IC22029 – Paycheck Protection Program

Team Members:

Po-Han Yen, Hsin Chen, Jinping Guo, Jing Lin

Introduction:

In this project, our team aims to analyze the Paycheck Protection Program (PPP) data provided by U.S. Small Business Administration (SBA). On the purpose of finding useful insights about why certain loan applications have been removed from the program. Additionally, we try to build predictive models to predict whether or not a loan application would be removed based on available information.

Problem statement:

Why cerain loan applications of the PPP program have been removed?

Can we predict whether or not a loan application would be removed?

Procedures:

1. Research and understand the background & the data of the program.

2. Clean the data:

Drop features that will not be used for our analysis.

Impute appropriate values for missing values.

Convert data to appropriate types.

3. Add additional salary data for our analysis.

4. Perform exploratory data analysis.

5. Perform explanatory data analysis.

6. Build predictive models.

7. Communicate findings and display final product.

Results & Conclusions:

1. Single Member LLC & Sole Proprietorship are most likely to be removed compare to other business types and the lower the jobs retained by the business the more likely for them to be removed, especially those with single employee.

2. Applications approved at the end of the program are three to fifteen times more likely to be removed than those approved earlier.

3. South-west districts are more likely to be removed than northern districts.

4. Applications that have servicing lenders located at Massachusetts have a remove rate of 79% which is at least 6 times more than those of other states.

Applications that have originating lenders located at Massachusetts have a remove rate of 43% which is at least 3 times more than those of other states.

Applications that have lender-Leader bank are most likely to be removed.

5. Mining, quarrying, and oil and gas extraction industries are twice more likely to be removed compare to other industries.

6. If the applied amount per employee is higher than the industrial and regional average salary, the application is twice more likely to be removed. Additionally, as the difference increases the remove rate increases.

7. Predictive models for small or large amount application respectively.