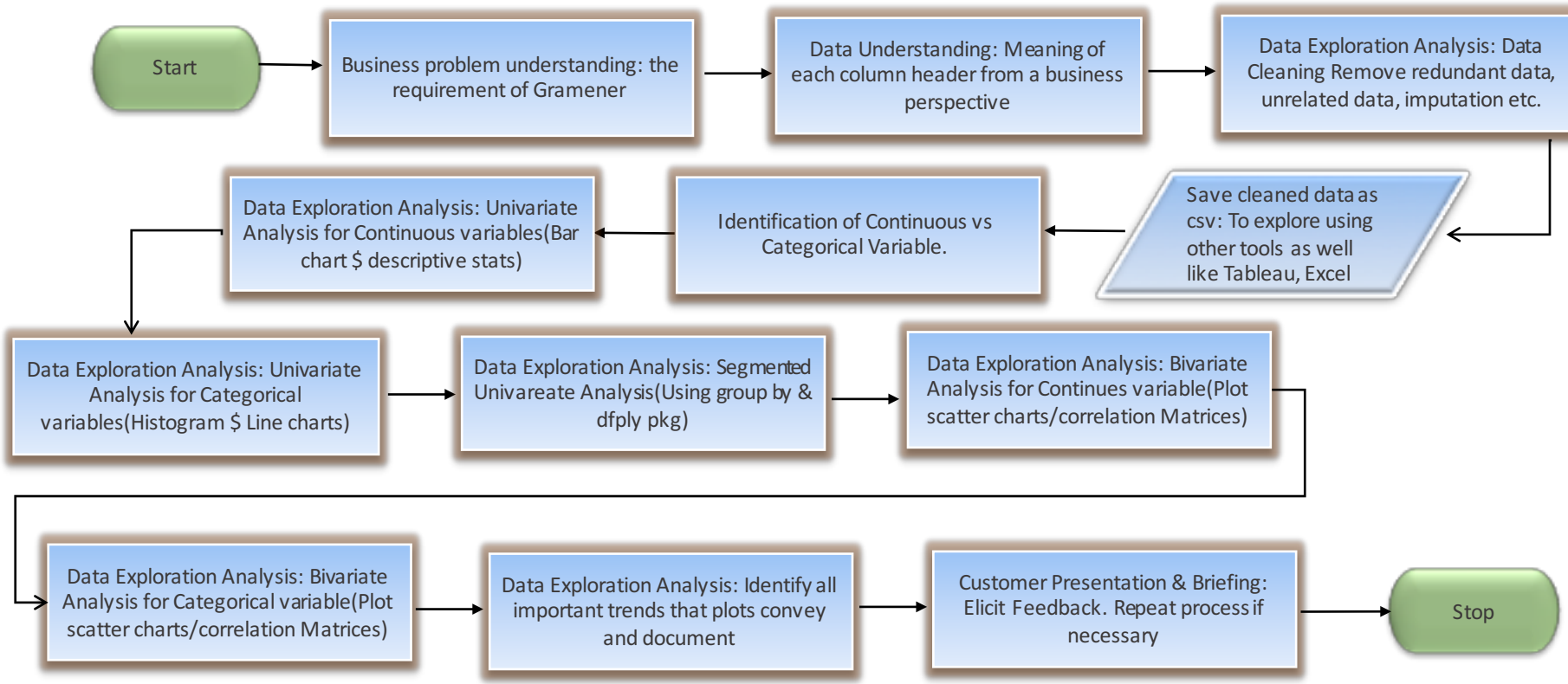
Gramener Case Study

Group Member:

1. Anand kumar
2. Vikas kumar singh
3. Narottam kumar

Problem solving methodology using CRISP-DM



Business Understanding

Business Objective:

- The finance company is looking for the attributes in a applicant profile which can help them is deciding whether to approve or decline the loan application.

Goal of Analysis:

- To find out the relation between the different attributes and their impact on loan default. And suggest which attributes contributes a significant difference in Loan Default.
- Identification of Loan Applicant traits that tend to 'default' paying back.
- Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena.
- Gramener may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants.

Data Understanding

Types of variables

- Customer (applicant) demographic
- Loan related information & characteristics
- Customer behavior (if the loan is granted)

Customer demographic
Employment length
Employment title
Annual income
Zip code
Description

Loan Information
Loan amount
Funded amount
Interest rate
Loan status
Loan grade

Customer Behavior
Employment length
Employment title
Annual income
Zip code
Description

Data Cleaning and Preparation

1. Identify all columns that don't have any other value other than NA
2. Remove rows and columns where NA values are more than or equal to 30%
3. Remove irrelevant columns.
4. Employment Length : Replace null value with 0.
5. Move all columns that don't change. Justification: There is no variance, it cannot help us to determine the reason for default. We can save memory and analysis, plotting and data frame transformations are faster.
6. Remove redundant columns. E.g. The purpose of loan is a drop down which is already a categorical variable. We don't need the title column as it becomes redundant. We can identify all such columns and remove them.
7. Identify all columns that don't have any other value other than NA and 0. Remove such columns.
8. Imputation: Identify all the NA values and replace them with appropriate value. We don't do this in the master frame but instead as and when that particular column is getting analyzed
9. Cast all continues variable to numeric variable.

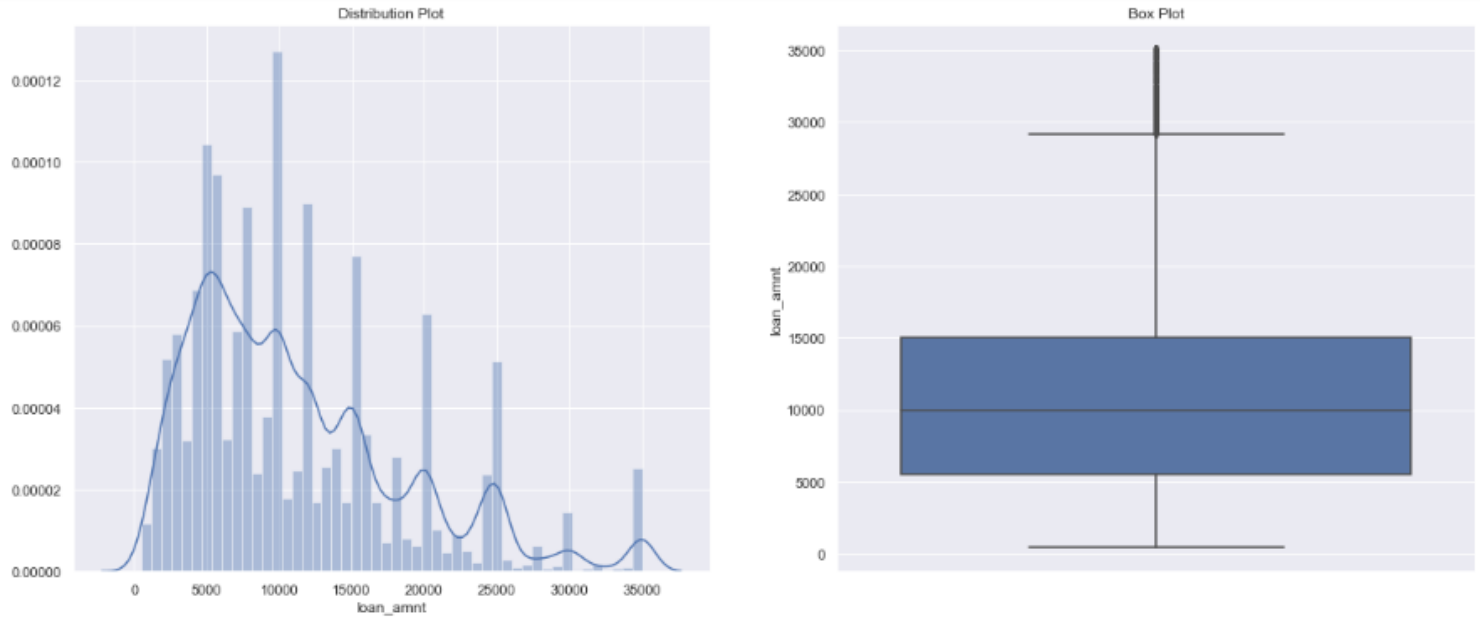
Analysis

- **Uni-variate Analysis**
  - **Continuous Variables**
    - Loan Amount
    - Interest Rate
    - Annual Income
  - **Categorical Variables**
    - Loan Status
    - Purpose of loan
    - Home Ownership wise Loan
    - Year wise Loan
    - Loan Term

Analysis

- **Bi-variate Analysis**
  - Correlation Matrix : All Continues (Numeric) Variables
  - Purpose of Loan vs. Loan Amount for each Loan Status
  - Employment Length vs. Loan Amount for different purpose of Loan
  - Location vs. Probability Charge Off
  - Grade / Sub grade vs. Probability Charge Off
  - Annual Income Range vs. Probability Charge Off
- **Derived Metrics Analysis**
  - Loan amount to Annual Income ratio
  - Extract Year & Month from Issue date
  - Change order of months from Jan to Dec, currently it's in alphabetical order(A-Z)
  - Create Bins for range of Loan Amount
  - Create Bins for range of Annual Income
  - Create Bins for range of Interest rates

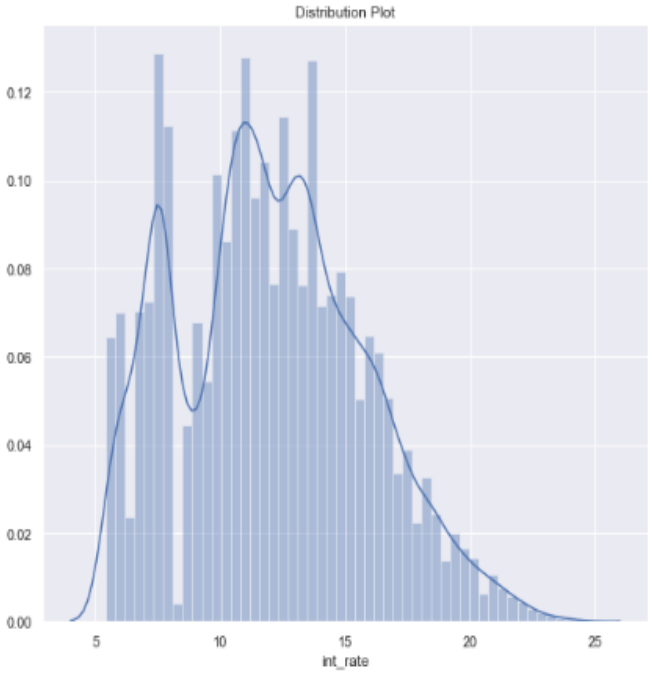
Univariate analysis - Loan Amount

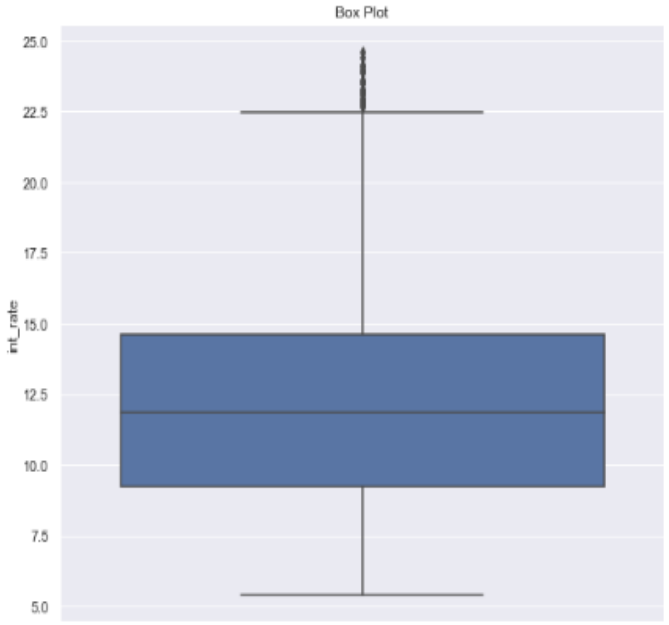


Insights: Most of the loan amounts are distributed between 6000 to 15000 USD.

We see that Most of the loan amounts are distributed between 6000 to 15000 USD.

Univariate analysis - Interest Rate

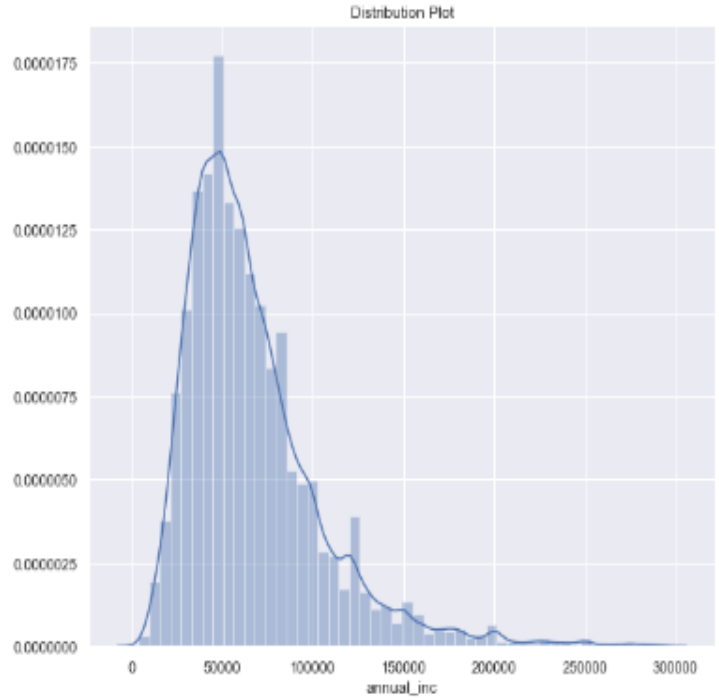


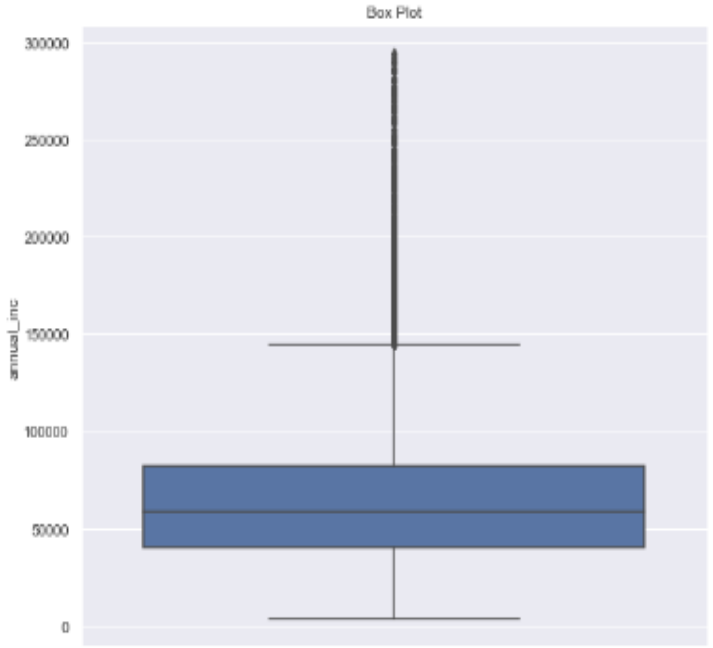


Insights: Most of the loans interest rates are distributed between 9% to 14.5%.

We see that Most of the loans interest rates are distributed between 9% to 14.5%.

Univariate analysis - Annual Income

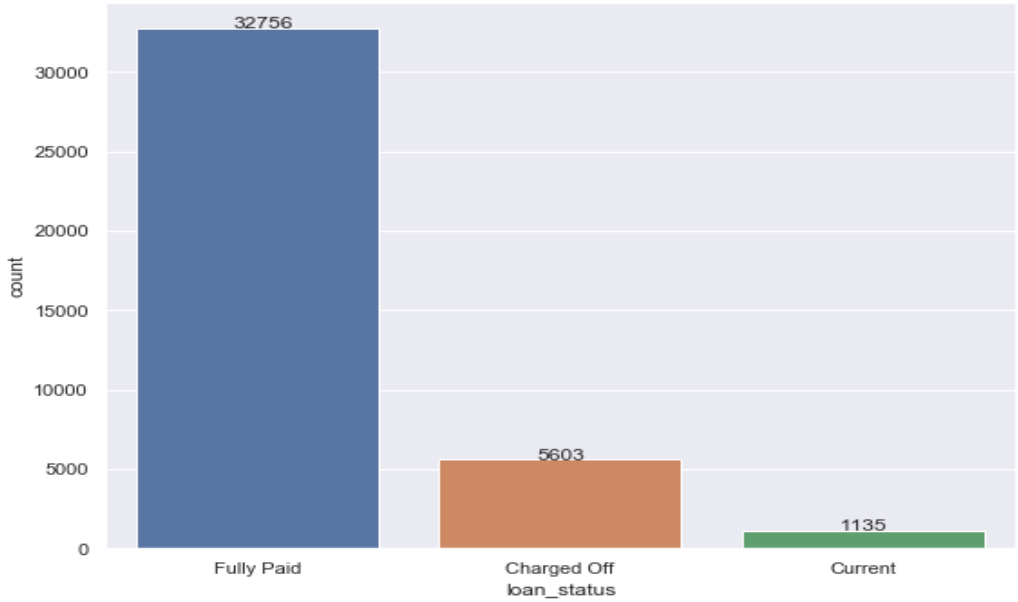


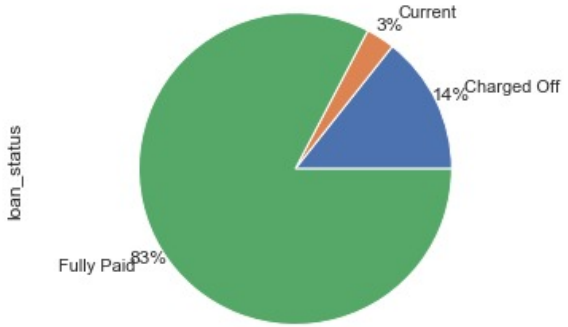


We see that Most of the applicants earns between 40000 to 80000 USD annually.

Insights: Most of the applicants earns beteen 40000 to 80000 USD annually.

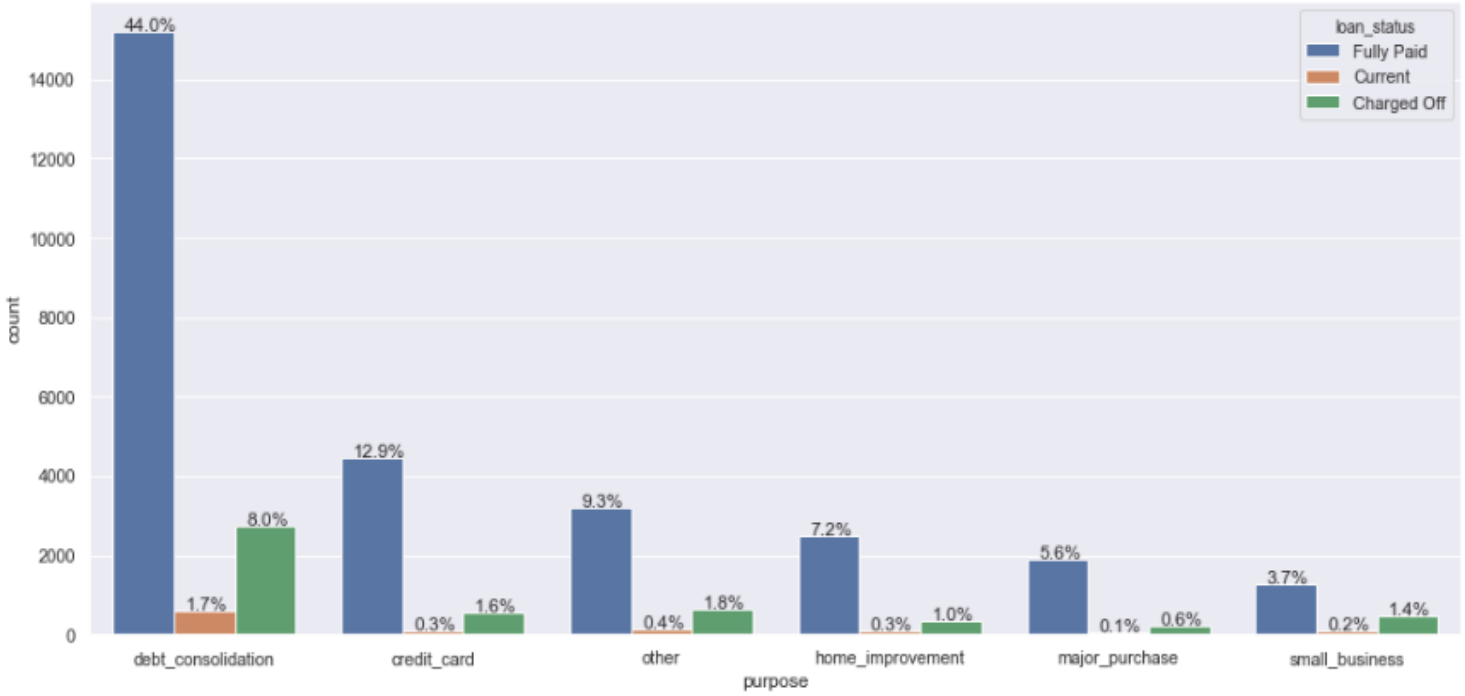
Univariate analysis - Loan Status





We see that : 14% of the applicants Charged off.

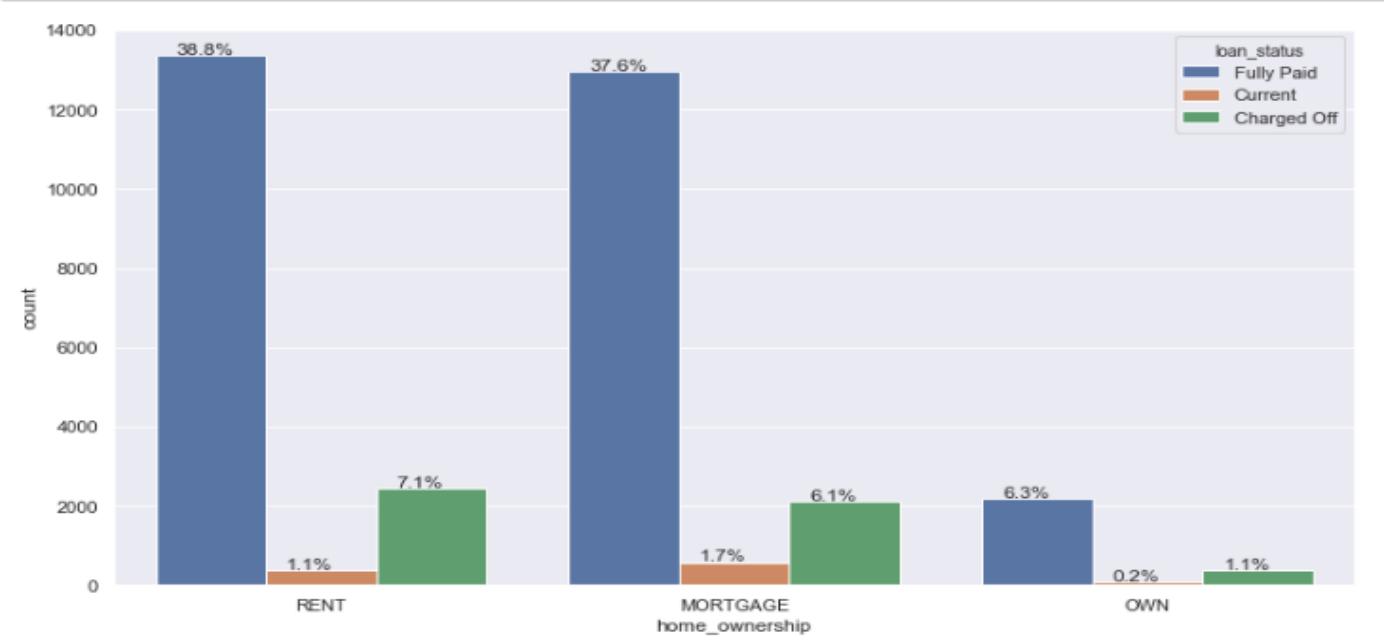
Univariate analysis - Purpose of loan



We see that about 53% of the applicant of loan taken for Debt consolidation

Insights: Approx 53% of the applicants applied loan for paying their other loans(Debt Consolidation).

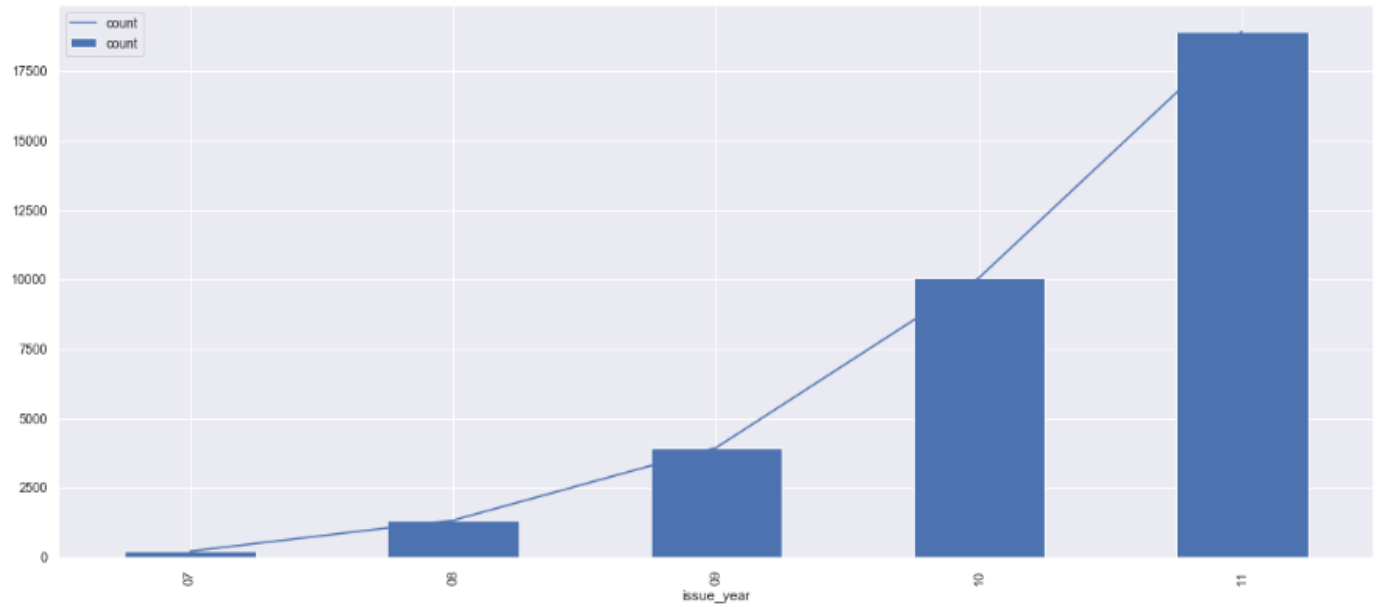
Univariate analysis - Home Ownership Vs Loan



Insights: 47% of applicants are living in rented home whereas 45% applicants were mortagaged their home.

We see that 47% of applicants are living in rented home whereas 45% applicants were mortgaged their home

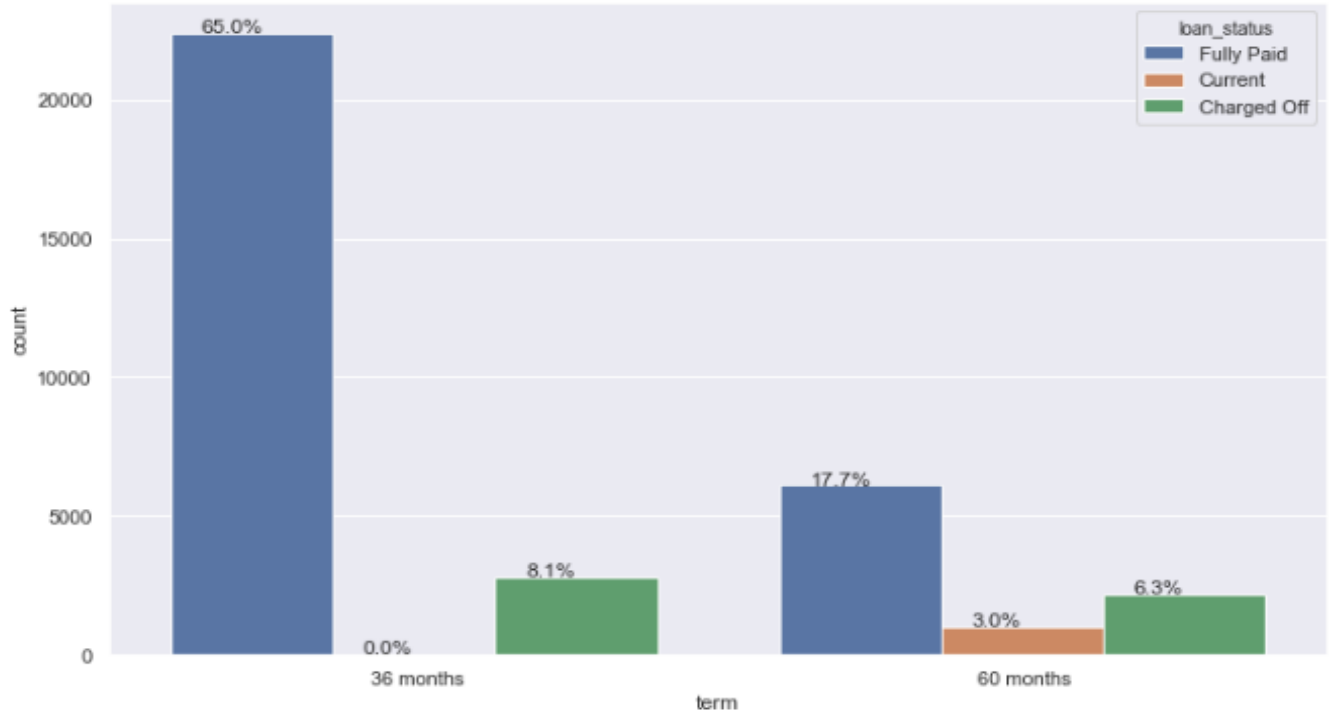
Univariate analysis - Year wise Loan



Insights: loan applicants are increasing year on year

We see that loan applicants increase during 2010 and 2011year.

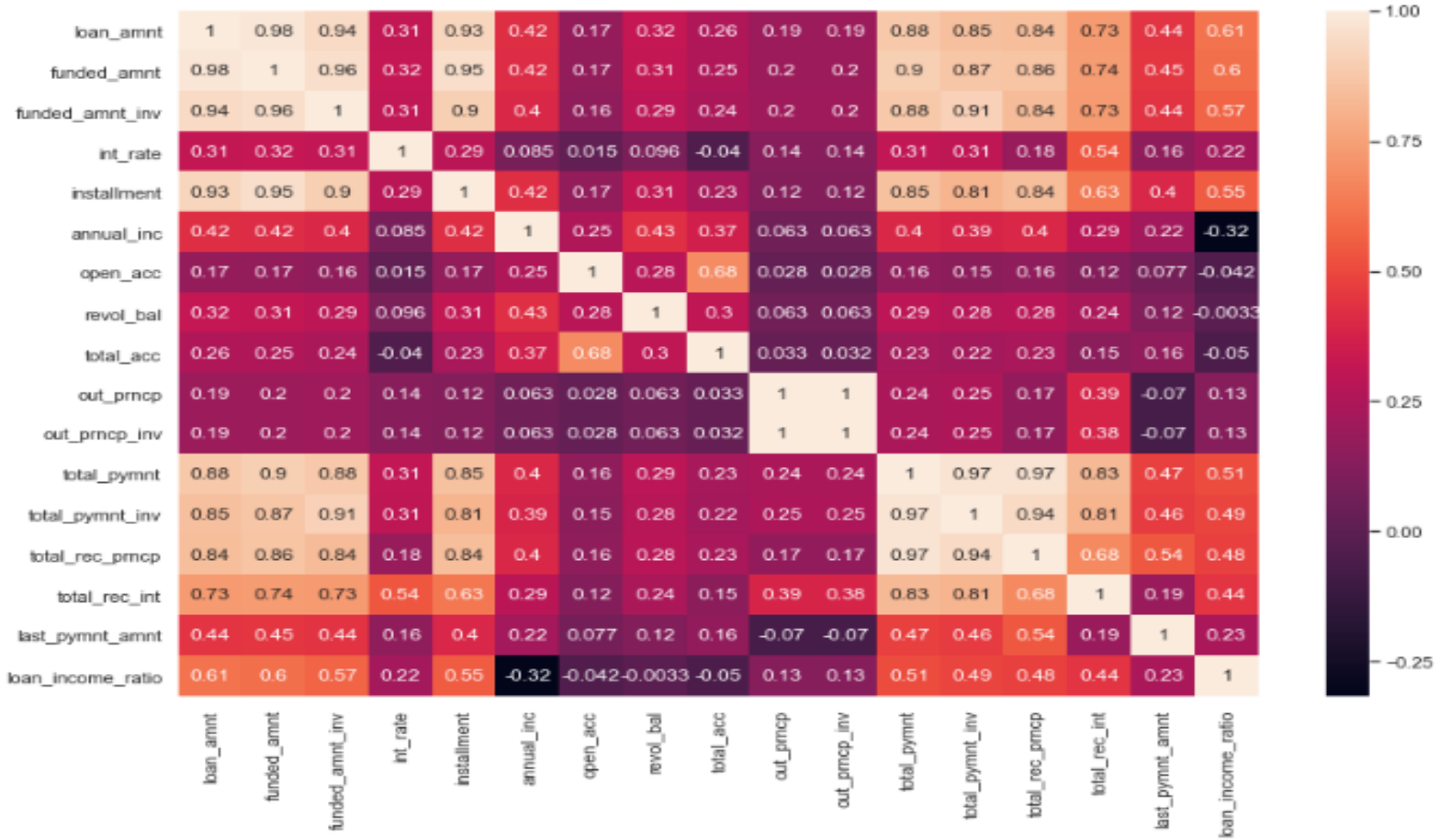
Univariate analysis - Loan Term



Insights: 73% of applicants applied loan for 36 months term period.

Insights: 73% of applicants applied loan for 36 months term period.

Bivariate Analysis correlation

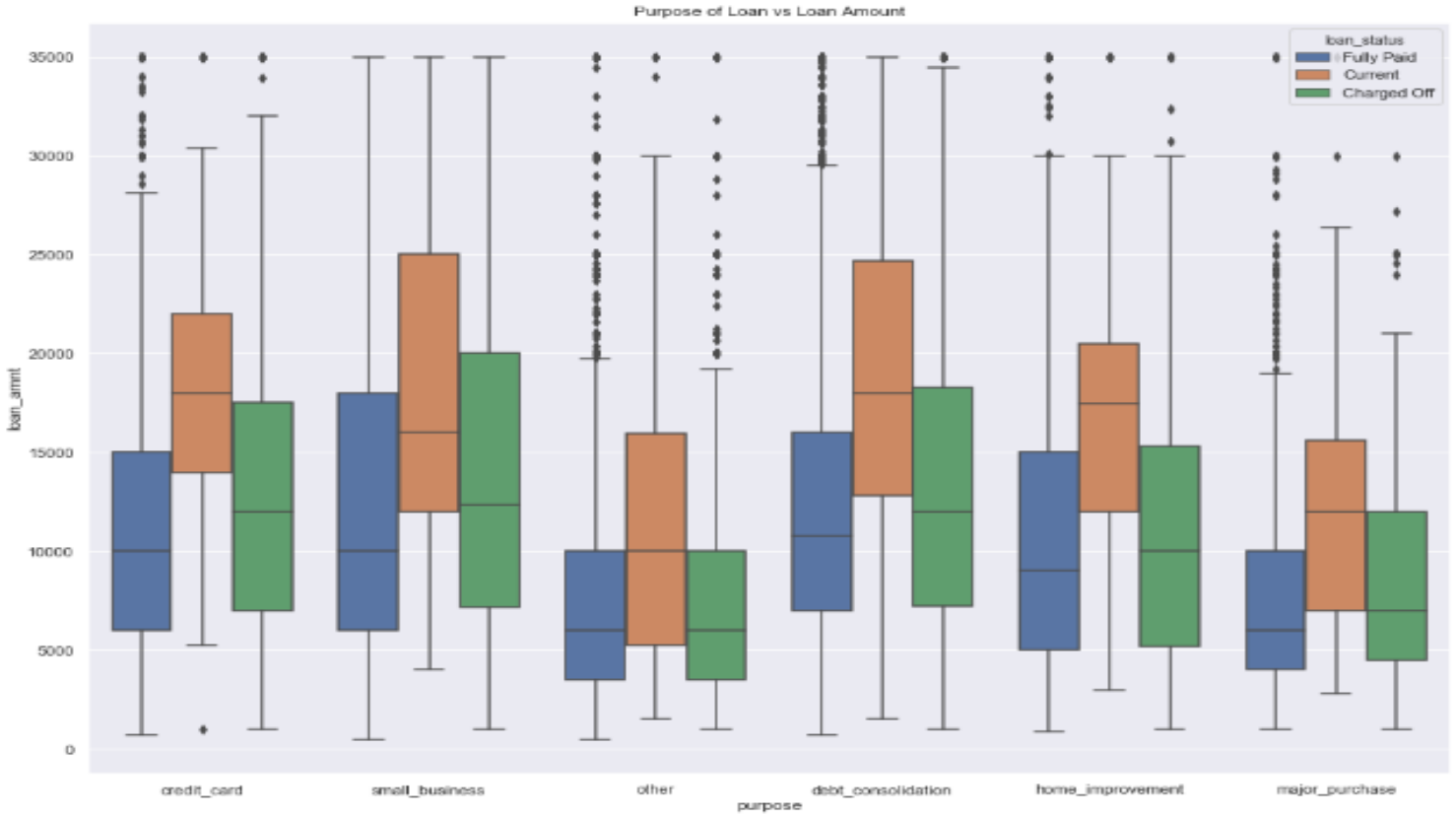


Bivariate Analysis correlation

We found there is strong correlation between below variable.

- Loan amount and funded amount
- Funded amount investor and total pay.
- Loan amount and total payment
- Principal received to date and Total payment
- Principal received to date total payment investor
- Installment and loan amount
- Installment and funded amount
- Installment and funded amount inv

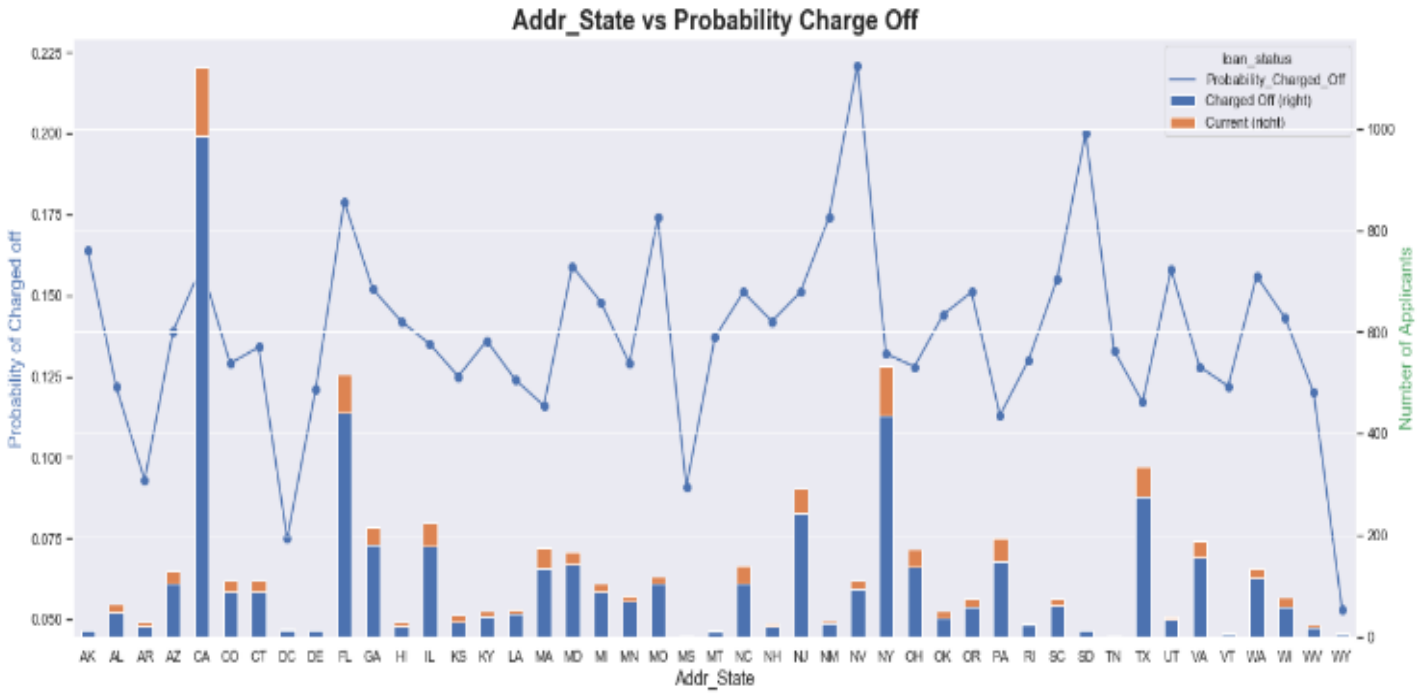
Bivariate Analysis - Purpose of Loan vs Loan Amount for each Loan Status



Insight : small business are having more number of chanrge off

We see that small business having more number of charged off

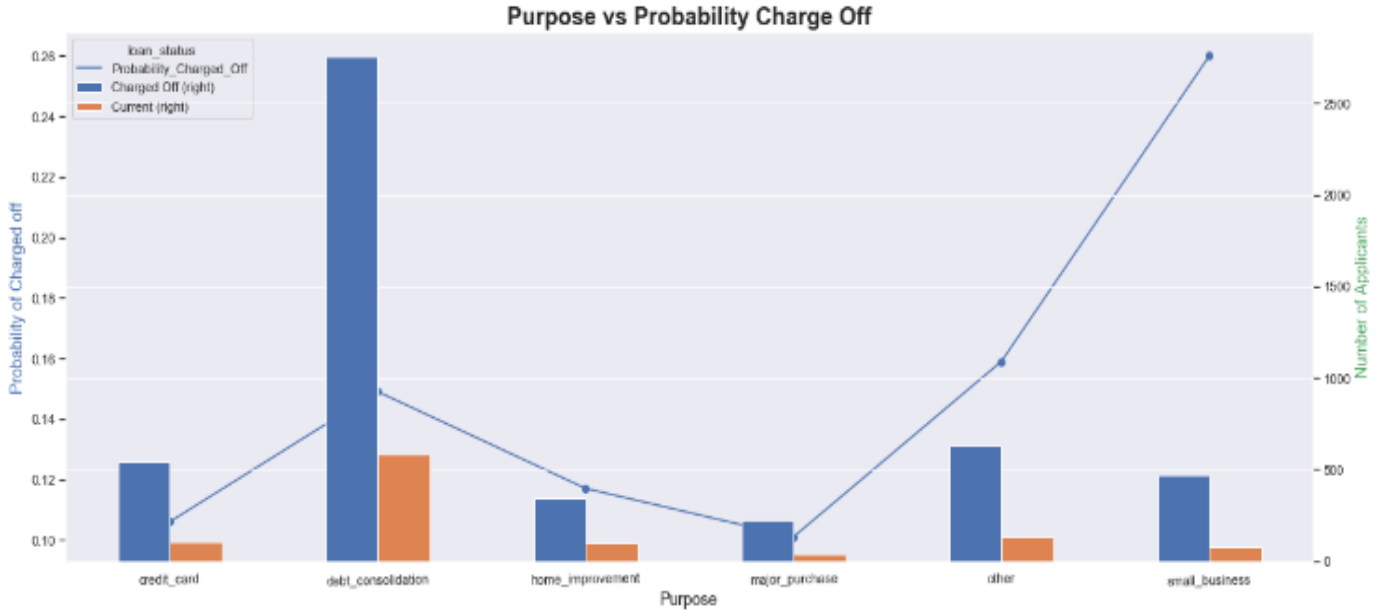
Bivariate Analysis - Location vs Probability Charge Off



Insights: There are multiple States/Provinces with high probability of charge,highest being 'NV' at 0.22

We see that multiple States/Provinces with high probability of charged off. Highest being 'NV' at 0.22

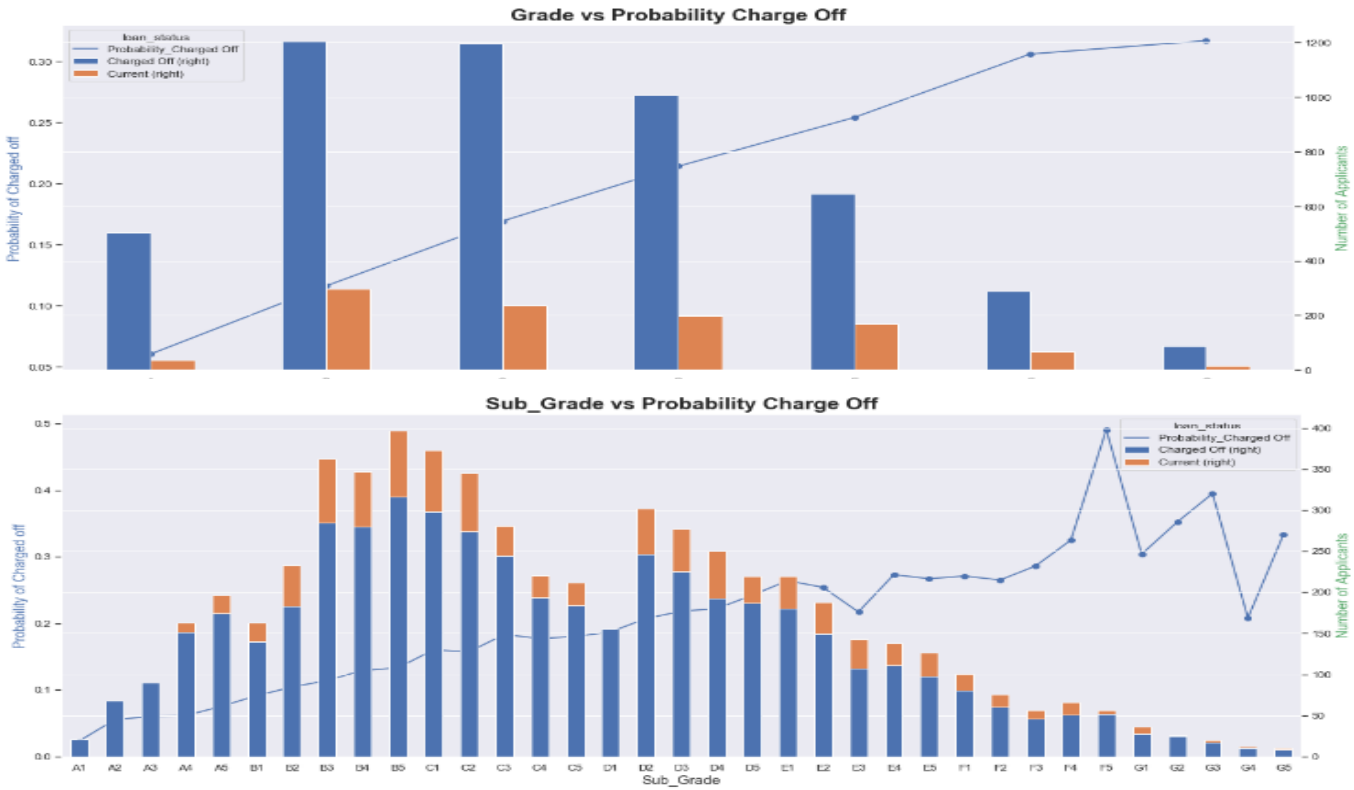
Bivariate Analysis - Loan vs Probability Charge Off



Insights: Applicants who has taken the Loan for 'small business' has the highest probabilty of charge off of 26%. So bank should take extra caution like take some asset or guarentee while approving the loan for purpose of 'small business'

Applicants who has taken the Loan for 'small business' has the highest probability of charge off of 26%. So bank should take extra caution like take some asset or guarantee while approving the loan for purpose of small business

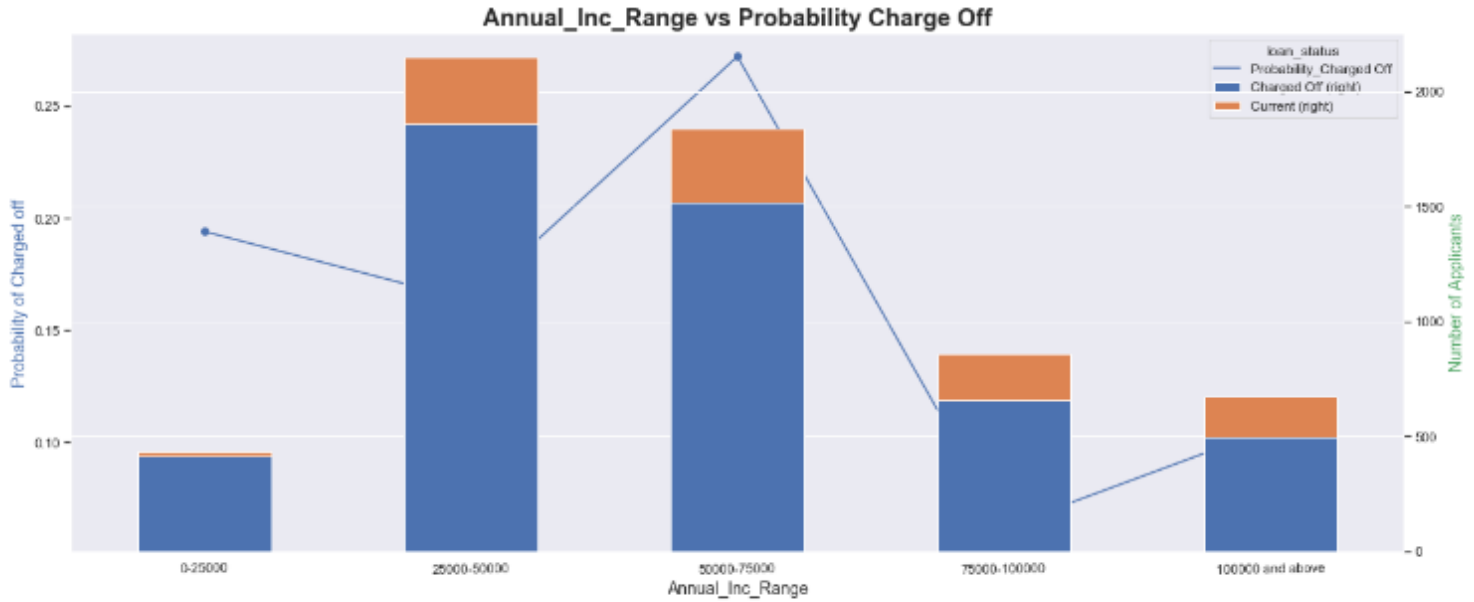
Bivariate Analysis- Grade/Sub grade vs Probability Charge Off



Insights: As we move from Grade A to G, probability that person will charged off is increasing.

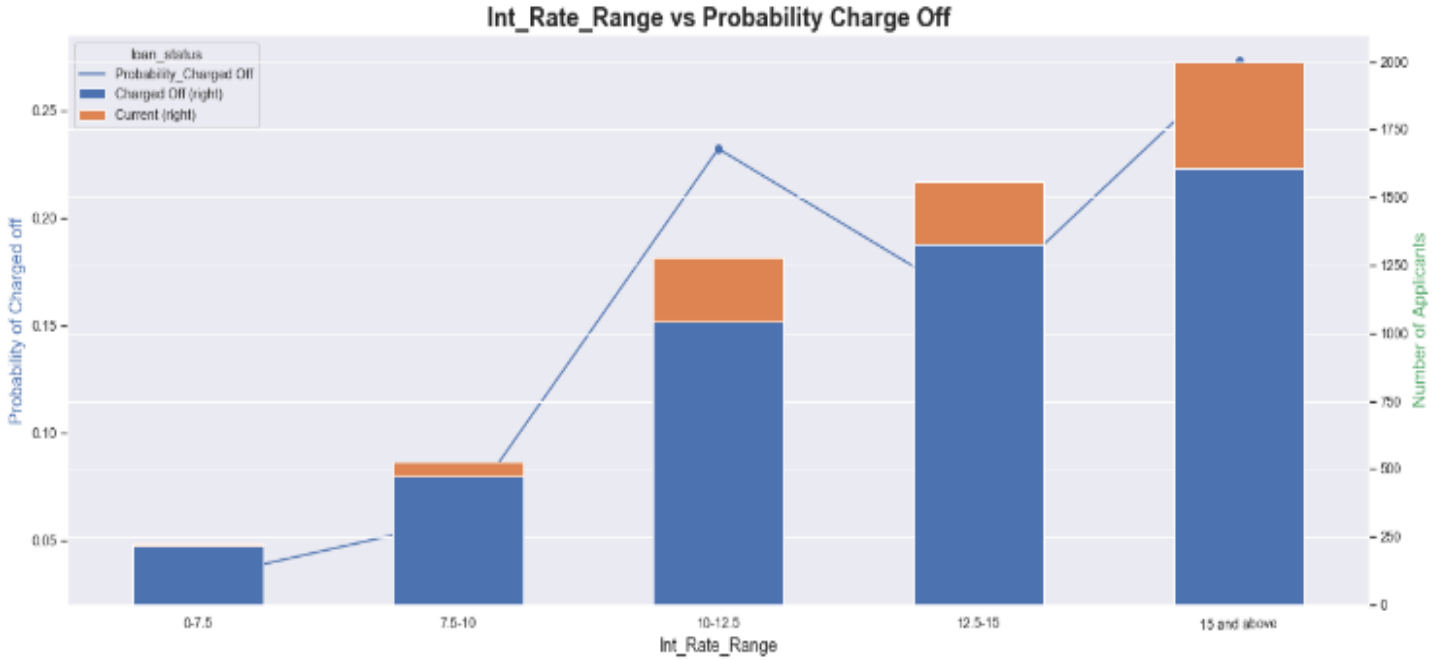
As we move from Grade A to G, probability that person will charged off is increasing

Bivariate Analysis- Annual Income Range vs Probability Charge Off



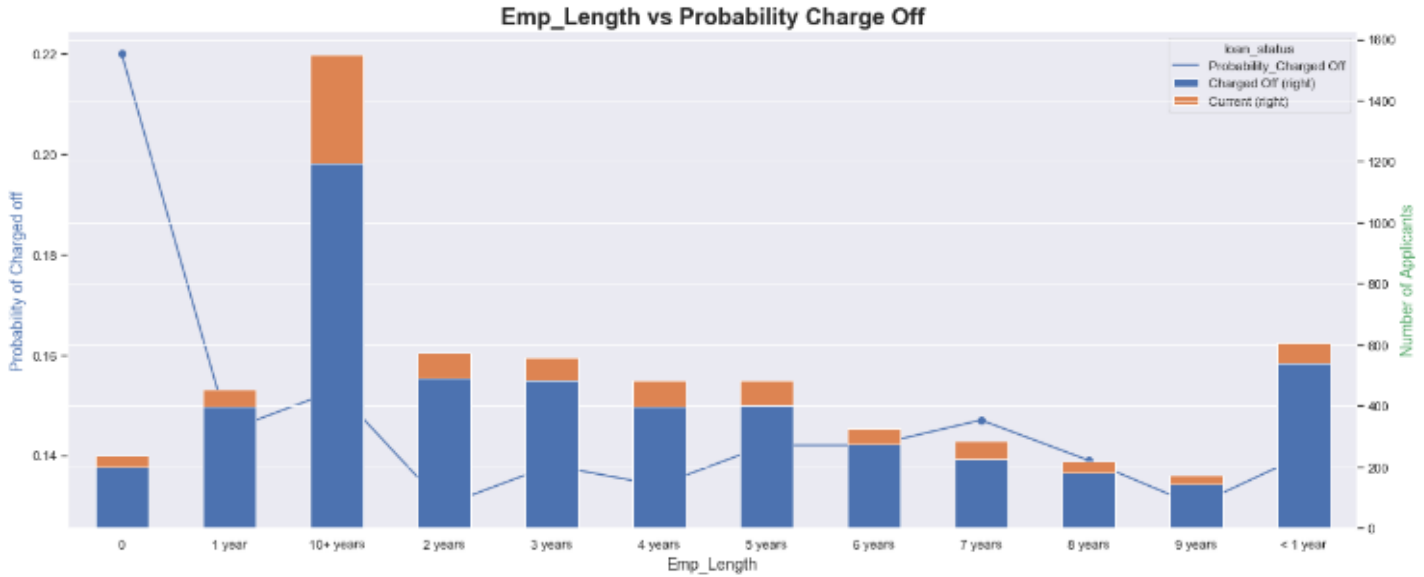
The probability that person will default is highest of 25% at (50000 to 75000) salary bracket.

Bivariate Analysis- Interest rate Range vs Probability Charge Off



As the interest rate is increasing the probability that person will default is increasing with highest of 28% at 15 & above bracket.

Bivariate Analysis- Employment Length vs Probability Charge Off



Insights: Applicants who are less than 1 year of experience are more probable of charged off..

Applicants who are less than 1 year of experience are more probable of charged off..

Conclusion:

- Giving loan to Depth consolidation is having higher risk as it has around 8% of charged-off
- Approving loan to applicant who are having rented or mortgage home is having more risk as we see 7 % and 8% applicant lying under charged off.
- We observe 36 month category loan is having 8.1 of charge off list.
- We see that small business having more number of charged off.
- We see that multiple States/Provinces with high probability of charged off. Highest being 'NV' at 0.22
- Applicants who has taken the Loan for 'small business' has the highest probability of charge off of 26%. So bank should take extra caution like take some asset or guarantee while approving the loan for purpose of small business.
- As we move from Grade A to G, probability that person will charged off is increasing
- The probability that person will default is highest of 25% at (50000 to 75000) salary bracket.
- Applicants who are less than 1 year of experience are more probable of charged off.