Analyzing Loan Default Risk Factors –Lending Club

Insights from Exploratory Data Analysis

Rajesh Rao and Akshit
07-Nov-2023

Data Analysis Presentation

- **Introduction**
  Overview and purpose of the data analysis

- **Data Overview**
  High-level summary of the key variables and metrics in the dataset

- **Data Preparation**
  Steps taken to clean, transform, and prepare the data for analysis

- **Univariate Analysis**
  Key insights from analyzing the distribution of individual variables

- **Segmented Analysis**
  Findings from analyzing key metrics segmented by various attributes

- **Bivariate Analysis**
  Results of analyzing relationships between pairs of variables

- **Insights and Recommendations**
  Suggested actions based on the key insights from the analysis

Data Overview

Dataset has over 39,000 loan records

The records are from 2007 to 2011 and detail borrowers' backgrounds, loan details, and repayment status.

Objective: "The goal is to identify patterns and factors contributing to loan defaults, marked as 'Charged Off' in our dataset.

This dataset will allow us to analyze patterns and factors associated with loan defaults to build a predictive model.

Data Preparation

Imported data into Pandas

Utilized Pandas for efficient data importation and preliminary analysis

Performed statistical summaries to grasp data attributes.

Cleaned the data

Removed columns with missing values and redundant fields to improve data quality

Omitted columns with over 95% missing values to ensure data quality.

Excluded columns with predominant zeros, reducing noise.

Removed transaction-specific and redundant fields irrelevant to the study

Transformed the data

Standardized date fields and converted data to numeric format for analysis
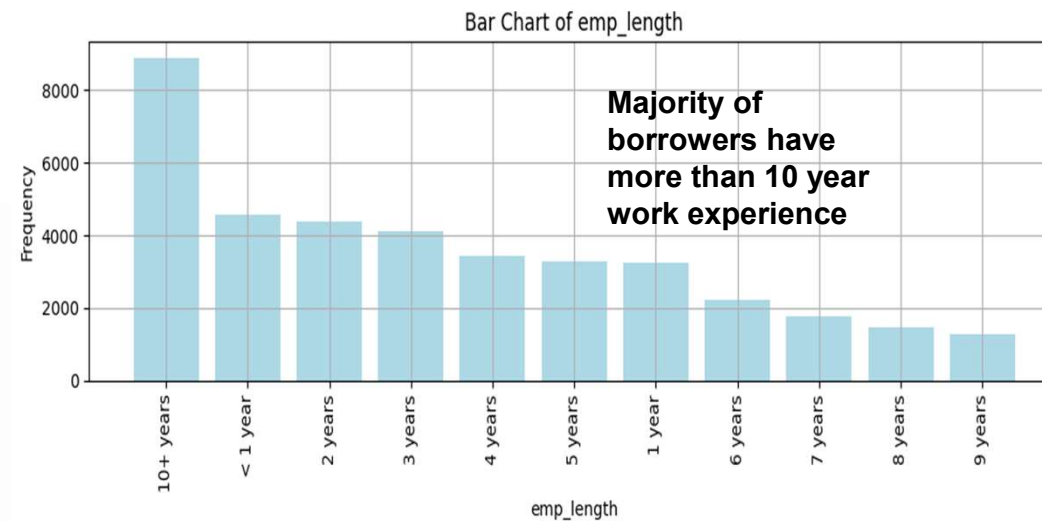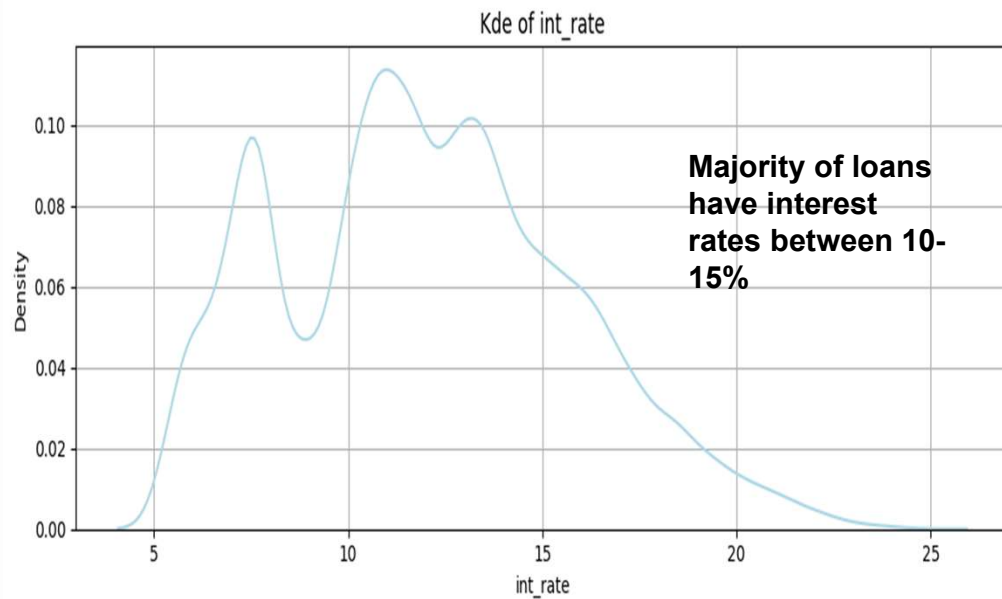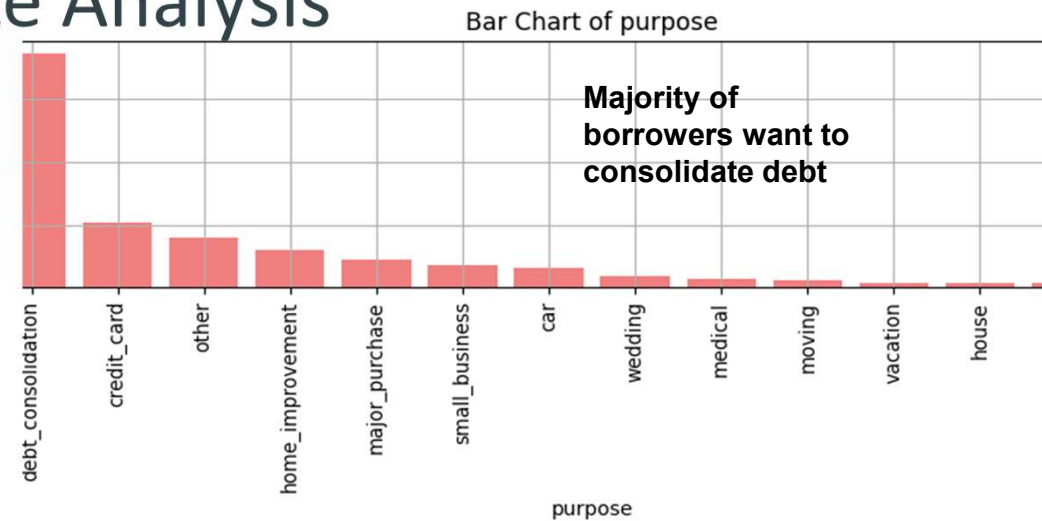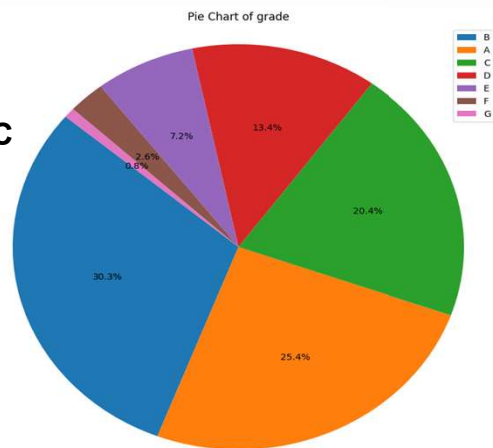
Standardized date fields for e.g issue_d and earliest_cr_line for temporal analysis.

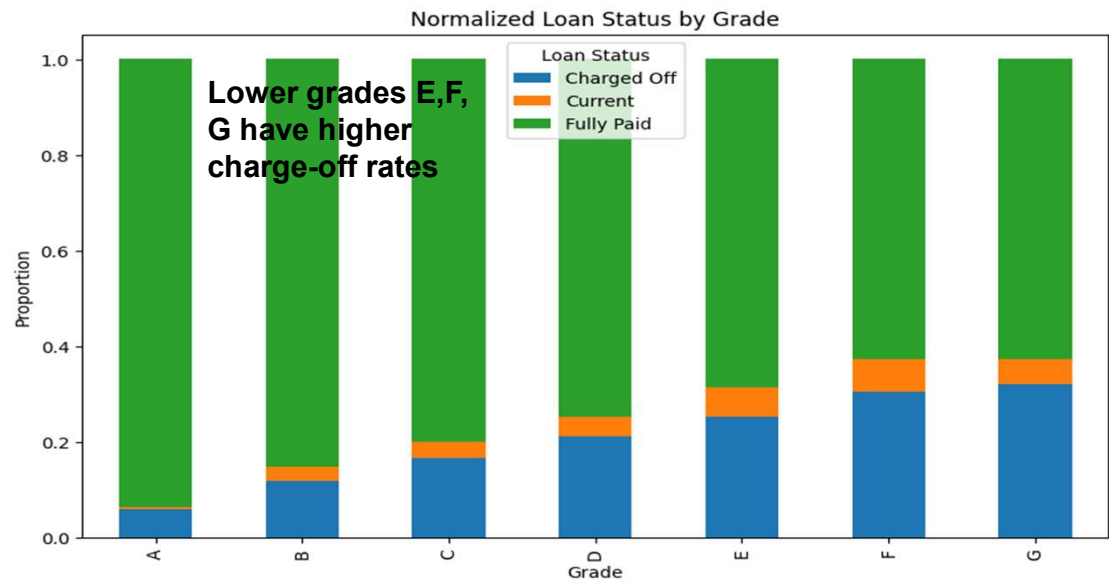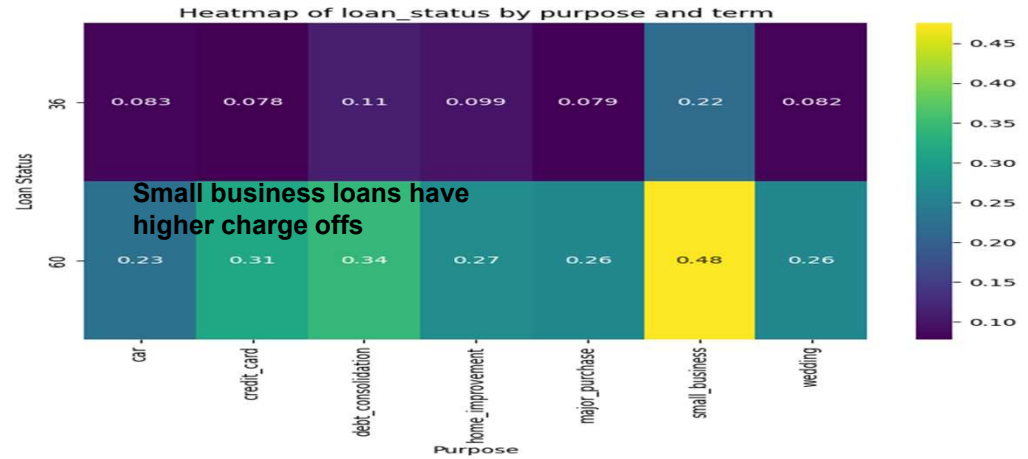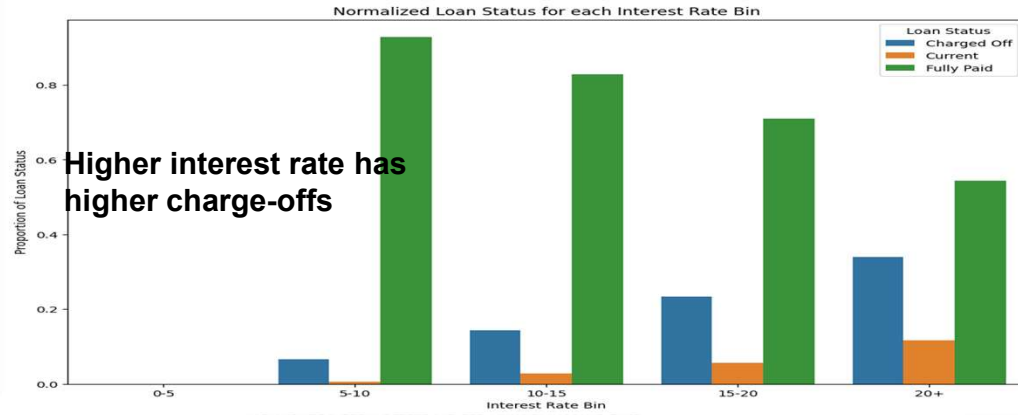Stripped non-numeric symbols from fields and converted to numeric for uniformity.

The data was successfully imported, cleaned and transformed for further analysis

Univariate Analysis

Pie Chart of grade

Nearly 75% of loans are provided to A,B,C grades

Bar Chart of purpose

Majority of borrowers want to consolidate debt

Kde of int_rate

Majority of loans have interest rates between 10-15%

Bar Chart of emp_length

Majority of borrowers have more than 10 year work experience

Segmented Univariate Analysis

Bivariate Analysis

60-month term exhibits more charge-off than 36-month term

Lower grades E,F, G have higher charge-off rates

Certain stages like NE, NV have significantly higher charge-off

Higher debt to income ration(DTI) leads to higher charge-offs
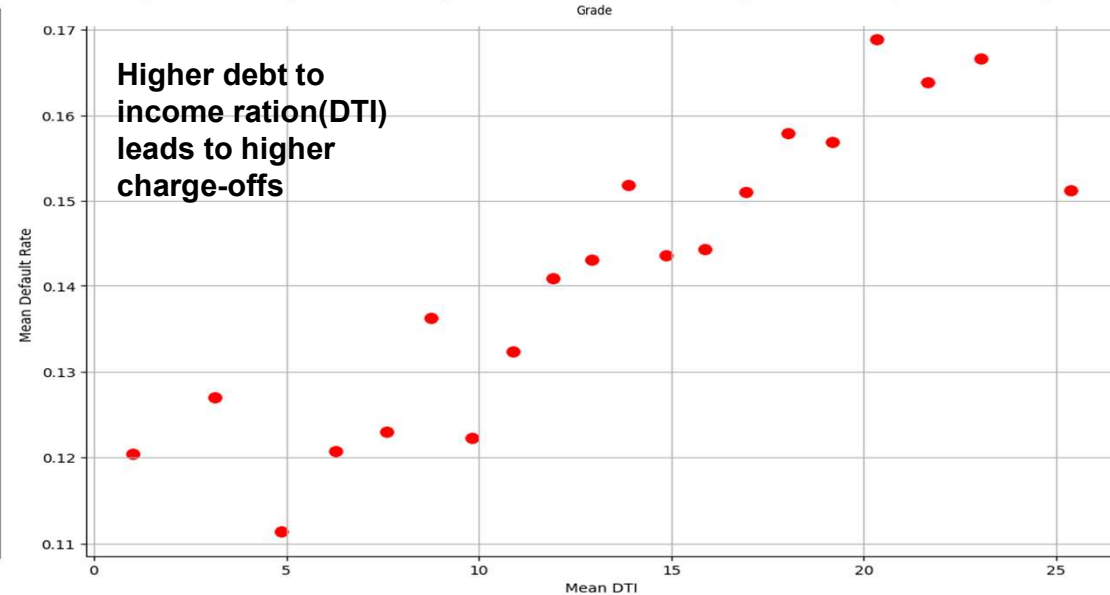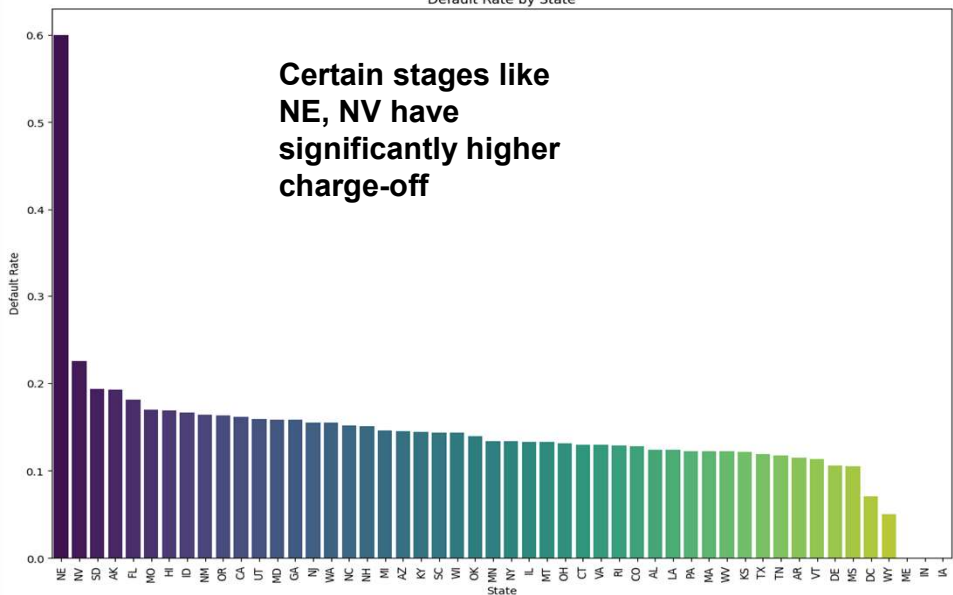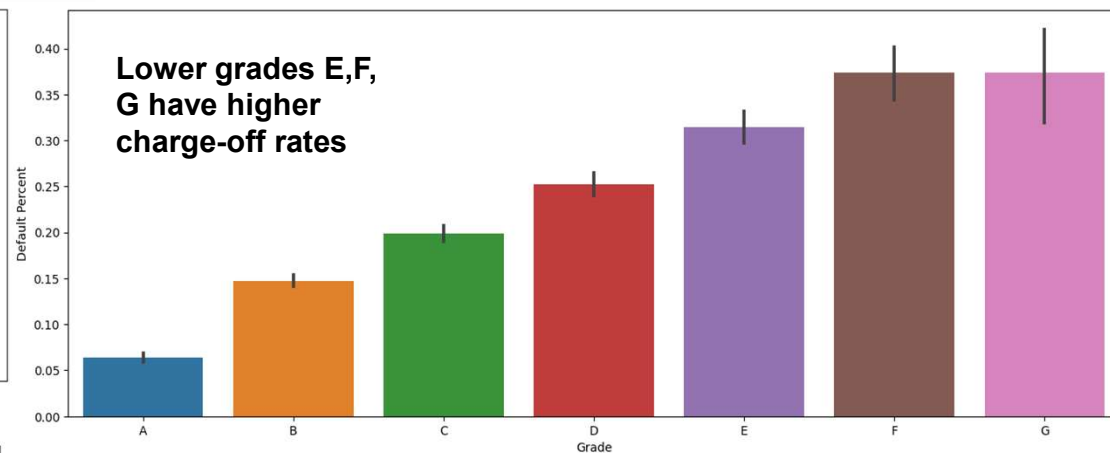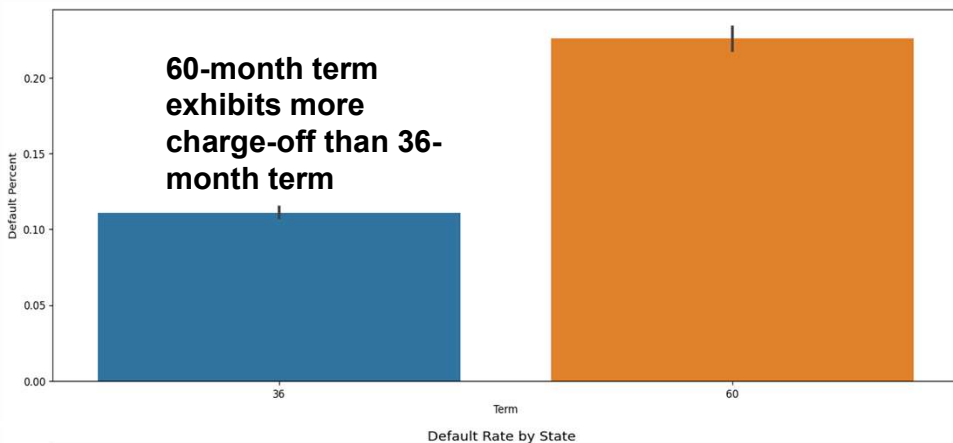
Key Findings

- **High default rates for small business loans**
  Due to variable success of small businesses

- **60-month loans originated in 2010-11 have higher defaults**
  Possibly due to economic conditions or longer repayment period

- **Higher loan amounts correlate with higher defaults**
  Increased repayment burden if finances change

- **Customers with high debt-to-income ratio default more**
  High DTI indicates financial strain

- **Lower income customers have higher defaults**
  Limited finances make repayment challenging

- **Lower credit grades have higher defaults**
  Indicate lower creditworthiness

Managing Lending Risk Recommendations

Higher risk loans

Implement stricter lending criteria and higher interest rates for higher risk loans like small business loans or loans with high DTIs or low income borrowers.

Lower credit grades

Require co-signers and charge higher risk premiums for borrowers with lower credit grades and higher DTIs

Longer loan terms and State Specific Strategies

Introduce early intervention strategies and graduated risk assessments for longer 60+ month loans. Develop state-specific strategies to combat charge offs

By tailoring lending practices to specific risk factors, banks can expand lending to higher risk groups while still managing overall credit risk.