Credit Card Fraud Detection

SUBMITTED BY:
SHASHWAT BHANSALI

AGENDA

- Objective
- Background
- Key Insights
- Cost Benefit Analysis
- Appendix:

  - Data Methodology

OBJECTIVE

- To stop the huge losses that are being incurred due to frauds and a manual credit card fraud detection system.
- To decrease losses due to credit card payment fraud for both merchants and issuing banks.
- To increase revenue opportunities for merchants.

BACKGROUND

- Multiple visual to view the different trend in the data.
- To detect frauds early and stop losses Machine Learning model has been built.
- A cost benefit analysis is also been done.

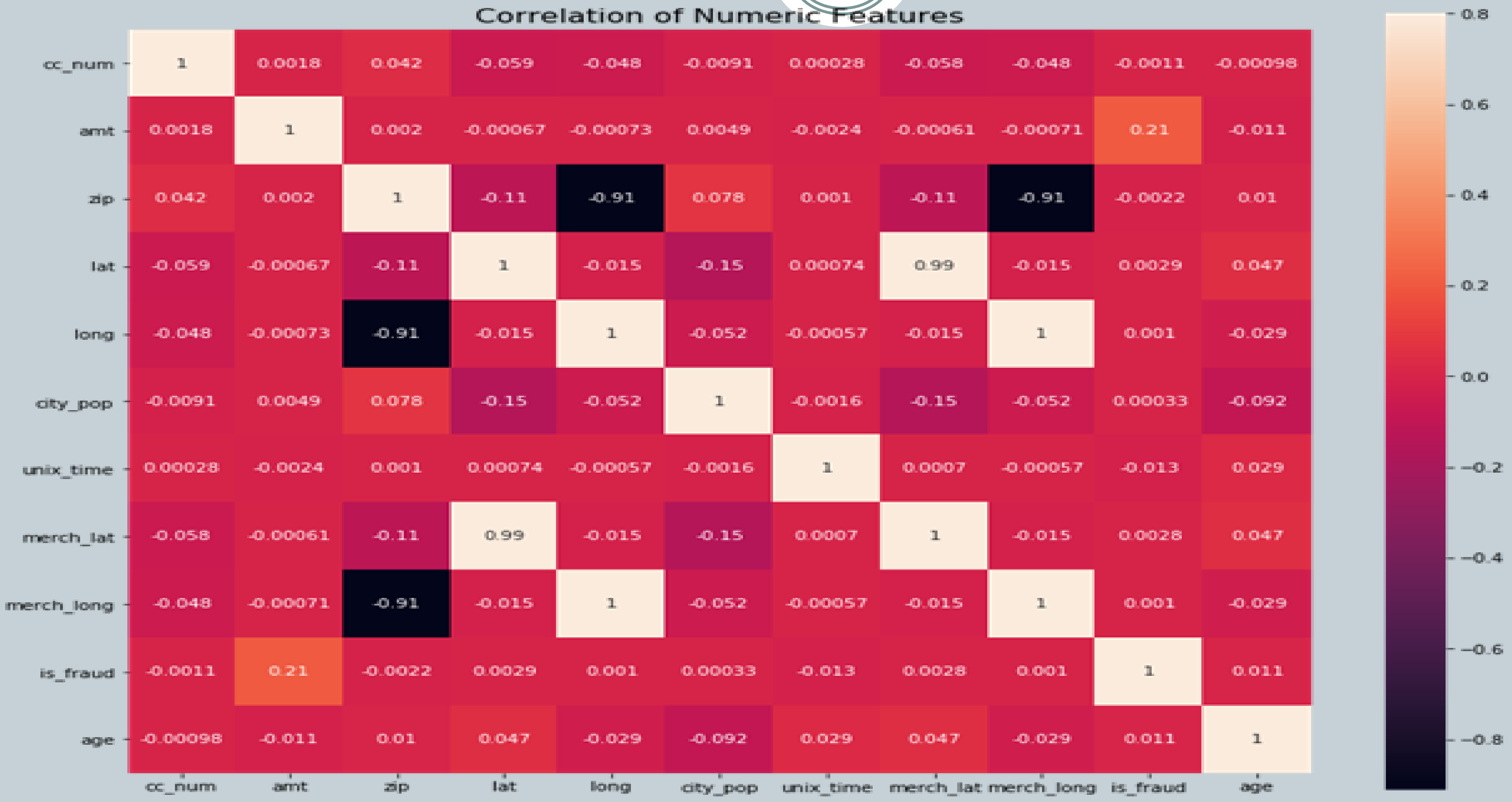
Data Understanding

- Data has 1852394 rows and on dropping one irrelevant column it make it 22 columns.
- It contains a total of 18,52,394 transactions, out of which 9,651 are fraudulent transactions.
- The given data has no null values
- The data set is highly imbalanced, with the positive class (frauds) accounting for 0.52% of the total transactions.

Exploratory Data Analysis

- Derived time, days and years from transaction time column.
- Converted data types into required format.
- Most females use more credit cards than males.
- According to the data Sunday and Monday are the days of week which have highest card transaction.
- 50% of our customers are from age group 35-55.
- Minimum age of customer is 14 and Maximum age of customer is 96.
- Highest number of transactions happened in the month of December where as lowest happened in February.

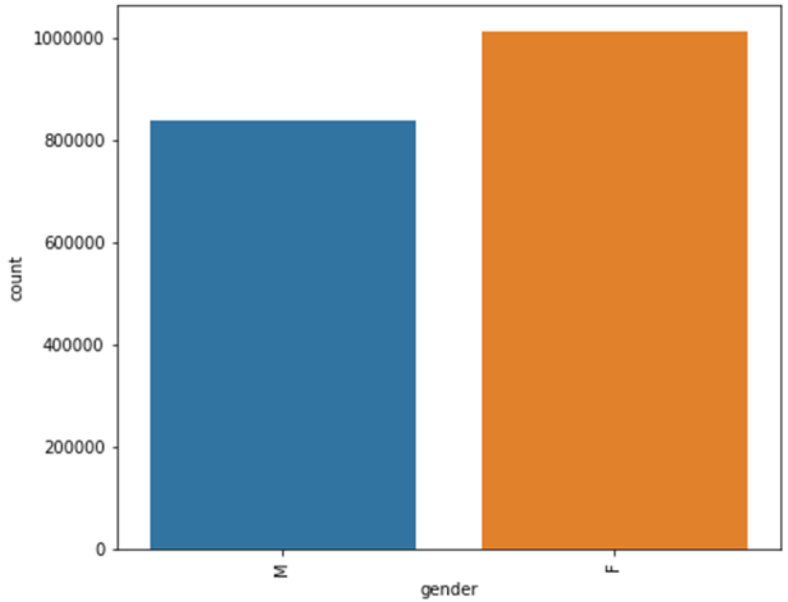
Correlation Heatmap



No highly correlated numeric data.

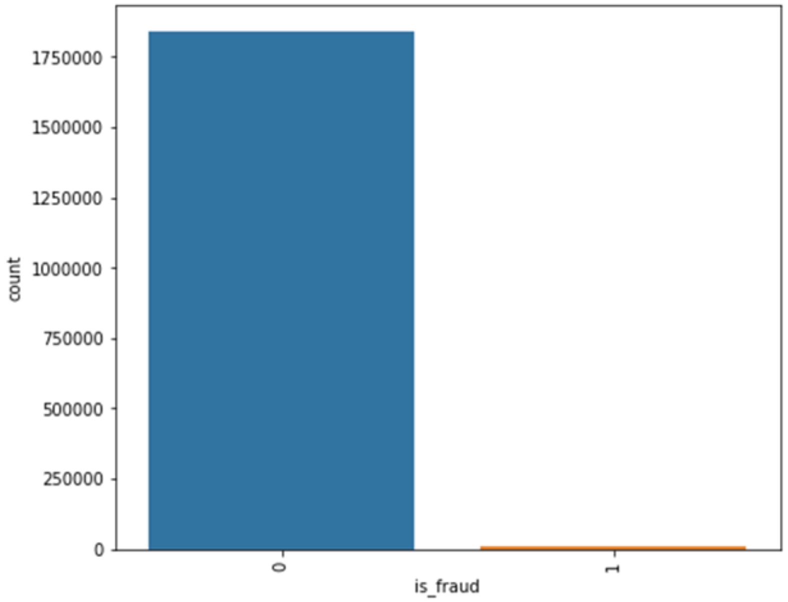
EDA- Univariate/Bivariate Analysis

Female vs Male users



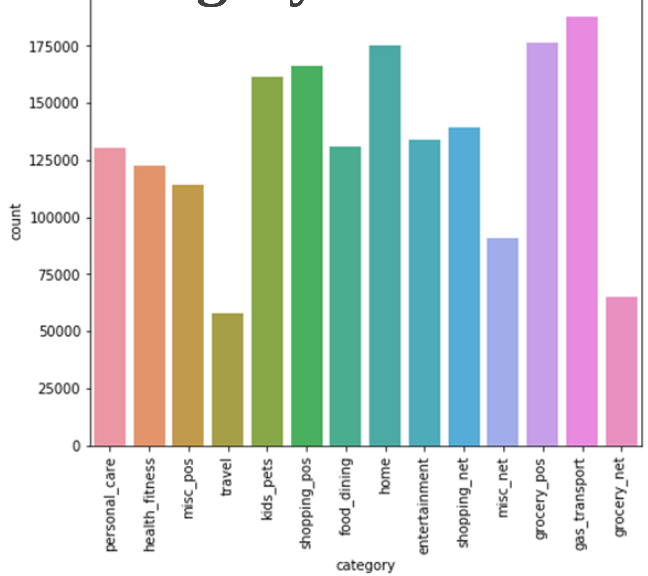
There are more female users than male users.

Fraud vs non-fraud



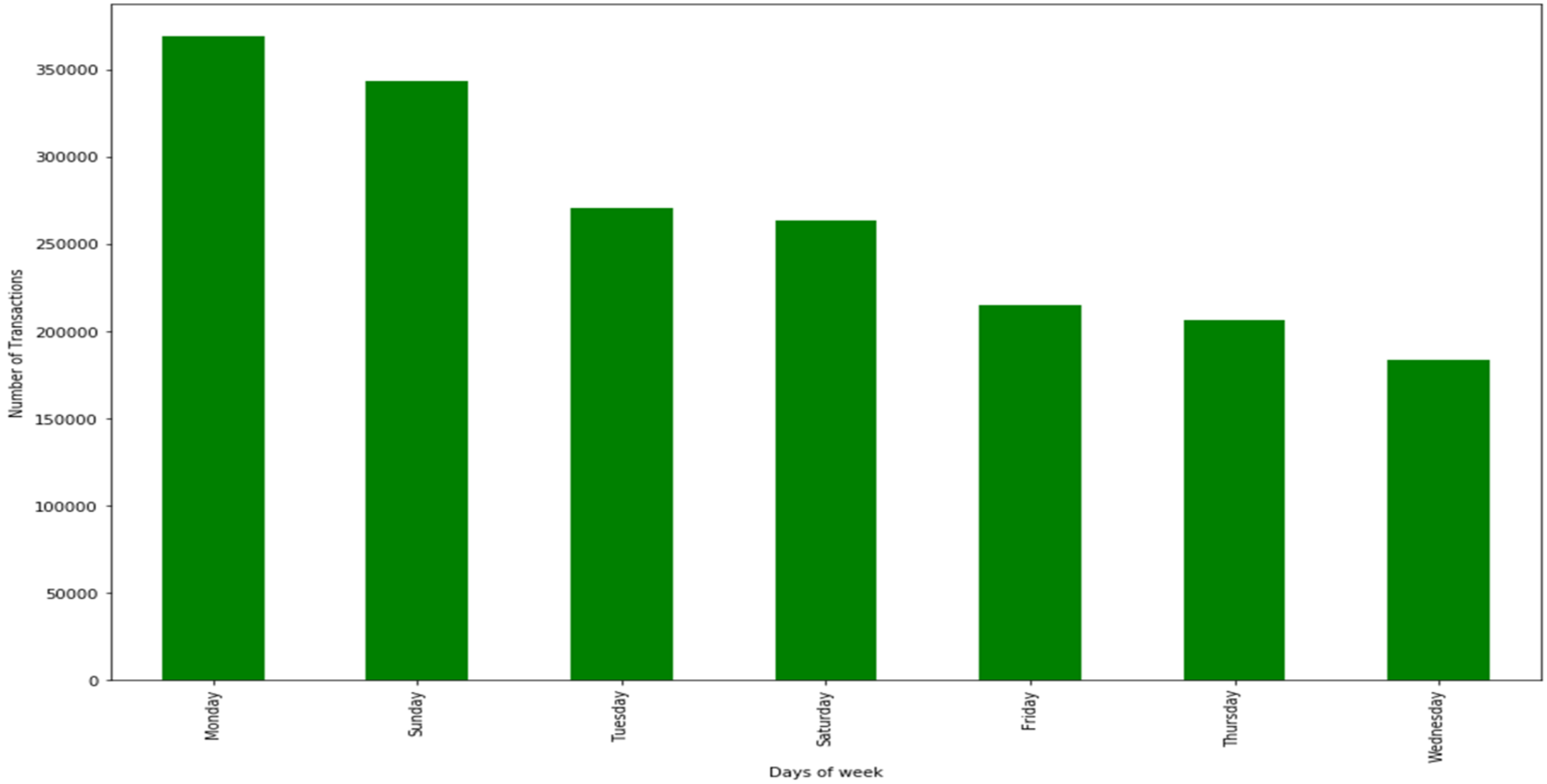
We can see the data is highly imbalanced.

Transaction Category



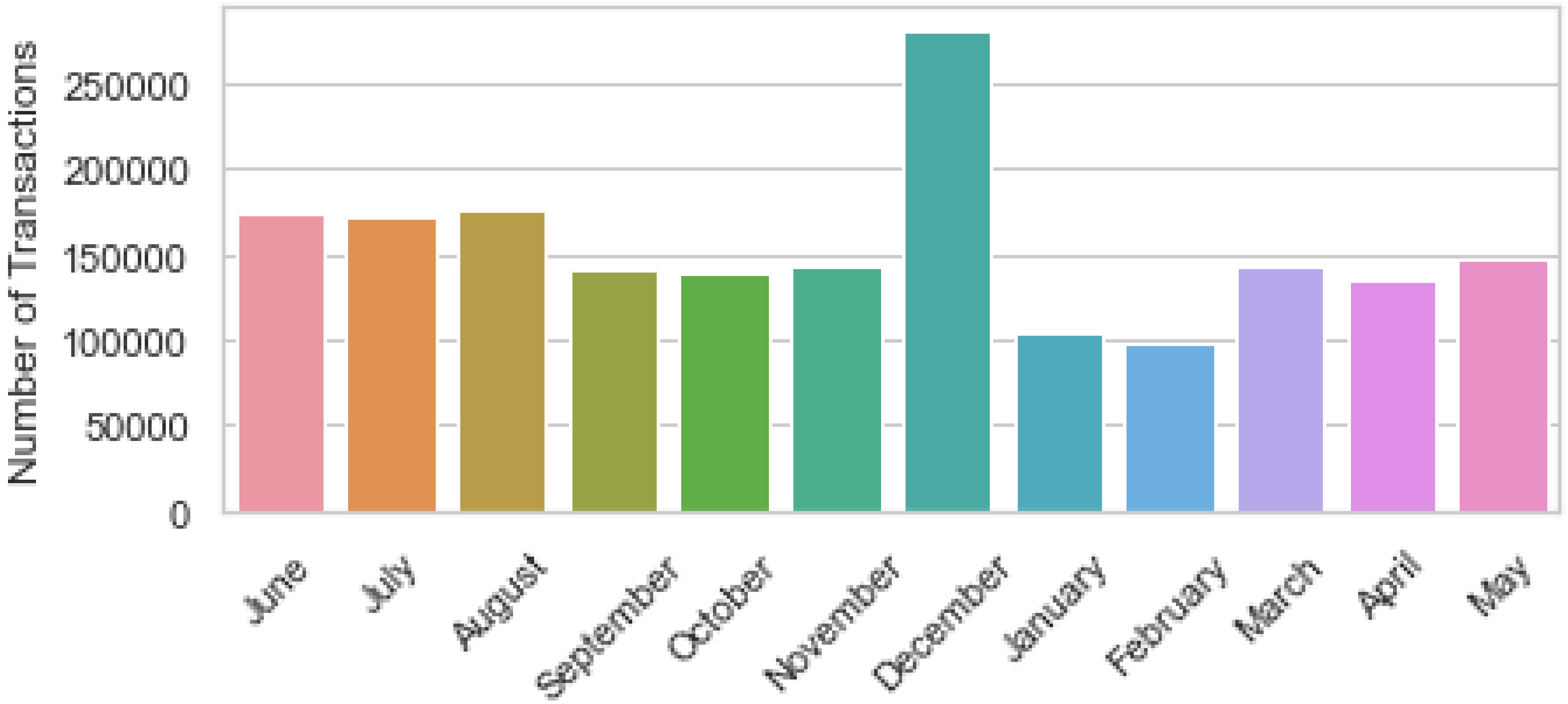
There are five categories whose transactions are higher than 150000

Distribution of transactions made on days of week



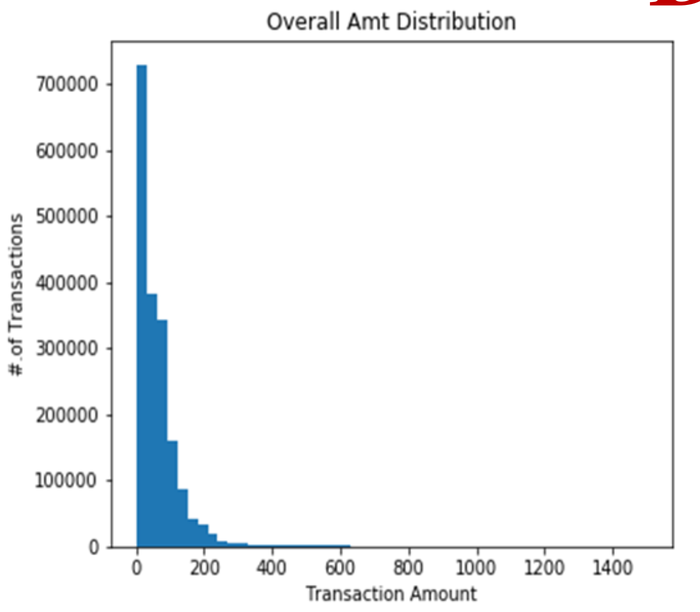
Sunday and Monday are two days of week when most of the transactions take place.

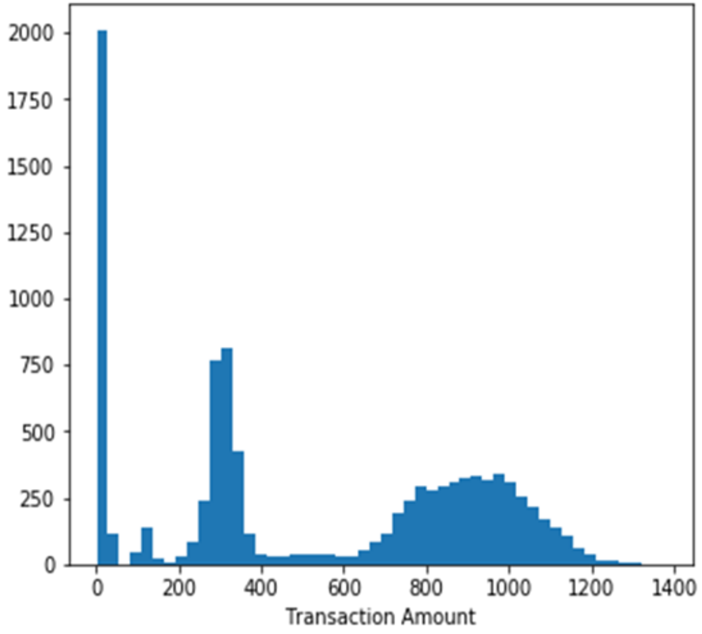
Distribution of transactions made in month of the year

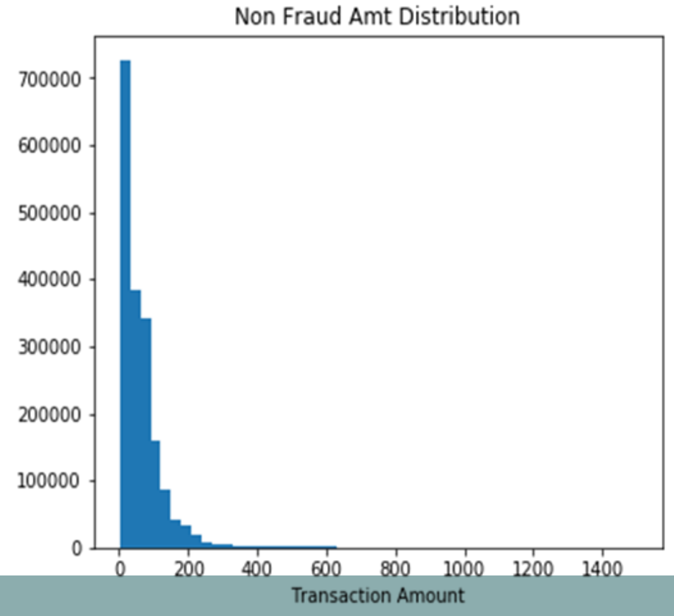


Maximum transaction has been done in the month of December.

Distribution of amount



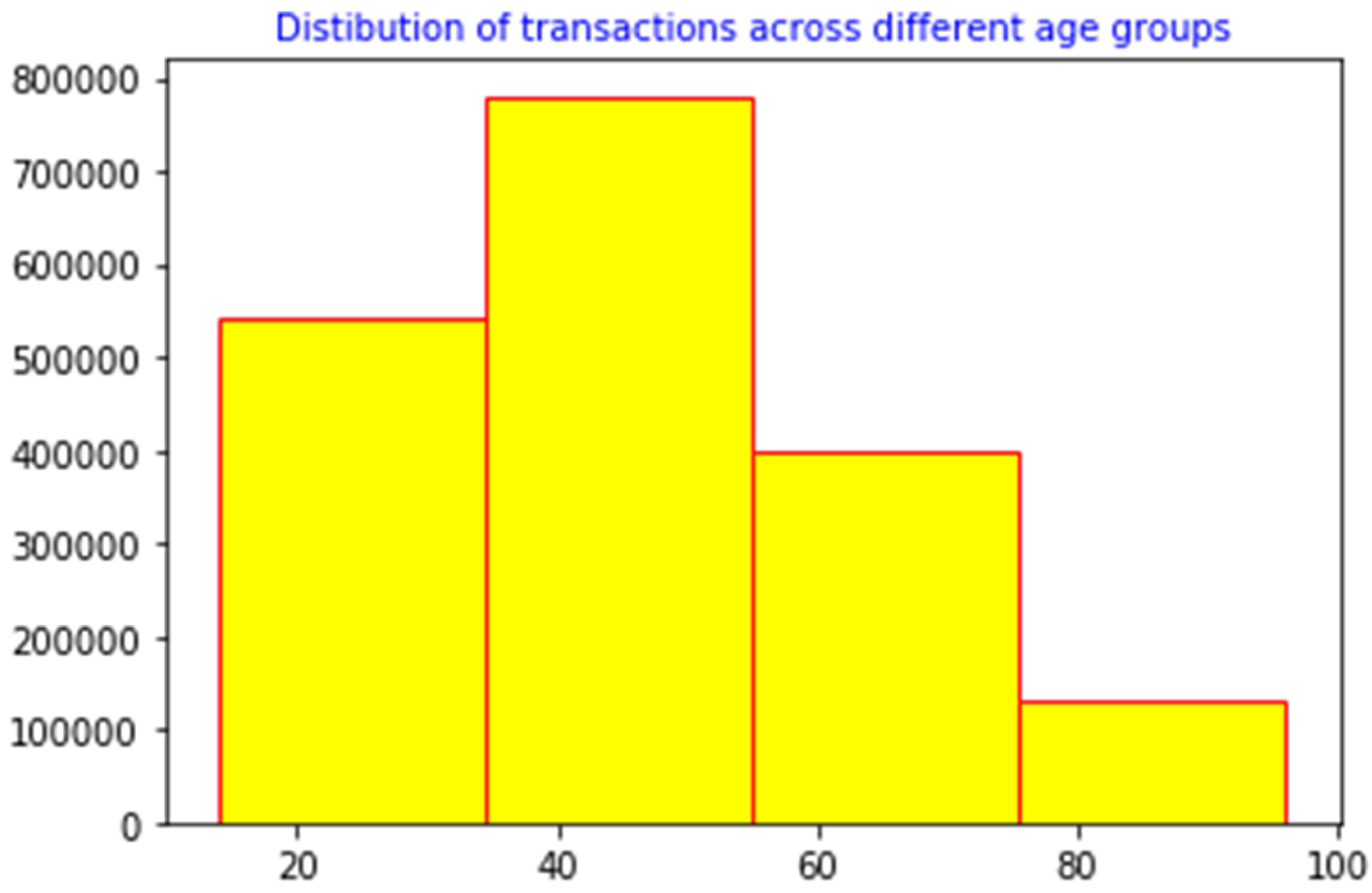




1. Fraud Transaction mean is way higher than non-fraud transaction.

2. Mean of Non Fraud Transactions: 67.6
3. Mean of Fraud Transactions: 530.6

Fraud among different age groups



Maximum of the transactions are made by the people in age group 35-55 approximately.

Key Insights

1) Transaction amount, category and gender are the most important variables.

2) Gas and transport, grocery and shopping are the top three categories.

Current Losses

- 77,183 credit card transactions per month.
- 402 fraudulent transactions per month.
- $ 530.66 amount per fraud transaction.
- Total costs incurred from fraud transactions is $ 213,392.22.

After Model Deployment

- 1720 fraudulent transactions detected by the model.
- $1.5 cost to provide customer support to these transactions that is $ 2,580.38 in total.
- 68 fraudulent transactions not detected by model which amounts to $ 35,908.09 loss.
- Total cost incurred after new model deployment is $ 38,488.46.
- Final savings after new model deployment is $174,903.76 that is reduction in losses by ~82%.

Appendix: Data Methodology

1. Different type of model where built on top of data dataset.
2. Class imbalance adjusted using Oversampling technique.
3. Due to extensive computational times when using Grid Search Cross Validation manual hyper parameter tuning was done.
4. Among the three used model building techniques decision trees and random forest techniques gives the best results.
5. Decision Tree model and Random forest model has the highest recall and accuracy.

THANK YOU