation, number of backers to the project, entrepreneurs and the time taken to back the project. Then divided them into successful and unsuccessful campaigns. After taking the difference between them and noting the characteristics of successful campaigns, it was noted that he was able to predict if the campaign obtained funding or not. Depending on the time taken for investors to back a project also determined whether it is successful or not. The earlier ones backed by investors eventually got the funding and the ones that lagged funding in the initial release date, were not able to receive funding. Our research is more inclined towards determining success of the campaign that directly affects retail participation. For that the author has considered a survey where relevant questions concerning awareness, interest and future trend is investigated.

The report “Success Drivers of online equity crowdfunding campaigns” (Anna Lukkarinen, 2016). This research very closely resonates with our topic of discussion to prove the success of equity crowdfunding campaigns by showing the changes in factors like investors and the

amount raised. Here, the data considered is of a platform from northern Europe where funding target, minimum investment, campaign duration, provision for financial are the independent variables. But none of the variables exhibit a strong positive relationship with the amount raised, in fact minimum investments is strongly negatively correlated with it. Also, the author finds that the decisions taken by angel investors is nothing in common to taken by the crowds.it was inferred that emotional criteria are of more important to crowd funders than financials. Early funding obtained, social media networks and the understanding of the product are the variables that lead to investor participation and amount of funding obtained. A hypothesis testing conducted in the report indicated that influence is a big factor in retail making a decision to invest. The author uses hypothesis testing in platform data set where it proves that total valuation and number of raises are in an increasing trend and increase with time that proves our point that retail participation is increasing.

1

3 Research Methodology

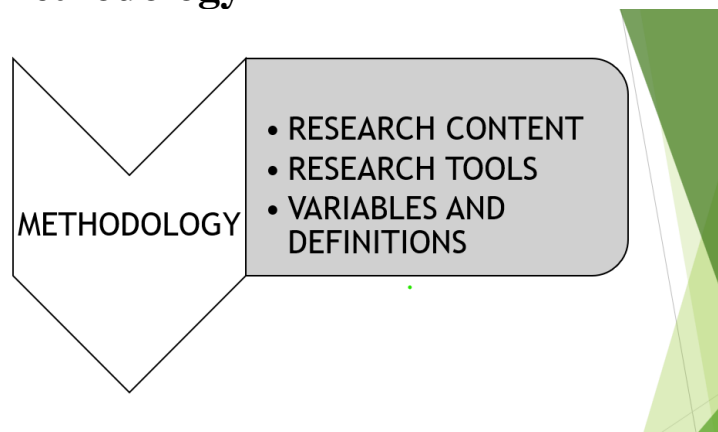


Figure 1 – Methodology

3.1 Research content:

To analyze the research, the author makes use of two data sets and a survey

The data is taken on 25th June 2023. (KingsCrowd, 2023)

1.The first data set contains data of platform and variables like Total valuation, number of raises and average raises per month. It consists of three columns and 325 observations.

2. The second data set deals with campaigns, 4144 campaigns consisting of funding status active, funded and non-funded data with variables like Analyst report, Platform, Valuation/capitalization, Amount raised, annual revenue, revenue multiple, Investors and average daily raise.

3.A survey to general public to know the view on equity crowdfunding as well as startup entrepreneurs to get a clearer vision of the future. Questions pertaining to the exposure of the investors to this alternative investment and whether this asset class is preferred for investment. With reference to the founders, if they are aware of financing through equity

¹ All data sets are taken from Kingscrowd.com that is publicly available on registering for trial

crowdfunding and would prefer to raise pre funding round through equity crowdfunding. The survey questionnaire was floated to those having some experience into finance.

3.2 Research variables and definitions

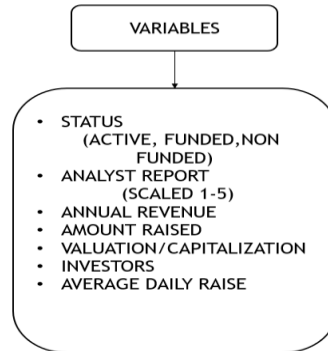


Figure 2 – Variables

Logistic Regression: It captures the changes; the independent variable has on the dependent variable. Here the qualitative variable is converted into binary form. A similar regression is done in the report “forecast success in equity crowdfunding” on the binary dependent variable (Aleksandrina Ralcheva, 2019). The data is tested, trained and used for multiple regression. (David W Hosmer, 2013)

Support Vector machine (V, 1998). The main function of Support vector machine is to find a hyperplane and divide. It separates two data points belonging to different class. The points closest to the planes are known as support vectors. It is able to separate data that is nonlinear in nature. (Alexandros Karatzoglou, 2006). We use R studio to ascertain the classification of our data. It shows how well the model’s predicted the outcomes. Funded or not funded.

Random Forest:It predicts the outcome of the campaign based on the instances provided. A higher accuracy score in random forest indicates a higher chance of predicting the correct outcome Gives accuracy based on the features that are most important for making decisions. Suitable for non-linear data. (Matthias Schonlau, 2020)

3.3 Research Tools

Data Collection and procedure

For the 1st data set.

- This is the data set of the five platform with columns consisting of total valuation, number of raises and amount raised monthly from 2018 to 2023.The analysis is performed in R programming
- A test is undertaken for hypothesis, we name it as (H1) for all 5 platforms and find the p value, H0= Null and Ha= Alternative hypothesis

Here the null hypothesis = Retail participation does not increase that is it is equal to 0

Alternative hypothesis=Retail participation increases with total valuation and number of raises

- The research proceeds by plotting a scatter plot to represent the relationship of variables
- The hypothesis test at 5% confidence level will prove that it is significantly different and can reject the null proving that retail participation increases.
- The summation of mean of every month is done to show the average total valuation of each platform and average number of raises
- The author checks the correlation between total valuation and number of raises.
- The research performs the augmented dickey fuller for each platform to check for unit root problem. If the p value exceeds the 5% confidence it means the data is trending for all platforms and is non stationary, that is it increases with time.

For the 2nd data set

In the second data set the author makes use of campaign data of the companies funded by the five platforms. So, the analysis is like a top-down approach. In order to gain more confidence on the analysis and prove our research question. The columns consist of amount raised, annual revenue, valuation/capitalization, revenue multiples, average daily raise, Investors, Status (Active, funded, not funded) and analyst report. We emphasize more on data visualization to prove our point. There are 4144 campaign data that is analyzed in R programming.

First begin by loading the data set in R and clean the columns for any non-numeric data using conversion function. After that the author makes use of data visualization to plot relevant parameters with each other in order to establish relationship

The plots that we consider are:

- Valuation /capitalization with the amount raised to ascertain whether with increasing capitalization, the amount raised also increases and if it does, it means that equity crowdfunding attracts startups to raise with a high valuation.
- Funding status to the frequency of times the particular status occurs. This is done to infer number of companies in each status. A higher number of funded status would indicate that it is favorable for entrepreneurs to opt for the online platform as the chances of getting funded are higher.
- Number of investors versus amount raised, if the amount raised increases with the number of investors it means there is an increase in retail participation
- Average daily raise over time to check the momentum of average funding daily funding over a period of time. An appreciating figure indicates increasing interest.
- Distribution of funding status by platform, a higher funded status in each platform unanimously indicates a higher probability of obtaining funding. Therefore, leading to more investors getting attracted to raise capital through equity crowdfunding.
- Annual revenue vs amount raised. A higher revenue should attract a higher amount raised to ascertain that with increasing revenues of a company there is higher investor participation.
- Average number of investors vs funding status. A higher number of investors in active and funded indicates wisdom of crowds.

- Average analyst report ratings vs status. A higher rating to the funded status would indicate that it is a key indicator that is seen by investors while investing. As it is a qualitative variable, a visualization chart would be apt at describing whether higher analyst report is associated with funding status and a key parameter taken in to account while making investing decision.
- Investor trend in each platform over time
- Analyst ratings vs investor. More number of investors in a campaign with higher analyst ratings indicate a conjunction with the above point as a key parameter of retail investors while making a decision.

The author performs a multiple regression of the amount raised as dependent variable and valuation, annual revenue, investors and average daily raise as independent variables to show whether these variables have an impact on the amount raised by the platform.

Further finding the F Stat, R square, Adjusted R square and p value.

As the data is qualitative in nature, status (funded, not funded), he makes use of logistic regression as it divides the qualitative data into binary (1,0)

Logit regression is performed using status (Funded=1, Not funded=0) as the dependent variable. Annual revenue, valuation/capitalization, Investors, average daily raise as independent variable to study the impact that these variables have on the funding status. Then separating the testing and training data in the ratio of 80% and 20 %. An accuracy test with F1, precision and recall values that measure the prediction of the conducted logistic regression. Also, the researcher performs logistic regression with each variable if there are any conflicting results. A variance inflation factor is performed to check for multicollinearity before performing random forest, SVM and PCA.

On seeing conflicting results and missing values, the researcher decides to use random forest and support vector machine as well as use the same parameters to check for accuracy. Also replacing the missing values with the mean and considering 80% training and 20 % test data.

Researcher then conducts SVM check for accuracy and again take the logistic multiple regression with the data that has been treated for the missing values with the mean.

A principal component analysis is done to further funnel the independent features that are able to contribute towards proving the research question

Lastly, creating a correlation matrix/heatmap to see how positively or negatively are the variables linked to each other.

Survey:

Survey is under taken as a secondary validation to our research to prove if equity crowdfunding is leading to retail participation and helping entrepreneurs to raise funds. As there are limited variables in the data set, the author conducts a survey as an additional validation to gain confidence on the existing variables as well as additional ones. For this, a survey is created on google form with a sample of 19 questions with a sample of 50 responses are collected by floating a questionnaire on social networking website and crowdfunding groups on LinkedIn

4 Design Specification

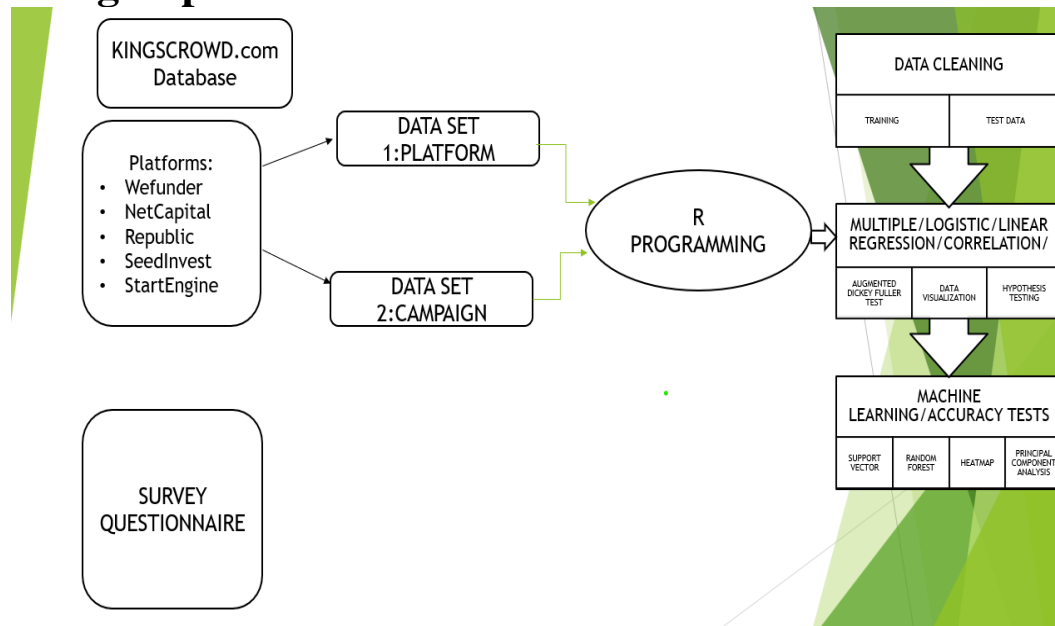


Figure 3 – Design/Architecture

In Figure 3, data is extracted from KINGSCROWD.com for five platform. The data was in unstructured format and had to be converted to structured formed as a csv file, file one has 325 observations for three variables where a hypothesis test is performed and the second data set consisting of campaigns with 4148 campaign history was copied and pasted as a csv file. It had to be cleaned for spaces, dates and formatted to be used for analysis in R. The research uses data visualization to create plots, as the second data set has qualitative variables like “Status” of funding we first perform multiple regression and then use training and testing data in 80% and 20% ratio the perform logistic regression, random forest and Support vector machine and then showing the results. To add confidence, a survey conducted to analyze the awareness, participation in equity crowdfunding and future trend of investment.

5 Evaluation

5.1.1 Analysis Of Dataset 1: Platform Based

The research analyzes the results and predict the data visualization results on the 1st data set and determine the results with the help of plots, machine learning tools and accuracy models in R

5.1.2 Data Visualization

The author plots the chart taking into account to show the total valuation and number of raises of the US based platforms from 2018 to 2023 for platforms SEEDINVEST, WEFUNDER, NETCAPITAL, STARTENGINE AND REPUBLIC

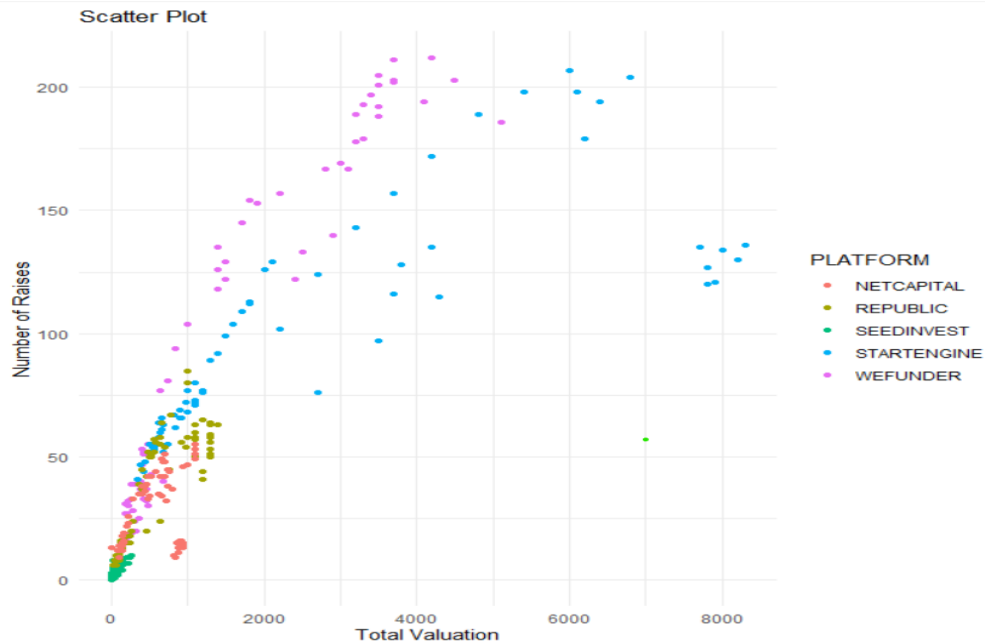


Figure 4 – Data Visualization Total valuation, Number of raises

In Figure 4, by plotting a chart in R, clearly showing the valuation since inception and number of raises progressing with time, showing an increased participation as well as new startups opting for equity crowdfunding

5.1.3 Analysis:

The researcher performs a Hypothesis test (Goethner et al., 2021) for each platform where logic is as follows:

Testing at 5% confidence

For $H_0 = 0$ Retail participation is not increasing with increase with total valuation and number of raises

$H_1 \neq 0$ Retail participation increases with increase in total valuation and number of raises

Hypothesis Test Results:

Table 1 – Hypothesis Test

	TESTING AT 5 % CONFIDENCE	
	TOTAL VALUATION P value	NUMBER OF RAISES P VALUE
WEFUNDER	3.01E-20	1.29E-26
STARTENGINE	3.97E-16	1.39E-11
SEEDINVEST	0.00050551	0.03018746
NETCAPITAL	6.90E-05	2.10E-25
REPUBLIC	4.52E-23	2.93E-12

The p value obtained for all platforms reject the null hypothesis and is significantly different from 0 proving that retail participation increases with increase in total valuation and number of raises.

Correlation

A correlation is conducted between total valuation and number of raises, that shows a strong correlation also a strong proponent of our research question, that, a higher number of raises on platform can command higher valuation.

[1] **0.8180542**

5.1.4 Augmented dickey fuller test:

Augmented dickey fuller test is done to test whether data is stationary or not. If it has a unit root problem or if the data is stationary. Dickey fuller test shows that the data is trending for all the platform. It proves that data is not stationary and has a unit root problem and is trending that proves total valuation and number of raises as trending. (Hornuf & Schwienbacher, 2018) If number of raises is increasing the funding obtained is increasing and earlier, the author proved during the hypothesis that retail participation increases if number of raises and total valuation uptick.

Table 2 – ADF test

	AUGMENTED DICKEY FULLER TEST P VALUE
WEFUNDER	0.6698
STARTENGINE	0.4141
SEEDINVEST	0.4218
NETCAPITAL	0.7693
REPUBLIC	0.4574

The P value data of each platform shows non stationary, having a unit root problem and is trending again inferring that total valuation is increasing with time, means it helps startups raise funds with the increasing valuation on platform

5.2 Data Set 2: Campaign Based

5.2.1 Data Analysis

The campaign data extracted for five platform is saved in a csv file and then it is filtered, cleaned and processed for gaps, dates and for converting non numeric data of the variables to numeric. Then plot the bar chart for companies in platform

The research looks at the funding status of campaigns in all platform. The below plot indicates that there are more funded companies than active and non-funded raises in 4144 campaign datasets, inferring a high probability of success on platforms. Our analysis revolves around dependent variables like Status, Annual ratings, Amount raised and independent

variables Amount raised, Valuation/capitalization, Investors, Annual revenue and average daily raise. A higher frequency indicates a higher probability of receiving funding, ultimately leading to successful funding and leading to retail participation

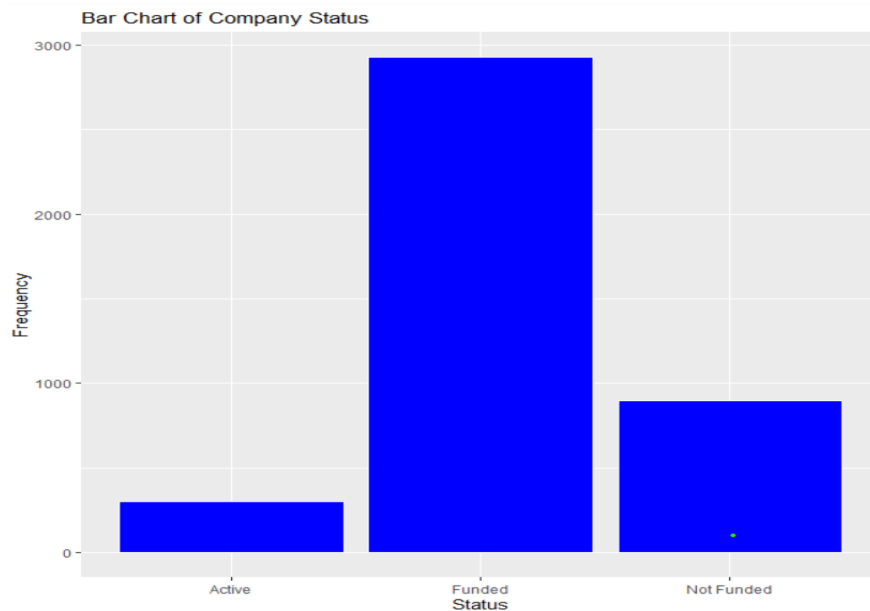


Figure 5 – Frequency of Funding Status

In figure 5. Here we show the visualization of the funding status according to the platform that shows unanimously all five platforms have a higher funded status than active and not funded. Thus, indicating that it is more favorable for a startup that comes to equity crowdfunding to obtain funding

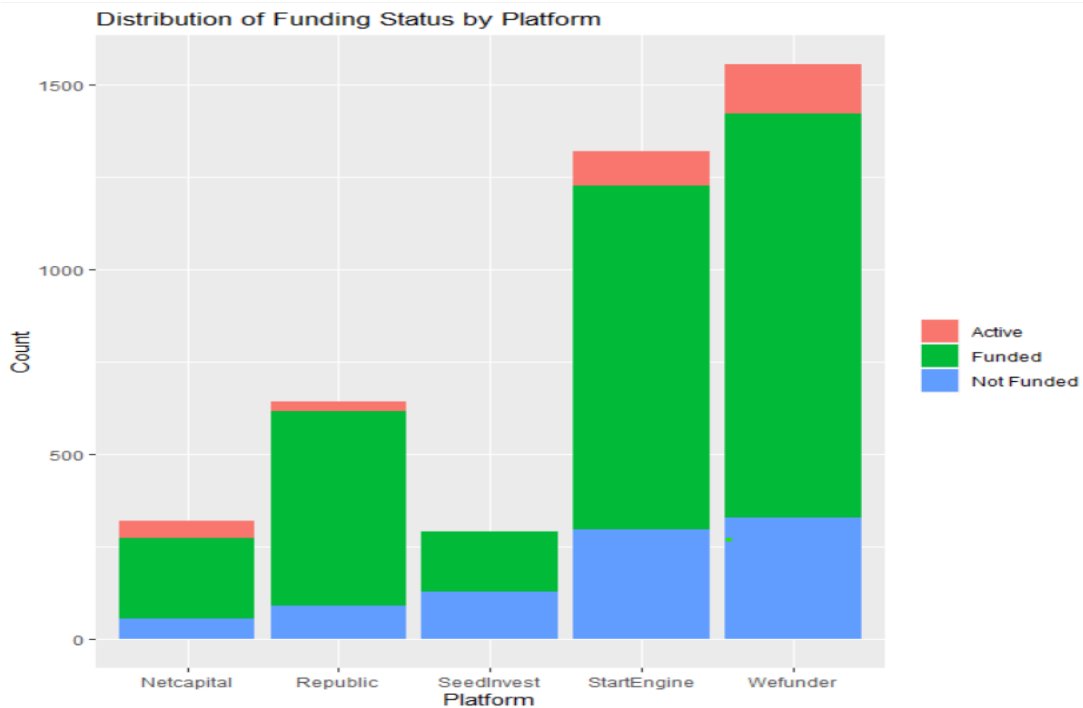


Figure 7 – Platform wise Funded data

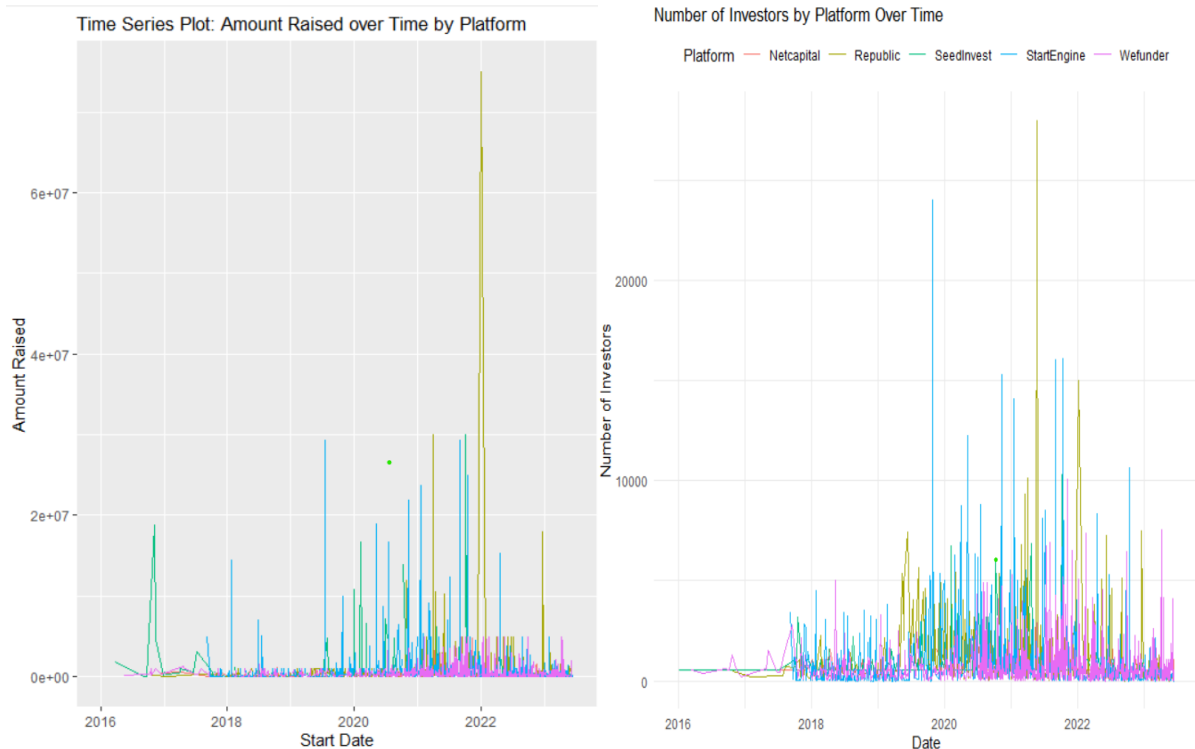


Figure 8 – Amount raised and number of investors over time

In figure 8, We plot the amount raised and number of investors by the campaigns over time according to platform and all of data points depict an increased trend, the objective of showing the funding and retail participation is to support that over the time from 2018 there has been a rise in the amount raised and number of participants month on month.

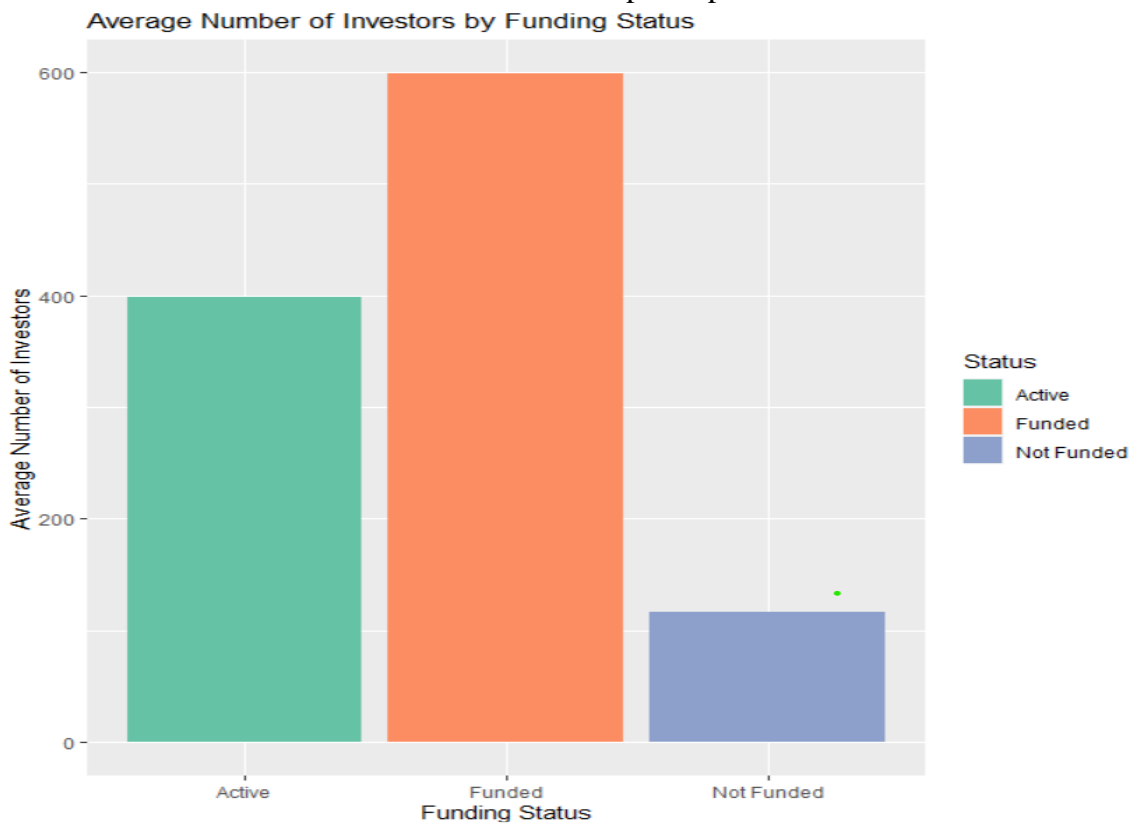


Figure 9 – Number of Investors vs Funding status

In Figure 9, the number of investors in the funded campaigns are higher in the funded as well in the active funds indicating it as an important variable on retail participation, the research also investigates if there is a strong relationship of other variables in determining the success of the equity crowdfunding platform. That signifies, the campaign receiving more investor participation, resulting in the status being funded.

In Figure 10, The amount raised and the number of investors possess a strong correlation, crowd participation increases as the amount raised increases



Figure 10 – Amount raised with the number of Investors

In Figure 11, we use log values of both average daily raise and analyst report, the average daily raise is higher when the analyst ratings go above 3, indicating a strong positive relationship, supporting our objective that investors consider ratings as a factor when making investment decision.

Then plotting visualization of Analyst ratings vs Average daily raise. Again, there is a positive mild correlation, the average daily raise increases as the analyst ratings increase

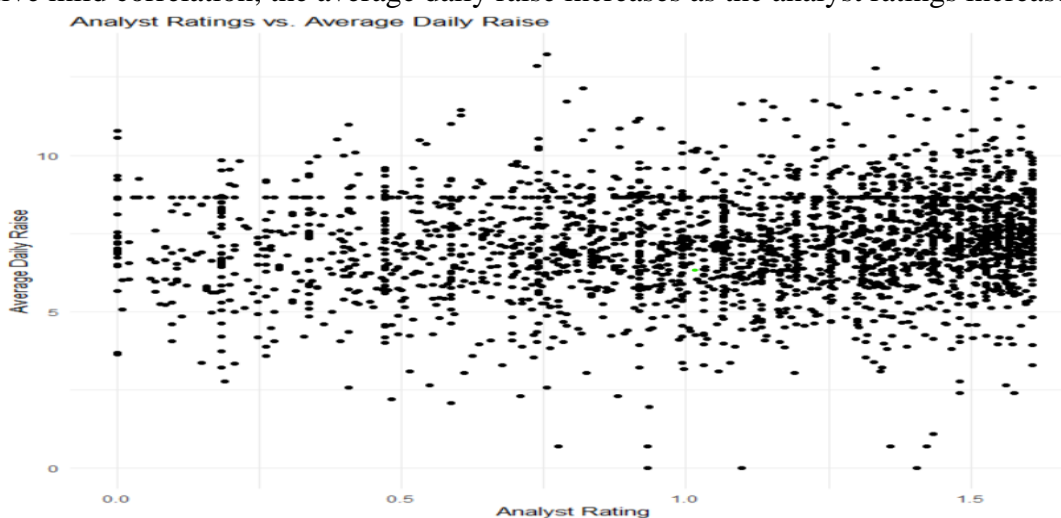


Figure 11 – Daily amount raised with respect to Analyst Ratings

The analyst report or ratings is a key indicator in determining the investment, also witnessed in the regression coming up next, Figure 12 indicates that the number of investor count increases with ratings.

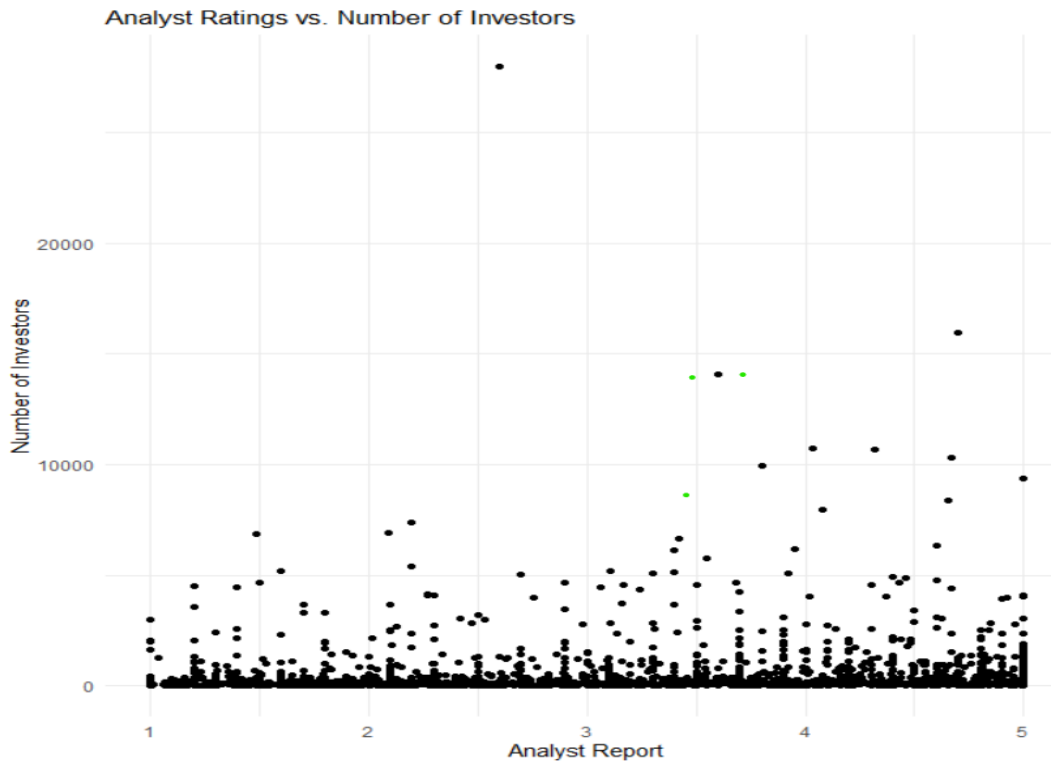


Figure 12 – Analyst Ratings with Number of Investors

5.2.2 Analysis

In the research report we performs a multiple regression with amount raised as the dependent variable and values of valuation/capitalization, investors, annual revenue and average daily raise as the independent variable and finds F stat to be significant , with positive correlation of 0.47 and adjusted R square of 47.26%.The P value is smaller than 5% inferring that the independent variables are able to explain the amount raised by the campaigns and all these variables helping in funding the entrepreneur and lead to success of crowdfunding campaign. P value of Annual revenue is not significant and is not able to explain the amount raised.

Table 3 – Multiple Regression-Amount raised

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.284e+05  3.008e+04  -4.268 2.03e-05 ***
Valuation...Cap  1.562e-03  2.261e-04   6.911 5.68e-12 ***
Annual.Revenue  1.010e-02  6.877e-03   1.468  0.142
Investors       9.297e+02  2.334e+01  39.824 < 2e-16 ***
Average.Daily.Raise  2.072e+01  1.131e+00  18.322 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1559000 on 3567 degrees of freedom
(556 observations deleted due to missingness)
Multiple R-squared:  0.4732,    Adjusted R-squared:  0.4726
F-statistic: 801.1 on 4 and 3567 DF,  p-value: < 2.2e-16

```

5.2.3 Variance Inflation Factor

The Variance Inflation Factor (VIF) is a measure used to assess multicollinearity in a multiple regression model. Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated, which can lead to unstable and unreliable estimates of the model coefficients. Here the variables show there is no multicollinearity and are independent of each other. This test is done to show that independent variables are true estimates without bias.

Valuation...Cap	Annual Revenue	Investors	Average Daily Raise
1.164822	1.049297	1.441030	1.465081

VIF < 5: Low multicollinearity, usually not a concern.

$5 \leq \text{VIF} < 10$: Moderate multicollinearity, may warrant further investigation.

VIF ≥ 10 : High multicollinearity, potentially problematic.

The researcher conducts a logistic regression by taking “status” into account, as it is a qualitative dependent variable, with valuation/capitalization, annual revenue, investors and average daily raise to transform the data for missing values, the author uses the mean values and take training and test data in the ratio of 80% and 20% and log values to conduct the logistic regression

Table 4 – Logistic regression (Train 80%/Test 20%)

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    8.73526    1.04980    8.321 < 2e-16 ***
log(Valuation...Cap) -0.47289    0.06585   -7.181  6.9e-13 ***
log(Annual.Revenue)  0.07263    0.03076    2.361  0.018208 *
log(Investors)     0.23860    0.06282    3.798  0.000146 ***
log(Average.Daily.Raise) -0.08502    0.05691   -1.494  0.135167
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1720.5  on 3301  degrees of freedom
Residual deviance: 1650.7  on 3297  degrees of freedom
AIC: 1660.7

Number of Fisher Scoring iterations: 5
    
```

In Table 4, the P value indicate that all variables except the average daily raise are significant, it means the status of the funding is greatly influenced by valuation/ capitalization, revenue, number of investors funding the project while making an investment decision for a campaign. All the variables except the average daily raise are able to explain the funding status and have an impact on it

The researcher performs the below accuracy tests that will determine the certainty of the positive outcomes. Also, it predicts the accuracy of the models.

Table 5 – Accuracy Table

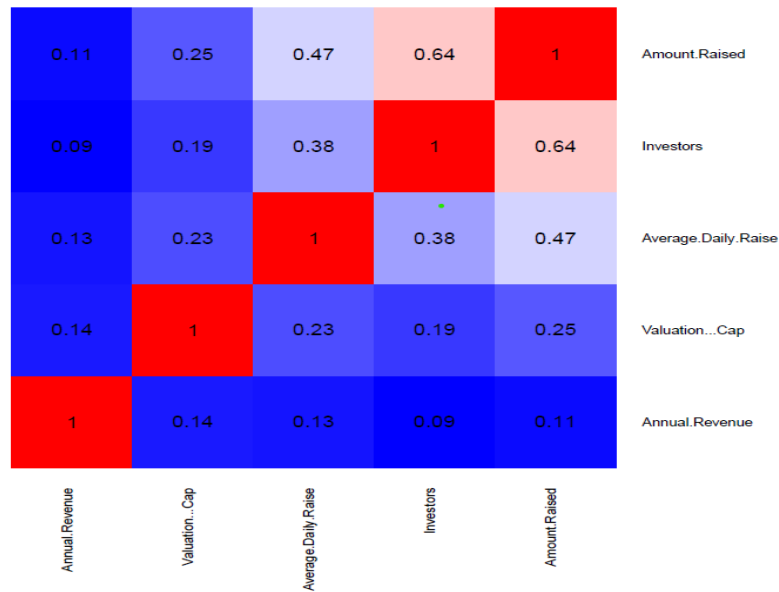
MACHINE LEARNING MODEL	ACCURACY
RANDOM FOREST	0.853
SUPPORT VECTOR MACHINE	0.719
ACCURACY	.707
PRECISION	1
RECALL	.707
F1 SCORE	.828

In table 5, the random forest and Support vector machine give an accuracy of 84.3% and 71.9% in predicting the accuracy of meaning the independent variables can predict 84.3 % percent of the status of funding and support vector machine can predict 71.9% of the funding status. This brings us to the conclusion that these variables help determine the success of the campaign and ultimately variables help as a guidance for investors and entrepreneurs. The accuracy, precision, recall and F1 score indicate there is high precision in prediction of outcomes. In this case, the funded status prediction is the positive outcome.

5.2.4 Correlation Matrix

Table 6- Correlation Matrix

	Valuation...Cap	Amount.Raised	Annual.Revenue	Investors	Average.Daily.Raise
Valuation...Cap	1.000	0.247	0.137	0.192	0.228
Amount.Raised	0.247	1.000	0.109	0.637	0.466
Annual.Revenue	0.137	0.109	1.000	0.088	0.133
Investors	0.192	0.637	0.088	1.000	0.377
Average.Daily.Raise	0.228	0.466	0.133	0.377	1.000



In table 6, all the variables have a positive relationship with each other, although the relationship is not very strong but they have a positive impact on amount raised and investor count. (Abdeldayem et al., 2021)

5.2.5 Principal Component Analysis

Table 7 – Principal Component Analysis

	PC1	PC2	PC3	PC4	PC5	PC6
Analyst.Report	0.092	0.942	0.295	0.110	-0.063	-0.015
Valuation...Cap	0.368	-0.057	-0.238	0.876	0.121	-0.149
Amount.Raised	0.563	-0.066	0.169	-0.104	0.270	0.753
Annual.Revenue	0.214	0.295	-0.859	-0.329	0.137	-0.036
Investors	0.517	-0.094	0.298	-0.302	0.372	-0.637
Average. Daily. Raise	0.476	-0.091	-0.011	-0.100	-0.867	-0.064

In table 7, Principal component analysis (Raflesia et al., 2023) is a dimensional reduction technique that allows reduction in number of features and presents only those features that are of highest relevance. Here, amount raised advocating the highest, followed by Investors, average daily raise of the campaign, annual revenue and valuation/capitalization.

5.2.6 Survey Results

Revenue multiple, number of investors, known industry, analyst ratings and conviction are the top indicators according to the respondents.

How would you decide to invest in startups? Based on what parameters? What parameters you think are most relevant?

49 responses

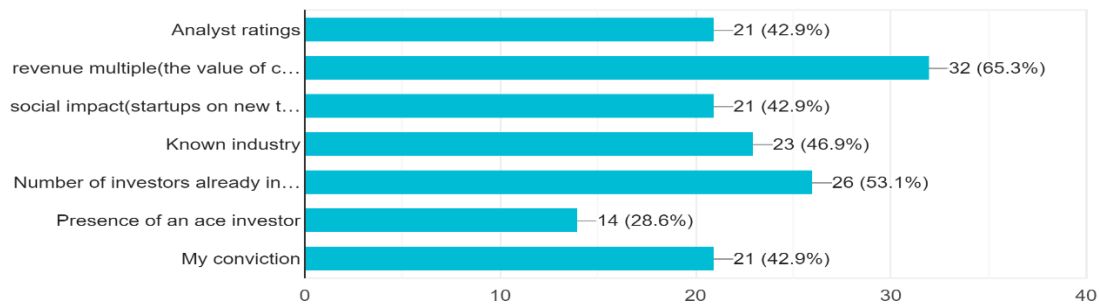


Figure 13 – Independent variables of Importance

The below chart is in agreement with the research on the variables, number of raises, that is showing a trending chart and valuation is increasing with the number of raises, a similar decision resonates in the survey that is in agreement with the investigation conducted through the dataset analysis.

If you decide to invest in a startup through any of the platforms, on what basis would you choose the platform?

46 responses

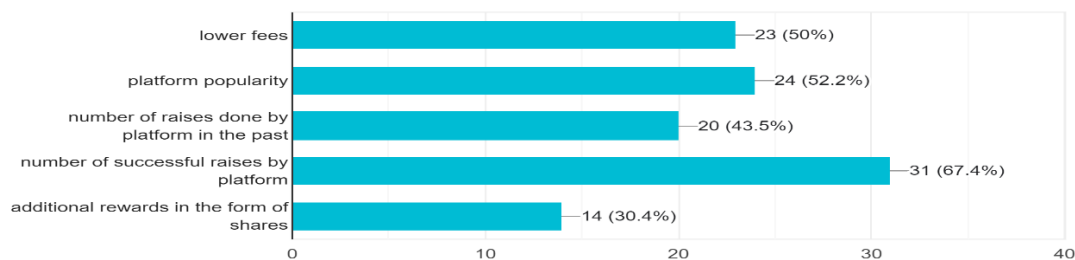


Figure 14 – Retail Investor platform selection criteria

A majority of the participants agree about the inclusion of the crowd in equity crowdfunding and the fact that it is going to help investors participate globally with an upward trend to continue in the near future.

Do you think equity crowdfunding would help participation of general public in return for equity of a company around the world?
50 responses

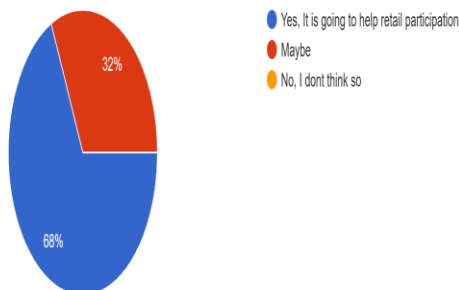


Figure 15 – Retail View

Do you think that raising funds by the way of equity crowdfunding will see an upward trend in future?
50 responses

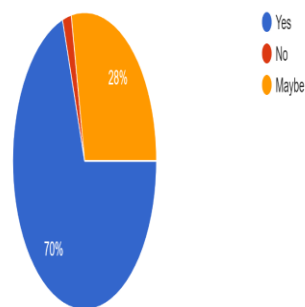


Figure 16 – Future trend of funding

6 Conclusion and Future Work

In conclusion, this report has substantiated its research question through data visualization and analysis of the success factors of the equity crowdfunding platforms, employing both platform and campaign data. The obtained results have been underpinned by robust models illustrating interdependencies among variables, validated using tests such as Hypothesis testing, Augmented Dickey Fuller, logistic multiple regression, data visualization tools, Random Forest, Support Vector Machines, and Principal Component Analysis (PCA). Remarkably, the Random Forest model exhibited the highest accuracy and the logistic regression is able to determine the variables that stimulate the funding status. Number of investors, analyst report, annual revenue and amount raised support the objective of showing increased participation. The research findings unequivocally demonstrate that the attributes considered wield a significant impact funding status.

The analysis meticulously examines key indicators such as campaign status, raised capital, analyst ratings, number of fundraising rounds, and investor count. The analysis deduces a compelling narrative: retail investor participation is witnessing an upward trajectory, and entrepreneurs are experiencing a favourable trend in securing funding. In essence, the analytical approach, augmented by machine learning models, serves as an effective barometer in gauging platform success, effectively reinforcing our research hypothesis.

Furthermore, our research adopts a holistic perspective, including a survey component as a validation measure. The incorporation of the survey is instrumental in providing an impartial outlook. A consensus among respondents looks forward to equity crowdfunding as the future for alternative investment and fundraising.

In a broader context, this research extends a roadmap for entrepreneurs, delineating factors that resonate with the general public. Entrepreneurs are thereby equipped to leverage these insights in their ventures. It's imperative to recognize that this research represents just a fraction of the broader landscape. Subsequent investigations, encompassing an array of variables, promise to unveil even more profound insights.

These are just selected variables available but with time this spectrum in future will broaden and unveil new attributes that would serve as a guidance to entrepreneurs to focus on, so that they are in a better position to concentrate on factors that contribute to a successful raise. A sentiment analysis could also reveal many insights about the behavior of the retail participants. A study could be undertaken globally across cultures to receive more insights on factors that influence the retail investors to invest in equity crowdfunding

7 Limitations

- Equity Crowdfunding is relatively new in United states with Jobs act amendment for regulatory crowdfunding coming into effect in 2012 with usage being started in 2013 (stemler, 2013)
- Limited variables available for campaigns that is free for analysis.

References

- Abdeldayem, M. M., Aldulaimi, S. H., & Alazzawi, A. (2021). Financial Analytics of Crowdfunding Campaigns: data visualization and implications in the Middle East. *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*. <https://doi.org/10.1109/icdabi53623.2021.9655954/>
- Estrin, S., Khavul, S., & Wright, M. (2021). Soft and hard information in equity crowdfunding: network effects in the digitalization of entrepreneurial finance. *Small Business Economics*, 58(4), 1761–1781. <https://doi.org/10.1007/s11187-021-00473-w>
- Goethner, M., Hornuf, L., & Regner, T. (2021). Protecting investors in equity crowdfunding: An empirical analysis of the small investor protection act. *Technological Forecasting and Social Change*, 162, 120352. <https://doi.org/10.1016/j.techfore.2020.120352>
- Hornuf, L., & Schwienbacher, A. (2018). Market mechanisms and funding dynamics in equity crowdfunding. *Journal of Corporate Finance*, 50, 556–574. <https://doi.org/10.1016/j.jcorpfin.2017.08.009>
- Jegelevičiūtė, S., & Valančienė, L. (2015). Comparative Analysis of the Ways Crowdfunding is Promoted. *Procedia - Social and Behavioral Sciences*, 213, 268–274. <https://doi.org/10.1016/j.sbspro.2015.11.536>
- KingsCrowd. (2023, May 6). *Home - KingsCrowd*. <https://kingscrowd.com/>
- Kleinert, S., Bafera, J., Urbig, D., & Volkmann, C. (2021). Access denied: How equity crowdfunding platforms use quality signals to select new ventures. *Entrepreneurship Theory and Practice*, 46(6), 1626–1657. <https://doi.org/10.1177/10422587211011945>
- Lo, P.-C., & San-Yih Hwang. (2017). Incorporating Comment Text into Success Prediction of Crowdfunding Campaigns. *PACIS*, 156. <http://aisel.aisnet.org/pacis2017/156>
- Raflesia, S. P., Lestarini, D., Kurnia, R. D., & Hardiyanti, D. Y. (2023). Using machine learning approach towards successful crowdfunding prediction. *Bulletin of Electrical*

- Engineering and Informatics*, 12(4), 2438–2445.
<https://doi.org/10.11591/eei.v12i4.5238>
- Reichenbach, F., & Walther, M. (2021). Signals in equity-based crowdfunding and risk of failure. *Financial Innovation*, 7(1). <https://doi.org/10.1186/s40854-021-00270-0>
- Ralcheva, A., & Roosenboom, P. (2019). Forecasting success in equity crowdfunding. *Small Business Economics*, 55(1), 39–56. <https://doi.org/10.1007/s11187-019-00144-x>
- Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support Vector Machines in R. *Journal of Statistical Software*, 15(9). <https://doi.org/10.18637/jss.v015.i09>
- Lukkarinen, A., Teich, J. E., Wallenius, H., & Wallenius, J. (2016). Success drivers of online equity crowdfunding campaigns. *Decision Support Systems*, 87, 26–38.
<https://doi.org/10.1016/j.dss.2016.04.006>
- Andersen, P. K. (2002). 3. Applied Logistic Regression. 2nd edn. David W. Hosmer and Stanley Lemeshow. Wiley, New York, 2000. No. of pages: xii+373. Price: £60.95. ISBN 0-471-35632-8. *Statistics in Medicine*, 21(13), 1963–1964.
<https://doi.org/10.1002/sim.1236>
- Block, J., Hornuf, L., & Moritz, A. (2017). Which updates during an equity crowdfunding campaign increase crowd participation? *Small Business Economics*, 50(1), 3–27.
<https://doi.org/10.1007/s11187-017-9876-4>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867x20909688>
- Vulkan, N., Astebro, T. B., & Sierra, M. F. (2016). Equity crowdfunding: A new phenomena. *Journal of Business Venturing Insights*, 5, 37–49.
<https://doi.org/10.1016/j.jbvi.2016.02.001>
- Brown, R., Mawson, S., Rowe, A. K., & Mason, C. (2017). Working the crowd: Improvisational entrepreneurship and equity crowdfunding in nascent entrepreneurial

ventures. *International Small Business Journal*, 36(2), 169–193.

<https://doi.org/10.1177/0266242617729743>

Stemler, A. (2013). The JOBS Act and crowdfunding: Harnessing the power—and money—of the masses. *Business Horizons*, 56(3), 271–275.

<https://doi.org/10.1016/j.bushor.2013.01.007>

Wu, Y., & Vapnik, V. (1999). Statistical learning theory. *Technometrics*, 41(4), 377.

<https://doi.org/10.2307/1271368>

Walthoff-Borm, X., Schwienbacher, A., & Vanacker, T. (2018). Equity crowdfunding: First resort or last resort? *Journal of Business Venturing*, 33(4), 513–533.

<https://doi.org/10.1016/j.jbusvent.2018.04.001>