Lending Club Case Study

by Pakki Harsha Prakash & Saikiran Andey

Problem Statement

Consumer finance company 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.

We are analysing the data for a consumer finance company which specialises in lending various types of loan amounts to urban customers.

We as Analyst, should help the consumer finance company to identify a pattern which indicates if a person is likely to 'default' or not .

Data Exploration

Brief discussion about obtained data :

> RangeIndex: 0 to 39716
> Columns: 111 entries
> Data types: float64(74), int64(13), object(24)

Missing Data*: It is observed that 'mths_since_last_delinq' column has ~60% missing values and 'Desc' has ~30% missing values.

*since the missing data is more than 25% of the given data we are dropping these respective columns in order to generate better insights
*Also note that defaulted loans have been marked as 1 and full paid loans as 0 in the analysis. Hence all loan_status plots are default_rate plots.
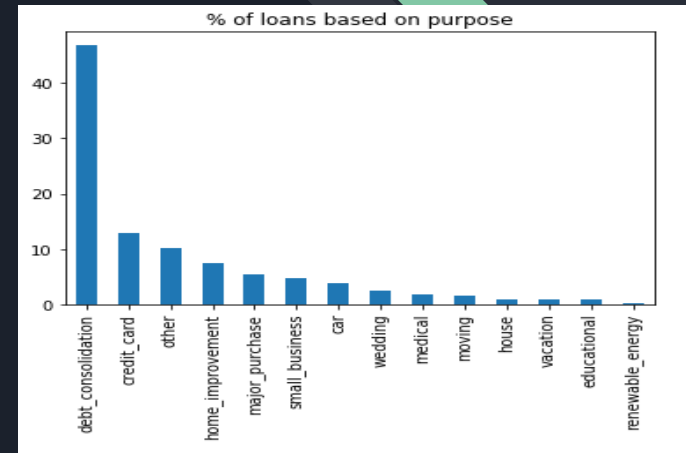
Univariate Analysis

We performed a univariate analysis on the loan status to understand the diversity of customers and purpose

> We observed that approximately 83% of the customers have fully paid their loan.
> We observed that majority of the customers prefer loan for debt_consolidation
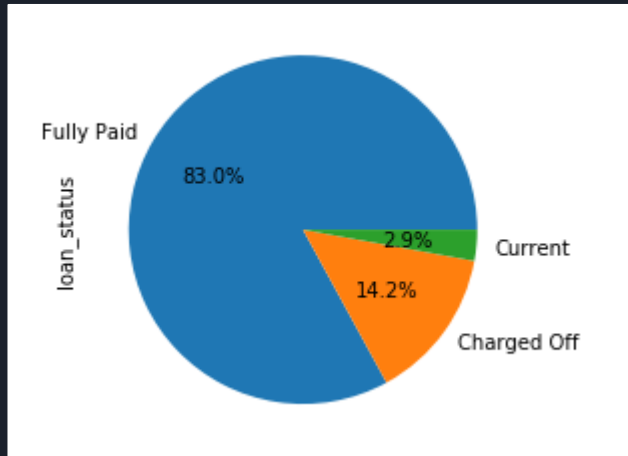
Univariate analysis on 'Loan_Status'

Univariate analysis on 'Purpose'

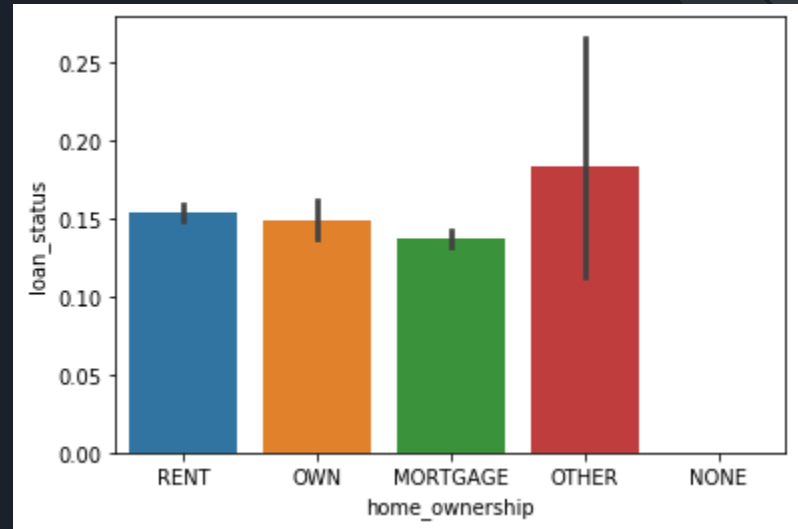# Analysis of default_rate against different categorical variables:

Description: Analysis performed against term duration and default_rate, we observe that it is higher for 60 month duration than 36 month duration
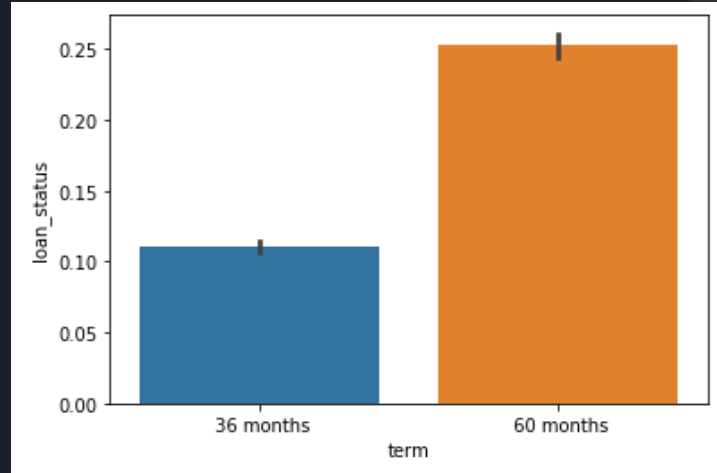
Default rate vs term duration

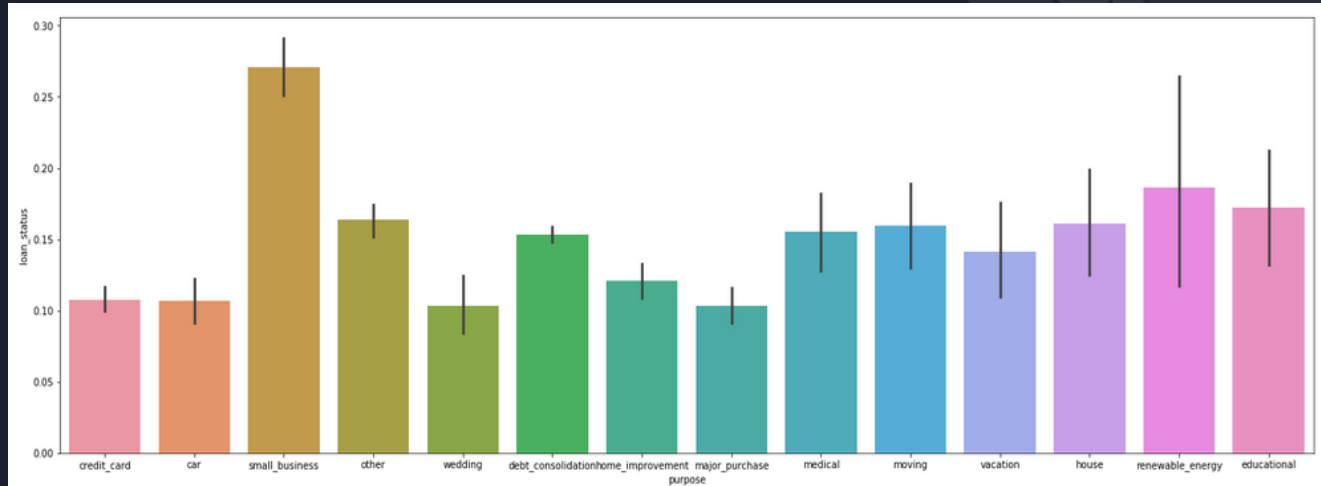# Analysis of default_rate against different categorical variables:

Description: Analysis performed against home ownership and default_rate, we observe that it is higher for others and rent

Default_rate vs home_ownership

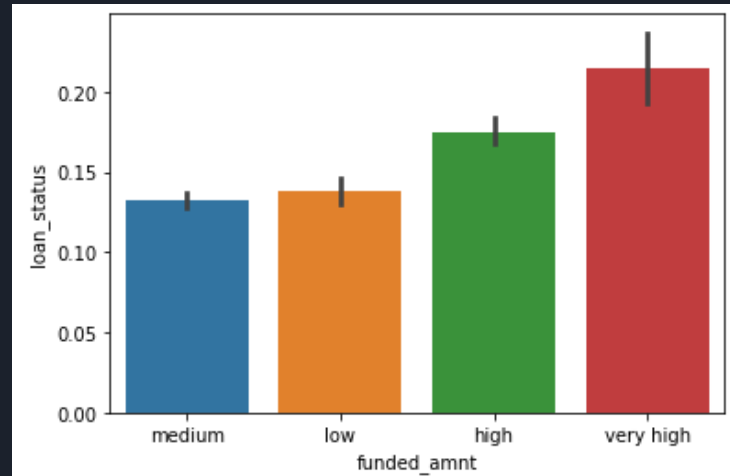# Analysis of default_rate against different categorical variables:

Description: Analysis performed against purpose and default_rate, we observe that it is higher for small business which in turn signifies that business related sectors are an important priority for classification

Default_rate vs Purpose

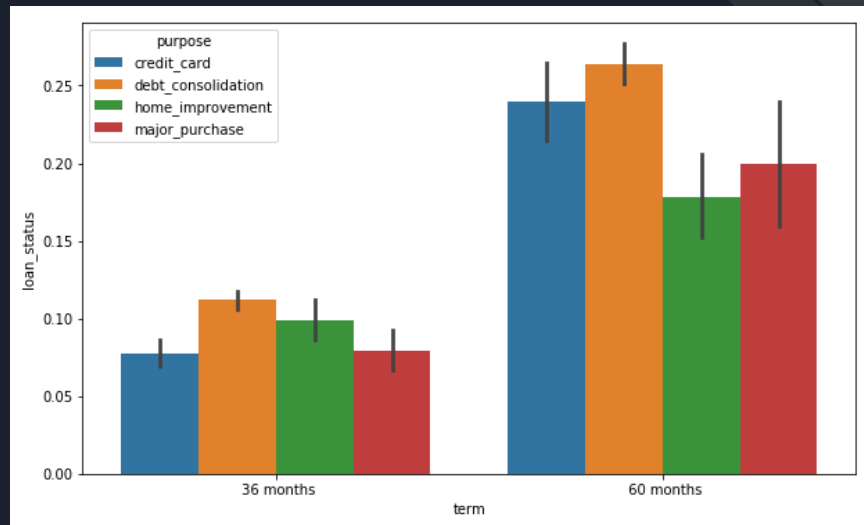# Analysis of default_rate against different categorical variables:

Description: Analysis performed against funded_amnt and default_rate, upon classification majority of the loans are registered for amount higher than 25000.

Default_Rate vs Loan_status

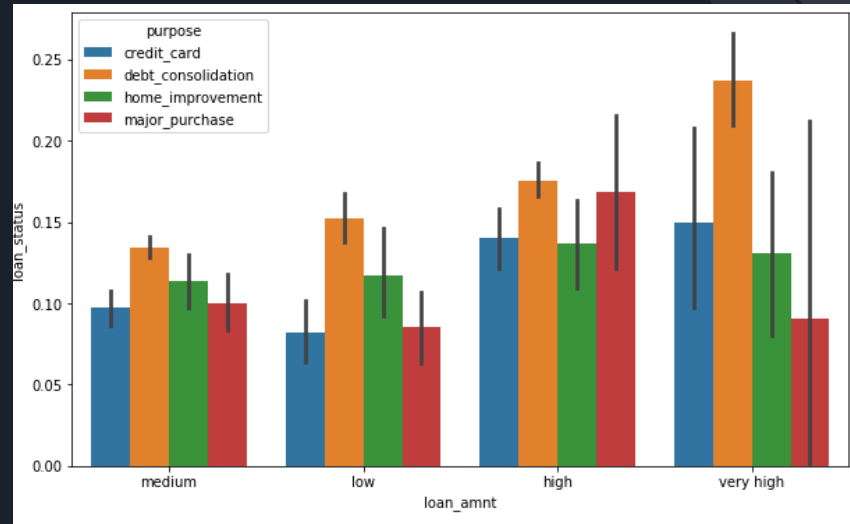# Plots for default rate segmented for different variables

Description: Segmented analysis performed against funded_amnt and term,which shows similar trends in both 36 month duration and 60 month duration.

Default_rate vs term

Plots for default rate segmented for different variables

Description: Segmented analysis performed against funded_amnt and loan amount,which shows similar trends in all the segments.

Default_Rate vs loan amount

Plots for default rate segmented for different variables
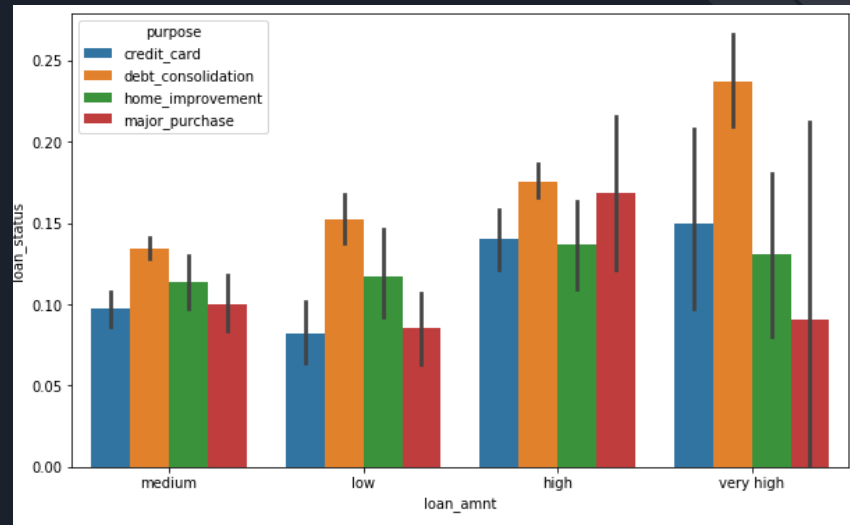
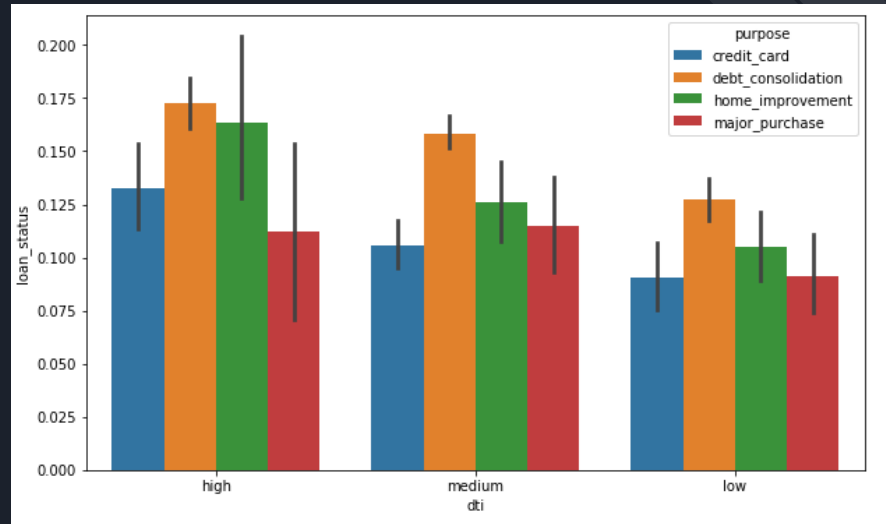Description: Segmented analysis performed against funded_amnt and loan amount,which shows similar trends in all the segments.

Default_Rate vs loan amount

Plots for default rate segmented for different variables

Description: Segmented analysis performed against funded_amnt and dti,which shows similar trends in all the three categories.

Default_Rate vs dti

Analysis insights and Observation

> It is observed that irrespective or the categories, all the segments which are prioritized follow a similar trend
>The default rate is higher for long term loans and customers opted for long term duration are more than short term duration.
>One interesting insight is that house ownership does not impact the default rate in customers.
>Another exception is that verified customers were defaulting more than unverified customers.
>Also, we can observe that small businesses and educational loans are the most defaulted, although they do not contribute as much to the revenue of Lending Club.
>The default rate is also directly proportional to the monthly emi payment , DTI of the person and salary of the person.

Thank You