Loan Data Analysis using Lending Club Loan Dataset

By,

Aniket Roy
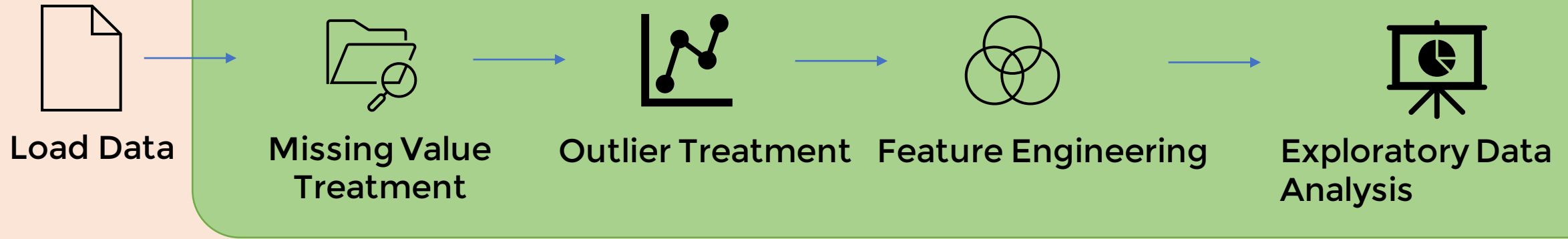
Srujana N

PROBLEM STATEMENT

- Lending Club is the largest online loan marketplace, offering products like
  - ➤ Personal Loans
  - ➤ Business Loans
  - ➤ Financing medical procedures
- Borrowers can apply and access these products through an online interface
- But lending loans to highly risky applicants can cause a very large credit loss
- Therefore, we must identify the driving factors that are strong indicators of a loan defaulter

OVERALL APPROACH

Load Data → Missing Value Treatment → Outlier Treatment → Feature Engineering → Exploratory Data Analysis

RECOMMENDATIONS

MISSING VALUE TREATMENT

ALL_UTIL	38577	100.00
TOTAL_REV_HI_LIM	38577	100.00
INQ_FI	38577	100.00
IL_UTIL	38577	100.00
MTHS_SINCE_LAST_RECORD	35837	92.90
MTHS_SINCE_LAST_DELINQ	24905	64.56
DESC	12527	32.47
EMP_TITLE	2386	6.19
EMP_LENGTH	1033	2.68
PUB_REC_BANKRUPTCIES	697	1.81
LAST_PYMNT_D	71	0.18
COLLECTIONS_12_MTHS_EX_MED	56	0.15

- Dropped all the columns that had at least 30% missing values as imputing these will result in bias/de-authentication of data

- Dropped all those rows whose respective columns had insignificant number of missing values. Dropping these will not result in data loss as the number is insignificant

OUTLIER TREATMENT

- Removing the outliers by making sure we are not losing a significant amount of data
- We may still have some outliers in the data, which is fine. Because, if we drop them, there will be a significant loss of data
- Following are some of the features where we performed outlier treatment,
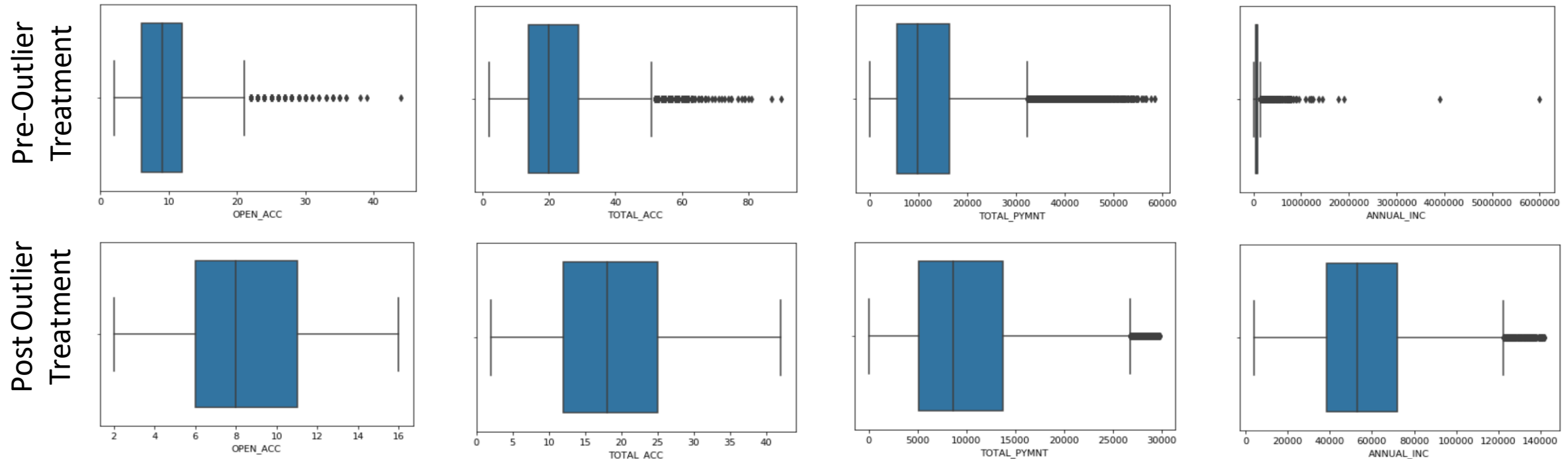
FEATURE ENGINEERING

- Continuous features have values that are absolute numbers and trying to compare the absolute numbers with "Defaulters" and "Non-defaulters" will be difficult and make no sense. Hence, we can segment these data-points within features such that each segments have data-points that are homogenous to each other, and the segments are heterogenous amongst each other

LOAN_AMNT	INT_RATE	INSTALLMENT	ANNUAL_INC	DTI
5000.0	10.65	162.87	24000.0	27.65
2500.0	15.27	59.83	30000.0	1.00
2400.0	15.96	84.33	12252.0	8.72
10000.0	13.49	339.31	49200.0	20.00
5000.0	7.90	156.46	36000.0	11.20

| LOAN_AMNT_SEGMENTS	INT_RATE_SEGMENTS	INSTALLMENT_SEGMENTS	ANNUAL_INC_SEGMENTS	DTI_SEGMENTS
3000-8000	11-15	100-500	0-50000	20+
0-3000	15+	0-100	0-50000	0-10
0-3000	15+	0-100	0-50000	0-10
8000-18000	11-15	100-500	0-50000	10-20
3000-8000	5-10	100-500	0-50000	10-20

EXPLORATORY DATA ANALYSIS

- Following plots depicts the key drivers for loan defaulting,
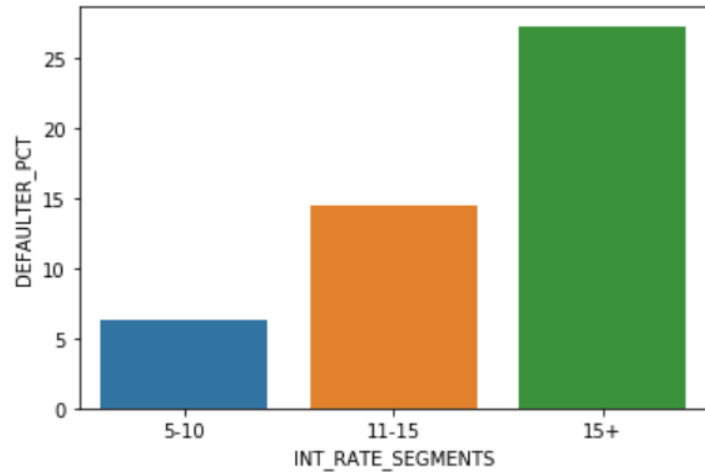
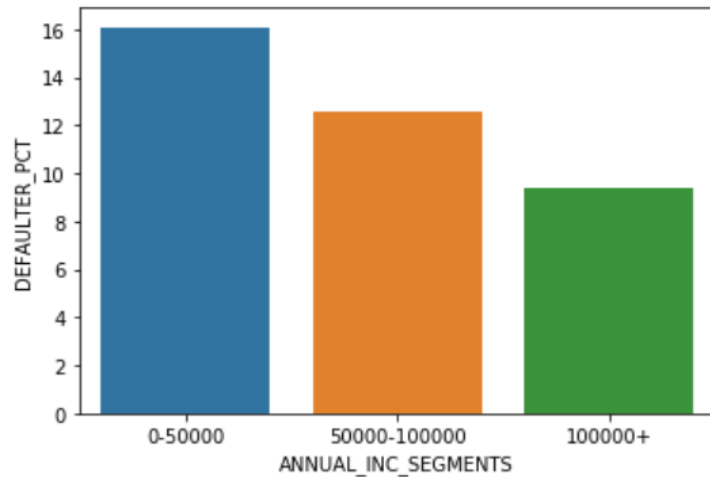**Higher the interest rates, higher are the chances of defaulting**
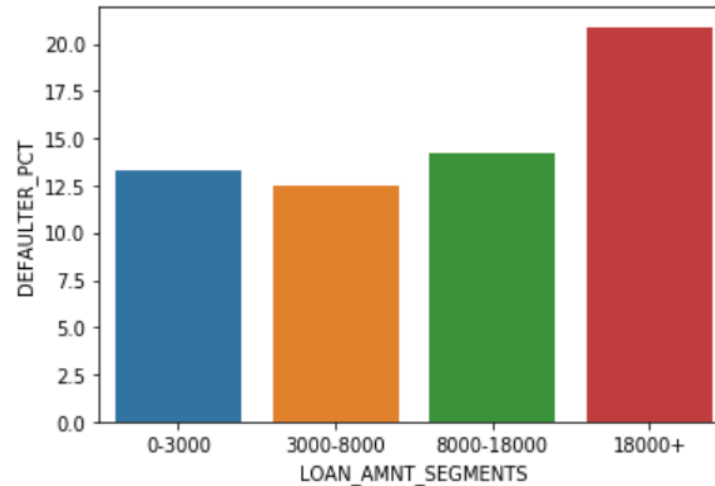
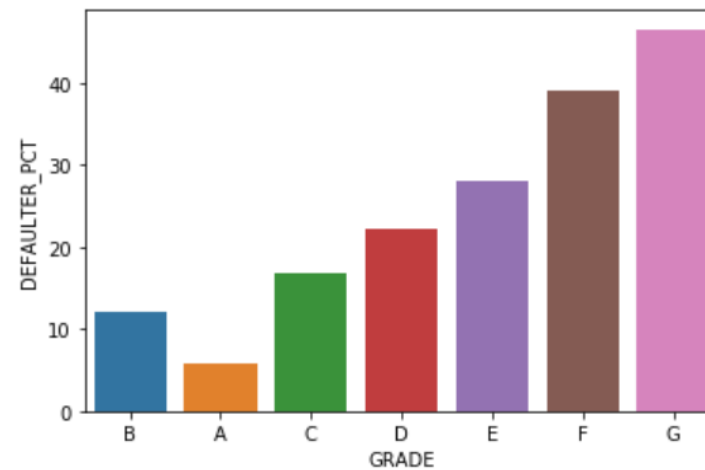**Higher the loan amount, higher are the chances of defaulting**
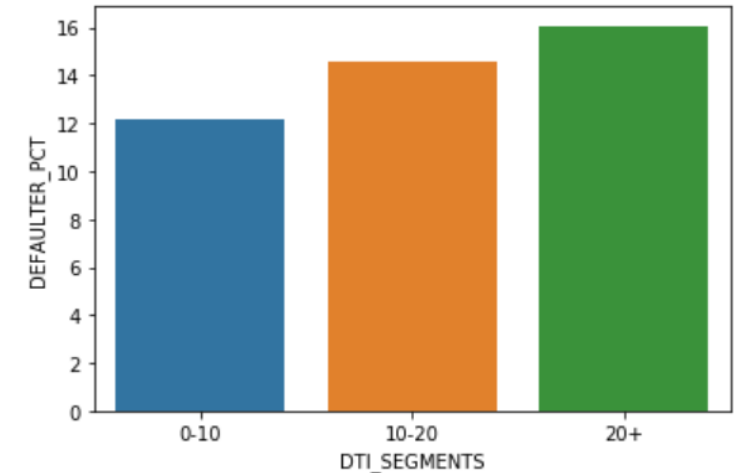
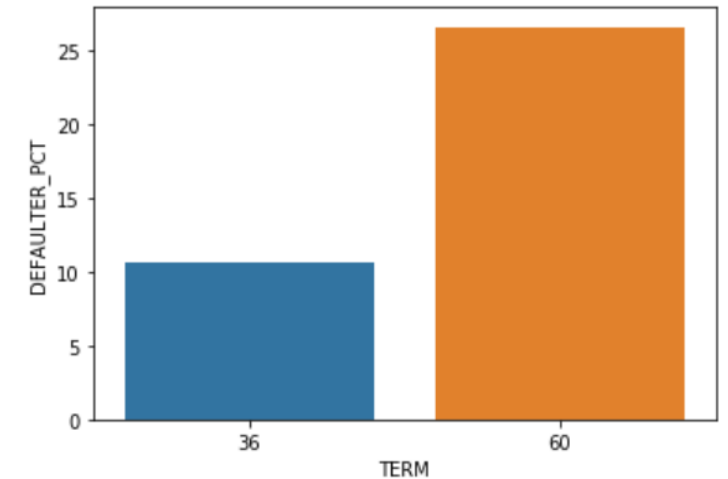**Higher the debt to salary ratio, higher are the chances of defaulting**

EXPLORATORY DATA ANALYSIS

Lower the salary, higher are the chances of defaulting

Higher the grade of debt, higher are the chances of defaulting

Higher the debt term, higher are the chances of defaulting

RECOMMENDATIONS

- As observed in the EDA, interest rates, loan amount, debt to salary ratio, salary, grade of debt and term of debt are the key drivers of loan defaulting

- Following are our recommendations,
  - To the applicants with low salary, high loan amount requirement, high debt to salary ratio, higher grade of debt and higher term requirement, we can approve a loan of an amount not to high but with a high interest as they are very high risky applicants
  - To the applicants with medium salary, medium loan amount requirement, medium debt to salary ratio, medium grade of debt and medium to low term requirement, we can lend a slightly higher loan amount and a slightly lower interest rate than the highly risky applicants
  - To the applicants with high salary, low loan requirement, low debt to salary ratio, low grade of debt and with a low term requirement, we can lend a higher amount as required with a standard or low interest rates as they are the least risky applicants