EDA Gramener Case Study

Loan Default Analysis

Group Members

Harkirat Dhillon

Shubhra Karmahe

Vartika Tewari

Yogesh BS Raju

Problem Statement

A **consumer finance company** which specialises in lending various types of loans to urban customers wants to identify **loan default patterns** . 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

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions like -

- **denying the loan**
- **reducing the amount of loan**
- **lending (to risky applicants) at a higher interest rate**

Problem Solving Methodology

1. **EDA** is used to understand how **consumer attributes and loan attributes** influence the tendency of default.
2. The dataset used for analysis contains the information about past loan applicants and whether they 'defaulted' or not for year **2007 to 2011**.
3. **Univariate Analysis** is used on major consumer and loan attributes.
4. On the basis of **Uni-variate Analysis**,we have identified **driver variables**.
5. **Bivariate Analysis** is used to find correlation between driver variables.
6. To get more insights,we have included **03 derived metrics**.

Assumptions:-

1. There are 54 columns in dataset having all observations as NA. They were excluded from EDA.
2. There are 06 columns having redundant values for all observations . They too were excluded .
3. For this anlysis,we have left outliers and missing values as it was in raw file.(i.e no treatment and value imputation)

Analysis Result

Driver Variables:-

Borrower Attributes:-

1. Grade & Sub-Grade
2. Loan Purpose
3. Annual income

Loan Attribute :-

1. Interest Rate
2. Loan Amount
3. Loan Payment Term
4. Public Record & Bankruptcies
5. Delinquent Status

Note :- DTI is one of the major driver variable for loan approval.The dataset used for EDA have data for approved loans only,So we are not considering it as driver variable for this analysis.

Derived Metrics :-

1. **Return on Investment** (ROI)

Business Insight - Tells about return on funds invested by investors.

2. **Is_bad** indicator

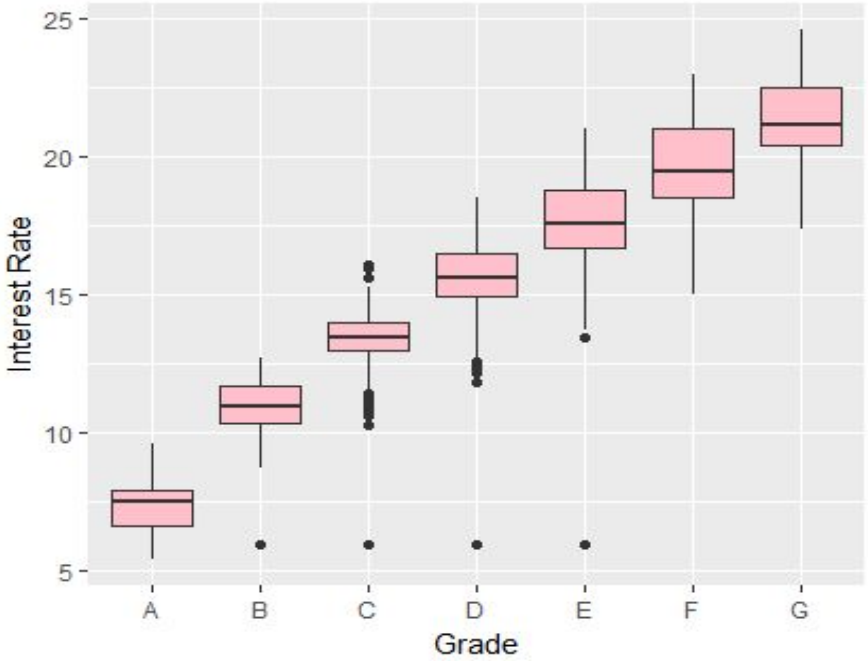
Business Insight - To predict whether the defaulted applicants were actually likely to default or not.

3. **Ratio of loan amount by annual income**

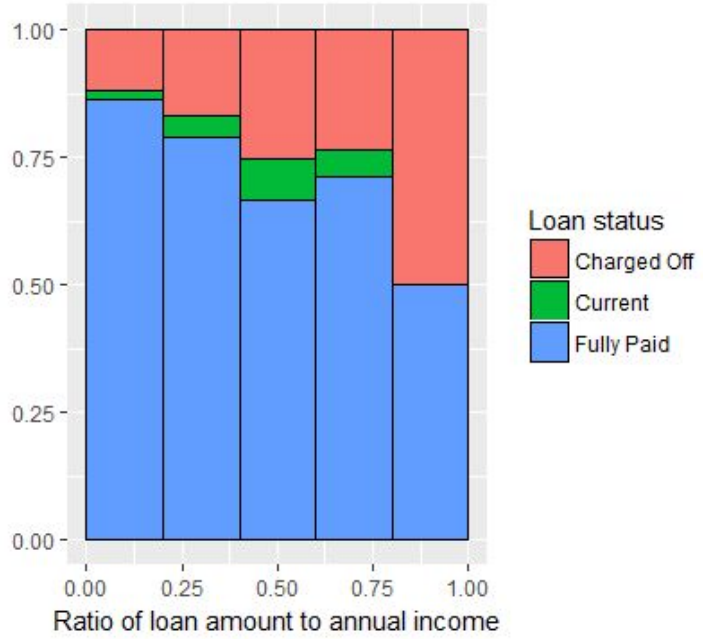
Business Insight - As the ratio of loan amount to annual income increases, chances of default too increases.

Visual Analysis - I

*The **interest rate** assigned depends on the **grade** assigned and the **higher the interest rate, the higher the default probability is**.*

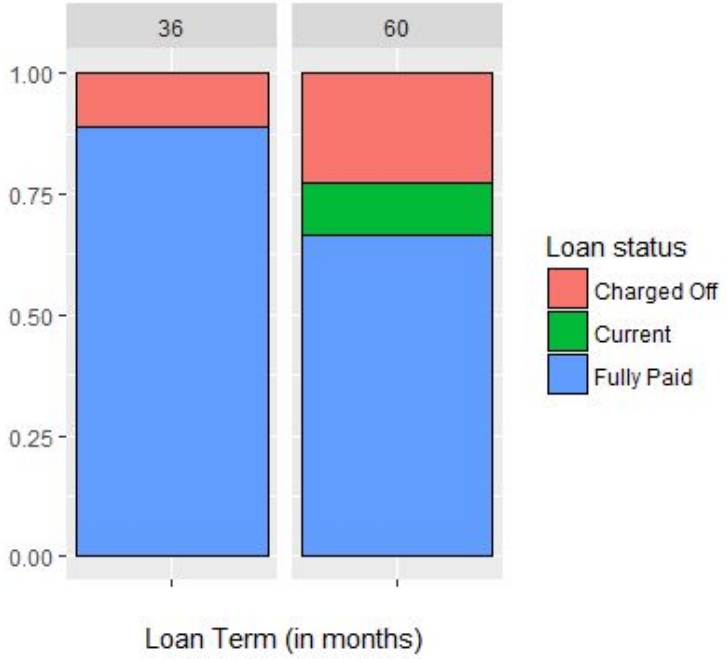


Higher** the ratio of **loan amount to annual income**, higher is the proportion of **defaulters.

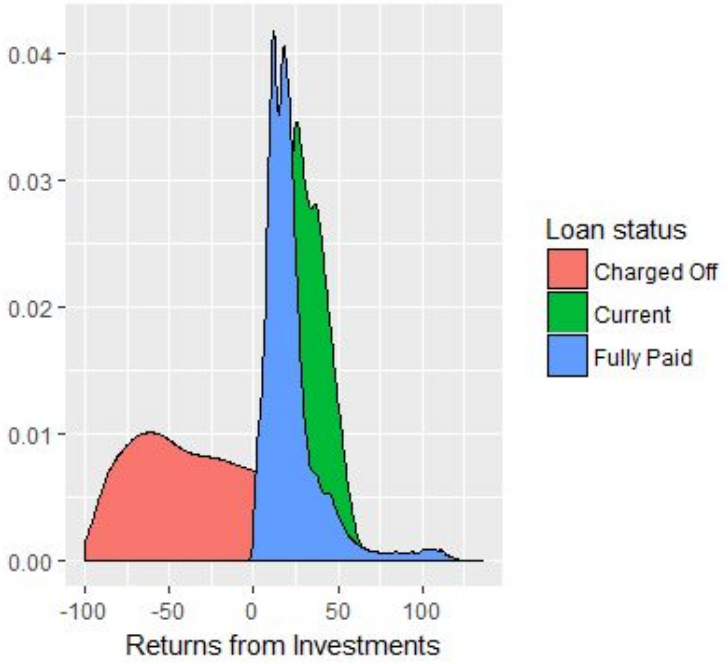


Visual Analysis - II

*Default rate is **high** for **longer loan terms** as installment amount to be paid is less as compared to **short term loan**.*

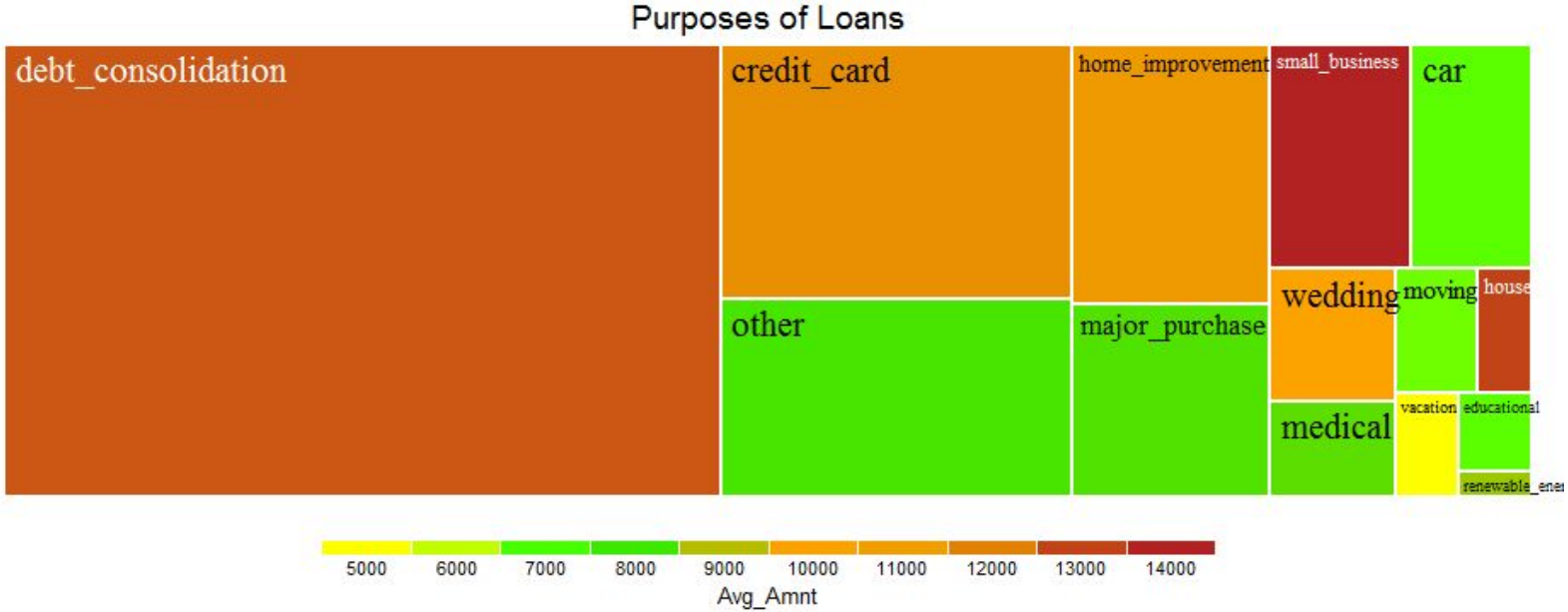


*Investors have a **negative ROI** for **charged off loans** (huge loss of money) & on the other hand ROI from fully paid loans is highest.*

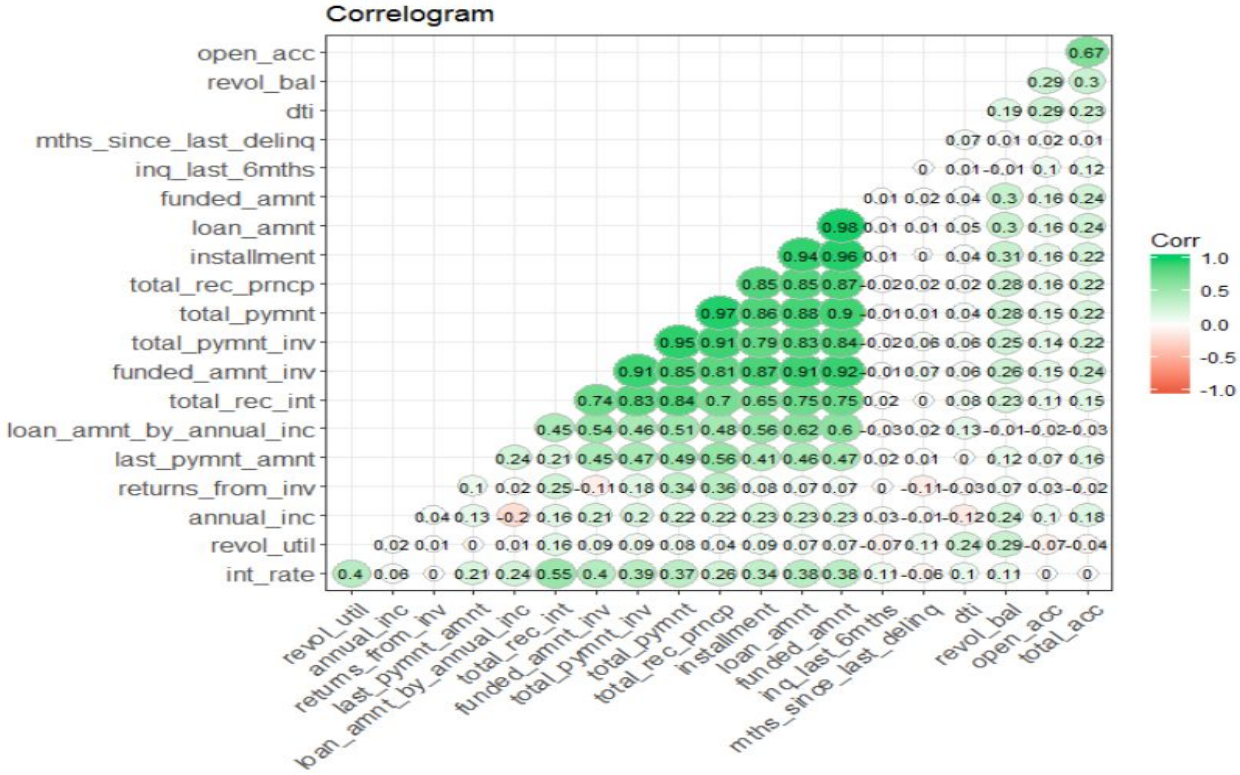


Visual Analysis - III

*Loan purpose is also a factor explaining default: wedding,**education** is the **less risky** loan purpose and **debt consolidation** is the **riskiest**.*



Visual Analysis - IV



Conclusion

Lending loans to **'risky'** applicants is the largest source of financial loss (called **credit loss**). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. We have understand how **consumer attributes and loan attributes** influence the **tendency of default.**

Driver variables (Grade,Subgrade,loan purpose,annual income,loan amount,interest rate,public records,payment terms) are the variables which are **strong indicators** of default. The company can utilise this knowledge for its portfolio and risk assessment.

Accordingly,they can decide on denying the loan, reducing the amount of loan or lending (to risky applicants) at a higher interest rate.

Peer-to-Peer lending model could possibly address the high interest rate paid by borrowers and low interest rate earned by investors.