Course 2 Task 2 --- Lessons Learned Report

This document will deal with the following:

Potential business value gained from this analysis

Main lessons learned

Recommendations based on findings

The data of this project included:

30,000 entries

Credit limit balances (between $10,000 and $1,000,000)

Marriage status (single, divorced, married, etc.)

Age (between 21 and 79)

Education level (university, high school, graduate, etc.)

Sex

Payment statuses between the months of April to September (see methodology document)

Bill amount between the months of April to September (see methodology document)

Payment amount between the months of April to September (see methodology document)

Loan result (default or non-default)

**Business Value Gained from Analysis** (in no order)

1. The average age of defaulters were 36 years old.
2. Nearly all of the defaulters were either married or single, which no major difference between the two. Divorce and “other” had the lowest default rates.
3. No clear gender-bias for defaulters, however women defaulted slightly more (about 13% more than that of men)
4. 50% of defaulters had university education, while only 19% and 31% of defaulters had high school and graduate school education, respectively
5. Regardless of month, the majority of defaulters had a payment status of 0 (revolving credit). However, in the month of September, payment status of 2 (payment delay of two months) nearly exceeded that of payment status of 0. In other words, the month of September is very important as it relates to whether or not someone will default.
6. 91% of defaulters were between the ages of 21 to 50, with 69% of defaulters being between the ages of 21 and 40. In other words, younger people tended to default more while older people did not. In fact, people between the ages of 61 and 80 had the lowest default rate of just about 1%.
7. Credit limits most associated with defaulting fell between $9,000 and $448,000. This means that people with balance limits within this range may be potentially more linked to defaulting.

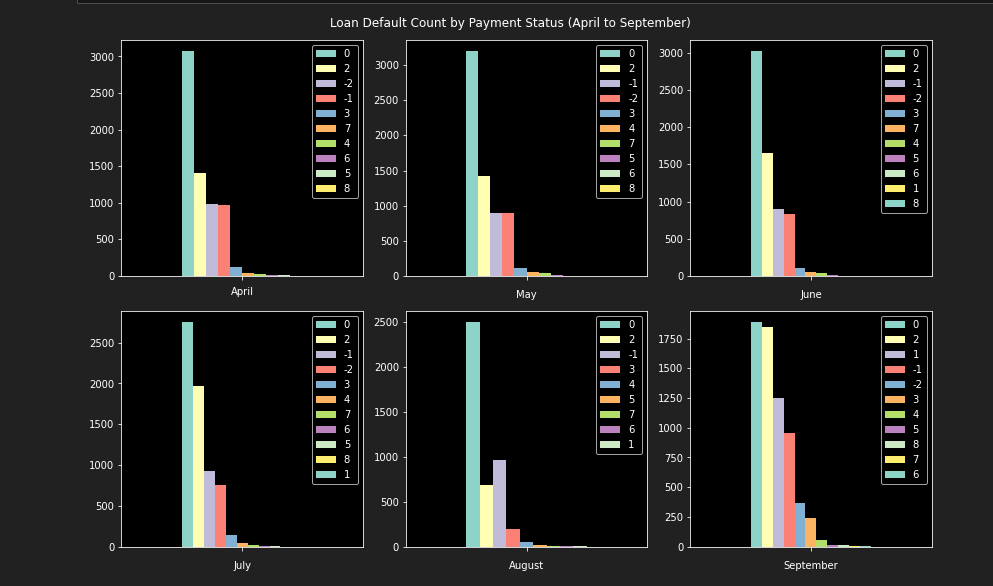
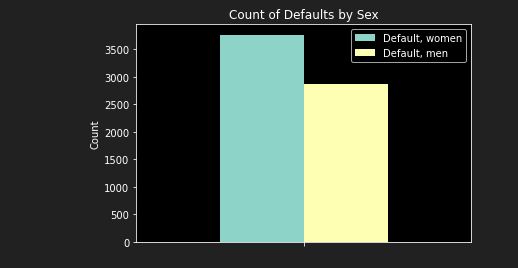
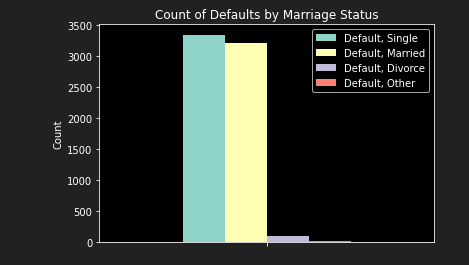
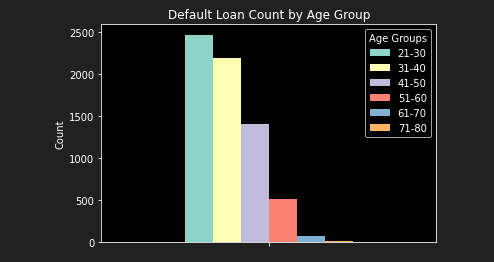
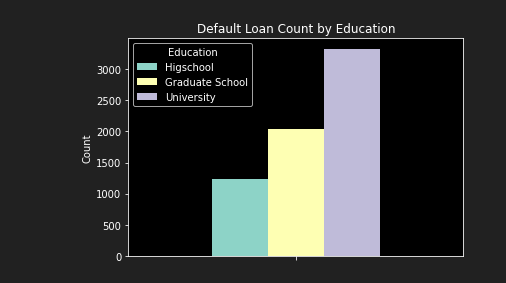
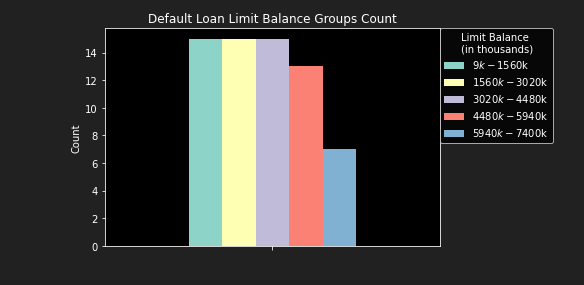
**Main Lessoned Learned** (in no order)

1. Majority (79%) of data features non-defaulters compared to just 21% of defaulters. In the future, it would be better for the data to have less of a stark difference. It would be better to have more data regarding defaulters for the sake of potentially observing more patterns.
2. Data cleaning is essential. The original data from the MySQL database was very messy, which makes it hard to properly analyze the data for any patterns and insights. So, in the future, improving the data collection process will improve business outcomes by making the process of analysis far more streamlined.

**Recommendations based on Findings**

1. Improve the data collection/entry process. Try to ensure that there are no duplicate rows or illegible/too complicated headers. Try to include more relevant data, such as employment status, income, or tax bracket. These three variables potentially may be insightful in helping to ascertain whether or not someone could default on a loan because they relate, at least somewhat, to whether or not someone has the financial means to make consistent payments.

**Visualizations**

****