LANDING CLUB CASE STUDY

Names:

Jeetender Kumar Batra

Jaswanth Kumar Daddla

Overview and Objectives

Overview

Landing Club is Consumer Finance Company which specializes in lending various types of loans to urban customers. This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile

Risks

- 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

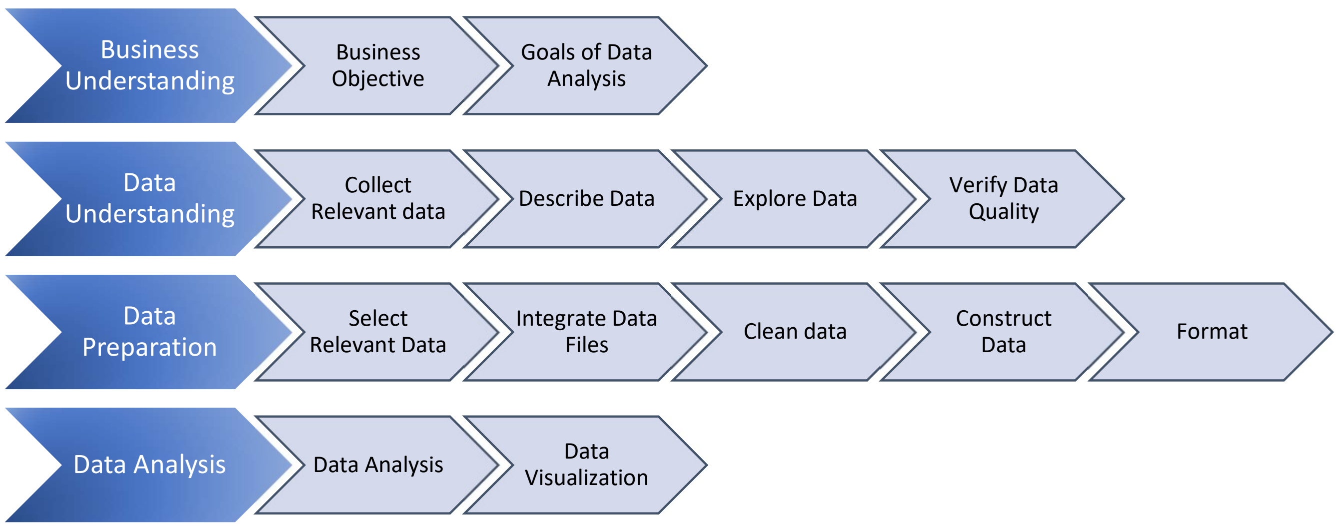
Business Objective

Business objective of analysis is to identify and highlight risky loan application. Lending loans to risky applicants is the largest source of financial loss aka 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. borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are

Goals of data analysis

- Driving Factors - Company wants to understand the driving factors (or driver variables) behind loan default
- Red flag Identifications – Identify risk loan application based on relevant consumer attributes and Loan Attributes

Methodology



Data Understanding

Data Sources

- Load loan.csv
    - Contains loan application, status, customer information

- Data_Disctionary.xlsx
    - Contains information related to columns or variables used in this loan.csv. This is meta-data for loan.csv

Data Preparation

Clean Data, Construct Data, Format

- **Data Source – loan.csv**
  - Data file contains fields related to Loan Status (Fully Paid, Charged Off, Current)
  - Current status can be filtered out as this is not relevant to this analysis.
  - Remove or drop columns having 90% missing values
  - Remove columns which are not relevant for analysis.
    - last_pymnt_d,last_credit_pull_d,title,url ,zip_code
    - addr_state,delinq_2yrs,earliest_cr_line,inq_last_6mths,open_acc
    - pub_rec,revol_bal,revol_util,total_acc,out_prncp,out_prncp_inv
    - total_pymnt,total_pymnt_inv,total_rec_prncp,total_rec_int
    - total_rec_late_fee,recoveries,collection_recovery_fee,last_pymnt_d,
    - last_pymnt_amnt,last_credit_pull_d,application_type

  - Handle NULL or NaN values in the columns
  - Check and clean the column data. Example int_rate columns contains %
  - Check any required derived column

Data Preparation

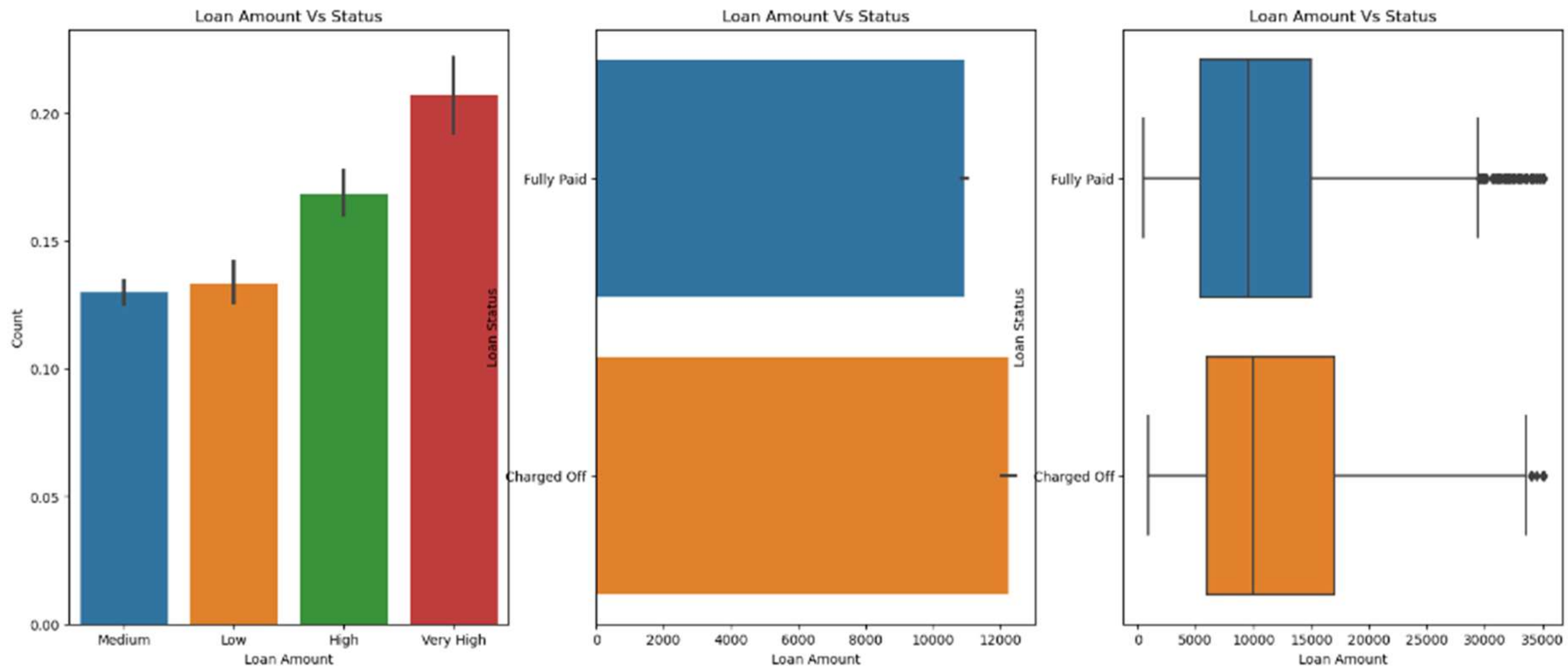
Clean Data, Construct Data, Format

- **Data Type Formatting**
  - Convert following columns to numeric type for standardization and easy analysis
  - Funded_amt(float)
  - Loan_amt(float)
  - Term(int)
  - Int_rate(float)

- **Category Type Formatting**
  - Convert following columns to category codes for standardization and easy analysis
  - loan_status
  - grade
  - sub_grade
  - verification_status
  - purpose
  - title
  - home_ownership
- **Derived Columns**
  - Add derived columns **issue_year and issue_month** from **issue_d** for EDA analysis

EDA Analysis

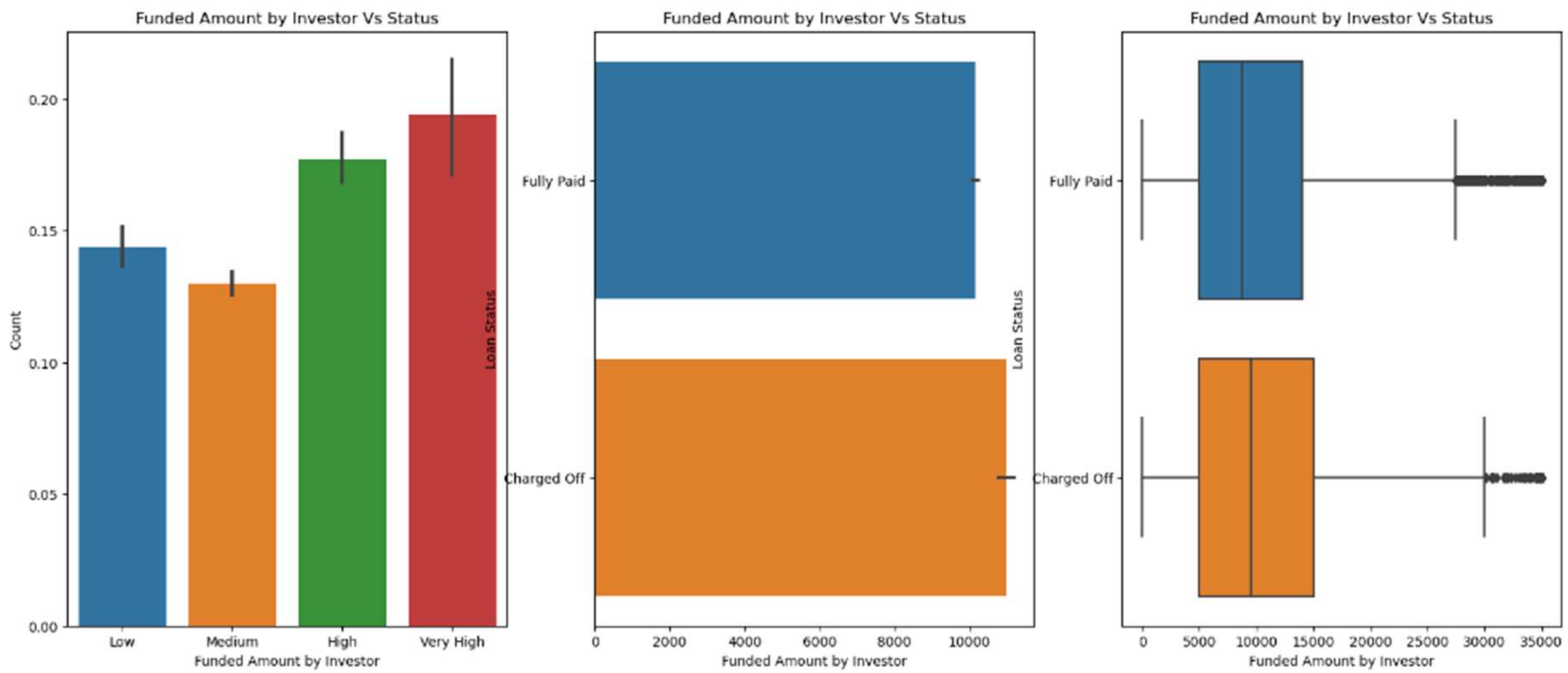
Driving Factors for default – Loan amount



Loan amount is the deriving factor for loan default. Higher loan amount increase the chance of loan default.

EDA Analysis

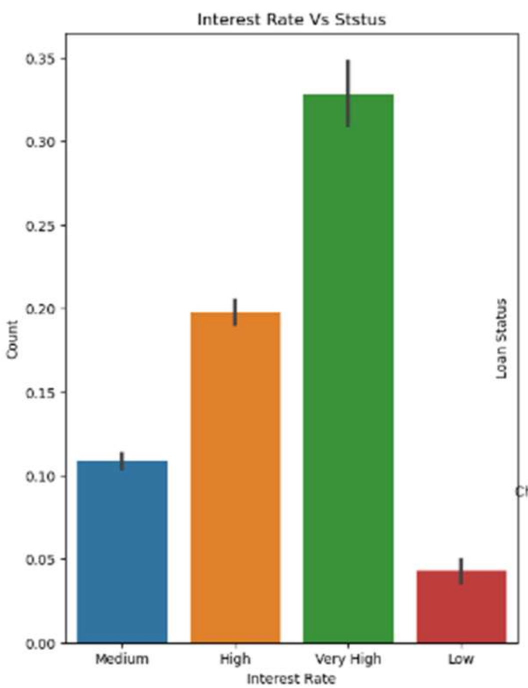
Driving Factors for default – Funded Amount by Investor

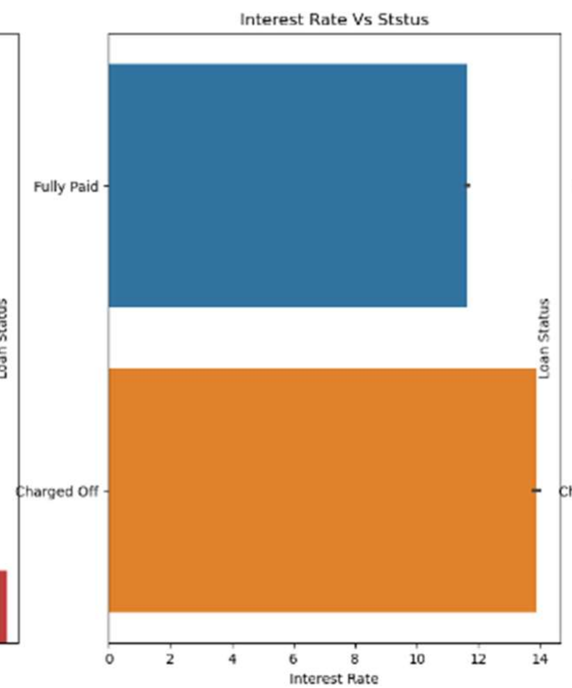


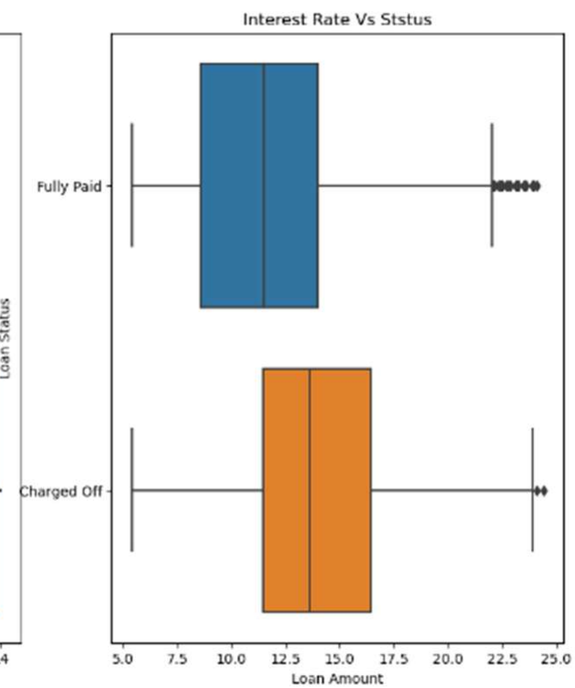
Fund by investor is the deriving factor for loan default

EDA Analysis

Driving Factors for default – Interest Rate



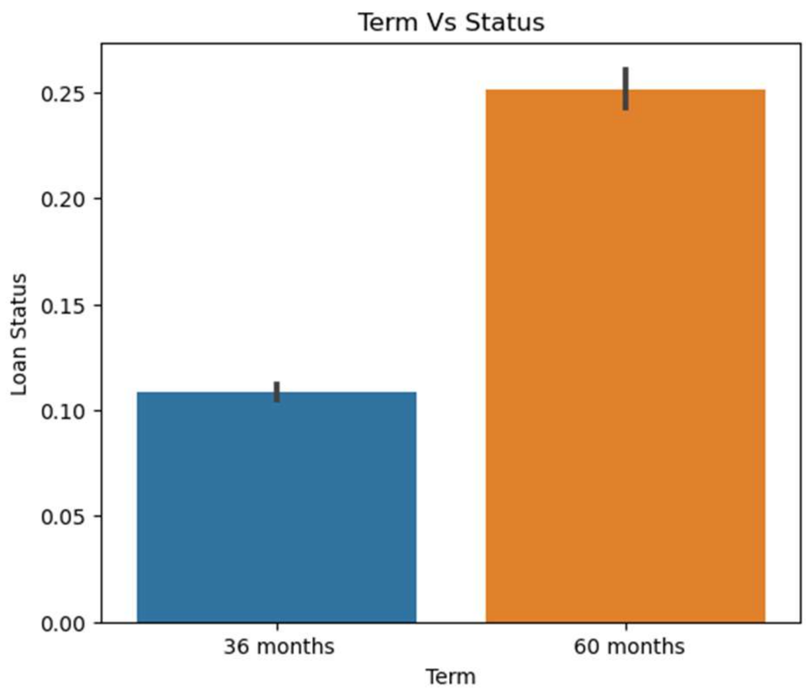




Higher Interest Rate is the deriving factor for loan default

EDA Analysis

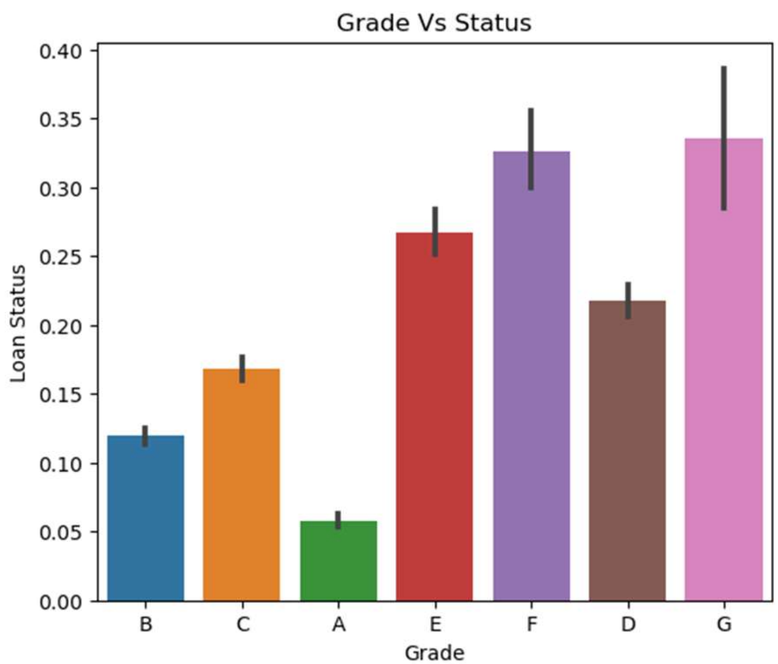
Driving Factors for default – Term



Longer Loan term is the factor of defaulting loan. 60 months has higher default rate.

EDA Analysis

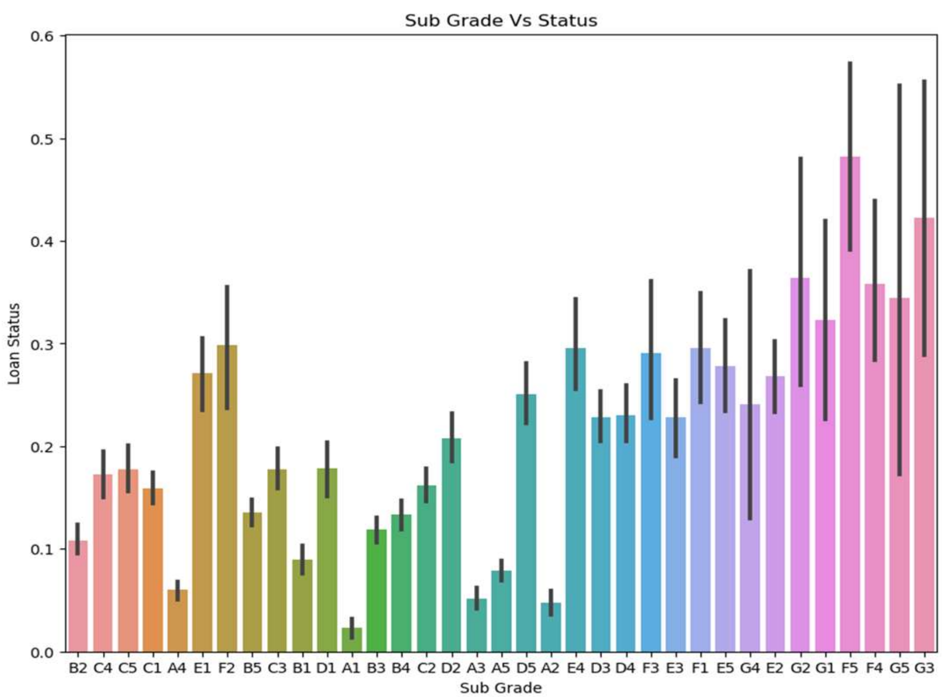
Driving Factors for default – grade



Grade is the factor of defaulting loan. A has lowest default rate.

EDA Analysis

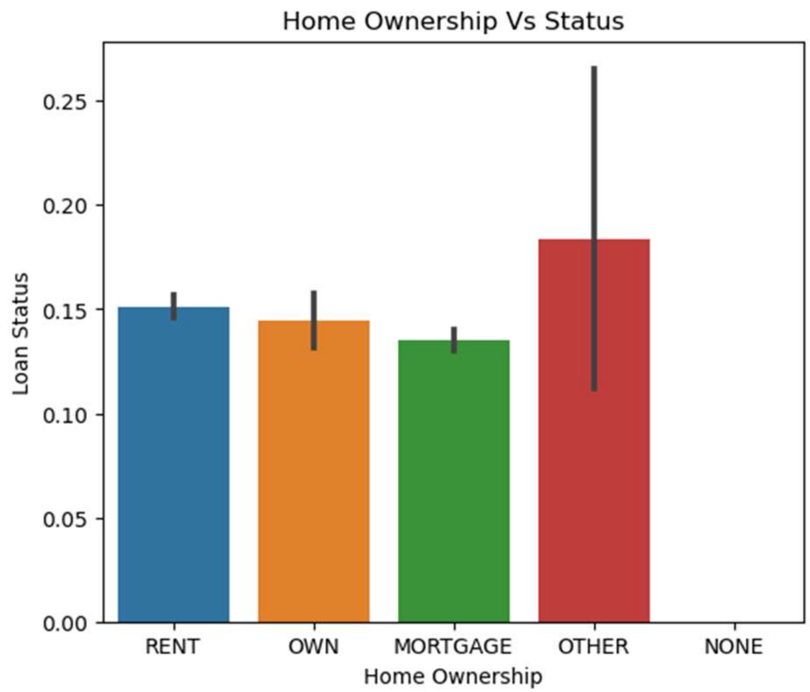
Driving Factors for default – Sub grade



Sub Grade is the factor of defaulting loan. F5 grade has highest default rate. A1 has lowest.

EDA Analysis

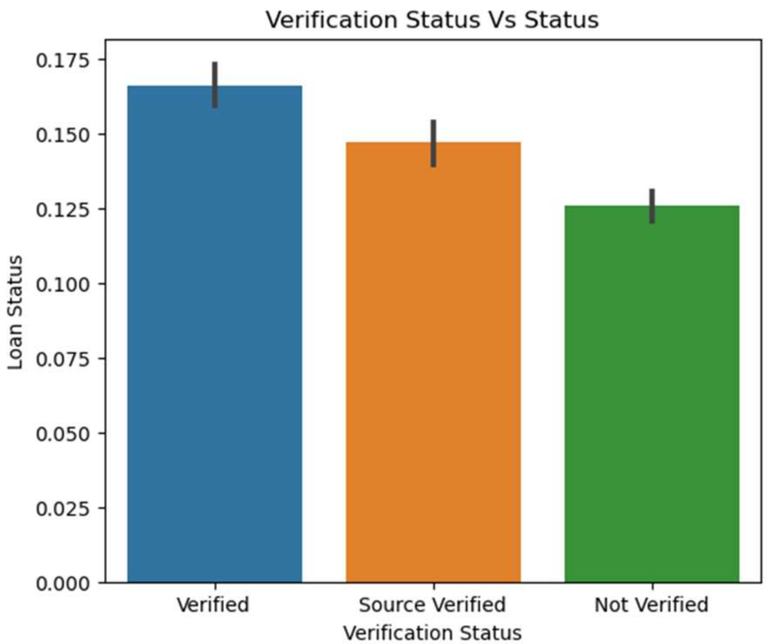
Driving Factors for default – Home Ownership



Home Ownership does not seems to be major factor for loan default as range is narrow.

EDA Analysis

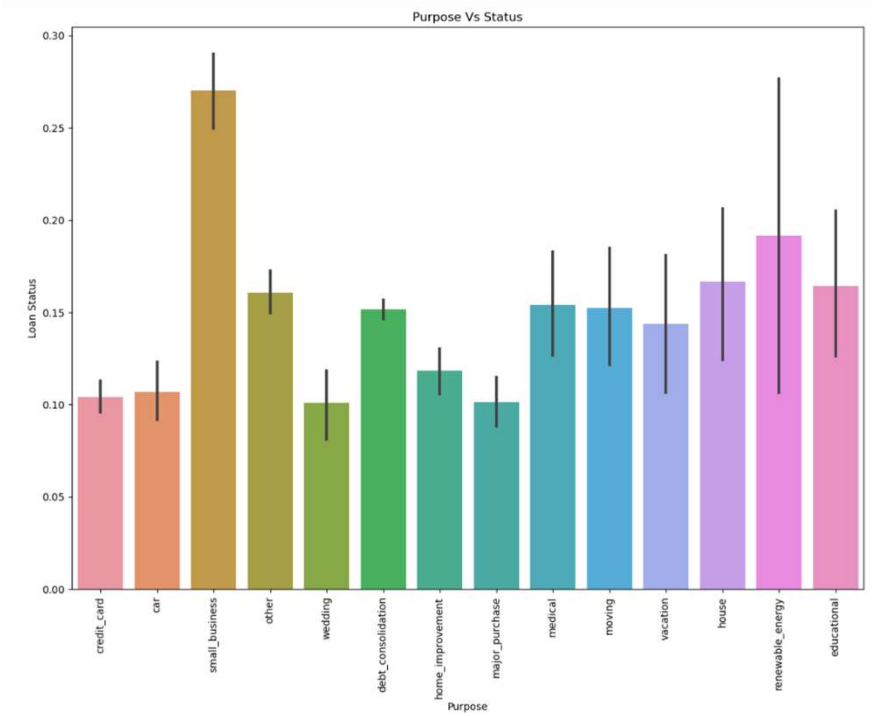
Driving Factors for default – Verification Status Code



Verification Status as Verified seems to show higher loan defaulting

EDA Analysis

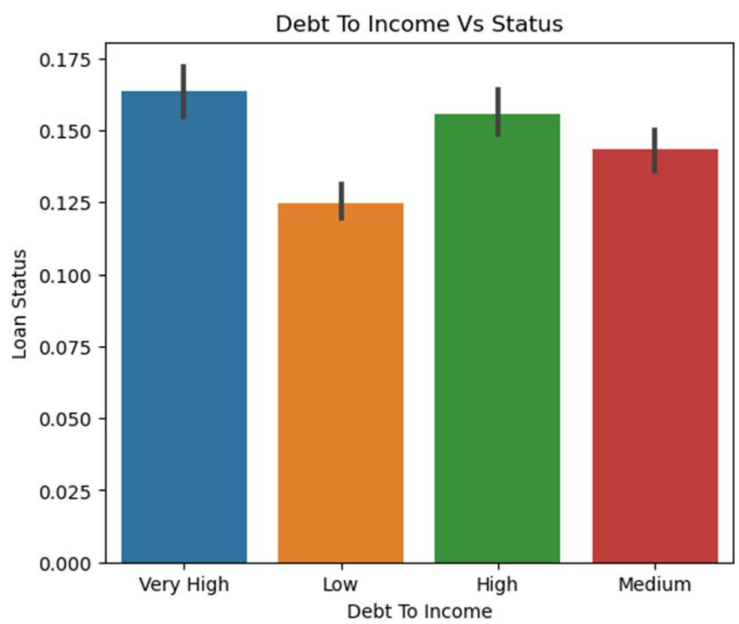
Driving Factors for default – Purpose



Purpose has impact on defaulting loan. Small Business has highest loan default rate.

EDA Analysis

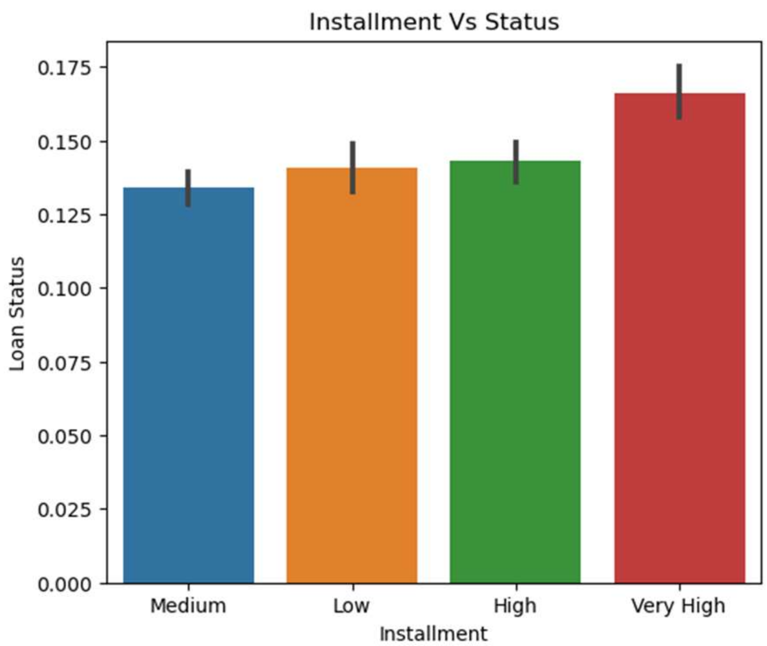
Driving Factors for default – Debt To Income



Debt to Income is the factor of defaulting loan. Higher debt to income results in higher rate of loan default.

EDA Analysis

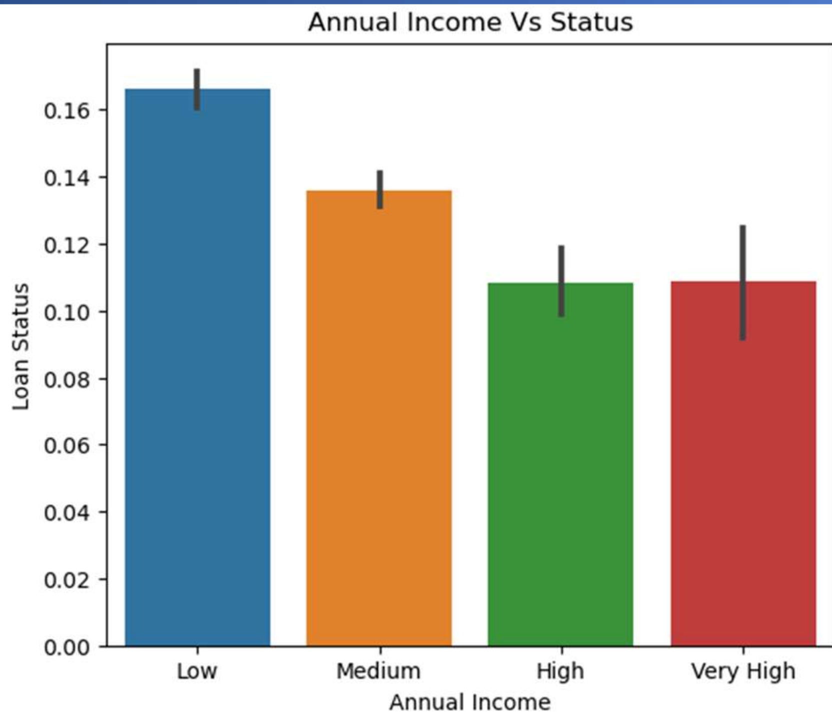
Driving Factors for default – Installment



Higher Installment is the factor of defaulting loan. Higher installment seems to have higher default rate.

EDA Analysis

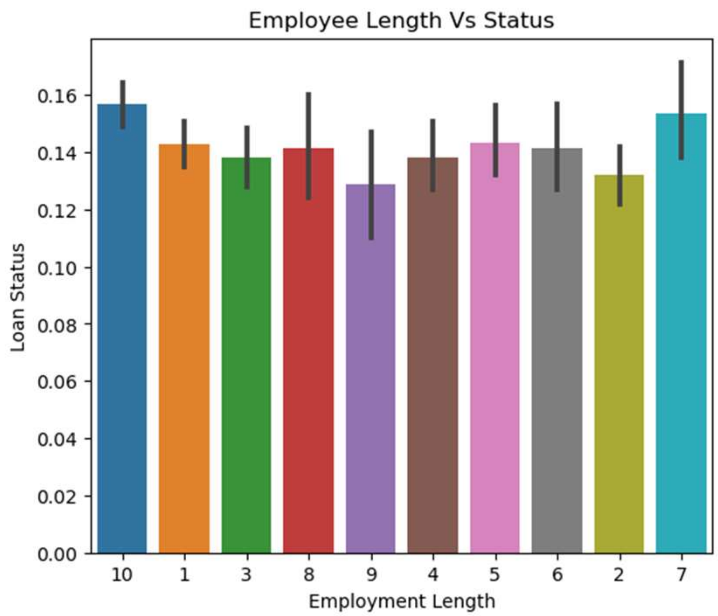
Driving Factors for default – Annual Income



Annual Income is the factor of defaulting loan. Low annual has higher loan default rate.

EDA Analysis

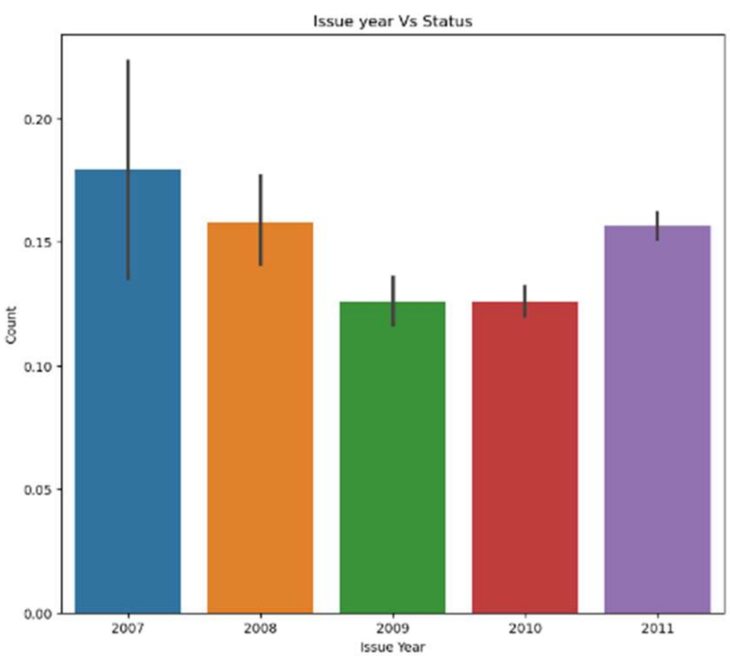
Driving Factors for default – Employee Tenure

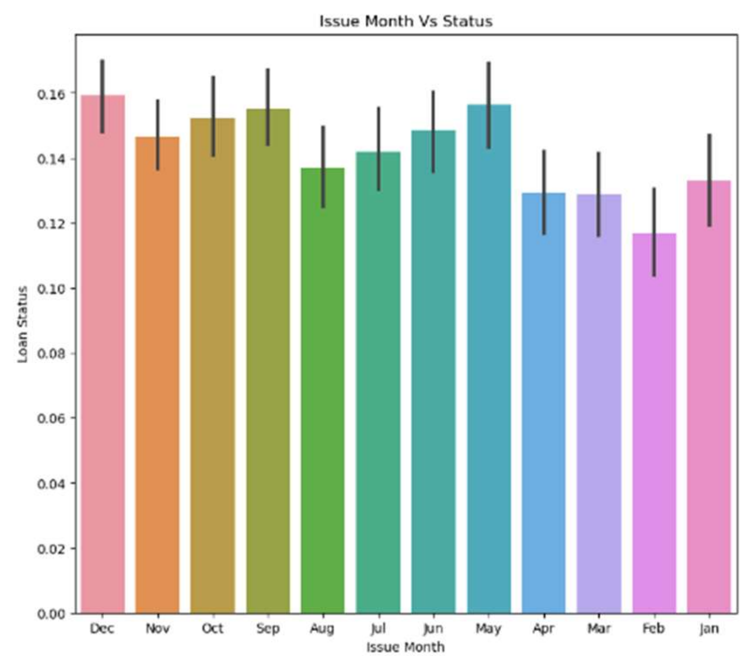


Employee Tenure does not seems to be major factor of defaulting loan.

EDA Analysis

Driving Factors for default – Issue Date (Month and Year)





As per analysis, issue_year(2007) has impact on defaulting loan. In 2011 loan default increased. Year can not be considered as major factor of loan default for future loans.

Conclusion

Conclusion

- As per analysis, following Consumer and Loan attributes are factors for loan default
  - Interest Rate
  - Grade/Sub Grade (Both are correlated so either of them can be considered)
  - Term (Loan term)
  - Loan Amount/ Funded Amount (Both are correlated so either of them can be considered)
  - Debt to Income
  - Installment
  - Annual Income
  - Purpose

LANDING CLUB CASE STUDY

Names:

Jeetender Kumar Batra

Jaswanth Kumar Daddla

Overview and Objectives

Overview

Landing Club is Consumer Finance Company which specializes in lending various types of loans to urban customers. This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile

Risks

- 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

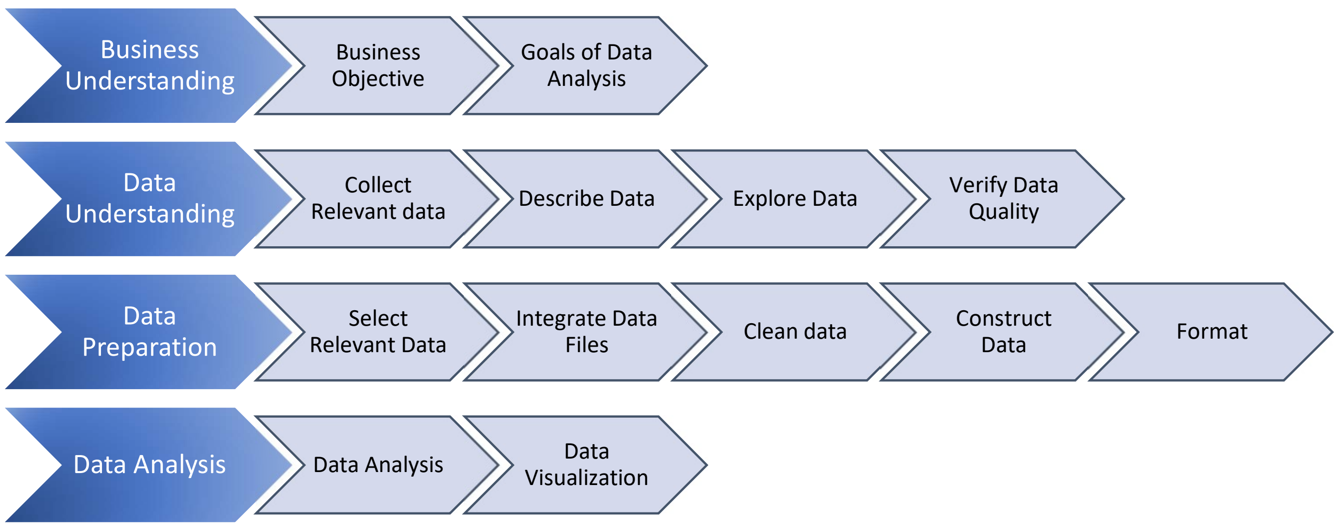
Business Objective

Business objective of analysis is to identify and highlight risky loan application. Lending loans to risky applicants is the largest source of financial loss aka 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. borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are

Goals of data analysis

- Driving Factors - Company wants to understand the driving factors (or driver variables) behind loan default
- Red flag Identifications – Identify risk loan application based on relevant consumer attributes and Loan Attributes

Methodology



Data Understanding

Data Sources

- Load loan.csv
  - Contains loan application, status, customer information

- Data_Disctionary.xlsx
  - Contains information related to columns or variables used in this loan.csv. This is meta-data for loan.csv

Data Preparation

Clean Data, Construct Data, Format

- **Data Source – loan.csv**
  - Data file contains fields related to Loan Status (Fully Paid, Charged Off, Current)
  - Current status can be filtered out as this is not relevant to this analysis.
  - Remove or drop columns having 90% missing values
  - Remove columns which are not relevant for analysis.
    - last_pymnt_d,last_credit_pull_d,title,url ,zip_code
    - addr_state,delinq_2yrs,earliest_cr_line,inq_last_6mths,open_acc
    - pub_rec,revol_bal,revol_util,total_acc,out_prncp,out_prncp_inv
    - total_pymnt,total_pymnt_inv,total_rec_prncp,total_rec_int
    - total_rec_late_fee,recoveries,collection_recovery_fee,last_pymnt_d,
    - last_pymnt_amnt,last_credit_pull_d,application_type

  - Handle NULL or NaN values in the columns
  - Check and clean the column data. Example int_rate columns contains %
  - Check any required derived column

Data Preparation

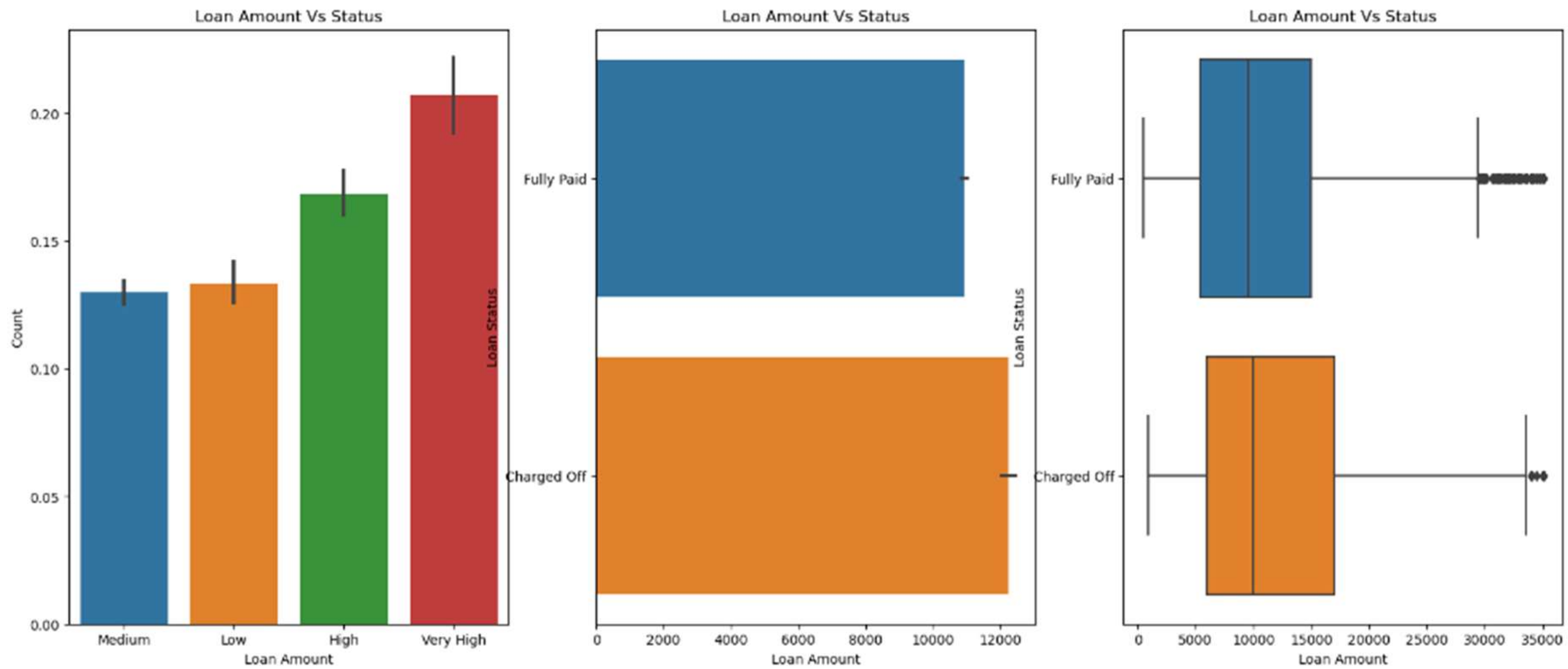
Clean Data, Construct Data, Format

- **Data Type Formatting**
  - Convert following columns to numeric type for standardization and easy analysis
  - Funded_amt(float)
  - Loan_amt(float)
  - Term(int)
  - Int_rate(float)

- **Category Type Formatting**
  - Convert following columns to category codes for standardization and easy analysis
  - loan_status
  - grade
  - sub_grade
  - verification_status
  - purpose
  - title
  - home_ownership
- **Derived Columns**
  - Add derived columns **issue_year and issue_month** from **issue_d** for EDA analysis

EDA Analysis

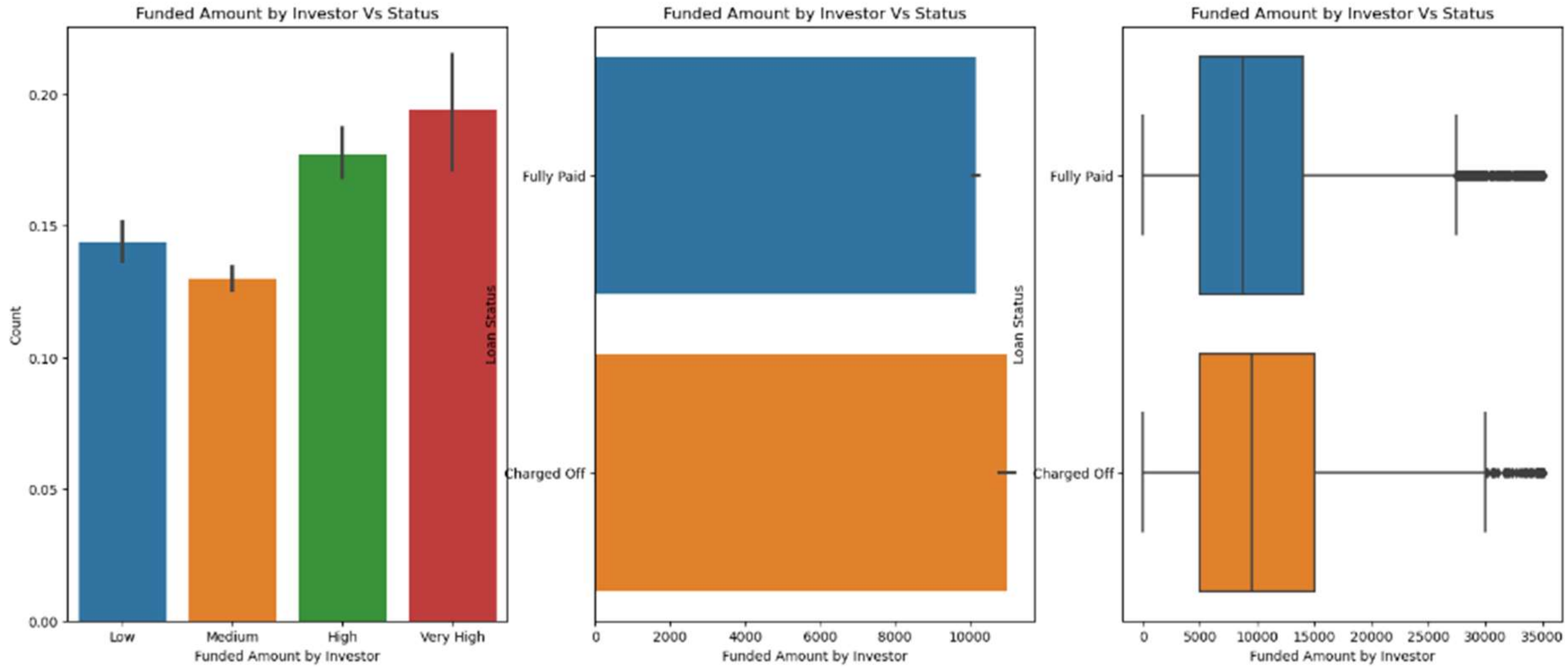
Driving Factors for default – Loan amount



Loan amount is the deriving factor for loan default. Higher loan amount increase the chance of loan default.

EDA Analysis

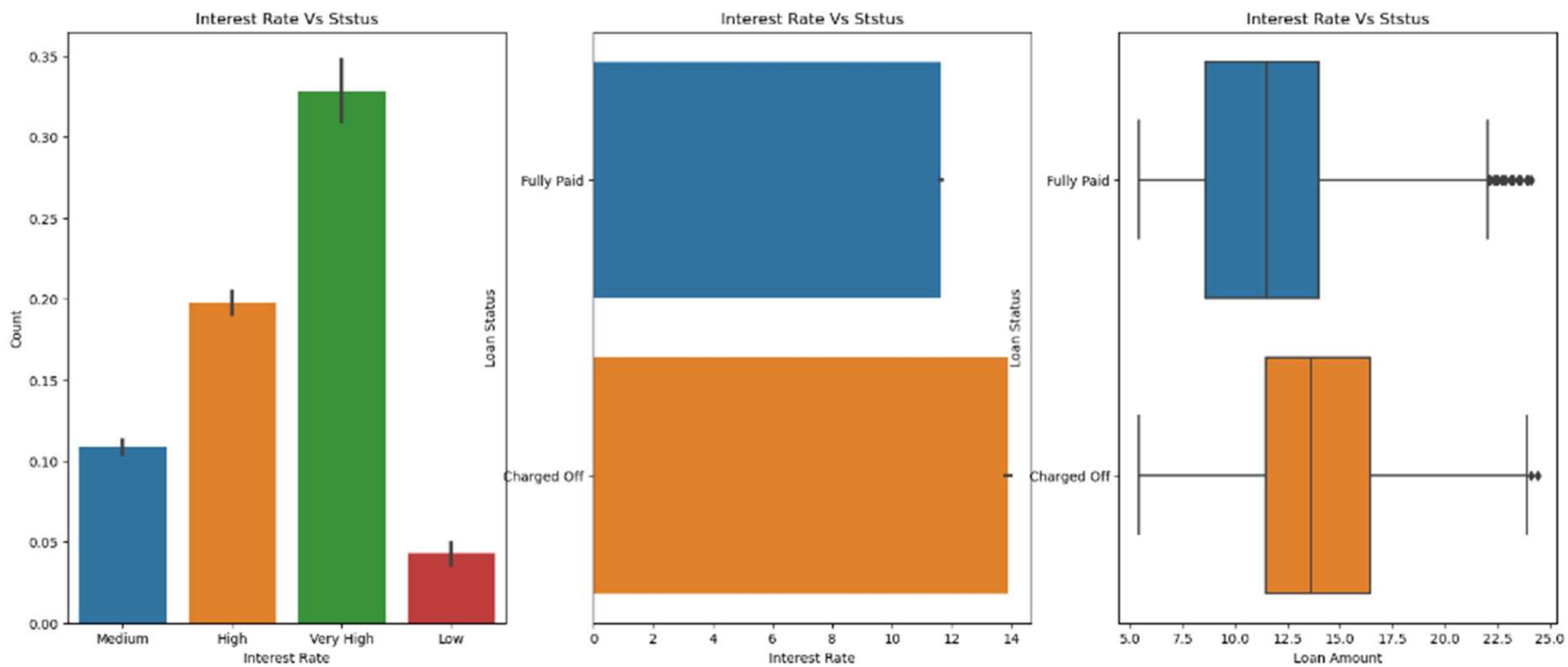
Driving Factors for default – Funded Amount by Investor



Fund by investor is the deriving factor for loan default

EDA Analysis

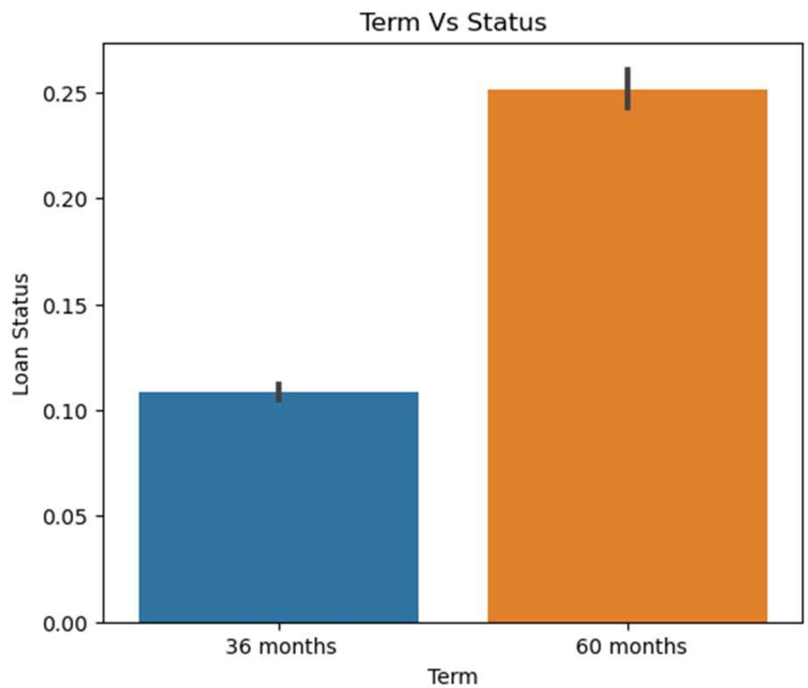
Driving Factors for default – Interest Rate



Higher Interest Rate is the deriving factor for loan default

EDA Analysis

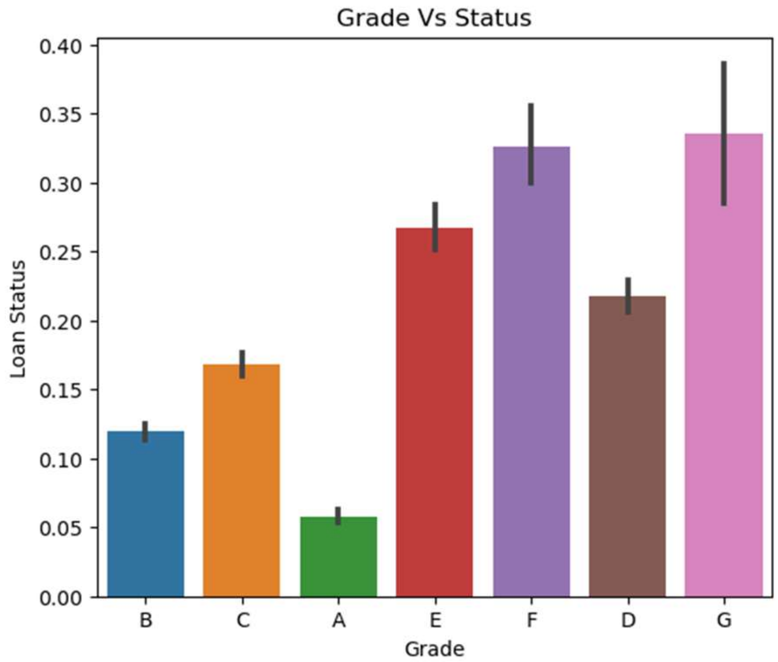
Driving Factors for default – Term



Longer Loan term is the factor of defaulting loan. 60 months has higher default rate.

EDA Analysis

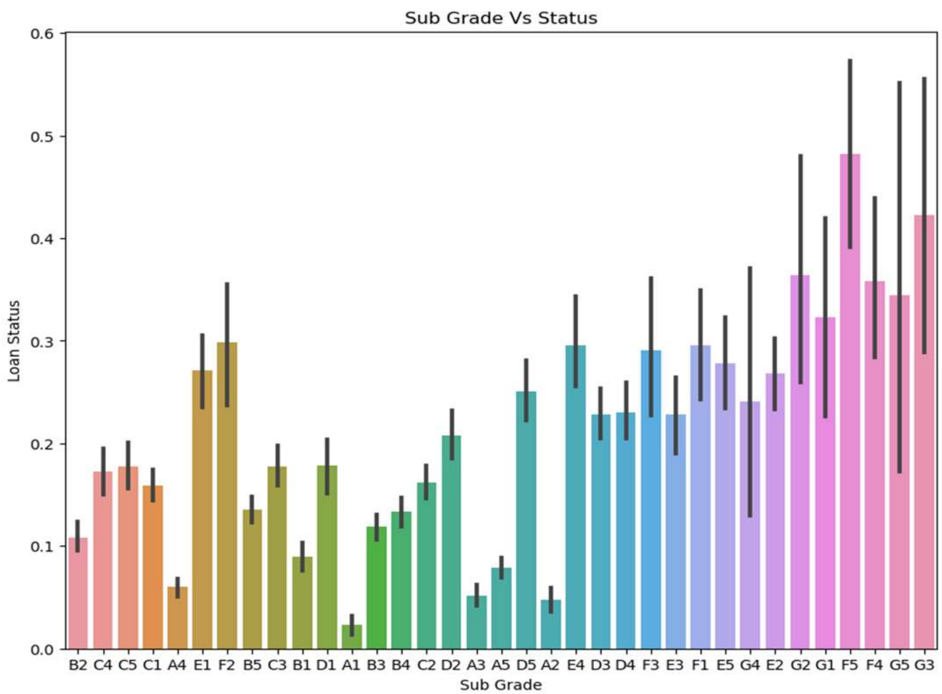
Driving Factors for default – grade



Grade is the factor of defaulting loan. A has lowest default rate.

EDA Analysis

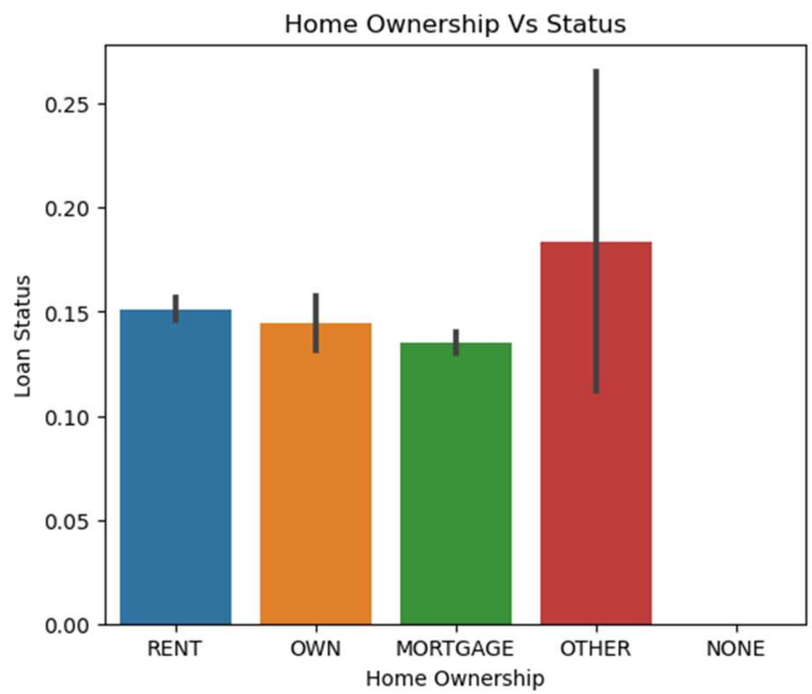
Driving Factors for default – Sub grade



Sub Grade is the factor of defaulting loan. F5 grade has highest default rate. A1 has lowest.

EDA Analysis

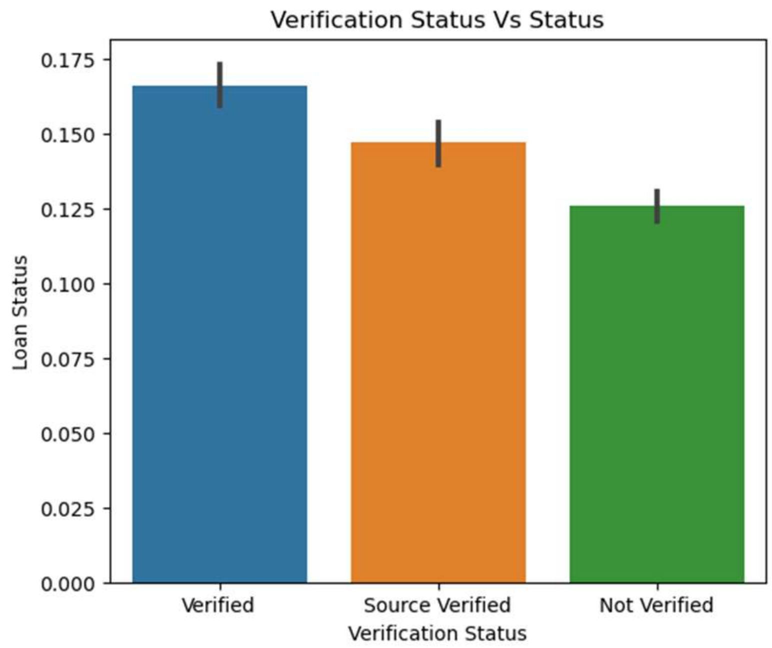
Driving Factors for default – Home Ownership



Home Ownership does not seems to be major factor for loan default as range is narrow.

EDA Analysis

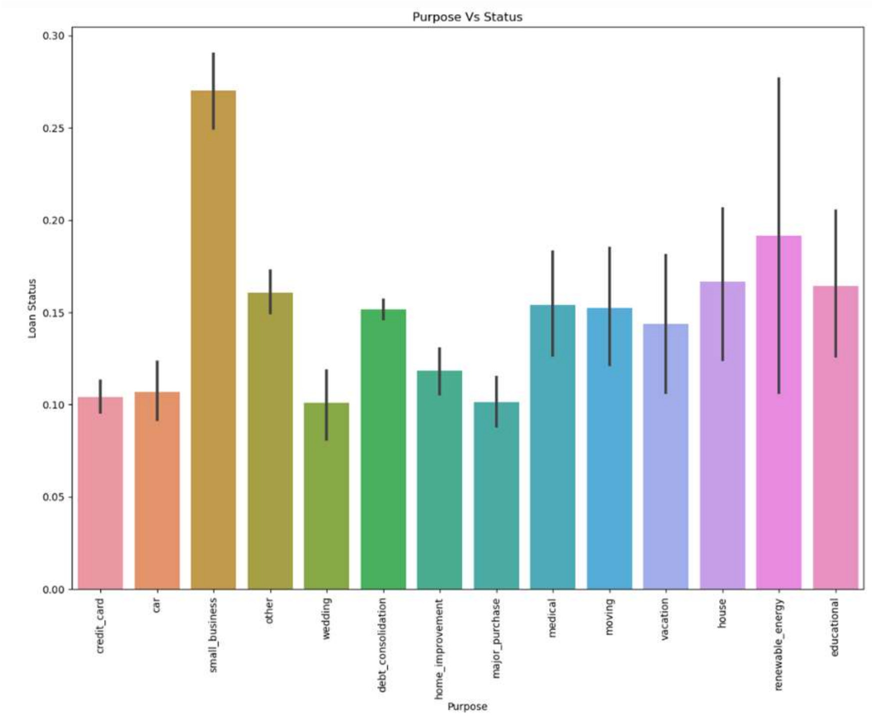
Driving Factors for default – Verification Status Code



Verification Status as Verified seems to show higher loan defaulting

EDA Analysis

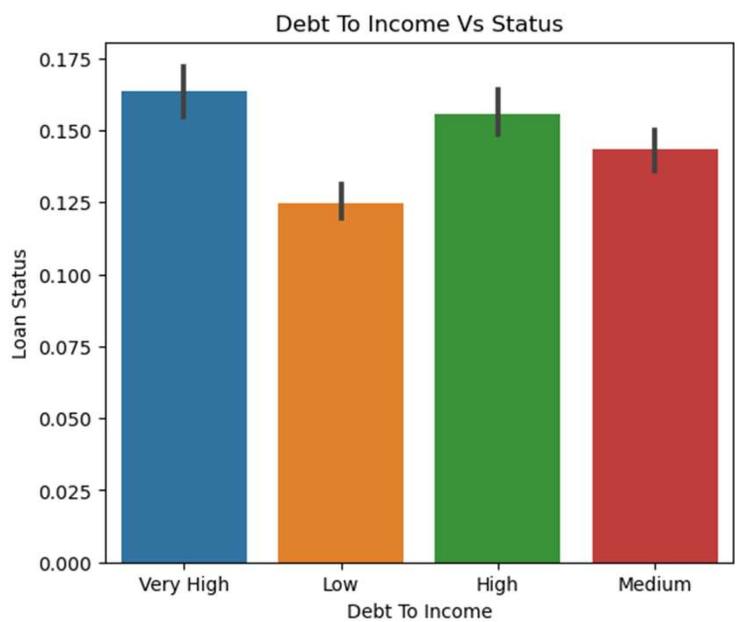
Driving Factors for default – Purpose



Purpose has impact on defaulting loan. Small Business has highest loan default rate.

EDA Analysis

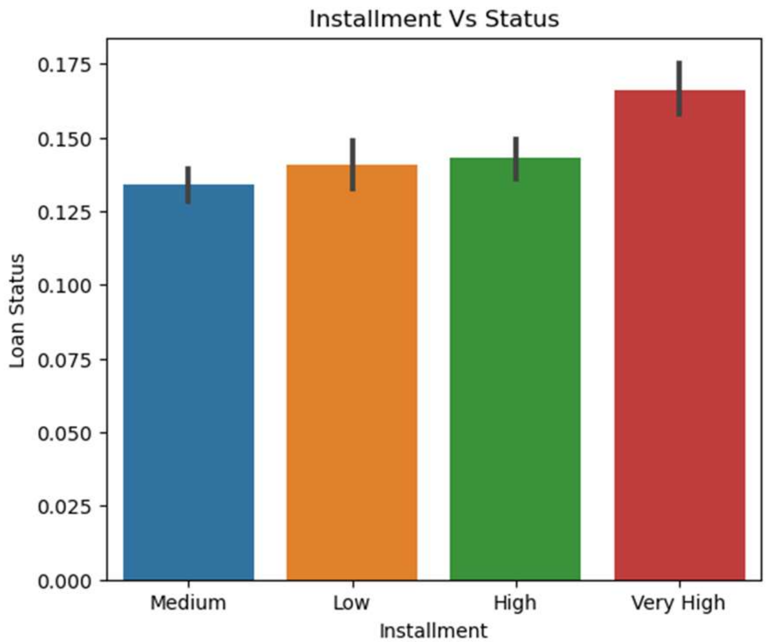
Driving Factors for default – Debt To Income



Debt to Income is the factor of defaulting loan. Higher debt to income results in higher rate of loan default.

EDA Analysis

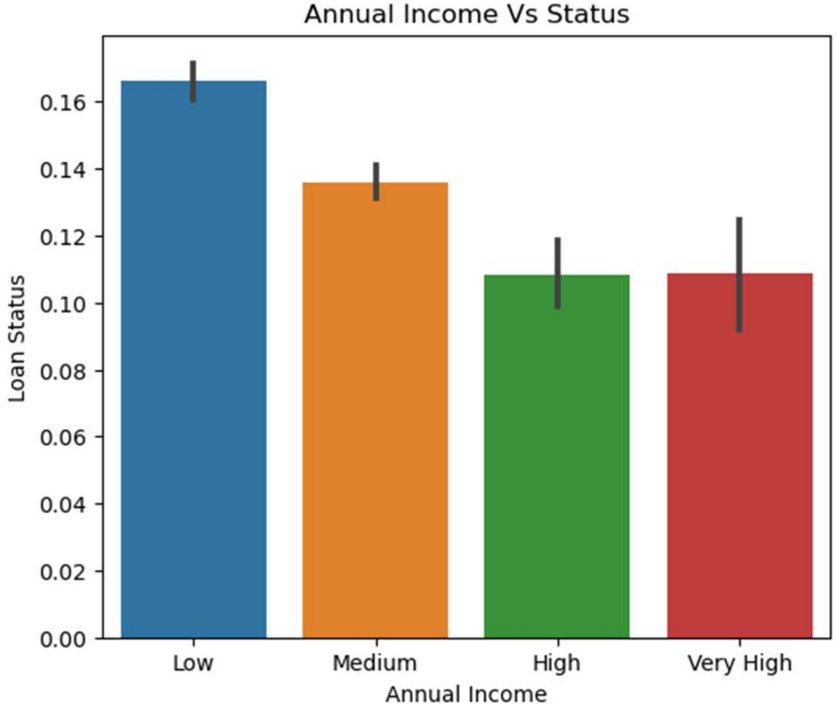
Driving Factors for default – Installment



Higher Installment is the factor of defaulting loan. Higher installment seems to have higher default rate.

EDA Analysis

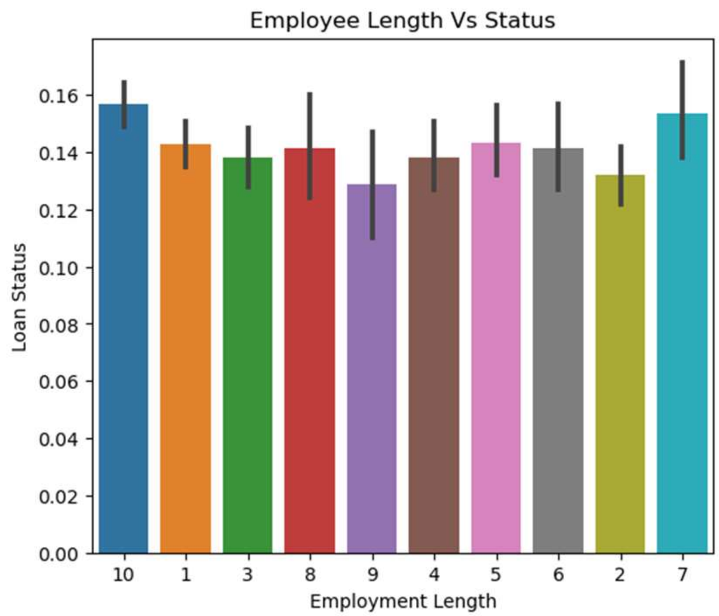
Driving Factors for default – Annual Income



Annual Income is the factor of defaulting loan. Low annual has higher loan default rate.

EDA Analysis

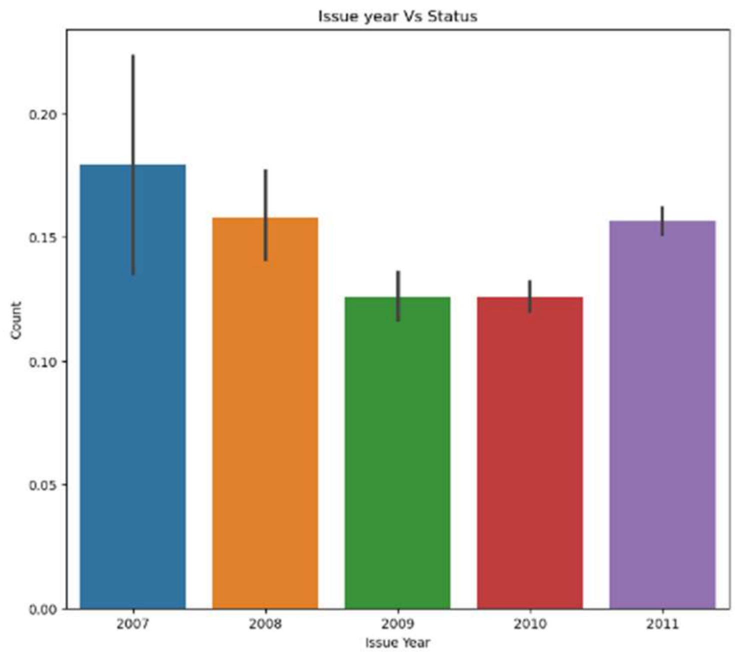
Driving Factors for default – Employee Tenure

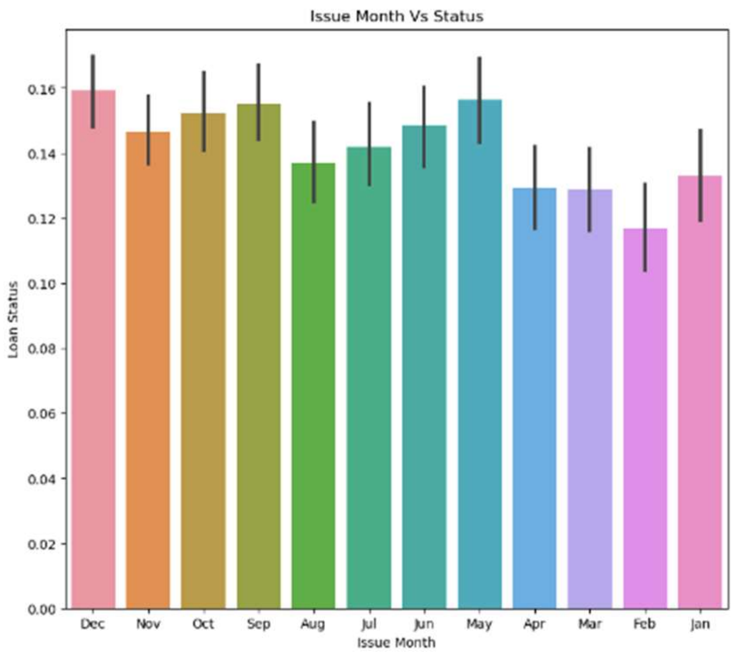


Employee Tenure does not seems to be major factor of defaulting loan.

EDA Analysis

Driving Factors for default – Issue Date (Month and Year)





As per analysis, issue_year(2007) has impact on defaulting loan. In 2011 loan default increased. Year can not be considered as major factor of loan default for future loans.

Conclusion

Conclusion

- As per analysis, following Consumer and Loan attributes are factors for loan default
  - Interest Rate
  - Grade/Sub Grade (Both are correlated so either of them can be considered)
  - Term (Loan term)
  - Loan Amount/ Funded Amount (Both are correlated so either of them can be considered)
  - Debt to Income
  - Installment
  - Annual Income
  - Purpose