Insights from Credit Card Fraud Analysis

By,
Madhumitha.R

AGENDA

- Objective
- Background
- Key Findings
- Recommendations
- Appendix:
  - Data sources
  - Data methodology
  - Data model assumptions

OBJECTIVE

- Build a Credit card Fraud Detection system for proactively monitoring and prevention of frauds.
- Analyze the Business Impact of fraudulent transactions.
- Recommendation of optimal ways for the bank to mitigate the fraud risks.

BACKGROUND

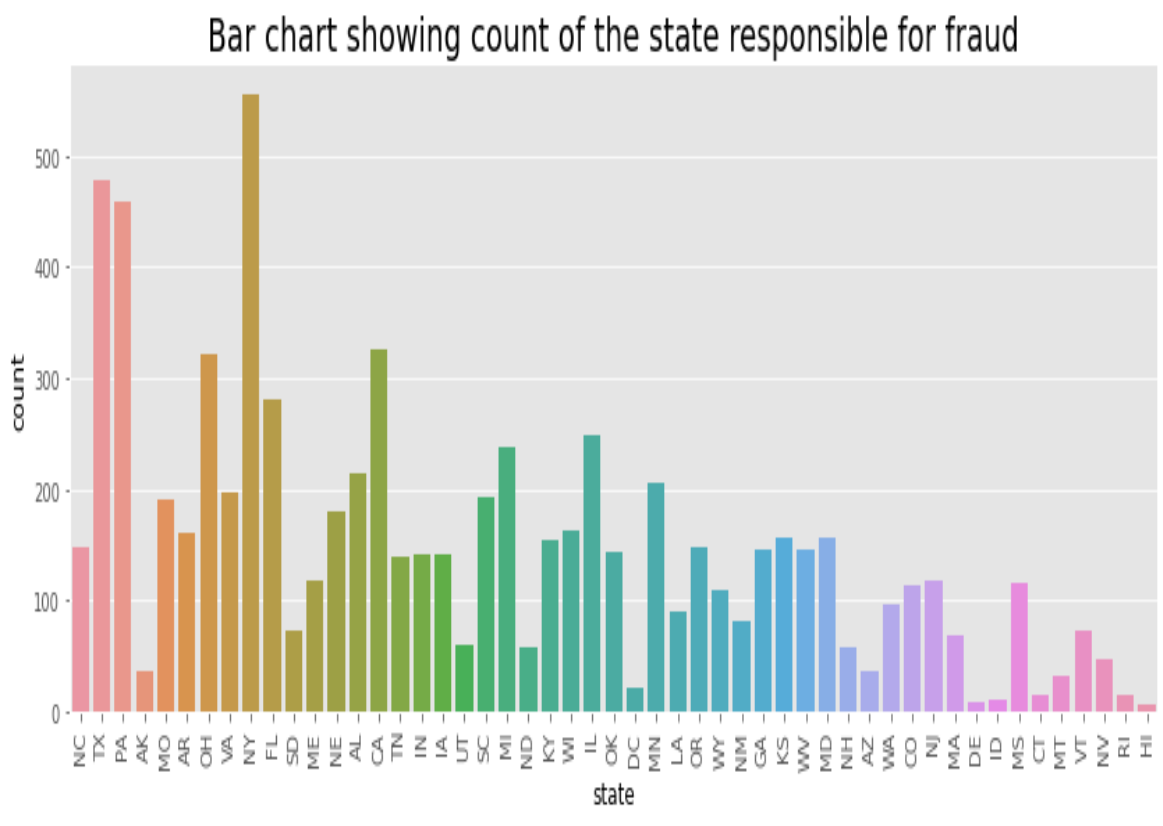
- Finex , a leading financial service provider based out of Florida, US has observed that large number of unauthorized transactions have been made.
- The bank has been facing a huge revenue and profitability crisis.
- It has become difficult for the bank to track these data breaches on time to prevent further losses.

Top 10 important features for fraud detection

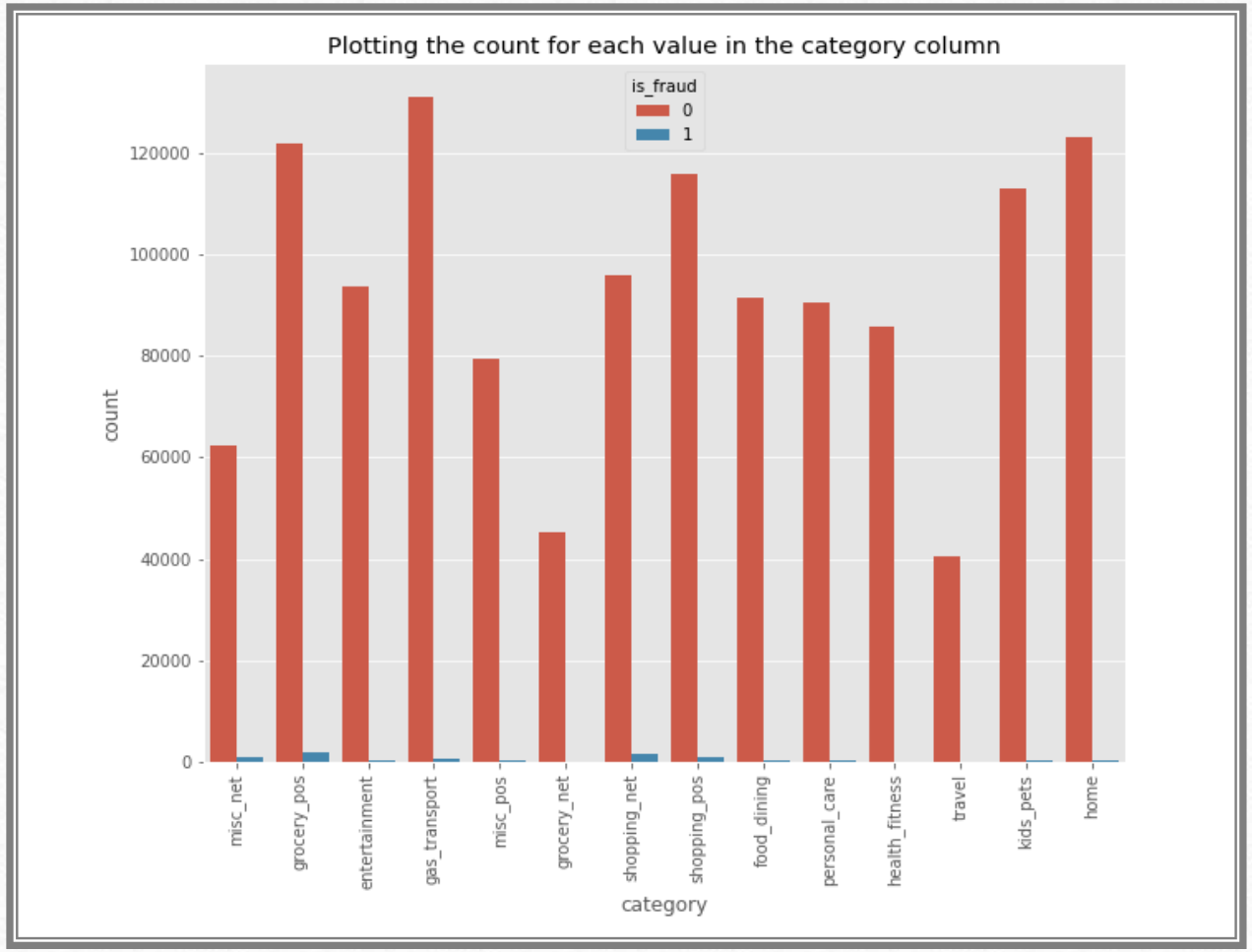
- The mean amount in fraudulent transaction is higher than normal transaction.
- The Fraud transactions tend to happen more at odd times(0-5,20-24 hrs.) Mostly fraud transactions happens more during night compared to daytime.
- Category, age, job, city, population of the city affect more chance of fraud.

Varname	Imp
amt	0.566028
trans_time	0.198514
category_gas_transport	0.025040
age	0.024016
category_grocery_pos	0.021293
category_shopping_net	0.020340
city_pop	0.015866
job	0.011650
city	0.011608
distance_people_to_merchant_co-ord(kms)	0.011252
category_misc_net	0.009632
trans_month	0.009617
state	0.009375

States with more Fraud transactions



- New York has the maximum number of fraud reported.
- Texas, Pennyslvania ,ohio,California has the next highest fraud cases reported.



Category where the frauds happen

- Generally the frauds are reported from credited card used at gas_transport ,grocery and shopping point of sales.
- Apart from these even shopping online, online miscellaneous. entertainment also have reported fraud transaction
- Surprisingly travel has very low fraud transaction reported.

RECOMMENDATIONS

- Build a machine learning model (using Random Forest)to track the fraud cases.
- Around 89% of all fraudulent transactions cases are correctly predicted using the fraud detection system .
- We make a cost saving of 88.7% after deploying the model for fraud analysis per month.

APPENDIX - DATA SOURCES

- Here is a snapshot of our data dictionary.
  - trans_date_trans_time
  - cc_num
  - merchant
  - category
  - amt
  - first
  - last
  - gender
  - street
  - city
  - state
  - zip
  - lat
  - long
  - city_pop
  - job
  - dob
  - trans_num
  - unix_time
  - merch_lat
  - merch_long
  - is_fraud
- The team used the following data sources:
  - Data provided by Upgrad Team from kaggle

APPENDIX - DATA METHODOLOGY

❖ We conducted a thorough analysis of the Credit Card Transactions Fraud Detection Dataset dataset. The process included

- Exploratory Data Analysis
- Train/Test Data Splitting
- Model Building or Hyperparameter Tuning (Random Forest ,Oversampling method with randomized Searchcv)
- Model Evaluation and Cost Based Analysis

Attached Files

- Cost Benefit Analysis:

  Cost+Benefit+Analysis test set.xlsx

- Methodology Document:

  Methodology Document_1.docx

- Random Forest Classifier Model:

  Capstone_project.ipynb

APPENDIX - DATA ASSUMPTIONS

- The data is assumed to be simulated credit card transaction dataset containing legitimate and fraud transactions from the duration 1st Jan 2019 - 31st Dec 2020.
- Assume that the Business stake holders want to find out the maximum number of fraud cases.
- The cost based analysis is done on the testing dataset.

THANK YOU