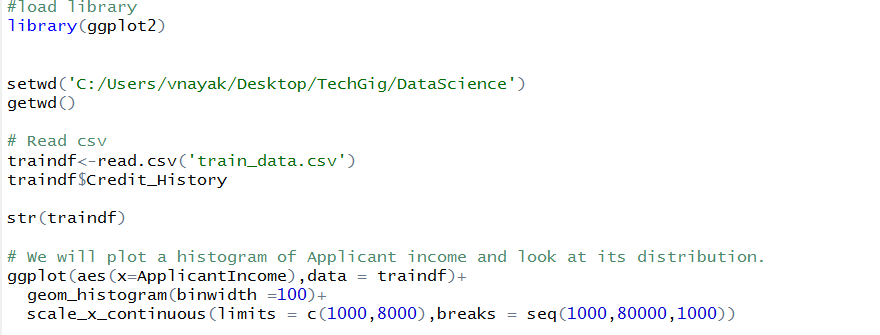
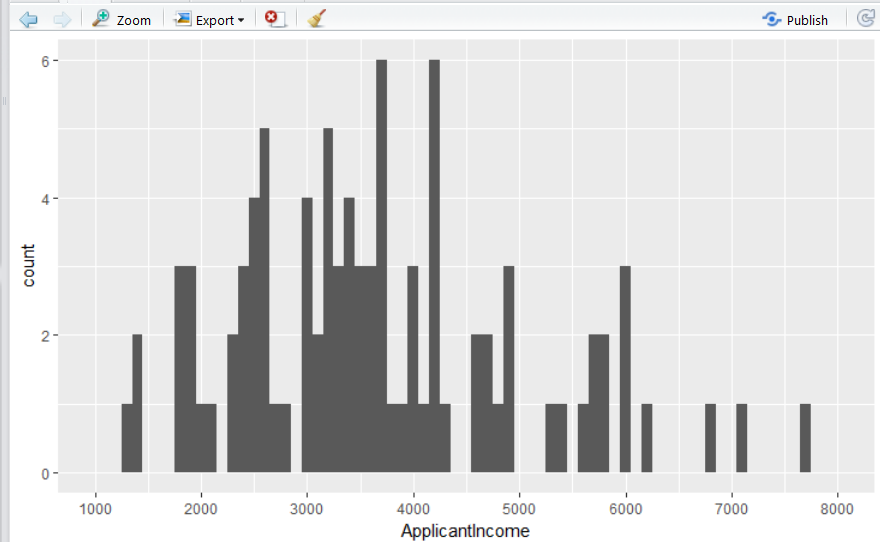
Data Analysis.

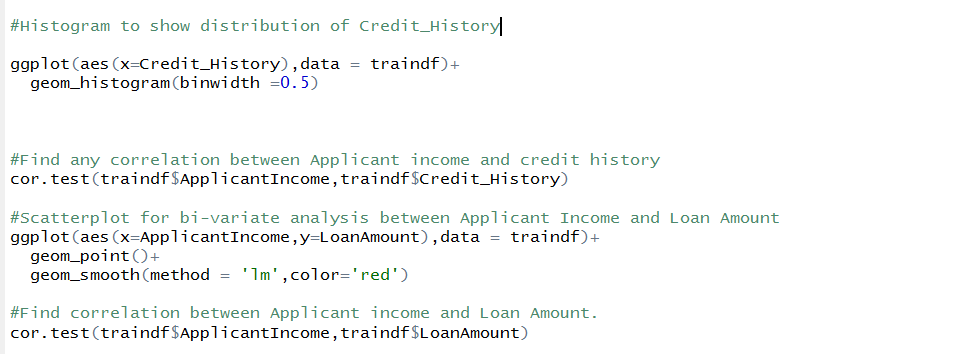
The first step towards creating a predictive model is to explore the data and understand the pattern of data if it is present. So I will start with exploratory data analysis. I will create few graphs using library ggplot2 and will try to find correlation between variables.

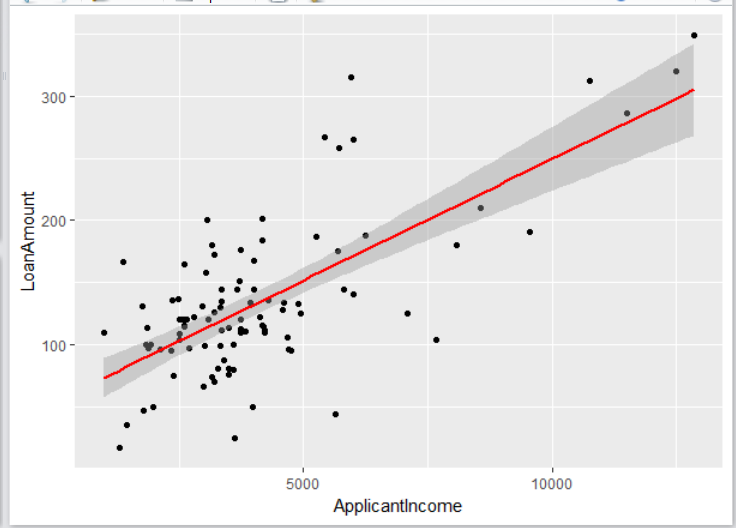
In the code below I performed some data exploration by creating histograms.





Then I explore more data by creating Scatter plot and finding correlation between variables.





After completing the data analysis process, I will start with creating a predictive model.

So I will take necessary steps to perform data cleaning.

* Find missing values
* Impute missing values using Knn
* Predict the missing values
* Convert variables to numerical features.

Run the basic model and find the important features. Include the important features only in the train and test data set.

Perform cross validation to train the data. Observe train and test error and tune it accordingly. Create the XGBoost model and run it.

Predict on the test data set.

Note: There is scope of increasing the accuracy of the model. We need to perform more data exploration and feature engineering.