

# **INVESTMENT ASSIGNMENT**

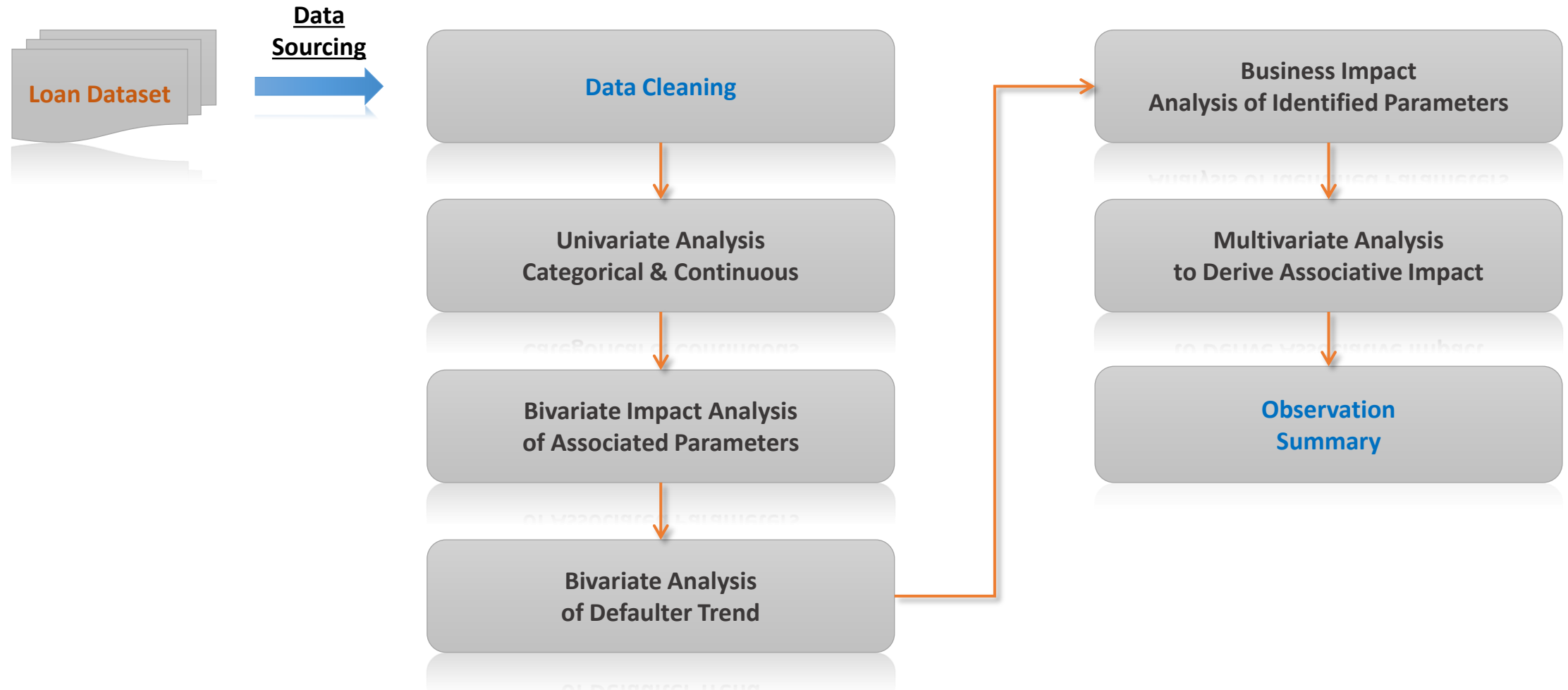
## **SUBMISSION**

Name: Fasal Shah and Mahadevaprasad Chandrashekariah

# Abstract

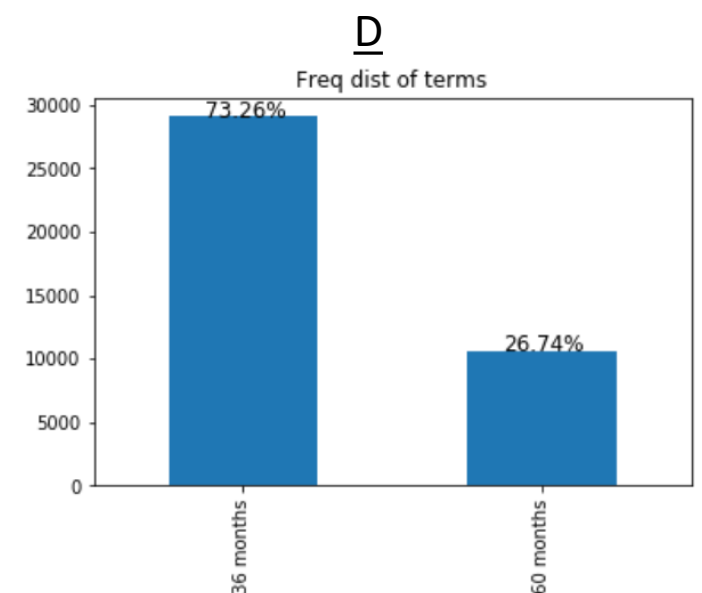
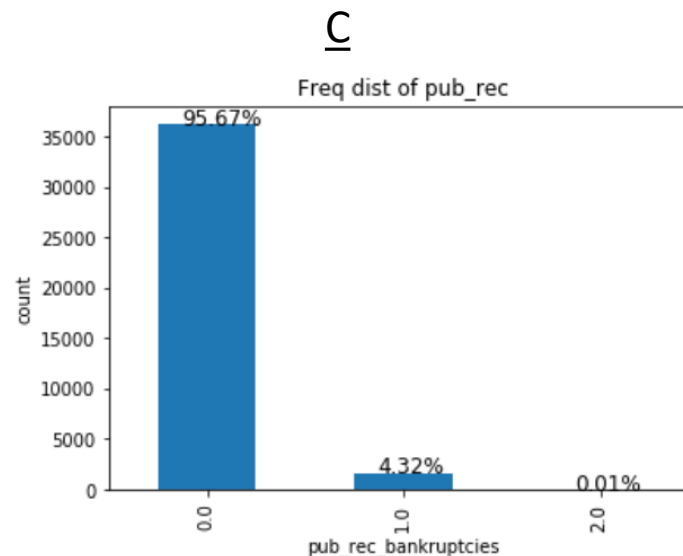
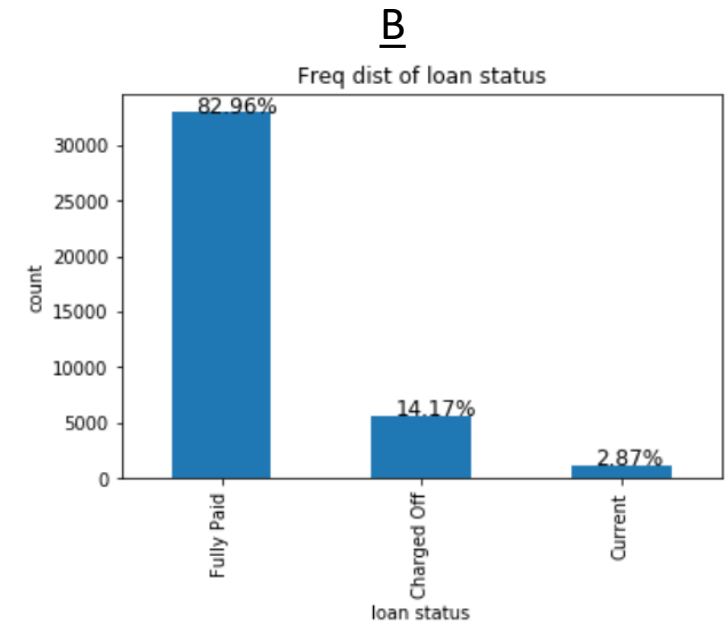
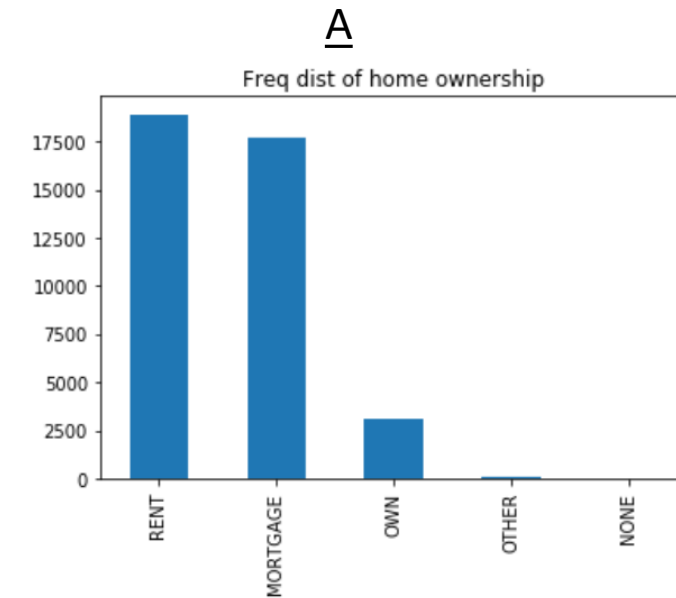
- Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- In this case study, we will understand how consumer attributes and loan attributes influence the tendency of default.
- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan.

# Problem Solving Methodology



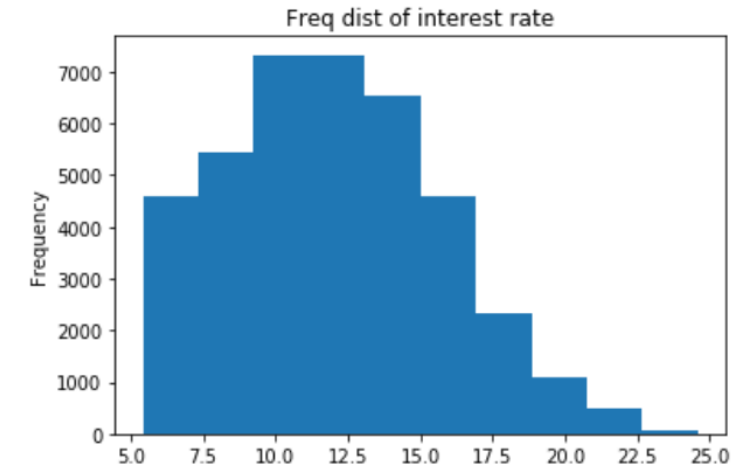
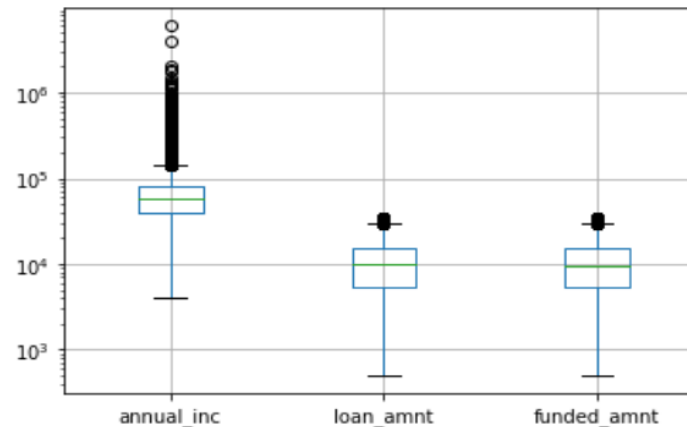
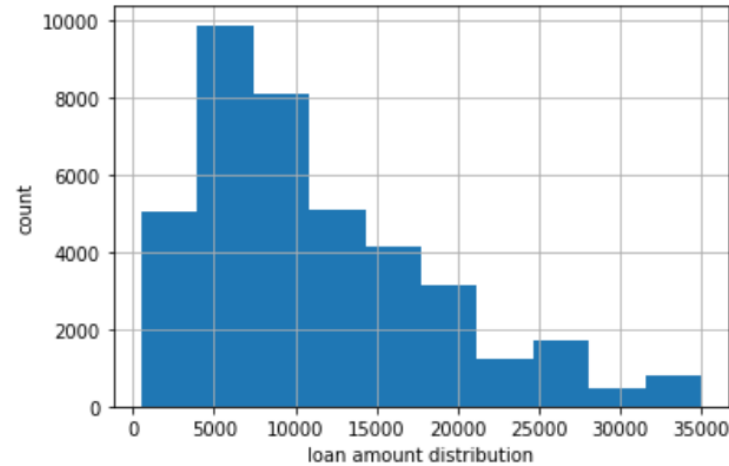
# Univariate Analysis: Categorical Data

- **Home Ownership (Plot A)**
  - Observed more than one type of borrowers. Use for further analysis.
- **Loan Status (Plot B)**
  - About 14.17% defaulters among the total borrowers.
- **Public Record Bankruptcy (Plot C)**
  - We can observe that 4.32% of the borrowers have 1 bankruptcy record. A small % 0.01% of the borrower's have been bankrupt twice.
- **Loan Term (Plot D)**
  - Observed two types of terms 60 months and 36 months. Analyze further.



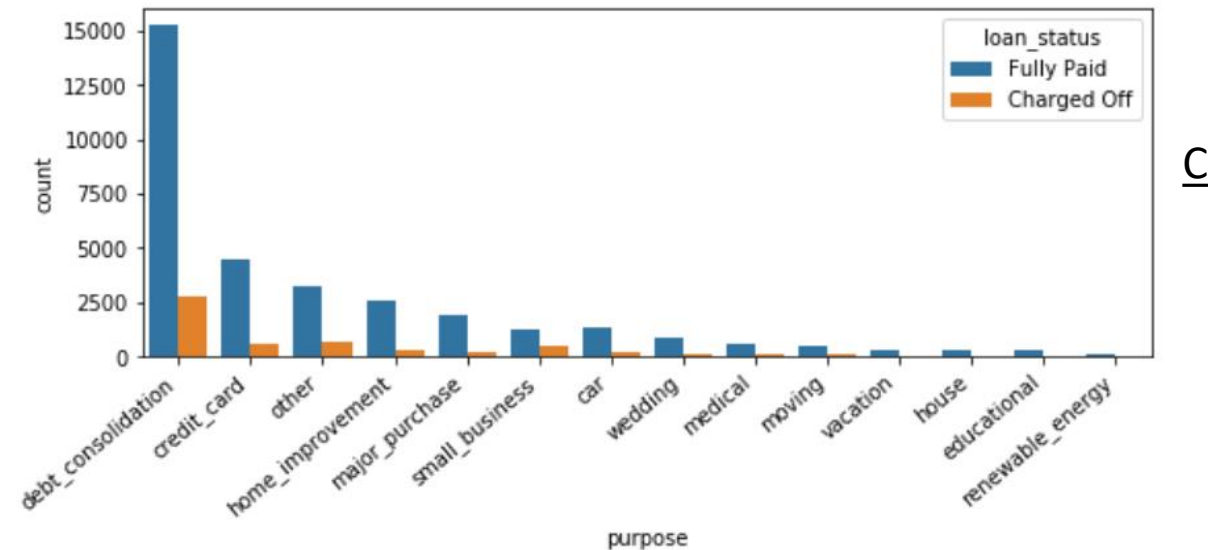
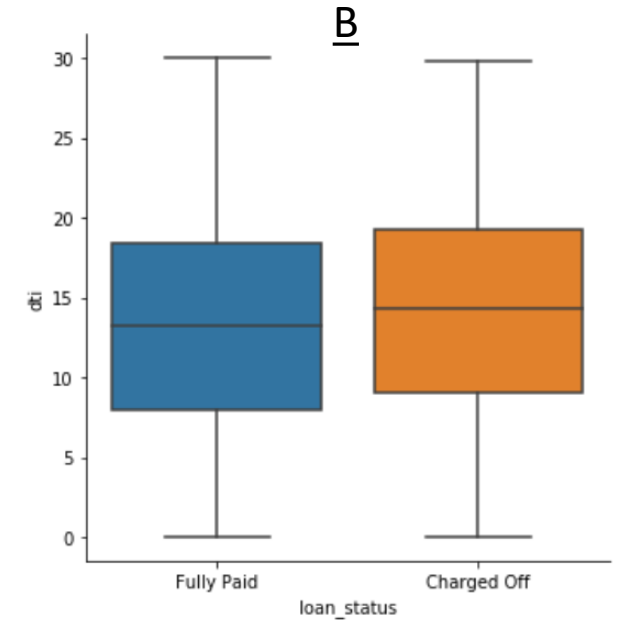
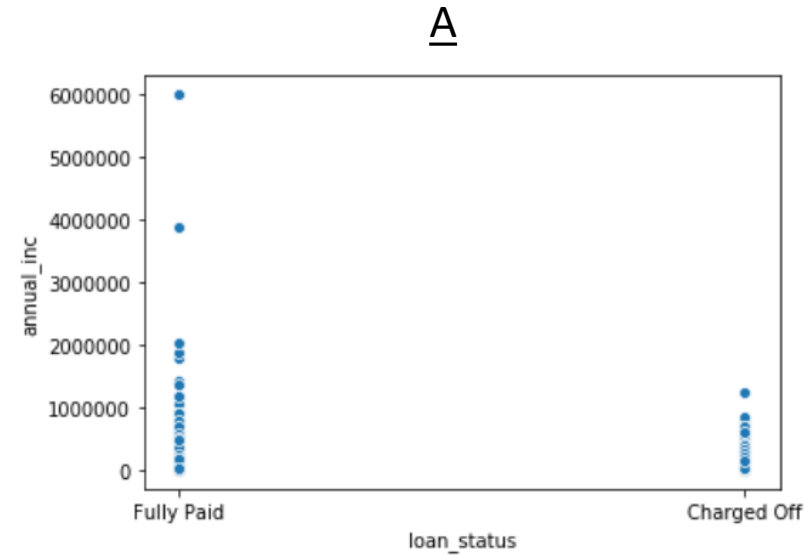
# Univariate Analysis: Continuous Data

- **Loan Amount Distribution**
  - Varying loan amounts observed. Lot of borrowers are availing loans between 5000 to 20000 USD.
  - Need to analyze their impact on other variables.
- **Interest Rate Distribution**
  - Majority of the loans are offered with interest rates ranging anywhere between 5% to 22.5%.
  - Analyze the impact of interest rate over defaulters.
- **Determine outlier impact**
  - Annual income has a lot of outlier. Not treating them as it can skew the data.
  - Loan amount and Funded amount have almost same distribution (as the correlation value is 100) . Using one of them can be sufficient for further analysis.



# Bivariate Analysis:

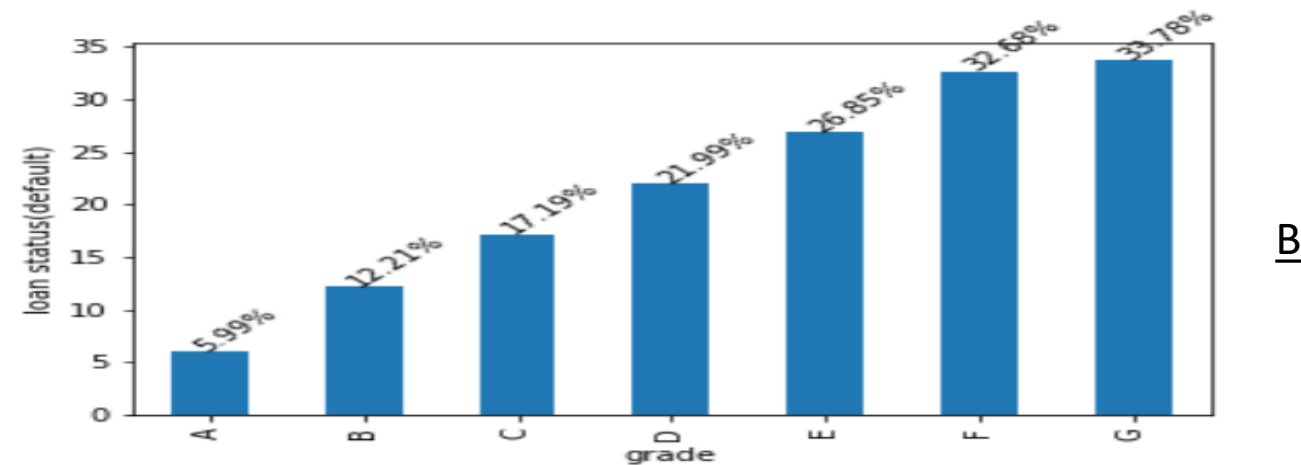
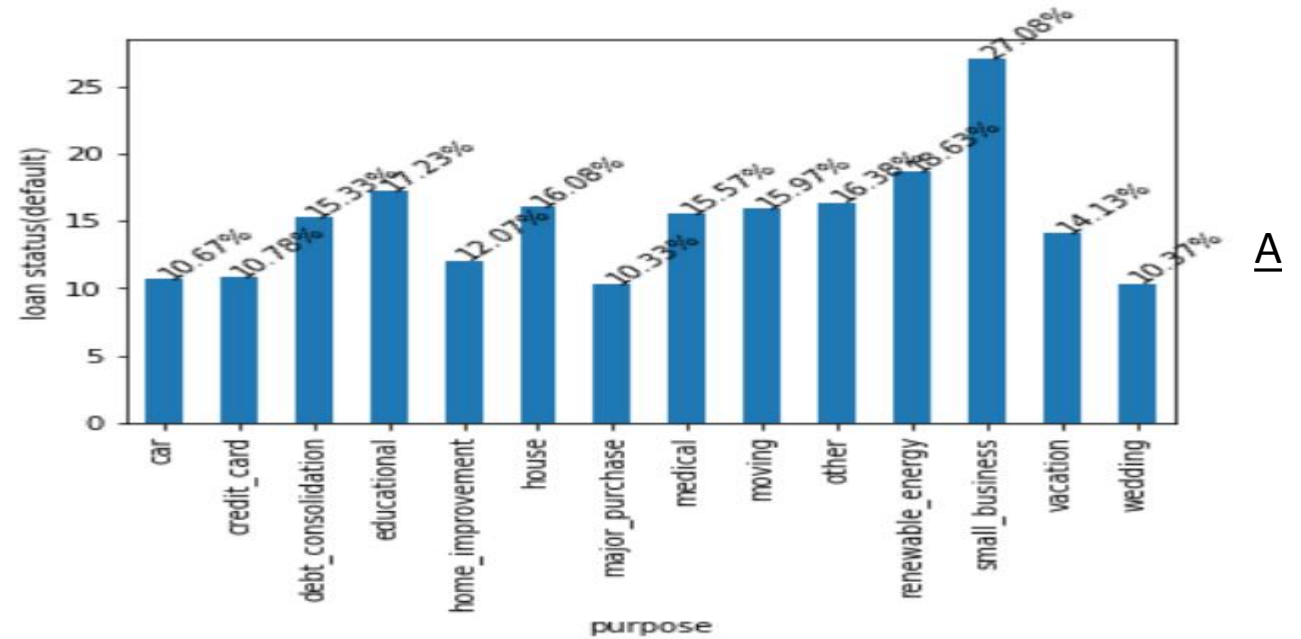
- *Note:* Removed the data with loan status as 'Current' before moving on with bivariate analysis.
- Impact of Annual Income on loan status (Plot A)
  - From the plot it can be observed that most of defaults are in the annual income bracket <100000. We will look at it in more granularity at a later stage.
- Impact of DTI over loan status (Plot B)
  - From the box plot it can be observed that defaulters range is only slightly shifted positively. Not significant delta to be considered for further analysis.
- Impact of purpose over loan status. (Plot C)
  - “Debt Consolidation” has the highest amount of defaulters as well as it is the largest in terms of sheer volume of applications. Investigate further.



# Bivariate Analysis: Defaulter Trend (I)

*Note: Analyze the defaulter data with respect to other parameters.*

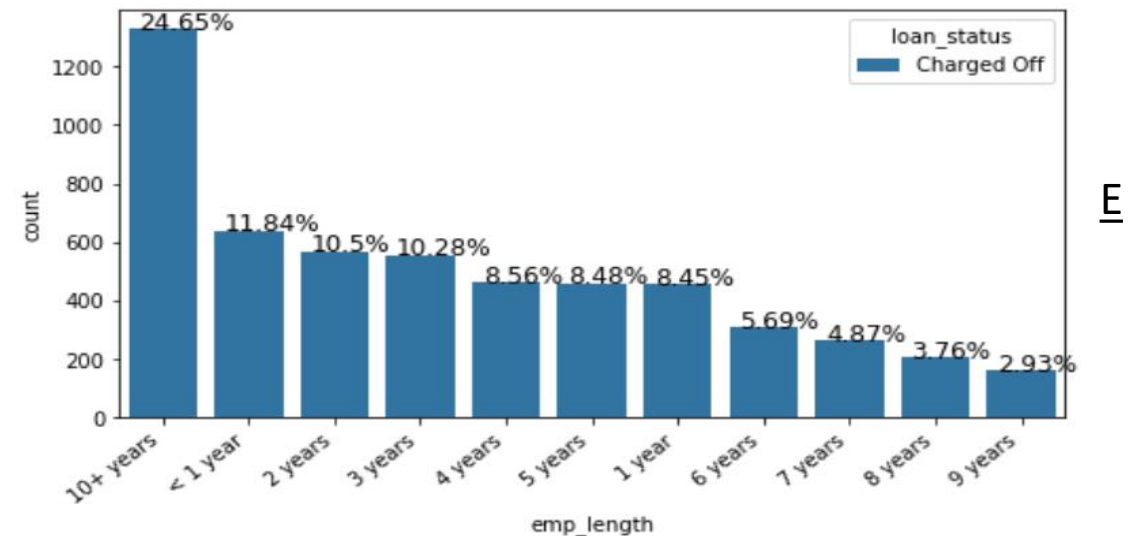
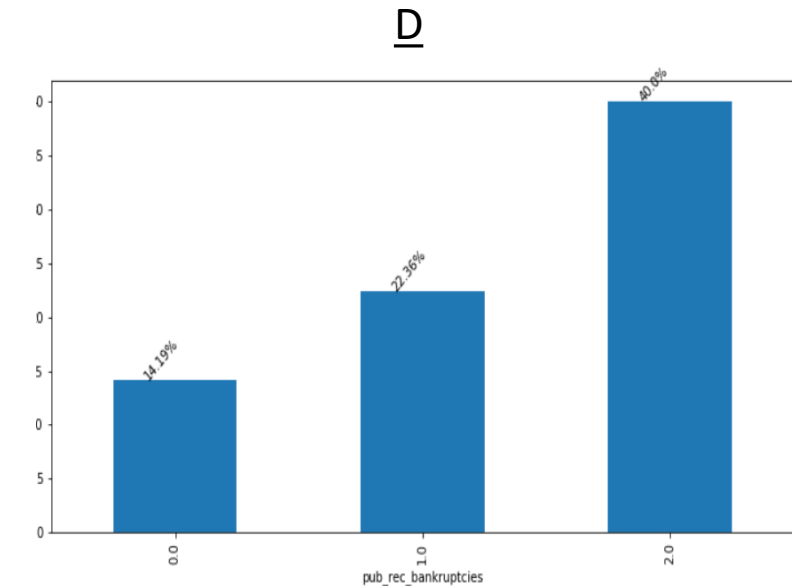
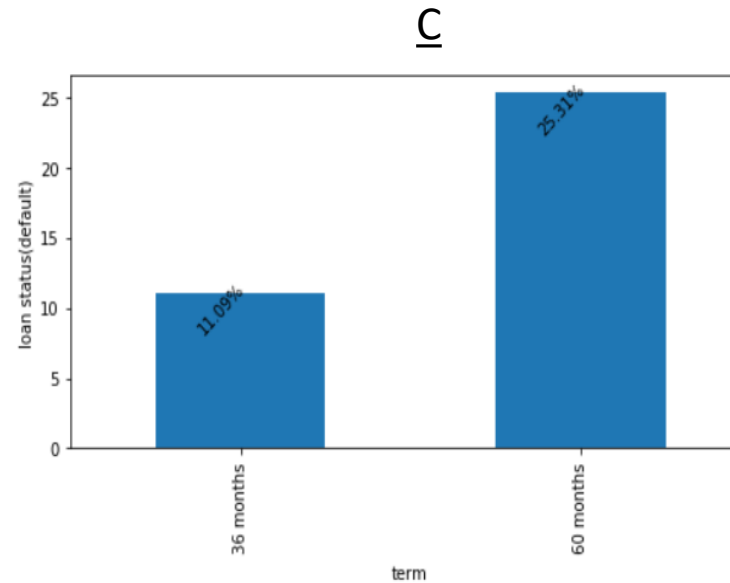
- Trend with loan Purpose (Plot A)
  - We can observe from the plot that “small business” has highest ratio defaulters followed with “renewable energy” and “educational” loans. “Debt consolidation” is not among the top three but since it has the highest volume it cannot be ignored.
- Trend with employee Grade (Plot B)
  - We can observe that employees with grade G, F, E and D have the highest ratio of defaulters.



# Bivariate Analysis: Defaulter Trend (II)

*Note: Continuation of previous slide.*

- Trend with loan Term (Plot C)
  - Defaulter ratio is higher with 60 months tenure. Out of ~10K loans with 60 months tenure nearly 25% of them have defaulted.
- Trend with Bankruptcy Data (Plot D)
  - Defaulter ratio increases as bankruptcy rate increases.
- Trend with employee length (Plot E)
  - Employees over 10+ years of experience have the highest defaulting ratio (nearly 25%).





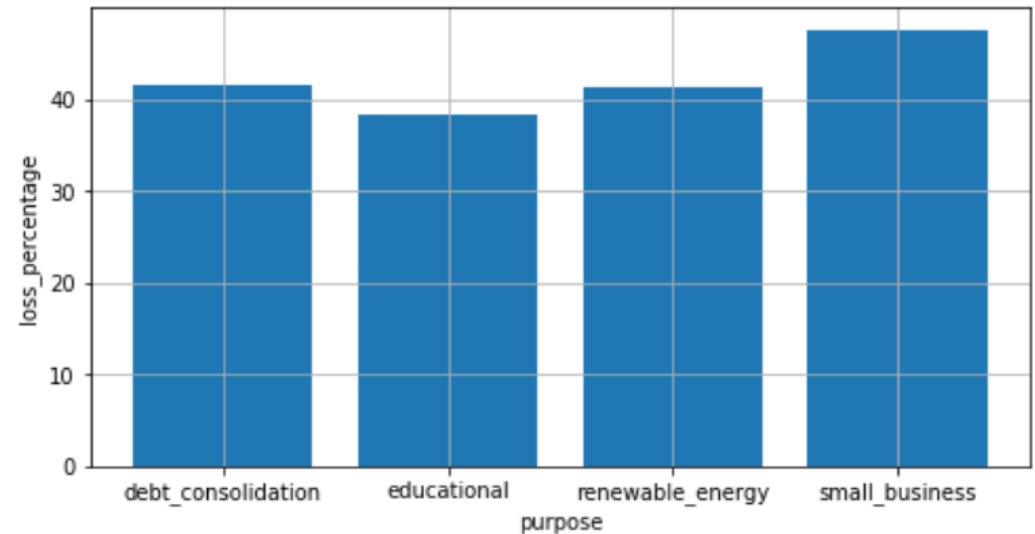
# Bivariate Analysis: Business Impact

- Analyze the business impact of loan purpose (identified top 4).
- Derive two new columns “loss\_percentage” and “loss\_value” based on columns “funded\_amnt” and “total\_pymnt” from the available data for defaulters.

\* Loss\_value does not account for the actual loss but represents the (total amount received by user) – (total amount paid by user (principal + interest)).

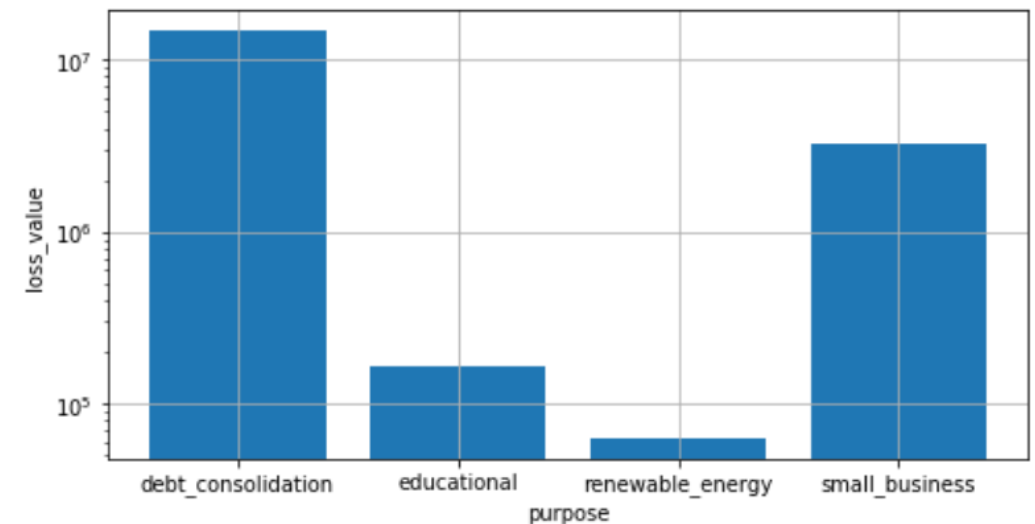
- From Plot A we can observe that “small business” has the highest percentage of loss per investment followed by “Debt Consolidation”, “Renewable Energy” and “Educational”.
- From Plot B we can observe that Debt Consolidation is the highest loss making purpose category among loans.

Average Loss Percentage/Purpose



A

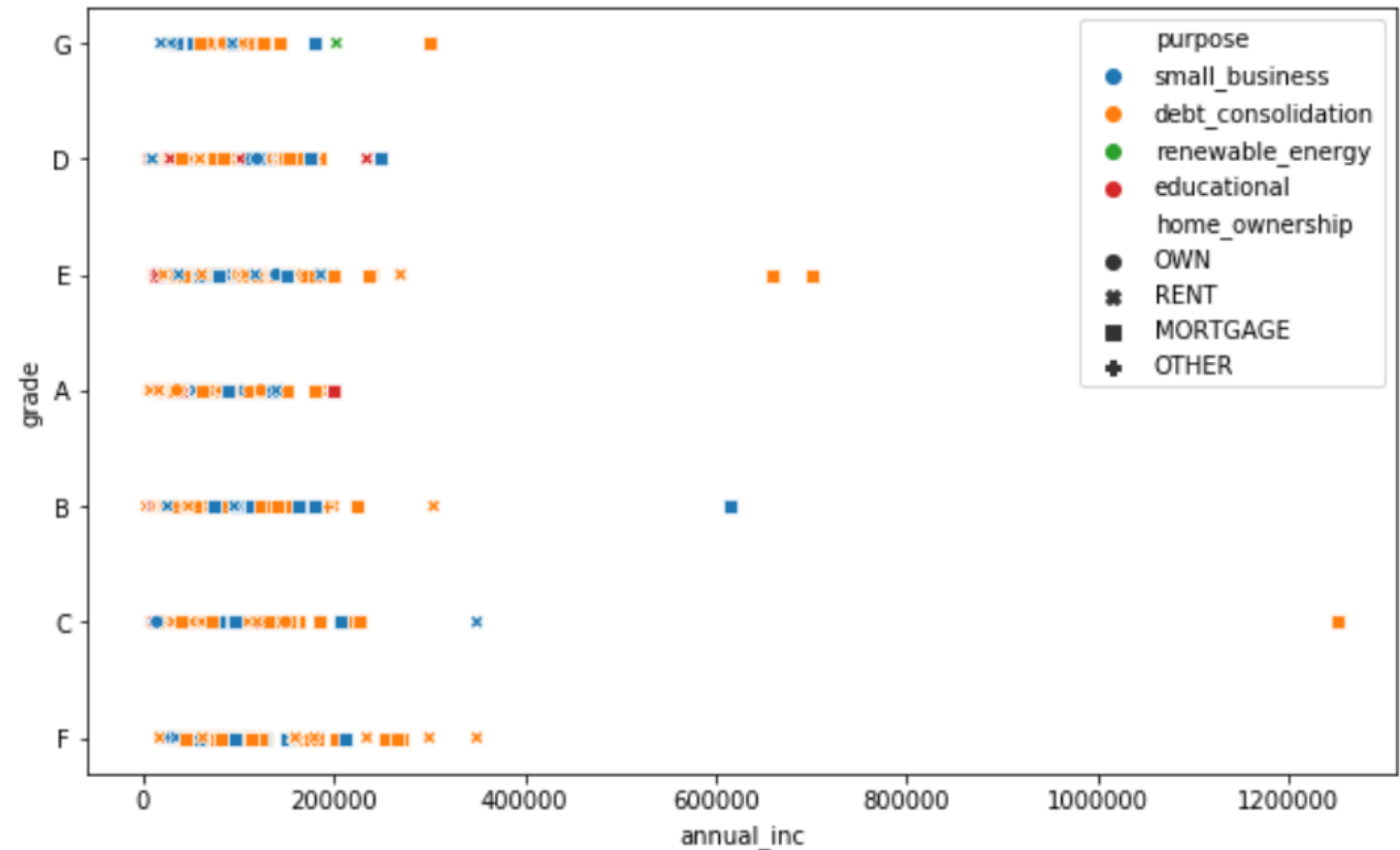
Total funded amount/Purpose (Log scale)



B

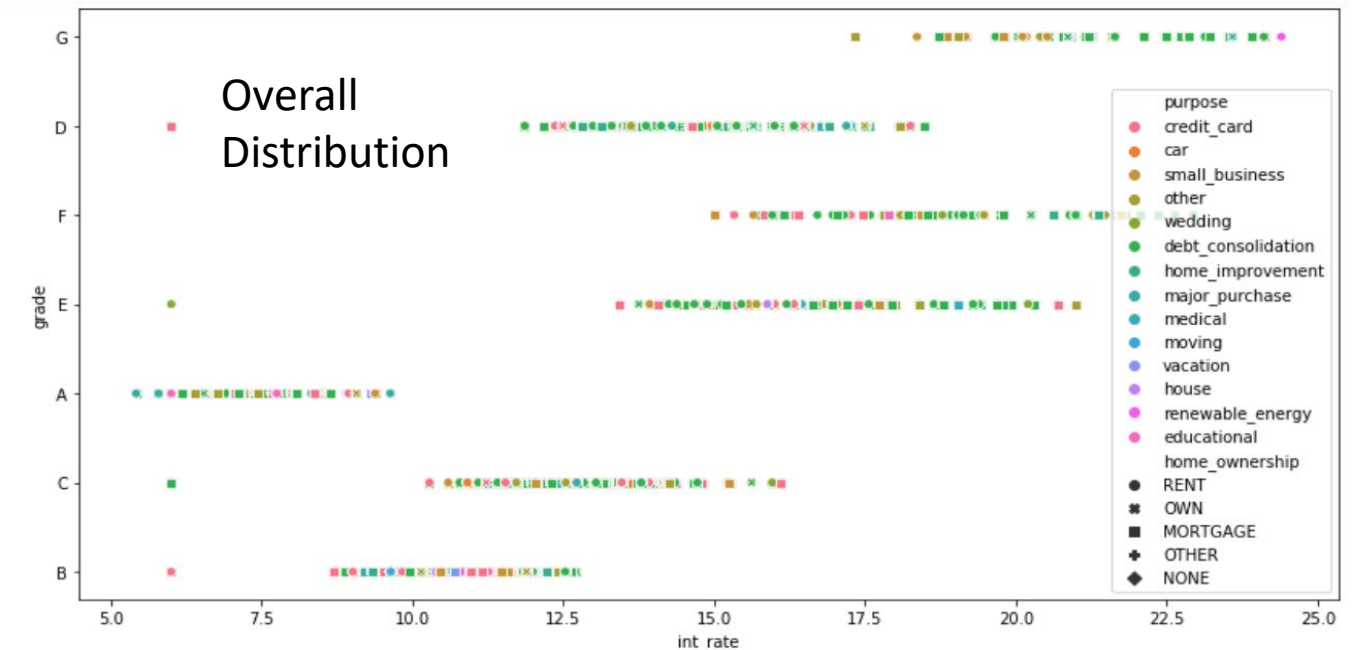
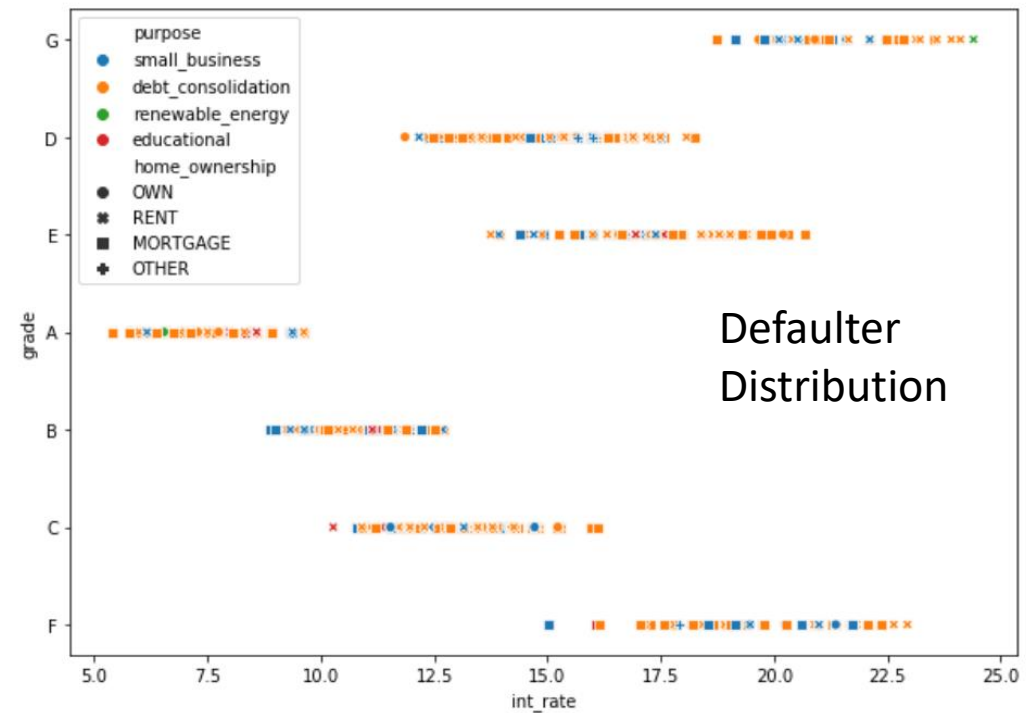
# Multivariate Analysis: Annual Income

- Impact of Annual Income on “Employment Grade”, “Home Ownership”, “Loan Purpose”.
  - From the plot shown we can observe that majority of the defaulters have an annual income of 200K or less across all grades.
  - There are a few with income range 200K to 400K but almost all of them are either having the home ownership status as “rented” or ‘mortgage’.
  - All the high income defaulters earling >400K are having the home ownership status as “mortgage”.
  - Annual income does not seem to make a huge impact across different employee grades or across the considered purposes.



# Multivariate Analysis: Interest Rate

- Impact of Interest Rate on “Employment Grade”, “Home Ownership”, “Loan Purpose” .
  - From the plots shown we can observe that there is a direct correlation between top defaulting grades and the applicable interest rates.
  - G, F, E and D had the highest ratio of defaulters in the order of highest to lowest. (refer to section **Bivariate Analysis: Defaulter Trend**)
  - Plot shows that the interest rates are also highest for G, followed by F, E and D respectively.
  - There is no significant impact of interest on purpose or home ownership status.



# Summary

- The key parameters impacting the lending company are Purpose, Annual Income, Interest Rate, Grade, Home Ownership, Term, Public Bankruptcy record and Employment Length.
- Loans applied with purposes as 'debt\_consolidation' and 'small\_business' are the highest loss making categories among defaulters.
- Grades G, F, E and D have been offered with high interest loans and also have the highest percentage of defaulters.
- Annual income criteria:
  - Income under 200K have the highest amount of defaulters
  - Income over 200K but under 400K are risky under home ownership types ('RENT' or 'MORTGAGE')
  - Income over 400K are risky if the home ownership type is ('MORTGAGE')
- Term with 60 months has the highest percentage of defaulters approximately 25% in the category.
- Public record of bankruptcies indicates default rate increases with increase in number of bankruptcies
- Employment length over 10 years have a high tendency to default with nearly 24% of them defaulting.