IIIT B / Upgrad - Lending Club Case Study

General Information:

- Analyst Name : Prashant Lotlikar & Amit Kumar
- Date Started : 02 Aug 2023
- Completion Date : 9 Aug 2023
- Data Set : As defined in Course link : https://learn.upgrad.com/course/4622/segment/38490/225106/688517/3483320
- Assignment type : EDA

Problem Statement

- Given the Loan data set for a Consumer finance company, the objective of the assignment is to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

- The company wants to understand the driving factors (or driver variables) behind loan default (loan_status = 'Charged Off'), i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Step 1 - Data Cleaning

- ➔ Downloading data set into a dataframe
- ➔ Understanding dataset and its key variables
- ➔ Identify and filter out key data columns
- ➔ Data Cleaning - Missing values, duplicates, data formatting errors, find synonyms, rename columns into meaningful labels,
- ➔ Creating new Columns b breaking up data into separate columns , data fields , etc.
- ➔ Identifying possible human errors in data set (if any)
- ➔ Deleting unwanted rows and columns
- ➔ Removing further columns basis on Domain knowledge. Like, removing the customer behaviour variables as these are not available at the time of loan application, and thus they cannot be used as predictors for credit approval. [As per feedback from UPGRAD live session on 6 Aug 2023]
- ➔ Few columns like int_rate, term and emp_length contains text. Convert into int/float for analysis

Conclusions - Step 1 - Data Cleaning

- Initial dataset size was (39717, 111)
- After Data Cleaning dataset size was (39717, 24)

Step 2 - Exploratory Data Analysis

- ➔ Identified "**loan_status**" as target variable
- ➔ Filtering only fully paid or charged-off. Converting to boolean and focus on " 1 --> Charged Off" for further analysis
- ➔ Identify columns to perform Univariate analysis.
- ➔ Perform Univariate analysis and look for patterns and report inferences if any. Ref.
- ➔ Perform Segmented Univariate analysis and look for patterns and report inferences if any. Ref.
- ➔ Perform Bivariate analysis and look for patterns and report inferences if any. Ref.
- ➔ Perform Correlation analysis

Conclusions - Step 2 - Exploratory Data Analysis

- Identified "loan_status" as target variable

Conclusions - Step 2 - Exploratory Data Analysis

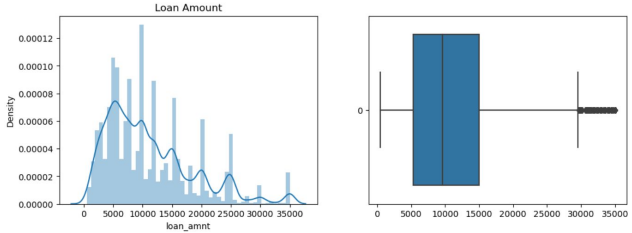
Perform Univariate analysis and look for patterns and report inferences if any. Ref.

1. Loan Amount Observations:

- Median of distribution is 10000

- 50 % people have taken loan between 5000 and 15000

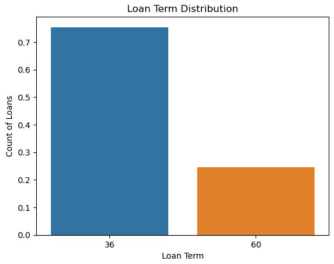
- Outliers to loan are after 29000, very few people have taken loans above 29000



2. Loan Term Observations:

- There are only 2 values here. 36 and 60 months

- 75 % borrowers have taken 36 month loan term, while only 25 % borrowers have taken 60 month loan term



Conclusions - Step 2 - Exploratory Data Analysis

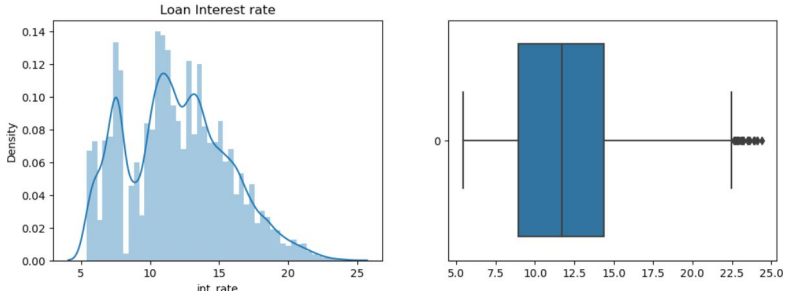
Perform Univariate analysis and look for patterns and report inferences if any. Ref.

3. Loan Interest rate Observations:

- Median of distribution is 11.71 %

- 50 % people have taken loan at int rate between 8.94 and 14.38

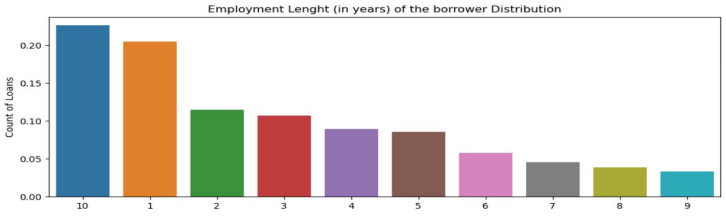
- Outliers to Loan Interest rate are after 22.5 % , very few people have taken loans above 22.5 %



4. Employment Length of the borrower Observations:

- Max loan applicants have employment history of more than 10 years , ie 22.6 %

- 50 % people have employment length between 2 and 9 years. Median is at 4 years

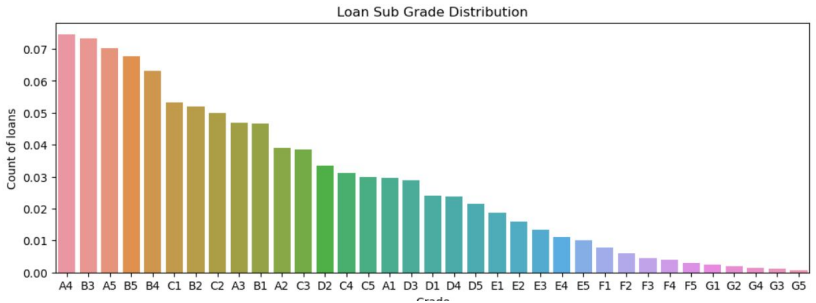


Conclusions - Step 2 - Exploratory Data Analysis

Perform Univariate analysis and look for patterns and report inferences if any. Ref.

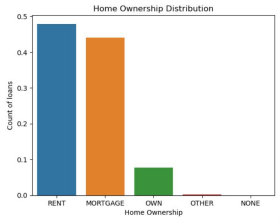
5/6. Employment Length of the borrower
Observations:

- Most borrowers fall under A & B grades



7. Home Ownership Observations:

- Most of the borrowers stay on rent / mortgage. Very few people own a home in the dataset



Conclusions - Step 2 - Exploratory Data Analysis

Perform Univariate analysis and look for patterns and report inferences if any. Ref.

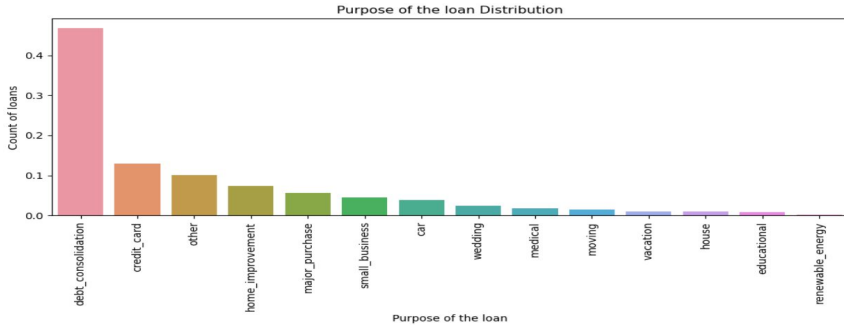
8. Annual Income Observations:

- 50 % people have annual income in range 404040 to 823040

```
In [37]: data['annual_inc'].describe()
Out[37]: count    3.857700e+04
         mean     6.877797e+04
         std      6.421868e+04
         min      4.000000e+03
         25%      4.000000e+04
         50%      5.886800e+04
         75%      8.200000e+04
         max      6.000000e+06
         Name: annual_inc, dtype: float64
```

9. loan Purpose Observations:

- Debt Consolidation and Credit card seem to be the primary purpose for the loans



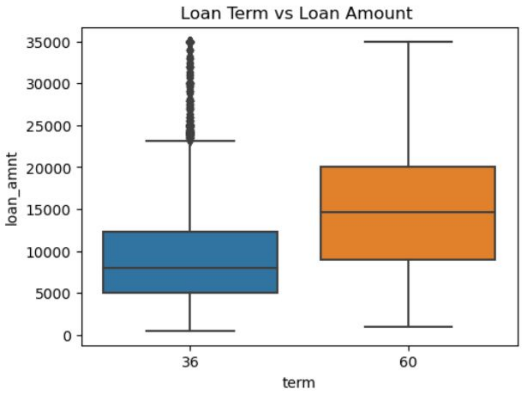
Conclusions - Step 2 - Exploratory Data Analysis

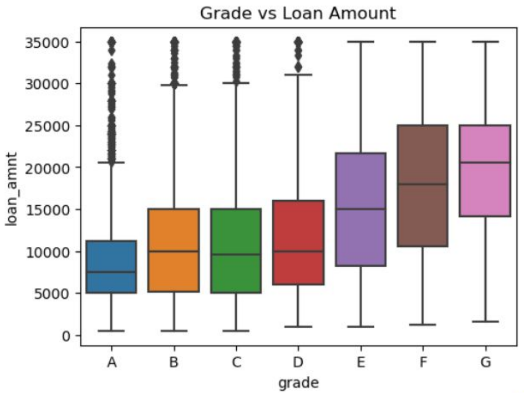
Perform Segmented Univariate analysis and look for patterns and report inferences

1. Loan Term vs Loan Amount & Grade vs Loan Amount - Observations:

- Loan amount correlates to higher tenure. i.e. Higher the loan amount, more the tenure.

- Grade 'F' and 'G' have taken maximum loan amount.





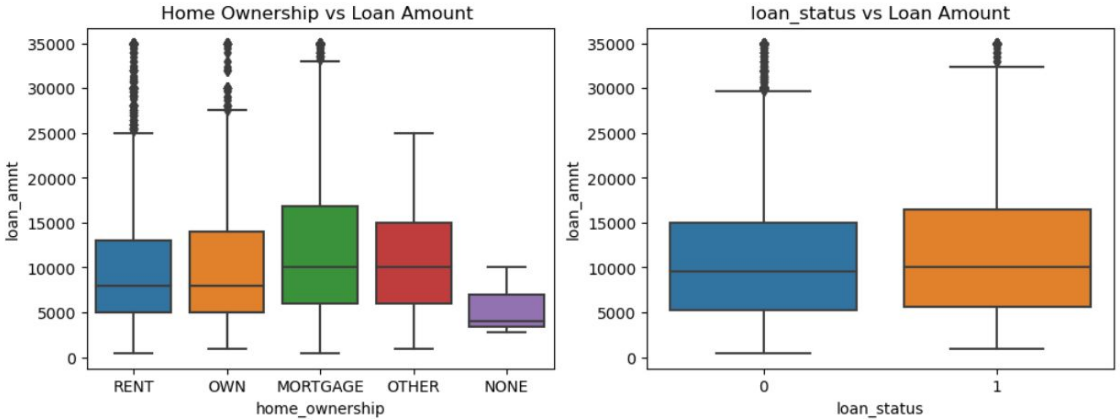
Conclusions - Step 2 - Exploratory Data Analysis

Perform Segmented Univariate analysis and look for patterns and report inferences

2. Home Ownership vs Loan Amount & Loan Status vs Loan Amount - Observations:

- Higher Loan amount correlates to Mortgaged borrowers

- Charged Off loans have slightly higher aloan mounts than Fully Paid.



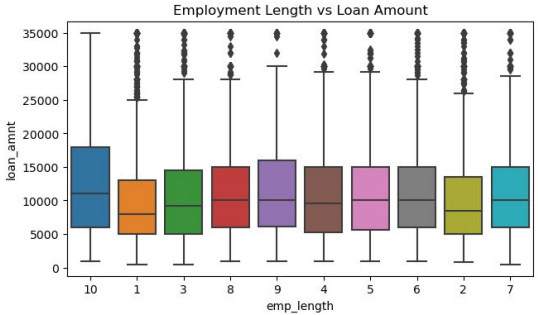
Conclusions - Step 2 - Exploratory Data Analysis

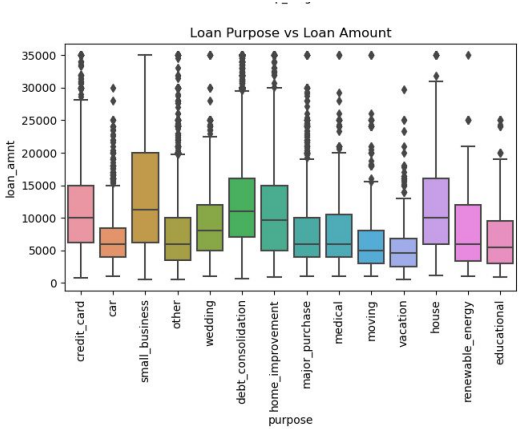
Perform Segmented Univariate analysis and look for patterns and report inferences

3. Employment Length vs Loan Amount & Loan Purpose vs Loan Amount - Observations:

- Loan amount correlates to higher emp length. i.e. 10 + years and above borrowers claim more loan amount

- Loan amount is most for Small Business and least for vacations





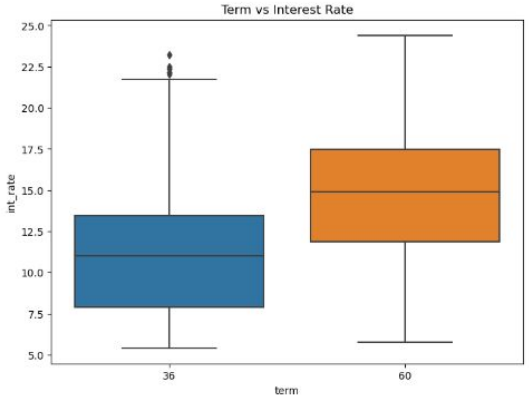
Conclusions - Step 2 - Exploratory Data Analysis

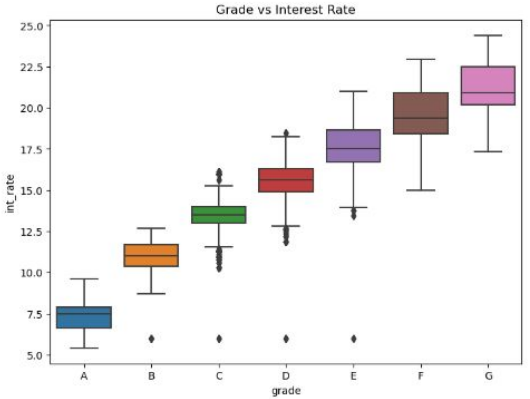
Perform Segmented Univariate analysis and look for patterns and report inferences

4. Term vs Interest Rate & Grade vs Interest Rate - Observations:

- Interest rates are higher for Higher tenure loans.

- And Also Interest Rates are Higher as Grades in descending order (G to A).





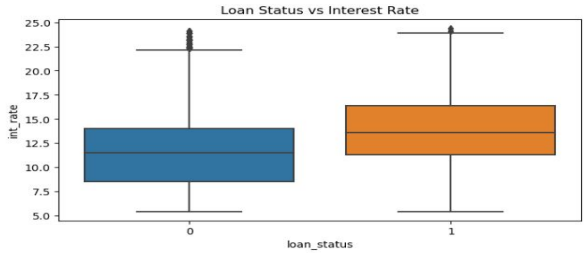
Conclusions - Step 2 - Exploratory Data Analysis

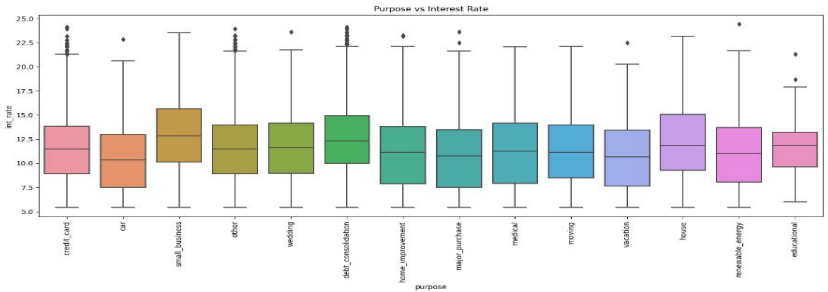
Perform Segmented Univariate analysis and look for patterns and report inferences

5. Loan Status vs Interest Rate & Purpose vs Interest Rate - Observations:

- Interest rates correspond to more defaults

- Small business followed by house loans make most defaults





Conclusions - Step 2 - Exploratory Data Analysis

Perform Bi Univariate analysis and look for patterns and report inferences

1. **Loan Amount vs Loan Status - Observations:**

- Higher the loan amounts, more the defaults

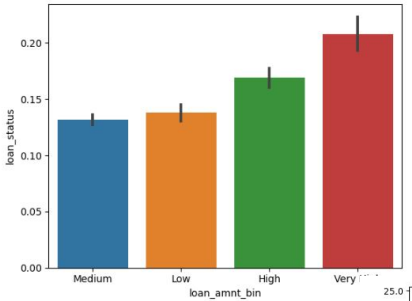
2. **Loan Term vs Loan Status - Observations:**

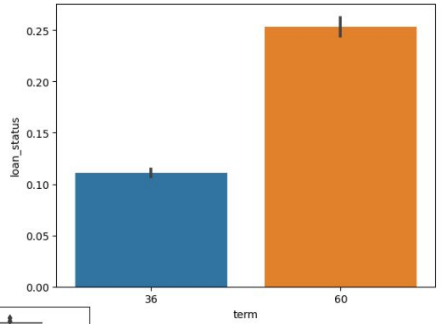
- More proportion of borrowers defaulted loan in 60 months term then 36 months

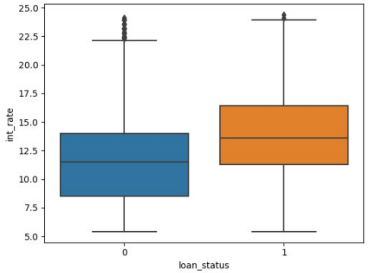
3. **Loan Int Rates vs Loan Status - Observations:**

- There are significantly more defaults at higher int. rates .

- This would imply that since the int rate is higher, load would be riskier, which correlates to more defaults







Conclusions - Step 2 - Exploratory Data Analysis

Perform Bi Univariate analysis and look for patterns and report inferences

4. Employment Length vs Loan Status - Observations:

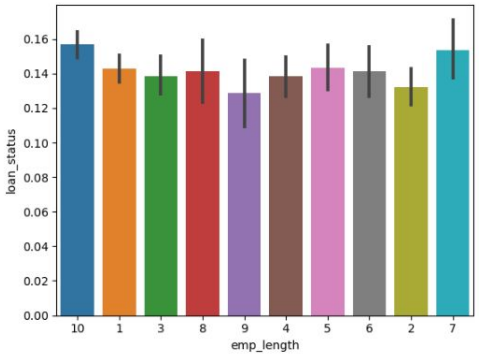
- It is observed that there is no impact of Employment Length on loan status. The spread is almost even

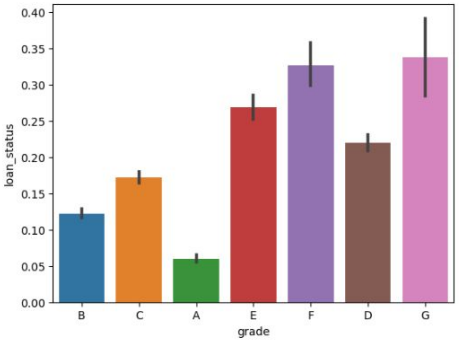
5. Loan Grades/Sub grades vs Loan Status - Observations:

- In the alphabetical order of loan grades and sub grades, Defaults increase.

- Ex. Loans if Grades A and Band less prone to defaults,

- whereas loans of grades E, F, G and H are more likely to be charged off





Conclusions - Step 2 - Exploratory Data Analysis

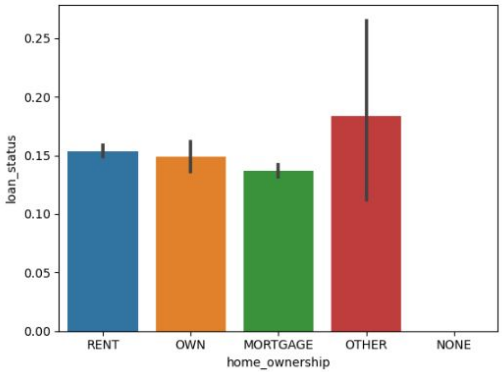
Perform Bi Univariate analysis and look for patterns and report inferences

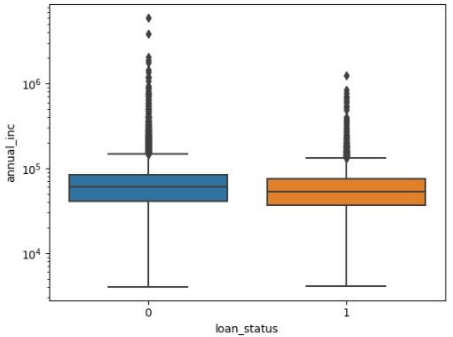
6. Home Ownership vs Loan Status - Observations:

- It is observed that there is no impact of Home Ownership on loan status. The spread is almost even

7. Annual Inc. vs Loan Status - Observations:

- Higher income borrowers usually have Fully Paid up Loans.





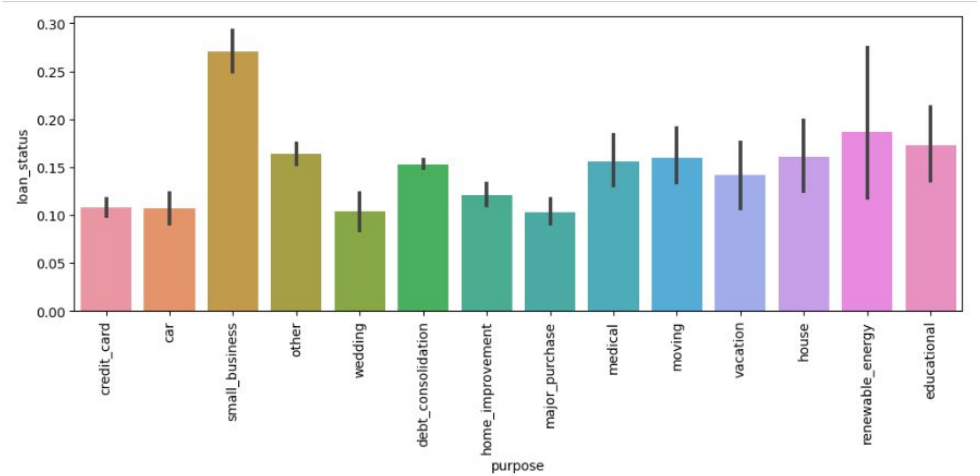
Conclusions - Step 2 - Exploratory Data Analysis

Perform Bi Univariate analysis and look for patterns and report inferences

8. Loan Purpose vs Loan Status - Observations:

- Loans taken for the purpose of small business have defaulted the most

- On the other hand, Loans taken for wedding, credit card, car and major purchase are least defaulted



Conclusions - Step 2 - Exploratory Data Analysis

Correlation Matrix Analysis Observations

- Int. Rate is correlated to the Installment amount and the loan amount

- loan amount, Funded Amount and Funded Amount Inv are highly correlated