Name:Ammu Ambika Ramachandran

Student ID: 23006390

Github : <https://github.com/Ammu-AR/Applied-Data-Science>

**CREDIT CARD FRAUD DETECTION**

I downloaded the Credit Card Fraud Detection dataset from Kaggle<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>). This dataset includes details about credit card transactions, such as the transaction amount, time, and whether the transaction was fraudulent or not. I performed an analysis of the data, made several assumptions, and visualized my findings using histograms, line plots, area plots, and a confusion matrix to explore the correlation between different variables.

**Dataset Overview:**

**Total Transactions:** 284,807 rows (transactions).

**Total Features:** 31 columns (including one target column).( 30 columnsare anonymized features like merchant card no etc and 1 column is the target variable, which shows whether the transaction is fraudulent or no

A collage of graphs

Description automatically generated

**Analysis 1 : Histogram Transaction Amount for Fradulent Transaction**

**Observations:**

*Most fraudulent transactions are relatively small, with 75% under 105.89. This pattern suggests that fraudsters may opt for smaller amounts to avoid detection. However, there is a broad range in amounts, from $0 up to 2,125.87, showing that while larger fraud attempts are less frequent, they do occur.*

**Analysis 2 :Line Plot - Frequency Of Fraudulent Transactions By Hour Of The Day**

**Observation:**

*Fraudulent activity peaks in the early morning hours, suggesting fraudsters may take advantage of lower oversight during this time. Midnight has the least activity, indicating a potential preference for times when fewer transactions occur.*

**Analysis3: Area Plot - Number Of Fraudulent Transactions By Amount Range**

**Observation:**

*This area plot shows that number of fraudulent transactions is higher for small amount and while the transaction amount is getting higher the number of fraudulent transactions is lowering. Fraudsters are preferring for small amount so that it is less noticeable easily. I chose this area plot because it is easy to know about the amount range of higher fraud transactions.*

**Analysis 4: A Confusion Matrix**

**Observation:**

*The model is very good at predicting class 0. It has a high number of true positives (32041) and very few false negatives (1). The model struggles with predicting class 1. It has a low number of true positives (56) and a higher number of false negatives (16).To improve the model's performance, especially for class 1, we could consideradjusting the threshold*

**Conclusion:**

The analysis reveals that fraudsters tend to target smaller amounts and specific times of day to minimize detection. The model’s performance could be improved by adjusting thresholds to better predict fraudulent transactions.