Stats 112 Final Project

Meghal Dubey, Sibo Guo, Humin Zhang, Julius Castro, Qinyi Chen, Luis Ceja Abrica

12/1/20

Introduction

Credit card churning is a method of gaming bonus incentives offered by banks and credit card companies through the practice of repeatedly applying and closing credit card accounts and only meeting the minimum spending requirments to earn the bonus incentives. These incentives include cash back, airline miles, airline companion passes, hotel loyalty points, and hotel free night certificates. Credit card churners cost banks and companies millions of dollars in profit. As a result, many companies have created systems to detect credit card churners and blacklist them.

Goal of the Study

1. The goal of this study is to identify factors that contribute to credit card churning.
2. Build a prediction model that can accurately classify potentional credit card churners.

Why This Dataset was Chosen?

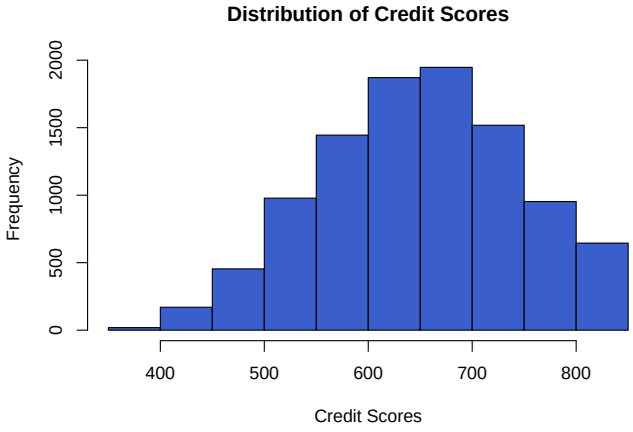
Our team has a common interest in the practical applications of data science and analytics in the finance industry and since this dataset encompasses several aspects of finance, it gives us the opportunity to understand a lot of the different problems that companies can solve in the industry using data science and analystics.

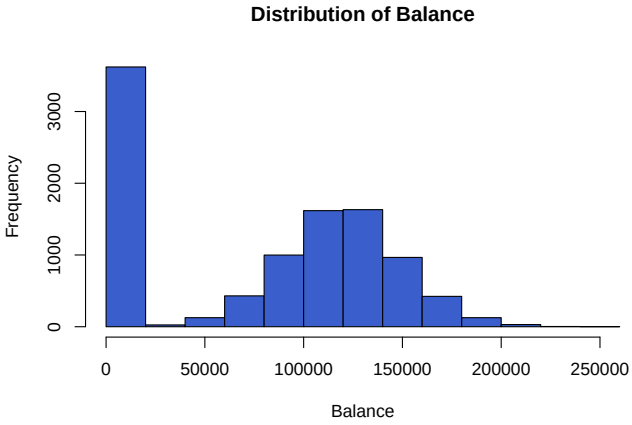
Questions About the Data

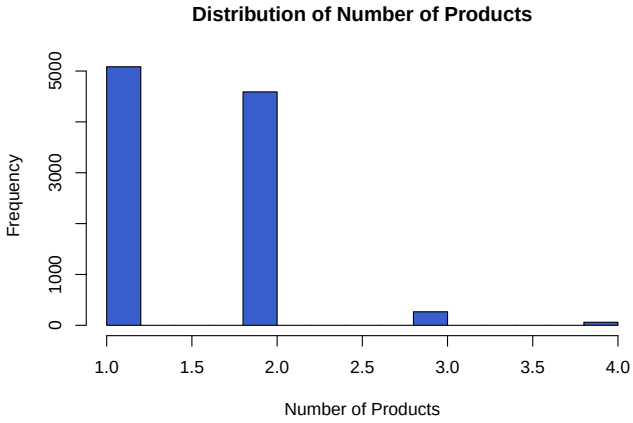
1. How are customers classified as Active Users?
2. Is Balance the average revolving balance in the customers account over a period of time or is it a balance in a single point in time.
3. When were credit scores recorded? Were they recorded when the customer applied for the credit card or were they recorded after?

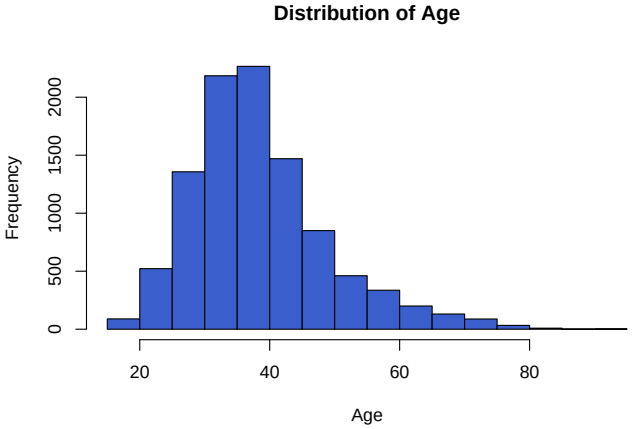
Exploratory Data Analysis

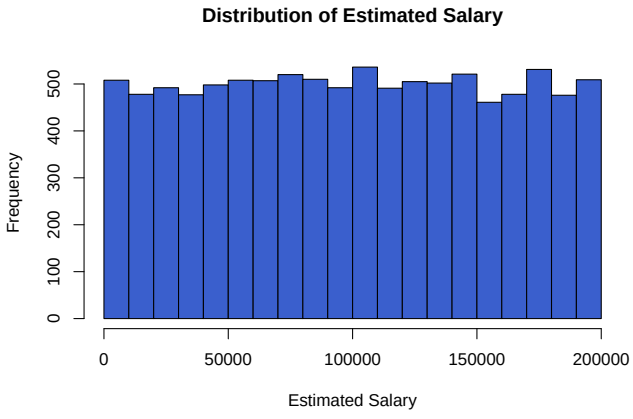
Numerical Variables

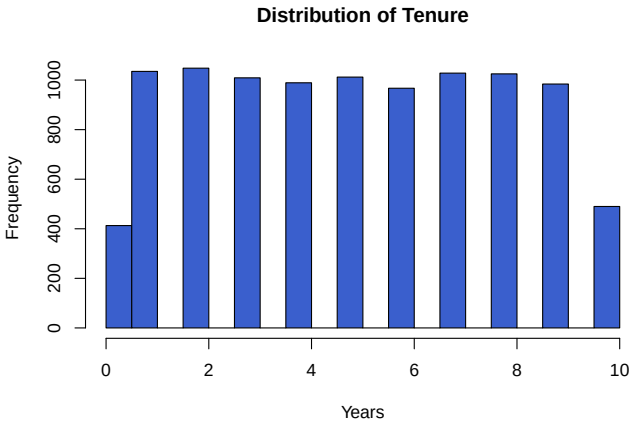




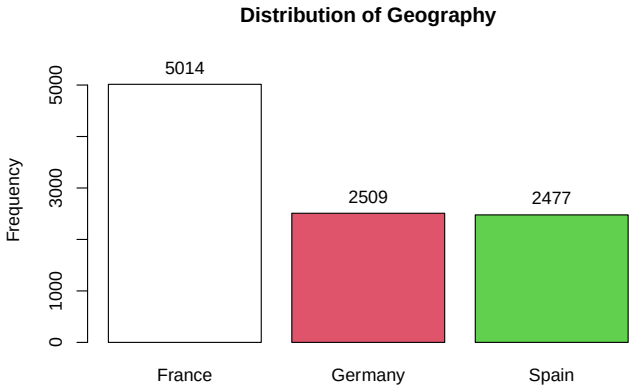




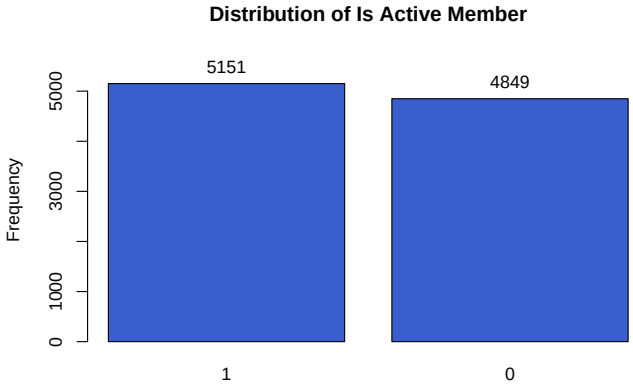




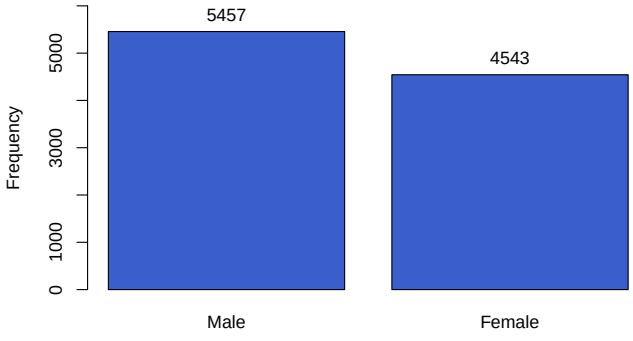
Catergorical Variables




```
## Credit_Card_Churning$Geography :
##         Frequency Percent Cum. percent
## France       5014    50.1         50.1
## Germany      2509    25.1         75.2
## Spain        2477    24.8        100.0
##   Total     10000   100.0        100.0
```

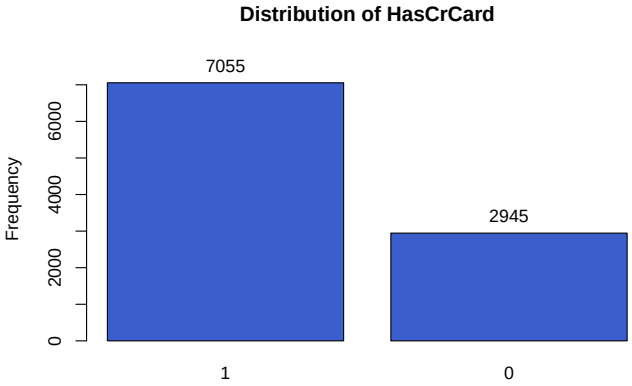



```
## Credit_Card_Churning$IsActiveMember :
##         Frequency Percent Cum. percent
## 1            5151    51.5         51.5
## 0            4849    48.5        100.0
##   Total     10000   100.0        100.0
```



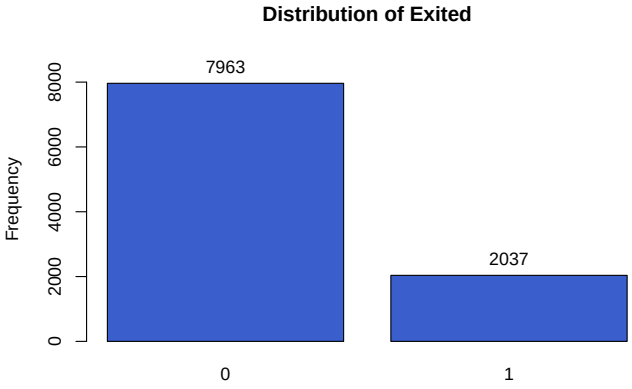

```
## Credit_Card_Churning$Gender :
##         Frequency Percent Cum. percent
## Male         5457    54.6         54.6
```

```
## Female       4543    45.4         100.0
##   Total     10000   100.0         100.0
```




```
## Credit_Card_Churning$HasCrCard :
##         Frequency Percent Cum. percent
## 1            7055    70.6         70.6
## 0            2945    29.4        100.0
##   Total     10000   100.0         100.0
```

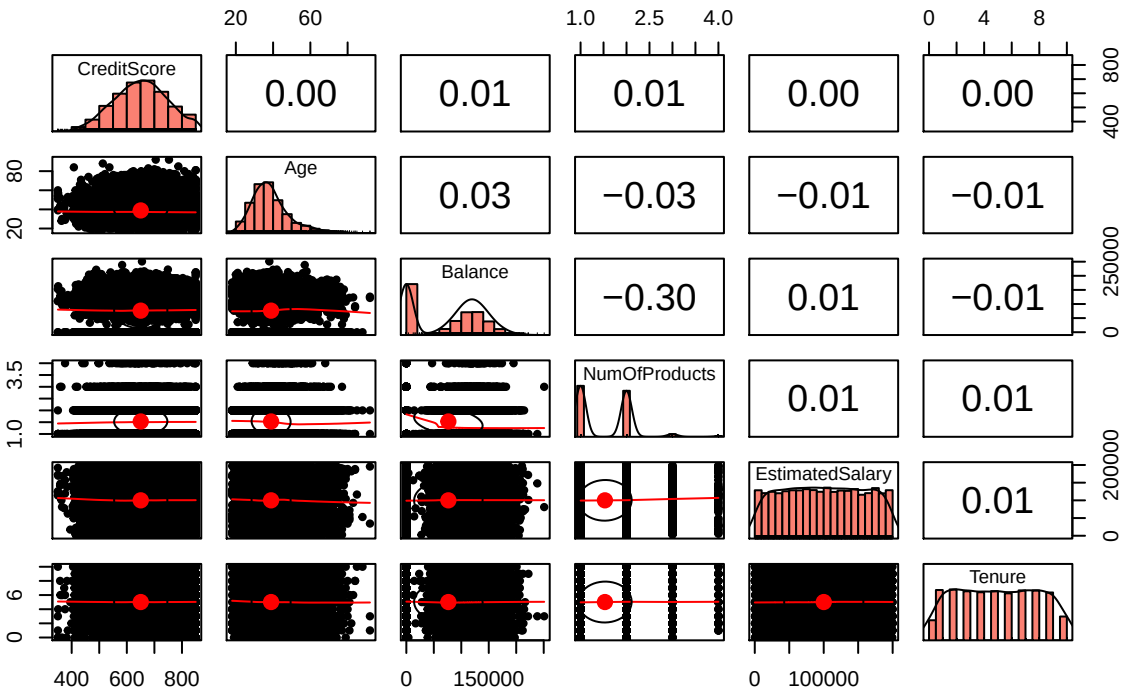
Outcome Variable




```
## Credit_Card_Churning$Exited :
##         Frequency Percent Cum. percent
## 0            7963    79.6         79.6
## 1            2037    20.4        100.0
##   Total     10000   100.0         100.0
```

Correlation Matrix

	CreditScore	Age	Balance	NumOfProducts	EstimatedSalary	Tenure
CreditScore	1.000	-0.004	0.006	0.012	-0.001	0.001
Age	-0.004	1.000	0.028	-0.031	-0.007	-0.010
Balance	0.006	0.028	1.000	-0.304	0.013	-0.012
NumOfProducts	0.012	-0.031	-0.304	1.000	0.014	0.013
EstimatedSalary	-0.001	-0.007	0.013	0.014	1.000	0.008
Tenure	0.001	-0.010	-0.012	0.013	0.008	1.000



Note: This plot shows how each numerical variable distributed and their correlation matrix based on Pearson correlation coefficient method.

Insights

The above EDA reveals the following insights:

- Credit Scores among customers are approximately normally distributed with a mean of 650.53 and a standard deviation of 96.65.
- 36.2% of customers carry a balance of $0 on their credit cards
- Age of Customers is skewed right with the average age of 39 years old, with the youngest customer being 18 and the oldest customer being 92
- About half of customers in the data set are in France, about a quarter from Spain, and a quarter from Germany
- 51.5% of customers are active users and Gender is split approximately evenly.
- Estimated Salary is uniformly distributed from 0 to 200,000 (Note: Estimated Salary is self reported and is not officially confirmed using formal documents)
- None of the variables are highly correlated with one another. The highest correlation is between Balance and NumOfProducts, which makes sense as customers who have more credit cards with the company tend to have higher balances.