Lending Club Case Study

SUBMISSION

Group case Study:
Members/Contributors
Name: Piyush Saini & Shaishaw Shashank

<Abstract>

Lending Club company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.The company receives a loan application, eventually company has to make a decision for loan approval based on the applicant's profile.

Lending Club company has few constraints over the analysis:

- Borrowers can easily access lower interest rate loans through a fast online interface.
- Identify the risky loan applicants.
- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- Disbursal of such loans can be reduced thereby cutting down the amount of credit loss.

<Problem solving methodology : EDA>

Identification of such applicants using EDA is the aim of this case study.

Step by approach to Perform Case study:

1. Data understanding : Reviewing the dates to visualise each column and select the objective target variables.
2. Data cleaning (cleaning missing values, removing redundant columns etc.): Cleanup the data set to make data more meaningful.
3. Data Analysis:
    a) Univariate analysis : Follow the checkpoints and appropriate analysis on the target variables.
    b) Bivariate analysis : Look at the relations between the column-set, for any meaningful information
    c) Derived metrics : Extract hidden informations out of the dataset for further analysis.
4. Plot the metrics to visualise the outcomes.
5. Recommendations

<Data Understanding and Data Cleaning>

The objective is to understand the dataset set with respect to provided data dictionary, and visualise the target variable and other important columns to be useful in further stages of analysis.

- Dropping of redundant, missing data and extra columns.
- Standardising the columns.
- Removing impurities (wrong values) from dataset to make it more meaningful.
- Deciding Target Variable : Observing the dataset we can clearly determine **loan_status** as our target variable

<Analysis>

Starting the analysis with univariate analogy on Target variable I.e loan_status in our case. Here we observed that any user/customer with Charged-off value is considered as defaulted.

By considering the above analogy we have started our analysis in the same direction.

- Lets observe total number of defaulted customers in the dataset Fig -1 ~ 15%
- We will approach our univariate analysis on different set of variables with respect to loan_status.
- We'll segregate the univariate analysis with multiple sets of variables and drive the correlation .
- Fico score is one of the deciding factor for loan amount, interest rate

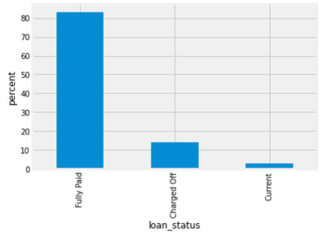


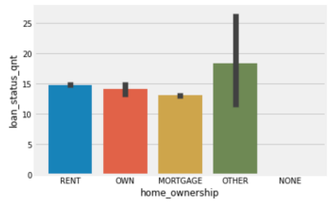
Fig.1: Loan Status Distribution for all set of customers

<Univariate Analysis>

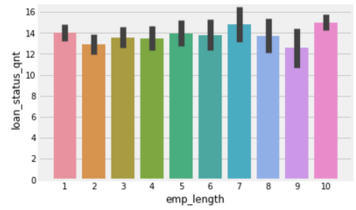
Deriving relation between set off variables and default trends.

- To visualise and observe the relation, we have some bar plots:

  1) First we observe the variables which doesn't have much impact for increase in default rate.



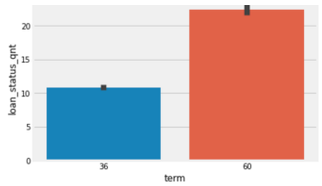
Home ownership does not have any impact on charged off rate



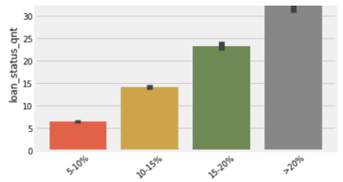
Employment term does not have any impact on charged off rate

* Now with these plots we can clearly right off that home ownership and employment term doesn't have much impact in the default or loan charged off.

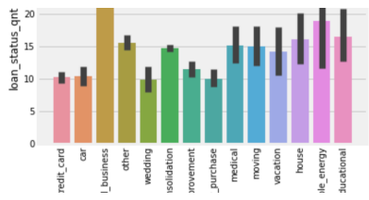
3) Let’s observe the relation between variable that have impact on deciding charged-off or default. (5-Plots - bivariate analysis with charged-off/default rate)



*** term compared with charged off rate higher term has major impact > 20% charged off rate**

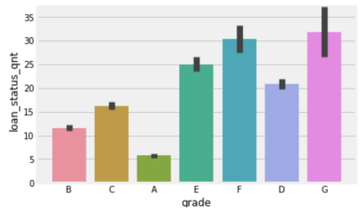


*** Charged off or loan default rate increases as rate of interest goes up**

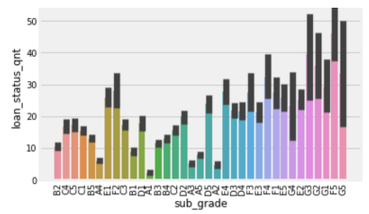


***Charged off or loan default rate is higher when loan is taken for small business**

* Fico score are used by many lenders, grading and sub grading is a most prominent driving factor as credit score. Lets observe Grade and sub grade together.



As grade is lowered charged off or loan default rate goes up significantly



With lower sub grade charged off or loan default rate goes up significantly

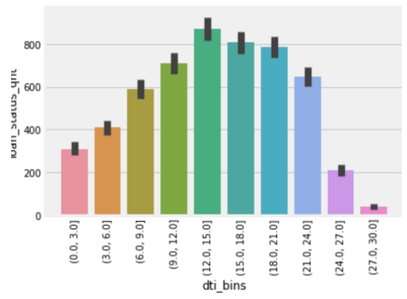
Checkpoint & Observations

We have analysed the impact of third pointer variables on charged-off /default rate, with this we concluded :

- All of the 3 variables (term, interest rate, loan purpose) are one of the indicating factor for charged off/default.
- Loans with higher terms are more likely to default.
- Loan with High Interest rate is another strong indication of default.

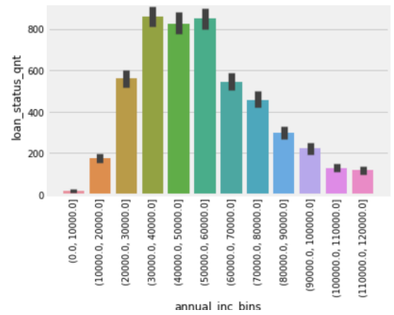
**) Resume Analysis with factors/variables derived from borrowers financial stability.

A) dti analysis vs loan default rate/charged-off



With increase in DTI loan default rate seems to go up till 24 and it decreases thereafter.

A) annual income analysis vs loan default rate/charged-off



Annual income does not much indication, however a high annual income group is less likely to default.

Multivariate Analysis

- As per our analysis with univariate and bivariate approach on different sets of variables, we didn't reach at conclusion on driving factors for default loans.
- Lets further resume with multivariate analysis and heat map plots.

A) Heat map on Grade/Purpose vs default rate:

Observation:
- Default rate is high in case of Grade G and purpose of loan as car, medical, small business and renewable energy loans

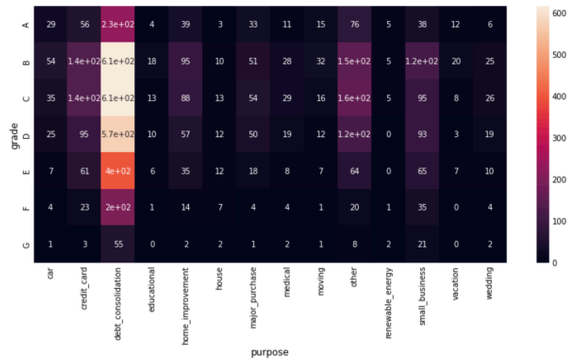


Fig: HeatMap

B) Heat map on interest rate and Purpose vs default rate: (Fig B)

Observation:
- Rate of default is very high when interest rate is high > 15 and loan type is educational and small business, renewable energy, medical or house.

C) Heat map on loan term and Purpose vs default rate: (Fig c)

Observation:
- Educational loan with 60 months term is really bad performer with43% default rate.
- Small business with 60 months term is also very high default rate 35%.

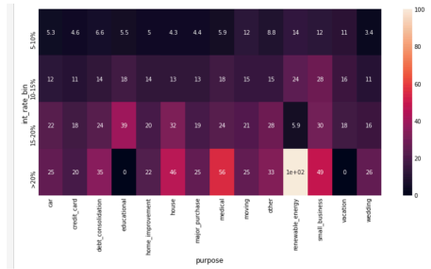


Fig b

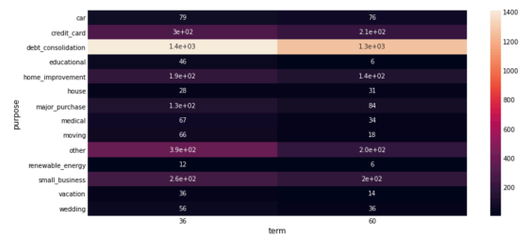


Fig c

D) Heat map on term and dti_bins vs default rate: (Fig d)

Observation:
- Loan default rate is consistently high in 60 months term.
- It increases further slightly with increase in DTI.

E) Heat map on term and loan_amt_bin vs default rate: (Fig e)

Observation:
- Loan default rate is consistently high in 60 months term.

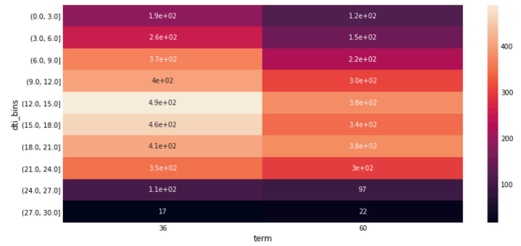


Fig: d

Fig: e

Recommendations:

With the case study analysis on different segmented of variables and plot, we observe few indications for a loan or borrower becoming default. However we've narrowed down our analysis to find out the most strong parameters as driving indications for default.

- Key Observations and Recommendations:

- *A high loan tenure is one of the most common indicator of loan default across all analysis.*
- *Low credit score (low grade and subgrade)with unsecured loans such as car, medical, small business and renewable energy loans have significant impact on loan defaults.*
- *High interest rate across all segments is primary indication of default.*
- *Increase in dti is more likely a risk of becoming default.*