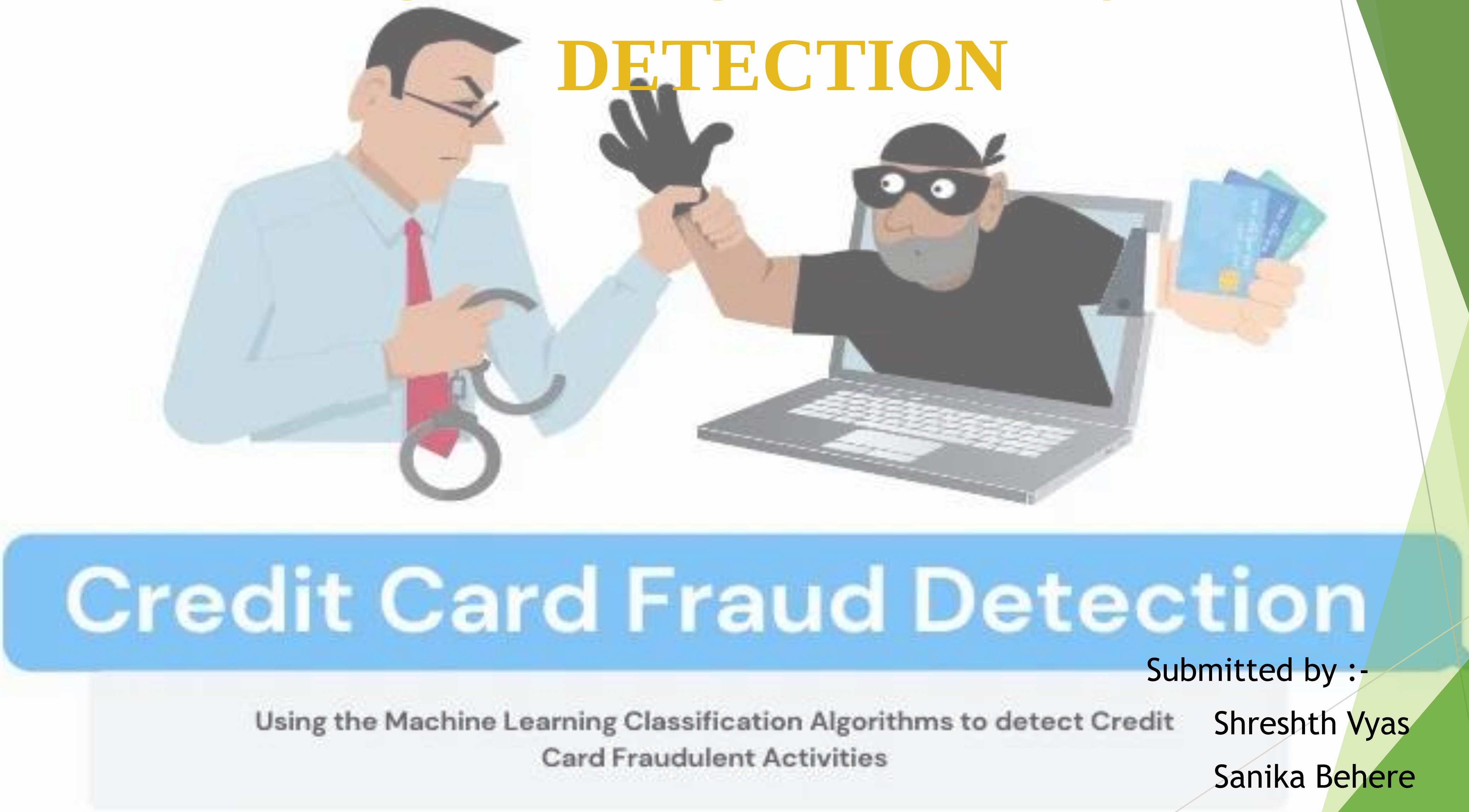
CREDIT CARD FRAUD DETECTION



Submitted by :-

Shreshth Vyas

Sanika Behere

Gaurav Singh Bisht

Agneda

- Objective/Problem Statement
- Background
- Problem Solving Approach
- Key Insights/ Visualization
- Cost Benefit Analysis
- Appendix:
    - Data Methodology
    - Attached Files

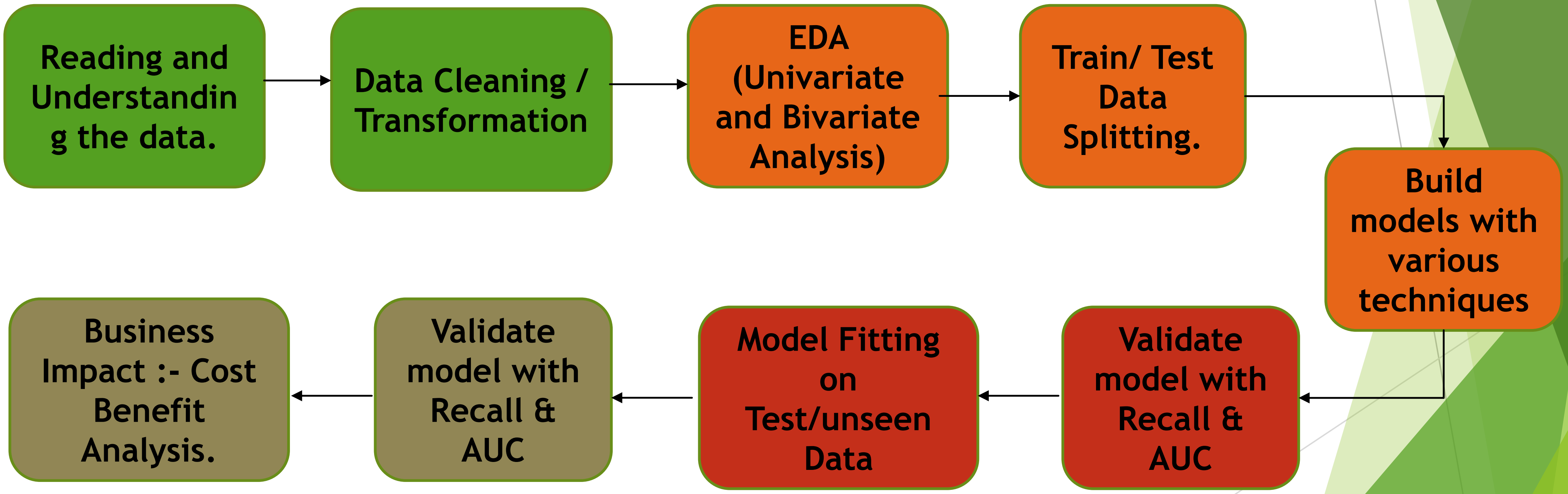
Objective/ Problem Statement

- As a part of the analytics team working on a fraud detection model and its cost-benefit analysis. We need to develop a machine learning model to detect fraudulent transactions based on the historical transactional data of customers with a pool of merchants.
- We have to analyse the business impact of these fraudulent transactions and recommend the optimal ways that the bank can adopt to mitigate the fraud risk.
- We need to put proactive monitoring and fraud prevention mechanisms in place
- Machine Learning help these institutions to reduce time consuming manual reviews, costly chargebacks and fees, and denial of legitimate transaction.

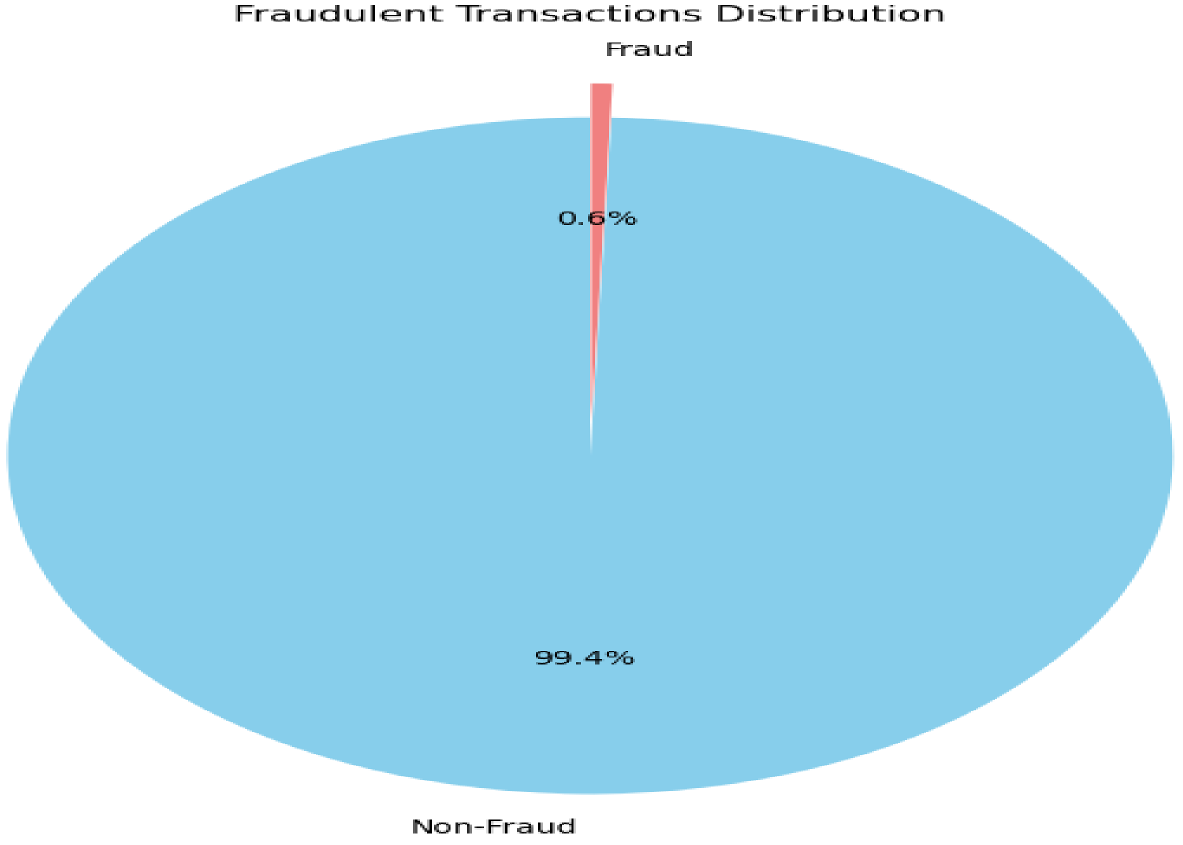
Background

- In recent times, the number of fraud transactions has increased drastically, due to which credit card companies are facing a lot of challenges. For many banks, retaining high profitable customers is the most important business goal.
- In terms of substantial financial loss, trust and credibility, banking fraud is a concerning issue for both banks and customers alike.
- With the rise in digital payment channels, the number of fraudulent transactions is also increasing as fraudsters are finding new and different ways to commit such crimes.
- We have performed the root causes analysis for the increasing number of frauds and high revenue loss, and you realized that building a fraud detection system using different machine learning techniques is quite important to identify such fraudulent activities at the right time and prevent them from happening.

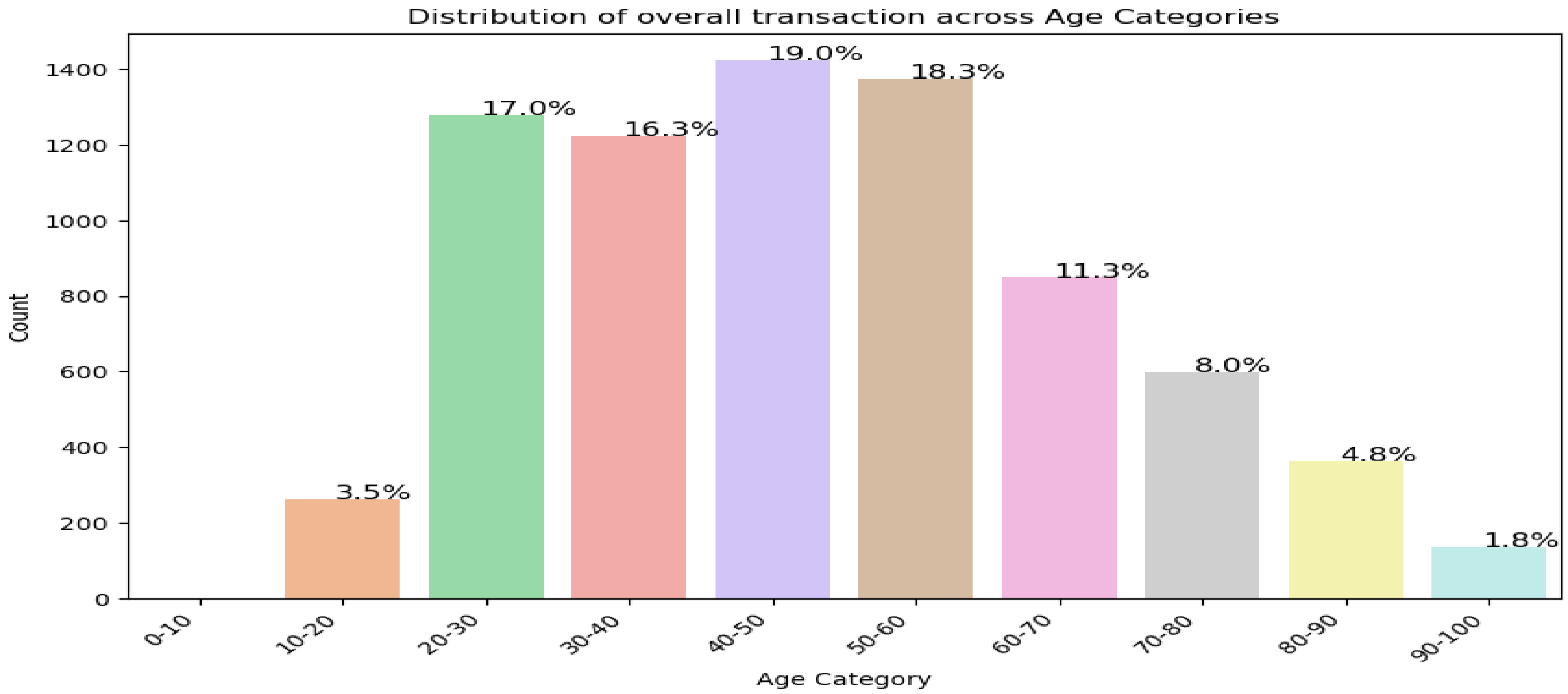
Problem- Solving Approach



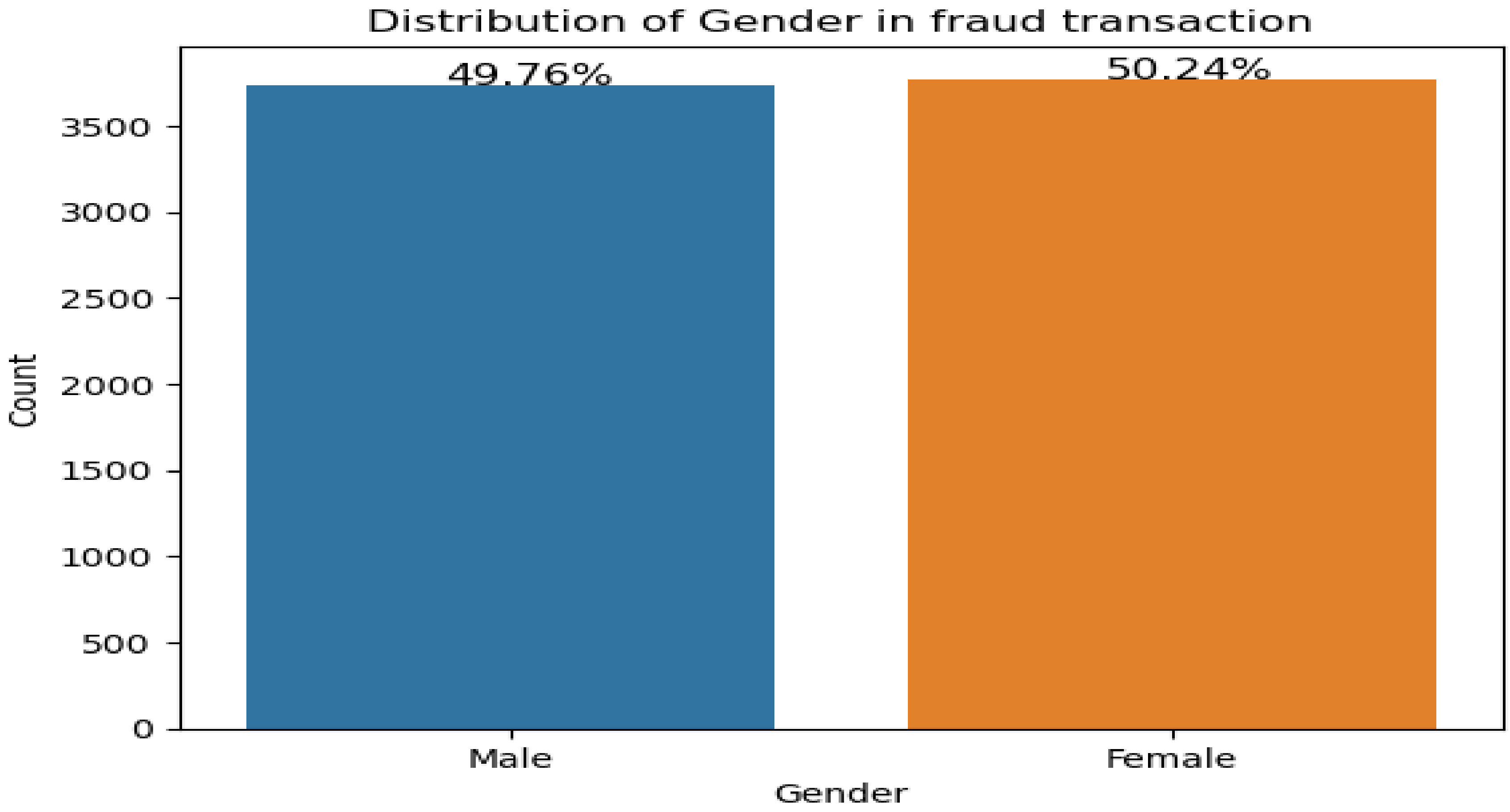
Key Insights/Visualization



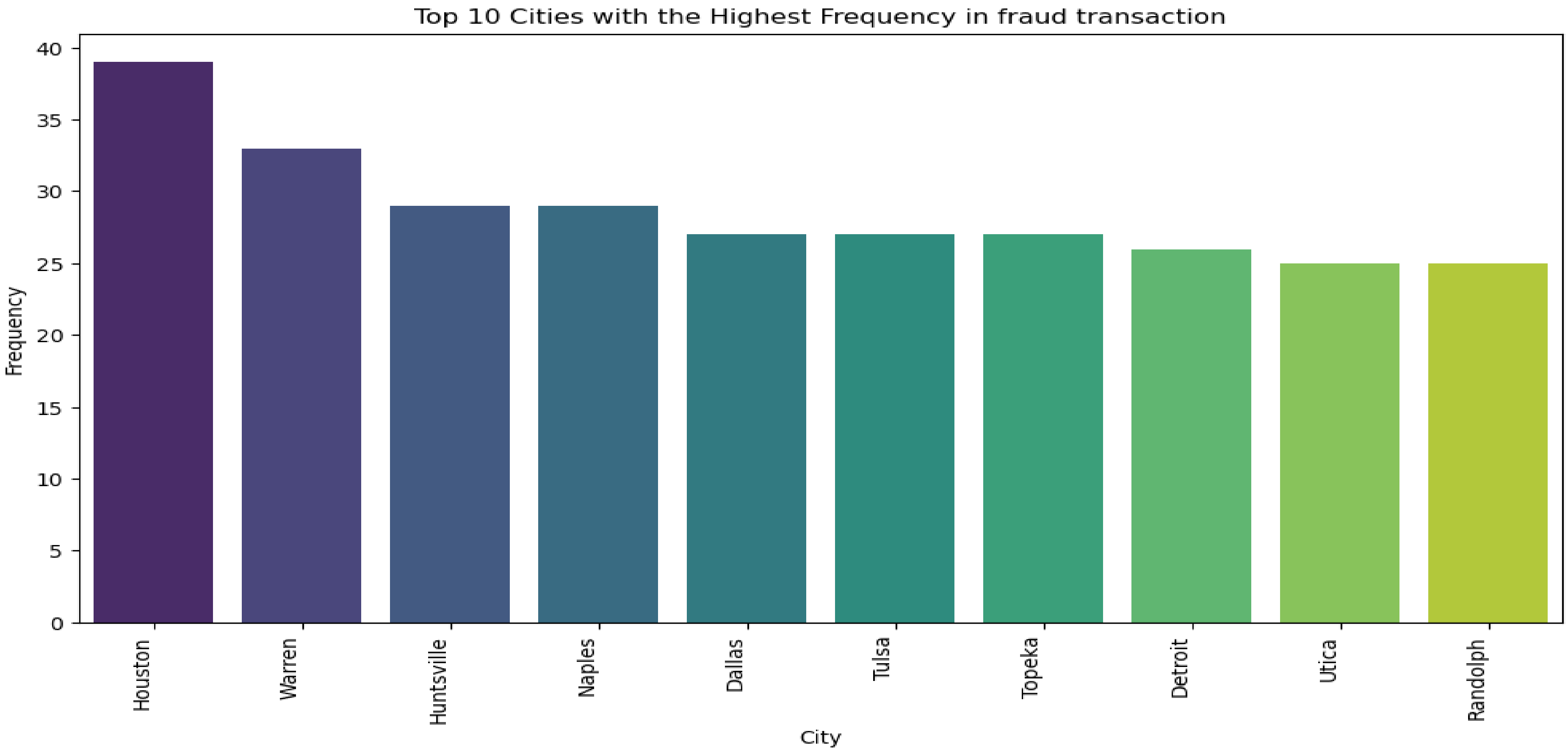
- Data set is highly imbalanced. Out of a total transaction 9651 are fraudulent i.e. 0.6% of the total transactions.



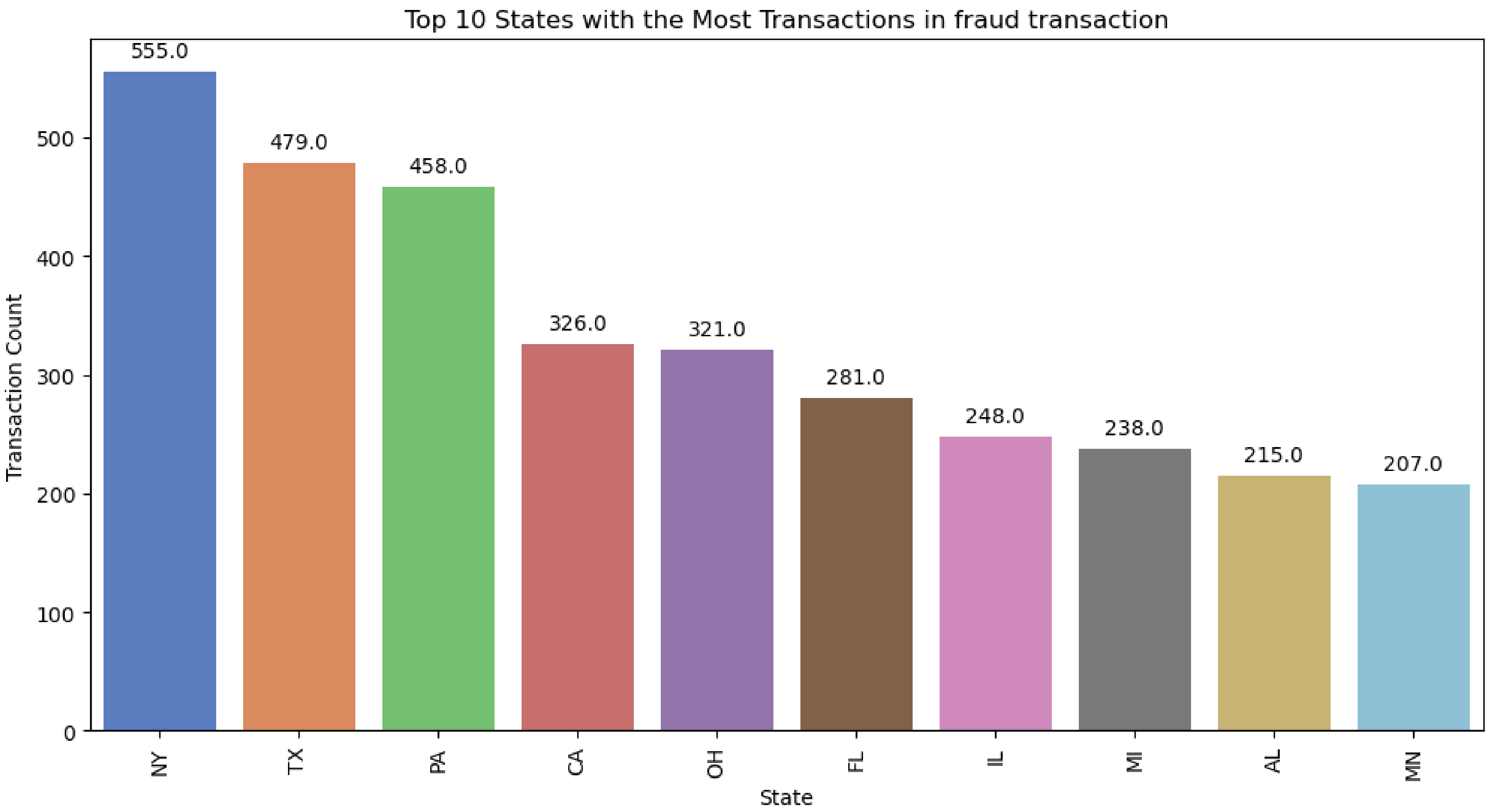
- Mid Age People between 30 – 60 have the most number of fraudulent activities.



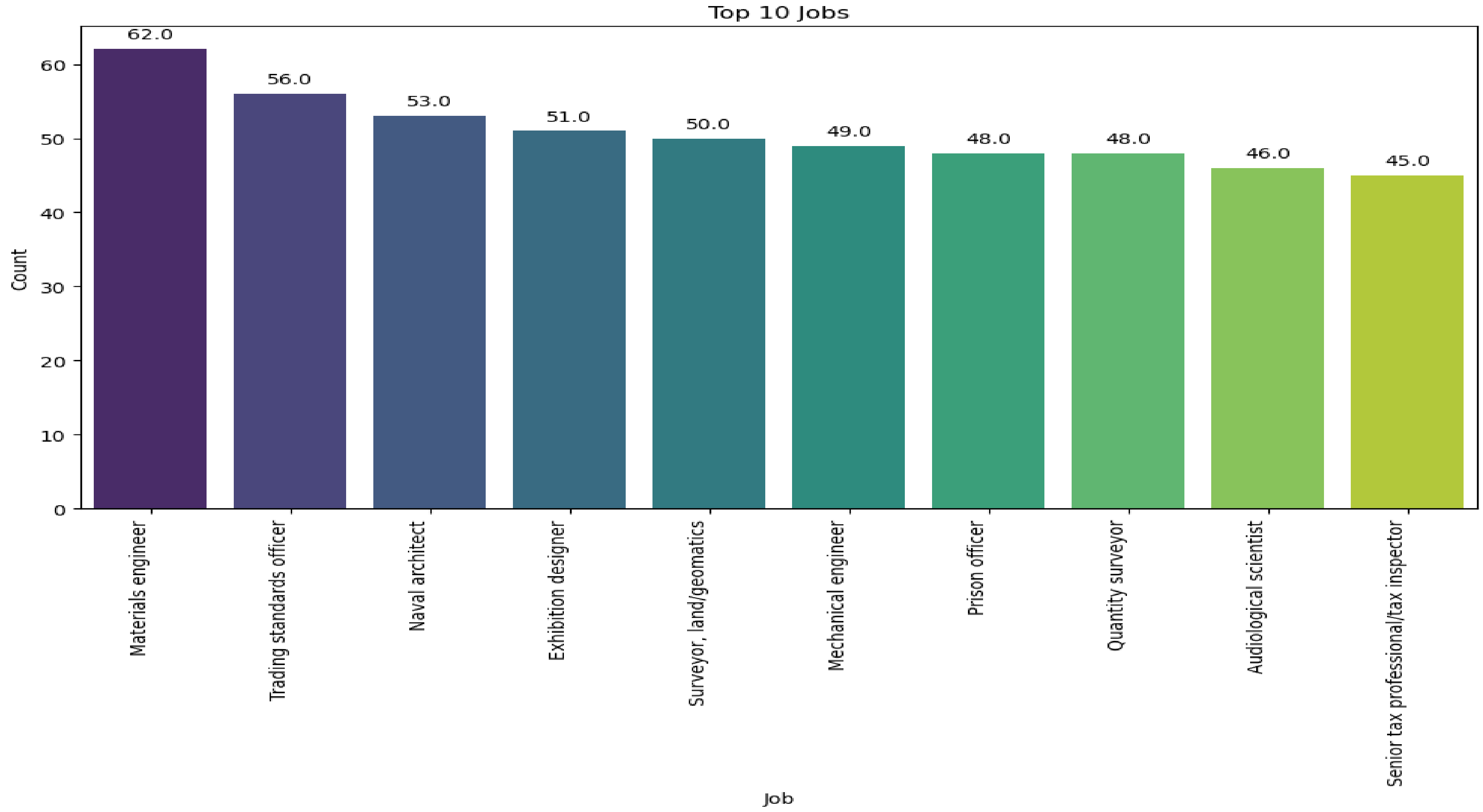
- Female having more Fraudulent transaction count over Man although the difference is not huge.



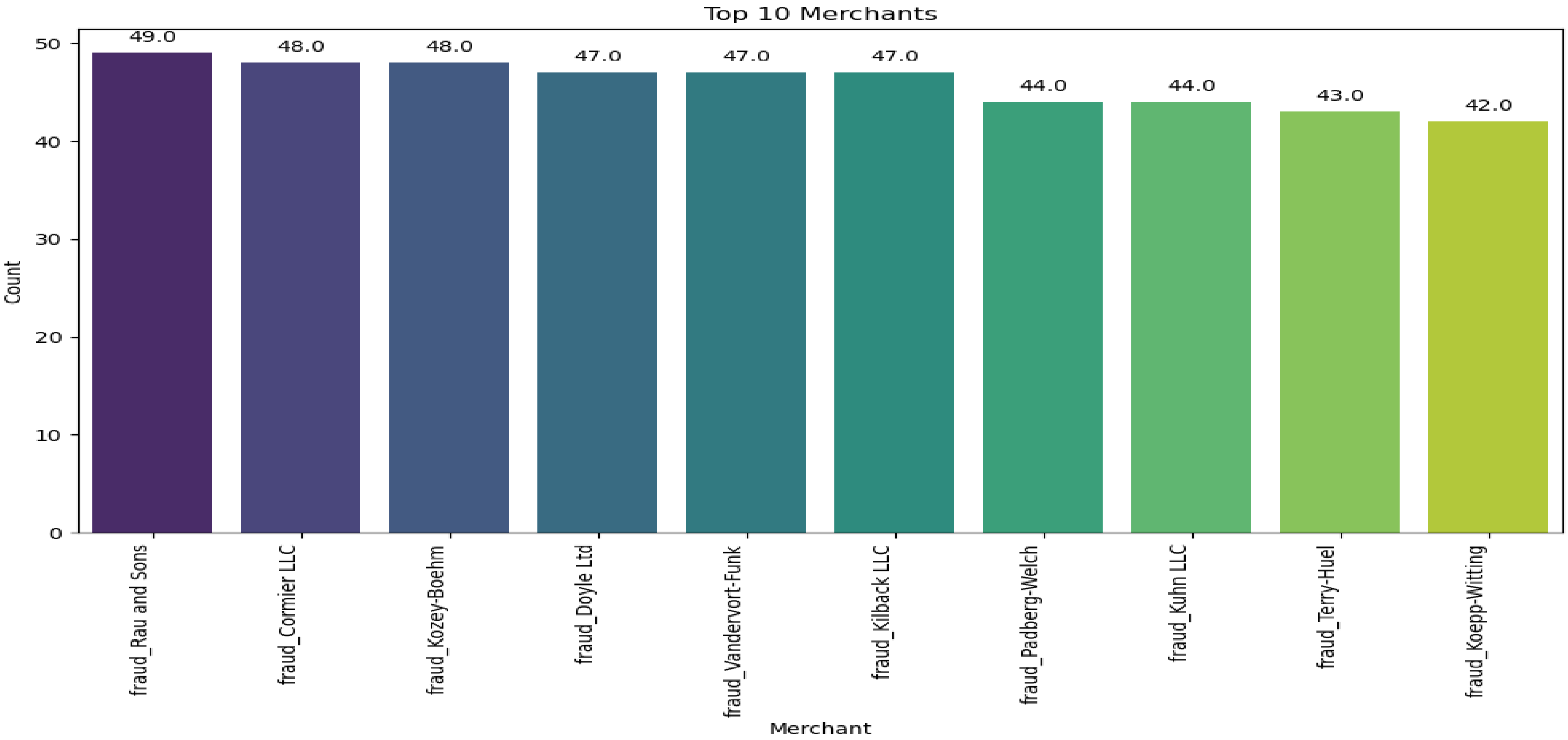
- Top 10 Cities with highest number of fraudulent activites.



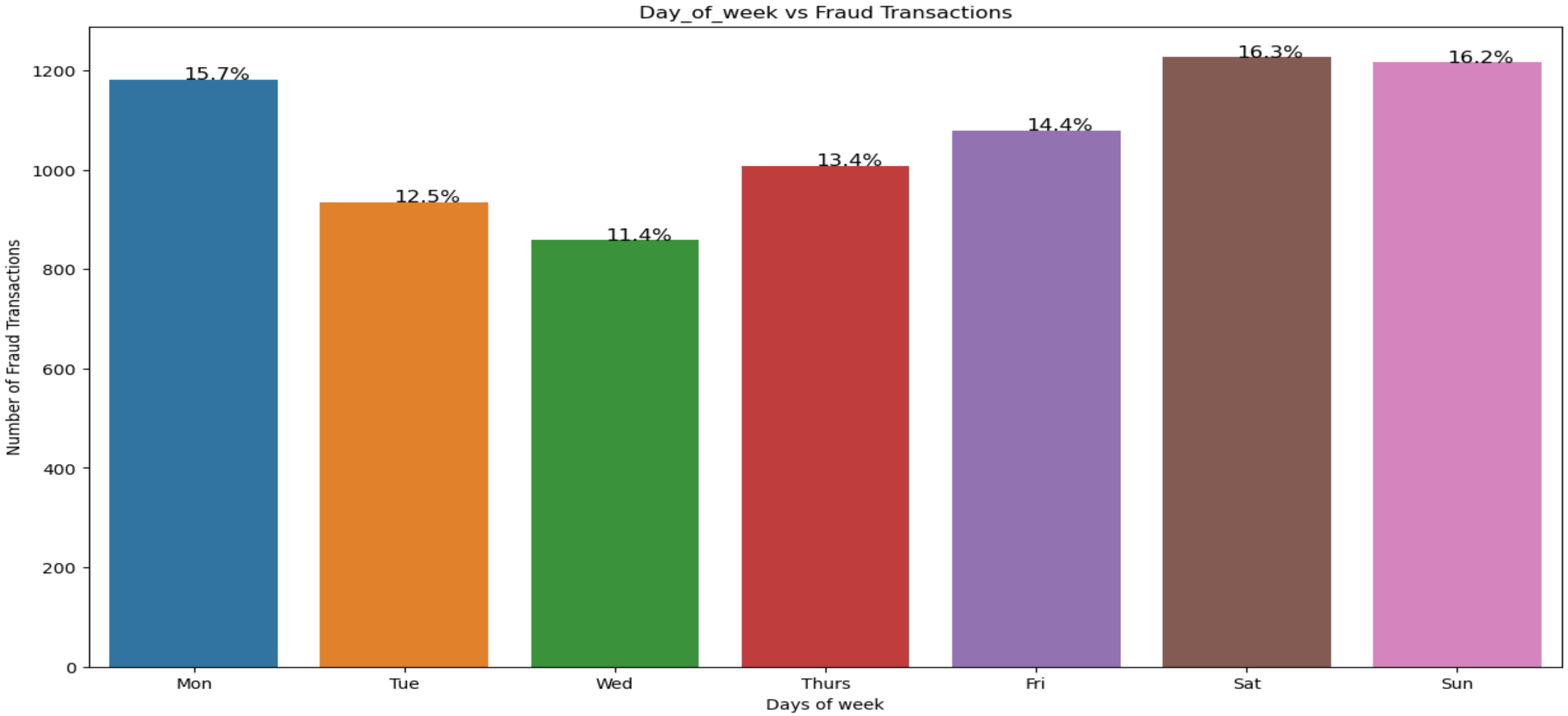
- These are the top 10 states where most number of fraudulent activities occured.



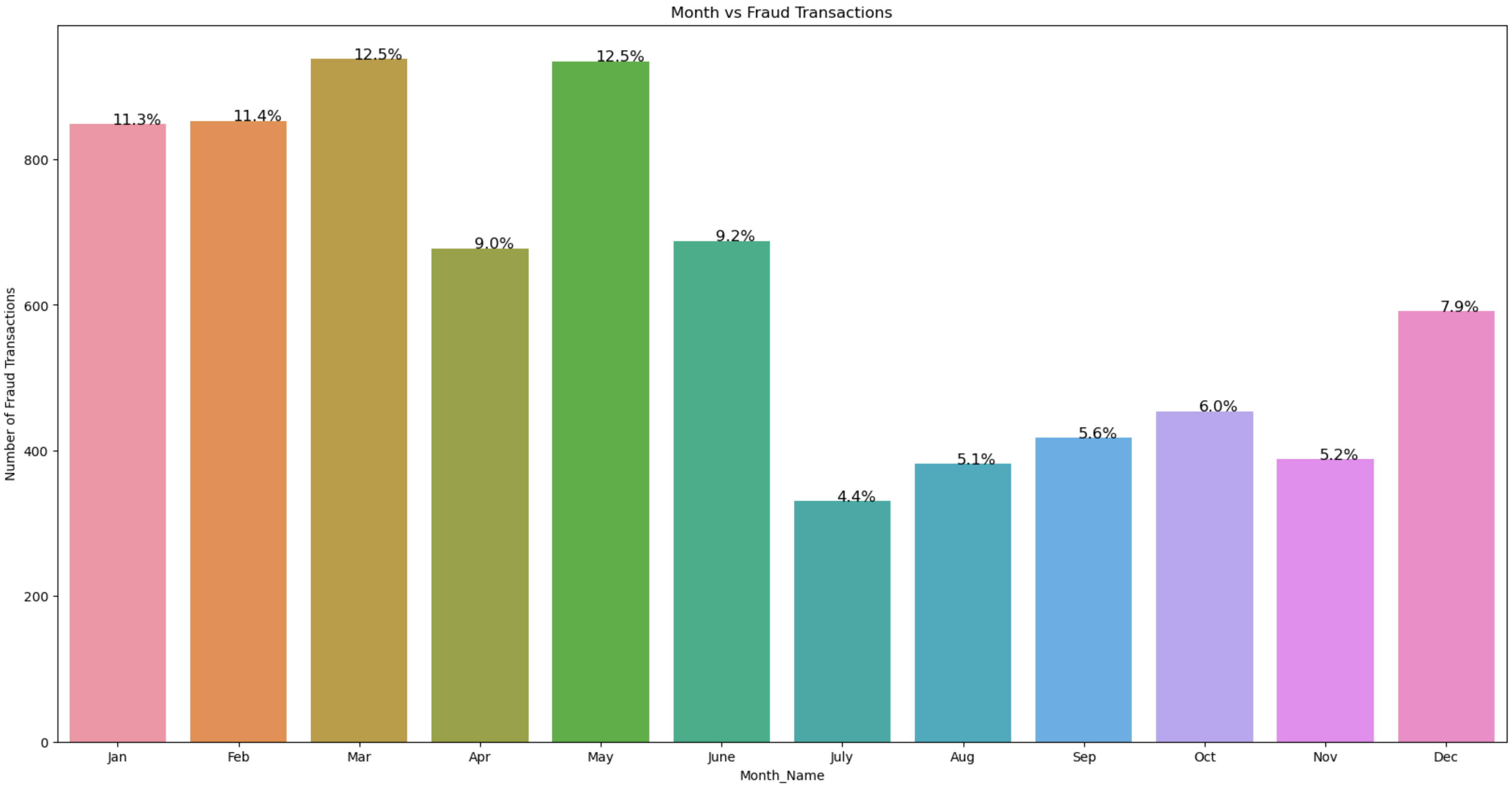
These are the top 10 jobs on whom most number of fraudulent activities occurred.



- These are the top 10 merchant on whom most number of fraudulent activities occurred.



- Most number of fraudulent activities occurred on Monday, Saturday and Sunday.



- Most number of fraudulent activities occurred in First 6 month.

Cost Benefit Analysis

- Part 1 – Cost Benefit Analysis

Cost Benefit Analysis		
S. No	Questions	Answer
a	Average number of transactions per month	77183
b	Average number of fraudulent transaction per month	402
c	Average amount per fraud transaction	531

❑ Part 2 – Cost Benefit Analysis

Cost Benefit Analysis		
S. No	Questions	Answer
1	Cost incurred per month before the model was deployed (b*c)	213392
2	Average number of transactions per month detected as fraudulent by the model (TF)	136
3	Cost of providing customer executive support per fraudulent transaction detected by the model	$1.5
4	Total cost of providing customer support per month for fraudulent transactions detected by the model (TF*$1.5)	204
5	Average number of transactions per month that are fraudulent but not detected by the model (FN)	16
6	Cost incurred due to fraudulent transactions left undetected by the model (FN*c)	8491
7	Cost incurred per month after the model is built and deployed (4+6)	8695
8	Final savings = Cost incurred before - Cost incurred after(1-7)	204697

Appendix : Data Methodology

- Multiple Machine Learning Model Classifier were built on the top of the provided dataset.
- Class Imbalance was adjusted using Adaptive Synthetic (ADASYN) sampling method.
- Manual Hyperparameter tuning is done due to extensive computational times when we are using Grid Search Cross Validation.
- Using ADASYN Sampling method in training the data ,we found the significant result, so we used ADASYN sampling method for unseen data.

Appendix : Attached Files

- Cost Benefit Analysis:

  Cost+Benefit+Analysis.xlsx.

- Multiple Machine Learning Model Deployments:

  project final.ipynb

- Video Submission Link :

  https://drive.google.com/file/d/1opC0m2A33xPSsGNy5yZUjY5vwnLmzshY/view?usp=sharing


THANK
YOU