

**Emine Akbulut,**

[akbulutemine7@gmail.com](mailto:akbulutemine7@gmail.com)

757-500-1667



# CREDIT CARD ANALYSIS OF PSPD BANK

MARCH 15, 2020

## Table of Contents

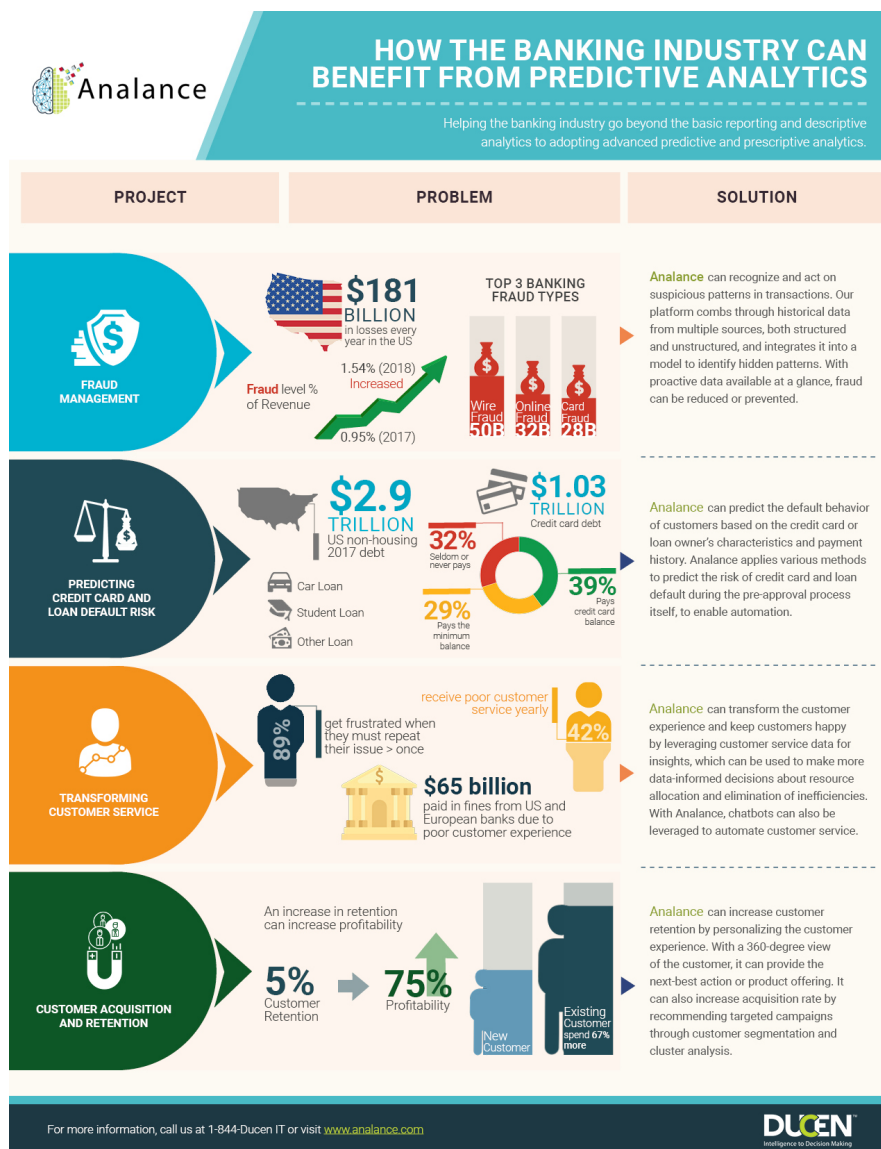
<b>About the analytics in credit card industry:</b>	<b>3</b>
<b>Problem Statement:</b>	<b>4</b>
<b>About the Dataset:</b>	<b>6</b>
<b>Data Preprocessing</b>	<b>7</b>
<b>Exploratory Data Analysis (EDA)</b>	<b>8</b>
Making estimations on the dataset	9
Creating the following summaries	9
Distribution of the City Wise Spending on each Product by Year	13
Distribution of the Segment Wise Spending:	14
Distribution of the Segment Wise Repayment:	14
Seasonal comparison of total customer spending, city wise	15
Comparison of seasonal customer spending for each product	15
Comparison histogram of log10 spend amount and repayment amount:	16
Correlation Between Features	16
<b>Feature Selection</b>	<b>18</b>
<b>Model Building / Machine Learning</b>	<b>19</b>
<b>Conclusion</b>	<b>20</b>

# About the analytics in credit card industry:

Analytics has penetrated every industry owing to the various technology platforms that collect information and thus, the service providers know what exactly customers want. The Credit Card industry is no exception. Within credit card payment processing, there is a significant amount of data available that can be beneficial in countless ways.

## *Understanding the customer behavior:*

The data available from a credit card processor identifies the types of consumer and their business spending behaviors. Hence, developing the marketing campaigns to directly address their behaviors indeed grows the revenue and these considerations will result in greater sales.



# Problem Statement:

In order to effectively produce quality decisions in the modern credit card industry, knowledge must be gained through effective data analysis and modeling. Through the use of dynamic data driven decision-making tools and procedures, information can be gathered to successfully evaluate all aspects of credit card operations. PSPD Bank has banking operations in more than 50 countries across the globe. Mr. Jim Watson, CEO, wants to evaluate areas of bankruptcy, fraud, and collections, respond to customer requests for help with proactive offers and service.

In this project, I will focus on the effect of the credit card limit on the credit card payments as well as customer spending. Predicting the patterns of spending and payment behaviors of customers is important for banks to provide more appealing offers to the customers, and to respond to customer requests more effectively. Although there are many factors that might be accounted for predicting customers' spending and payment behaviors, in this project, I will focus on the effect of credit card spending limit. Understanding the role of credit card spending limit will give a valuable input to PSPD bank to see how different amounts of limit affect how much customer spends and how much they pay their credit card debts. Furthermore, illustrating the relationship between credit card spending limit and the payment and spending behaviors of customers will provide evidences for banks to consider in determining the credit card limits for customers from different backgrounds and different spending and payment behaviors.

In analyzing this relationship, I will first look at bilateral correlation between credit card spending limit and credit card payment as well as between the spending limit and the amount of spending with the credit card. After doing the correlational analyses, I will develop machine-learning models. Machine-learning models will help PSD bank to correctly predict how credit card spending limit affects the customer spending and payment behaviors. The analyses will reveal important recommendations that PSPD bank and its CEO might want to consider.



# About the Dataset:

The data for the analyses were acquired from Kaggle.com - Credit Card Exploratory Data Analysis.

The data contain three separate files which are ;

- Customer Acquisition: At the time of card issuing, company maintains the details of customers. It has 100 observations and 8 variables including No, Customer, Age, City, Product, Limit, Company, Segment.
- Spend (Transaction data): Credit card spending for each customer. It has 1500 observations and 5 variables including SI No, Customer, Month, Type, Amount.
- Repayment: Credit card Payment done by customer. It has 1523 observations and 3 variables including SI No, Customer, Month, Amount.

The dataset used for this project can be acquired using the following link: <https://www.kaggle.com/darpan25bajaj/credit-card-exploratory-data-analysis#Repayment.csv>

# Data Preprocessing

Data preprocessing is a crucial step to be taken for a data analysis since the quality of data affects the quality of the models in predicting the concepts we attempt to explain. Hence, before jumping into performing analysis and running models, several different data preprocessing steps have been taken in order to inspect the data and look for the missing values. These steps are also important in terms of cleaning the data to make it ready for further analysis. The following steps have been taken in this section to make the data ready before feeding it into the machine learning models:

Steps taken to get the data ready for further analysis:

1. All the packages necessary for carrying out the analyses throughout this project have been imported.
2. “Credit Card Exploratory Data Analysis” dataset was read into a data-frame.
3. A general insight about the dataset was developed through checking the head, shape, info, and description of the data-frame. A preliminary analysis of the head and info of the data-frame showed that there are some columns with 23 numbers of missing values. Checking the description of the data-frame revealed that the numerical attributes looks normal (within accepted ranges) but there are some outliers were detected and they are managed. Unique values of id attribute were counted to see how many unique individuals we are dealing with.

The following section includes the results of the various exploratory data analyses These analyses have been performed to uncover the relationships between variables.

# Exploratory Data Analysis (EDA)

This section focuses on exploratory data analysis which allows us to understand relationships between the variables before moving onto more complex data analysis. I took a glance of the data in the previous section and will explore the data with the help of graphs in order to illustrate the relationships between variables. The amount of customer spending and the amount of credit card repayment are my target variables in this project. In predicting these two features, I will focus on the effect of credit card spending limit. The following graphs have been produced to explore the relationship between these three features as well as some other important features in the data.



1. Making estimations on the dataset
  - Estimation of Customer Age Mean
  - Estimation of Spend Amount Mean
  - Estimation of Repayment Amount Mean
2. Creating the following summaries
  - How many distinct customers exist?
  - How many distinct categories exist?
  - What is the average monthly spending by customers?
  - What is the average monthly repayment by customers?
  - Which city is having maximum spending?
  - Which age group is spending more money?
3. Distribution of the City Wise Spend on each Product by Year
4. Distribution of the Segment Wise Spend
5. Distribution of the Segment Wise Repayment
6. Create graphs for
  - Seasonal comparison of total spending, city wise
  - Comparison of seasonal spending for each product
  - Comparison histogram of log10 spending amount and repayment amount
7. Correlation Between Features



### Making estimations on the dataset

- Customers age mean: 48.4
- Spend amount mean : 159944.32
- Repayment amount mean : 163321.57

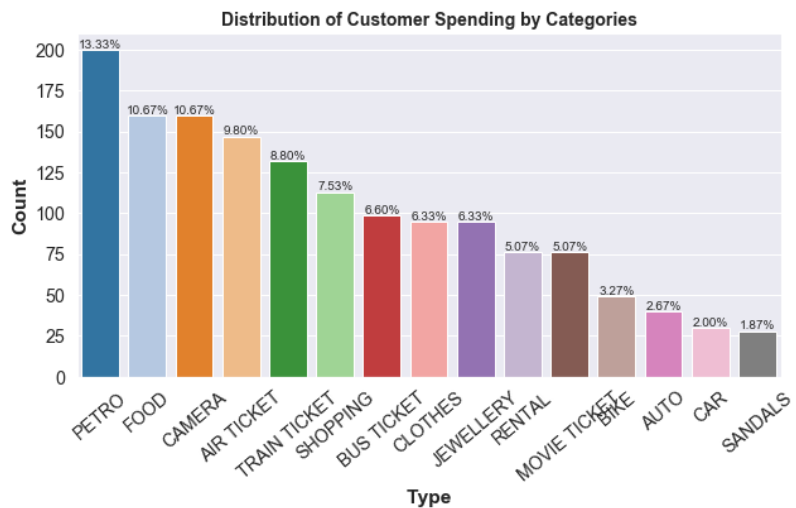
### Creating the following summaries

- Distinct Customers : Number of distinct customers is 100 .
- Distinct Categories:

*# Customers from Different Spend*

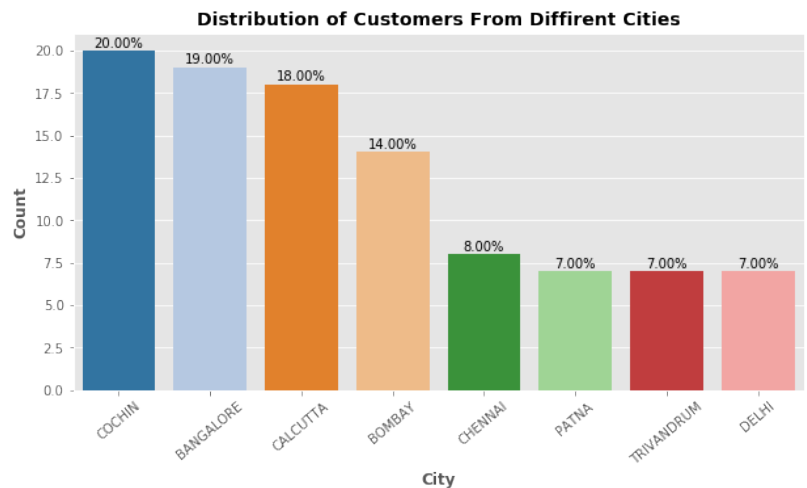
*Type:*

This graph illustrates the distribution of overall customer spending based on different categories. As the graph shows, the highest amount of spending has been made over petroleum products. Food and come next as 10.67 percent of all customer spending has been made per each. The lowest amount of spending has been made on sandals. Only 1.87 percent of overall spending has been made over this product.



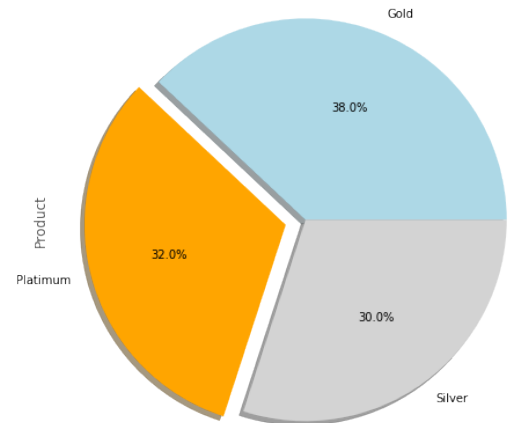
*# Customers from Different Cities:*

This histogram graph demonstrates that Cochin has the highest number of customers in the country while the least amount of customers using credit cards are reported in Delhi.



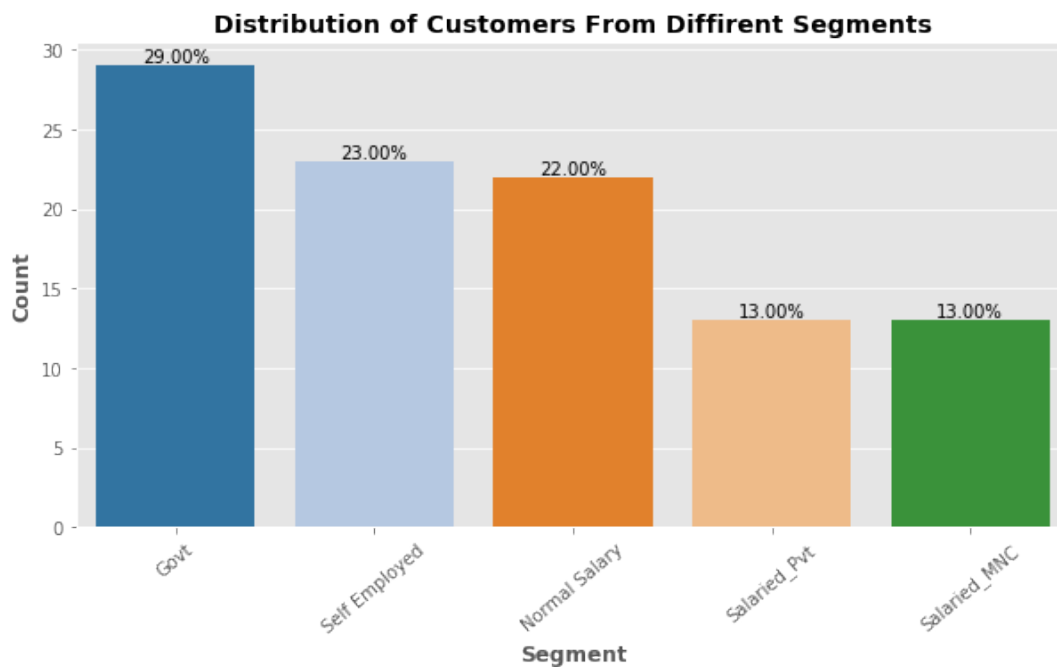
### # Customers from different products:

This pie chart shows that there is not a very significant difference on the usage of different credit card products in percentage. It reports that the number of customers using Gold credit cards is higher than those using Silver and Platinum products.



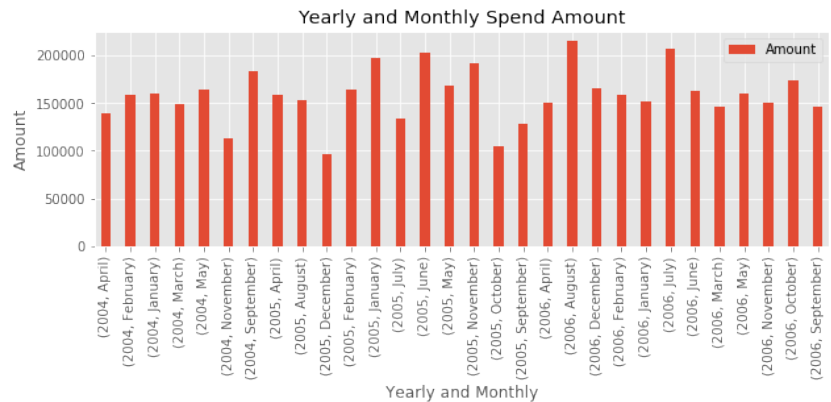
### # Customers from different segments:

This bar graph illustrates economic status of the customers in the sample. It shows that the number of customers working for government (29%) is higher than people from other occupation groups.



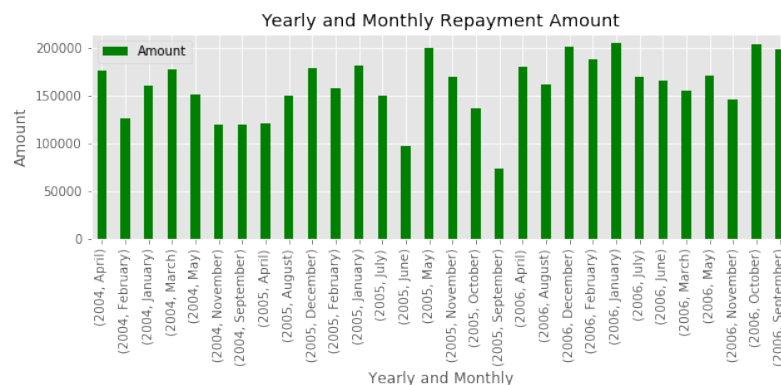
- The average monthly spending by customers:

The above graph demonstrates the two-year distribution of customer spending. The time range covers from April 2004 to September 2006. It suggests that the highest amount of customer spending is reported on August 2006. The least amount of customer spending is on December 2005. The graph below shows the same two-year distribution for repayment by customers. It demonstrates that the highest amount of repayments is reported on October 2006 and January 2006. The least amount of repayment by customers is reported on September 2005. Thus, the dates of the highest and lowest customer spending do not match with the dates of the highest and lowest repayment by customers.



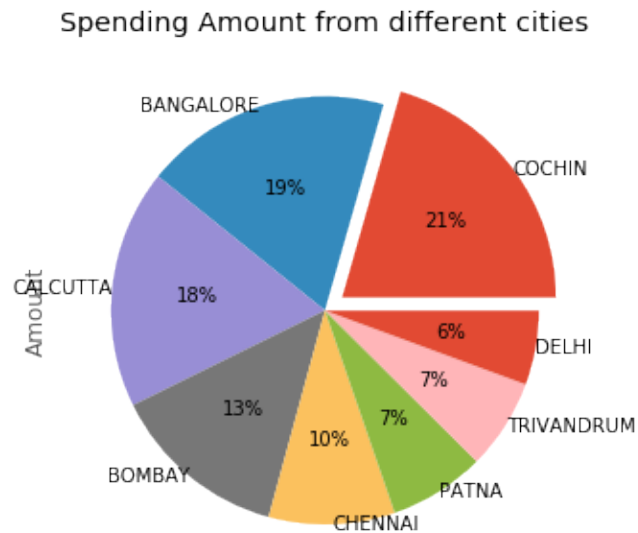
- The average monthly repayment by customers:

The graph above demonstrates yearly and monthly distribution of credit card payments. It demonstrates that the highest amounts of credit card payments have been made in October 2006 and January 2006. The lowest amount of payment is reported on September 2005. The evidence in this graph and the previous graph suggest that we don't see an overlap in terms of the dates of highest credit card payments and highest customer spending.



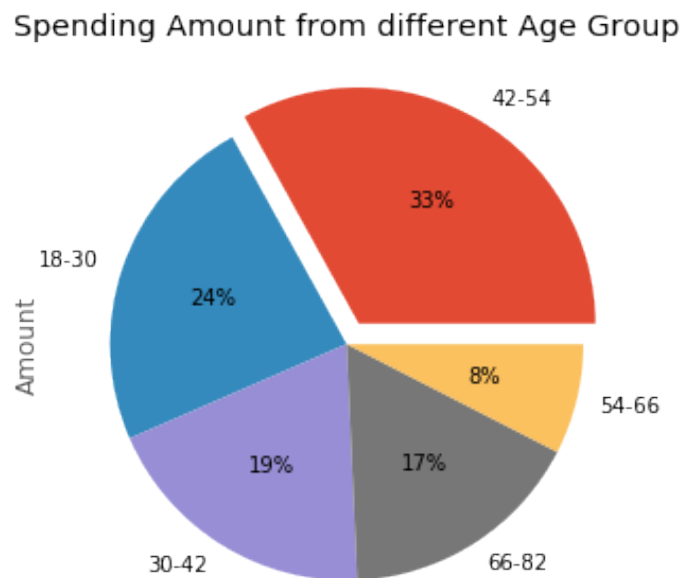
### •Cities Spending Ratio :

In the pie chart below, I have illustrated the amount of customer spending in different cities. It reports the highest customer spending is produced in Cochin, which is not surprising because it includes the largest amounts of customers in the sample.

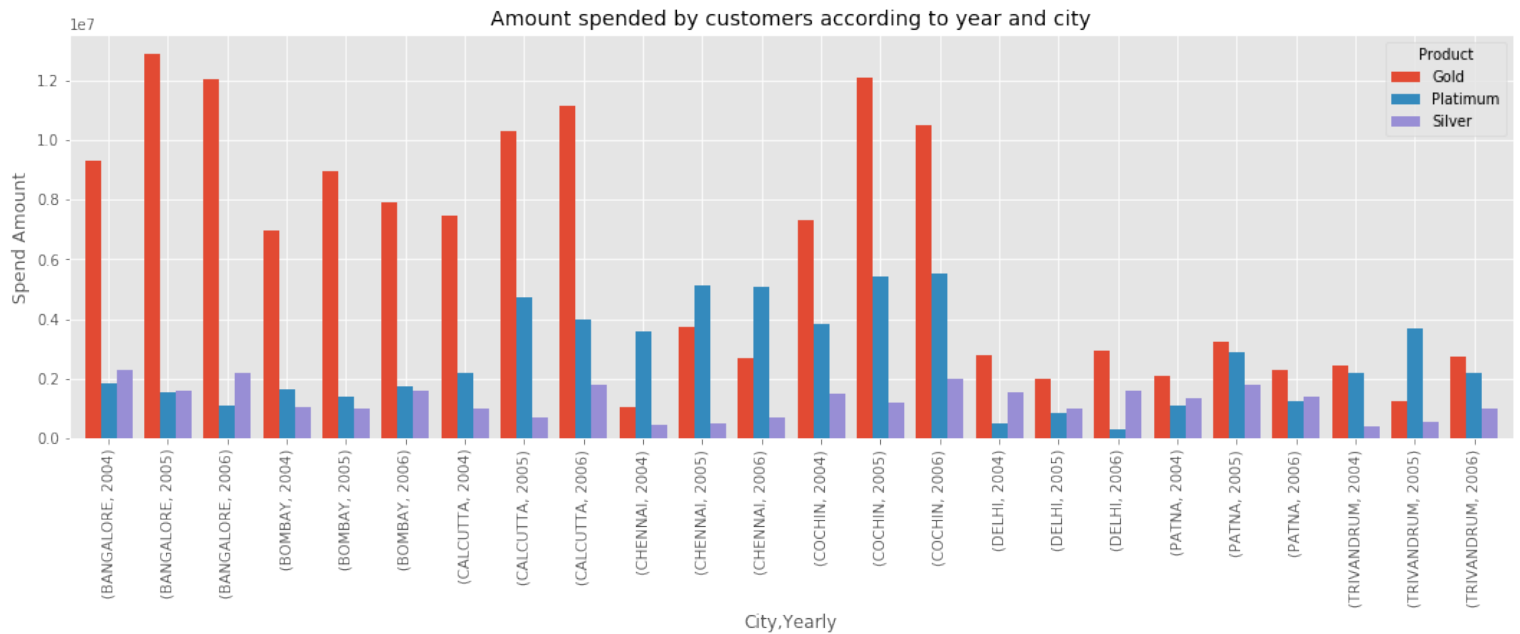


### • Age group is spending more money:

From the pie chart shown below we can say that age group 42 - 54 is spending more money than people in other age groups.

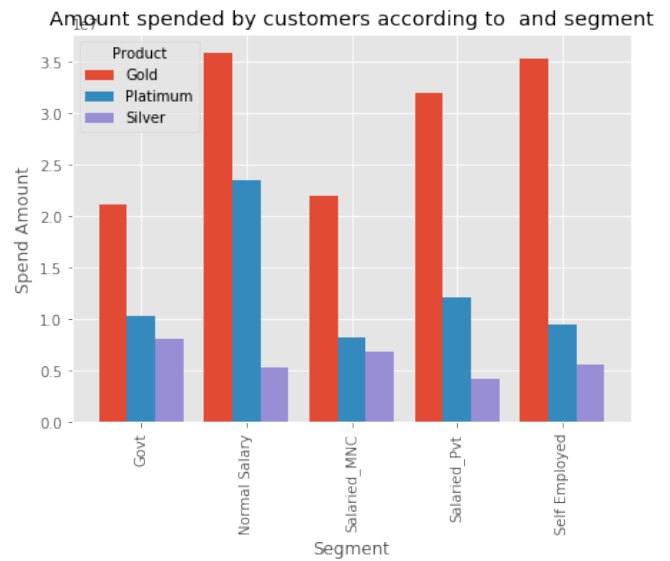
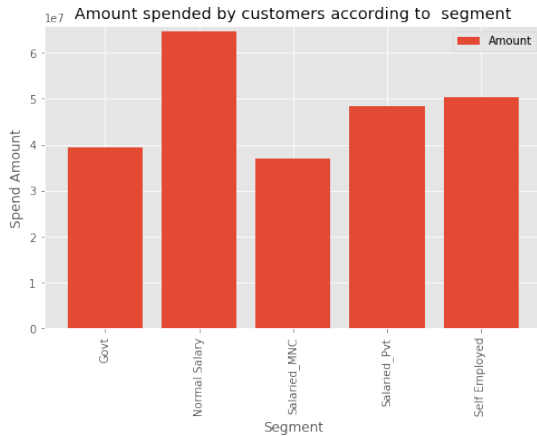


## *Distribution of the City Wise Spending on each Product by Year*



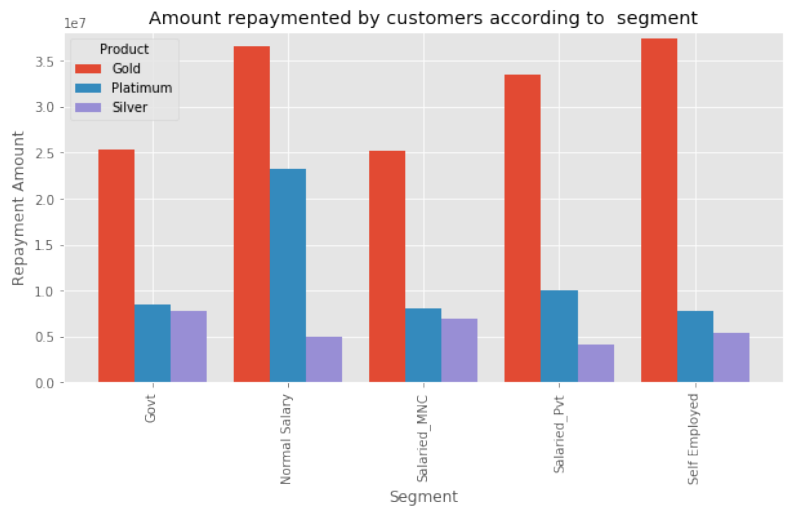
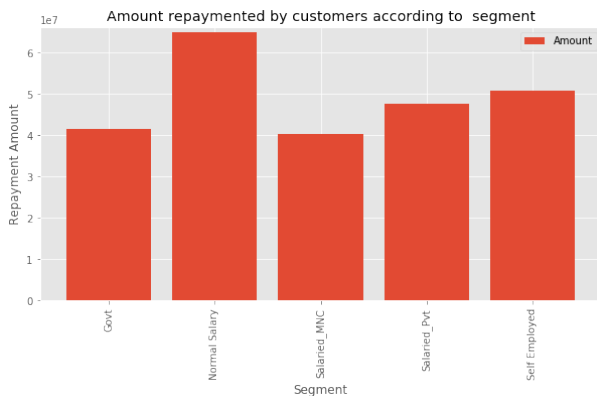
This graph above demonstrates that customers with gold credit cards make higher spending than those with platinum and silver. The only two cities where customers with platinum card has made more spending than those with gold cards are Trivandrum, in 2005, and Chennai in 2004. It also reports that the highest amount of customer spending has been made in Bangalore in 2005.

### Distribution of the Segment Wise Spending:



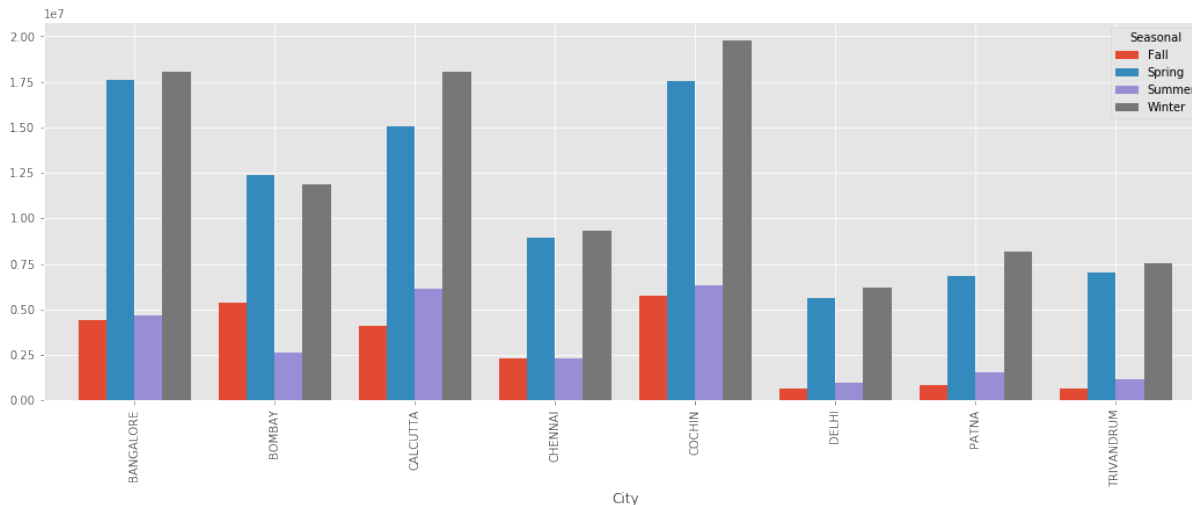
The two graphs above show the distribution of customer spending based on different economic segments. Both graphs suggest that people with normal salary are making the highest spending compared to customers in the other economic segments, particularly those having gold and platinum credit cards. However, the amount of spending made by the customers at the normal salary segment who use silver credit cards is not at the highest level, compared to customers at other economic segments who use silver cards. The graph on the left suggests that customers who work in government jobs and use silver cards make higher amounts of spending than those with silver cards at other economic segments.

### Distribution of the Segment Wise Repayment:



The two graphs above illustrate the distribution of credit card payments by economic segments. Consistent with the previous two graphs demonstrating the distribution of customer spending, customers with gold and platinum cards who are at normal salary and self-employed segments make higher credit card payments than those at the other economic segments. Those who are self-employed, however, make slightly higher credit card payments than those who are at normal salary segment.

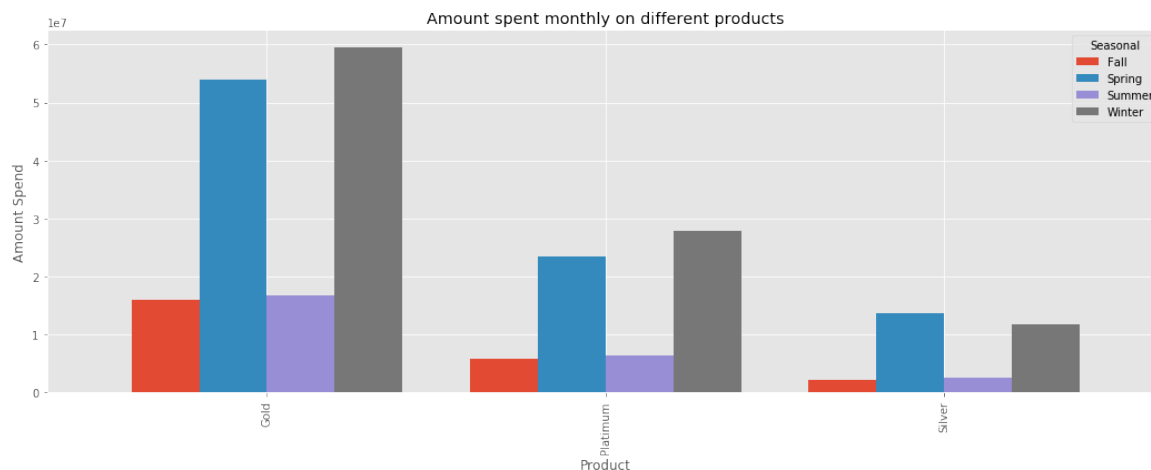
### Seasonal comparison of total customer spending, city wise



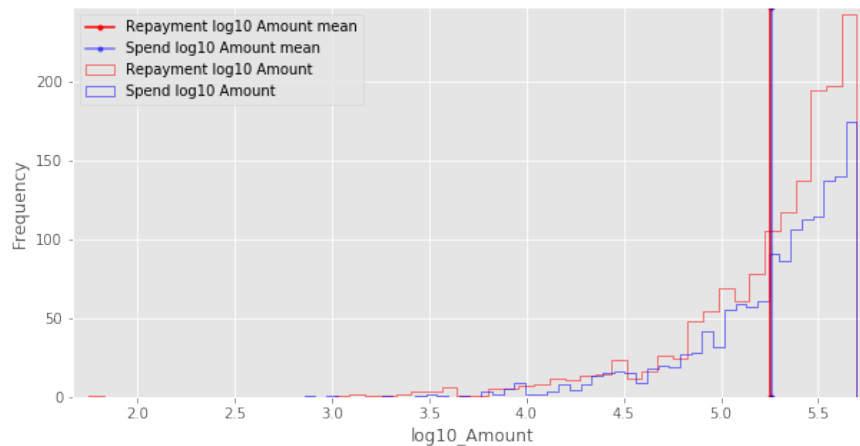
The graph above illustrates the seasonal distribution of customer spending on different cities. It reports that the winter season sees the highest amount of customer spending, except for the city of Bombay where the customer spending in spring is higher than the other seasons. The highest winter season spending has been made in Cochin, the highest spring season spending has been made in Cochin and Bangalore, the highest fall spending has been made in Bombay and Cochin. Finally, the highest summer spending has been made in Cochin and Calcutta.

### Comparison of seasonal customer spending for each product

In the graph below, I have made a comparison across three products based on season. It shows that the highest customer spending with gold and platinum cards has been made in the winter season. As for the customer spending with the silver card, the highest spending has been made in the spring season.



### **Comparison histogram of log10 spend amount and repayment amount:**



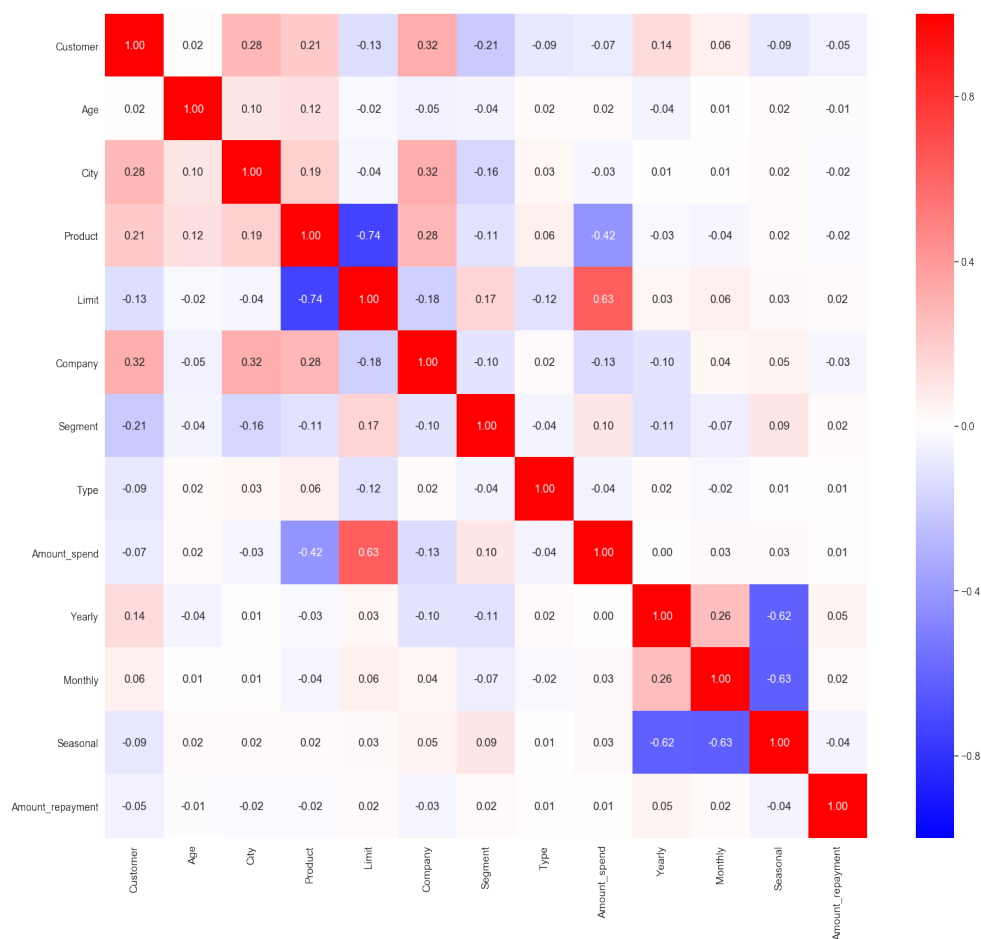
In the graph above, I have demonstrated the changes in the amounts of credit card payments and customer spending. The graph shows that as the customer spending increases, the amounts of credit card payments increases.

### **Correlation Between Features**

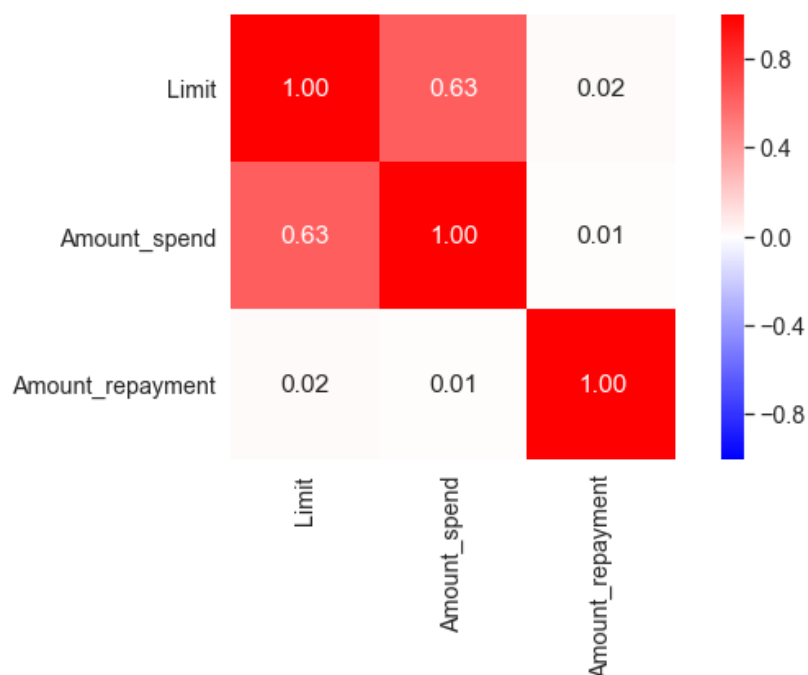
Since the data includes many features, the entire dataset was divided into two parts to create a correlation matrix by using the heatmap that is illustrated below. Features with object data type were converted into categorical data type and the correlation matrix below includes those features with categorical codes.



The first heatmap illustrates the relationships between the variables. While the red color indicates a positive correlation between the two variables, blue color indicates negative correlation. The magnitudes of correlations have also been shown on the heatmap. In the second heatmap, I looked at only the variables of interest in my project, which are credit card spending limit, customer spending, and credit card payments.



The second heatmap demonstrates that there is a positive and significant correlation between credit card limit and customer spending. The correlation coefficient is reported 0.63. However, the heatmap below shows that credit card payment isn't significantly correlated with credit card limit ( $p=0.02$ ) and



# Feature Selection

# **Model Building / Machine Learning**

# Conclusion