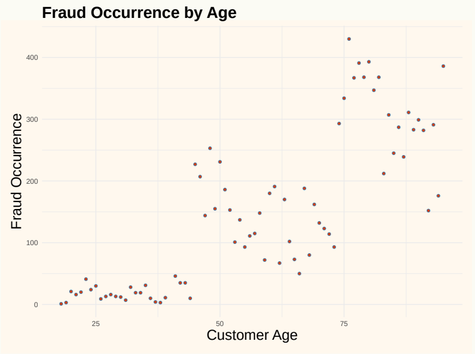
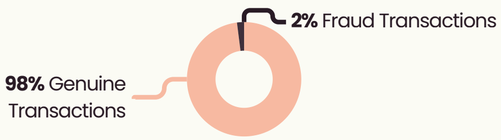
Guardian of the Vault: Cutting-Edge Model to Detect Fraud

Project Rationale:

Identify fraudulent transactions for the Bank by **enhancing** the bank's existing **fraud detection model** to **accurately** identify fraudulent transactions, thereby **minimizing financial losses** and **safeguarding customer confidence**.

Overview of Dataset:

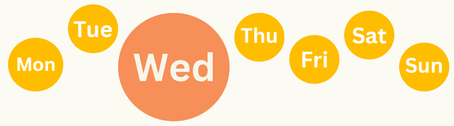




500,000 synthetically generated bank transactions, **highly imbalanced** with **98% genuine** and **2% fraudulent** transactions

72 **Average Age** of Customers with Fraud Transaction

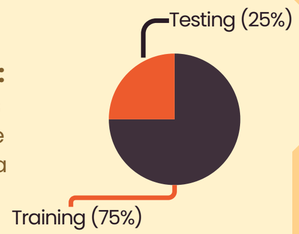
4% of total **10,000 customers** observed have been affected by fraud



1 in 5 fraud cases occured on Wednesday

Methodology:

1. **Data Transformation:**
   - Considered **missing agent** categories as **offline** transactions.
   - Transformed **Dates** into **day** of the **week**
2. **Data Partitioning:** **75%** of the data was used for **training** the model and **25%** data for **testing**



3. **Data Preprocessing:**
   - Features such as **Normalisation**, **Zero Variance Filtering**, **Up-sampling imbalanced** class
4. **Model Training:**
   Training various machine learning models of classification
   - Trained data on **LightGBM**, **XGBoost**, **Logistic Regression, Random Forest & KNN**
5. **Model Evaluation:**
   - **XGBoost** model performs best in correctly identifying **fraud transactions** based on **Test Data**

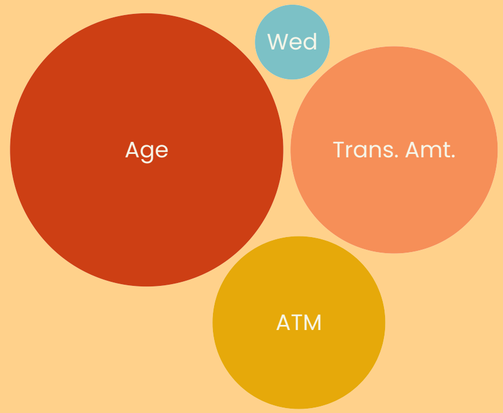
Assumptions:

- ***Stationarity:*** *Statistical properties of the data are assumed consistent over time*
- ***Feature Relevance:*** *Only the most significant features are selected to improve model accuracy*
- ***0.5 Cut-Off Value:*** *Threshold set as the classification boundary for detecting fraud*
- ***Scalability:*** *Large datasets can be handled efficiently without compromising performance*

Results:

~ Scores are calculated based on Test Data

Prominent Factors influencing Fraudulent Transactions



Fraud Transactions tends to **increase** with both **Age** and **Transaction Amount**. Additionally, fraud incidents are **more likely** to **occur** on **Wednesdays**. **For Seniors**, using **ATMs** also presents a higher risk of fraud.

Higher Risk associated → **Senior individuals, ATM Withdrawals, Median Amt. & Wednesday**

Lower Risk associated → **Balance, other days, other platforms**

Performance Indicators

Sensitivity - 78%

Model is **efficient** in **Predicting Fraudulent** Transactions with a sensitivity score of **78%**

ROC AUC - 87%

Model achieved an **AUC of 87%,** demonstrating **strong** overall **accuracy** in **differentiating** between **fraudulent** and **genuine** transactions

Specificity - 77%

It **correctly** identified **77%** of **genuine** transactions, **reducing incorrect flags** and boosting **monitoring** efficiency

Recommendations:

VAULT

Vigilant

Tiered Response system will classify transactions into **lower & higher risk** & **trigger actions** accordingly

Lower Risk: Push Notifications

Higher Risk: Call Customer

Adaptive

Delay Option for High Risk Transaction

- Offer customers the **option to enable a short delay** (e.g. 30 mins) for high risk transaction, during which they can **confirm** the **transaction**.

User-Linked

Gamification - customers could **earn points**, **discounts** or **rewards** for using **secure banking practices**

Trusted Security

Ensuring all these measures work together to create a robust, **reliable**, and **secure environment** for customers