Lending Club – Case study

SUBMISSION

Name: Ashwin Prabhu

Nishant Kumar Behl

Lending Club – Case Study

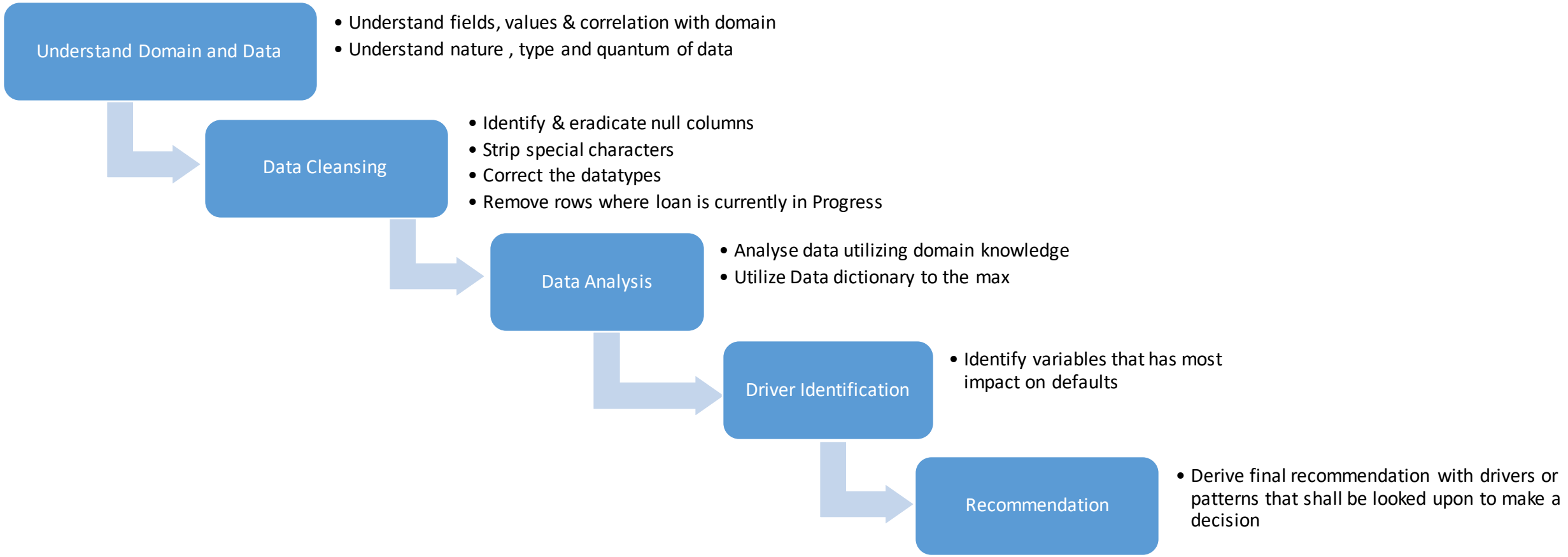
BUSINESS OBJECTIVE: When the company receives a loan application, the company would like to identify risky loan applicants, take a decision for loan approval keeping in mind its portfolio and risk appetite.

BUSINESS STRATEGY: Identify driving factors and patterns that indicate applicant is likely to default, which will in turn be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

APPROACH:

- Develop Domain understanding
- Understand data dictionary
- Perform data sanitization
- Perform EDA on sourced data:
  - Univariate analysis
  - Bivariate analysis
- Carve out 5 variables that has most impact on default

Analysis Approach



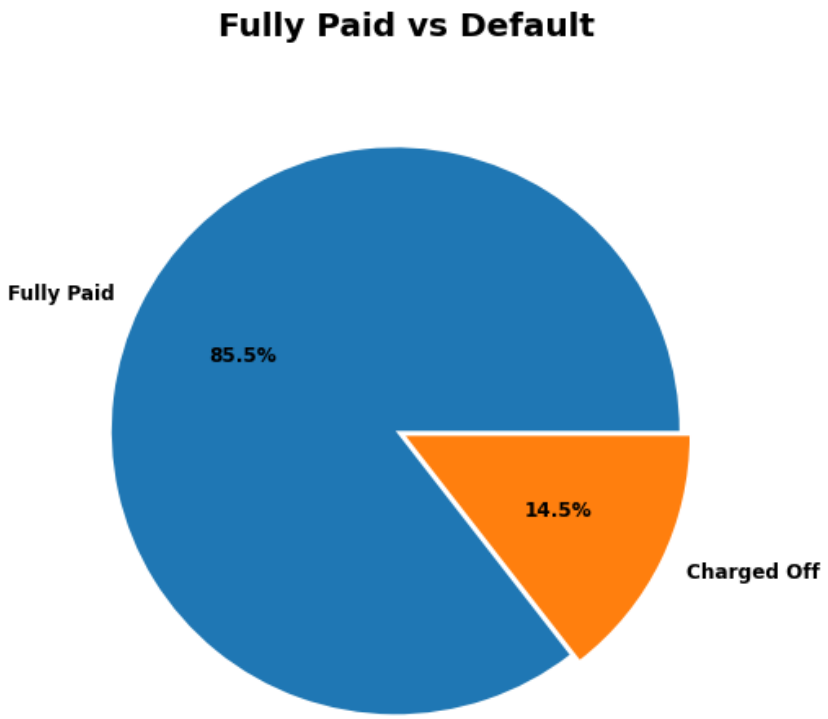
Data Cleansed

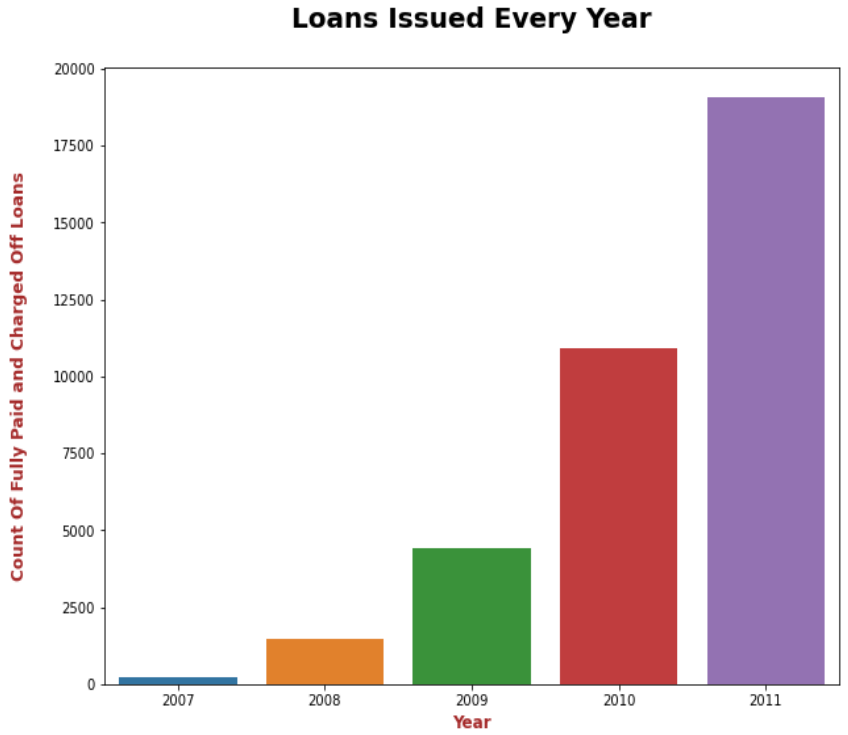
- 56 columns dropped as these columns had 100% null values
- Columns with high volume of null values have been dropped
- Columns with only single\unique value has been dropped e.g.,policy_code, initial_list_status
- Columns with redundant information or found not to be useful for analysis have been dropped
- Rows with loan_status “Current” have been dropped as loan tenure has not completed yet
- Special characters from below columns have been stripped off e.g; int_rate , revol_util

FINAL SHAPE: (36105, 20)

High Level Data Understanding

Default Overall ratio	Default ratio over years

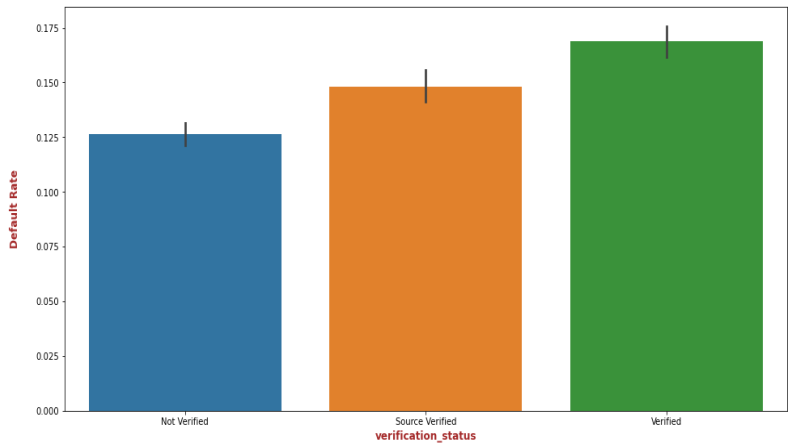




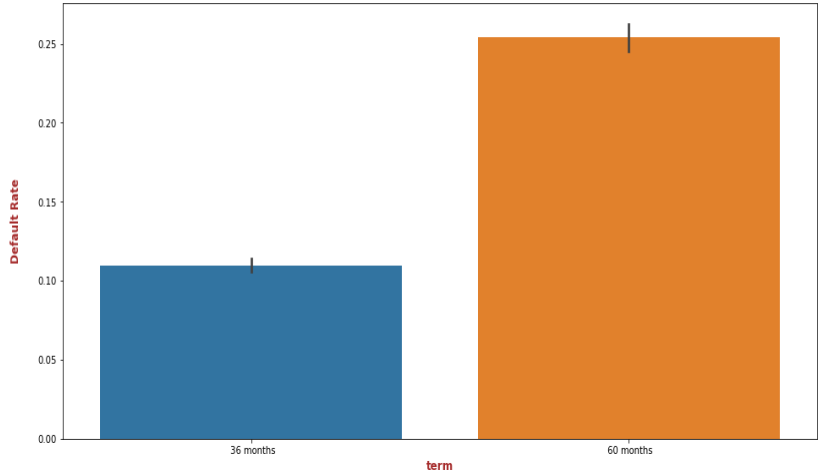
- 14.5% loans are being defaulted
- 85.5 % loans are with full payment

- Total Loan amount disbursed every year is increasing at high rate

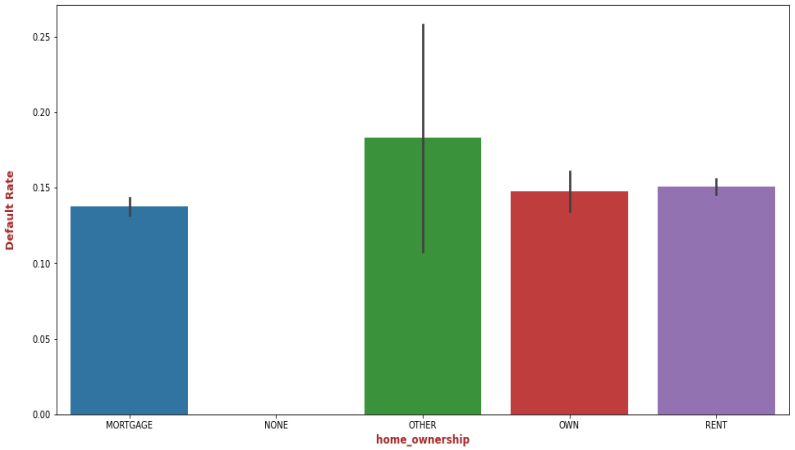
Analysis & Insights - Driving Factors



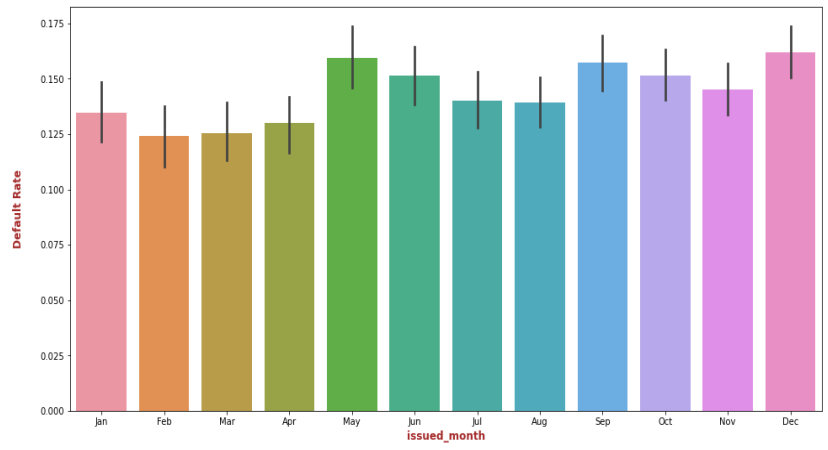
- Large number of verified accounts are still defaulting
- Inference cannot be drawn with this distribution



- Long tenure has more chances of default hence more riskier

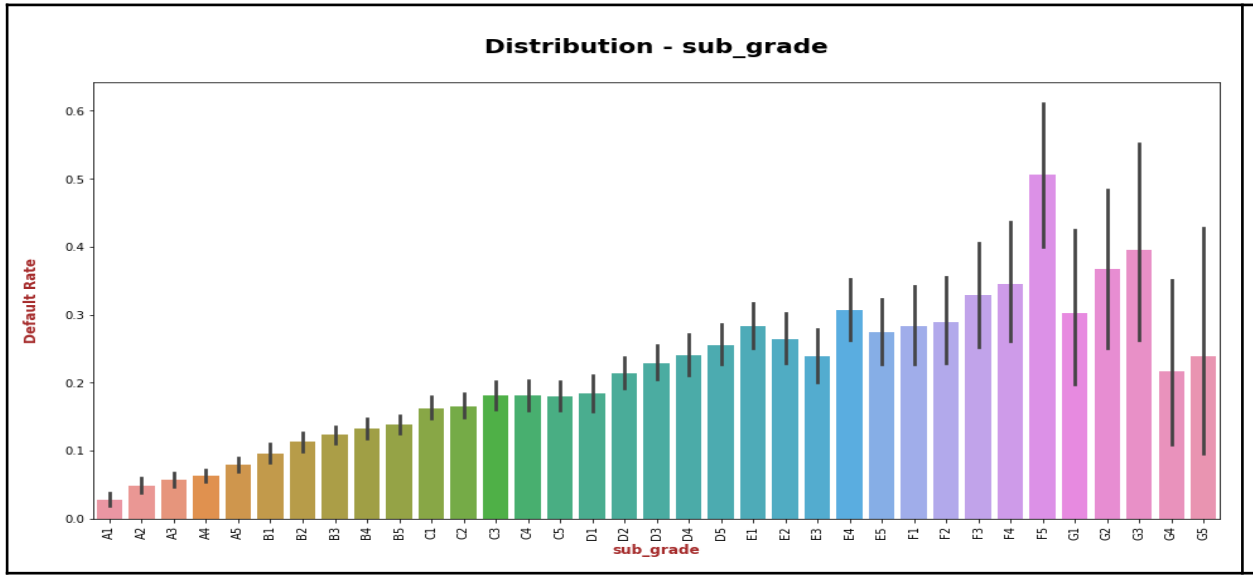


- Spread of default loan among 'home_ownership' categories remain uniform

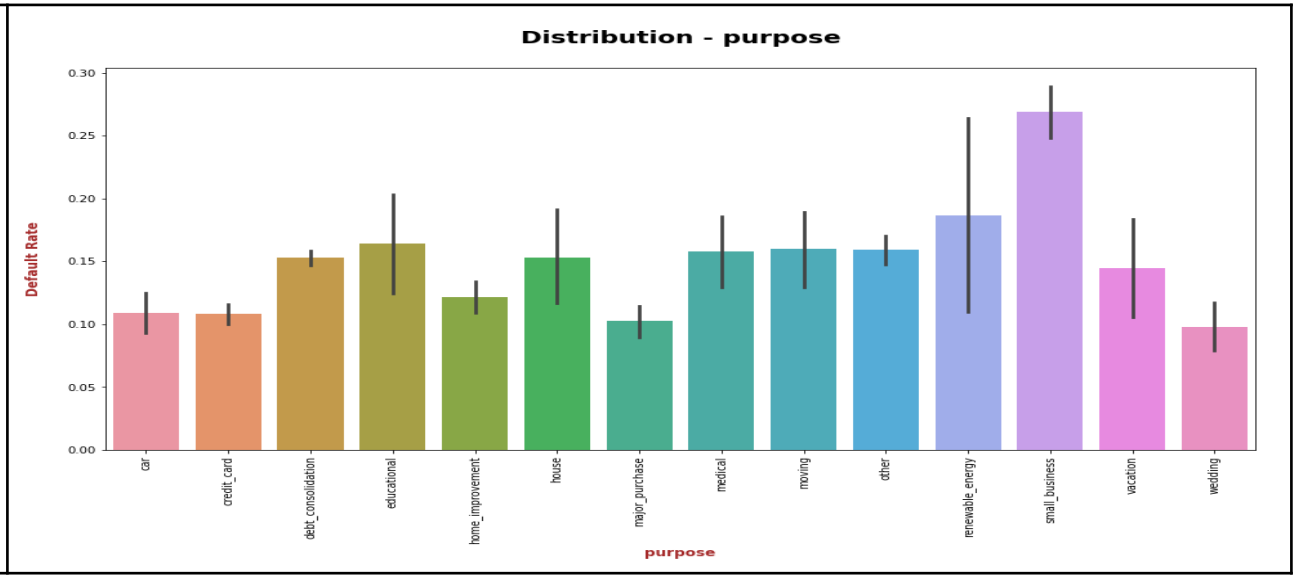


- Loan taken in May, Sep and Dec month seems to have higher default rate
- Inference cannot be drawn from this variable as default rate appears consistent post May

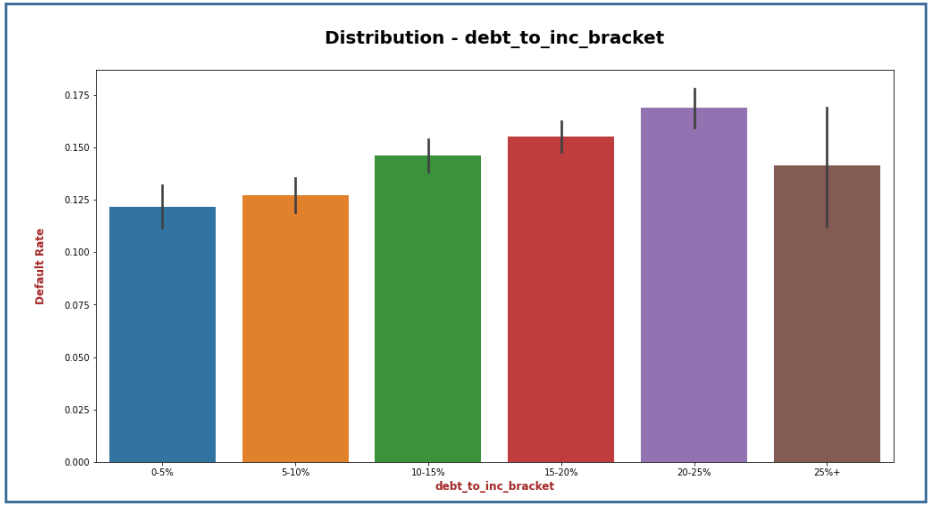
Analysis & Insights - Driving Factors



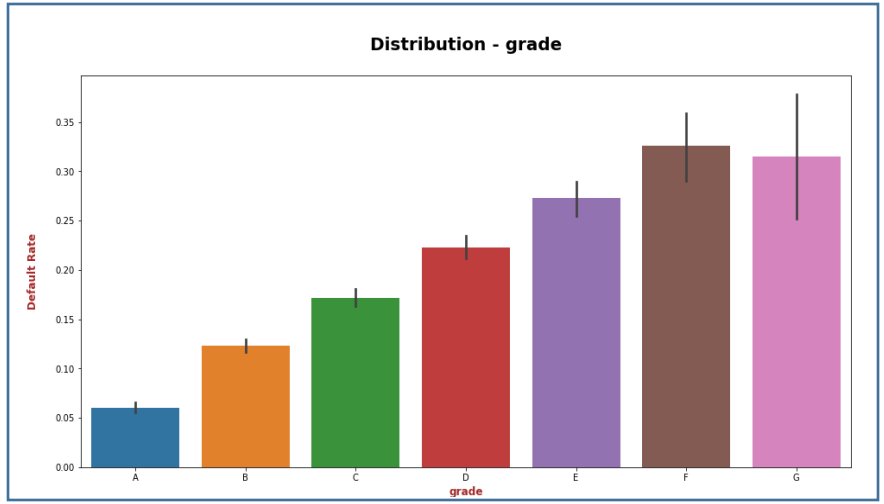
G2,F5 and G3 are usually higher default rate



Small business have higher default rate

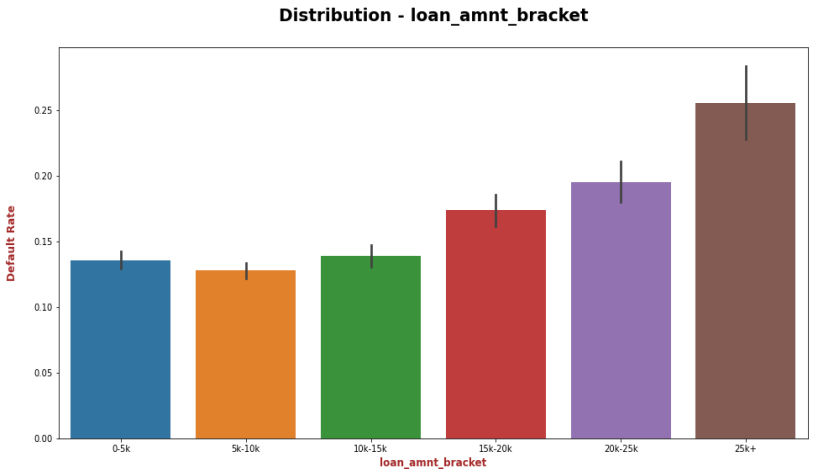


- Higher Debt to Income ratio, higher is the risk of default

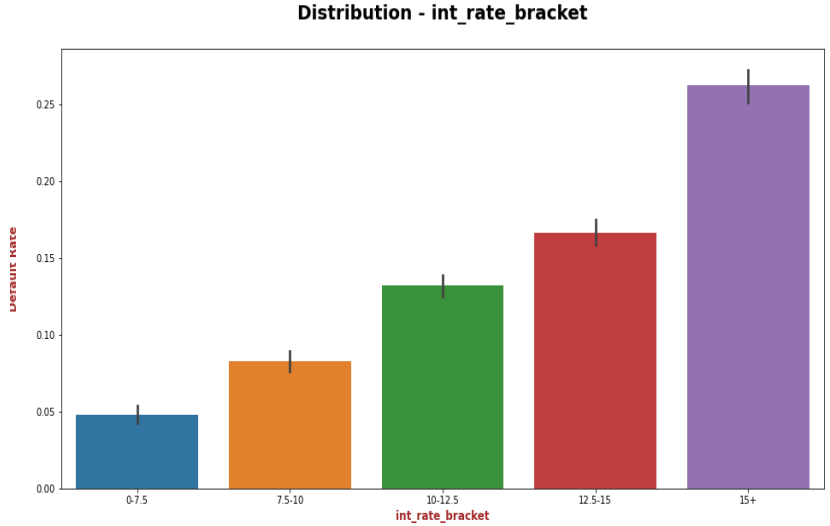


- Higher grades-E,F,G are showing more default rates
- Higher grades seems to carry higher risk of default

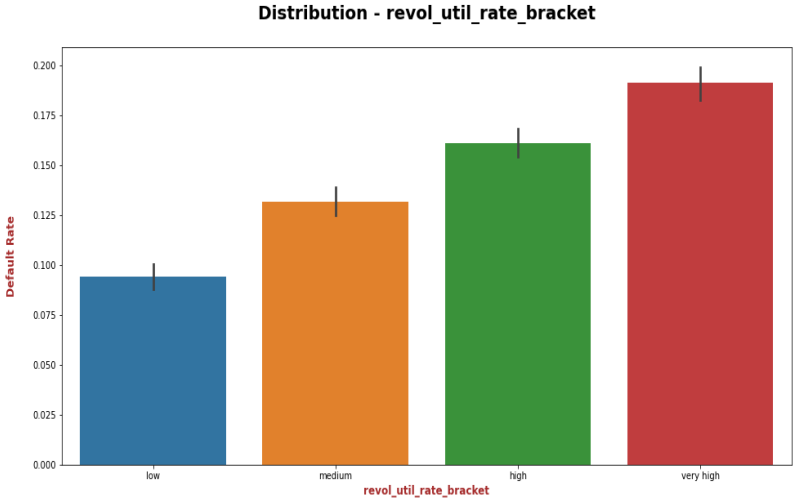
Analysis & Insights - Driving Factors



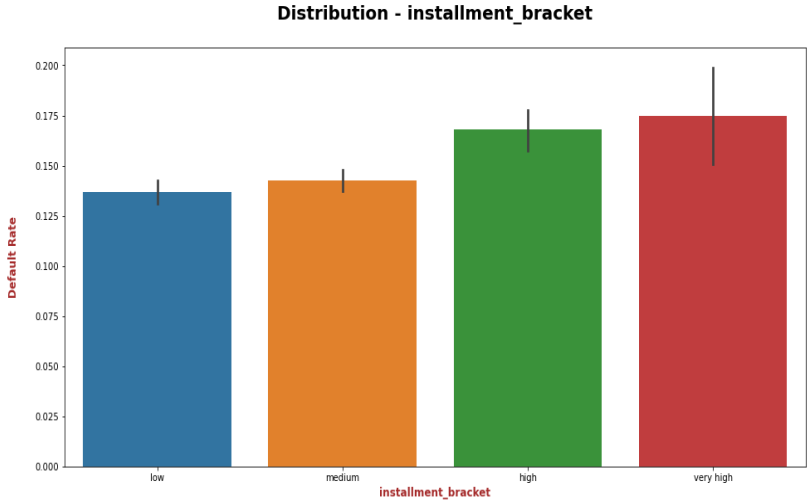
- Higher the amount of loan, higher is the risk of loan being default



- Default rate is higher where interest rate is higher



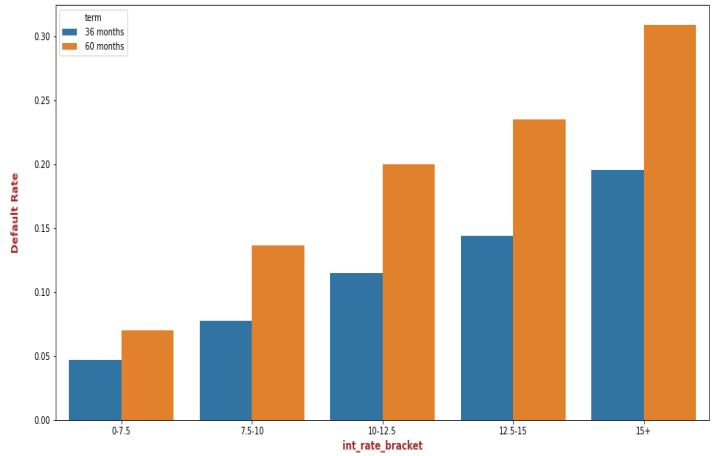
- Larger revolve rate has higher risk of loan being default



- Higher instalment amount tends to pose higher risk

Analysis & Insights - Driving Factors

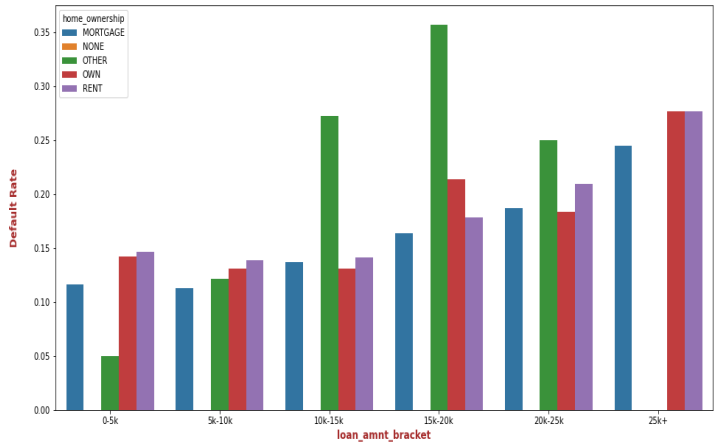
Loan Default rate(%) wrt 'int_rate_bracket' variable for hue 'term' in the data using barplot



Interest rate & Term

- The default rate of higher int_rate is higher for longer term loans.

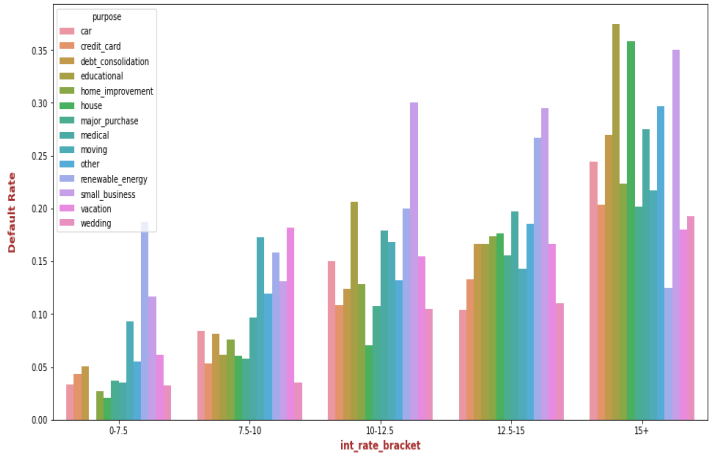
Loan Default rate(%) wrt 'loan_amnt_bracket' variable for hue 'home_ownership' in the data using barplot



Loan amount & home ownership

- Default rate for borrowers with home_ownership of category Mortgage, Own and Rent increases as loan amount increases.
- Home ownership as others still has higher to default in certain brackets

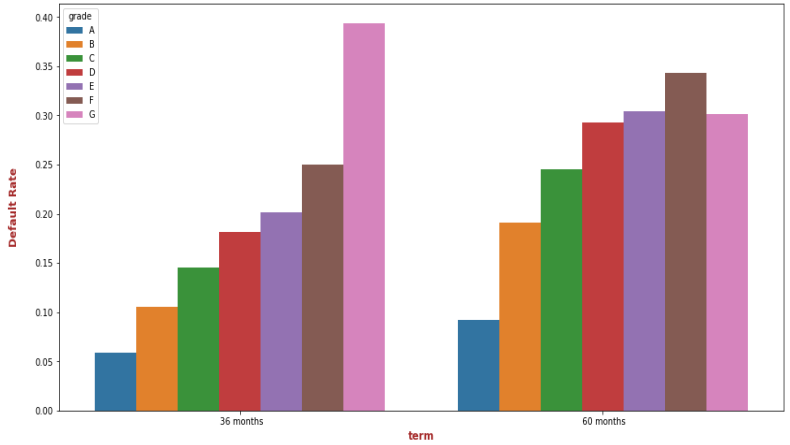
Loan Default rate(%) wrt 'int_rate_bracket' variable for hue 'purpose' in the data using barplot



Interest rate & purpose

- There is a high default rate for borrowers with higher int_rate for education, house, small business, renewal energy

Loan Default rate(%) wrt 'term' variable for hue 'grade' in the data using barplot



Term & Grade

- Grade G has higher default rate for short term loans
- Grade E,F,G loans shows higher default rate for longer term loans

Recommendation

In the light of in-depth analysis Lending club is recommended to consider 5 major driving factors to predict a potential loan default

- **Grade/Interest Rate –** Borrowers who are been charged with higher grade/interest rate are more like to default the loan
- **Home Ownership –** Borrowers with home ownership status declared as others have higher tendency to default
- **Purpose –** Borrower with higher default rate generally quote purpose as ‘small business’, ‘medical’, 'educational' ,'Renewable Energy' and 'Debt Consolidation'
- **Duration of loan term –** Borrowers with longer loan tenure are more likely to default
- **Loan Amount –** Borrowers applying for larger loan amount are more likely to default