Lending Club Case Study

Submitted By:
Dilip Chauhan
Digdarshan Sahoo

Contents

Case Study Overview

ABOUT:

• A Consumer finance company is into various types of loans to urban customers.

• The company is trying to reach a decision whether to give a loan to a certain individual or not, thus reducing two types of risks:

- If the person is likely to repay the loan, then not approving the loan will result in loss of business
- If the person is not likely to repay the loan, then approving the loan will result in financial loss

Case Study Overview

RISK ASSESSMENT:

- A company can either choose to accept or reject the loan application. While accepting the loan application, there are various scenarios
    - The customer has fully paid the loan
    - There is a current loan on going
    - Charged off: the customer has missed or delayed some payments for an existing loan

OBJECTIVES:

- Based on data analysis, we try and predict which parameters about the customer such as age, monthly income, loan amount may be possible indicators to indicate as to which customer should be given a loan.

Reading the Data and Understanding

- The Data frame is read and columns are analysed
- Total rows and columns are 39717 and 111 respectively.
- There are a mix of columns ranging from categorical, numerical, date etc
- Our target columns is **loan status** which we need to compare and check for defaulters across various parameters.

Data Cleaning – Key Considerations

• **NULL Values:** Several columns such as verification_status_joint, annual_inc_joint etc. consist of only NULL values and are dropped

• **Single values:** initial_list_status, application_type, acc_now_delinq have only one NON Null values and hence these are dropped

• **Non relevant columns:** Columns which may not be useful for our analysis can be dropped as they provide no value

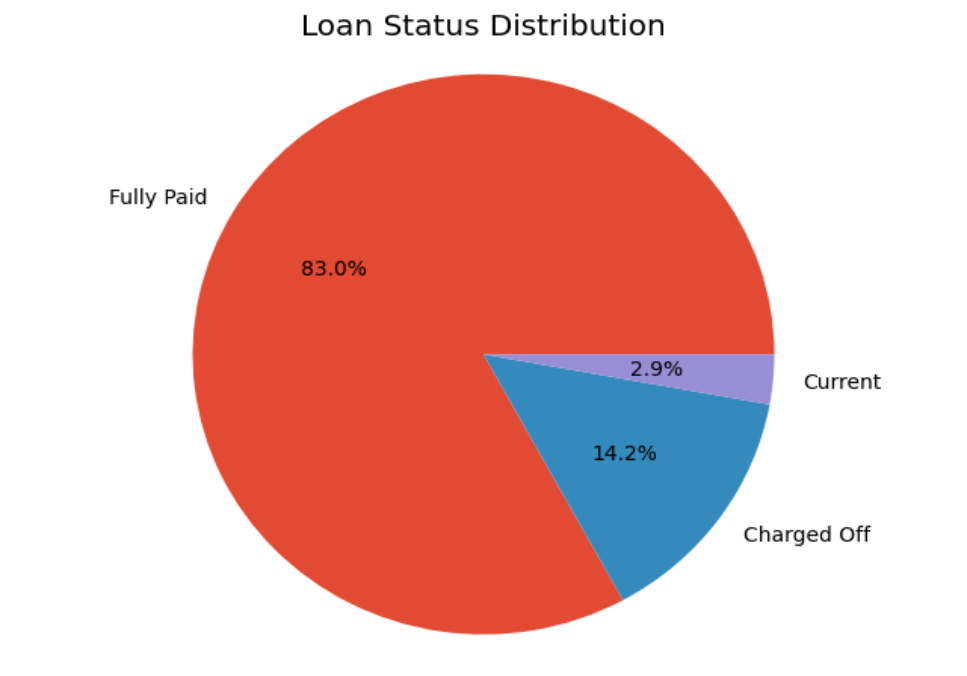
• **Outliers:** We analyse data for outliers to see if certain values may affect our overall interpretation of a certain column

• **Derived Columns:** Some columns such as issue_d may be further used to find year wise or month wise data which may be more relevant in scenarios.

We only consider data for loan status not equal to current which means we are only checking data for those loans which are not currently ongoing.

Analysis

Checking Categorical variable loan_status to see number of charged off customers:



- We consider only **charged off** data

Analysis Contd ..

Though most of the data look similar below are some observations for bivariate Analysis

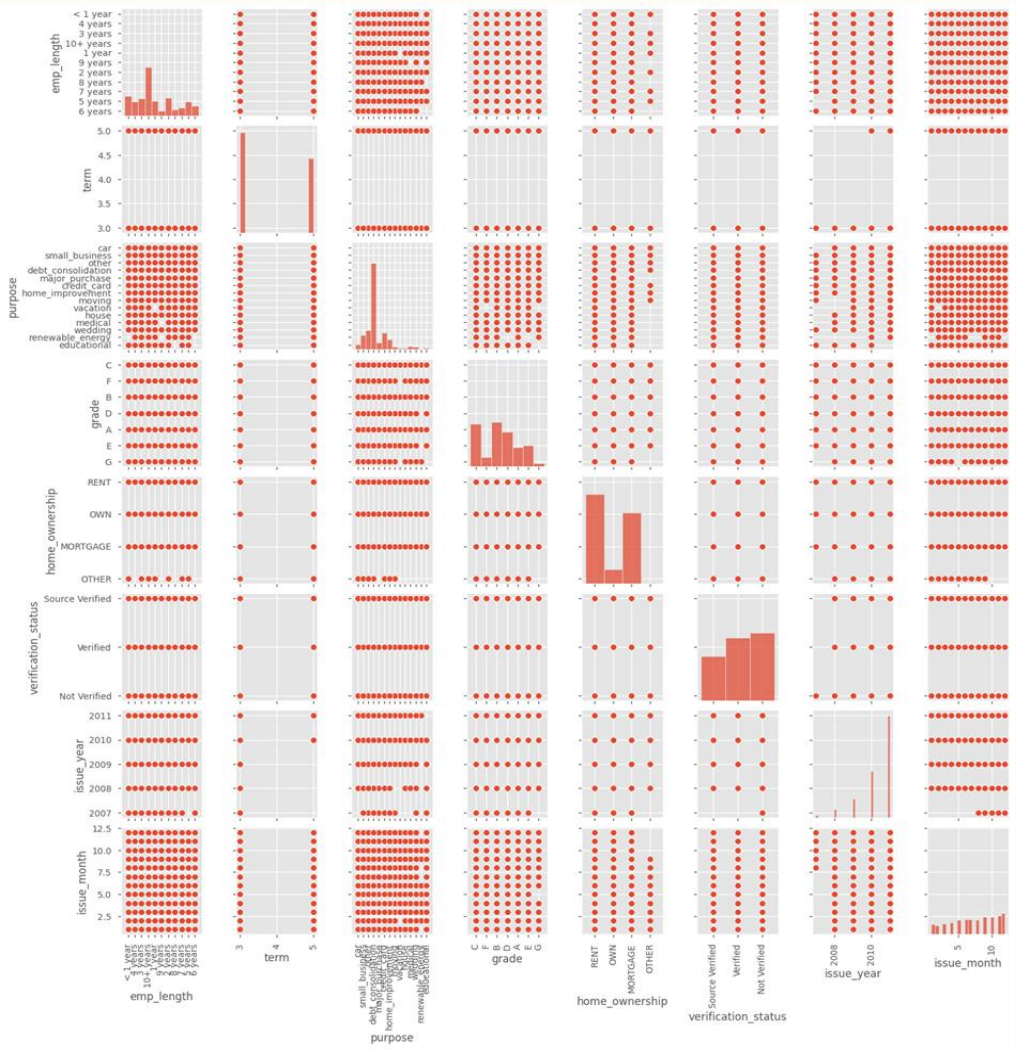
- Customers with 3 years Term have defaulted the most

- Number of Instalment increases with Issue year, Funded amount Invoiced and interest Rate,

- Stagnant loan disbursement amount in initial years

- Loan amount increases by ~5k in 2011 onwards

- Year 2007-2008 annual income of customers are more or less same; even after 2011, with upward trend in year 2009.



Analysis Contd..

Analysis Summary - *Drivers* where customers have defaulted

• **Loan Issue Month:** Customers defaulted where *'Month of Loan Issue' is December.*

• **Loan Amount:** Customers defaulted where '*Loan Amount'* falls in the range of 6k-8k.

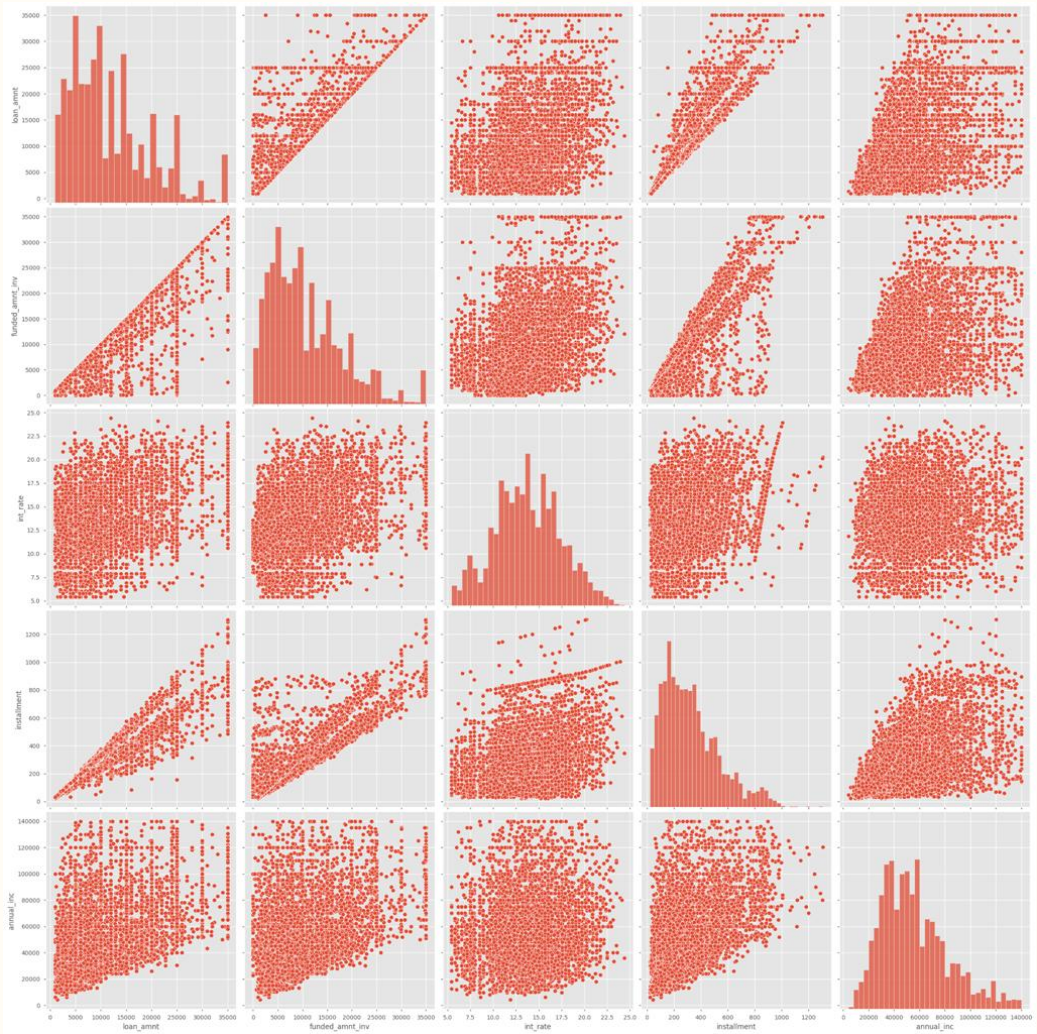
• **Rate of Interest:** Customers defaulted where *'Rate of Interest'* falls between 11% to 13%.

• **Loan Amount Invoiced:** Customers defaulted where 'Loan Amount Invoiced' falls in the range of 6k-8k.

• **Debt Obligation(dti):** Customers defaulted where 'Debt Obligation(dti)' ration falls between *11 to 17.*

• **Number of Credit Line:** Customers defaulted where their 'No of Credit Line (open_acc)' falls between *6 to 9.*

• **Revolving Line Utilization:** Customers defaulted where their 'Revolving Line Utilization rate(revol_util)' falls between *65 to 85.*



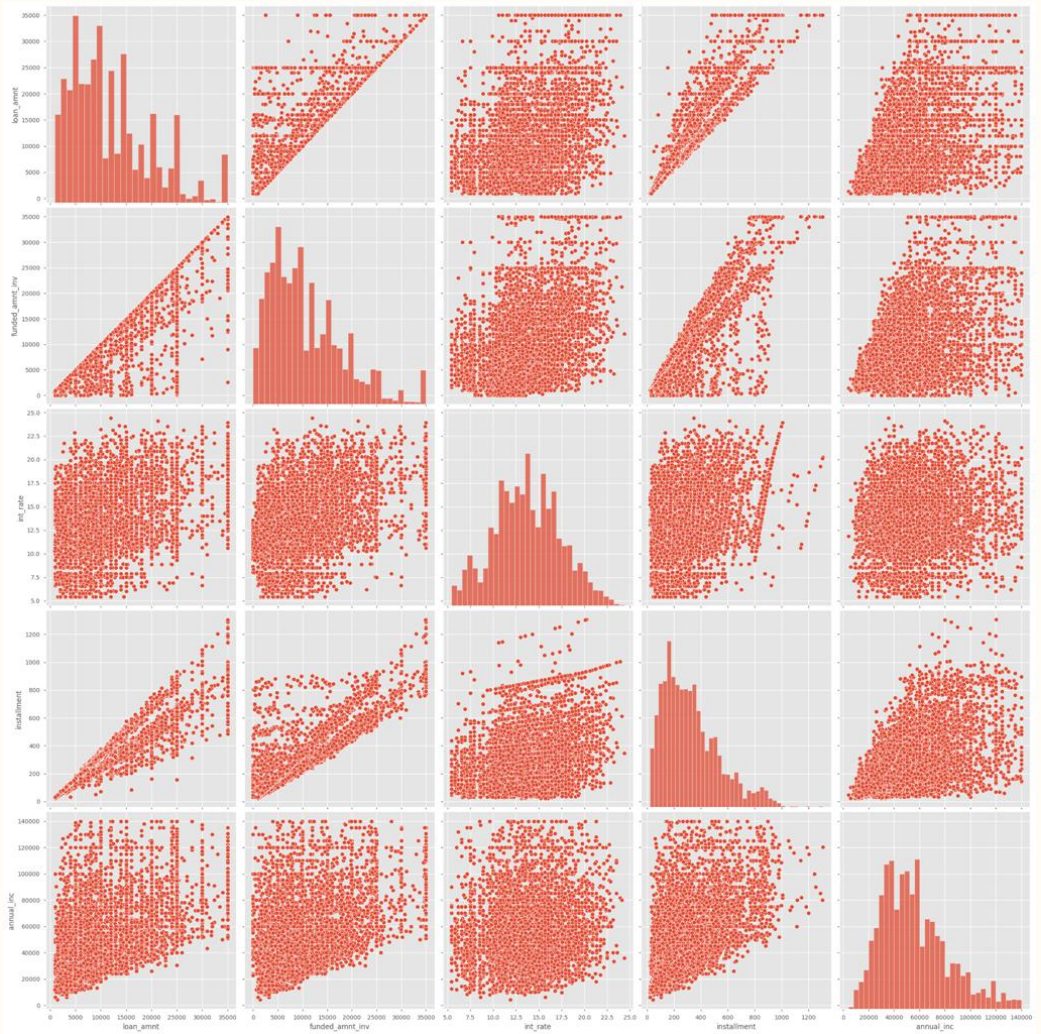
Analysis Contd..

Analysis Summary: Further Observations:

- **'Loan Amount'** decreases with an increase in *'Invoiced Funded Amount'.*

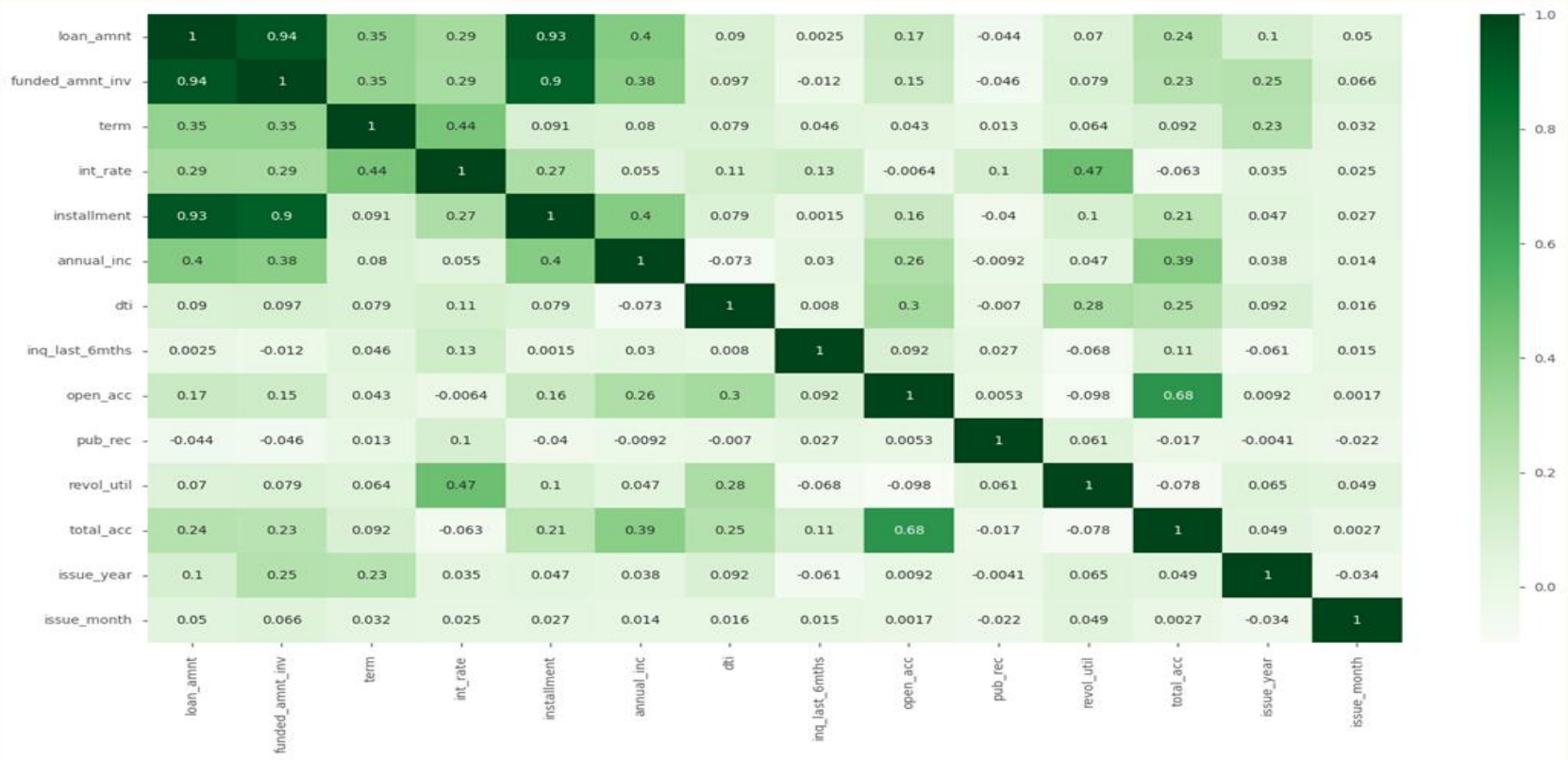
- **'Number of Instalments'** increases with an increase on in '*Invoiced Funded Amount*'.

- '**Loan Amount'** increases with increase in *'Number of Instalments'*



Analysis Contd..

- A heat map is drawn to check correlation between various columns in the dataset:



Key Observations

Scenarios where customers defaulted:

• Customers with **'Mortgages home'** and higher annual income in year 2008 - *defaulted the most*.

• Customers with **'Mortgages home'** and higher annual income in year 2011 - *defaulted the most.*

• Customers with **'Other home'** with higher Loan Amount in year 2011 - *defaulted the most*.

• For customers with **'Mortgages home'**; Funded Invoiced Amount, Number of Instalments, Rate of Interest, Loan Amount, Revolving Line Utilization Rate rapidly increased every year. Thus increasing the chances of rate of 'Defaults'

Scenarios where loan was fully paid

• In year **2007** customers with **'Highest Income**' FULLY_PAID the loan *in 'renewable energy*' sector.

• In year **2009** customers with **'Highest Income**' FULLY_PAID the loan *in 'renewable energy'* sector.

Key Observations Contd...

Customers with highest income lived in all types of homes.

- Customer's Annual Income is highest where they have a **'mortgaged home'** and loan application was verified.
- Customer's Annual Income is highest where they have a **'Other home'** and loan application was Source verified.
- Customer's Annual Income is highest where they have a **'Own home'** and loan application was verified.
- Customer's Annual Income is highest where they have a **'Rented home'** and loan application was verified.

Conclusion & Recommendations

Drivers for "Loan Default" and "not to Default"

Who is **NOT likely** to Default?

- Customers who 'FULLY_PAID' their loans are **less** likely to *Default.*

Who is **likely** to Default?

- **"Grade":** Customers with assigned Loan in ***"Grade B".***
- **"Loan Amount":** Customers whose loan fall in range of ***6k-8k***.
- **"House Ownership":** Customers staying in ***Rented or Mortgaged*** home.
- **"Loan Term Period":** Customers with average ***3 years of loan term***.
- **"Years of Experience":** Customers with work experience of ***10+ years***.
- **"Loan Verification“:** Customers whose loan application is ***'not verified'***

References

- https://www.youtube.com/watch?v=fiQhxn9RjEQ
- https://www.youtube.com/playlist?list=PLupD_xFct8mFDeCqoUAWZpUddeqmT28_L
- https://pandas.pydata.org/docs/user_guide/basics.html
- https://matplotlib.org/stable/tutorials/pyplot.html#sphx-glr-tutorials-pyplot-py
- https://seaborn.pydata.org/tutorial/introduction.html

Thank You