Lending Club Case St

Key Insights of EDA

Author 1: Sumegh Singh Chouhan

Author 2: Ananth

Date: 25-12-2024

em Statement:

ess Understanding:

rk for a consumer finance company which specialises in lending types of loans to urban customers. When a person applies for a ere are two types of decisions that could be taken by the company:

n accepted: If the company approves the loan, there are 3 possible scenarios described below:

Fully paid: Applicant has fully paid the loan

Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet co

Charged-off: Applicant has not paid the instalments in due time for a long period of ti has defaulted on the loan

rejected: The company had rejected the loan (because the candidate does not meet their requirem n was rejected, there is no transactional history of those applicants with the company and so e with the company (and thus in this dataset)

ess Objectives:

mpany wants to understand the driving factors (or driver variables) behind loan default, i.e. the varia ndicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Loading and rstanding:

oading

dataset using pandas.

atistics

rows: 39,717

columns: 111

et Used:

ionary.xlsx

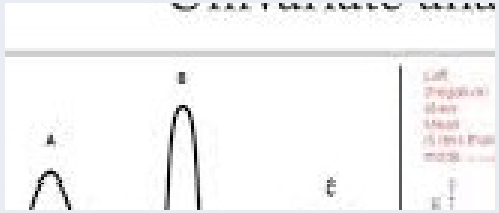
Modules of the Pro

1. Data Cleaning & Identifying Outliers

2. Analysing Data Imbalance

3 De

4. EDA-Univariate Analysis Insights



5. EDA-Segmented Univariate Analysis Insights

6 A

8. Results

9.Recommendations

Cleaning & Identifying Outliers :

Inspection- Displayed the first few rows of the
d observed various data types including floats,
d objects.

alue Check- Checked for missing values in the
d identified columns with a significant number of
ues.

emoval- Removed those Columns which had more
of missing values or were irrelevant to the problem
i.e. came into existence after loan approval or was
to the borrower)

val- Duplicate rows were removed.

lue Treatment- Null values were replaced with
r Numerical Columns) and Mode (For Categorical

Data types- Converted the data types of Interest
Revolving Line Utilization Rate Columns' values from
oat type using astype() function and removed the '%'
() function.

7. Converted date columns to datetime form
columns to datetime format using to_datetim

8. Performed Sanity Check on the data- Ou
from the Annual Income and the revolving b
the Column "Account Type" was removed
value throughout.

9. Removed some instances where categ
are really low compared to total population-
(i) The data about the following states wer
dataset as their value counts were really l
total population-

'MT', 'WY', 'AK', 'SD', 'VT', 'MS', 'TN', 'IN'

(ii) Some values which were 'NONE' and
Home Ownership Column were also remove

-variate Analysis Insights:

alysis was performed on the dataset
a Imbalance
ed Metrics like ‘delinquency_in_2years’,
_occured’, 'month’, ‘Year’

0% of the loans were fully paid.
oans were taken on month of November and
r of loans were taken in Feb.
lmost getting doubled year after another 2007
percentage of loans whereas 2011 have highest
s implies that year after another this company is

5% of folks have no records of bankruptcies and
have no previous delinquency.
olks applying for loan are from CA, NY.
oans are for paying other debts, after that loans
r home and business improvement.
0% of the borrowers are not verified, some are
very few are sourced verified.
g in rented and mortgage houses have majority of

9. People working for more than 10 percentage of loan(~25%), People wo 3 Years contribute for 33% of the l because they are borrowing mon career growth.

10. More than 73 percent borrowers of 36 months, this might be to sa funded amount.

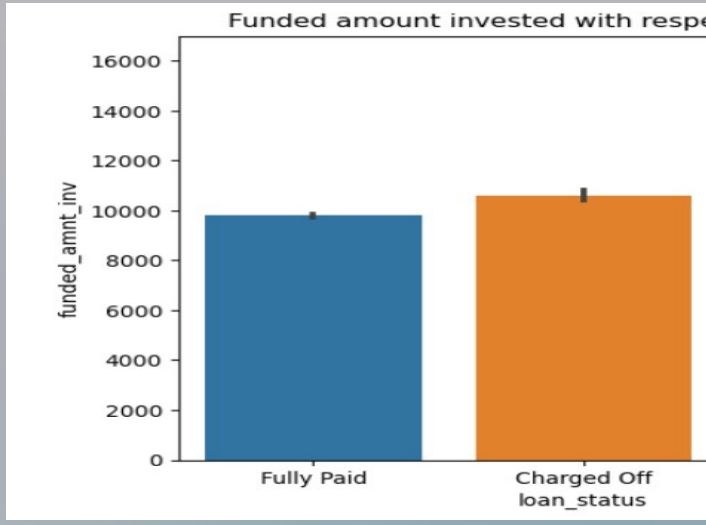
11. Loan Amount, Funded amoun amount by investors follow a s Mostly values of these three varia $5000 to $15000, this may represent investing their money in this compa amount Is fulfilled by investors onl will be seen in bivariate.

12. Most of the borrowers having between $50000 and $70000 are organization for loan where peak li very few investors have annual s $200000

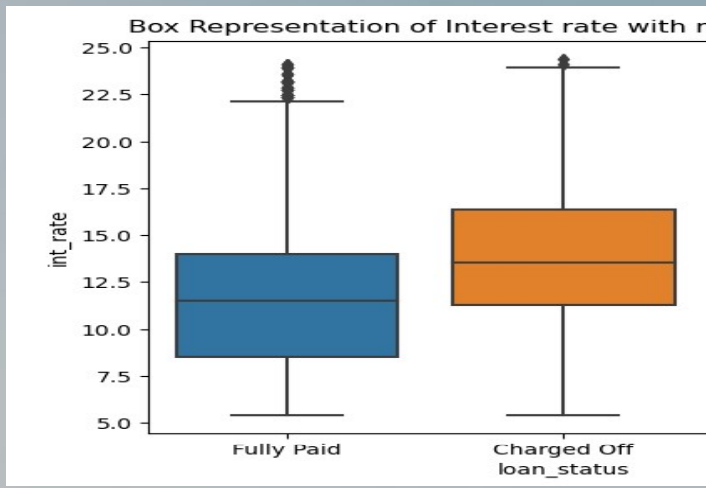
-Segmented
ariate Analysis
ts:

off borrowers have a higher funded amount, interest
installments compared to fully paid ones.

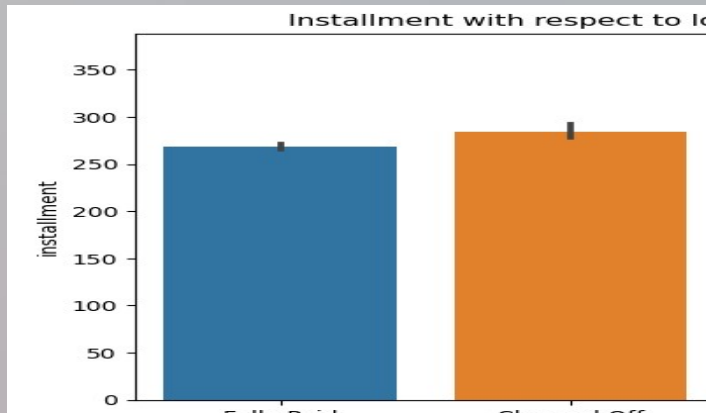
- Bar Plot of. Funded Amount by Investo



- Box Plot of Interest Rates w.r.t. Loan



- Bar Plot of Installment w.r.t. Loan Sta



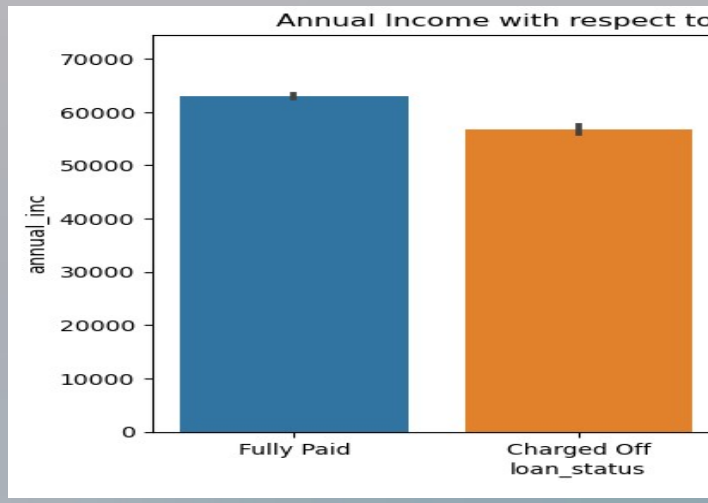
-Segmented
ariate Analysis
ts:

ome depicts lower chances of defaulting .

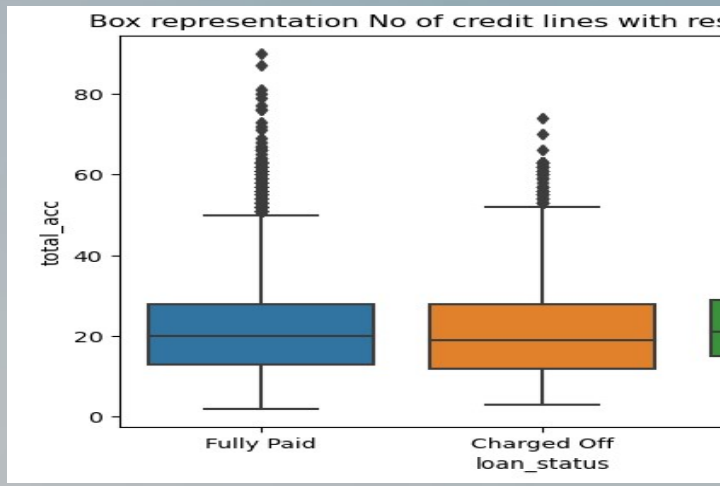
borrowers have a slightly higher number of credit
e compared to charged off borrowers. Also, the
outliers are more in the case of Fully Paid borrowers.
t take more loan but pay their debt.

s with Higher Debt-to-Income Ratio are more likely to
ers.

- Bar Plot of Annual Income w.r.t. Loan S



- Box Plot of Number of Credit Lines w.r



- Bar Plot of Debt-to-Income Ratio w.r.t.

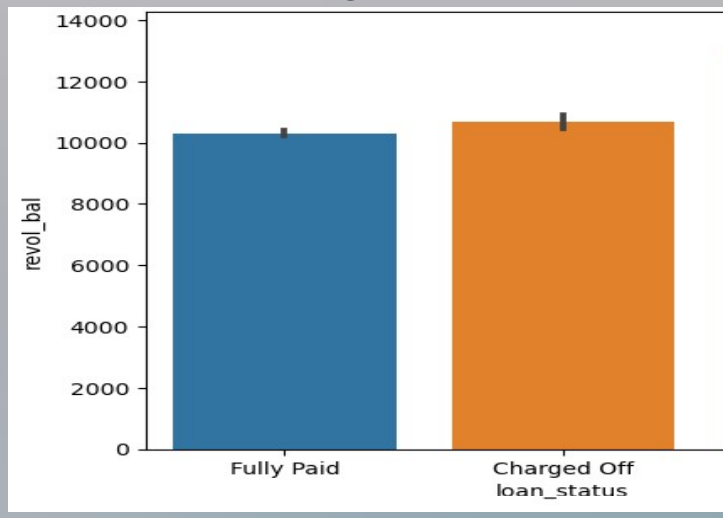
-Segmented
ariate Analysis
ts:

who are likely to default have slightly higher average
alance than fully paid ones.

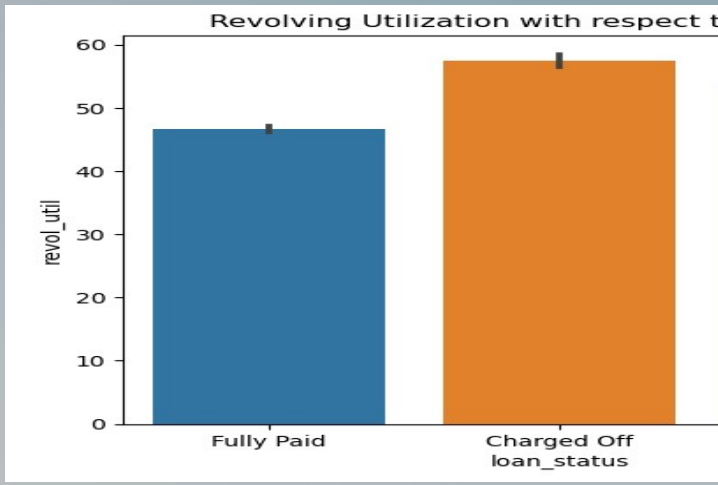
with Higher value Revolving Line Utilization rates are
to be defaulters.

with a term of 60 months are more likely to default
ners with a term of 36 months.

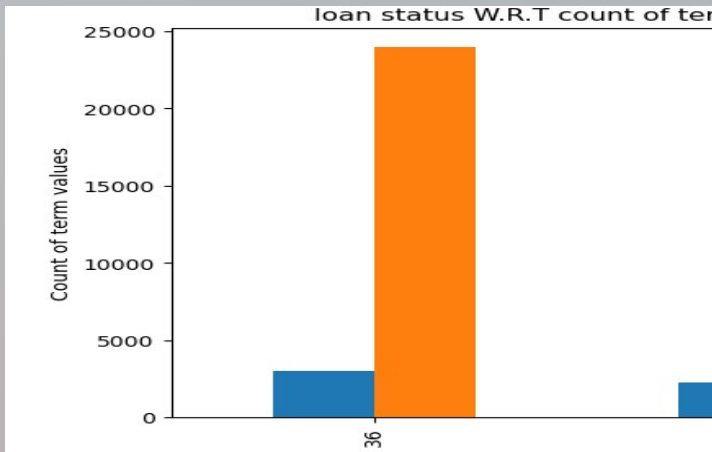
- Bar Plot of Revolving Balance w.r.t. Lo



- Bar Plot of Revolving line utilization ra



- Bar Plot of Loan Status w.r.t Term.

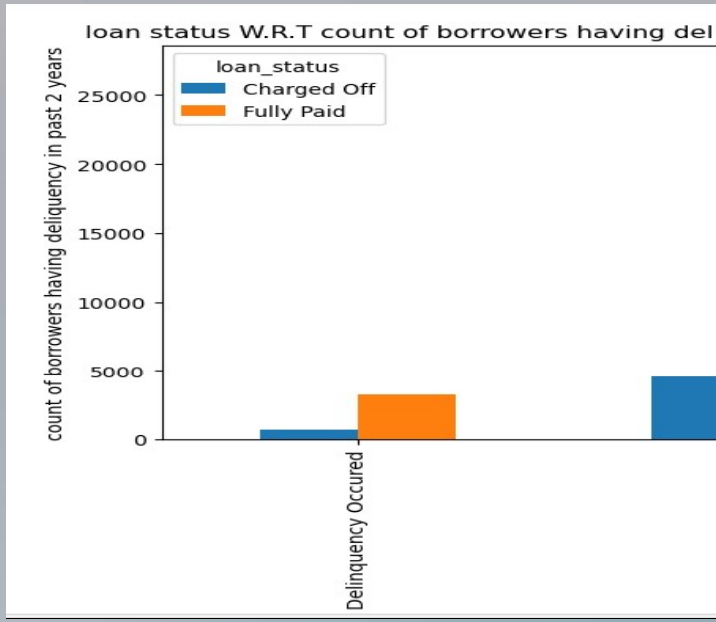


-Segmented
ariate Analysis
ts:

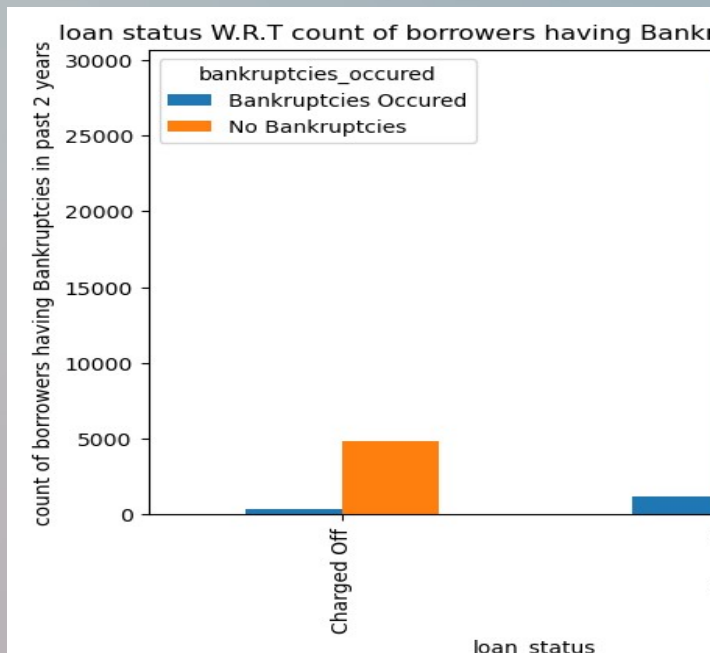
past Delinquency has charged off to fully paid ratio
ch is higher than that of non-delinquent borrowers
nargin.

earlier Bankruptcies records are more like to be
that non-bankrupt borrowers with a small margin.

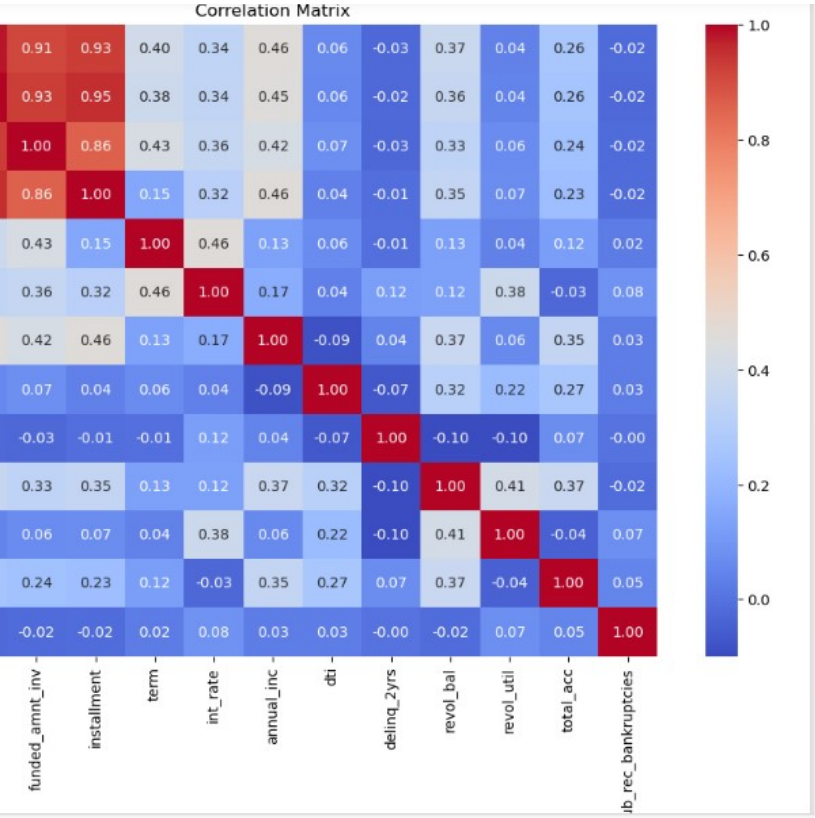
- Bar Plot of Loan Status w.r.t. Borr
  delinquency in past 2 years



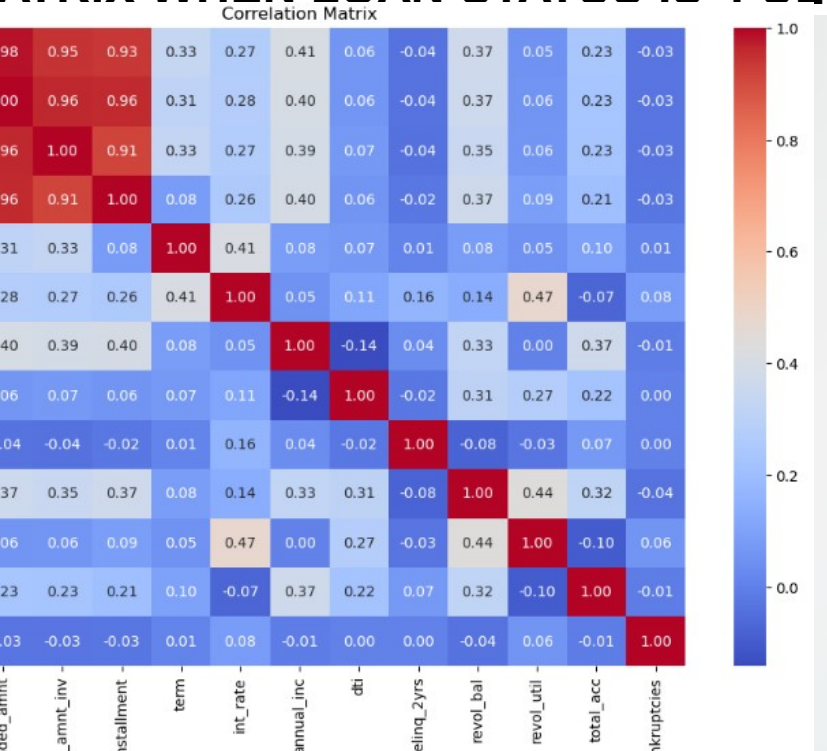
- Bar Plot of Loan Status w.r.t. Borr
  bankruptcies in the past 2 years.



ATRIX WHEN LOAN STATUS IS 'CHARGED-OFF



ATRIX WHEN LOAN STATUS IS 'FULLY-PAID'



EDA – Bi-variate Insights:

Key Comparisons:

Consistency in High Correlations: Both fu

off loans show consistent high correlations between l

amount, funded amount by investors and installment,

fundamental loan structure remains consistent regard

Term and Interest Rate: The correlation betwe

rate is slightly higher in charged-off (0.46) compared

(0.41), suggesting that longer terms might be slightly

higher interest rates in charged-off loans.

Revolving Balance and Utilization: The co

fully paid loans (0.44) than in charged-off loans (0.41

revolving balances are more critical in fully paid loans

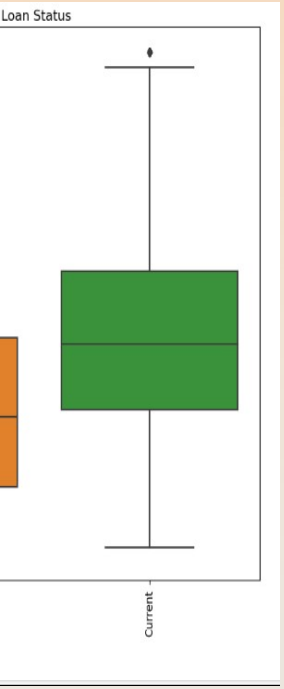
Annual Income: The moderate correlation betwe

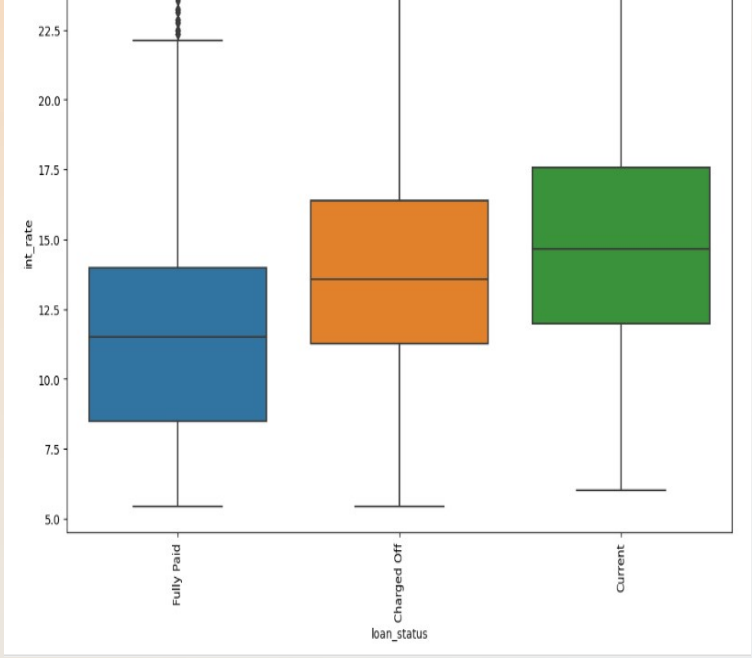
loan amount is stronger in charged-off loans (0.46) co

loans (0.41), suggesting that higher income borrower

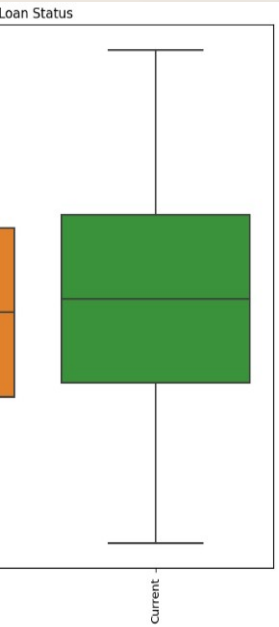
loans and are more likely to not repay them.

EDA – Bi-variate Insights:

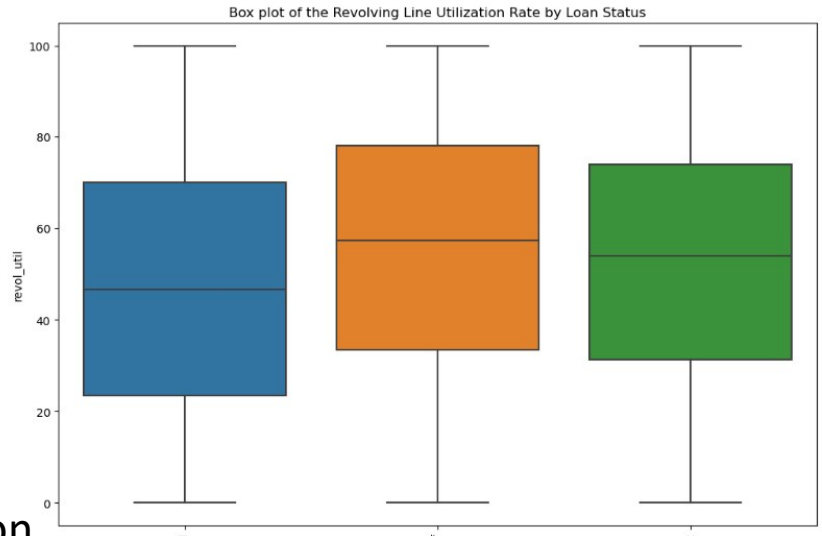




2) Box Plot of Interest Rate by Loan Status



3) Box Plot of Debt-to-Income Ratio by Loan Status



olving Utilization

1) "Charged Off" loans received ***more amounts*** compared to "Fully Paid" Loans. ***variability in funded loan amounts more*** in "Charged Off" loans as compared to "Fully Paid" Loans. The presence of ***more and higher outliers*** in "Fully Paid" Loans indicate a significant shift in lending policy or borrower behaviour for "Fully-Paid" Loans.

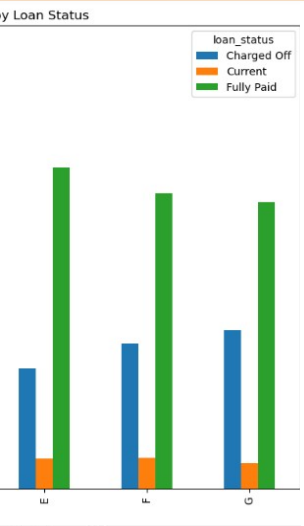
2) "Charged Off" loans had ***higher interest rates*** compared to "Fully Paid" Loans.

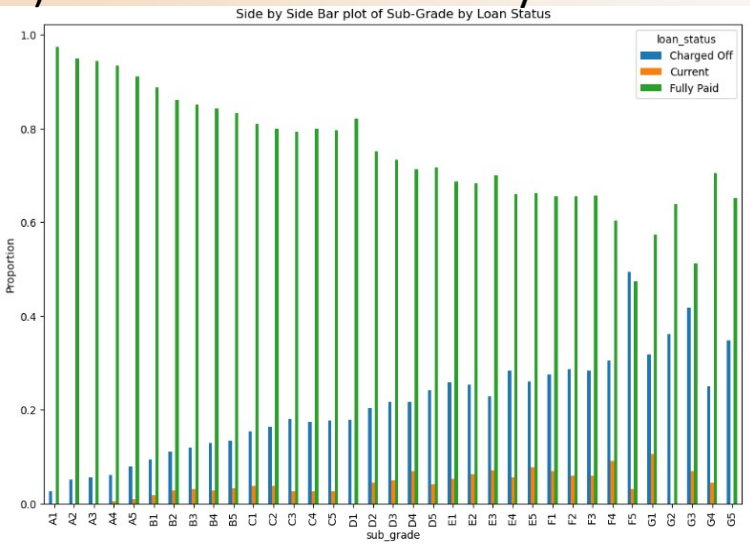
3) "Charged Off" loans had ***slightly higher Debt-to-Income Ratio*** compared to "Fully Paid" Loans.

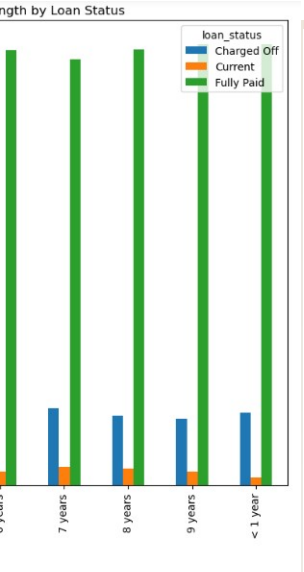
4) "Charged Off" loans had ***higher Revolving Utilization Rates*** compared to "Fully Paid" Loans.

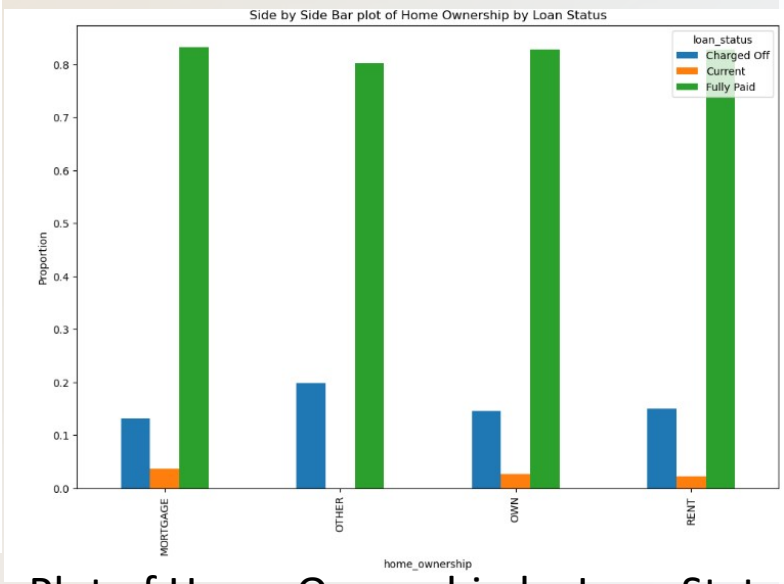
oan Status

2) Bar Plot of Sub-Grade by Loan Status



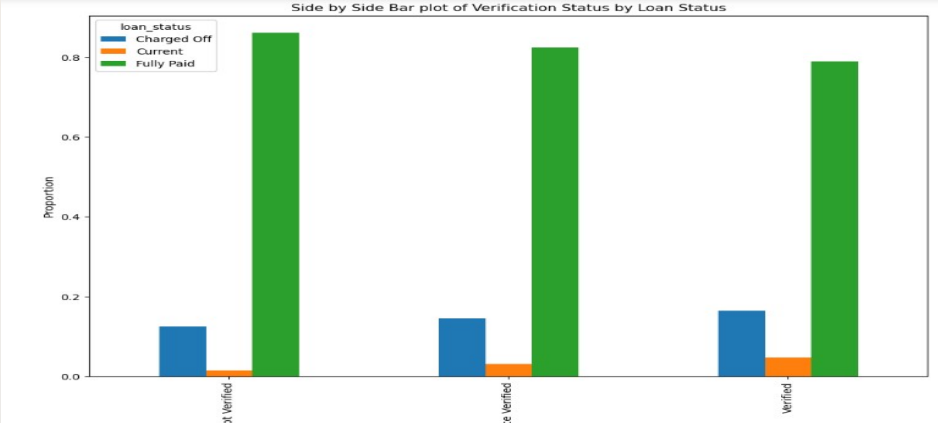






ent Length

4) Bar Plot of Home Ownership by Loan Status



Plot of
us by

EDA – Bi-variate A

Insights:

1) Overall loan grades are predictive
with ***higher grades*** indicating ***lower***
outcomes.

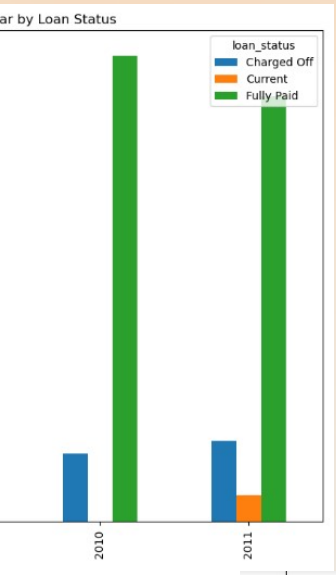
2) Sub-grade is a strong predictor of lo
with ***higher sub-grades*** (e.g., A1, A2)
better outcomes than lower sub-grade

3) While employment stability gen
with better loan performance, those
of employment length (Very new like
long-term like 10+ years) might presen

4) Homeownership status impacts loa
with those ***owning homes or hav***
performing ***better than*** others.

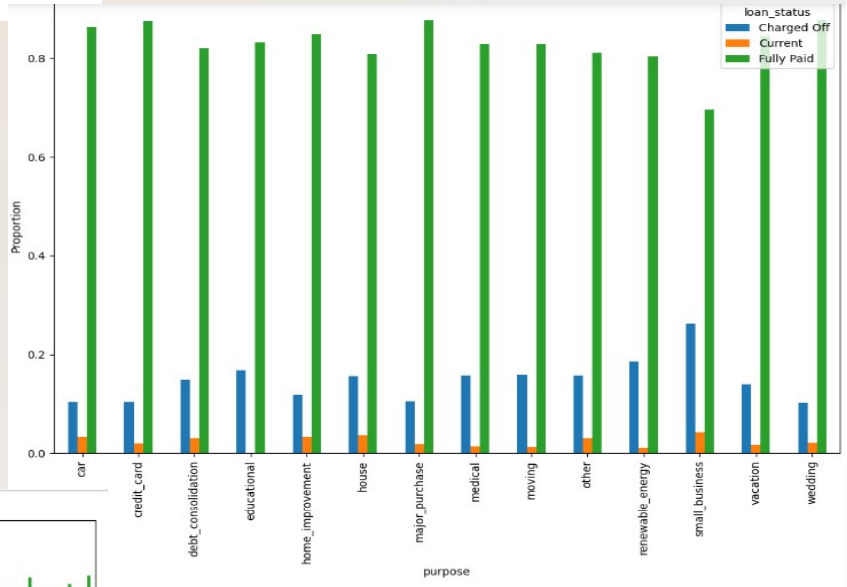
5) Verification status is a significant
performance, with ***verified loans***
outcomes.

EDA – Bi-variate Insights:



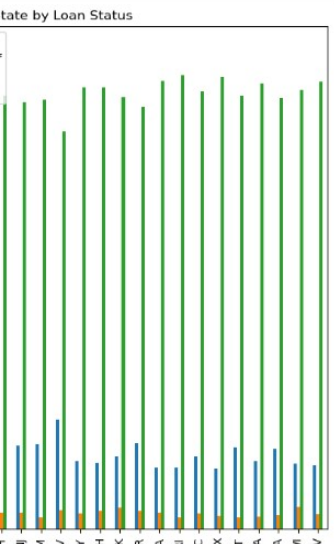
6)Side-by Side Bar Plot of Issue Year by Loan Status

6) Loans issued in ***2007*** and ***2011*** ***proportion of charged-off loans*** cor years. This could reflect broader ecor during those years, such as the finan 2007-2008 and its lingering effects.



Bar Plot of an Status

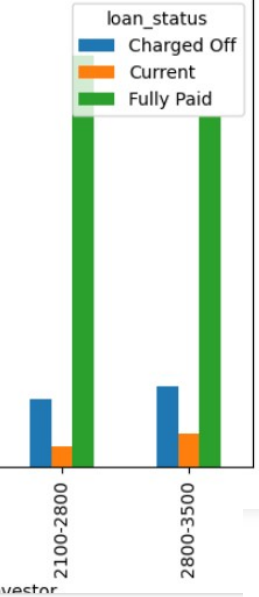
7) Loans for ***renewable energy*** and ***purposes*** have a relatively ***higher charged-off status*** compared to oth credit card loans, which have a high fully paid statuses.



8)Side-by Side Bar Plot of Address-State by Loan Status

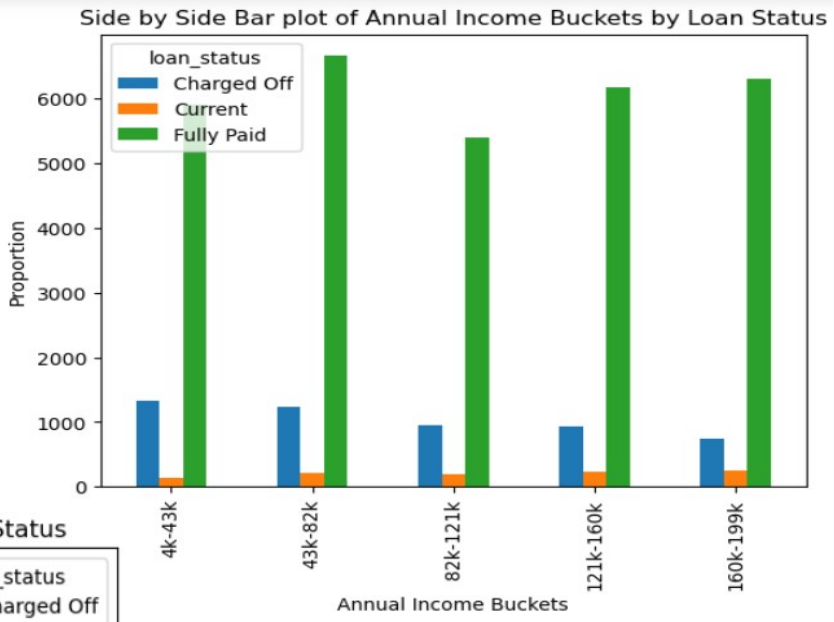
8) States like ***NV and FL*** have a ***high charged-off loans*** compared to ot potential economic stress or higher ris

EDA – Bi-variate Insights:



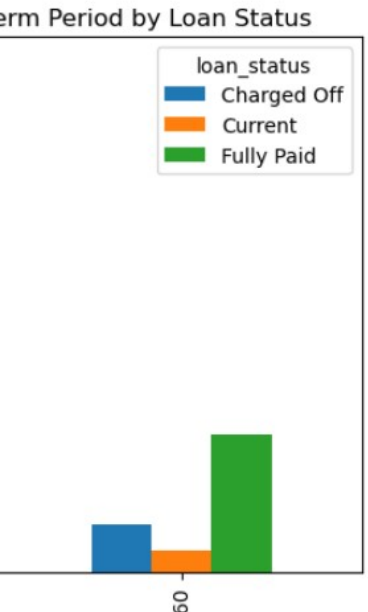
9)Side-by Side Bar Plot of Funded Amount by Investors by Loan Status

9) Borrowers with Funded amount range of ***$2800-$3500*** are ***most likely*** charged off to fully paid ratio of 22% have a ratio between 14-17%,wh bucket has the ***least probability to d*** small margin.



Plot of
Loan Status

10) As the ***annual income increas*** ***defaulting decreases***. Income range ***$*** ***charged off / fully paid ratio*** of ***22*** income ***$160k to $199k*** the ratio is ***11%***



11)Side-by Side Bar Plot of Term Period by Loan Status

11) For the term period of ***36 months*** ***to fully paid ratio is 1/10*** but in case o ***ratio is 1/3***. This indicates that custor period of ***60 months*** are ***very likely to***

S:

repayors than defaulters among *higher annual income*

e likely to be repayors than those who rented their

repayors than defaulters among people with *higher*

in file.

d loans are more likely to be repayors than borrowers

rce verified loans.

DEFAULTERS:

- There were more loan defaulters than loan repayors among those borrowers

  amount, interest rate and installments.

- There were more defaulters than repayors among people with *higher Debt-t*

- There were more defaulters than repayors among people with *higher averag*

  revolving balance and revolving line utilization rates.

- There were more defaulters than repayors among customers with *a term of*

  with a term of 36 months.

- There were more defaulters than repayors among borrowers *with more than*

- As the *grade and sub-grade decreases*(From A1,A2 to F5,G5) the borrower i

  defaulter.

- People with *past Delinquency* were more likely to be defaulters than non-de

- People with *earlier Bankruptcies records* are more like to be defaulters than

- There were more loan defaulters than repayors among borrowers who were

  2011

- There were more loan defaulters than repayors among borrowers who took

  energy and *small business purposes* compared to those who took *car or cred*

- There were more loan defaulters than repayors among borrowers who belon

  FL compared to those from other states.

mendations:

alysis, we are now able to predict whether a Client will repay the loan or not. The following are our recom

ded amount and installments where interest rate have a *higher dependency on loan status*. To avoid defaul

high amount should not be given to *low annual income* personal with *income range $4000 to $43000.*

ans should be avoided for *low income* people who opt *for high amount loans* as they have really *high risk* c

d *for lower sub-grade(F5, G5) loans*, have a *higher risk* of getting *charged off.*

n taken by borrowers who have *more credit lines* are *more likely to repay the loan back* maybe from gettin

ns for *small business are really risky*, a thorough *background check must be done* for them before approvin

term period of 60 months, have *high probability of getting charged off*, hence *if loan amount/ Interest is l*

ney can be asked *to repay* the loan within a *term period of 36 months*.

nployment history of *more than 10 years*, have a *higher risk* of getting *charged off.*

gher Debt-to-Income Ratio, have a *higher risk* of getting *charged off.*

gher average revolving balance and revolving line utilization rates, have a *higher risk* of getting *charged off*

Delinquency in the past 2 years, have *a higher risk* of getting *charged off.*

present in *the earlier Bankruptcies records,* have a *higher risk* of getting *charged off.*

lied for a loan in *2007* or *2011,* have a *higher risk* of getting *charged off.*

ong *to States* like *NV* and *FL,* have *a higher risk* of getting *charged off.*

lusion:

loration:

e data about the Lending Club's loan offers and the factors impacting the approve or deny the loan application.

nsights:

ends and patterns related to loan offers for further analysis and decision-making.

endations:

commendations to minimize the possibility of financial losses in the approval of loans.

THANK YOU!