Cracking the Code: Predicting Loan Defaults with Lending Club Data

By:
Sanskar Gupta: sanskar752000@gmail.com
Mahima Awasthy: awasthymahima@gmail.coml

1

Introduction

Let’s check out what it is all about.

Lending Club *is a peer-to-peer lending platform that connects borrowers with investors. The company was founded in 2007 and has since facilitated over $50 billion in loans.In this presentation, we will analyze the data provided by Lending Club to identify the types of borrowers who are most likely to default on their loans.*

“

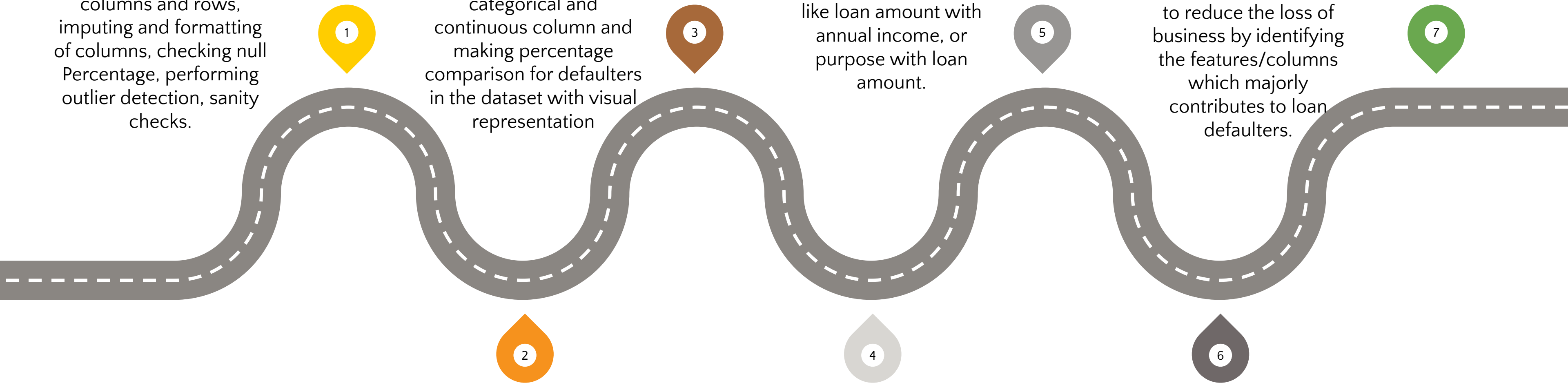
This is a Dataset Overview

- Dataset contains 39717 loans and 111 features
- Dataset contains record from year 2007 till 2011
- Loan status is the target variable.
- Other important feature includes loan term, grade, verification status, purpose, loan amount, interest rate, installments, annual income, DTI, etc.

Exploratory Data Analysis

EDA involves analyzing the dataset to discover patterns, anomalies, and relationships. It helps to identify missing values, outliers, and correlated features. EDA techniques include descriptive statistics, visualizations, and hypothesis testing.

Roadmap - Problem Solving Methodology



Data Cleaning
Removing the null the columns and rows, imputing and formatting of columns, checking null Percentage, performing outlier detection, sanity checks.

Data Understanding
Working with the dataset to figure out the important features or columns that can be used for better analysis

Univariate Analysis
Analysing each categorical and continuous column and making percentage comparison for defaulters in the dataset with visual representation

Segmented Univariate Analysis
Analysis on categorical and continuous variable by forming the bins, measuring that effective proportion leading to charged off

Bivariate Analysis
Analysing two variables, like loan amount with annual income, or purpose with loan amount.

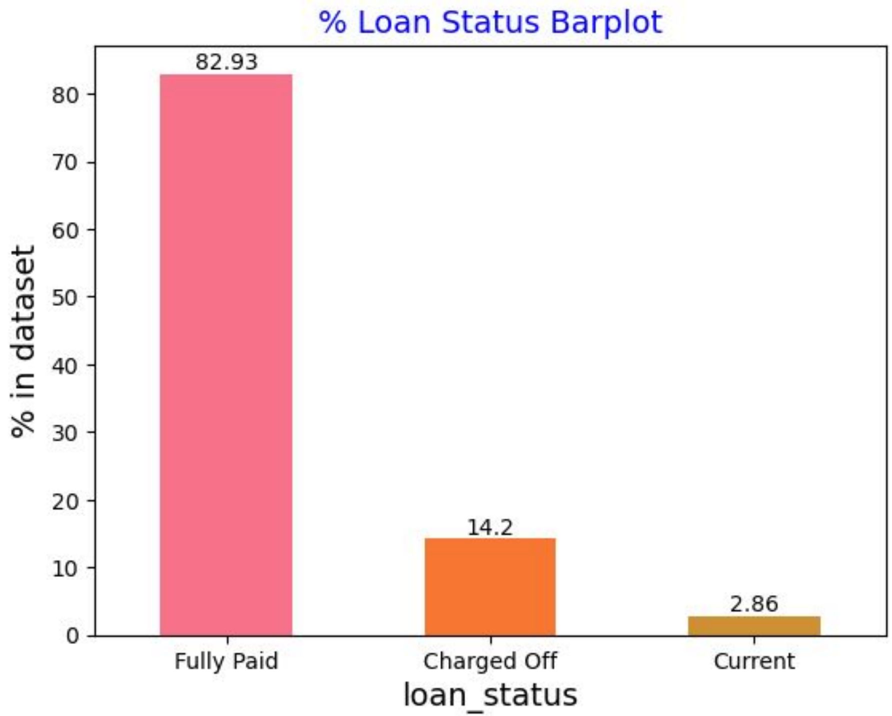
Multivariate Analysis
Analysing more than two variables to get better insights and better prediction the borrower likely to default.

Recommendations
After analysing the data gave recommendations to reduce the loss of business by identifying the features/columns which majorly contributes to loan defaulters.

Analysis: Loan Status % in Dataset

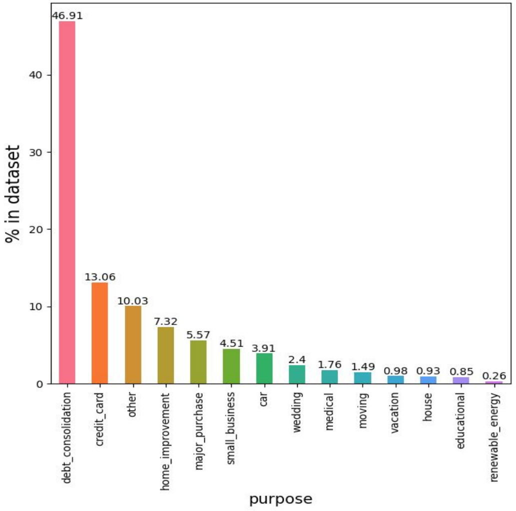
- 82.93% of borrowers have fully paid the loan amount.
- 14.2% of borrowers got Charged Off/defaulted.
- 2.86% of borrowers are still paying their loan.

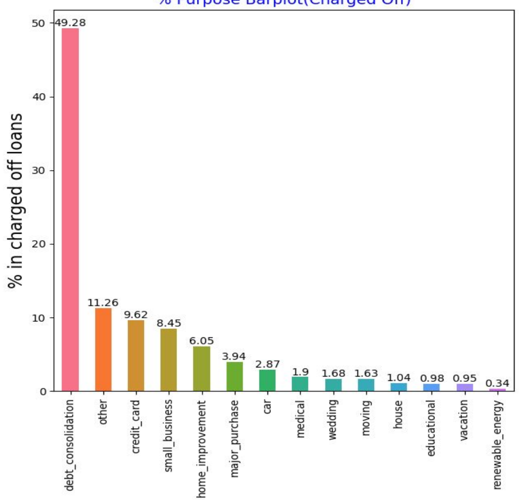
Moving forward we will be focusing on Charged Off borrowers.

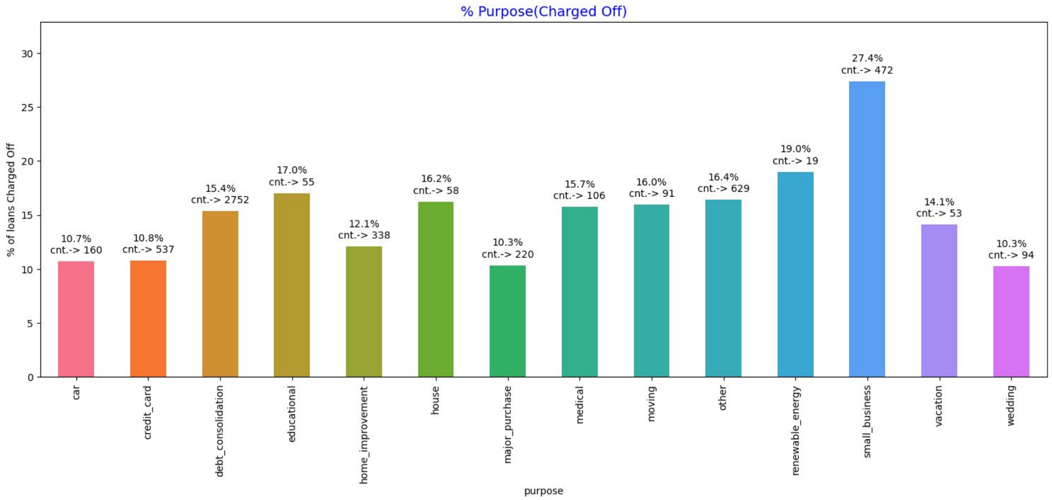


Analysis: Purpose % in Dataset

- % of loans taken for small business among the charged off loans got doubled, which means – loans taken for the small business can be risky and lead to defaulters
- Also % of loans taken for debt consolidation is high which represents that most of the borrower aren't able to manage their finance and pay their bills causing them to default

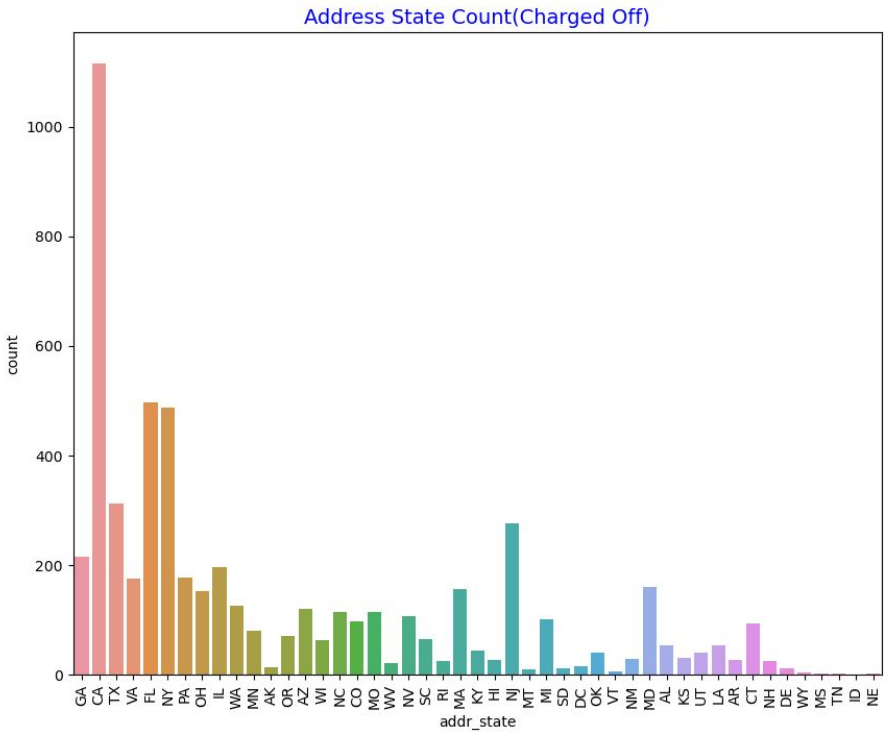






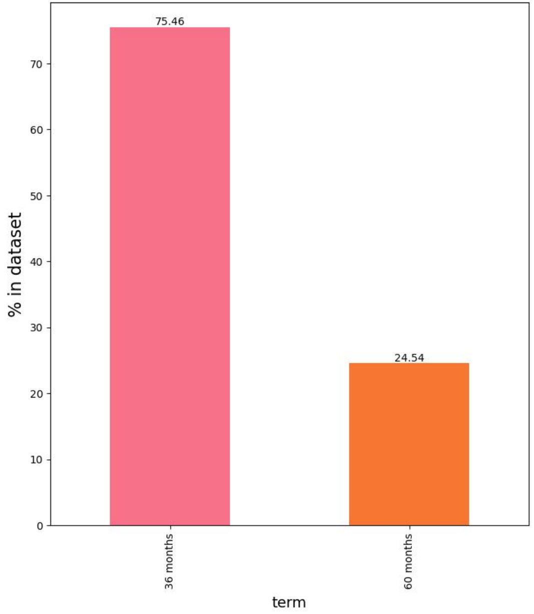
Analysis: Address State in Dataset (Charged Off)

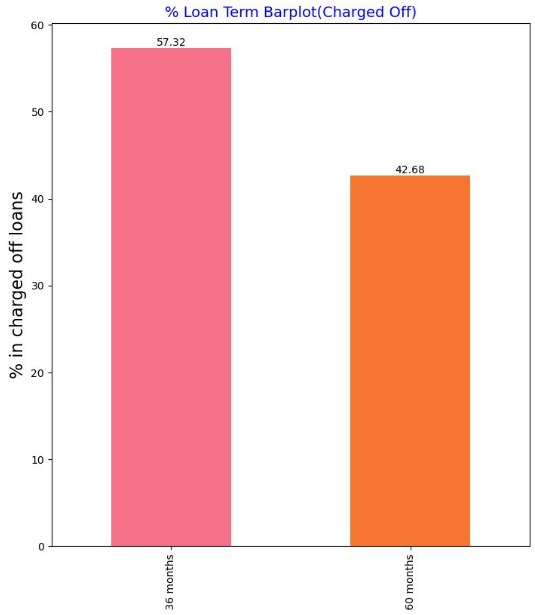
- Loans which are provided in the states of California, Florida, and New York have higher default rates due to the high cost of living in those areas, making it difficult for people to manage their finances.
- Lending Club should assess the annual income of people living in those states before granting any loans. If the annual income is low, then they should consider offering a loan at a higher interest rate or should reduce the approved loan amount.



Analysis: % Loan Term in Dataset (Charged Off)

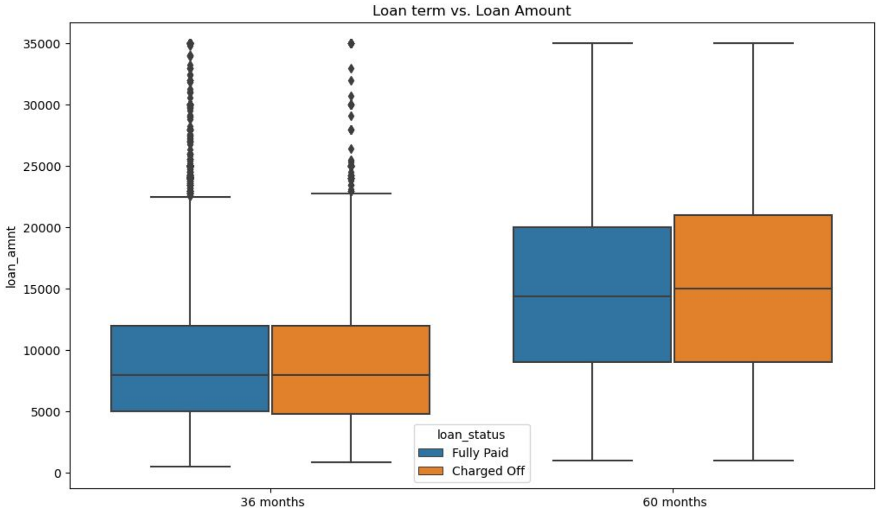
- Among Charged Off loans % of 60 months term rises to around 42.68% which is high.
- It leads to the observation that loans taken for higher term have a higher chance of default.

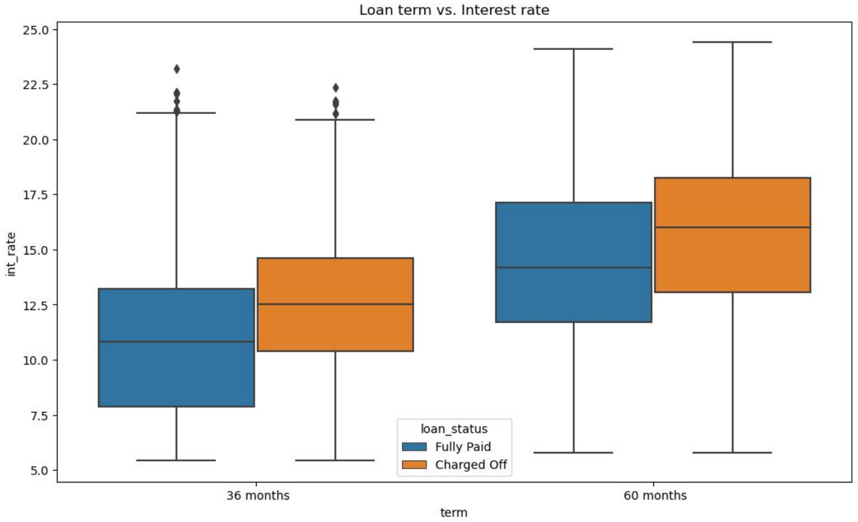




Analysis: Loan term vs. Interest Rate and Loan Amount

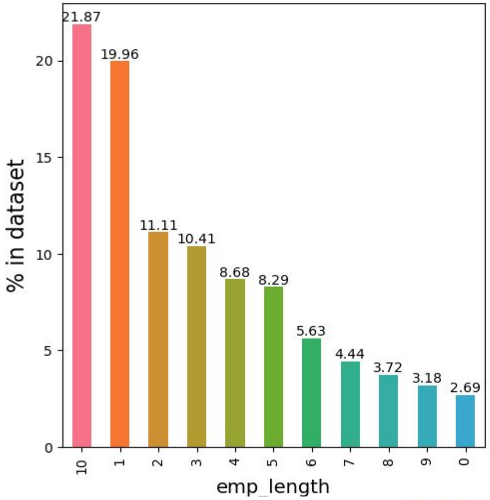
- Based on the graph at the top, those who have taken larger loan amount have opted 60 months term
- Based on the graph at the bottom, shows that these people with 60 months term pay higher interest.
- So the borrower are taking higher amount loan at higher interest for 60 months term, face difficulties paying back the loan.

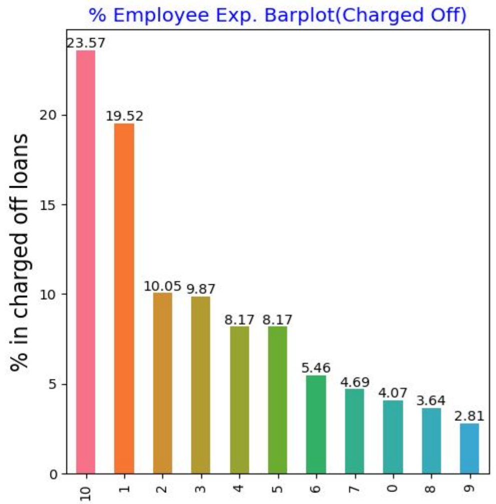


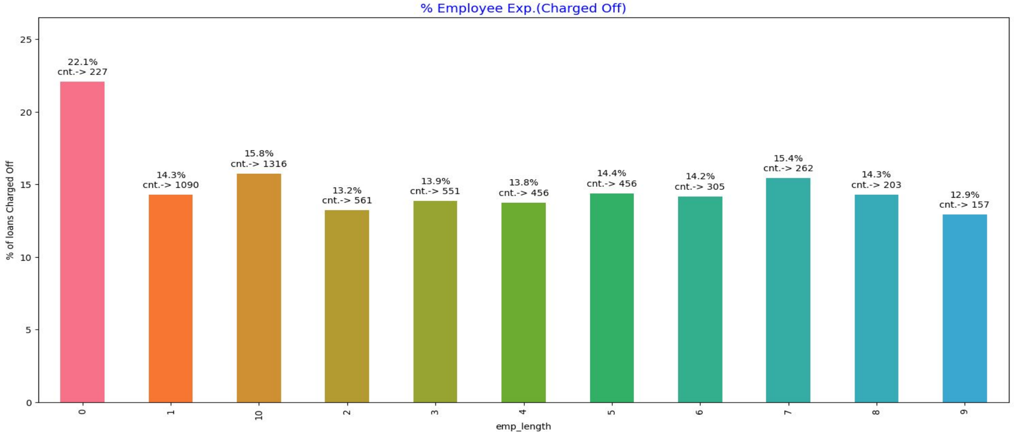


Analysis: % Employee experience in Dataset (Charged Off)

- If we look at the barplot at the top it shows among Charged Off loans, borrower having 0 years of experience rises to 4.07%, just got doubled compare to complete dataset.
- If we look at the barplot at the button it shows the higher percentage of defaulted borrowers have 0 years of experience.
- Lending Club should prevent giving loans to people who have 0 or near to 0 years of experience.

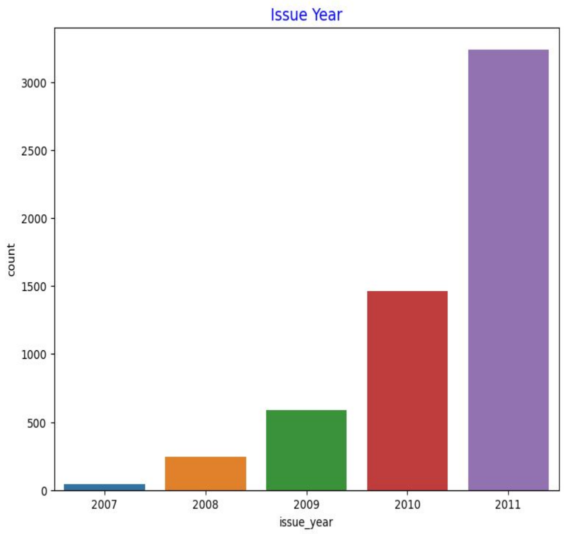


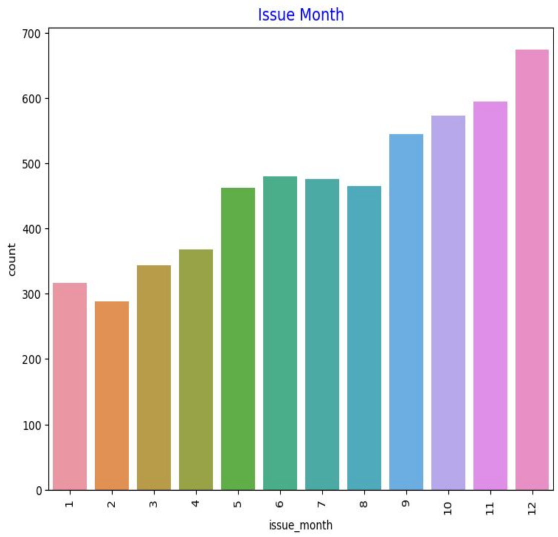




Analysis: Years and Months (Charged Off)

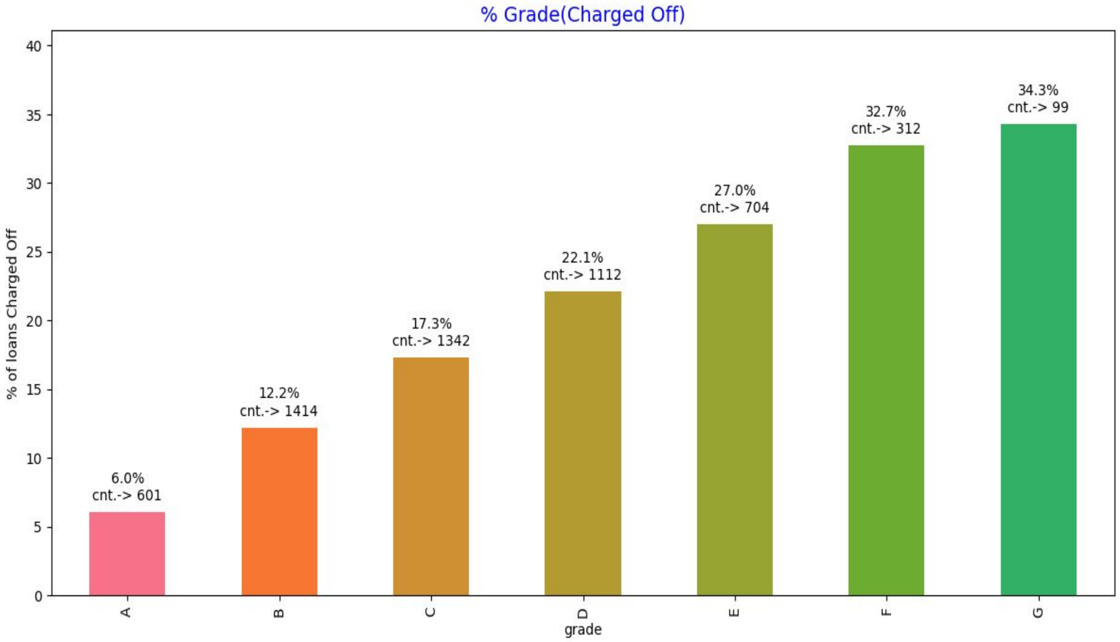
- During 2011 the number of defaulters was high, this could be due to the US still recovering from the severe recession that had occurred within the past two to three years.
- The number of defaults in December was especially high, this could be attributed to the festive season and borrowers spending more money on expensive gifts for Thanksgiving, which may have led to financial imbalances.





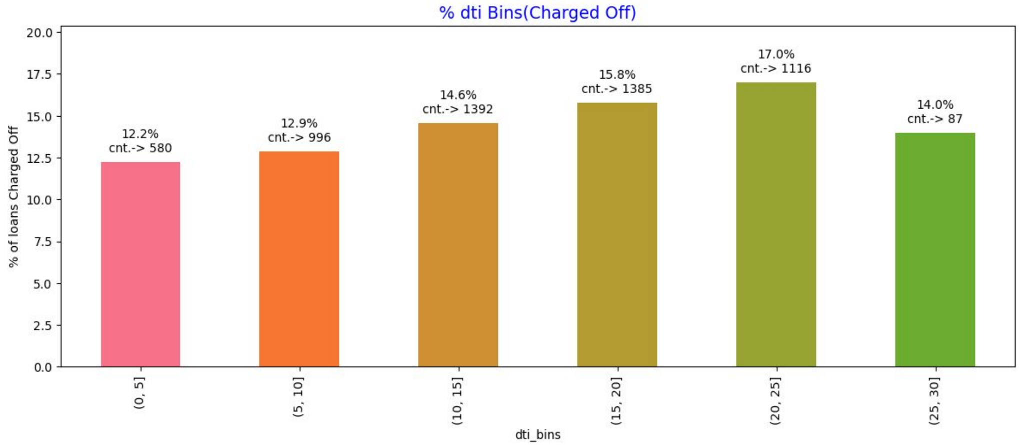
Analysis: Grades in Dataset (Charged Off)

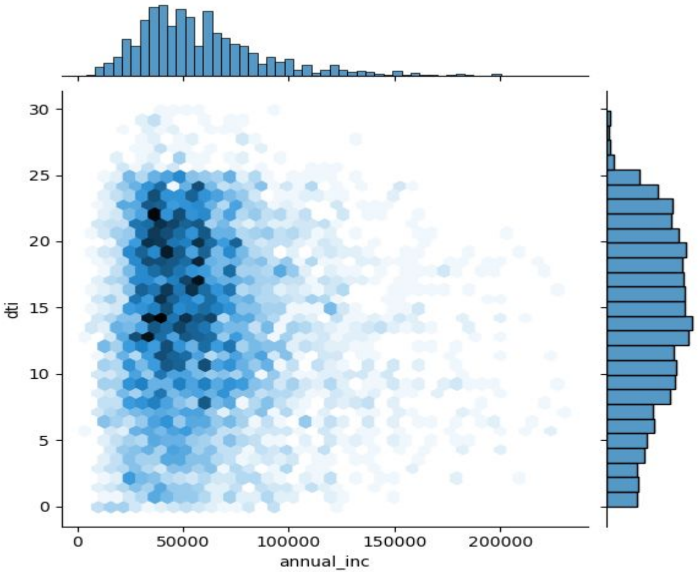
- Grade represents the credit worthiness of the borrower.
- We can see from the graph that most of the defaulters are grouped in D, E, F and G grades.
- Lending Club should prevent giving loan to these types of borrowers or should charge them higher interest rate.



Analysis: Annual Income compared to DTI in Dataset (Charged Off)

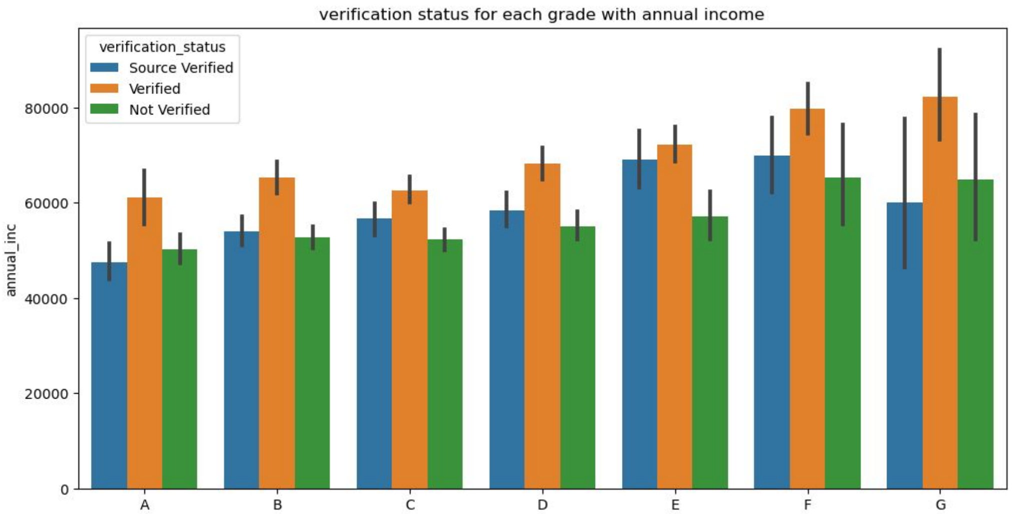
- DTI represents the debt to income ratio which explains percentage of debt with respect to annual income.
- Based on the graph shown at the top, higher the dti, higher will be the defaulters. DTI after 20 shows high default percentage.
- Based on the graph shown at the bottom, we can say that borrowers having annual income less than $50000 forms the larger group of defaulters having DTI of around 20-25 which is risky and shows that these borrowers have taken loans for which their income is not suitable to pay back to lenders.
- Lending Club Should limit the amount of loan granted to these types of people.

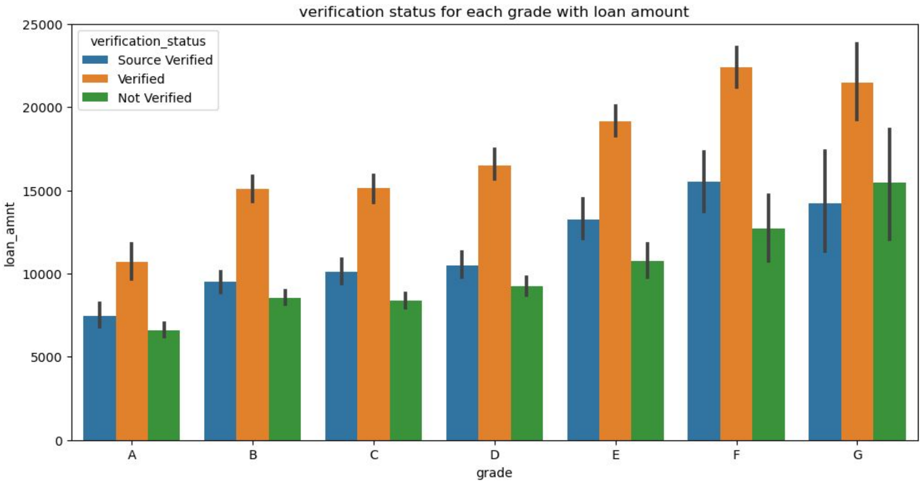




Analysis: Verification status for Grades with Annual Income and Loan amount(Charged Off)

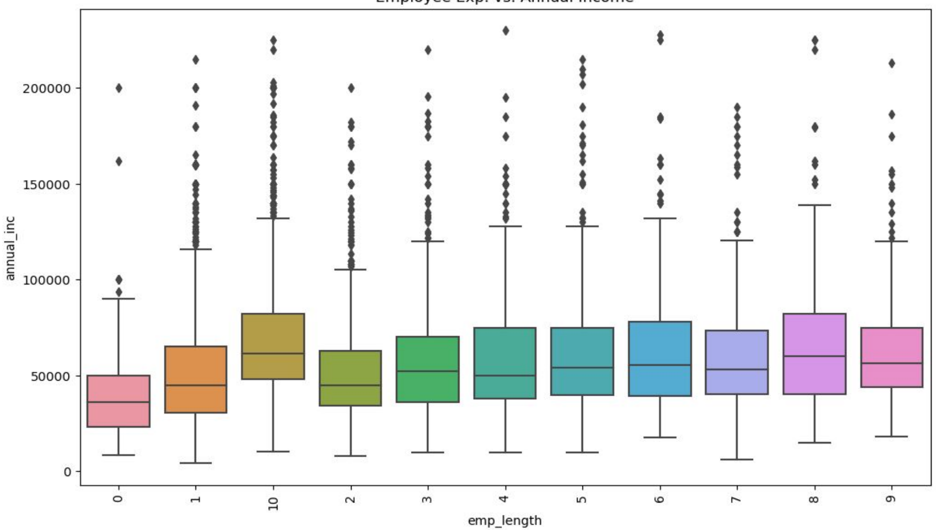
- Based on the graph shown at the top, lower grades Like E, F, and G who are not verified, have higher annual income which could be false information given by the borrower.
- The graph at the bottom shows that these non verified, low grade borrowers have taken loan of larger amounts which is causing them to default as we have just stated that their income source is not verified.
- Lending Club should source verify these types of low grade borrowers, before granting the loan as they are likely to default.

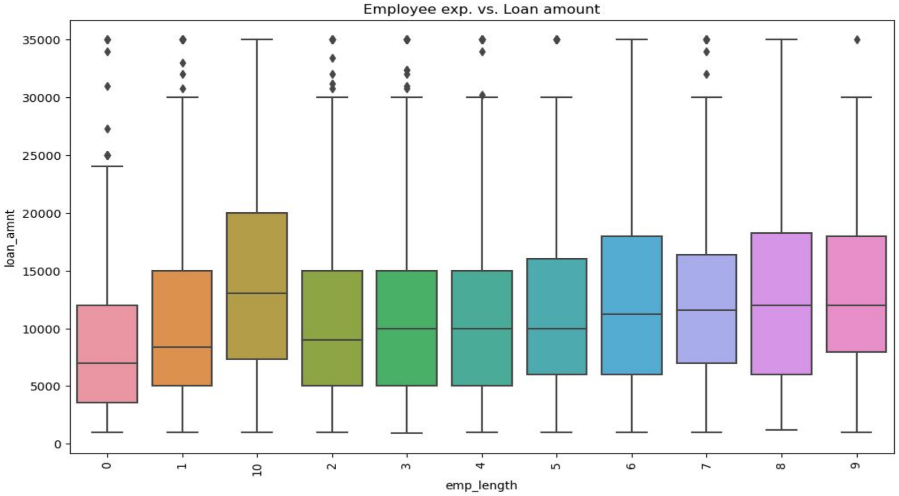




Analysis: Employee Exp. compared to Annual Income and Loan Amount (Charged Off)

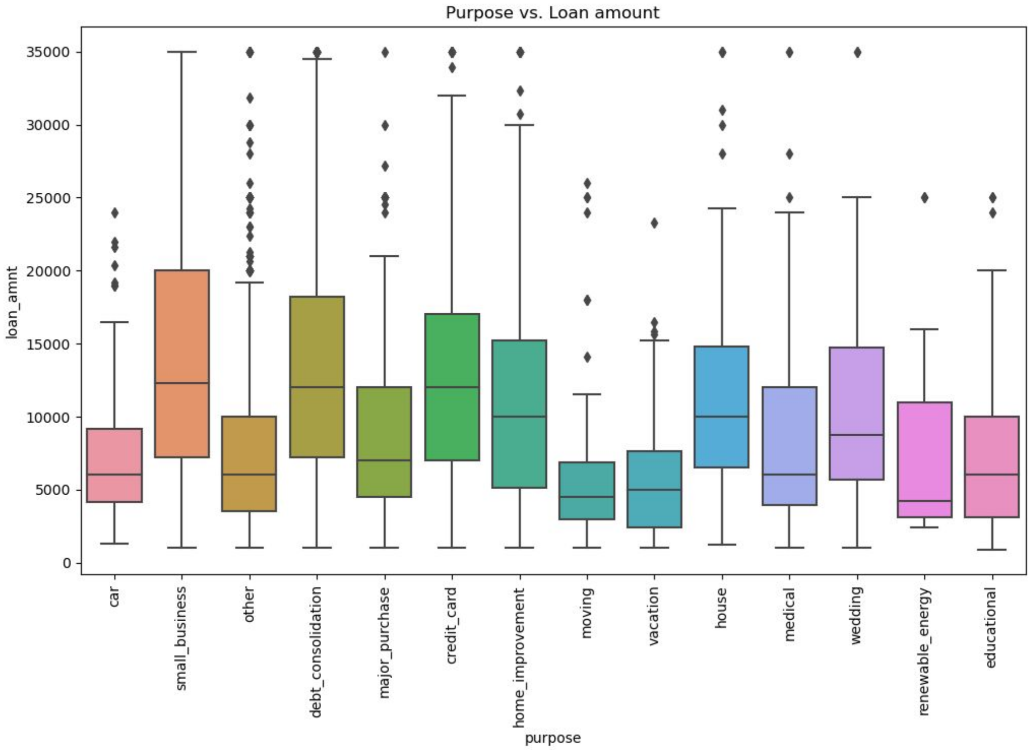
- Based on the graph shown at the top, borrower with 0 years of experience have less annual income which is less than $50000.
- But if we look at the graph at the bottom, it shows that borrower with 0 years of experience have took loan of larger amount of upper extreme around $24000 which is risky, likely causing them to default.
- Lending club should restrict on the amount of loan given to the borrower having 0 or near to 0 years of experience.





Analysis: Purpose compared to Loan amount (Charged Off)

- Our earlier analysis on Purpose in previous slide showed that most of the defaulter percentage are the people taken loan for the small business purpose.
- Now with the graph in this slide we can clearly see that people taken loan for small business, took larger amount loan and we have seen that borrower who have less annual income did falls in higher DTI bin means having they took larger loan, in our previous slide, causing them to default.
- Lending Club should give more attention to these types of borrower taking loan for this purpose and should charge them higher interest or limit the amount of loan that can be available for them.



Recommendations

- Lending Club should source verify the annual income of the borrower. Loan should not be approved or if approved should be very less in amount for non-verified borrowers.
- They should not grant higher amount of loan to the people who have 0 or near to 0 experience
- They should charge higher interest rate or reduce the amount of loan or prevent giving loans to lower grade borrower such as E, F, and G.
- If the loan is taken for the purpose of small business or debt consolidation higher interest should be charged or limit the amount of loan available to these types of borrower.
- DTI should be measured more closely before approving the loan. We found in the analysis that after DTI of 20 chances of loans default gets high.
- They should limit the amount or increase the interest rate whenever in any year there is financial crisis happening and this decision should be maintain for upto 2-3 years.
- They should reduce the loans taken with higher interest for 60 months, because these are likely to get default.