Leading Club Case Study

By

Siddharth Das

Shivil Mangal

Jerrin Xavier

Mahesh Vemula

Business Objective

- *One of the largest online consumer finance company which is facilitating with personal loans, home loans, business loans, and financing of medical procedures would like to understand the ‘risky’ loans/applicants which would be the largest source of financial/credit loss i.e. when the borrower refuses to pay or runs away with the money owed.('charged-off’/ 'defaulters’).*

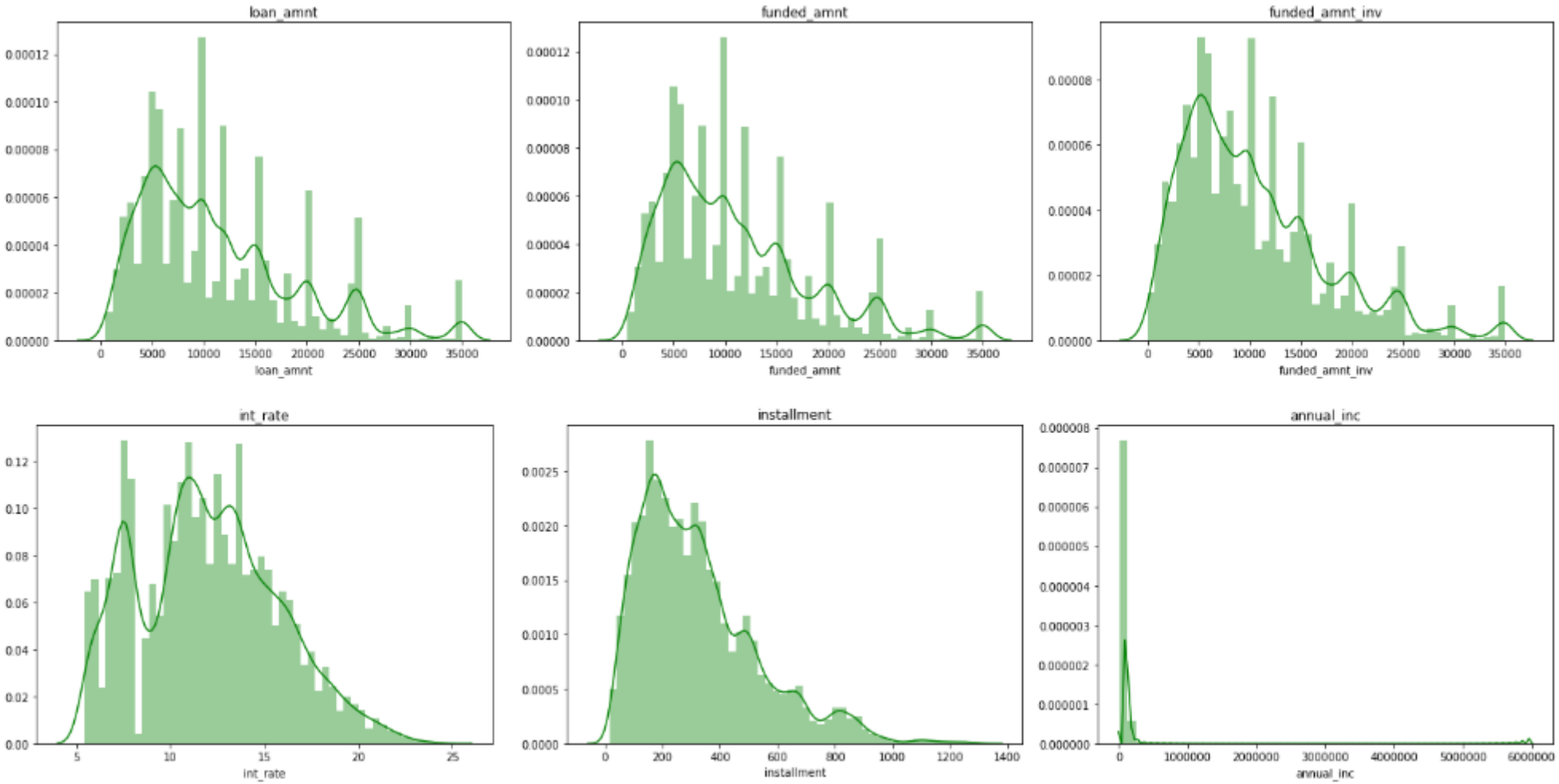
- *If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.*

- *In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.*

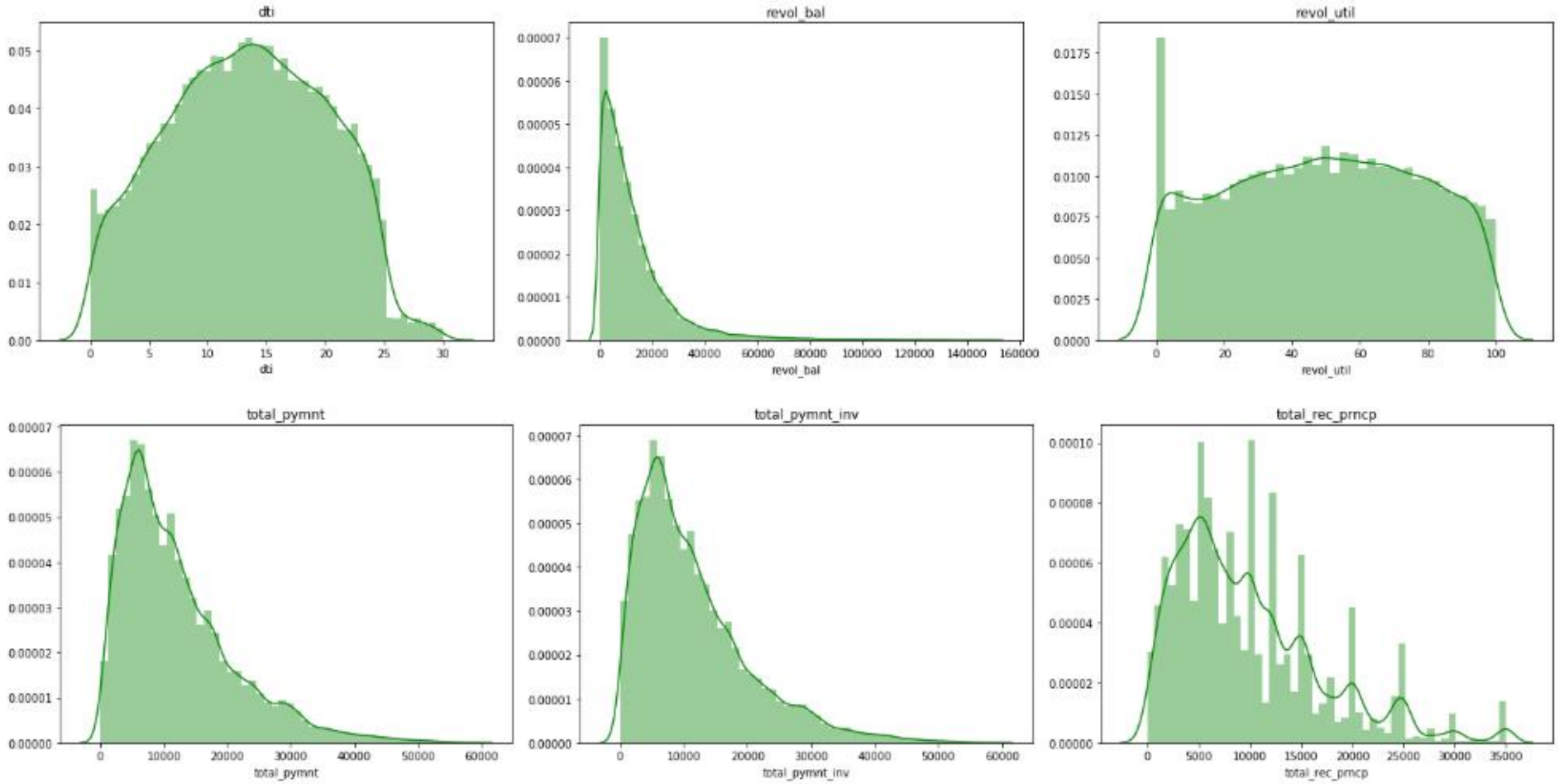
Analysis Approach

- Understanding of the Data
- Data Cleaning by removing NAN Columns > 30% and other non-important columns
- Understanding Number of Numerical and Categorical Features/Variable
- Data Distribution Analysis by plotting Graphs for both Numerical and Categorical Variables.
- Created required Derived Variables like Percentage of Bad Loans,
- Applying Univariate and Bivariate Analysis on Data, to understand the Factors driving to a Bad Loan.

Numeric Data Distribution Analysis - 1



Numeric Data Distribution Analysis - 2



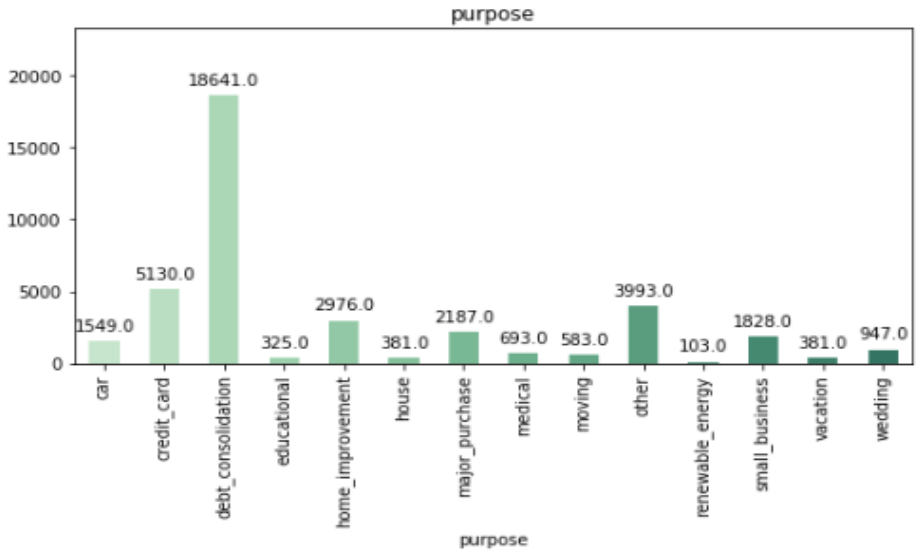
Observations from Numeric Data Distribution Analysis

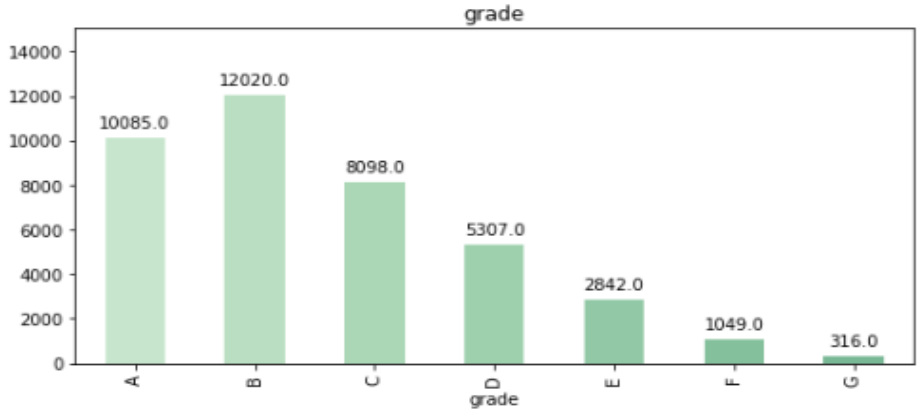
- *We can see from the distribution plot that loan_amnt, funded_amnt and funded_amt_inv has similar distribution.*

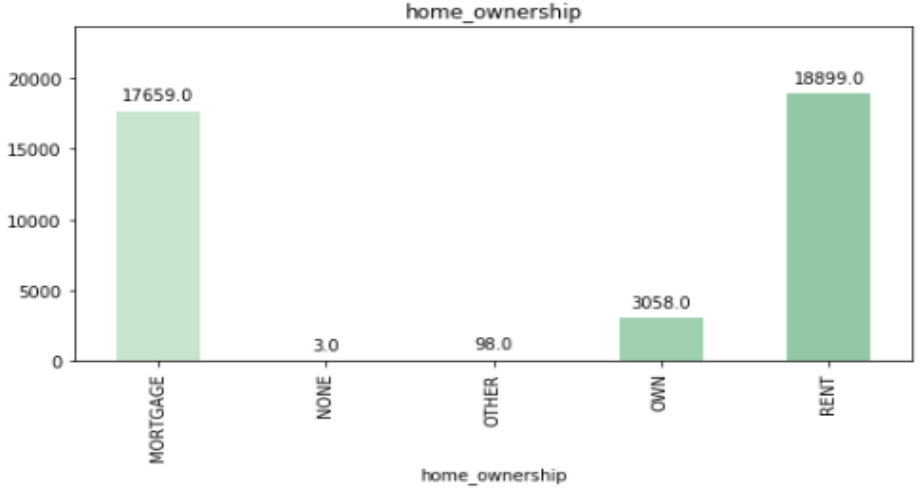
- *Income distribution is highly skewed toward the left.*

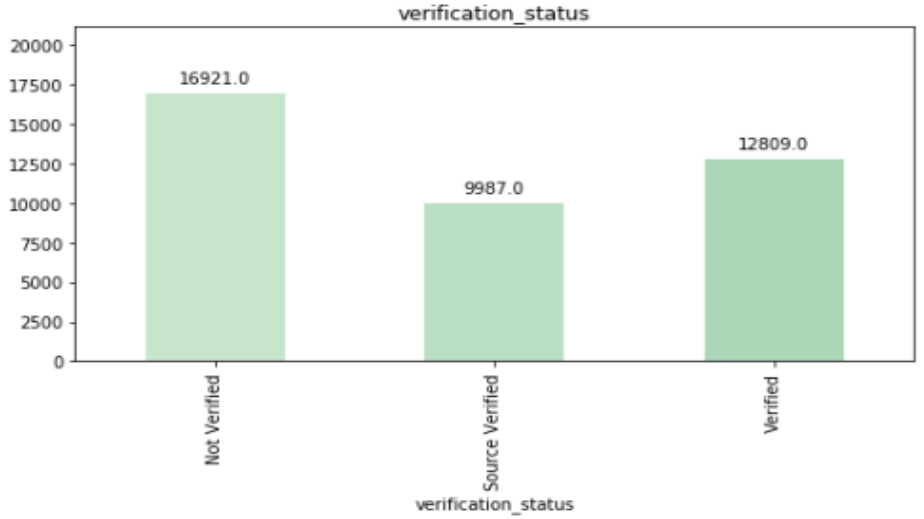
- *Most of the plots follows normal distribution.*

Categorical Data Distribution Analysis - 1

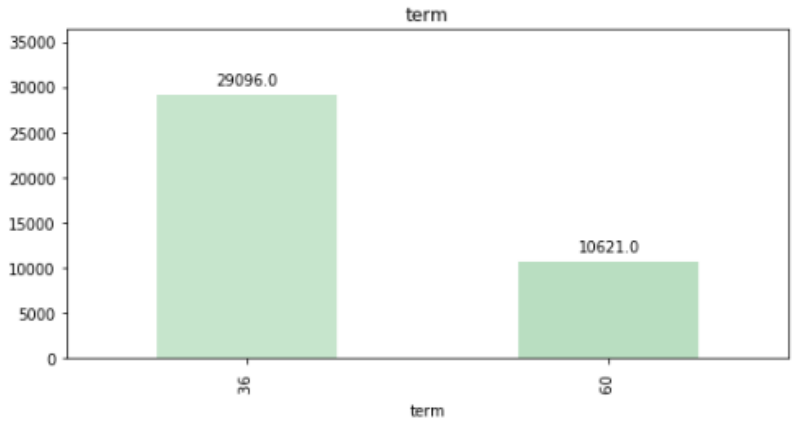


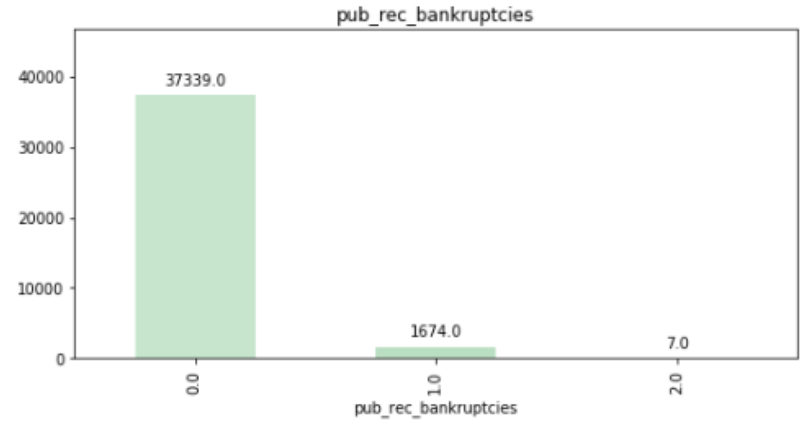


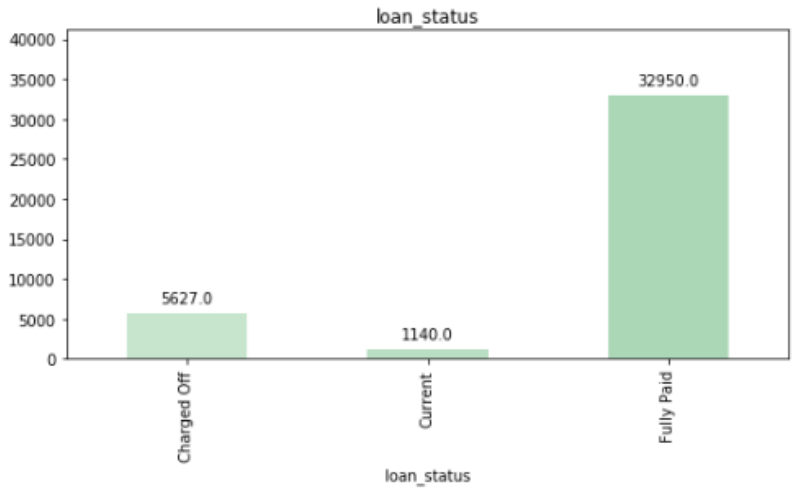


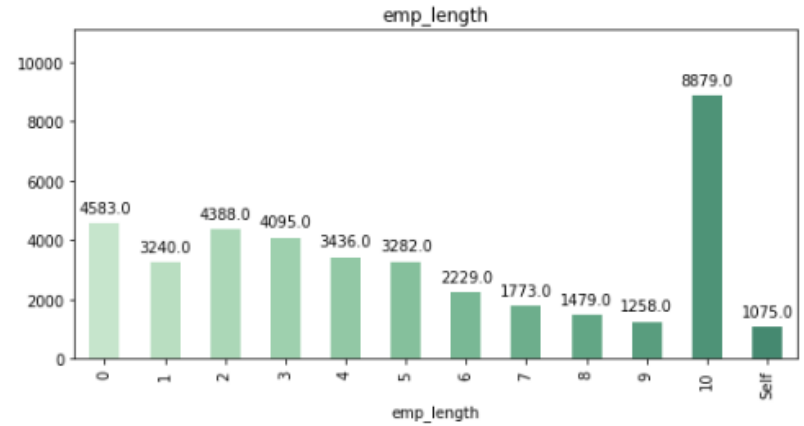


Categorical Data Distribution Analysis - 2







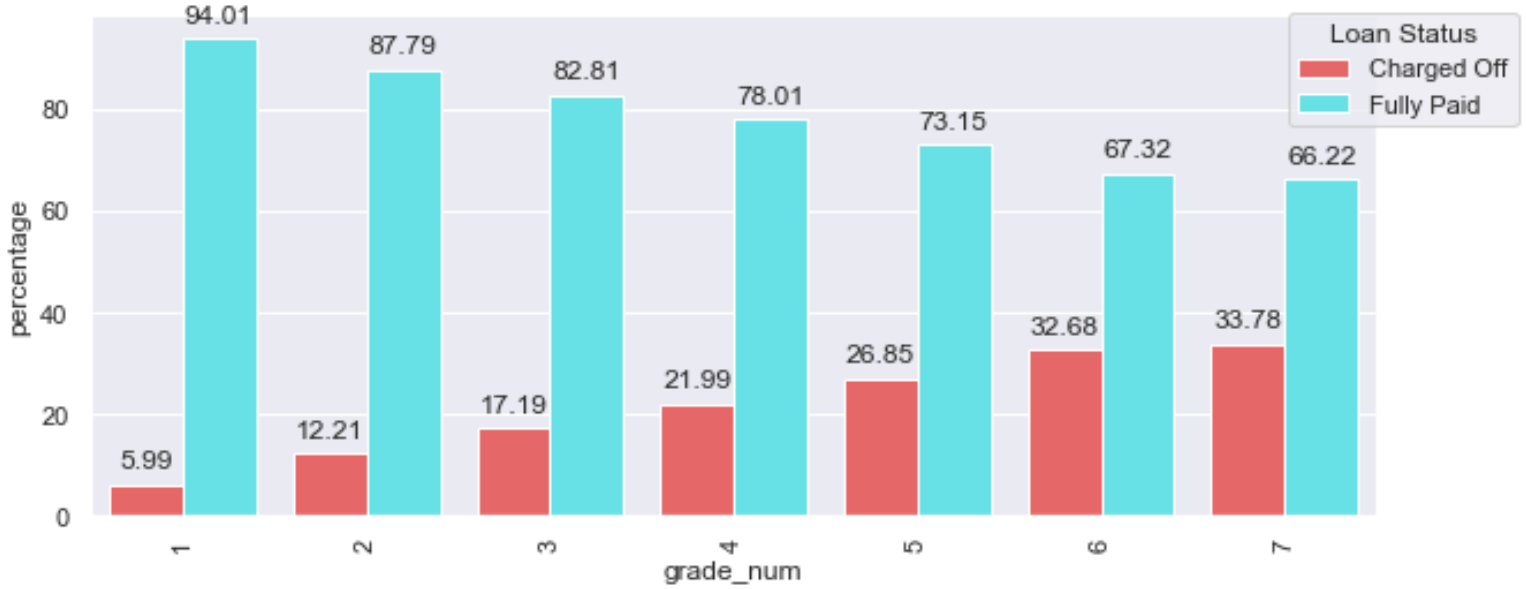


Analysis on Income Range

Insights:

- *Charged off loans are very high in very low-income range and very high-income rage.*
- *The default rate is almost 19% in people with an income of less than 20000 and people with income around million.*

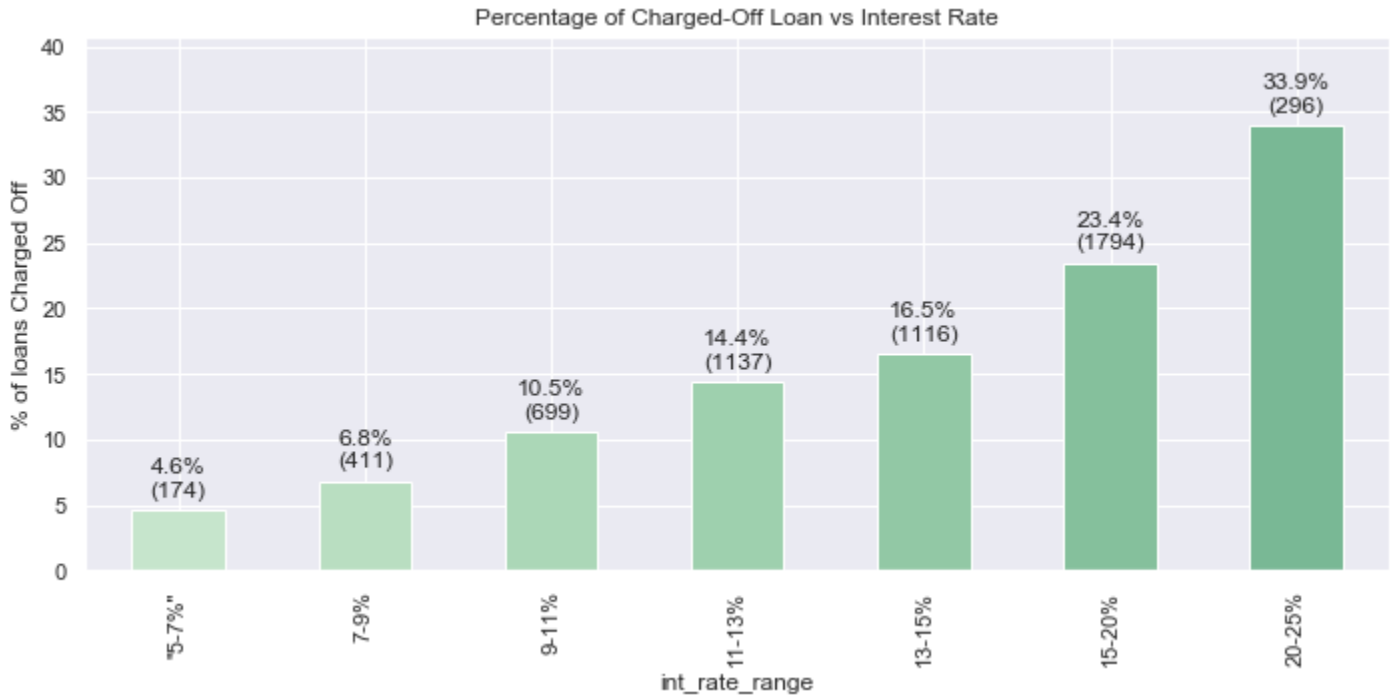
Analysis on Loan Term



Insights:

- *Loans are either given for 3 year or 5 year duration.*
- *Around 30,000 (74%) loans are given for 3 years.*
- *Out of total charged off loans 57% belongs to 36 months Term and 43% belongs to 60 month Term.*
- *Despite of giving more number of loans for 36 months, more money is disbursed for 60 months term.*
- *25% of Loans are getting charged-off for 60 month term, as compared to 11% for 36 month term.*

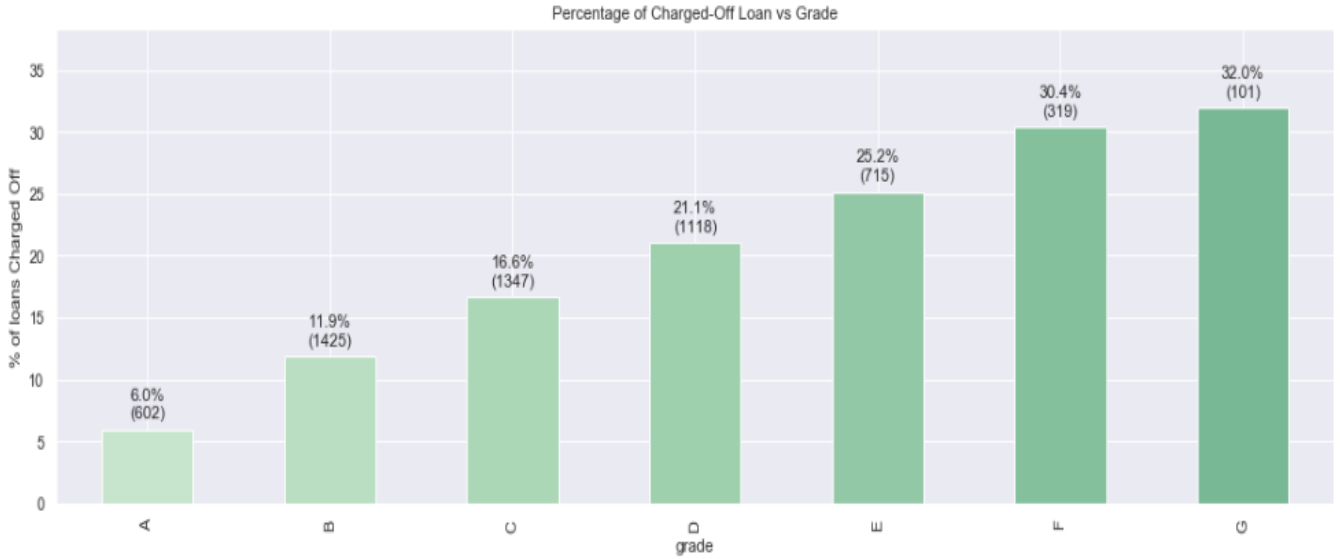
Analysis on Interest Rate

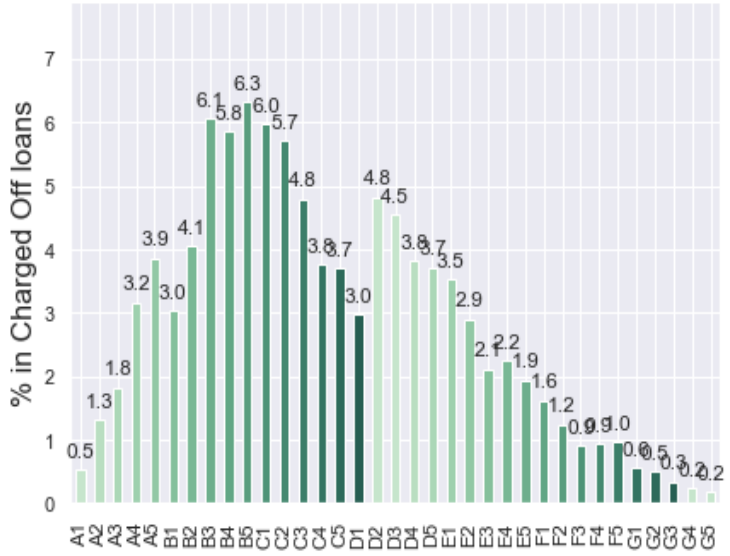


Insights:

- *Higher the interest rate for the loan, higher the probability of getting the loan charged-off*

Analysis on Loan Grade

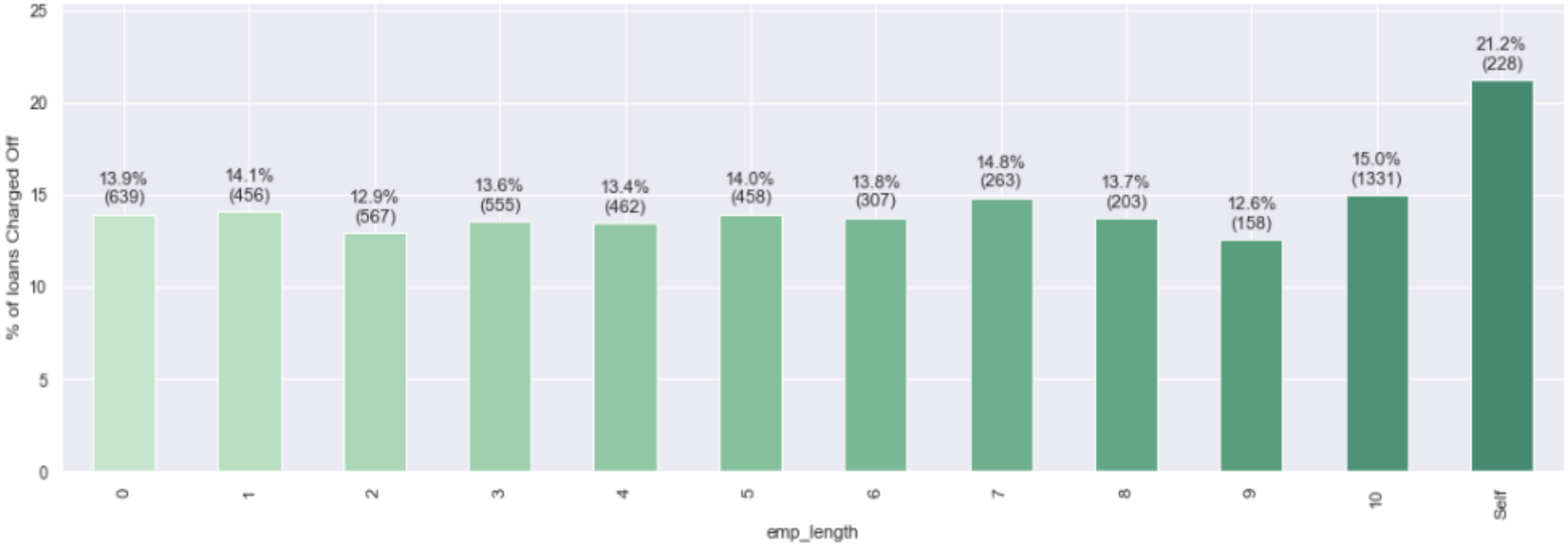




Insights:

- *Lending Clubs grading system able to recognise the customers who have a high probability of default.*
- *Grade A and B loans are safe.*
- *Grade D, E, F, G loans are prone to risk.*
- *About 30% of all loans in Grades F and G see a default.*
- *Mean and Median values are approximately equal and monotonically increasing as we move from A to G.*
- *As the grades increasing then risky becomes the loan and hence the interest rate is increasing.*

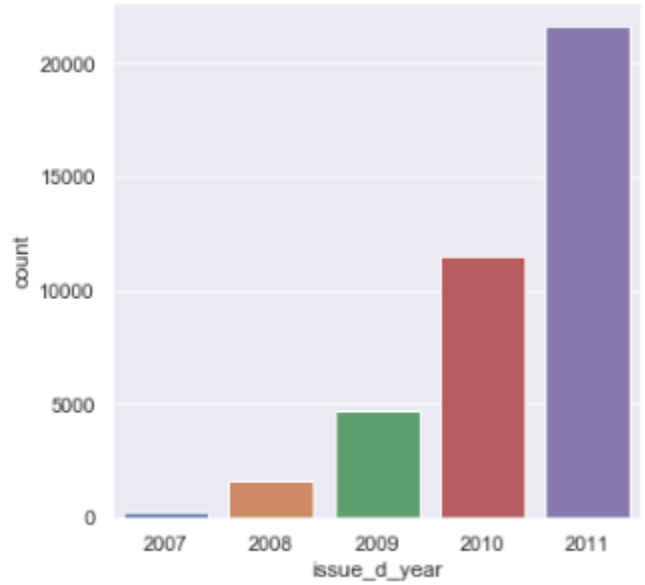
Analysis on Employment Length

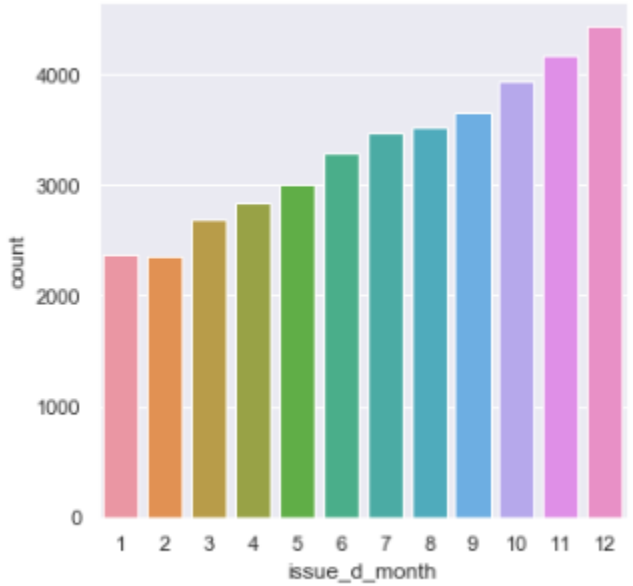


Insights:

- *If you are a self-employed person there is a high chance for default even though only 2.7% is the loan takers are self-employed.*

Analysis on Loan Issue Date

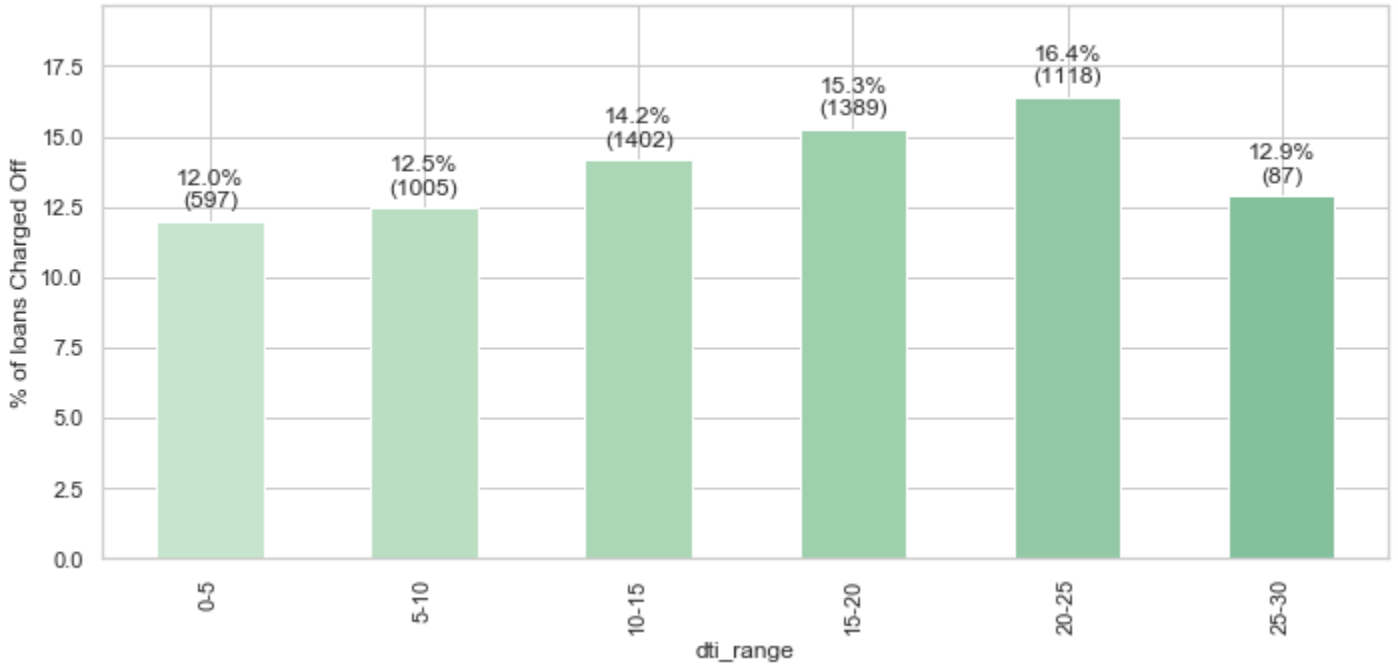




Insights:

- *Number loan issued significantly increases from 2007 to 2011.*
- *December is the month of the year where the maximum number of loans are being issued. It could be probably company trying to meet targets in the later months of the year.*
- *Year of loan doesn't have any impact on default.*

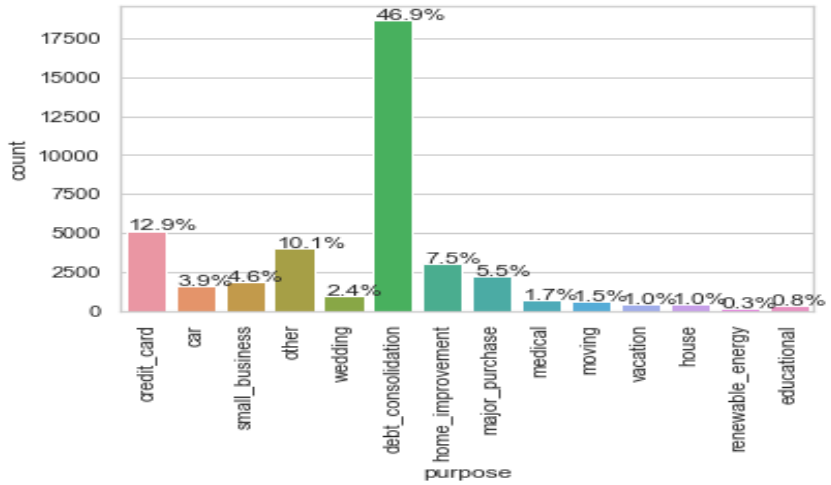
Analysis on Debt-To-Income Ratio



Insights:

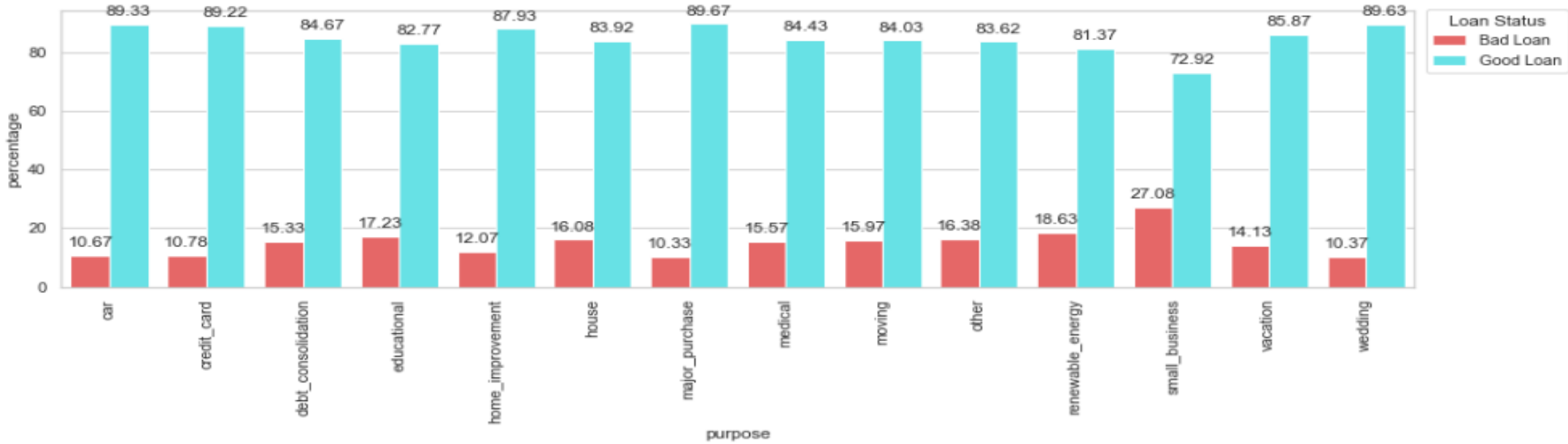
- *Initially when we started the analysis, dti (Debt to Income Ratio) seems like a very important factor, if debt to income ratio is more the customer is most likely to default the loan.*
- *The fully paid loans has lower dti as the debt amount has decreased as the customer paid their debt.*
- *Contrary, based on our analysis it seems like dti doesn't follow a pattern for good or bad loan. We can see this trend based on the plots which are plotted against dti*

Analysis on Purpose

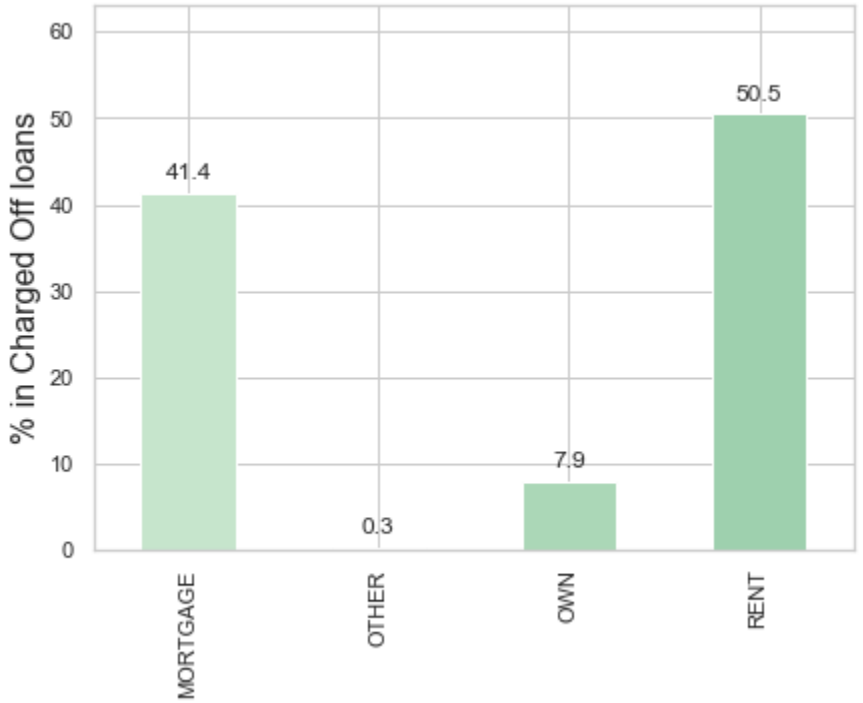


Insights:

- *Around 50% of loans are taken for Debt Consolidation, i.e. with this borrowed money the customers want to preclose all other loans.*
- *Higher Loan Amount are taken for Debt Consolidation and small business.*
- *Around 25% loans taken for Small Business defaults*



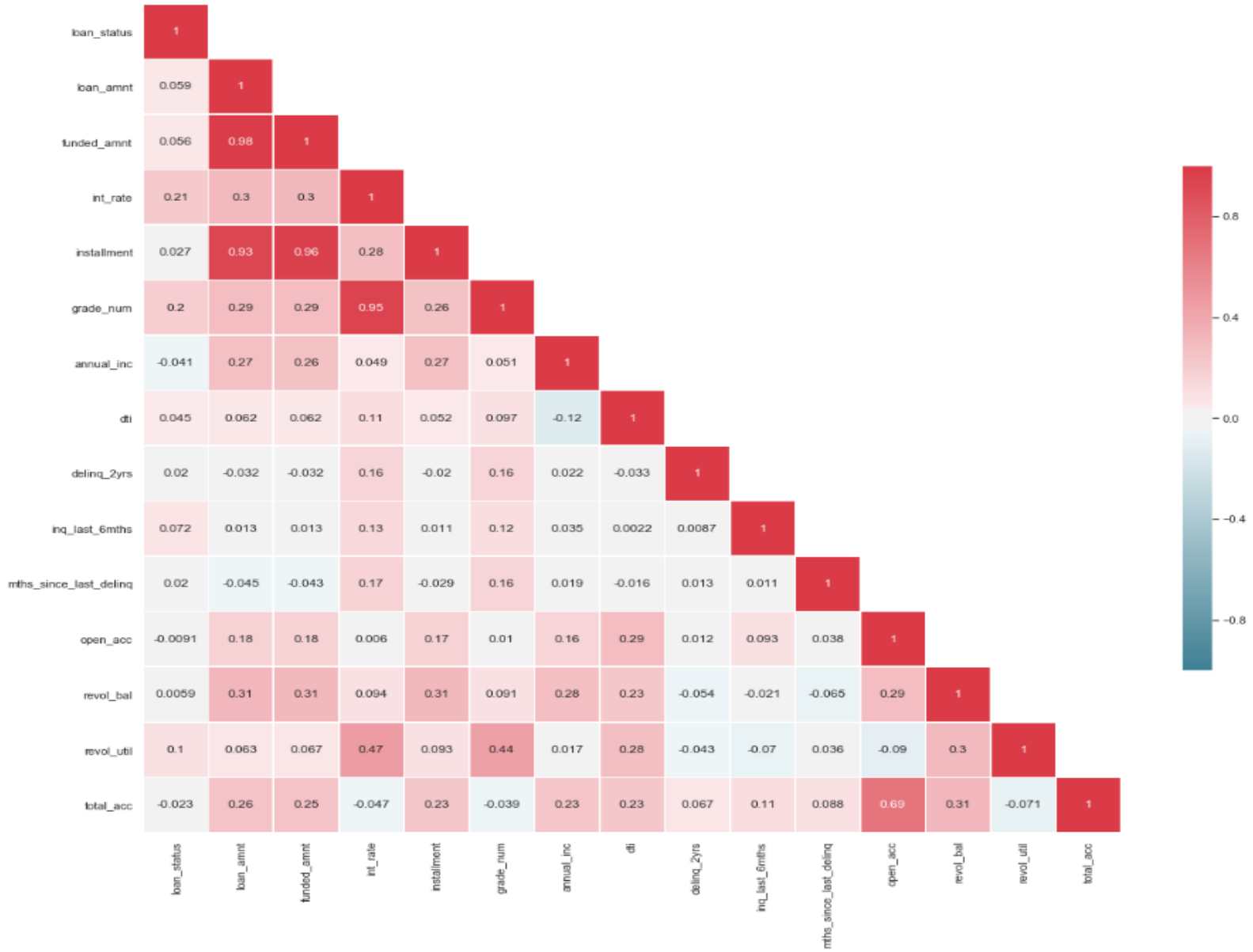
Analysis on Home Ownership



Insights:

- *If the customer is either paying mortgage there is higher change for charge-off as a considerable amount of income is spent on rent/mortgage*

Correlation Matrix



Conclusion

From the overall EDA Analysis for Lending Club Case Study, we consider the below are the Driving Factors for a Bad (Charged-Off/Defaulter) Loan.

- **Loan Grade:**
  Higher the loan grade higher is the interest rate and higher the defaults.
  About 30% of all loans in Grades F and G see a default
- **Interest Rate:**
  Higher the interest rate for the loan, higher the probability of getting the loan charged-off.
  Around 34% loans are defaulted for interest rate > 20%
- **Home Status:**
  If the customer is either paying mortgage or rent there is higher chance for charge-off as a considerable amount of income is spent on rent/mortgage
  More than 40% times the loan is defaulted if the borrower has mortgage/rent on house
- **Term:**
  25% of Loans are getting charged-off for 60 month term, as compared to 11% for 36 month term.
- **Purpose:**
  Around 50% of loans are taken for Debt Consolidation, and around 25% loans taken for Small Business defaults
- **Employee Length:**
  23.7% of the of loan taker has an employment length of more than 10 years.
  If you are a self-employed person there is a high chance for default even though only 2.7% is the loan takers are self-employed.