Gramener Case Study

Background

- Gramener is a consumer finance company which specialises in lending various types of loans to urban customers.
- As part of this case study, we want to analyse the loan applications that are risky.
- Two **types of risks** are associated with the bank's decision:
  - If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  - If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company.
- When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

Background

- **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below.
  - **Fully paid**
  - **Current**
  - **Charged-off**
- **Loan rejected**
  - Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company

Business Objectives

- Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).
- The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.
- In this case, the customers labelled as 'charged-off' are the 'defaulters’.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- As part of this case study Gramener wants to understand the **driving factors (or driver variables)** behind loan default.

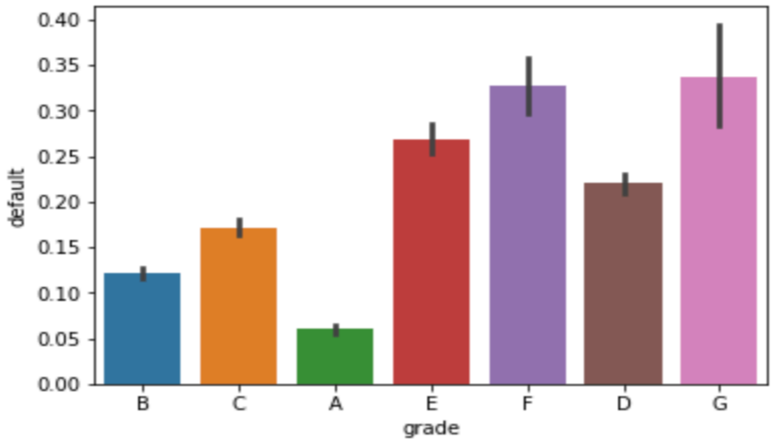
Analysis

- Data Understanding.
  - There are totally 39717 rows in the dataset
  - There are totally 111 columns in the data set
  - The data contains the loan applications of many customers.

- Data cleaning
  - Out of the provided columns, only 54 columns seem to be in a good state usable for our analysis. We can remove the rest of the columns.
  - Also some columns like annual_salary, interest rate needs to be grouped (derived metrics) before they can be consumed for analysis.
  - loan_status seems to be the dependent and final analysis variable

Analysis and visualizations

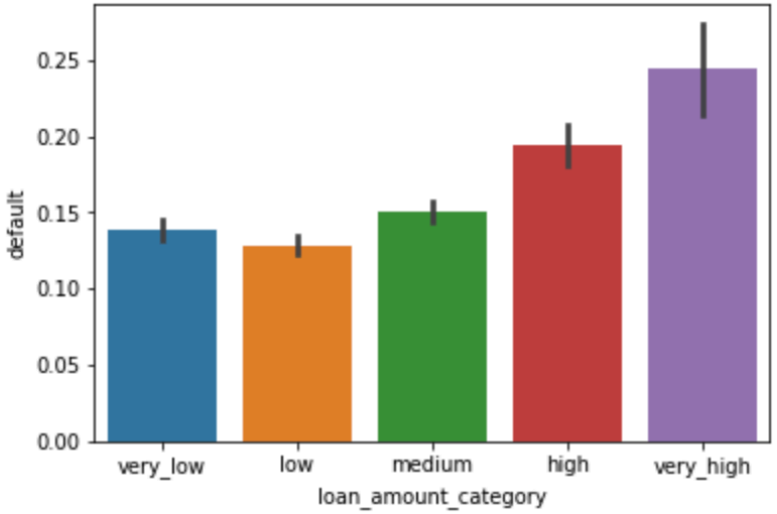
- We can use seaborn for plotting graphs for univariate and bivariate analysis
- Lets see the plot for grade grade by default rate.



- the above graph shows that the default rate goes higher in this order G > F > E > D > C > B > A

Analysis and visualizations

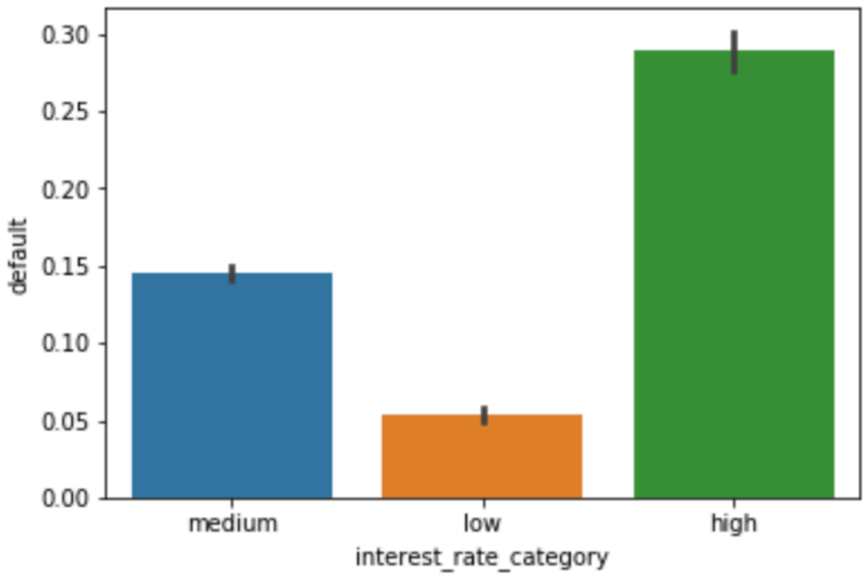
- Lets see graph of loan_amount_category vs default rate



- From the above analysis it seems that the default rate is high for higher loan amounts

Analysis and visualizations

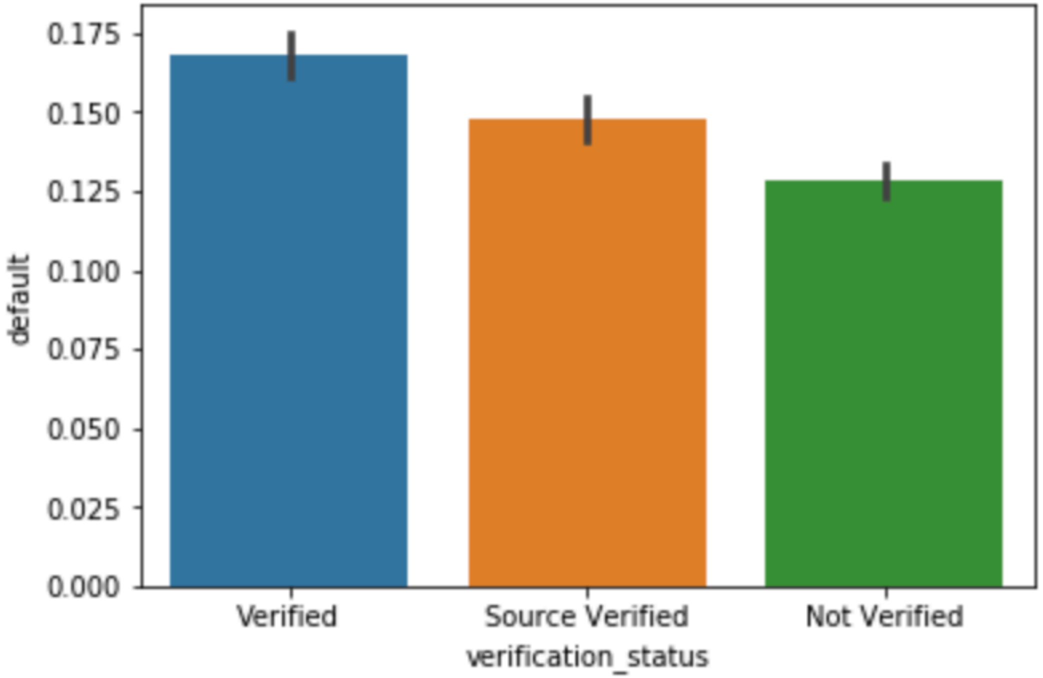
- Lets see graph of interest_rate_category vs default rate



- The above plot shows that high interest rates have higher default rates followed by medium and then low

Analysis and visualizations

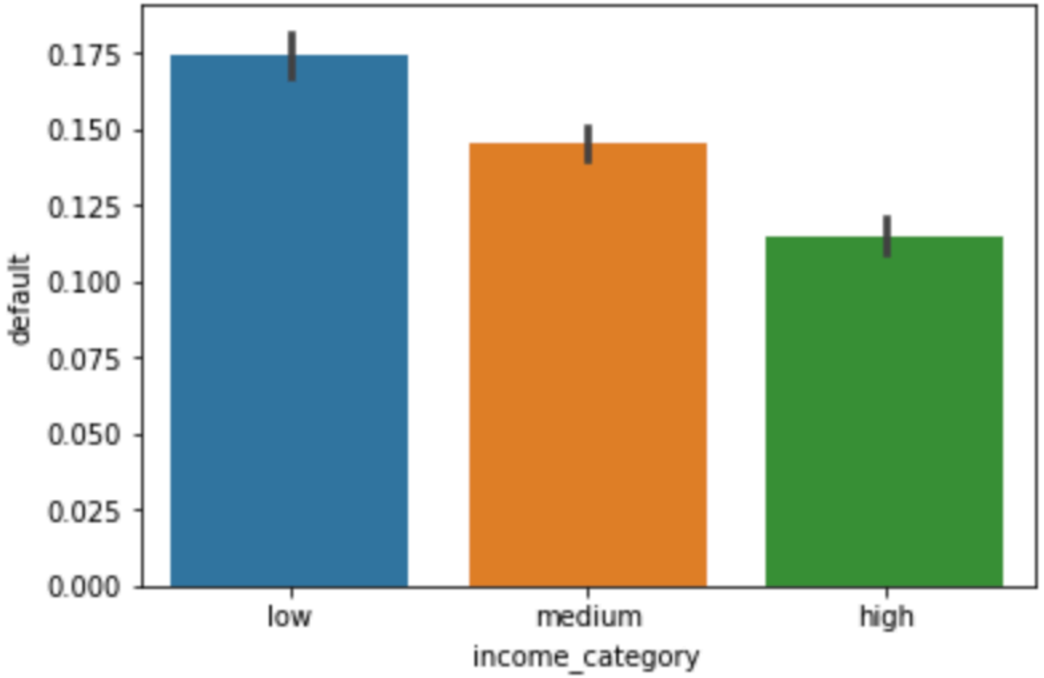
- Lets see graph of verification_status vs default rate



- verified loans have higher default rates

Analysis and visualizations

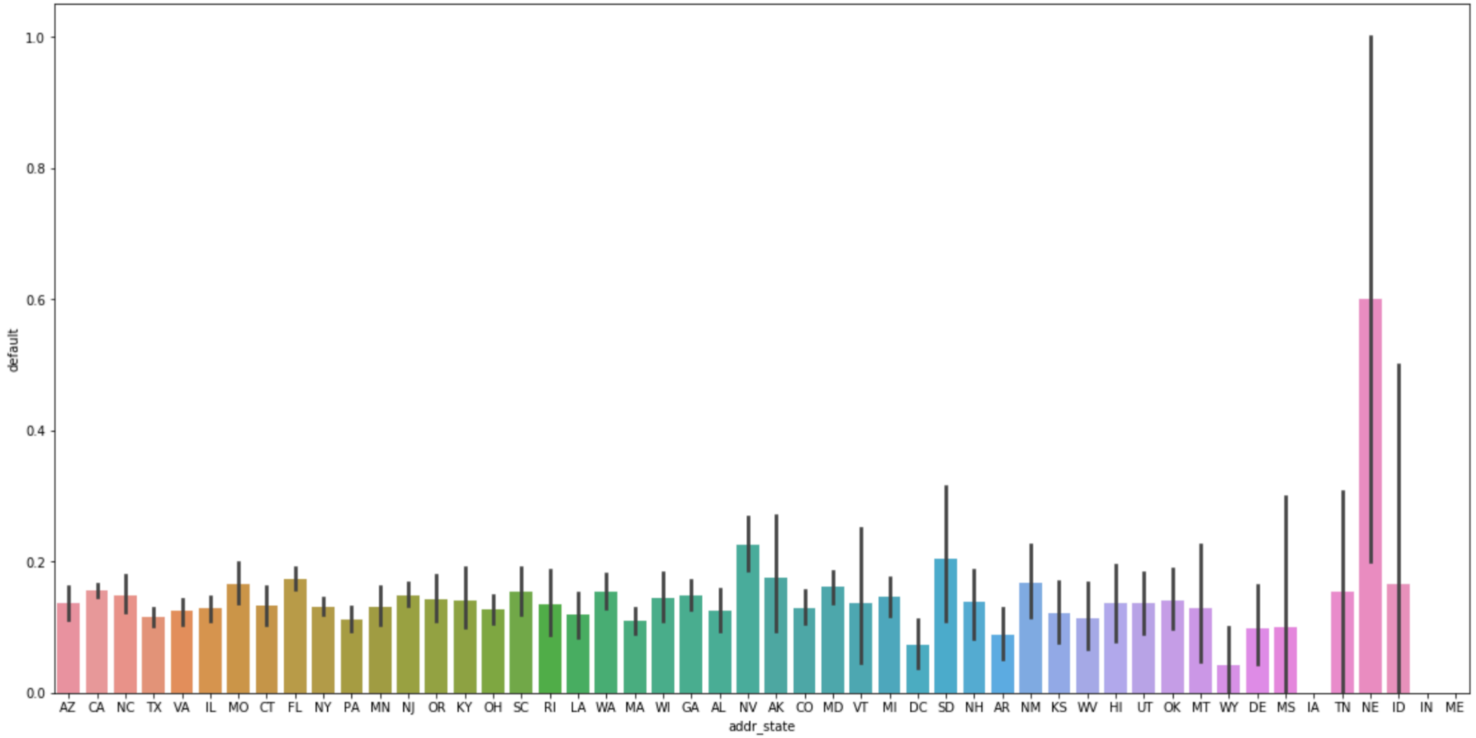
- Lets see graph of income_category vs default rate



- As expected lower income have higher default rates

Analysis and visualizations

- Lets see graph of state vs default rate



- Clearly Loan granted NE (Nebraska) has higher chance of default.

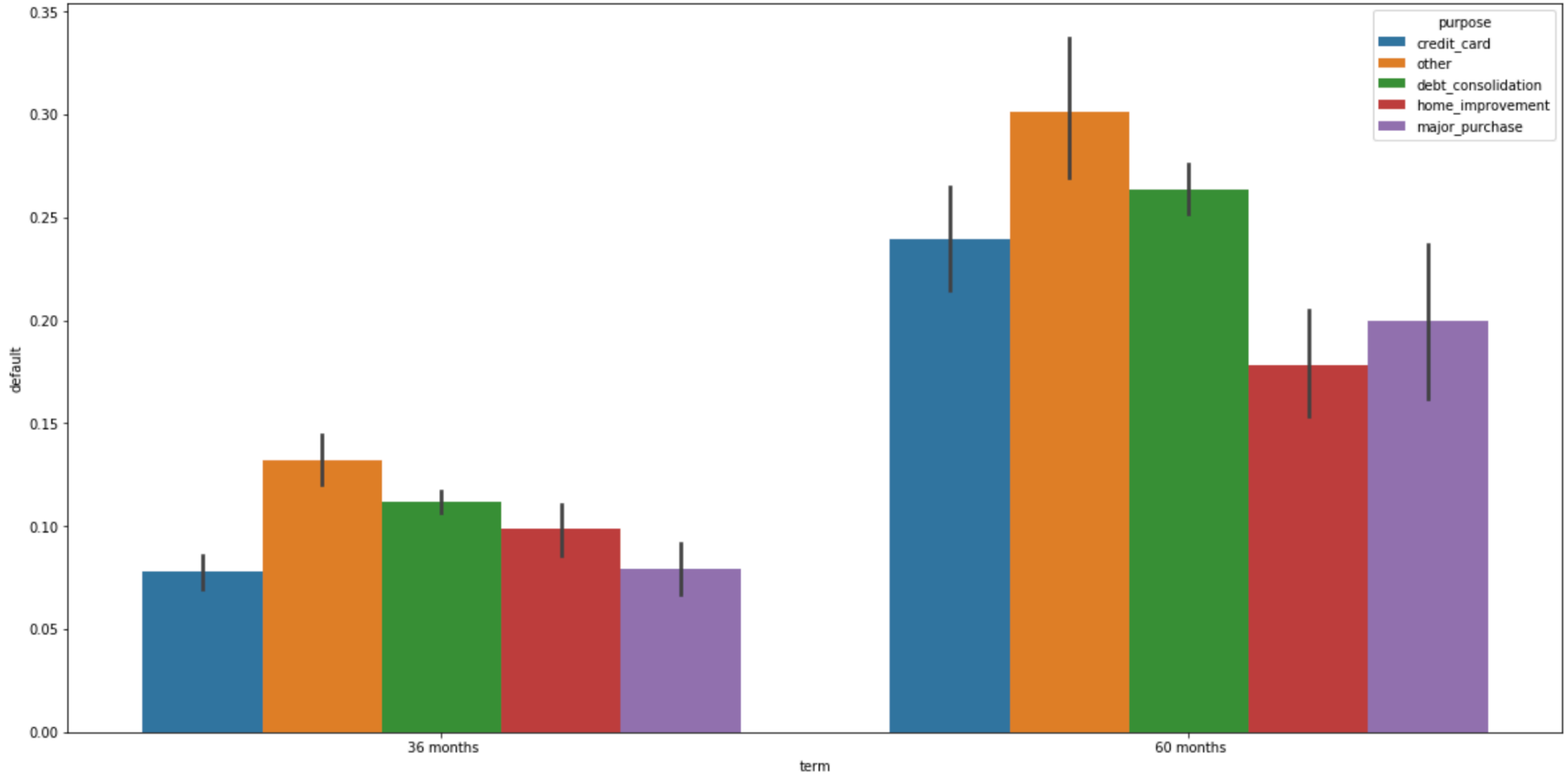
Segmented univariate analysis

- Top 5 purpose of loans are
  1) **debt_consolidation**
  2) **credit_card**
  3) **Other**
  4) **home_improvement**
  5) **major_purchase**

- Lets do some analysis keeping the purpose as a factor for loan applications

Analysis and Graphs

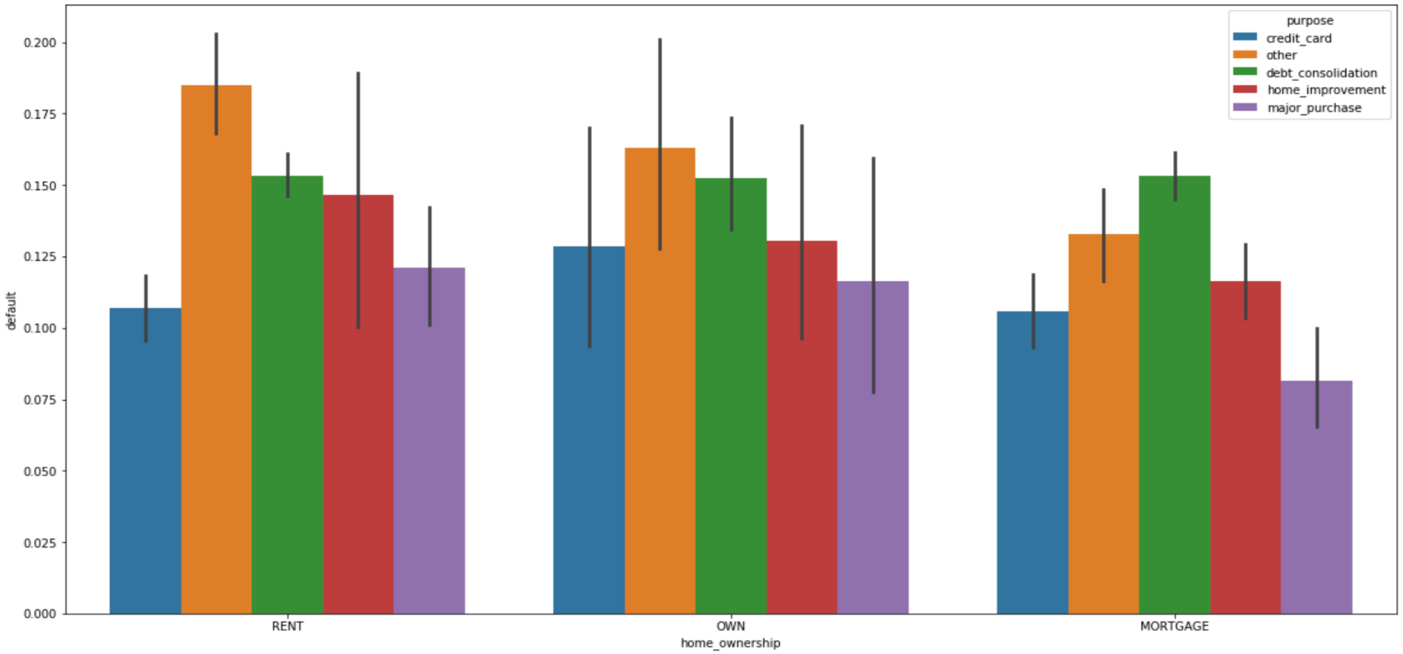
- Lets see graph of term vs default rate against the purpose.



- The above graph shows that 60 months terms have greater default rates and in that too debt_consolidation has the highest default rates across different purpose

Analysis and Graphs

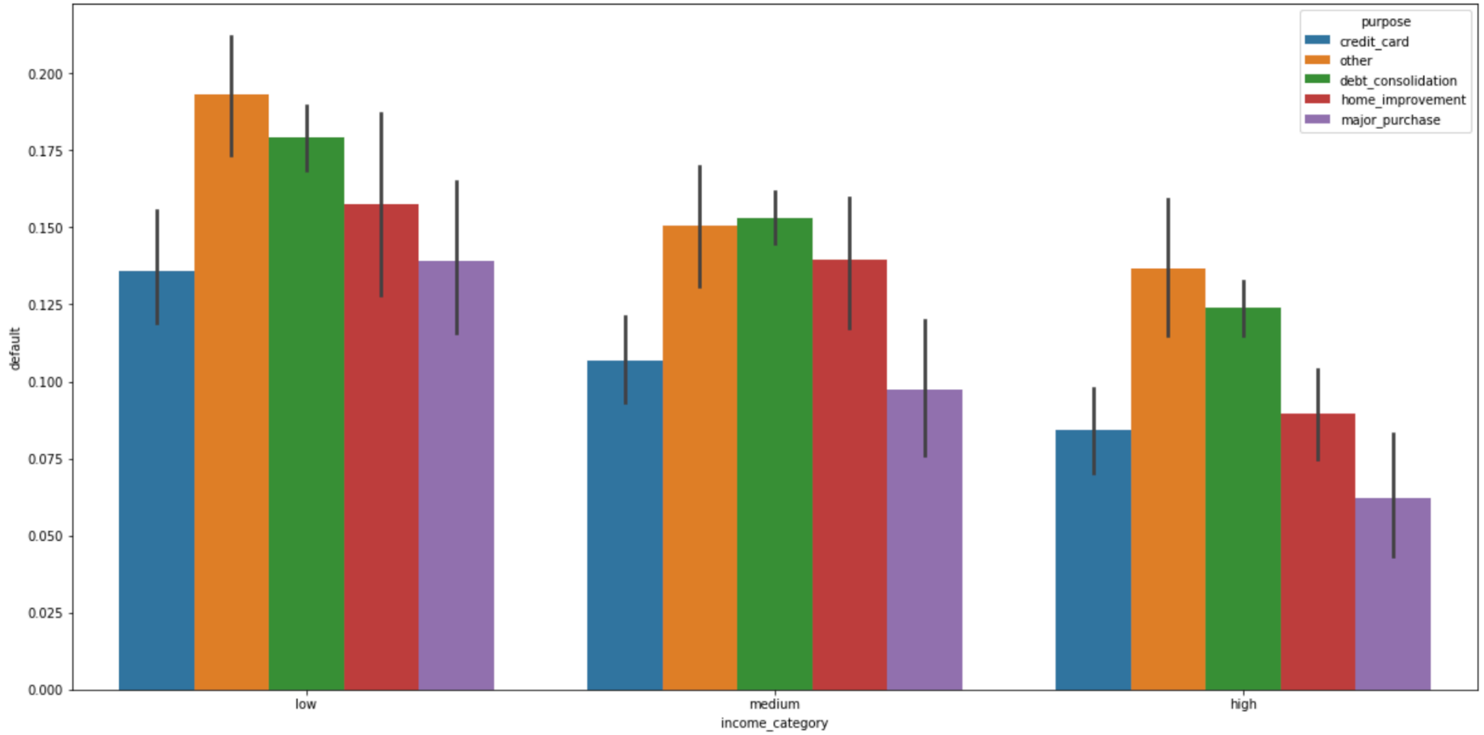
- Lets see graph of home_ownership vs default rate against the purpose.



- It is clear that rented owners have a higher default rate across different purpose

Analysis and Graphs

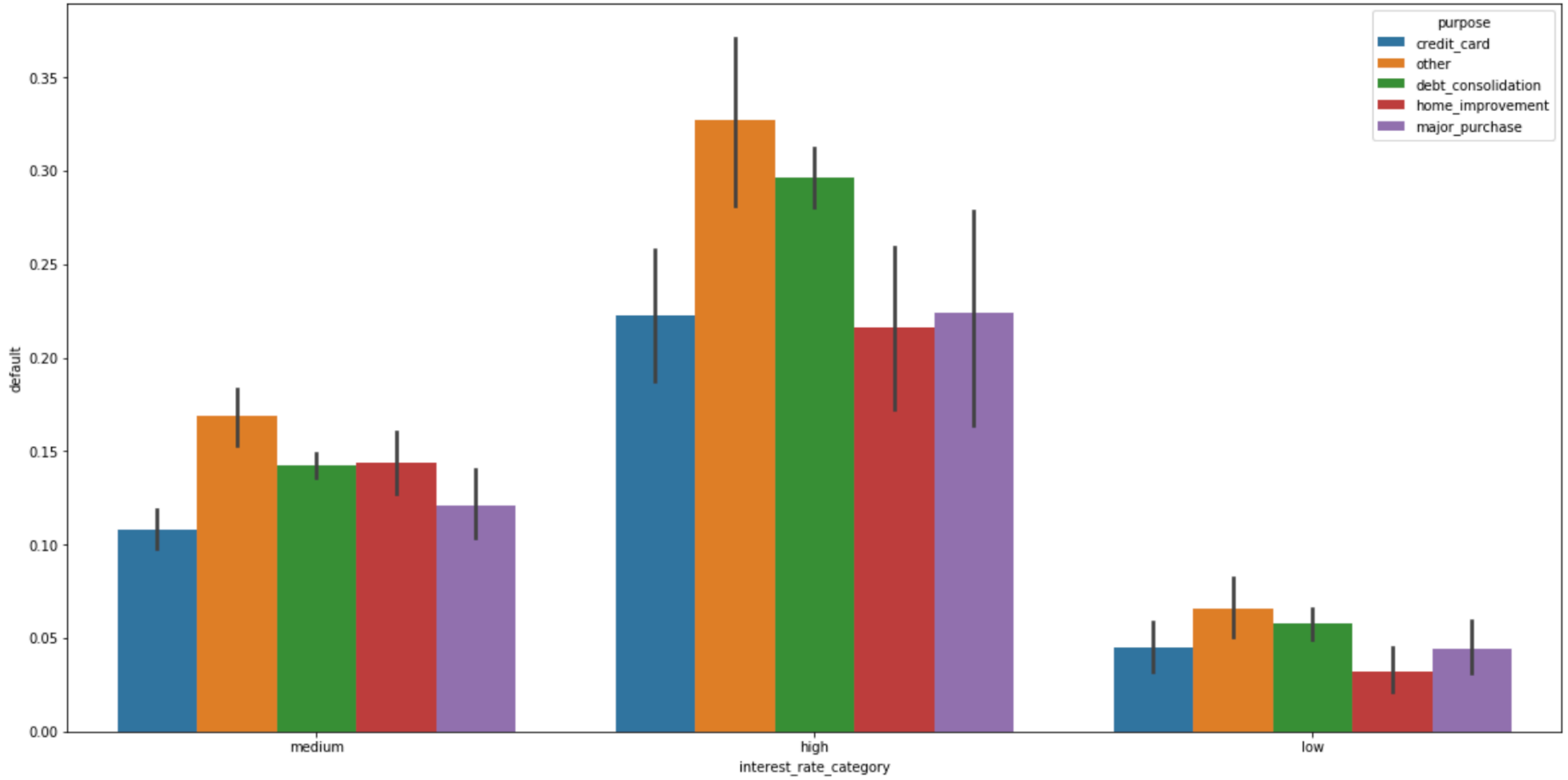
- Lets see graph of income_category vs default rate against the purpose.



- It is clear that lower income ranges have a higher default rate across different purpose

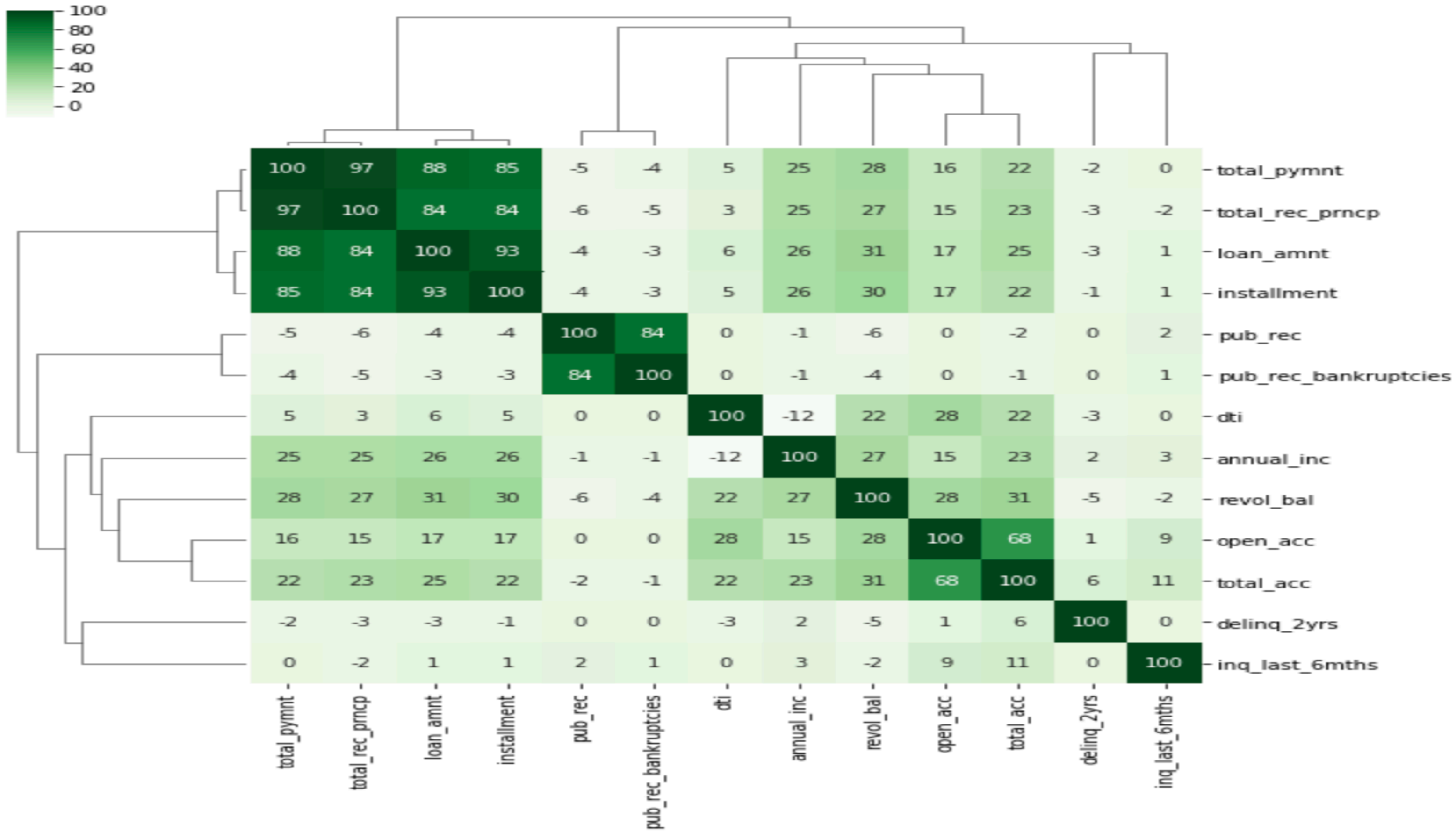
Analysis and Graphs

- Lets see graph of interest category vs default rate against the purpose.



- It is clear that higher interest rate have a higher default rate across different purpose.

Bivariate Analysis



Correlation - Positive

"pub_rec" and "pub_rec_bankruptcies" has positive correlation of +86. Which means a applicant/customer who has derogatory public records has a high chance of having huge number of public record bankruptcies.

Correlation – Negative

"pub_rec" and "pub_rec_bankruptcies" has negative correlation with "total_pymnt" as -4 and -3 respectively. Which means a applicant/customer who has derogatory public records or public record bankruptcies has a very less chance of repaying the loan amount. Leading them to be a defaulter. So avoid giving loan to these applicants.

Conclusion

- From the case study we have analysed some of the variables that are important for Gramener to consider before granting loan.
    - Annual Salary
    - Interest rate
    - Grade and sub grade
    - Loan amount type
    - Verification status
    - Purpose

- Also we analysed how the default rates vary across different purposes for which loan is applied and due diligence needs to be given to that as well