Exploratory Data Analysis

Of

Credit Card Customer Attrition

BY

Isaiah Wright

5/9/2022

Fundamentals of Data Science and Analytics – CS345

Professors: Dr. Adam Albina, Dr. Stephen Shea

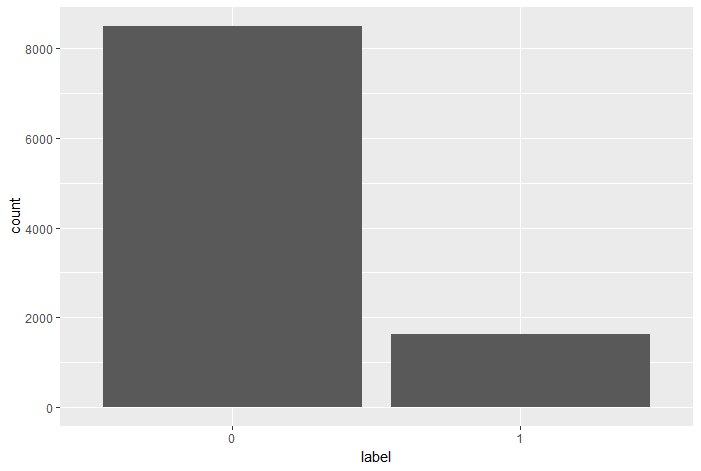
INTRODUCTION

The data used in this analysis is from a credit card company provider. The data consists of their clients and relevant information about them. The data includes information of each client’s number, age, gender, education level, marital status, income category, card category, dependent count and credit limit. The company lately has been losing customers and unsure why that is happening they have kept track of all customers that have left the company and have them as Attrited customers in their data those who haven’t left have been tagged as existing customers. The goal of the data analysis is to figure out if we can predict whether a customer will leave or stay. This also includes finding which variables correlate with customers who stay and customers who leave, to give the company a better understanding of their customer base.

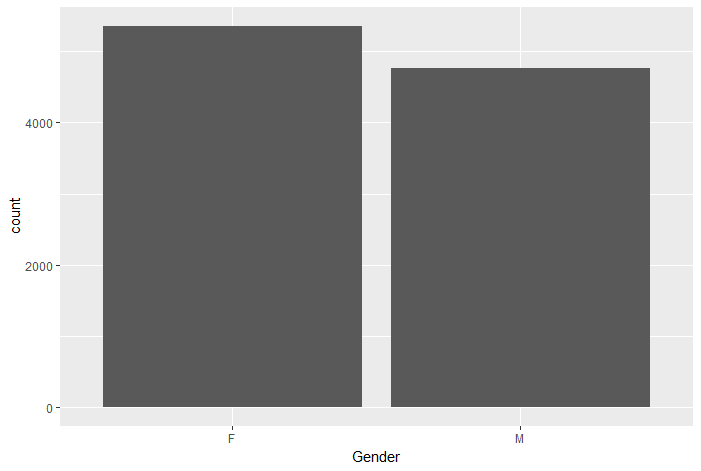
Basic Data Cleaning and Preprocessing

The data itself was relatively complete with no missing values and all data types weren’t violating any data types the only inconvenience came from the client number and the attrition flag. The client number interfered with modeling so it was removed and the rest of the data was kept in the data frame (line 59). The attrition flag was a categorical variable, to make it an actual number, an ifelse was ran to turn each category in 0 and 1 (line 23) and renamed to label (line 24). Those who have left get a 1 and those who haven’t 0.

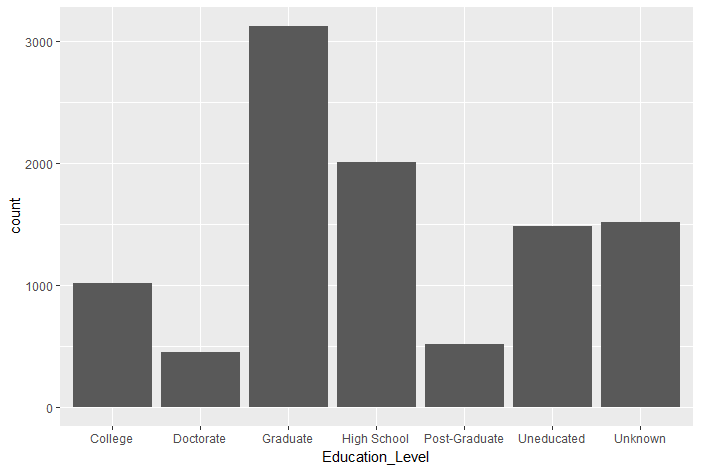
Univariate Exploration:



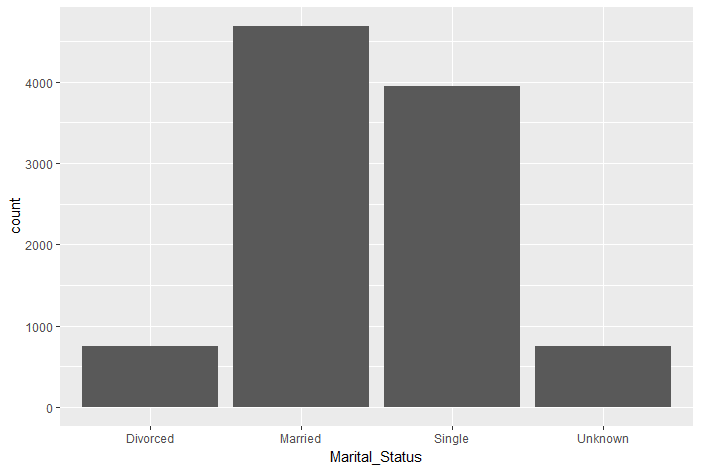
(Line 28)



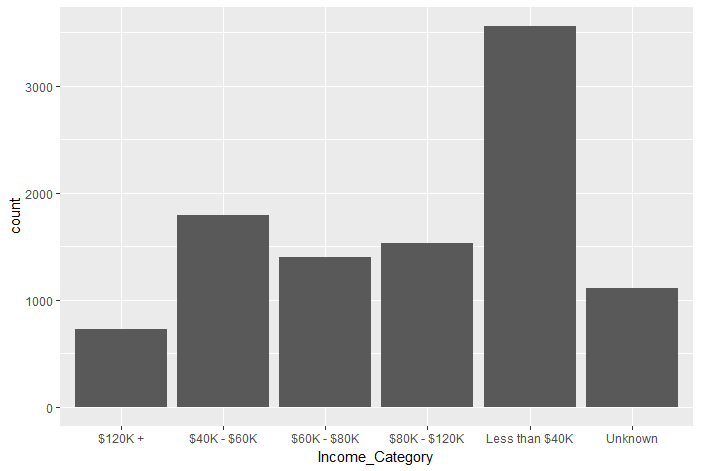
(Line 30)



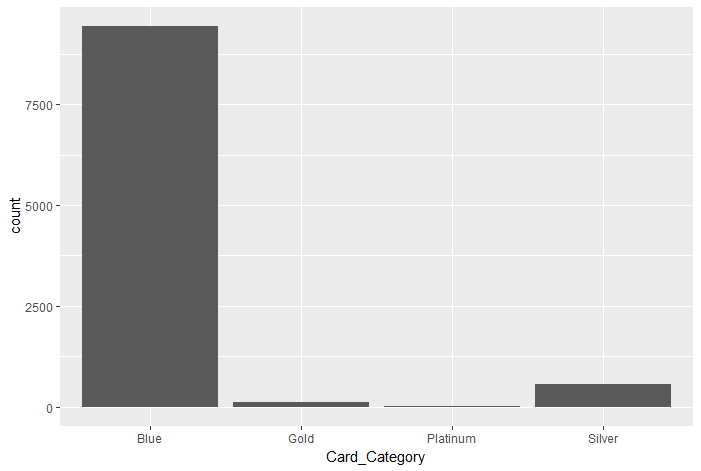
(Line 32)



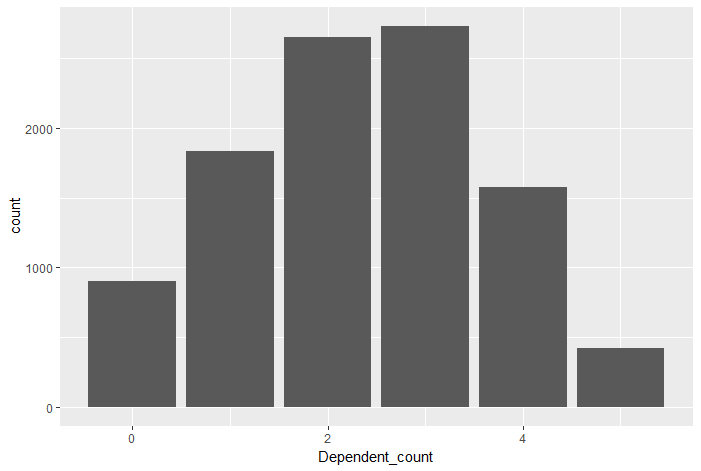
(Line 34)



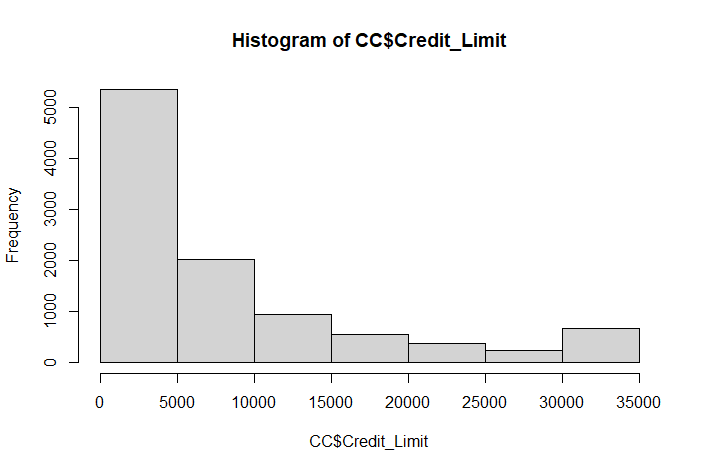
(Line 36)



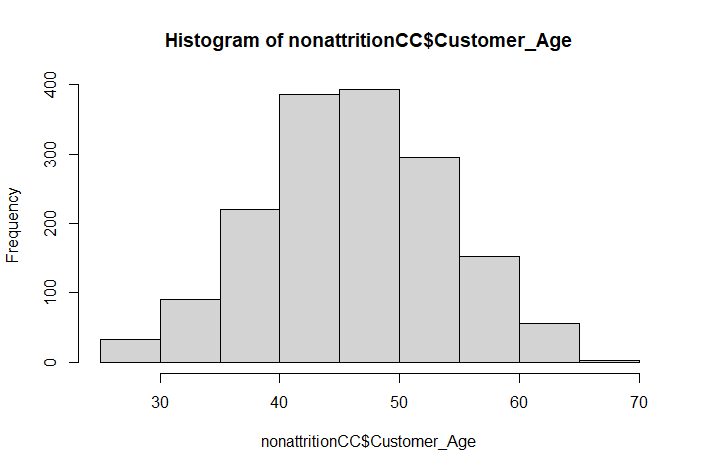
(Line 38)



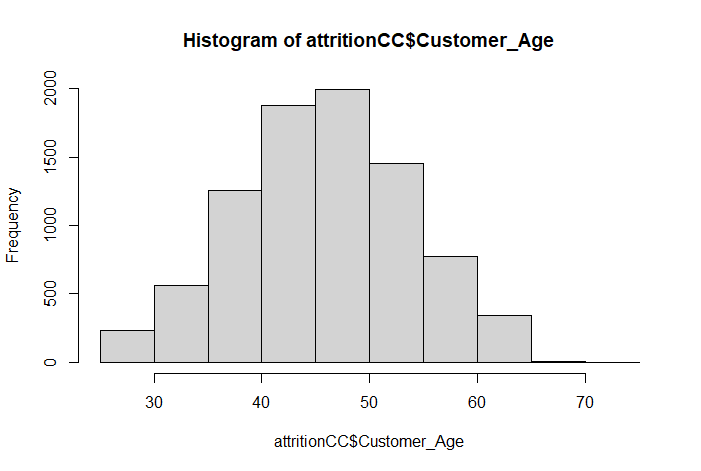
(Line 40)



(Line 42)

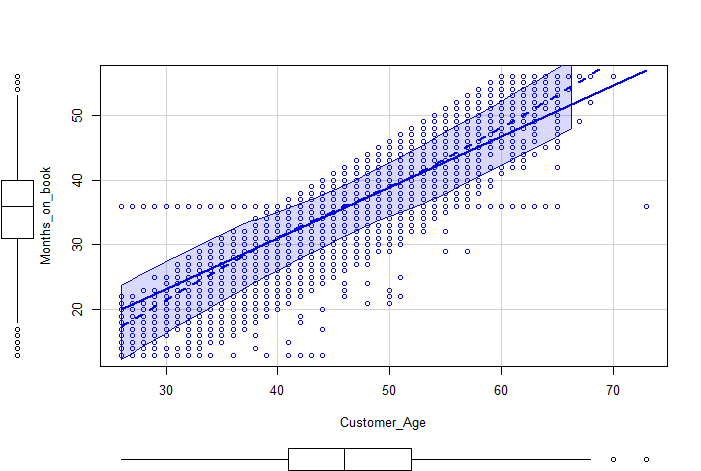


(Line 43)

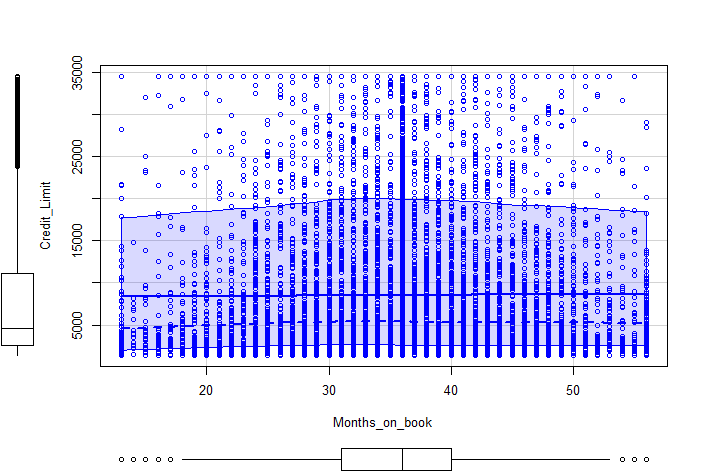


(Line 44)

Bivariate Exploration

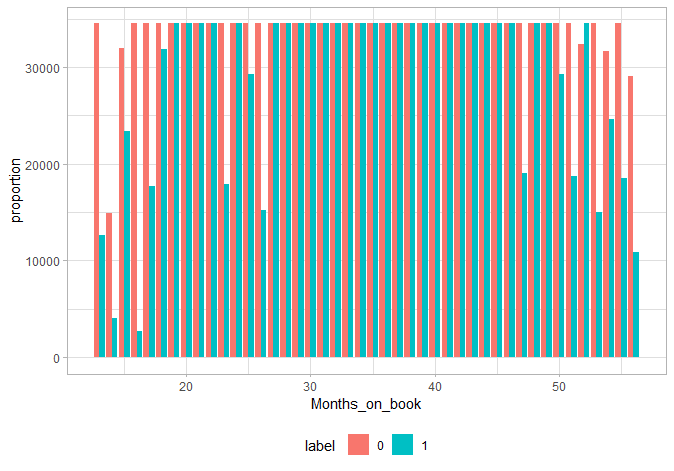


(Line 47)



(Line 48)

Multivariate Exploration



(Line 54)

Conclusion

In this analysis, I have found that most of the customers are Women, college graduates, and are married. Many of the customers claim 2 people insinuating Customers have families this coincides with many of the customers being in their 40’s or 50’s. I noticed that many people don’t make more than 40,000. What I found interesting is that there is an obvious correlation between customer age and how long they have been with the company. This could stand to reason that customer who are older are more likely to stay than those who are younger.

Statement

I made a model that incorporates Customer Age, Gender, Dependent count, Education Level, Marital Status, Income Category, Card Category and Credit Limit. I chose this because it yielded the highest adjusted r – value and I plan to test this model by backwards testing