Case Study: Lending Club

Exploratory Data Analysis

Vaibhavi Khachane
Date: 29 May 2024

Table of Content

Introduction

- **Brief overview of Lending Club-**
  - In this case study, we'll delve into **LendingClub's** loan data, explore key factors influencing loan defaults, and propose strategies for mitigating credit losses
- **Importance of analyzing loan defaults-**
  - **Risk Management**: LendingClub faces the challenge of optimizing loan approvals while minimizing risks associated with loan defaults.
  - **Default Prediction**: By analyzing historical data, we can identify patterns and factors that contribute to loan defaults.
  - **Strategic Decision-Making**: Insights from this analysis help LendingClub make informed decisions about loan underwriting, risk assessment, and portfolio management.
  - **Investor Confidence**: Investors rely on accurate default predictions to allocate funds effectively and manage their investment portfolios.

Business Context & Problem Statement

- A consumer finance company specializing in lending loans to individuals faces the challenge of optimizing loan approvals while minimizing risks associated with defaulters.
- The task at hand is to analyze historical data related to loan applications.
- The goal is to identify the key factors that drive loan defaults.
- By understanding these factors, the company can make informed decisions to improve their loan approval process and reduce the risk of default.

Table of Content

Data Understanding:

- **Dataset**: Historical loan data from Lending Club
- **Attributes**: Borrower information, loan details, credit history, and repayment status
- **Loan outcome categories**: Fully Paid, Current, Charged-Off, and Rejected

Exploratory Data Analysis (EDA)

- Visualize distributions, correlations, and key statistics
- Identify patterns related to loan defaults
- Explore loan amounts, credit grades, employment details, etc.

Key Factors Influencing Loan Defaults

- Conduct detailed analysis of consumer and loan attributes
- Uncover strong indicators of default
- Discuss findings from EDA

Mitigating Credit Loss

- Strategies to minimize impact of defaults
- Optimizing loans to high-risk applicants
- Enhancing overall lending portfolio performance

Python Code for Loan Default Prediction

- Handling missing values
- Outliers
- Univariant Analysis
- Bivariant Analysis

Conclusion

- Summary of insights gained
- Recommendations for risk management

Technologies Used

- Python
- Python Libraries:
  - NumPy
  - Pandas
- Data Visualization
  - Matplotlib
  - Seaborn
- Editor:
  - Jupyter Notebooks
- Approach:
  - Exploratory Data Analysis (EDA)

Data Cleaning and Manipulation

1. Loading data from loan CSV
2. Checking for null values in the dataset
3. Checking for unique values
4. Checking for duplicated rows in data
5. Dropping records
7. Data Conversion
8. Outlier implementation

Data Dictionary of imp used columns

- **loan_amnt:** The listed amount of the loan applied for by the borrower.
- **annual_inc:** The self-reported annual income provided by the borrower during registration.
- **int_rate:** Interest Rate on the loan.
- **loan_status:** Current status of the loan.
- **purpose:** A category provided by the borrower for the loan request.
- **grade:** LC assigned loan grade.
- **home_ownership:** The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
- **dti:** A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

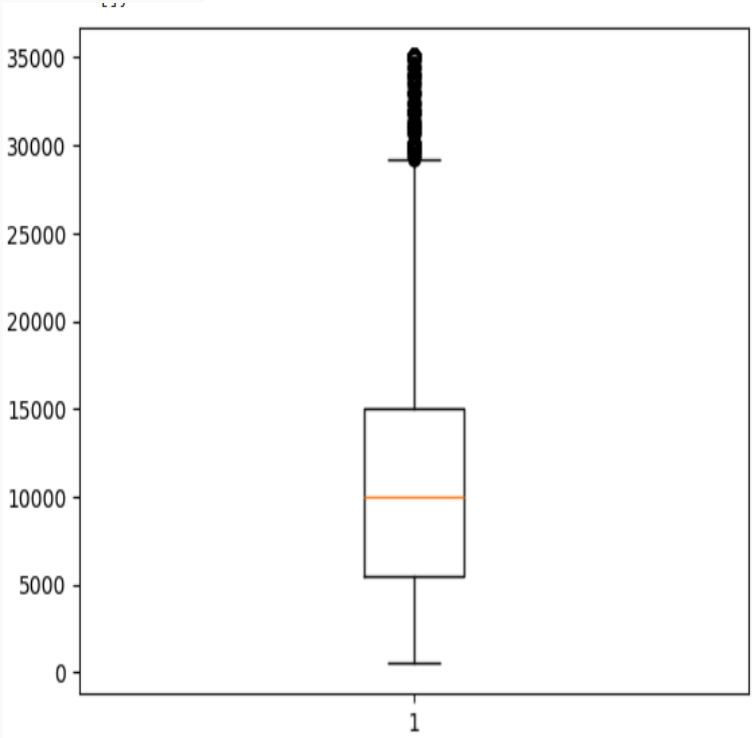
Import the relevant Python libraries

```
# Import the necessary Python libraries
import pandas as pdimport numpy as np
import matplotlib.pyplot as plt
import seaborn as sea
# Optional for better styling
import datetime as dt
import seaborn as sns
import warningswarnings.filterwarnings('ignore')
```

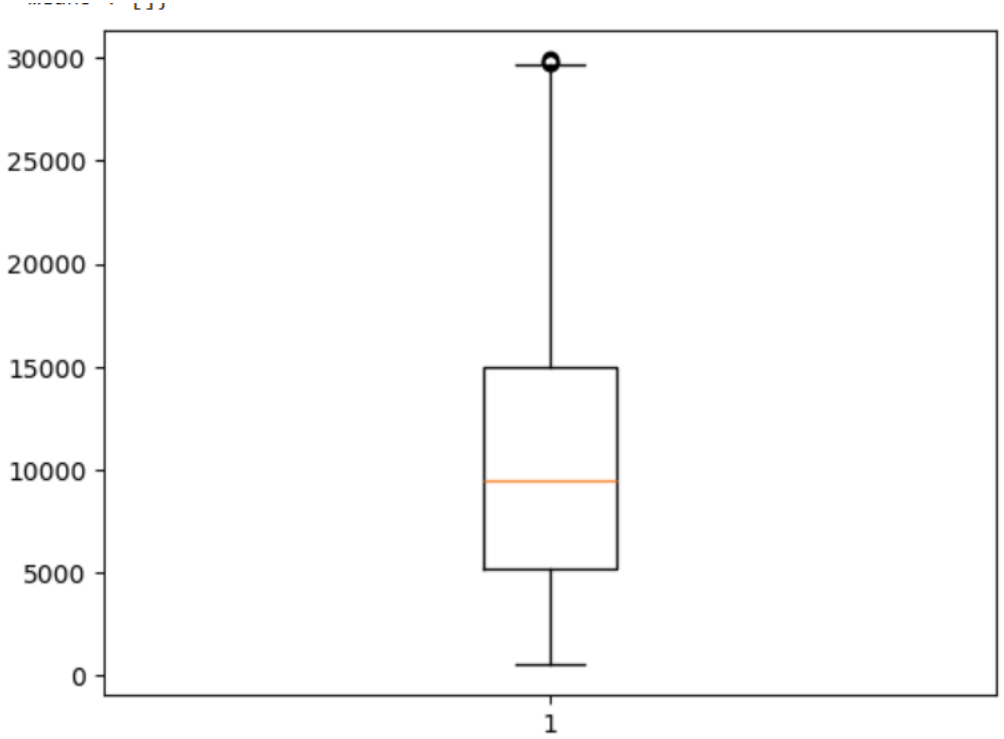
outliers

 # Outliers for loan_amt

Before outliers removal



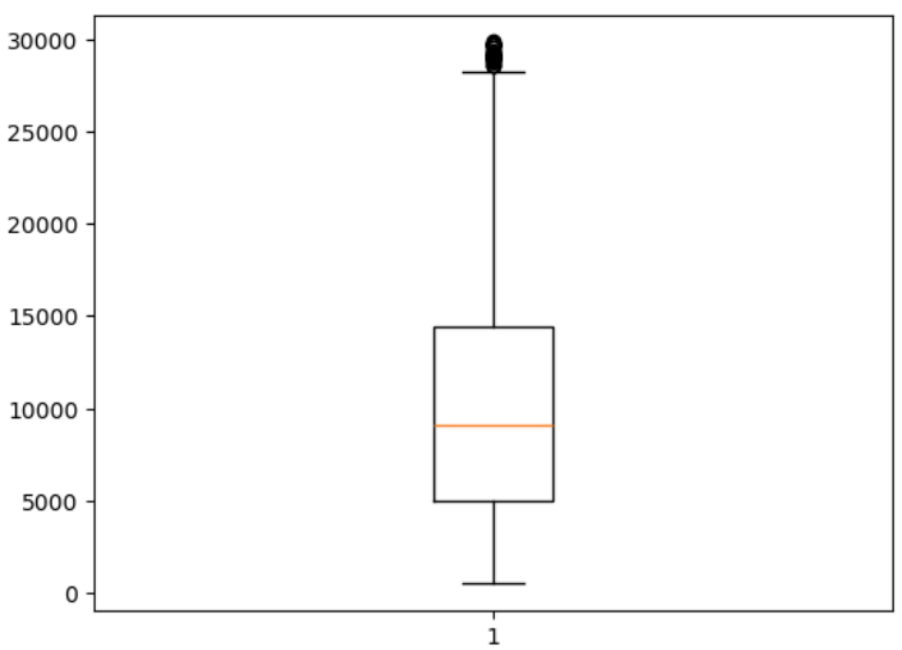
After outliers removal



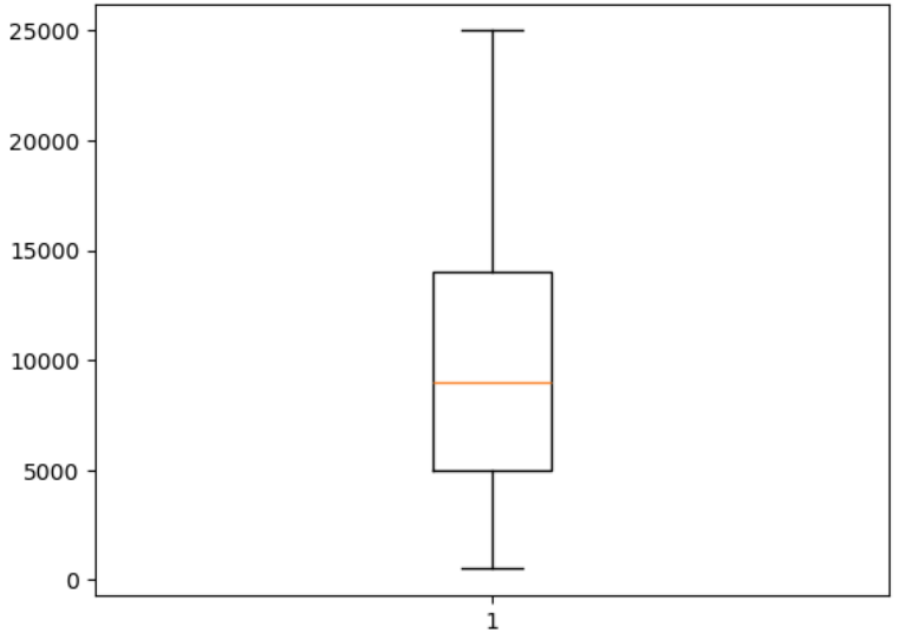
outliers

 # Outliers for funded_amnt

Before outliers' removal

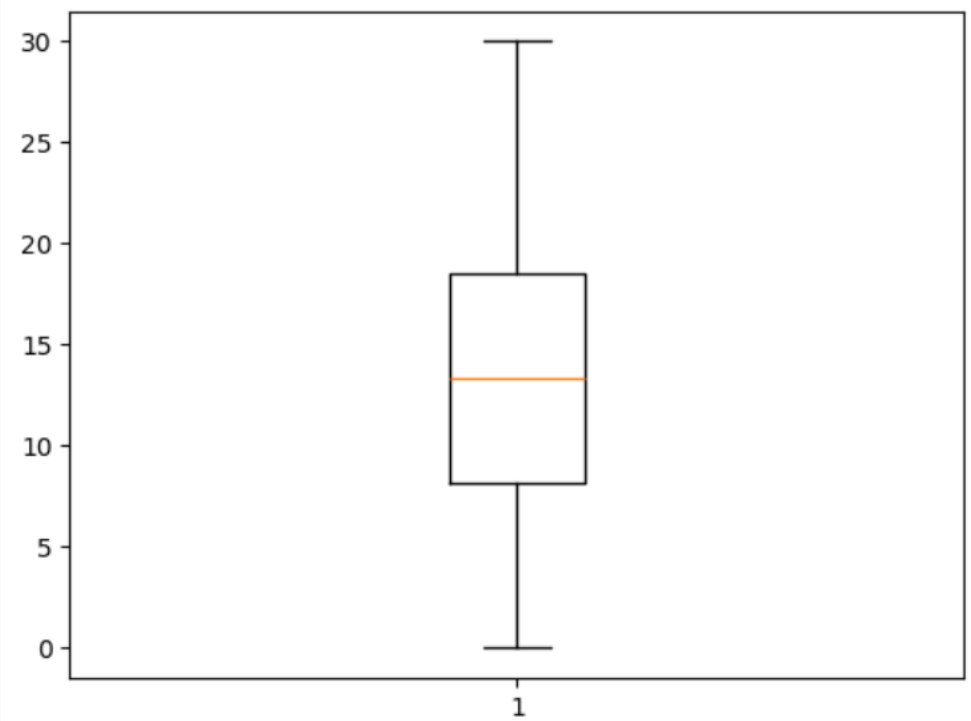


After outliers' removal



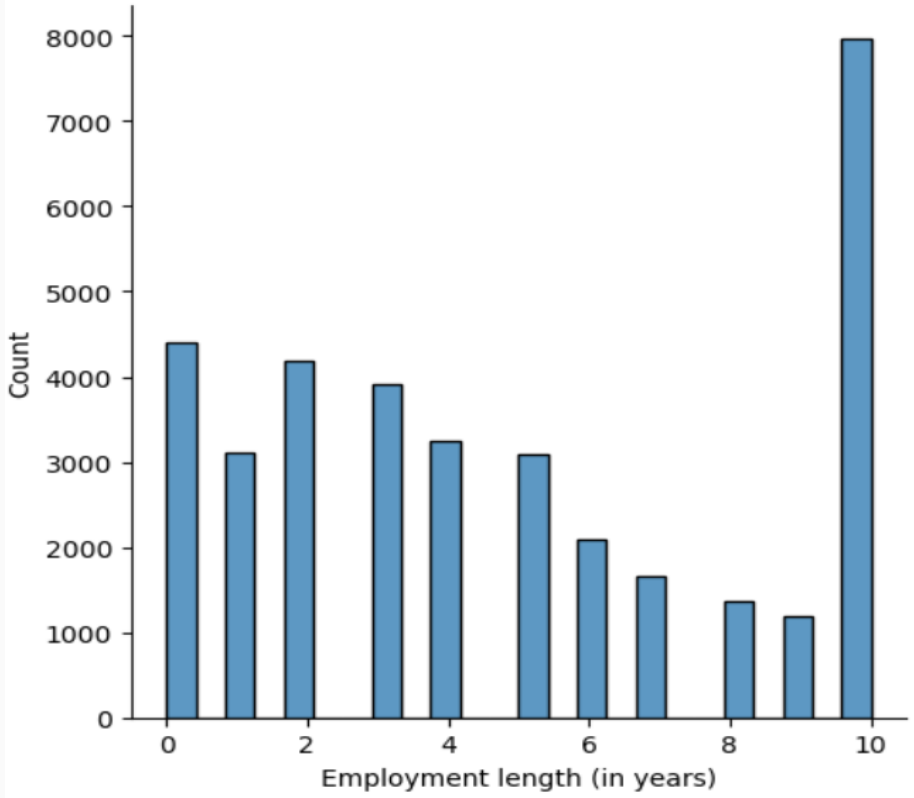
outliers

#outliers for **dti** which seems as expected no need to handle



Univariant Analysis

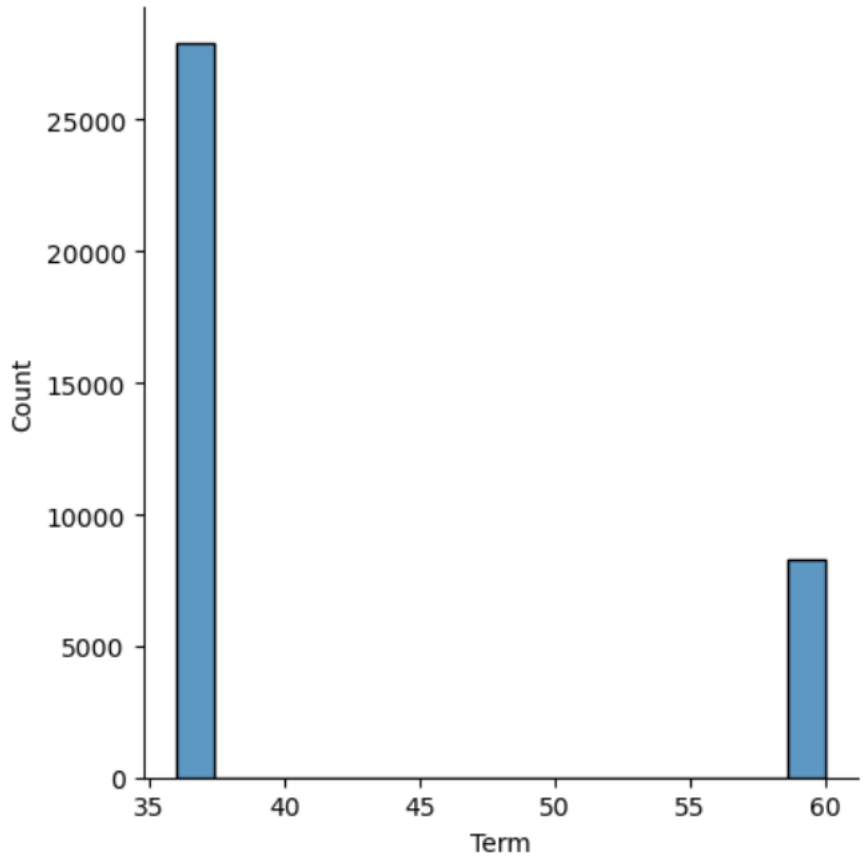
Univariate analysis looks at one feature at a time to summarize and find patterns in the data.



There are approximate 8000 employs who have taken the loan having tenure greater than 10 years

Univariant Analysis

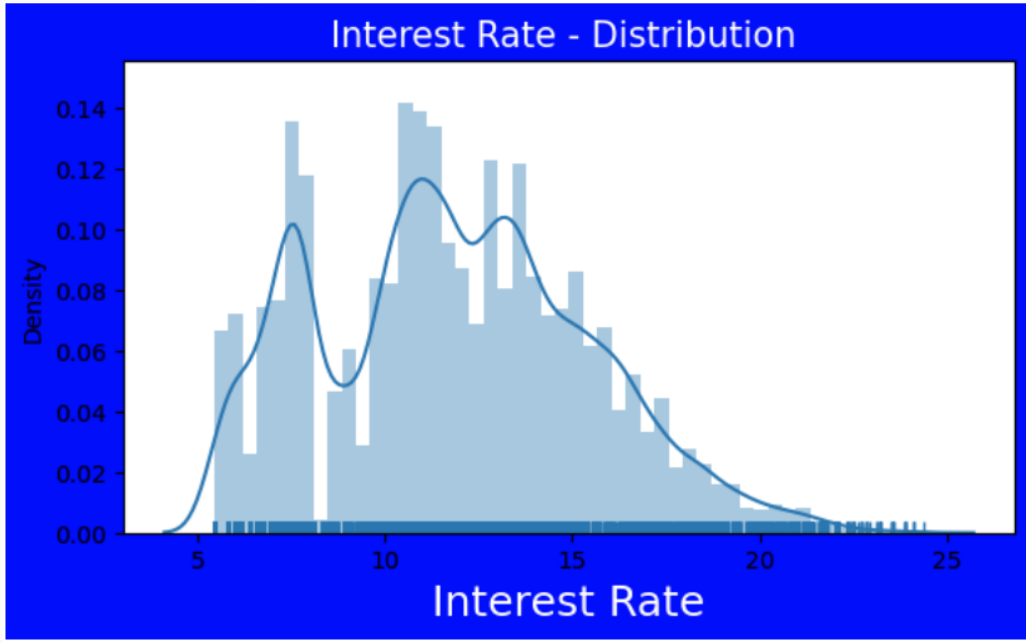
Univariate analysis looks at one feature at a time to summarize and find patterns in the data.



No of employess who taken the loan for Term of 36 months is greater than 60 months

Univariant Analysis

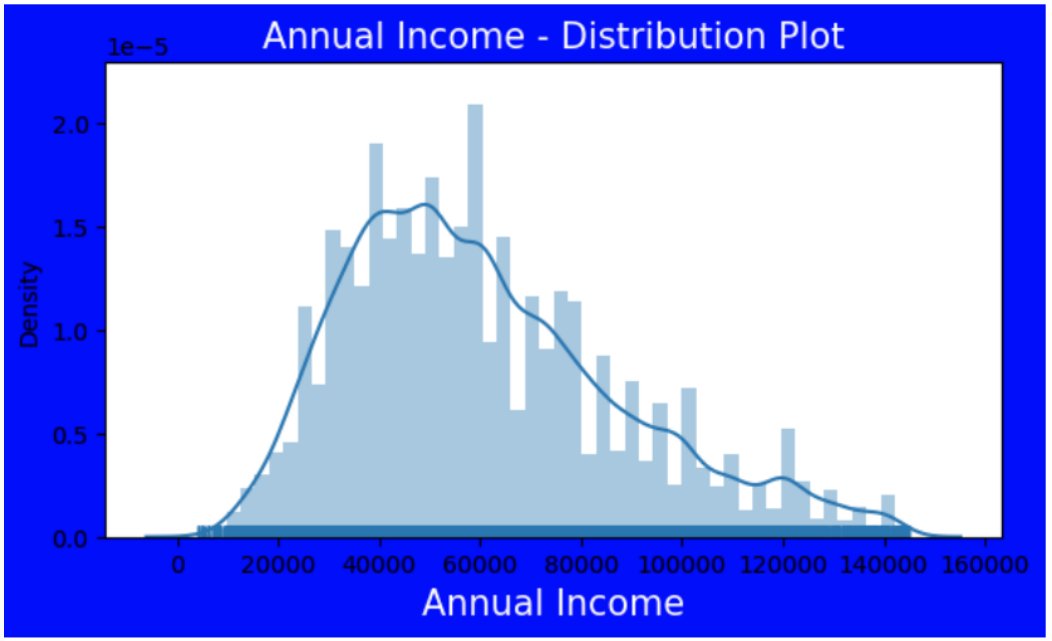
Univariate Analysis on Interest rate-



Maximum loans have interest rate between 9% - 14%.

Univariant Analysis

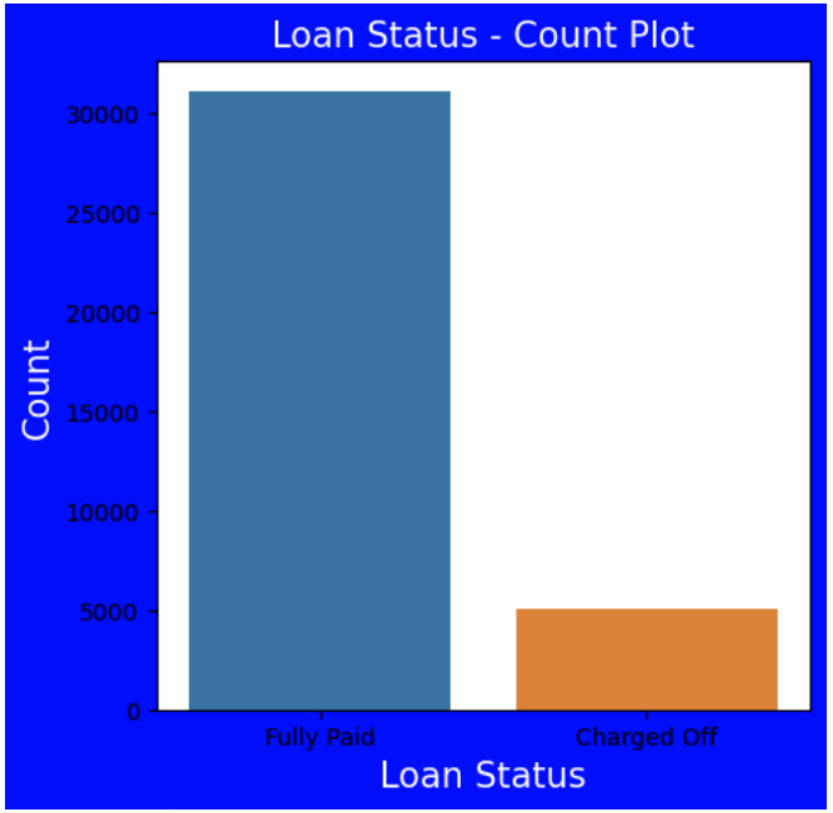
Annual Income distribution and box plot - Quantitative Variable Analysis



Maximum loans have been taken by the borrower having annual income of 40K - 80K.

Univariant Analysis

Loan Status Count Plot - Unordered Categorical Variable Analysis

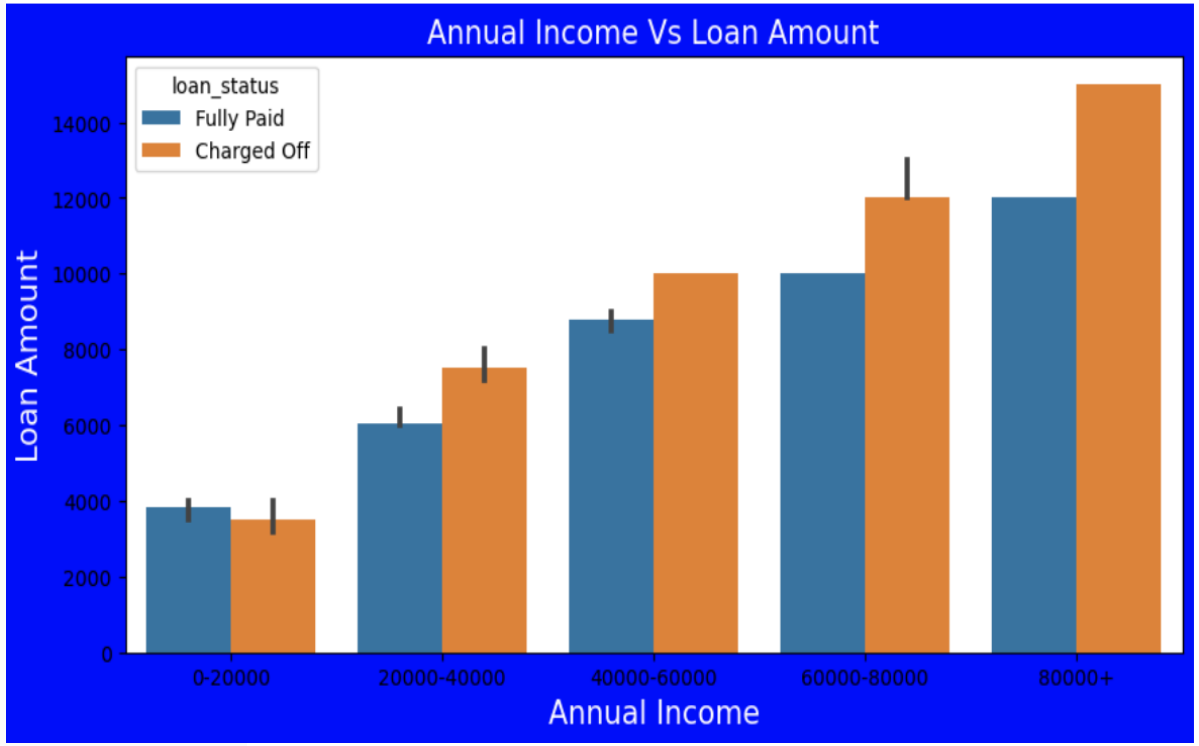


Approximately 15% of total loans are charged off.

Byvarient Analysis

Bivariate analysis explores the relationship between two different variables.

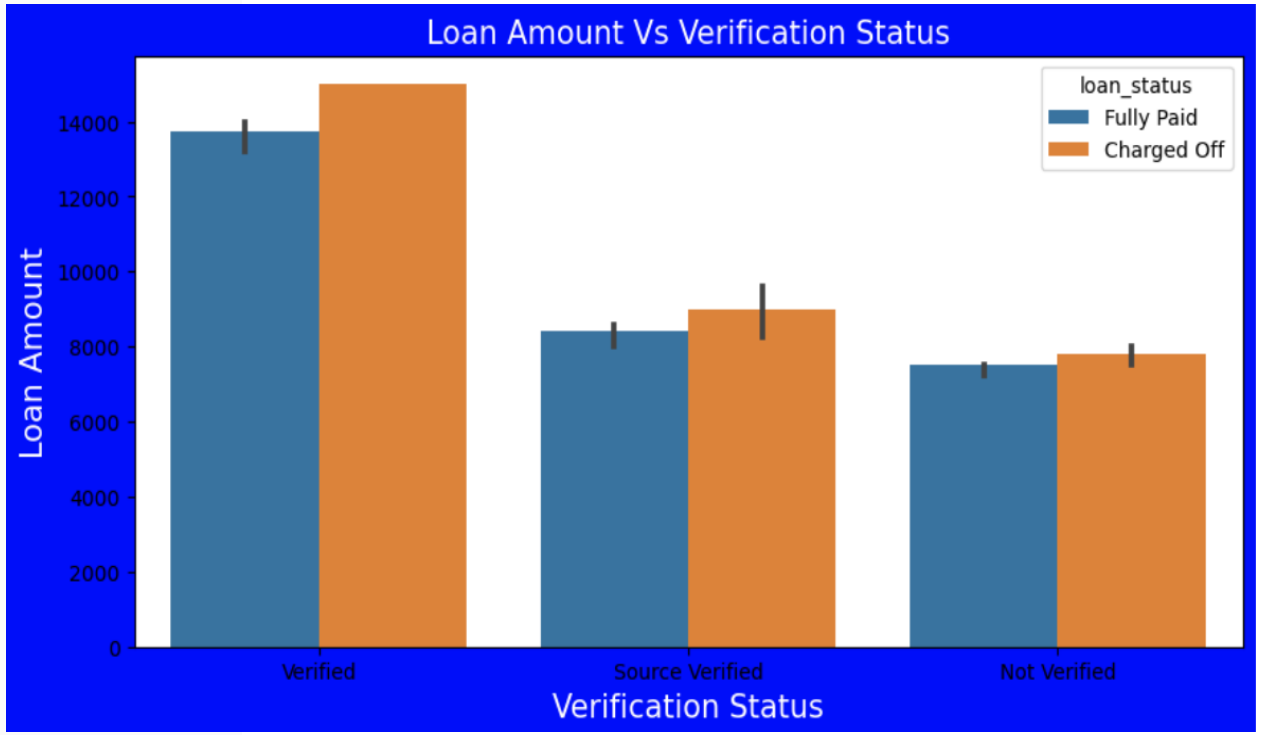
Annual Income Category and Loan Amount



This implies that people who borrow more money are more likely to default, which belong to the same income category

Byvarient Analysis

Loan Amount and Verification Status



Verified borrowers can get more money.
Chargeoff loans are bigger than fully paid ones.

Thank You