Lending Club Case Study

Date: 21st May 2024

Gaurav Kumar
Dinkar Navaneetha

Agenda

- Introduction
- Approach
- Data Cleaning and Preparation
- Univariant Analysis
- Bivariant Analysis
- Segmented Analysis
- Multivariant Analysis
- Correlation Analysis
- Business Insights and Conclusion

Introduction

- **Problem Statement**: You work for a consumer finance company specializing in various types of loans for urban customers. The company must decide on loan approvals based on the applicant's profile, balancing the risk of losing business by rejecting good applicants and the risk of financial loss by approving those likely to default. The provided data includes information on past loan applicants and their default status. The goal is to identify patterns indicating a likelihood of default to inform decisions such as loan denial, loan amount reduction, or higher interest rates for risky applicants. This case study involves using Exploratory Data Analysis (EDA) to understand the impact of consumer and loan attributes on default tendencies.
- **Objective:**
    - Understand and identify driving factors behind loan defaults.
    - Provide insights for risk assessment.
    - Identify patterns indicating a high risk of loan default.

Approach

Data Read	Data Cleaning	Data Conversion/ inconsistency	Derived Columns	Identify Outlier and Removal	Univariant & Bivariant Analysis	Segmented Analysis	Multivariant & Correlation Analysis
• Import required library in jupyter notebook • Read the data using panda pd.read_csv() • Check its info()	• Check for Duplicate rows • Check for null values in each column • Remove null values column and columns which has more than 40% null values. • Remove column having text or description and columns which not use full for analysis	• Check for columns data type • Convert required columns to numeric by removing additional characters. • Convert columns to float and round off their values. • Imputing the driving columns having null values.	Create derived column required for analysis	Identify the outlier in the dataset and key driving factor and remove outliers from them	• Perform univariant by distribution and box plot and Bivariant analysis by bar plot on key columns • Create a Charged off proportion columns and performed analysis with it • Note the observation s from the plots.	• Modify the Loan Status column and perform analysis on “purpose”. • Identify top relevant purpose and perform analysis it with other key columns or driving columns.	• Performing multivariant analysis on few specific column relevant for defaulter analysis. • Identify the key columns and perform correlation analysis on them.

Data Cleaning and Preparation

Data Cleaning :

- After analyzing the dataset which has 111 columns and 39717 rows ,checked for duplicate rows and no duplicate rows found in the provided data set
- Checked for null values columns and there are plenty of null columns , removed the null value columns and then remaining column count 57
- Remove the columns which has text or descriptive details and not required for analysis.
- Removed the columns which has single value son whole data set.
- Remove rows where the loan status is current as its doesn't add any value in analysis with respect to objective.

Data Conversion / remove inconsistency :

- Check the column datatype and its unique values.
- Remove unnecessary character from the columns which can be used as numeric, like emp_length , int_rate, term etc.
- Imputing null values with “Not Known” in pub_rec_bankruptcies column.

Derive Columns :

- Derive month and Year column from issue date and issue year

Outliers' identification and its removal :

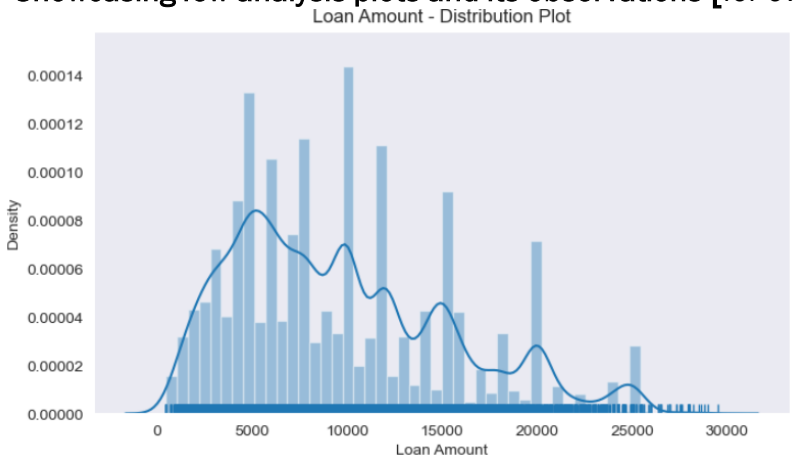
- Check identifier for key columns using box plot and remove the identifier using IQR formula with 1.5 as threshold value.

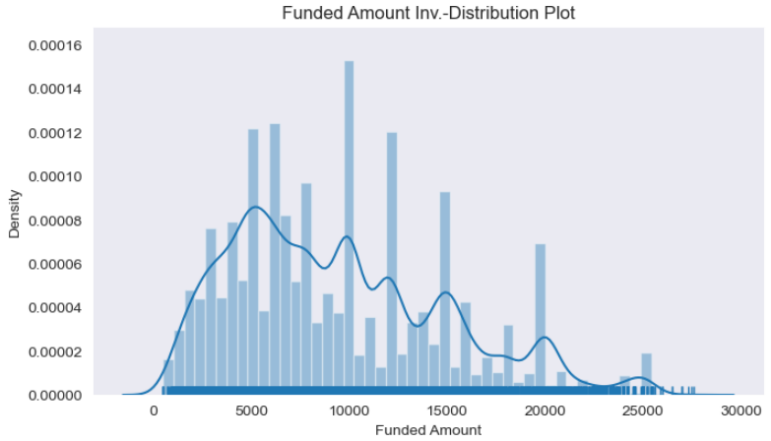
Bucketing variables for plots :

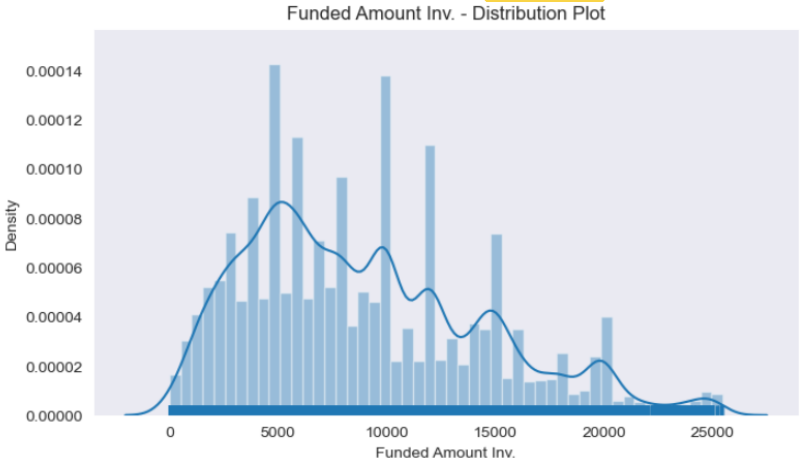
- Create bucketing on interest rate , annual income , dti and loan amount.

Univariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]

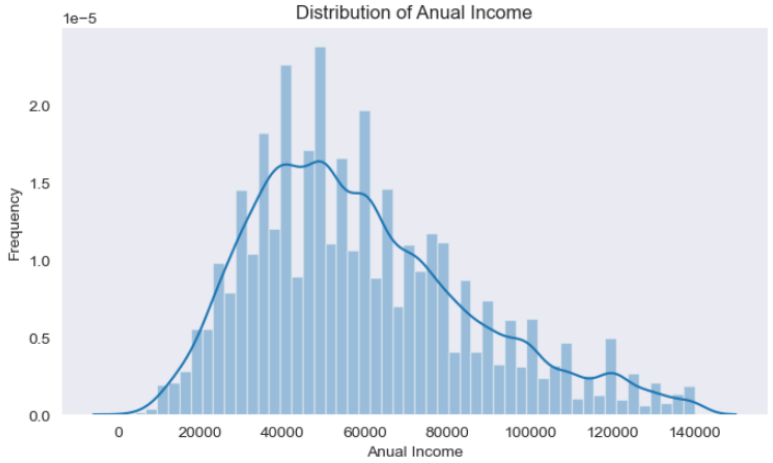


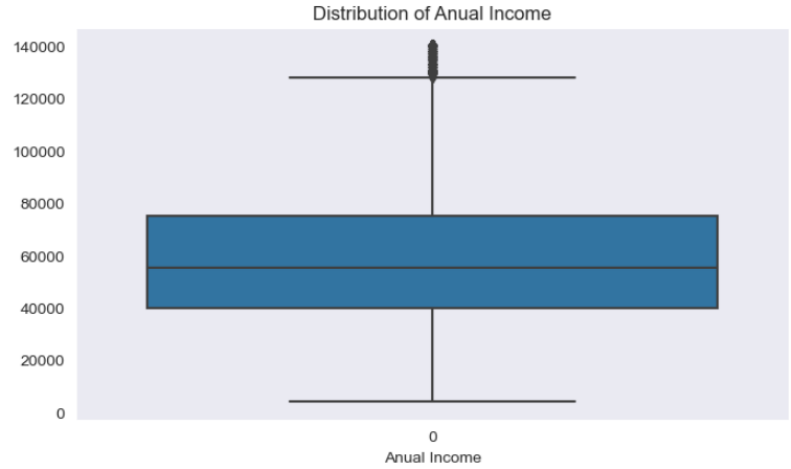




Analsyis of loan_amnt , funded_amnt and funded_amnt_inv

- Distribution of amounts for all three looks very much similar
- Most of the loan amount applied was in the range of 5k-14k.
- Max Loan amount applied was ~27k or ~28k.



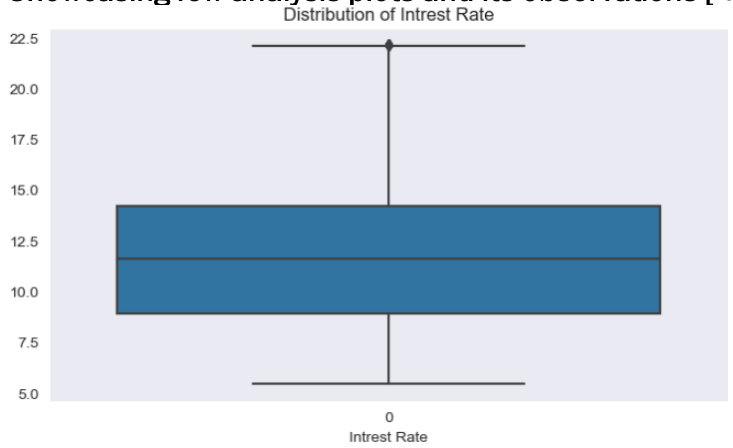


Analysis of Annual Income

1.Annual income of most of applicants lies between 40k-75k

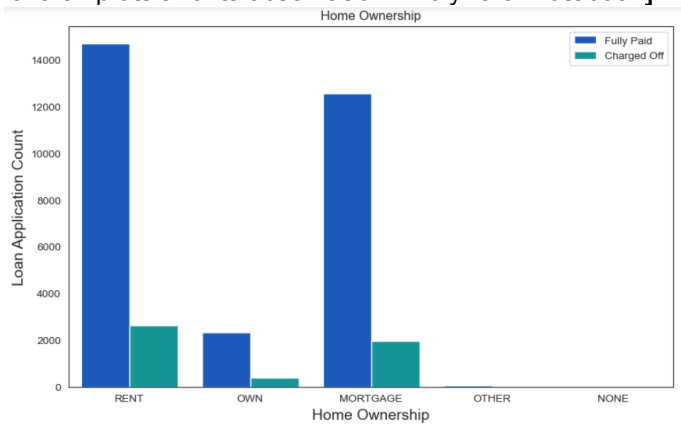
Univariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



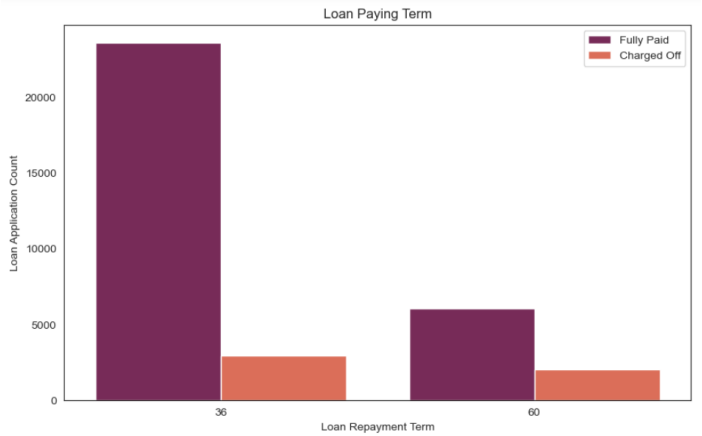
Analysis of Interest Rate

- Most of the applicant's rate of interest is between in the range of 8%-14%.
- Average Rate of interest of rate is 11.7 %.



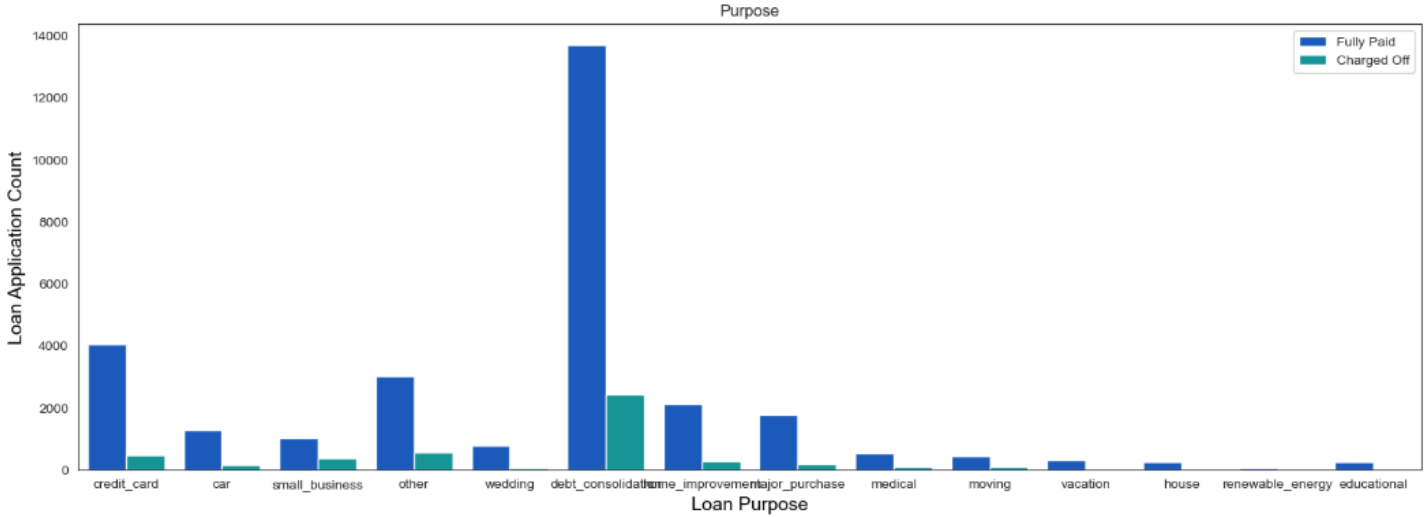
Analysis of Home Ownership

- Majority of loan applicants are either living on Rent or on Mortgage , and majorly these people defaulted the loan



Analysis of Loan Re-Paying Term

- Those who had taken loan to repay in 60 months had more % of number of applicants getting charged off as compared to applicants who had taken loan for 36 months

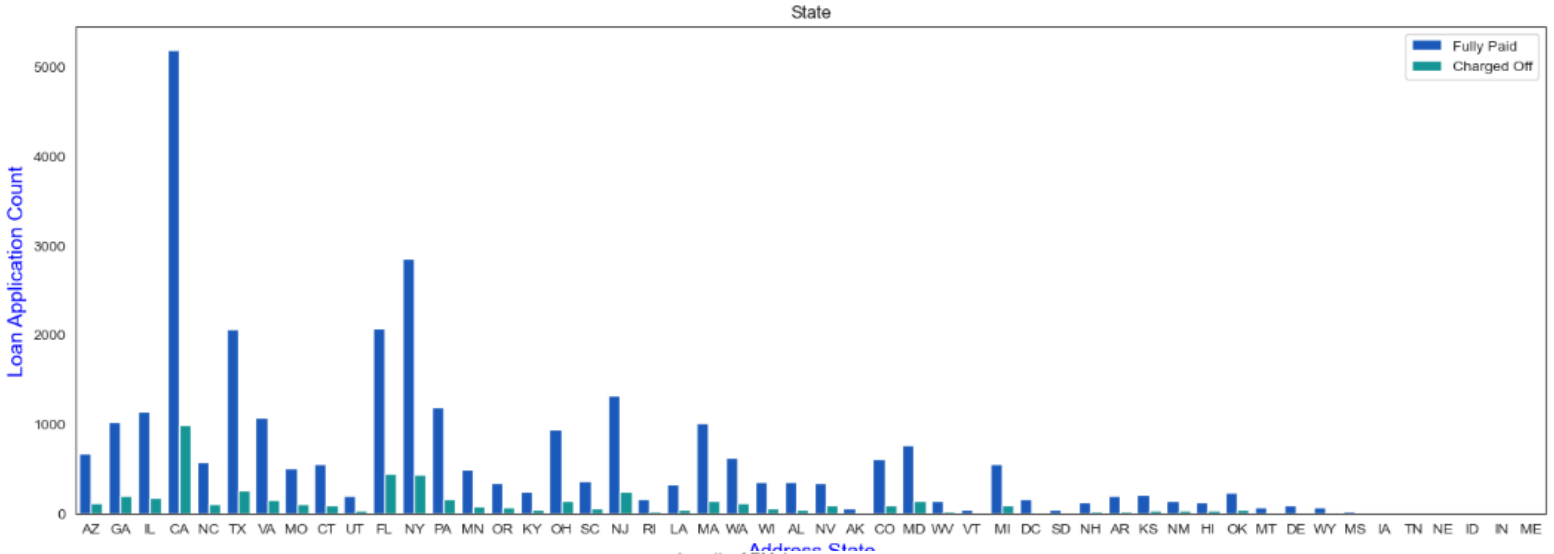


Analysis of Loan Purpose :

1.Most of the applicant are for debt_consolidations and compared to other these people defaulted the loan

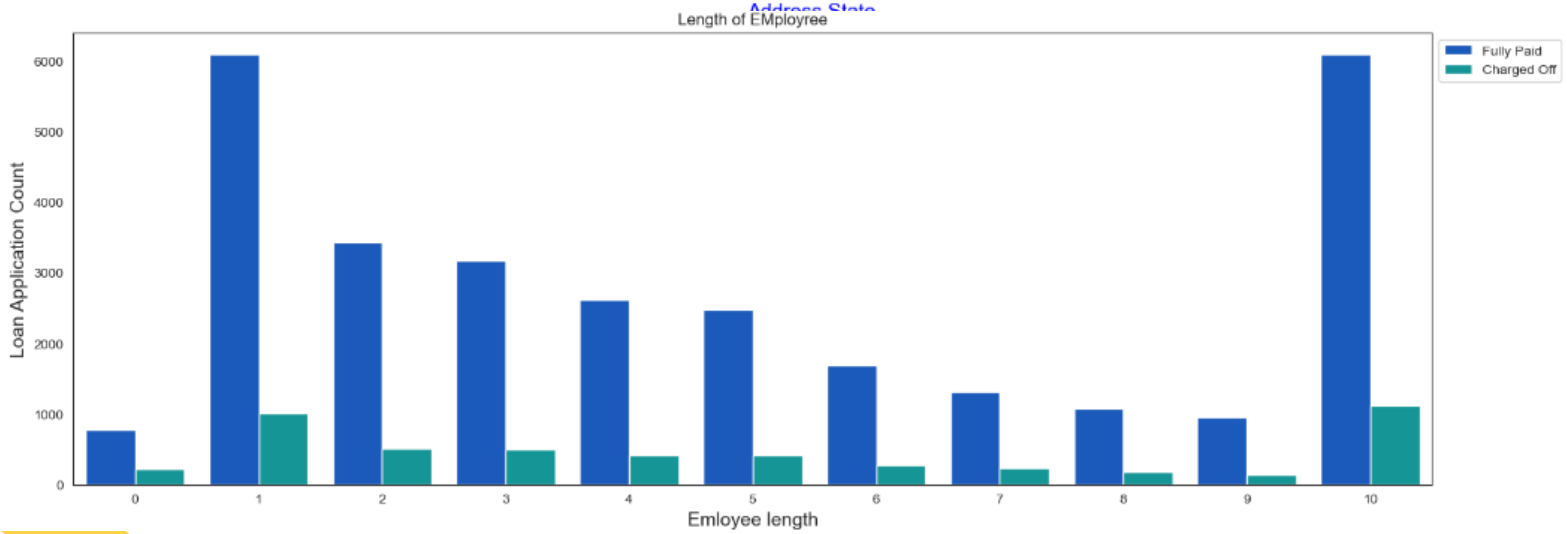
Univariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



Analysis of Address State

- Most of the applicant are from CA and maximum people who defaulted are also from CA

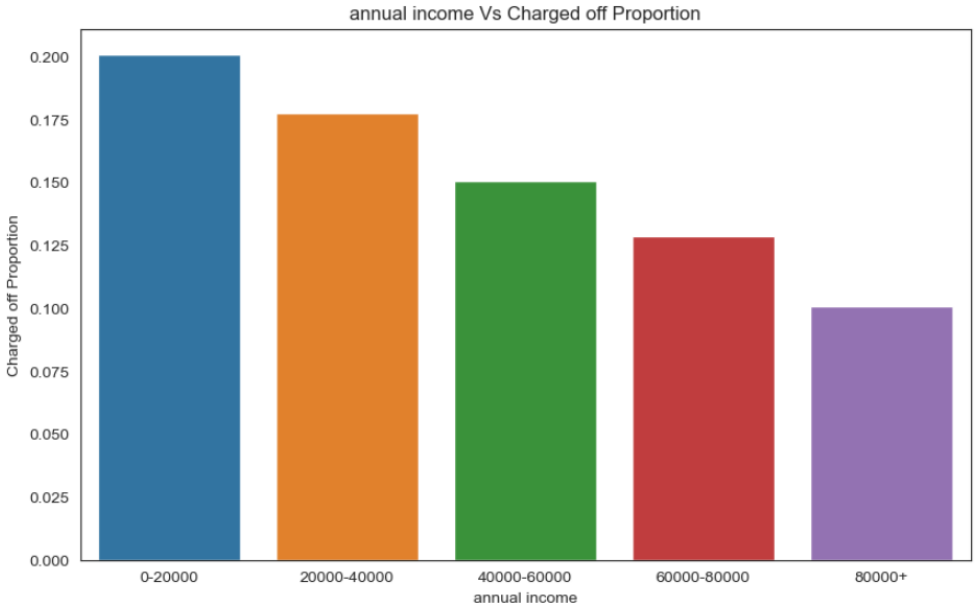


Analysis of Employee Length:

1.Most of the loan applicants are having 10+ yrs of Exp

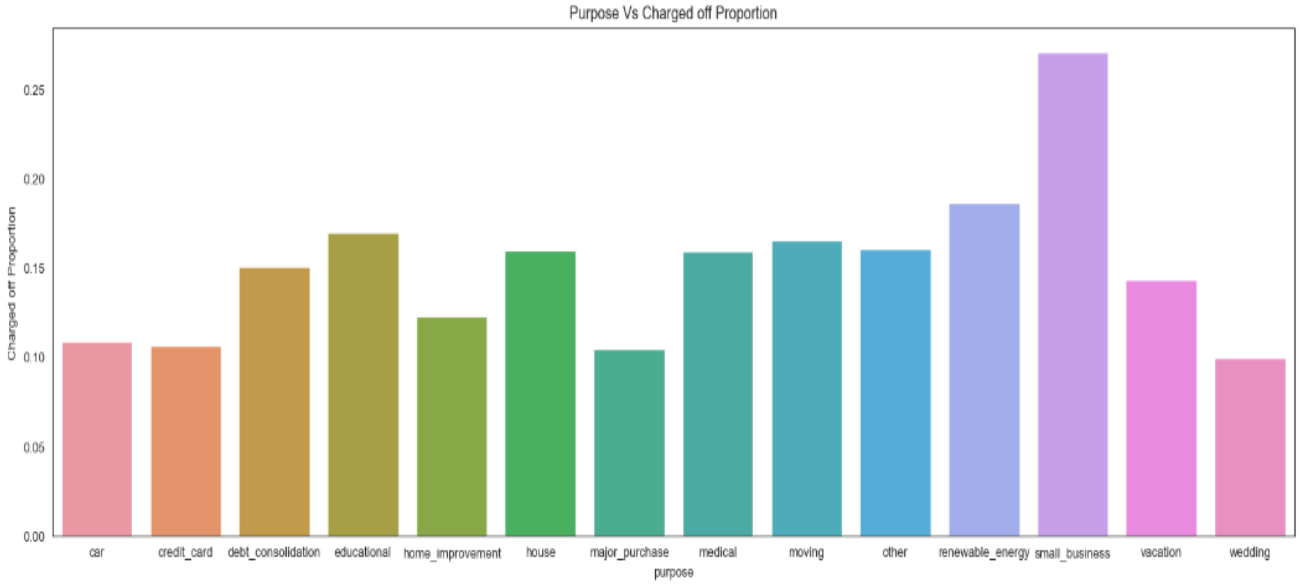
Bivariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs annual income**

- Income range 80000+ has less chances of charged off.
- Income range 0-20000 has high chances of charged off.
- With increase in annual income charged off proportion got decreased. So, annnual income is inversely proportional to charged off loan

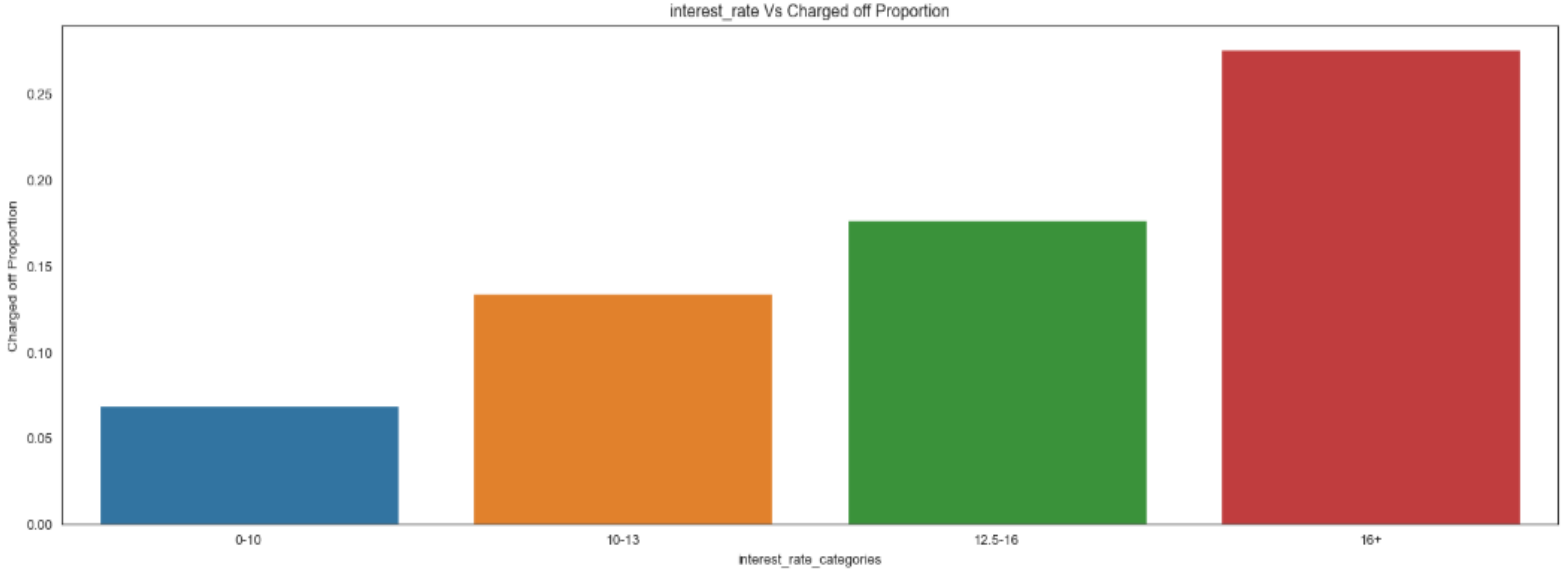


Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs annual Purpose**

- loan taken for Small business have higher charged off proportion, more tendency to get charged off

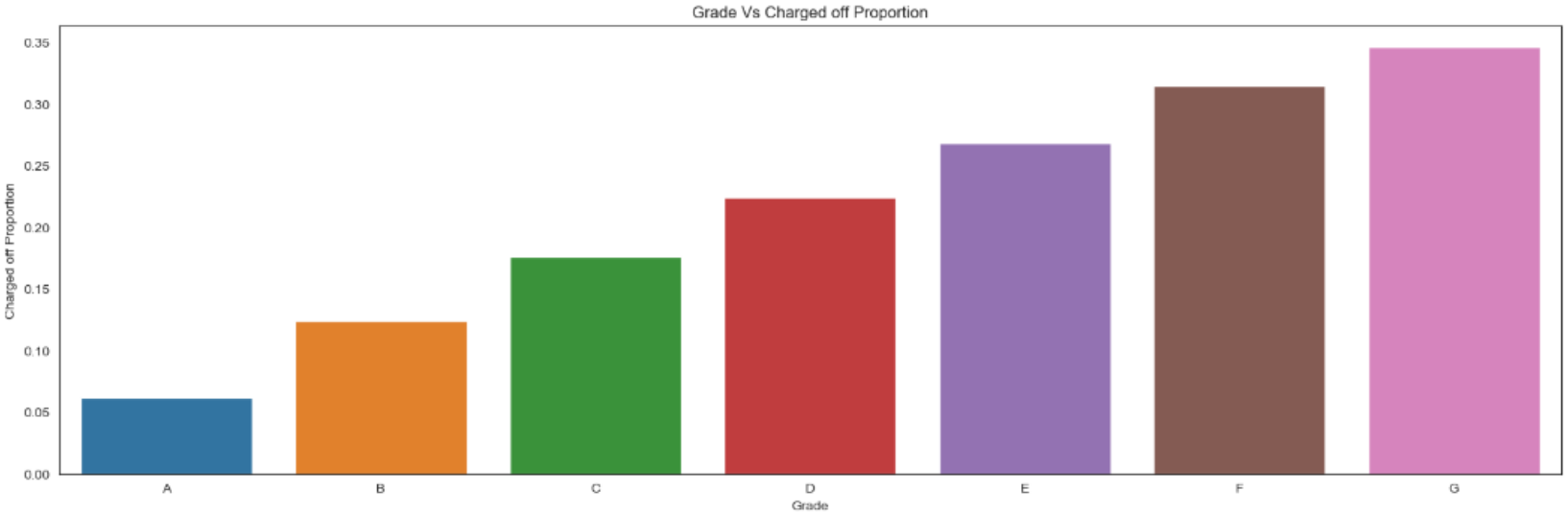
Bivariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs Interest Rate**

- Loan with higher interest , 16%+ rate have higher rate of getting charged off
- Increase in interest rate directly proportional to increase in getting charged off

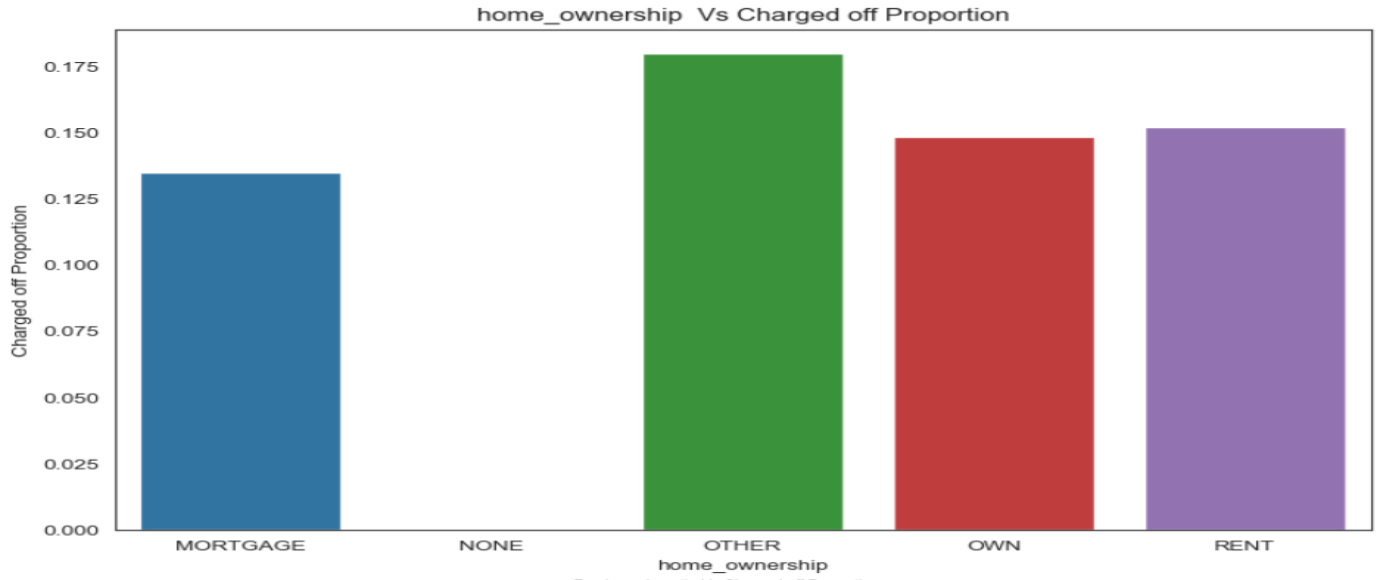


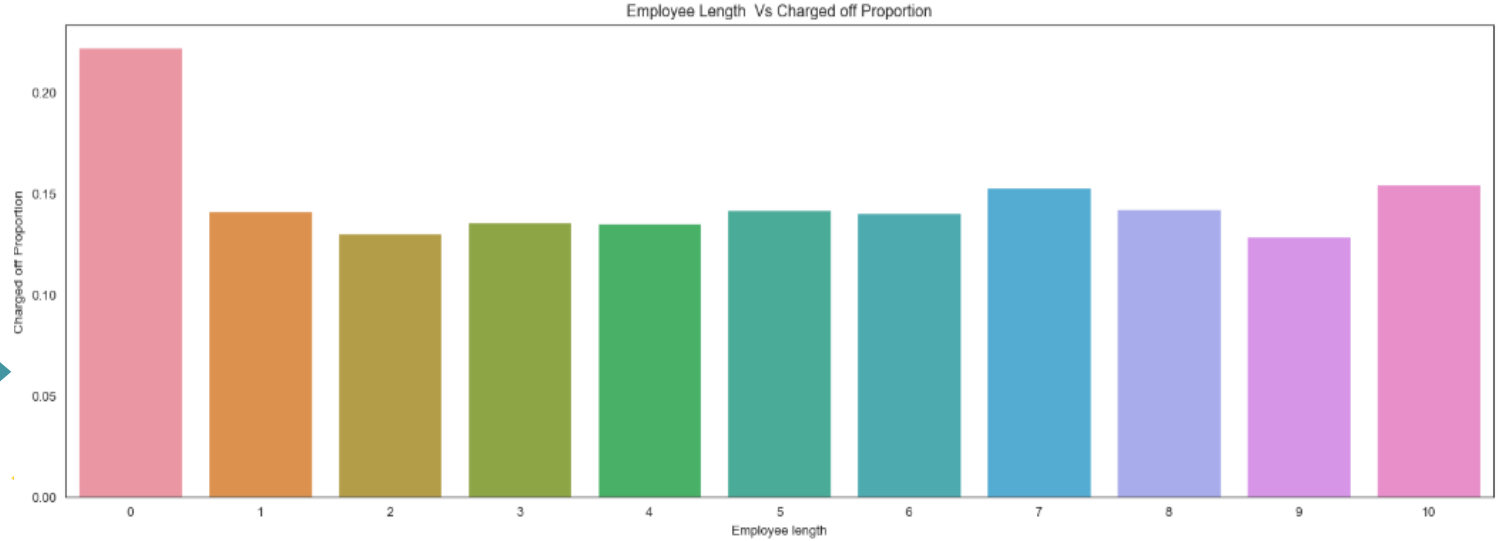
Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs Grade**

1.Grade A has least chances of getting charged off and Grade F &G has high chance of getting charfe doff
2.tendency of getting charged off is increasing with grades moving from A to G

Bivariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]





Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs Home Ownership**

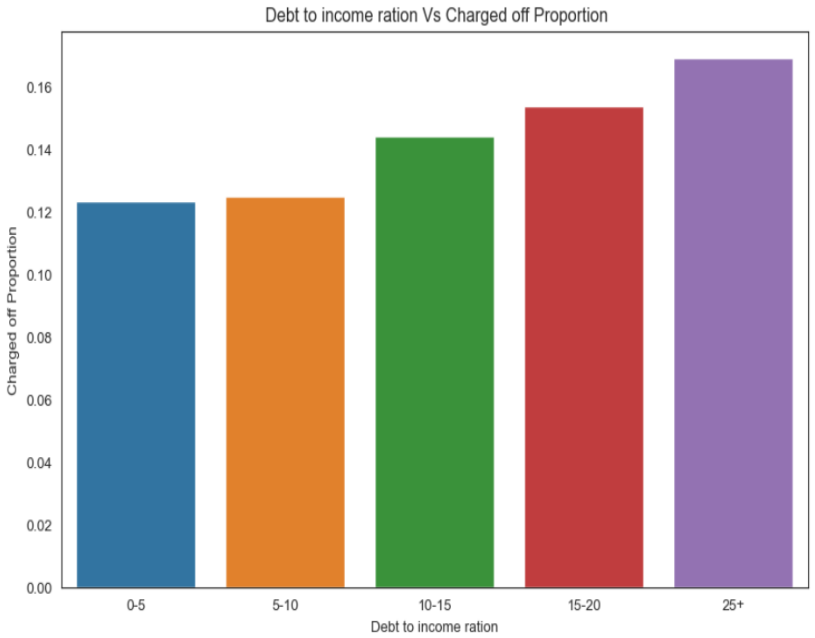
- Not significant conclusion through home owner and loan defaulter. but certainly those who are not owning the home is having slightly high chances of loan defaults

Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs Employee Length**

- Employee with less than 1 year of experience or not working has high tendency to get loan defaulted , as they less experience or just started the job so don't have much amount or not having fixed source of income

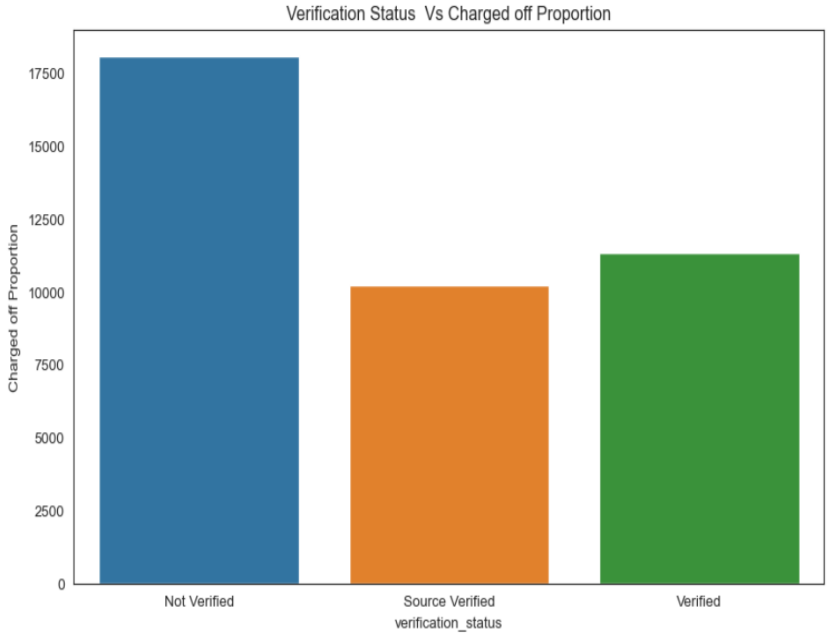
Bivariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



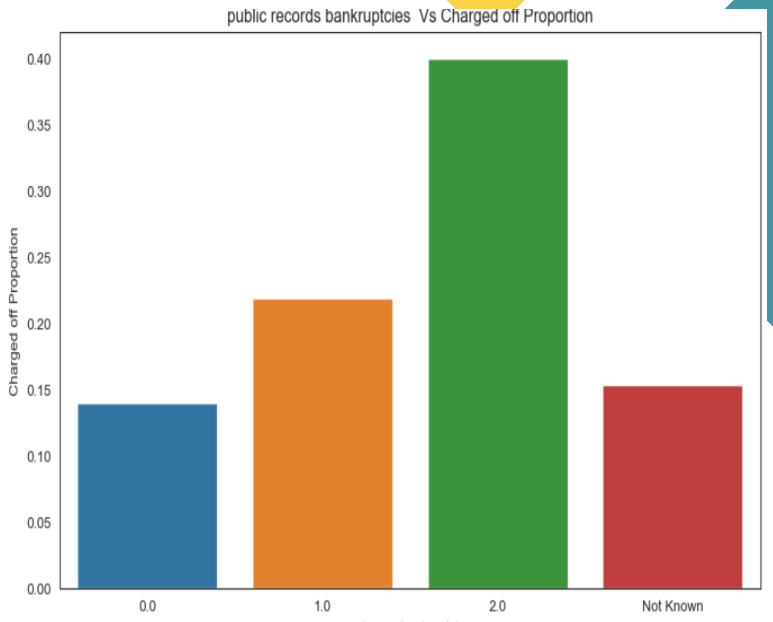
Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs DTI**

- High DTI value having high risk of defaults



Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs Verification Status**

- Loan which are not verified having high chances of getting defaulted

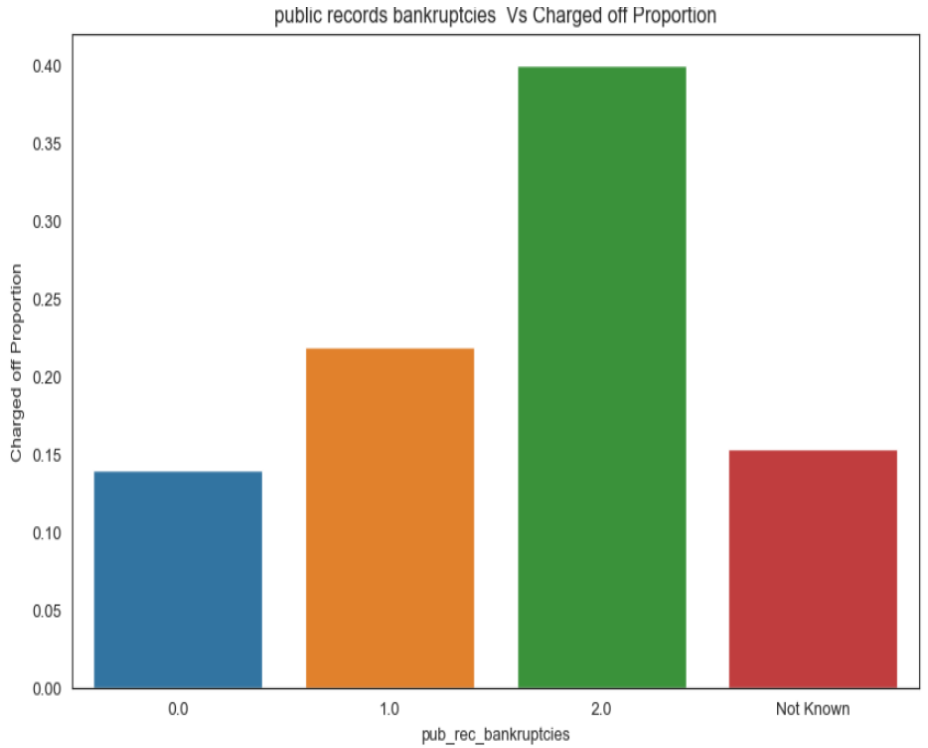


Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs pub_rec_bankruptcies**

- Those who already have pub_rec_bankruptcies value 1 or 2 , have higher charged off proportion than who do not have pub_rec_bankruptcies.
- Not known is the column for which we do not have any information about borrower.
- This also makes sense that who has defaulted before has more chances of defaulting in future as well.

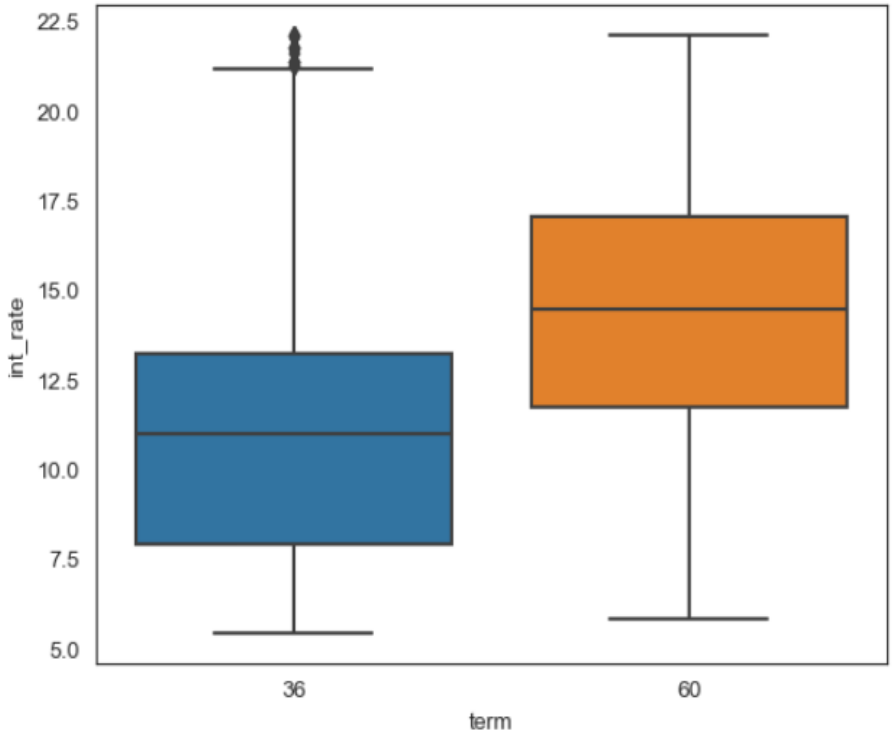
Bivariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



Analysis of **Chargedoff_Proportion [(charged off loan / (Charged off loans + Fully Paid loans))] vs** pub_rec

- Those who already have pub_rec value 1 or 2 , have higher charged off proportion compared to who don't have any public records
- As the number of public records increase , tendency to get charged off increases

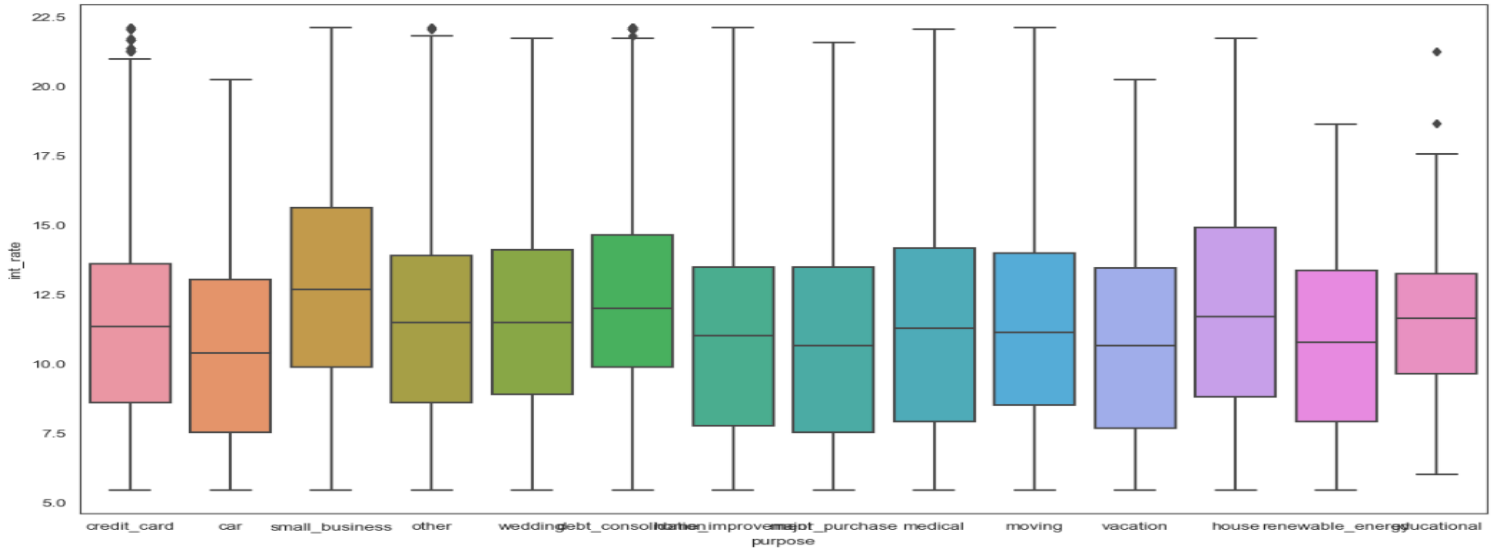


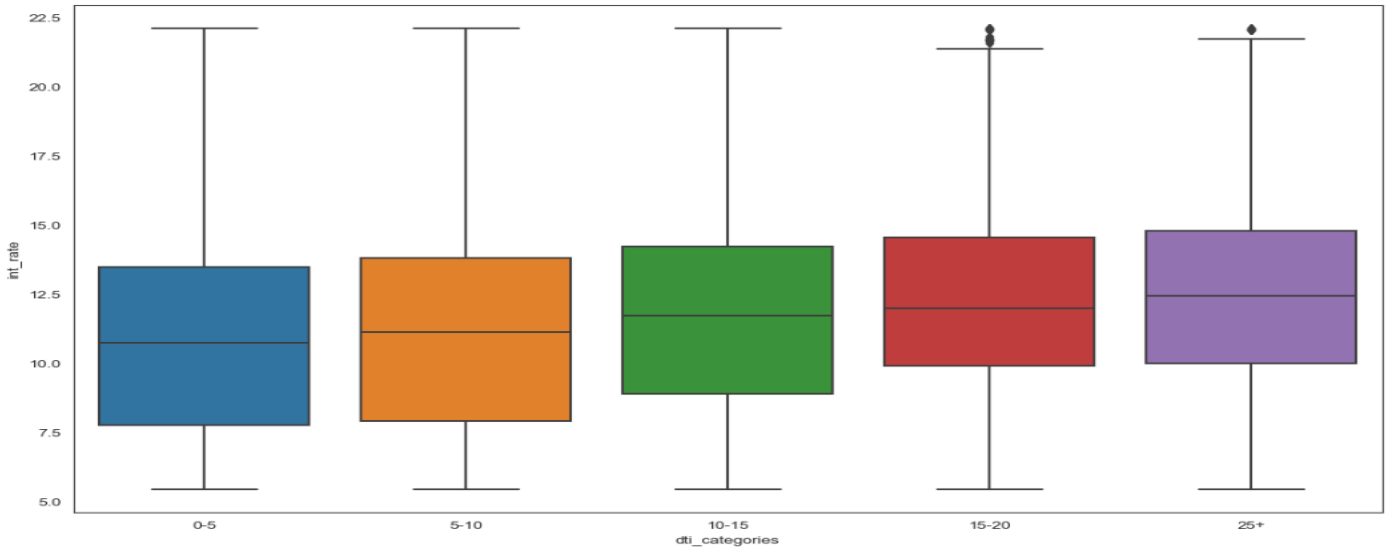
Analysis of **Term vs Interest Rate**

- Avearge intrest rate is higher for 60 months loan term.
- Most of the loans issued for longer term had higher intrest rates for repayement.

Bivariant Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]





Analysis of **Purpose vs Interest Rate**

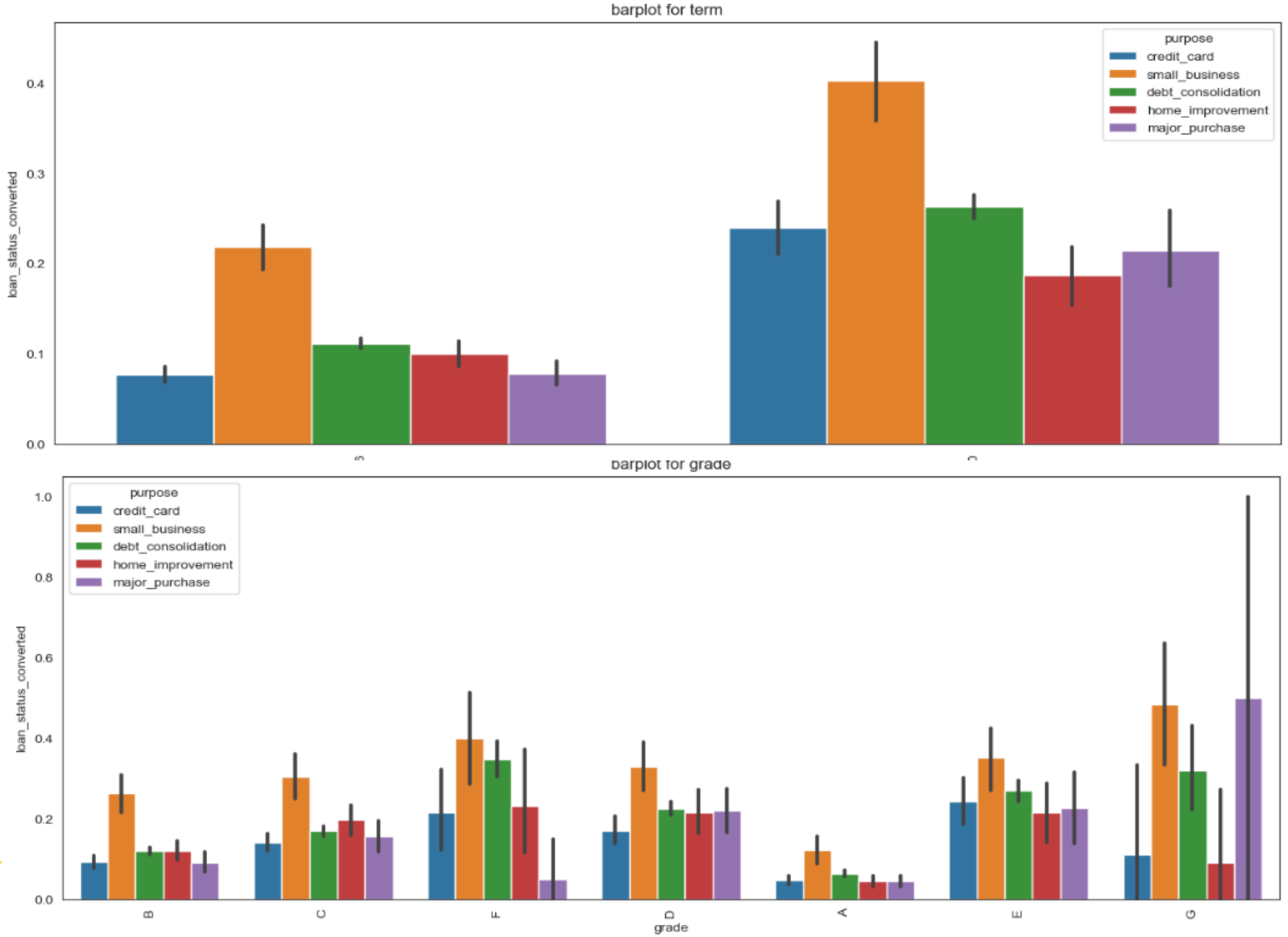
- average interest rate is highest for small business purpose.
- Loans taken for small business purposes had to repay loan with more interest rate as compared to other.

Analysis of **DTI categories vs Interest Rate**

- DTI is low enough you may get a lower interest rate.
- Plot shows no significant variation but there is slight increase in interest rate with increase in DTI.So if DTI is low enough you may get a lower interest rate

Segmented Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



Analysis of Loan Status vs Term

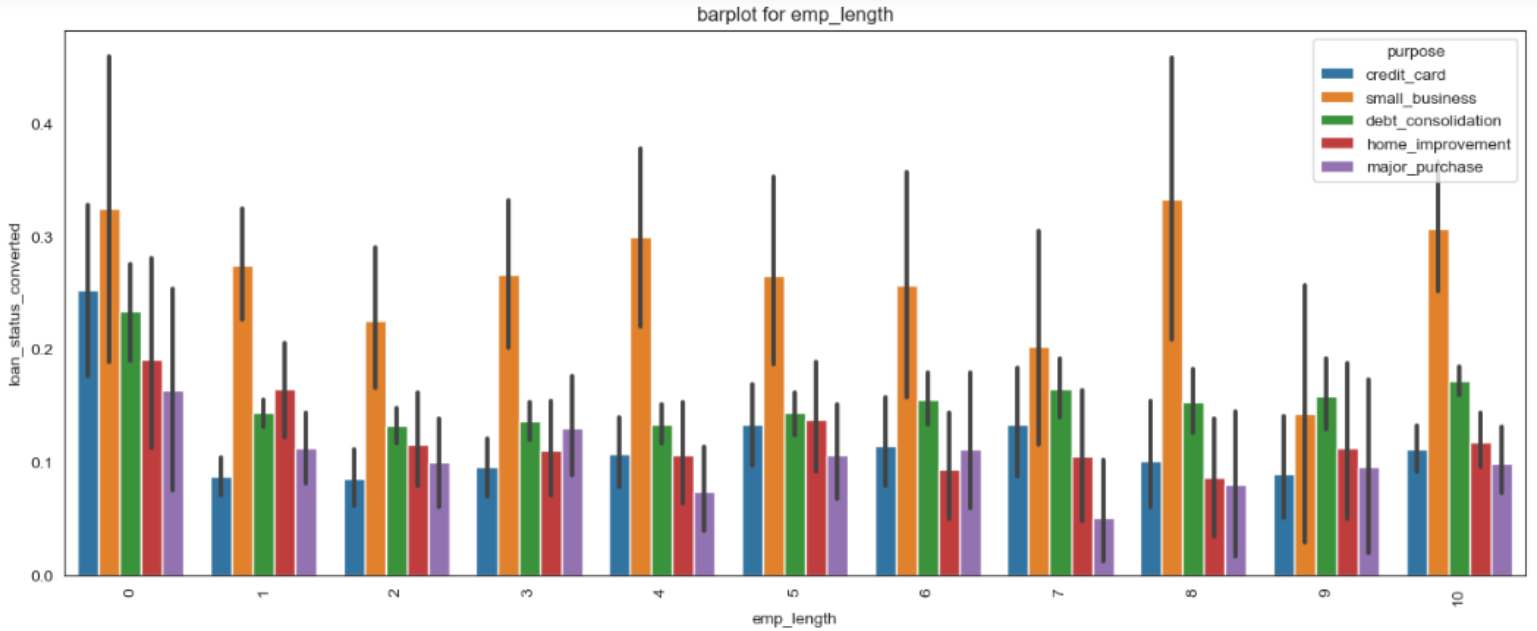
- Small business has the most loans defaults, followed by debt_consolidation for both terms 36 months and 60 months.

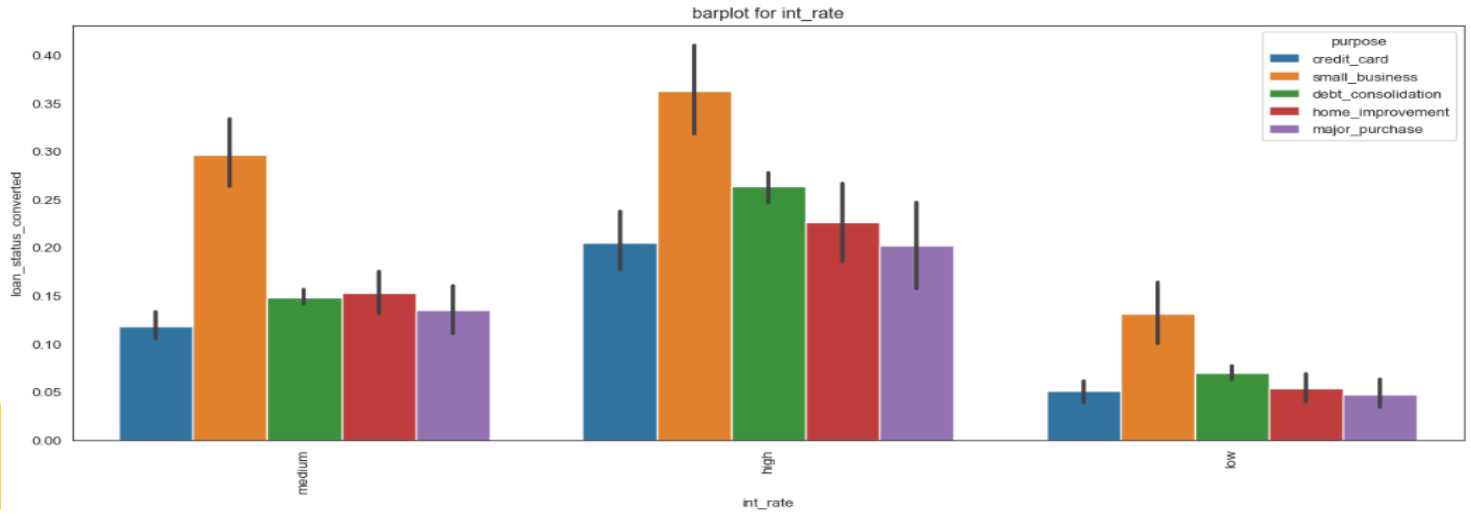
Analysis of Loan Status vs Grade

- Small Business has the most loan defaulters in maximum of the grade

Segmented Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]





Analysis of Loan Status vs Emp Length

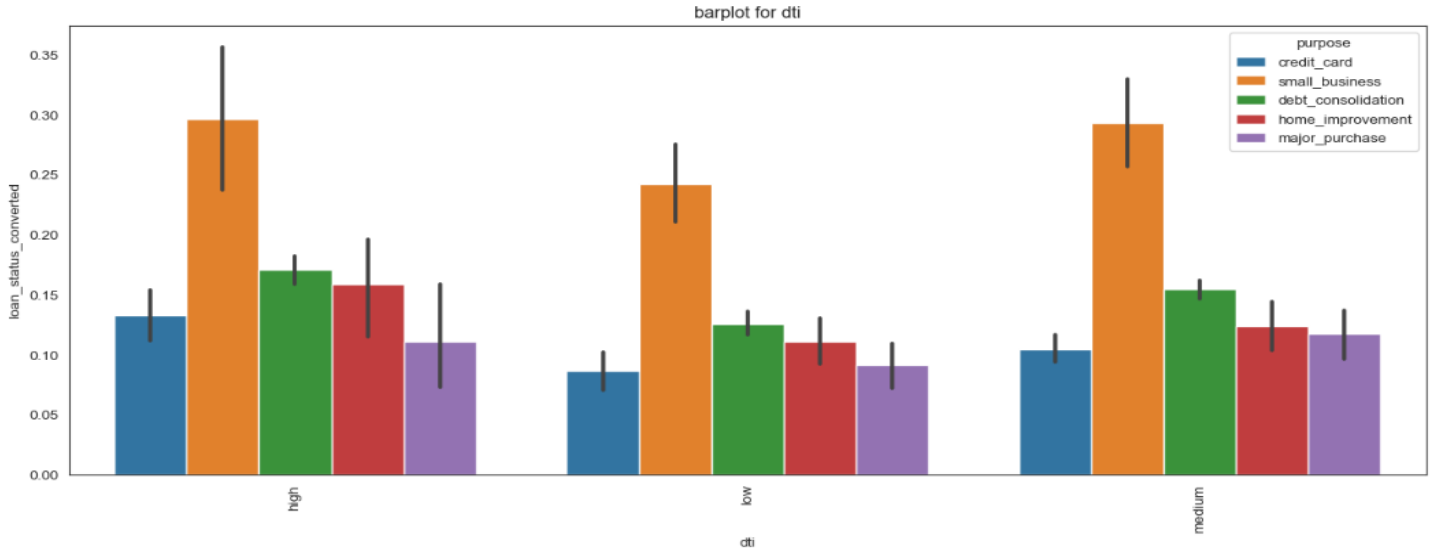
- For all Employee length small business loan is getting defaulted.

Analysis of Loan Status vs Interest Rate

- Small Business has the most loan defaulters in all intrest rate range

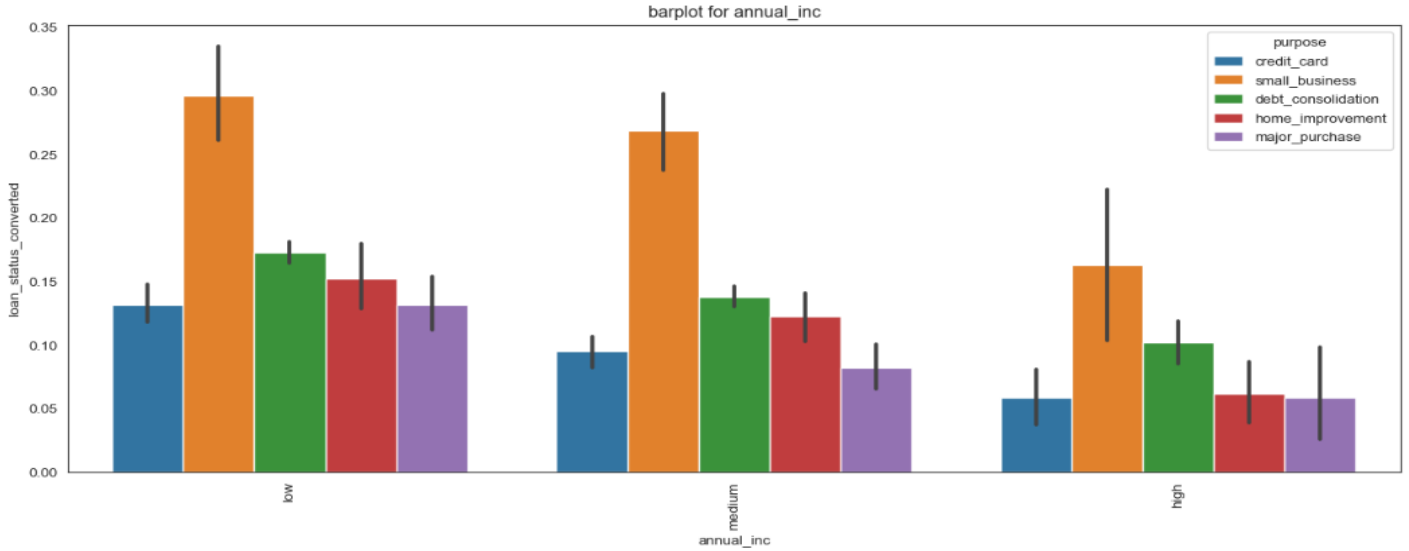
Segmented Analysis

Showcasing few analysis plots and its observations [for overall plots and its observation kindly refer notebook]



Analysis of Loan Status vs DTI

- Small Business has the most loan defaulters in all dti range

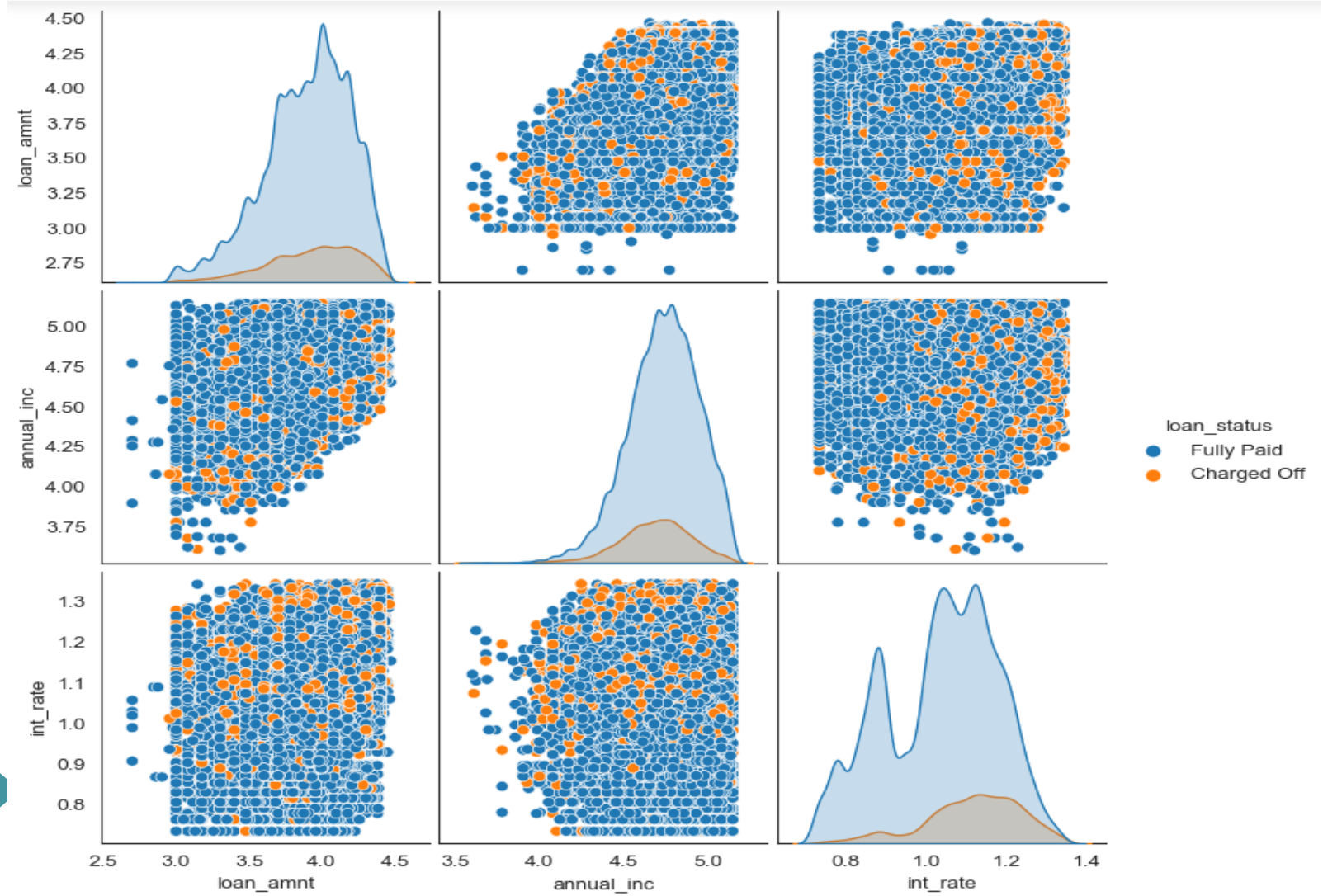


Analysis of Loan Status vs Annual Income

- Small Business has the most loan defaulters for each category of income range

Multivariant Analysis

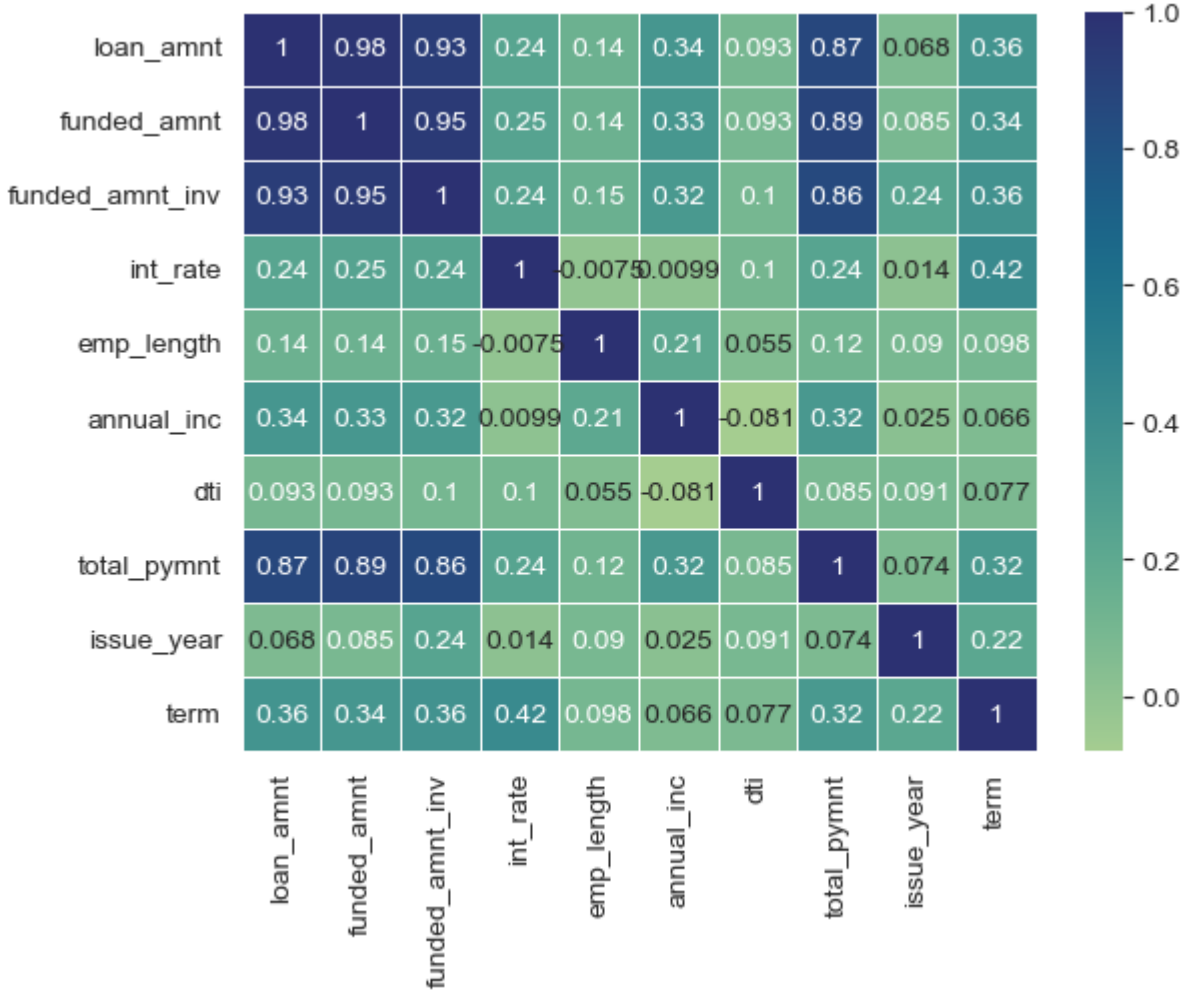
Showcasing few analysis plots and its observations



- Higher is the charged off ratio for higher interest rate.
- Slightly Higher the loan amount for a higher the annual income .
- With increase in loan amount there is an increase in interest rate

Correlation Analysis

Showcasing few analysis plots and its observations



- Annual Income to dti are negatively correlated
- Loan Amount,Investor Amount and Funding Amount are strongly correlated
- Term has a strong correlation with loan amount
- Term has a strong correlation with interest rate
- Positive correlation between Annual Income and employment years (emp_length)
- annual income has a strong correlation with loan_amount
- Positive correlation between annual income and funded amount that means people with high income gets high funded amount
- Positive correlation between annual income and total payment

Business Insights and Conclusion

Business Insights

1.Loan Amount and Default Risk: Higher loan amounts are associated with increased risk of default. Most loans are in the range of 5k-14k, but loans closer to the maximum (27k-28k) exhibit higher default rates.

2.Interest Rates: Loans with higher interest rates (16% and above) have a significantly higher chance of being charged off. The average interest rate for defaulted loans is higher than for fully paid loans.

3.Annual Income: Applicants with lower annual incomes (0-20k) have higher default rates, while those with incomes above 80k have a lower propensity for default. Annual income is inversely proportional to the likelihood of loan default.

4.Loan Purpose: Loans taken for small businesses and debt consolidation have higher default rates compared to other purposes.

5.Home Ownership: Majority of the loan defaulters are either renting or have a mortgage. Homeownership slightly reduces the likelihood of default.

6.Employment Length: Applicants with less than 1 year of employment experience have a higher tendency to default. Longer employment length correlates with lower default rates.

7.Debt-to-Income Ratio (DTI): Higher DTI values correlate with a higher risk of loan default. A low DTI ratio is associated with lower interest rates and better repayment behavior.

8.Verification Status: Loans that are not verified have higher default rates. Verification status plays a crucial role in assessing risk.

9.Public Records and Bankruptcies: Applicants with public records or previous bankruptcies have a higher chance of defaulting. Historical financial behavior is a strong predictor of future default risk.

10.Geographical Insights: States like CA, NV, and FL have higher default rates, suggesting regional economic factors may play a role.

Recommendations

1.Stricter Loan Approval for High Amounts: Implement more rigorous screening for higher loan amounts to mitigate default risks.

2.Interest Rate Management: Adjust interest rates based on risk assessment. Higher-risk applicants should be offered loans at higher rates with caution or additional guarantees.

3.Income-Based Screening: Focus on applicants with stable and higher incomes. Introduce minimum income thresholds for loan approval.

4.Purpose-Specific Strategies: Tailor loan terms and interest rates based on the purpose, especially for high-risk purposes like small business loans.

5.Employment Verification: Emphasize employment stability by requiring a minimum employment length for loan approval. Consider offering better terms to those with longer employment histories.

6.Reduce Unverified Loans: Increase efforts to verify applicant information. Unverified loans should either be limited or subject to higher interest rates and stricter terms.

7.Regional Risk Adjustments: Consider regional economic conditions and adjust loan terms and risk assessments accordingly.

Conclusion & Driver Variables

- The exploratory data analysis has highlighted several key factors influencing loan default risks, such as loan amount, interest rates, annual income, loan purpose, home ownership, employment length, and public records.
- By understanding these factors, the company can better assess risk and make informed decisions on loan approvals, terms, and interest rates.
- Implementing these insights and recommendations will help minimize financial losses due to loan defaults and improve the overall profitability and sustainability of the lending business.

Driver Variables Identified

1. **Loan Amount**
2. **Interest Rate**
3. **Annual Income**
4. **Loan Purpose**
5. **Home Ownership**
6. **Employment Length**
7. **Debt-to-Income Ratio (DTI)**
8. **Verification Status**
9. **Public Records and Bankruptcies**
10. **Geographical Location**

These variables are critical in driving the risk of loan default and should be closely monitored and used in predictive modeling for risk assessment.

Thank you