



# Recommendation System Using Credit Card Transactions

## Capstone 2

Fariha Baloch  
Springboard - Data Science



# Problem Statement

Create a recommendation system using credit card transactions records. The data contains transactions from multiple people and contains GPS records. The aim is to provide the customer following

- a. The recommendation system should recommend similar merchants to the customer based on similarities between the merchants
- b. The recommendation system should recommend similar merchants to the customer based on similarities between users. The recommendation system should consider current customer's location as well in order to recommend the nearest similar merchants.



# Data Mining

- The data for this project came from [FinGoal](#).
- FinGoal analyzes consumer spending to give radically insightful recommendations built on the experience of other savvy consumers (source: FinGoal.com).
- The data was anonymized and converted to json format by the company. The file size was 48.2MB.
- Used `pd.read_json` to read the file into a data frame. The file had 29 columns and 90086 rows of transactions.



# Data Wrangling

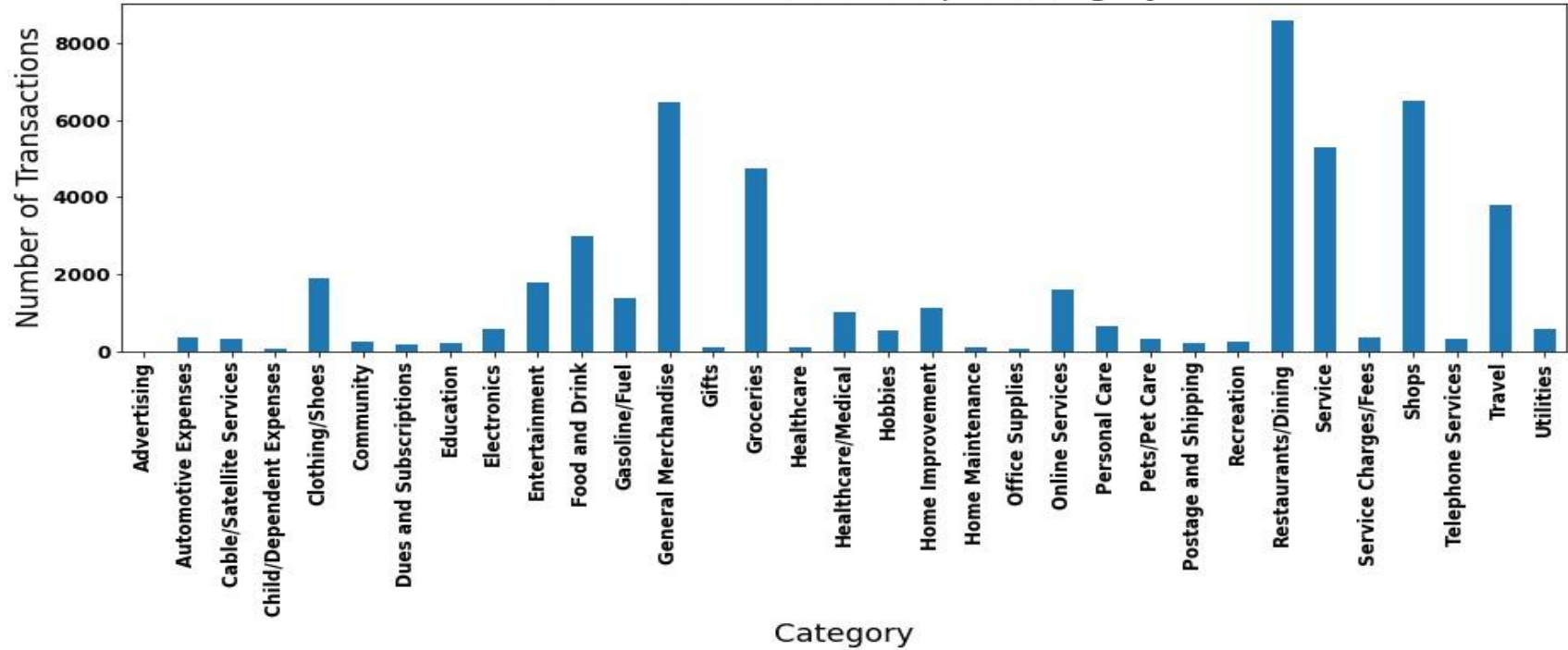
- Dropped columns with no values and columns that were not important to the recommendation system
- Used columns like zip\_code and city to fill NaNs of each other
- Multiple uid's were missing accountid. Therefore assigned random unique numbers to their accountid in order to get rid of the NaNs in accountid columns
- Category column had values that were related to expenditures by a customer, for example, PayChecks/Salary. All such values were deleted from the data by deleting the rows with those specific column values
- In the ammountnum column 75% of dollar amount spent by consumers was less than \$52. A few of the consumer purchases were > \$6000 and were considered outliers therefore those rows (44) were deleted



# Exploratory Analysis

- To see how the transaction were divided into categories and states a few bar charts were created
- Several date/time graphs were plotted to understand how the data changed from month to month

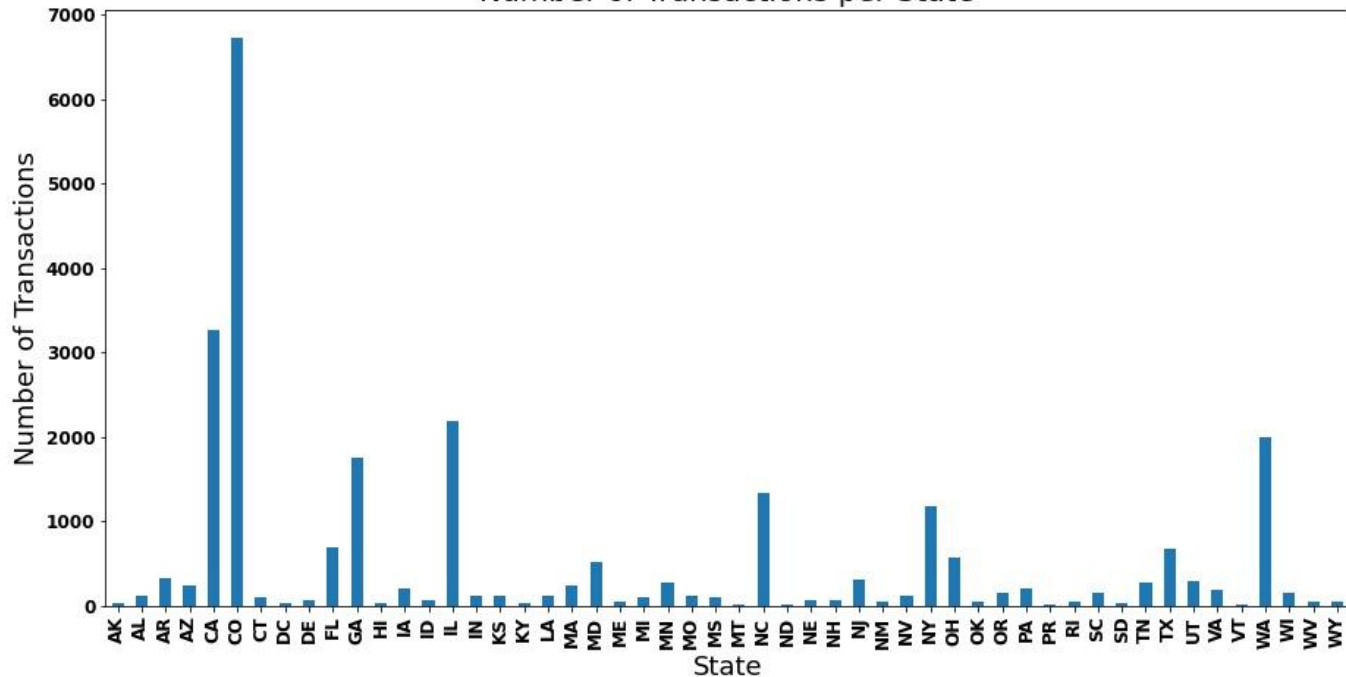
Number of Transactions per Category



Restaurants/Dining have most transactions (>8k)

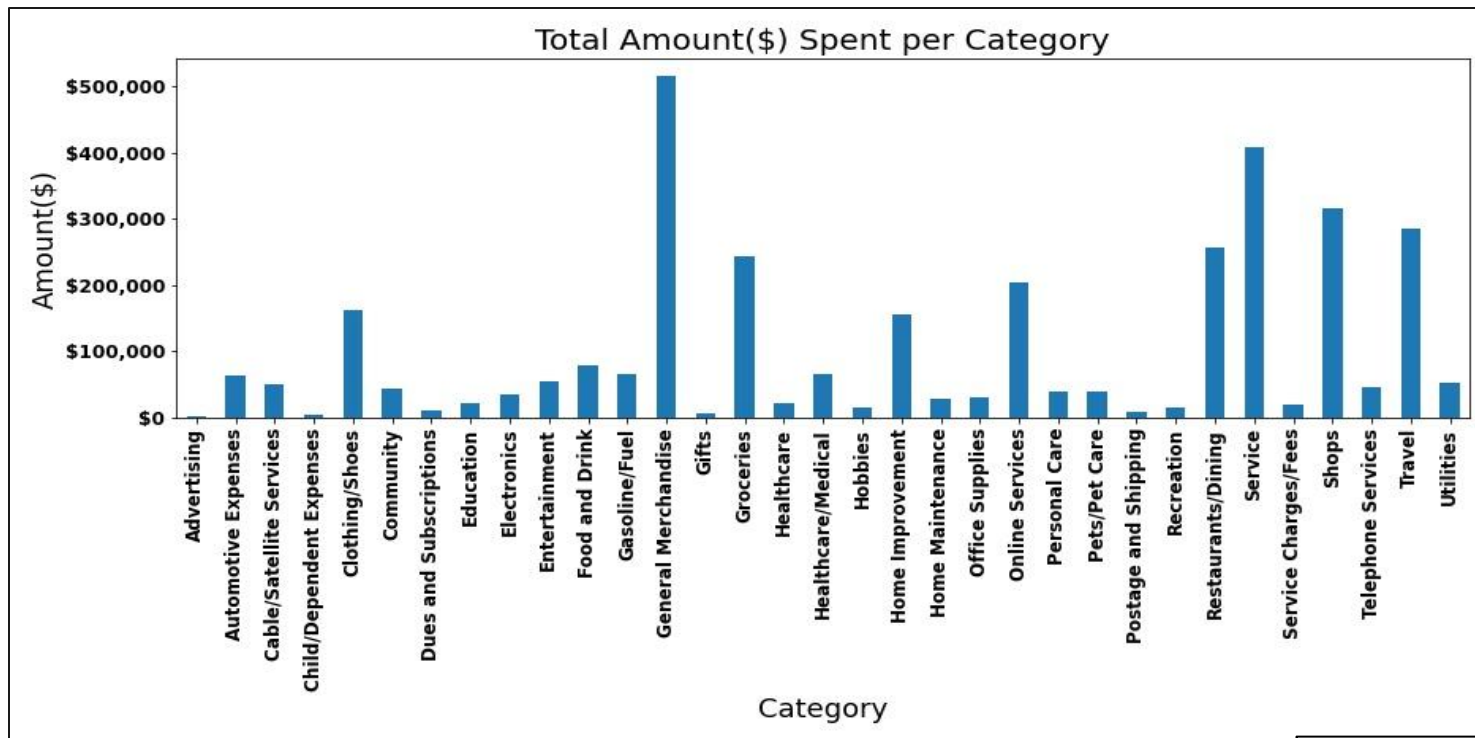
General Merchandise, Shops, Service and Groceries follow next

Number of Transactions per State



CO has the most transactions(~7k)

Majority of the states have less than 500 transactions

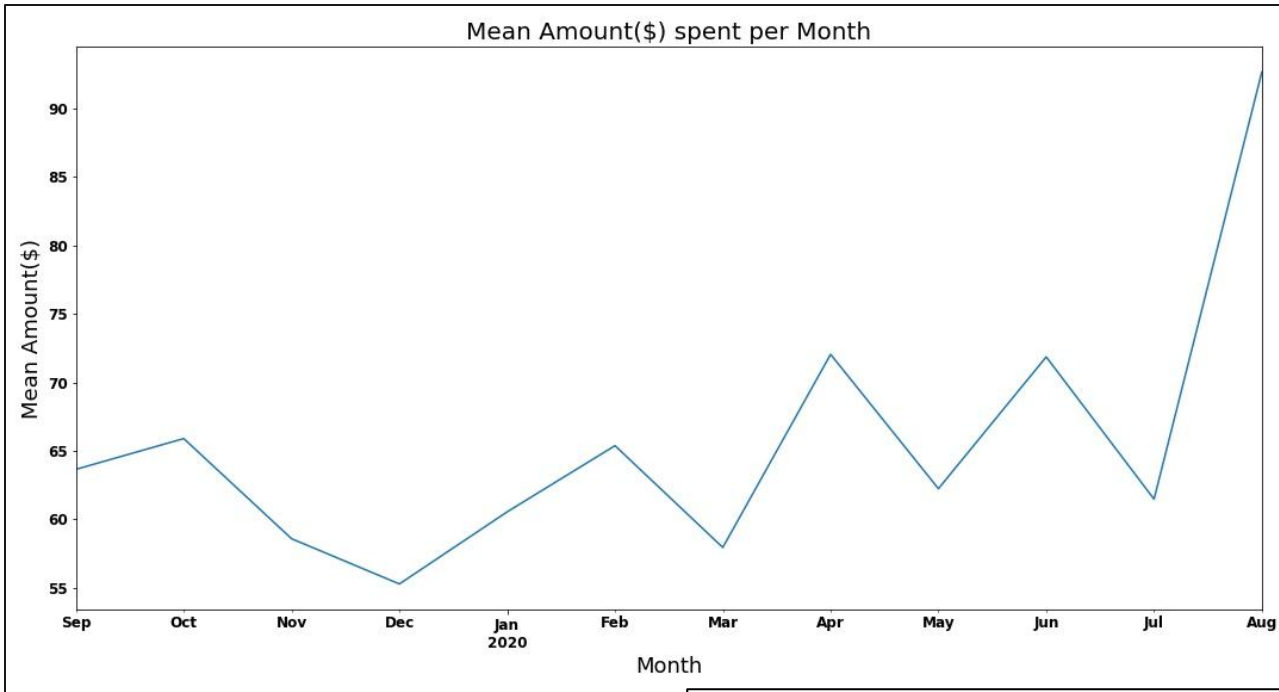


General Merchandise brought in close to \$500k

Service, Shops, Travel, Restaurants/Dining and Groceries are some categories where customer spent between \$250k - \$400k



## DateTime Analysis

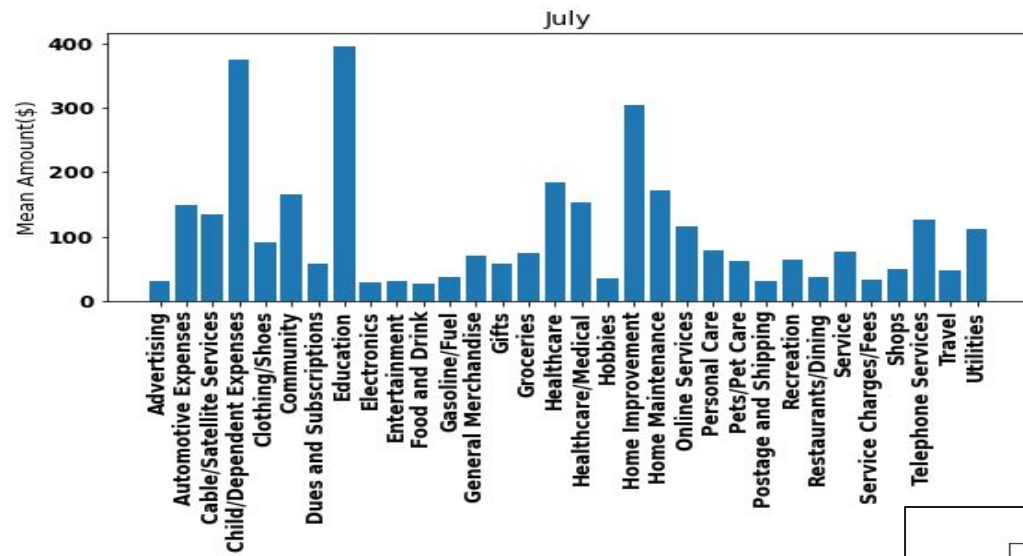


Clear spikes in the data can be noticed in different months

In March a huge dip occurs - Due to Corona sales went down

In April the data shows a high mean amount, however most of the transactions are in "Office Supply" category. Corona did affect this month as well but not in all categories

In August there is a huge mean amount increment from July - Users spent in all categories



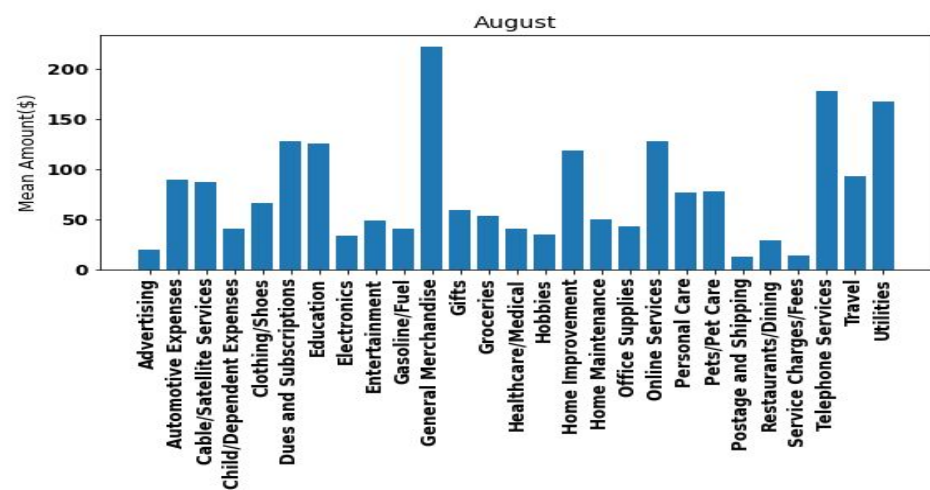
Total Transactions July - 19666  
Total Transactions Aug - 2090

Total Transactions Sum July - \$120,8925  
Total Transactions Sum Aug - \$193,644

General Merchandise - mean amount increased in August compared to July (~\$100 to ~\$200) - reason might be the start of school year

Child Dev Expenses decreased in Aug (~\$400 to ~\$50) - reason might be the start of school year

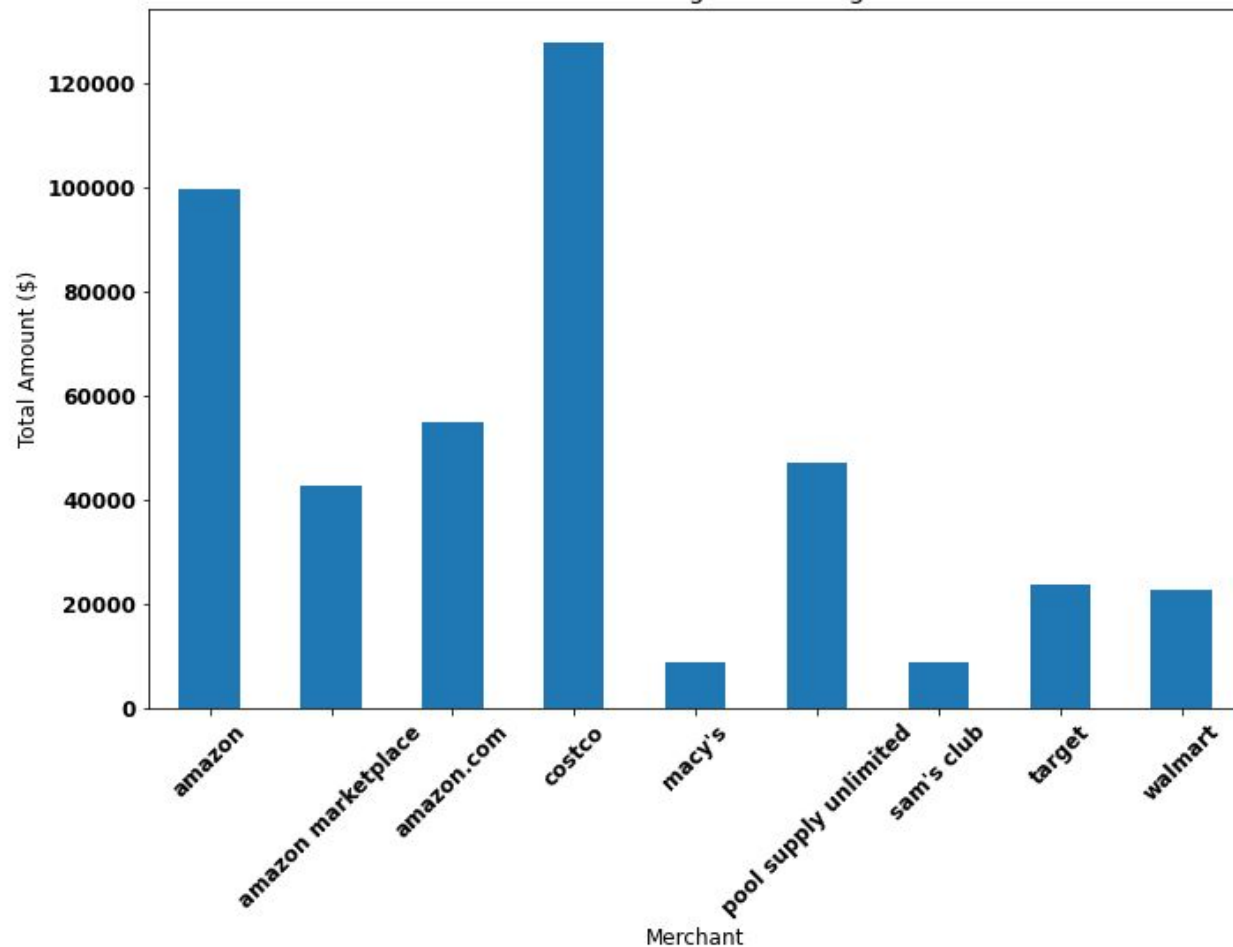
Overall there was increase in spending in Aug from amount(\$) perspective. However the number of transaction were lesser than July ....Therefore a spike in Aug Mean Amount



---

# Backup Slides

General Merchandise Highest Earning Merchants



Users spent \$128K at Costco (highest amount in the General Merchandise category)

All the top names in this category are well known except for 'pool supply unlimited'.

Upon investigations it was found that some high dollar amount in 'pool supply unlimited' transactions were duplicates.

