

Lending Club Case Study

Ritesh Lath

20-Nov-24

Problem Statement & Expected Result

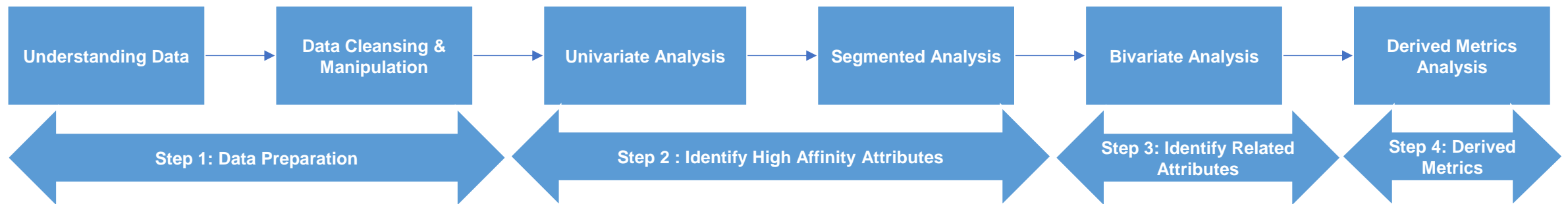
Problem Statement

The largest online loan marketplace company facilitating in loans need to reduce its financial losses due to defaulting borrowers.

Approach

Perform Exploratory Data Analysis (EDA) on the loans information and identify the features that can help to identify risky customers and thus potentially reduce loan defaults.

EDA Process



1. Perform Univariate Analysis by Percentage Of Charge Off Loans
2. Perform Bi-variate Analysis By Loan Amount (Concentration Of Risk)

Expected Result

A list of features that can be used to identify high risk customers

Step 1: Data Preparation

Data Set Statistics

No. Of Rows: 39717

No. Of Attributes: 111

No. Of Attributes (All Nulls) : 54

Data Set Cleaning & Standardization

- Remove all unwanted attributes (drop all the attributes for which there is no data)
- For Surviving Attributes
 - Drop if not of significance
 - Fill Missing Values Or Standardize it

Sr. No	Attribute Name	Action Taken
1	Publicly Recorded Bankruptcies	No Action taken
2	Months Since Last Record	No Action since sourced publicly
3	Last Payment Date	No Action taken
4	Next Payment Date	No Action taken
5	Publicly Recorded Bankruptcies	No Action taken
6	Months Since Last Record	No Action since sourced publicly
7	Loan Description	Dropped as it a very verbose
8	Collections In Past 12 Months Excluding Medical	Dropped since zeros or missing
9	Number of charge offs in last 12 months	Dropped since zeros or missing
10	Tax Liens	Dropped since all are zeros
11	Number of accounts narrower is delinquent	Dropped since all are zeros
12	Application Type	Dropped since all are 'Individual'
13	Delinquency Amount	Dropped since all are zeros

Sr. No	Attribute Name	Action Taken
14	Employment Title	Missing Values Filled With "Unknown"
15	Employment Length	Missing Values Filled With "Unknown"
16	Months Since Last Delinquency	Missing Values Filled with '0'
17	Interest Rate	Standardized to float removing %sign
18	Revolving Credit Utilization Rate	Missing Values filled with '0' Removed row with revolving balance of 49238 (Fully Paid) Standardized to float removing '%' sign
19	Last Credit Pull Date	Removed row where loan status is Fully Paid for missing value record Standardized to Date
20	Loan Amount Bucket	Derived Column in Steps of 5K
21	Monthly Income Bucket	Derived Column in Steps of 10K
22	Monthly Income	Derived Column from Annual Income

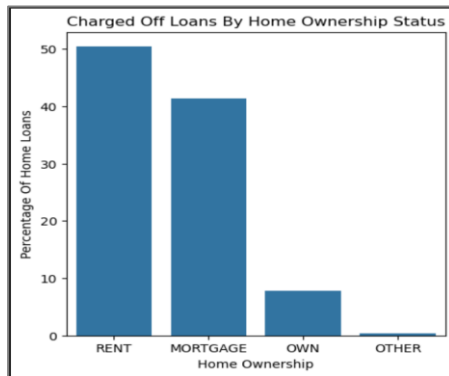
Step 2: Univariate & Segmented Analysis

Univariate and Segmented Analysis Results

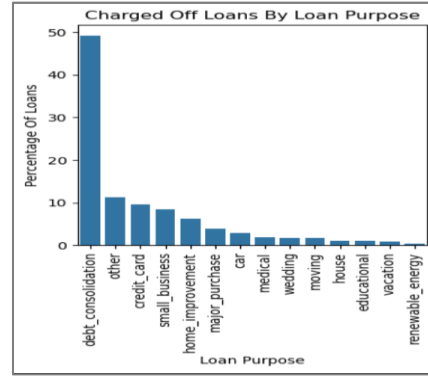
Below Attributes Are High Affinity Attributes:

- Home Ownership (Rent & Mortgage)
- Purpose Of Loan *Debt / Credit Card
- State Of Customer (CA / FL / NY / TX / NJ)
- Employment Length
- Monthly Income
- Loan Grade
- Loan Amount
- Last Credit Pull Date (Recent Credit Pull Date)

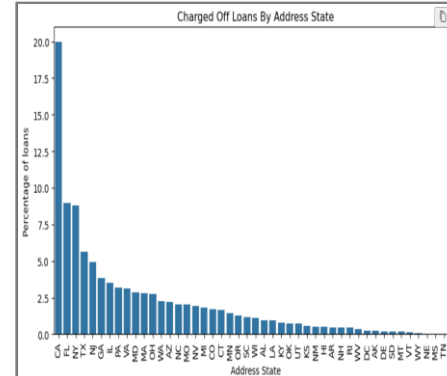
Home Ownership



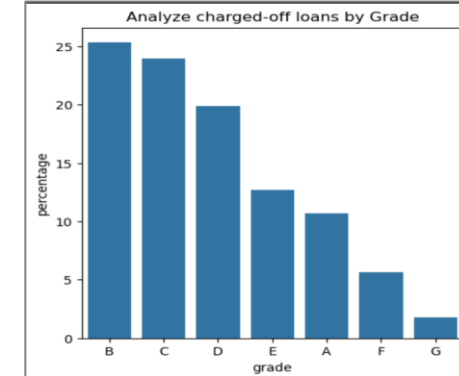
Purpose



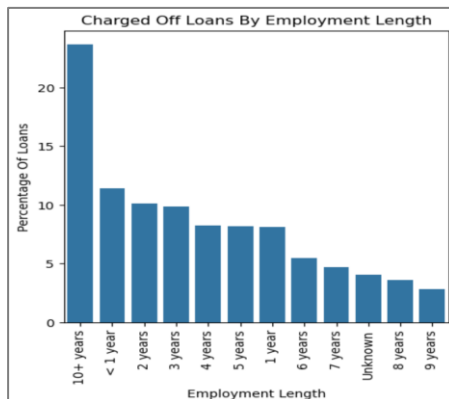
Address State



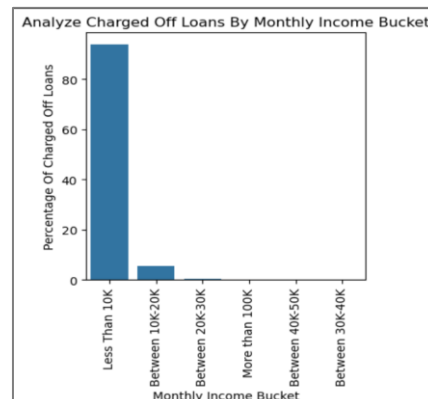
Loan Grade



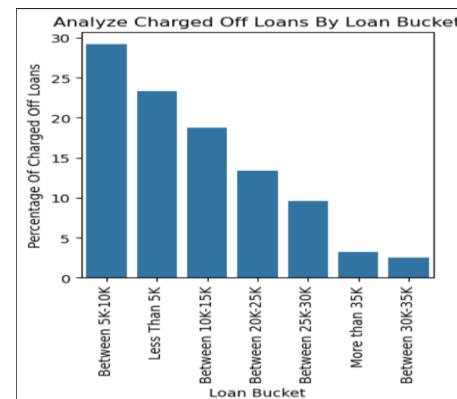
Employment Length



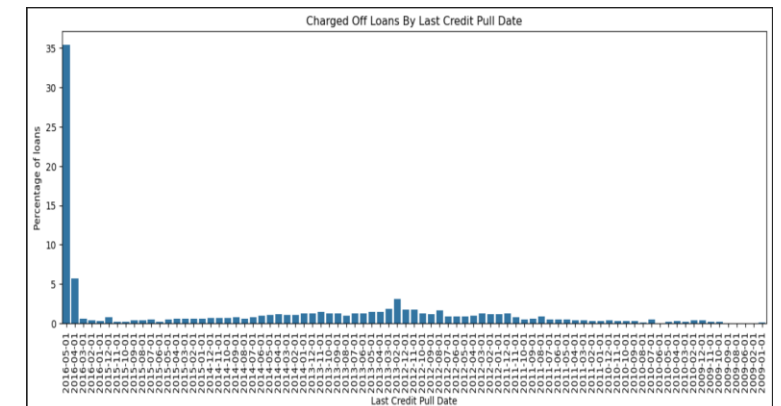
Monthly Income



Loan Amount



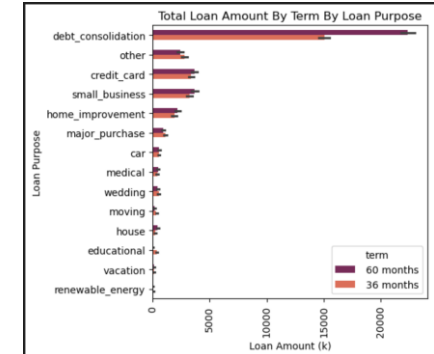
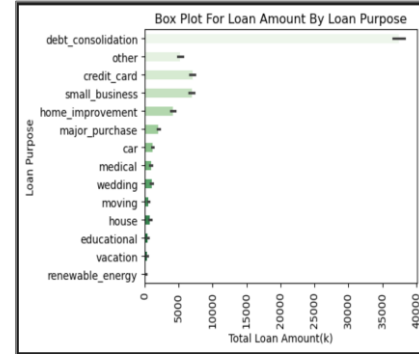
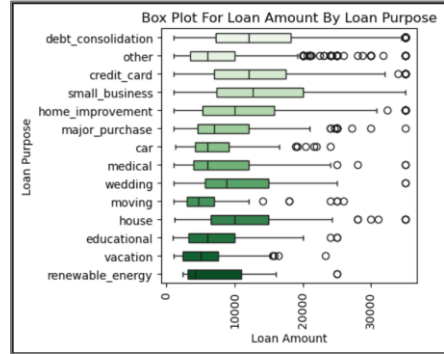
Last Credit Pull Date



Step 3: Bivariate Analysis (Purpose, Home Ownership, Address State)

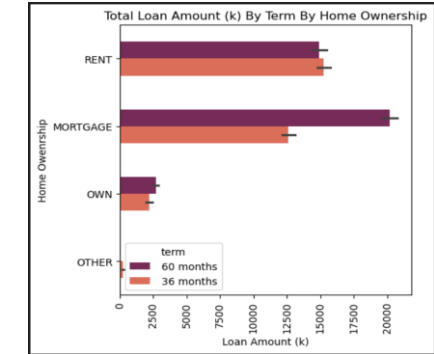
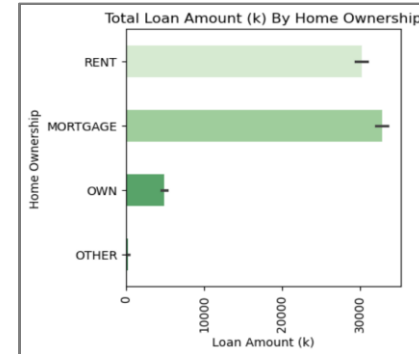
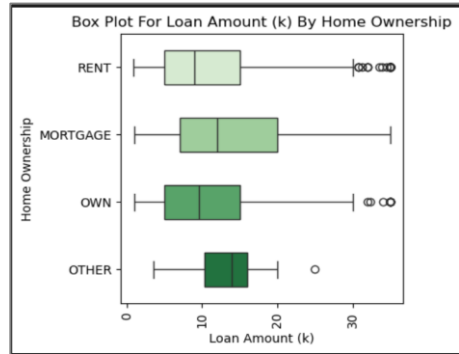
Loan Purpose

- Debt Consolidation
- Credit Card
- Small Business



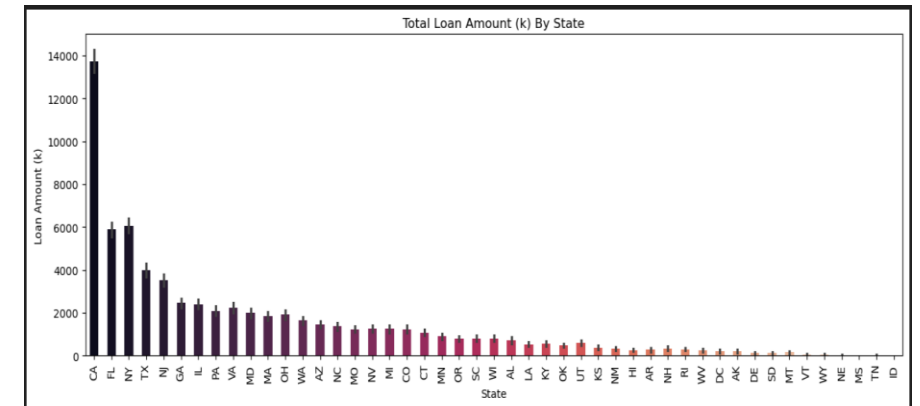
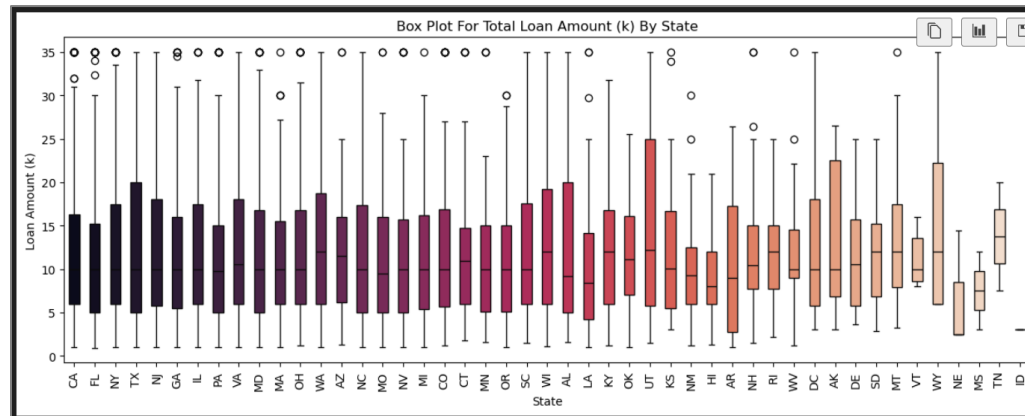
Home Ownership

- Rent
- Mortgage



Address State

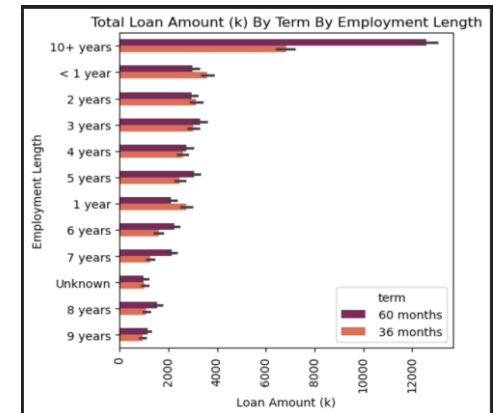
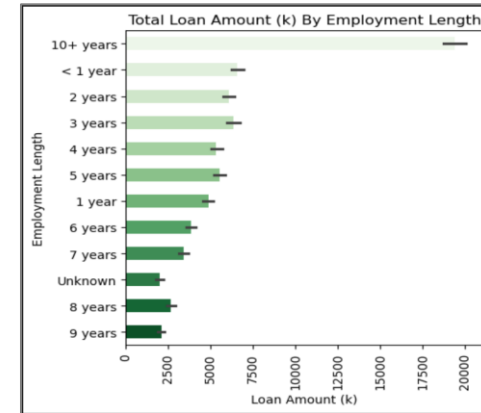
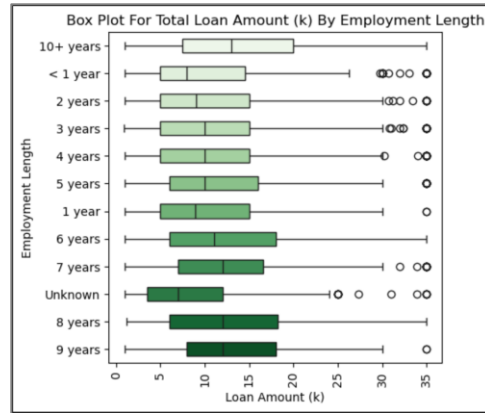
- CA
- NY
- TX
- FL
- NJ



Step 3: Bivariate Analysis (Employment Length, Monthly Income Bucket)

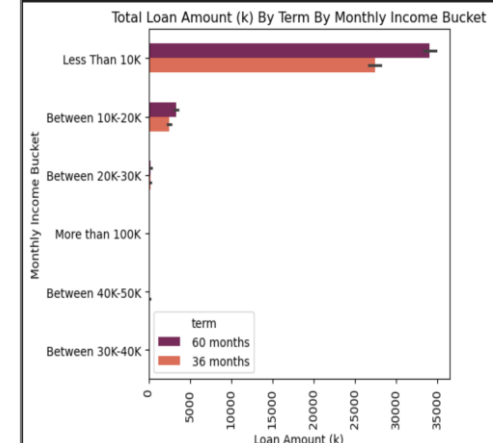
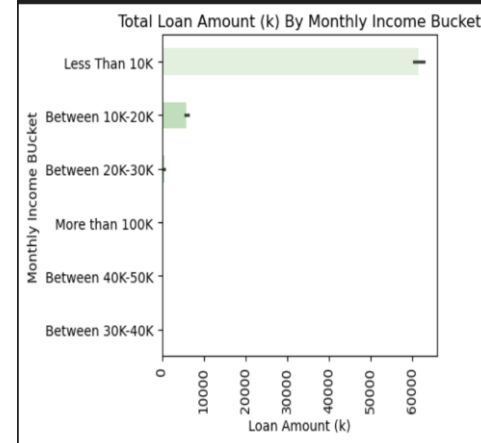
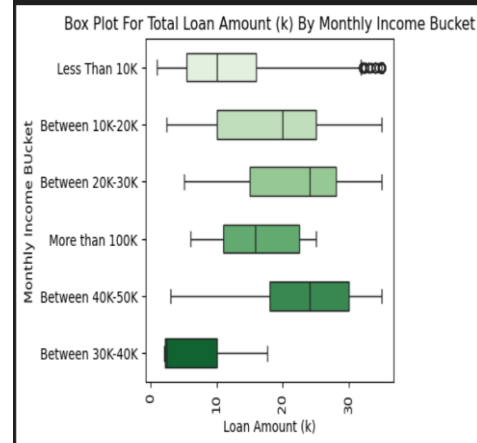
Employment Length

- 10 + years
- <1 year
- 2 year
- 3 year
- 4 year



Monthly Income

- Less Than 10K



Conclusion

As per EDA findings the below attributes and its values point to a potential risky customer:

- **Address State:** CA, NY, TX, NJ, FL
- **Monthly Income:** Less Than 10K (Granular affinity is between Less Than 5K)
- **Home Ownership:** Rent or Mortgage
- **Purpose Of Loan:** Debt Consolidation, Credit Card
- **Employment Length:** More Than 10 years