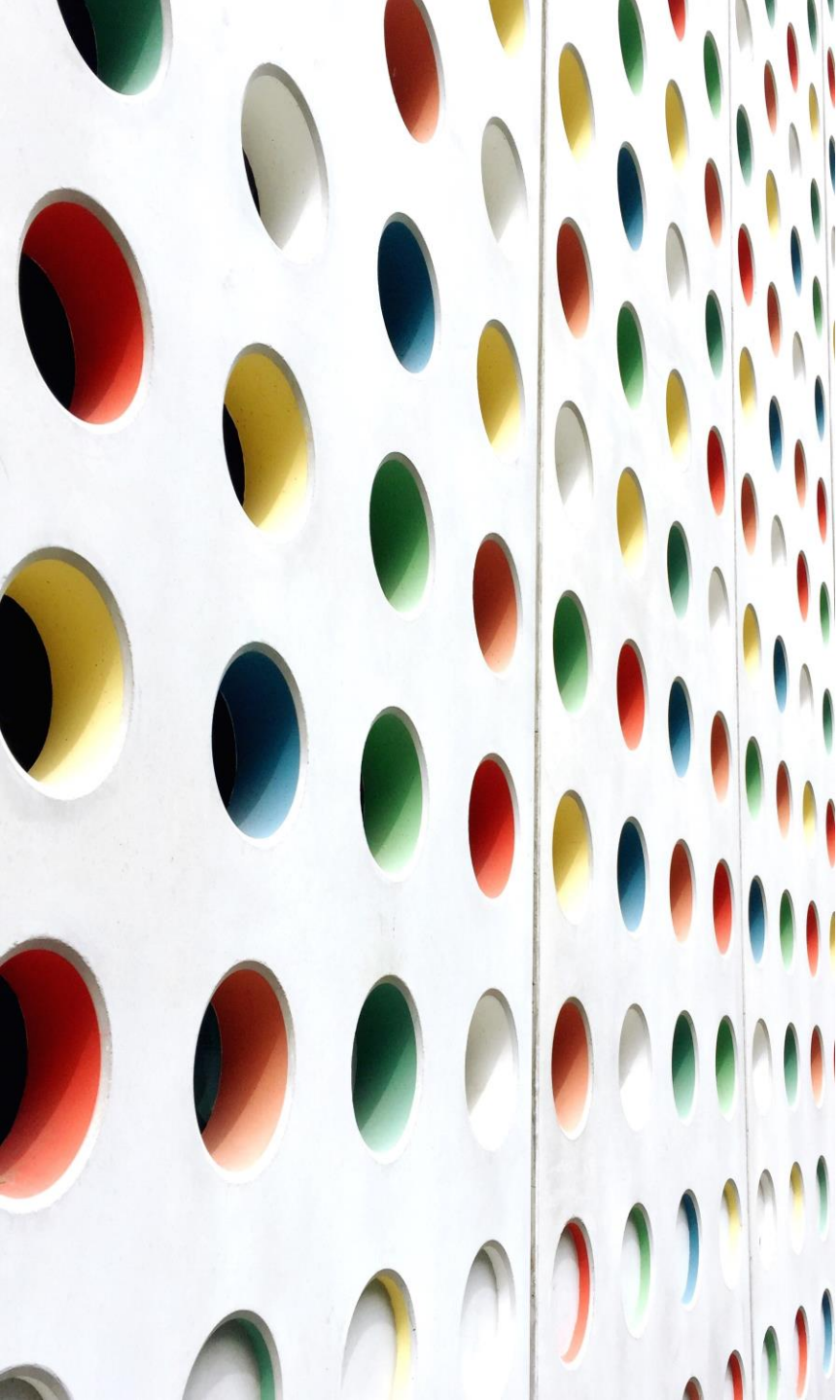


Lending Club Case Study

BANSIDHAR MOHAPATRA

Introduction to the Lending Club Case Study Project

In the world of finance, the ability to transform raw data into actionable insights is not just a competitive advantage—it's a game-changer. As we dive into this case study on Lending Club, we uncover how innovative data analysis is reshaping lending practices, driving financial inclusion, and redefining the future of personal finance."

Analysis - Abstract

The Objective of this case study is to implement EDA technique on a Banking System problem and understand the insights and present in a business first manner via presentation.

Techniques:-

- Data Cleaning
- Data Derivation and metadata generation.
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Various analysis reports

Data Cleaning

Find Metadata from the data given

```
metaDict={}
metaDict["NumberOfColumns"]=loan_data.shape[1]
metaDict["NumberOfRows"]=loan_data.shape[0]
metaDict["NumberOfMissingValues"]=loan_data.isnull().sum().sum()
metaDict["NumberOfUniqueValues"]=loan_data.nunique().sum()
metaDict["NumberOfDuplicates"]=loan_data.duplicated().sum()
```

```
metaDict
```

```
{'NumberOfColumns': 111,
 'NumberOfRows': 39717,
 'NumberOfMissingValues': 2263364,
 'NumberOfUniqueValues': 416801,
 'NumberOfDuplicates': 0}
```

Dropping Columns with only one values.

```
loan_data=loan_data.drop(['tax_liens', 'delinq_amnt', 'charg
```

Rechecking the Unique value Count

```
print(loan_data.nunique().sort_values(ascending=True))
```

```
term                      2
pub_rec_bankruptcies      3
loan_status               3
verification_status       3
pub_rec                   5
home_ownership            5
grade                     7
inq_last_6mths            9
delinq_2yrs              11
emp_length               11
purpose                  14
sub_grade                35
open_acc                 40
addr_state               50
issue_d                  55
```

Removing and fixing the null values

```
# Checking for missing values across the dataframe
print(loan_data.isnull().sum().sort_values(ascending=False)
```

```
desc                      12940
emp_length                 1075
pub_rec_bankruptcies        697
last_pymnt_d                71
revol_util                  50
last_credit_pull_d           2
out_prncp_inv                0
open_acc                     0
```

Removing duplicate rows from the dataframe

```
loan_data = loan_data.drop_duplicates()
print(loan_data.shape)
```

Removing columns which has more than 50% null values

```
loan_data=loan_data.drop(null_perc[ null_perc > 50 ].index, axis=1)
```

ReCheck the Null values Column percentage

```
null_perc = round(100*(loan_data.isnull().sum()/len(loan_data.index)), 2)

null_perc[ null_perc > 0 ]
```

```
emp_title                   6.19
emp_length                  2.71
desc                       32.58
title                       0.03
revol_util                  0.13
last_pymnt_d                0.18
last_credit_pull_d          0.01
collections_12_mths_ex_med  0.14
chargeoff_within_12_mths    0.14
pub_rec_bankruptcies        1.75
tax_liens                   0.10
dtype: float64
```

```
# Checking values in pub_rec_bankruptcies columns for feasibility of inserting null values
print(loan_data.pub_rec_bankruptcies.value_counts())
```

```
0.0    37339
1.0     1674
2.0        7
Name: pub_rec_bankruptcies, dtype: int64
```

The column 'emp_length', If we fix the null values it will impact the business information so we are removing it.

```
loan_data = loan_data.dropna(subset=['emp_length'])
print(loan_data.shape)
```

```
(38642, 42)
```

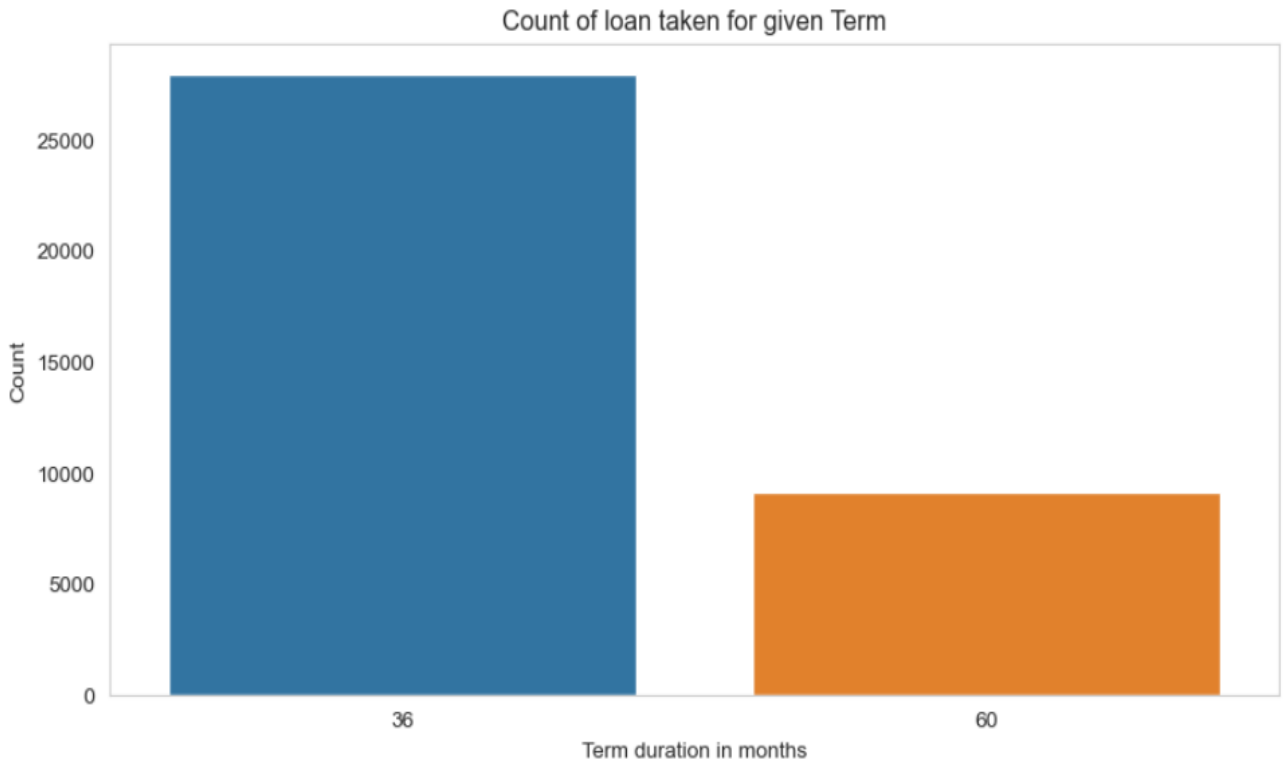
The column 'revol_util', 'last_pymnt_d', will have no impact on the analysis, if deleted

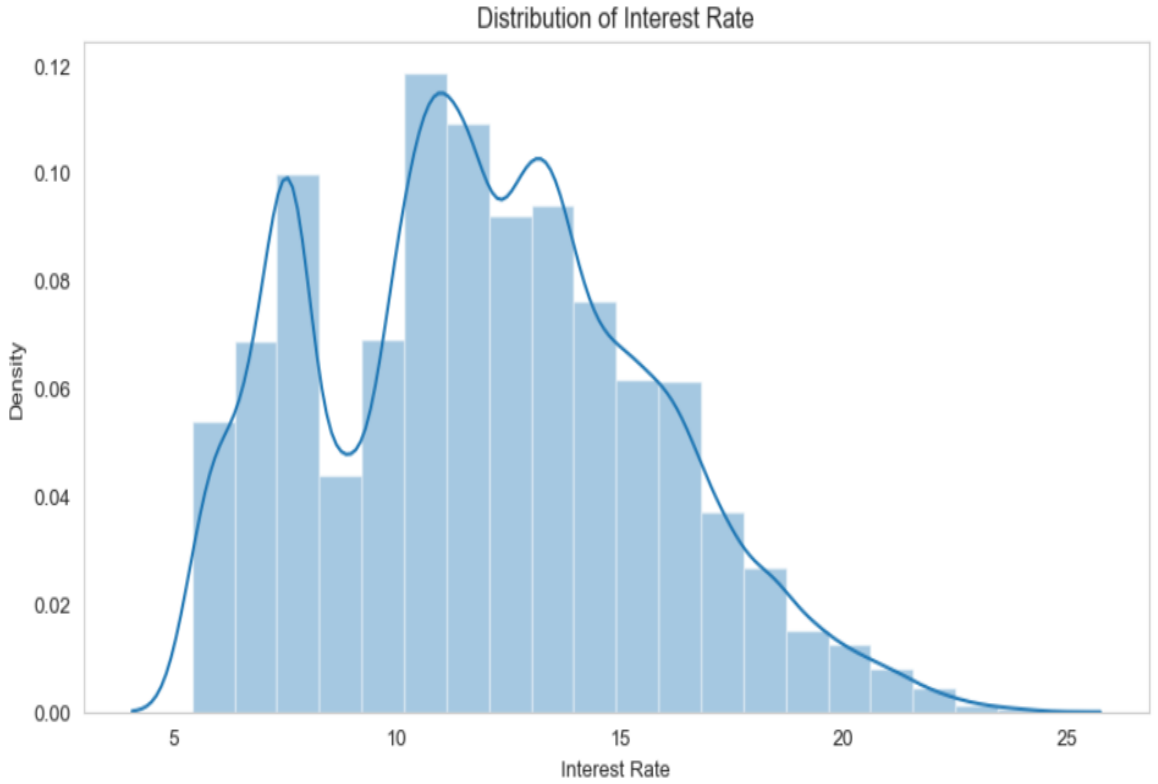
The column pub_rec_bankruptcies the data mostly has value 0 which can be subsitued in place of null as that

Term and Interest Rate

Loan Term: This analysis shows that the Loans taken for 36 month term are much more than 60 months and have lower chance of defaulting.

Interest Rate: Here the analysis shows the count of loan taken varies with interest rate showing peak around in 5-15 bracket and decreasing slowly where as the chance of defaulting increases with interest rate.

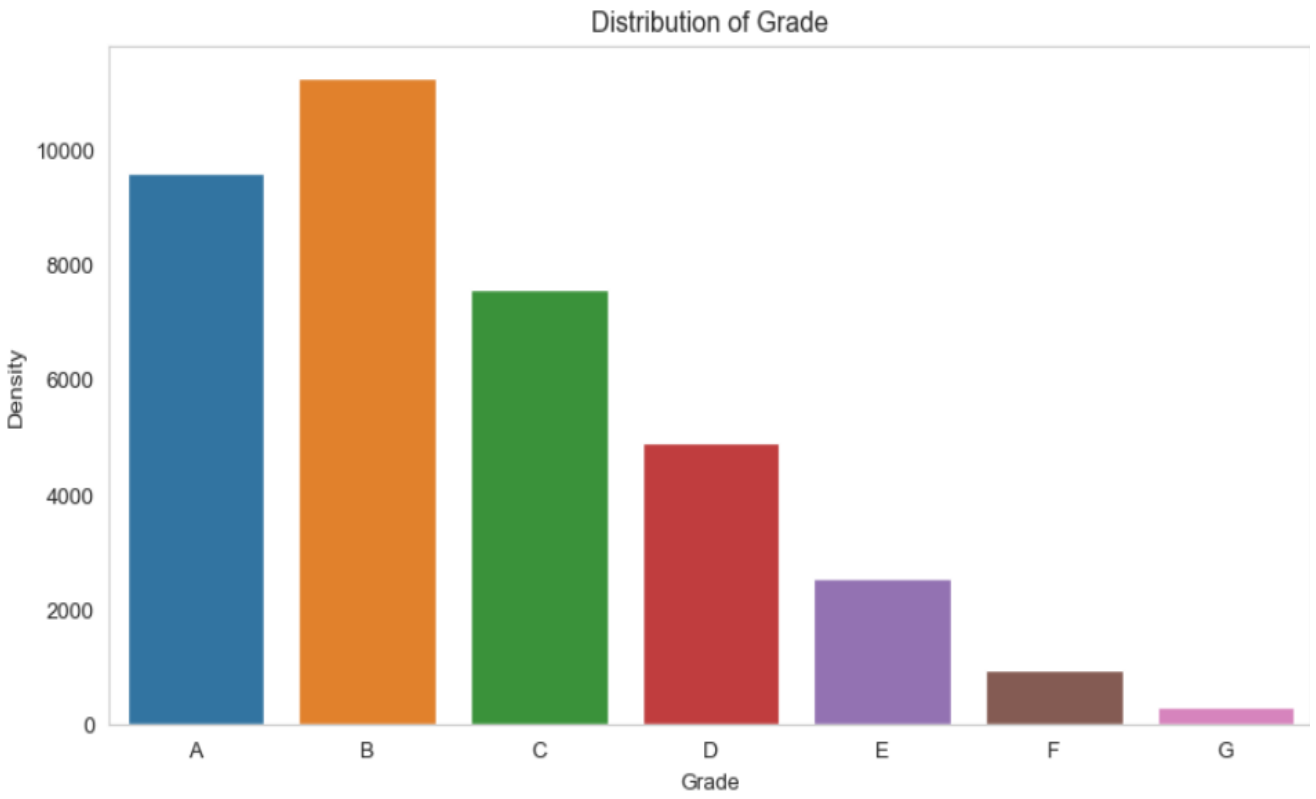


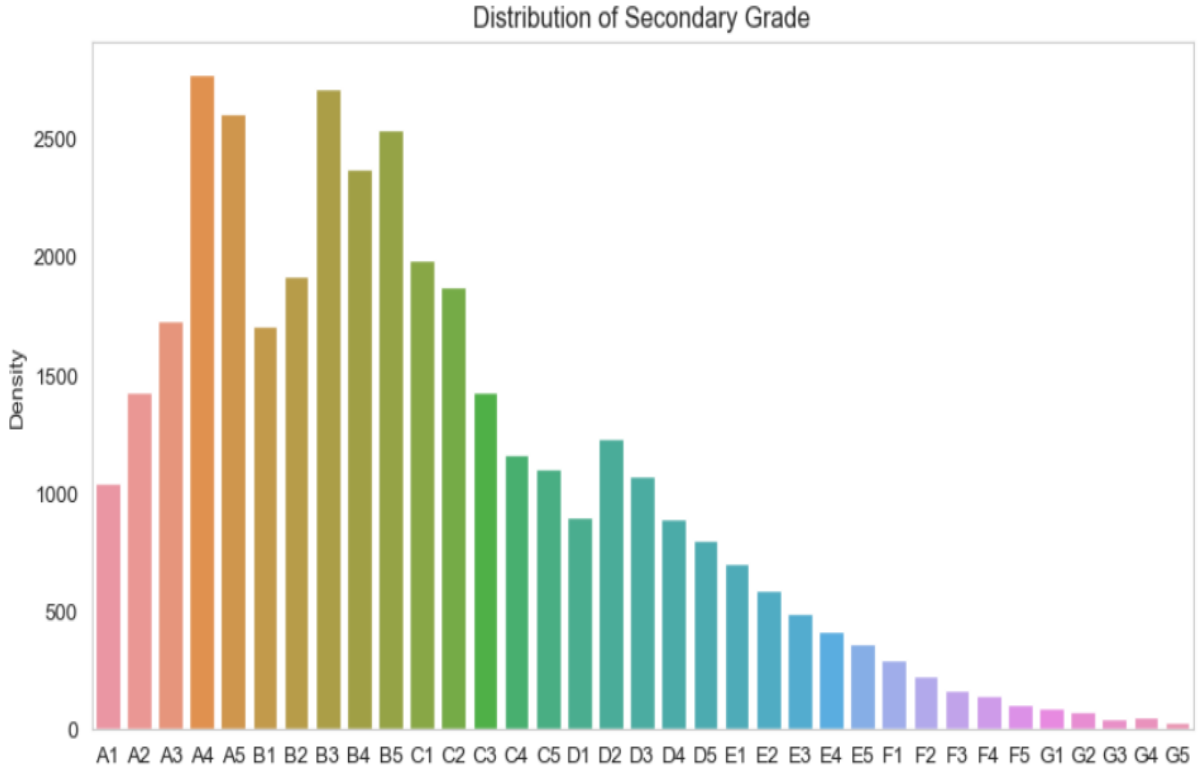


“Distribution of Grade” vs “Distribution of Secondary Grade”

Distribution of Grade: In this analysis I want to depicts that the loan approved are majorly of higher distribution of grade as they are of low risk soi it low chance of defaulting.

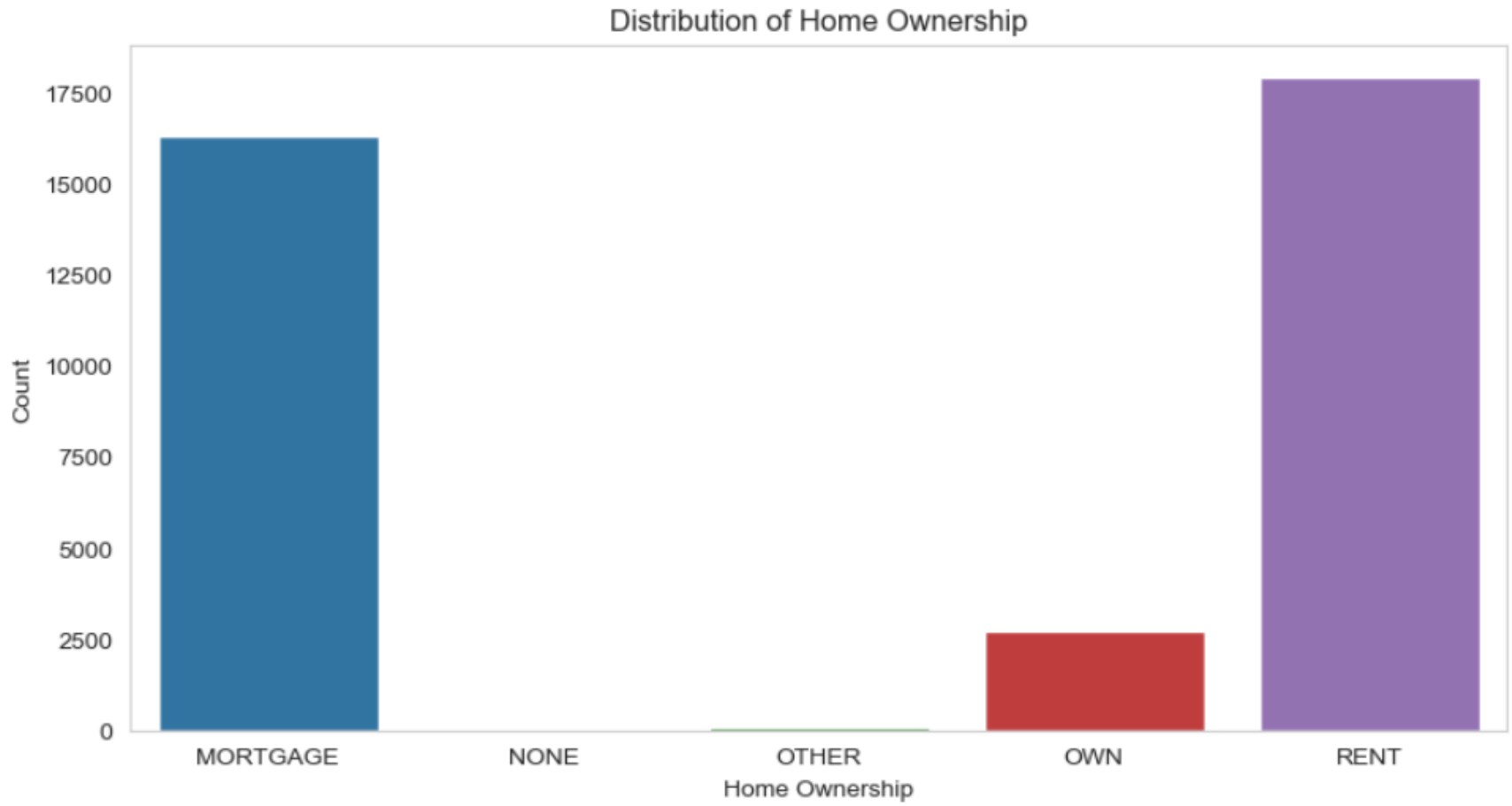
Distribution of Secondary Grade : In this analysis I want to depicts more insight that the loans within grade are more skewed towards lowered “Distribution of Secondary grades”





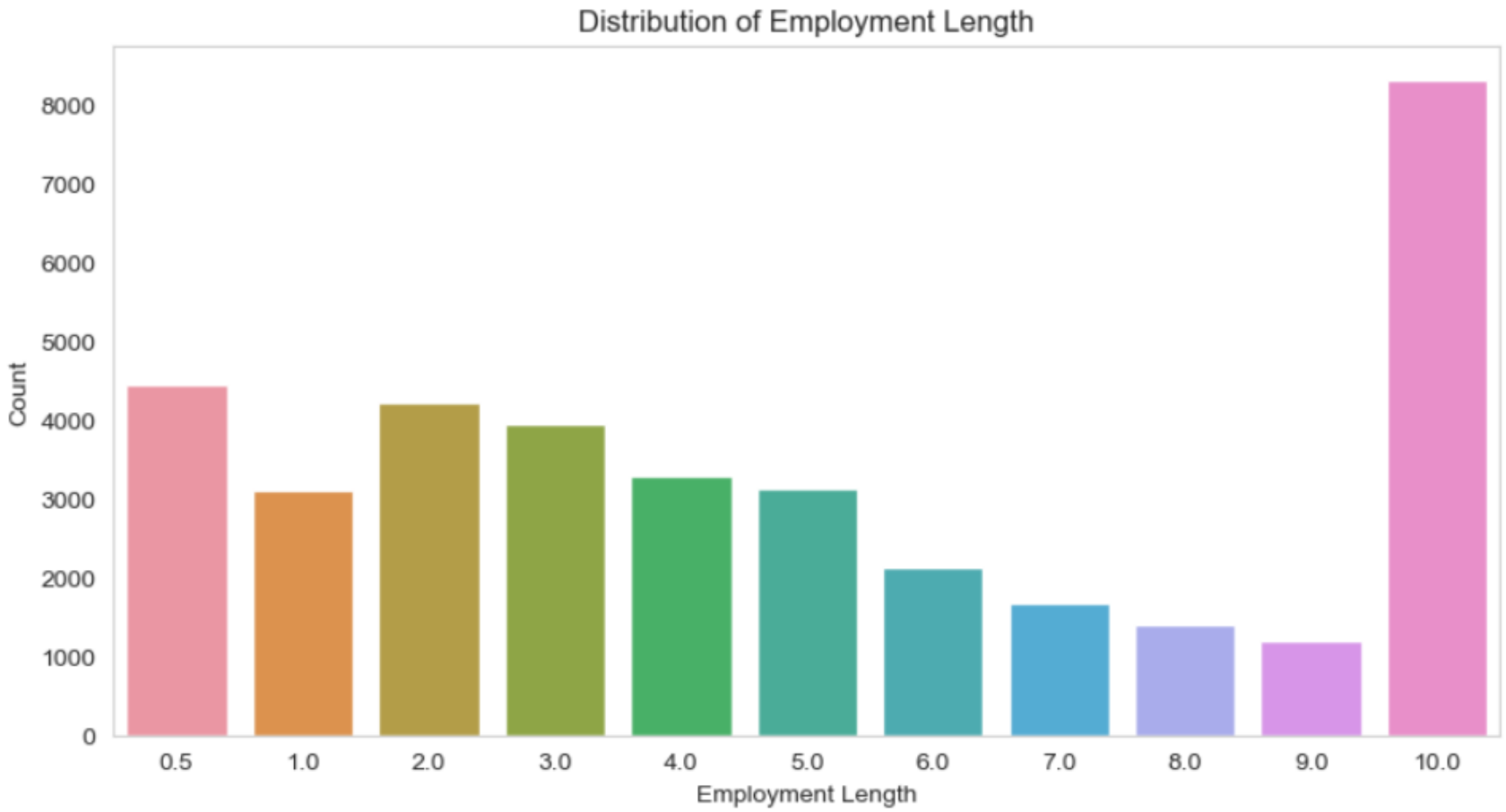
Loan against Home Ownership

With the majority of borrowers lacking property ownership and either renting or managing mortgage obligations, understanding their unique financial challenges is crucial. This demographic shift underscores the need for tailored financial solutions that address their specific economic realities.



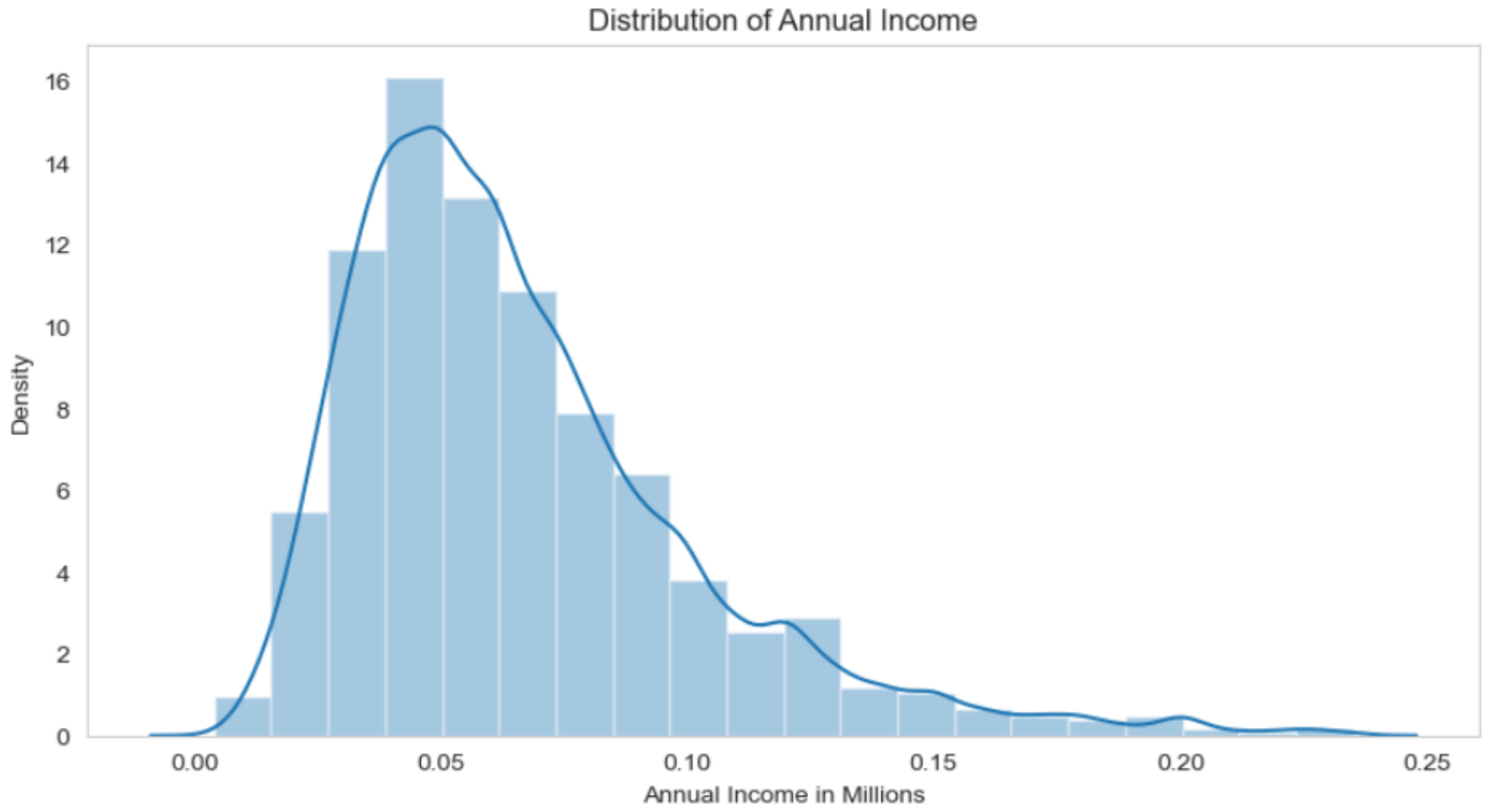
Loan Distribution of Employment length

"Employment length serves as a key indicator in loan distribution analysis, highlighting how borrower job stability influences lending decisions and risk assessments."



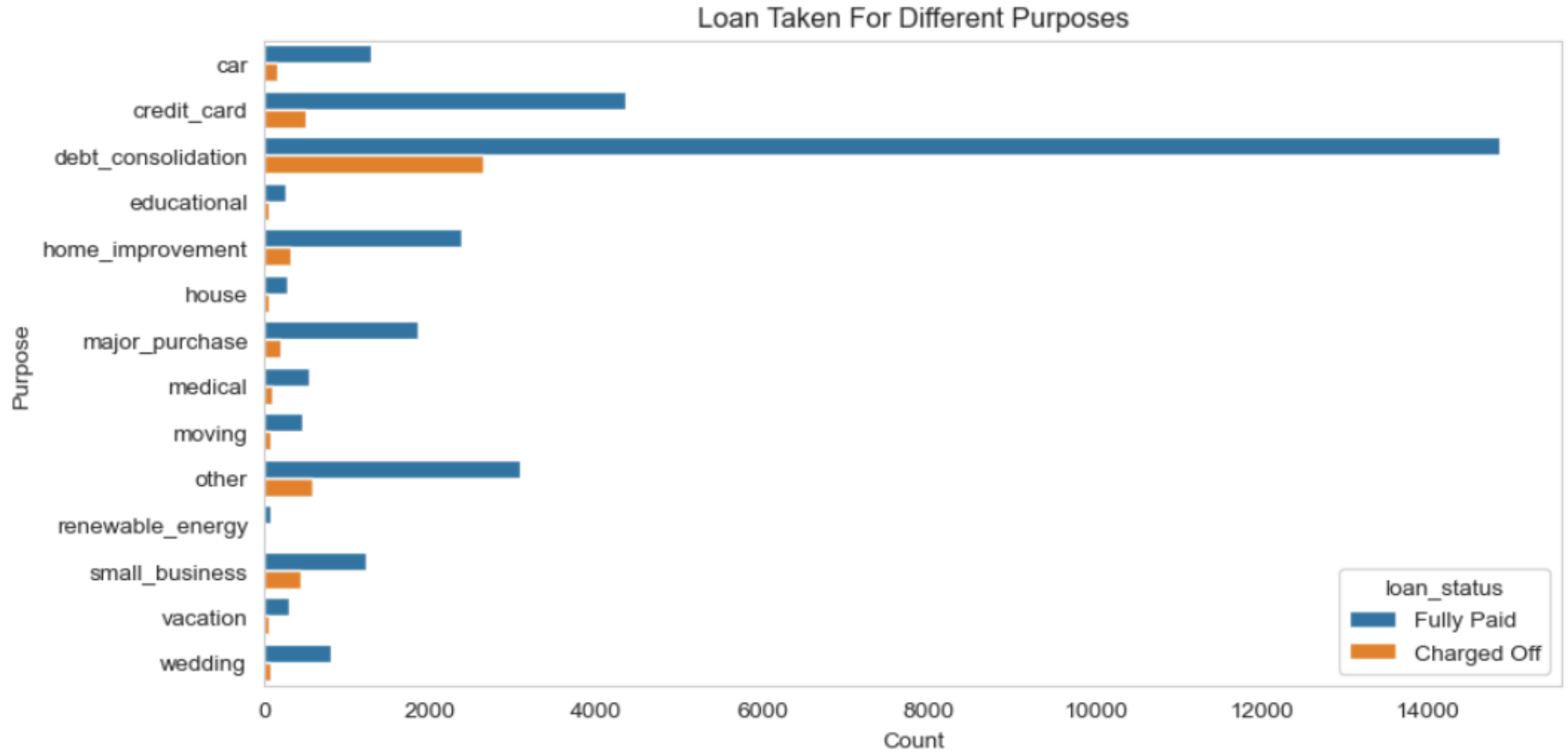
Distribution of Annual Income

"Analyzing loan distribution based on annual income underscores the importance of aligning financial products with borrowers' income levels, ensuring that lending decisions are both fair and feasible."



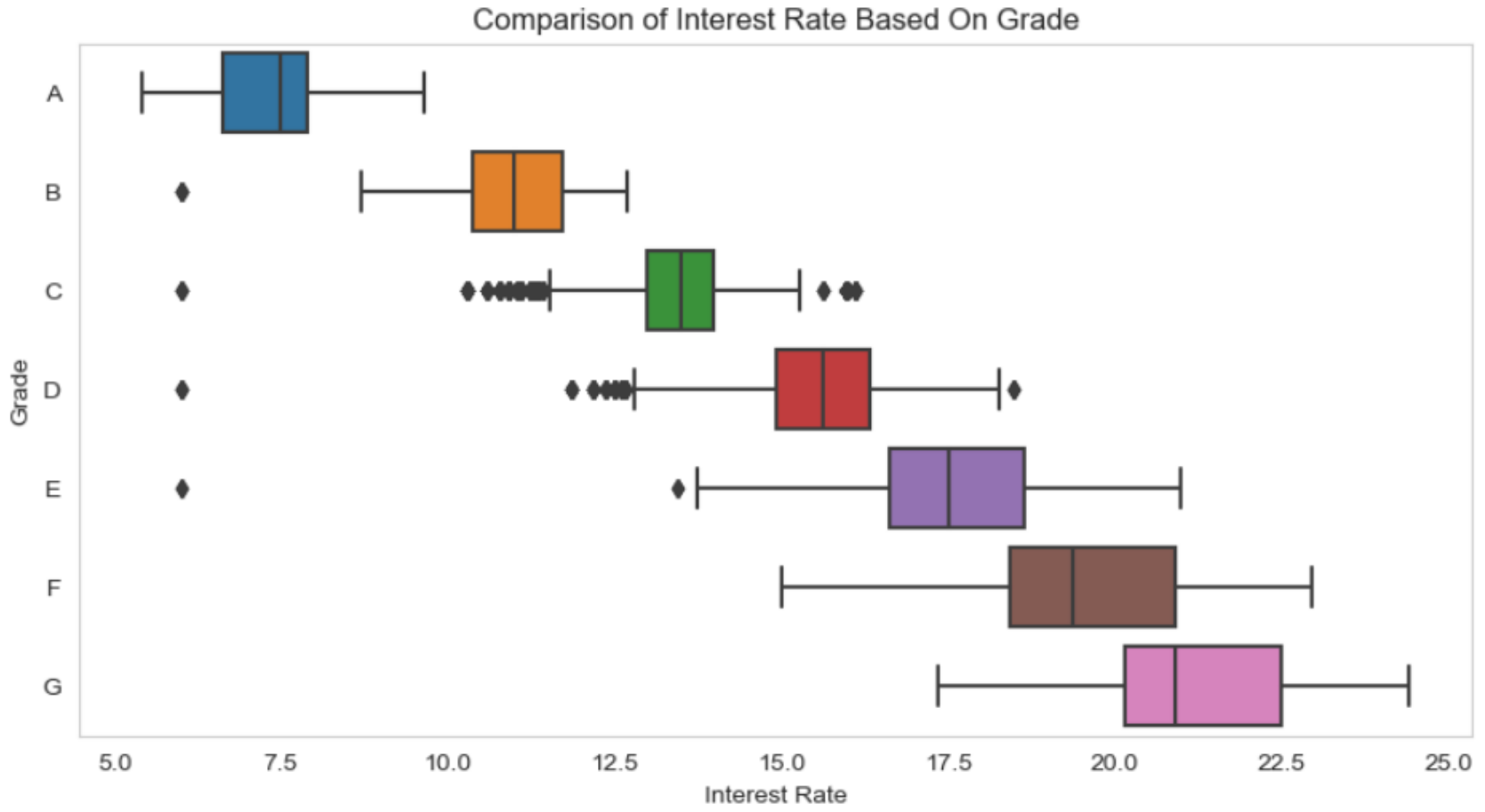
Loan Taken for Different Purpose

"Analyzing loan purposes sheds light on how financial products are utilized to fulfill various life needs, offering a deeper understanding of borrower priorities and informing more strategic lending practices."



Comparison of Interest Rate Based on Grade

- Loan grades serve as a critical measure of risk, with higher risk levels often leading to increased interest rates as lenders adjust for potential uncertainties.
- The correlation between loan grades and interest rates highlights how lenders manage risk, with higher grades indicating lower risk and consequently, lower interest rates.



Recommendations

Loan Amount and Term :- Larger loan amounts and longer terms may increase default risk, especially if not matched with the applicant's financial capacity.

Employment Status and Income Stability :- Stable employment and a steady income reduce the risk of default. Unstable or insufficient income increases it.

Loan Purpose :- Loans for purposes such as debt consolidation might have different risk profiles compared to loans for discretionary spending.

Borrower's Assets :- Assets like savings or property can serve as a buffer against default. The lack of assets might increase risk.

Economic Indicators :- Broader economic conditions, such as unemployment rates and economic growth, can impact default risk.

Previous Credit History : A history of late payments, charge-offs, or bankruptcy can indicate a higher risk of future default.