

Lending Club Case Study

Topics

1. Business Understanding & Objectives
2. Data Understanding
3. Data Preparation
4. Univariate Analysis
5. Bivariate Analysis
6. Multivariate Analysis
7. Outcomes of analysis

Topics

1. Business Understanding & Objectives
2. Data Understanding
3. Data Preparation
4. Data Analysis
5. Outcomes of analysis

Business Understanding

The Consumer finance company lending loans to urban customers. The company wants to given loan as much as possible.

The company also wanted to avoid giving loans to defaulters as much as possible.

Objectives

The company wanted to minimize issuing loan to possible defaulters

The company wants to increase the profit by increasing interest rates for high risk customers.

Data Understanding

The data provided contains the information about the past loan applicants.

The data provide the loan status information Fully Paid, Current and Charged-Off.

Data Preparation



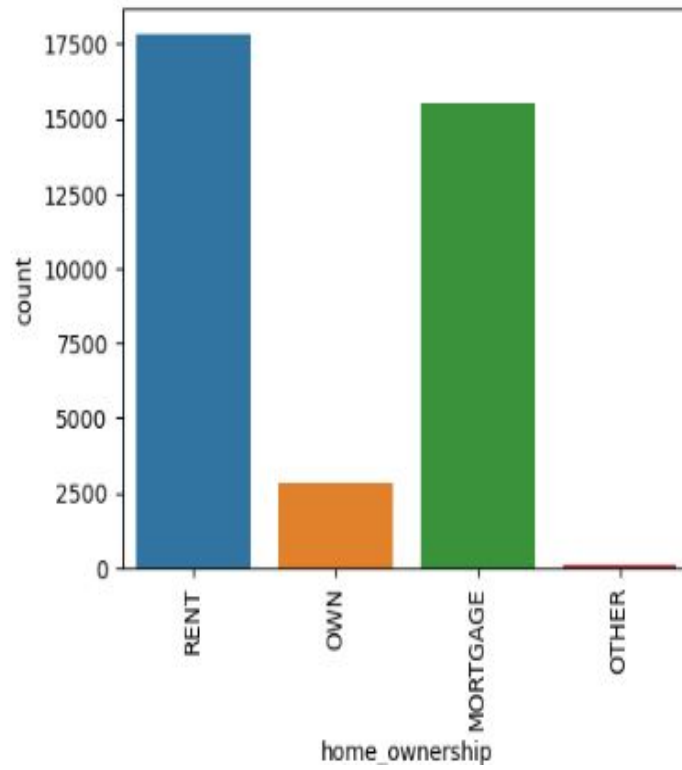
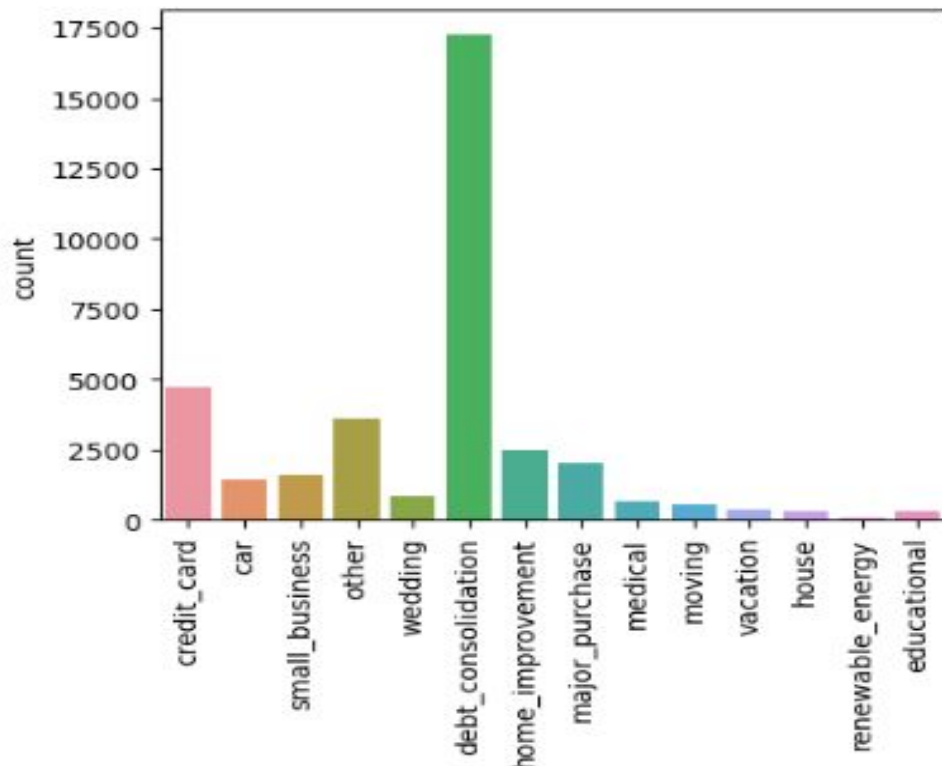
Data Analysis

1. Loan dataset is of size 39717 * 111. Analysing 111 columns is huge task. Remove unwanted portions of this dataset.
2. Remove columns which are having more than 30% of empty values
3. Some of columns are having NULL values. Either impute with meaningful values or remove columns if not needed.
4. Segregate continuous column to categorical column such as column 'dti','annual_inc' etc
5. Segregate categorical column to continuous column such as column 'grade' etc
6. Format some of the column to convert to integer such as int_rate or term etc
7. Drop unwanted columns

```
colToDrop = ['member_id', 'url', 'title', 'emp_title', 'zip_code', 'tax_liens', 'acc_now_delinq',  
'chargeoff_within_12_mths', 'delinq_amnt', 'application_type', 'policy_code', 'initial_list_status',  
'installment', 'pymnt_plan', 'collections_12_mths_ex_med', 'total_pymnt', 'out_prncp',  
'total_rec_prncp', 'total_rec_int', 'last_pymnt_d', 'last_pymnt_amnt', 'last_credit_pull_d']
```

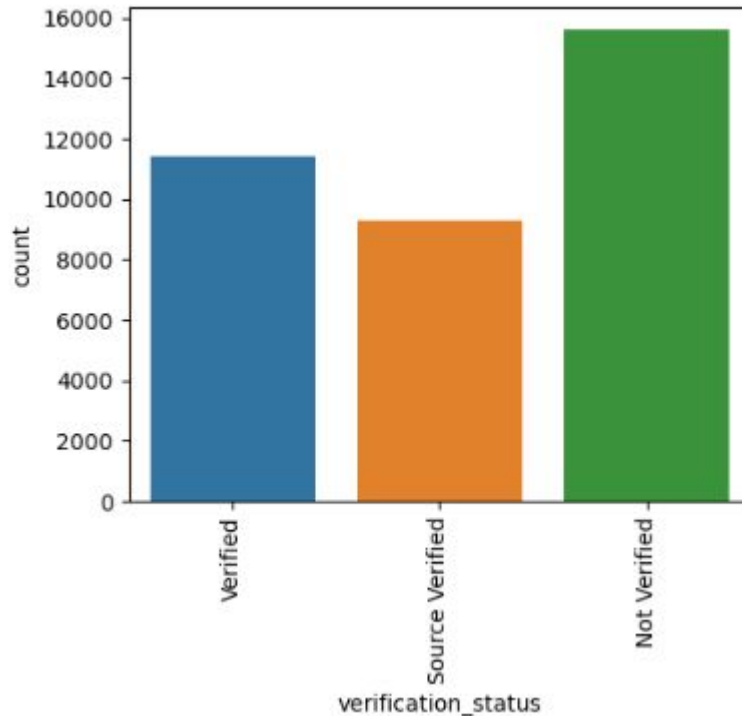
Univariate Analysis : Categorical Columns

1. People who are in Rent or their home is under Mortgage are major loan seekers
2. Top reasons for purchasing loans are Debt consolidation and Credit Card



Univariate Analysis : Continuous/Categorical Columns

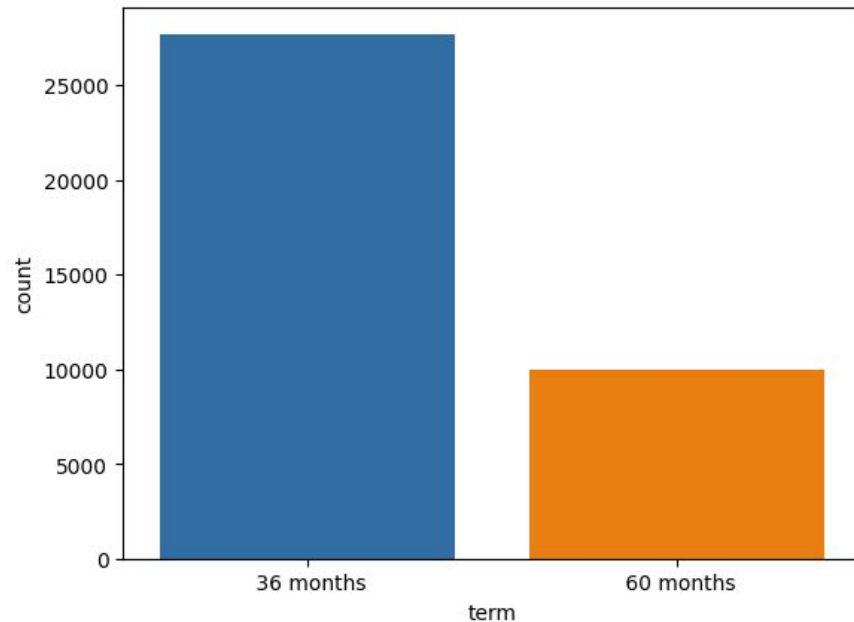
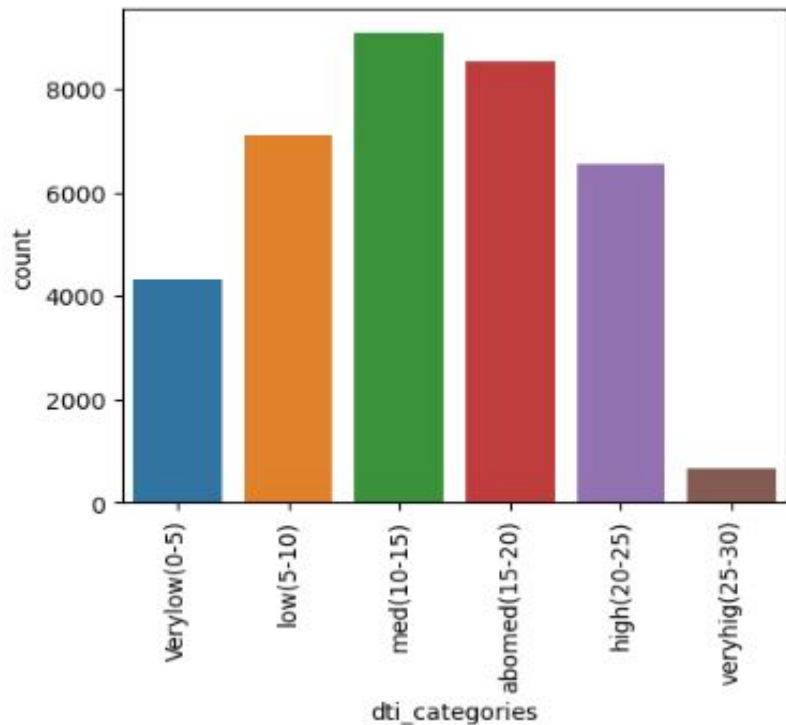
3. Among loan applicants still major portion of people are not yet verified.



Univariate Analysis : Continuous/Categorical Columns

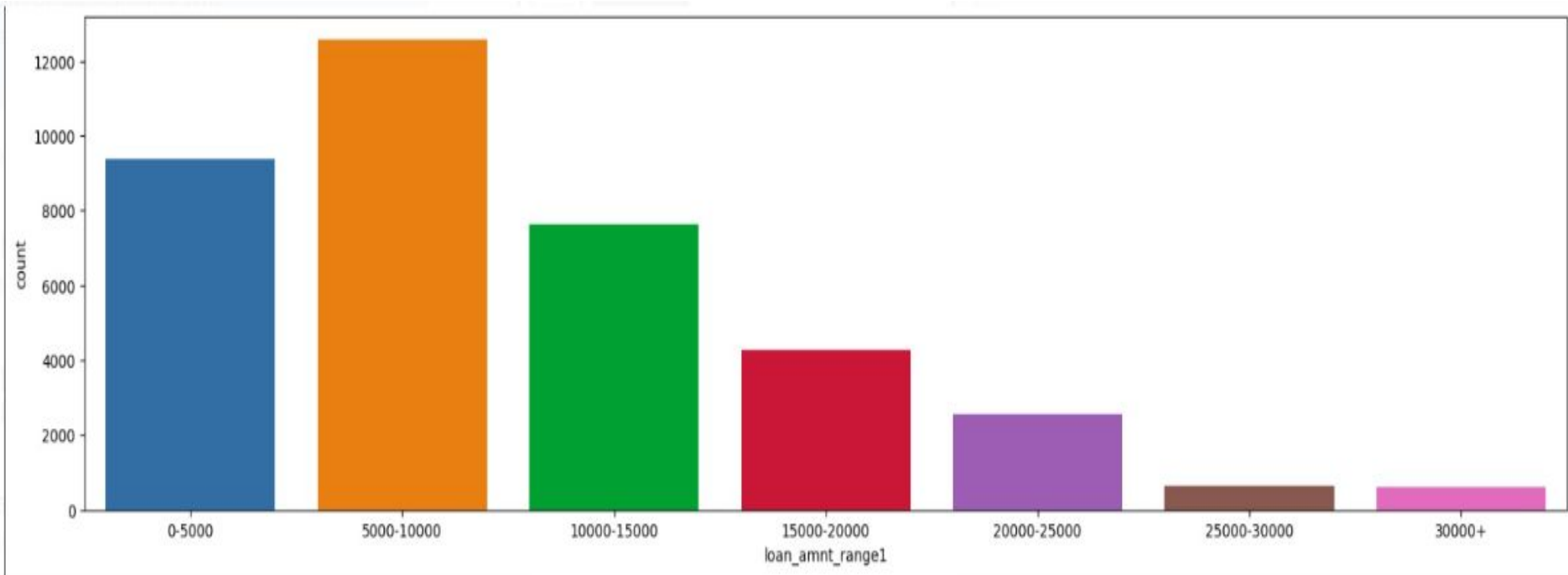
4.DTI ratio of majority people lies between 10 to 20

5. Most loans are of 36 months term



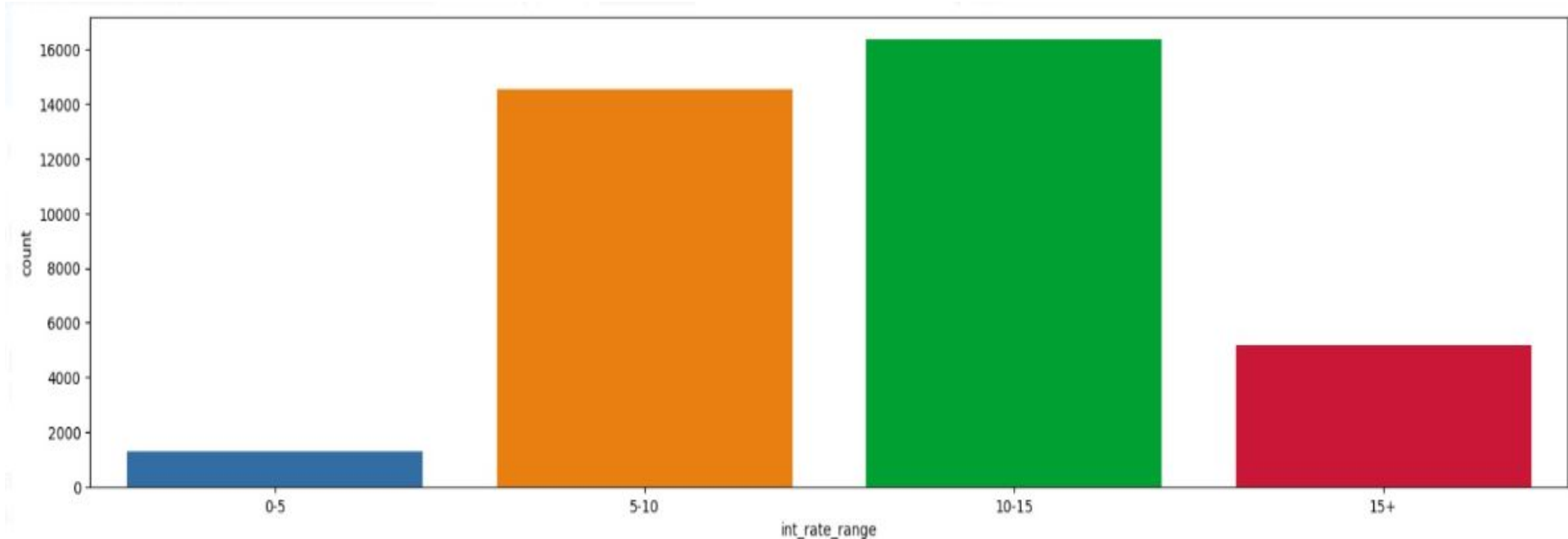
Univariate Analysis : Continuous/Categorical Columns

6. Most loan amount are from 5000 - 15000



Univariate Analysis : Continuous/Categorical Columns

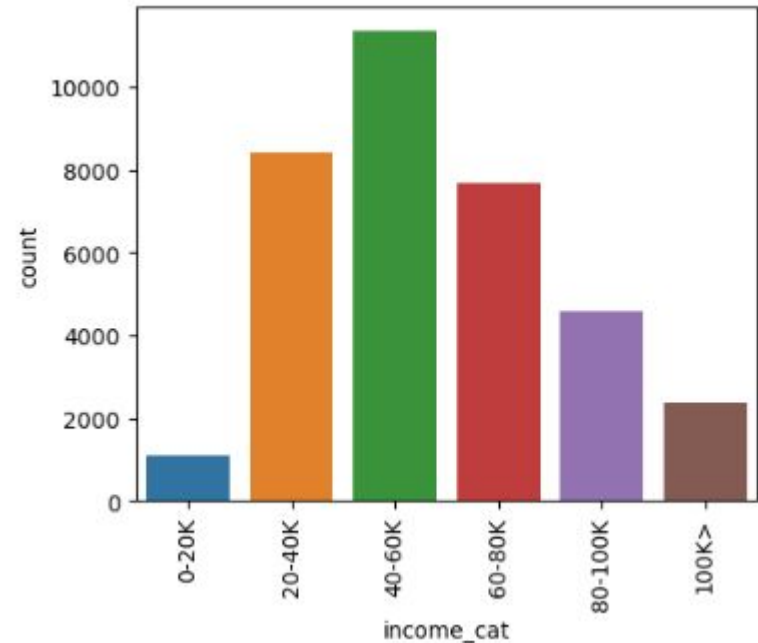
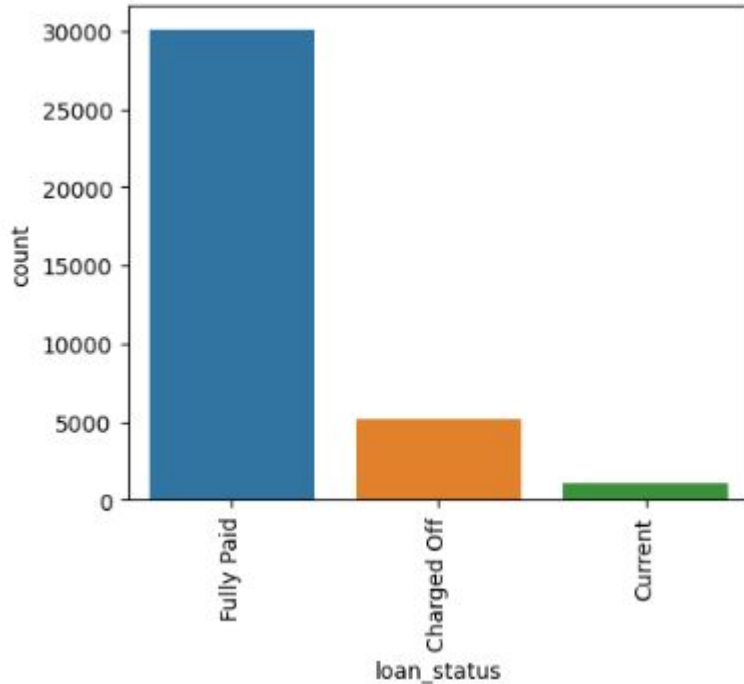
7. Most interest rate ranged from 10% - 15%



Univariate Analysis : Continuous/Categorical Columns

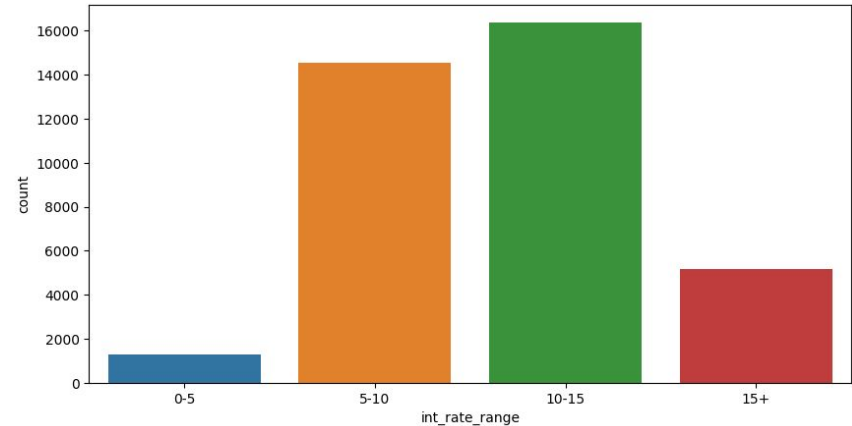
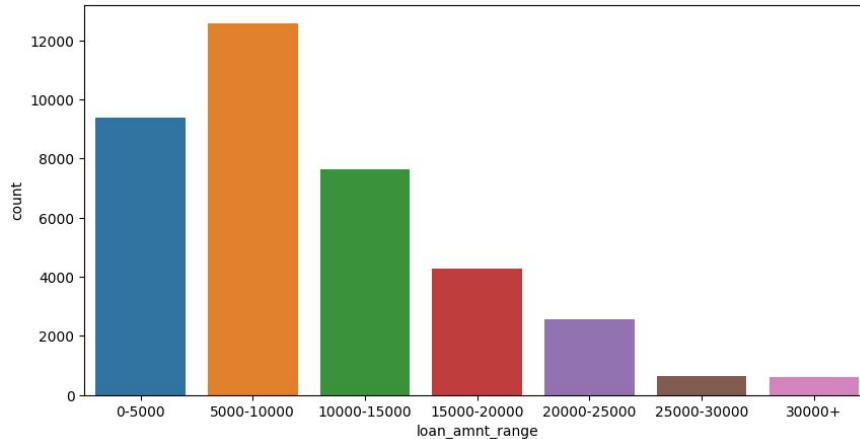
8. Majority of Loans are Fully Paid

9. Majority of People have annual income ranging from 40k - 60k



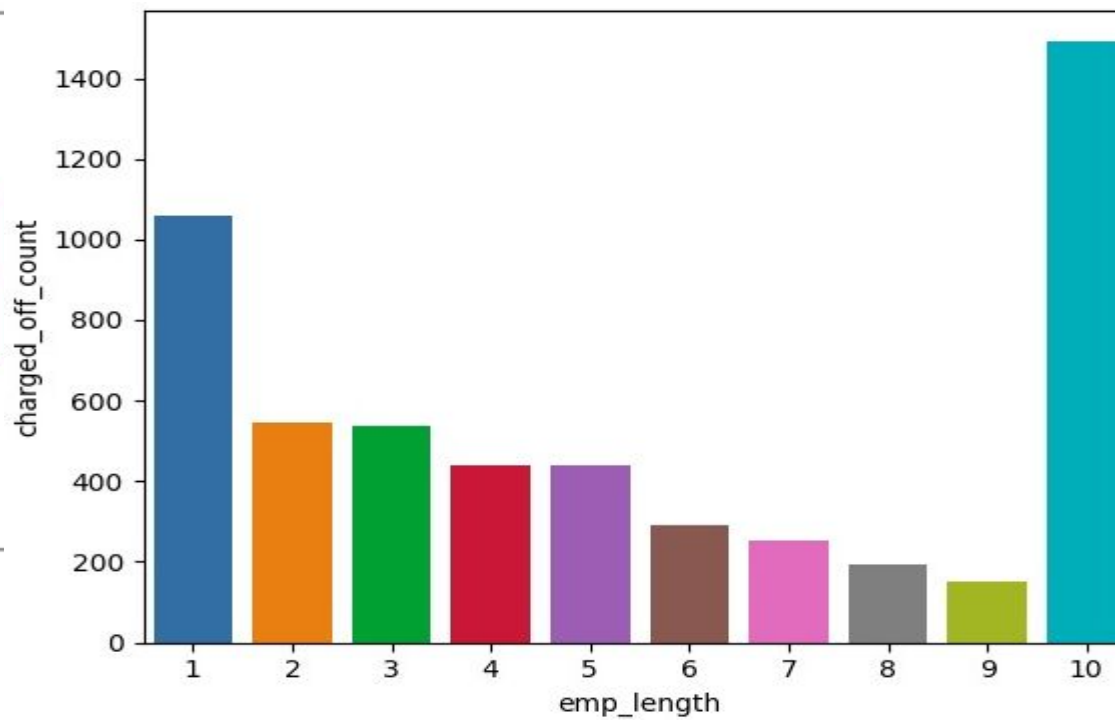
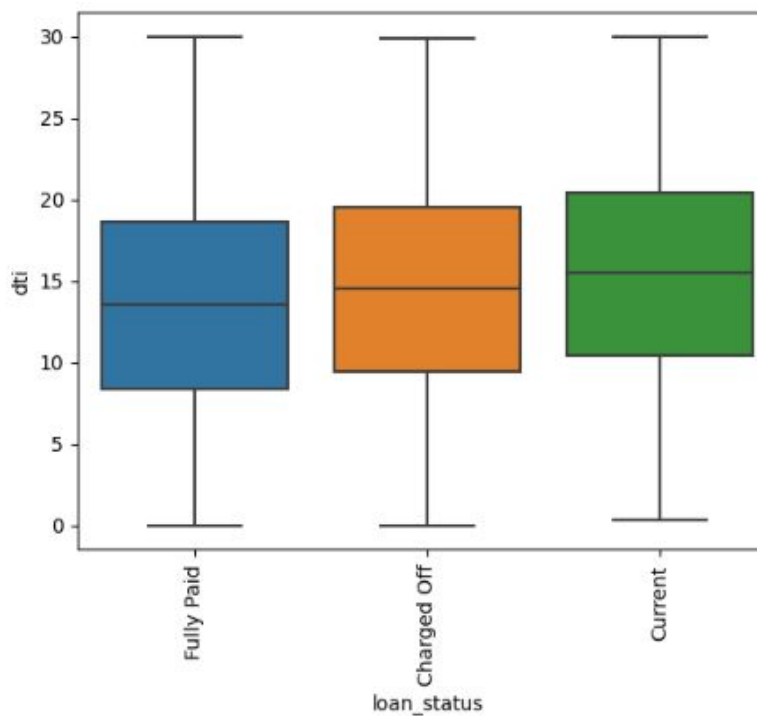
Basic data Insights

1. Most loan amount are from 5000 - 15000
2. Most interest rate ranged from 10% - 15%



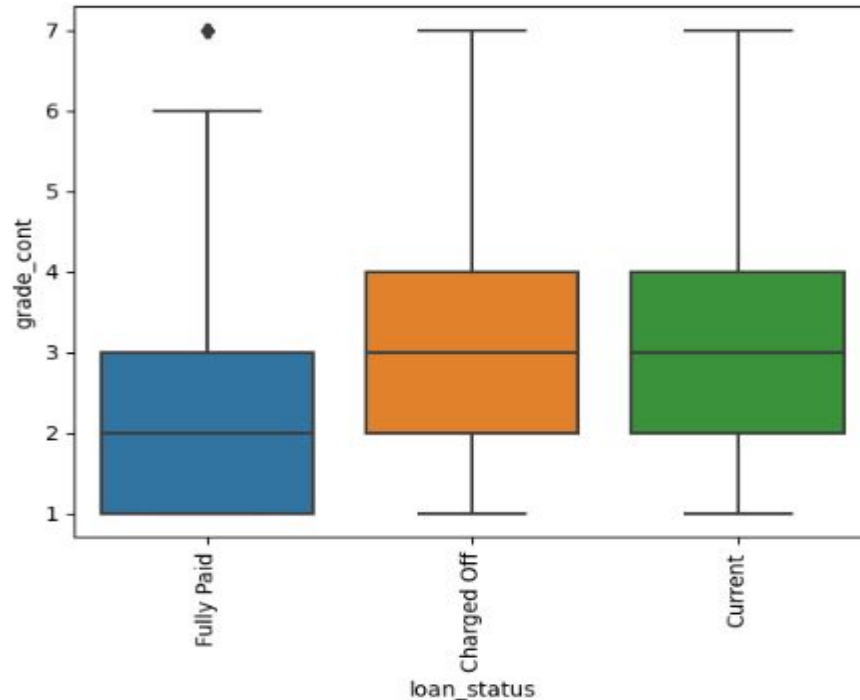
Bivariate Analysis: Categorical vs Continuous columns

1. Emp_length with more that 10+ have more defaults
2. People with dti of 10 to 20, likely to be 'charged off'



Bivariate Analysis: Categorical vs Continuous columns

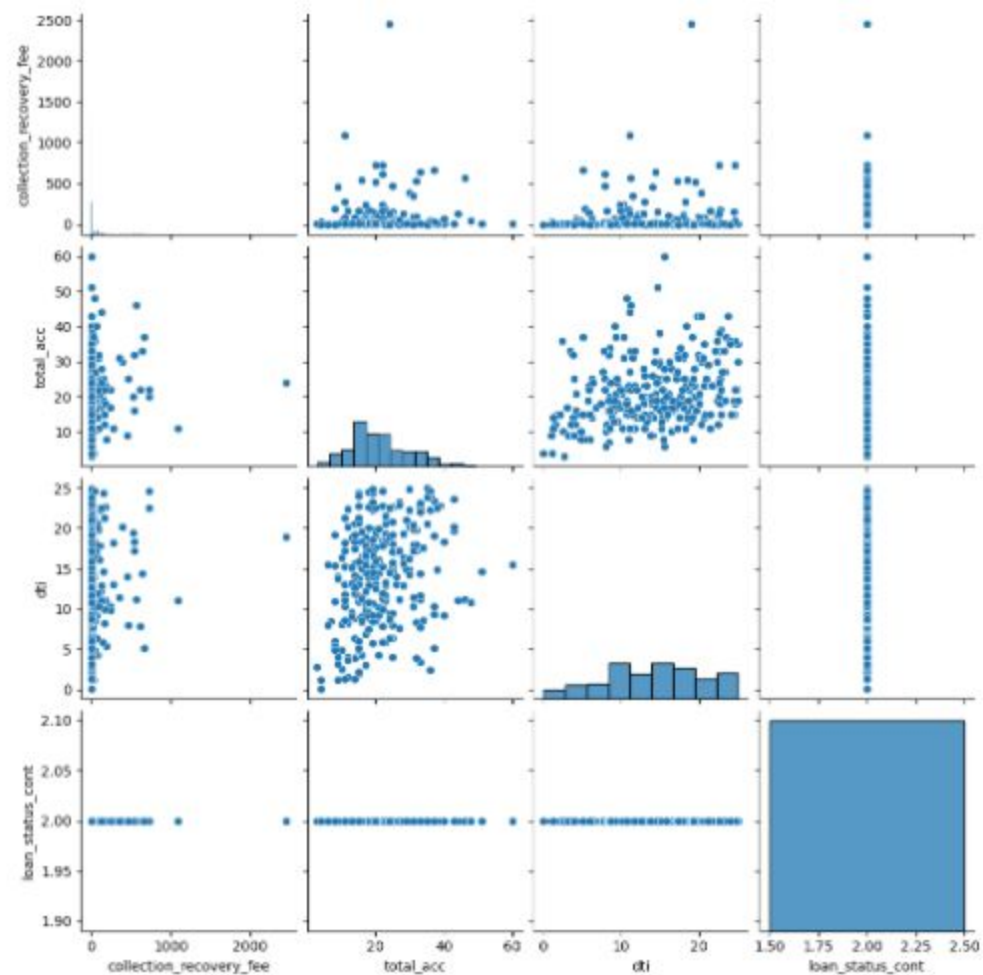
3. Among 'Charged Off' candidates, people who have secured loan grade as B, C or D (2=B, 3=C, 4=D in y axis) are more likely to default than other categories



Bivariate Analysis: Continuous vs Continuous columns

Pair plots showing

1. Collection Recovery fee is high whenever dti ratio lies b/w 10 to 20
2. total credit lines are high whenever dti ratio lies b/w 10 to 20
3. Loan status count=2 stands for 'Charged Off' account, which is clearly high whenever people either paid Collection Recovery fee or have pub_rec or high dti ratio

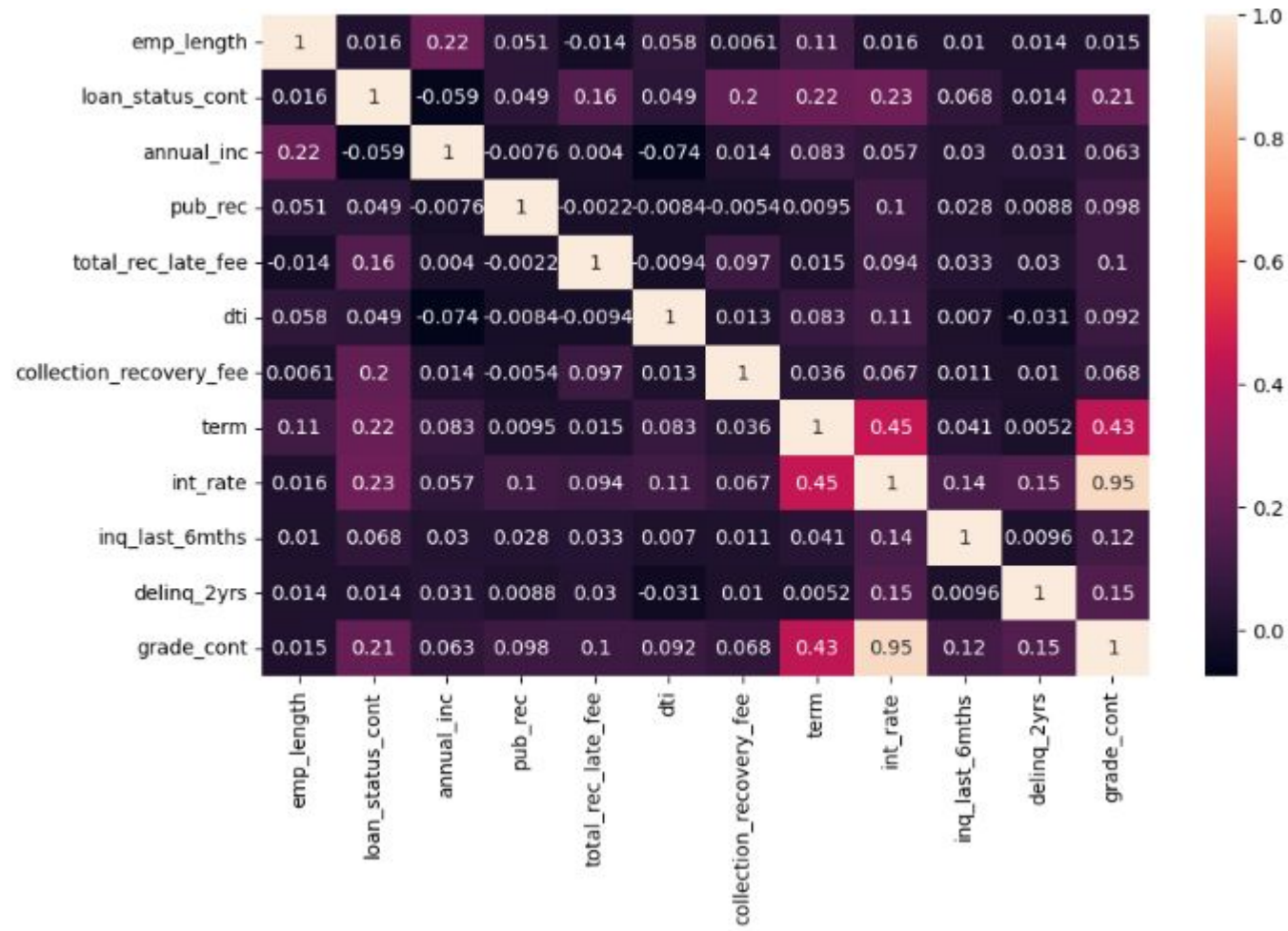


Multivariate analysis

1. People who enquired inquired 1 or more times i.e (columns : inq_last_6mths), in last 6 months have higher probability of defaulting loan
2. People whose Verification status as 'Source Verified' are less likely to default than people with other verifications status



Multivariate Analysis using Heat map



Outcomes of Analysis

1. Majority of 'Charged Off' loans belongs to Loan Grade B, C or D
2. People who are paying 'collection_recovery_fee' are more likely to get into 'Charged Off' the loan
3. People who have Derogatory public records are more likely to get into 'Charged Off' the loan
4. People with dti ratio lies b/w 10 to 20 are likely to 'Charged Off' the loan
5. People who enquired inquired more than 1 times i.e (columns : inq_last_6mths), in last 6 months have higher probability of defaulting loan
6. People whose Verification status as 'Source Verified' are less likely to default than people with other verifications status