PS6$_Z$hang

Haotian Zhang

March 2025

1 Data Clean

First, I replace all the "#N/A" in the data set with `NA`, and I used the `na.omit` function in R to clean all the `NA` values in my data set. Then, I deleted the column that I am not going to use. After that, I also distinguished the income levels and transfered the education levels in the data to facilitate classification and visualization. I also added latitude and longitude to the dataset based on the postcode
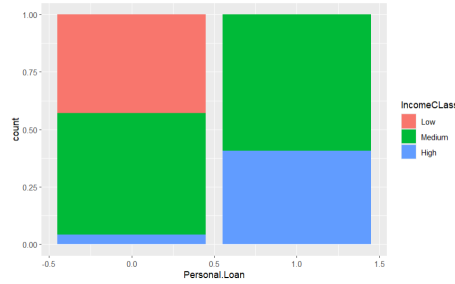
2 Visualization

Figure 1: Relationship Between Income Class and Personal Loan Acceptance

The stacked bar chart illustrates the relationship between income class and personal loan acceptance. It reveals that individuals in the low-income class are far less likely to accept a personal loan, as indicated by their absence in the loan-accepting group. In contrast, the medium-income class constitutes the largest proportion in both groups, suggesting that individuals in this category may have a balanced financial capacity that makes them more open to taking loans. Additionally, the high-income class has a significantly larger share among those who accepted the loan compared to those who did not, indicating that individuals with higher incomes may have greater confidence in their ability to repay loans or may qualify for better loan terms. This pattern suggests that

income level plays a crucial role in personal loan decisions, with higher-income individuals being more inclined to take advantage of such financial opportunities.
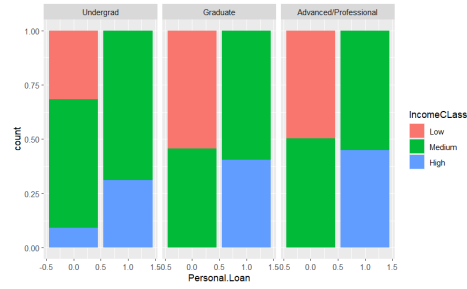

Figure 2: Education, Income, and Loan Acceptance

The bar chart illustrates the relationship between education level, income class, and personal loan acceptance. The data is grouped into three educational categories: Undergraduate, Graduate, and Advanced/Professional. Across all education levels, individuals in the medium-income class (green) form the majority of both those who accept and decline personal loans. The low-income class (red) is more prevalent among those who do not take personal loans, particularly at the undergraduate level, indicating that financial constraints may be a limiting factor for loan acceptance. The high-income class (blue) increases in proportion among loan acceptors across all education levels, suggesting that higher-income individuals are more inclined to take loans regardless of their educational background. Additionally, as education level increases, the distribution of income classes remains relatively stable, implying that income level, rather than education, is the primary factor influencing personal loan decisions.
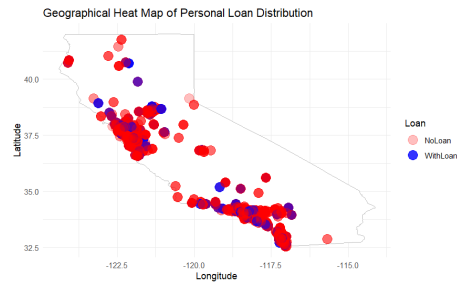

Figure 3: Geographical Loan Distribution in California

The geographical heat map illustrates the distribution of personal loans across different locations in California. The red points represent individuals who did not take a loan (NoLoan), while the blue points indicate those who accepted a personal loan (WithLoan). The clustering of red and blue points suggests that loan decisions are concentrated in certain regions, with a notable

density in metropolitan areas such as the San Francisco Bay Area, Los Angeles, and San Diego.