Lending Club Case Study

Submitted by

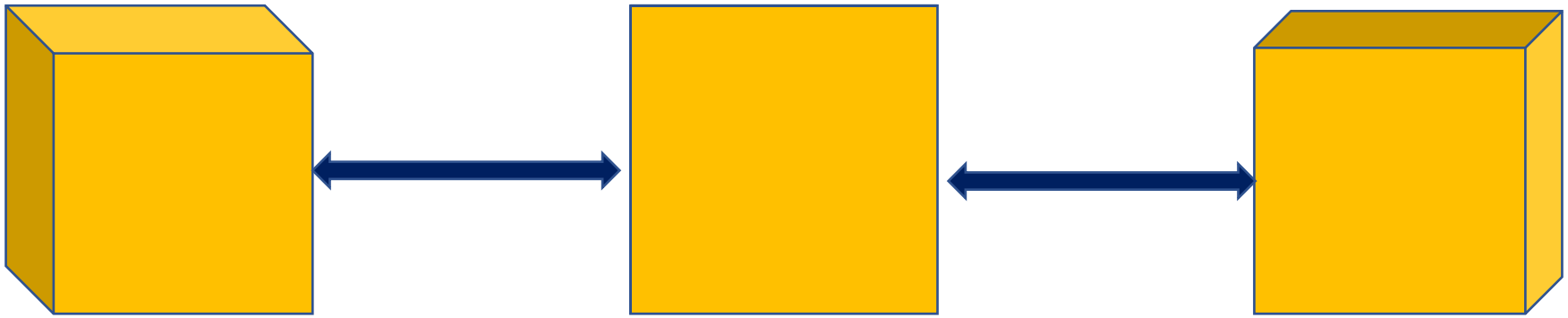
Nagendra J

Table of content

Since 2007, Lending club is a consumer finance company that offer various types of loans for urban clients. The product serves the need by providing personal loan, Business loan, Auto Refinance and Patient Solutions. Other products include:

- Business loans
- Business banking
- Institutional banking
- Institutional investment
- LC notes
- Other types of loans

It also provides a better platform for the investors to gain better profit for their investment. Lending club receives the loan application, verifies the authenticity of the client, offers the loan with attractive interest rates.



Business overview (problem statement):

As mentioned in the workflow, the company receives a loan application, the company must decide 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

Objective

Business objective:

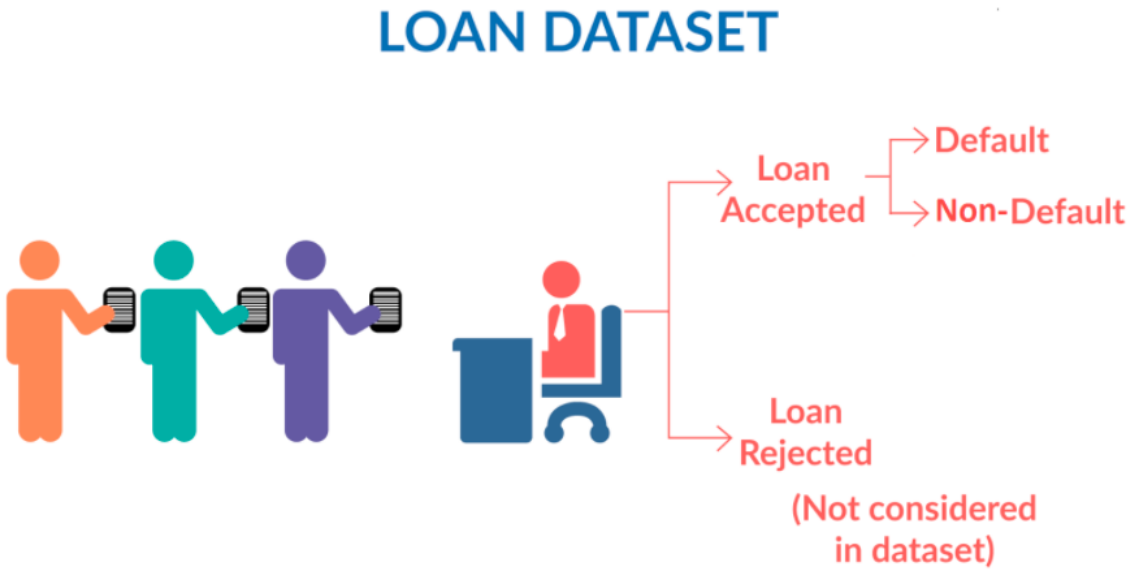
Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

- Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
- 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.
- To develop your understanding of the domain, you are advised to independently research a little about risk analytics (understanding the types of variables and their significance should be enough).

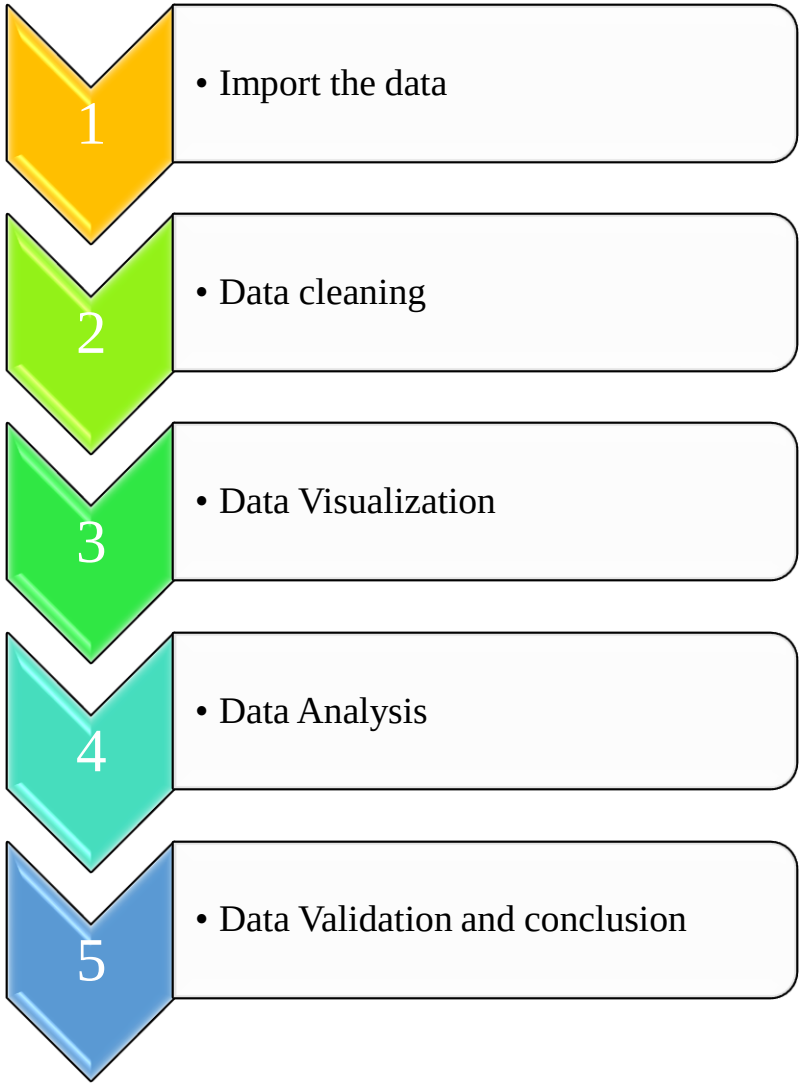
LendingClub

Case study objective:

1. Applying the techniques in EDA
2. Understanding of risk analytics in banking and financial services.
3. To minimize the risk of losing money while lending to customers.



Case study Work flow



Data Analysis

1. The data in the .csv format was imported to jupyter notebook and was further analysed for the missing data.
2. The inadequate columns were identified and filtered out from the main data frame – data cleaning.
3. Understanding the corelation between the given data sets.
4. Data analysis using univariate and bivariate analysis was conducted - EDA.
5. Data visualization of the derivatives.

Data cleaning

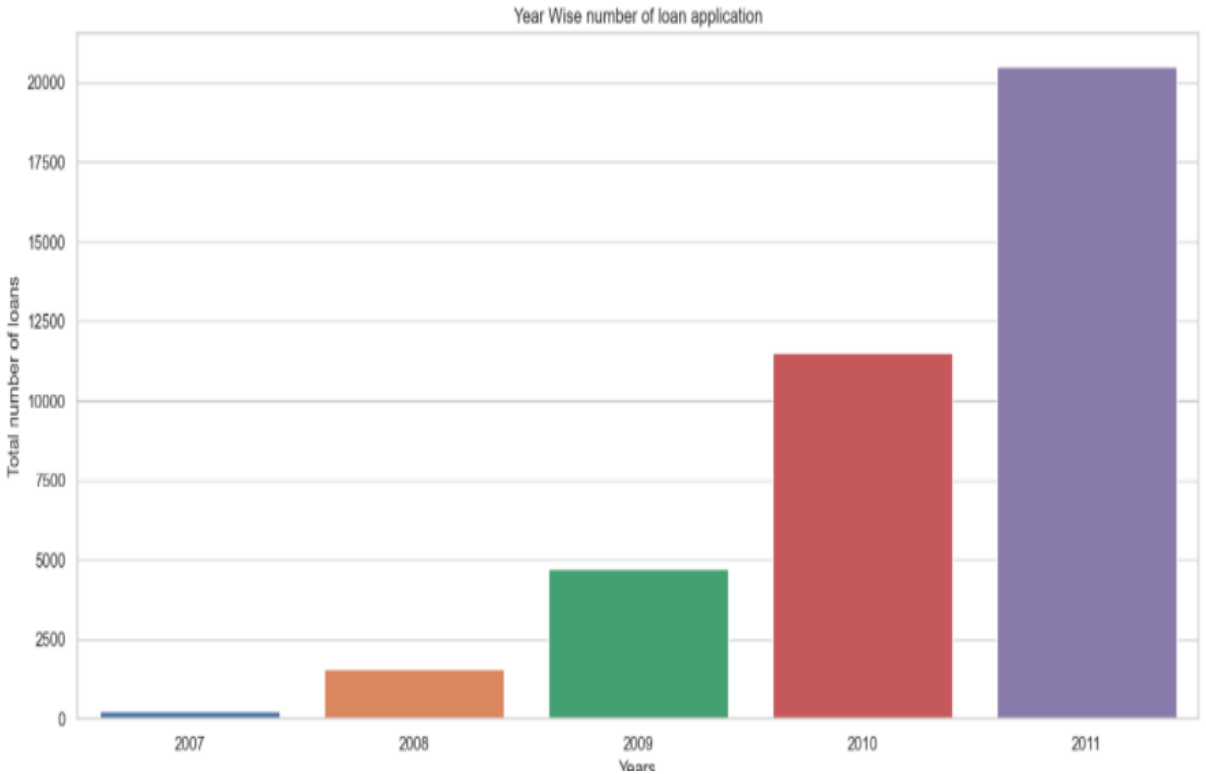
Data frame: loan_df

	Count	Percentage
mths_since_last_delinq	25682	64.662487
mths_since_last_record	36931	92.985372
next_pymnt_d	38577	97.129693
mths_since_last_major_derog	39717	100.000000
annual_inc_joint	39717	100.000000
dti_joint	39717	100.000000
verification_status_joint	39717	100.000000

The columns with 60% and above with NaN entries were sorted and deleted from the main data frame: loan_df

LendingClub

EDA

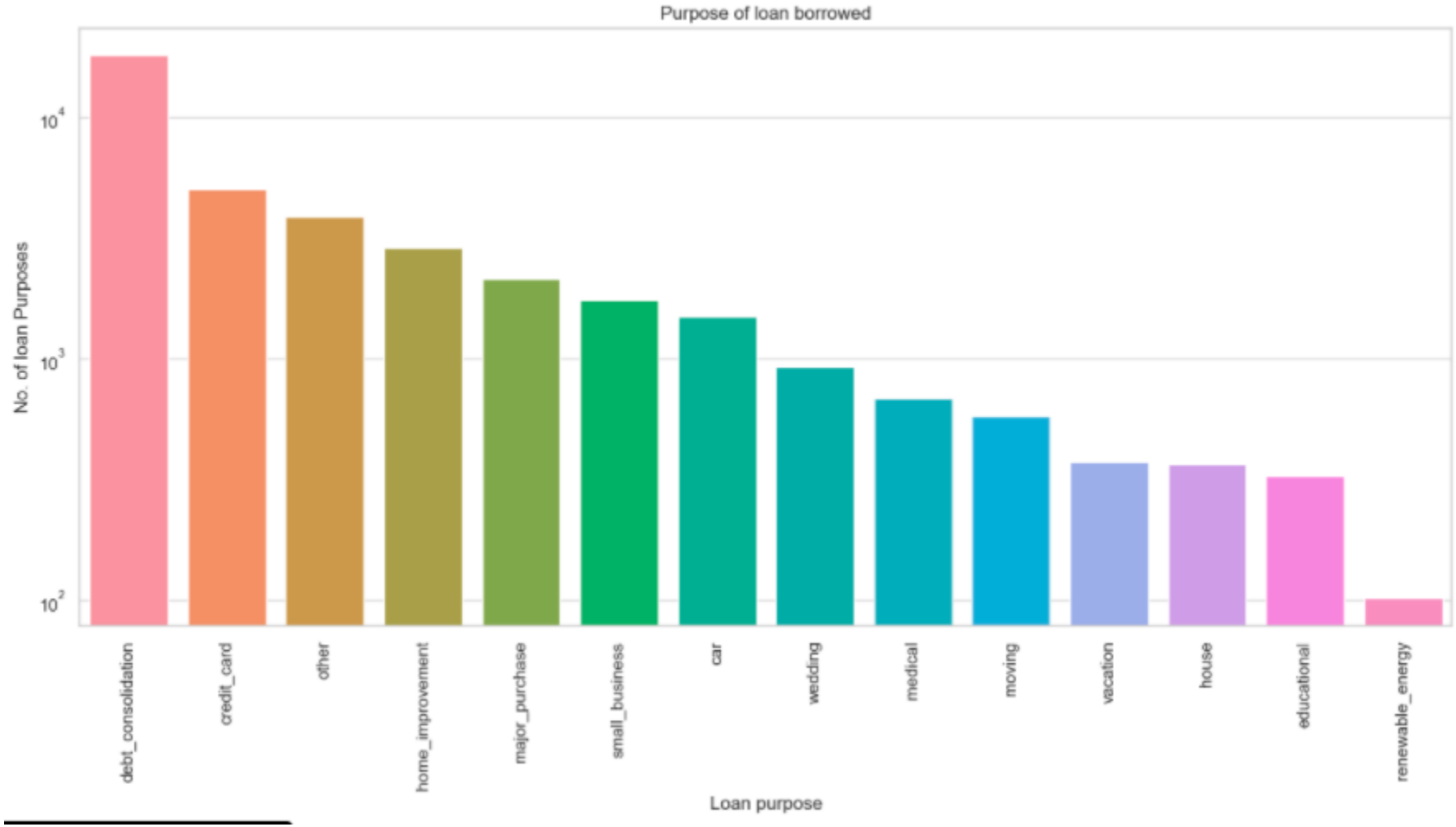


The loan application was found to be maximum for the year 2011. Hence, the data related to data related to 2011 was analysed.

Observation: Application >20000 was noticed

LendingClub

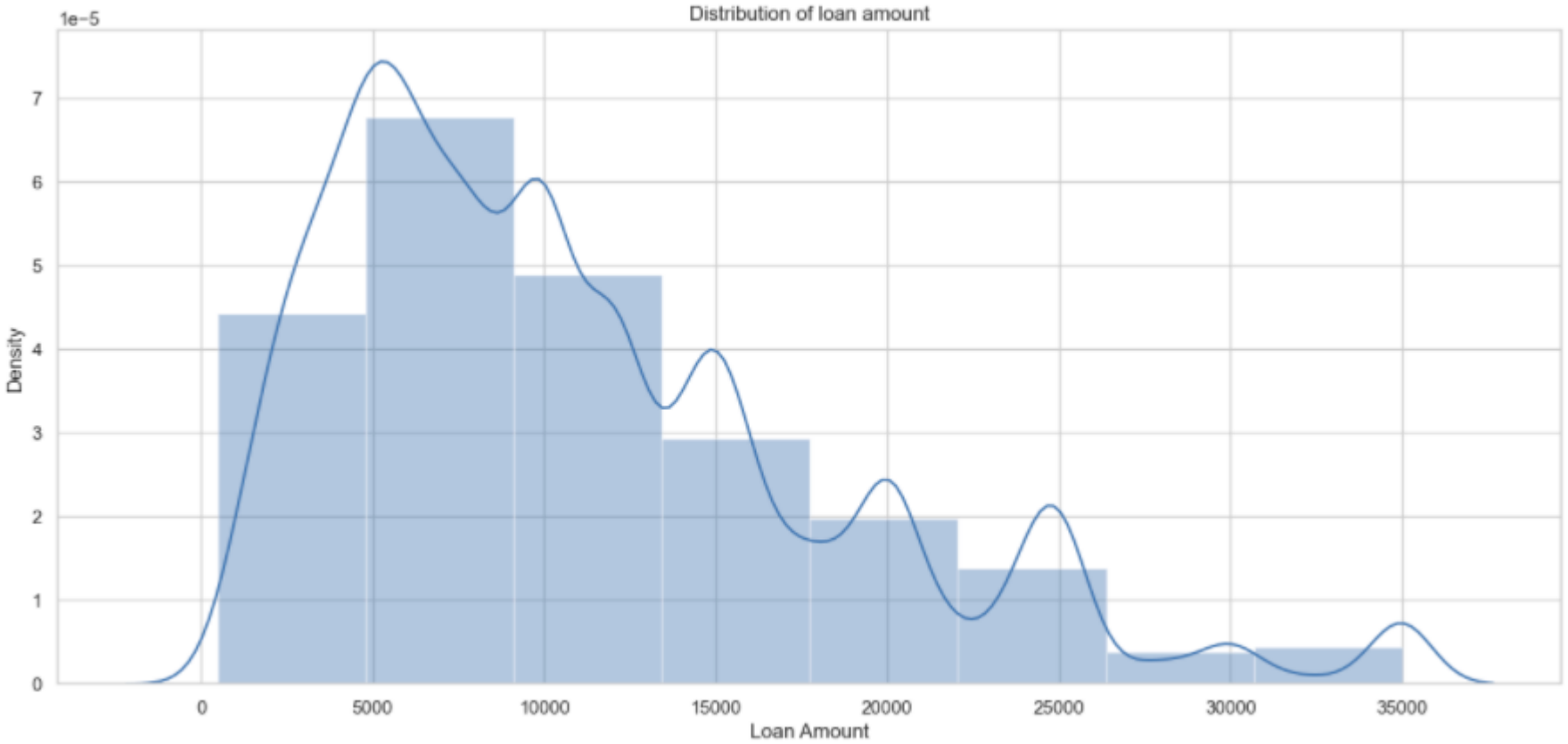
EDA



A plot of loan purpose and the number of application is analysed

Observation: Maximum loan purpose is for Debt Consolidation

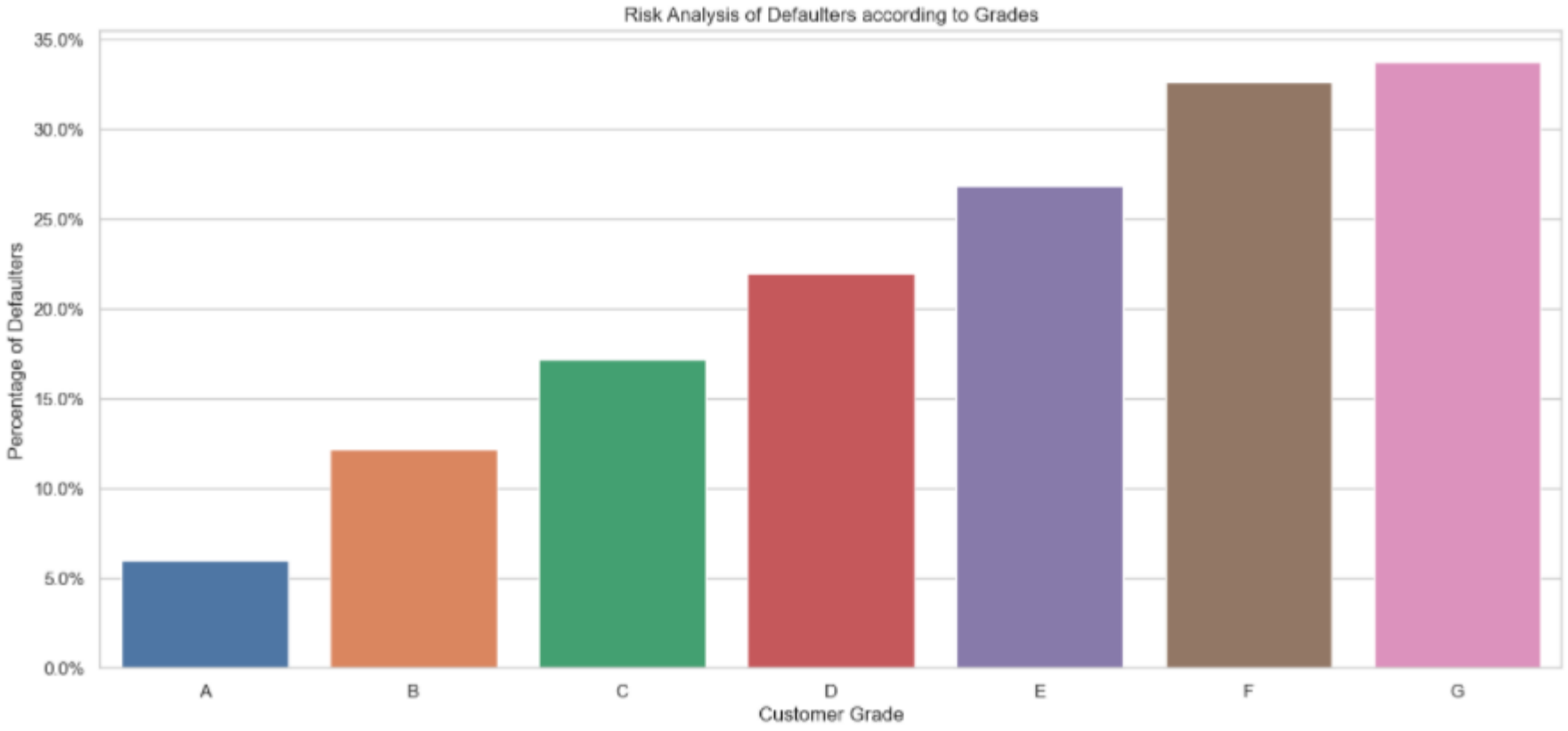
EDA



A plot of loan amount and the density of application is analysed

Observation: Most of the clients were looking for the loan amount ranging from $5000 to $10000

EDA

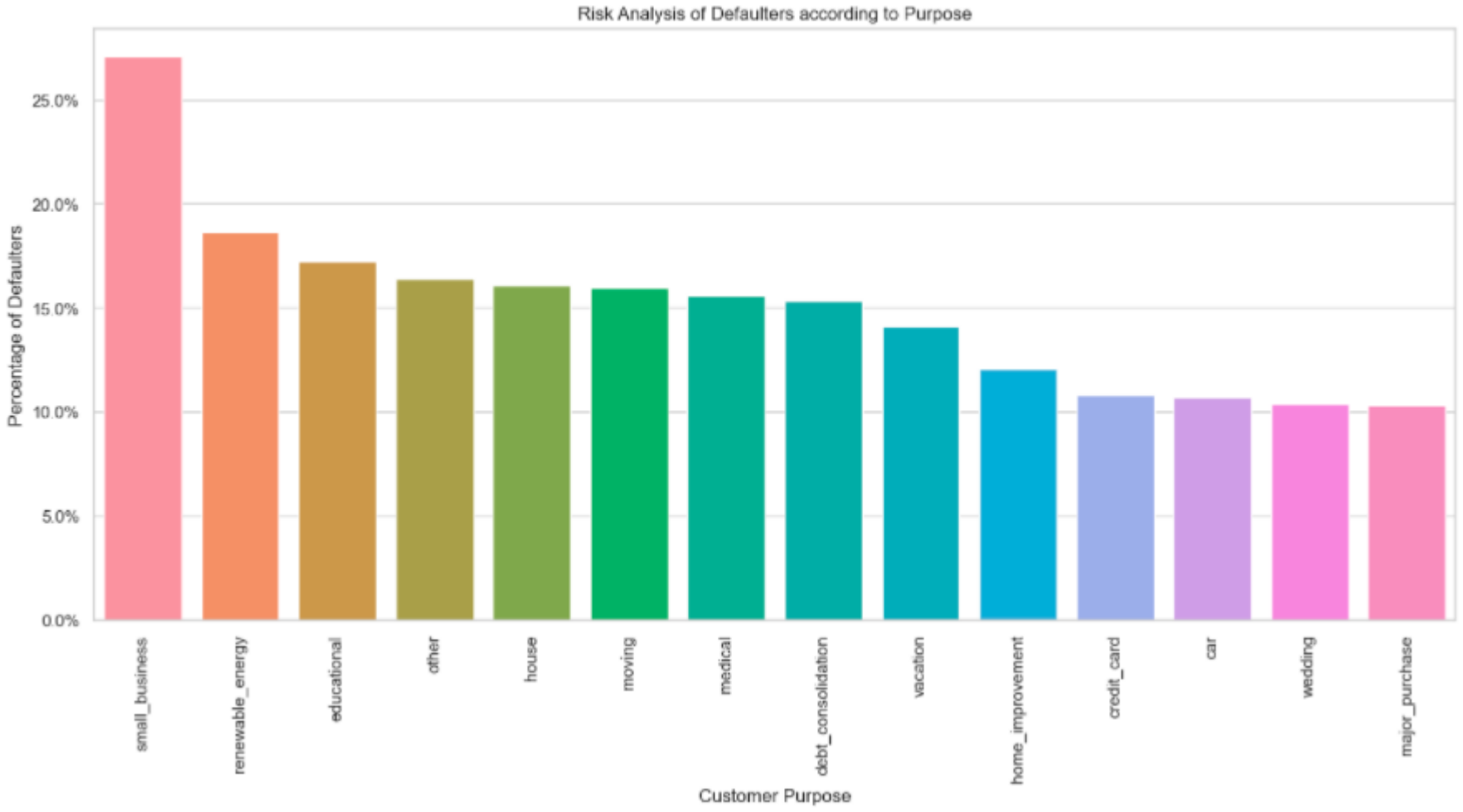


A plot of customer grade and percentage of defaulters is analysed

Observation: Highest number of defaulters were from Grade- G with 34%

Note: The grading is done based on the defaulters percentage.

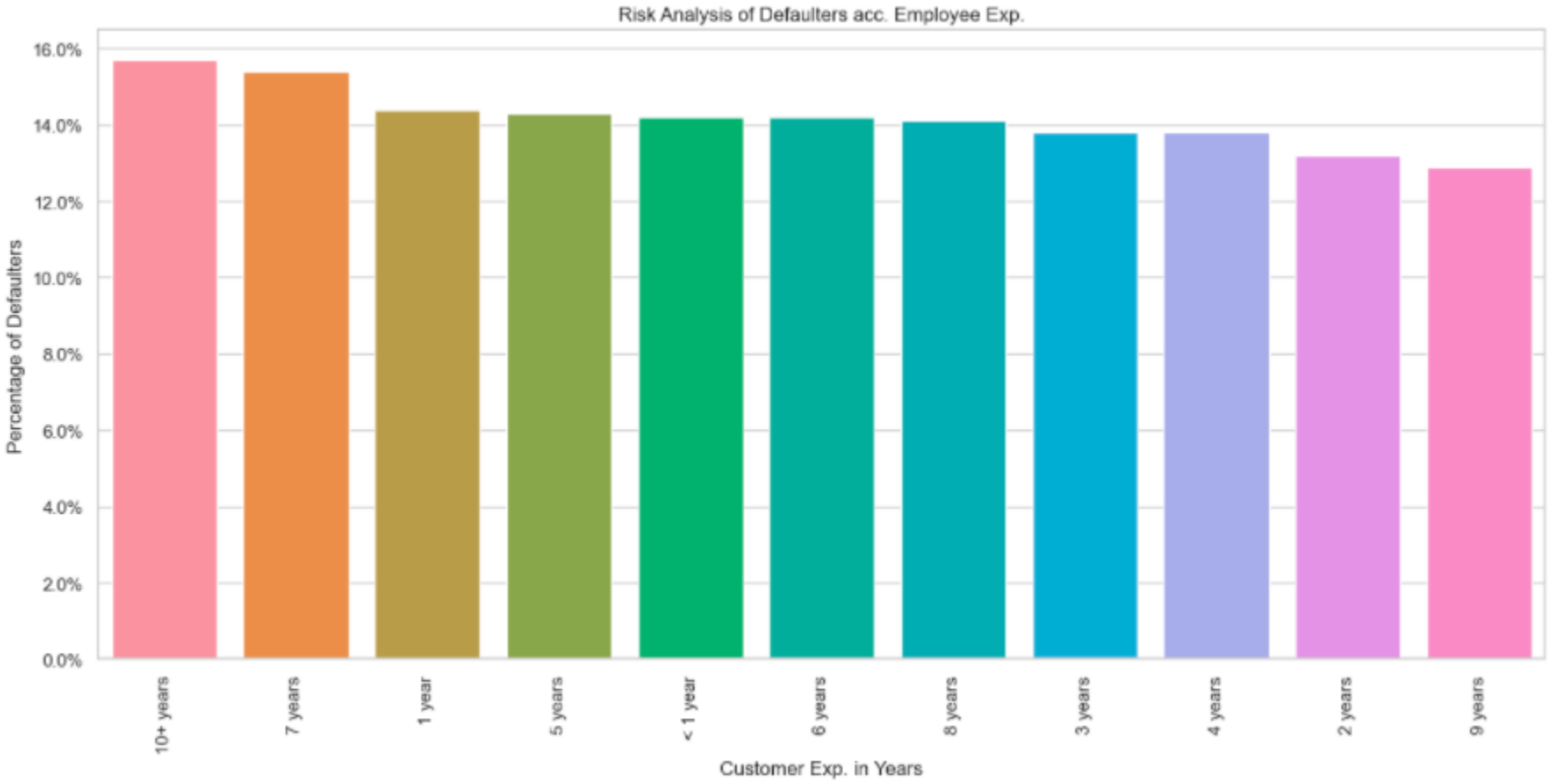
EDA



A plot of customer purpose and percentage of defaulters is analysed

Observation: Customer who opt loan for "small business" often exhibit high risk with 27.08%, followed by "renewable energy", 'education", "other" and so on as seen in the graph

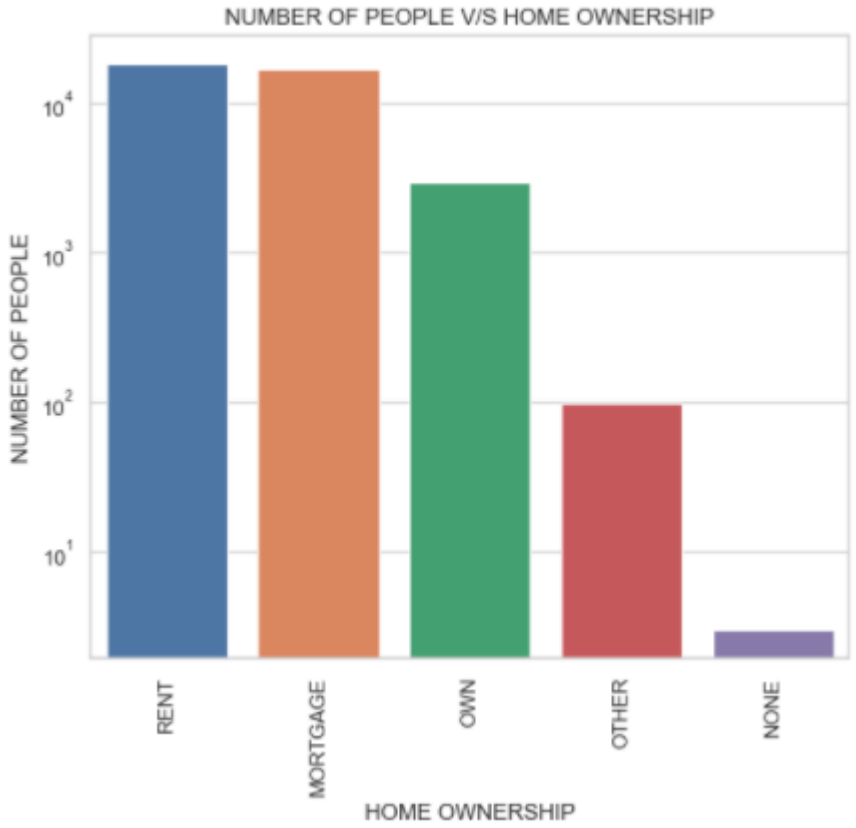
EDA



A plot of Risk analysis of defaulters according to employee experience is analysed

Observation: Employees with employee experience of 10+ year often have higher percentage of defaulting.

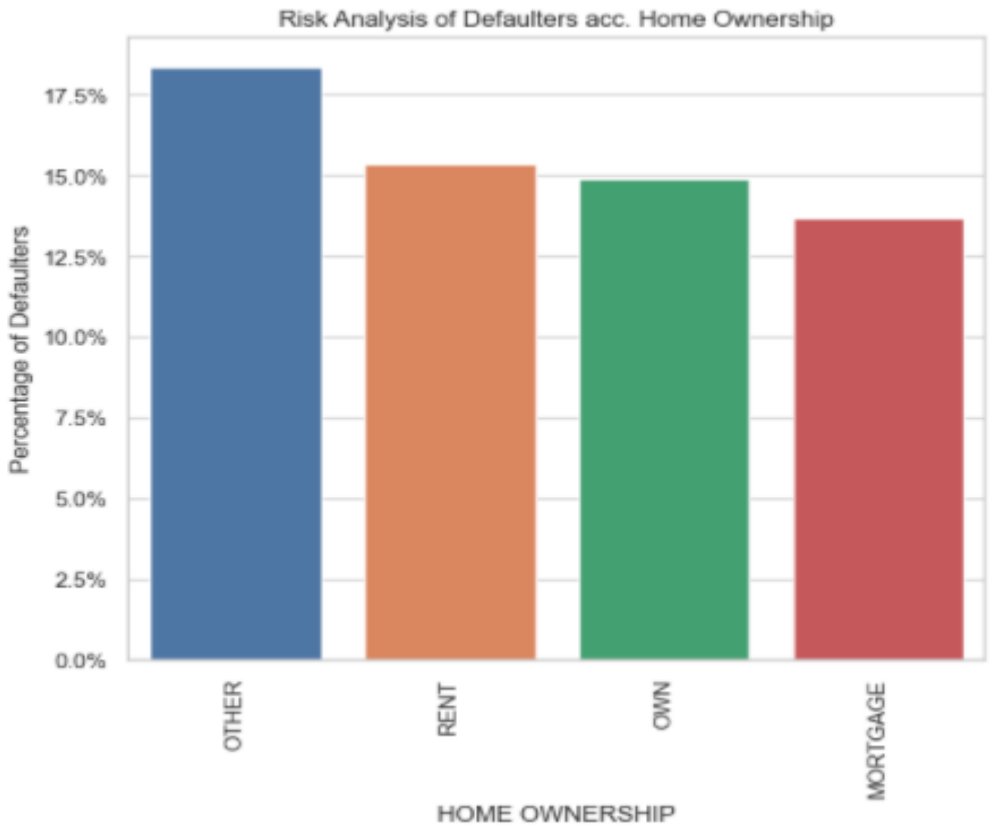
EDA



A plot of number of applications received vs home ownership

Observation: More applications were received from the clients staying in Rented property.

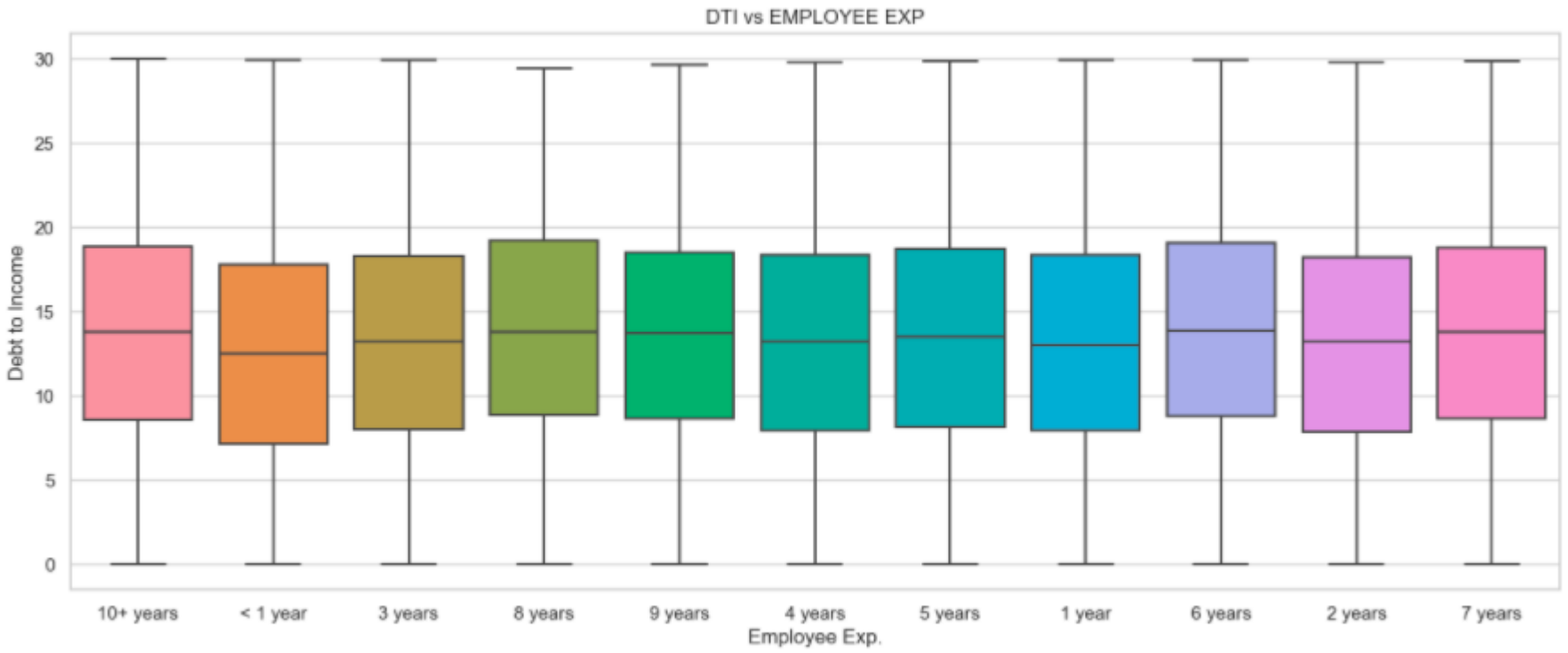
EDA



A plot of Risk analysis vs home ownership

Observation: People staying in rented community often exhibit high risk up to 15.4%

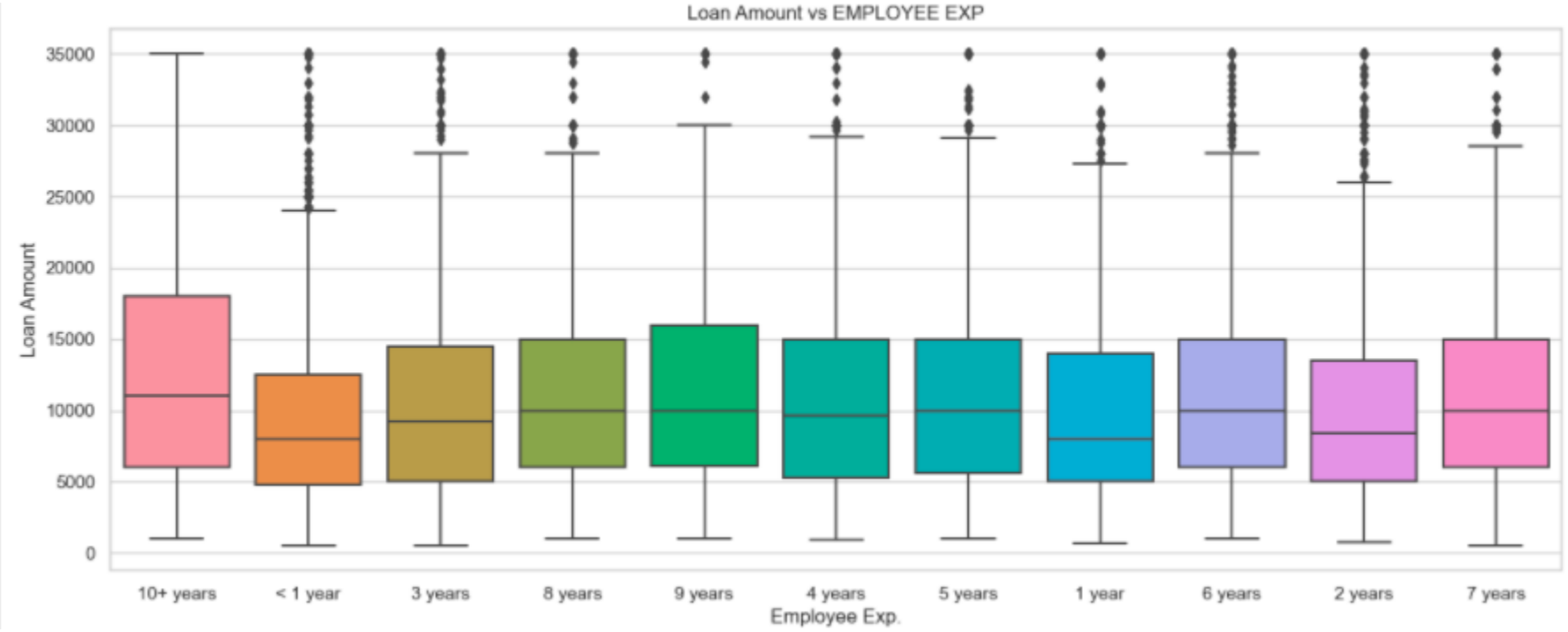
EDA



The plot shows the dti vs Employee experience.

Observation: Employee experience with 10+ years exhibit higher debt to income

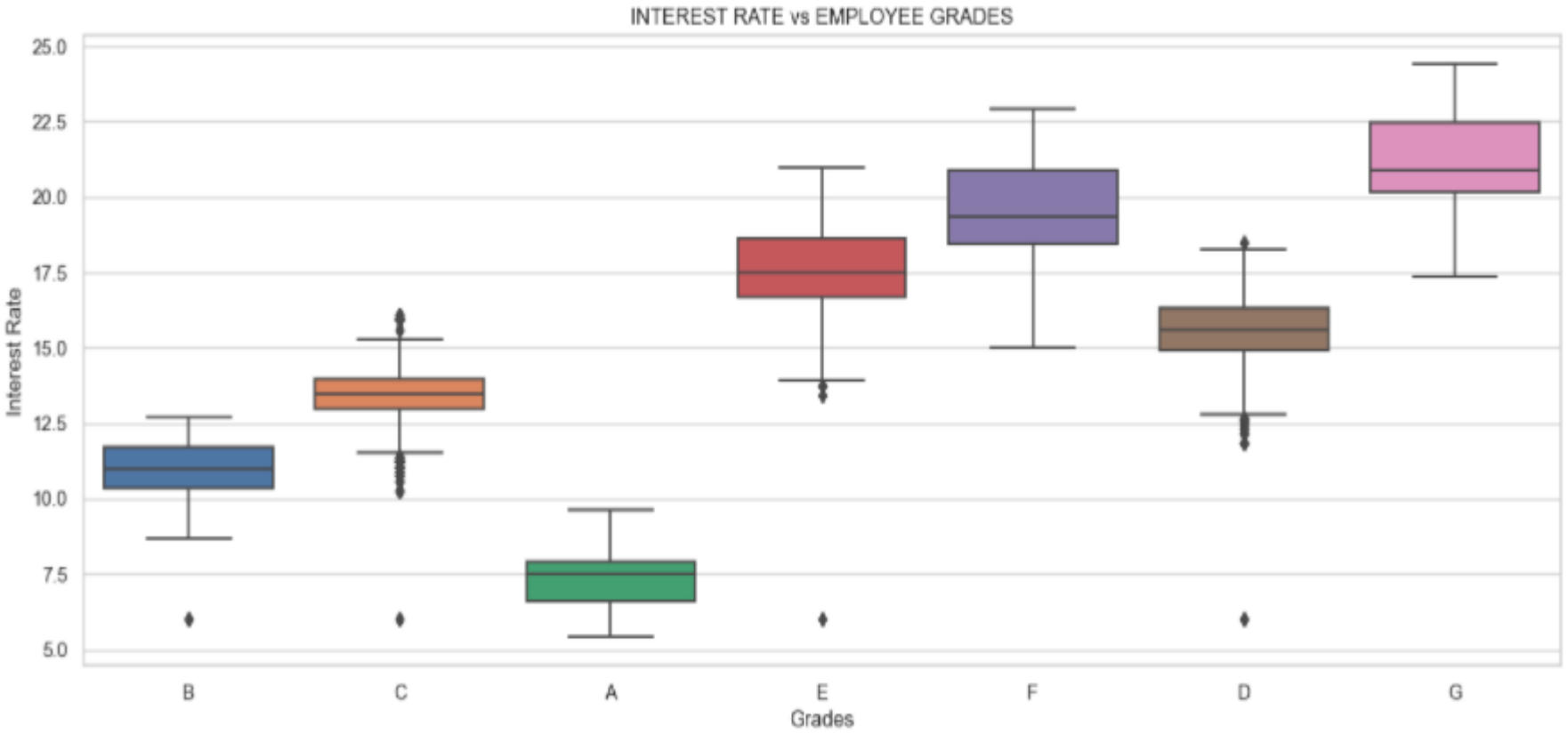
EDA



The plot shows employee experience vs the loan amout.

Observation: From the graph, it is evident that the work group of 10+ years of experience go for larger loan amount.

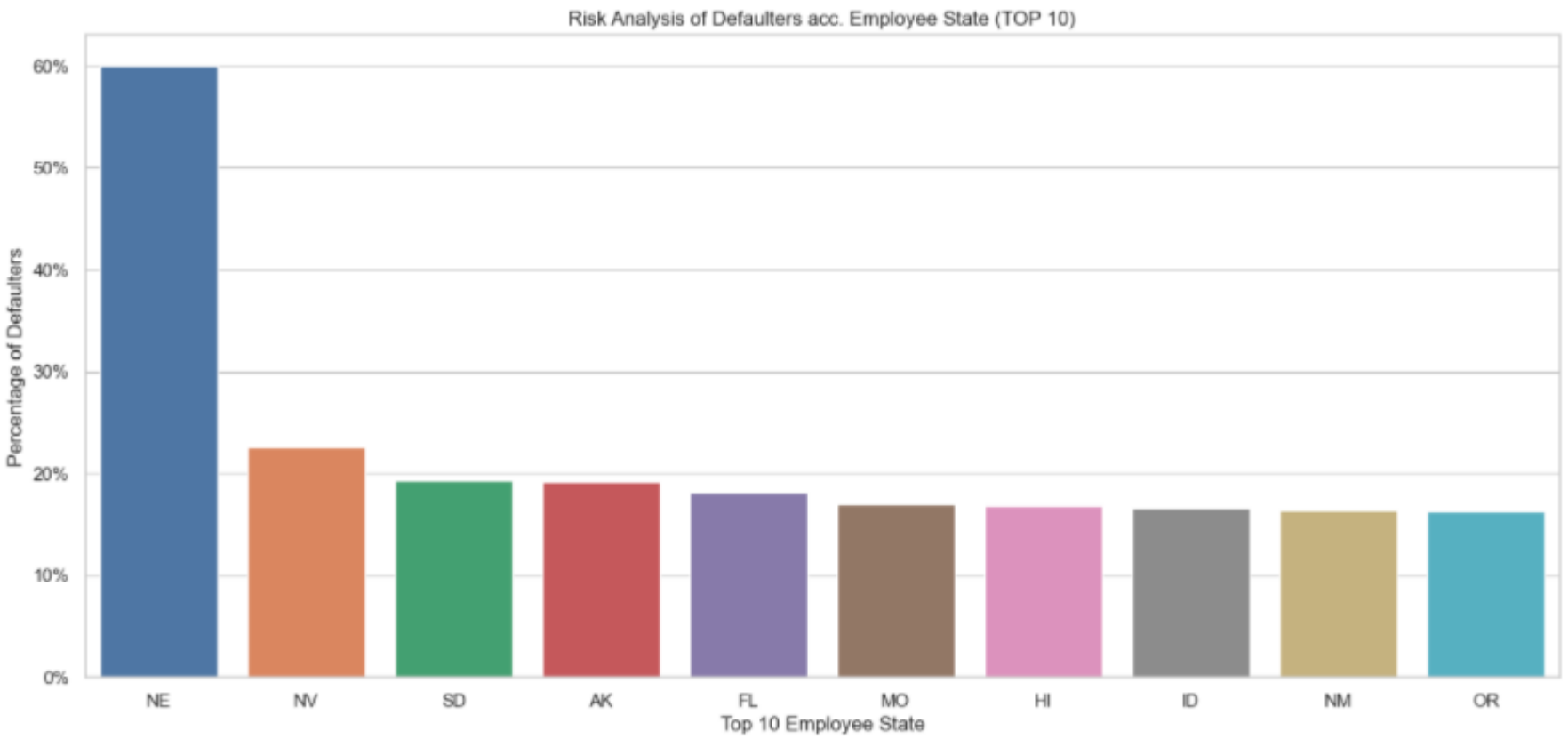
EDA



The plot shows Interest rate vs Employee grades.

Observation: Grade "G" has highest interest rate as they are high risk (defaulters) customers. Grade 'A' has lowest interest rate as they are low risk..

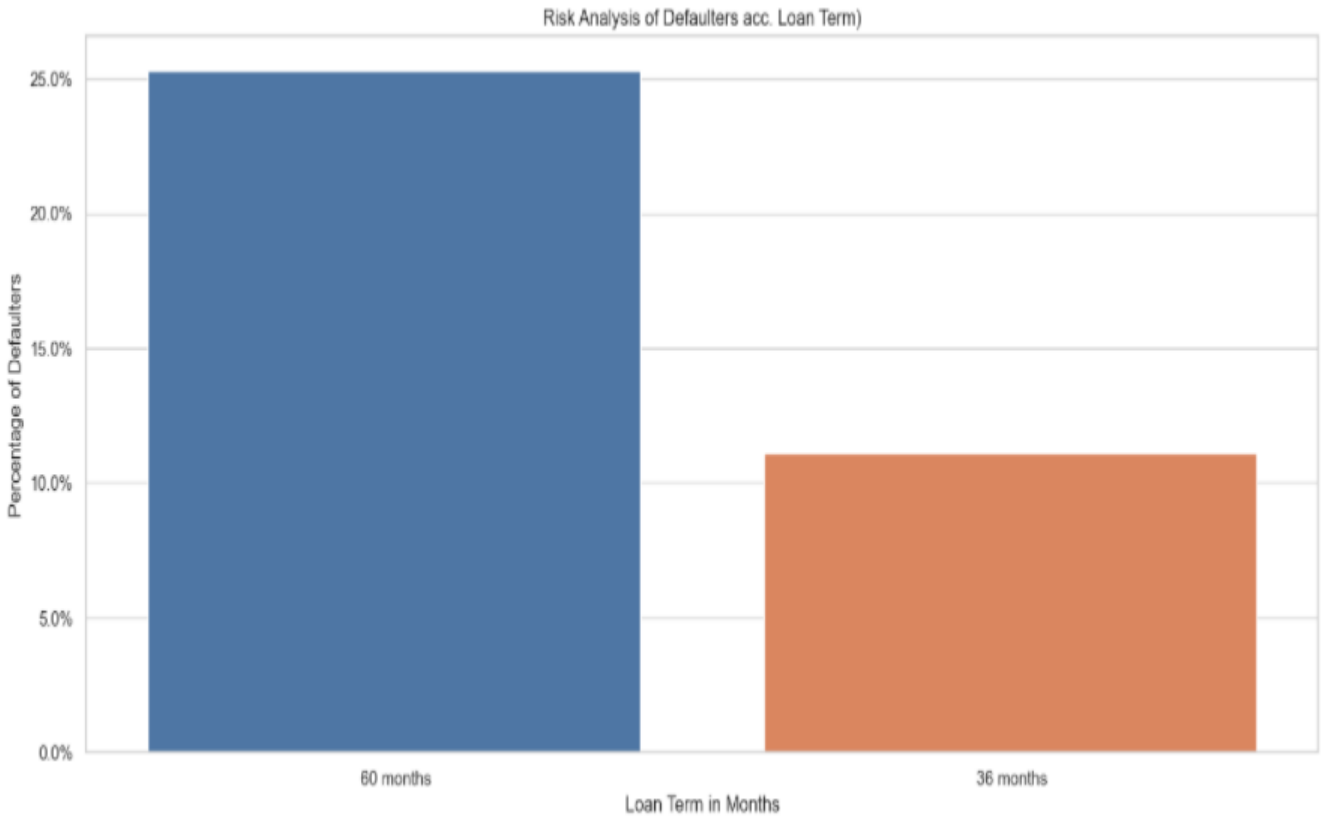
EDA



The plot shows risk analysis of defaulters according to employee state of residence

Observation: The loan application received from the location NE has high risk of 60% default.

EDA



The plot shows risk analysis of defaulters according to loan term

Observation: The loan application with loan tenure of 60 months have risk of 25% defaulter.

LendingClub

Observations

The following observations have high risk of loosing the money:

1. Loan application with longer tenure exhibit high risk
2. The loan application received from the location NE has high risk of 60% default.
3. Applicant with Grade "G" are high risk (defaulters) customers and hence are charged with higher interest rate so that the customer can deny the proposal, or the company can recover more money from such clients.
4. Employees with employee experience of 10+ year often have higher percentage of defaulting (15.7%) and such employees also have higher debt to income.
5. Application from residence type “Rent” are of higher risk with 15.4% defaulters.
6. Higher defaulters are seen for the loan applicants those have chosen the reason for debt_consolidation, which is more often noticed for employees with 10+ years of experience.

Conclusion:

The following conclusion are arrived based on the observation.

1. The loan application received from the location NE is to be rejected or needs thorough scrutiny.
2. Applicant with Grade "G" are to be rejected or needs thorough scrutiny.
3. Employees with employee experience of 10+ year are to be rejected or needs thorough scrutiny.
4. Application from residence type “Rent” are to be rejected or needs thorough scrutiny.

Technologies used:

Library	Vesrsion
notebook	6.5.2
matplotlib	3.6.2
notebook_shim	0.2.2
numpy	1.24.1
openpyxl	3.0.10
pandas	1.5.2
numpy	1.24.1
plotly	5.13.0

Acknowledgment

I would like the extended support of the following esteemed organizations and members.

1. LendingClub
2. M/s Upgrad
3. Faculty members of IIIT-B
4. Supporting team

LendingClub

Thank You