



Lending Club Case Study

- An Exploratory Data Analysis Approach

Authors:

Ashutosh Kumar Sahu

Mohammed Saqib Rahman



Objective

The objective of this case study is to implement EDA technique on a real-world problem and understand the insights and present in a business first manner via presentation.

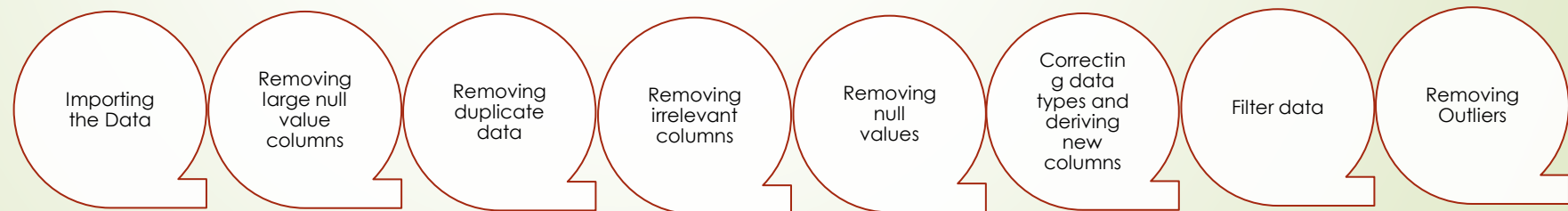
Benefits of the case study:

- The goal is to identify patterns in loan applicants that indicate the likelihood of defaulting on loans.
- Gives idea about how EDA is used in real life business problems.
- Decision-making involves balancing potential business losses from declined loans versus financial losses from approved loans that default.

Business Understanding:

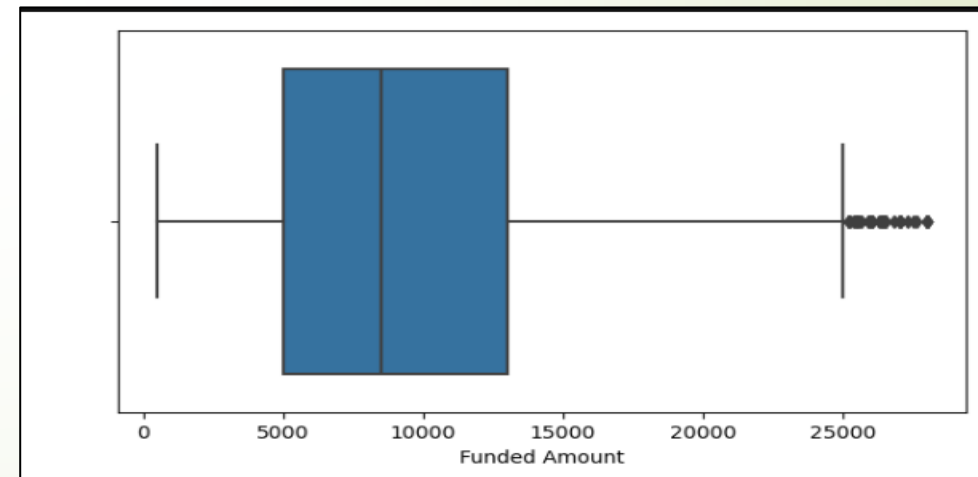
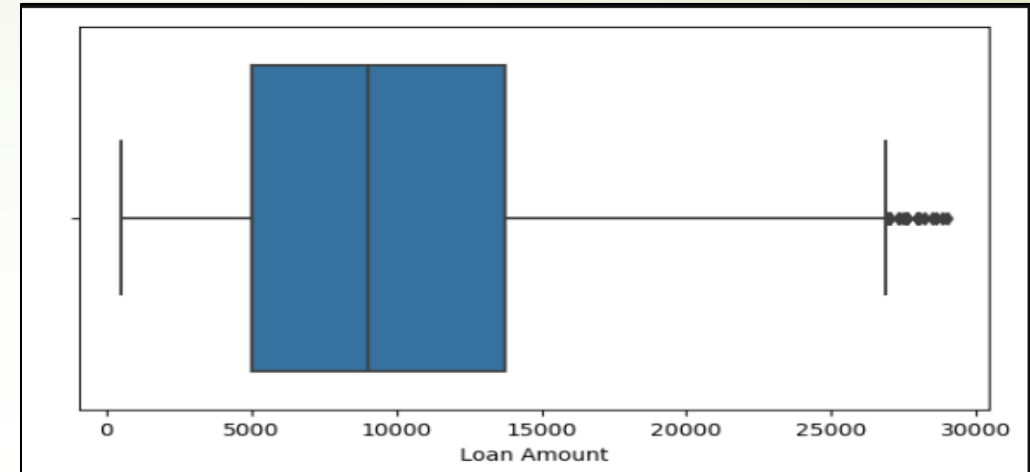
- The business objective is to take a decision whenever they receive a loan application whether to reject or approve based on certain variables
- Dataset Details:
 - The data given below contains information about past loan applicants and whether the 'defaulted' or not.
 - Data has details regarding approved loan not the rejected ones. It has 3 status of loans which is Fully paid, Current and Charged-Off.

Data Clean-up and preparation process:



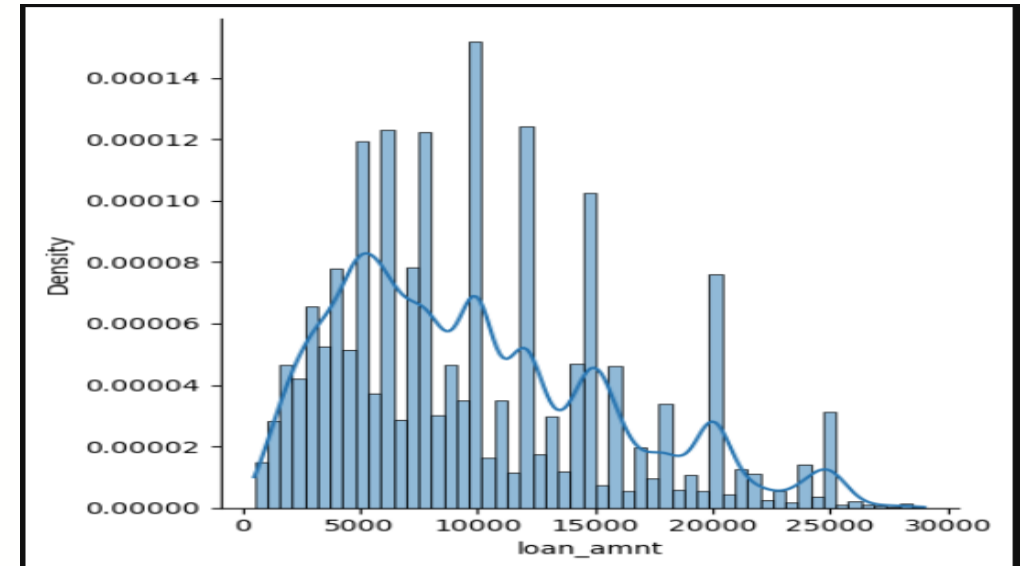
Loan Status and Amount

- **Loan Amount:** It varies from 5000 to 14000 with a median of 8000.
- **Funded Amount:** It varies from 5000 to 14000 with a median of 8000. Funded amount is majorly small and very few clients have taken large loans and larger it goes we have higher chance of defaulting.



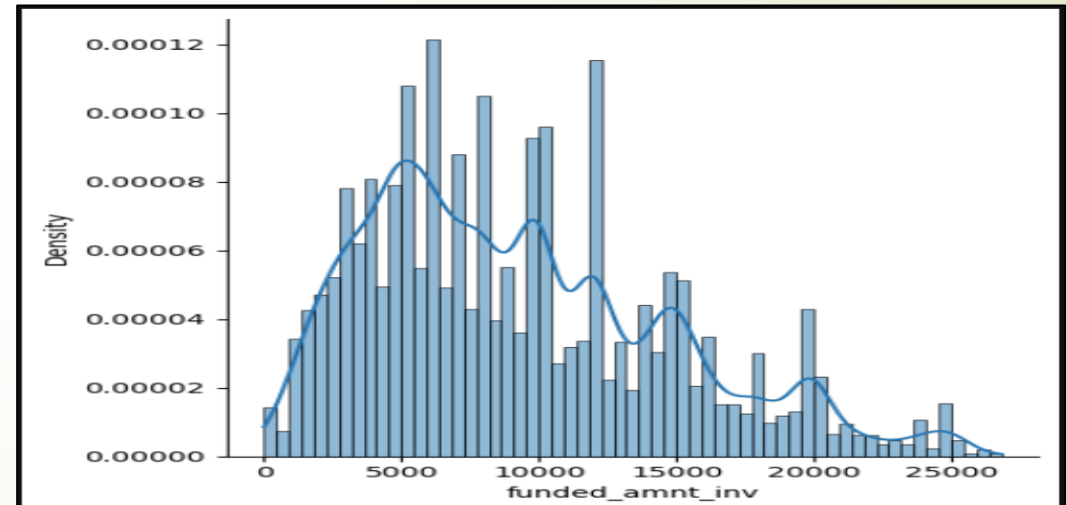
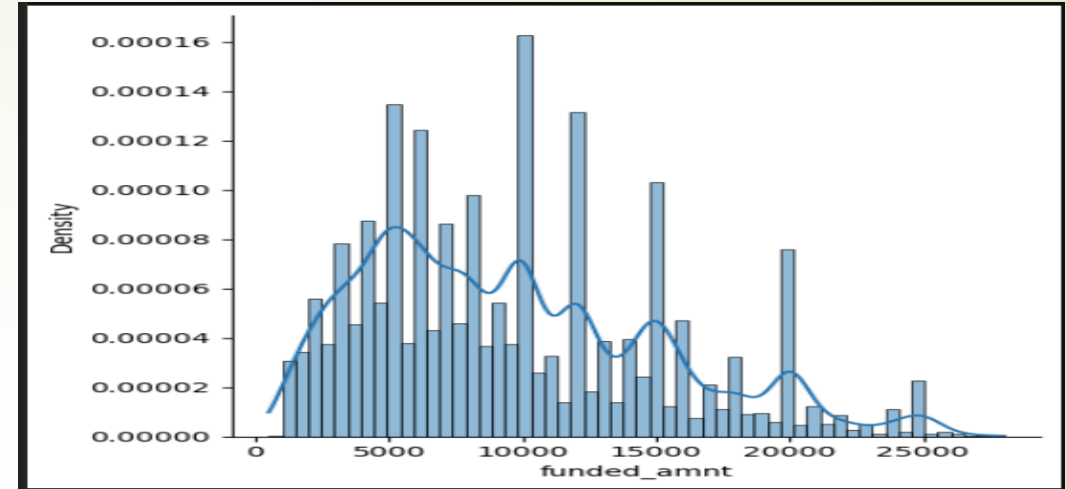
Loan Amount Density

- Most of the loan amount was applied between 5k to 14k
- Max loan amount applied was ~27k



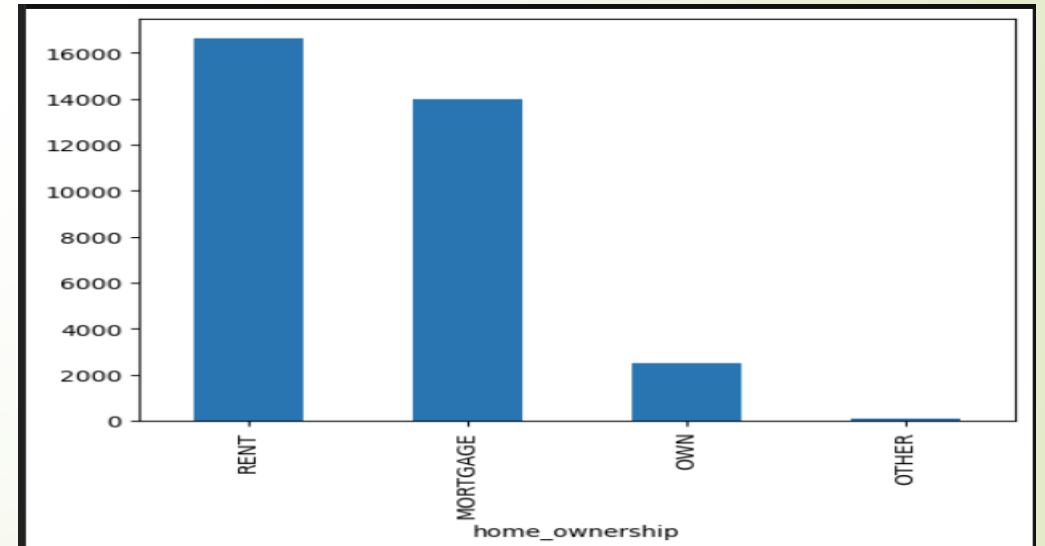
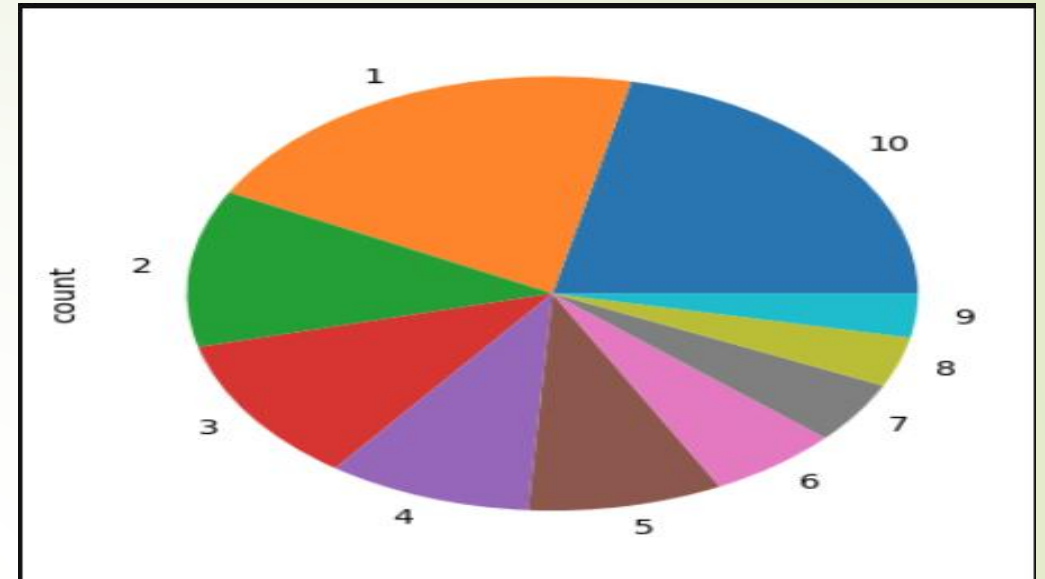
Funded Amount and Funded Amount Invested

- **Funded Amount:**
 - You can see the Funded amount trend is non-linear. The highest is 10000.
- **Funded Amount Invested:**
 - You can see the Funded amount invested is similar Funded amount.



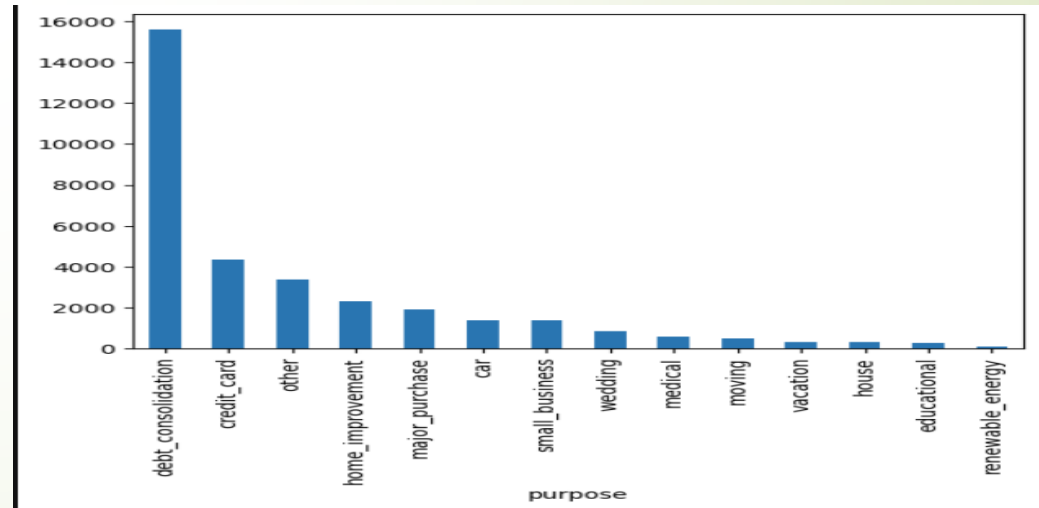
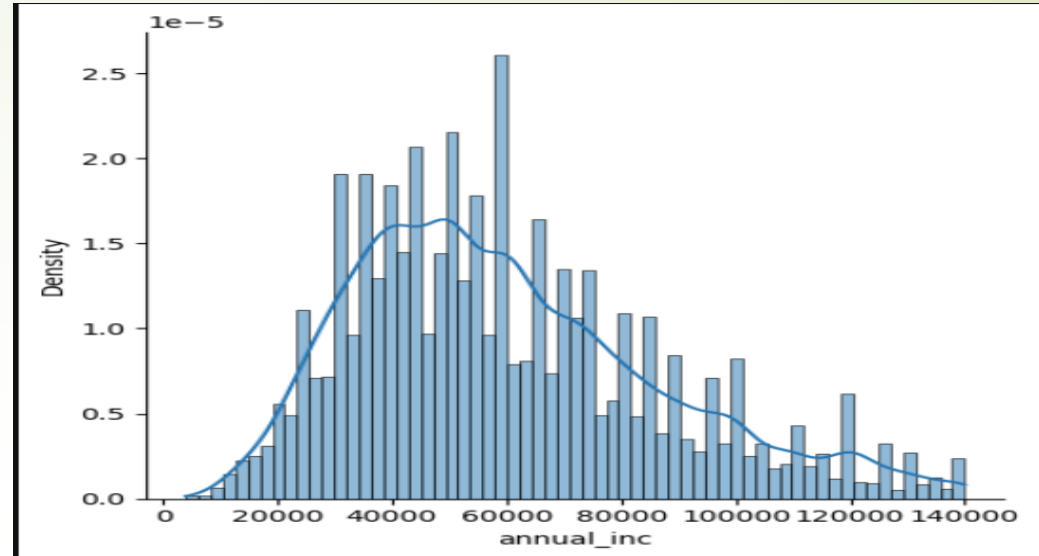
Employment Length and Home ownership

- **Employment Length:** Majority of clients have 10 years of experience.
- **Home Ownership:** Majority of clients are lacking ownership of any property and are on rent or mortgage and have a higher chance of defaulting.



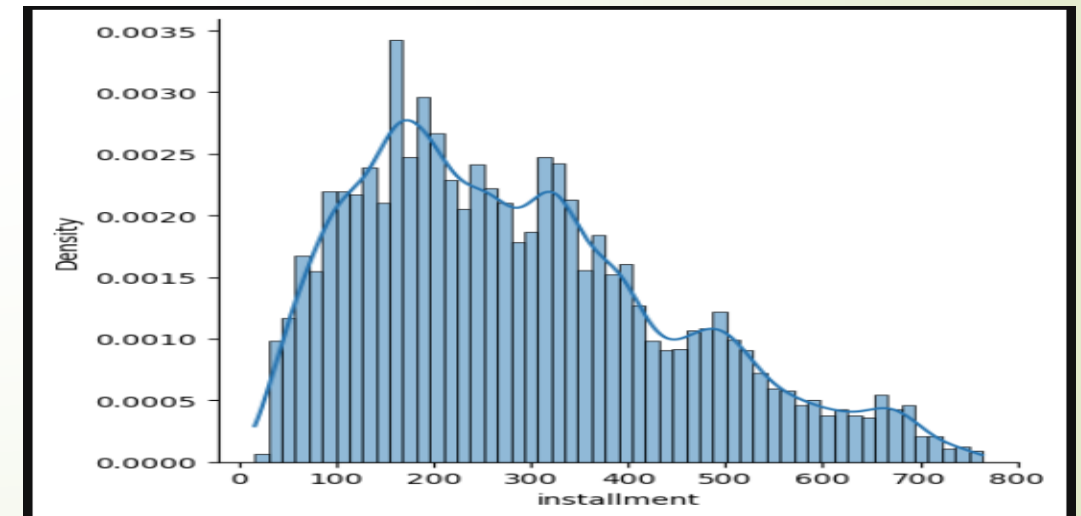
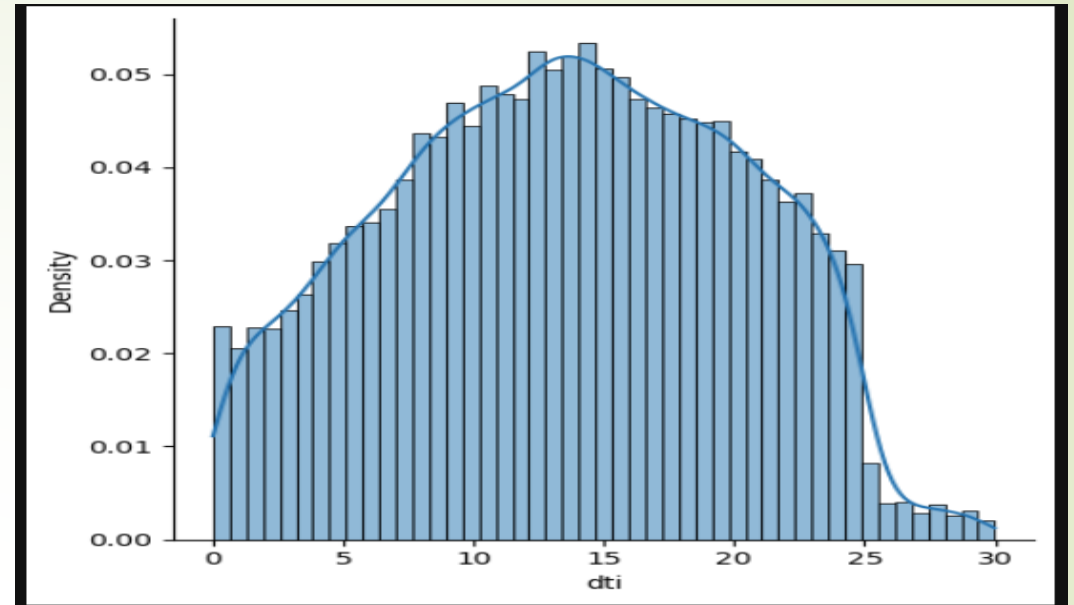
Annual Income & Purpose

- **Annual Income :** The Majority of clients have low annual income compared to rest and income lower than 50k has higher chance of defaulting.
- **Purpose:** Loans are taken mostly for debt consolidation followed by credit card payment. Whereas the debt consolidation has highest fully paid loan but also has highest defaulted loans as well.



DTI ratio & Installments

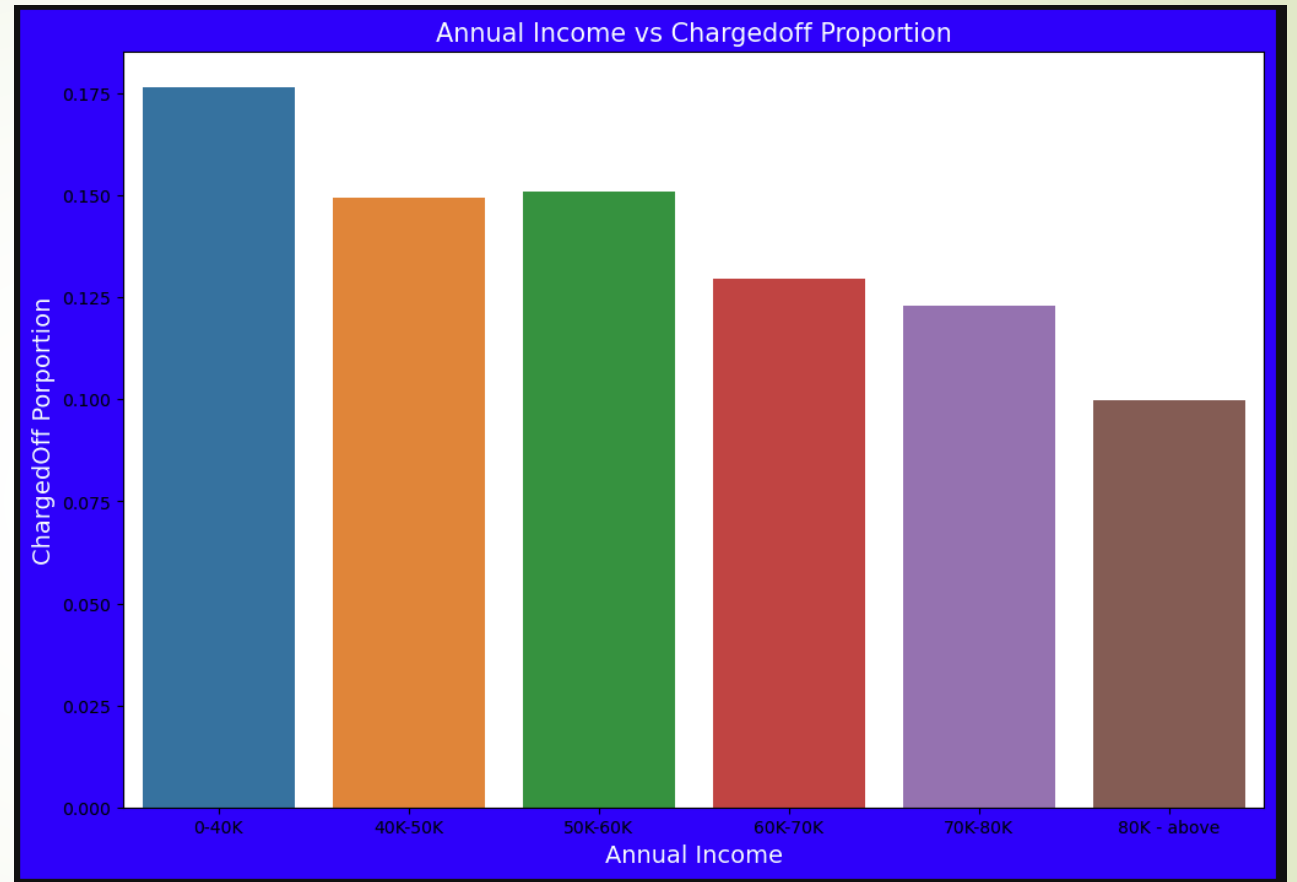
- **DTI:** The large percentage of Clients have a large Debt to Income ratio which shows that lending to such clients can be very risky.
- **Installments:** Majority of clients have 260 months of instalment period.



Annual Income and Charged-off Proportion

- **Observation:-**

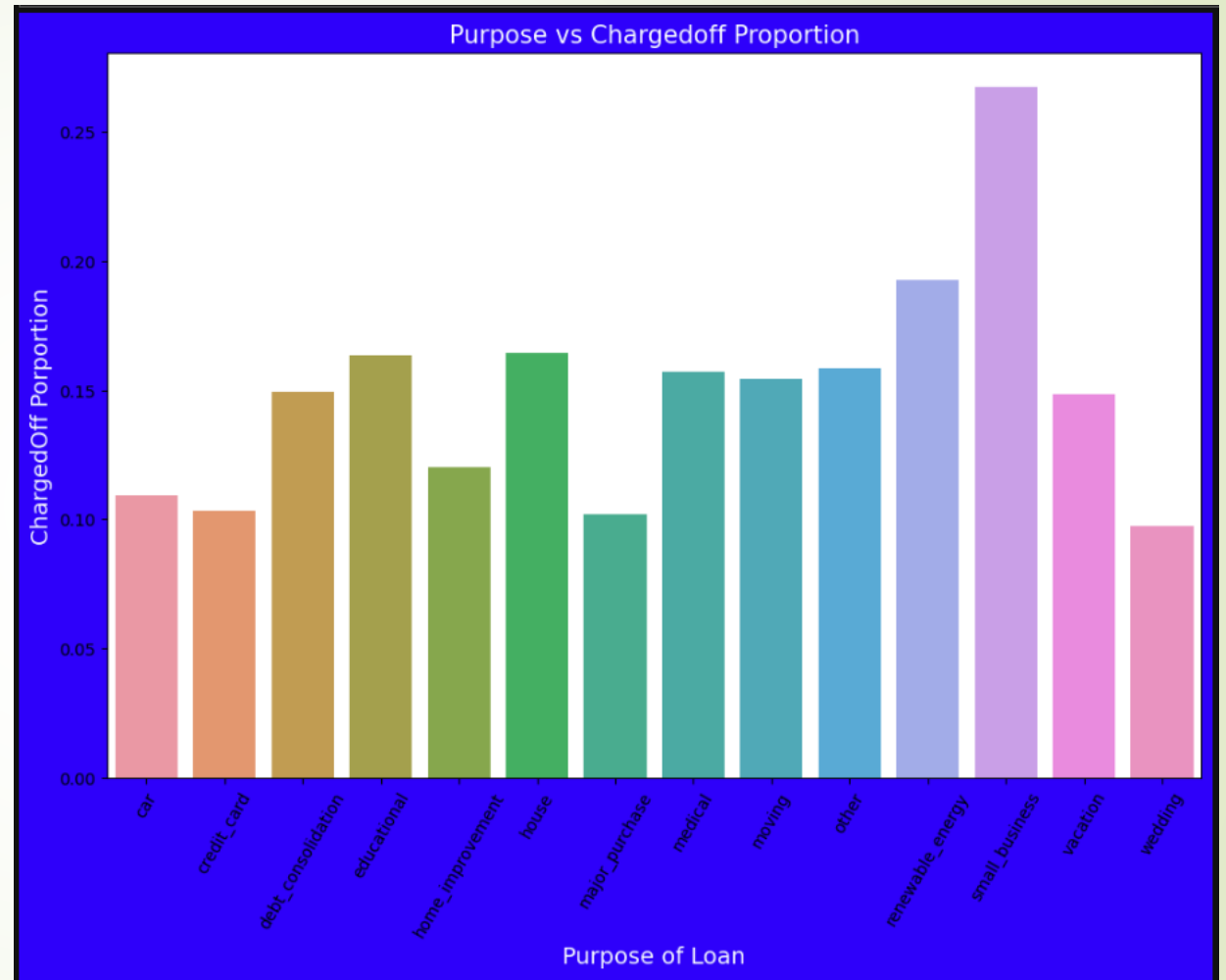
- Income range 8000+ has less chances of charged off
- Income range 0-2000 has high chances of charged off
- With increase of annual income, charged off proportion got decreased.



Purpose vs Chargedoff Proportion

■ Observation:-

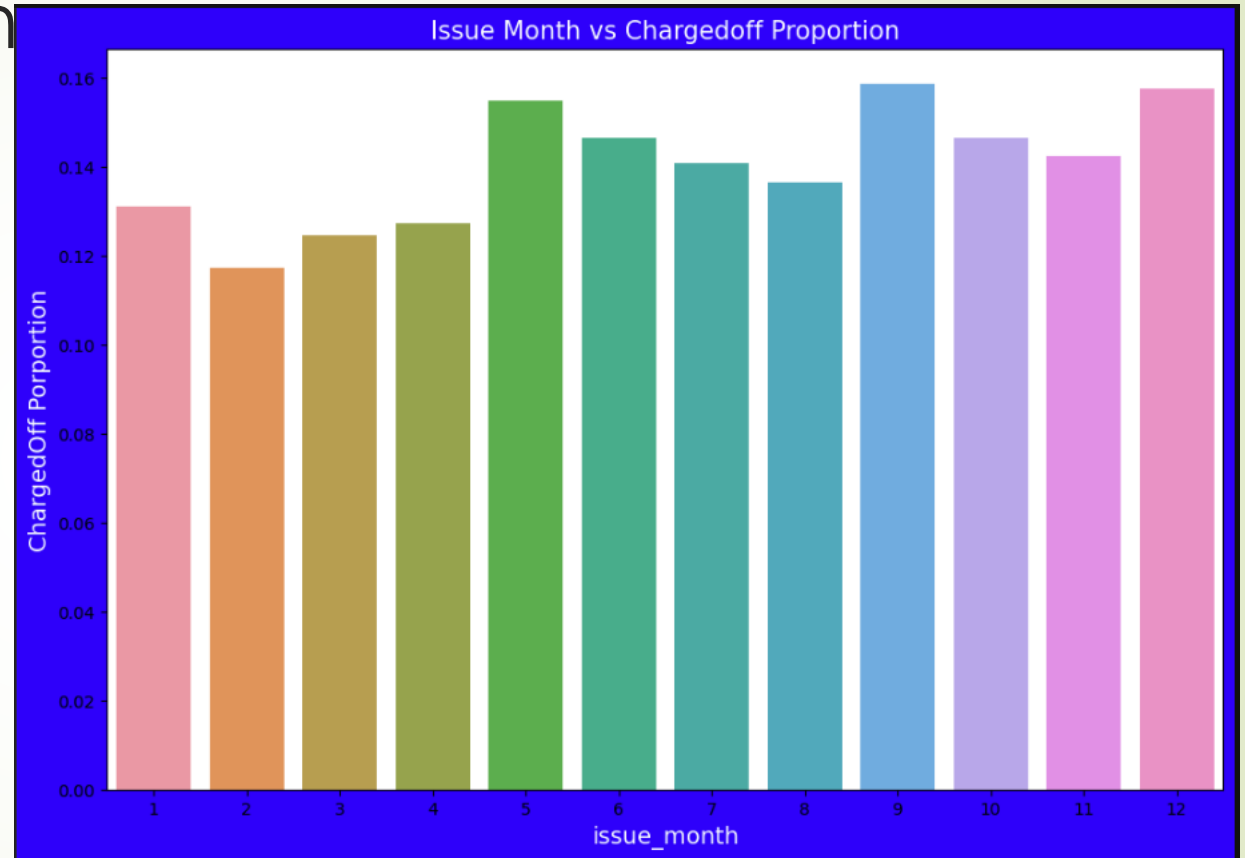
- The person having home loan is having low chances of defaults.
- The Person having loan for small business are having high chances of loan defaults



Issue Month vs Chargedoff Proportion

- **Observation:-**

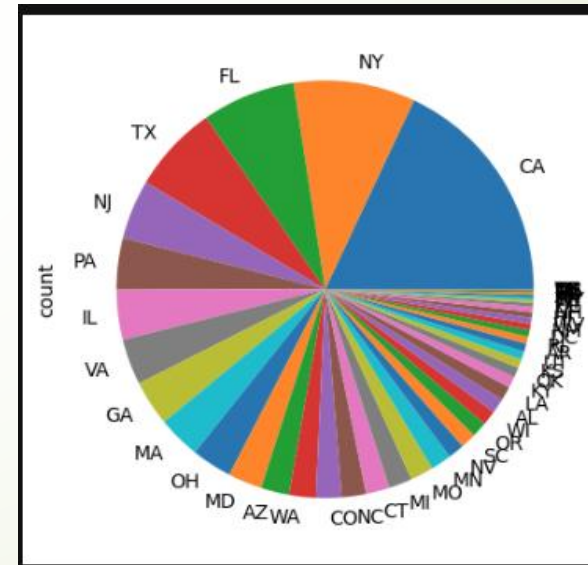
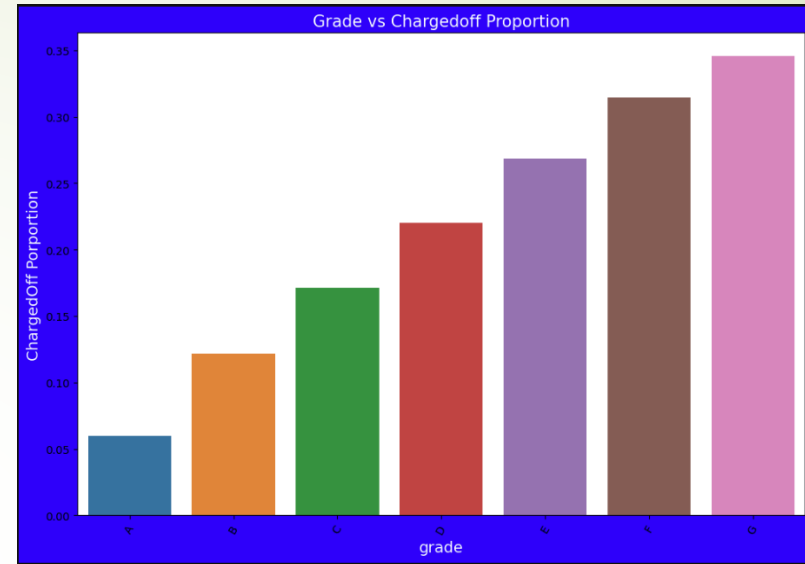
- Loan issued in the month of May; September & December is having high no of loan defaults.
- Majority of loan defaults are coming from the month of September to December.



Grade and State

- **Observation:-**

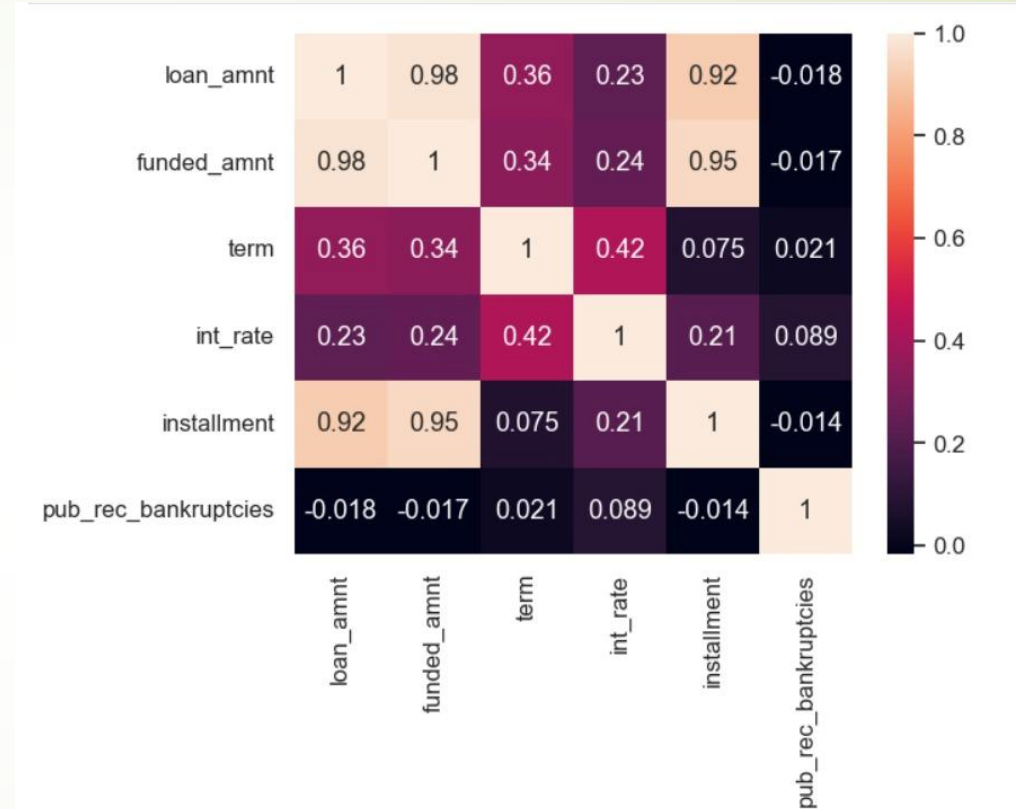
- Grade G is having high loan defaults.
- Grade A is having lowest loan defaults
- Most of the applications are from State "CA"



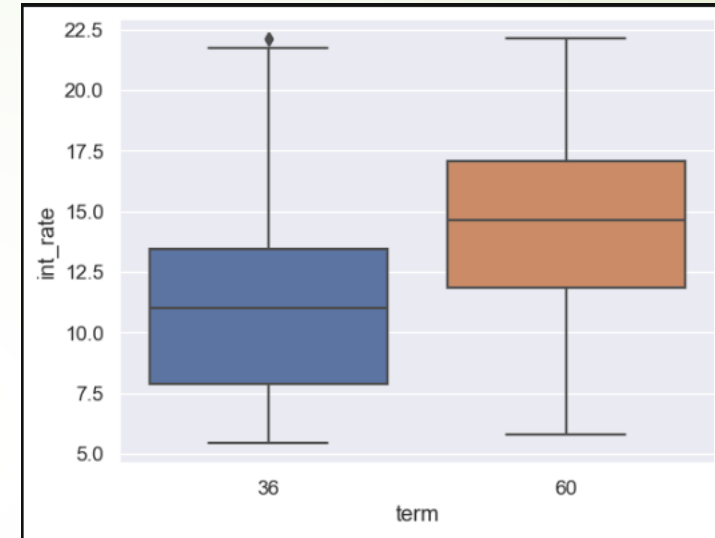
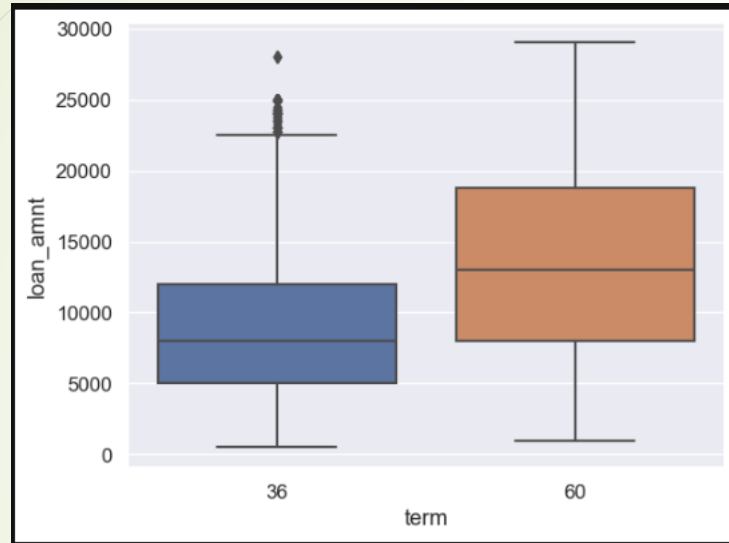
Heatmap Correlation:

- **Observation:-**

- Negative Correlation
 - Loan_amnt has negative correlation with pub_rec_bankruptcies
- Strong correlation
 - term and loan amount are in strong co relation
 - term and intrest rate are in strong co relation



Location Based



- As per the above graph on left, with in increase of term loan amount increases.
- As per the above graph on right, with in increase of term loan interest rate increases.



Recommendations:

Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:

- 1. DTI
- 2. Grades
- 3. Employment length
- 4. Annual income
- 5. Purpose

Other considerations for 'defaults' :

- 1. Borrowers are from large urban cities like California, New York, Florida etc.
- 2. Borrowers having annual income in the range 50000-60000.
- 3. Borrowers having Public Recorded Bankruptcy.
- 4. Borrowers with working experience 10+ years.