LENDING CLUB CASE STUDY

By – Rajesh Suresh Thakur

Problem Statement:

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company

- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Things to Consider for analysis:

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:

    - **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)

    - **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

    - **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

2. **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Analysis Approach

High level data analysis:

1. First looked data at using df.shape, df.info, df.head, df.tail, df.dtypes etc commands to get a high level overview of number of records, number of rows, no. of columns, types of data types used in column. It will help me and provide me a point to start.
2. On running df.shape command I understood there are 39717 Rows and 111 columns present in given dataset.

Data Cleaning: Next, I proceeded to find if any NULL values in present in dataset.

1. First checked my target Variable which is 'loan_status" and it had no NULL values.
2. Second I removed all current running loans as they are not required for analysis.
3. Third I checked all other columns and removed other columns where there are many NULL values as part of data cleaning.

Data Conversions-

1. 'Int_rate' column had % sign so removed it. Likewise 'term' column had Years word and + and > and < signs so removed them.
2. Dropped few columns where all values are same and unique value came as 1, which means all rows have same value. That kind of columns are not useful for analysis.
3. After all cleaning and conversions got (37544, 35) rows and columns finally.

Finding and removing outliers-

1. Run the commands to find out outliers in our data and identified that few columns like ['loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'term', 'int_rate', 'installment'] have few outliers in data.
2. Executed the commands to find IQR and deleted those records which are outliers.

Plotting charts

1. After cleaning of data and removal of all outliers I proceeded with plotting the graphs and drawing some business conclusions.
2. Refer below section for detailed analysis and conclusions.

1) Charged Off Loans Vs Purpose of Loans

Charged Off Loans by Purpose of Loan
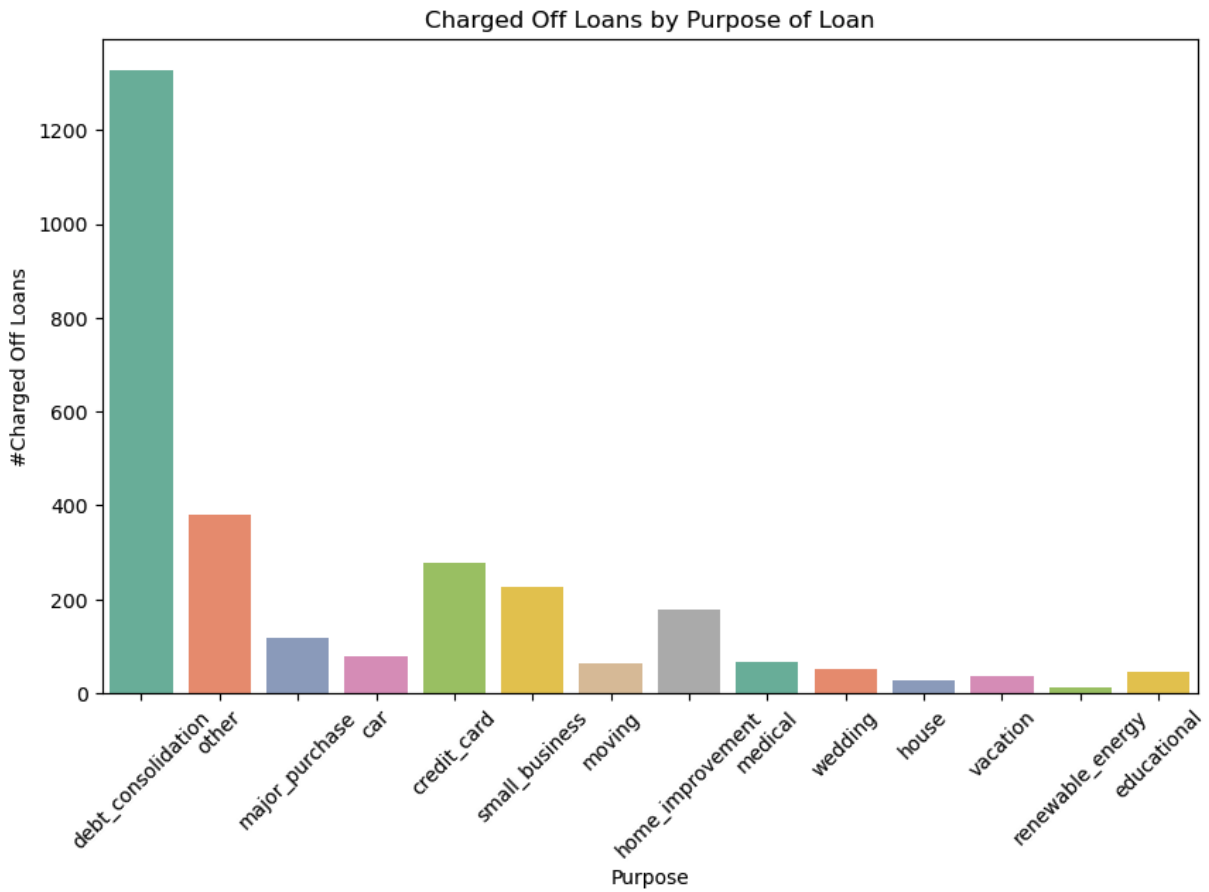
Conclusions

1) By looking at above histogram we can draw a conclusion that people taking loans for debt reconsolidation have highest chances of defaulting the Loans.

2) Charged Off Loans Vs Annual Income Range

Charged Off Loans by Annual Income Range
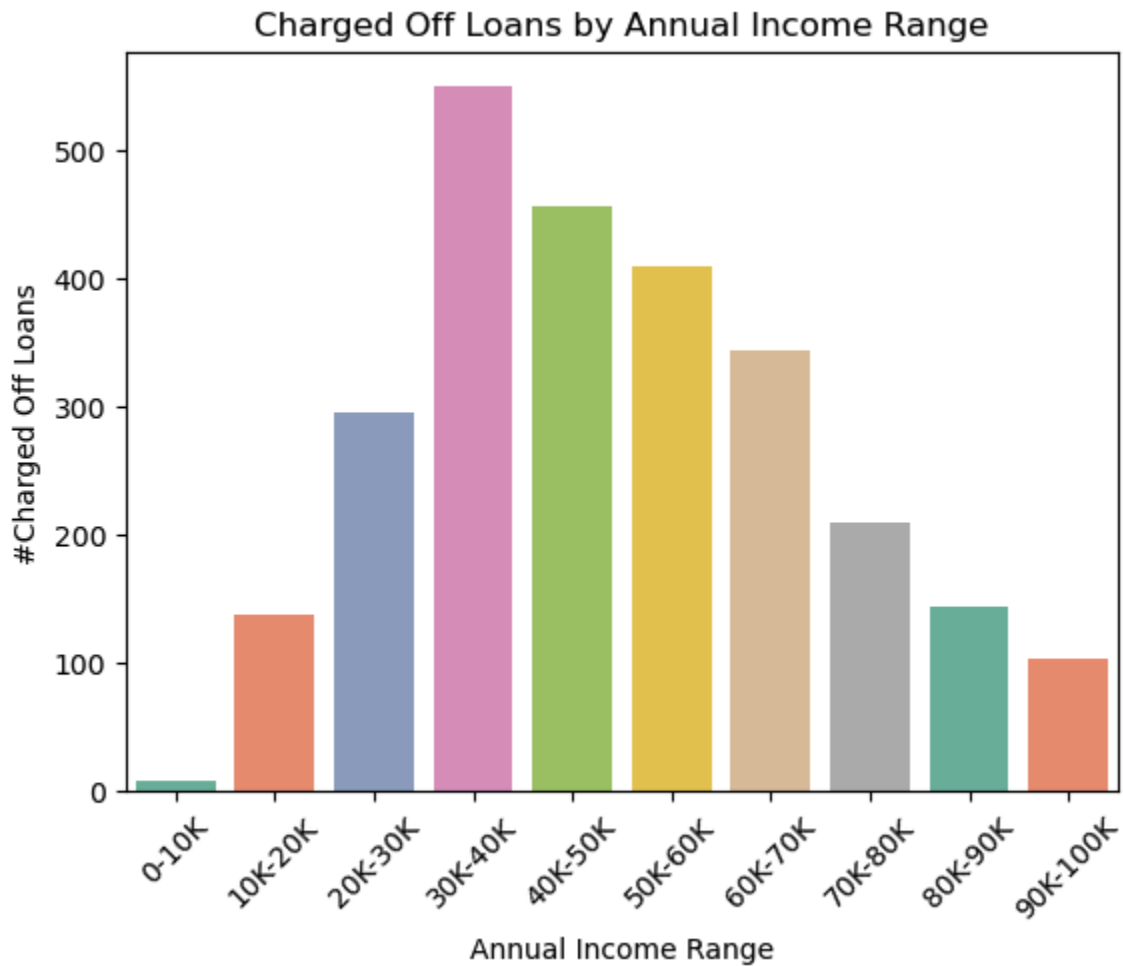
Conclusions

1) Here I can draw a conclusion that people having salary in range of 30-40K have highest chances of defaulting the Loans.
2) Also, as the salary increases the chances of loan default is getting reduced. Because a person with high salary will be able to manage his finances better.
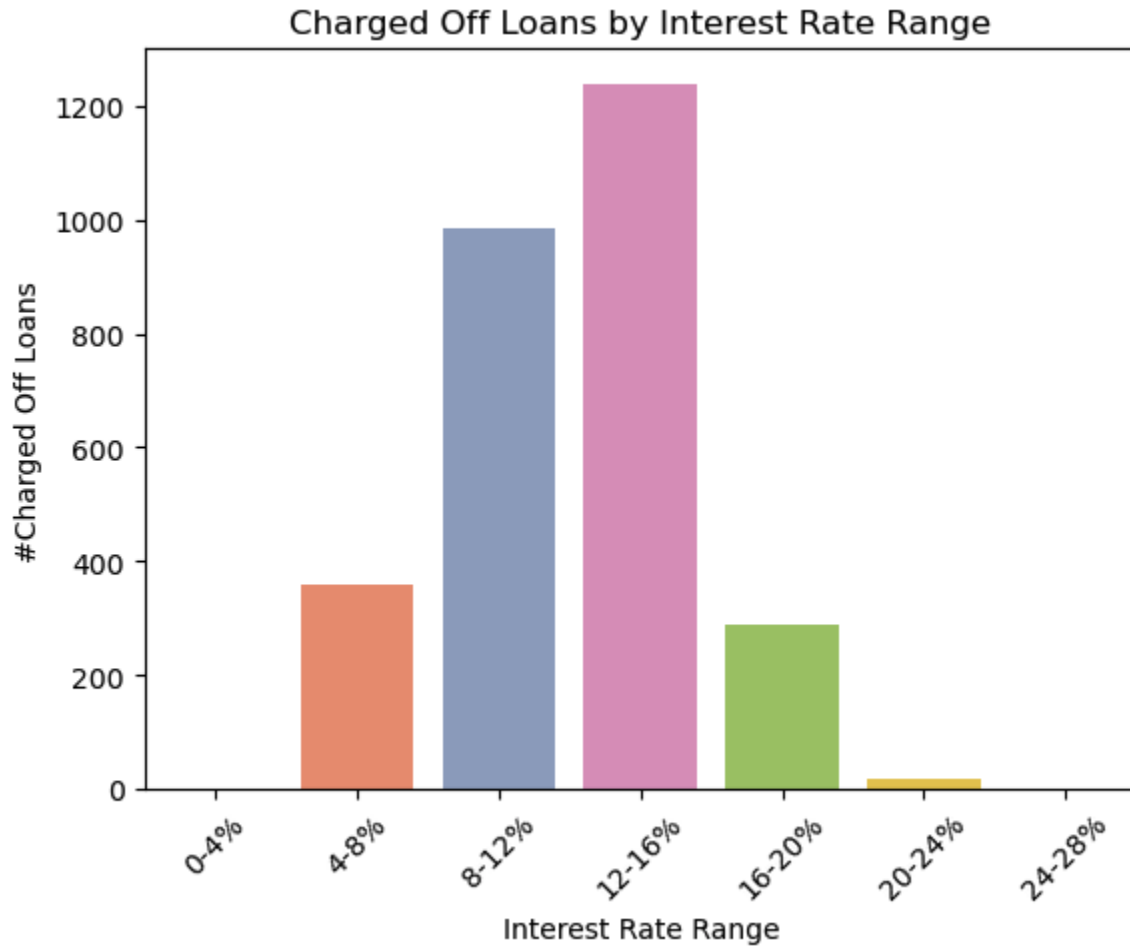
3) Charged Off Loans Vs Home Ownership Status
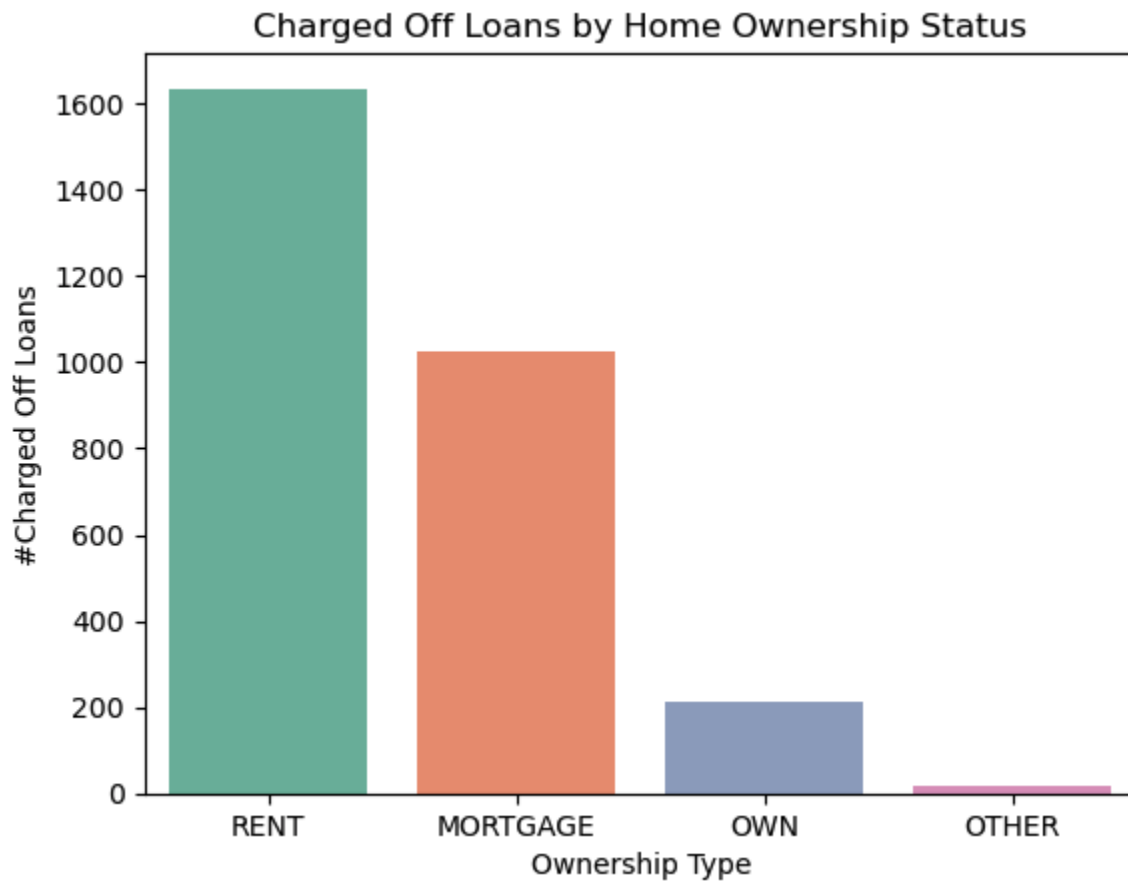
Conclusions

1) Here we can draw a conclusion that people living on Rent has highest chances of defaulting the Loans.
2) And People who have own house has very less chances of defaulting on loan.
3) Another conclusion comes is people having salary less than 50K and taking specifically personal loans are riskier for bank.

4) <u>Charged Off Loans Vs Interest Rate Ranges</u>

Charged Off Loans by Interest Rate Range

<u>Conclusions</u>

1) Here we can draw a conclusion that people having Interest rate in range of 12-16% have highest chances of defaulting the Loans.
2) Another conclusion we can draw is these rates belongs to Personal loans. So, people taking personal loan have high chances of defaulting.
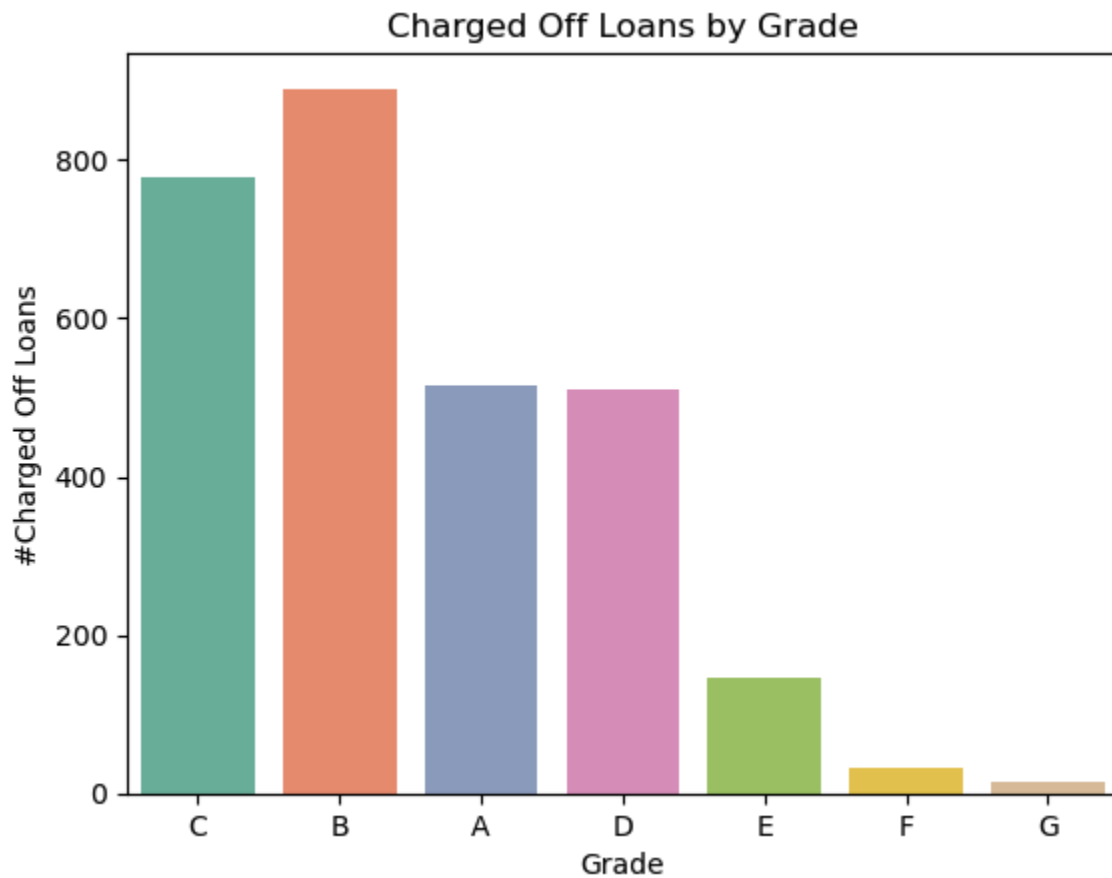
5) Charged Off Loans Vs Loans by Working Experience

Charged Off Loans by Experience (Binned)
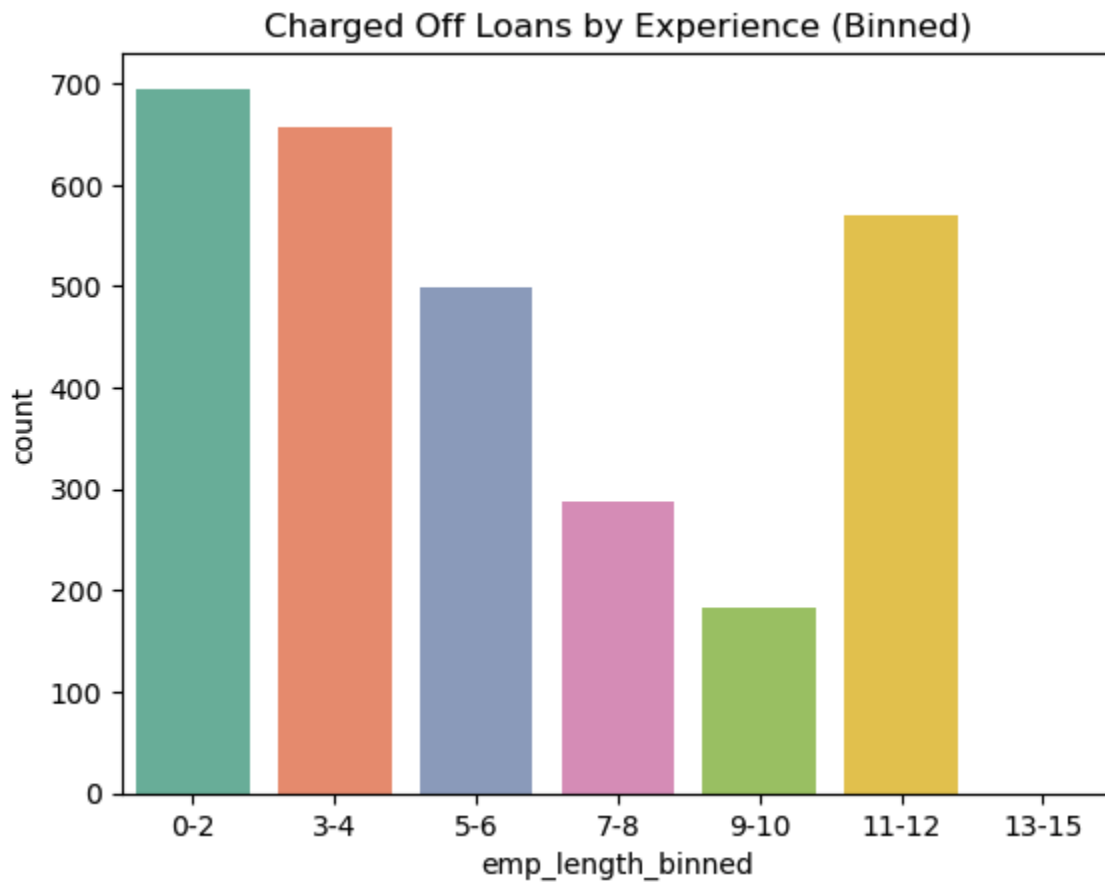
Conclusions

1) From the above chart we can conclude that within 0-2 years of experience people has most of the "Charged OFF" Loans.
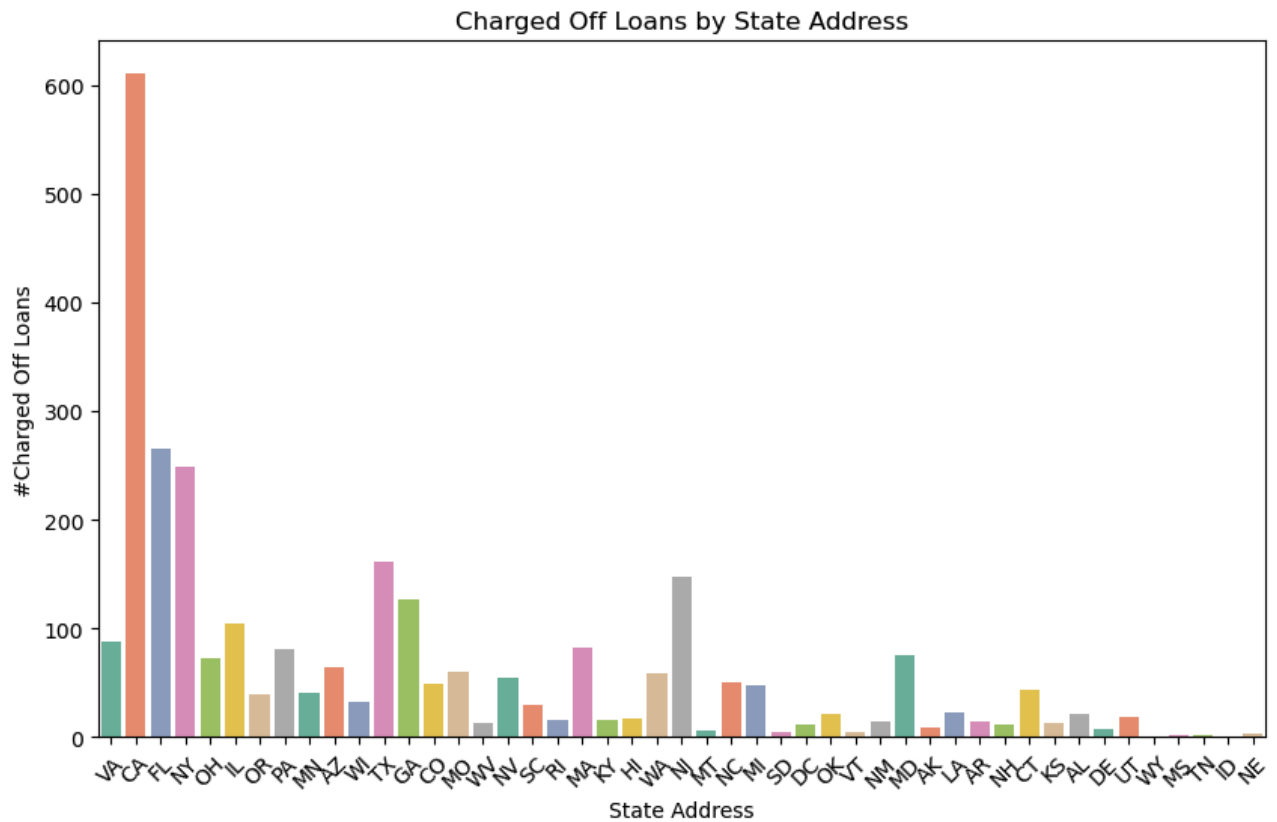2) And as experience increases chance of defaulting loans decreases.

6) Charged Off Loans Vs Employee Grade

Charged Off Loans by Grade

Conclusions

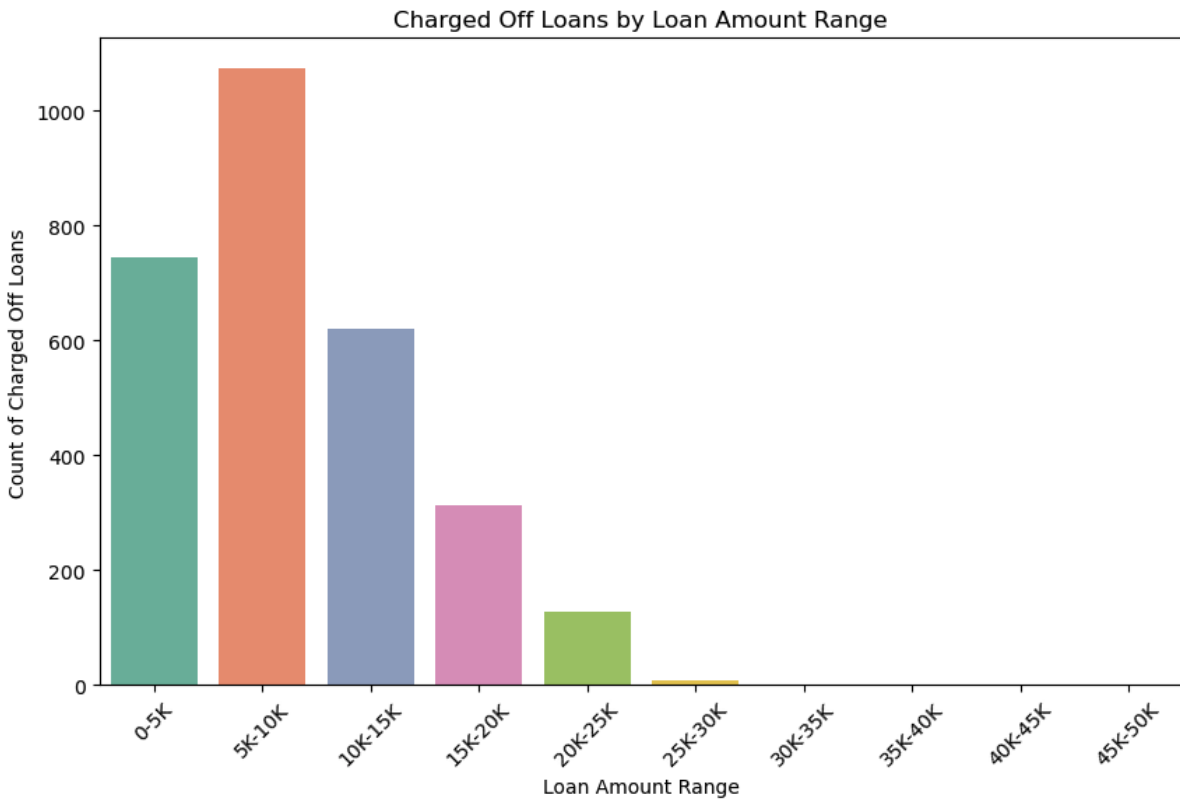1) From the above chart we can conclude that people with grade 'B' have defaulted the Loans most of the time.

7) Charged Off Loans Vs State Address Code

Charged Off Loans by State Address

Conclusions

1) From the above chart we can conclude that people with state Address code as '**CA**' has defaulted the Loans most of the time

8) **Charged Off Loans Vs Loan Amount Range**

Charged Off Loans by Loan Amount Range

Conclusions

1) From the above chart we can conclude that people taken loans in range of 5K-10K has defaulted the Loans most of the time.

Final Conclusion Analysis-

1) People taking loans for debt reconsolidation have highest chances of defaulting the Loans.
2) People having salary in range of 30-40K have highest chances of defaulting the Loans.
3) Also, as the salary increases the chances of loan default is getting reduced. Because a person with high salary will be able to manage his finances better.
4) People living on Rent has highest chances of defaulting the Loans.
5) And, People who have own house has very less chances of defaulting on loan.
6) Another conclusion comes is people having salary less than 50K and taking personal loans are riskier for bank.
7) People having Interest rate in range of 12-16% have highest chances of defaulting the Loans.
8) Another conclusion we can draw is higher rates belongs to Personal loans. So, people taking personal loan have high chances of defaulting.
9) People having working experience in range of 0-2 years has most of the "Charged OFF" Loans.
10) And as experience increases chance of defaulting loans decreases.
11) People with grade 'B' has defaulted the Loans most of the time.
12) People with state Address code as 'CA' has defaulted the Loans most of the time
13) People taken loans in range of 5K-10K has defaulted the Loans most of the time.

THANKS