**Problem Statement**

The goal of this project is to analyze the "Credit Card Approval - With Target" dataset and identify the factors that affect credit card payments. By examining various attributes such as annual income, job type, number of children, car ownership, house ownership, and employment, we aim to determine which characteristics are associated with a higher likelihood of timely credit card payments. The insights gained from this analysis can be utilized by banks to make informed decisions regarding credit card approvals.

**Description/Outline**

The project involves analyzing the "Credit Card Approval - With Target" dataset, which contains information about individuals' attributes and their credit card payment status. The dataset includes variables such as annual income, job type, number of children, car ownership, house ownership, and employment. The objective is to investigate the relationship between these factors and the payment status of credit cards.

By examining the dataset, we will explore patterns and trends in the data to identify factors that have a significant impact on credit card payments. This analysis will involve data visualization techniques, such as plotting different attributes against the payment status, to uncover any correlations or trends. The ultimate aim is to provide insights to banks on the characteristics of individuals who are more likely to make timely credit card payments.

**Research Questions to Analyze**

1. Does age or gender have an influence on the credit card loan payment?

2. Does owning a car or house have an influence on the credit card loan payment?

3. Does having children have an influence on the credit card loan payment?

4. Does being employed have an influence on the credit card loan payment?

5. Does job type and annual income have an influence on the credit card loan payment?

By answering these research questions, we can gain a better understanding of the factors that influence credit card payments and provide valuable insights to banks for their credit card approval processes.

All the factors were analyzed against credit status provided in the dataset set.

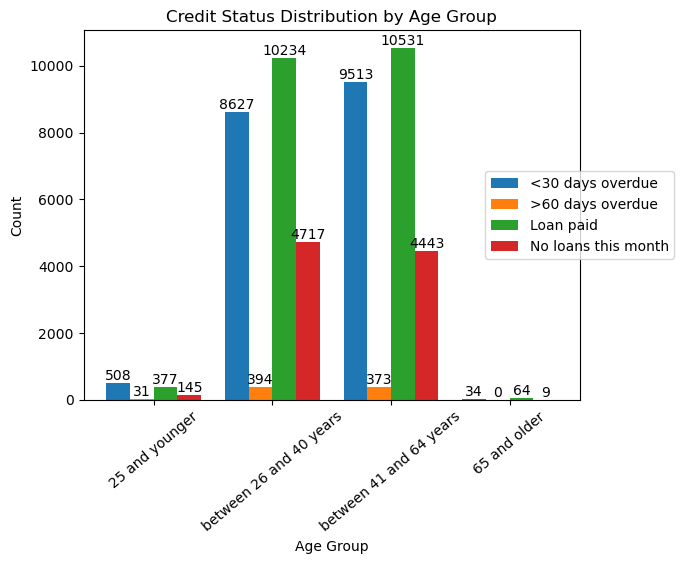
Credit Status: The dataset contained 4 categories of credit status - <30 days overdue, >60 days overdue, loans paid, no loans this month.

**Age and Gender**

*Age*

To understand the relationship between age and credit status, the dataset was further divided into four age groups categories:

* 25 and under
* Between 26 and 40 years
* Between 41 and 64 years
* 65 and older



Analysis: Overall, very few people from all age groups have low numbers who fall under the “>60 days overdue category” with noone in the age group 65 and older falling under this category. Most significant trend is seen in ‘loan paid’ credit status where age groups “between 26 and 40 years” and “between 41 and 64 years” have the highest number of people in this category compared to age “25 and younger” have more number of people in ‘<30 days overdue’ compared to loan paid category.

Statistical test: Conducted chi-square test to test the following hypothesis

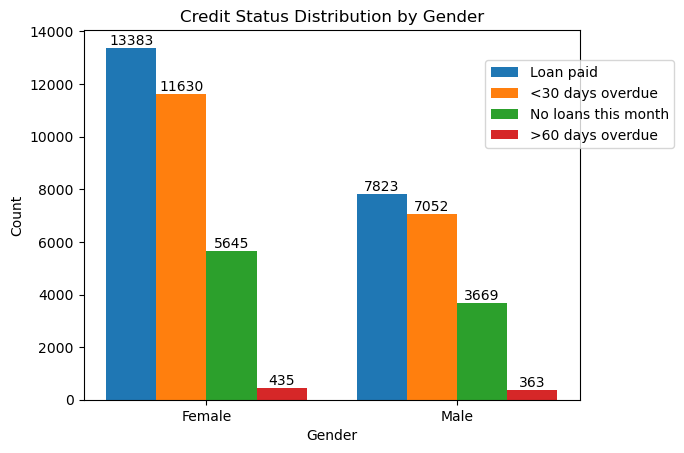
Null hypothesis: There is no relationship between age and credit card payment status

Alternative hypothesis: There is a relationship between age and credit card payment status

Result: Reject the Null Hypothesis (p value is less than 0.05 for all age groups)

*Gender*

To understand the relationship between gender and credit status, the dataset was analyzed between male and female groups.

**

Analysis: While the credit status between the male and female gender groups are similar, ‘loan paid’ is the highest, followed by ‘<30 days overdue’, ‘no loans this month’ and ‘>60 days overdue’, respectively. However, the difference between the ‘<30 days overdue’ and ‘no loans this month’ vary greatly between the two gender groups.

Statistical test: Conducted chi-square test to test the following hypothesis

Null hypothesis: There is no relationship between gender and credit card payment status

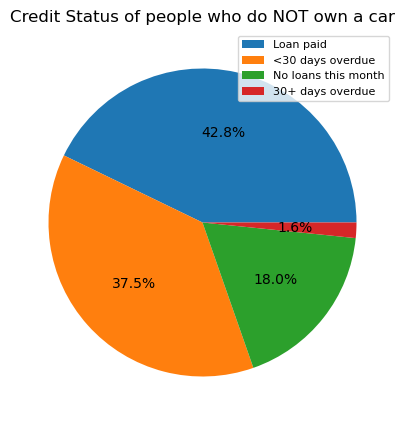
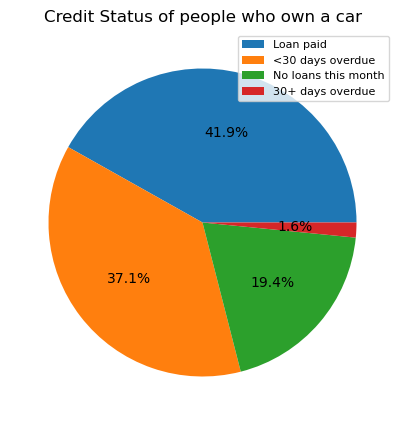
Alternative hypothesis: There is a relationship between gender and credit card payment status

Result: Reject the Null Hypothesis (p value is less than 0.05 for all gender)

**Owning a car and a house**

*Owning a car:* To understand the relationship between credit status and the following:

* People who own car(s)
* People who do not own car(s)



Analysis: Similar patterns were observed in both ‘people who own a car’ and ‘people who do not own a car’

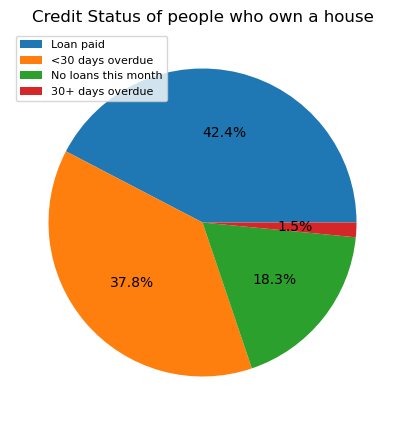
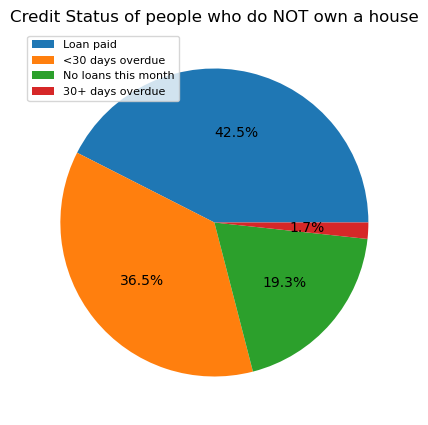
Statistical test: Conducted Chi-squared test on the percentage of Credit Status of car owners and non-car owners to test the following hypothesis

Null hypothesis: There is no relation in Credit Status of car and non car owners

Alternative hypothesis: There is a relation in Credit Status of car and non car owners Result: Accept the Null Hypothesis (Critical value > Chi-square statistic and p-value close to 1)

*Owning a house:* To understand the relationship between credit status and the following:

* People who own house(s)
* People who do not own house(s)



Analysis: Similar patterns were observed in both ‘people who own a house’ and ‘people who do not own a house’

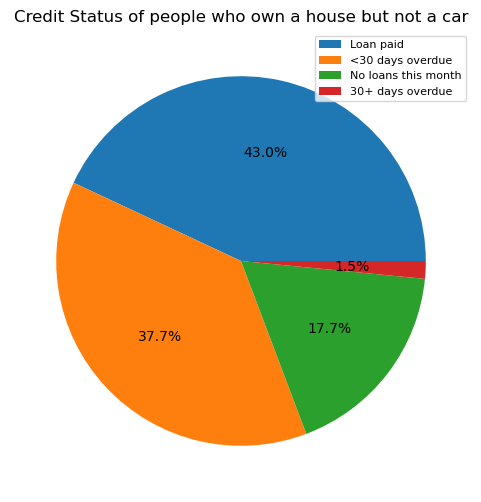
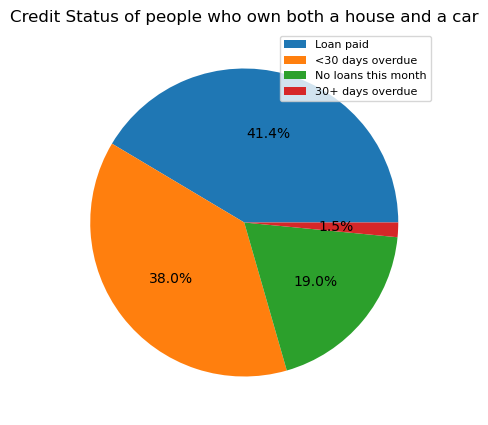
Statistical test: Conducted Chi-squared test on the proportion of Credit Status of homeowners and non-homeowners to test the following hypothesis

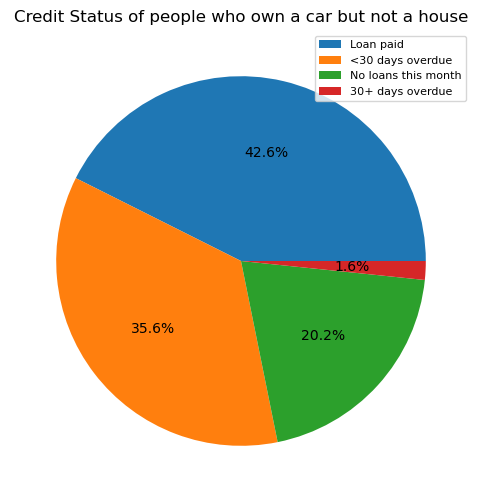
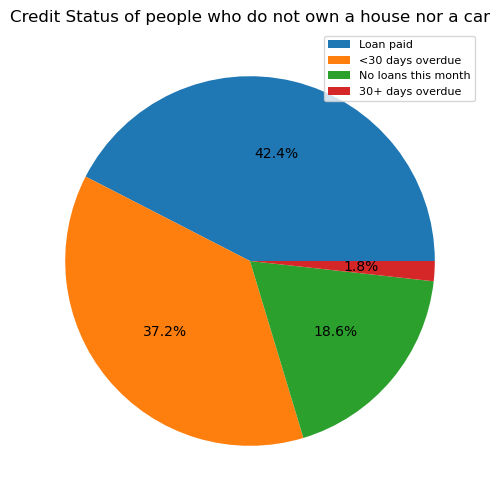
Null hypothesis: There is no relation in Credit Status of homeowners and non-homeowners

Alternative hypothesis: There is a relation in Credit Status of homeowners and non-homeowners Result: Accept the Null Hypothesis (Critical value > Chi-square statistic and p-value close to 1)

*Owning a car and a house:* To understand the relationship between credit status and the following:

* People who own car(s) and own house(s)
* People who do not own car(s) nor own house(s)





Analysis: Similar patterns were observed in both all the categories as seen in the above pie charts

Statistical test: Conducted Chi-squared test on the proportion of Credit Status of car and homeowners and non-car and homeowners to test the following hypothesis:

Null hypothesis: There is no relation in Credit Status of car and homeowners and non car and homeowners

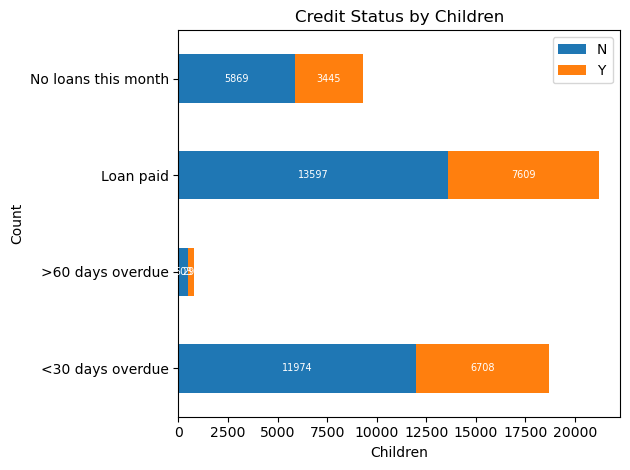
Alternative hypothesis: There is a relation in Credit Status of car and homeowners and non car and homeowners

Result: Accept the Null Hypothesis (Critical value > Chi-square statistic and p-value close to 1)

**Having children**

To understand the relationship between credit status and the following:

* People who have children
* People who do not have children



Analysis:The values in the DataFrame represent the counts or frequencies of each combination of 'Children' and 'Credit Status'. Based on the values provided, it seems that there are two levels for the 'Children' variable: 'N' and 'Y', which likely represent 'No children' and 'Children' respectively. In this analysis, it can be said that people who have no children have more percentage than others who paid their loans.

Statistical test: Conducted chi-square test to the following hypothesis:

Null hypothesis: There is no relation between credit status and having children

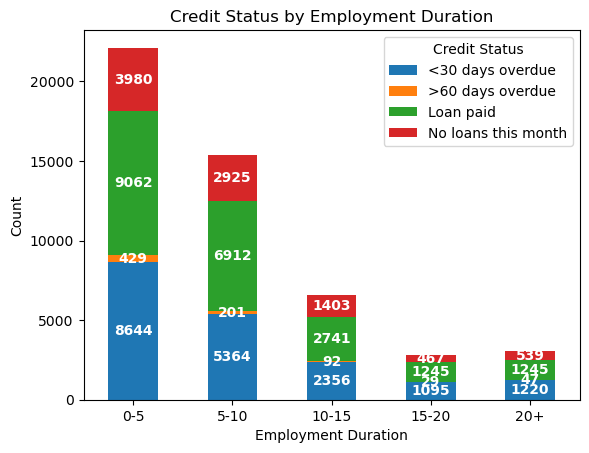
Alternative hypothesis: There is a relationship between credit status and having children

Result: Accept the Null Hypothesis (p-value >0.05)

**Being employed**

To understand the relationship between credit status and the following:

* People who are employed
* People who are not employed



Analysis: It has been analyzed that the relationship between credit status and employment duration using the provided dataset. Here are the key insights:

1.Default Rates:

•Default rates decrease as employment duration increases.

•Individuals with longer employment durations (20+ years) have the lowest default rates.

2.Loan Repayment:

•The likelihood of successful loan repayment increases with longer employment durations.

•Individuals with shorter employment durations (0-5 years) have higher loan repayment rates.

3.No Loans:

•The count of individuals without loans varies across different employment duration ranges.

•The number of individuals without loans generally increases with longer employment durations.

Statistical test: Conducted chi-square test to the following hypothesis

Null hypothesis: There is no relation between years employed and credit status

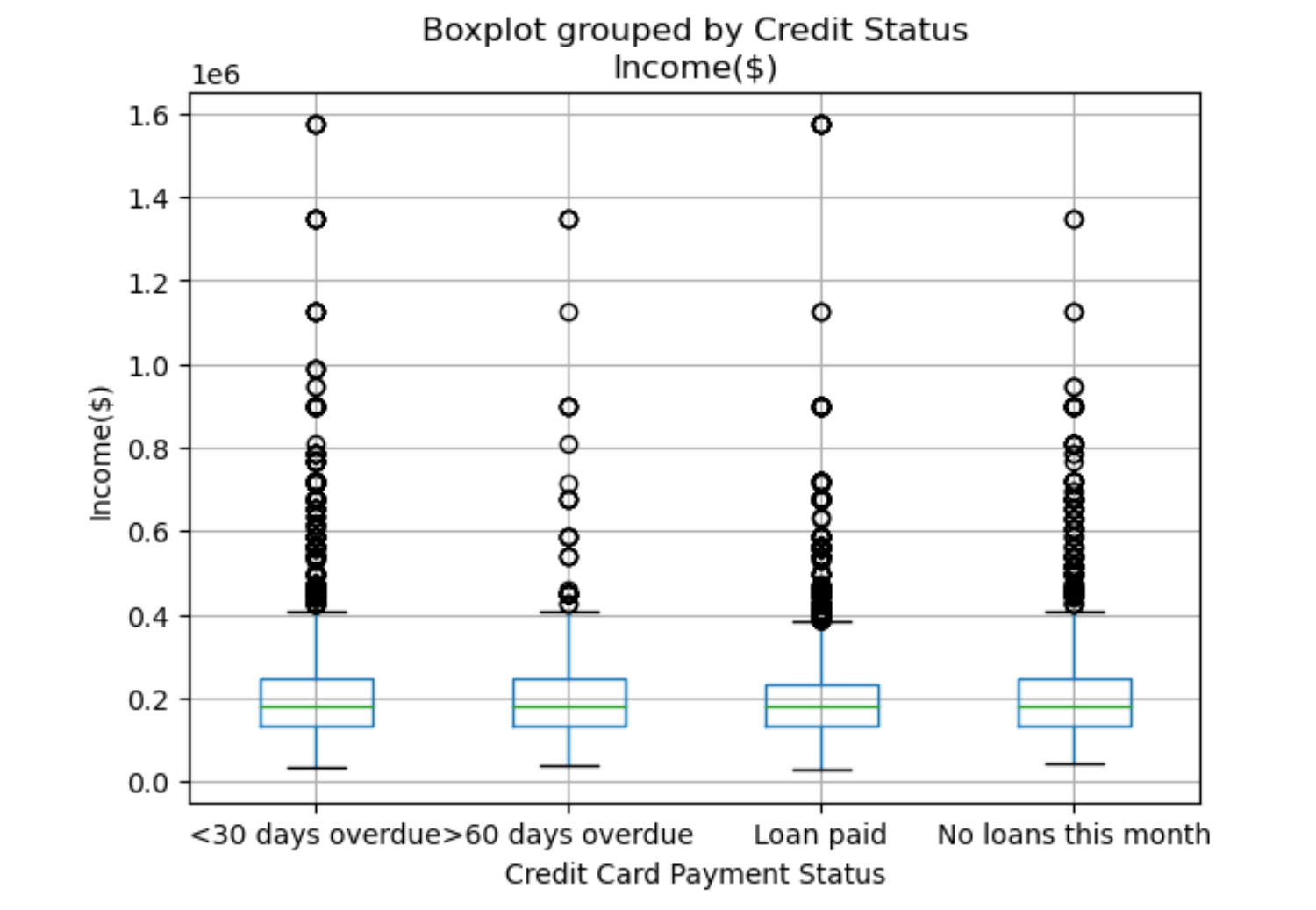
Alternative hypothesis: There is a relation between years employed and credit status

Result: Reject the Null Hypothesis (p-value <0.05)

**Job type and annual income**

*Annual income*

To understand the relationship between credit status and the various income levels



Analysis: As observed this data has lots of outliers with most medians in the $190,000 income per year. Although highest incomes are both spread among the over 30 days overdue and loan paid, less than 60 days and no loans this month has many outliers as well. The median, max and min for all 4 payment status are in a similar range.

Statistical test: Conducted Chi-square test to test the following hypothesis

Null hypothesis: There is no relation between income and credit card payment

Alternative hypothesis: There is a relation between income and credit card payment

Result: Reject the Null Hypothesis since p value is less than 0.05 for all income levels. 9.538811367010588e-160<0.05

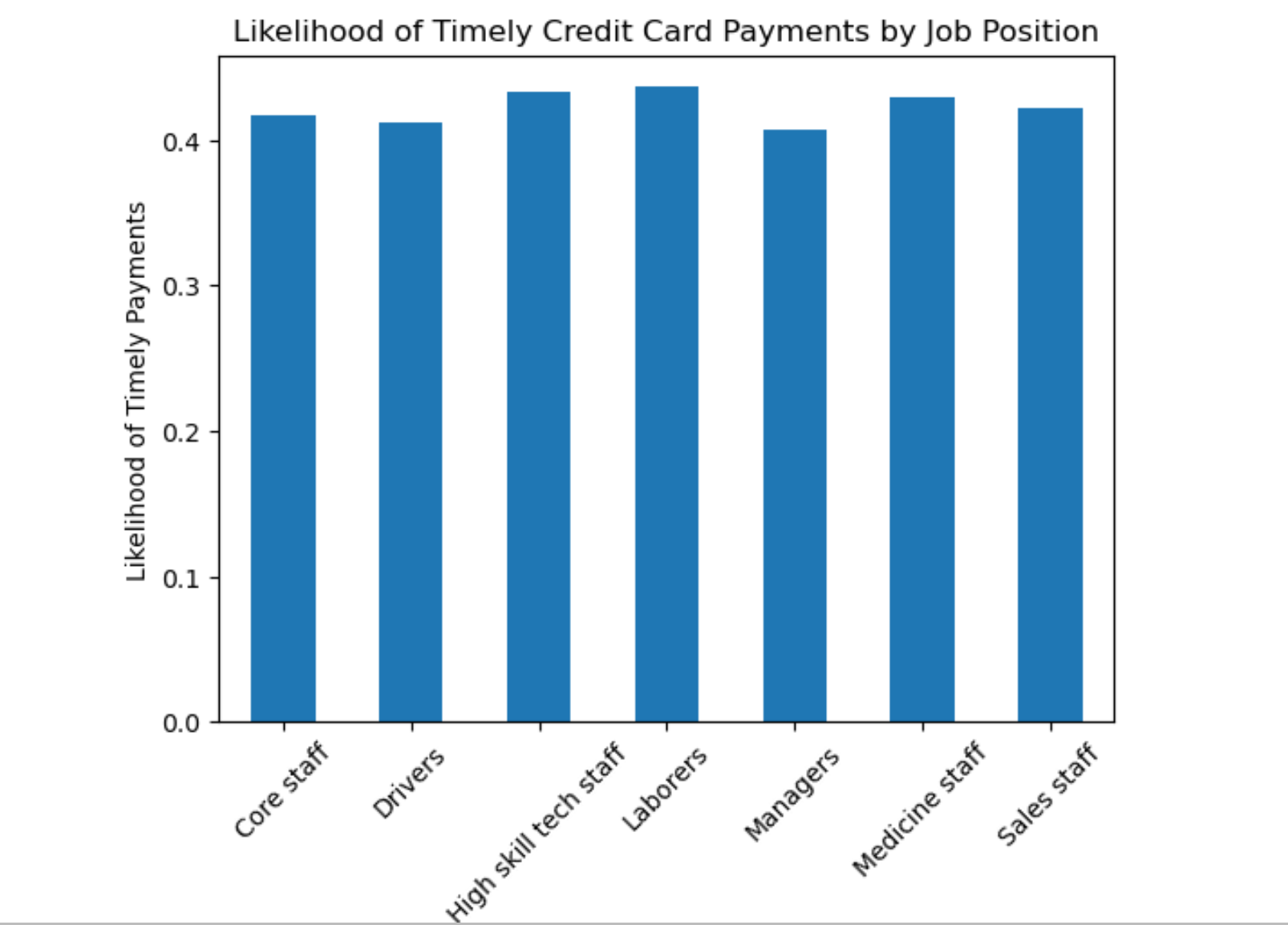
*Job type*

To understand the relationship between credit status and various income levels and various job types:

* Core staff
* Drivers
* High skill tech staff
* Labourers
* Managers
* Medicine staff
* Sales staff

Vs

* Loan paid



Analysis: This graph was done by looking into the percentage of likelihood of loan paid based on the job type. Based on this analysis it can be observed that, although all the positions over the 40% change of loan paid, managers have the lowest with 40.7%, meanwhile, laborers had the highest, with 43.6%. For this analysis, to simplify the analysis and have more concise data, the 7 jobs with the most participants were chosen, laborers were at the highest with over 12,000. Some limitations found in this data was the lack of job description, therefore causing some issues understanding the nature of each job.

Statistical test: Conducted Chi-square test to test the following hypothesis

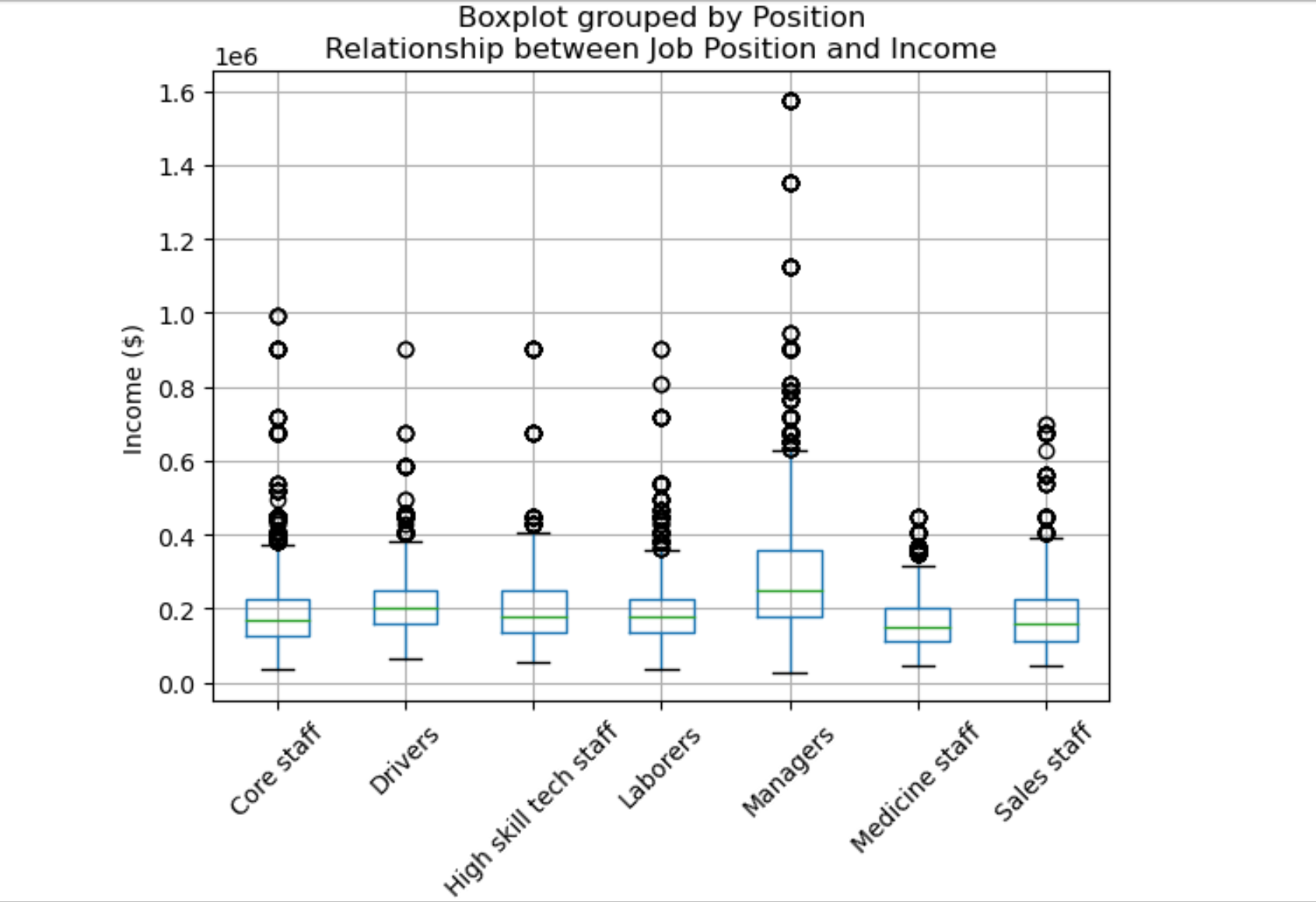
Null hypothesis: There is no relation between job type and timely payments

Alternative hypothesis: There is a relation between job type and timely payment

Result: Reject the Null Hypothesis since p value is less than 0.05 for all job types. 9.88292888262165e-07<0.05

Income vs Job

To understand the relationship between income and top 7 job types.



Analysis: this box plot was created to observe which positions had the highest income and to compare to the previous bar graph. As observed, managers have the highest median and outliers but were also the positions with the least percentage of loan paid. Managers also had the biggest outliers and the biggest maximum and minimums.

Conducted Chi-square test to test the following hypothesis:

Null hypothesis: There is no relation between job type and annual income

Alternative hypothesis: there is a relation between job type and annual income

Result: Reject the Null Hypothesis since p value is less than 0.05 for all job types and income levels). 0.0<0.05.

***Reference***

https://www.kaggle.com/datasets/laotse/credit-card-approval