# CREDIT CARD FRAUD DETECTION

**Submitted By:**

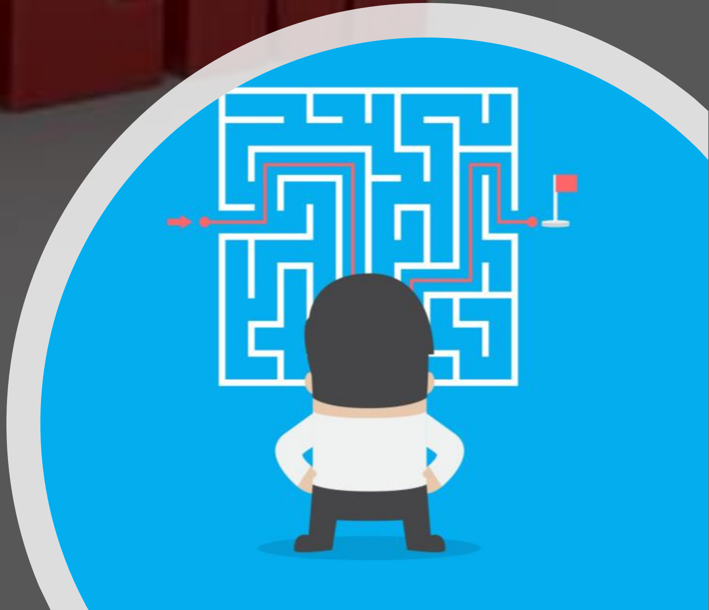
Suzanna Michelle Young

Deepak R

Varsha B S

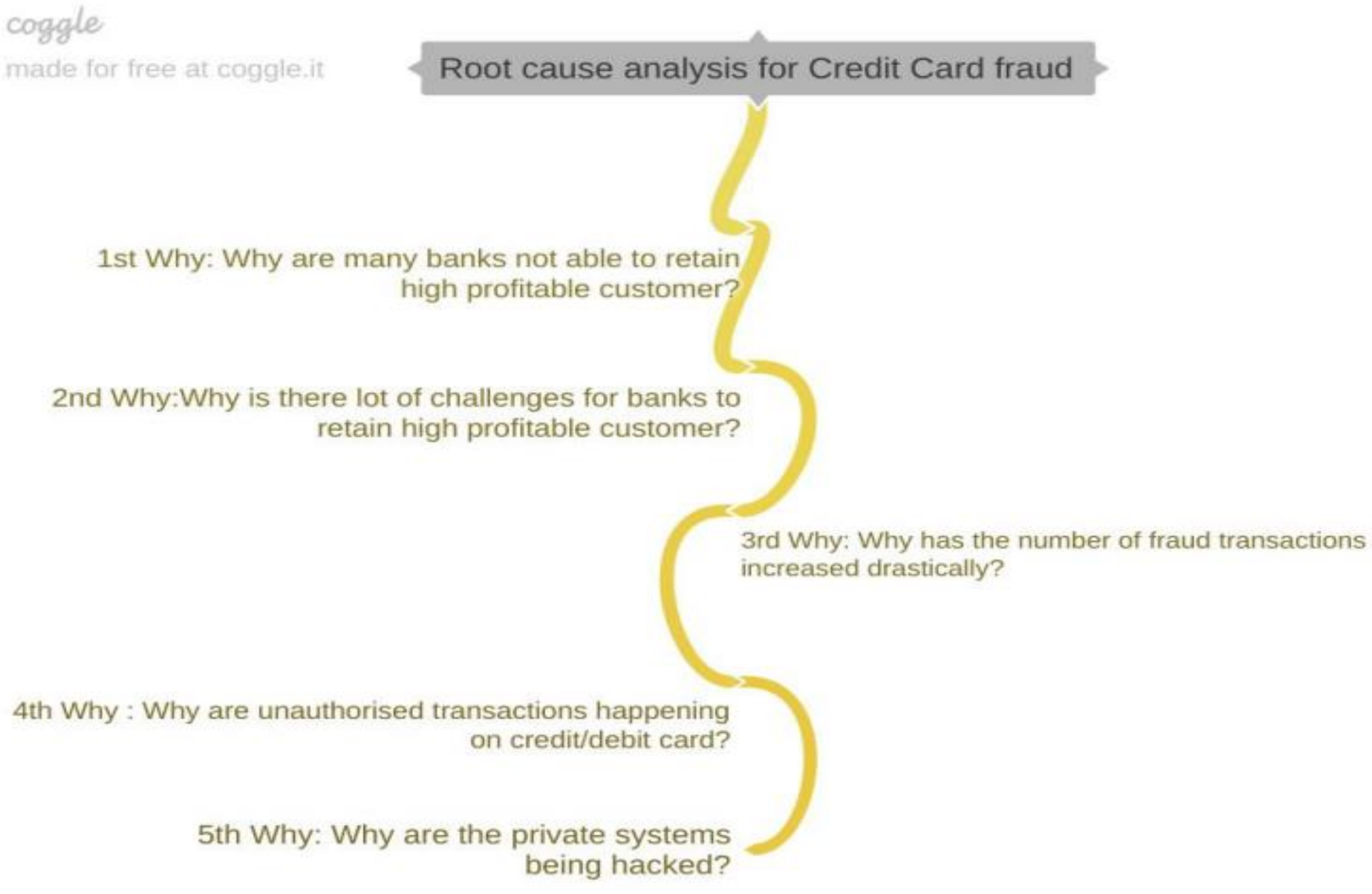
# BACKGRAOUND OF PROBLEM

- Huge revenue loss due to unauthorised transactions during non-peak and odd hours of the day.
- Source: Stolen/lost cards, ATM skimming at various POS terminals.
- Stopping the fraudsters/fraudulent activity is highly impossible as they actively find alternate ways.
- Extra level authentication is complex & tedious job.
- Goal is to detect and prevent the fraudulent transactions before the financial crisis.



# ROUTE CAUSE ANALYSIS

- 5 WHYs
- Sensitive credit card details of customers is being accessed.
- Private systems are hacked, stolen/lost cards are misused, ATM skimming at POS terminals.



Root cause analysis for Credit Card fraud

1st Why: Why are many banks not able to retain high profitable customer?

2nd Why:Why is there lot of challenges for banks to retain high profitable customer?

3rd Why: Why has the number of fraud transactions increased drastically?

4th Why : Why are unauthorised transactions happening on credit/debit card?

5th Why: Why are the private systems being hacked?

# UNDERSTANDING THE PROBLEM & IMPACT

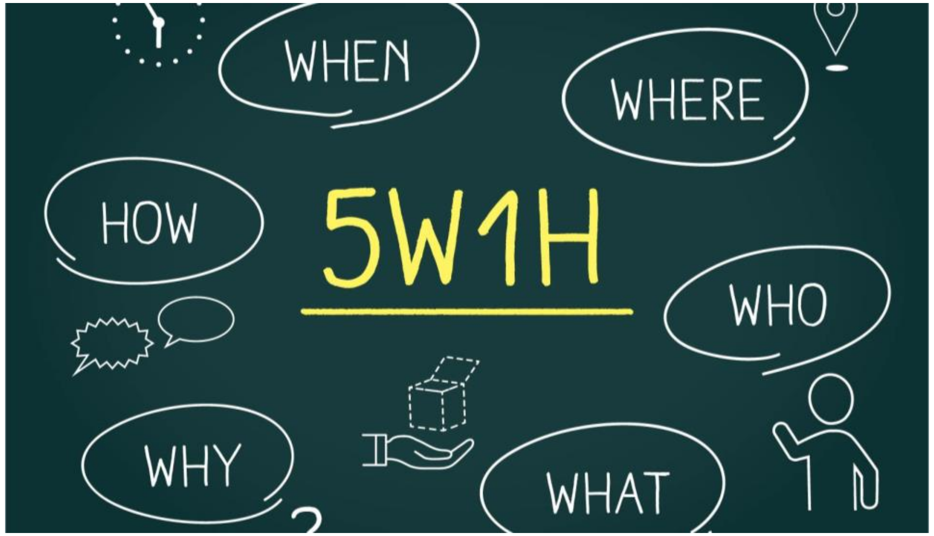
1. Who is involved in the process?
    - Fraudsters, skimmers who get unauthorized access to a customer's credit card info.

2. What do they do with it?
    - Commit fraudulent transactions: alter genuine cards, create counterfeit cards, steal credit cards, fraudulent telemarketing.

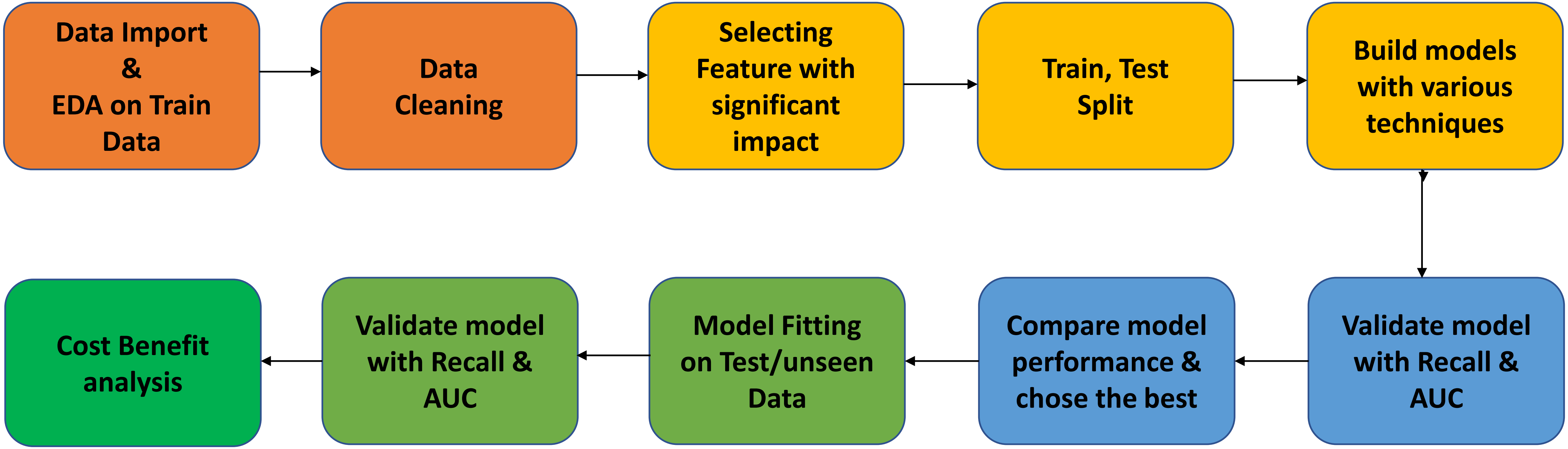
3. Where do the transactions happen?
    - Most of them happen online and sometimes offline.
    - Online frauds are hard to detect due to difficulty in amassing evidence, time & resource constraints.

4. When does it happen?
    - Most happen at non-peak and odd hours of the day when user in-active or asleep.

5. How is business affected by this?
    - As bank need to repay customer the total transaction amount in case of fraudulent transaction
    - Banks suffer from substantial financial loss, lose trust & customer's credibility.



# PROCESS FLOW



Data Import & EDA on Train Data

Data Cleaning

Selecting Feature with significant impact

Train, Test Split

Build models with various techniques

Validate model with Recall & AUC

Compare model performance & chose the best

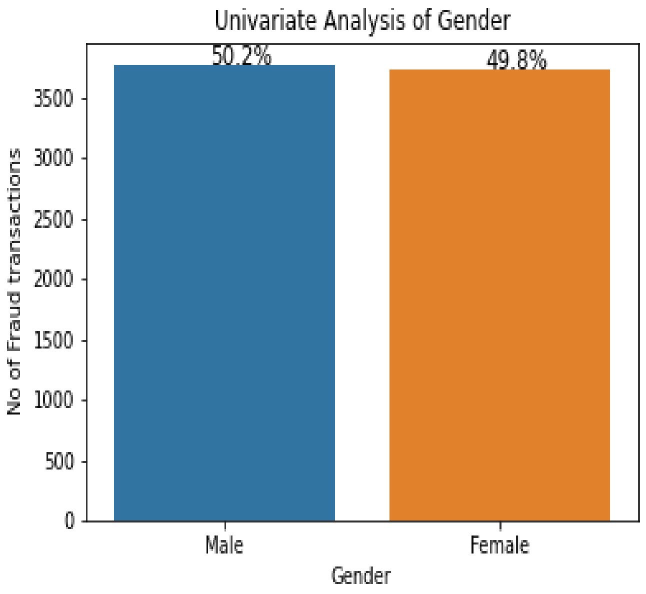
Model Fitting on Test/unseen Data

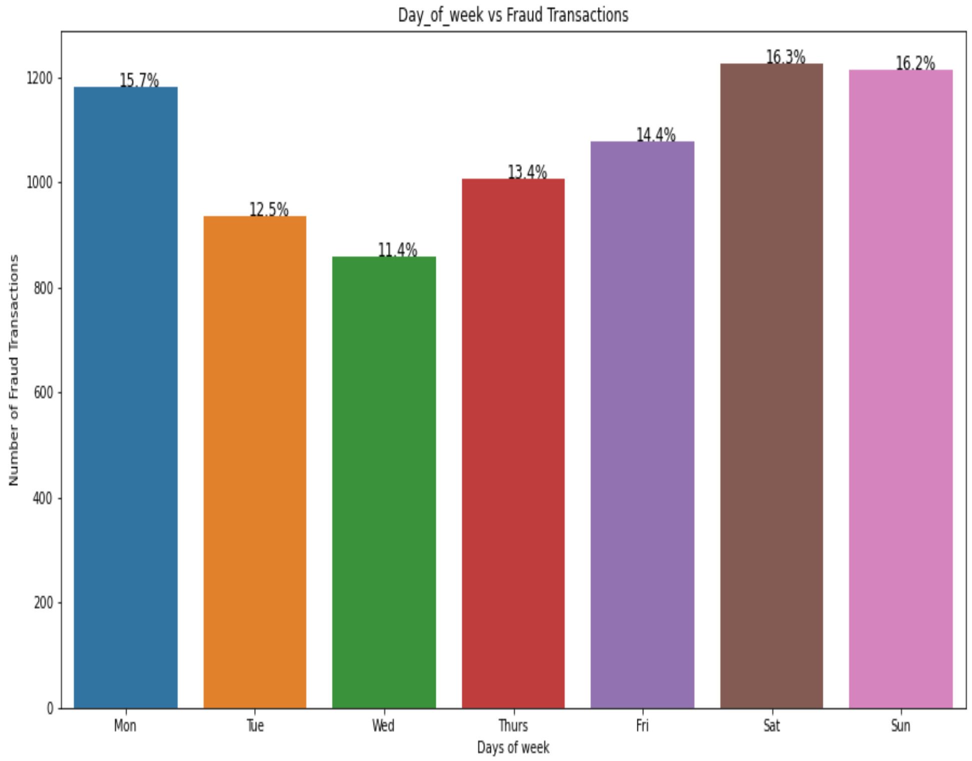
Validate model with Recall & AUC

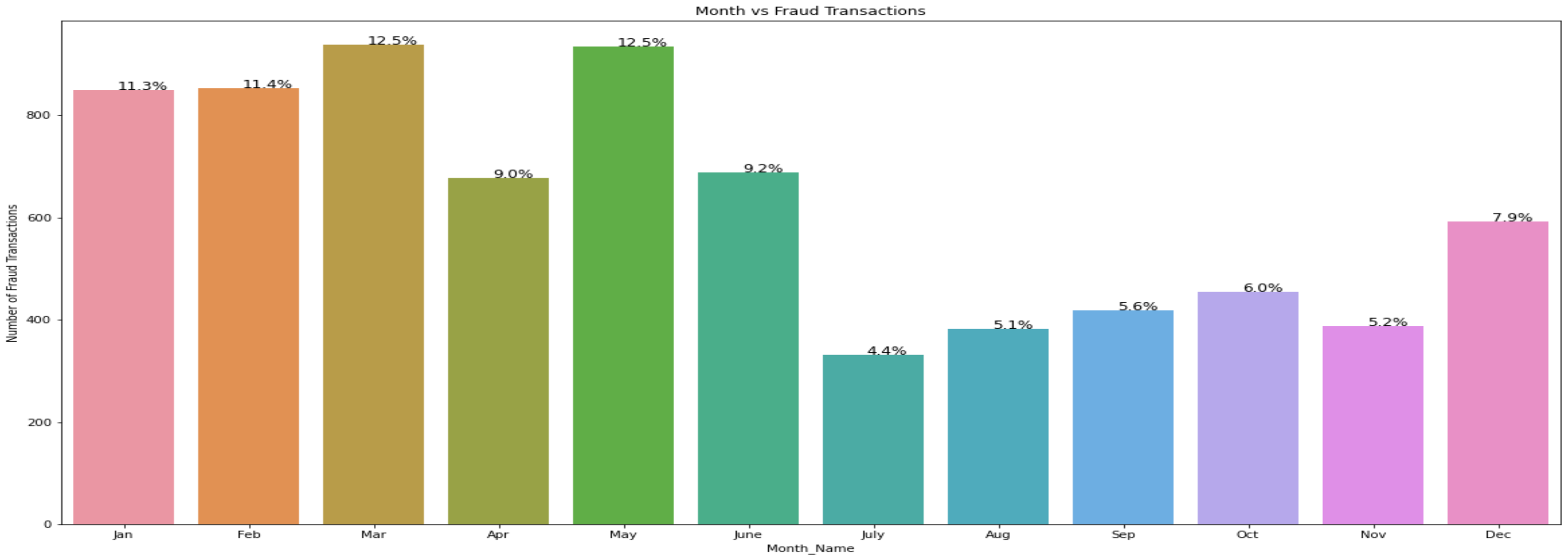
Cost Benefit analysis

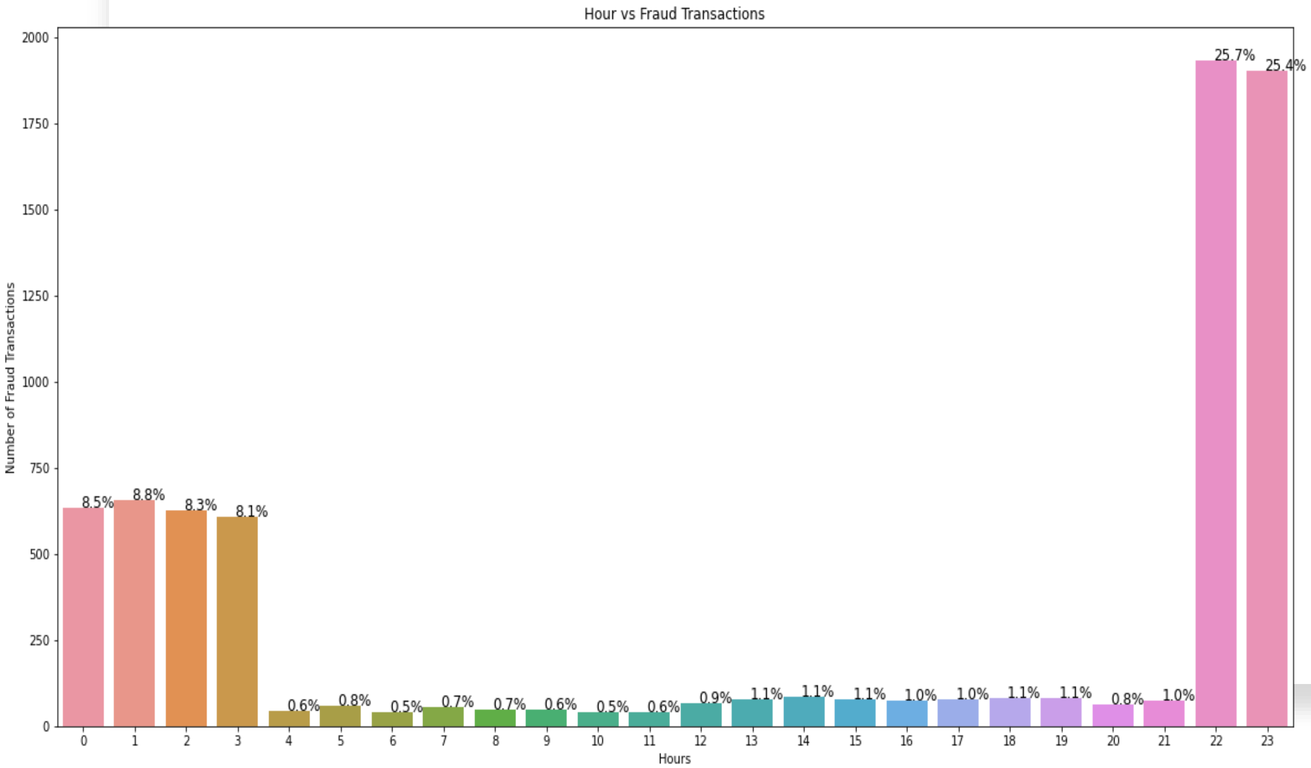
# CHARACTERISTICS OF FRAUDULENT TRANSACTIONS

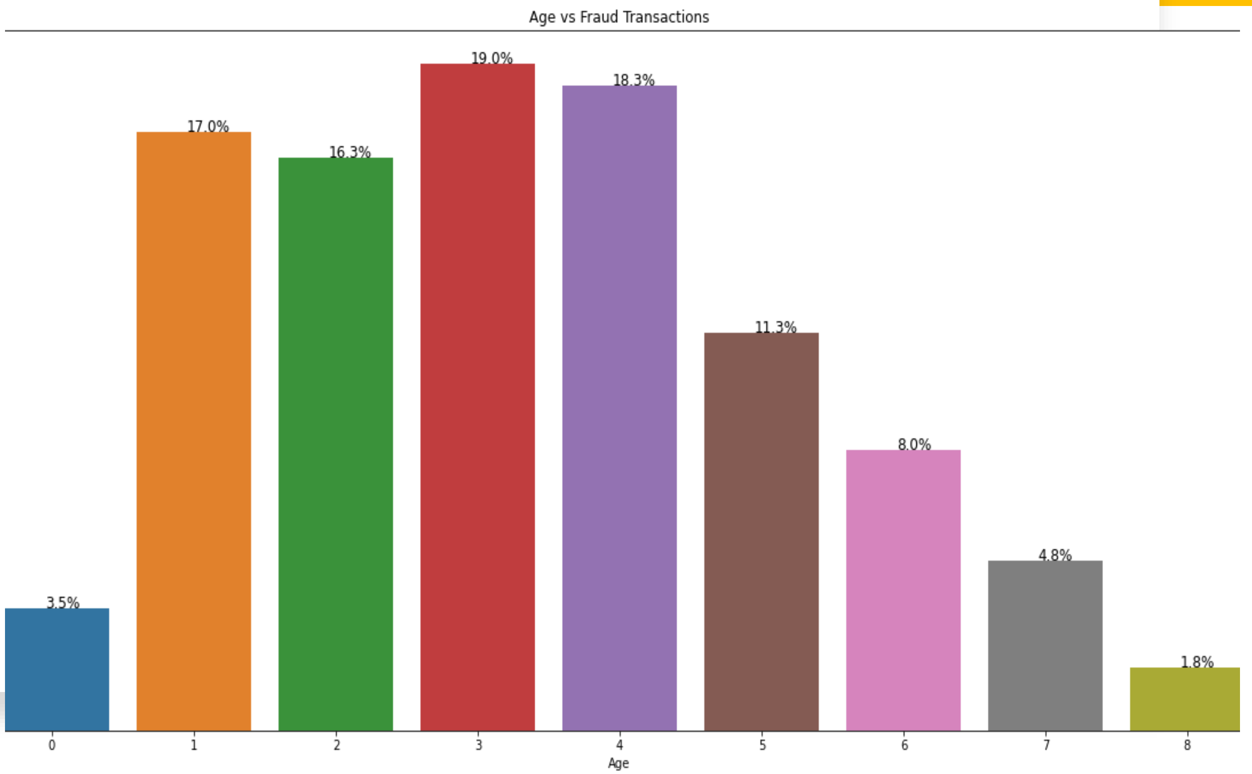
- Gender v/s fraudulent transactions
- Day of the week v/s fraudulent transactions
- Month v/s fraudulent transactions











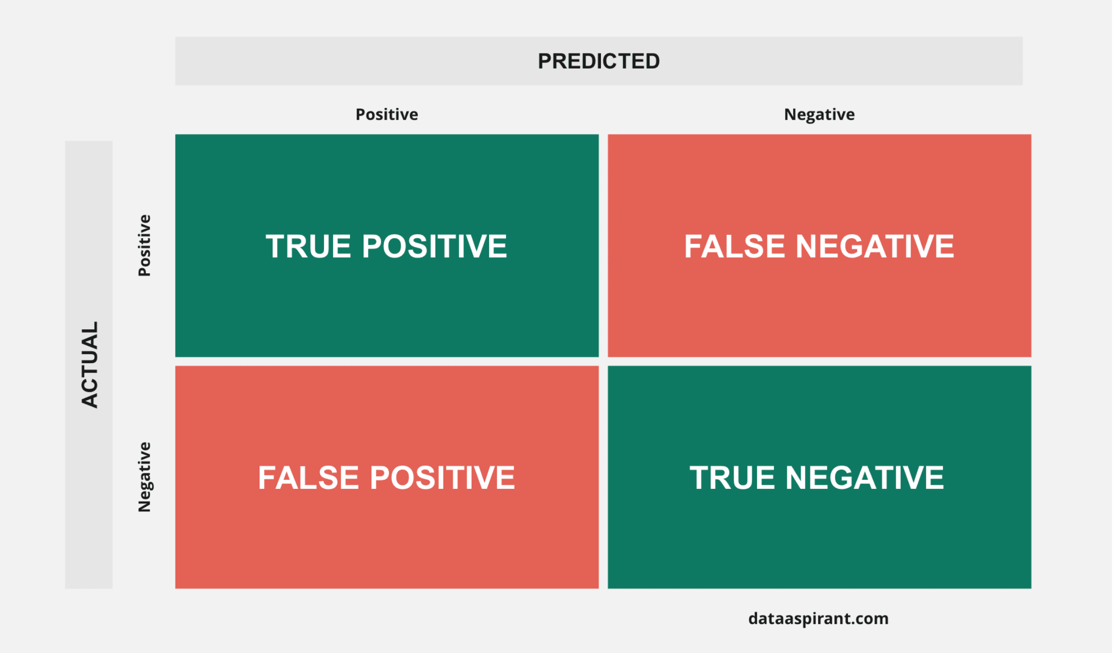
# CHARACTERISTICS OF FRAUDULENT TRANSACTIONS

- Hours of day v/s transactions
- Age v/s fraudulent transactions

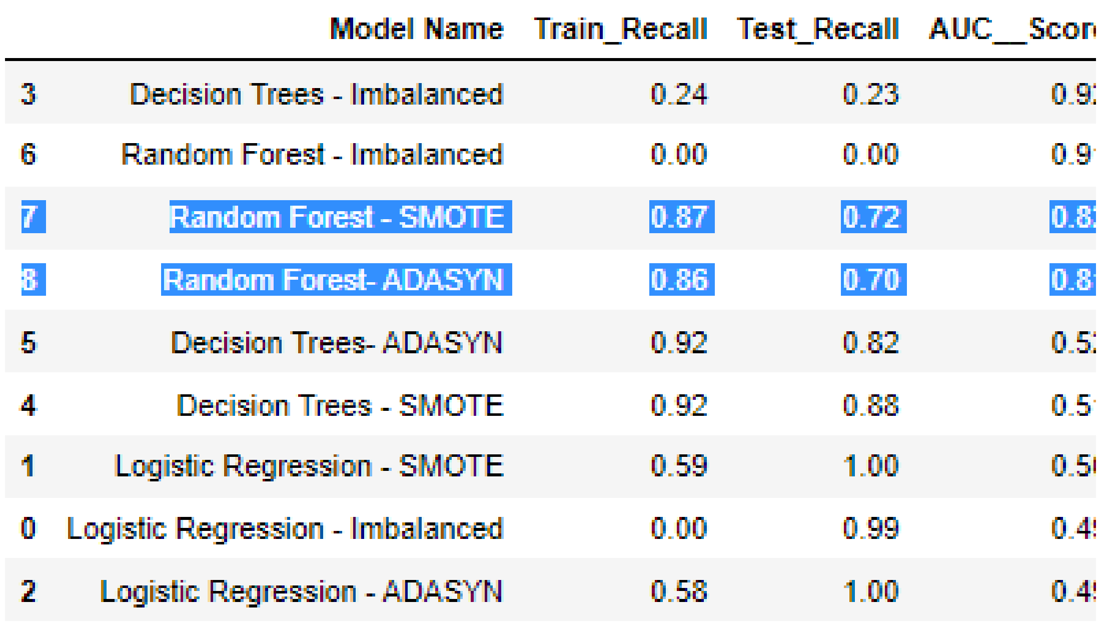
# MODEL PERFORMANCE ON TRAINING DATA

Evaluation metrics: Recall, AUC

1. Why Recall?
   - Most likely to predict actual Positive cases(Fraud transactions).
   - Very helpful to reduce financial loss.
   - Recall = TP / (TP + FN)
   - High Recall refers to low false negative score.
2. Why AUC?
   - Best to measure overall performance of a binary classification problem.
   - Best to select optimal threshold value of a model.
   - Plots TPR v/s FPR.



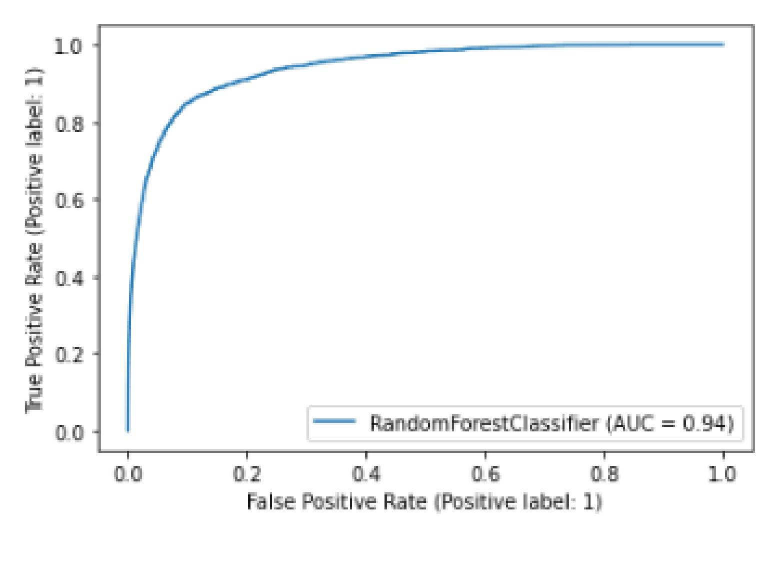
dataaspirant.com



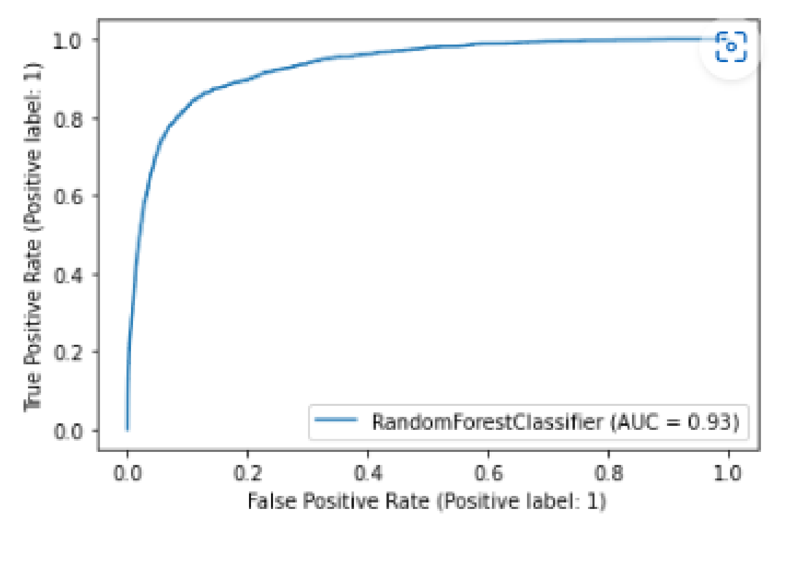
| | Model Name | Train_Recall | Test_Recall | AUC__Scor |
|---|---|---|---|---|
| 3 | Decision Trees - Imbalanced | 0.24 | 0.23 | 0.9 |
| 6 | Random Forest - Imbalanced | 0.00 | 0.00 | 0.9 |
| 7 | Random Forest - SMOTE | 0.87 | 0.72 | 0.8 |
| 8 | Random Forest- ADASYN | 0.86 | 0.70 | 0.8 |
| 5 | Decision Trees- ADASYN | 0.92 | 0.82 | 0.5 |
| 4 | Decision Trees - SMOTE | 0.92 | 0.88 | 0.5 |
| 1 | Logistic Regression - SMOTE | 0.59 | 1.00 | 0.5 |
| 0 | Logistic Regression - Imbalanced | 0.00 | 0.99 | 0.4 |
| 2 | Logistic Regression - ADASYN | 0.58 | 1.00 | 0.4 |

# MODEL PERFORMANCE ON TEST/UNSEEN DATA

- Trade off b/w Random Forest SMOTE & Random Forest ADASYN.
  - Still choose Random Forest ADASYN why?
  - Impact from 1% increase in Recall score.
  - 1% of overall transaction = 18,524 (Train+Test)
- Assumption: Financial impact of 1% fraud in 1% of overall transaction
  - 1% of Fraudulent transactions = 0.01% overall transaction = 185 transactions.
  - Avg amt for 185 transaction = (avg amount/transaction) * $185 = $98,050



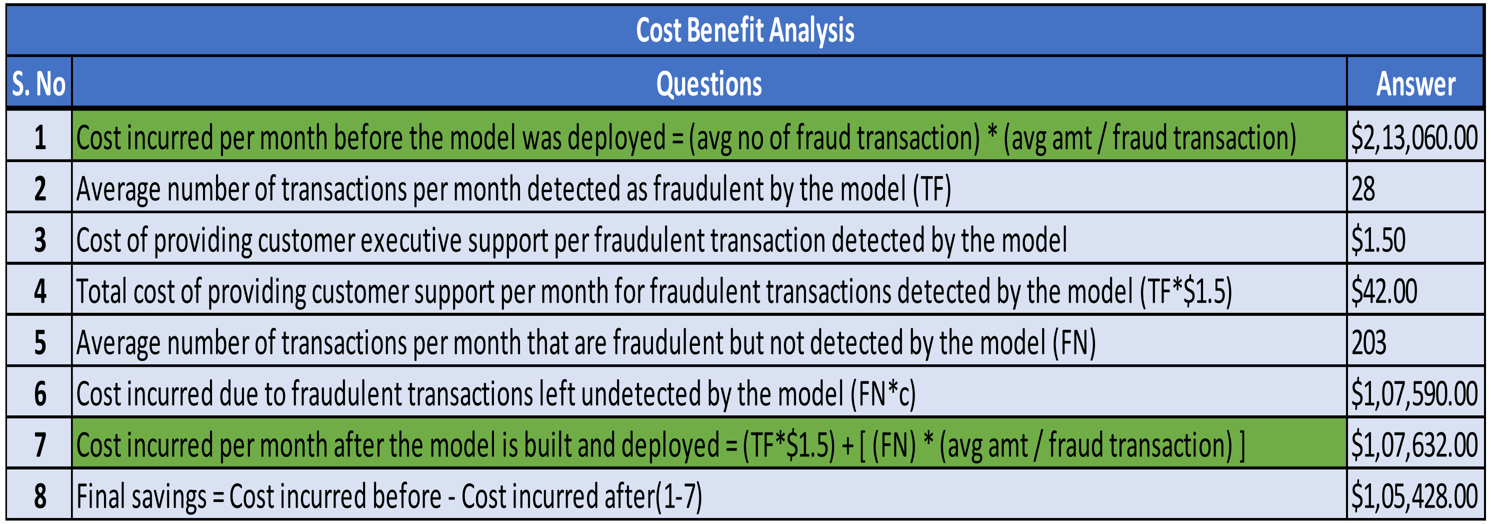
Random forest SMOTE - Metrics : Recall :0.86 AUC :0.94



Random forest ADASYN - Metrics : Recall :0.87 AUC :0.93

# COST BENEFIT ANALYSIS

| Cost Benefit Analysis | | |
|---|---|---|
| **S. No** | **Questions** | **Answer** |
| **a** | Average number of transactions per month | 77183.00 |
| **b** | Average number of fraudulent transaction per month | 402.00 |
| **c** | Average amount per fraud transaction | 530.00 |



| Cost Benefit Analysis | | |
|---|---|---|
| S. No | Questions | Answer |
| 1 | Cost incurred per month before the model was deployed = (avg no of fraud transaction) * (avg amt / fraud transaction) | $2,13,060.00 |
| 2 | Average number of transactions per month detected as fraudulent by the model (TF) | 28 |
| 3 | Cost of providing customer executive support per fraudulent transaction detected by the model | $1.50 |
| 4 | Total cost of providing customer support per month for fraudulent transactions detected by the model (TF*$1.5) | $42.00 |
| 5 | Average number of transactions per month that are fraudulent but not detected by the model (FN) | 203 |
| 6 | Cost incurred due to fraudulent transactions left undetected by the model (FN*c) | $1,07,590.00 |
| 7 | Cost incurred per month after the model is built and deployed = (TF*$1.5) + [ (FN) * (avg amt / fraud transaction) ] | $1,07,632.00 |
| 8 | Final savings = Cost incurred before - Cost incurred after(1-7) | $1,05,428.00 |

THANK
YOU