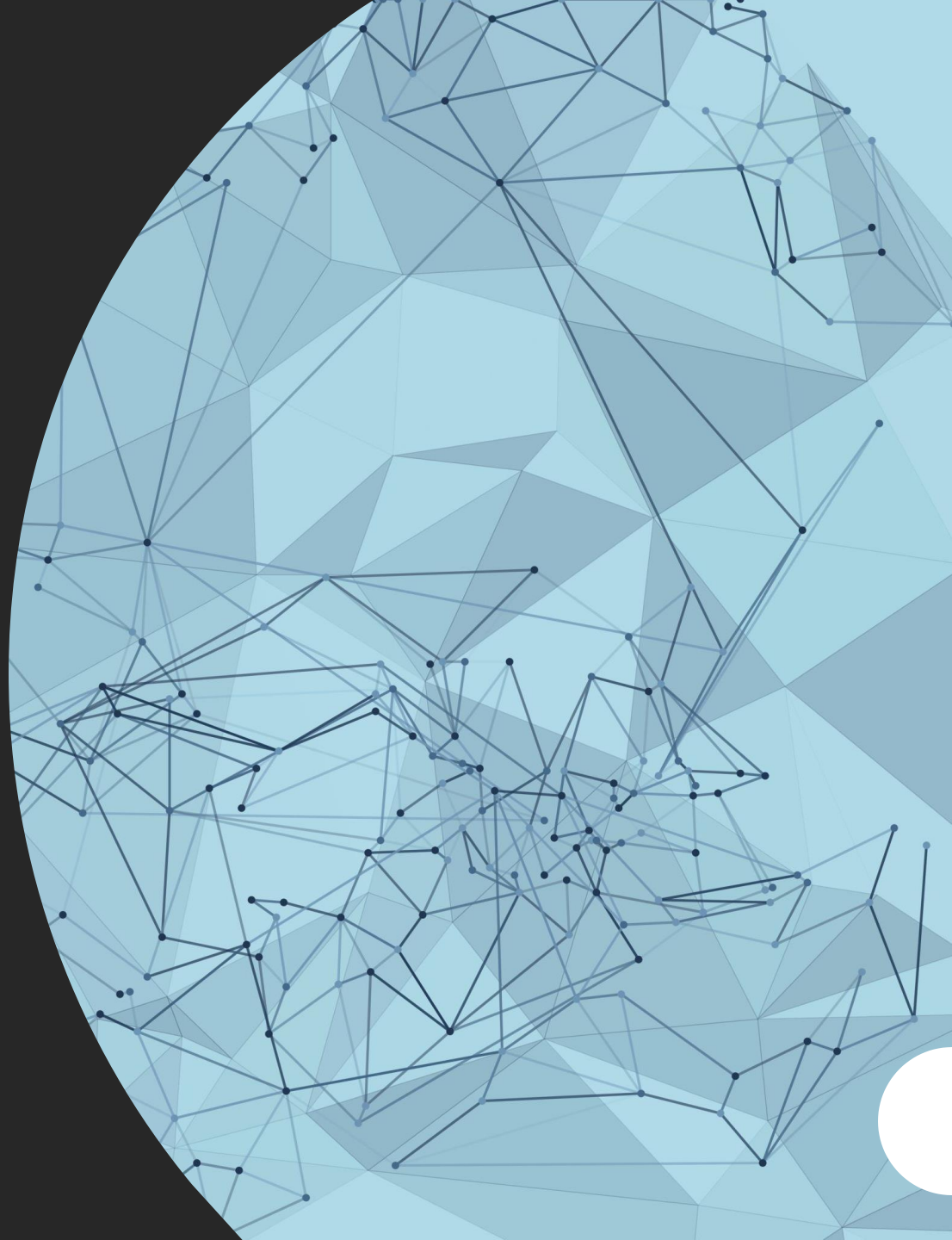
Explorative Data Analysis on Online fraud Detection (Blossom Bank).



Introduction

- **Fraud** is intentional deception to secure unfair or unlawful gain, or to deprive a victim of a legal right. Fraud can violate civil law , criminal law. It may cause no loss of money, property, or legal right but still be an element of another civil or criminal wrong.

According to the FBI's 2017 Internet Crime Report, the Internet Crime Complaint Center (IC3) received about 300,000 complaints. Victims lost over $1.4 billion in online fraud in 2017. According to a study conducted by the Center for Strategic and International Studies (CSIS) and McAfee, cybercrime costs the global economy as much as $600 billion, which translates into 0.8% of total global GDP. Online fraud appears in many forms. It ranges from email spam to online scams. Internet fraud can occur even if partly based on the use of Internet services and is mostly or completely based on the use of the Internet.

- ***Blossom Bank, commonly known as BB PLC, is an international financial services company with its headquarters in London, UK. It provides retail and investment banking, pension management, asset management, and payments services.***

- ***PROBLEM STATEMENT***
- ***Blossom Bank is attempting to forecast online payment fraud by creating a machine learning model.***

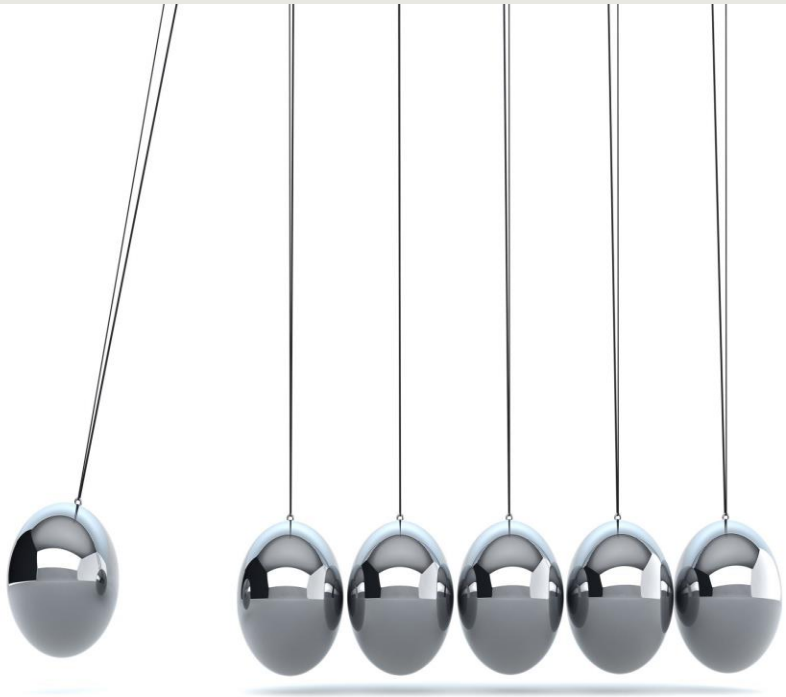
• Dictionary of Data

- Step: Represents a unit of time where 1 step equals 1 Hour;
- Type: Type of online transaction;
- Amount: The amount of the transaction;
- The nameOrig: The customer initiating the transaction;
- The oldbalanceOrg: The balance prior to the transaction;
- The newbalanceOrig: The balance following the transaction;
- NameDest: Refers to the recipient of the transaction;
- OldbalanceDest: The recipient's initial balance before the transaction;
- NewbalanceDest: The recipient's new balance after the transaction;
- IsFraud: Fraudulent transaction;

Insights.

- The data provided is from a financial institution which comprises of both normal transaction and fraudulent transactions.
- From the Dataset we can deduce that the most occurring transaction is "CASH_OUT" with 36%, closely followed by "PAYMENT" with 34% while "DEBIT" is the least occurring transaction.
- The dataset contains 1,048,575 records with 10 fields/columns.
- There are less Fraudulent transactions when compared to fraudulent transactions, this observations shows that large percentage of the trasactions are not frauds.
- It was also observed that frauds majorly occurs in cash_out and transfer.
- It was also observed that cash_out and transfer has highest amount which could be responsible for why frauds occurs in these transaction mode.
- The investigation shows that the common thing among the fraudulent transactions is that the initial balance of recipient before transaction, after transaction and initial balance of recipient before transaction are all 0 and the time taken is mostly high.
- We also noticed from the dataset that the data is imbalanced, therefore we need precison and recall in addition with accuracy to select the best model.

Recommendations:



- On transfer, there should be a limit to what one can transfer on a daily basis and also the number of transfers one can make daily, in order to reduce the fraud rate.
- At no point should the balance of the account be equal to zero i.e. There should be a stipulated minimum account balance at all times.
- On Transfer transactions, there should be more security, i.e. 2-way factor security or a one time OTP.
- Improve controls by implementing continuous auditing and monitoring.
- Provide management with immediate notification when things are going wrong.

Refrences:

- Fraud - Wikipedia
- 2017 Internet Crime Report Released – FBI