# Consumer Credit Analysis

Team: Credit Safari (Dominic, Taryn, Tiffany)

## Project Description

According to Experian, the average FICO score in the US is 715. While there’s no such thing as the average consumer, there are plenty of quantifiable trends and data that the three credit bureaus use to determine creditworthiness. In this project, we’ll examine which demographic and psychographic factors are associated with higher average credit scores and contrast them with factors associated with lower average credit scores. We’ll apply the core concepts learned in class so far, such as data cleaning and statistical analysis, to get to the bottom of how things like payment history, debt-to-income ratio, and age of accounts influence a person’s credit score.

## Research Questions:

1. What are the top 3 positive behaviors associated with high credit scores?
2. What are the bottom 3 negative behaviors associated with lower credit scores?
3. What geographic area has the highest average credit scores?

## Purpose

1. Determine if publicly available data could show positive and negative behaviors that impact credit score
2. Determine if there’s a geographic area where people with higher credit scores tend to congregate
3. Display results in a professional data analysis fashion

## Methods and Results

We started our analysis by first obtaining data sets through Kraggle.com. We ensured that each was sufficiently large to generate a statistically meaningful result. After obtaining each data set, we first searched for null or missing values and discarded them. We then plotted outliers in the data and discarded them as well.

**Datasets Used:**

a. Credit\_score.csv

b. loan\_data\_2015.csv

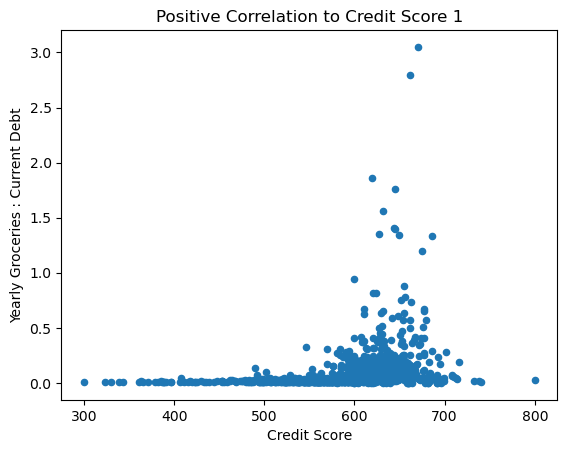
c. train.csv

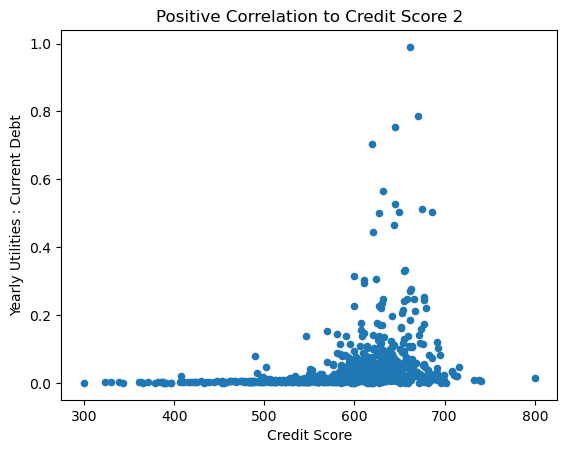
Once the data was clean, we performed a linear correlation analysis on 40 different dimensions to determine what, if any, correlation they had with credit scores. These dimensions included debt-to-income ratio, yearly expenditures, yearly savings, and more financial information. This correlation analysis yielded our top three and bottom three factors that influence credit scores:

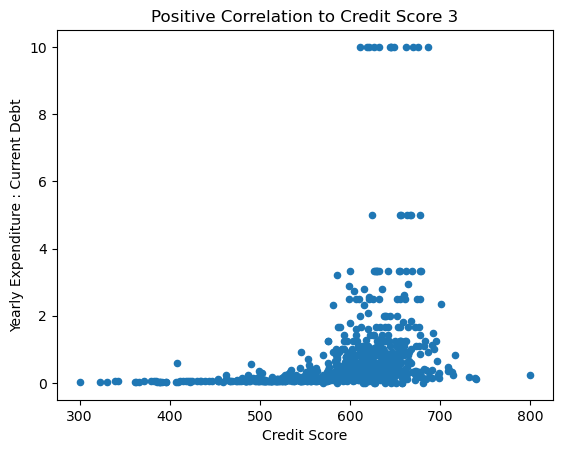
**Strongest Positive Corr**

| **Yearly Groceries: Current Debt** | 0.286999 |
| --- | --- |
| **Yearly Expenditures: Current Debt** | 0.286999 |
| **Yearly Utilities: Current Debt** | 0.303482 |

From this result, we can see that the ratio of one's yearly spending on groceries, utilities, and general expenditures to current debt has a positive correlation with higher credit scores. That is to say that people with a higher ratio tend to have a higher credit score, while those with a lower ratio tend to have a lower credit score. This is illustrated further in the following graphs:



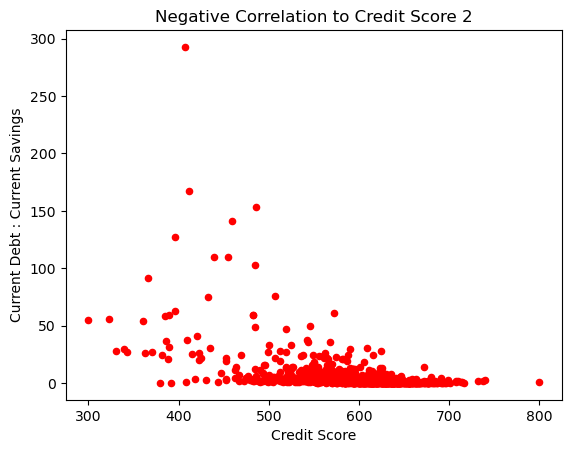
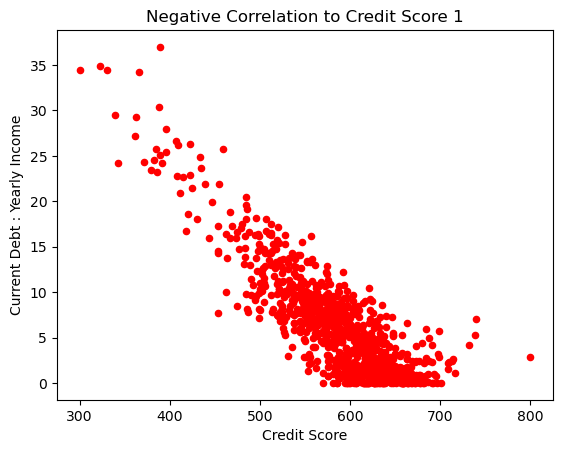


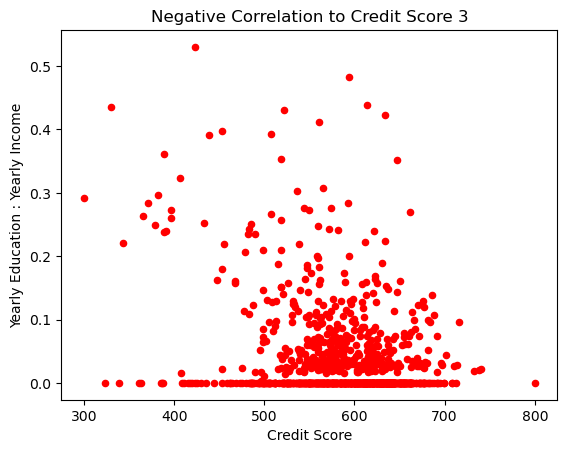
****

**Strongest Negative Corr**

| **Current Debt: Yearly Income** | -0.857900 |
| --- | --- |
| **Current Debt: Current Savings** | -0.454068 |
| **Yearly Education: Yearly Income** | -0.303268 |

On the other side of the coin, the three factors with the strongest negative correlation to credit scores are the current debt-to-yearly income ratio, current debt-to-savings ratio, and yearly education-to-income ratio. People with lower credit scores tend to have higher or less favorable ratios than those with higher ones.

****



Moving to our next research question, we performed a Spearman correlation analysis on income to credit score and debt to income to credit score ratios to factor in non-linear correlations. In order to project credit score quintiles from smaller dataset to larger data set, we first took the results of the Spearman analysis to use debt to income as a means to project credit score quintiles onto larger dataset. We then took summary statistics of the D-to-I quintiles, then took mean and standard deviations of each quintile. Once we determined there was overlap, we assumed the means were the maxes of each quintile.

As a result of this analysis, we discovered that Arkansas, New York, and Washington DC are the geographic areas with the highest average credit score.

## Conclusion

Our analysis of credit score data uncovered the three most prominent positive and negative factors that affect credit scores. From these factors, we can conclude that things like having a favorable debt to income ratio and a favorable debt to savings ratio are more common in people with higher credit scores. As a whole, people with higher credit scores tend to congregate in geographic regions that include Arkansas, New York, and Washington DC.