

Aliah University



2-Year Master of Science in Statistics

Project on “IPO shares in Business Market”

Project By: Abraha Shaik
Masters in Statistics, 2023

Supervised By:
Dr. Hare Krishna Maity
Assistant Professor
Aliah University, Kolkata

Acknowledgement

I would like to express my special thanks of gratitude to my supervisor and mentor **Dr. Hare Krishna Maity, Assistant Professor**, for his enormous support and guidance in helping me complete this project on **IPO shares in Business Market**. He has played a great role in completion of my project, his undivided attention has pushed all way through the project. Without his support this would not have been possible. Sir, has also guided me through the technical aspects and providing me with good supporting sources to complete my work. I would also like to thank my parents for their huge support and encouragement. I would also like to thank my friend for the support and helping me all the time.

Yours faithfully
Abraha Shaik

Introduction

Initial Public Offering is a significant event in a company's journey towards growth and expansion. It is an opportunity for a company to raise funds and expand its operations by issuing shares to the public. The process of IPO in India is governed by the Securities and Exchange Board of India (SEBI), and it involves several steps and regulations that companies must comply with. Understanding the IPO process steps is crucial for companies planning to go public, investors looking to invest in IPOs, and professionals in the finance industry. The initial public offering process is a crucial step for companies seeking to raise capital and expand their business operations. Going public can provide access to a larger pool of potential investors, including institutional investors, mutual funds, and individual investors. This can help the company raise significant amounts of capital and provide liquidity for existing shareholders, such as private investors, allowing them to fully realise gains from their investment.

Everyone knows that investors have a wide range of options when it comes to investing in an upcoming IPO. For the IPO subscription, various slots are offered for different kinds of investors. Each category has a set quota or percentage of the total number of shares that the company wants to list. Depending on the category you applied for, you'll be assigned a certain number of shares. Let's take a look at all the various ways individuals, institutions, and others may participate in a company via the IPO. An investor is someone who puts money into a business in the hopes of seeing a return on that money. In order to make a profit and achieve key financial goals such as saving for retirement, paying for a child's education, or just growing wealth over time, investors use various financial instruments.

Investors:

Qualified Institutional Brokers (QIBs) are non-institutional investors who are registered with SEBI. As an initial public offering nears, underwriters sell significant amounts of shares at a profit to qualified investors in order to raise the necessary cash. The SEBI mandates a 90-day lock-in period for businesses that want to allot more than 50% of their shares to QIBs. **HNIs** are high-net-worth individuals (II) who invest more than Rs. 2 lakhs in a single investment. The NIIs should also be consulted if an institution wants to subscribe for more than 2 lakhs. Investors get their shares regardless of how well the IPO does. **Non-institutional investors** that apply for shares via the book-building procedure up to 2 Lac only are known as retail investors and may be individuals, NRIs, or HUFs. In comparison to institutional investors, their purchasing power is very low, and they wind up paying large trading commissions or fees. However, this charge is removed if they invest online, but owing to a lack of market understanding, these investors choose that path. Retail investors, on the other hand, will be able to purchase 35% of the stock.

Underpricing and Overpricing:

Underpricing is an IPO technique where a company lists the stocks at a price below the original face value. As a result, the stock value increases from the IPO date to the first-day closing price. Primarily, corporations underprice IPOs to promote stock demand. Various IPOs are released every week, and the demand varies. Firms try to stand out; low-priced stocks could potentially achieve that objective. Sometimes, it is mere publicity. Low-priced stocks can act like free advertisement—it is more likely to be covered by reporters and speculators. As a result, the price rises steeply on the first day. Sometimes, investors

lack knowledge about a particular company. Usually, informed buyers trade long-term. Uninformed investors bid randomly and lose money. But once the latter loses money, they stop bidding.

Companies tend to overprice their IPOs for a variety of reasons, including to maximize their fundraising and to create a sense of hype and demand around their shares. However, overpriced IPOs can be detrimental to investors who end up paying more for shares than they are actually worth. **Overpricing** the IPO can lead to a rapid fall in prices, even though the higher price benefits the underwriting bank issuing the stock since it only makes money on the initial issue. This is a good way to gain more capital, but it is not risk free. By overpricing an IPO underwriters will have difficulty fulfilling their obligation to sell shares. Furthermore, if all shares are sold, there is a chance that the share price will plummet on the first day of trading, resulting in a loss of value. The most efficient IPO pricing method is to establish an offering price that is low enough to attract interest, yet high enough to raise a decent amount of capital for the company.

Any individual who is an adult and is capable of entering into a legal contract can serve the eligibility norms to apply in the IPO of a company. However, there are some other inevitable norms an investor needs to meet.

The eligibility criteria are :

- It is required that the investor interested in buying a share in an IPO has a **PAN card** issued by the Income Tax department of the country.
- One also needs to have a valid **Demat** account.
- It is not required to have a trading account, a Demat account serves the purpose. However, in case an investor sells the stocks on listings, he will need a trading account. It is often advised to open a trading account along with the Demat account when an investor is looking forward to investing in an IPO for the first time.

The present study offers a distinct contribution to IPO literature in general and survival in particular. Apart from issuers, investors can evaluate the issue, market, and company specific factors in order to ensure that their decision to invest in an issue should turn out to be profitable in the aftermarket. In practical terms, the findings of this study can inform public policy decision makers who are concerned with regulating the market. In other words, the study would provide a base for the regulators and policy makers to update their laws and formulate such kind of policies that would not only create a lucrative and more sustained market but will also protect the interest of investors in the aftermarket.

In nutshell, the significance of analysing the most fitted IPO is immensely fruitful for every associated party of an IPO. The survival profile of IPOs varies across the several industries as well. The findings of this study will have fruitful implication for the issuers, investors, regulators, and the entire capital market as they can evaluate the future prospects of IPOs and can take rational decisions accordingly. Going public is an important phase in the life cycle of a company. The first stage in a company is generation of an entrepreneurial idea or concept that is initially nurtured with private equity capital. Then, at a subsequent stage in its development, the firm attempts to raise additional capital through an IPO.

Data and Methodology

The data in this project is obtained from the author Bala Baskar(Kaggle), Kaggle and from Moneycontrol. This IPO dataset contains the data about companies that did their initial public offering in the Indian capital market. Also, this dataset will help us to analyse the market sentiment over time based on IPO listing and which industry had good positive sentiment over time. This dataset has well defined timely company IPOs with each containing its gains through time. I have personally adjusted some data to make it more reliable and to go with the analysis process. The dataset also contains an additional column of fitted gain values, to make it easy to check the accuracy for the analysis. To make it more understandable the dataset has company names so that it would be easy to go through their sectors.

The dataset contains 15 variables and 262 no of datas in each variable. On successful completion of the IPO process, a private company becomes a public limited company. The IPO listing allows a company's shares to be traded on the stock exchanges. IPO investments can also be a good way to realize profits in a short term. When the demand for an IPO is high, it lists on the stock exchanges at a premium. At times these premiums provide a substantial return to the investors. Through any given IPO, a set number of shares of a certain value are allotted to members of the public. You, the investor, may want a specific number allotted to you, but this is as per the discretion of the company allotting the shares.

Introducing with the variables

- Date: Date at which IPO got listed.
- IPO Name: Name of the company.
- Issue Size (in crores):The amount expected by company in turn selling their shares.
- QIB: xTimes subscribed by Qualified Institutional Bidders.
- HNI: xTimes subscribed by High Net worth Individuals.
- RII: xTimes subscribed by Retail Individual Investors.
- Total: Total subscription.
- Issue: Issue price/share in Indian Rupees.
- Listing Open: Open price on listing day in Indian Rupees.
- Listing Close: Close price on listing day in Indian Rupees.
- Listing Gains %: Percentage gain on listing day (Open - Issue price).
- CMP: Current Market price in Indian Rupees as on 31-07-2021.
- Current Gains %: Percentage gain on listing day (CMP - Issue price).

Data Preprocessing

- I have added two additional column into the dataset for better and efficient analysis.
- "Sectors" and "difference", the first one defines the industrial sector of which the IPO Company belongs to and the latter defines the fitted values after modelling with keeping explanatory variable as "Listing Close"
- I have also created a seperate sheet for the counting of IPOs for each year and sector wise
- There were no missing values in the data, so no exclusion of data was needed.

Now, for our analysis steps I have decided to go through numerous statistical methods within my knowledge to interpret on the basis of my objectives. Here I have taken Linear Regression Model into consideration for my initial analyzation process. As we all know regression analysis is more versatile and has wide applicability. Regression analysis allows you to understand the strength of relationships between variables. Using statistical measurements like R-squared / adjusted R-squared, regression analysis can tell you how much of the total variability in the data is explained by your model. In simpler terms, if you give a regression model 50 features, you can find out which features are good predictors for the target variable and which aren't. Regression analysis can give a confidence interval for each regression coefficient that it estimates.

With each summarising of linear model, we can obtain its significant codes to know about the significancy of the factors in that model and also its model accuracy with the help of R-Square values. A fitted value is a statistical model's prediction of the mean response value when you input the values of the predictors, factor levels, or components into the model.

Now as mentioned earlier to know about the quality of prediction we consider doing RMSE, MAE and MAPE in which Root Mean Square Error or root mean square deviation is one of the most commonly used measures for evaluating the quality of predictions. It shows how far predictions fall from measured true values using Euclidean distance. To compute RMSE, calculate the residual (difference between prediction and truth) for each data point, compute the norm of residual for each data point, compute the mean of residuals and take the square root of that mean.

RMSE is commonly used in supervised learning applications, as RMSE uses and needs true measurements at each predicted data point. Root means square error can be expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}} \quad (1)$$

where N is the number of data points, y(i) is the ith measurement, and yhat (i) is its corresponding prediction.

And on the next case, I used Mean Absolute Error which is a measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a regression model. The MAE loss function formula:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y(i) - \hat{y}(i)| \quad (2)$$

where n is the number of observation in the dataset, y(i) is the true value and $\hat{y}(i)$ is the predicted value.

Thereafter to measure the forecast of the error I used Mean Absolute Percentage Error where it is used to forecast error, probably because the variable's units are scaled to percentage units, which makes it easier to understand.

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (3)$$

where n is the number of fitted points. A_t is the actual value, F_t is the forecast value.

Now, I have also considered of taking Generalised Linear Model into process. Because as I have considered taking “difference” as my response variable it is difficult to process linear model into it, so applying GLM is a good choice. GLM models allow us to build a linear relationship between the response and predictors, even though their underlying relationship is not linear. This is made possible by using a link function, which links the response variable to a linear model. Unlike Linear Regression models, the error distribution of the response variable need not be normally distributed. The errors in the response variable are assumed to follow an exponential family of distribution (i.e., normal, binomial, Poisson, or gamma distributions).

Now to show how well the model predicts the response variable with only the intercept, we use null deviance. The size of the **null deviance** can determine how well the model explains the data. For example, if the null deviance is very small, it most likely means the model is doing well explaining the data. The null deviance helps us to understand if additional independent variables are needed based on the degree of freedom and serves for comparing how much the model has improved by adding the predictors or independent variables. The **residual deviance** tells us how well the response variable can be predicted by a model with p predictor variables. The lower the value, the better the model is able to predict the value of the response variable.

Pseudo-R-squared values are used when the outcome variable is nominal or ordinal such that the coefficient of determination R^2 cannot be applied as a measure for goodness of fit.

Now for visual representation I have also used descriptive methods like **Box-Plot, Sigmoid Curve**. Etc.

A box and whisker plot—also called a box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum. In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum. The method to summarize a set of data that is measured using an

interval scale is called a box and whisker plot. These are maximum used for data analysis. A box plot is a chart that shows data from a five-number summary including one of the measures of central tendency. It does not show the distribution in particular as much as a stem and leaf plot or histogram does. But it is primarily used to indicate a distribution is skewed or not and if there are potential unusual observations (also called outliers) present in the data set. Boxplots are also very beneficial when large numbers of data sets are involved or compared.

A sigmoid function is a mathematical function with a characteristic “S”-shaped curve or sigmoid curve. It transforms any value in the domain to a number between 0 and 1. We typically denote the sigmoid function by the greek letter σ and define as:

$$\sigma(x) = \frac{1}{1 + e(-x)} \quad (4)$$

where x is the input to the sigmoid function and e is the Euler’s Number.

A logistic growth curve is an S-shaped (sigmoidal) curve that can be used to model functions that increase gradually at first, more rapidly in the middle growth period, and slowly at the end, levelling off at a maximum value after some period of time. The initial part of the curve is exponential; the rate of growth accelerates as it approaches the midpoint of the curve. At the midpoint ($K/2$), the growth rate begins to decelerate but continues to grow until it reaches K which is called the “Carrying Capacity” for the environment.

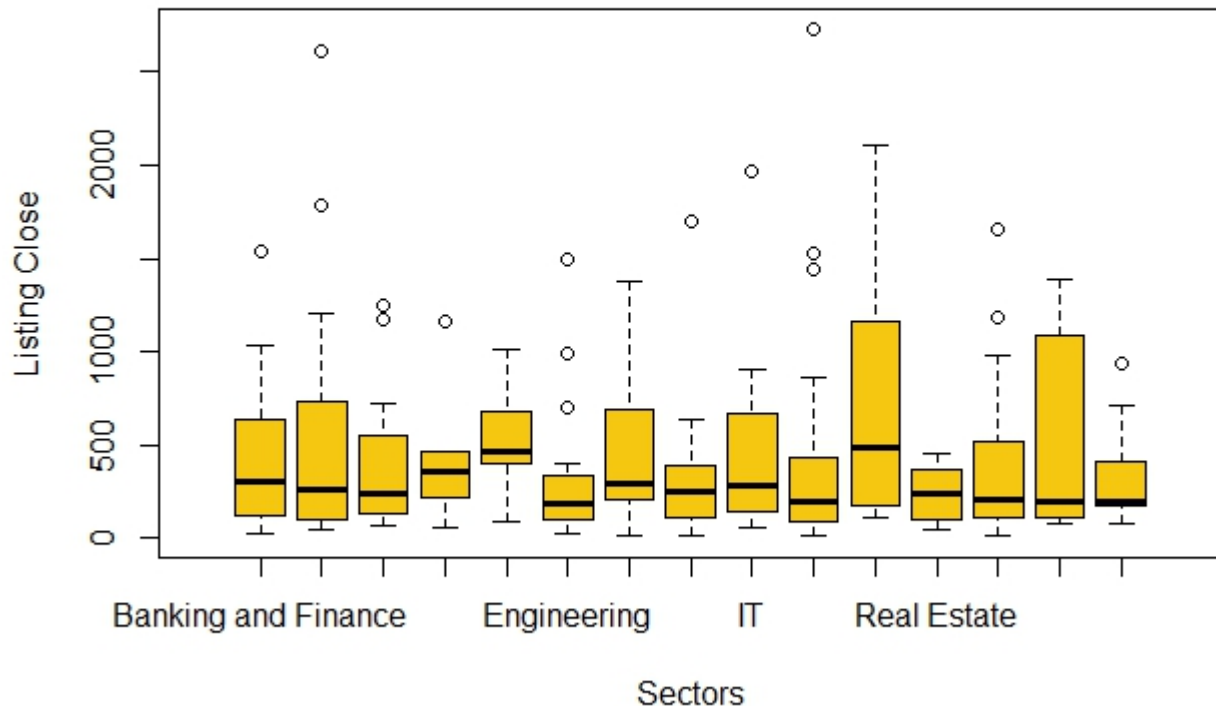
Objectives

- To check the significant factors for the reason behind IPO being Overpriced and on which factors does the Listing Gains depend.
- Market Sector wise Listing Open price comparison.
- Using statistical methods to ensure the accuracy of the predictive model.
- To visualise the effect of factors on Listing Gains.

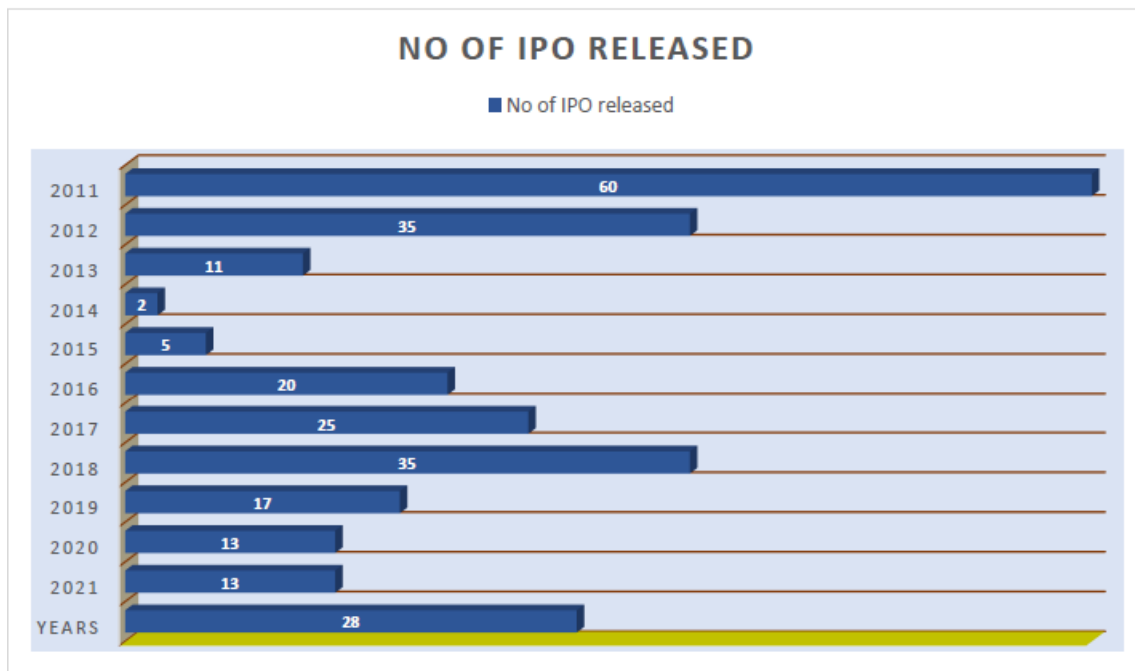
The main idea of evaluating and analyzing the datas is to check which factors are responsible for higher profit or higher loss. If with the analyzing we can come to fact that there are some things to look over before investing or buying stocks in IPO. And with the help of visualization it can easily be understood with diagramatic representation of the increase in number of IPO releasing with year and at what year the stocks have been in bad performance.

Data Analysis and Visualization

Here we can completely come to the part where it can be found out about the variability of Listing Open price for different Sectors in the market by looking to its median, range and the quartiles. This comes to the fact that one of the objectives is achieved.

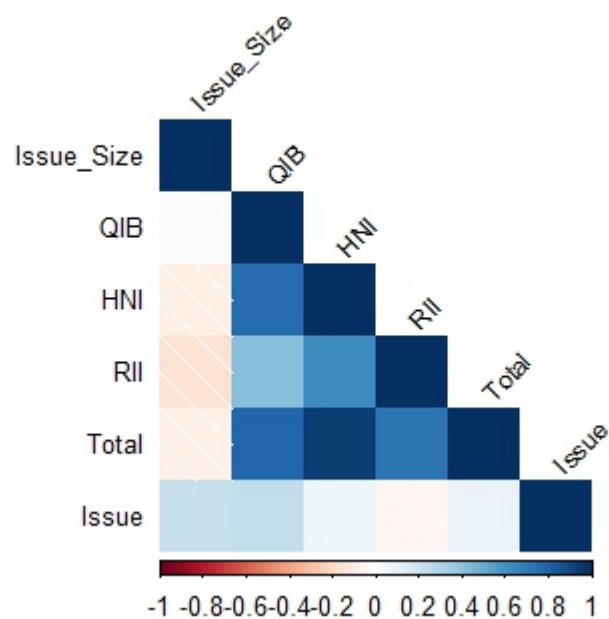


We can also use some other diagrammatic representations to look into different other objective like the no of IPOs released with time (per year) using bar plot.



"Here we have a bar plot which signifies number of IPOs released per year (2011-2021). We can understand 2011 has the highest no of IPOs in these 10yrs span followed by 2012, 2018"

For more precise data visualization I have considered finding correlation between variables (Issue Size, HNI, QIB, RII, Issue, Total) with the help of correlation plot in R.



With this plot we can look that the lighter shades represent less correlation between the variables mentioned for each and the darker shades shows higher correlation with white shade having 0 correlation.

Now to complete other objectives we need to go through modelling of dataset with appropriate variables.

Firstly, to create a linear model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (5)$$

using response variable or Y as Listing Close and others as explanatory variables or X (QIB, Total, HNI ,RII, Issue Size, Issue, Listing Gains) we can see the following result,

Let us consider this model as “Model 1”,

Variable	Values
Model	Model 1
Adjusted R-Square	0.8283
Multiple R-Square	0.8421
% of model efficiency	86%

“Model 1” has 4 significant factors these are,

Factors	p_values_range
Listing Gains	[0,0.001]
Issue	[0,0.001]
QIB	[0.001,0.01]
Total	[0.01,0.05]

With this we can come to the fact that Listing Close is dependent on few highly significant variables which are Listing Gains, Issue and QIB causing its changes in market.

Now to find RMSE, MAE and MAPE we used the fitted values and found out our result,

Model	Model_1
RMSE	165.920500
MAE	93.848390
MAPE	2.066678

Again, we took another with keeping response variable as Listing Gains and explanatory variables as Total, HNI, RII, QIB, Issue, Issue Size, Listing Open.

Let us also represent this model also as “Model 2”,

Model	Model_2
Adjusted R-Square	0.335
Multiple R-Square	0.3885
% of model efficiency	33%

This model seems to have 4 significant factors,

Factors	p_value_range
Listing Open	[0,0.001]
Issue	[0,0.001]
Total	[0.001,0.01]
Issue Size	[0.01,0.05]

Now, we will be performing Generalised Linear Model,

$$h\mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} \quad (6)$$

keeping “difference” as response variable and explanatory variables as, QIB, Total, RII, HNI, Issue, Issue Size, Listing Open, Listing Gains.

For Model 3 we get the following summary values,

Model	Model_3
Null Deviance	358.25
Residual Deviance	204.21
AIC	250.21

The significant factors for this model are,

Factors	p_value_range
Listing Gains	[0,0.001]
Listing Open	[0,0.001]
Issue	[0,0.001]
Total	[0.01,0.05]
RII	[0.05,0.1]

Now to find the accuracy of this model, I have used the fitted values and converted into **confusion matrix**,

CM	X_0	X_1
X_0	121	69
X_1	28	44

With this we calculate the accuracy by,

$$= (121 + 44)/262 = 0.6297 = 62.97\%$$

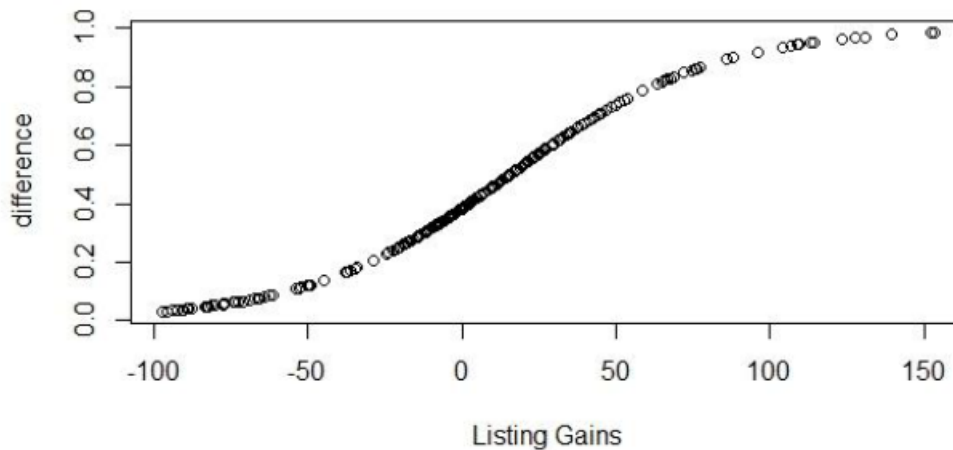
This shows us that the model we have created is **62.97%** accurate upto my knowledge.

Pseudo-R-squared values are used when the outcome variable is nominal or ordinal such that the coefficient of determination R^2 cannot be applied as a measure for goodness of fit. It explains how well the regression model is fitted for observed data.

Pseudo R-Square values = 0.4299669

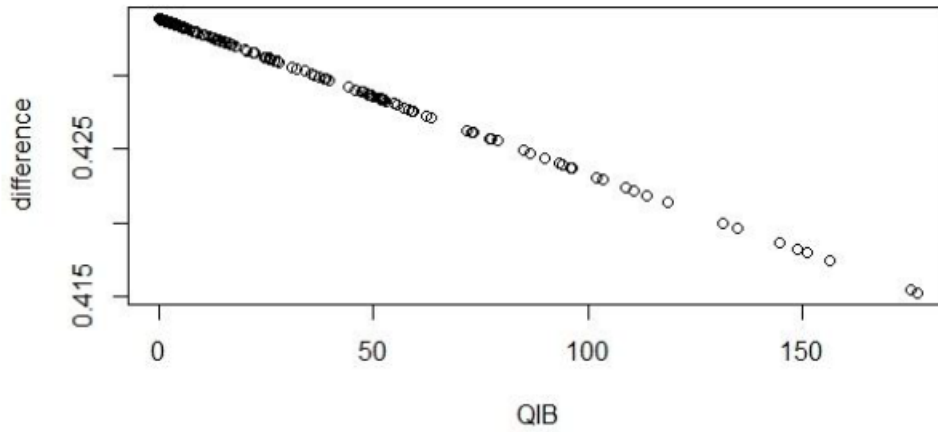
Lastly, plotting a **sigmoid curve** for “difference” with respect to Listing Gains, QIB, HNI, RII.

- difference ~ Listing Gains



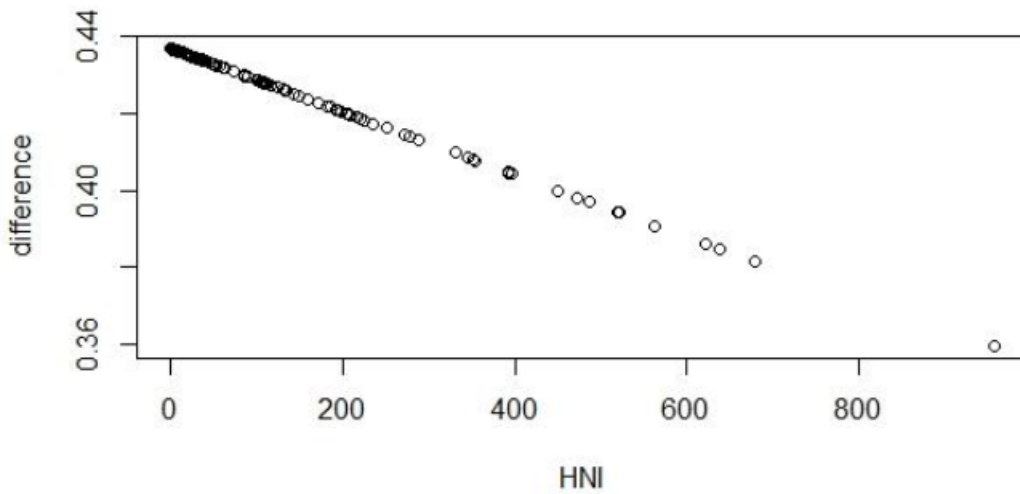
From the above plot we can ensure that the predicted values of the model with respect to the Listing Gains is gradually increasing in its exponential growth stage forming a slope like structure.

- difference \sim QIB



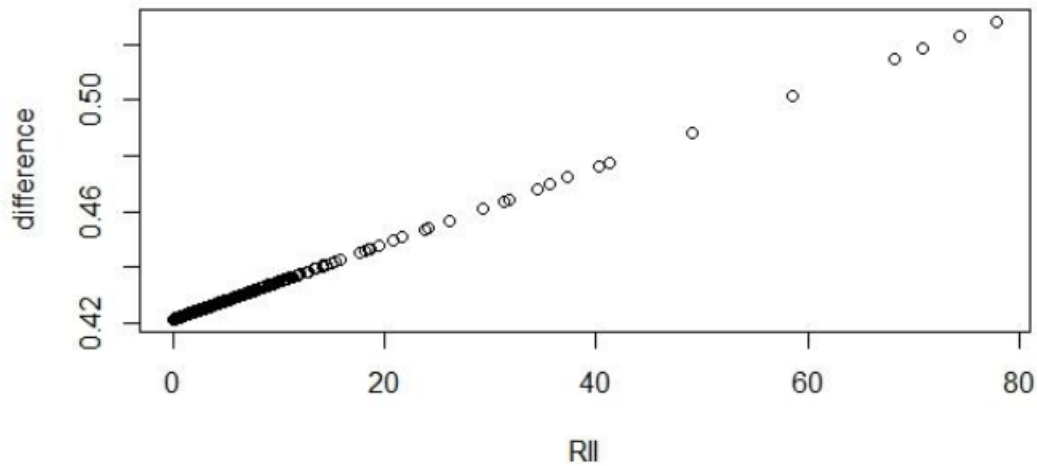
In the above diagram, we can actually see about the stiff decrease of values with increase in QIB, which defines that with increase in QIB the predicted values decrease.

- difference \sim HNI



Here also we can look its gradual decrease of fitted values with respect to the HNI values. This depicts that with increase in HNI our difference values decrease.

- difference \sim RII



But in this case, there has been sharp increase of values of difference as compared to the RII values. And look that there has clustering of values more between 0 and 19 of RII.

Conclusion

With this we come to the conclusion that Stock Market seems to be a very risky place to invest and IPO being its one way of investing has its own way of running into the market. Through some factors affecting the Investment price and gaining through it must be thoroughly checked. From this project we had concluded the risking factor affecting the IPO, the variables profiting the Return on Investment (ROI). With the help of statistical modelling the relationship of variables with different factors has been noted and based on our analysis one can easily think of entering into the stock market with the help of analysis performed. Visualization of factors in this project can also help you make understand the risking or the profiting factor easily.

This project has considered the important reason needed to think of in the market analysis and the performance is completely dependable on the company listing their prices, therefore the modelling has been done on the basis of company's listings. We should always invest with taking into mind the risks of the stock market as the stock exchanges keeps on dipping and rising.

Reference

- An Introduction to Statistical Learning with Applications in R
- Peter Dalgaard, Introductory Statistics with R
- Performance Analysis of IPOs in the Indian Market
- Regression Analysis by Example ,Fifth Edition, Samprit Chatterjee and Ali S. Hadi
- balabaskar indian ipo dataset 2010 2021