Providing Solutions to Help Business Grow

Strengthen **fraud detection** and prevention capabilities to **reduce financial losses**, and uphold the bank's reputation.

Introduction

- The company is facing a growing challenge with fraudulent transactions.
- Our current detection system is struggling to adapt to evolving fraud patterns, leading to financial losses, and erosion of customer trust.

Data

500,000 May - July 2024
FROM TRANSACTION RECORDS (SYNTHETICALLY GENERATED)

13 Variables
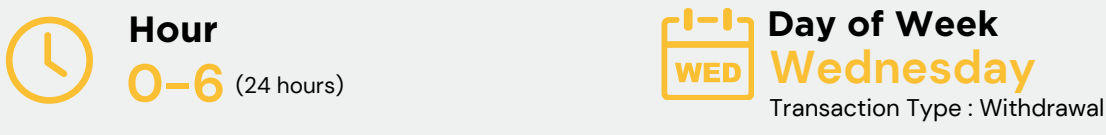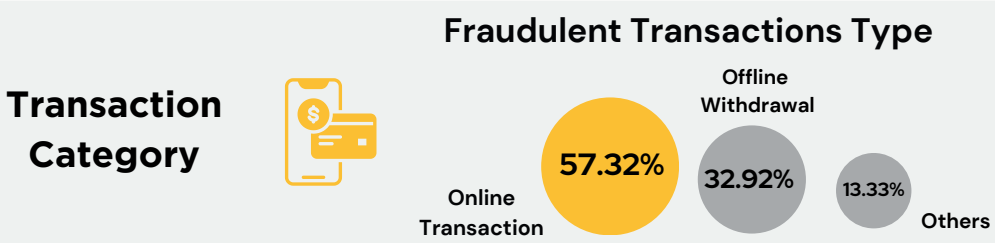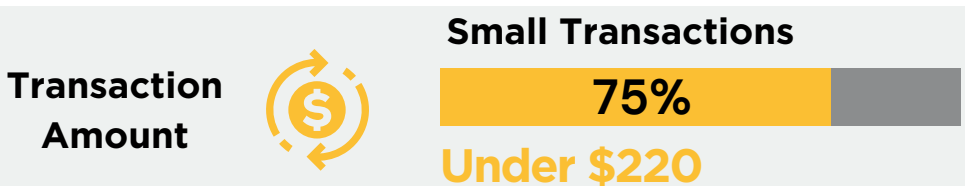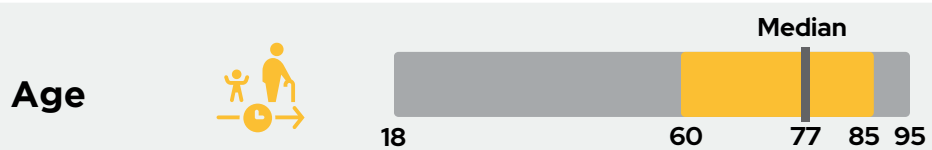IDs, Date, Amount, Location, Type, Balance, Joint Flag, Agent
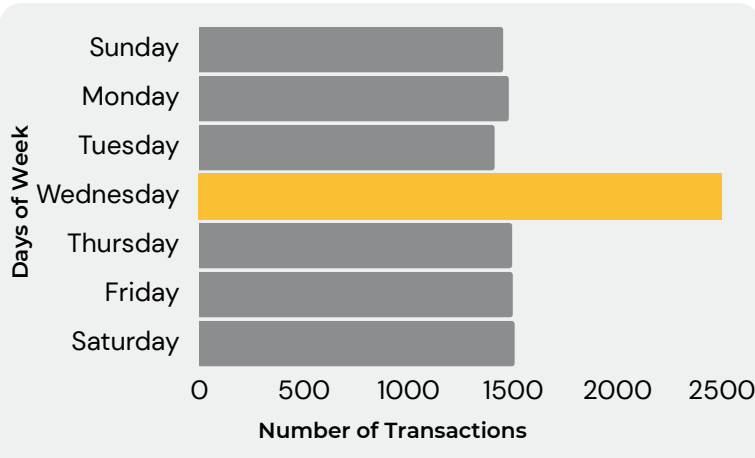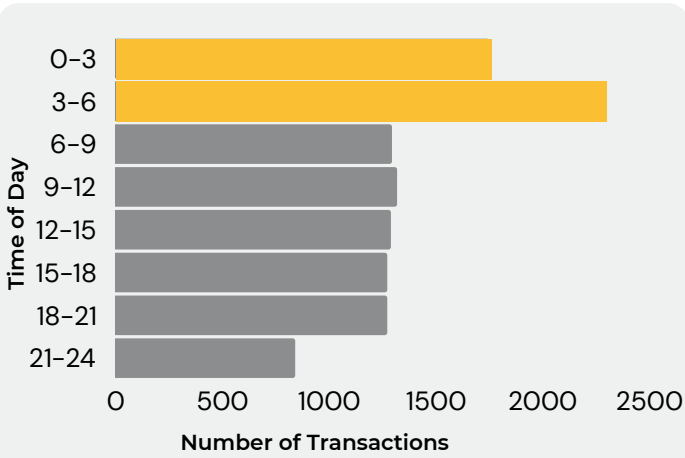
2.2% Fraud

4.7M Loses

WHAT WE SAW

- Fraudulent transactions typically involve smaller amounts. This suggests fraudsters aim to blend in with legitimate activity, exploiting patterns to avoid detection.
- Fraudulent activity is concentrated during specific days and hours, particularly between 0 - 6AM. This timing suggests fraudsters exploit periods of low customer activity and limited bank oversight
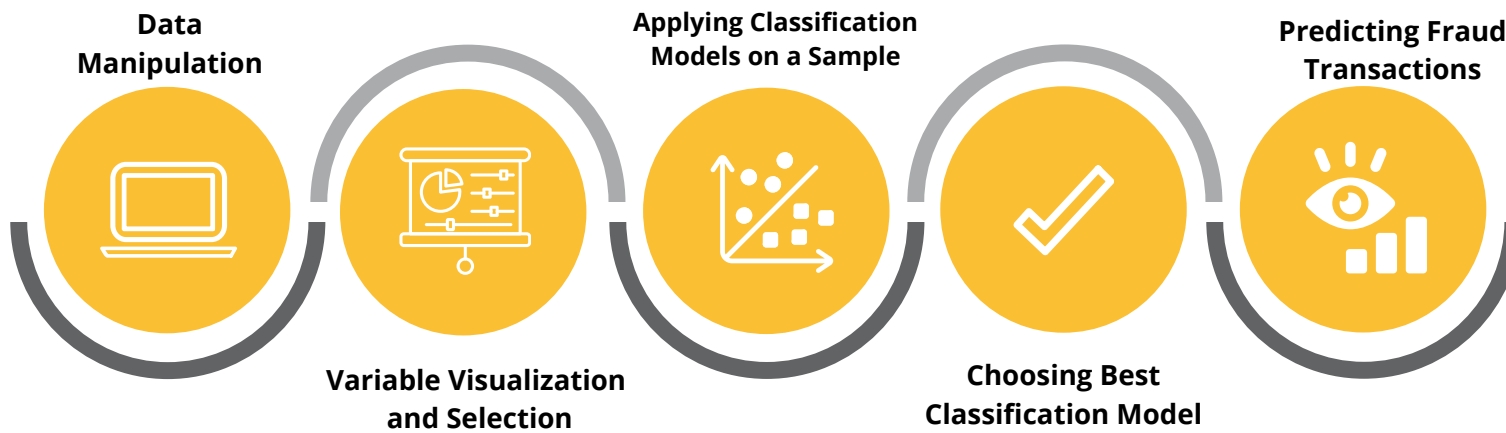
Patterns of Fraudulent Transaction

Age
Median
18   60   77   85   95

Transaction Amount
Small Transactions
75%
Under $220

Transaction Category
Fraudulent Transactions Type
Online Transaction **57.32%**
Offline Withdrawal **32.92%**
Others **13.33%**

Hour **0–6** (24 hours)
Day of Week **Wednesday**
Transaction Type : Withdrawal

Transaction Distribution by Time and Day of Week

63% Fraudulent Transactions at 0 - 6 hrs are on account of customer aged 70+

WHAT WE DID

Data Manipulation → **Variable Visualization and Selection** → **Applying Classification Models on a Sample** → **Choosing Best Classification Model** → **Predicting Fraud Transactions**
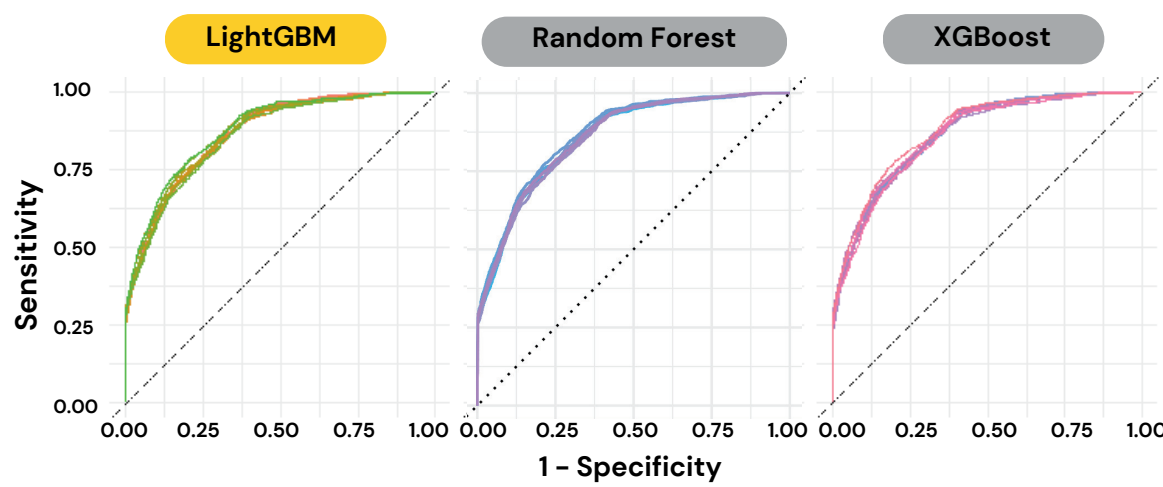
Assumptions
Independent Observations
Threshold Value 0.5
Scalability & Stationarity

Data Preparation
Bivariate Analysis
Data Transformation for Transaction Date
Data Split - 75% Train Data & 25% Test Data
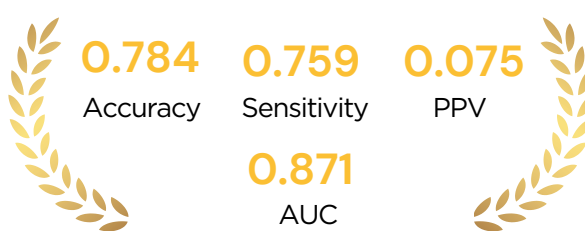
Model Comparisons - Top 3
ROC Curve by Model and Fold

LightGBM   Random Forest   XGBoost

LightGBM Performance on Test Data

Models	Accuracy	Sensitivity	PPV
LightGBM	0.785	0.758	0.076
XGBoost	0.774	0.761	0.073
Random Forest	0.817	0.709	0.083

LightGBM Performance on Test Data

LightGBM
0.784 Accuracy
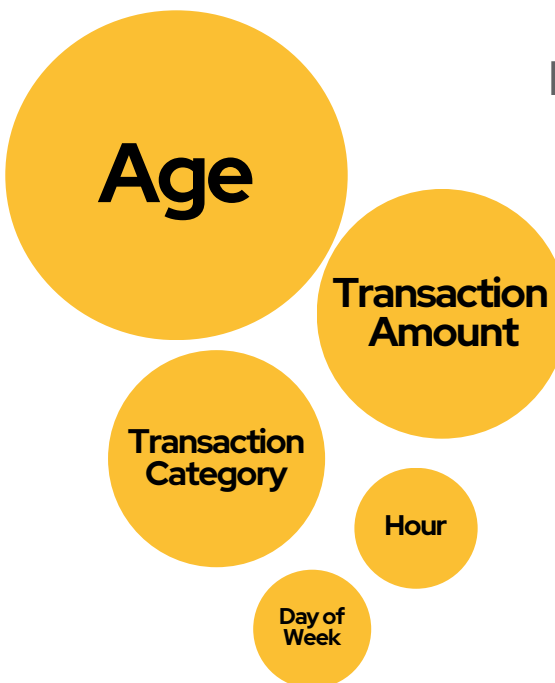0.759 Sensitivity
0.075 PPV
0.871 AUC

CONFUSION MATRIX

True Positive 2137 1.7%	False Positive 26274 20.4%
False Negative 680 0.5%	True Negative 95909 77.3%

*Target class : Fraud Transactions

WHAT CAN WE DO

Key Variables

Age
Transaction Amount
Transaction Category
Hour
Day of Week

Level of Importance

Key Finding Flag accounts for further investigation rather than immediate blocking transactions, as Positive Predictive Value (PPV) is low.

- Perform status (alive/not alive) to update bank records for all customers above the age of 77.
- Provide flexible option to set up OTP and limited banking services for 65 and above customers during hours of maximum fraudulent transactions.
- Localised geo-fencing based on transaction history through merchant location for elderly customers.
- Send regular fraud pattern updates to customers so they are always aware. Send summary statement especially on Wednesday for customers to flag any fraud transactions.

References :
- Sohony, I., Pratap, R., & Nambiar, U. (2018). Ensemble learning for credit card fraud detection. Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, 289–294. https://doi.org/10.1145/3152494.3156815
- Fayzrakhmanov, R., Kulikov, A., & Repp, P. (n.d.). The Difference between Precision-recall and ROC Curves for Evaluating the Performance of Credit Card Fraud Detection Models. http://dx.doi.org/10.25673/5577

Authors:
Dicky Samudra (dsam041); Jiawei Xu (jxu987), Md Rajib Hossain (mhos144), Taha Sameer (tsam755)
The University of Auckland - BUSINFO 704 Predictive Business Analytics – Group 26 – 2024 Quarter 3

*The data that used for this project is synthetically generated