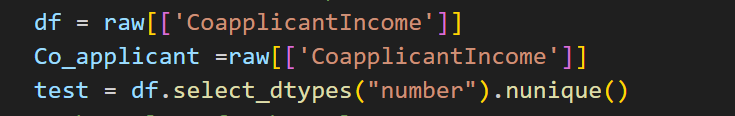
MLG 382

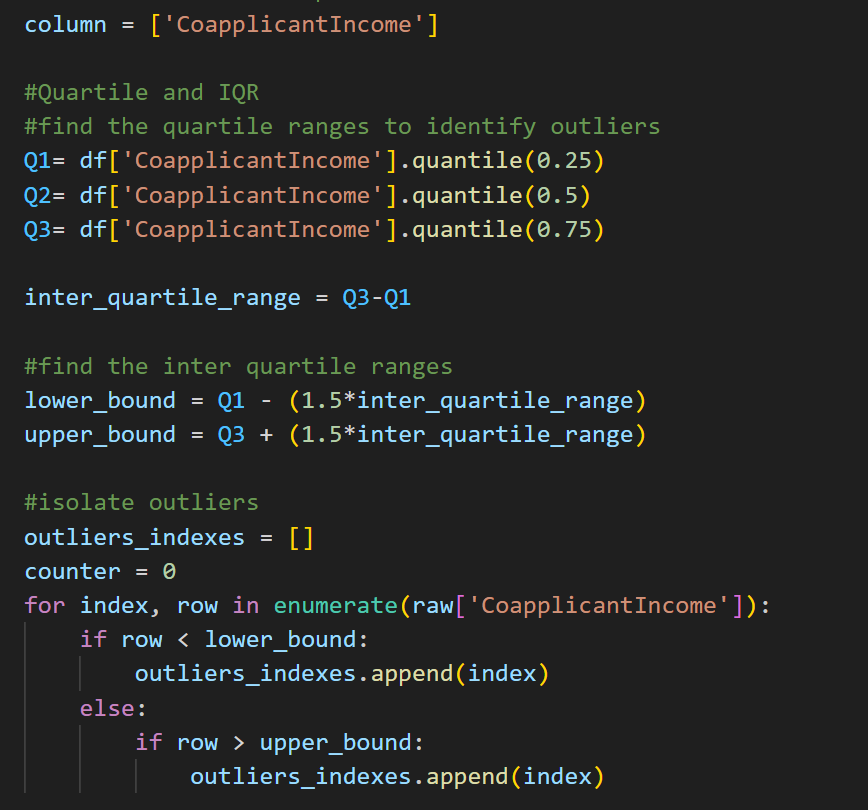
# Co-applicant data summary

## Univariate analysis

To understand the numeric data provided in the co-applicant data column, first, the column was isolated and then tested to see whether there were any missing data entries.



We found that there were no missing entries. Then after the data analysis and interpretation phase began, a box plot was chosen to see how the co-applicant income data is spread and to observe any significant outliers.



A computer screen shot of text

Description automatically generated

A screenshot of a computer screen

Description automatically generated

From these results, we notice that of all the data we had, outliers made up nearly 3% of the overall data and to make sure our algorithm worked well we would have to remove the outliers

A black screen with colorful text

Description automatically generated

With the outliers dropped we were then able to create two bar graphs that represent the data before and after we dropped the outliers as well as relay any statistical informationA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Bivariate Distribution

We plot a new box blot that represents the data with respect to the target of Lao Status

A screen shot of a computer code

Description automatically generated

A screenshot of a graph

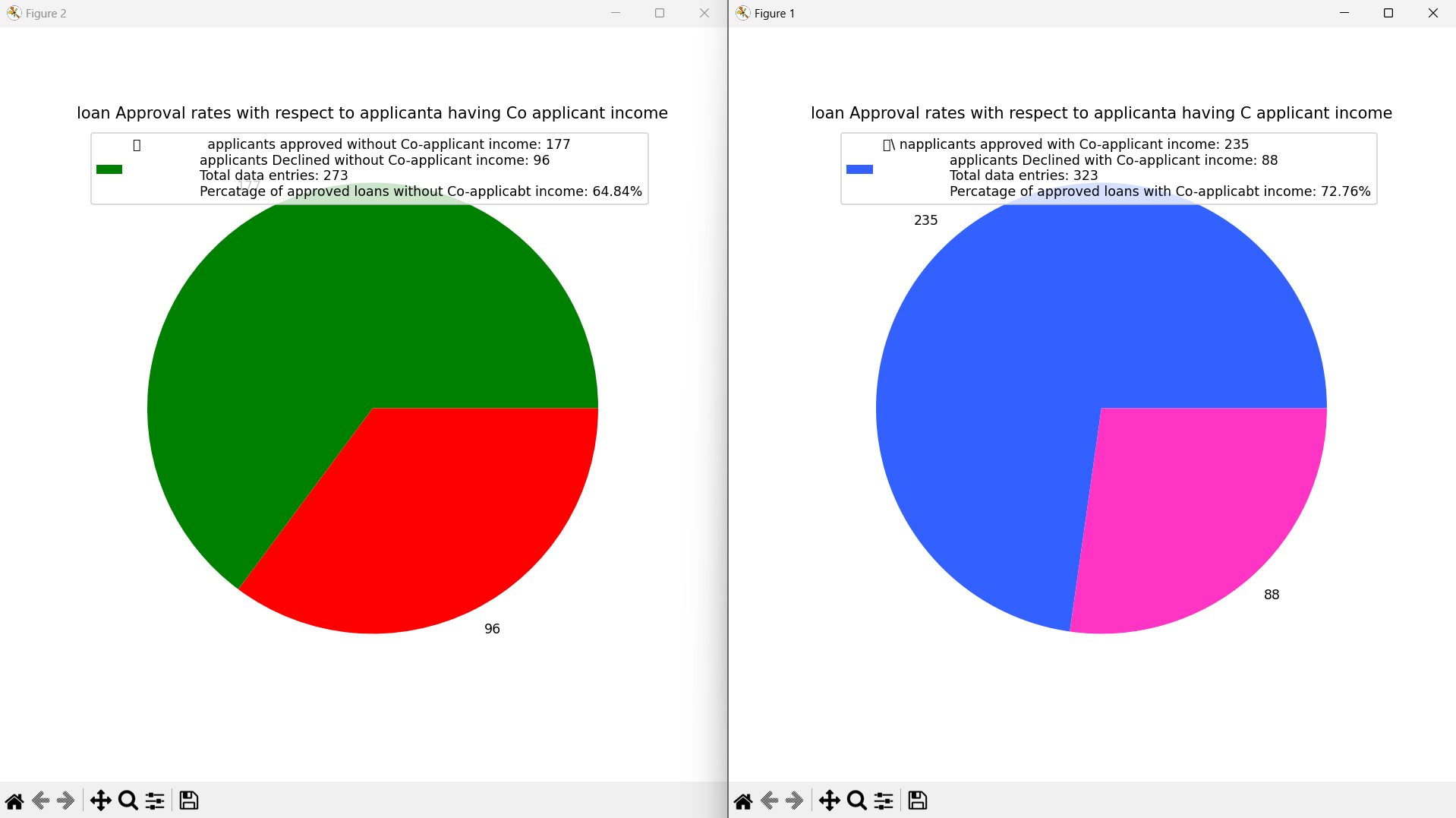
Description automatically generated

Lastly, we do the same with a histogram.

A screenshot of a graph

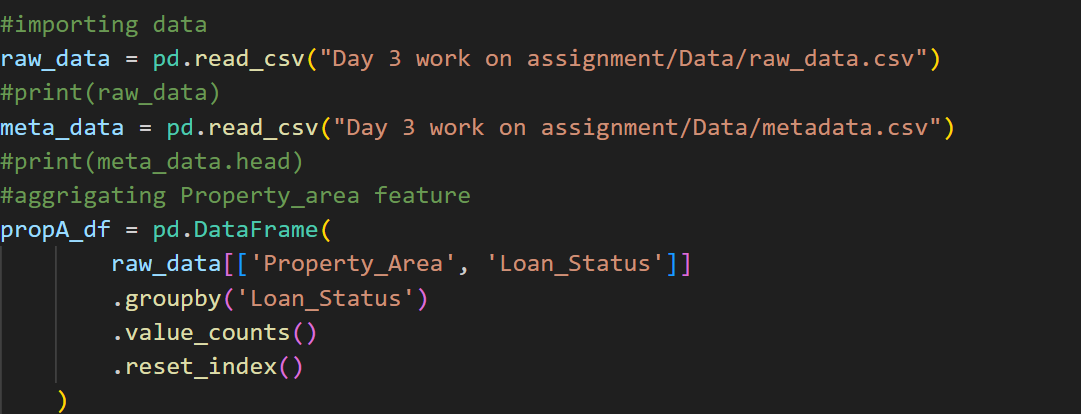
Description automatically generated

In addition to the metrics shown above, we also added pie charts that represent the amount and percentage of applicants who got their loans approved without having Co applicant income as compared to those that did



# Property area analysis

The Property Area column of the data set provides categorical data on areas in which the applicants live. Analysing this data was simple,



First, we imported and aggregated the data.

Thereafter, the data was plotted on a bar graph to compare the distribution of loan approvals and denials concerning the area in which the applicant lived.

A graph of different colored squares

Description automatically generated

A screenshot of a computer

Description automatically generatedFrom the graph, we see that those living in semi-urban areas had higher chances of getting their loans approved unlike those living in urban and rural areas. The percentages of which can be represented on the following pie charts