315 Final Project: Exploring the Drivers of Credit Card Usage

Bhavya Jain, Christian Angelov, Raka Mazumder, Christina Kim

I. Dataset Introduction

Our dataset, called Credit Card Customers, consists of 10127 rows, with each row representing a person/client. Our dataset includes 23 columns consisting of various demographics such as age, number of dependents, education level, marital status, and income category. The rest of the columns refer to various credit card statistics, which are the focus of our final project. These include transaction amount, transaction count, credit limit, revolving balance, average card utilization ratio, and more. Our data is sufficiently large and complex enough for us to answer our three research questions as we have plenty of predictors (demographics) and response variables (credit card statistics) that are of interest to us. By doing so, we hope to gain a deeper understanding of the relationship between various socioeconomic factors and credit card usage.

II. Research Questions

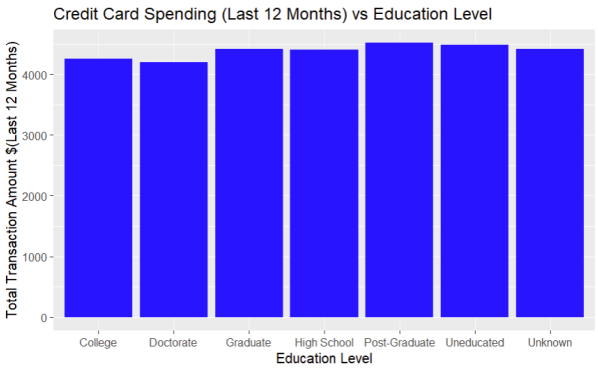
Our three research questions include the following:

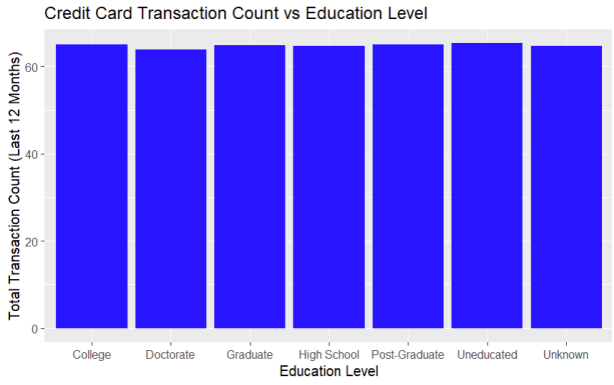
1) How does education level affect credit card transaction amounts/counts?
2) Can credit card utilization ratio be predicted by demographics or other baselines?
3) Do people with higher education status or income have longer relationships with the bank and higher credit limits?

By analyzing these questions, we hope to discover any insights or patterns that can better allow banks to meet the needs of different customer groups. This report aims to develop a deeper understanding of the interplay between socioeconomic factors and credit card usage. Considering this information can ultimately lead banks and other financial institutions to create a personalized banking experience, resulting in higher customer satisfaction.

III. Research Question #1: How does income status and education level affect credit card transaction amounts/counts?

Bar Charts: Mean Credit Card Spending and Mean Transaction Count vs Education Level





The two bar charts illustrate mean credit card spending and mean transaction counts vs education level. The first bar chart has an x-axis of education level, such as College and High School, with a y-axis of the average total transaction amount. The second bar chart also has an x-axis for education level but a y-axis for the average total transaction count. We see very little difference in terms of credit card spending and transaction count for the last 12 months across members of different education levels. For instance, Post-Graduates and Undeducated people spend a little more on credit cards, but the difference is negligible, as confirmed by the ANOVA tests we ran. Regarding transaction counts, the difference is even more negligible as it is hard to distinguish at all between the education groups. As we mentioned earlier, similar transaction counts are not surprising, but the spending is rather surprising. We might be able to infer that wealthier individuals are saving more or that they are spending in methods other than credit cards.

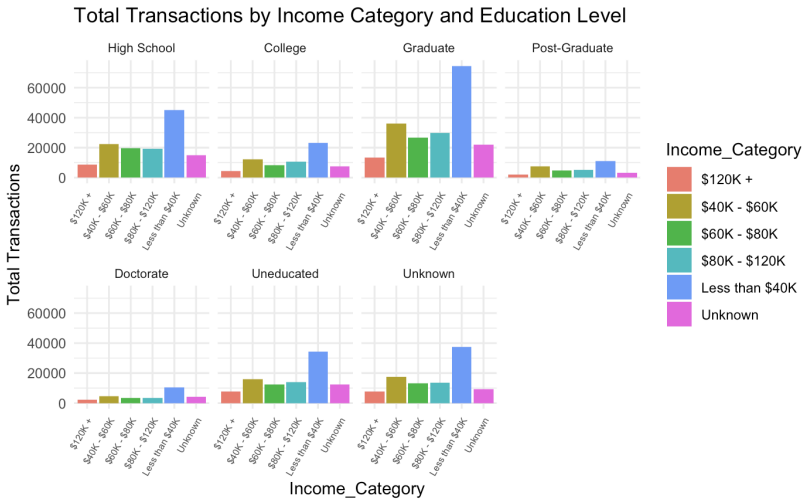
ANOVA Test: Transaction Amount and Transaction Count on Education Level

```
                    Df    Sum Sq  Mean Sq F value Pr(>F)
Education_Level      6 5.937e+07  9895758   0.857  0.525
Residuals        10120 1.168e+11 11541462
                    Df  Sum Sq Mean Sq F value Pr(>F)
Education_Level      6     942   157.0   0.285  0.944
Residuals        10120 5578095   551.2
```

To further examine the mean transaction amount and education level, we conducted an ANOVA test. Our null hypothesis is that all education levels have an equal average transaction amount, and the alternative hypothesis is that at least one group has a statistically significant difference from the other education levels. As we can see in our ANOVA test output, we obtained a p-value of 0.525, which is larger than an alpha value of 0.05, which means that we failed to reject the null. By failing to reject the null, we cannot confirm that there is a statistically significant difference between education level and mean credit card spending in the last 12 months. This is surprising as we expected there to be a difference as one could logically conclude that higher education results in higher income which could cause higher spending.

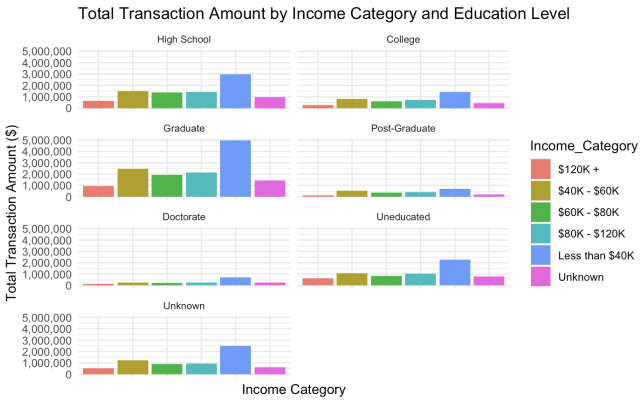
To further examine the mean transaction count and education level, we conducted an ANOVA test. Our null hypothesis is that all education levels have an equal transaction count, and the alternative hypothesis is that at least one group has a statistically significant difference from the rest. As we can see in our ANOVA test output, we obtained a p-value of 0.944, higher than 0.05, which means that we failed to reject the null. By failing to reject the null, we cannot confirm that there is a statistically significant difference between education level and mean credit card spending in the last 12 months. This is not as surprising as people across all education and income levels could be expected to make a similar amount of transactions as everyone needs the same basic necessities (i.e., food, clothing, bills). Discrepancies in income might have been seen in terms of higher spending but not more total transactions.

Faceted Bar Chart: Total Transactions by Income Category and Education Level



This is a faceted bar chart where the following education levels are shown: College, Doctorate, Graduate, High School, Post-Graduate, Uneducated, and Unknown. The y-axis illustrates the total number of transactions recorded for that category. Meanwhile, the x-axis represents each income category depending on a specific education level. For all education levels, individuals with incomes less than $40K had the highest total number of transactions, with Graduates having the highest transaction volume. Graduates stand out as having the highest number of transactions across all income categories. This may be due to a larger population of Graduates in the dataset or a high frequency of Graduates utilizing credit services across all income levels. Doctorate and Post-Graduate groups have relatively low transaction counts across the income groups, on the other hand, which may either be due to a smaller population size or conservative spending habits. They may not have to actively engage in credit transactions compared to individuals at other education levels. Based on the figure, it does not look like with the increase in education level, there's a clear trend in the increase or decrease of total transactions across different income categories.

Faceted Bar Chart: Total Transaction Amount by Income Category and Education Level

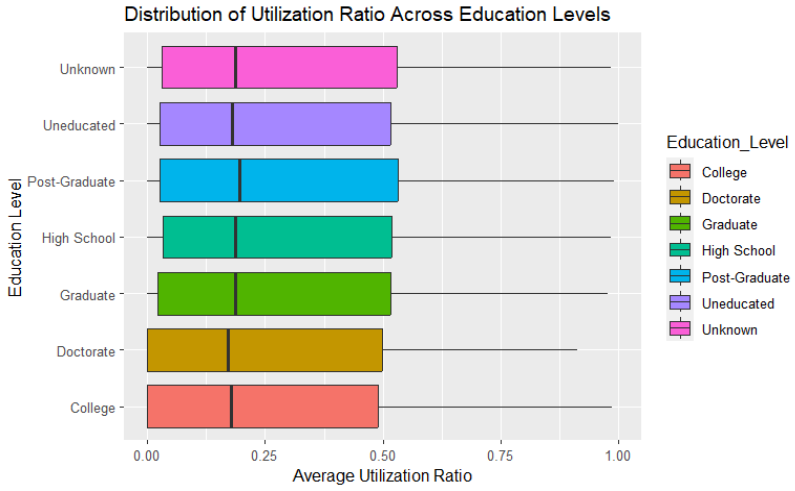


This is a faceted bar chart where the following education levels are shown: College, Doctorate, Graduate, High School, Post-Graduate, Uneducated, and Unknown. The y-axis illustrates the total transaction amounts for that category. Meanwhile, the x-axis represents each income category depending on a specific education level. Compared to the previous graph that depicted the total number of transactions per category, this figure illustrates total transaction amounts per category, which is critical when analyzing financial behavior. The Graduate level for the category less than $40K had the highest total transaction amount, $5,000,000, which illustrates the high amount of credit card use within this demographic. Individuals with a High School education level who also have a salary of less than $40K also had a notable total transaction amount, potentially indicating higher reliance on using a credit card for daily expenses. Meanwhile, people with a Doctorate degree had significantly lower transaction amounts compared to other education levels, which may indicate a small sample size or a common spending behavior that does not result in large total purchase amounts using a credit card. When observing activity in both the Unknown and Uneducated levels, individuals who have an income of less than $40K also had the highest total transaction amounts for the two categories. It does not look like, with the increase in education level, there's a clear trend in the increase or decrease of total transaction amounts across different income categories. This can indicate factors beyond education, such as local economic conditions and availability of credit, that can significantly impact financial behavior.

IV. Research Question #2: Can credit card utilization ratio be predicted by demographics or other baselines?

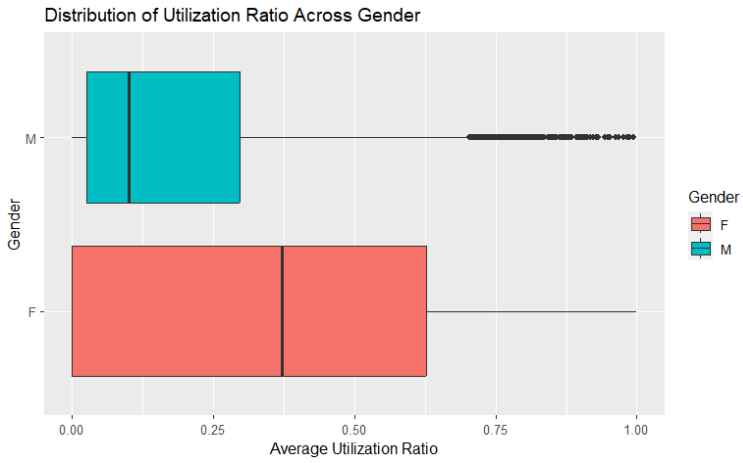
Credit card utilization, which measures the ratio of credit card balances to credit limits, provides insights into how individuals manage their available credit. High utilization rates could indicate financial strain or mismanagement, while low utilization rates might suggest responsible credit usage. The three key demographics that we identified as baseline demographics to try and predict the utilization ratio are gender, education level, and income category. In order to know if demographics are an effective way to predict utilization ratio, we must first see if there is a difference between different categories of each of the key demographics. To compare the spread of the utilization ratio of different categories in each demographic, we chose to plot three comparative box plots. This can help us compare the spreads of categories in each key demographic. Box plots are appropriate since we are not particularly interested in the shape of the spread, but simply if the spread ranges and medians are different or not.

Box Plot: Distribution of Utilization Ratio Across Education Levels



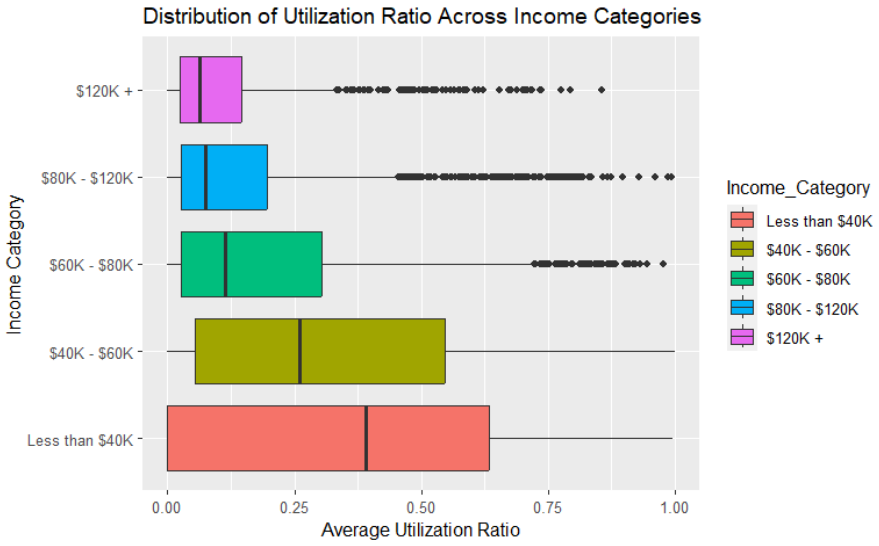
The box plots for different education levels show similar distributions of credit card utilization ratios. The medians and interquartile ranges (IQR) appear to be aligned across various education levels, indicating that education level may not significantly influence credit card utilization behavior. This suggests that individuals with different education levels tend to manage their credit card debt similarly.

Box Plot: Distribution of Utilization Ratio Across Gender



In contrast, the box plots for gender reveal a somewhat notable difference in credit card utilization ratios between males and females. Females exhibit a wider interquartile range and a higher median compared to males. This indicates that, on average, females tend to have higher credit card utilization ratios than males.

Box Plot: Distribution of Utilization Ratio Across Income Categories



The box plots reveal a notable trend across income categories, with lower income categories exhibiting both a wider range and higher median of credit card utilization ratios. This trend suggests that income level indeed plays a significant role in determining credit card utilization behavior. Individuals in lower income categories may have financial strain or financial mismanagements compared to those in higher income categories, leading to a more varied distribution of credit card utilization ratios.

To verify that education level has little to no effect on utilization ratio and that Gender and Income Category do, we ran ANOVA tests on each demographic's category.

ANOVA Test: Demographic Category

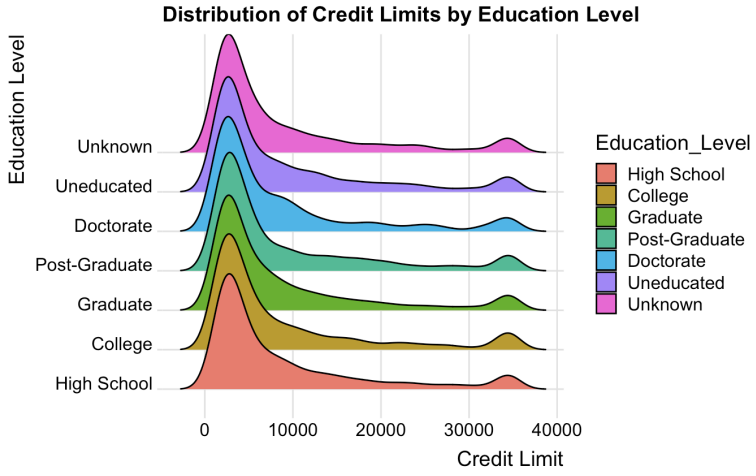
```
                 Df Sum Sq Mean Sq F value Pr(>F)
Education_Level     6    0.3 0.04464   0.587  0.741
Residuals       10120  769.4 0.07602
               Df Sum Sq Mean Sq F value Pr(>F)
Gender          1   51.2   51.17   721.1 <2e-16 ***
Residuals   10125  718.5    0.07
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                  Df Sum Sq Mean Sq F value Pr(>F)
Income_Category    5   86.3  17.256   255.6 <2e-16 ***
Residuals       10121  683.4   0.068
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of the ANOVA tests confirm and provide further statistical support for the inferences drawn from the box plots. The ANOVA test for education level shows that the p-value (0.741) is not significant ($p > 0.05$). This indicates that there is no significant difference in average credit card utilization ratios across different education levels. Thus, education level does not appear to be a significant factor

influencing credit card utilization. In contrast, the ANOVA test for gender yields a highly significant p-value ($p < 0.001$), indicating that there is a significant difference in average credit card utilization ratios between males and females. This supports the observation from the box plots that females tend to have higher credit card utilization ratios compared to males. Similarly, the ANOVA test for income category also produces a highly significant p-value ($p < 0.001$), indicating significant differences in average credit card utilization ratios across income categories. This aligns with the observation that income level appears to have a notable impact on credit card utilization behavior, as evidenced by the varied box plots.

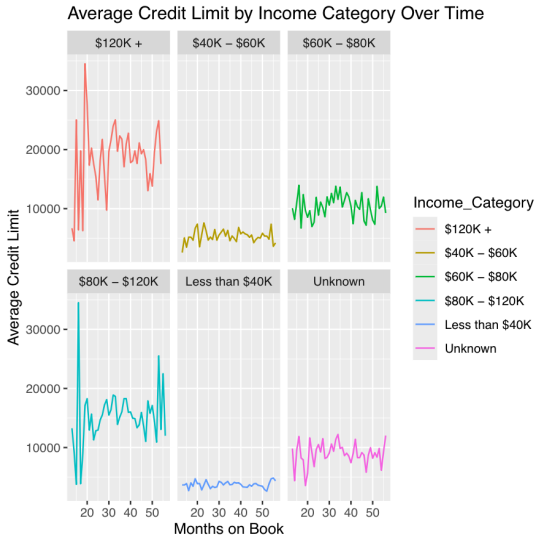
V. Research Question #3: Do people with higher education status or income have longer relationships with the bank and/or higher credit limits?

Ridgeline Plot: Credit Limit Distribution by Education Level



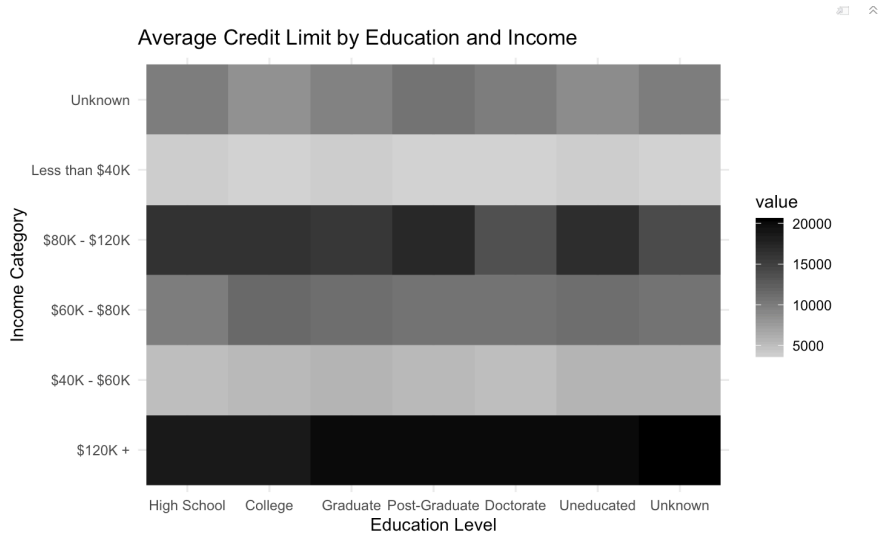
In the above Ridgeline plots, we examine how the distribution of credit card limit, on the x-axis (dollars ranging from 0 to 40000), differs between education levels, on the y-axis (College, Doctorate, Graduate, High School, Post-Graduate, Uneducated, and Unknown). Our initial hypothesis is that those with higher levels of education tend to have higher credit card purchasing limits due to a corresponding higher level of income. In examining the plots above, however, we see similar distributions across each of the education levels, as they are all unimodal and right-skewed. The mode for each of the distributions occurs around 4000 dollars, and the outliers on the right tail of the distributions each appear to peak slightly at around 35000 dollars. This finding is contrary to our preliminary assumption that individuals with higher levels of education tend to have higher spending limits.

Faceted Line Chart: Average Credit Limit by Income Over Relationship Duration



We now examine the above faceted line chart to determine the potential association between income category and average credit limit by months on book, which represents the individual's period of relationship with the bank. Each of our line charts have average credit card limit (dollars ranging from 0 to 35000) on the y-axis and months on book (ranging from 0 to 60) on the x-axis and are each grouped by income category (Unknown, Less than $40k, $40k-$60k, $60k-$80k, $80k-$120k, $120k+). Our primary hypothesis is that those with higher levels of income have higher credit card limits and that longer relationships with the bank also encourage individuals to increase credit card limits. From the charts above, we see that there generally is a trend of higher credit card limits for higher income individuals. For under $40k, the credit card spending limit range is the lowest at below 5k, for $40k-$60k, the range increases to around $3k-$7k, and for $60k-$80k, it increases further from $7k-$15k. Each of these income categories are also fairly consistent across months on book. For $80k-$120k, however, we note a higher variance in credit card limit, as the range is much wider for earlier months on the book, from $5k-$35k in the first 20 months and becoming more consistent in later months later at $10k-$25k. Finally, for $120k+, we notice the same general spike with overall more variance. Thus, our first hypothesis was correct, as higher income individuals have higher credit card spending limits. However, we noticed the opposite trend for months on book, as there was overall a consistency for spending limit regardless of relationship with the bank. This is with the exception of individuals making $80K+, however, as they are more likely to have higher spending limits in the first 20 months.

Heatmap: Relationship Duration and Credit Limit by Education and Income



We finally summarize our findings from the prior two plots into a heatmap for this research question in order to see the overall trend between education level (on the x-axis) and income level (on the y-axis), filled in with credit card limit (value shaded from light gray to black, representing 5k-20k). This graph corroborates our findings from earlier that education level does not tend to have any strong association with card limit, as the shades of gray are more or less consistent horizontally across the x-axis, signifying similar credit card limits across varying education levels. The disparity occurs when we look vertically across the y-axis, as we see that higher levels of income are associated with higher spending limits on credit cards, with income of under $40k shaded light gray across each education level (indicating card limit of around $5k) while income of $120k+ is shaded black across each education level (indicating card limit of around $20k). This demonstrates that income disparities are not necessarily correlated to the level of education, according to this dataset.

VI. Conclusion

Regarding our first question, “How does income status and education level affect credit card transaction amounts/counts?, our findings state that education level alone is not a determinant of credit card use. Income level and other factors, such as financial habits, demographic details, and personal preferences, play a larger role in credit card transactions. Hence, analyzing other datasets that explore these various factors in more detail can provide more information in assessing consumer and financial behavior.

For our second question, “Can credit card utilization ratio be predicted by demographics or other baselines?” Our analysis reveals that gender and income category significantly influence credit card utilization ratios, whereas education level does not. Females tend to have higher utilization ratios compared to males, and as income increases, utilization ratios decrease. Therefore, demographics,

particularly gender and income, serve as predictive factors for credit card utilization ratios. This could be helpful information when issuing credit to bank customers.

When answering our third question, “Do people with higher education status or income have longer relationships with the bank and higher credit limits?”, our findings suggest that while education status may not significantly impact credit card limits or the duration of banking relationships, income level plays a crucial role. Our graph also suggests that longer banking relationships do not necessarily correlate with increased credit limits for most customers, except possibly in the initial stages for higher-income earners. As a result, banks should consider these dynamics and income levels when designing their credit policies.

As mentioned earlier, examining factors such as financial habits and personal preferences can provide us with a more comprehensive understanding of credit card use. Taking into account the attitudes towards savings versus spending can also be interesting when gathering insights into consumer behavior. Hence, we can ask the following question, “How can we analyze inherent patterns and trends in credit card data to distinguish consumers with a tendency to save versus spend?” Leveraging advanced statistical techniques like predictive modeling can be used to forecast outcomes in this scenario. This enables us to conduct a deeper analysis of influential factors that influence whether an individual is likely to save or spend in a situation, making it useful for banks and other financial institutions to consider when tailoring their services to customers.