Lending Club Case Study

POONAM RATHOD & PRASEETHA KUMARSINGH

Business Understanding

- **BACKGROUND**

- Lending Club is a consumer finance company which is the largest online loan marketplace, facilitating personal loans, business loans etc. Borrowers can easily access lower interest rate loans through a fast online interface.

- **PROBLEM STATEMENT**

- When the company receives a loan application, the decision for loan approval has to be made by the company, based on the applicant's profile.

- Two scenarios that are likely to occur –
  - case1 - Applicant likely to repay the loan. If loan not approved, Loss of business for the company
  - case2 - Applicant not likely to repay the loan ("Risky applicants"). If loan approved, financial loss (called credit loss) for the company

Objective

- The objective is to use EDA and provide insights which can be used by the company to make decisions on loan approval/rejection.

  ➢ Identify the driving factors/driver variables behind loan default , i.e. the variables which are strong indicators of default.
  ➢ Identify patterns which indicates if a person is likely to default(risky applicant).
  ➢ Provide insights and observations using data visualization

  Advantages :
  ➢ This gives us a good understanding of how EDA can be used in Real life problems
  ➢ It gives us a basic idea of risk analytics in banking and financial firms
  ➢ It gives us how data can be visually presented to explain the analysis

Data Understanding

- The dataset provided contains complete loan data for all loans issued through time period 2007-2011.

- The data dictionary is also provided which describes meaning of variables in the dataset.

- The dataset has consumer attributes and loan attributes. EDA will help to understand how they influence the tendency of default.

- The dataset contains data only for accepted loans and does not contain data of rejected loans

- In Loan accepted case , the 3 possible values of Loan status are
    - Fully Paid – Applicant has fully paid the loan (principal and interest)
    - Current – Applicant is in the process of paying the loan in installments . Hence these applicants cannot be labelled as defaulters.
    - Charged-off – Applicant failed to pay installments in due time for a long duration and hence has defaulted on the loan

Data Preparation

- **Our Approach : Data Cleaning process to prepare data for better analysis**

1 • Handling Null values. Drop columns which has all Nulls.

2 • Dropping columns which has high percentage of missing values(> 50%).

3 • Dropping data which has Loan status = 'Current' because these applicants are in the process of paying the loan.

4 • Dropping Columns which has mostly zero values

5 • Dropping Columns which have non-relevant data

6 • Dropping rows for columns which have very few null values

7 • Checking and Removing Outliers in data

8 • Dropping columns which have only one single unique value

Analysis

- **Analysis approach:**

|	
1	• Segmentation of Data
2	• Deriving necessary metrices
3	• Univariate Analysis
4	• Segmented Univariate Analysis
5	• Bi-Variate Analysis
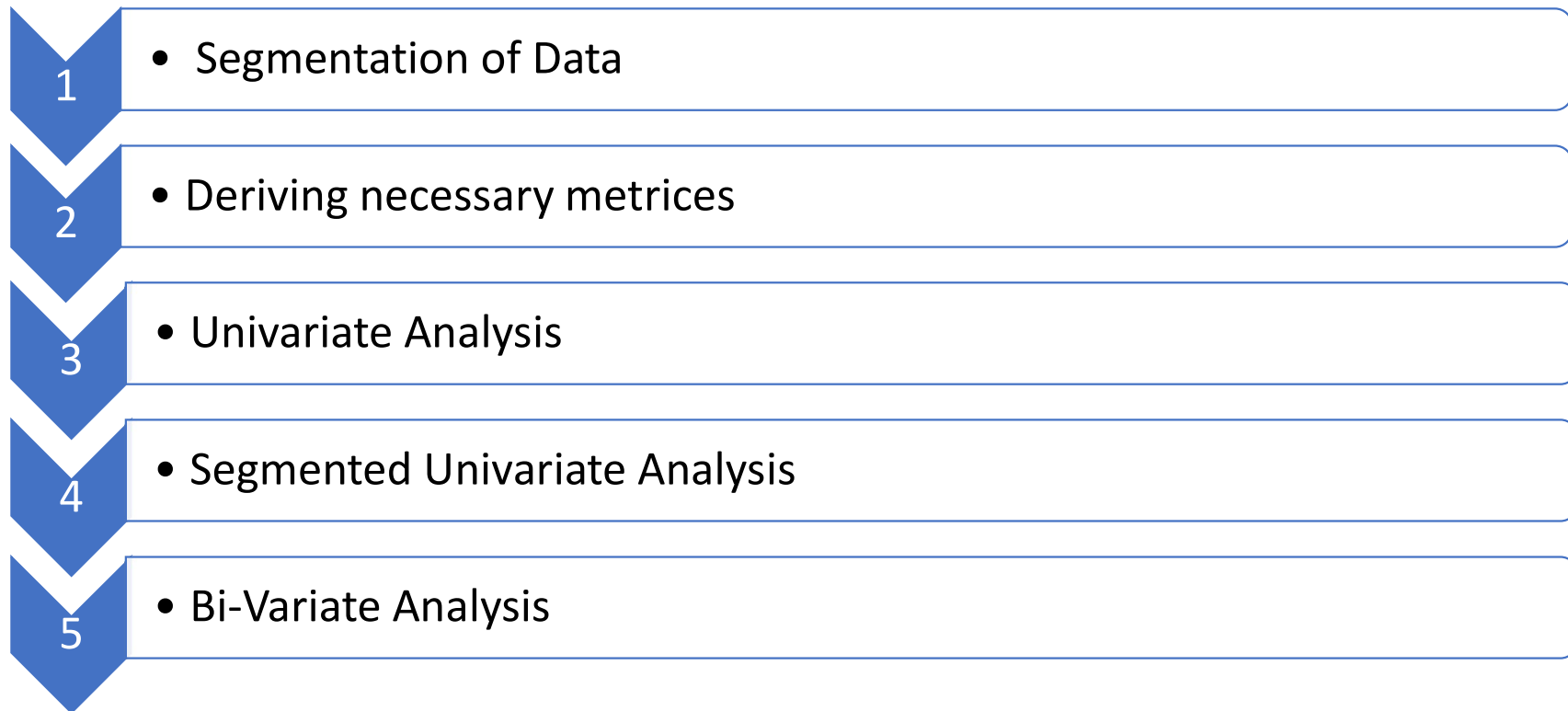
Loan Status
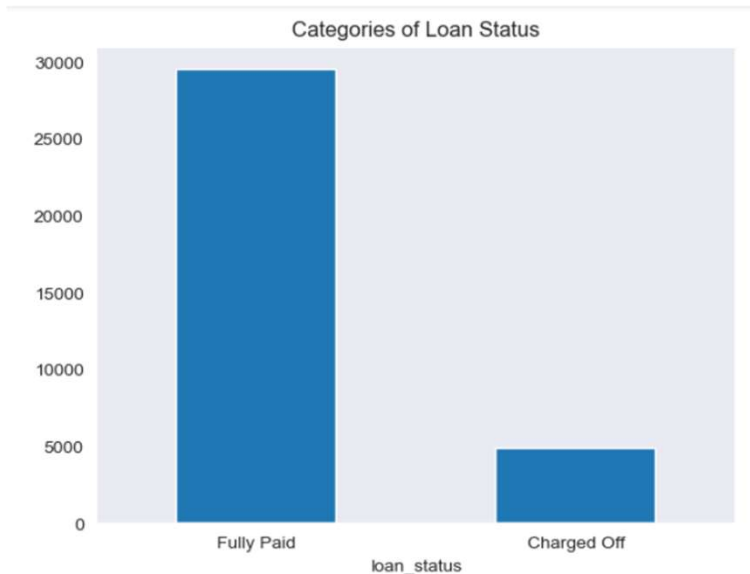
Term of Loan

- There are very few defaulters compared to non-defaulters among the loan applicants
- Customers with Charged off (Defaulters) are less than 5000 i.e., ~ 15 %.
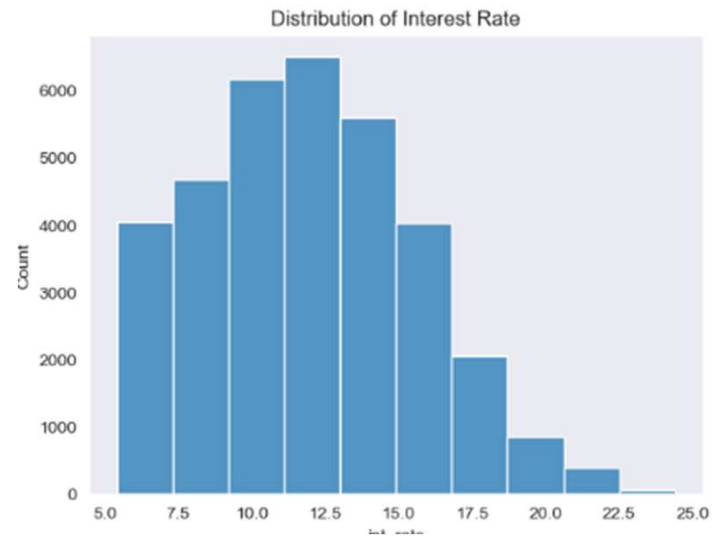- Customers with Fully paid (Non-defaulters) are about 3000 i.e., ~ 85 %.
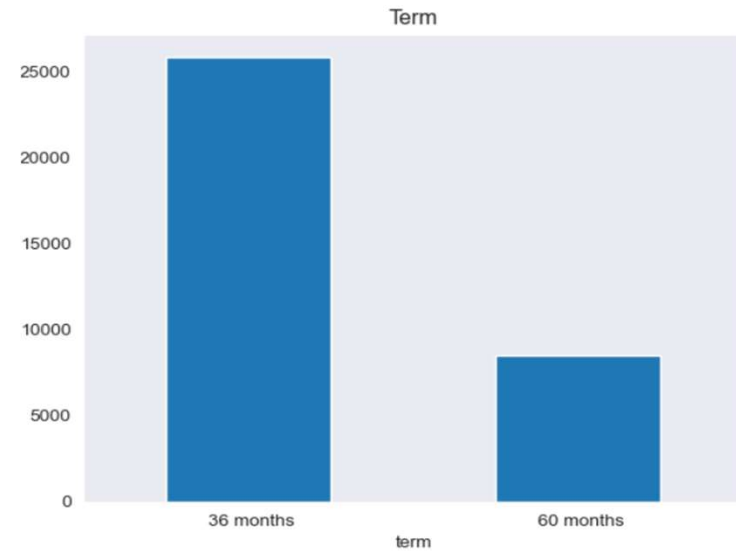
- More than 50% of loan applicants have opted for 36 months term compared to 60 months term.

Loan Amount

Interest Rate

Distribution of Loan Amount

Distribution of Interest Rate
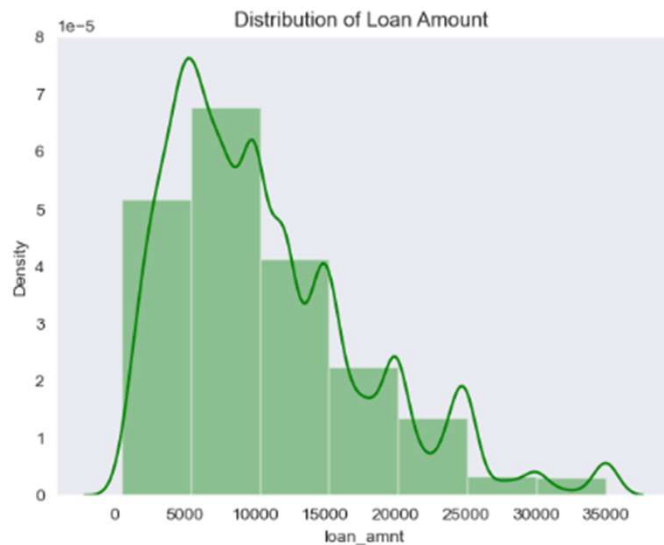
- Max Loan amount is dispersed in 5K-10K range
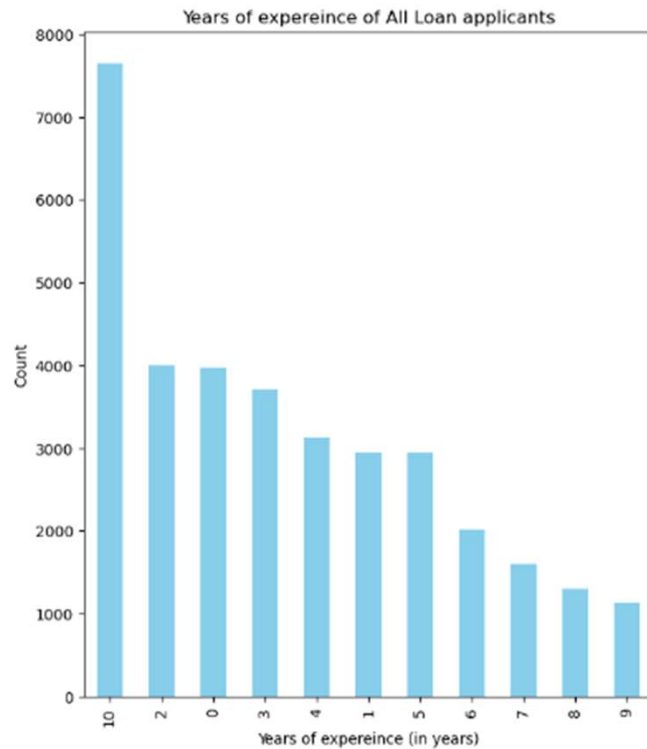- Loan amount is not very high

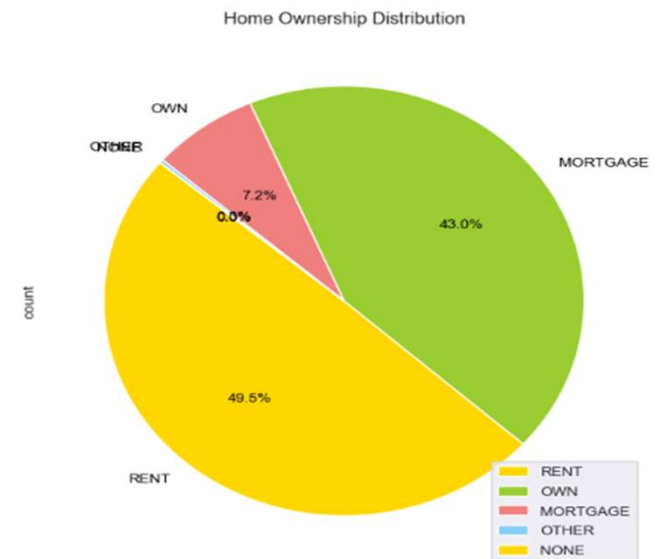- Most of the approved loans incurred an maximum interest range of 10-12.5% interest
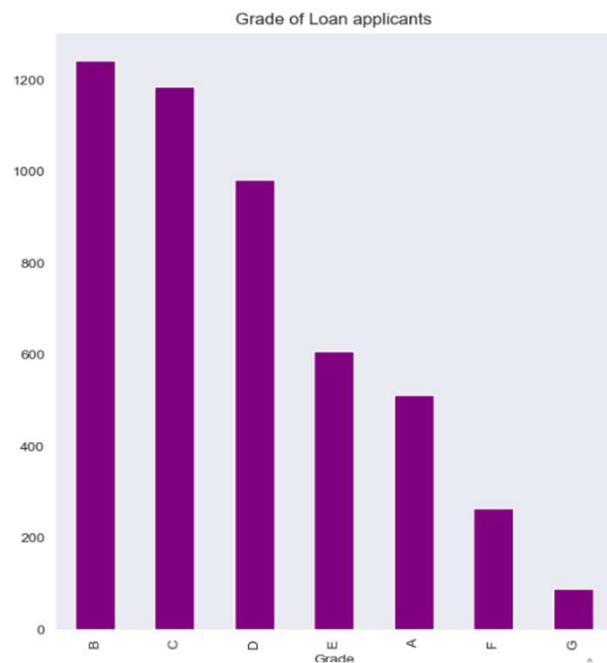
Applicants work experience   Home Ownership

- **Maximum Loan applicants have 10+ years of job experience**
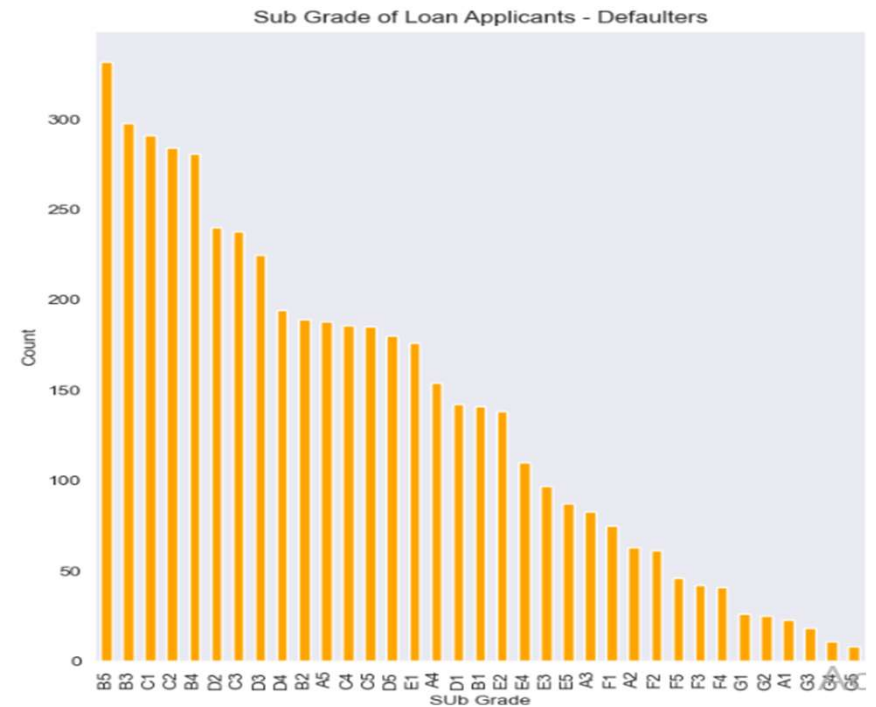

- **Maximum loan applicants have either rented property / mortgaged property.**
- **Very small applicants have a own house (7.2%)**
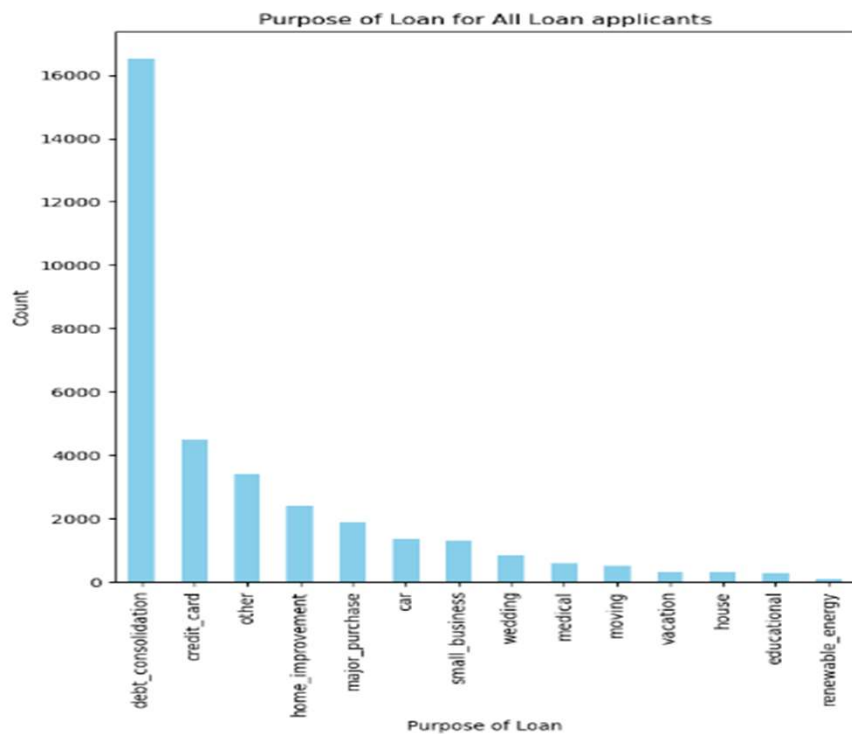
Grades (Defaulters)

Grade of Loan applicants

- **Defaulters are mostly B,C,D Grade**

Subgrades (Defaulters)
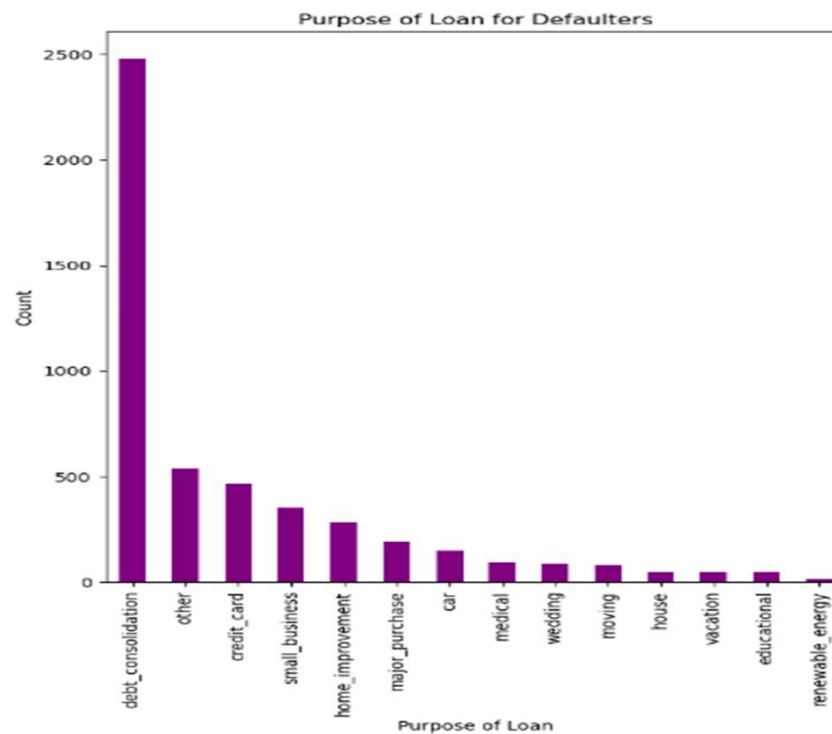

Sub Grade of Loan Applicants - Defaulters

- **Majority of the loan applicants are from Grade B sub categories [B3,B4,B5,C1,C2,C3,D3,D4]**
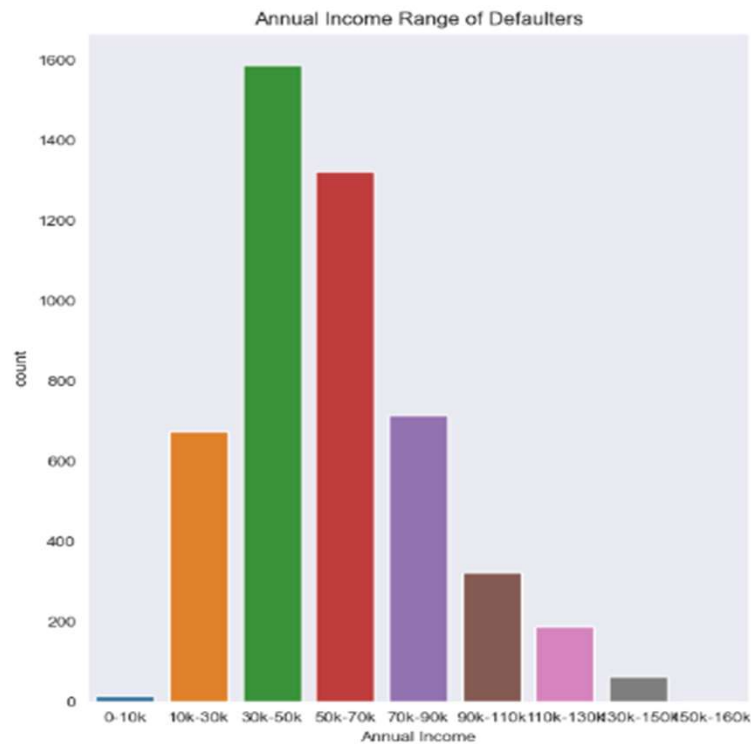
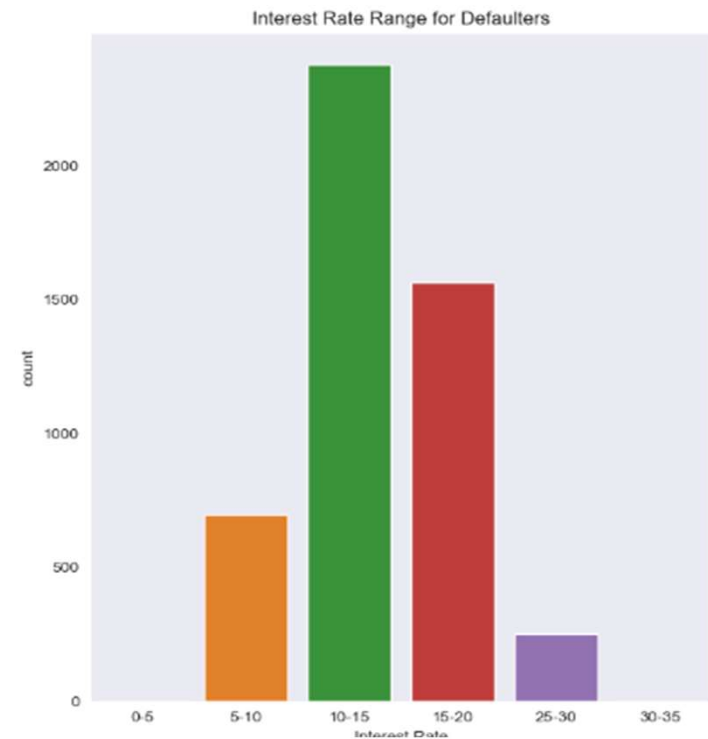# Purpose of Loan (All applicants)

Purpose of Loan (Defaulters)

- Maximum Loan applicants took loan for debt consolidation purpose
- Maximum Defaulters also took loan for debt consolidation purpose
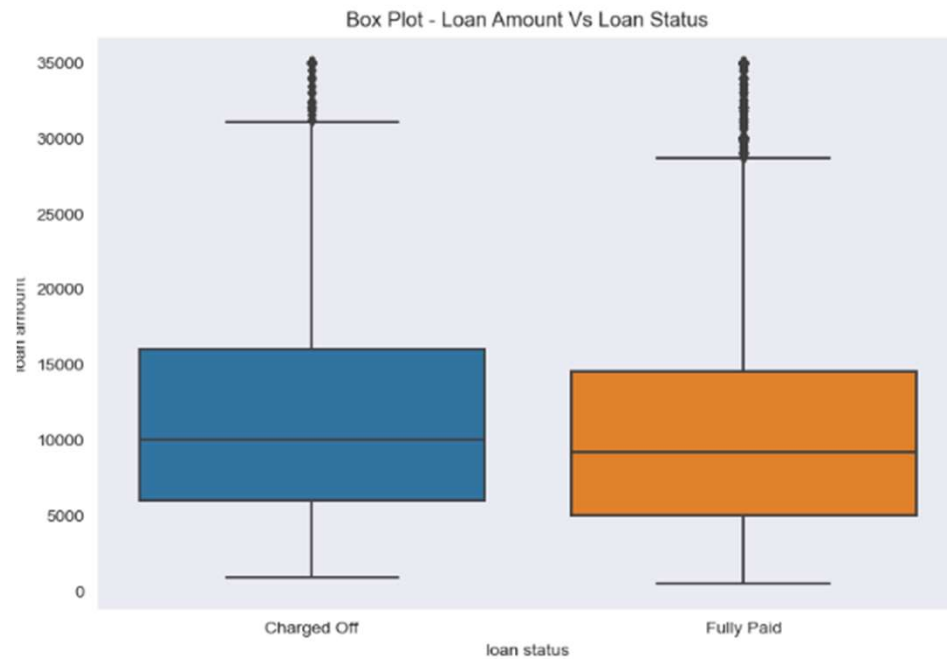
Annual Income (Defaulters)

Interest Rate (Defaulters)

Annual Income Range of Defaulters

Interest Rate Range for Defaulters

Maximum defaulters are in the 30-50K and 50-70K Income Range and 10-20% Interest Loan range

Loan Amount vs Loan Status

Box Plot - Loan Amount Vs Loan Status

Maximum defaulters are in the 30-50K and 50-70K Income Range and 10-20% Interest Loan range

Purpose vs Annual Income

Loan Status -Purpose vs Annual Income
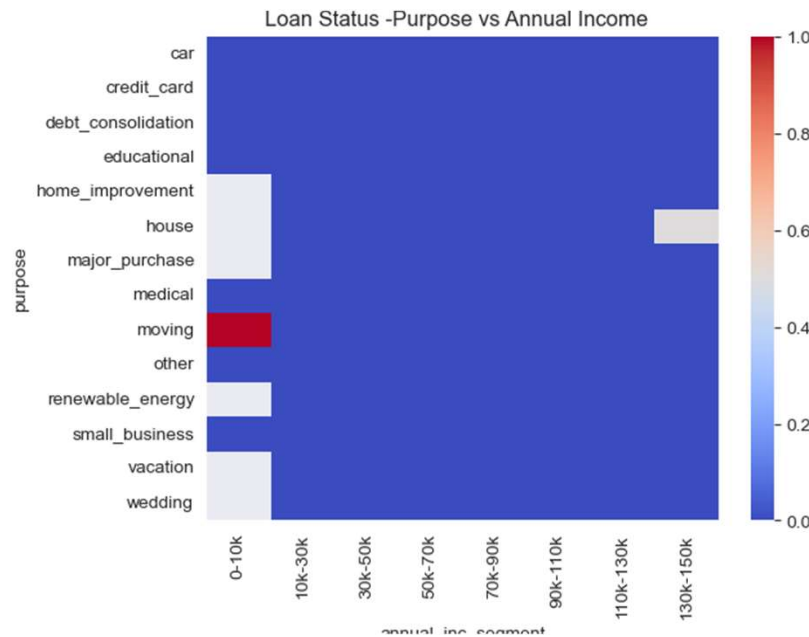
-Loan applicants in LowerIncome Range mostly availed loans for wedding, vacation, renewable energy, moving, major_purcahse, house and home improvement and have 50% chance of defaulting. Highest default is for loans taken for moving purpose

 Highest default in lower income is moving

Loan applicantis in the Higher income range tend to default ONLY for the purpose of house.

Purpose vs Loan Amount

Purpose vs Loan amount for Loan status
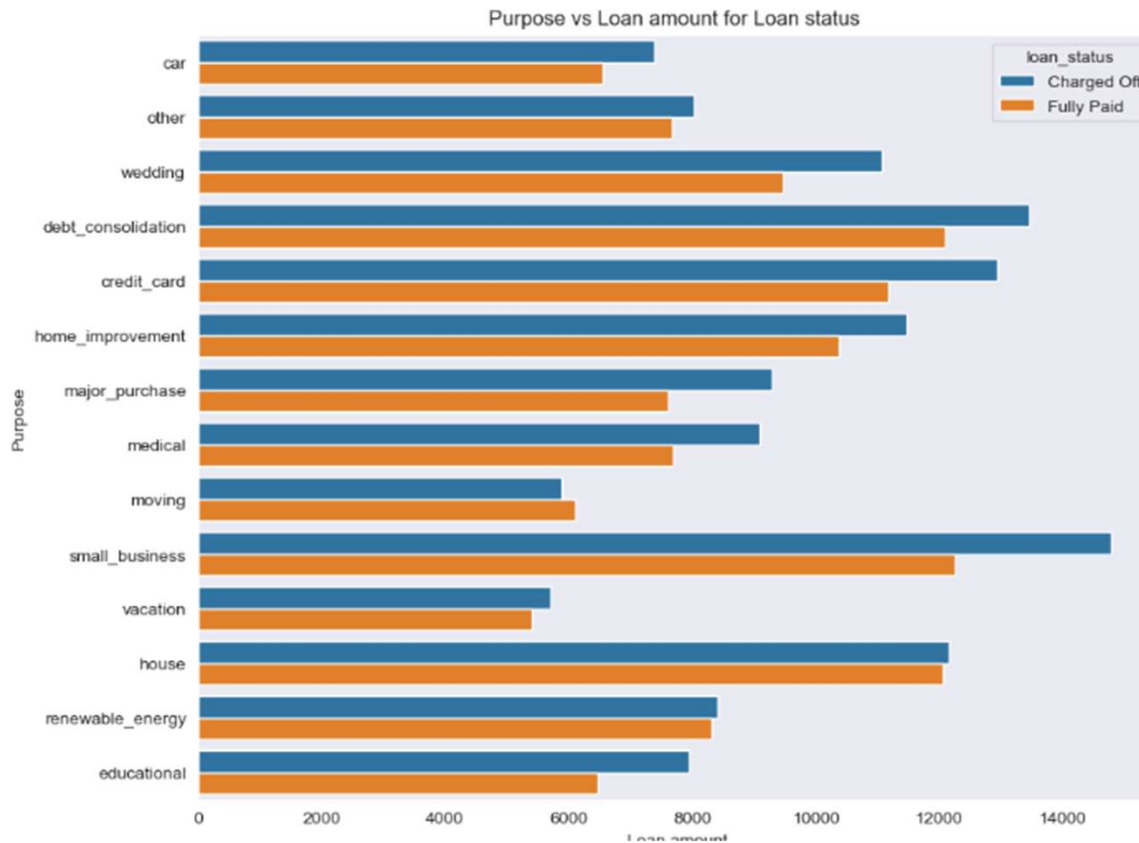
- **Maximum defaulters have availed Higher Loan amounts taken for small business, Debt Consolidation ,House, Credit Card, Wedding & Home improvement**

- **Highest loan amount approved for small business have maximum defaulters**

Open Account vs Annual Income

Loan Status - Open account vs Annual Income
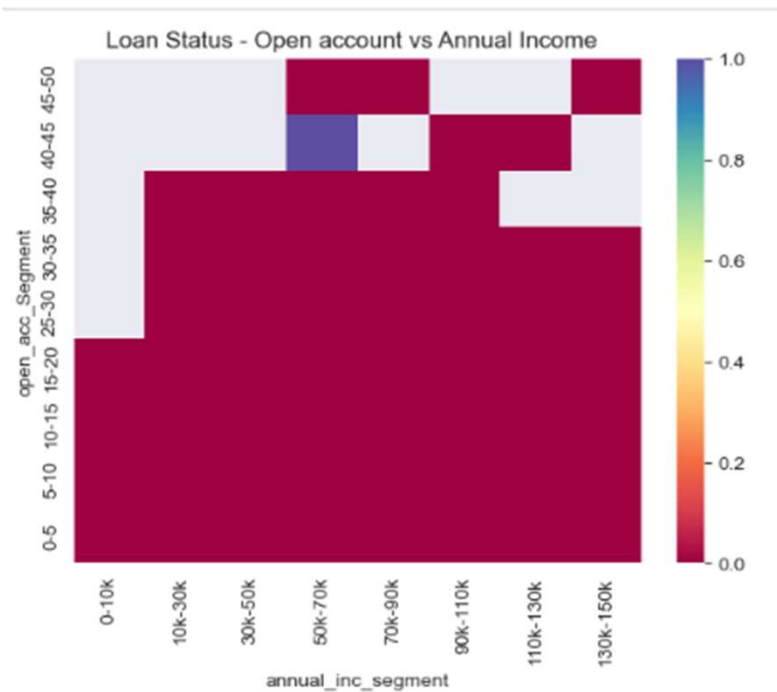
- Loan applicants of Mid Income Range 50K-110K , usually have less open accounts . But if they have more open accounts (40-45) , they are MOST likely candidates to default
- Loan applicants of lower income range (<50K) mostly tend to have more open accounts , they have a 50% chance to default
- Loan applicants of Higher income range(> 110K)-very few tend to open accounts. But if they have more open accounts (>35) they have 50% likely chances to default

Home Ownership vs Annual Income based on Defaulters
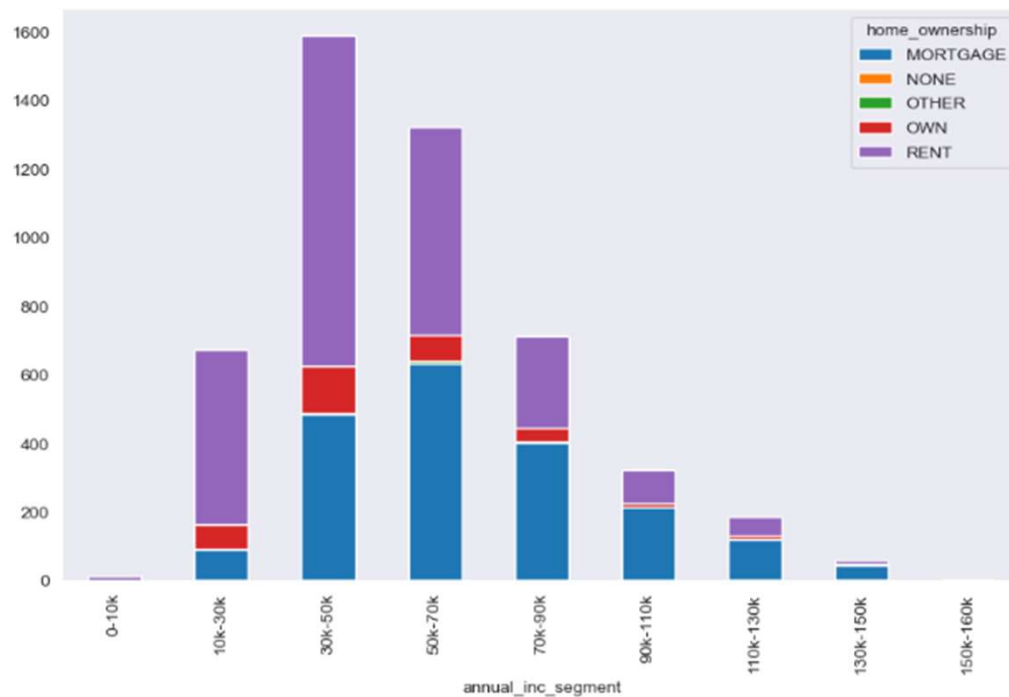
- **Most of the defaulters were in the rented category in home ownership and Income Range 30K-50K, 50K-70K**

Analysis Summary

High Risk of defaulters :

- **Lower Income Range (<30K) and Loan Purpose is Moving**
- **Higher income range tend to default ONLY when the Loan purpose is House.**
- **Mid Income Range (50K-110K) and has more open accounts**
- **Mid Income Range (30K-70K) with Home ownership as "Rented"**
- **Mid Income Range (30K-70K) with 10-20% Interest Loan range**
- **High experience of Job experience (10+years)**
- **Higher Loan amount -> high defaulting**
- **Grade – B,C,D and Sub Grades - B3,B4,B5,C1,C2,C3,D3,D4**
- **Loan Purpose – Debt consolidation**
- **Trend to default loan in the last few months of the year (Nov-Dec)**

Moderate Risk of Defaulters:

- **Lower Income Range (<30K) and Loan Purpose is Moving and Purpose is wedding, vacation, renewable energy, moving, major purchase, house and home improvement**
- **Large number of open accounts (Any income Range)**
- **Home ownership – Rented, Mortgaged**