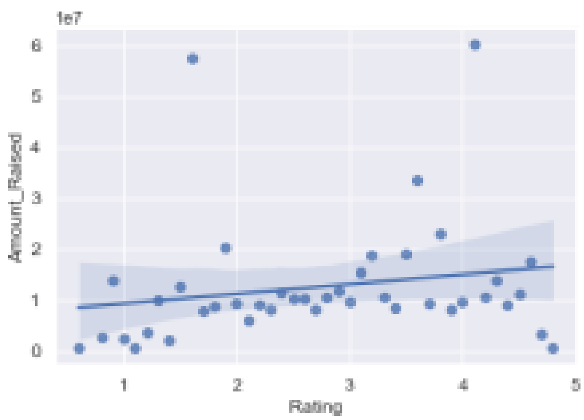
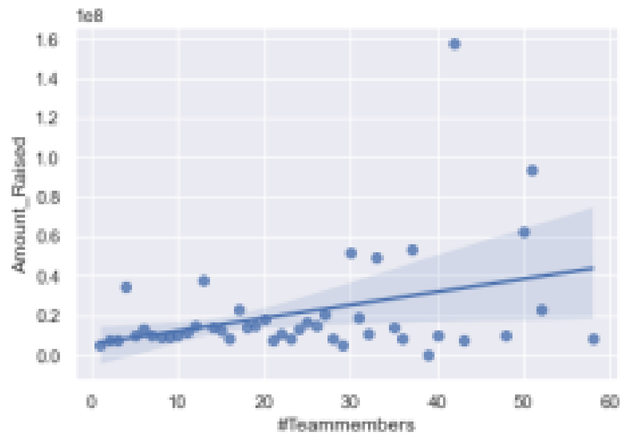
Determining Parameters of Success on Initial Coin Offerings (ICOs)

The final project comprises four parts. It analyses the effect of a number of variables on the success and performance of Initial Coin Offerings through descriptive and predictive analysis. The analysis conducted on the data is hereby discussed in detail:

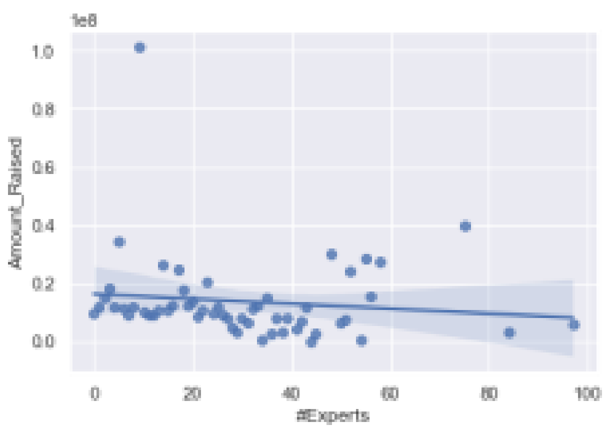
Part I

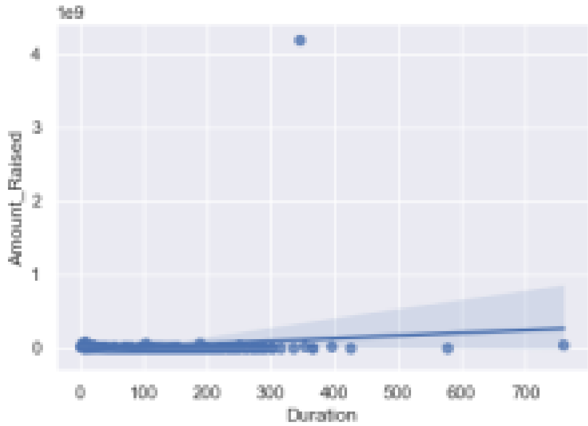
The data on the ICO is analysed through graphs to visualize the connections amongst different variables and how they affect whether the ICO was successful or not. Through data visualization, it is observed that ICO success is largely dependent on the amount raised during the offering and the soft cap of the ICO. The Amount Raised also increases with Ratings provided by the Experts and the Number of Team Members.

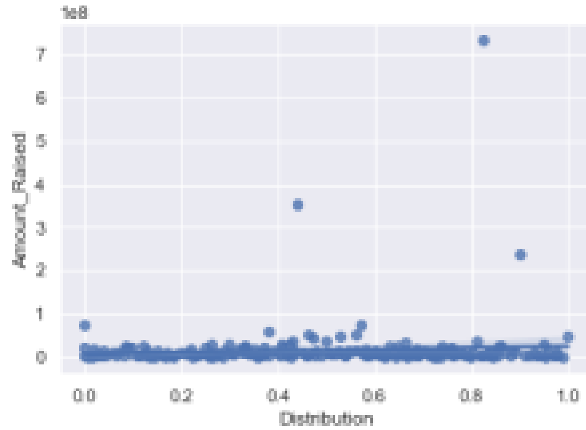




The Amount Raised, however, decreases with an increase in the number of Experts providing the Ratings. There is no visible change in the Amount Raised with respect to the Duration of the ICO and the Distribution (the ratio of the tokens offered in sales to the total supply of tokens).





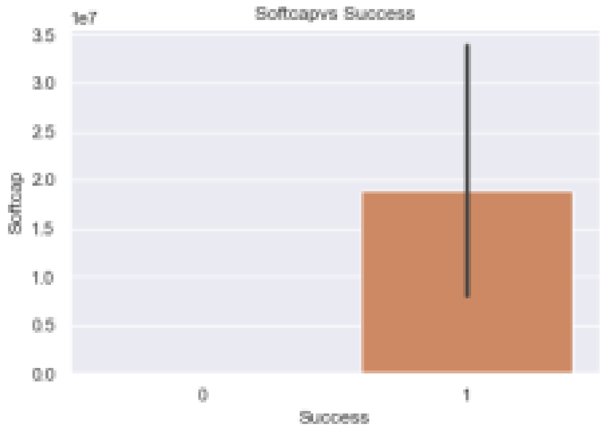


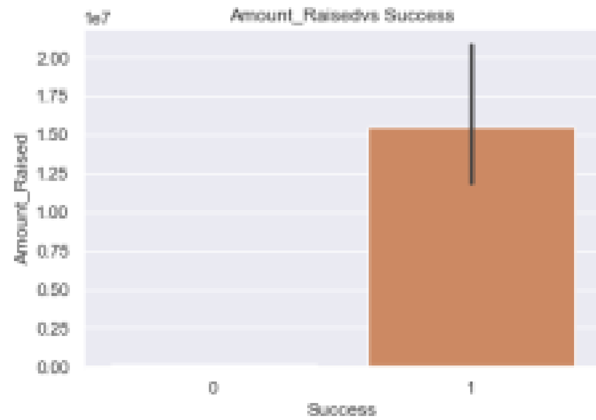
There are numerous missing values in the data set. All the missing values are treated in the most suitable way. Numeric missing values, including Duration, Number of Team Members, Number of Experts, etc. are treated by putting in the average values from the data set. The missing values in Softcap and Hardcap are simply replaced with zeroes, assuming that the ICO is not launched at all if the minimum amount is not met. If there are missing values in the Restricted Countries, it implied that there are no restricted countries for that ICO: so essentially a new country category is derived as “None” to treat the missing values in that field. To treat the missing values in the Token field, the rows are dropped.

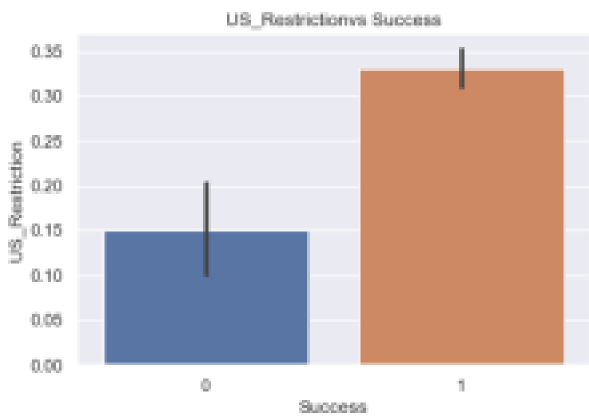
The data is then split into a 70:30 train/test ratio for regression purposes with the independent variables as those factors which visually appear to have a substantial effect on the Amount Raised, namely 'Token', 'Softcap', 'Hardcap', 'Start', 'End', 'Quarterstart', 'Duration', 'Country', 'ERC20', 'Rating', and '#Experts. The dependent variable is the Amount Raised. Dummy variables are created for categorical variables.

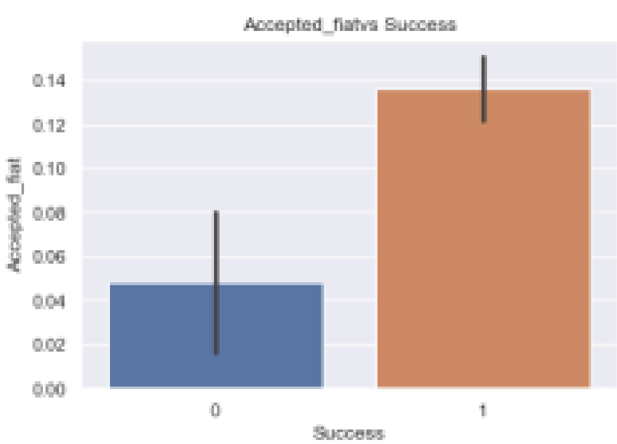
Part II

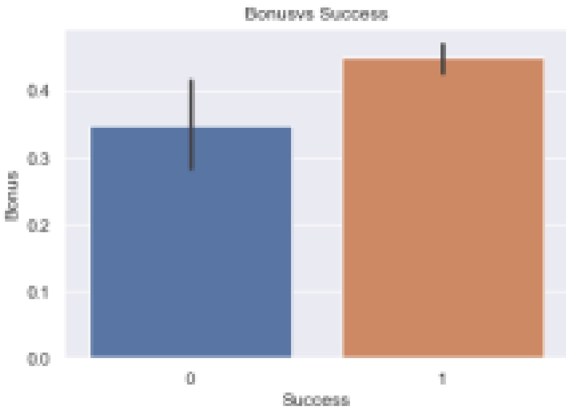
It is given that two factors determine the success of an IPO at the same time: the ICO should ascend to the SoftCap or if it fails to do so the Amount Raised should be greater than 0.5 million dollars. Otherwise, the ICO is deemed unsuccessful. Therefore, a new variable is determined as “Success” which gives a binary value of 1 if the conditions above are met. The IF function is used to test the values and generate a new column of Success for each ICO.
Then a graphical analysis is made of the variables to determine which variables are the most important in deciding the success of the ICO.

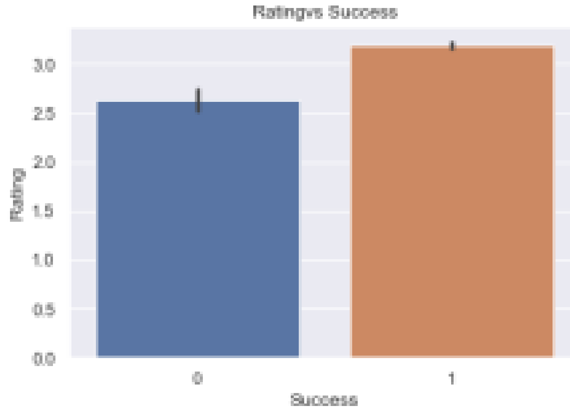

Softcapvs Success
1e7
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Softcap
0
1
Success

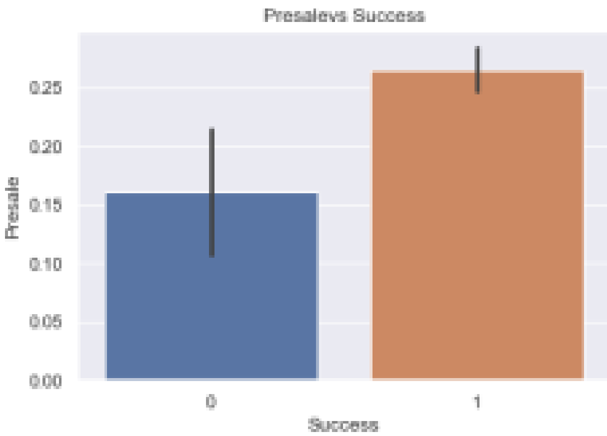

Amount_Raisedvs Success
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Amount_Raised
0
1
Success

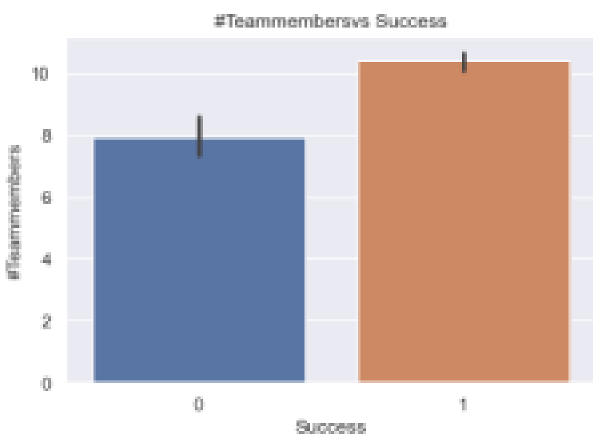

US_Restrictionvs Success
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
US_Restriction
0
1
Success

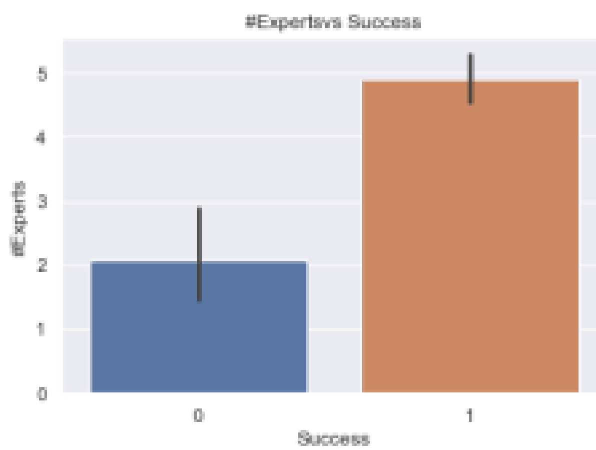

Accepted_fiatvs Success
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Accepted_fiat
0
1
Success





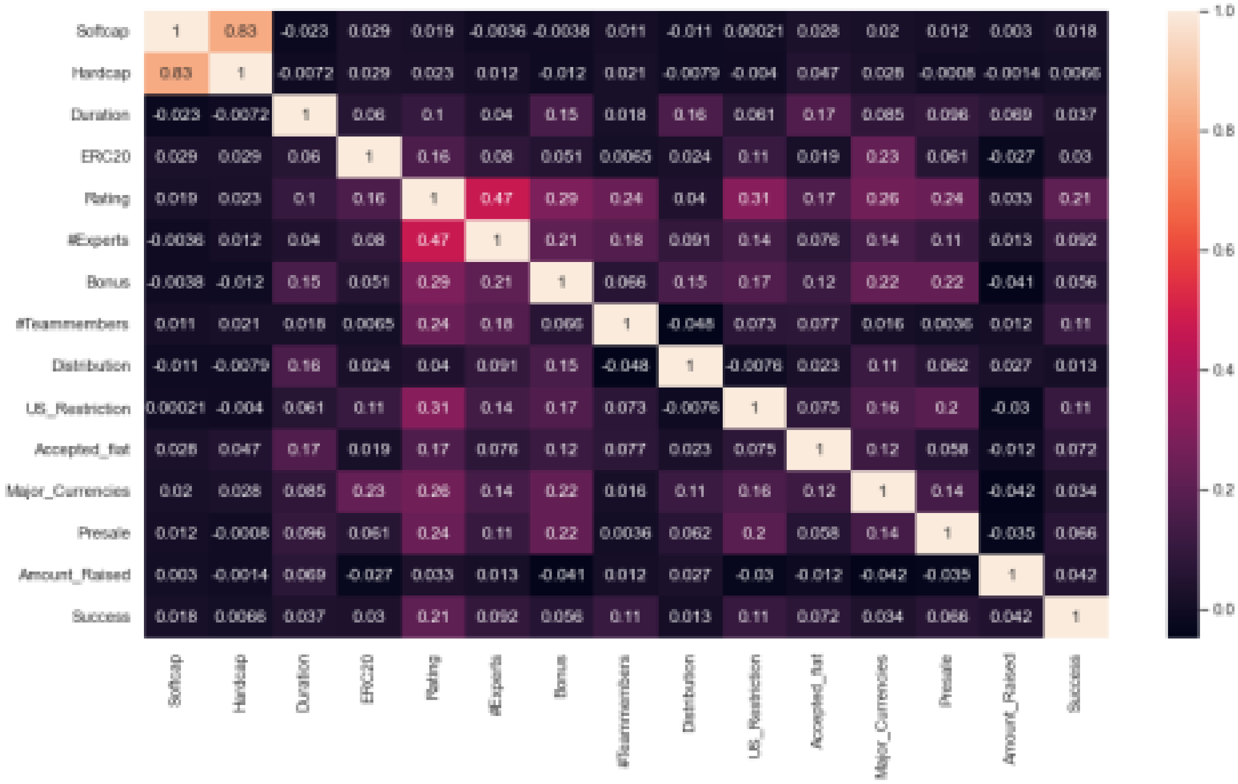






Needless to say, Amount Raised and Softcap are the most important ones because the Success criterion is derived from the two variables. Other variables which appear to increase the success of the ICO are Accepted Fiat, PreSales, Number of Experts, Number of Team Members, Ratings, Number of Bonuses, and the existence of US Restrictions.

A heatmap is generated to visualize the correlation among all the variables. Except for the correlation between SoftCap and HardCap, there is a negligible correlation between any pair of variables.



Using the significant variables of **Softcap, Rating, #Experts, Bonus, #Teammembers, US_Restriction, Accepted_fiat, Presale, and Amount_Raised**, K-Means clustering is done to determine the success of the ICOs. Using the Elbow Method, the optimal number of clusters is shown to be 2. Indeed, as seen in the table below, the mean values vary the most significantly when it comes to the aforementioned variables.

Mean Values of Variables for the 2 Clusters:

	Cluster 0 Cluster 1
SoftCap	19695759.93 15571701.31
HardCap	92190969.77 93671788.03
Duration	62.28 44.86
ERC20	0.81 0.71
Rating	3.68 2.73
Experts	8.84 1.51
Bonus	0.74 0.21

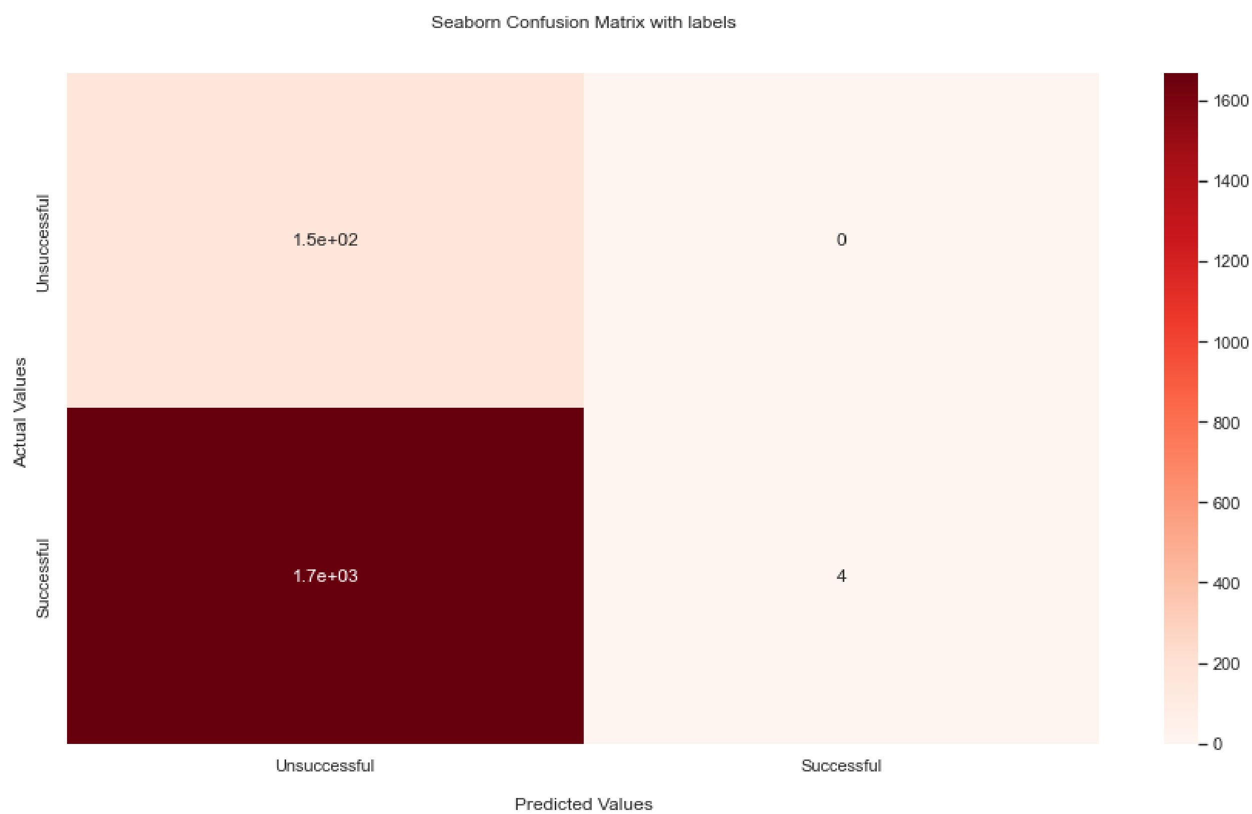
#Team Members 12.05 8.76

Distribution 0.54 0.50

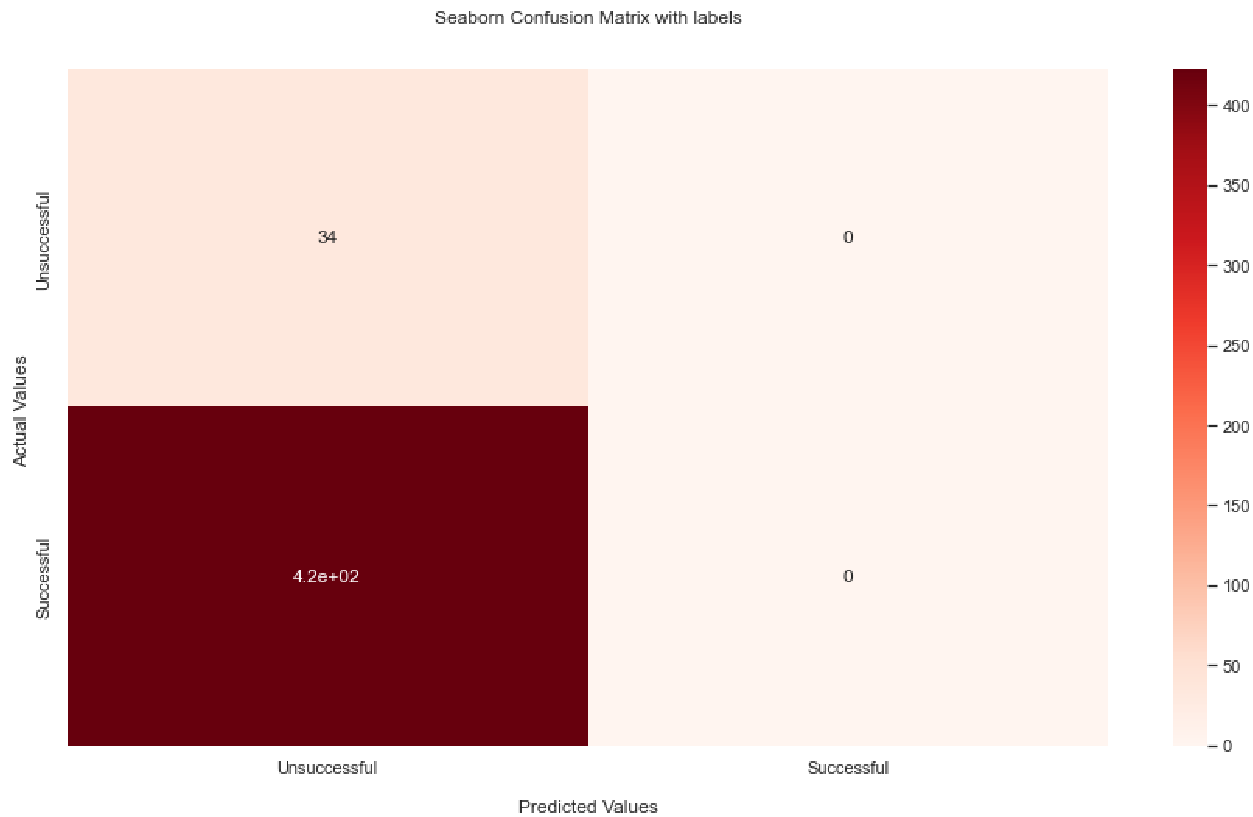
US_Restriction	0.56 0.13
Accepted Fiat	0.24 0.05
Major Currencies	0.97 0.79
PreSale	0.48 0.08
Amount Raised	9528035.70 17822122.15
Success	0.96 0.88

Confusion Matrices

The confusion matrices are created to determine the TP, TN, FP, and FN for train and test datasets both. The false-negative rate is very high but there are no false positives for the train data.



For the test data confusion matrix below, the TP rate is zero, while TN and FN are very high. The classification accuracy is close to 7%, which is very low for this model. Thus we can conclude the model performance indicates that the model did not accurately classify the ICOs in the two clusters.



The results of the confusion matrices essentially show that it is very difficult to correctly predict that an ICO will be successful based on the high number of False Negative values. However, it is commendable that no unsuccessful ICOs were predicted otherwise.

Part III

In this part, the two datasets of the data concerning the ICOs and the 1-week, 1-month, and 6-month Average Returns are reconciled so that all the ICOs for which the Average Returns are available, all of the data can be matched to run a regression and estimate the Average Returns of the ICOs, using the linear regression model, for the complete data set. A new excel file, ICO_Combined, was created and regression was run only on the significant variables from Part II. The missing values were also treated the same way as in the previous parts.

Using the independent variables (Softcap, Hardcap, Duration, ERC20, Rating, Experts, Bonus, Team members, Distribution, US_Restriction, Accepted_fiat, Major_Currencies, Presale, Amount_Raised), three linear regressions are conducted on each of the AR7 (Avg Returns in 1 week), AR30 (Avg Returns in 1 month), and AR180 (Avg Returns in 6 months). Initially, the linear regression was run simply on the Average Returns (AR) as the dependent variable but it showed that none of the variables was significant. So, a new binary variable was generated which gave a value of 1 for a positive AR and a value of 0 for a negative AR. Then the regressions were run for each AR7, AR30, and AR180. The results obtained were as such:

Significant Variables that explain the trend in

- AR7: None
- AR30: Distribution
- AR180: Duration, Rating, #Experts, and #Teammembers

The equations determined for the average returns will be as such:

- AR7 = 0.3638
- AR30 = 0.1576 + 0.3352* Duration
- AR180 = 0.1434 + 0.0013* Duration +0.0739* Rating - 0.0077* #Experts -0.0088* #Teammembers

Then a new data frame was created which, using the aforementioned equations, predicted the estimated 1-week, 1-month, and 6-month average returns for all of the ICOs in the original ICO_Data dataset.

Part IV

In this part, some new data is collected to understand which other variables can determine the success of an ICO. The data has been taken from Github: A user from Github named: PhilippeFerreiraDeSousa The user completed some whitepaper analysis on ICO success.

URL:
https://github.com/PhilippeFerreiraDeSousa/ICOSuccessPrediction/blob/master/src/python/coincheckup/dataset.csv data set name = dataset.csv. These additional variables are as follows:

- Returns on the ICOs in the first 45, 90, and 200 days
- Market Capitalization
- 24-hour Volume
- Circulation of Supply
- Total Supply
- Brand Buzz
- Social Media presence
- Average Volume
- Age of the ICO (in months)
- Winning Months (the months in which the ICO performed the best)

The missing values are also eliminated successfully or replaced with average values. The metric of ICO Success is chosen as such: the Market Cap should be higher than the Average Market Cap. Again, K-Means clustering is used and the Elbow Method is employed to determine the optimal number of clusters, i.e. 2.

The results were as such:

Mean Values of the Variables for the Two Clusters

	Cluster 0 Cluster 1
Price	6.71 98.72
45d	30.46 288.49
90d	-29.57 -0.47
200d	-51.04 -49.52
MktCap	122.81 148.45
24hVol	129.38 109.89
Circ_Supply	147.80 93.19
Total_Supply **Brand_Buzz**	7.515066e+10 1.077545e+11 33.59 3.98
Social	56.41 39.99
Avg_mo	1583922.73 14834955.42
Winning_months	11.49 48.35
Success	0.25 0.35

The Success mean is significantly different for the two clusters. The most visible difference is in winning months. Where the mean winning months are higher, the Success rate is clearly higher. As was predicted, a higher MktCap also indicates higher Success. Similarly, successful ICOs also have a much higher total supply. When the average trading volume is higher, again the ICOs are more probably successful. Successful ICOs show higher returns over the 45, 90, and 200-day periods. It is also curious that the higher-priced ICOs are also, on average, more successful. This may be explained by a lower circulation of expensive coins, so the standard deviation of Circulation of Supply is 2.25 for Cluster 0 and 1.74 for Cluster 1, thereby showing the lower volatility of circulation for high-priced ICOs. The price pattern of ICOs is similarly more volatile, as indicated by the standard deviation, for Cluster 0, compared to Cluster 1.