

EXPLORATORY DATA ANALYSIS

CASE STUDY : LENDING CLUB

Sandeep Kumar
Sandeep Malagi

BACKGROUND

This Case Study is part of Graded Assignment under 'Exploratory Data Analysis' Module. We are expected to analyze the given dataset and do various steps involved in Data Understanding, Data Cleaning, Data Visualization & Data Summarization

BUSINESS UNDERSTANDING

- Case Study consist data analysis part for given dataset. As part of this exercise, Financial Firm have provided a dataset which is having Loan Application and Loan Status data from past.
- Before we deep dive into analysis lets try to understand how Financial firm Loan scenarios works.

Financial Firms gives Loan to Individuals. Once Loan given to Applicant then there are 2 types of Scenarios

- Scenario 1 : Applicant Paid the Loan Amount Fully on time
- Scenario 2 : Applicant become defaulter and not paying loan amount which leads to Credit Loss for Financial Firm

BUSINESS OBJECTIVE

Now as we understand that Financial Firms have challenges in evaluating Loan Application and choosing the one which is having least probability to become defaulter (or charged off) as it leads to Credit loss for Financial Firm.

Business wants us to Analyze the data and all possible variables available in Dataset and give recommendation about the attributes which are having relation with Charged Off cases. This analysis will help Financial Firm to take decision what to do if in future they receive Loan Application which is having similar variable values.

Core Objective

1- Reduce Credit Loss.

DATA UNDERSTANDING

Total No of Columns : 111

Total No of Rows : 39717

Dataset consists 3 Type of Data about customers

- **Demographic Attributes**
 1. Address
 2. Zip Code
 3. Annual Income
 4. Employment Title
- **Loan Specific Attributes**
 1. Loan Amount
 2. Funded Amount
 3. Rate of Interest
 4. Loan Tenure/Term
- **Behavioral Attributes**
 1. Loan Purpose
 2. Application Type
 3. Recoveries
 4. Earliest Credit Line

DATA CLEANING PROCESS



Null Check

- Removed Columns that were having more than 90% null values.



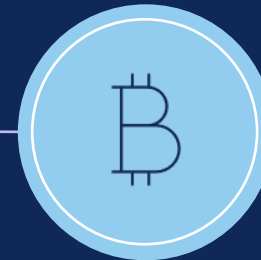
Single Value Columns

- Removed all those columns which were having Single values for all rows e.g. 'Policy Code' was having all values as '1'



Behavior data columns

- Removed behavior data columns as it was not adding value to analysis e.g. last_payment_date



Data Type Correction/Conversion

- Removed '%' from Interest_Rate,
- Issue_date column converted into date format



Derived Columns

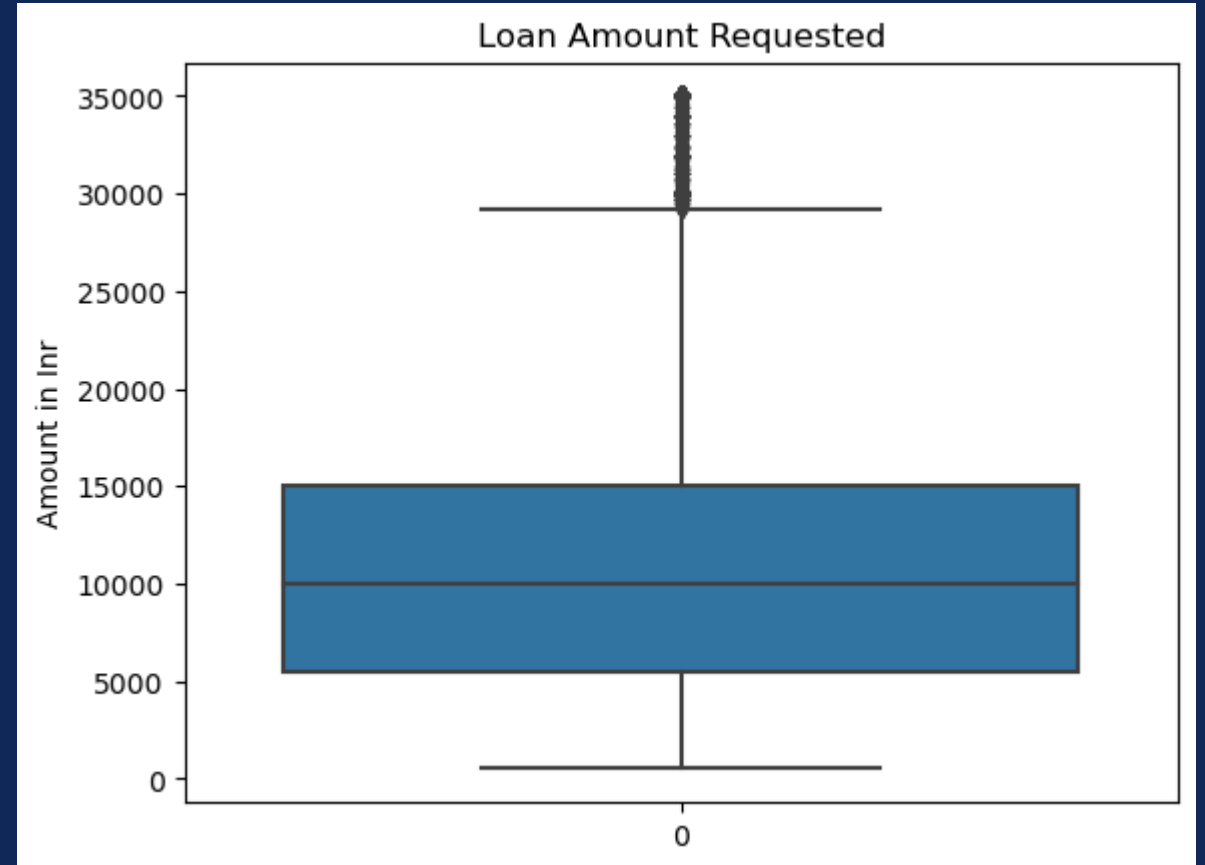
Created a derived column which shows ratio between Loan Amount Requested and Loan Amount funded

DATA ANALYSIS

Loan Attribute: Loan Amount Requested

Observation

- Median of loan Amount Requested is 10000 inr.
- There are very few people who have requested Loan amount more than 30000 inr.

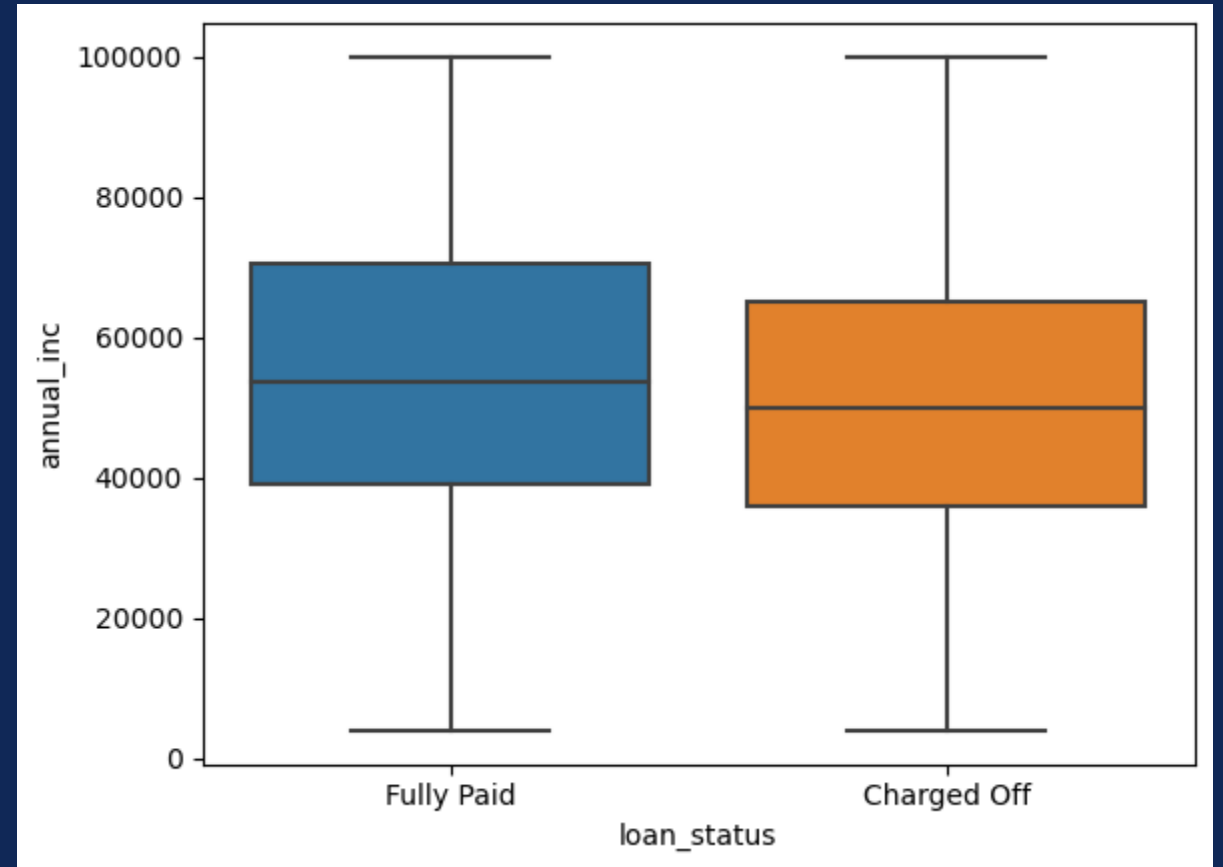


DATA ANALYSIS

Loan Attribute: Loan Status vs Annual Income of Applicant

Observation

- Charged-Off Loan cases belongs to Low Median Annual Income Group.



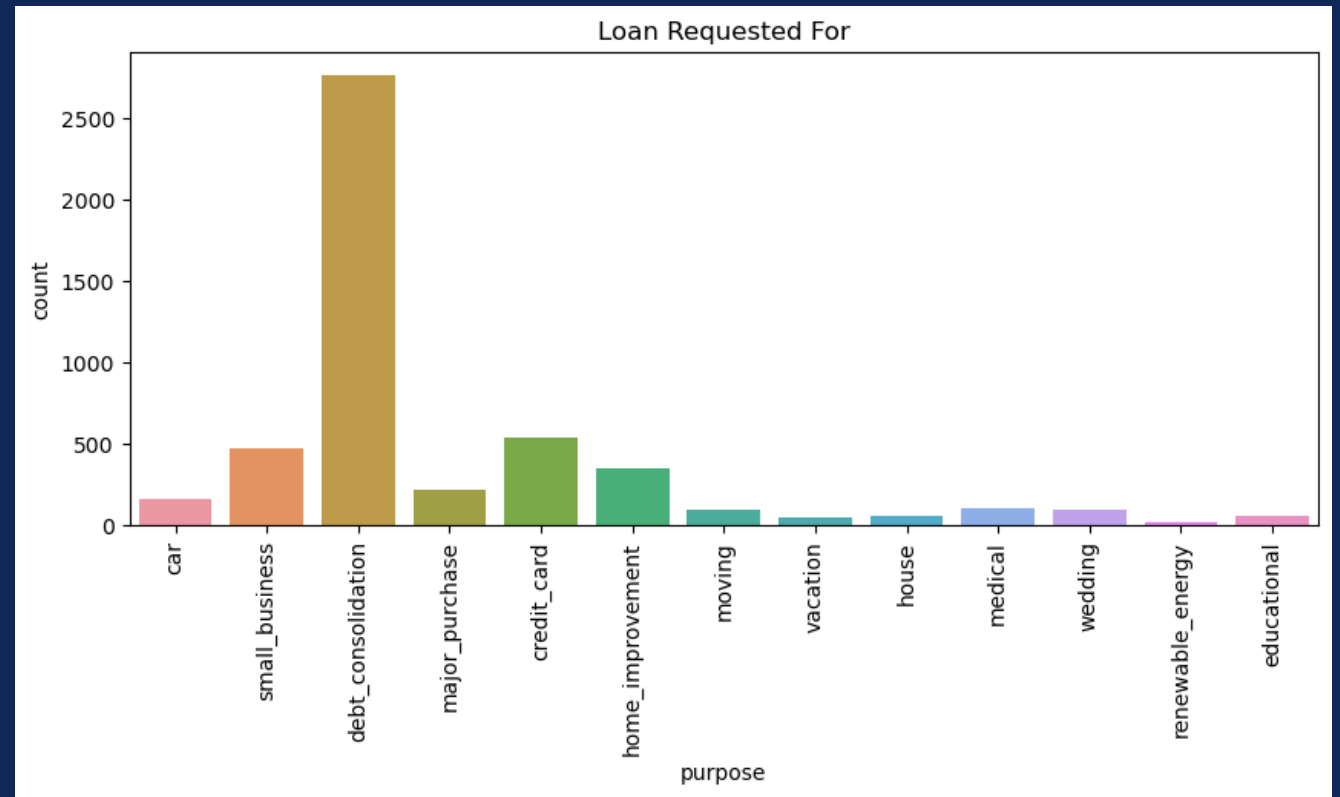
DATA ANALYSIS

Analyzing Charged-Off Loan Cases against Loan Purpose

Observation:

Top 3 **Loan Purpose** which are highly probable for Charged Off use case

- Debt Consolidation
- Credit Card
- Small Business



DATA ANALYSIS

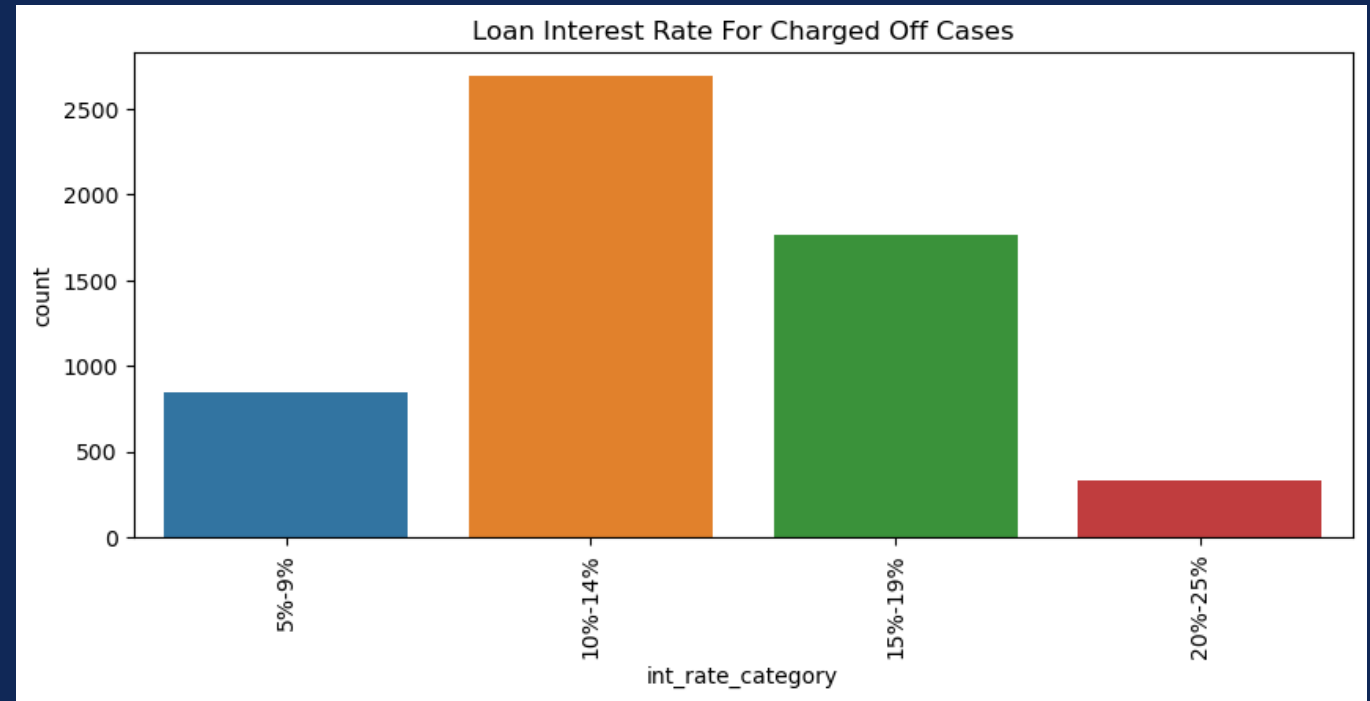
Analyzing Charged-Off Loan Cases against Interest Rate

Observation:

Majority of Charged-Off cases were falling in the following range of Interest Rates:

- 10% - 14%
- 15% - 19%

High Interest rates leads to high probability of a loan to be charged-off



DATA ANALYSIS

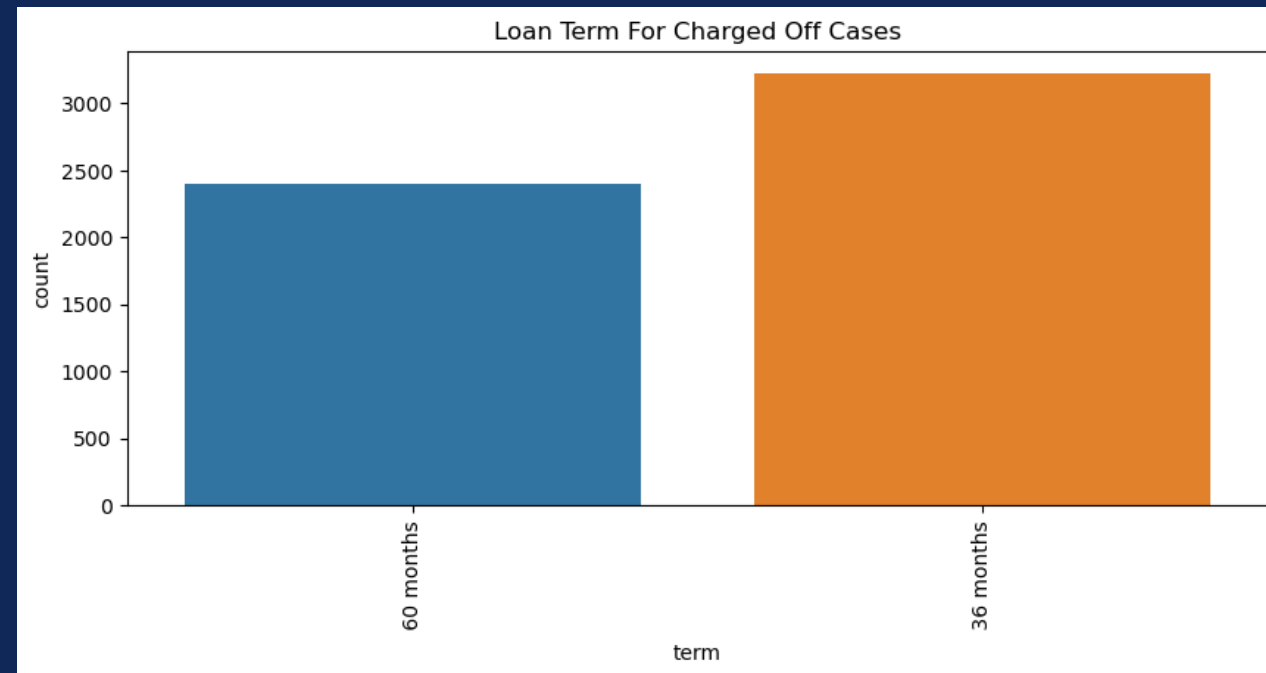
Analyzing Charged-Off Loan Cases against Loan Term/Tenure

Observation:

A strange behavior observed here

- 36 Months Term Loan is having more chances to be Charged-off cases than 60 Months Term Loan

It could be because shorter tenure loan have High EMI value which leads to Charged-off cases.



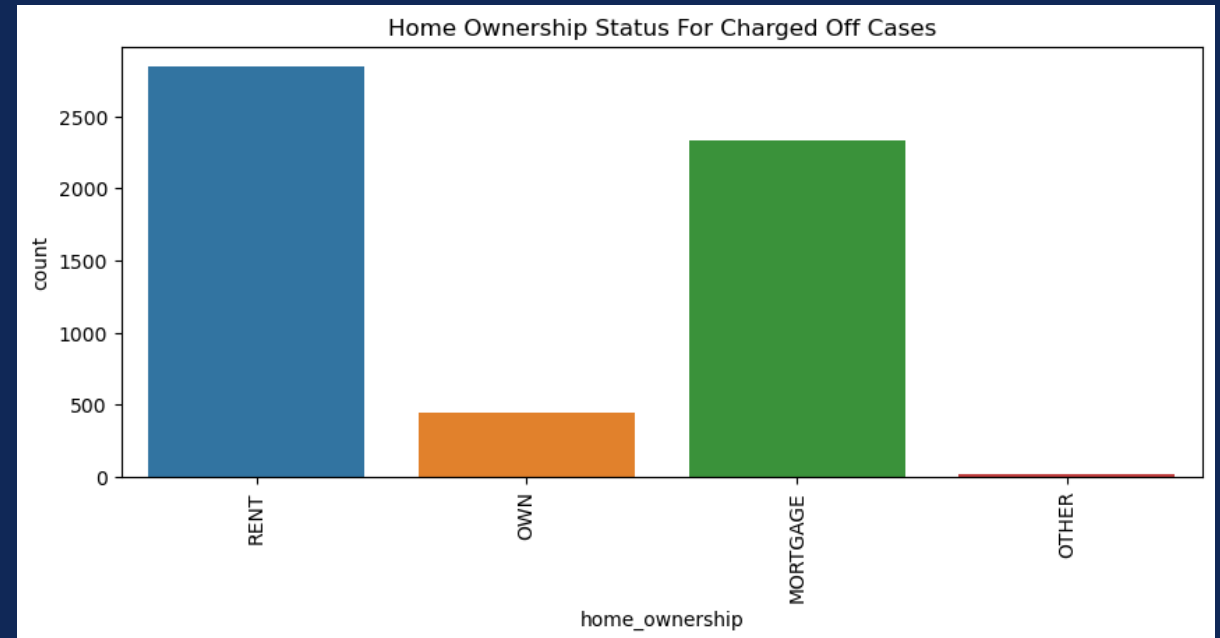
DATA ANALYSIS

Analyzing Charged-Off Loan Cases against Loan Applicant's Home Ownership

Observation:

- Loan Applicant which own the houses are least prone to be Charged-Off
- Loan Applicant's who are having 'Rented House' or 'Mortgage' are more prone to be Charged-off cases.

It could be because people have to pay RENT and Loan EMI both.

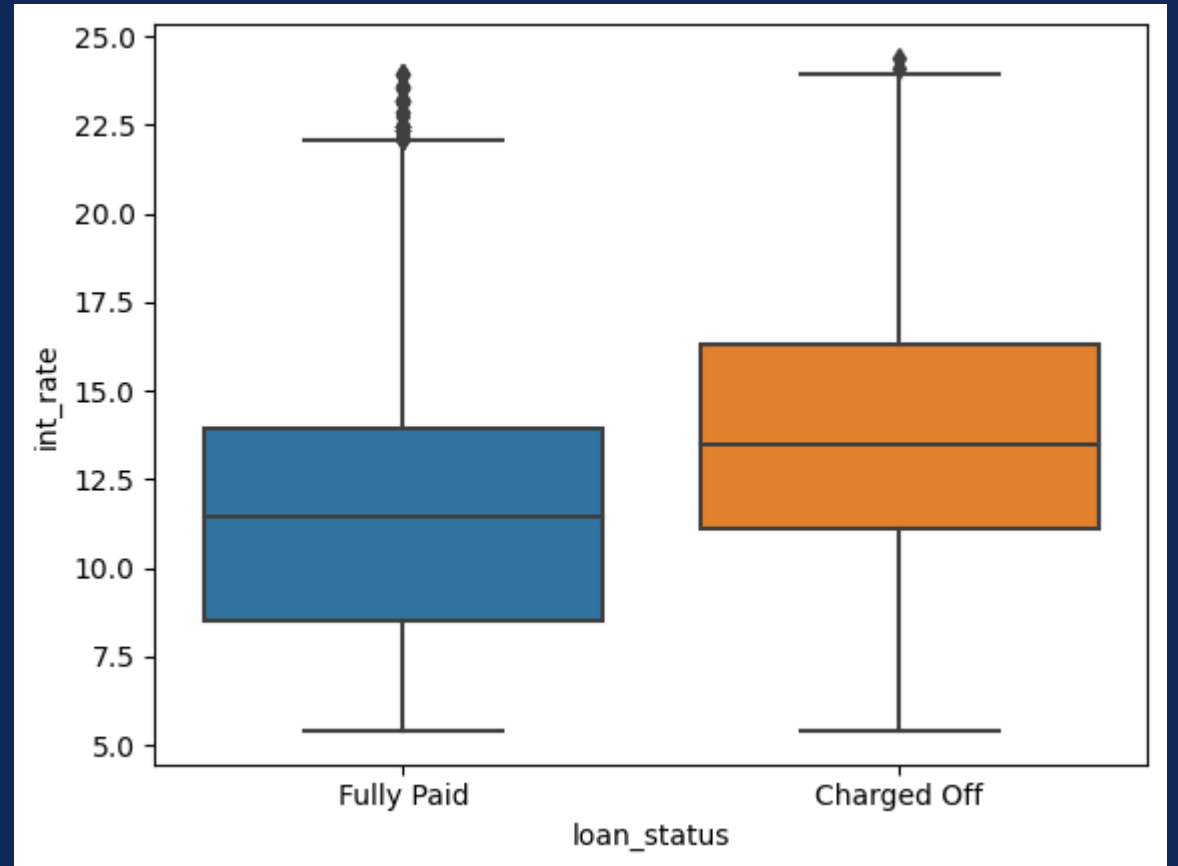


DATA ANALYSIS

Loan Attribute: Loan Status vs Loan Interest Rates

Observation:

- Charged-Off Loan cases belongs to cases where average Loan Interest rate is more than 12%.

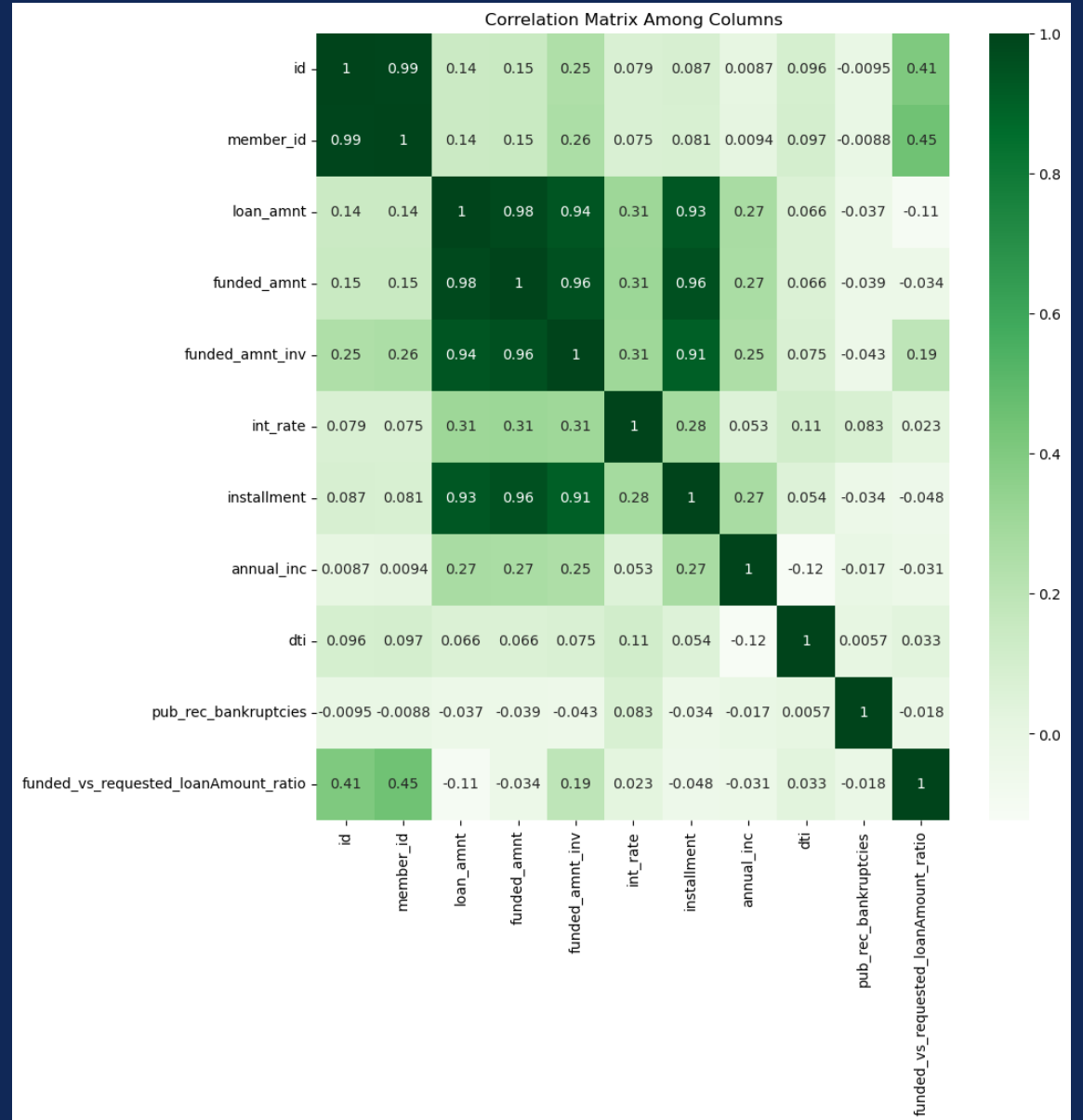


DATA ANALYSIS

Loan Attribute: All Numeric Columns
Type : Checking Correlation

Observation:

- A High Correlation between Loan Requested & Loan Approved (Funded) by bank
- A High Correlation between Installment & Loan Approved (Funded) by bank



CONCLUSION

Type of Analysis	Observation
Loan Status Analysis	<ul style="list-style-type: none">• Approx 85-86% Loan Applicant has Fully Paid their loan.• Approx 14-15% Loan has been charged-off
Loan Status (Charged-Off) vs Annual Income of Applicant	<ul style="list-style-type: none">• Average Annual Income of Charged-Off loan Applicant's is less than Fully-Paid Loan Applicants.• Average Annual Income for Charged-off Loan Applicant's is below 50K INR
Loan Status (Charged-Off) vs Loan Purpose	Top 3 Loan Purpose which are highly probable for Charged Off use case are <ul style="list-style-type: none">• Debt Consolidation• Credit Card• Small Business
Loan Status (Charged-Off) vs Loan Interest Rate	Majority of Charged-Off cases were falling in the following range of Interest Rates: <ul style="list-style-type: none">• 10% - 14%• 15% - 19%
Loan Status (Charged-Off) vs Loan Term/Tenure	36 Months Term Loan is having more chances to be Charged-off cases than 60 Months Term Loan
Loan Status (Charged-Off) vs Home Ownership of Loan Applicant	<ul style="list-style-type: none">• Loan Applicant which own the houses are least prone to be Charged-Off• Loan Applicant's who are having 'Rented House' or 'Mortgage' are more prone to be Charged-off cases.