Lending Club Case

EDA Case Study

06-Sep-2022

Agenda

Introduction
Primary Goals
Data Gathering and Cleansing
Univariant Analysis
Bivariant Analysis
Summary
Team Definition

INTRODUCTION

Introduction

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.

PRIMARY GOAL

Primary Goal

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

1. 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.

Primary Goal Ctd.

2. 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)

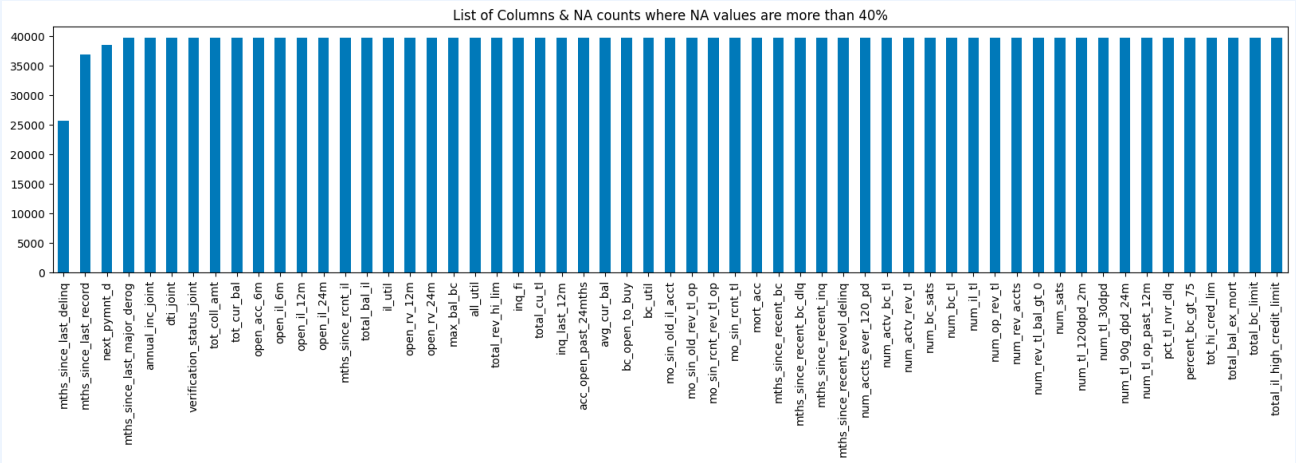
DATA GATHERING and CLEANSING

Dataset Details

Data received from Upgrad platform in form of csv file with below details:

- Total No. of rows: 39717
- Total No. of columns: 111

Analysis of data column which has NA or missing values more than 40%.



Dataset Cleansing

Removing these columns as it is difficult to impute the missing values.

```
Number of Columns dropped       :  57

Old dataset rows,columns (39717, 111)
New dataset rows,columns (39717, 54)
```

Removing the columns which has unique value is only 1.

```
1  data.drop(labels = list(unique.index), axis =1, inplace=True)
2  print("So now we are left with",data.shape ,"rows & columns.")
✓  0.3s

So now we are left with (39717, 45) rows & columns.
```

Removing of irrelevant columns.

```
1  not_required_columns = ["id","member_id","url","zip_code"]
2  data.drop(labels = not_required_columns, axis =1, inplace=True)
3  print("So now we are left with",data.shape ,"rows & columns.")
✓  0.4s

So now we are left with (39717, 41) rows & columns.
```

Dataset Cleansing Ctd.

Removing those records are less than 0.75 in purpose of loan.

```
So now we are left with (35621, 41) rows & columns.
['credit_card' 'car' 'small_business' 'wedding' 'debt_consolidation'
 'home_improvement' 'major_purchase' 'medical' 'moving' 'vacation' 'house'
 'educational']
```

Creating the bin for defining the range in loan amount, annual income, interest rate.

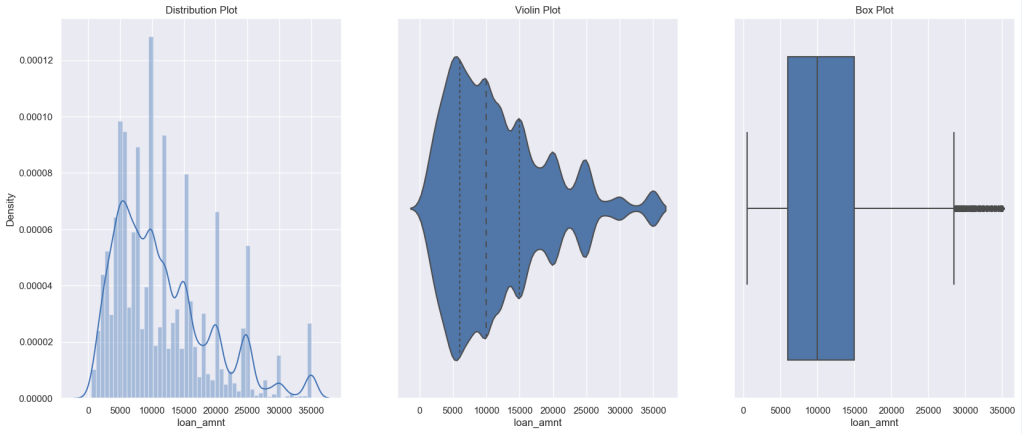
Cast all continuous variables to numeric so as to find the correlation between them.

UNIVARIANT ANALYSIS

Univariant Analysis

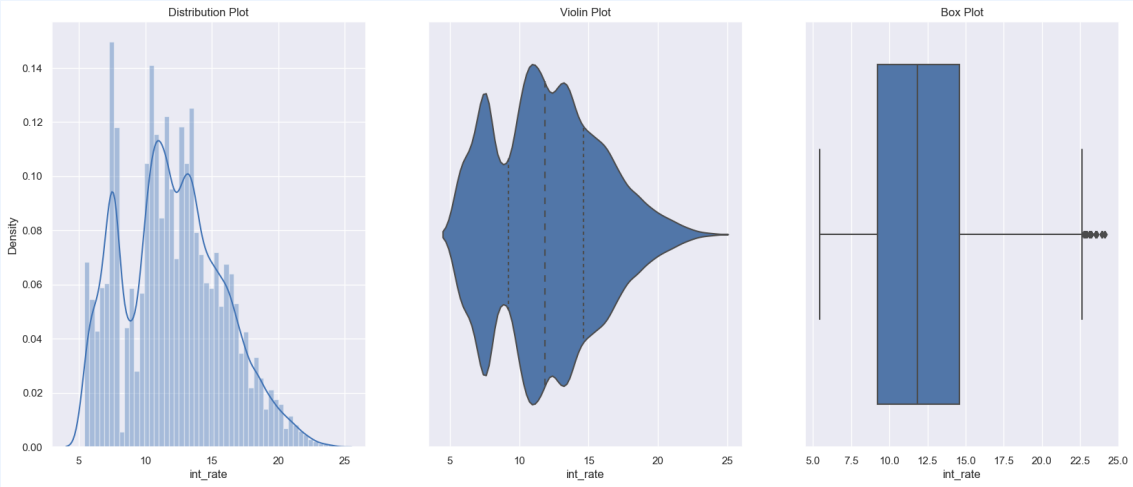
Univariate analysis is the simplest form of analysing data. Uni means one, so in other words the data has only one variable. Univariate data requires to analyse each variable separately.

Analysis for loan amount is given below:



Univariant Analysis Ctd.

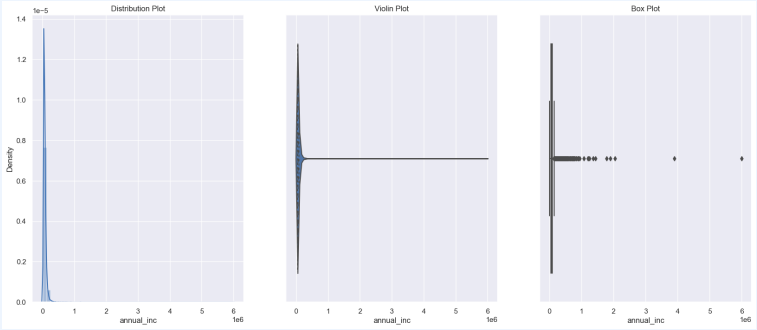
Analysis for Interest Rate is given below:



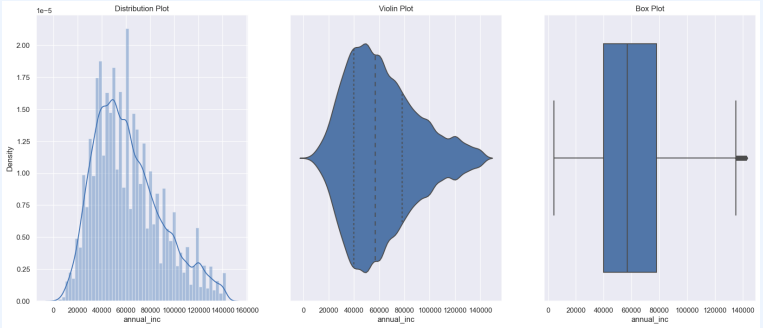
Analysis for Annual Income before and after removing outliers.

Univariant Analysis Ctd.

Before removing outliers.

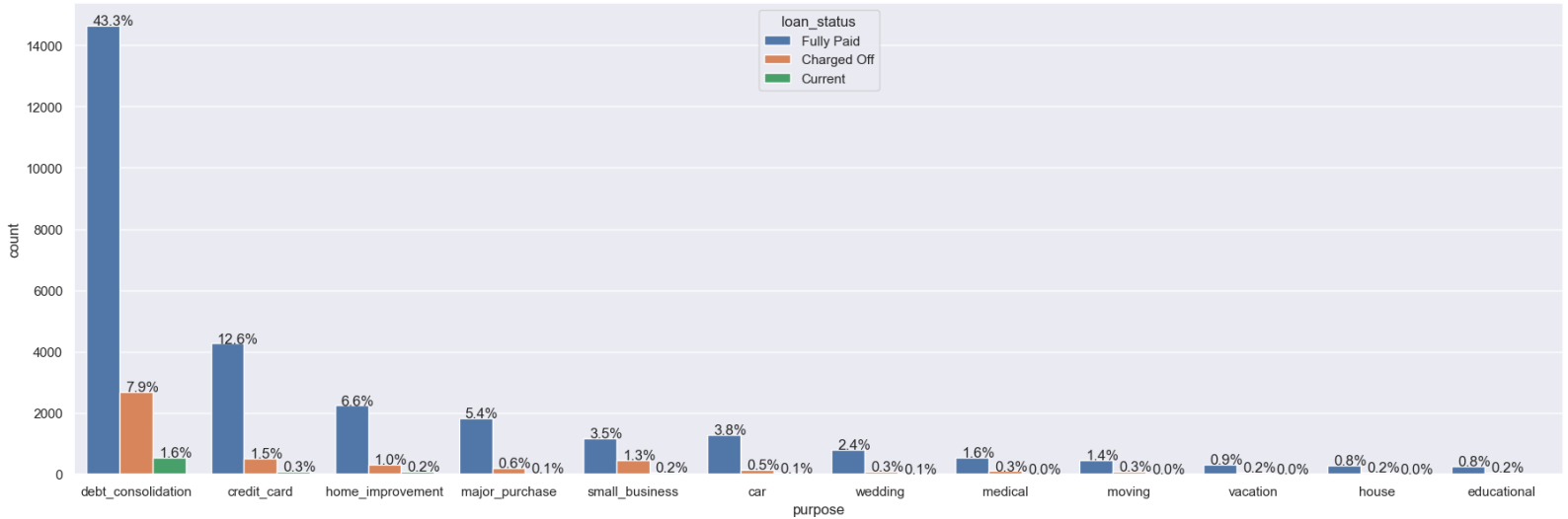


After removing outliers.



Univariant Analysis Ctd.

Analysis for Purpose of loan is given below:

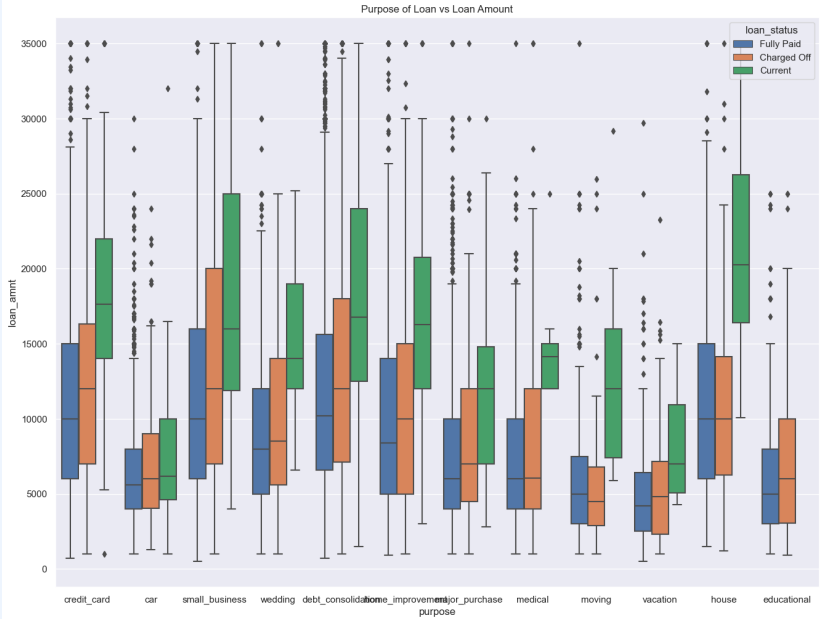


BIVARIANT ANALYSIS

Bivariant Analysis

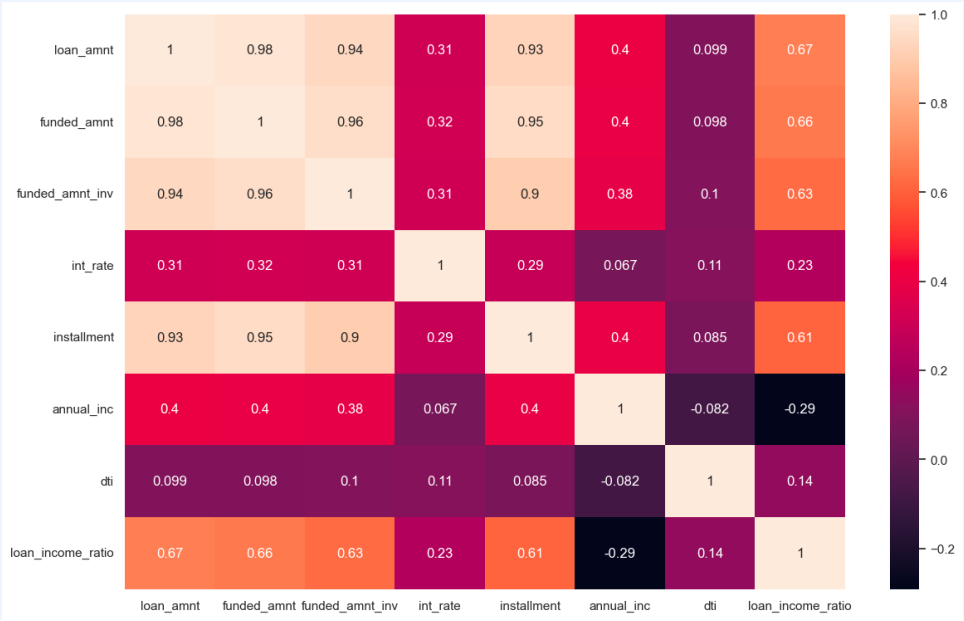
Bivariate/Multivariate Analysis finds out the relationship between two or more than two variables.

Analysis for purpose of loan and loan amount is given below:



Bivariant Analysis Ctd.

Relation of continuous variables and heatmap:



SUMMARY

Summary

- Most of the loan amounts are distributed between 5000 to 15000 USD.
- Most of the loans interest rates are distributed between 8% to 15%.
- Most of the applicants earns between 35000 to 80000 USD annually.
- Approx. 52% of the applicants applied loan for paying their other loans(Debt Consolidation)
- Approx. 48% of applicants are living in rented home whereas 42% applicants were mortgaged their home.
- Loan applicants are increasing year on year, approx. 47% of loan applicants received loans in 2011.
- Approx. 70% of applicants applied loan for 36 months term period.
- Heatmap tell that how 'loan_amnt', 'funded_amnt' & 'funded_amnt_inv' are closely interrelated.

TEAM DEFINITION

Team Definition

Contact Details

Name	Contact Information	Photo
Himanshu Pandey	Email: himsoftware@gmail.com	
Avinash Saraswat	Email: avinash.saraswat93@gmail.com	

Thank You!