**02/20/2023 Week 1 Biweekly Report**

**Data Collection**

* In order to solve the problem, there is a need to get transactional data.Uniquely identifiable transaction id, transaction amount, customer name, and credit card number used for the transaction.These columns would provide enough information to identify if a customer has used a credit card for a fraudulent purpose. After reading through several data sets, I identified a data set that would align with the problem I am trying to solve.
* The data set I procured from Kaggle has the following information.

<https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv>

**Columns information**

* There are about 1.3Million rows and 46 columns. I am planning on using this data set to tackle the problem. The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be a fraud. This model is then used to identify whether a new transaction is fraudulent or not.
* The columns are
  + - index - Unique Identifier for each row
    - trans\_date\_trans\_time - Transaction DateTime
    - cc\_num - Credit Card Number of Customer
    - merchant - Merchant Name
    - category - Category of Merchant
    - amt - Amount of Transaction
    - first - First Name of Credit Card Holder
    - last - Last Name of Credit Card Holder
    - gender - Gender of Credit Card Holder
    - street - Street Address of Credit Card Holder
    - city - City of Credit Card Holder
    - state - State of Credit Card Holder
    - zip - Zip of Credit Card Holder
    - lat - Latitude Location of Credit Card Holder
    - long - Longitude Location of Credit Card Holder
    - city\_pop - Credit Card Holder's City Population
    - job - Job of Credit Card Holder
    - dob - Date of Birth of Credit Card Holder
    - trans\_num - Transaction Number
    - unix\_time - UNIX Time of transaction
    - merch\_lat - Latitude Location of Merchant
    - merch\_long - Longitude Location of Merchant
    - is\_fraud - Fraud Flag <--- Target Class
* Installing the Required softwares to my system.
  + - · Python – 3.x
    - · NumPy – 1.19.2
    - · Pandas
    - · Seaborn
    - · Scikit-learn – 0.24.1
    - · XG Boost
    - · Matplotlib – 3.3.4
* Preparing the base work for the project.

**Next Steps**

* In the next step I am planning on deep diving into the data set to see how dirty the data is and what steps I can take to clean up the data and identify any outliers.