

LENDING CLUB CASE STUDY

Exploratory Data Analysis Assignment

Group members:

- ☐ Shraddha Jaiswal
- ☐ Sanat Srivastava

BUSINESS OBJECTIVE

A leading consumer finance company, which facilitates disbursement of personal loans, business loans, and loans for other purposes, via an online marketplace wants to identify “risky” loan applicants so that it can reduce disbursement of bad loans. At the same time legit applications shall not be rejected. Thus the key objectives are to:

- Study the data available for past loans
- Apply Exploratory Data Analysis techniques and identify patterns which indicate if a person is likely to default.
- Use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc

APPROACH

The entire study was broken down into simple steps, which were further divided into smaller sub steps.

The broader steps are shown below (in the order of their occurrence)...



DATA ASSESSMENT



Load data file and examine rows/columns using **read_csv** method.

Understand columns using the Data Dictionary

Obtain info on dataframe using **info** method such as the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values)

Using the **describe()** method calculated some statistical data like percentile, mean and standard deviation of the numerical value.

DATA CLEANING

- Drop columns with high null values
- Drop unique columns
- Drop columns not relevant for analysis
- Drop rows where loan status is “Current”

DATA PREPARATION



Converting text values to numeric values, e.g. `int_rate`, `loan_status`, `term` , etc.

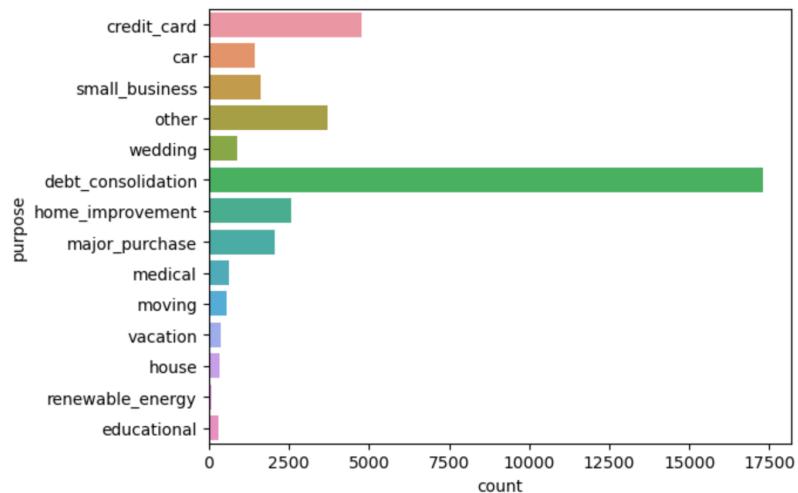
Setting correct data type on numeric columns, e.g. `emp_length`, `loan_amnt`, `annual_inc`, etc.

Outlier treatment

Creating data groups on continuous data columns, e.g. `annual_inc`, `loan_amnt`, `dti`

DATA ANALYSIS

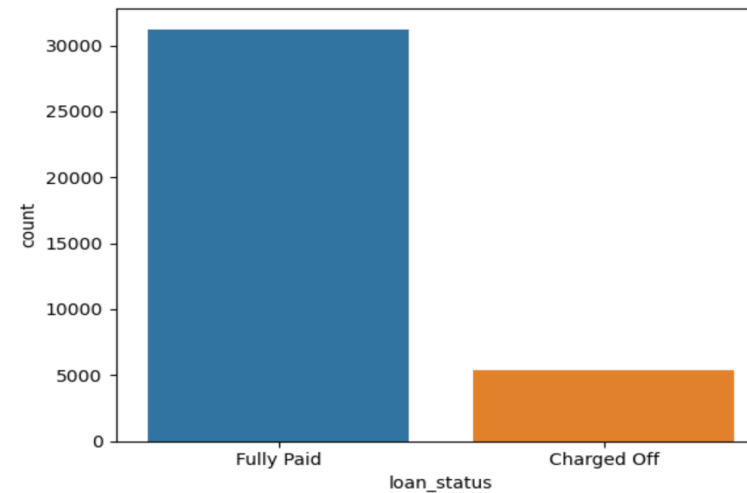
Loan Purpose



Observation:-

Nearly 50% of the loans were taken for debt consolidation. The next big head was that of credit card.

Charged off Loan vs Fully Paid Loans

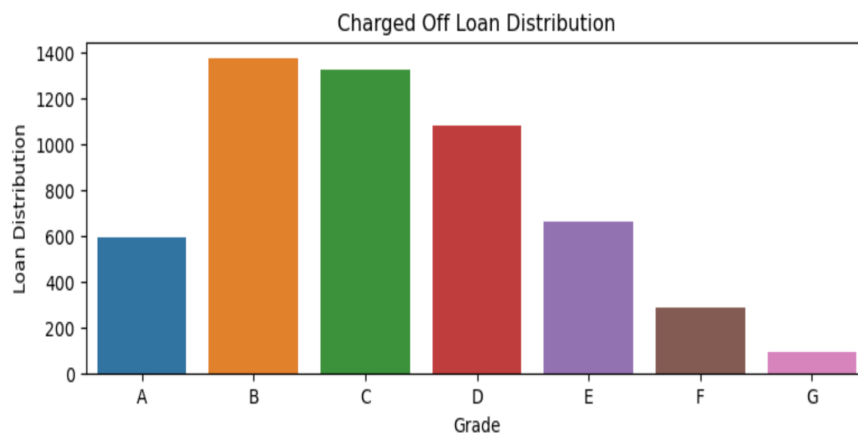


Observation:-

Nearly 14% of loans turn out to be bad loans.

DATA ANALYSIS

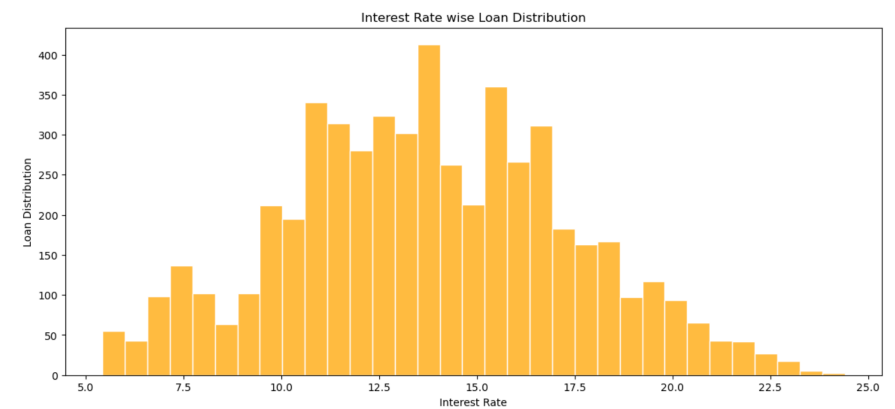
Charged off loan distribution / Grade



Observation:-

Loan default rate is high for Grades B, C & D.

Interest rate wise loan distribution

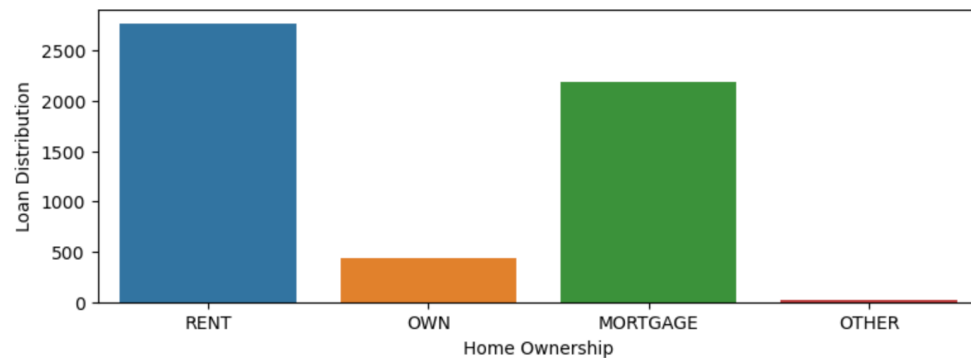


Observation:-

The graph indicates that loans taken at higher interest rates are more likely to default.

DATA ANALYSIS

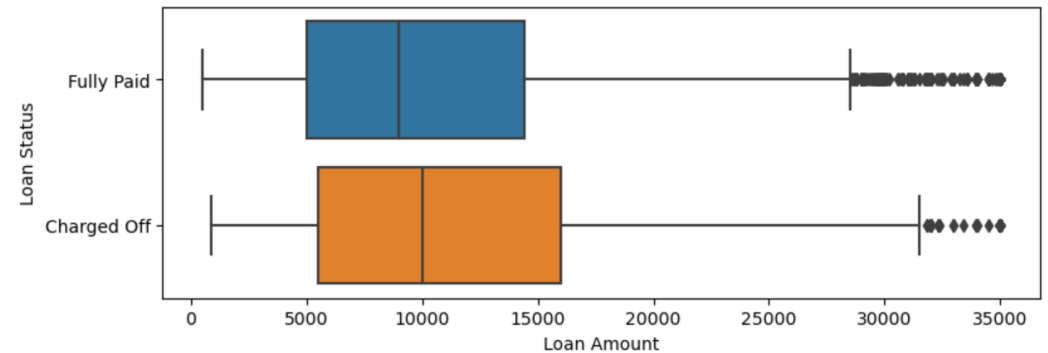
Charged off loan distribution / Home Ownership



Observation:-

People who live in Rented accommodation or have their homes mortgaged are more likely to default.

Loan Status / Loan Amount

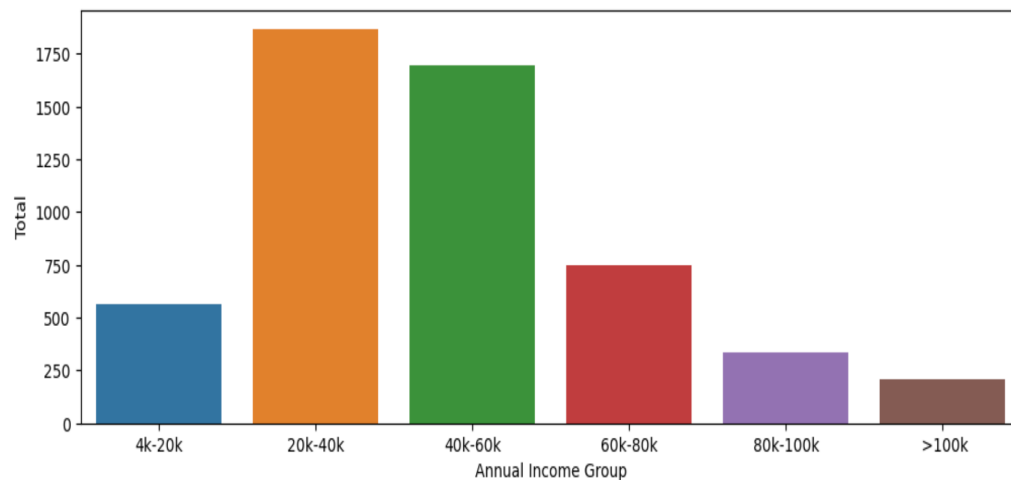


Observation:-

It can be inferred from the above graph that if the loan amount is higher, chances of default are more.

DATA ANALYSIS

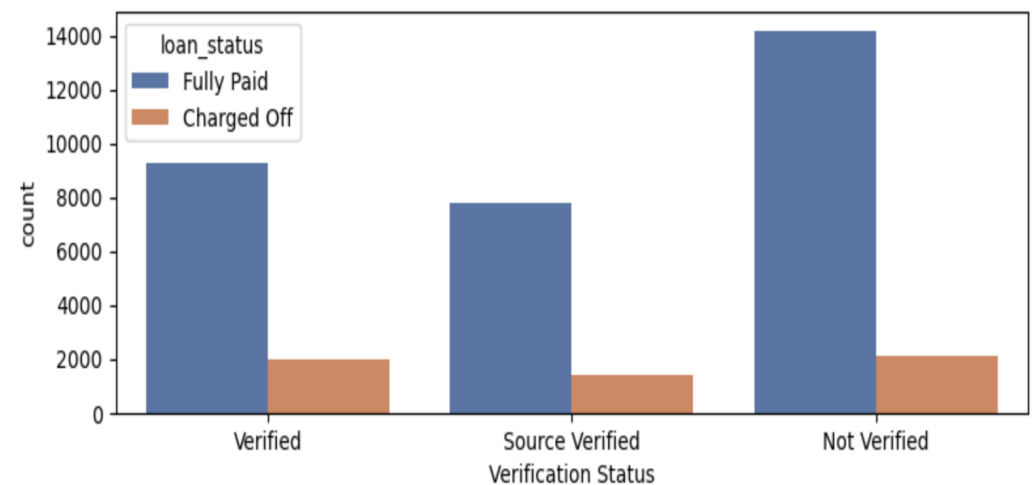
Monthly income group vs Charged off loans



Observation:-

Maximum loan charge off happened in the annual income group ranging from 20k to 60k. Higher the annual income, lesser the default.

Verification status and its impact on Charged off loans

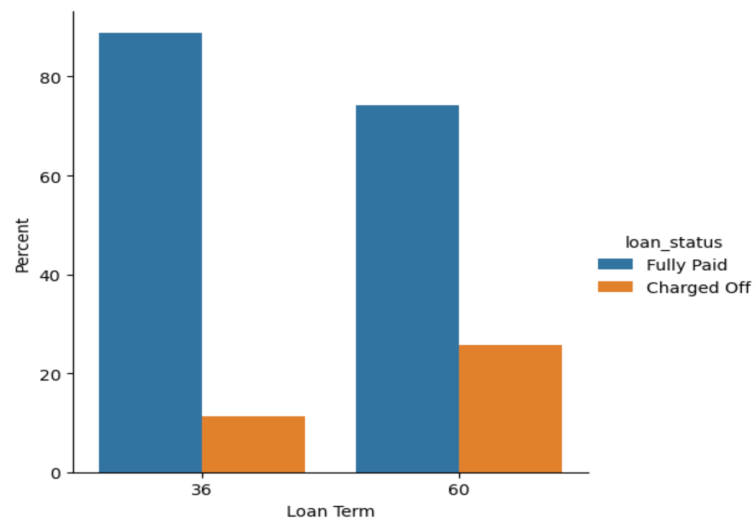


Observation:-

Not verified loans default the max. However, there are large number of verified loans also which have defaulted. Verification process is not robust.

DATA ANALYSIS

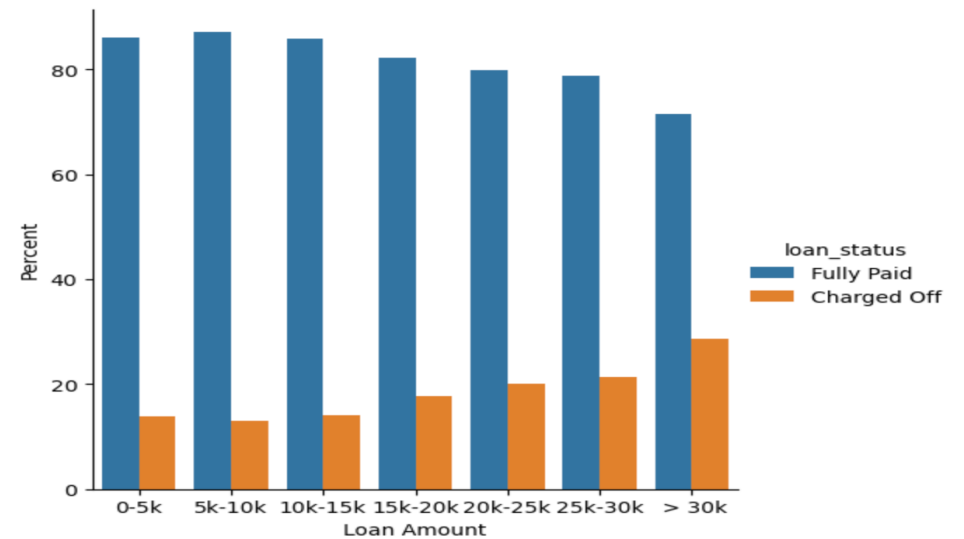
Loan Term and its effect on Loan Status



Observation:-

Loans given for longer terms are more likely to get charged off.

Loan Amount and its effect on Loan Status

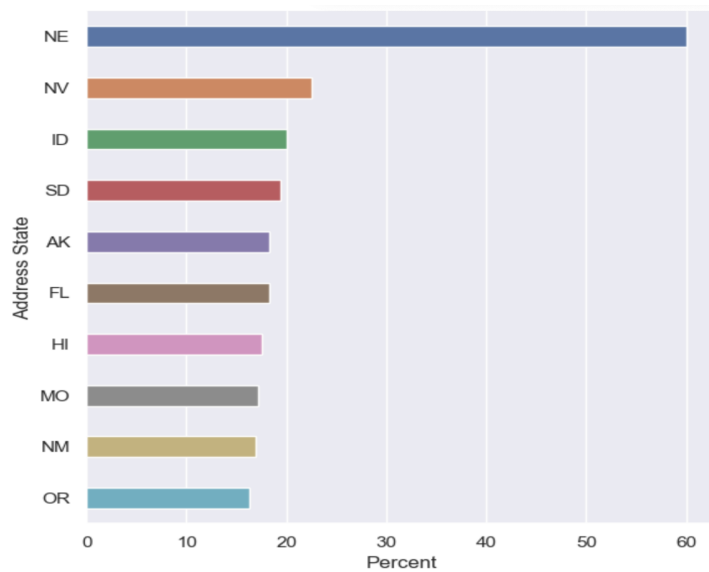


Observation:-

Likelihood of loan default increases with loan amount.

DATA ANALYSIS

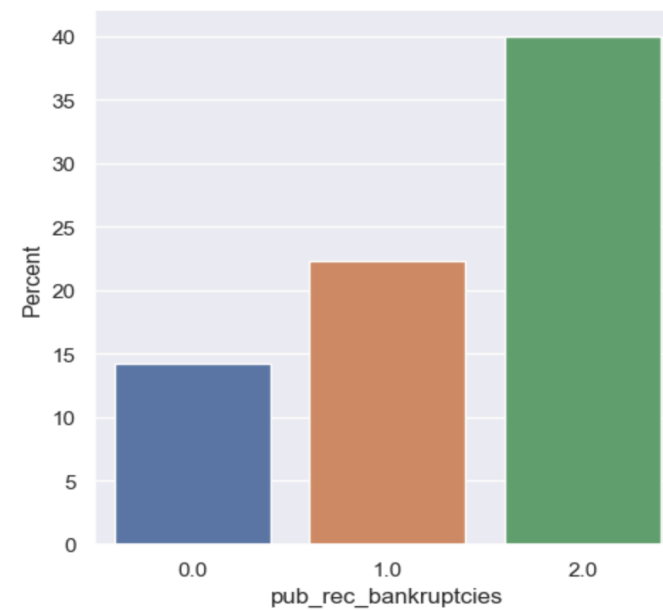
State wise data for max loan defaults



Observation:-

Maximum loan default comes from states : NE, NV, ID, SD, AK, FL.

pub_rec_bankruptcies vs charged off loans

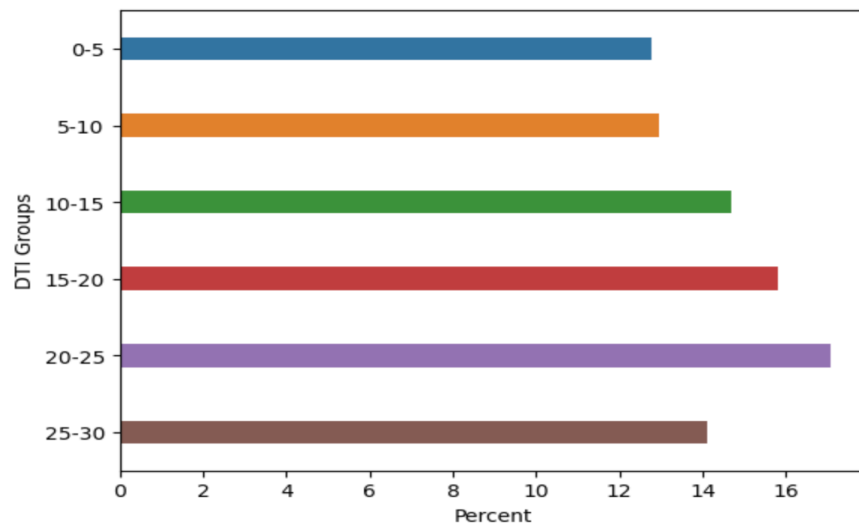


Observation:-

Higher the public records for bankruptcies, greater is the risk of loan default.

DATA ANALYSIS

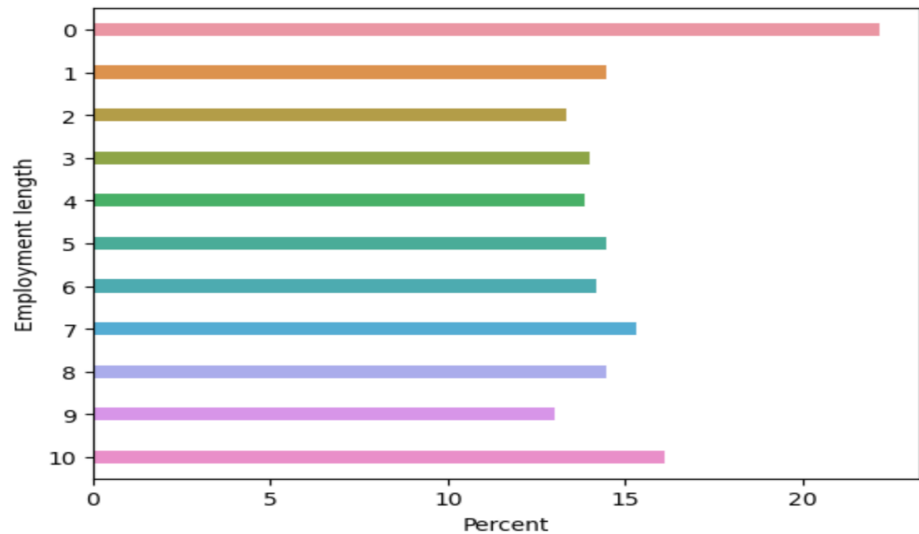
DTI Groups vs Charged off Loans



Observation:-

Increase in Debt to Income ratio may lead to higher default rate.

Employment length vs Charged off Loans

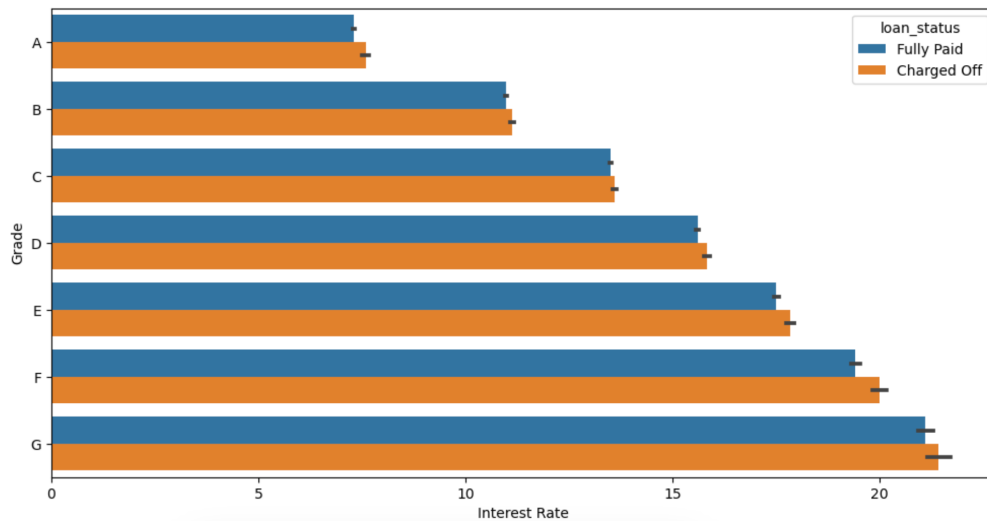


Observation:-

Unknown Employment Duration has max charged off loans.

DATA ANALYSIS

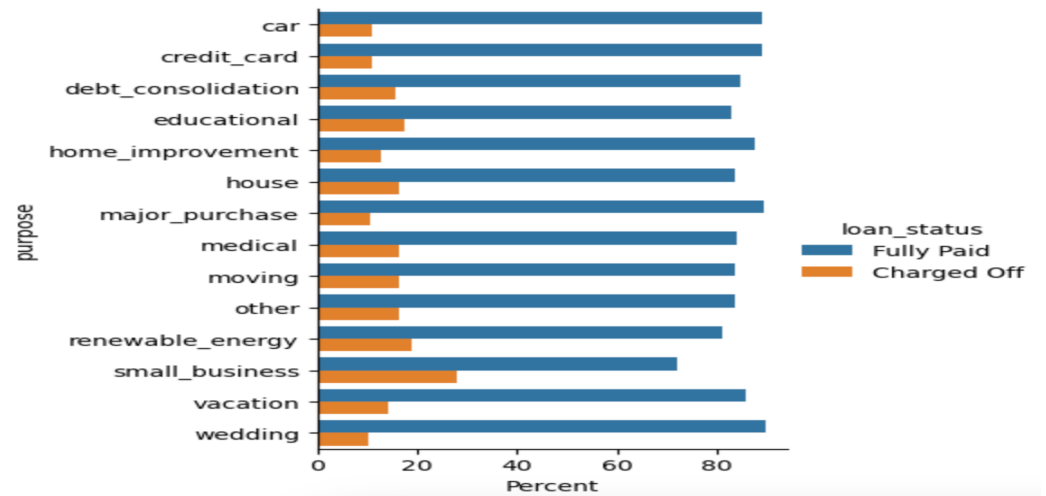
Grade & Interest Rate effect on loan status



Observation:-

Higher the Grade & Interest Rate, higher the Charged Off loans rate.

Loan purpose vs Loan status

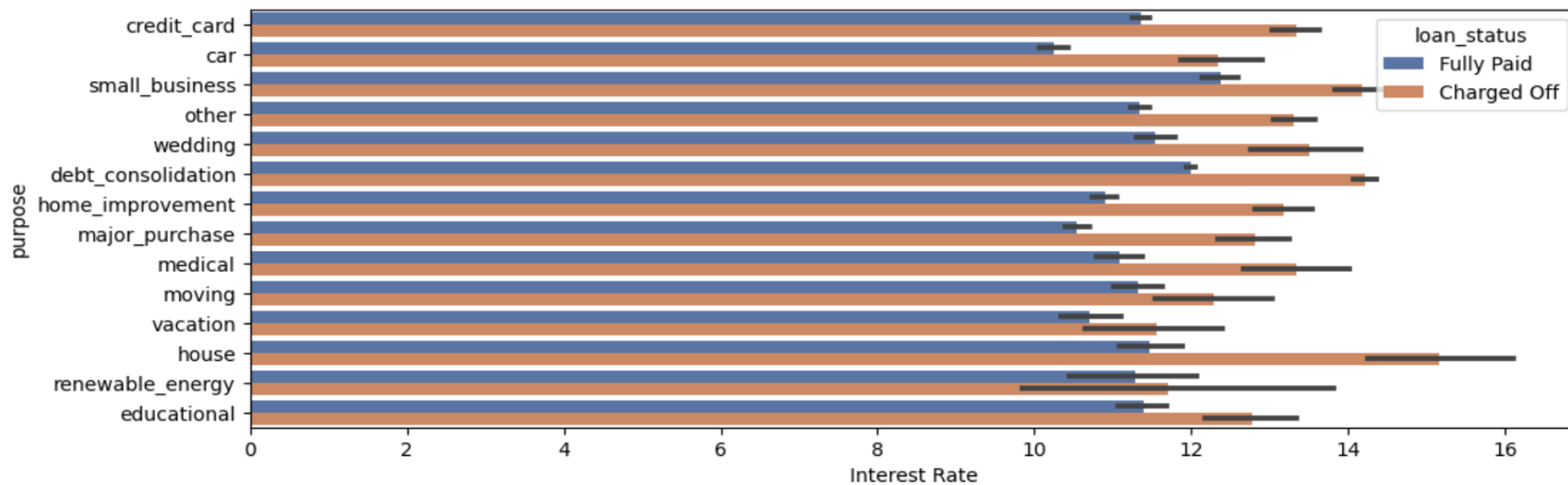


Observation:-

Small Business loans are more likely to get charged off.

DATA ANALYSIS

Loan purpose & Interest Rate's impact on Loan Status

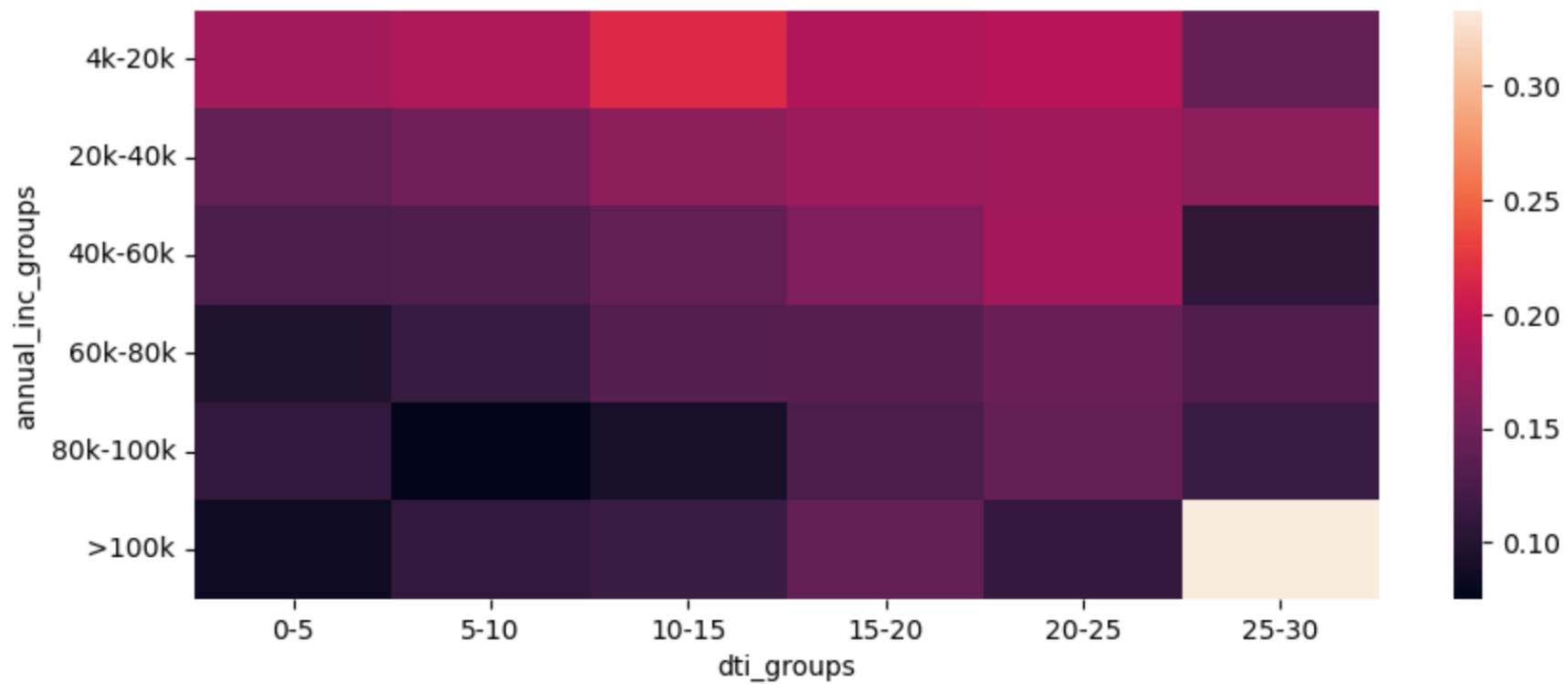


Observation:-

Small business and House purpose loans were given at very high interest rates, which led to high percentage of default. Interest rate has huge impact on loan default.

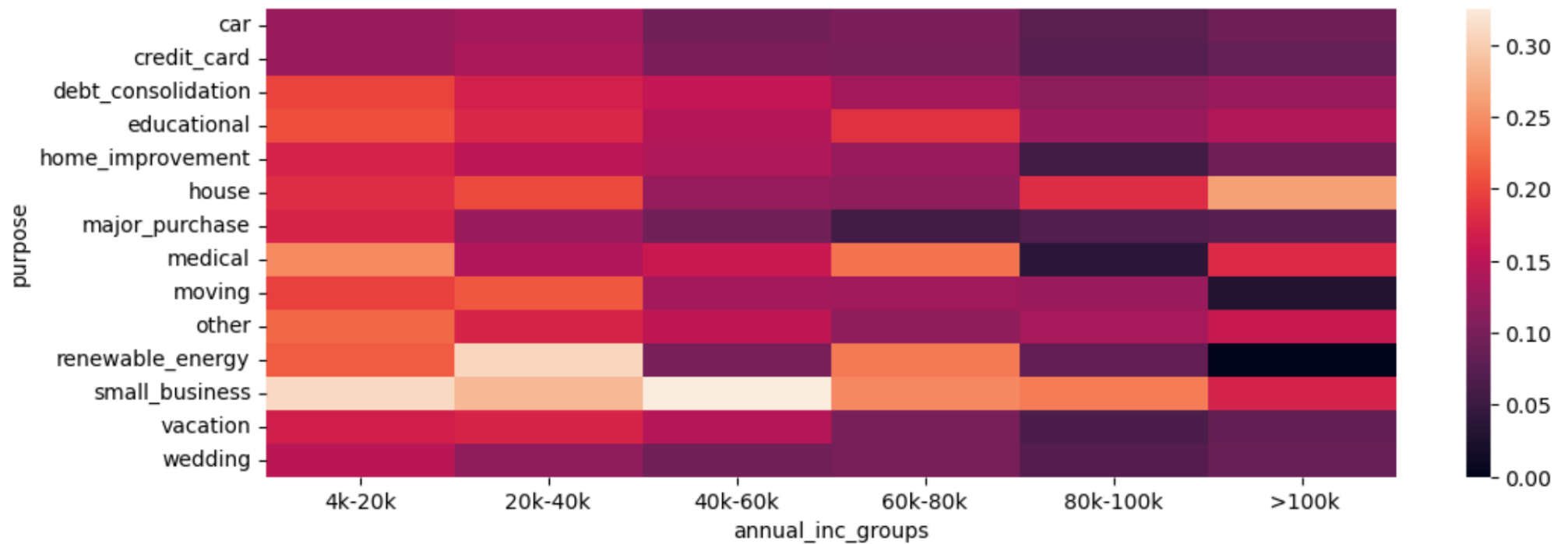
DATA ANALYSIS

Annual Income Groups – DTI Groups heatmap



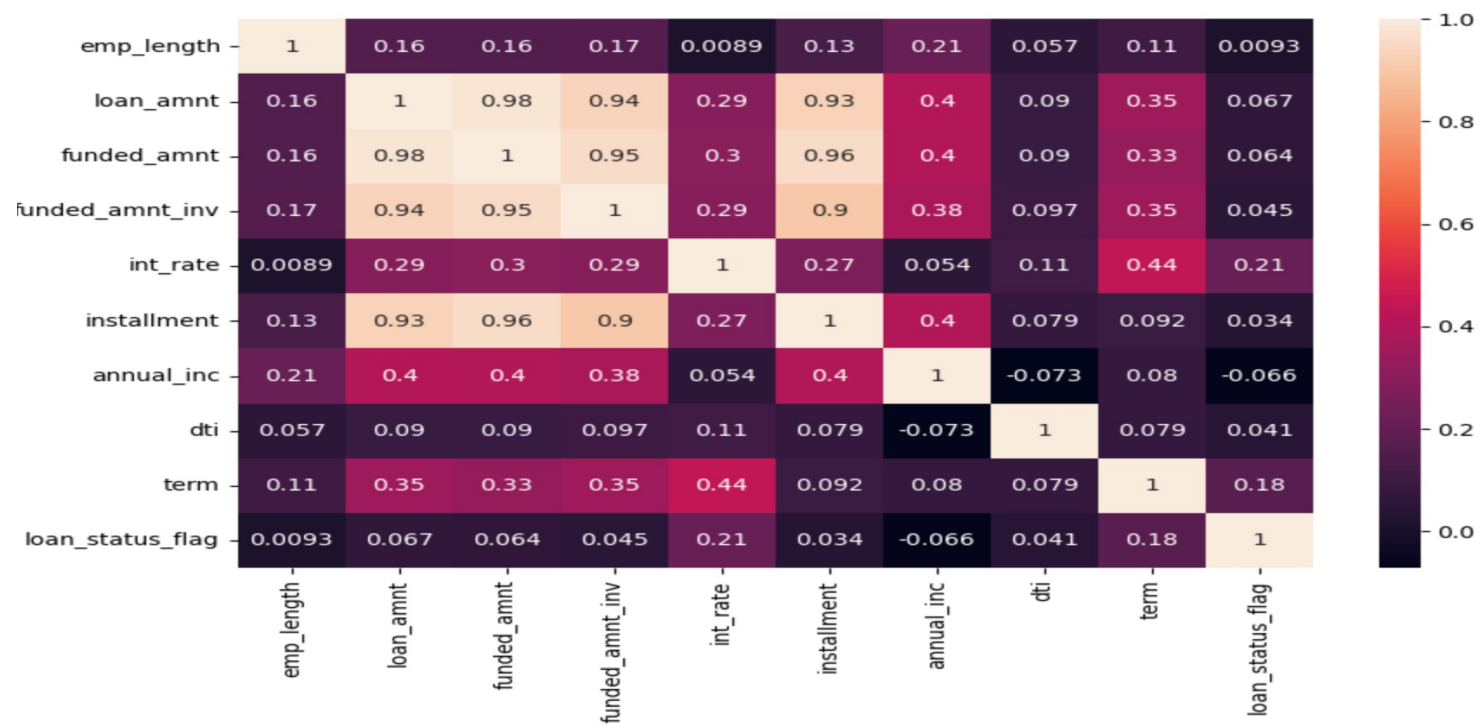
DATA ANALYSIS

Annual Income Groups – Loan purpose heatmap



DATA ANALYSIS

Heatmap of correlation among numeric variables



CONCLUSION / RECOMMENDATION

- Loan default rate is high for Grades B, C & D.
- Higher the Grade & Interest Rate, higher the Charged Off loan rate.
- People who live in Rented accommodation or have their homes mortgaged are more likely to default.
- Caution has to be exercised while giving loans in the lower annual income group specially at higher interest rate.
- Not verified loans default the max. However, there are large number of verified loans also which have defaulted. Verification process needs to be improved to reduce loan default.
- Loans given for longer terms are more likely to get charged off.
- Likelihood of loan default increases with loan amount.
- Maximum loan default comes from states : NE, NV, ID, SD, AK, FL.
- Higher the public records for bankruptcies, greater is the risk of loan default.
- Increase in Debt to Income ratio leads to higher default rate.
- Unknown Employment Duration has max charged off loans.
- Small business and House purpose loans were given at very high interest rates, which led to high percentage of default. Interest rate has huge impact on loan default.