CREDIT CARD FRAUD DETECTION

Mohammed AlQasmi

ABOUT PROJECT

Credit card fraud is a term for fraud committed using a payment card, credit card fraud can be authorized, where the genuine customer themselves processes a payment to another account which is controlled by a criminal, or unauthorized, where the account holder does not provide authorization for the payment to proceed, and the transaction is carried out by a third party. In 2018, banks and card companies in the United Kingdom prevented £1.66 billion in unauthorized fraud in 2018. That is the equivalent to £2 in every £3 of attempted fraud being stopped.

It is important that banks and credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

ABOUT DATA

SOURCE

The dataset has been collected and analysed during a research collaboration of Worldline and the Machine Learning Group (http://mlg.ulb.ac.be) of ULB (Université Libre de Bruxelles) on big data mining and fraud detection.
The data itself was acquired from Kaggle. (https://www.kaggle.com/mlg-ulb/creditcardfraud)

CONTENTS

The dataset contains 284,807 records with 31 features.

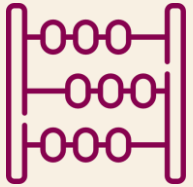
DATASET

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we couldn't get the original features and more background information about the data. The only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset, and 'Amount' is the transaction Amount. Feature 'Class' is the response variable, and it takes value 1 in case of fraud and 0 otherwise.

HOW MANY FRAUDULENT TRANSACTION IN THE DATASET?

LABEL DATA

Labelled Data into Normal Transactions and Fraud Transactions.



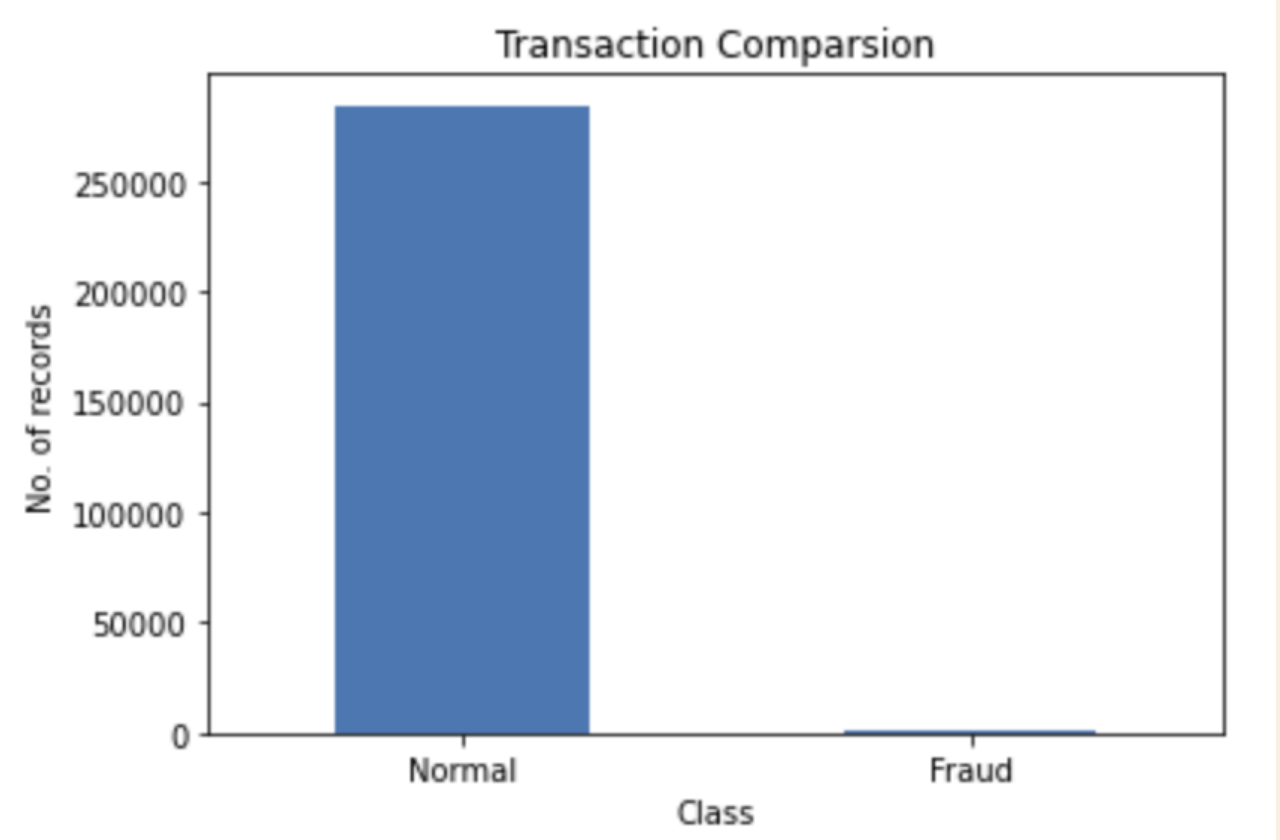
COUNT FRAUD TRANSACTION

Normal Transactions: 284315
Fraud Transactions: 492

COMPARE TRANSACTION

To view difference in records.

TRANSACTION COMPARISON



WHAT'S THE AMOUNT OF FRAUD TRANSACTIONS COMPARED TO NORMAL TRANSACTIONS?

FRAUD TRANSACTIONS INFO

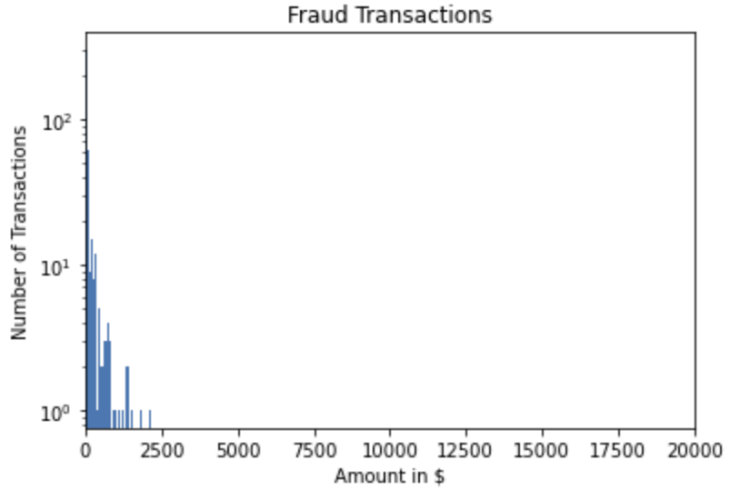
- **No. of Records:** 492
- **Transaction Average "Mean":** 122.21$
- **Maximum Amount:** 2,125.87$

NORMAL TRANSACTIONS INFO

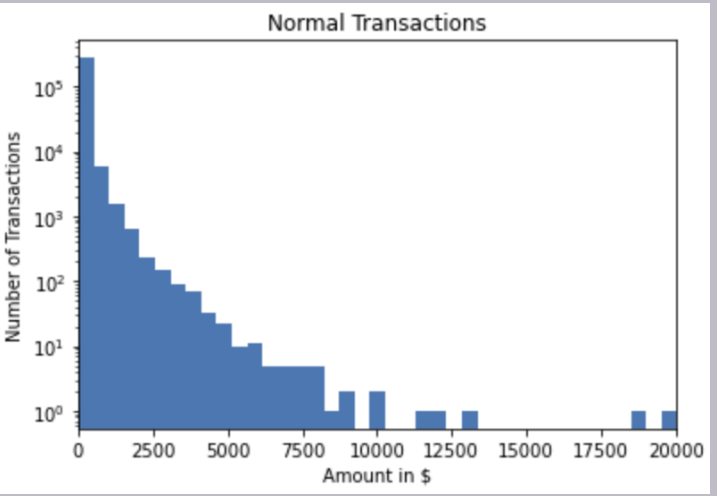
- **No. of Records:** 284,315
- **Transaction Average "Mean":** 88.29$
- **Maximum Amount:** 25,691.16$

WHAT'S THE AMOUNT OF FRAUD TRANSACTIONS COMPARED TO NORMAL TRANSACTIONS?

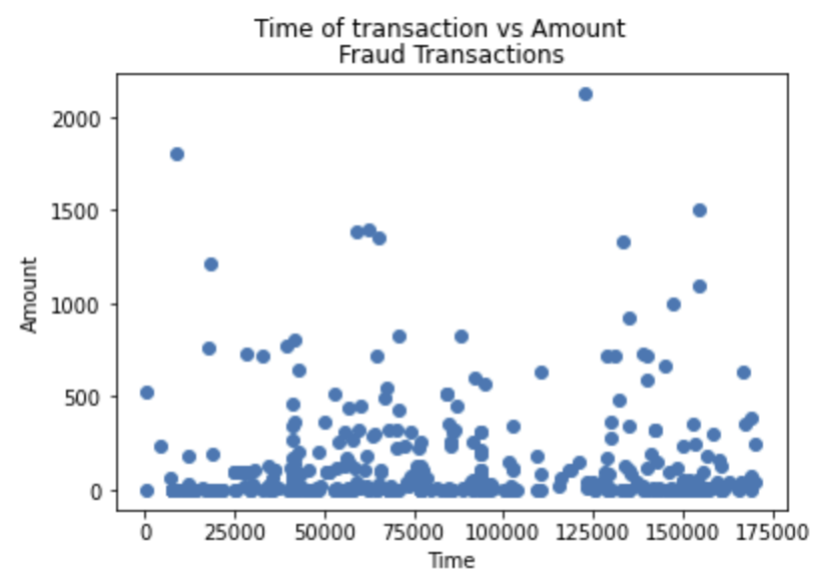
FRAUD TRANSACTIONS INFO

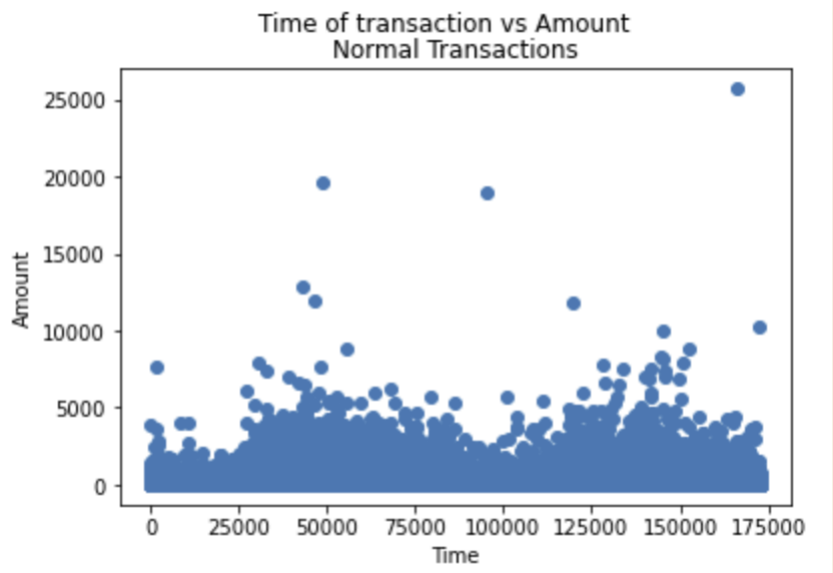


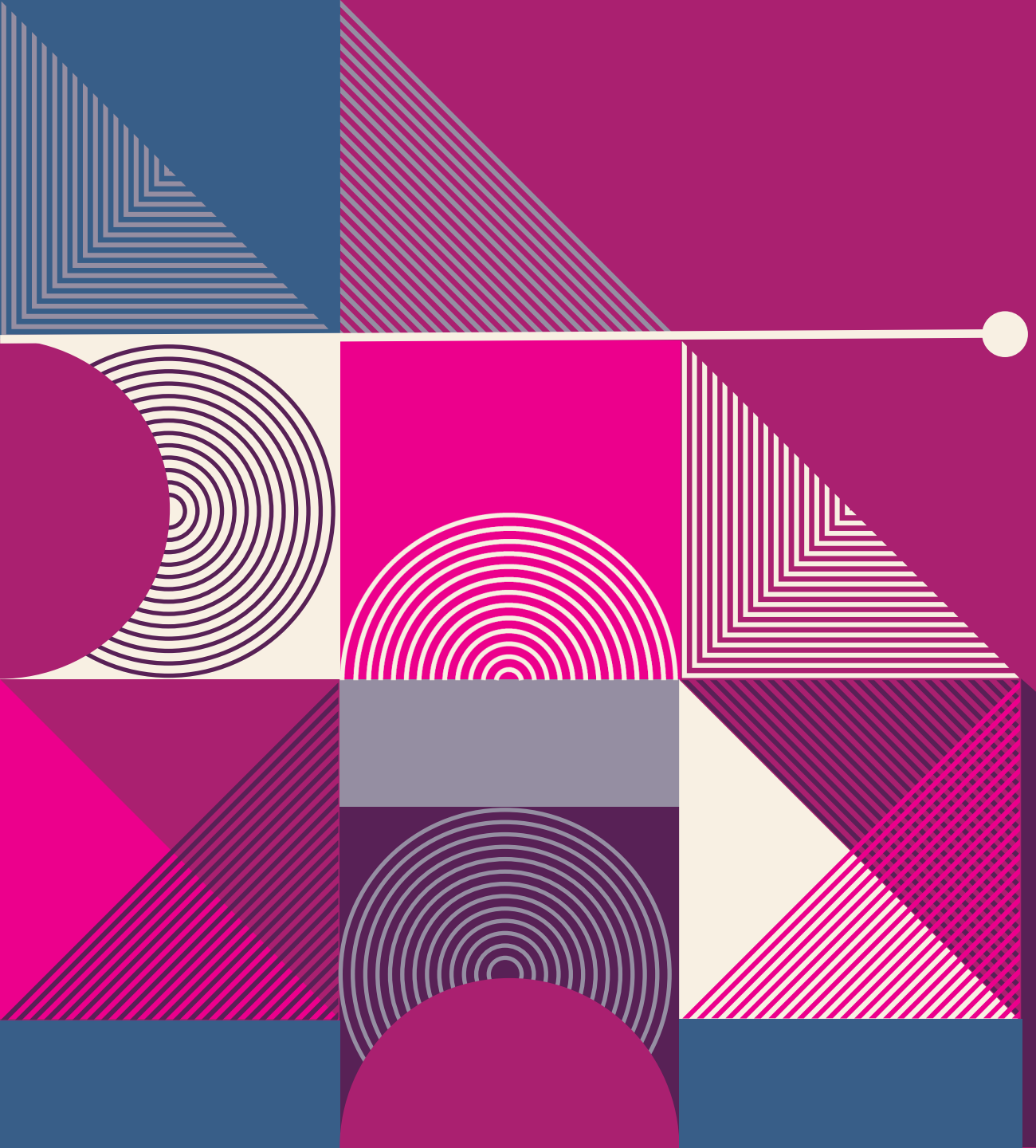
NORMAL TRANSACTIONS INFO



DOES FRAUD TRANSACTIONS HAPPEN MORE OFTEN DURING CERTAIN TIMES?







PREDICTING FRAUD

DATA

BALANCING DATA

Since data is imbalanced due to the fact that we have a lot more valid transactions compared to fraud we have to balance the data. We balance data by splitting fraud transactions then taking a random sample of 492 normal transactions then merge it with the fraud transactions.

FINAL SAMPLE

Final sample consists of a balanced 984 record and 31 features dataset.

PREPARING DATA FOR MODELS

80%	20%
Training Data	Testing Data
787 Record	197 Record

MODELS ACCURACY

LOGISTIC REGRESSION

Accuracy on Training data:

0.940279542566709

Accuracy score on Test Data:

0.9187817258883249

K-MEANS

Accuracy on Training data:

0.4515569795255547

Accuracy score on Test Data:

0.45296864576384255

SUMMARY

After evaluating both models, We see the Logistic regression is better than K-Means in predicting credit card fraud.

THANK YOU