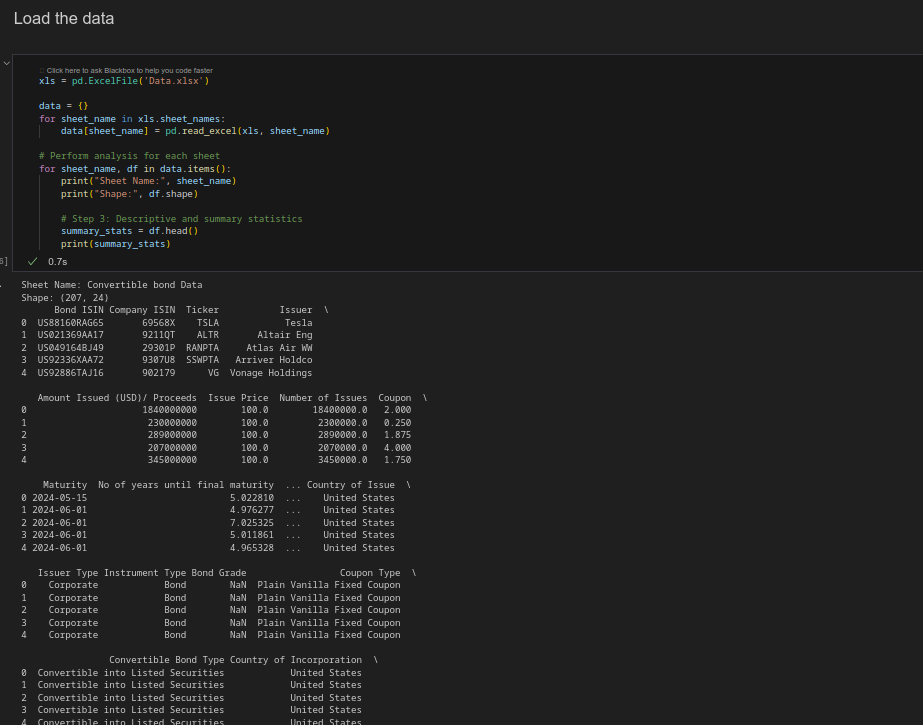
**Data Overview**

The data source is a excel file that basically consisted of various sheets that are used to perform the describe and univariate analyses that are required. In this prospects, the data used are from the features that are required to perform the analyses. This comes in handy with the actual data transformations that are required to give prospects and prediction on public offering placement over private placements. In this case the below analytic display the functionalities of the code as it structures the major functionalities that are to determine the essential company choosing

**Data Summary**

The features on the data have numerical data formats so basically I instantiated the actual mean and standard deviation of the features getting the quartile values and minimum.



**Data Cleaning**

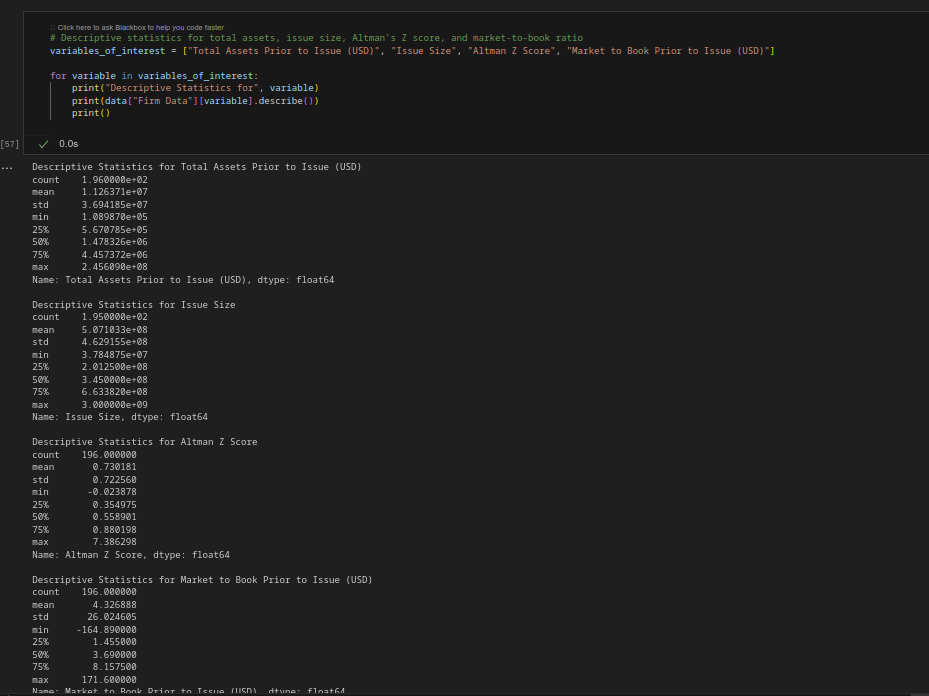
The data had missing values and as so, we evaluated the -inf to get the mean of the data and fill in the values that were missing. The technique basically evaluated the important features that were needed to perform the modelling of the data using unique features for the logistic regression. The filling of the missing values with mean is appropriate because it categories a better functionality that removing the rows with missing values.

**Data standardization**

Numerical features are standardized to have a 0 mean value and a standard deviation of 1 which basically structures the essential preprocessing for the optimal running of the learning algorithms.

**Descriptive analysis**

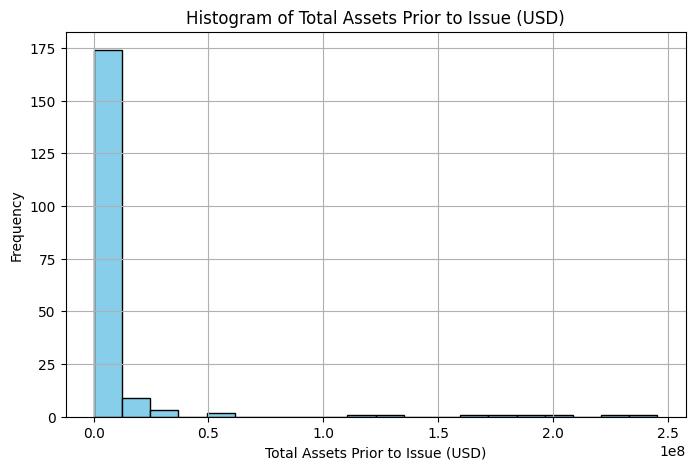
The analysis gave the features and type of data for the Altman Z-score features that were necessary for the modelling of the data.



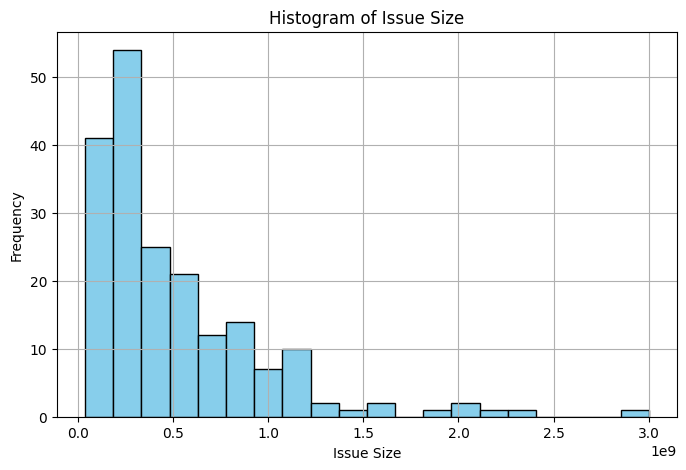
The essential features dtype data of the columns is essential for the realization of the actual model standardization for prediction and modelling.

**Visualization.**

The histogram below visualizes the firm data based on the various frequencies that are in the continuous variables



As Altman z-score features are described I also visualzied the issue size below



**Modelling**

Using statsmodel.api I was able to optimize the logistic regression results which basically comprised of the content below as a result. Their performance of the data was measured by the various total assets, issue size and market to book prior issue using an Altman z score

Optimization terminated successfully.

Current function value: 0.500178

Iterations 6

Logit Regression Results

==============================================================================

Dep. Variable: Placement Type No. Observations: 195

Model: Logit Df Residuals: 191

Method: MLE Df Model: 3

Date: Sun, 18 Feb 2024 Pseudo R-squ.: -0.01413

Time: 11:28:05 Log-Likelihood: -97.535

converged: True LL-Null: -96.176

Covariance Type: nonrobust LLR p-value: 1.000

===============================================================================

coef std err z P>|z| [0.025 0.975]

-------------------------------------------------------------------------------------------------------

Total Assets Prior to Issue (USD) 3.126e-10 4.88e-09 0.064 0.949 -9.26e-09 9.89e-09

Issue Size 1.016e-09 4.38e-10 2.320 0.020 1.58e-10 1.87e-09

Altman Z Score 1.2148 0.345 3.526 0.000 0.540 1.890

Market to Book Prior to Issue (USD) 0.0034 0.008 0.452 0.651 -0.011 0.018

===============================================================================

**Model Assessment**

By defining the independennt variables and dependent variables I was able to fit the model regression and got the summary based on the current function value and analysed the other features.

Optimization terminated successfully.

Current function value: 0.486423

Iterations 6

Logit Regression Results

==============================================================================

Dep. Variable: Placement Type No. Observations: 195

Model: Logit Df Residuals: 190

Method: MLE Df Model: 4

Date: Sun, 18 Feb 2024 Pseudo R-squ.: 0.01376

Time: 11:28:05 Log-Likelihood: -94.852

converged: True LL-Null: -96.176

Covariance Type: nonrobust LLR p-value: 0.6185

===============================================================================

coef std err z P>|z| [0.025 0.975]

-------------------------------------------------------------------------------------------------------

const 0.9213 0.387 2.379 0.017 0.162 1.680

Total Assets Prior to Issue (USD) -7.314e-10 4.74e-09 -0.154 0.877 -1e-08 8.55e-09

Issue Size 4.266e-10 4.53e-10 0.942 0.346 -4.61e-10 1.31e-09

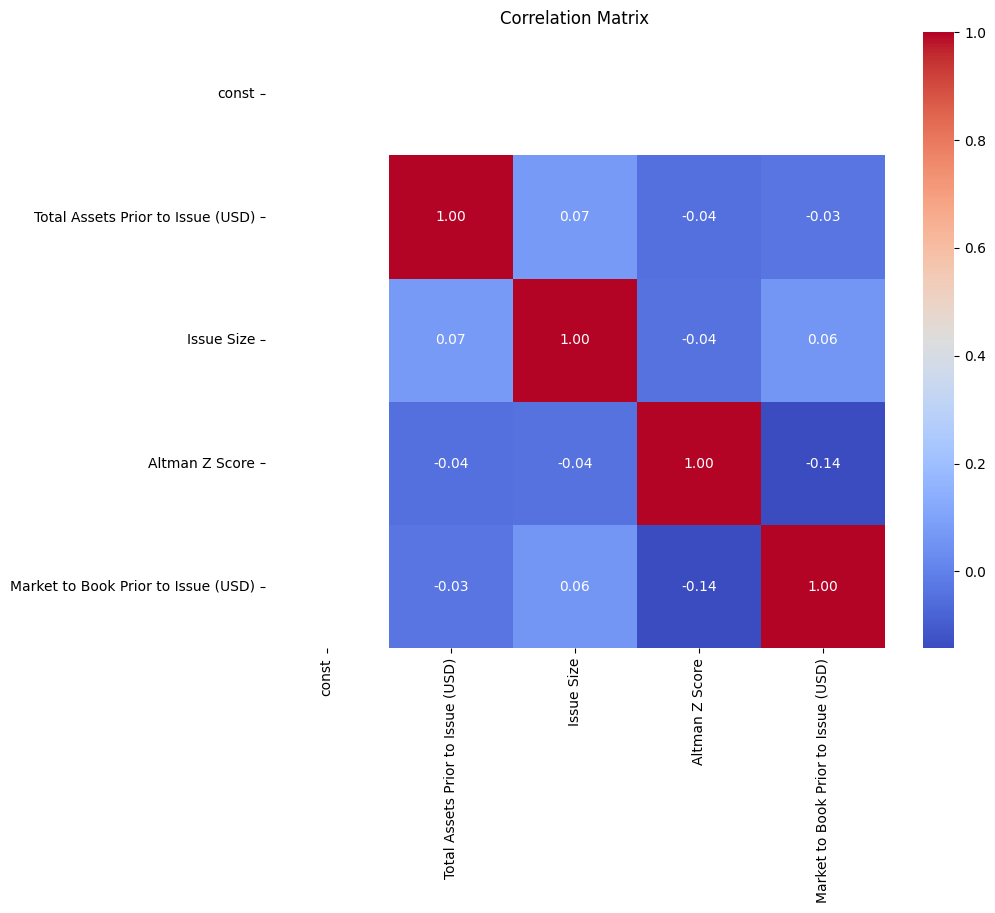
Altman Z Score 0.4479 0.407 1.101 0.271 -0.349 1.245

Market to Book Prior to Issue (USD) -0.0002 0.008 -0.029 0.977 -0.015 0.015

===============================================================================

**Correlation matrix.**

By analysis the mode for both the Logistic Regression and the SGD Classifier we were able to get the accuracy of the two models and generated a correlation matrix



The actual accuracy of the two models are:

Logistic Regression Accuracy: 0.8205128205128205

SGDClassifier Accuracy: 0.6666666666666666

**Feature Distribution**

The feature distribution of the data basically was determined by creating a function that will help in realizing the various market to book prior issue, the Altman Z-Score, the total assets prior issue, and the distribution constants

