**Project Proposal**

• The dataset chosen for analysis is the Credit Card Fraud dataset, sourced from Kaggle. Also as already discussed this data is synthetic data so the personally identifiable information is not real data. Here is the link to the dataset: [Credit Card Fraud Dataset] <https://www.kaggle.com/datasets/kelvinkelue/credit-card-fraudprediction?select=fraud+test.csv>

• The business value derived from our findings and predictions could be substantial for financial institutions and credit card companies. By accurately predicting instances of fraud, these entities can mitigate financial losses, enhance security measures, and improve customer trust. The modeling work aims to answer key business questions such as:

- What are the patterns and characteristics associated with fraudulent credit card transactions?

- How can we identify potentially fraudulent transactions in real-time?

- What measures can be implemented to prevent fraudulent activities effectively?

• The objective is to predict instances of credit card fraud. Specifically, the target variable will be the "Is\_fraud" attribute, which serves as a binary classification indicating whether a transaction is fraudulent (1) or legitimate (0).

• The dataset comprises 555,719 rows and 22 columns. Each row represents a single credit card transaction, while the columns contain various features such as

Trans\_date\_trans\_time: Timestamp of the transaction (Date and Time).

Cc\_num: Unique customer identification number (Int).

Merchant: The merchant involved in the transaction (String).

Category: Transaction type (String).

Amt: Transaction amount (In Dollars).

First: Cardholder's first name (String).

Last: Cardholder's last name (String).

Gender: Cardholder's gender (Categorical M/F).

Street: Cardholder's street address (String).

City: Cardholder's city of residence (String).

State: Cardholder's state of residence (String).

Zip: Cardholder's zip code (Int).

Lat: Latitude of cardholder's location (Float).

Long: Longitude of cardholder's location (Float).

City\_pop: Population of the cardholder's city (Int).

Job: Cardholder's job title (String).

Dob: Cardholder's date of birth (Date).

Trans\_num: Unique transaction identifier (String).

Unix\_time: Transaction timestamp (Int).

Merch\_lat: Merchant's location (latitude)(Float).

Merch\_long: Merchant's location (longitude)(Float).

Is\_fraud: Fraudulent transaction indicator (1 = fraud, 0 = legitimate). This is the target variable for classification purposes.