Lending Club Case Study

Presenter :

Name : Mayank Aditya

Course : PGP - ML & AI

Batch : September 2021

What is lending club?

- Lending Club is the world's largest online loan marketplace, offering personal loans, business loans, and medical procedure funding.
- Through a quick internet interface, borrowers can readily acquire lower interest rate loans.
- Lending to 'risky' applicants, like most other lending organization's, is the most common source of financial loss (called credit loss).
- The amount of money lost by the lender when a borrower refuses to pay or flees with the money owed is referred to as a credit loss. In other words, defaulting borrowers do the most financial harm to lenders.

Problem Statement!

Past loan applicants are listed in the loan data, along with whether or not they 'defaulted.' Our goal is to find patterns that can suggest whether a person is likely to fail on a loan, which can then be used to make decisions like refusing the loan, lowering the loan amount, lending (to risky applicants) at a higher interest rate, and so on.

Loan Data Features

1. **Loan characteristics** such as loan amount, loan term, and purpose, which display information about the loan and will aid us in the discovery of loan default.

2. **Demographical Data** : **Age**, **employment position**, and **relationship status** are examples of demographic variables that reveal information about the borrower profile that can be relevant to us.

3. **Behavioral Variables** such as **next payment date, EMI, delinquency** which shows the information which is updated after providing the loan.

Loan Data Analysis

With the help of EDA, we have divided our loan data analysis into 3 parts.

1. **Univariate Analysis** (Categorical/Continuous Features)

2. **Bivariate Analysis** (Box Plots)

3. **Multivariate Analysis** (Correlation Heatmap)

Univariate Analysis Part 1

1. **Image 1 :** Depicts that there are around ~5000 defaulters in the data set. The ratio of defaulters is ~14%

2. **Image 2 :** From this visualization we can clearly state that there more defaulters when the loan term is 60 months.

3. **Image 3 :** We can clearly infer that that the home ownership has no correlation with defaulters. As the graph is same for rent, own and mortgage homes.

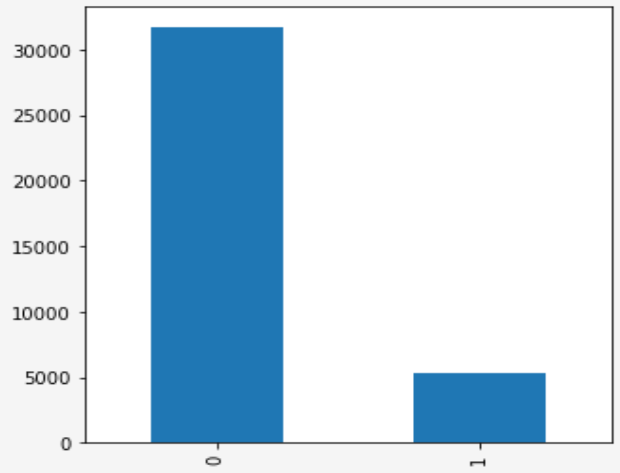


Image 1 : **Defaulter Vs Non Defaulter**

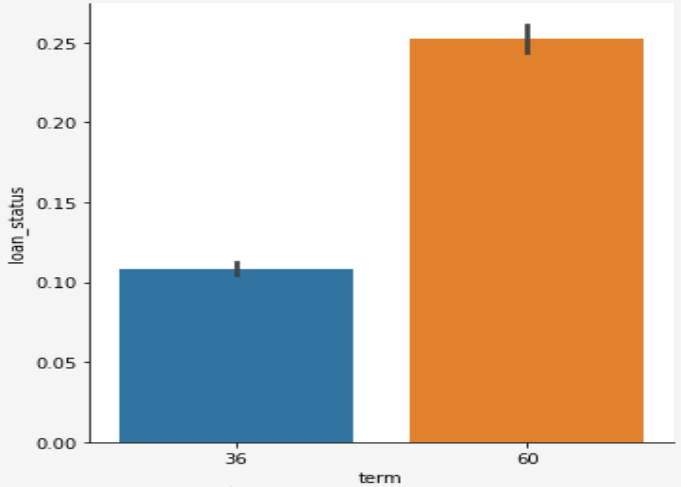


Image 2 : **Defaulter Vs Loan Term**

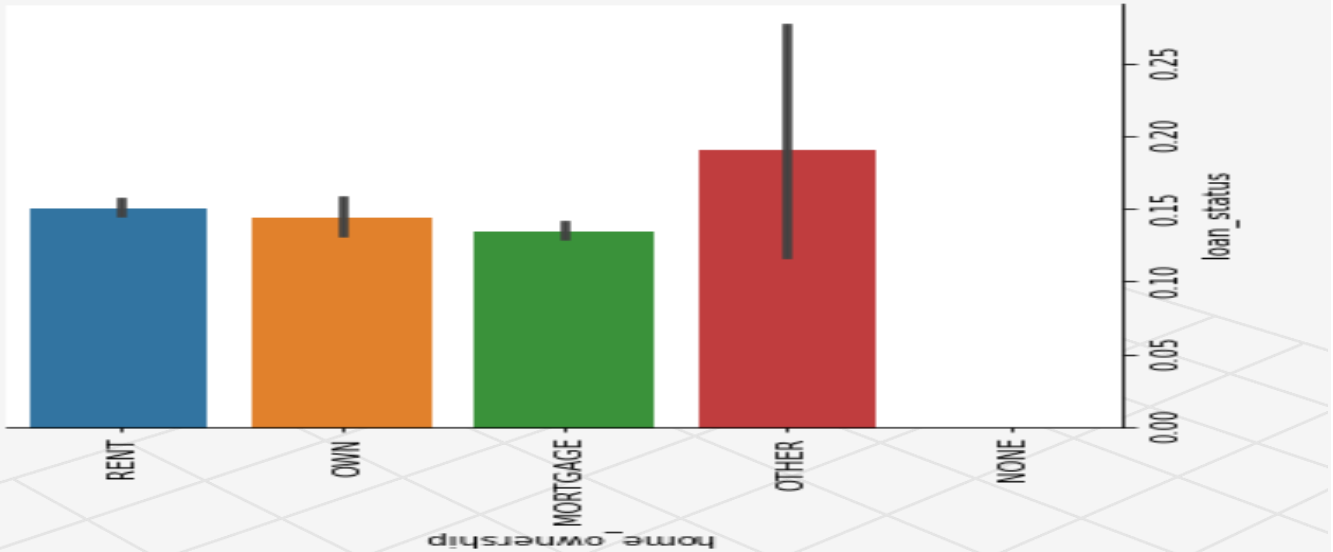


Image 3 : **Defaulter Vs Home Ownership**

Univariate Analysis Part 2

1. **Image 1 :** Defaulter Vs Grade graph can clearly depicts that defaulters are more in E , F & G grade.

2. **Image 2 :** From plot for 'purpose' we can infer that the defaulters rate is nearly constant for all purpose type except 'small business', hence rate will depend on purpose of the loan.

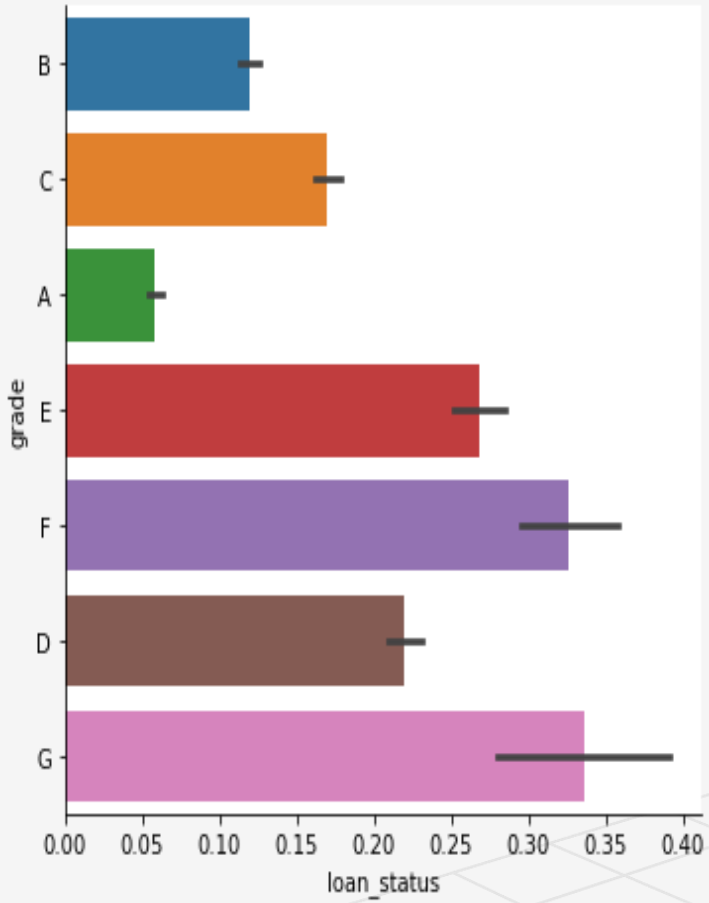


Image 1 : **Defaulter Vs Grade**

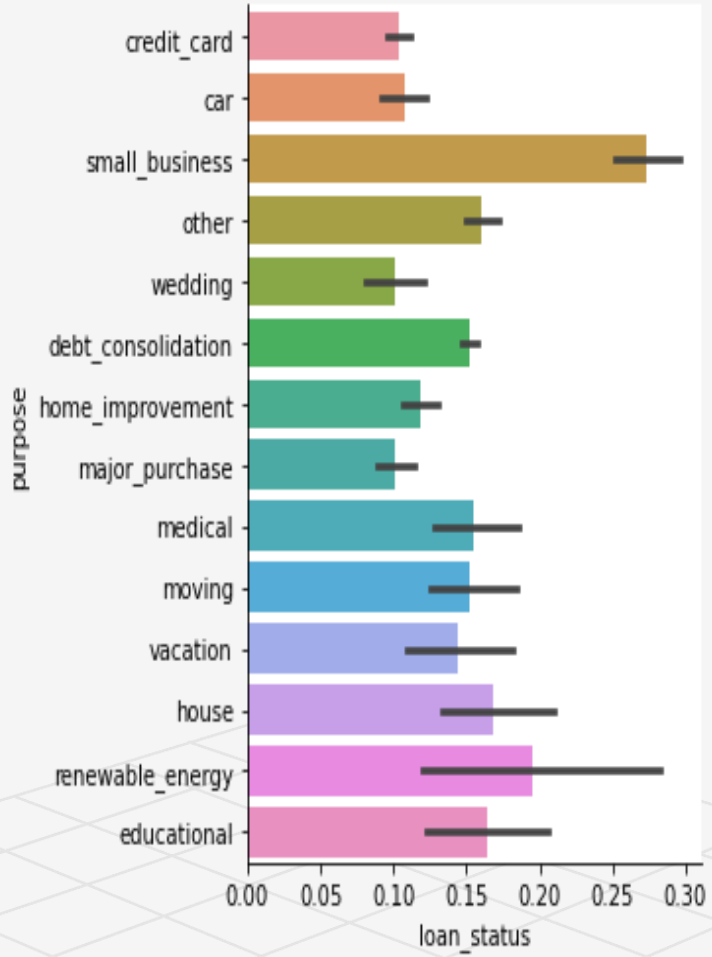


Image 2 : **Defaulter Vs Purpose**

Univariate Analysis Part 3

1. **Image 1 :** We may deduce from the following plot for 'annual inc range' that the defaulters rate is dropping as annual income increases, and hence rate will be dependent on annual inc range feature.

2. **Image 2 :** We may deduce from 'Loan amount range' is that the defaulters rate is increasing with Loan amount range, and hence rate will be dependent on Loan amount range feature.

3. **Image 3 :** We may deduce from the following plot for 'interest rate range' that the defaulters rate decreases interest rate decreases, hence rate will be dependent on 'interest rate range' feature.

4. **Image 4 :** We may deduce from the given plot for 'DTI Range' that the defaulters rate is increasing in with DTI range values, and so rate will be influenced by DTI range feature.

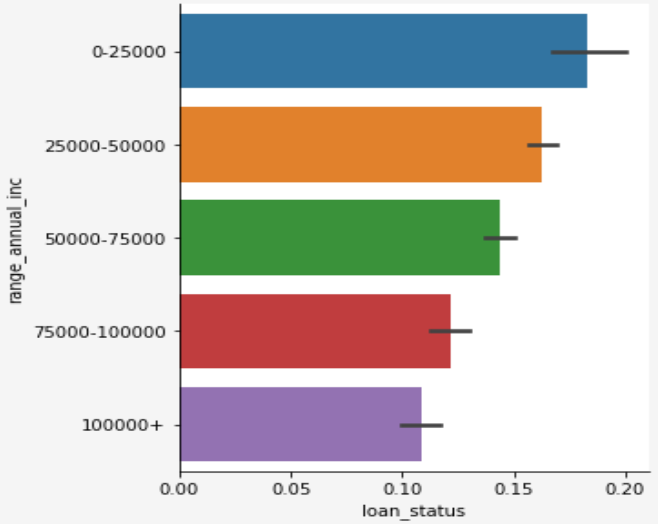


Image 1 : **Defaulter Vs Annual Income Range**

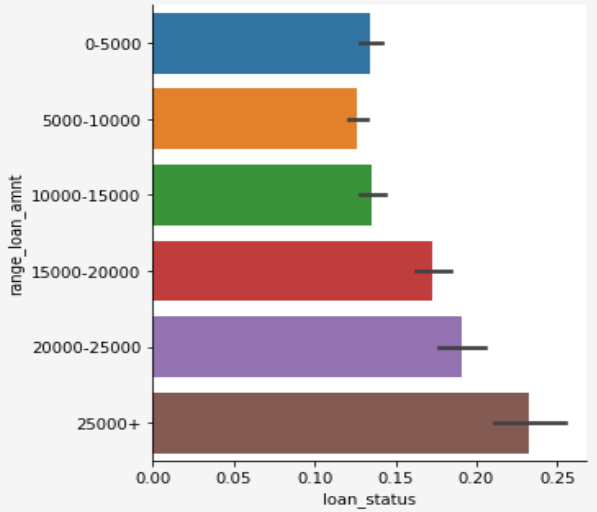


Image 2 : **Defaulter Vs Loan Amount Range**

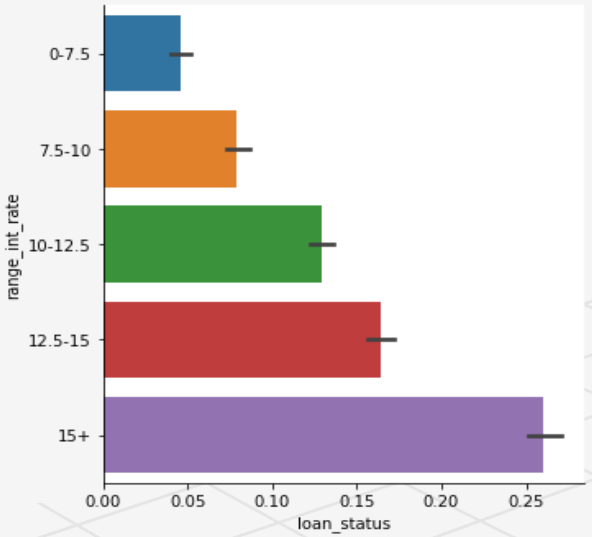


Image 3 : **Defaulter Vs Interest Range**

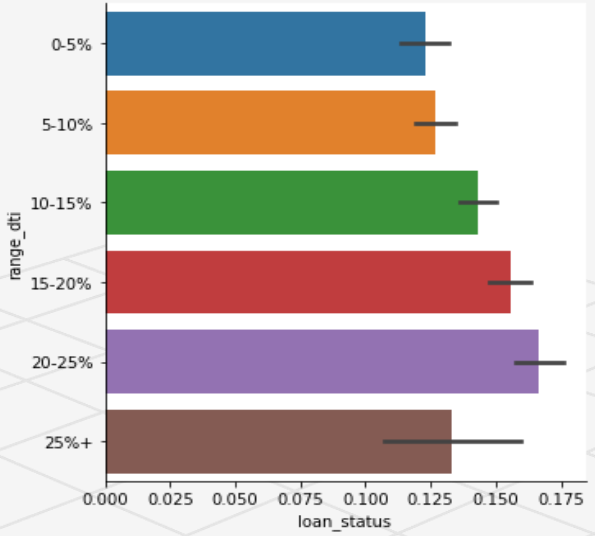


Image 4 : **Defaulter Vs DTI Range**

Bivariate Analysis Part 1

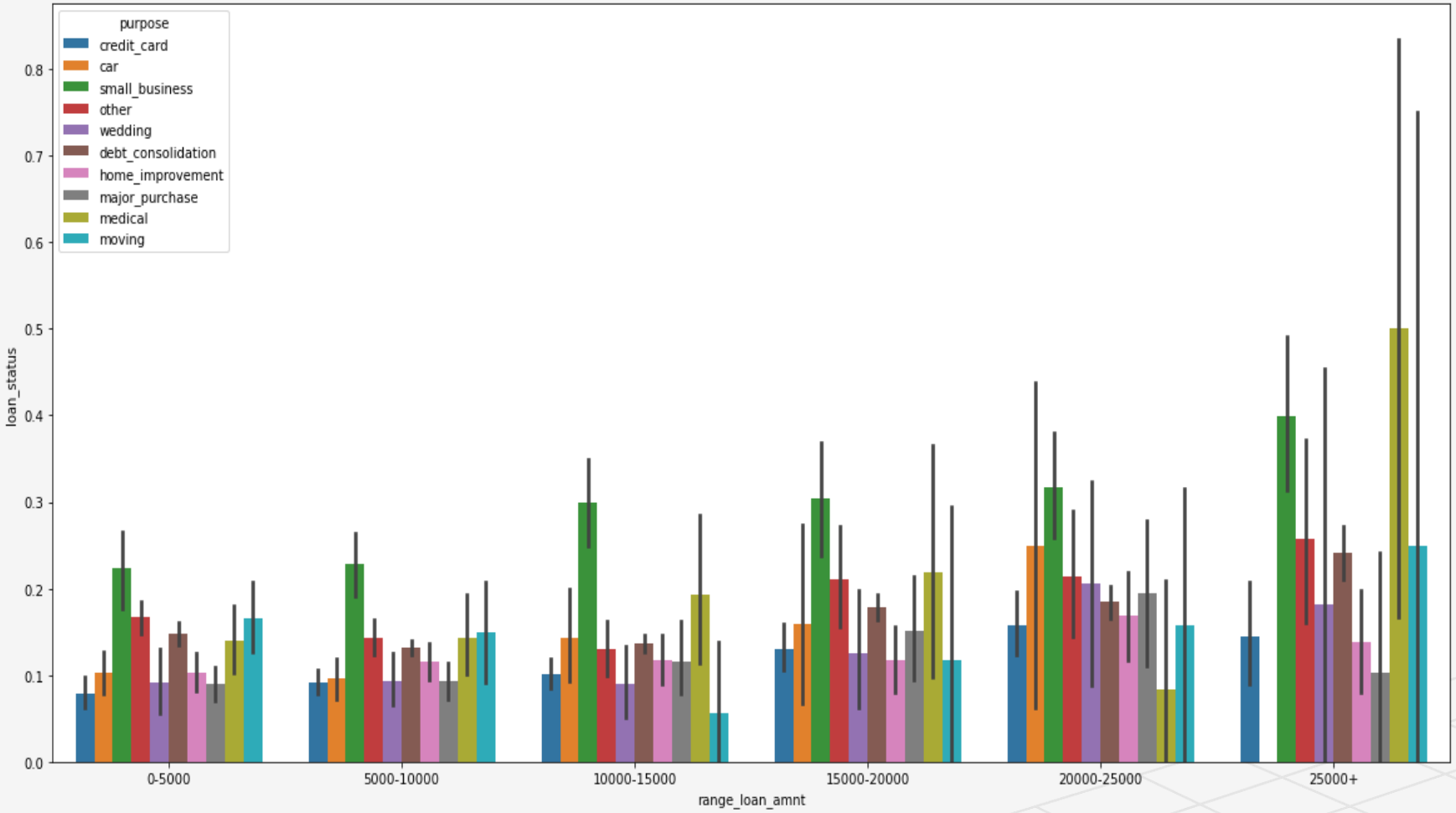
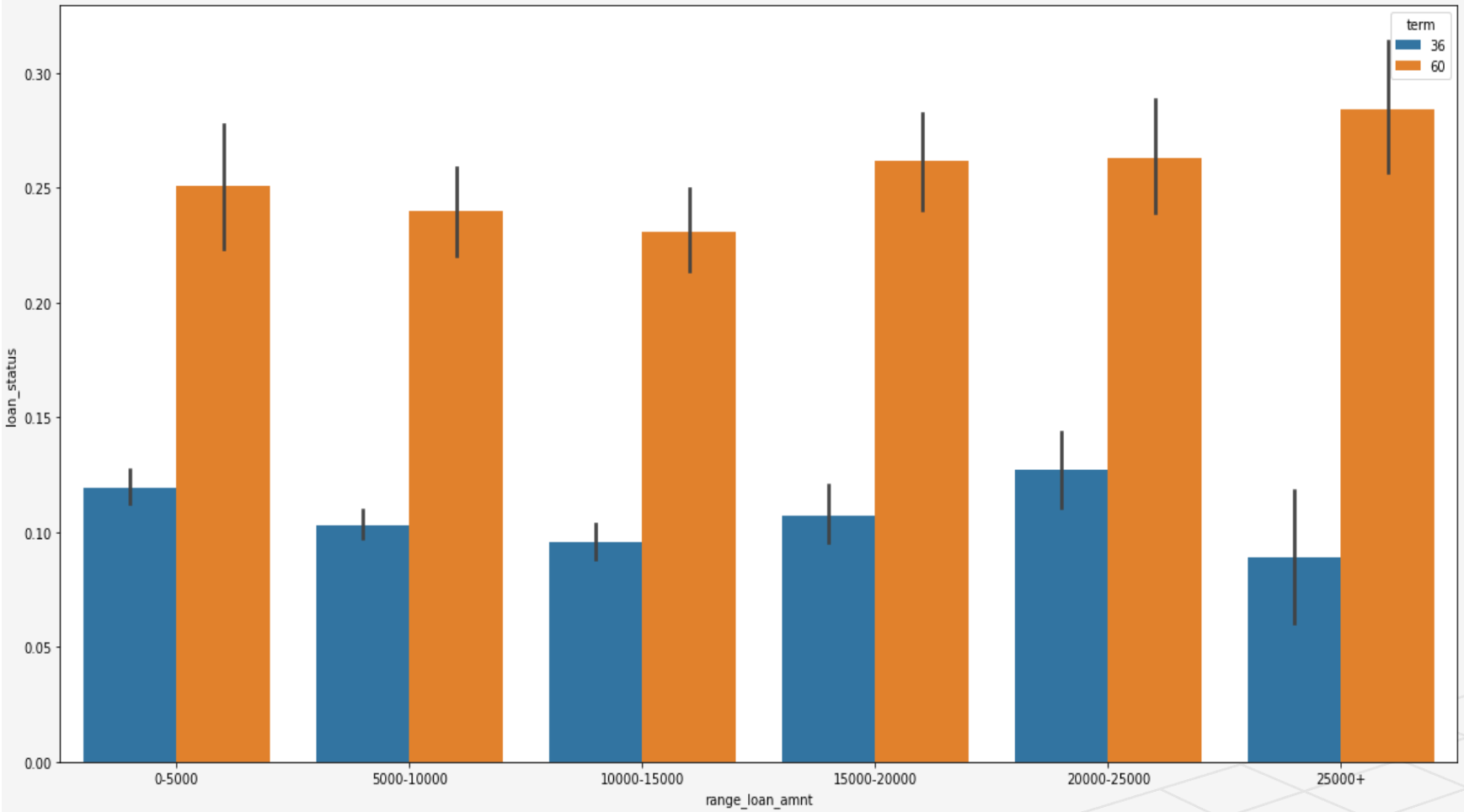


Image 1 : Defaulter Vs Pupose Vs Loan_Amount_Range

1

As we can see straight lines on the plot, default ratio increases for every purpose with respect too loan_amount_range. We can concur that the defaulters are more in small business and medical purpose with 25000+ loan amount.

Bivariate Analysis Part 2



1 *As we can see straight lines on the plot, default ratio increases for every term with respect to loan_amount_range.*

Image 1 : Defaulter Vs Loan Amount Range Vs Term

Bivariate Analysis Part 3

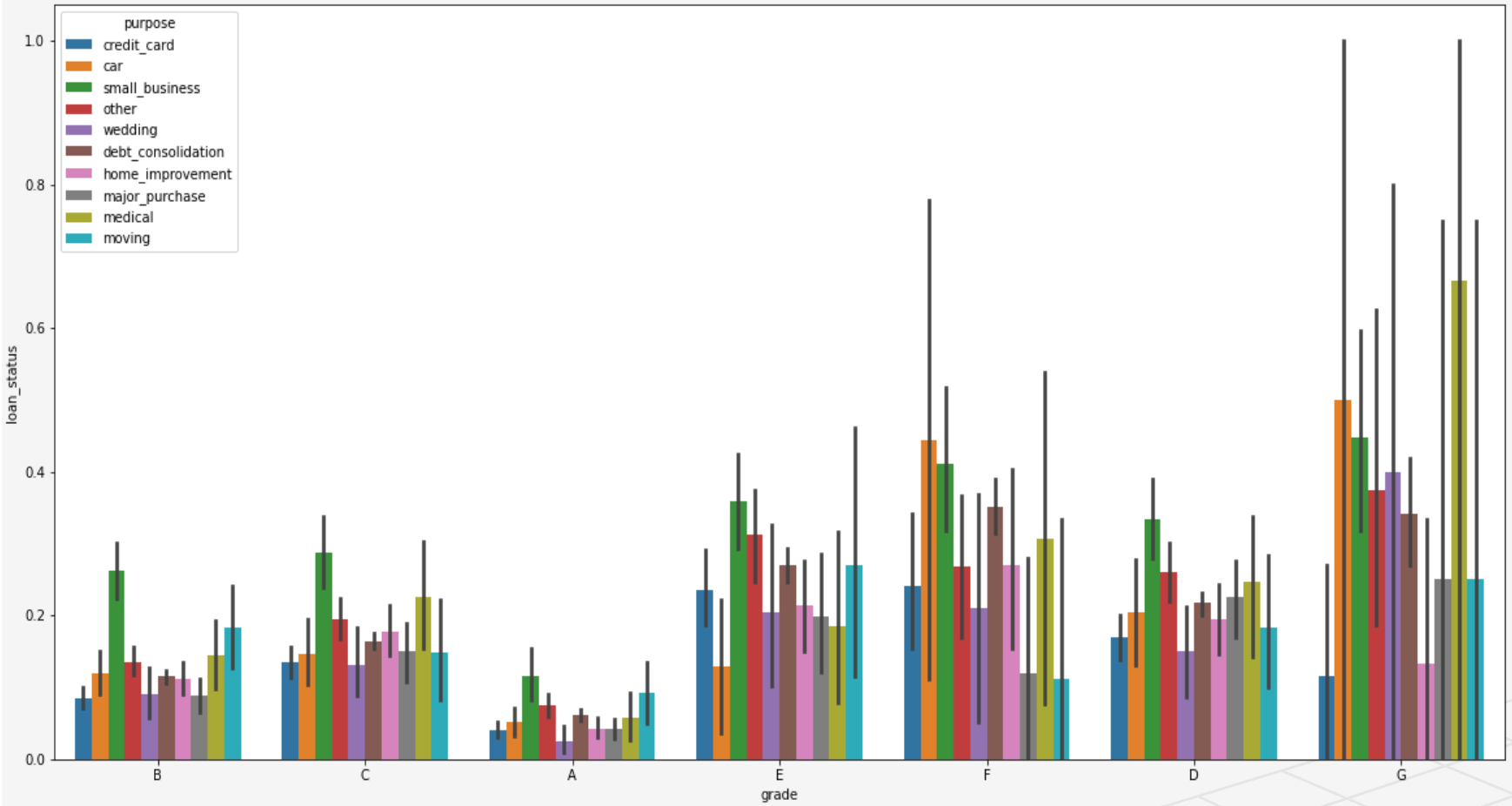


Image 1 : Defaulter Vs Purpose Vs Grade

1

As we can see straight lines on the plot, default ratio increases for every purpose with respect to grade. Top defaulters being E, F, G grades.

Bivariate Analysis Part 4

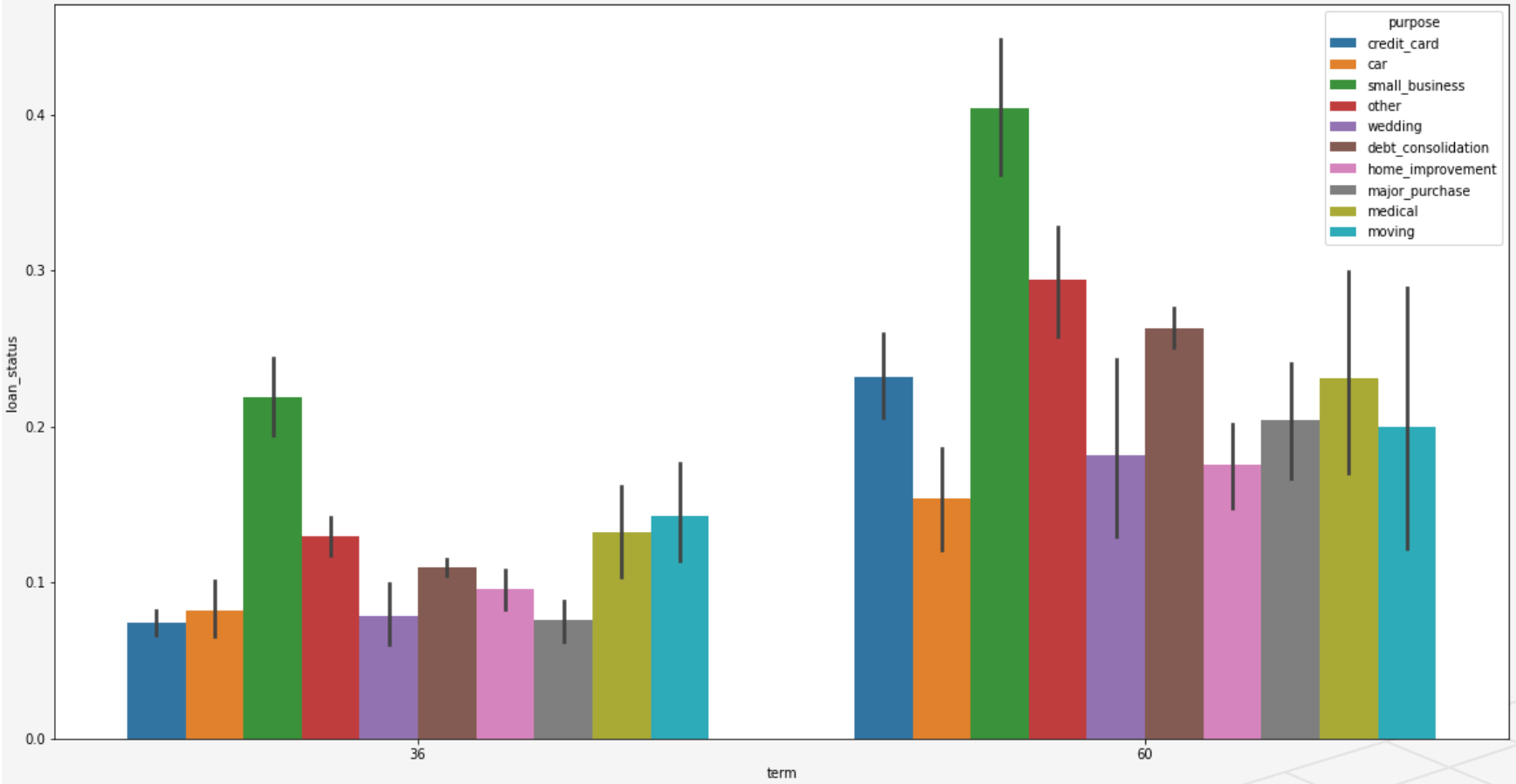


Image 1 : Defaulter Vs Purpose Vs Term

1

As we can see straight lines on the plot, default ratio increases for every purpose with respect to term. Top defaulter being small business with 60 months term.

Multivariate Analysis

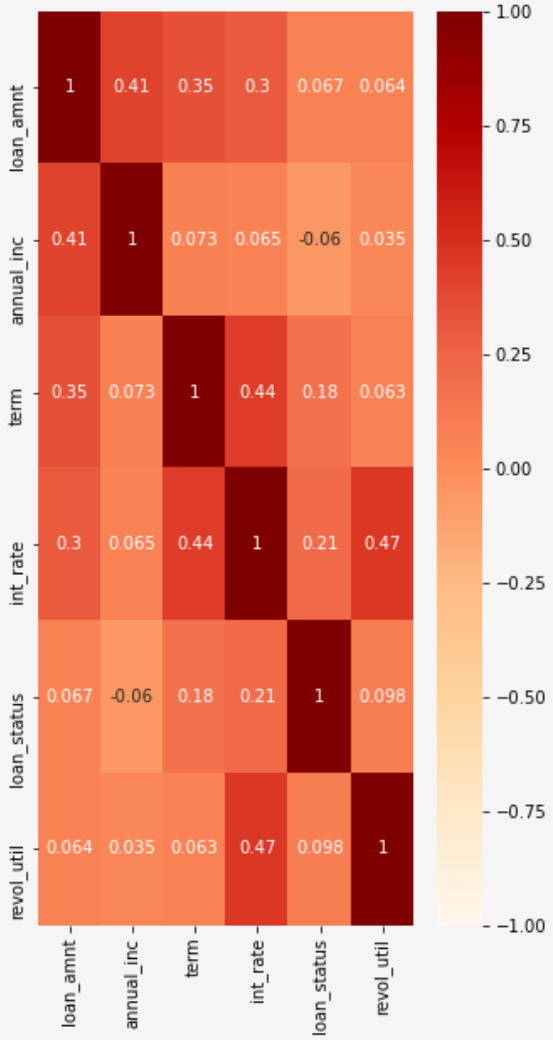


Image 1 : Correlation Heatmap

1

Through this heat map we have tried to show the correlation between loan_amount, annual_income, term, interest_rate, loan_status, revolving_line_utilization_rate.

Final Findings

From the various analyses we have performed, we can concluded that the defaulter ratio is mostly based on below variables :

- **Term**
- **Grade**
- **Purpose**
- **Interest_Rate**
- **Annual_Income**
- **Loan_Amount**
- **DTI Range**

Thank You !

Presenter :

Name : Mayank Aditya

Course : PGP - ML & AI

Batch : September 2021