

Lending Club Case Study



Lokesh Chenta

Mohammed Aamish

22 June 2024

Problem Statement and Objective

- This project is for a large financial company that provides various types of loans to applicants.
- The purpose of this project is to minimize credit loss to the company due to loan default by borrowers.
- A data set is available that contains historic data with details (both loan related and borrower related data) for each loan provided by the company.
- The objective is to use data analytics to analyze this loan data set and determine the factors(variables) that influence loan default.
- These factors will help the company when issuing new loans to predict the risk and minimize the risk by deciding whether to accept or reject the new loan application.

Analysis Approach

- Data Cleaning
 - Handling Nulls
 - Removing unwanted columns
 - Converting to right datatypes
- Data Enriching
 - Data Imputation
 - Derived Metrics
- Data Analysis
 - Univariate Analysis
 - Bivariate Analysis
- Observations
- Summary

Data Cleaning

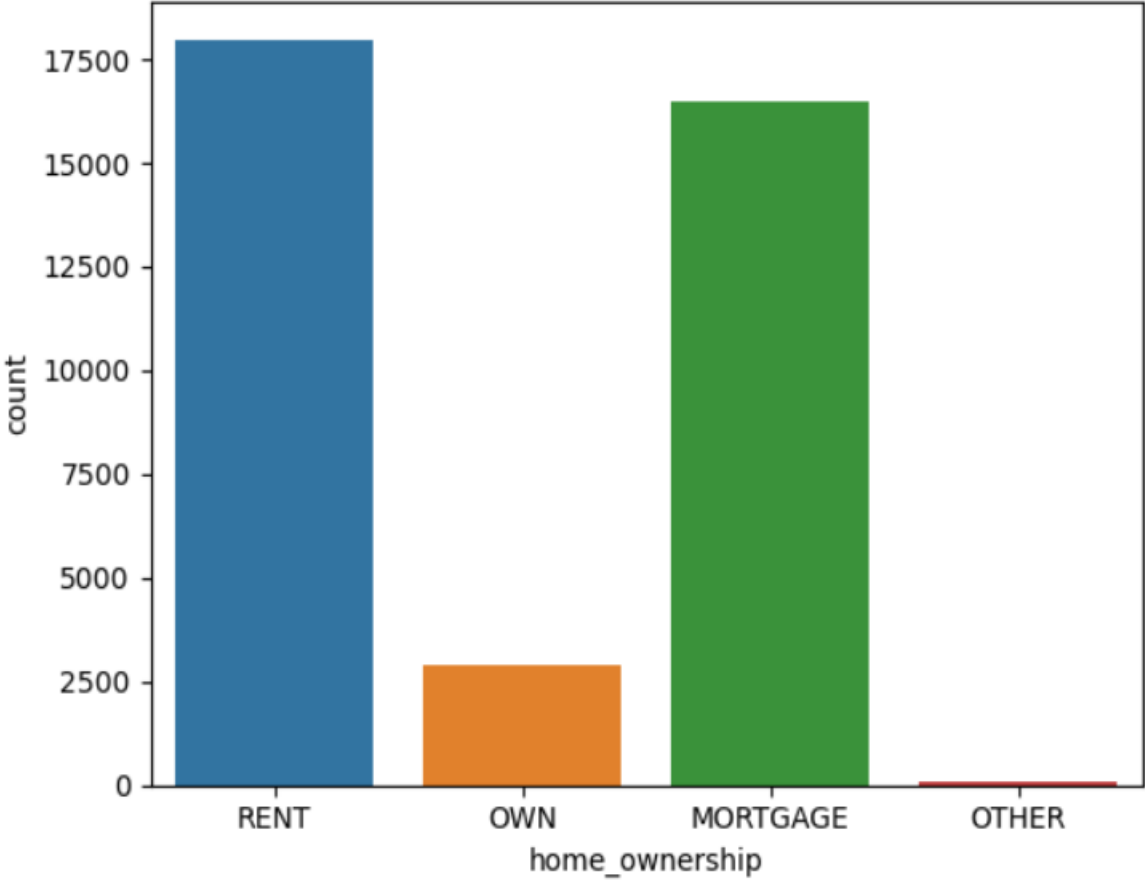


- Dropped columns that contain all missing(null/NA) values.
 - Eg: annual_inc_joint, dti_joint
- Dropped columns with large percentage of missing values.
 - Eg: next_pymnt_d(97%), mths_since_last_record(92%)
- Verified rows with all missing values – found none.
- Dropped columns that seem not relevant for any analysis.
 - Eg: id, url, title, out_prncp_inv
- Dropped columns with single value only.
 - Eg: policy_code, application_type, pymnt_plan

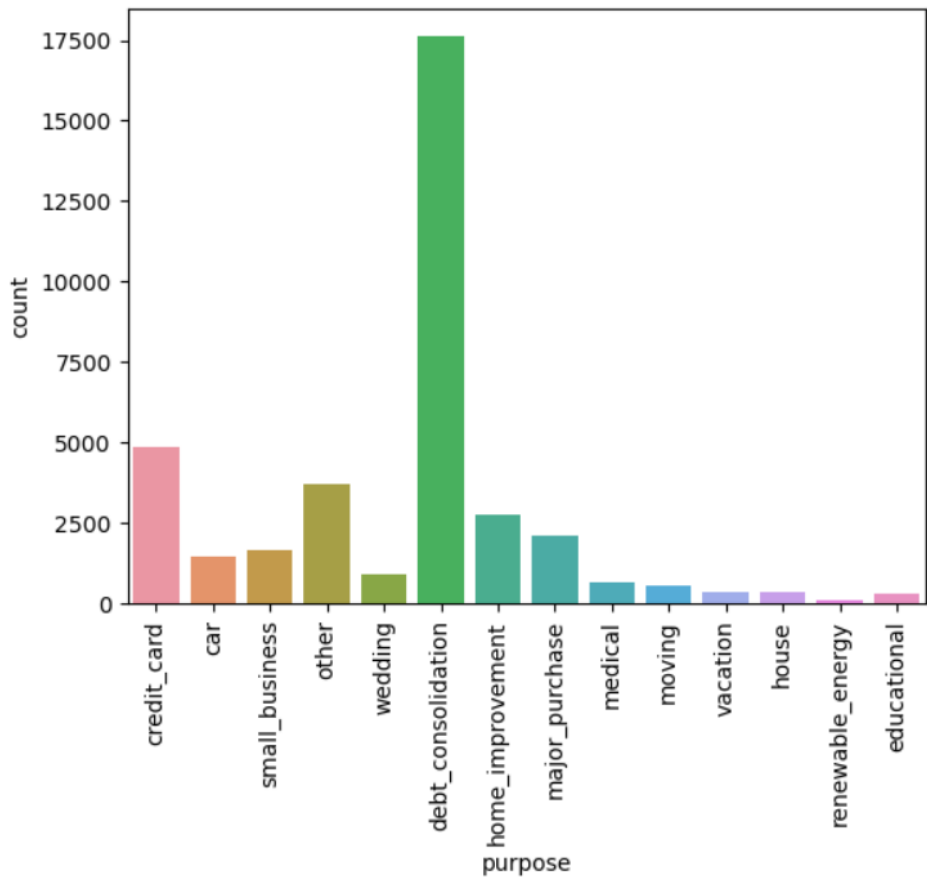
Data Enriching

- Converting columns to right data types
 - Eg: int_rate(from object to float), revol_util
- Data Imputation for certain columns with very less % of missing values
 - Eg: emp_length
- Creating derived metrics
 - Eg: issue_year, issue_month from issue_d

Data Analysis – Default Loans

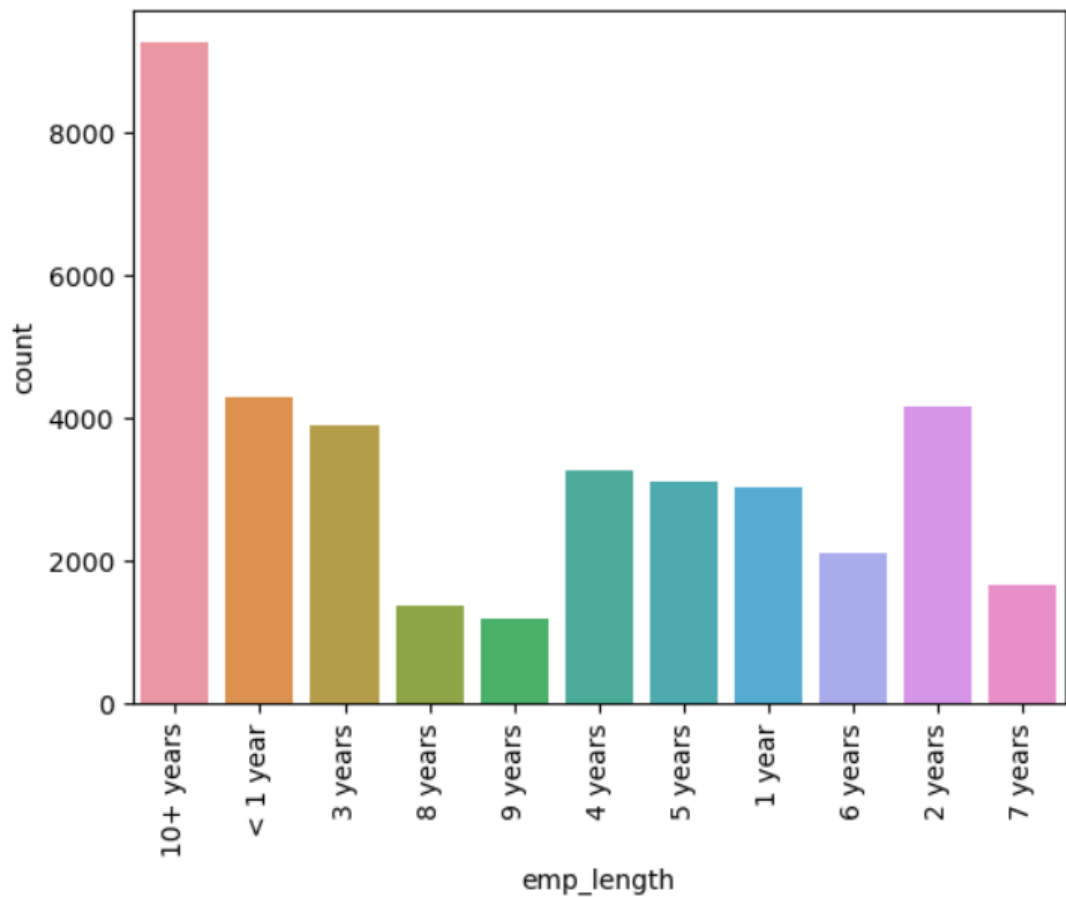


Loan defaults are high in case of applicants with home ownership category of RENT and MORTGAGE.

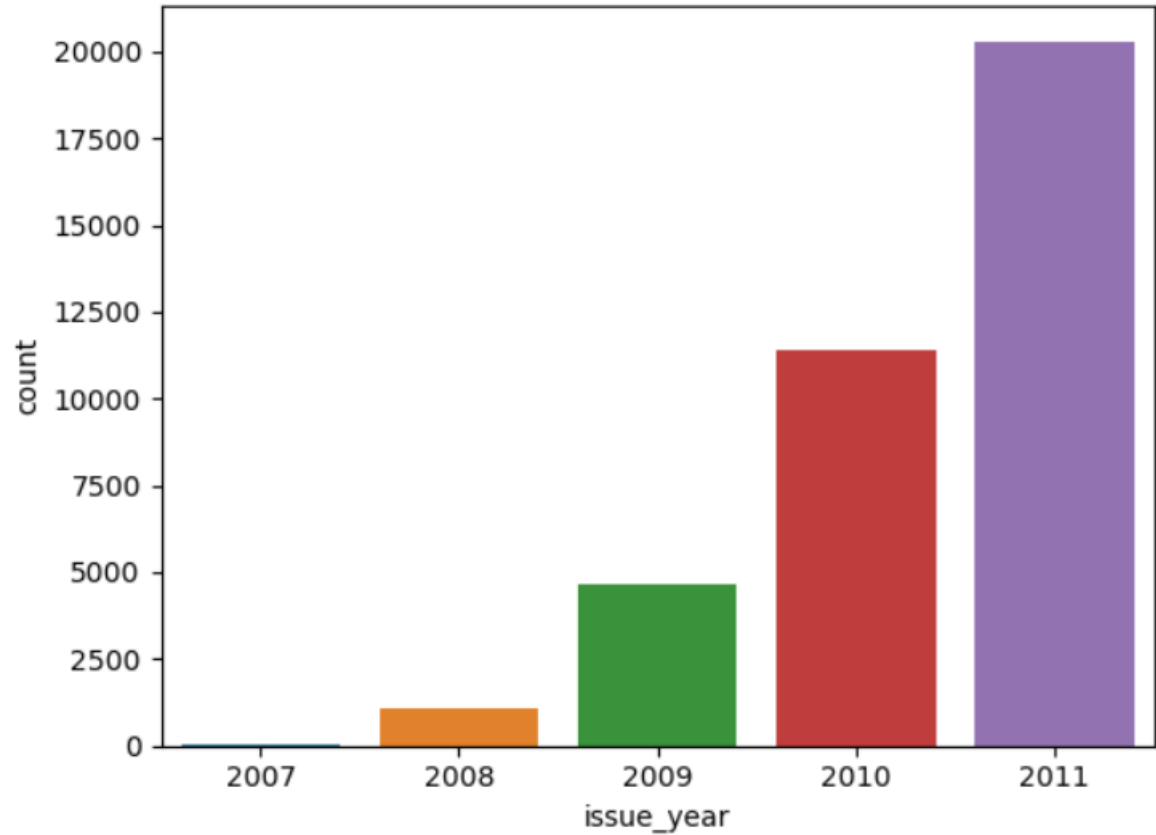


Loan defaults are high in case of loans taken for the purpose of debt consolidation and credit card.

Data Analysis – Default Loans



Loan defaults are high in case of applicants with employment experience of more than 10 years.

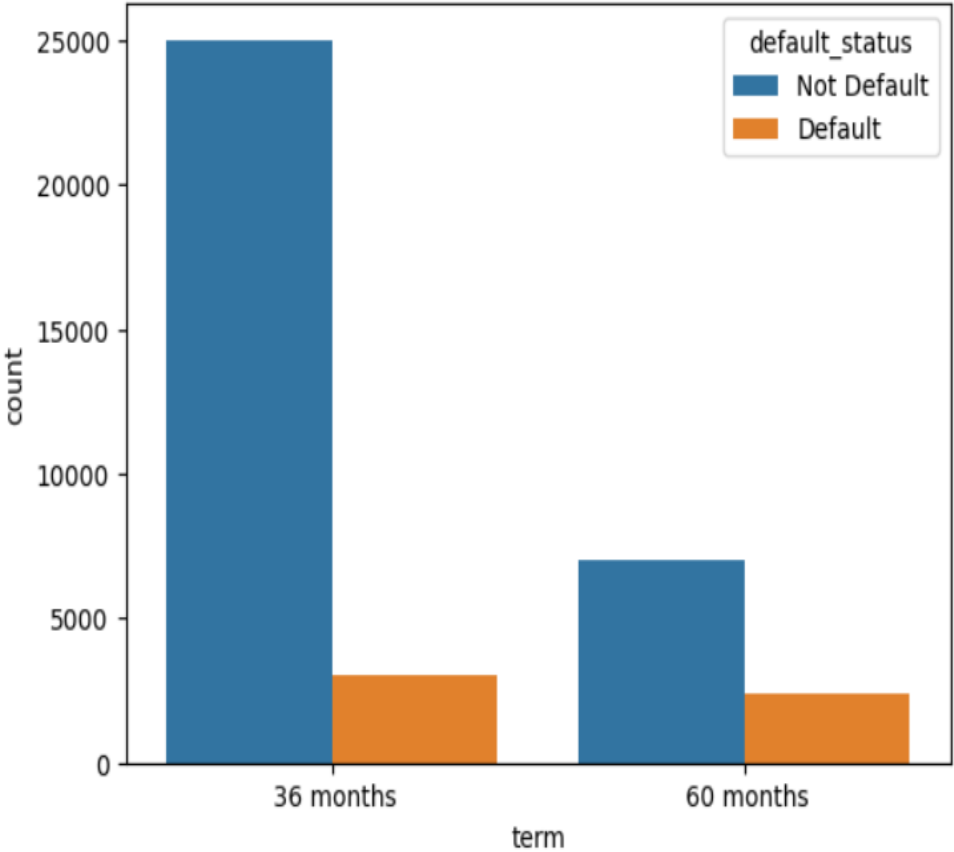


Loan defaults are almost doubling every year.

Data Analysis – All Loans

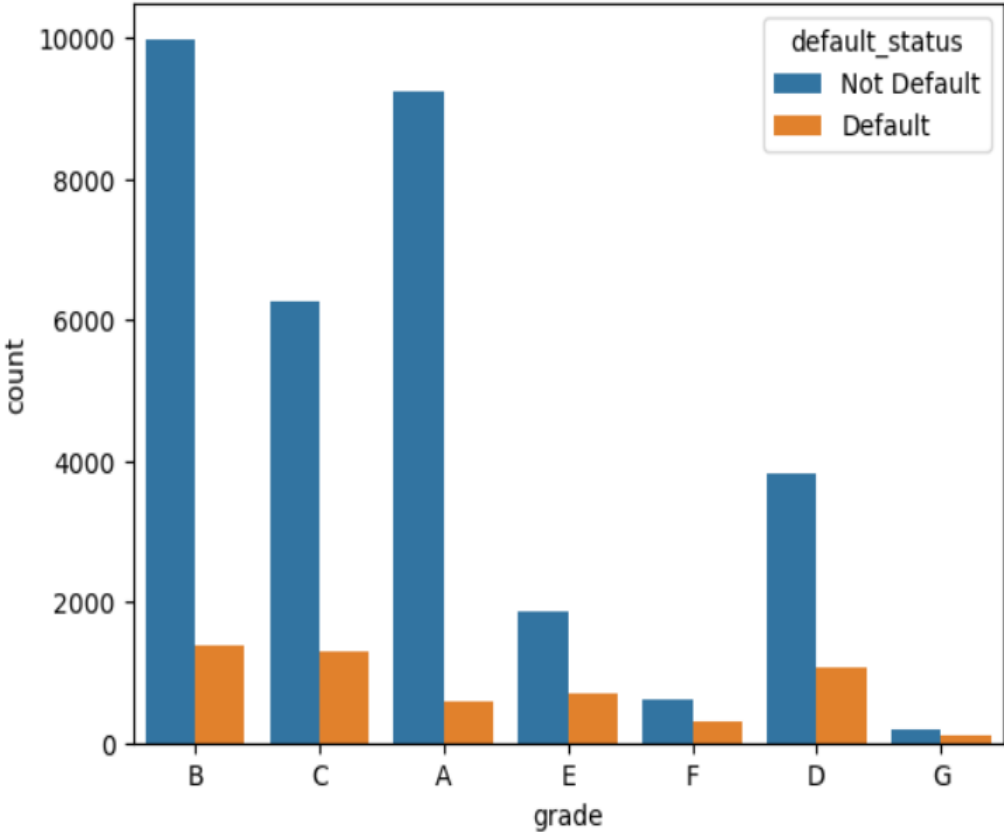


Loan Counts by Loan Term



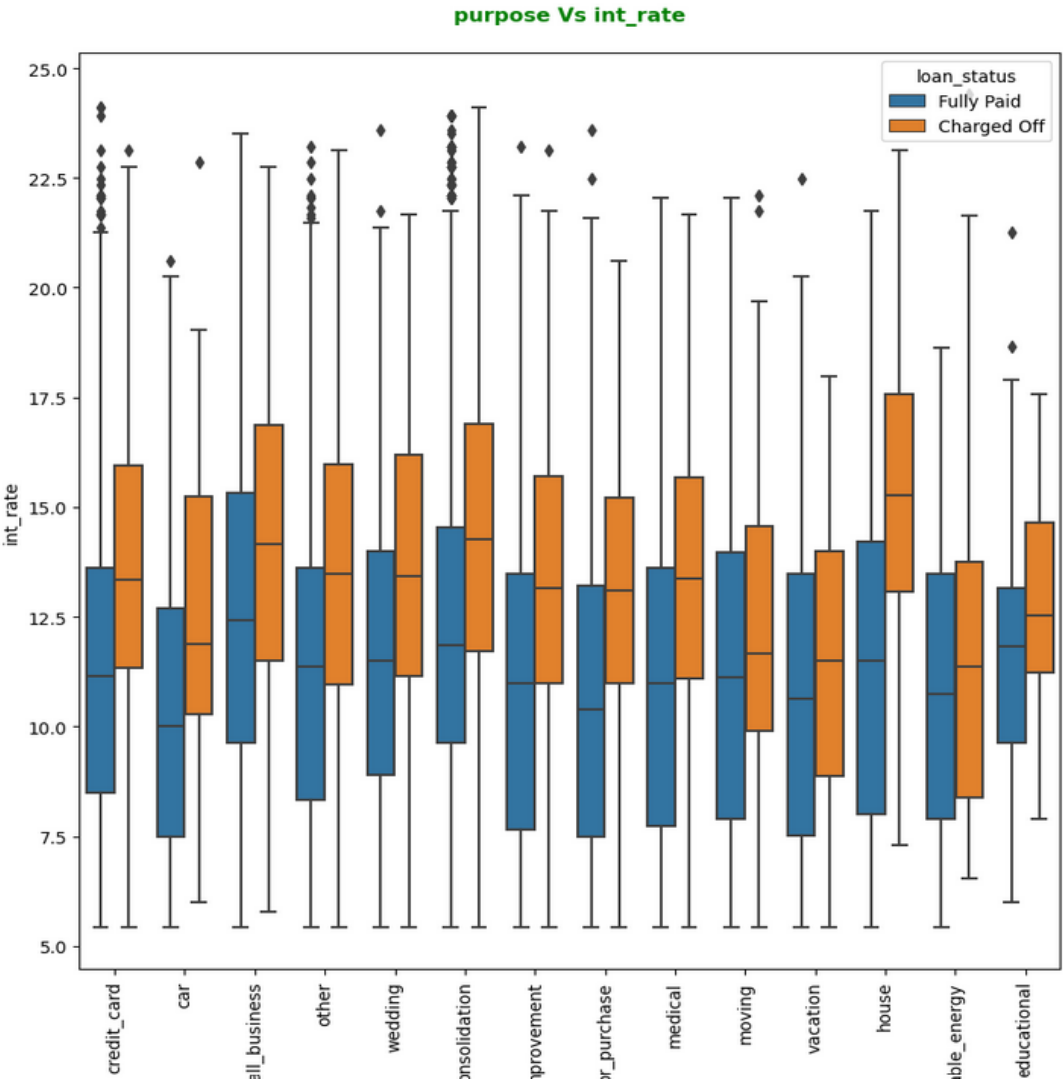
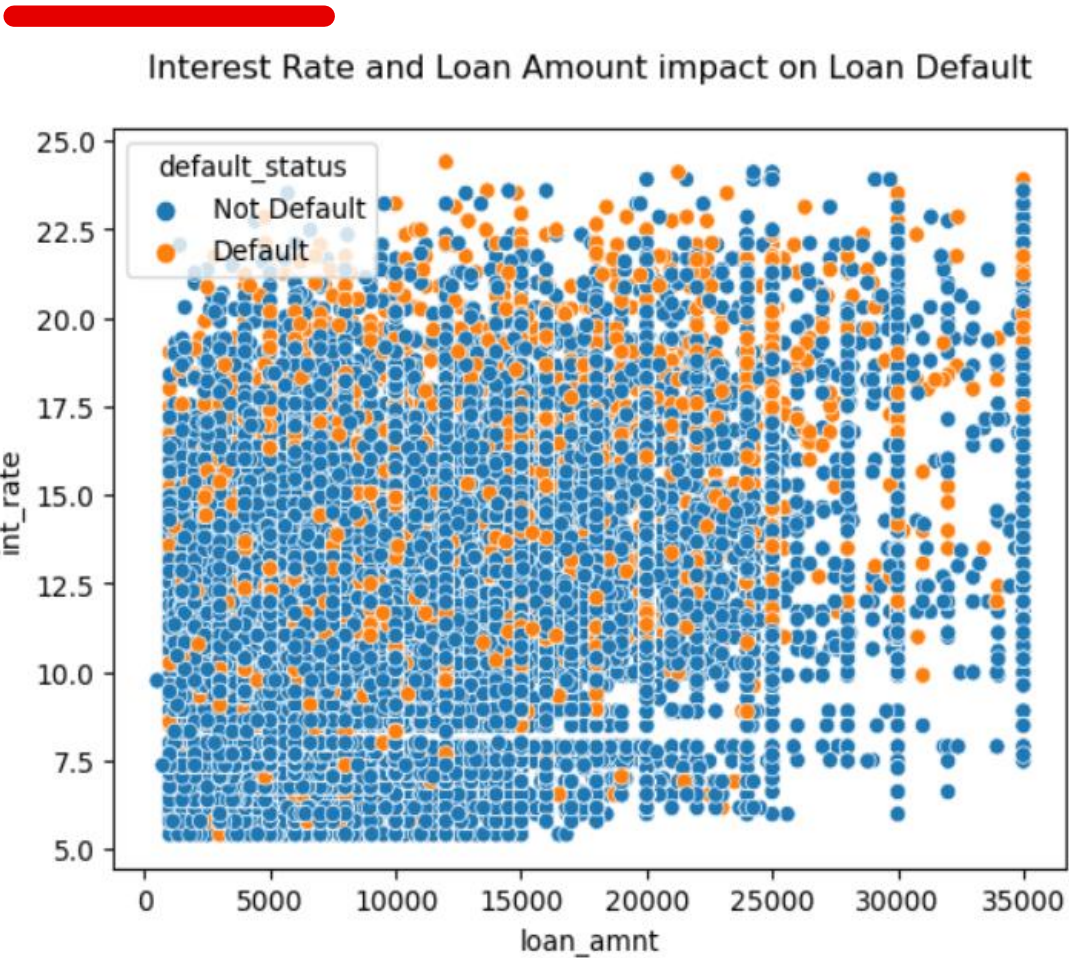
The proportion of Loan defaults is more for the term of "60 months".

Loan Counts by Grade



The risk of Loan default increases as the grade decreases - lower grade applications (D, E, F, G) have more defaults.

Data Analysis – All Loans



Higher loan defaults are seen as int_rate crosses 15% and loan_amnt crosses 25000.

In case of loans taken for purpose of "house", the median interest rate for default loans is very high (at 15%) compared to other purposes.

Observations



UniVariate Analysis

- Loan defaults are high for applicants with home ownership category of RENT and MORTGAGE.
- Loan defaults are high for loans taken for the purpose of debt consolidation and credit card.
- Loan defaults are high for applicants with employment experience of more than 10 years.
- Loan defaults are almost doubling every year.
- The proportion of Loan defaults is more for the term of "60 months".
- The risk of Loan default increases as the grade decreases - lower grade applications (D, E, F, G) have more defaults.
- Loan defaults increase as the interest rate increases (after 12%).

Observations



BiVariate Analysis

- In case of loans taken for the purposes "moving", "vacation", 'renewable_energy', the Loan defaults are happening even for smaller loan amounts.
- Higher loan defaults are seen as Interest rate crosses 15% and Loan amount crosses 25000.
- In case of loans taken for purpose of "house", the median interest rate for default loans is very high (at 15%) compared to other purposes. So, house loans are less likely to go to default even at slightly higher interest rates.

Summary

The following are the key driving factors(driver variables) behind loan default

- Loan Related Variables
 - Interest Rate
 - Loan Term
 - Loan Purpose
- Application Related Variables
 - Grade
 - Employment Experience