Lending Club Case Study

Group:
Harish Dambalkar
Sumeetkumar Tripathi (Facilitator)

10th May 2023

Analysis for Lending club

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

As a part of this analysis, we analyse the given dataset to find the details of the borrowers who failed to repay the loan.

This is done based on the available data and considering various factors like employment length, term of the loan, purpose of loan etc.

Analysis for Lending Club

Analysis of the dataset and arriving at a conclusion

Steps performed for the analysis.

1. Performing high level analysis of the dataset.
2. Cleaning the dataset.
3. Checking for any outliers in the data.
4. Univariate analysis.
5. Bivariate analysis.
6. Conclusion from the analysis

High level analysis

- The given dataset is having 110 columns and 39717 entries in those columns.
- Some of the columns are of less significance for the analysis. So we will perform the cleaning of the data and arrive at the final dataset which is ready for analysis.

Cleaning the dataset

As a part of cleaning we performed following steps

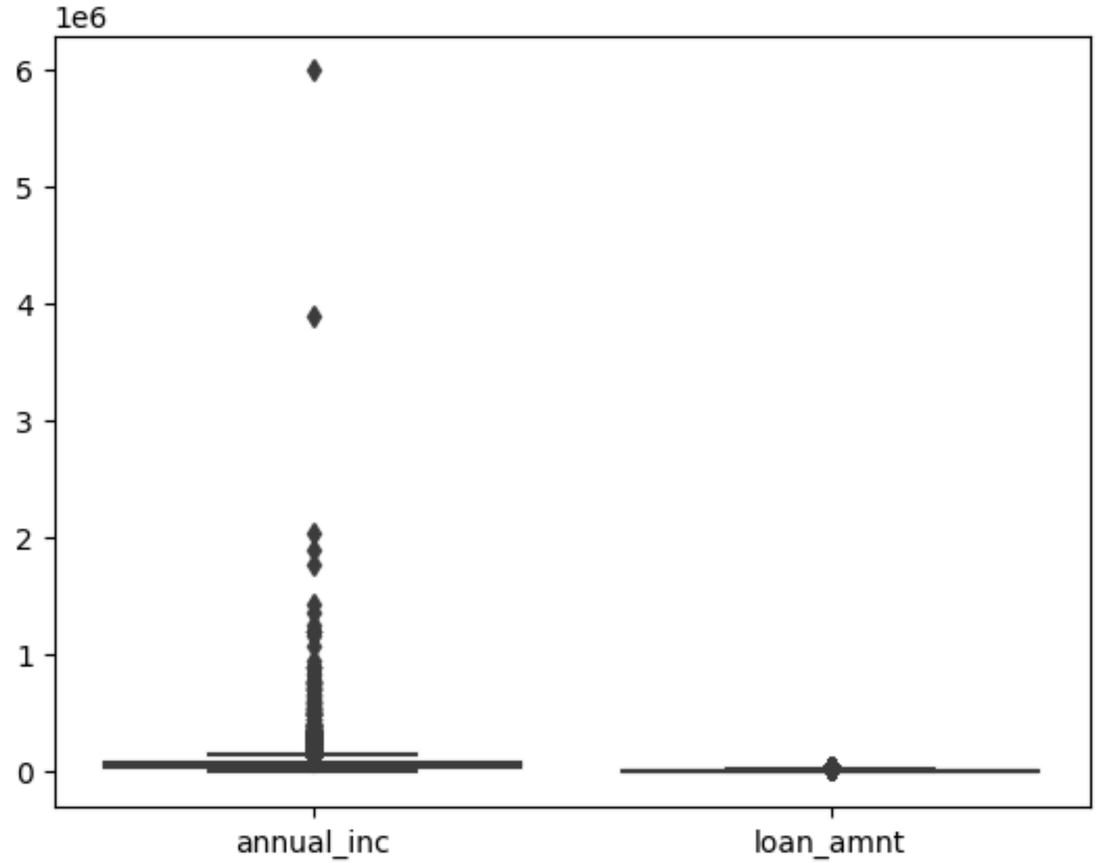
- Removing the columns having null values
- Finding and removing the columns which have unique or single values in it.
- Removing the columns which are of less significance. Like ‘Id’, ‘member_id’, ‘url’ etc.
- Removing columns with large amount of missing data. Since data imputation is not possible.
- Removing columns who’s data is not available at the time of loan application.
- And also removing the current ongoing loans.

Checking Outliers

Outliers in columns: annual_inc, loan_amnt

- In the annual_inc column we have some outliers detected. So we will remove those.
- Outliers were removed using the 99 percentile of data.

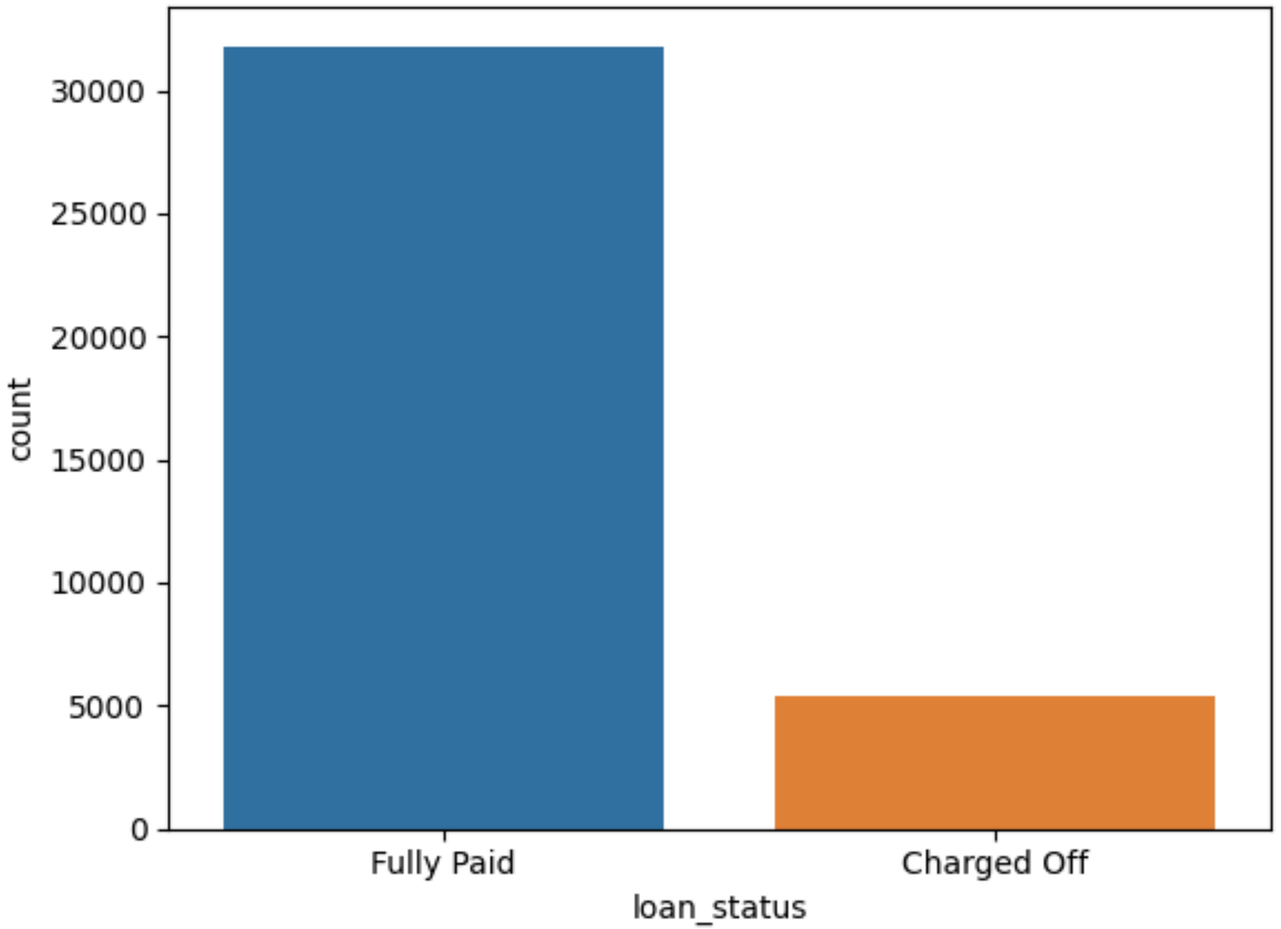
```
In [20]: # Plotting the boxplots for both the columns

         sns.boxplot(df[['annual_inc', 'loan_amnt']])
         plt.show()
```



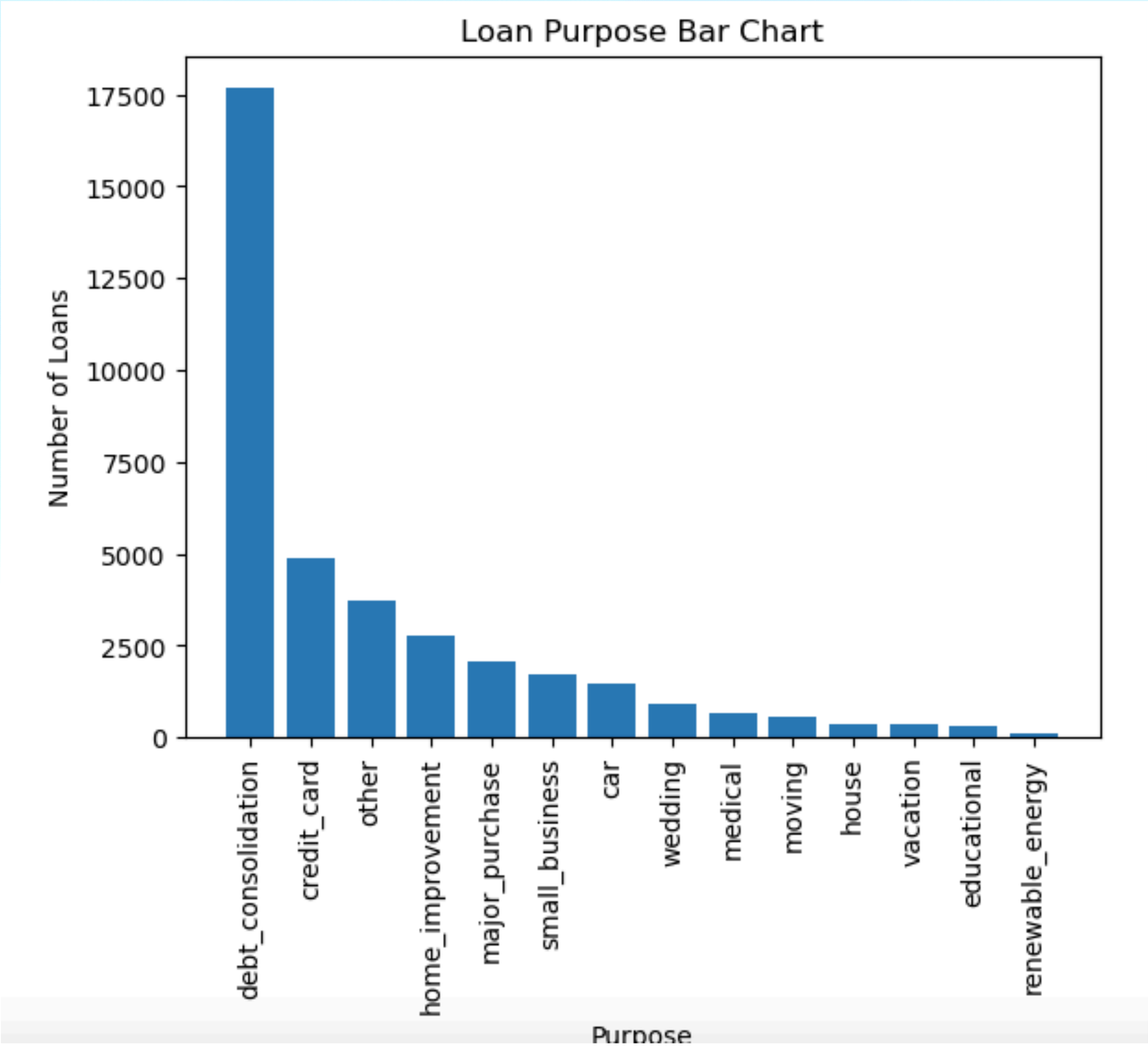
Univariate analysis

- Univariate analysis performed to find the count of “Fully Paid” and “Charged Off” loans.

Count of Full paid vs Charged off loans



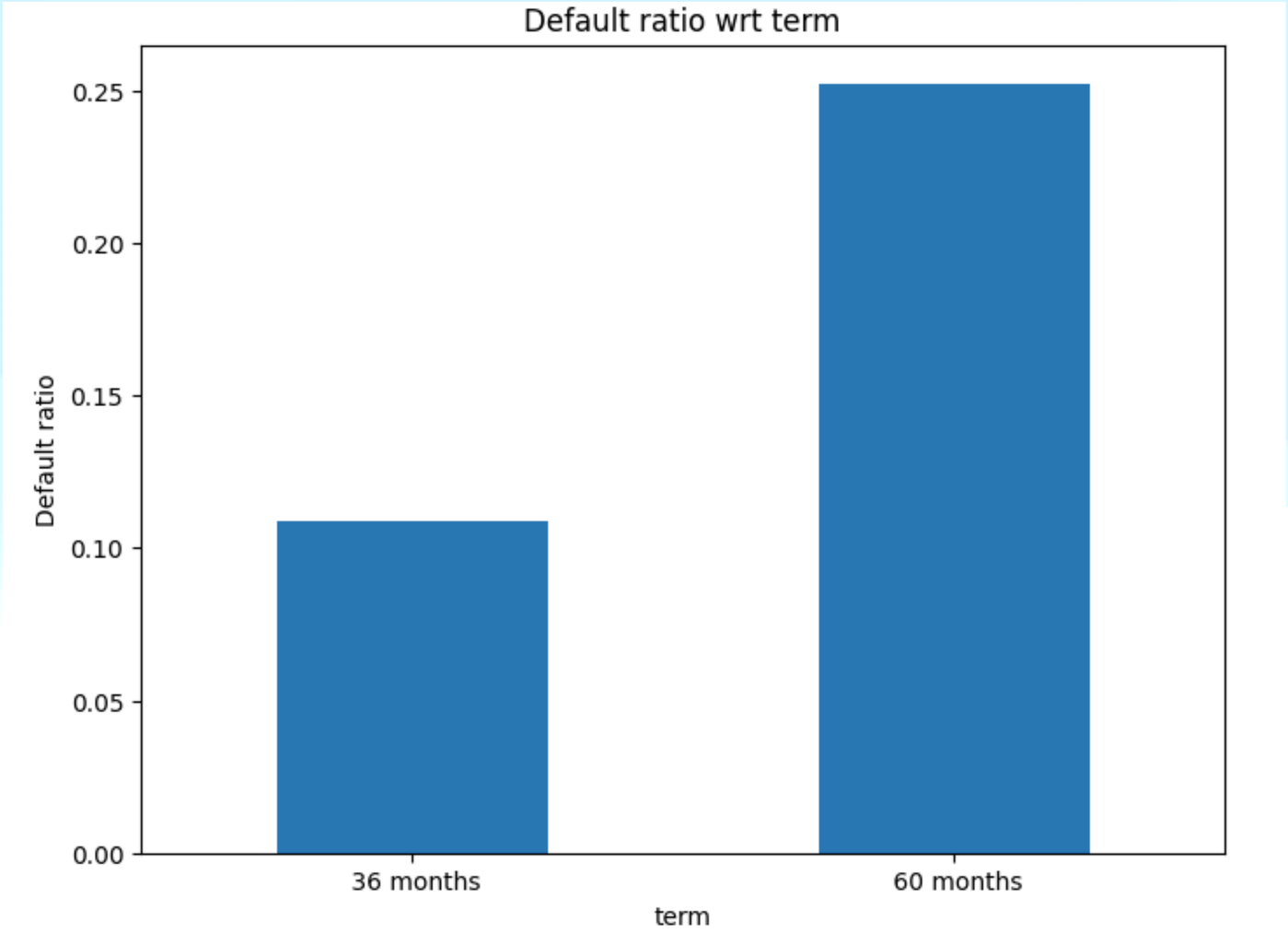
- Univariate analysis to find the number of loans under various categories



Bivariate analysis

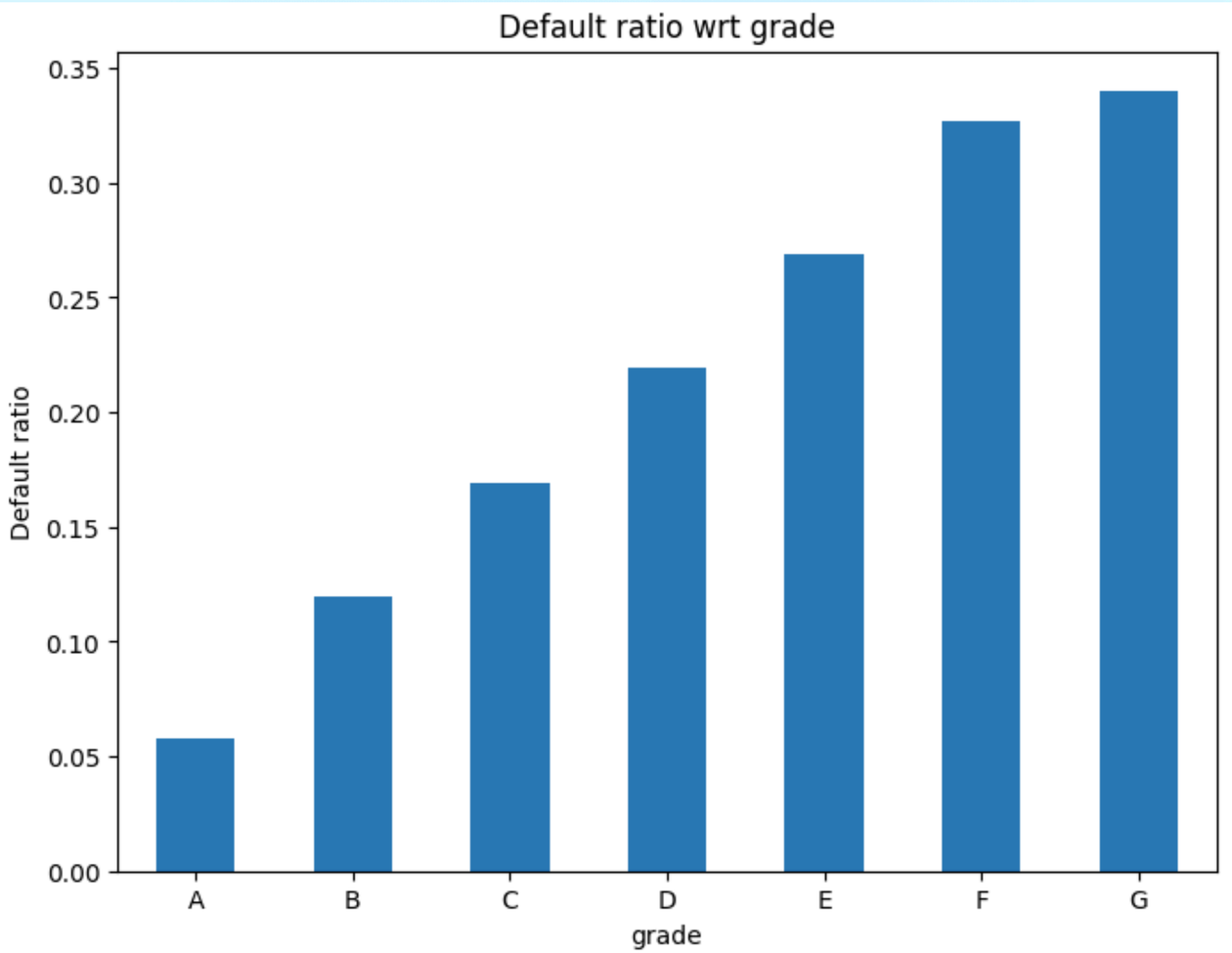
Analysis for “term” column

- Analysis shows that the loans with “60 months” term have higher chances for default.



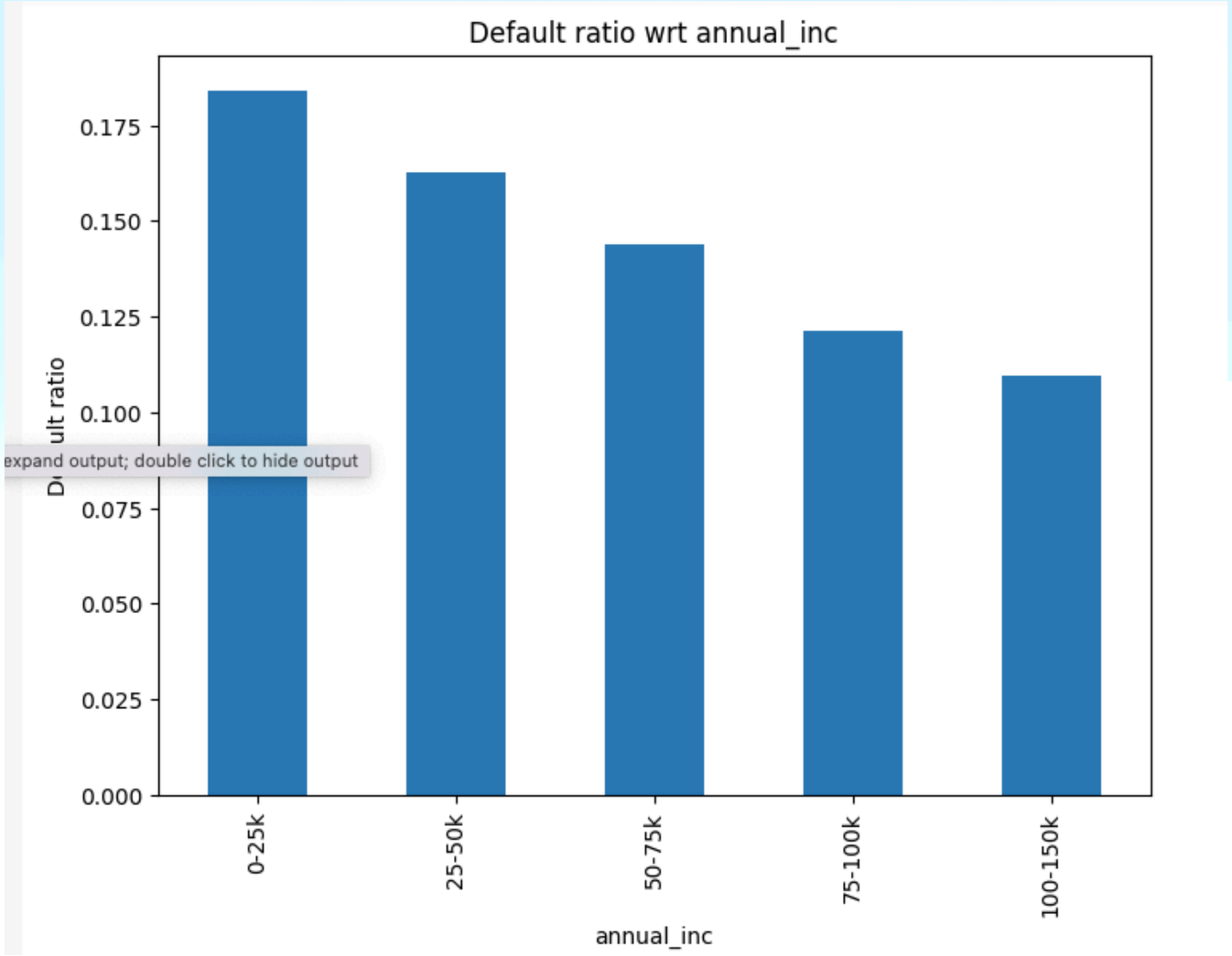
Comparison against the grade of the loan

- Analysis shows the loans with grade “F” and “G” have higher chances of default.



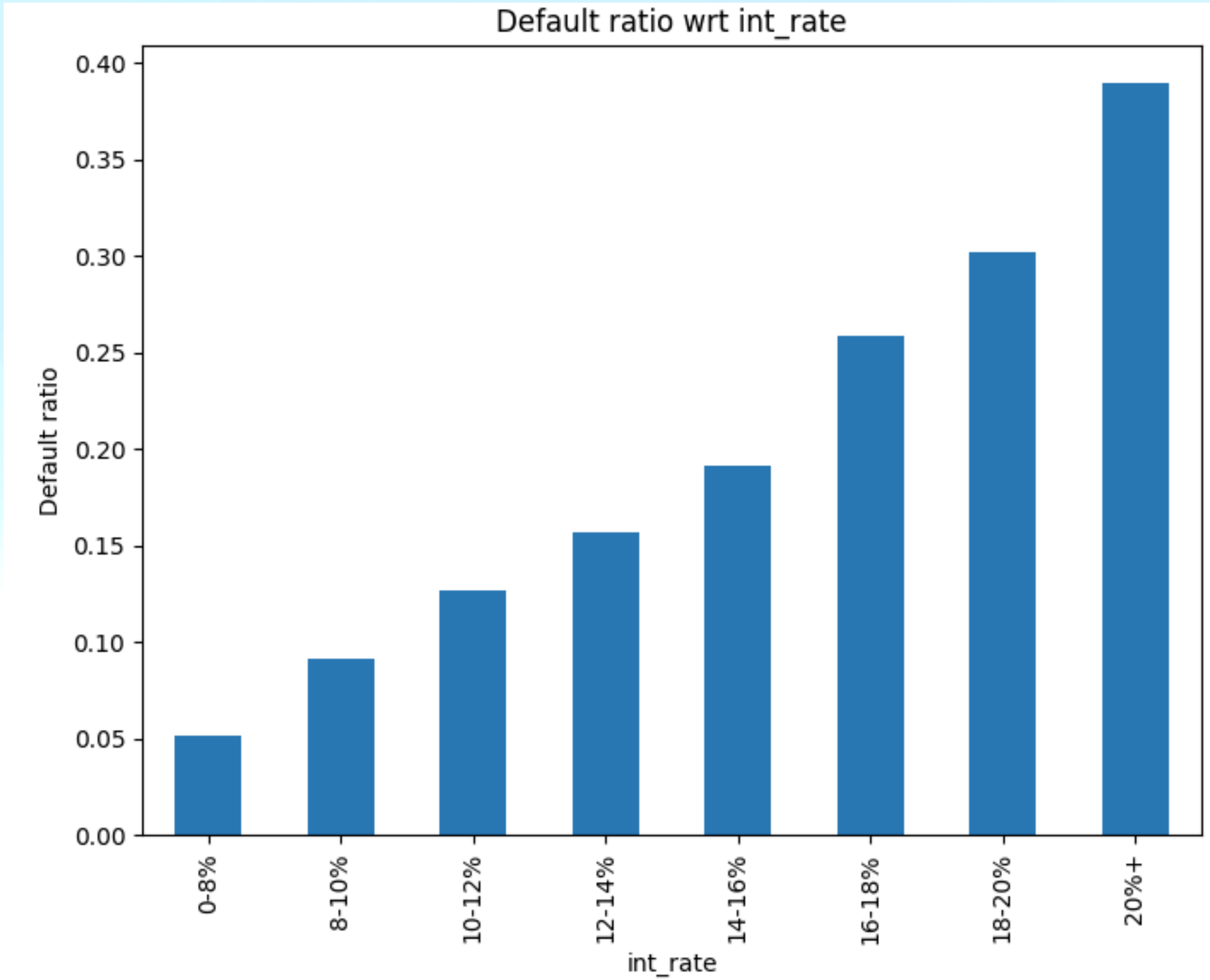
Comparison against the annual income

- This graph shows, as the income increases the default ratio decreases. Which says the loans given to less income groups will associate a risk with it.



Comparison against the loan interest rate

- This shows as the loan interest rate increases, the chances of defaulting the loan also increases.



Final observations

Following observations were made based on the different analysis done.

- If the term of the loan is higher, (Ex. 60months) the default rate is more likely to increase.
- The default rate will increase if we go down the grade. The default rate is lowest for grade A and the highest for grade G.
- For smalll business the default rate is the highest. If a person is applying for a loan for a small busniess then that person is more likely to default.
- Default rate is increasing with the decrease in annual income. People with less annual income are more likely to default.
- Default rate is increasing with increase in loan amount. Amounts greater than 15000 are facing the highest defaults.
- Higher interet rates will alwayts increase the defaults.
- Issuing loans in state NE can be risky.
- Higher installments will increase the defaults.
- Default rate increases with increase in dti ratio.