**Credit Card Customer Churn Analysis  
Team Members:** William Cutrone, Edyson Katuuo, Subhashree Mallick, Jessica Hunter

Our objective was to evaluate credit card customer data to discover trends that may or may not cause people to leave their credit card service. Customer churn is one of the most important and challenging problems for businesses such as credit card companies and service providers worldwide. Understanding customer churn metrics can help businesses be able to predict customers at risk for churning and improve customer retention. Some common reasons for customer churn include lack of usage, poor service, and ability to find a better price elsewhere. Regardless of the reasoning, it costs more to acquire new customers than it does to retain existing ones. This has a direct impact on operating costs and marketing budgets within the company. Because of the high importance of better understanding customer churn within a business, businesses are investing more time and effort in finding out the reasons for churn within their organizations, how they can better predict churn, and what they can do to minimize the customer churn. Our project will explore some consumer data and see how we can leverage data insights and predictive modeling in order to better understand churning and improve customer retention.

To carry out our project, we used Python, Pandas and Matplotlib to analyze our data. We started with a CSV file called “bank\_churners.csv” found on Kaggle.com. We cleaned the data by removing all customers with a zero utilization ratio as keeping these customers led to misrepresentations of the data. After this, we deleted two unnecessary columns. Lastly, we took the first, large data-frame which contained all customers and split that into two smaller data-frames, one with the churning customers and one with the existing customers. The original data-frame had about 10,000 total customers. The churning customers made up a little over 1,600 of these total customers, accounting for about 16% of the customers in the dataset.

After studying the data, we realized there are two major questions to answer. The first is: What trends may or may not be associated with customer churn? And the second is: “What does the profile look like for a customer who is at risk for churning? To answer these questions, we made either a boxplot or a bar-graph for each of the variables in the dataset, which are represented by the columns in the data-frame. We made bar-graphs for the categorical variables and box-plots for the quantitative variables. For the quantitative variables, we also performed the independent t-test for each one to see if the means between the two groups (existing and churning customers) was statistically significant. The variables that we looked at were age, gender, income category, education level, dependent count, months spent with the credit card company, card type, total months inactive, total revolving balance, monthly credit limit, total transaction amount and total transaction count.

**Question 1: Which trends may or may not be associated with customer churn?**

Out of this long list of variables, the plots and statistical analysis showed that only three of these variables produced a statistically significant difference between the two types of customers. These three variables were months inactive (out of 1 year), total transaction amount and total transaction count. For the variable months inactive, the mean for the existing customers was 2.27 and the median was 2. For the churning customers, the mean was 2.73 and the median was 3. The p-value for months inactive was 2.3e-35. In addition , when making a histogram to see the distribution of the groups, the plot represented a “semi” normal distribution (it was slightly skewed left for both types of group). For total transaction amount the mean and median values for the existing customers was $4861.27 and $4,081 respectively, and the churning customers had values of $3,148.97 and $2,361 respectively. The p-value for these groups was 9.9e-51. However, when making the histogram for this variable, the plot did not look anything like a normal distribution. In fact, there were two distinct peaks, one at the low end of spending and one at the high end of spending. For total transaction count, the mean and median values for existing customers was 68 and 70. And for the churning customers, those values were 45 and 43 respectively. And finally the p-value for this variable was 6.9e-212. And for this histogram, the plot looked very much like a typical normal distribution (bell-shaped curve). After learning which values had statistically significant effects, it made sense as to why those variables stood out. It makes sense that customers who spend less money on their card, make fewer transactions, and have longer periods of inactivity are more likely to discontinue their credit card.

**Question 2: What is does the profile look like for a customer who is at risk for churning?**

After discovering which factors were statistically significant, we then felt like we were able to determine what the profile of a churner would look like. We concluded that a customer who is at significant risk for churning would have a total transaction count of 51 or less, and have a months inactive value of 3 or greater. To determine the cutoffs, we decided to base it upon the 0.75 quartiles for both total transaction count and months inactive. We thought that this value would make sense as a cutoff point as it would capture many of the customers who are contained in those variables. As for total transaction amount, we decided to not make a prediction based of this variable due to the appearance of the histogram. Due to the histogram being completely unlike any normal distribution, we thought that it would not make sense to make predictions. Therefore, we decided to make our profile and our predictions based off of only total transaction count and number of months inactive. Finally, when we went back to the data-frame of existing customers and went to look for customers who met both parameters for transaction count and months inactive, we found that 714 (or about 1%) of the existing customers were at high risk for churning. Therefore, if we were asked to make a recommendation to the credit card company of which clients to be most concerned about churning, we would identify the 1% of the customers we found that met our determined profile.