Credit Card Fraud Detection for Banks and Retailers

Metis Business Project

By: Marjan Rezvani

Abstract

The goal of this project was to do data analysis to help providing proper ways to determine whether a given credit card transaction will be fraudulent or not. one of the most important responsibilities that a bank or financial institution has is to protect the integrity of the institution by working hard to protect the financial assets that it holds. Bank fraud can be defined as an unethical and/or criminal act by an individual or organization to illegally attempt to possess or receive money from a bank or financial institution. In this project, I have done different data analysis and visualization using card transaction data to identify high risk of being fraudulent transactions. I built an interactive dashboard to visualize and communicate my results using Tableau.

Design

This project organized to help providing proper ways to detect credit card fraudulent transactions. The data is provided by Capital one bank in 2016. I worked on main dataset to collect information, analyzes the data gathered and extracts the required features.
My goal is to provide data and analyze it to help us make models to separate fraud transactions from non-fraud transactions. These processes would help us:

- How to recognize fraud transactions.
- How to choose a solution for credit card fraud detection.
- Help to manage and track high-risk credit card fraud alerts and subsequent processing.

Data

For this project, first I used the whole dataset contains a year credit card transaction made by Capital One in 2016, including 641914 transactions and 29 features. Then I sampled data to make it smaller to be able to be imported in Excel and Tableau, which includes 1300 instances and 29 columns. Some of numerical attributes are like available money, credit limit and categorical attributes like merchant name and transaction type.

Algorithms

- Feature Engineering

Out of the 641914 transactions in the dataset, 10892 were fraudulent which means the True frauds account for 1.7% of all transactions.

- Sampling data to make it possible to be imported in Excel and Tableau.
- Cleaning data in Excel

There are also features that describe customers' behavior are added, besides being fraud and not fraud. Some of numerical attributes are like available money, credit limit and categorical attributes like merchant name and transaction type. I also have few attributes such as echoBuffer, merchantCity, merchantState, merchantZip, posOnPremises, recurringAuthInd, which totally have missing values that I dropped these columns.

- Identify different functions and conditional functions in Excel and Google sheets.
- Sorting and Filtering data in excel
- Making Pivot tables in sheets
- Making different charts

- Handle date and datetime features

  There are some datetime features we have in our dataset such as:

  - accountOpenDate
  - transactionDateTime
  - currentExpDate
  - dateOfLastAddressChange

- Feature Selection

Feature selection is important step in project. By Feature selection we can find important features which are most effective in explaining the relationship between variables and results.

- Integrate a Fraud Prevention Solution:

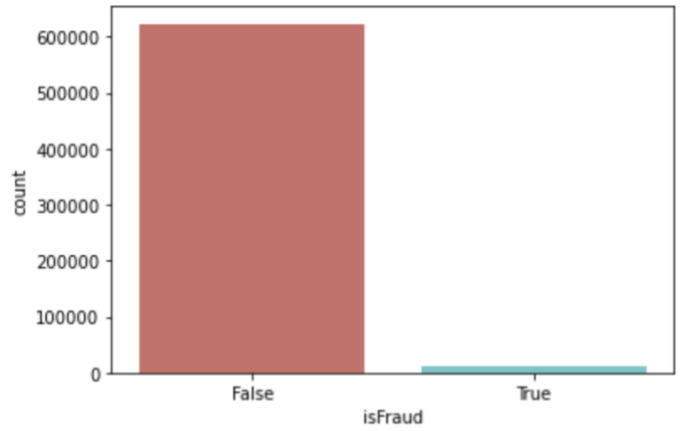
- Gather the user's card and transaction data.
- Data is compared with previous transactions.
- Then block risky users whose transactions have been fraudulent in the past.

- Visualization in Tableau

Tools

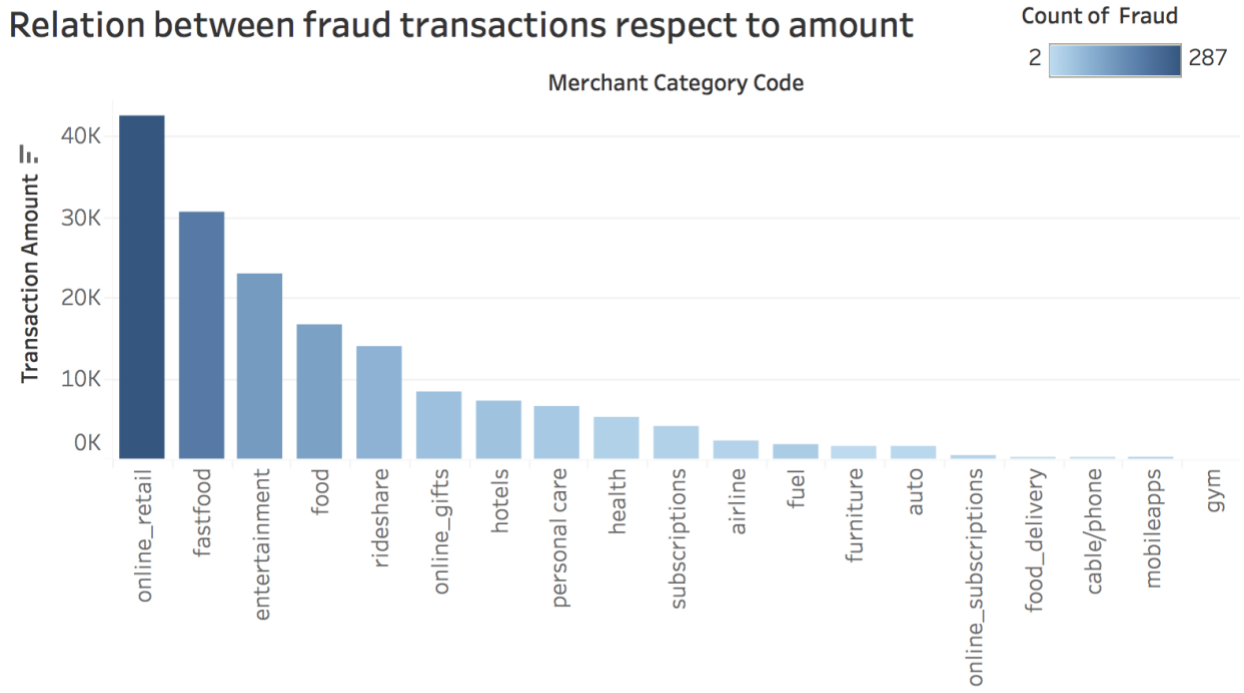
- Data manipulation and sampling: NumPy and Pandas
- Data cleaning and analysis: Excel, Google Sheets
- Visualizations: Excel, Google Sheets, Tableau

Communication

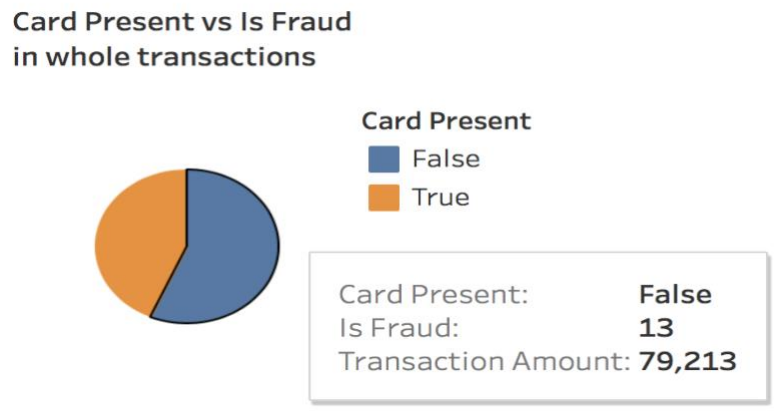



```
False    622954
True      10892
```

This graph shows that the number of fraudulent transactions is much lower than the legitimate ones. (around 1.7%)

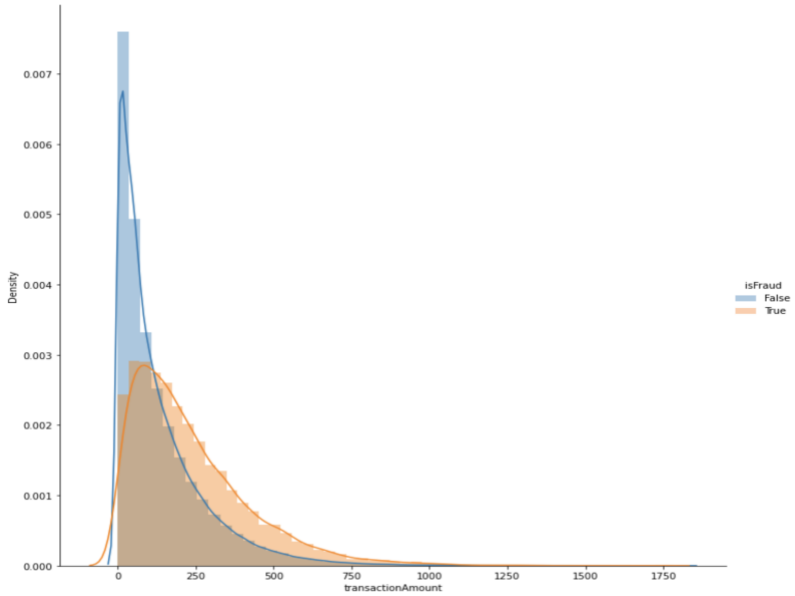


Sum of Transaction Amount for each Merchant Category Code. Color shows count of Is Fraud.

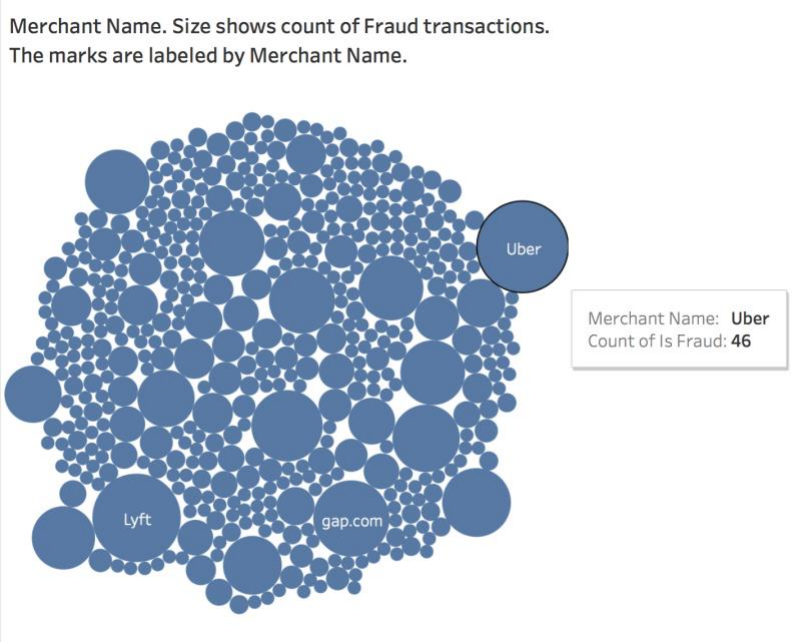


Card Present (color) and sum of Transaction Amount and being Fraud(size).

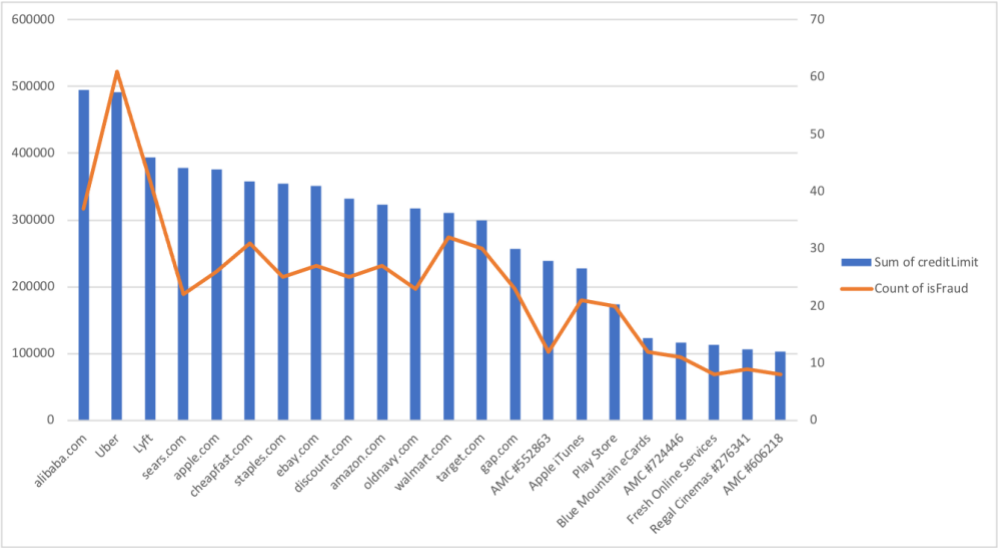
The credit card fraud that is most prominent for online merchants is known as card-not-present fraud or CNP fraud.
CNP fraud is possible when physical card is not required to complete an online transaction.



Most of the transaction have amount less than 1000 approximately and all of the fraud transactions have transaction amount less than 1750.

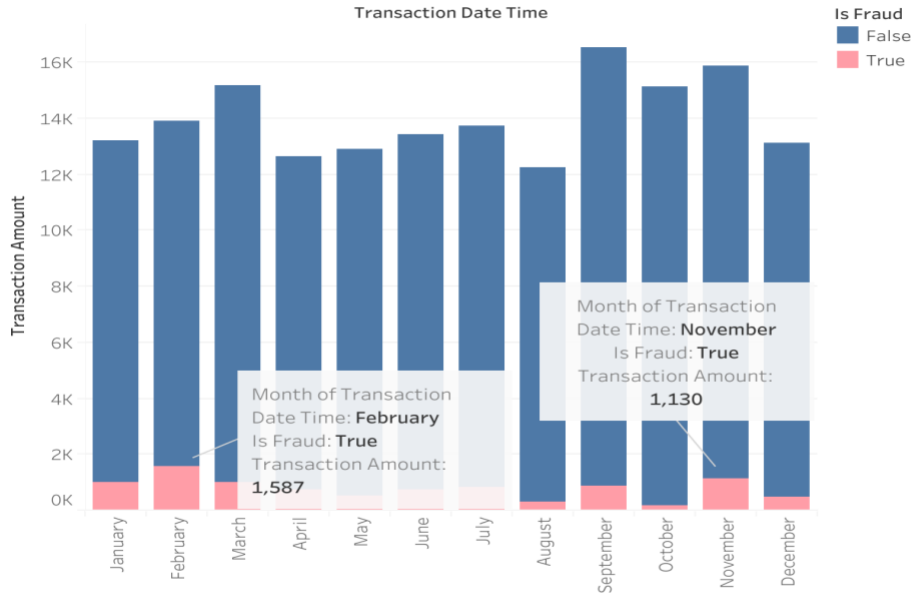


Respectively, most fraud transactions have happened in some companies where can be paid online, such as Uber, Lyft and gap.



It shows Credit limit of Merchants vs Fraud transactions

Fraud and Non_Fraud Transactions per each Month of 2016



Sum of Transaction Amount for each Transaction Date Time Month. Color shows details about Is Fraud. The view is filtered on Is Fraud, Transaction Date Time Month and sum of Transaction Amount.