**Course 2 Task 2: Lessons Learned Report – Michelle Giniewicz**

I began my initial analysis of the Credit One data by looking at the customer demographics within this dataset of almost 30K customer records. Most customers in this dataset are female (60%) and most customers are single (53%) followed by married (46%). Additionally, most of the customers have a university education level (47%) and are between 21-33 years of age (48%). Finally, most of the customers in this dataset have a limit balance between $100K-$250K (36%) followed by a limit balance of $20K-$60K (24%).

After performing Exploratory Data Analysis (EDA) on the Credit One data, we see that 22% of Credit One customers will default on their loans. After analyzing customer demographics of those who default, we can see that half of all customers who default are Single and half of all customers who default have University education. However, as noted above, these groups make up a higher proportion of the overall dataset. Therefore, without additional analysis, it is not currently possible to determine if there is statistical significance in the default rates of customers who are Single and/or have University education.

**Customer Marital Status grouped by Sex and Default**

A picture containing chart

Description automatically generated

**Customer Education grouped by Sex and Default**

Chart, bar chart

Description automatically generated

The main lessons that I learned from is experience is that you cannot just look at counts of data to determine if a customer group is more likely to default. Instead, you should look at the proportions of the overall dataset. For example, female customers make up 12.6% of the 22% that default. It may initially seem that female customers are more likely to default than male customers. However, when we look at the overall proportion of customer sex without our dataset, we can see that 60% of the customers are female, so this seems to be as expected.

Table

Description automatically generated

I also learned that the data features about bill and pay amounts are not relevant at this stage of analysis. It seems that the only features that were potentially significant in my initial analysis were Sex, Age, Education, Marital Status, and Limit Balance. However, it is possible that through more detailed analysis, these features may prove to be valuable in determining whether a customer will default or not.

Finally, I would recommend that as a next step, we build models to try to determine which of these features can determine whether a customer will default on their loans or not. For example, a model can help predict if single customers or customers with university education level are really more likely to default on their loans, or if there are other features that are more likely to determine default rates. We can then use these models to verify whether we should approve future customers for loans or not.