LENDING CLUB CASE STUDY SUBMISSION

Group Members:
1. **Piyush Kadve (Group Facilitator)**
2. **Raghul Adithya**

Case Study

1. Write all your code in one well-commented Python file; briefly mention the insights and observations from the analysis

2. Present the overall approach of the analysis in a presentation

- Mention the problem statement and the analysis approach briefly

- Explain the results of univariate, bivariate analysis etc. in business terms

- Include visualisations and summarise the most important results in the presentation

Data Cleaning Steps

- **Delete columns**: Deleted columns having all rows as null.

- **Remove outliers**: Remove high values that would disproportionately affect the results of your analysis.

- **Missing values:** Treat missing values with appropriate approach.

- **Filter rows:** Filter by segment, filter by type of loan and income to get only the rows relevant to the analysis.

Analysis

- We performed **data cleaning and processing** on the Loan dataset:
  - Imputed the NA values for all the variables
  - Imputed data in couple of new columns:
    - Date, amount, Rate of interest.
    - Ratio of funded amount and annual income

- During **univariate analysis** we have created:
  - Histograms and Bar charts to check out the distribution of all the driver variables
  - Box plots to detect the Outliers
  - Performed the Multivariate analysis to understand how different variables interact with each other.
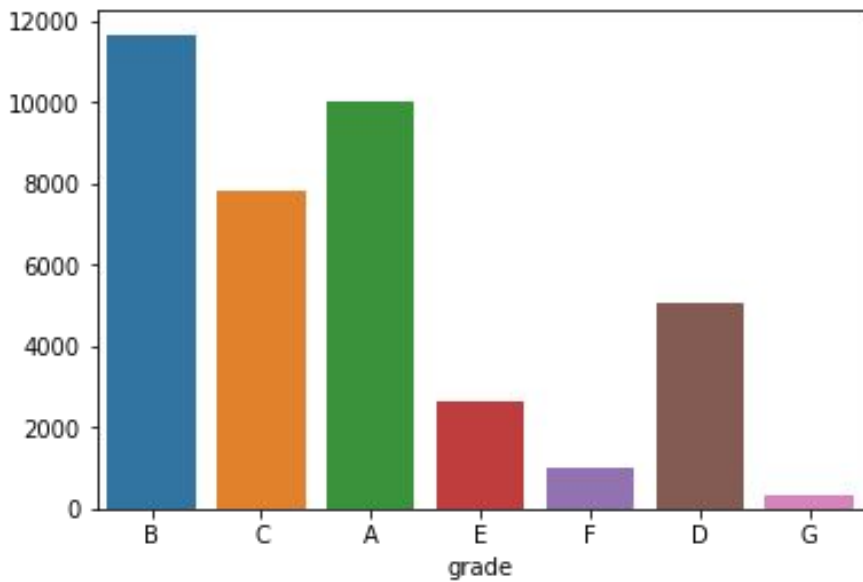
Loan Status Analysis
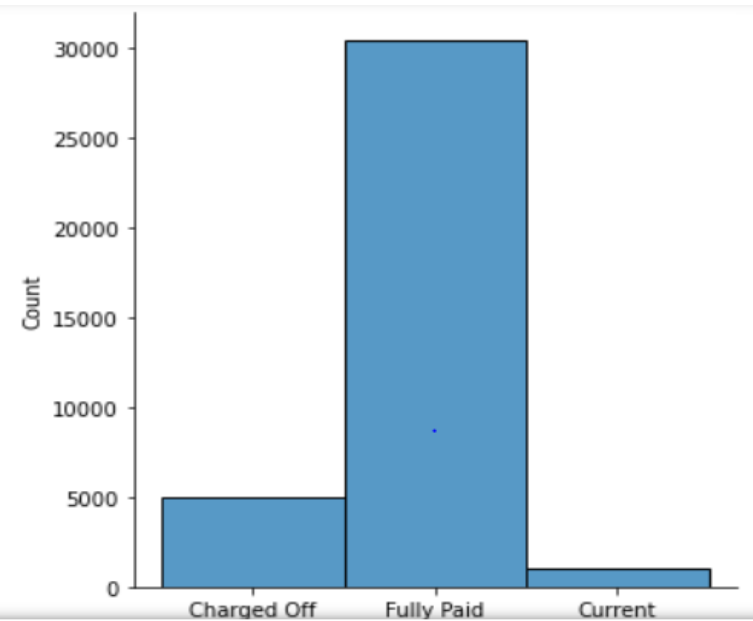
Observations:

1. Most of the loans are Fully Paid.
2. About 14% of loan are having status as defaulters.
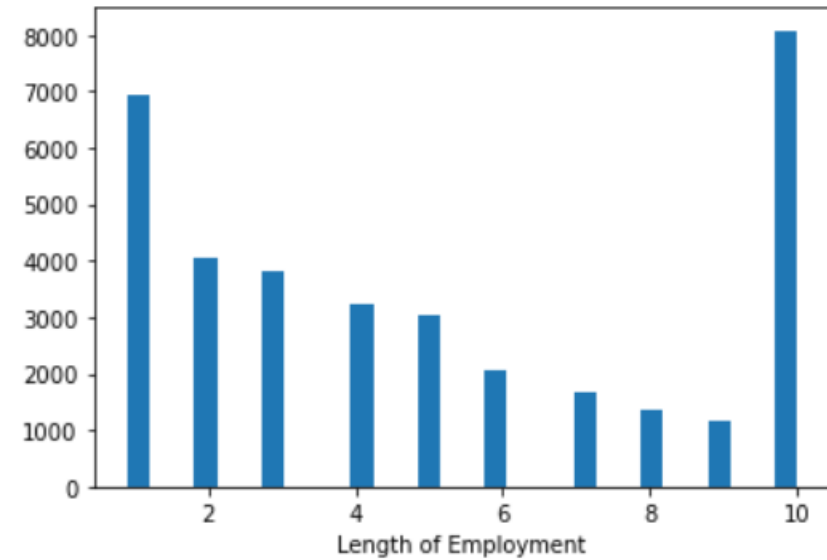
Loan Grade Analysis

Observation:

Most of the loans have grade of A and B. Therefore stating most of the loans are high graded loans.

Loan Applicants work Experience Analysis

Observation:

You can see people tend to avail loans when they have just joined an organization and started working.

❖ We believe people who have less then 2 years of experience have needs and usually will take loans for their personal use.

❖ People who have more than 10 years of experience have usually higher lot of cluster i.e. all those who have 10, 12, 15, 18 years of experience are counted in this group.

❖ It is safe to say lower age leads to higher changes of someone availing loan.
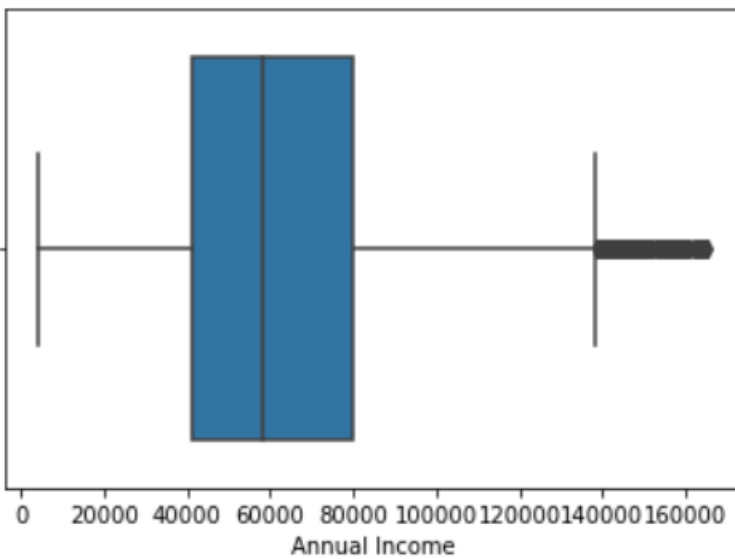
Loan Amount Analysis
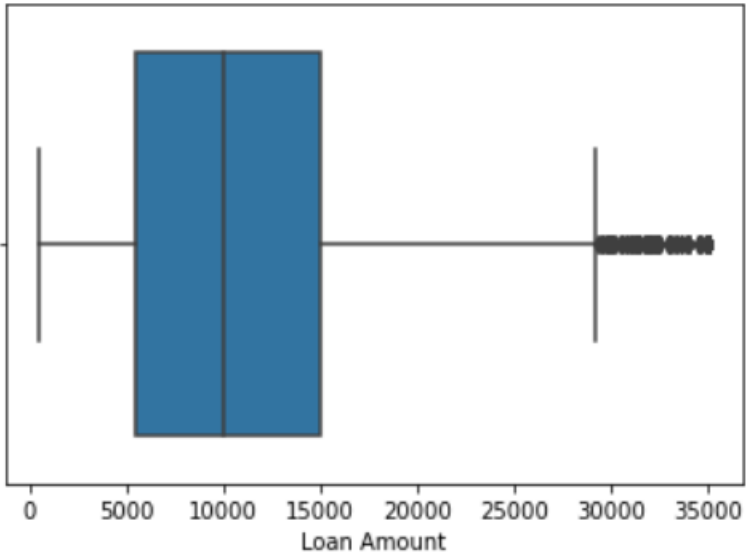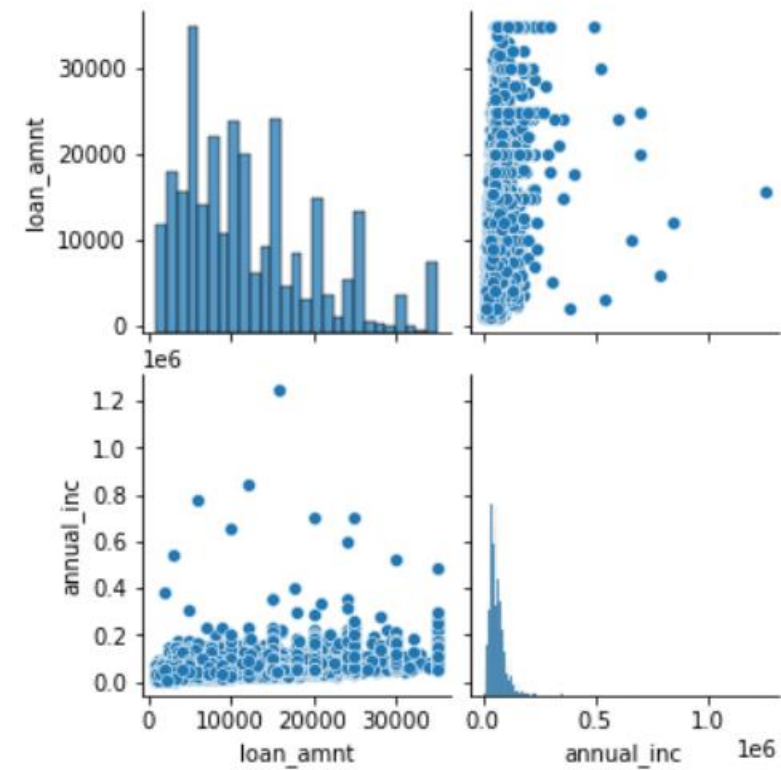
Observation:

1. Funded amount is left skewed. Most of the loan given is below 10000$.
2. We see loans are given to people irrespective of checking their annual income, which can lead to increase in number of defaulters.
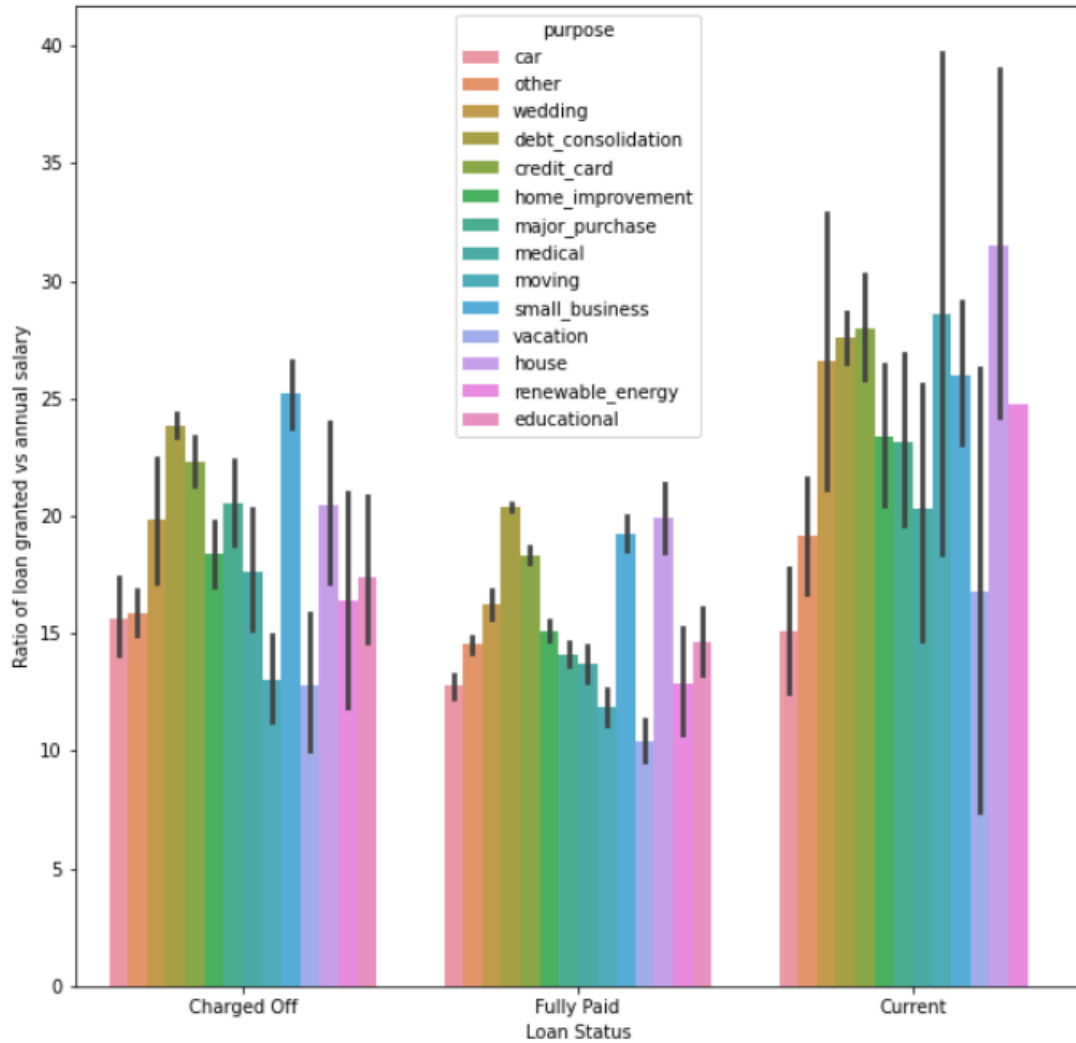
Income Analysis for the loan applicants

Observation:

❖ Most of the applicants have income between 40000$ to 80000$.
❖ There are 1/36 applicants having income way above normal spectrum and they are considered as outliers.

Purpose for applying Loan Analysis

Observation:

- This chart can give us insights that most common type of loan is small businesses which is also a volatile type of loan. If the business fails then there is no way someone follows through and pays entire loan.
- The second insight is home loan is second most common type of loan and it is very secure return on investment.

Final Conclusion

➢ About 14% of loan are having status as defaulters.

➢ Low grade loans have high tendency to default. Grading system is working as expected.

➢ It is safe to say lower age leads to higher changes of someone availing loan.

➢ We see loans are given to people irrespective of checking their annual income, which can lead to increase in number of defaulters. Extra attention must to given in this aspect.

➢ Most common type of loan is "small businesses" which is also a very volatile type of loan. If the business fails then there is no way someone follows through and pays entire loan.

➢ The second insight is home loan is second most common type of loan and it is very secure return on investment.

Best Regards,
Piyush & Raghul.