Lending Club Case Study

GROUP MEMBERS :

- INDRASHIS CHATTERJEE
- RANJAN CHOUDHURY

Problem Statement

The company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss).

The aim is to identify patterns which indicate if a person is likely to 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.

In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Business Understanding

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**

Business Understanding

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

▶ **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

▶ **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)

Business Understanding

LOAN DATASET

Loan Accepted → Default

Loan Accepted → Non-Default

Loan Rejected

(Not considered in dataset)
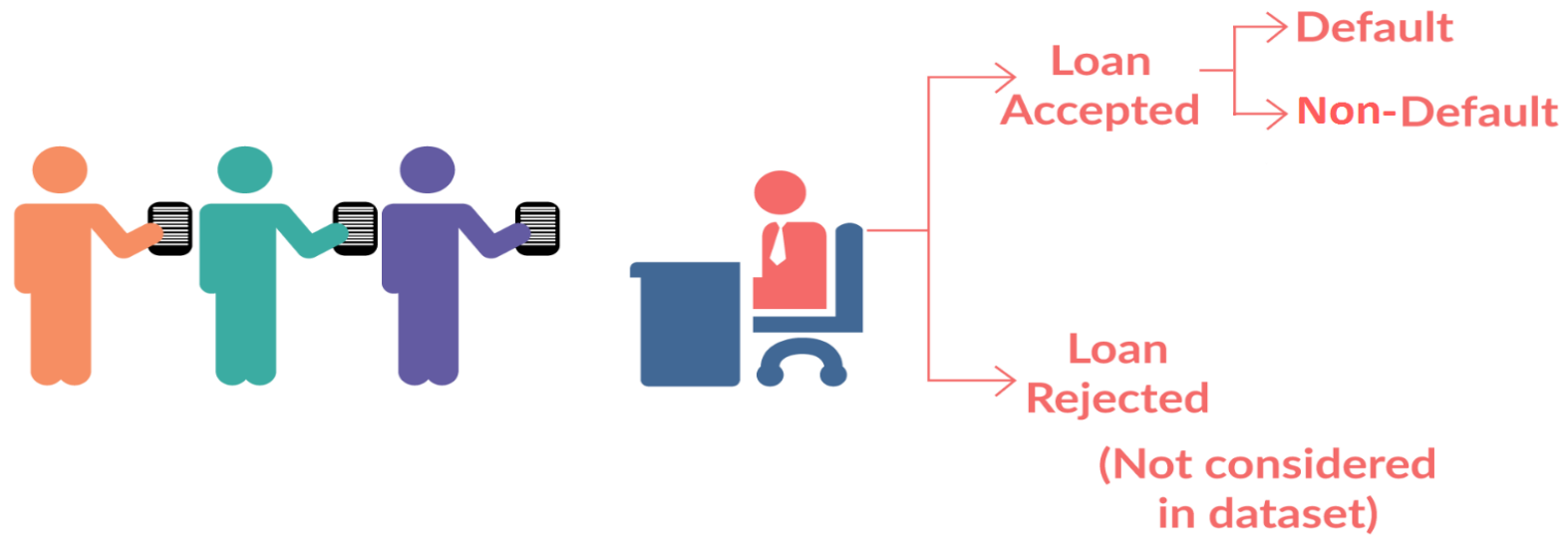
Methodology
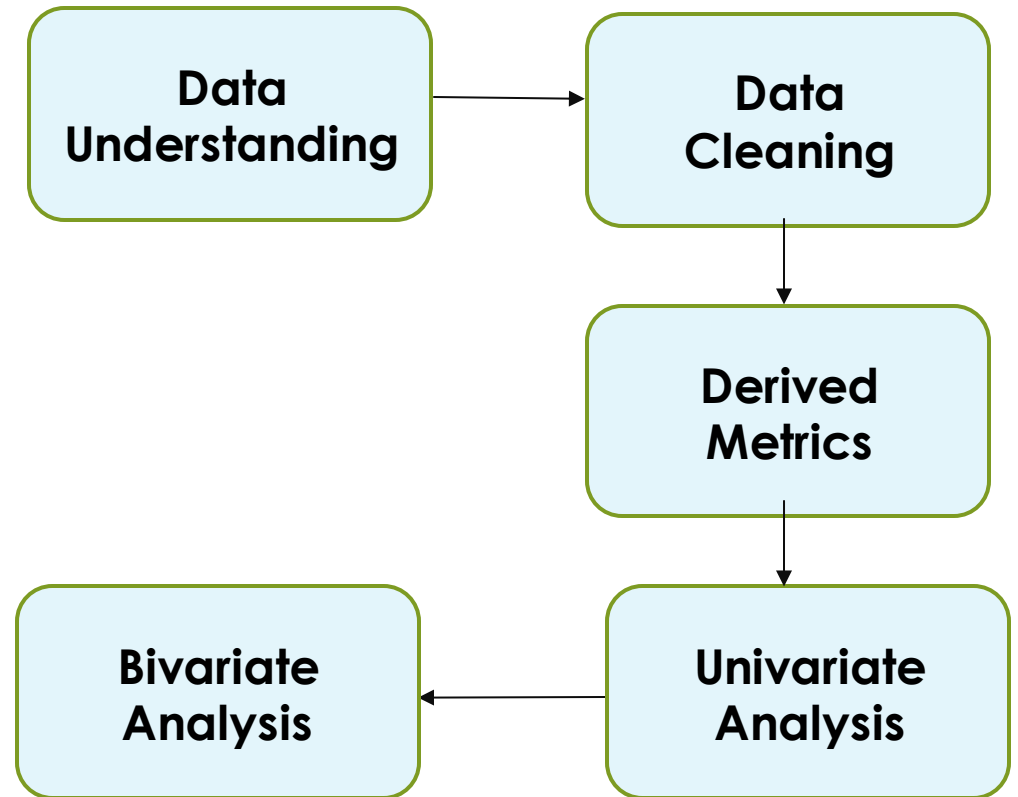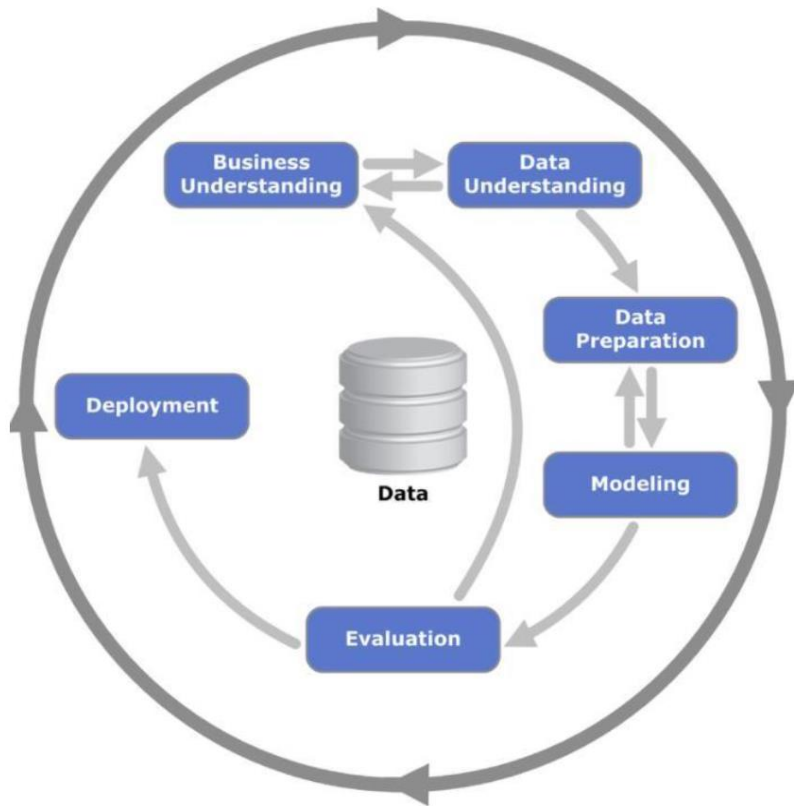
Data Understanding & Cleaning

- The provided data had a lot of null values. Columns with more than 60% null have been dropped and are also imputed wherever applicable

- Few of the columns had malformed data. Those were standardized.

- Business driven, data driven and type driven derived columns are created as required.

- Checked for duplicate data throughout if any in the dataset. There were no duplicate rows

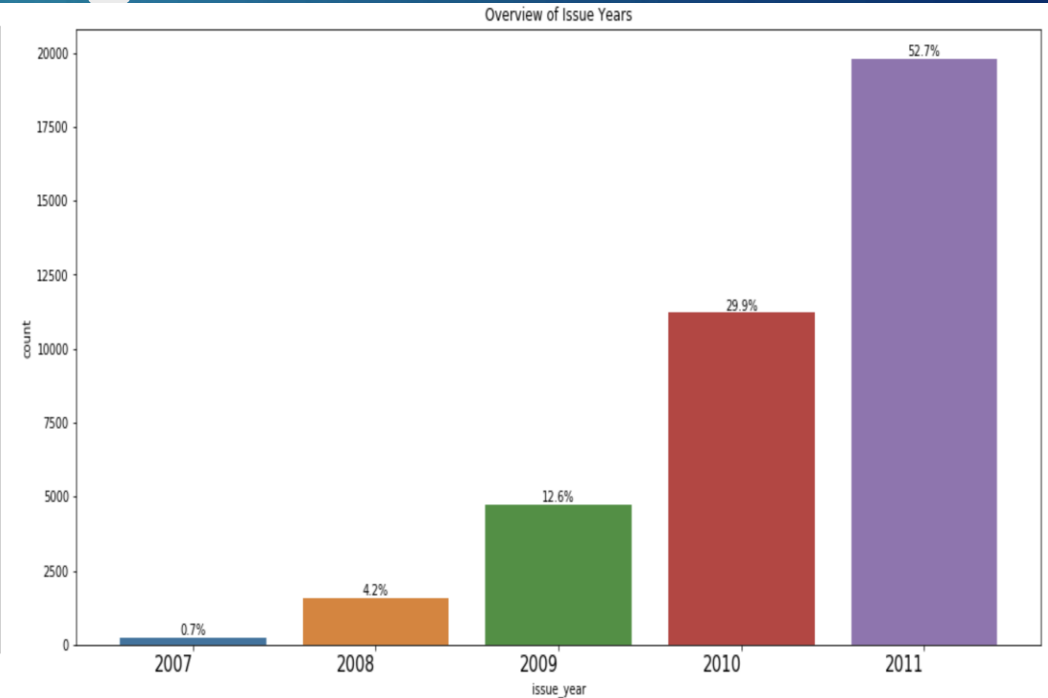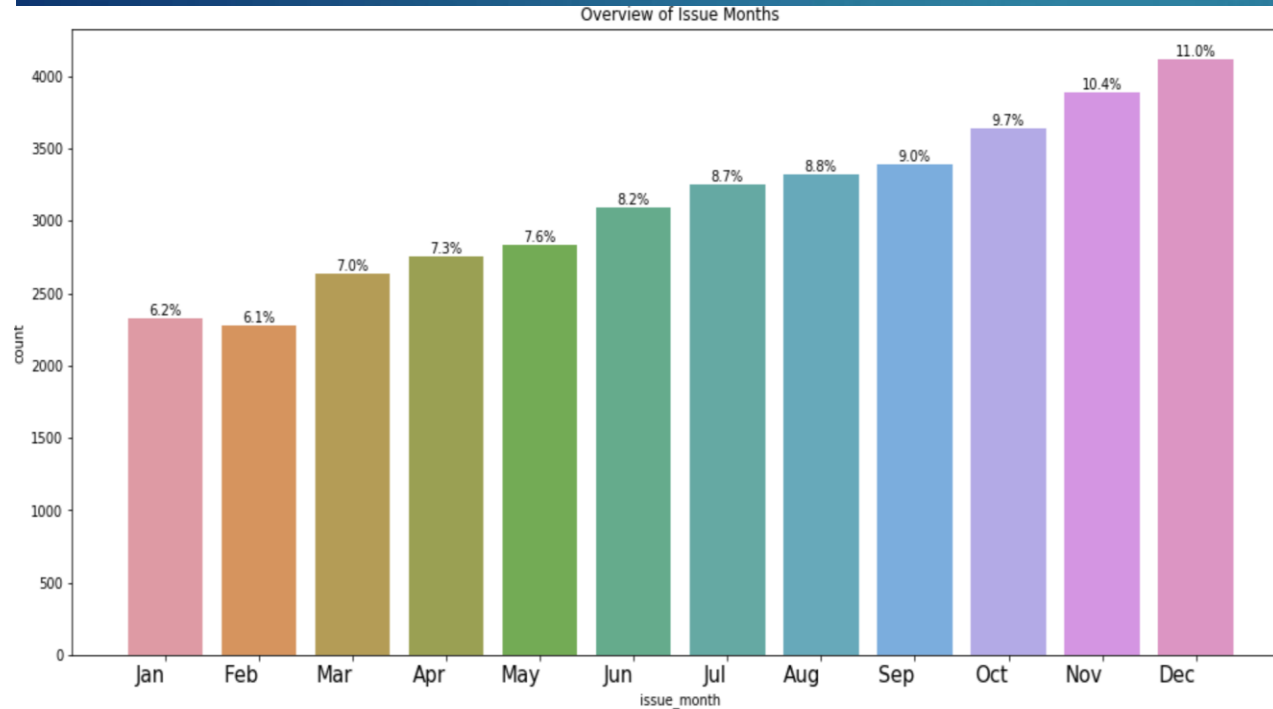- After understanding and cleanup, proceeded with data analysis

Overview of Lending Club

14.4% of total borrower accounts for defaulter

Average profit is of 9.51%

Overview of Lending Club

Overview of Issue Months

Overview of Issue Years
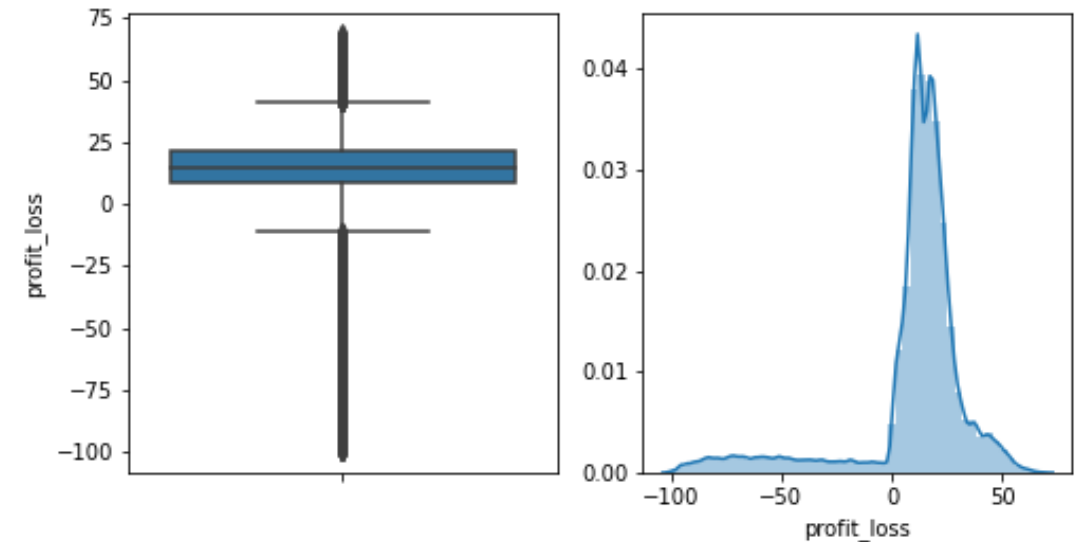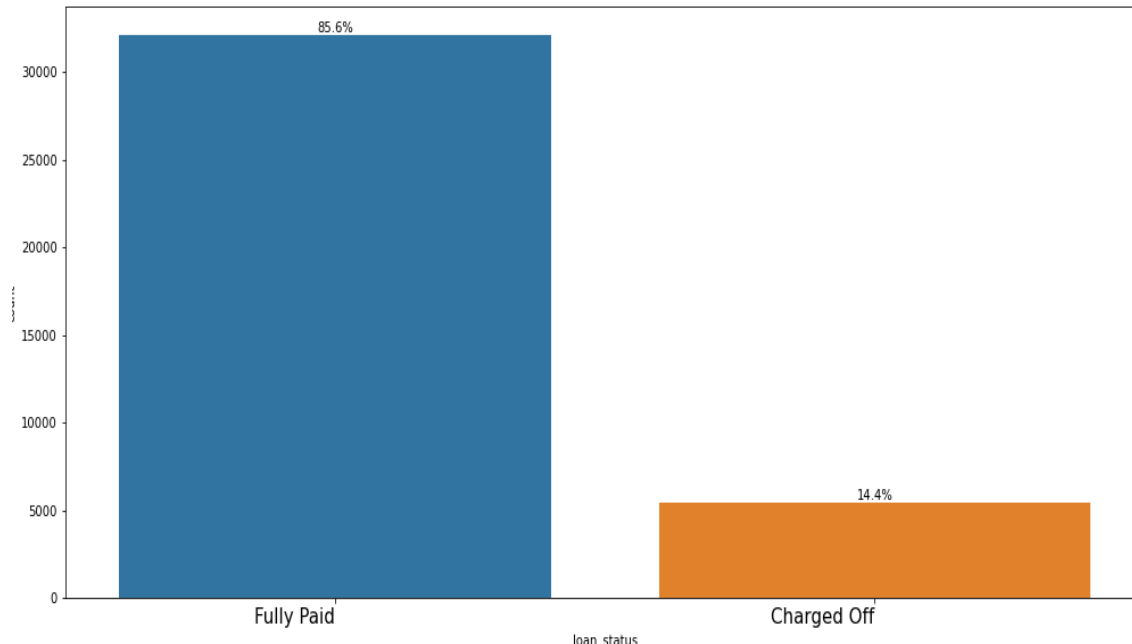
Trend of taking highest loans in the month of Dec (may be due to their festivals). The trend increases from Jan - Dec
There is a increasing trend in number of loans every year.

Impact of Interest Rate and Grade

Increasing rate of interest shows more loan defaulter. Loans with greater interest might be a risk.

Lower grade of loan with increasing rate of interest shows higher chances of defaulter.
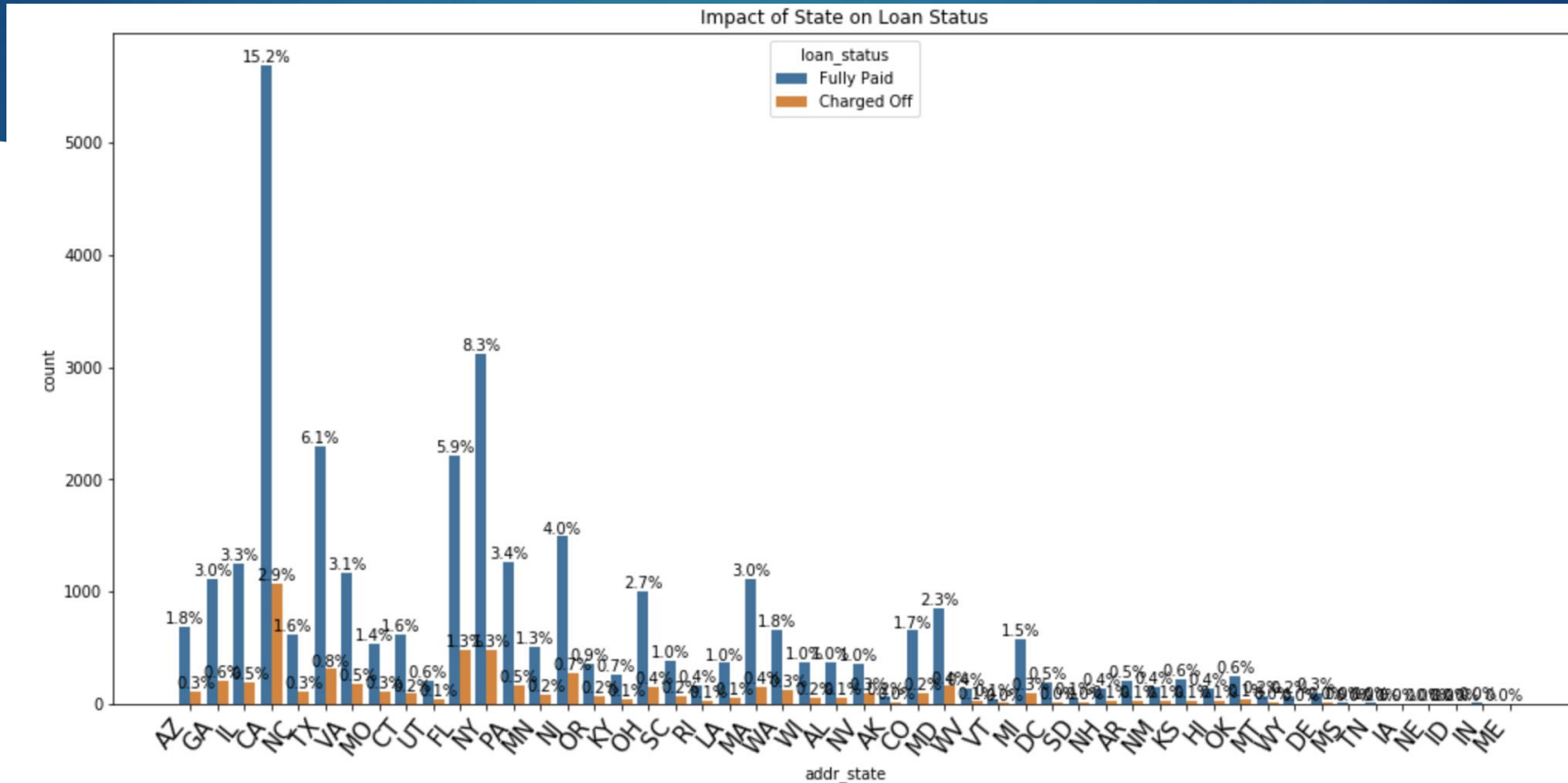
Impact of Homeownership and Experience

Mortgaged and rented homeowners shows a greater number of loan with more defaulters when interest rate is medium to high.

Seniors accounts for more defaults but juniors with rented house also accounts to more default than mortgaged

Impact of State on loan status
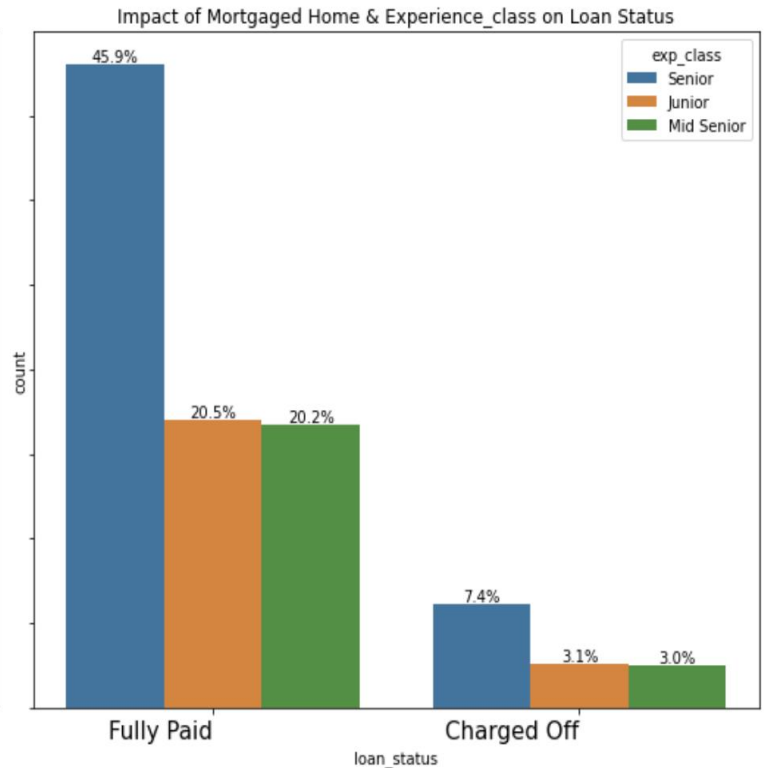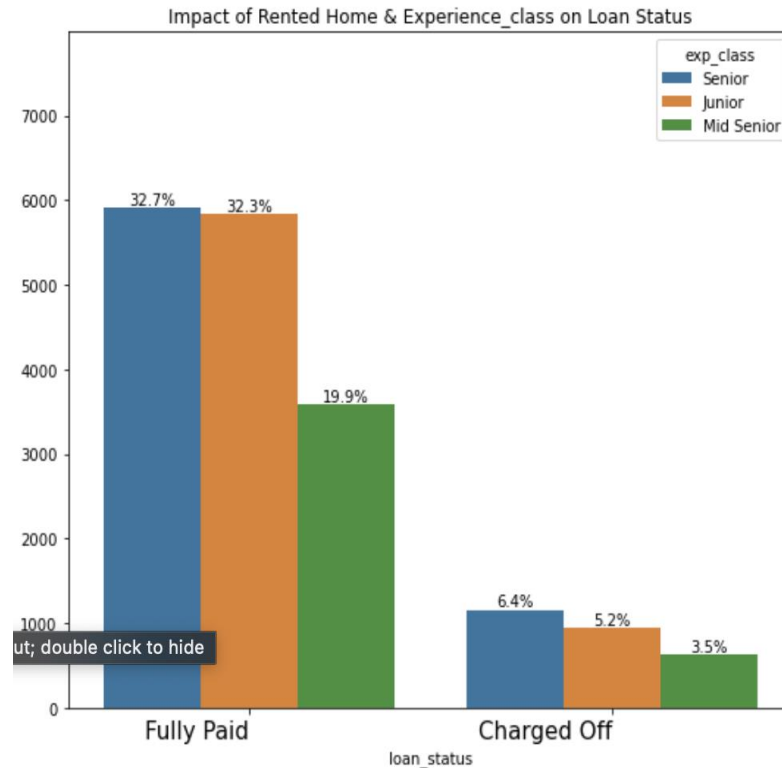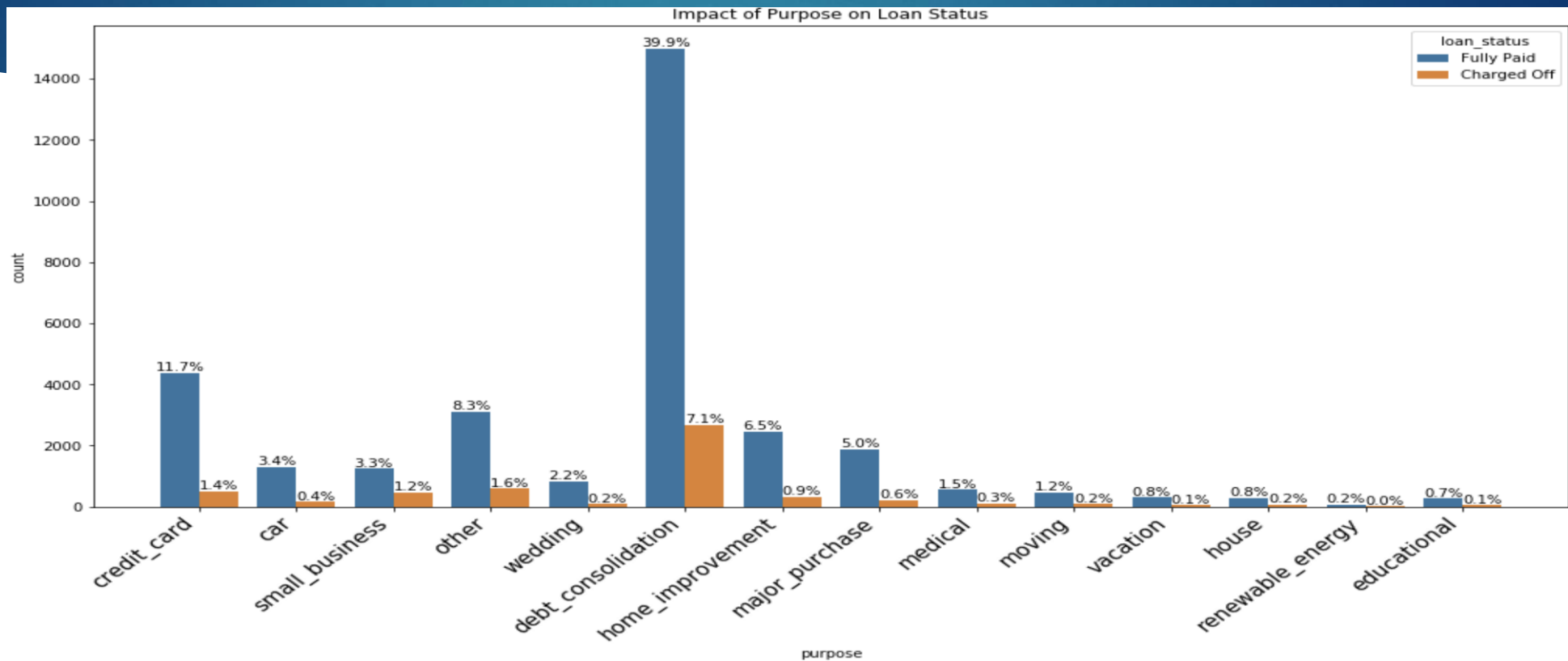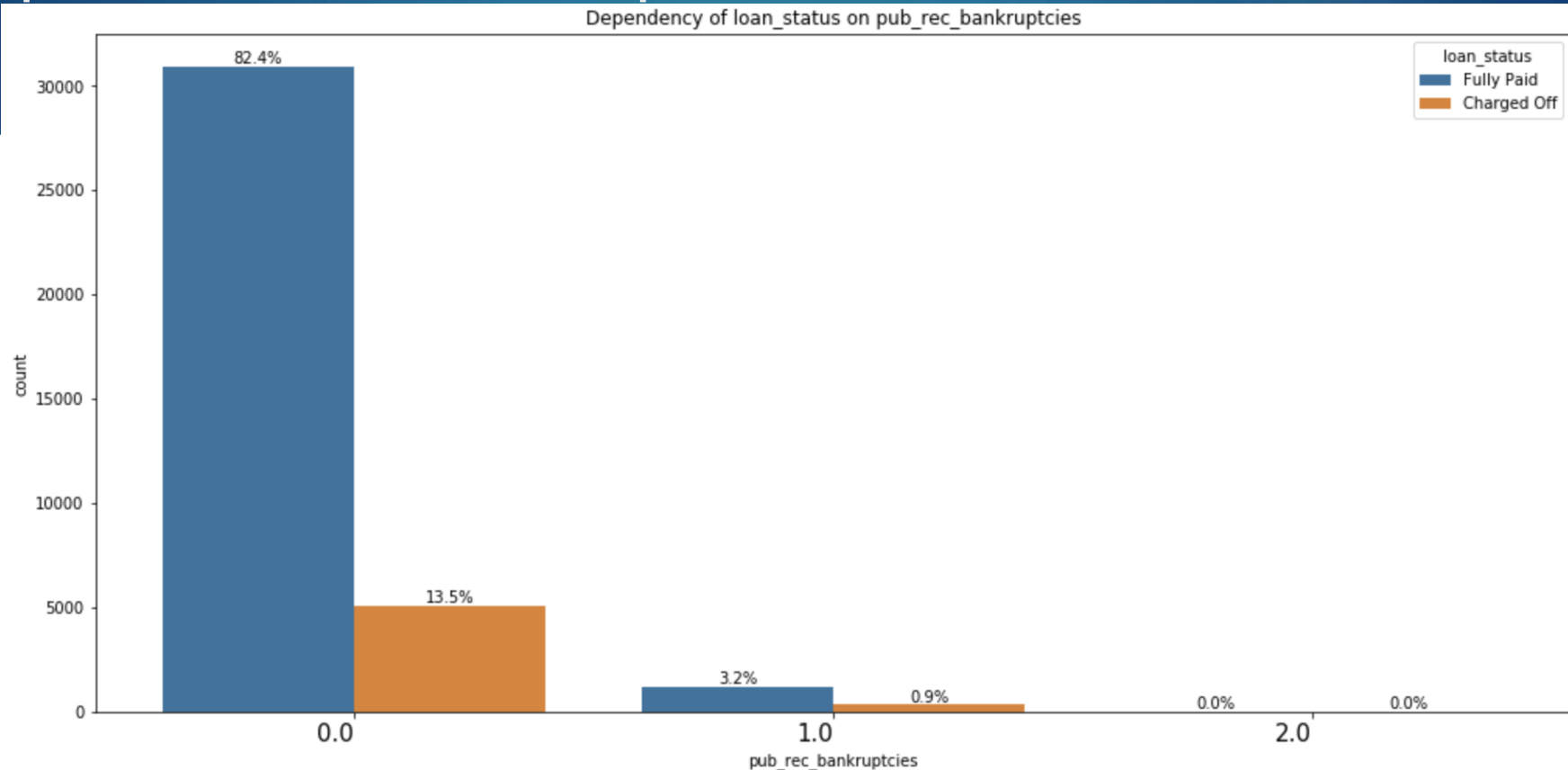

Impact of State on Loan Status

Number of borrower from CA is most and it also has the highest number of defaulter. Borrower from state CA needs to be scrutinized properly

Impact of Purpose on loan status

Impact of Purpose on Loan Status

The purpose for debt_consolidation attracts more & also results more defaulters too
Borrower taking debt_consolidation loans at a higher interest

Impact of Bankruptcies Record

Dependency of loan_status on pub_rec_bankruptcies

pub_rec,pub_rec_bankruptcies has impact on loan status if compare ratios of fully paid to charged off it is more for 0 & eventually decreases i.e. is more defaulter cases

Conclusion

1. The provided data has lot of null and missing values. Columns with more that 50% null have been dropped and null have been enriched wherever applicable.

2. Few of the columns had malformed data. Those were standardized.

3. Business driven, data driven and type driven derived columns are created as required.

4. After data understanding and cleanup, univariate and bivariate analysis have been done.

5. Few of the strong driver variables identified are as follows:
   I. Increasing rate of interest shows more loan defaulter. Loans with greater interest might be a risk.
   II. Number of borrower from CA is most and it also has the highest number of defaulter. Borrower from state CA needs to be scrutinized properly.
   III. Lower grade of loan with increasing rate of interest shows higher chances of defaulter.
   IV. Loan with 60 months term shows greater number of defaulter when rate of interest is high.
   V. Mortgaged and rented homeowners shows a greater number of loan with more defaulters when interest rate is medium to high.
   VI. most of the people have taken loan for debt consolidation with high interest rate.

Conclusion

1. The provided data has lot of null and missing values. Columns with more that 50% null have been dropped and null have been enriched wherever applicable.

2. Few of the columns had malformed data. Those were standardized.

3. Business driven, data driven and type driven derived columns are created as required.

4. After data understanding and cleanup, univariate and bivariate analysis have been done.

5. Few of the strong driver variables identified are as follows:
   I. Increasing rate of interest shows more loan defaulter. Loans with greater interest might be a risk.
   II. Number of borrower from CA is most and it also has the highest number of defaulter. Borrower from state CA needs to be scrutinized properly.
   III. Lower grade of loan with increasing rate of interest shows higher chances of defaulter.
   IV. Loan with 60 months term shows greater number of defaulter when rate of interest is high.
   V. Mortgaged and rented homeowners shows a greater number of loan with more defaulters when interest rate is medium to high.
   VI. most of the people have taken loan for debt consolidation with high interest rate.

Driver Variables

1. Int_rate
2. Grade
3. dti
4. Home_ownership
5. Emp_length
6. Pub_rec_bankruptcy
7. Addr_state

Thank You