BROOM SOLUTIONS: CREDIT CARD FRAUD DETECTION

TEAM 2 – DUKSS ANALYTICS:
DHARMI, SHWETA, KEERTHANA, URVAJ, SUNVID

AGENDA

- BUSINESS PROBLEM
- DATA
- METHODOLOGY
- SOLUTION

BUSINESS PROBLEM

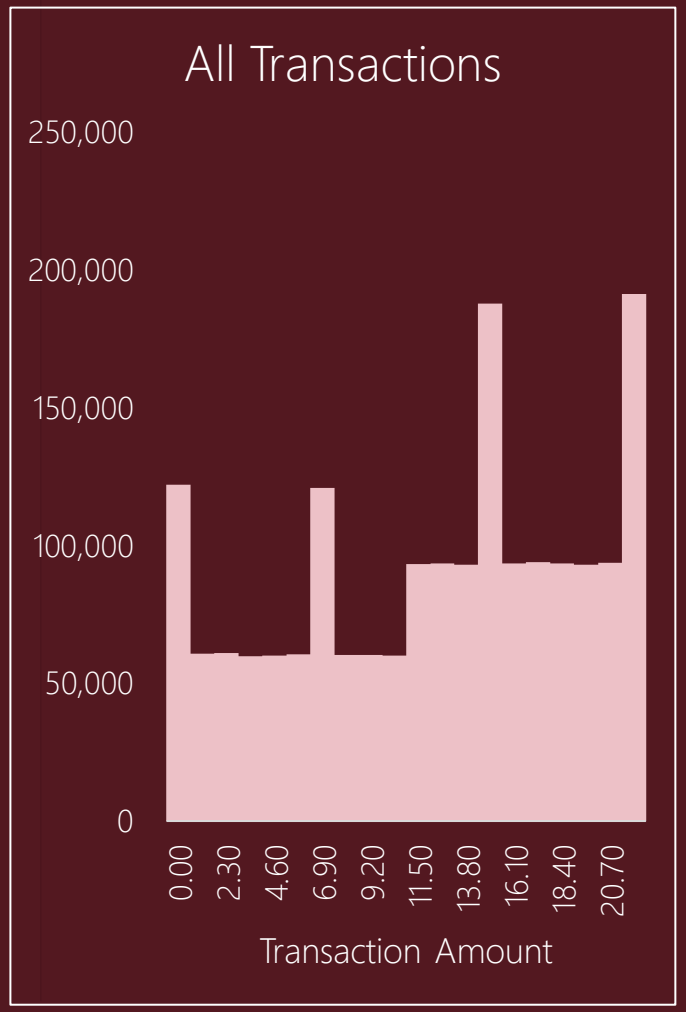
Broom Solutions' current model is ***flagging many non-fraudulent transactions as fraudulent.*** As a result, their ***retail clients are frustrated*** with the ***loss of revenue*** *and* ***customers.***

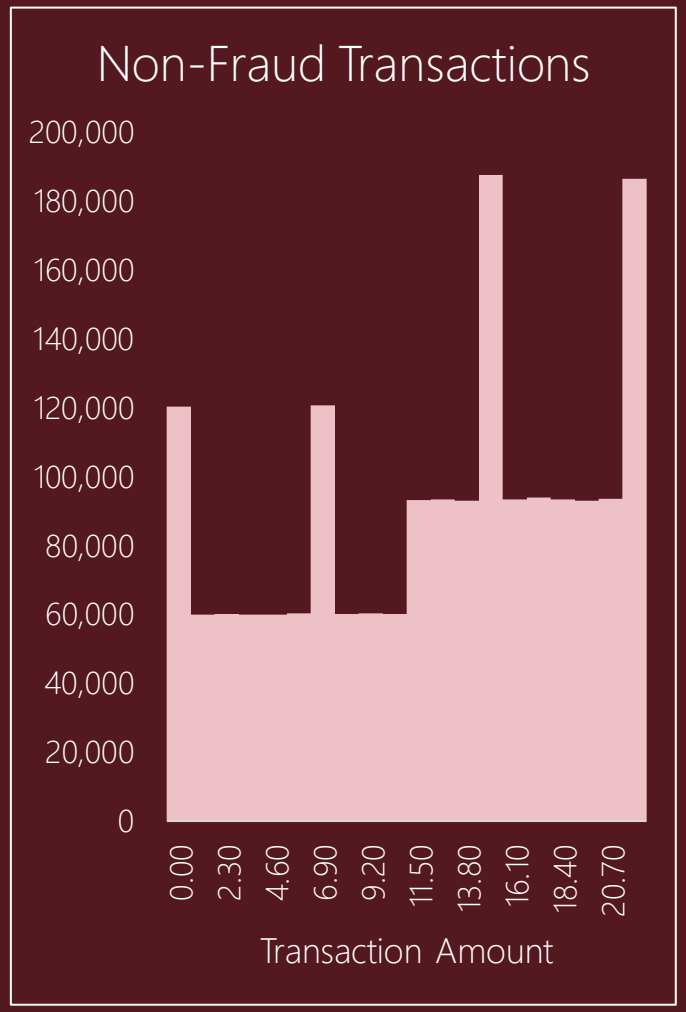
DATA

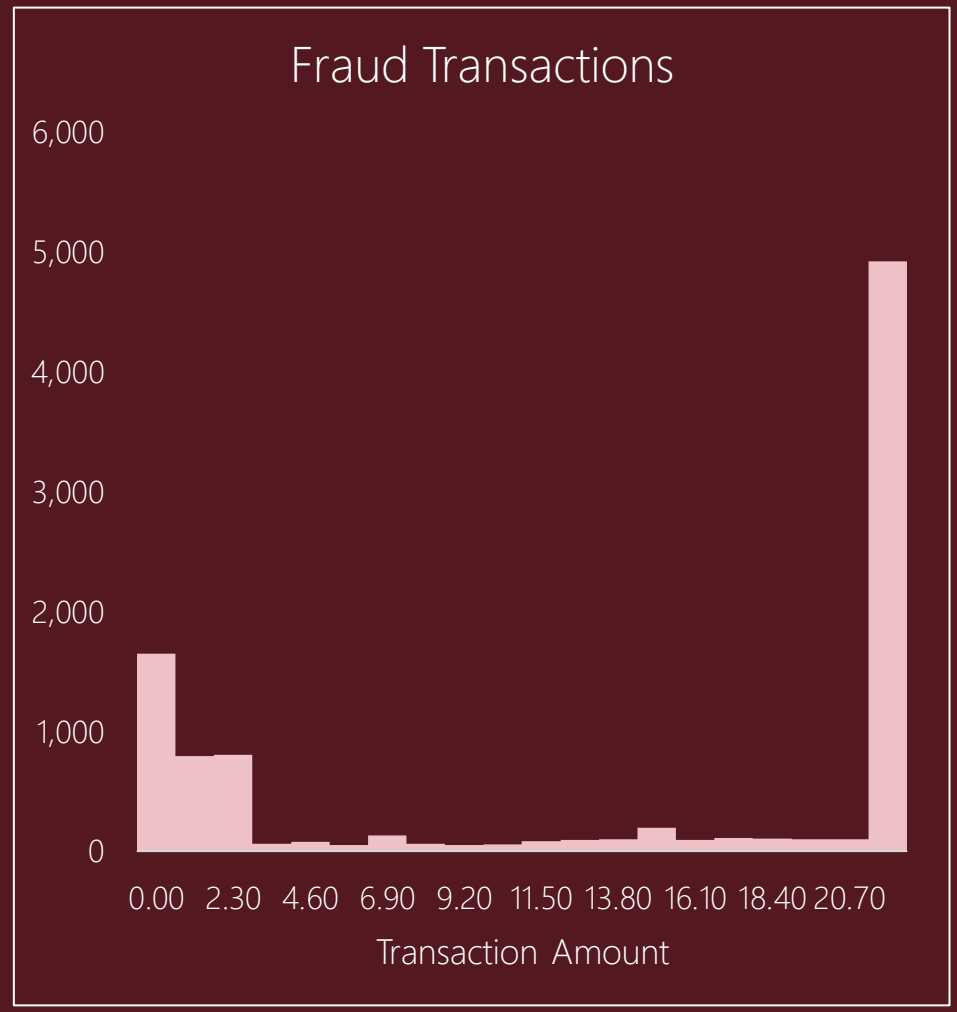
"It is a capital mistake to theorize before one has data."

- Sherlock Holmes in "A study in Scarlet" by Arthur Conan Doyle

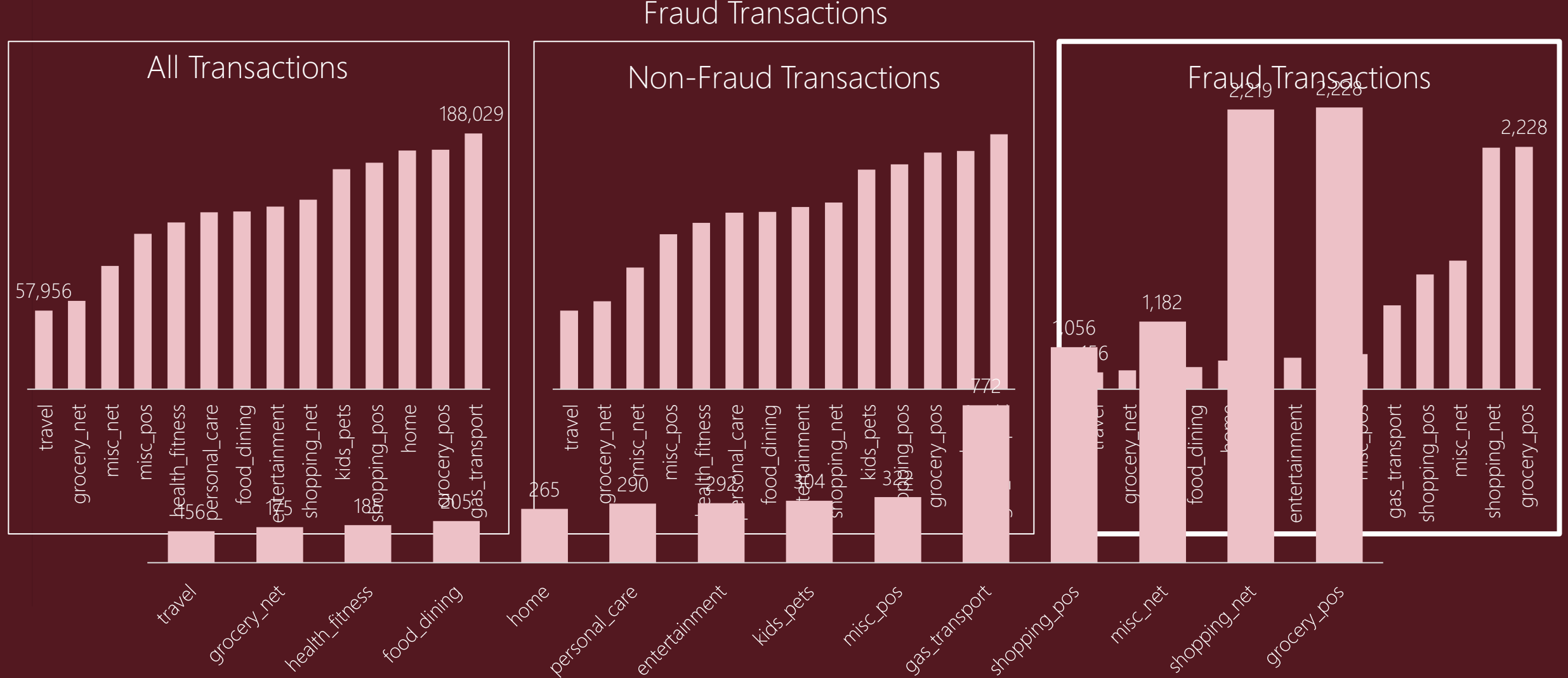
TRANSACTION PER HOUR



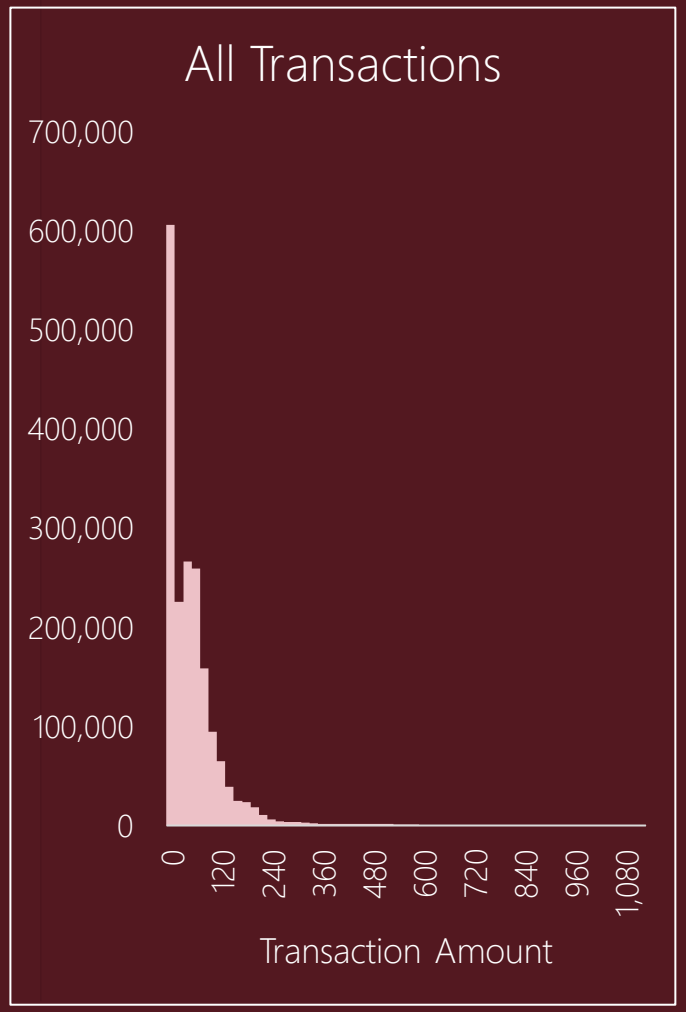


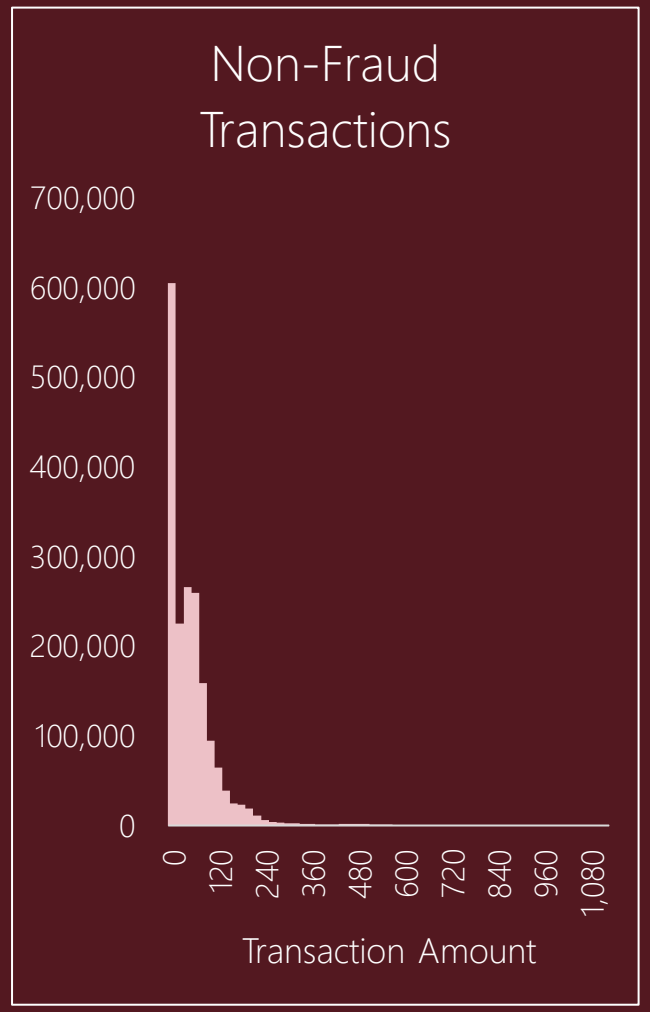


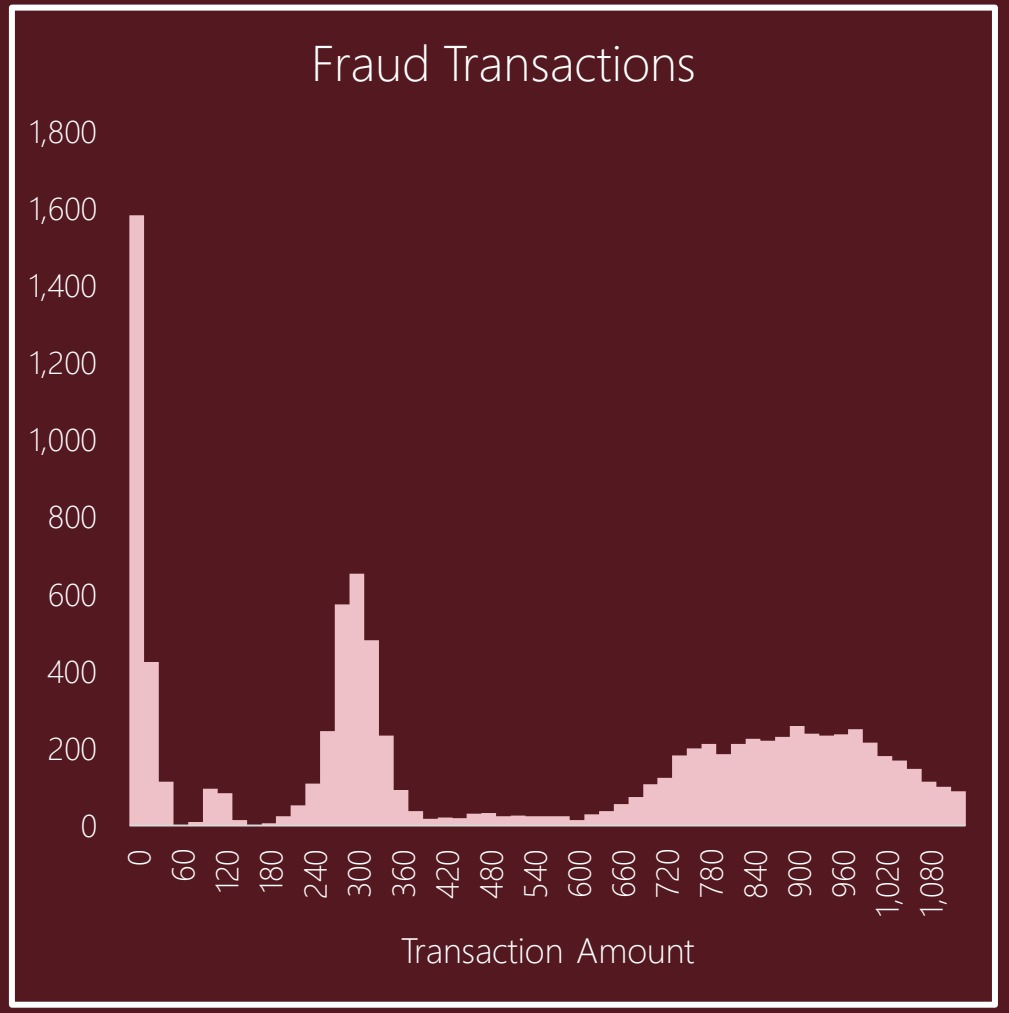
TRANSACTIONS PER CATEGORY



TRANSACTION AMOUNT







FRAUD DETECTION DATA SUMMARY

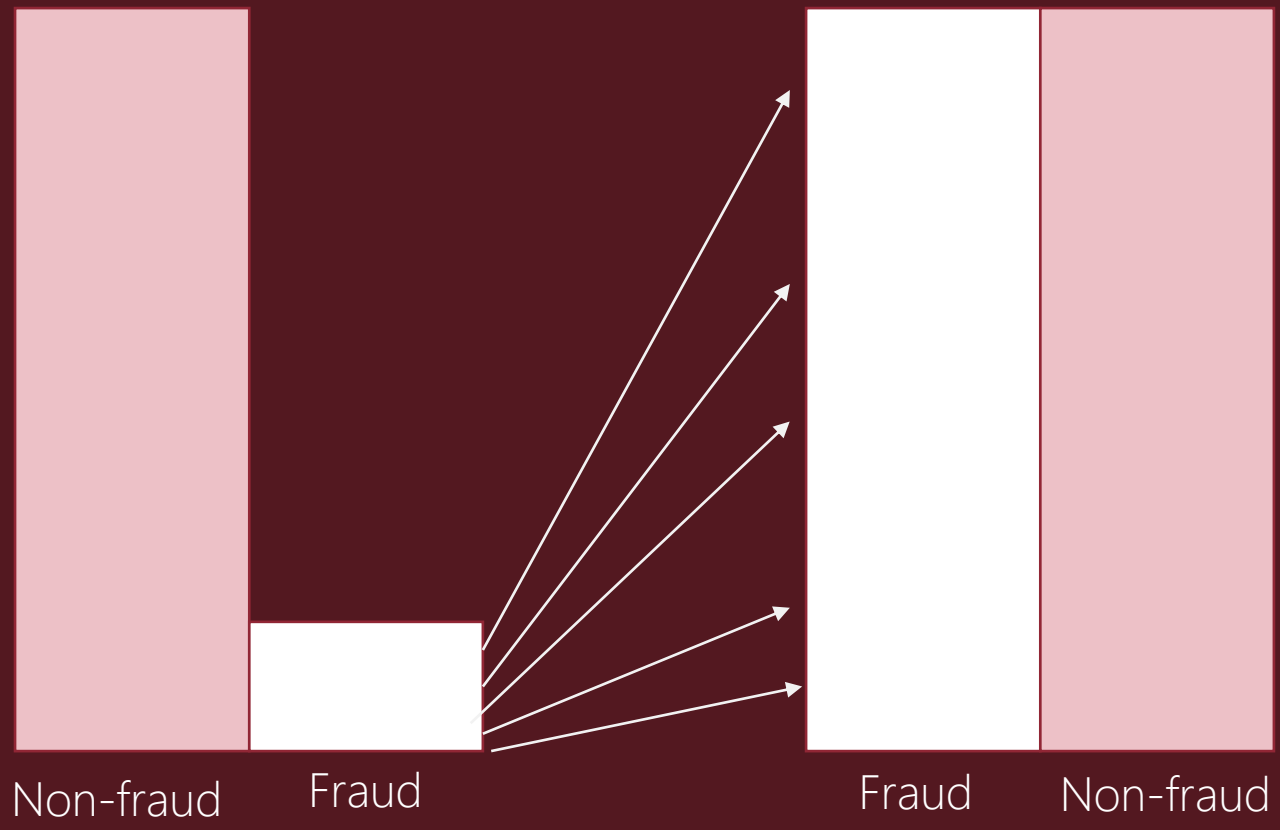
Number of transactions	18,52,394
Number of non-fraud transactions	~99%
Number of fraud transactions	~0.5%
21 features	1 target variable

- No null values in the data
- The data is highly imbalanced with ~0.5% of the transactions being fraud.

METHODOLGY

OVERSAMPLING

Since the data is imbalanced with ~0.5% of the transactions being fraud, oversampling was done to balance the data categories better.



MODELLING

LOGISTIC REGRESSION

	Precision	Recall	F1-score
0	0.81	0.90	0.85
1	0.88	0.79	0.83

DECISION TREE

	Precision	Recall	F1-score
0	0.85	0.93	0.89
1	0.92	0.84	0.88

RANDOM FOREST

	Precision	Recall	F1-score
0	0.86	0.92	0.89
1	0.92	0.85	0.88

XGBOOST

	Precision	Recall	F1-score
0	0.91	0.95	0.93
1	0.95	0.91	0.93

MODELLING

XGBOOST

	Precision	Recall	F1-score
0	0.91	0.95	0.93
1	0.95	0.91	0.93

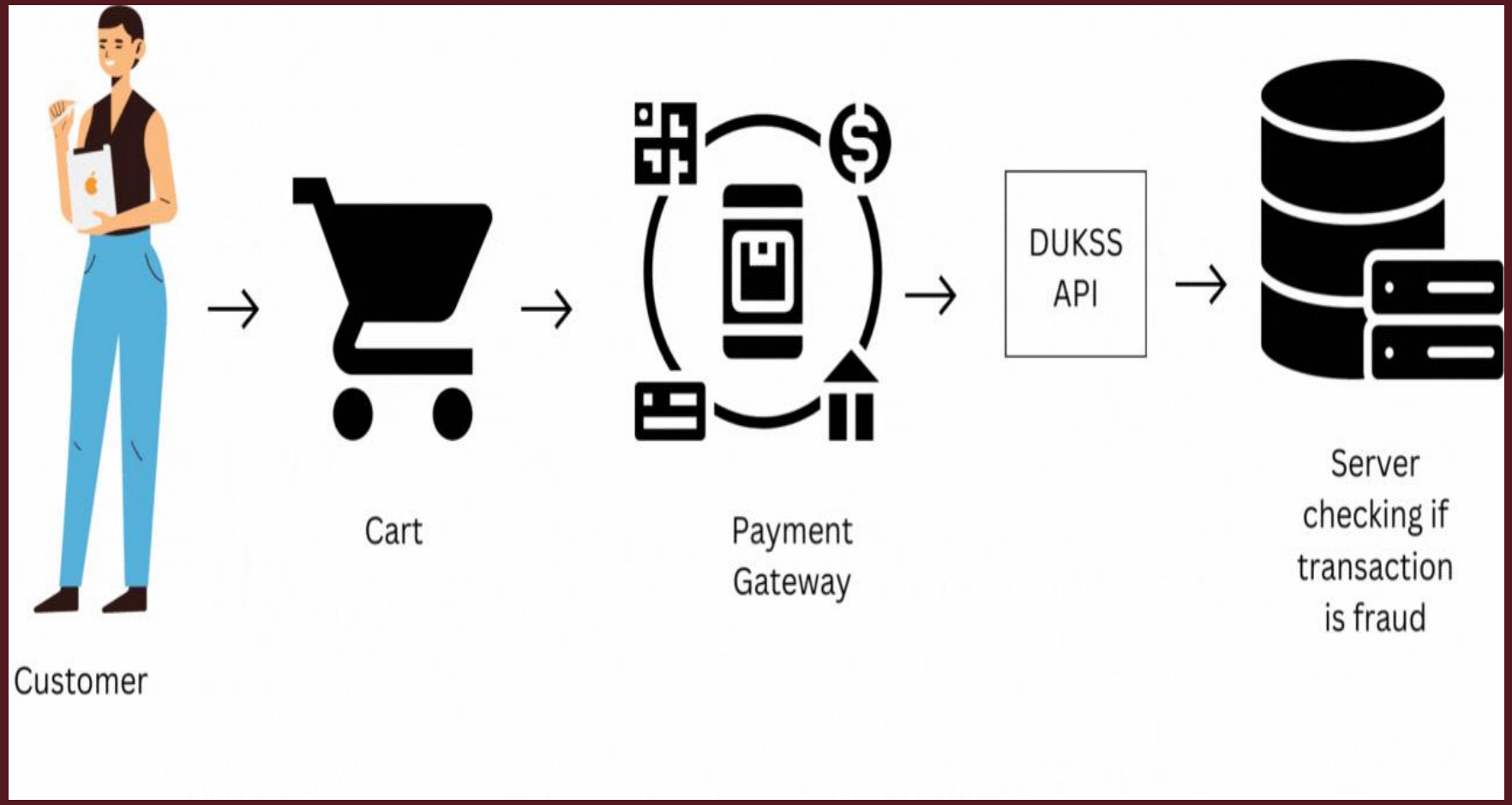
$$Precision = \frac{True\ Positives}{True\ positives + False\ positives}$$

$$F1 - score = 2(\frac{precision\ * recall}{precision + recall})$$

- XGBoost is the model that identifies a higher number of false positives, which is our main business problem
- Precision improvement can tell us if the algorithm performance improved.

SOLUTION

WHY IMPLEMENT OUR MODEL?



- Our model lets us use data like TC40s, SAFE reports, and early dispute notifications
- If considering a fraud, will perform 3DSecure to enable customer to verify.

CURRENT SITUATION

- Lost customer lifetime value.
- Wasted acquisition costs.
- Brand Damage for retailers

ECONOMIC IMPACT

- 8000 – falsely identified fraud customer
- 6700 – after model, number of falsely identified customers

- Revenue lost for falsely identifying - **$160K**
- With our model, you will be losing - **$125K**

- **Results could be seen within 45 days**

THANK YOU