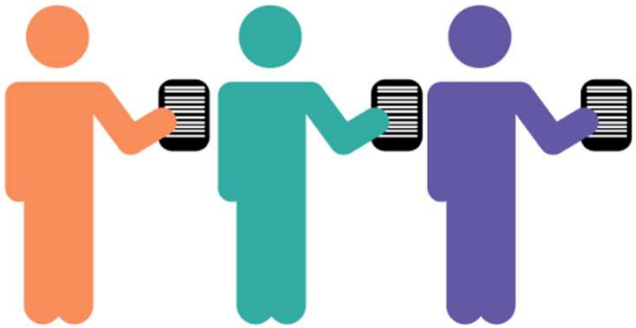
Presentation on Loan EDA case study

Narendhra Nadh
Neerja Mudgal

Purpose

Loan risk analysis will help the company which specializes in lending various types of loans to urban customers. When the company receives a loan application, make a decision for loan approval based on the applicant's profile. Which control loss of business to the company and avoid financial losses of the company

LOAN DATASET



Loan Accepted → Default

Loan Accepted → Non-Default

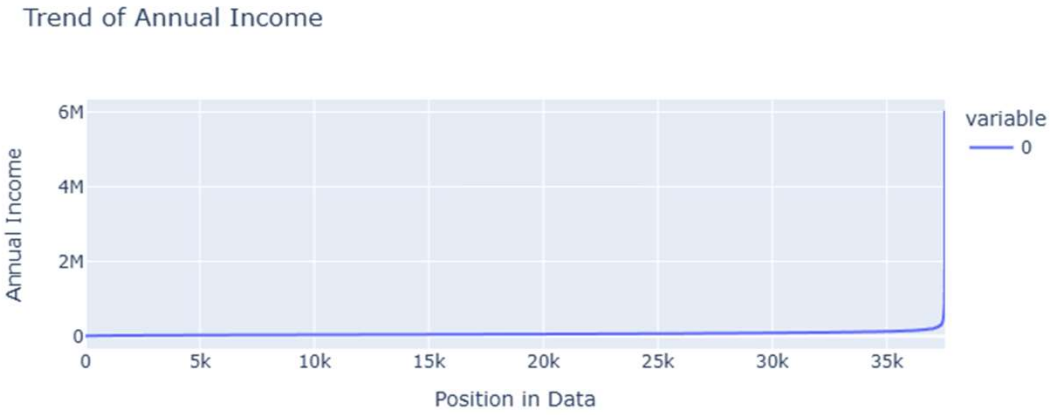
Loan Rejected

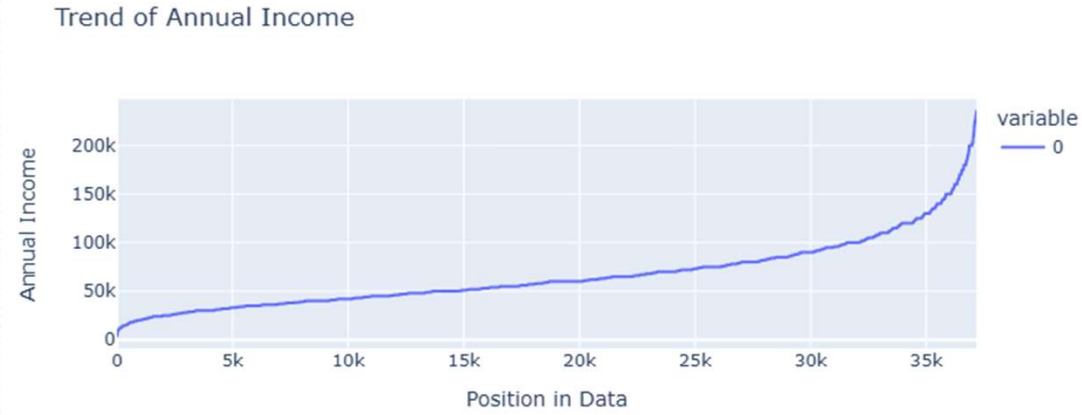
(Not considered in dataset)

Steps of analysis

1. Data Understanding and Sourcing
2. Data Wrangling
3. Check for Data quality issue
4. Data Visualization and analysis based on Univariate and Bivariate analysis and correlation
5. Conclusion

Data Analysis in continuous application data


Trend of Annual Income
6M
4M
2M
0
Annual Income
0
5k
10k
15k
20k
25k
30k
35k
Position in Data
variable
0


Trend of Annual Income
200k
150k
100k
50k
0
Annual Income
0
5k
10k
15k
20k
25k
30k
35k
Position in Data
variable
0

Univariate Analysis

Univariate Analysis is a type of data visualization where we visualize only a single variable at a time. Univariate Analysis helps us to analyze the distribution of the variable present in the data so that we can perform further analysis.

Bivariate analysis

Bivariate analysis is the simultaneous analysis of two variables. It explores the concept of the relationship between two variable whether there exists an association and the strength of this association or whether there are differences between two variables and the significance of these differences.

The main three types we will see here are:

Categorical v/s Numerical

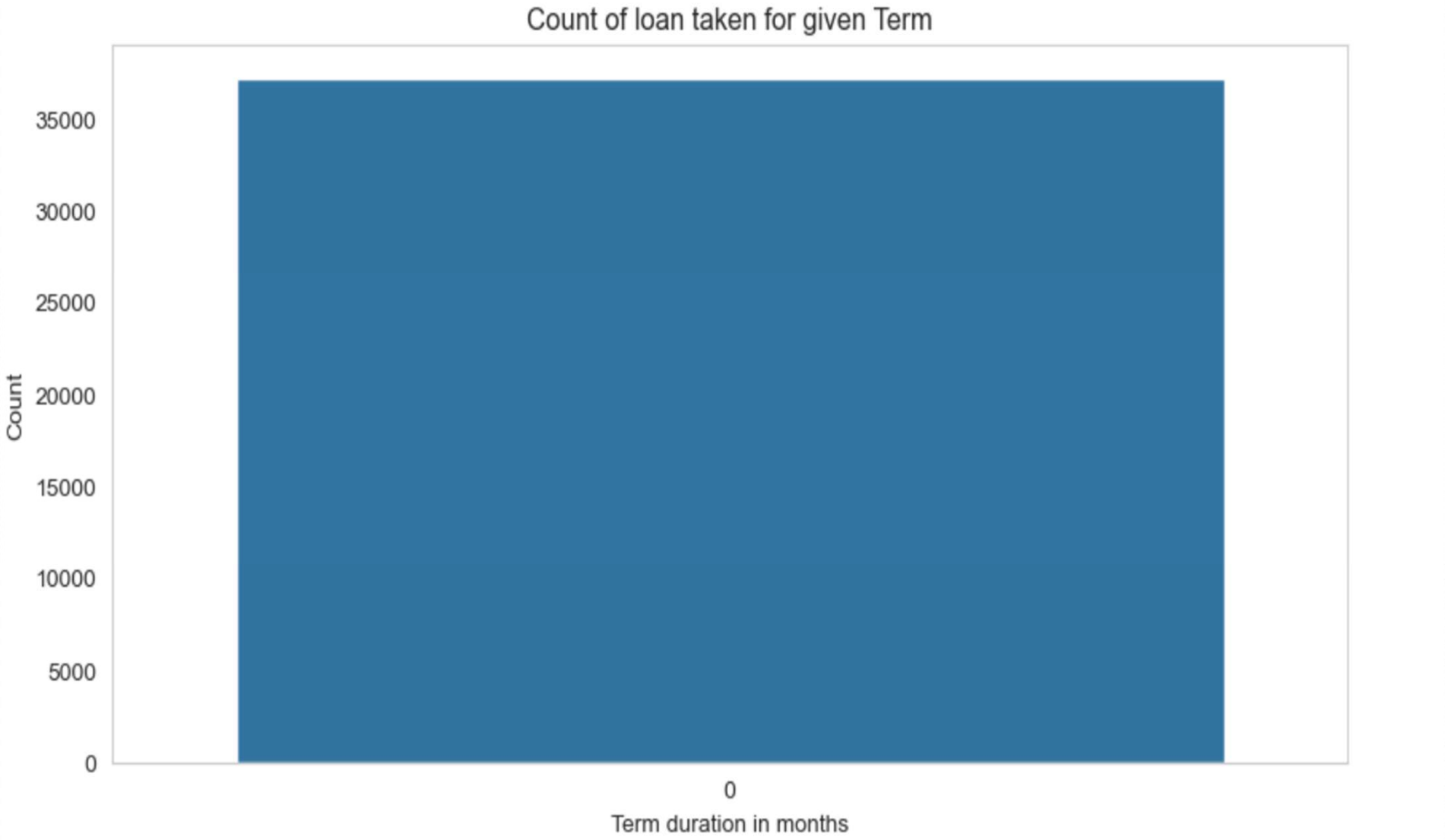
Numerical V/s Numerical

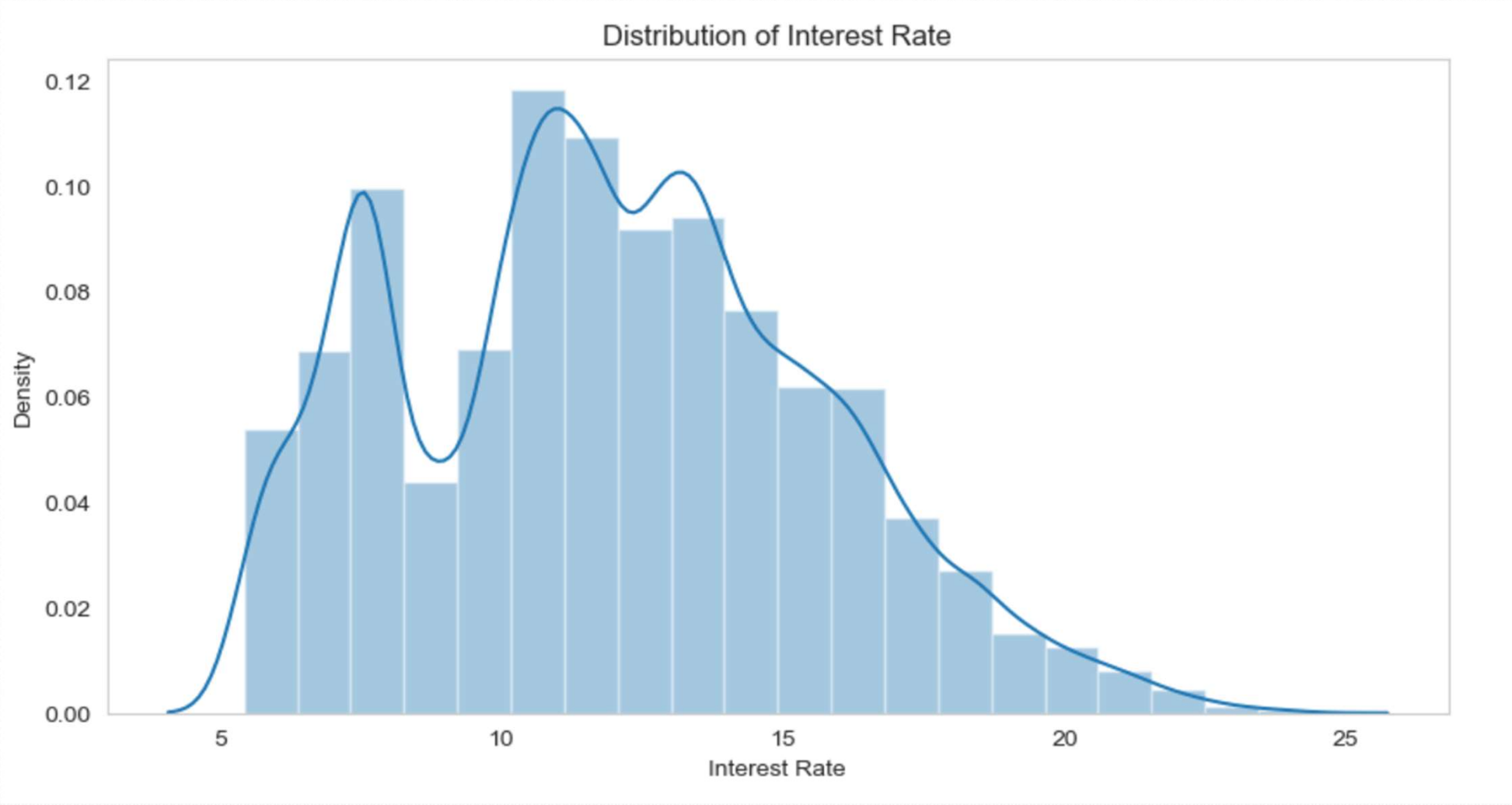
Categorical V/s Categorical data

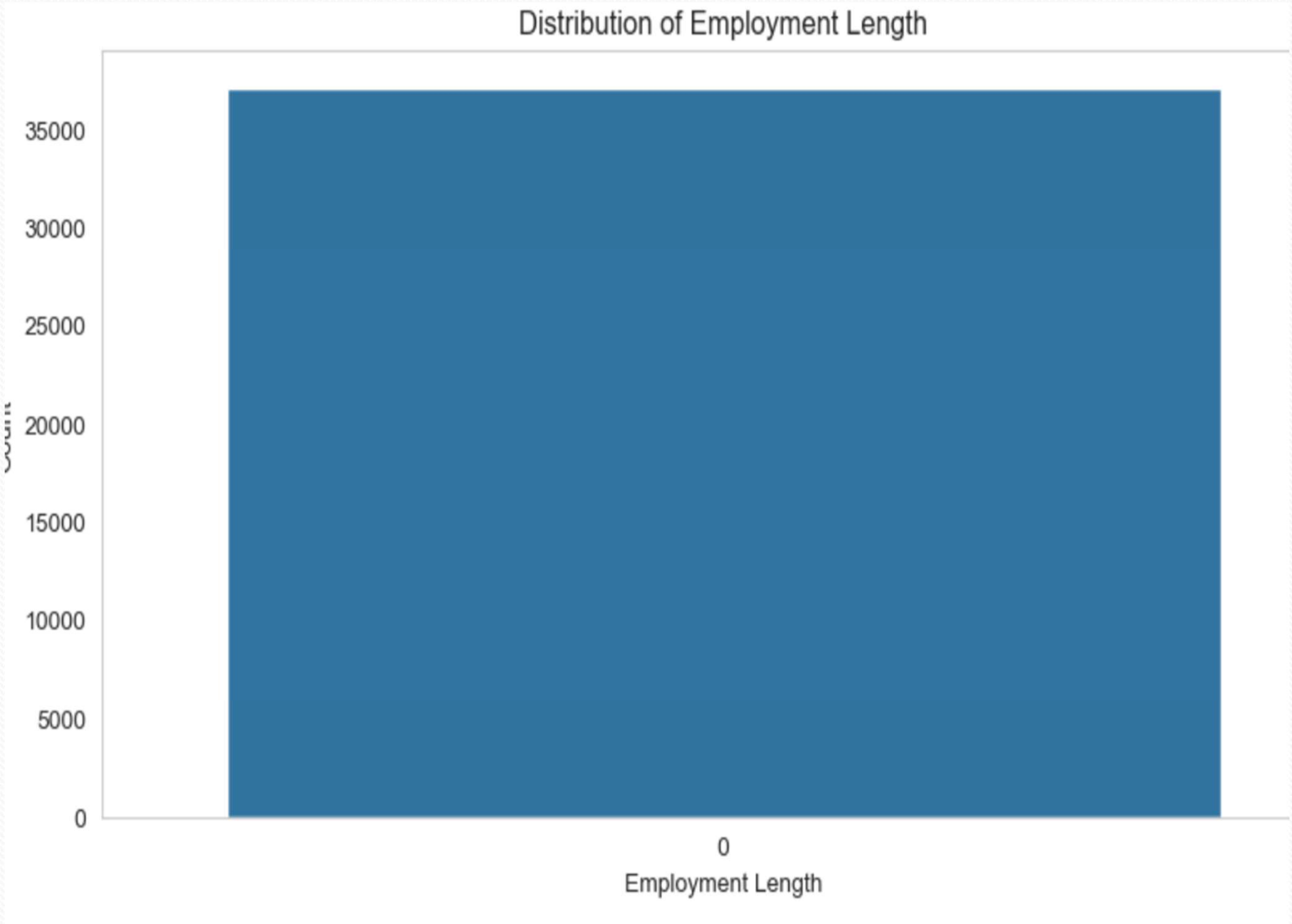
Segmented univariate

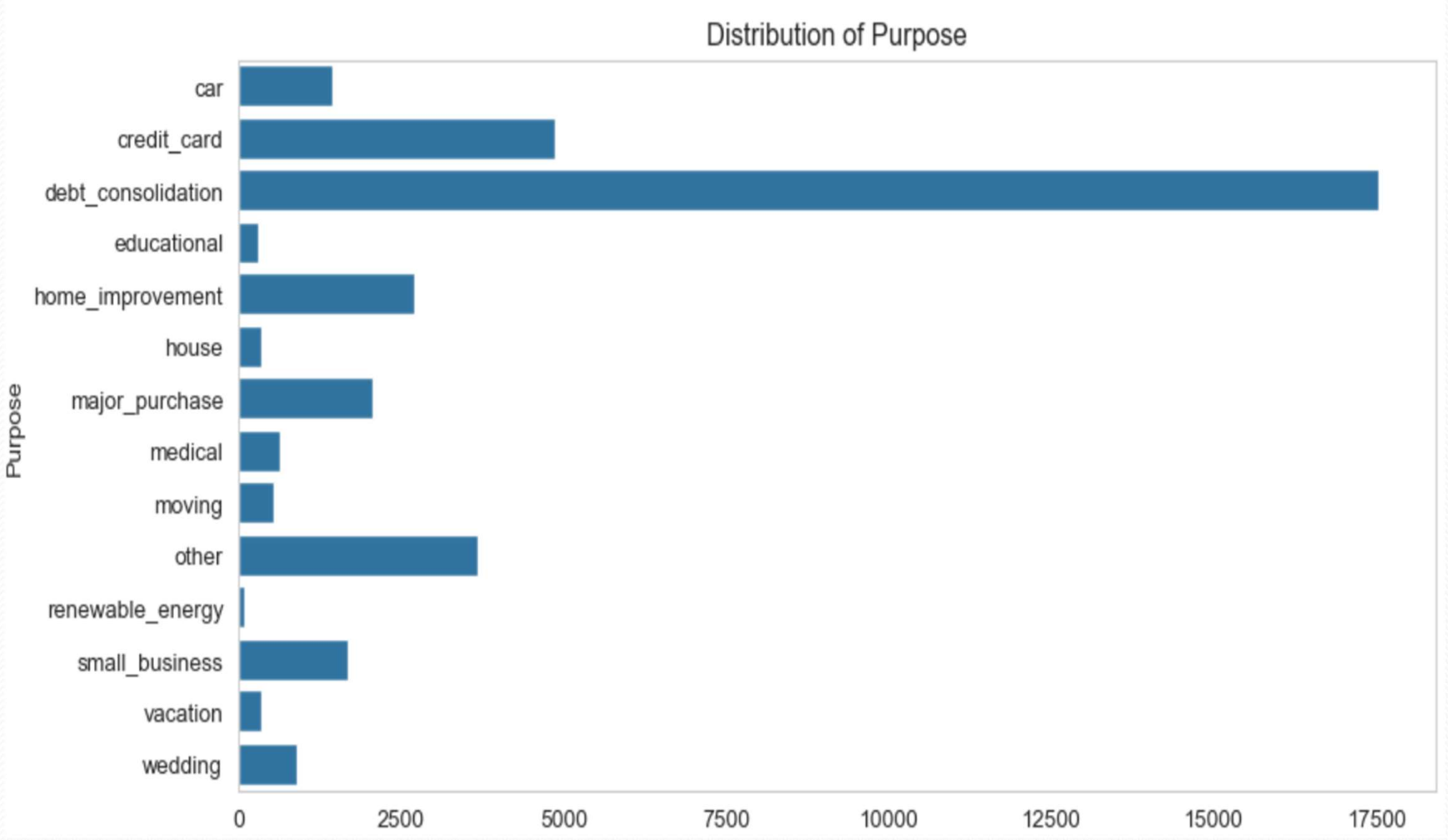
Analysis can help to identify differences and similarities within the data across different segments, and can provide insights into the factors that drive customer behavior or product performance.

Univariate Analysis

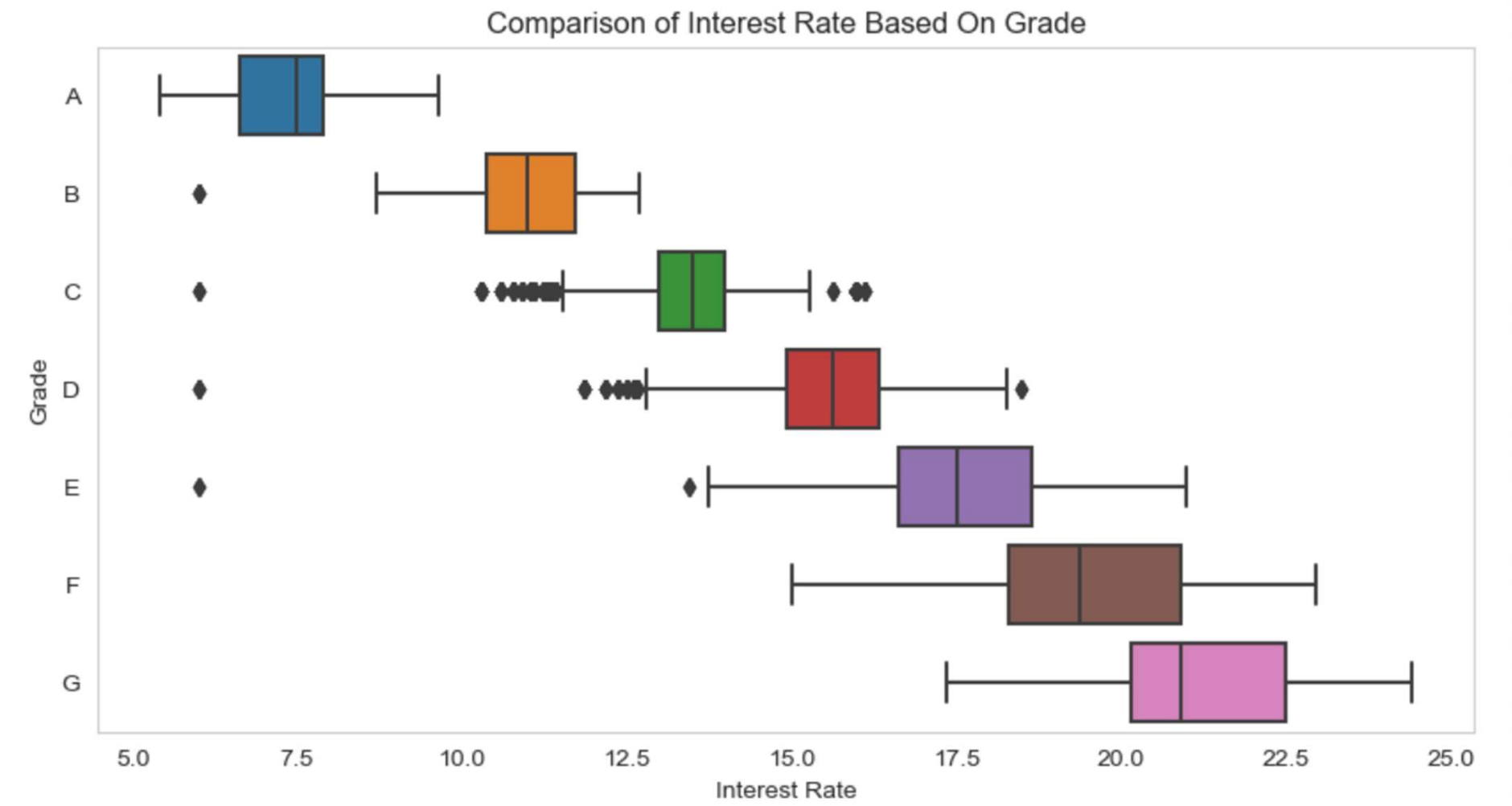


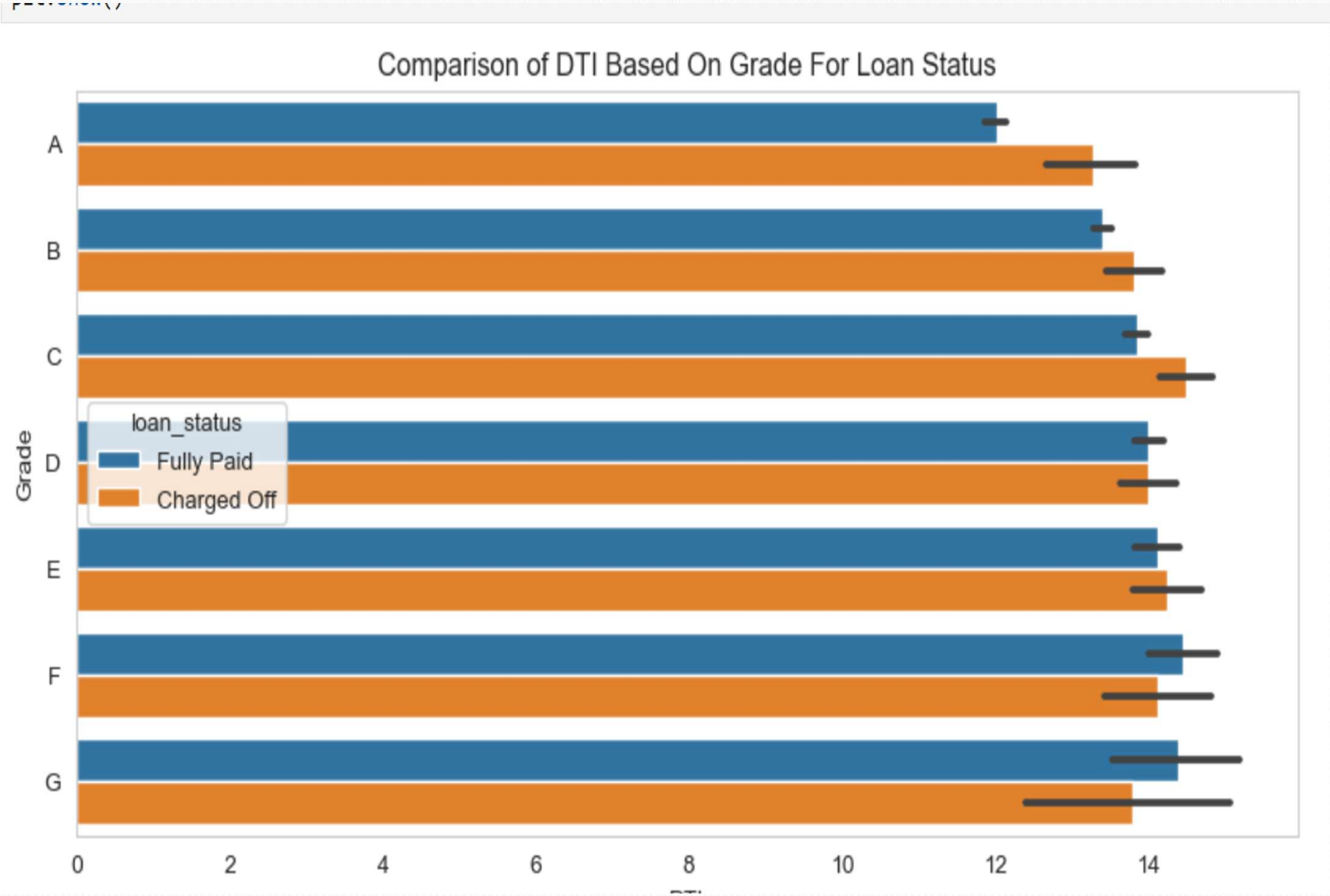

Distribution of Interest Rate
Density
0.12
0.10
0.08
0.06
0.04
0.02
0.00
5
10
15
20
25
Interest Rate


Distribution of Employment Length
35000
30000
25000
20000
15000
10000
5000
0
0
Employment Length

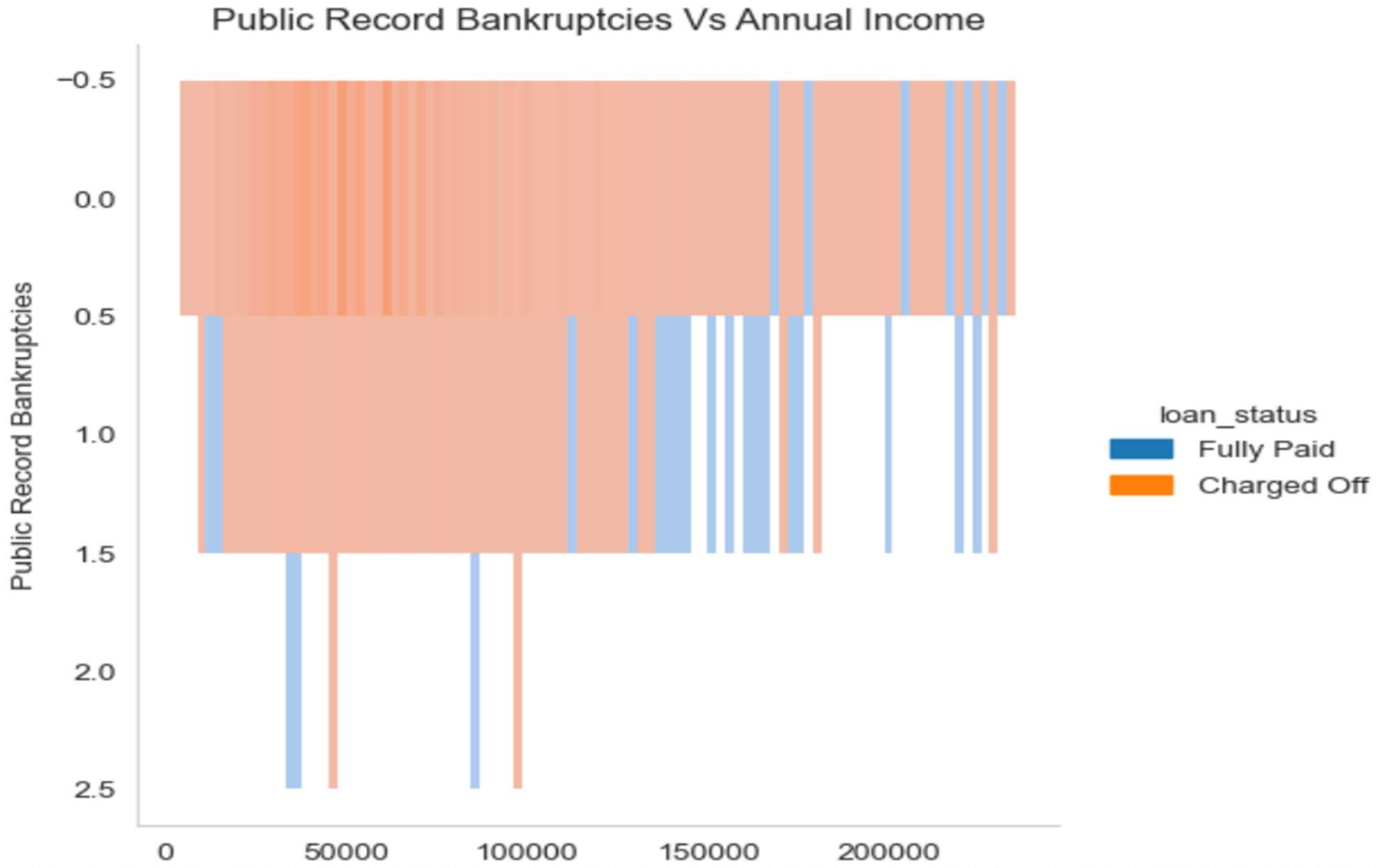

Distribution of Purpose
Purpose
car
credit_card
debt_consolidation
educational
home_improvement
house
major_purchase
medical
moving
other
renewable_energy
small_business
vacation
wedding
0
2500
5000
7500
10000
12500
15000
17500

Bivariate Analysis

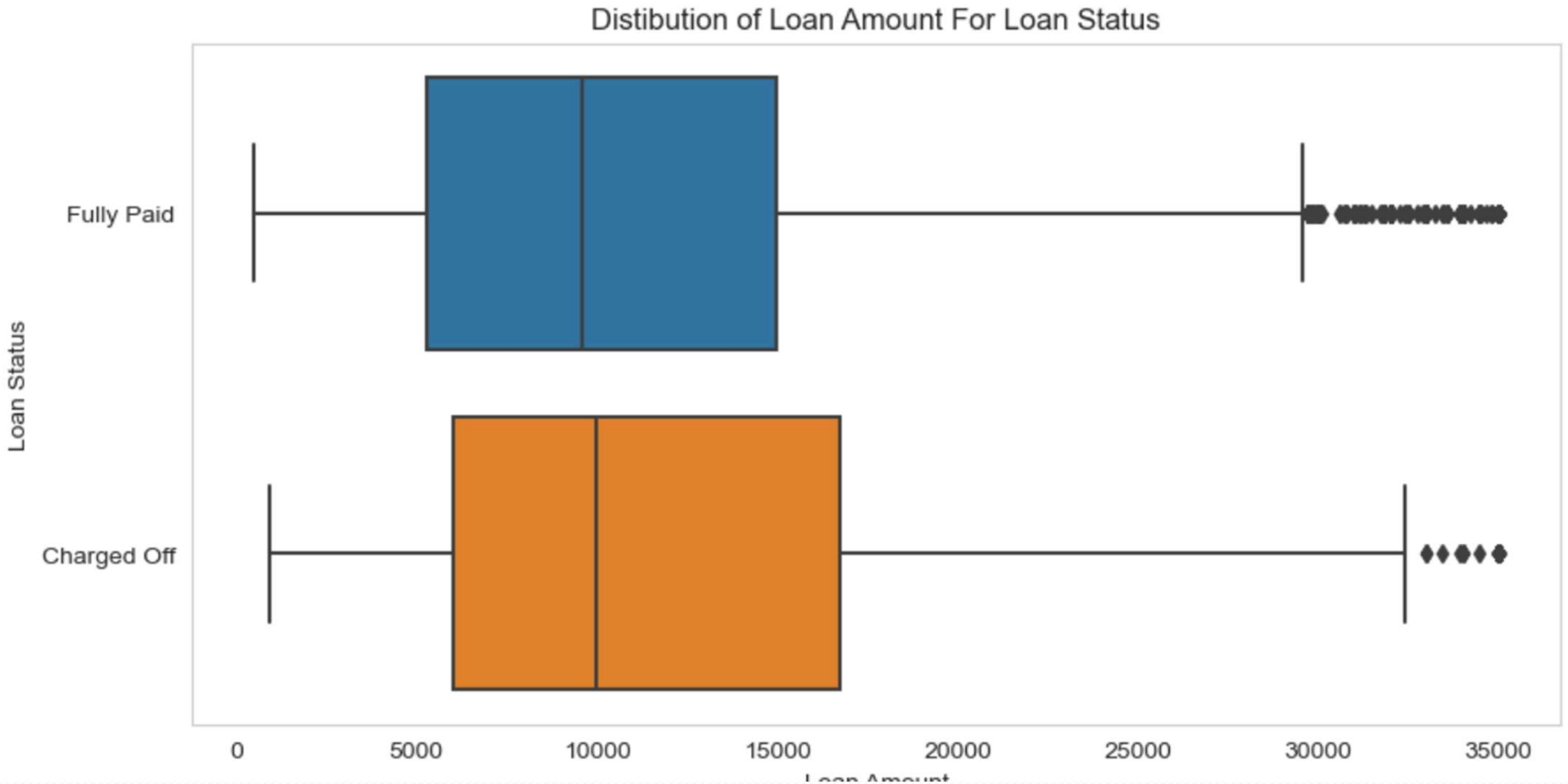


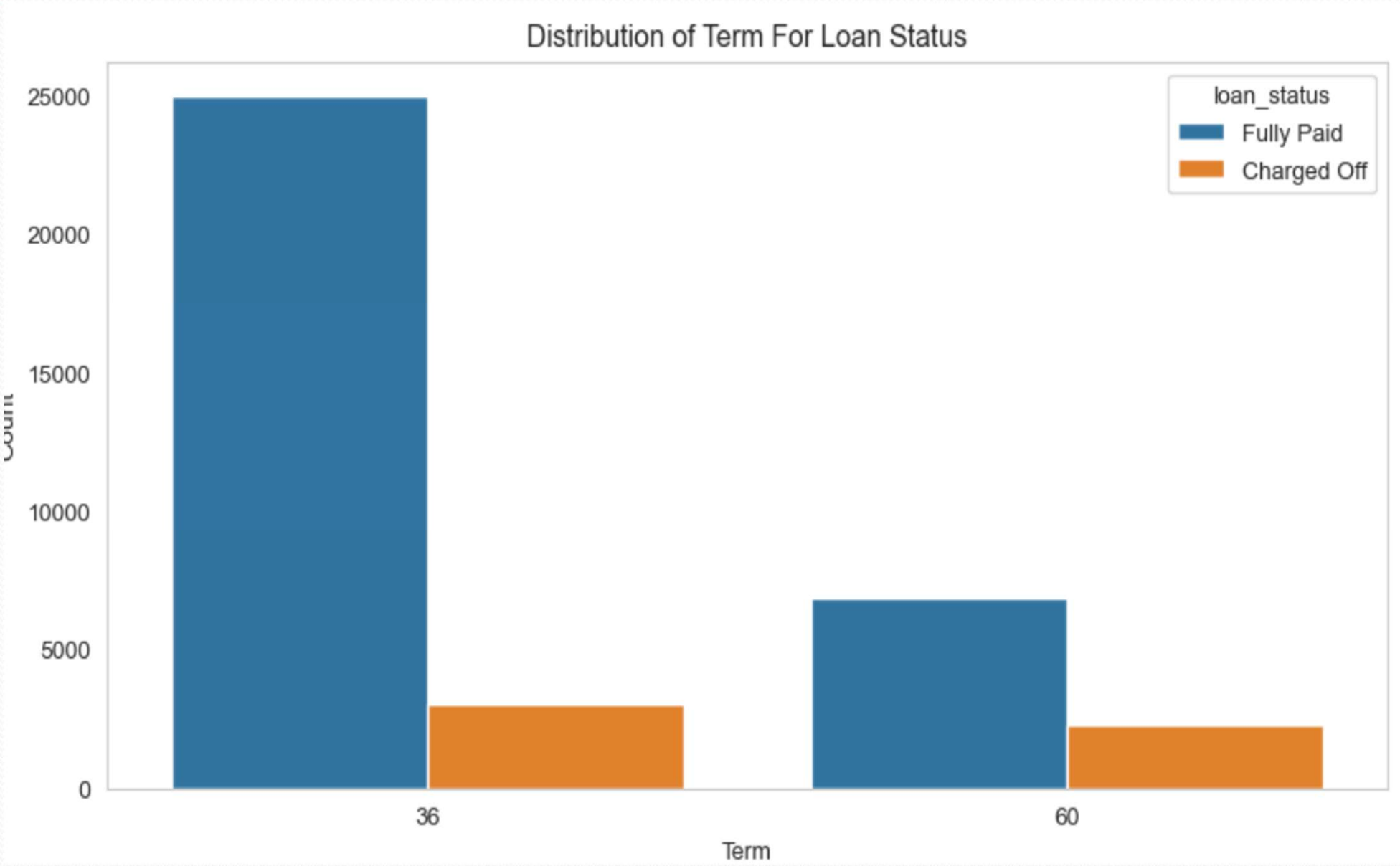

Comparison of DTI Based On Grade For Loan Status
Grade
A
B
C
D
E
F
G
loan_status
Fully Paid
Charged Off
0
2
4
6
8
10
12
14

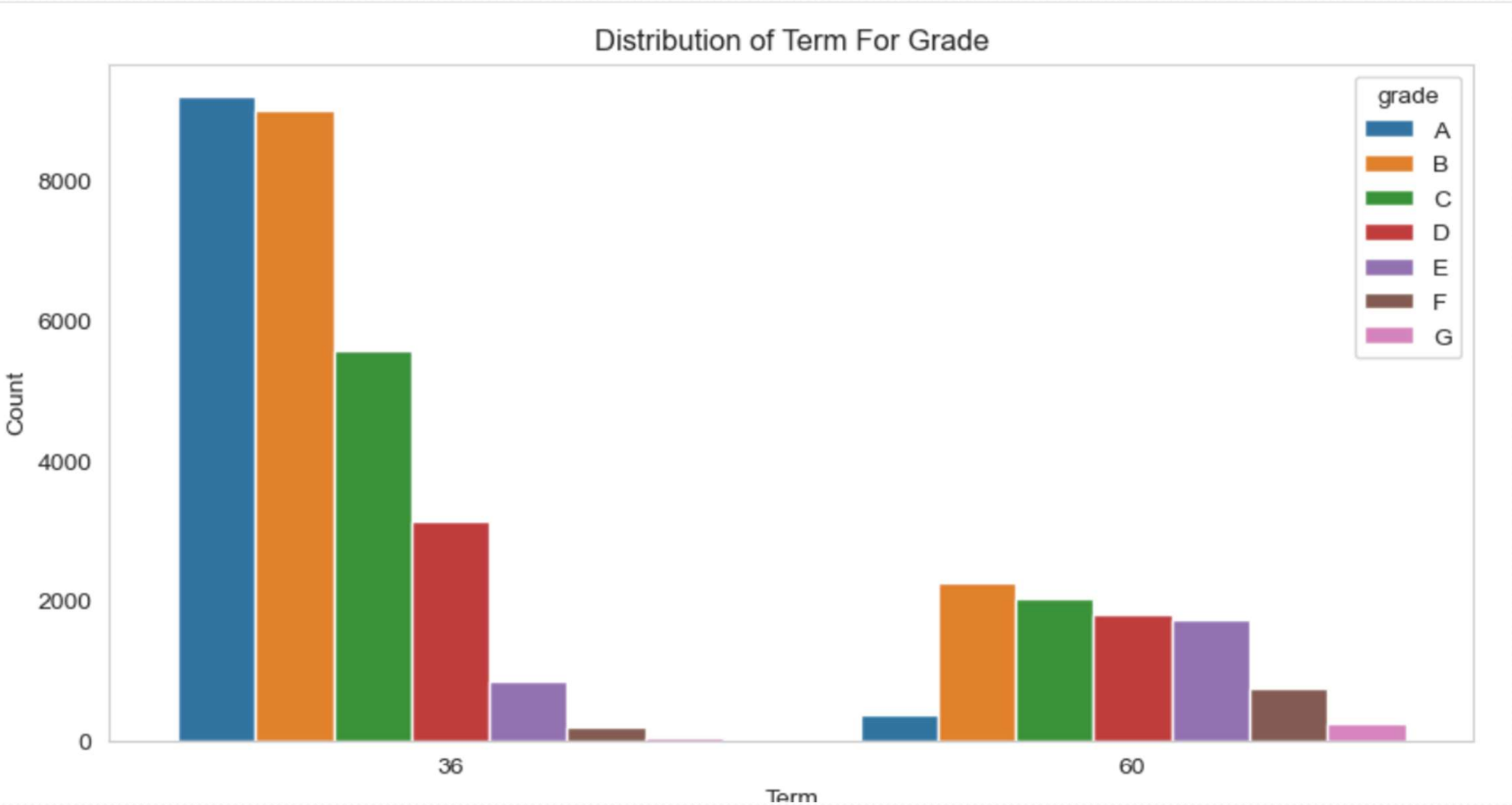
<Figure size 1000x500 with 0 Axes>



Segment Analysis




Distribution of Term For Loan Status
loan_status
Fully Paid
Charged Off
25000
20000
15000
10000
5000
0
Count
36
60
Term


Distribution of Term For Grade
grade
A
B
C
D
E
F
G
8000
6000
4000
2000
0
Count
36
60
Term

Case Summary

- All the below variables were established in analysis of Application dataframe as leading to default. Checked these against the Approved loans which have defaults, and it proves to be correct.
- The data given below contains information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.