Lending Club Case Study
Exploratory Data Analysis (EDA) for a Financial Company

Table of Content

Introduction

Lending Club is a peer-to-peer lending platform that connects borrowers with investors. As a financial company,

it is essential to thoroughly analyze the data to make informed decisions about loan approvals and investments.

In this case study, we will perform Exploratory Data Analysis (EDA) on Lending Club's dataset to gain valuable insights.

Objective

The objective of this analysis is to understand the patterns, trends, and characteristics of borrowers and loans,

which will help in making data-driven decisions.

Data Overview

- The dataset contains information about loans, including borrower details, loan amount, interest rates, loan term, etc.
- It includes both approved and rejected loan applications.

EDA Steps

1. Data Cleaning and Preprocessing

- Handle missing values: Identify and address missing values in the dataset. This ensures that the data is clean and ready for analysis.
- Data types: Check and correct data types to ensure consistency and accuracy.
- We have identified and removed values having more than 60% of null values.
- Identified 35 columns for our analysis. Identified outliers on annual income and removed.
- Added derived columns loan issue year, month, annual income bucket, interest rate group, loan amount group.
- We have filled emp interest rate and revolution utility columns with medium. Public record bankcorospy we filled with 0

2. Descriptive Statistics

- Calculate summary statistics like mean, median, mode, range, and standard deviation for relevant variables (e.g., loan amount, interest rates).

3. Univariate Analysis

a. Loan Amount Distribution

- Visualize the distribution of loan amounts to understand the range of loans requested by borrowers.

b. Interest Rates

- Analyze the distribution of interest rates to identify common rates and outliers.

c. Loan Term

- Examine the distribution of loan terms (e.g., 36 months, 60 months) to understand the most popular loan duration.

d. Purpose of Loan

- Explore the purposes for which borrowers are seeking loans (e.g., debt consolidation, home improvement, etc.).
-

4. Bivariate Analysis

a. Loan Amount vs. Interest Rates

- Analyze the relationship between loan amounts and corresponding interest rates. This can provide insights into risk assessment.

b. Loan Status vs. Purpose

- Investigate how the purpose of the loan relates to loan status (e.g., fully paid, charged off, etc.).

5. Time Series Analysis (if applicable)

- If the dataset includes a time component (e.g., application dates, approval dates), analyze trends over time.

6. Risk Assessment

- Identify factors that may contribute to loan default or higher interest rates. This could include variables like credit score, employment length, and de income ratio.

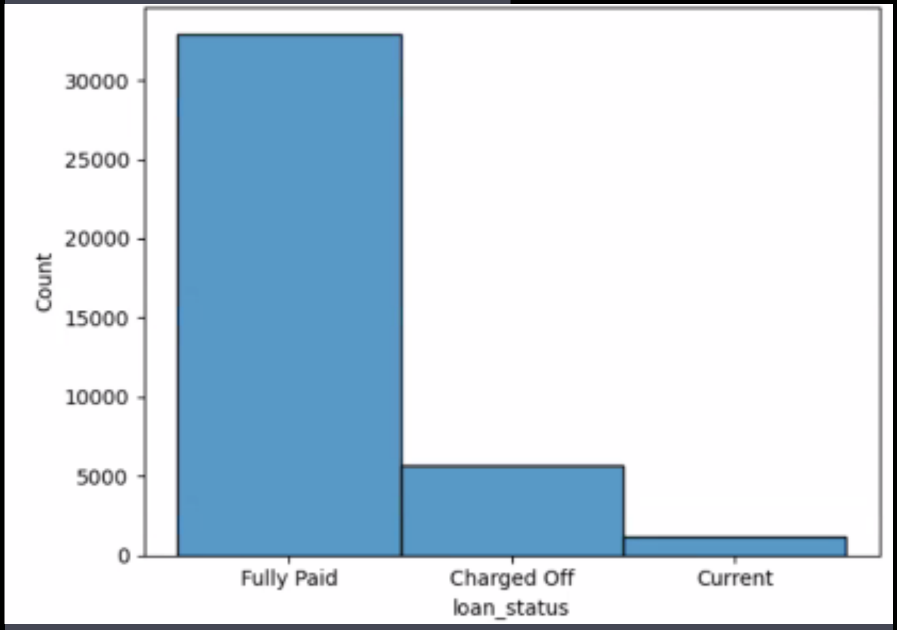
7. Data Visualization and Reporting

- Use visualizations (e.g., histograms, bar charts, scatter plots) to present key findings.
- Summarize insights and recommendations for decision-making.

Insights - Loan Status

Univariate Analysis:

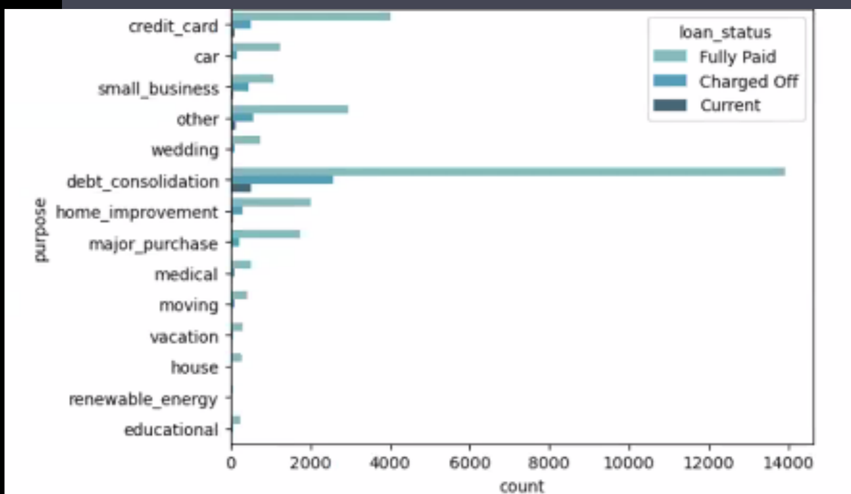
* Observed 83% loan in fully paid status
* Observed 14% loan in charged off status
* Observed ~3% loan in current status

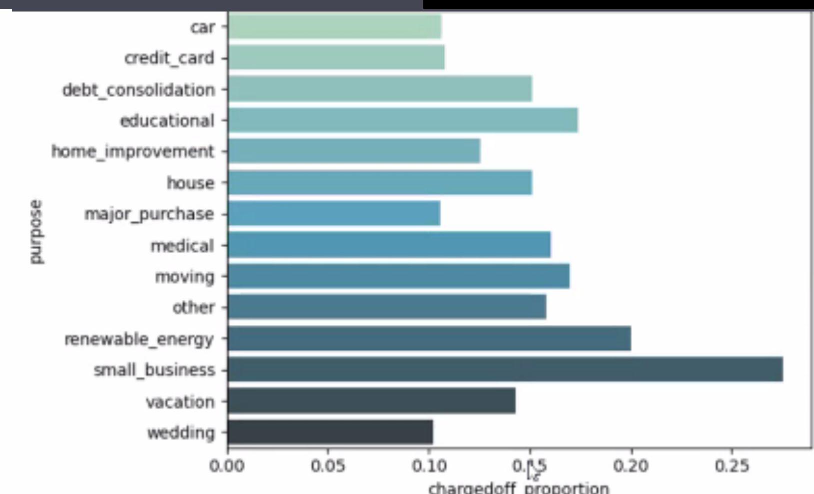


Insights - Loan Purpose

Univariate/Bivariate Analysis:

* Observed more number of applicants takes loan for debt consolidation, credit card and other reasons.
* Small Business has got more number of Charge off percentage as compared to other loans.

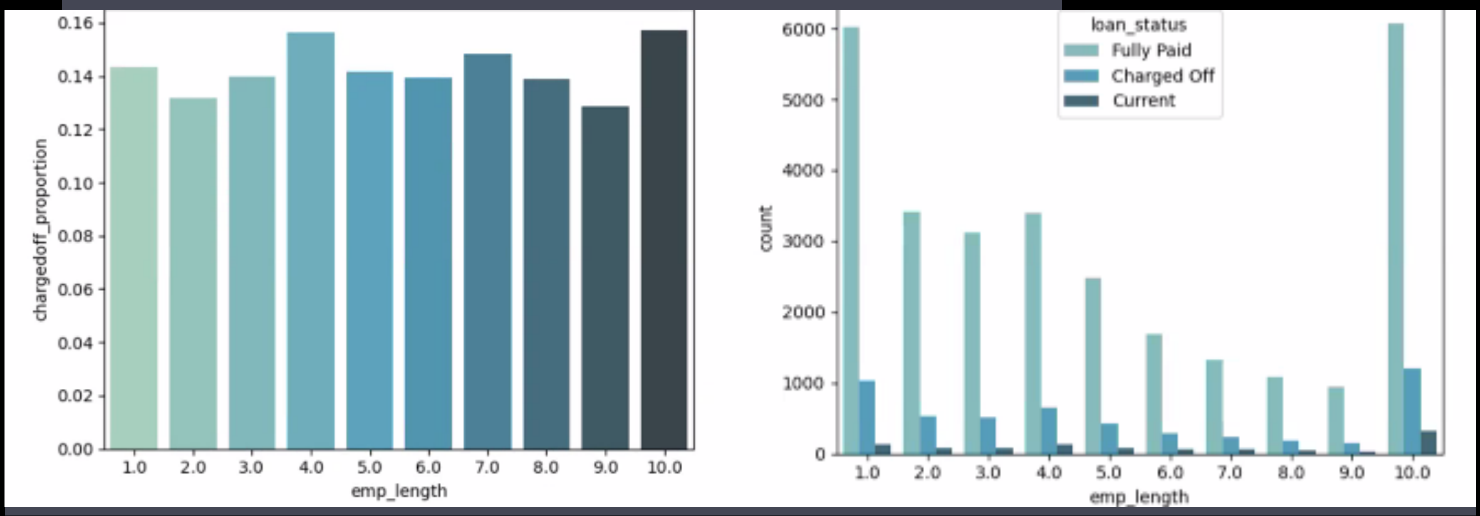




Insights - Employee Experience

Univariate/Bivariate Analysis:

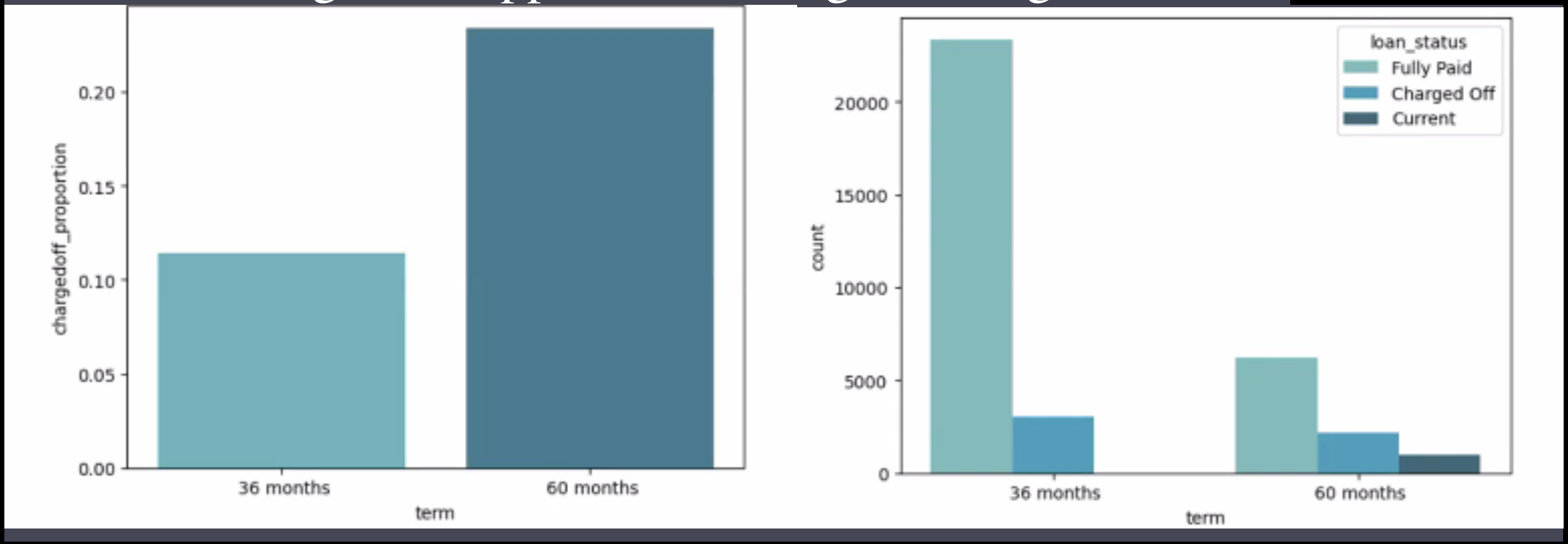
* Observed employee experience is not making much difference on finding chance of defaulter's
* Observed charged off proportion is similar for all the experience categories.



Insights - Loan Term Length

Univariate/Bivariate Analysis:

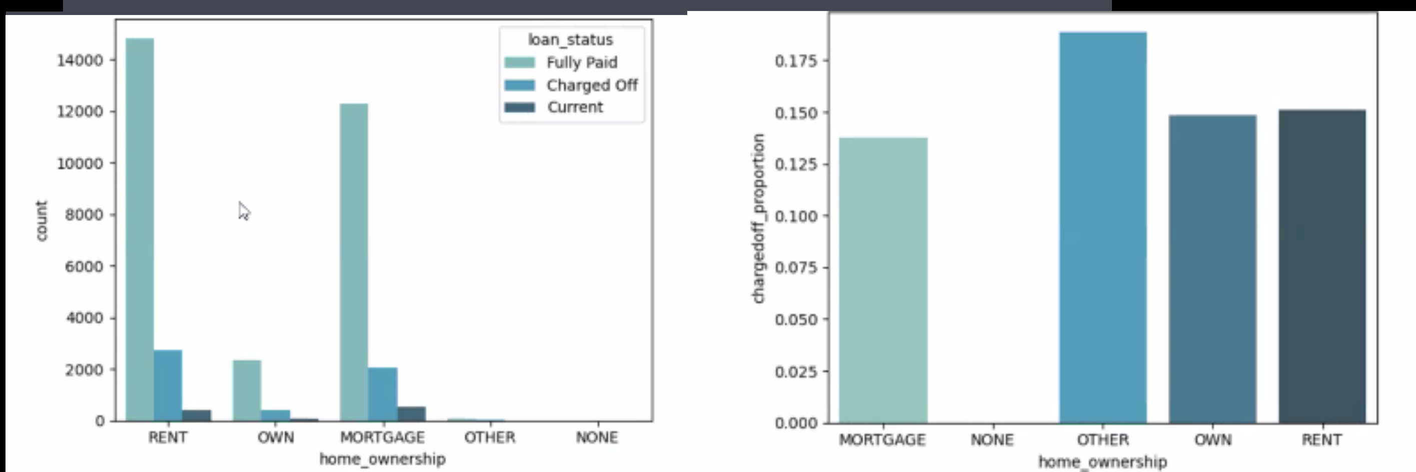
* Observed charged off applicants are high on long term loans



Insights - Home Ownership

Bivariate Analysis:

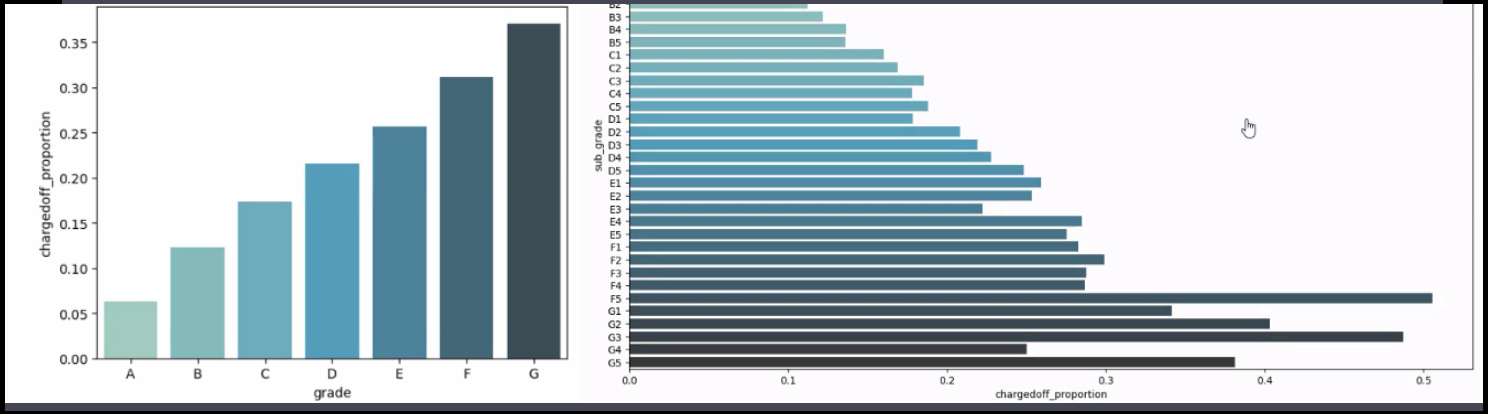
* Observed high number of applicants are people who are currently staying on rent or mortgaged their property
* Observed Charged off possibilities are almost similar for all home ownerships.
* Other category applicants has slightly higher possibility to become charged off.



Insights - Applicants Grade,Subgrade

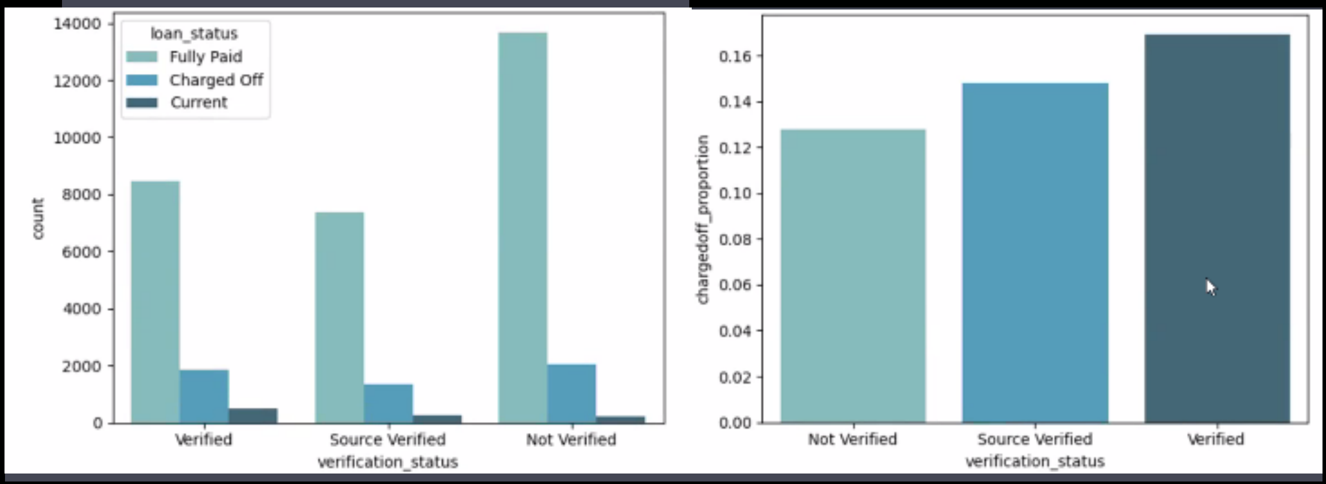
Bivariate Analysis:

- Observed Charged off applicant counts are increasing from Grade A to G
- Observed sub grade also shows the same trend of grade.
- Charged off applicants percentage is higher in Sub grades of G and F categories compare to A, B grades.



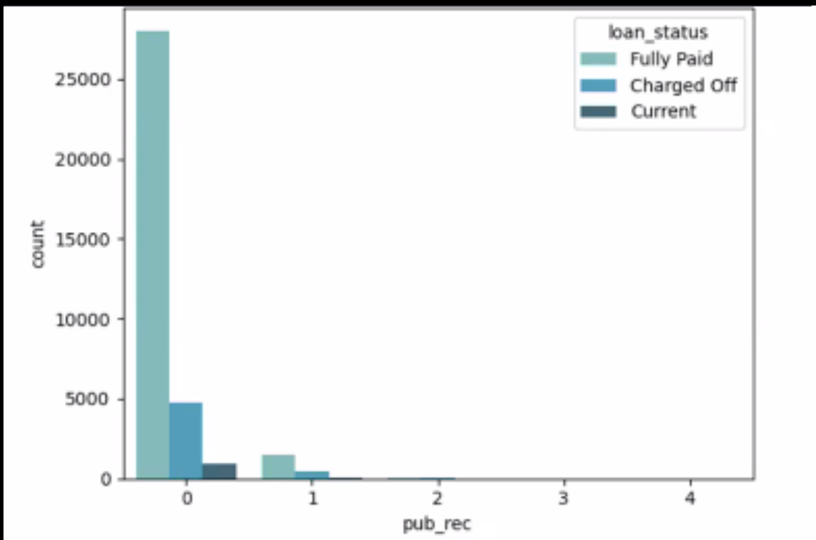
Insights -Verification Status
Univariate/Bivariate Analysis:

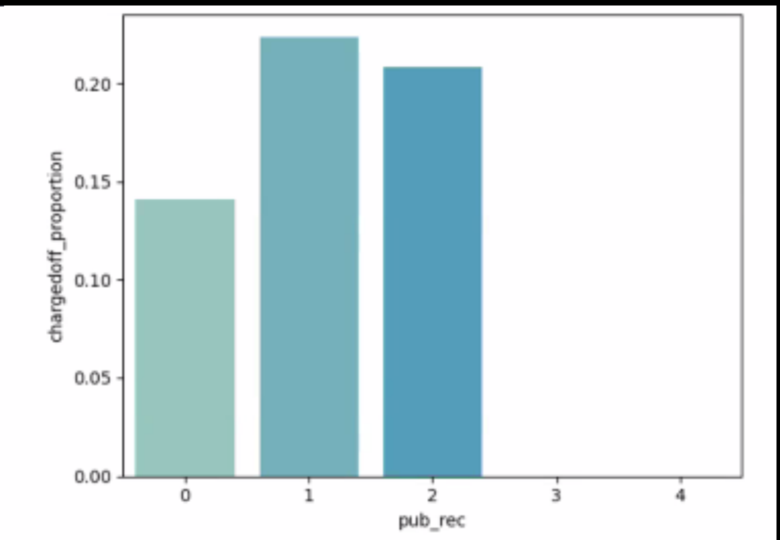
* Observed charged off applicants are avilable on all verficication status.
* Charge off is independent of verification status.



Insights - Derogatory Public Record Univariate/Bivariate Analysis:

* Observed clearly that if there is one or more derogatory public record there more chances of being defaulter.

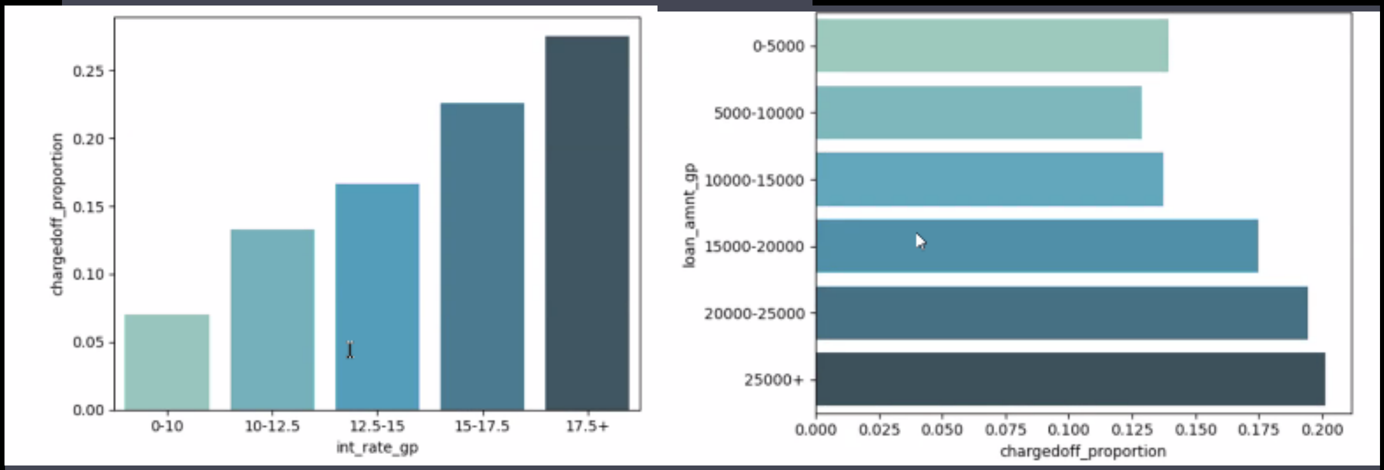




Insights - Int Rate/Loan Amount

Bivariate Analysis:

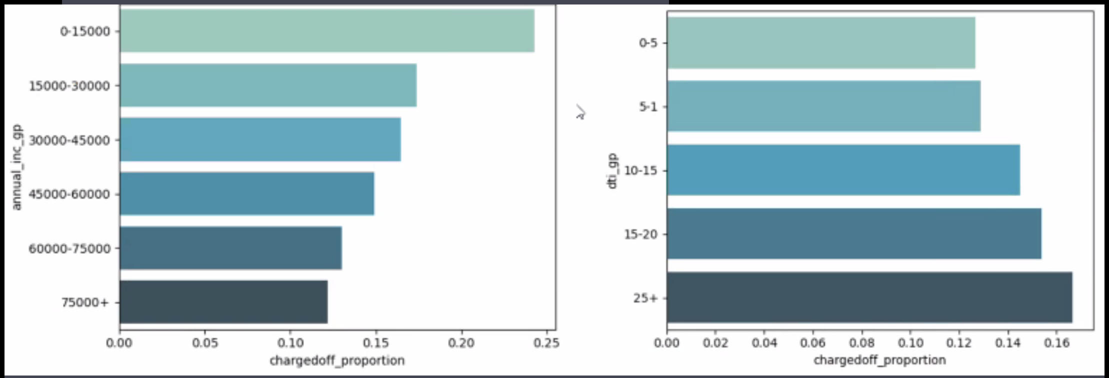
- Observed percentage of charged off applicants increases while interest rate increases.
- Observed defaulters will be more on higher interest rate.
- Higher the loan amount possibility of charge off is more.



Insights - Annual Income/monthly Debit Ration

Bivariate Analysis:

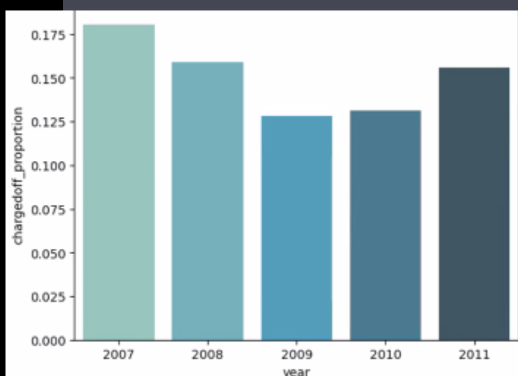
- Observed charged off applicants percentage increases while annual income decreases.
- Possibility of charged off applicants are high around 15000 annual income applicants.
- Observed Charged off applicants are getting increases while monthly debt payment ratio increses.
- Annual Income has outliers so we removed it.

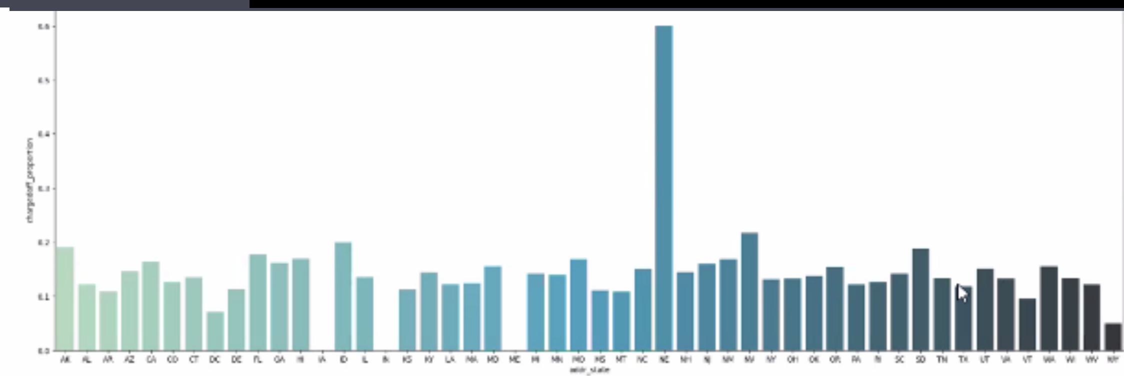


Insights - Years and State

Bivariate Analysis:

* Observed charged off proportions has not significant difference between 2007 to 2011
* Observed charged off applicants are high on state ‘NE'

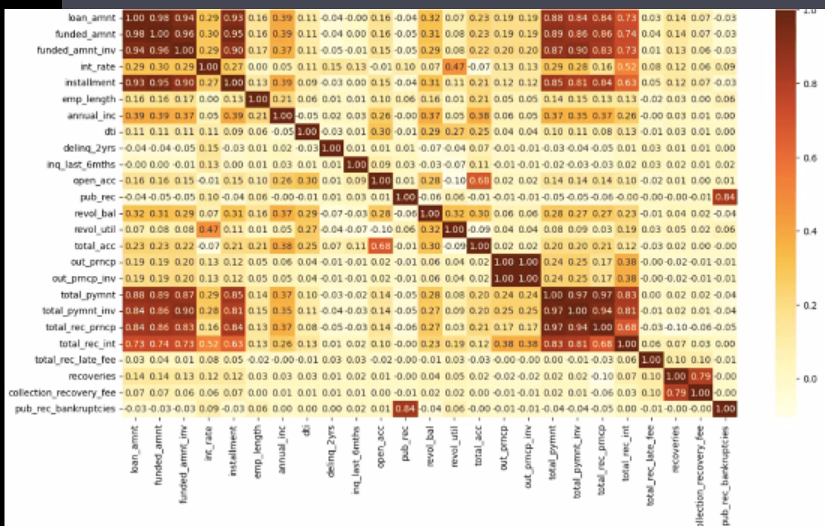




Insights - Other Factors

Bivariate Analysis:

* Observed loan amount, funded amount and funded investor amount are having close corelations and looking similar.
* Same applicable for payment recieved attributes (total_pymnt, total_pymnt_inv. total_rec_prncp)
* We are not analyzing the other factors due to close correlations.



Conclusion

Through this EDA, we have gained valuable insights into the lending patterns and borrower characteristics. These insights can inform loan approval st
risk assessment, and investment decisions for Lending Club.

* We have identified when there is higher rate of interest and loan amount is high , they're more number of defaulters.
* If annual income is less the chances of defaulter is higher.
* When monthly debit ratio increases the defaulter rate increase.
* F&G Grades have more number of defaulters.
* If Loan term is more there are more chances of getting the defaulter.
* Public record is direct indicator for finding defaulters.
* Small Business loan borrower has more chances of being defaulter.
* Borrower from State NE has more defaulters so need to be causation while lending them money.
* Other than above individuals we should focus on applicants other than above mention categories so that we can get good business.