

LENDING CLUB CASE STUDY

Navaneetha Krishna A

Loan
Lending

EDA

Process

Feature Engineering:

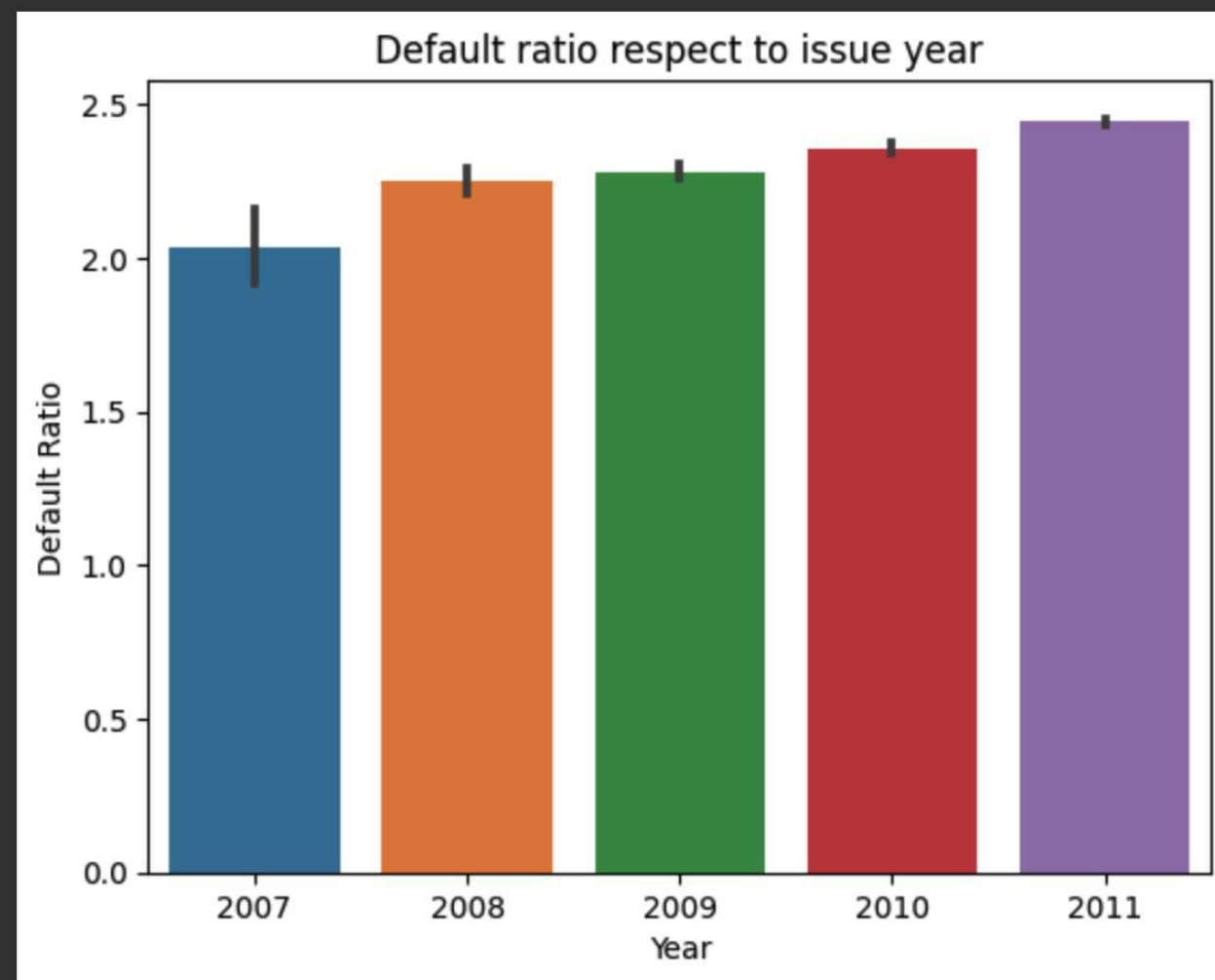
- Create new features or transform existing ones based on domain knowledge and analysis insights.
- Examples include debt-to-income ratio, loan amount to total income ratio, or employment length

Missing Data Imputation:

- Address missing data by filling in or imputing values to maintain data completeness
- Ensure data quality and integrity for accurate analysis.

Default ration with respect to Issue Year

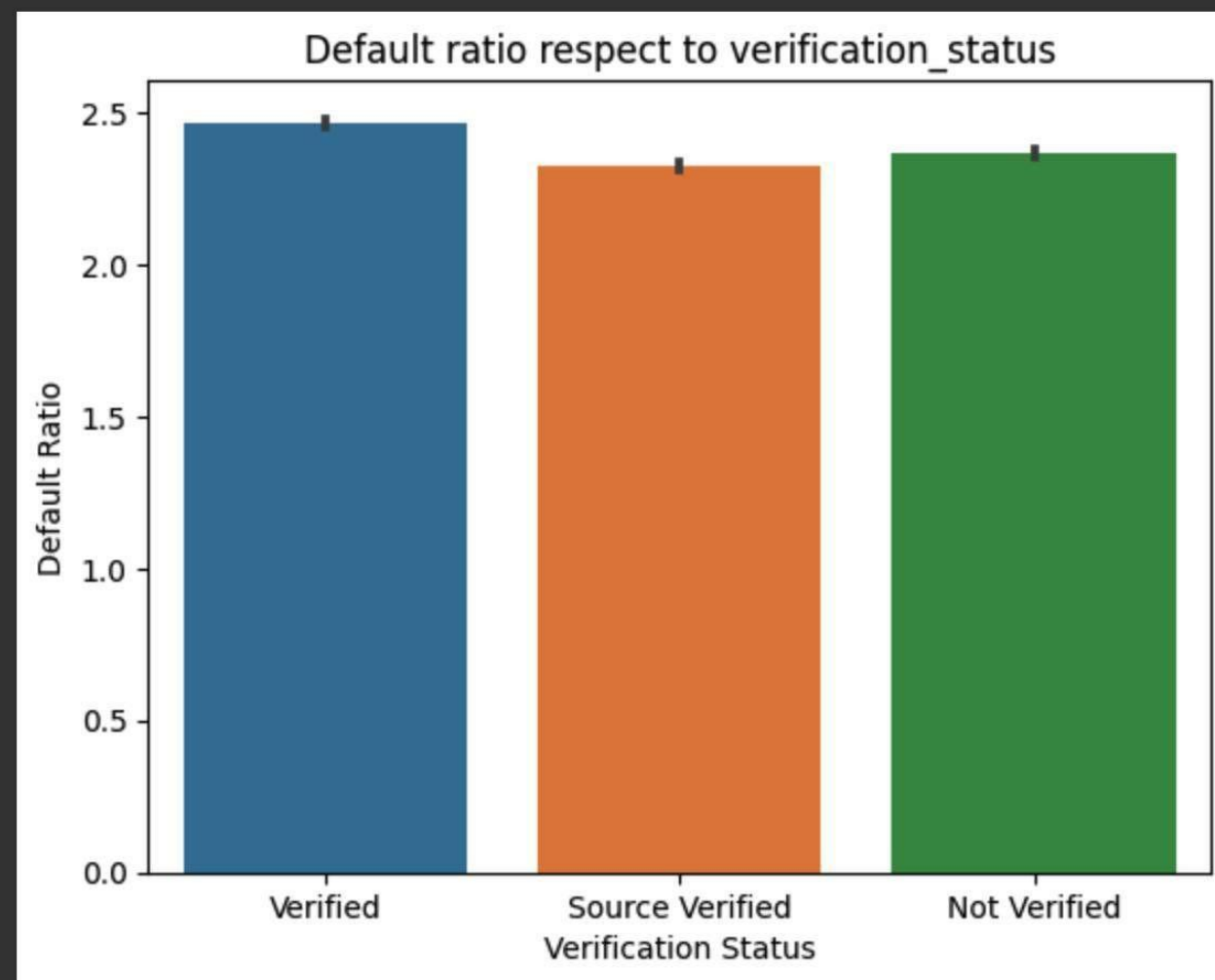
```
sns.barplot(x='year',y='dti_log',data=df1_filtered)  
plt.title('Default ratio respect to issue year')  
plt.xlabel('Year')  
plt.ylabel('Default Ratio')  
plt.show()
```



Here We can visualize that the default ratio is increasing with respect to the year. Comparing 2007 and 2011, we can see 86.13% increase

Default ratio with respect to Verification Status

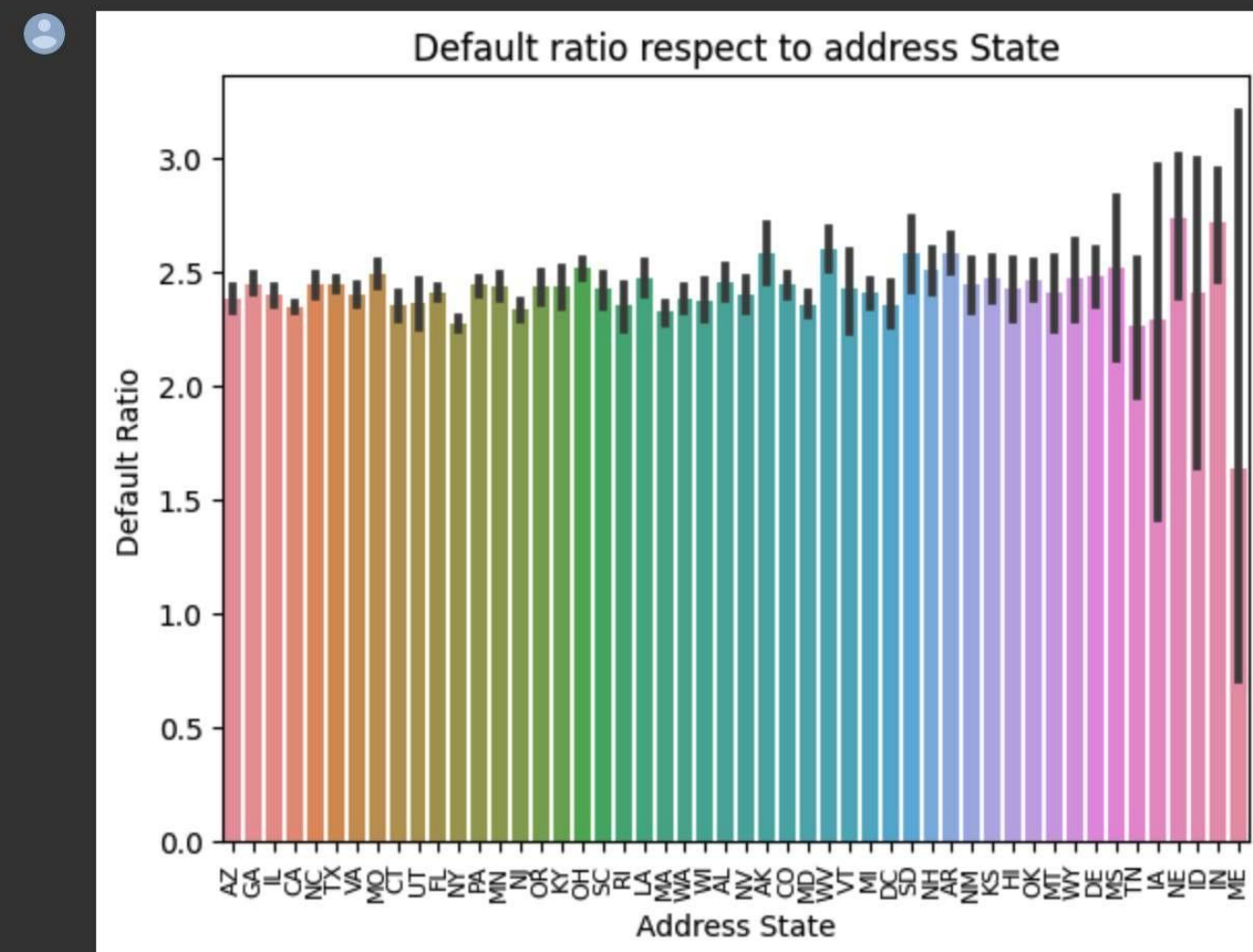
```
sns.barplot(x='verification_status',y='dti_log',data=df1_filtered)  
plt.title('Default ratio respect to verification_status')  
plt.xlabel('Verification Status')  
plt.ylabel('Default Ratio')  
plt.show()
```



In the above graph we can see that the default ratio for Verified status is the highest whereas Source verified has the least

Default ratio with respect to address state

```
▶ sns.barplot(x='addr_state',y='dti_log',data=df1_filtered)
plt.title('Default ratio respect to address State')
plt.xlabel('Address State')
plt.ylabel('Default Ratio')
#plt.figure(figsize=(15, 10))
plt.xticks(rotation='vertical')
plt.tick_params(axis='x', labelsize=8)
plt.show()
```

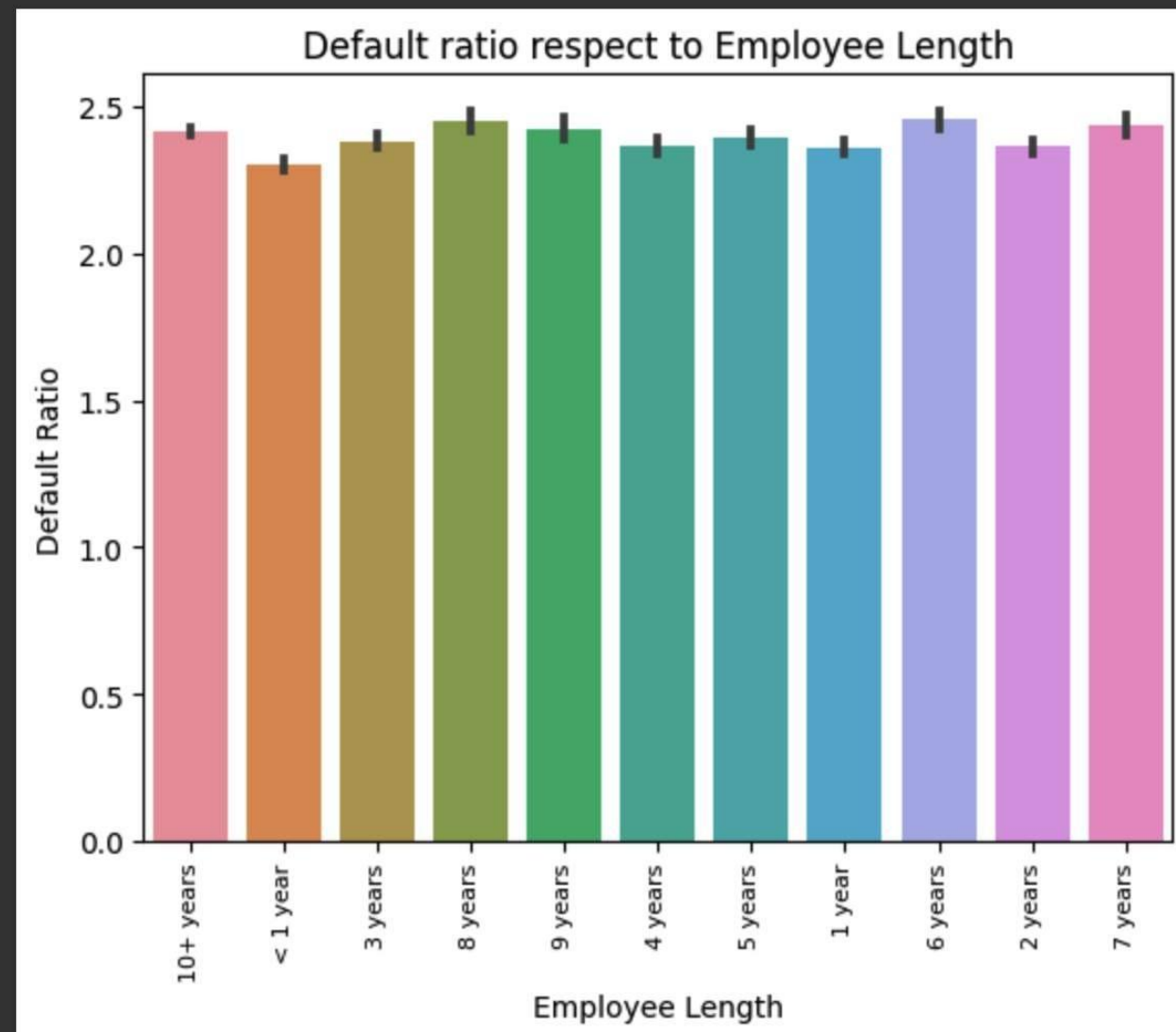


In this visualization, the state with IA & ID has the largest default ratio

Default ratio with respect to the employee work age

```
sns.barplot(x='Emp_len', y='dti_log', data=df1_filtered)
plt.title('Default ratio respect to Employee Length')
plt.xlabel('Employee Length')
plt.ylabel('Default Ratio')
plt.xticks(rotation='vertical')
plt.tick_params(axis='x', labelsiz=8)

plt.show()
```

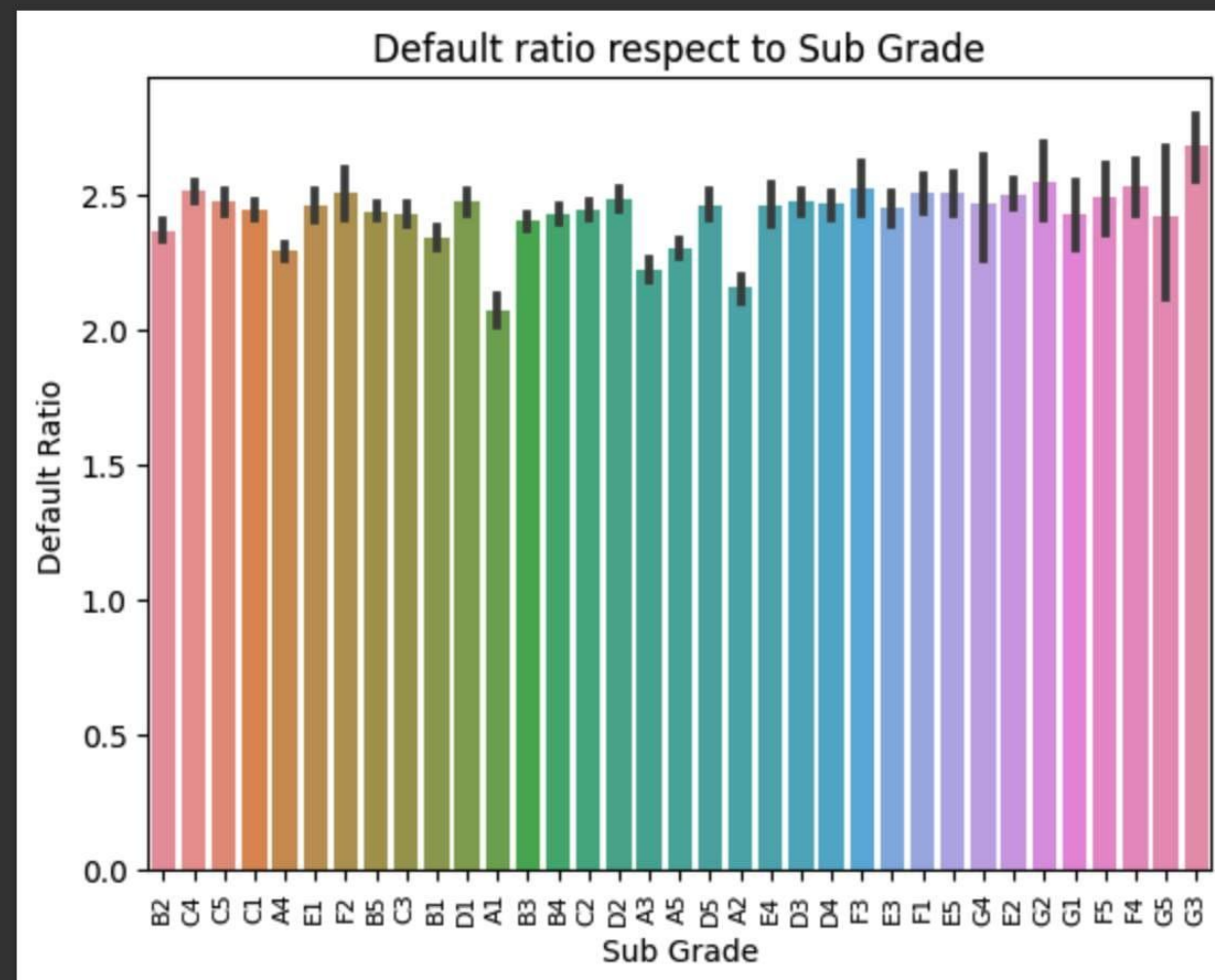


As we can see here, There is no much impact of employee length with default ratio but for Analyzation, employees with 6,9 and 10 years+ have more default ratio

Default ratio with respect to Sub Grade

```
sns.barplot(x='sub_grade', y='dti_log', data=df1_filtered)
plt.title('Default ratio respect to Sub Grade')
plt.xlabel('Sub Grade')
plt.ylabel('Default Ratio')
plt.xticks(rotation='vertical')
plt.tick_params(axis='x', labelsize=8)

plt.show()
```



G3 sub graded people have the highest default rate as we can see here

Categories of loan

```
import warnings

# Ignore all warnings
warnings.filterwarnings("ignore")

categories = ['<5000', '5000-10000', '10000-15000', '15000-20000', '20000-25000', '25000-30000', '30000-35000']

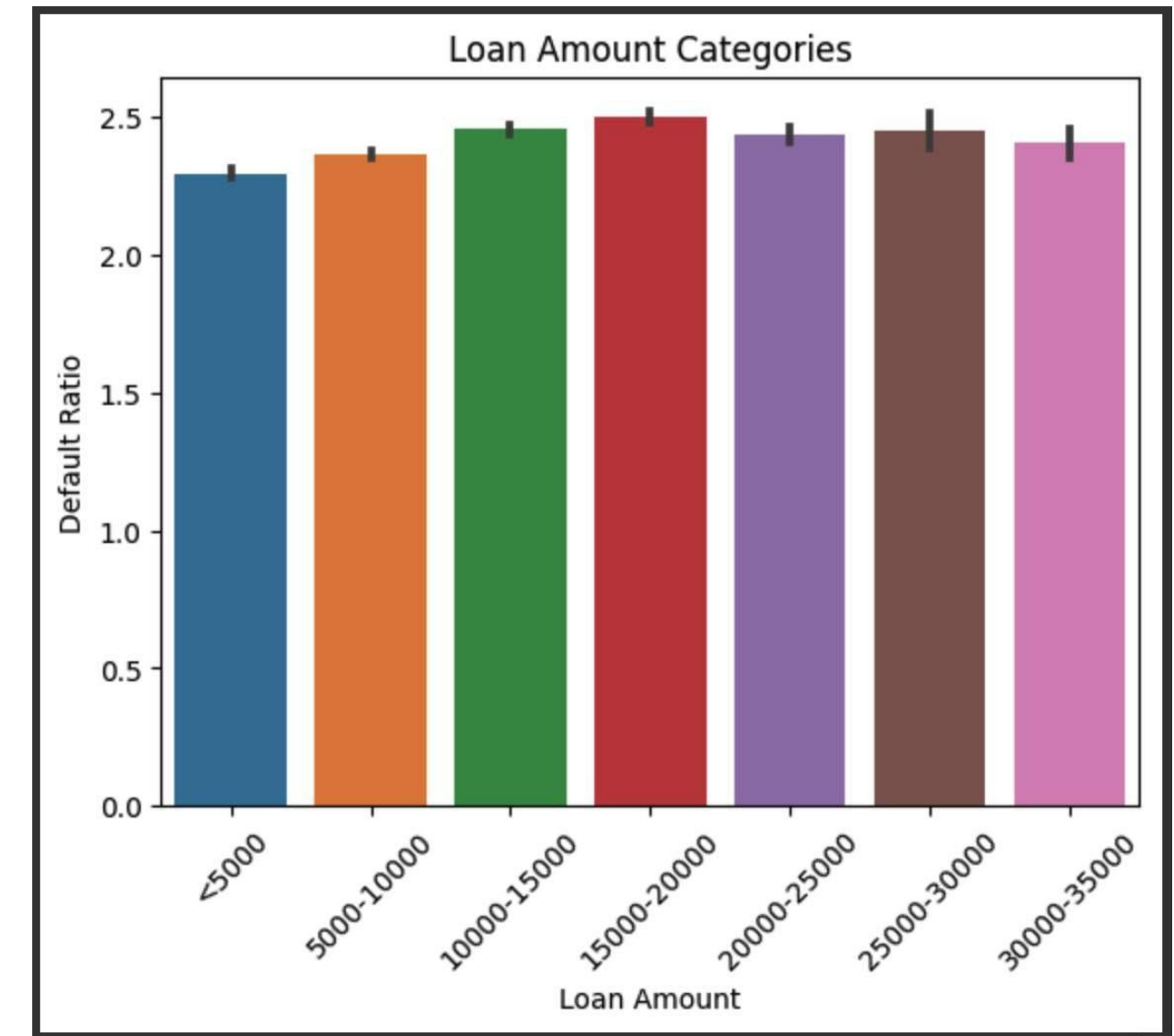
# Create a new column for loan amount categories
df1_filtered.loc[:, 'loan_amount_category'] = pd.cut(df1_filtered['loan_amnt'], bins=[0, 5000, 10000, 15000, 20000, 25000, 30000, 35000], labels=categories)

# Plot the categorized loan amounts
sns.barplot(x='loan_amount_category', y='dti_log', data=df1_filtered)

# Customize the plot
plt.xlabel('Loan Amount')
plt.ylabel('Default Ratio')
plt.title('Loan Amount Categories')

# Rotate x-axis labels for better readability
plt.xticks(rotation=45)

# Display the plot
plt.show()
```



The people with less than 5000 amount category's default ratio is least as compared to other categories followed by 5000-10000 and 30000-35000 category

Default ratio with respect to annual Income

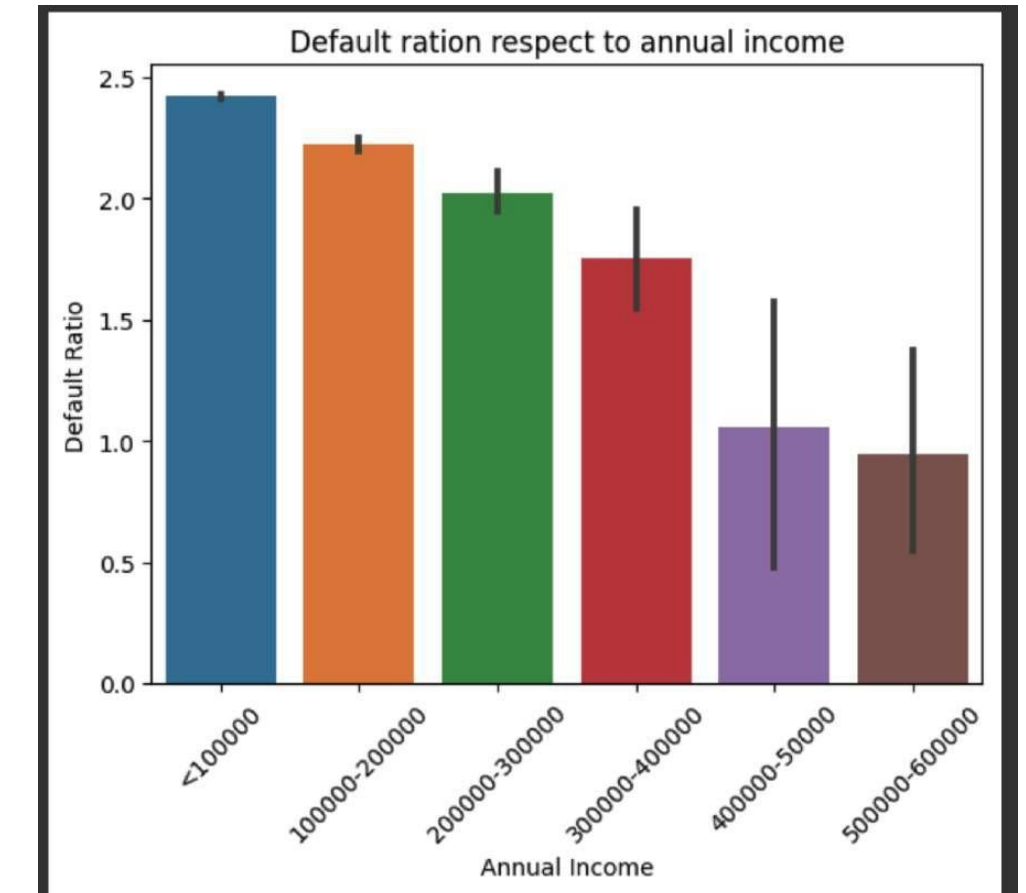
```
categories_i= ['<100000', '100000-200000', '200000-300000', '300000-400000', '400000-500000', '500000-600000']
df1_filtered['annual_inc_category'] = pd.cut(df1_filtered['annual_inc'], bins=[0,100000,200000, 300000, 400000,500000,600000], labels=categories_i)

sns.barplot(x='annual_inc_category', y='dti_log', data=df1_filtered)

# Customize the plot
plt.xlabel('Annual Income')
plt.ylabel('Default Ratio')
plt.title('Default ration respect to annual income')

# Rotate x-axis labels for better readability
plt.xticks(rotation=45)

# Display the plot
plt.show()
```



Default ratio is highest for people whose annual income is less than 100000, here we can analyze that as the salary of people is rising their default ratio is becoming less

Default ratio with respect to Grade

```
[ ] df1_filtered['annual_inc'].max()

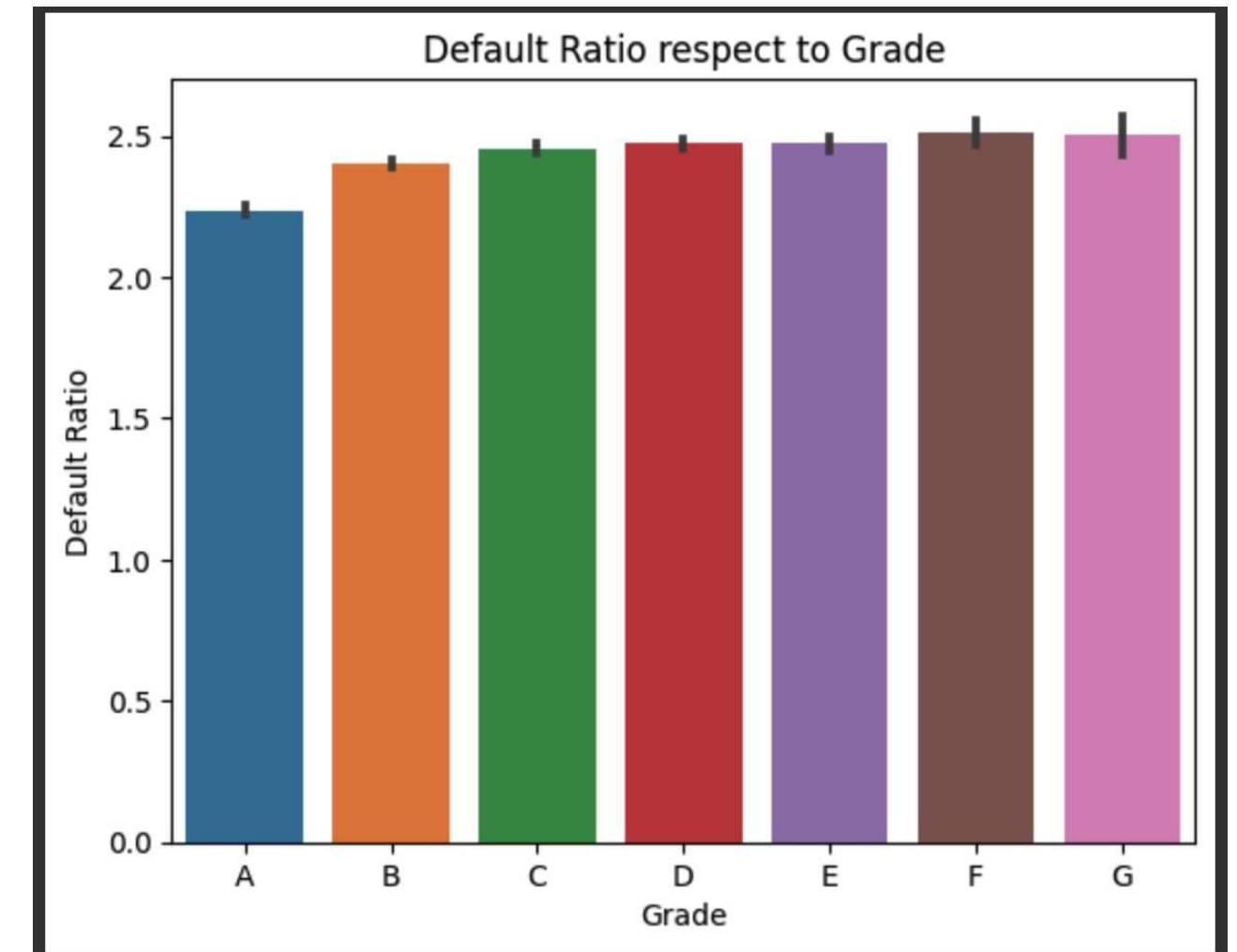
6000000.0

# Sort the 'grade' column in alphabetical order
sorted_grades = sorted(df1_filtered['grade'].unique())

# Create the bar plot using sorted grades
sns.barplot(x='grade', y='dti_log', data=df1_filtered, order=sorted_grades)

# Customize the plot
plt.xlabel('Grade')
plt.ylabel('Default Ratio')
plt.title('Default Ratio respect to Grade')

# Display the plot
plt.show()
```

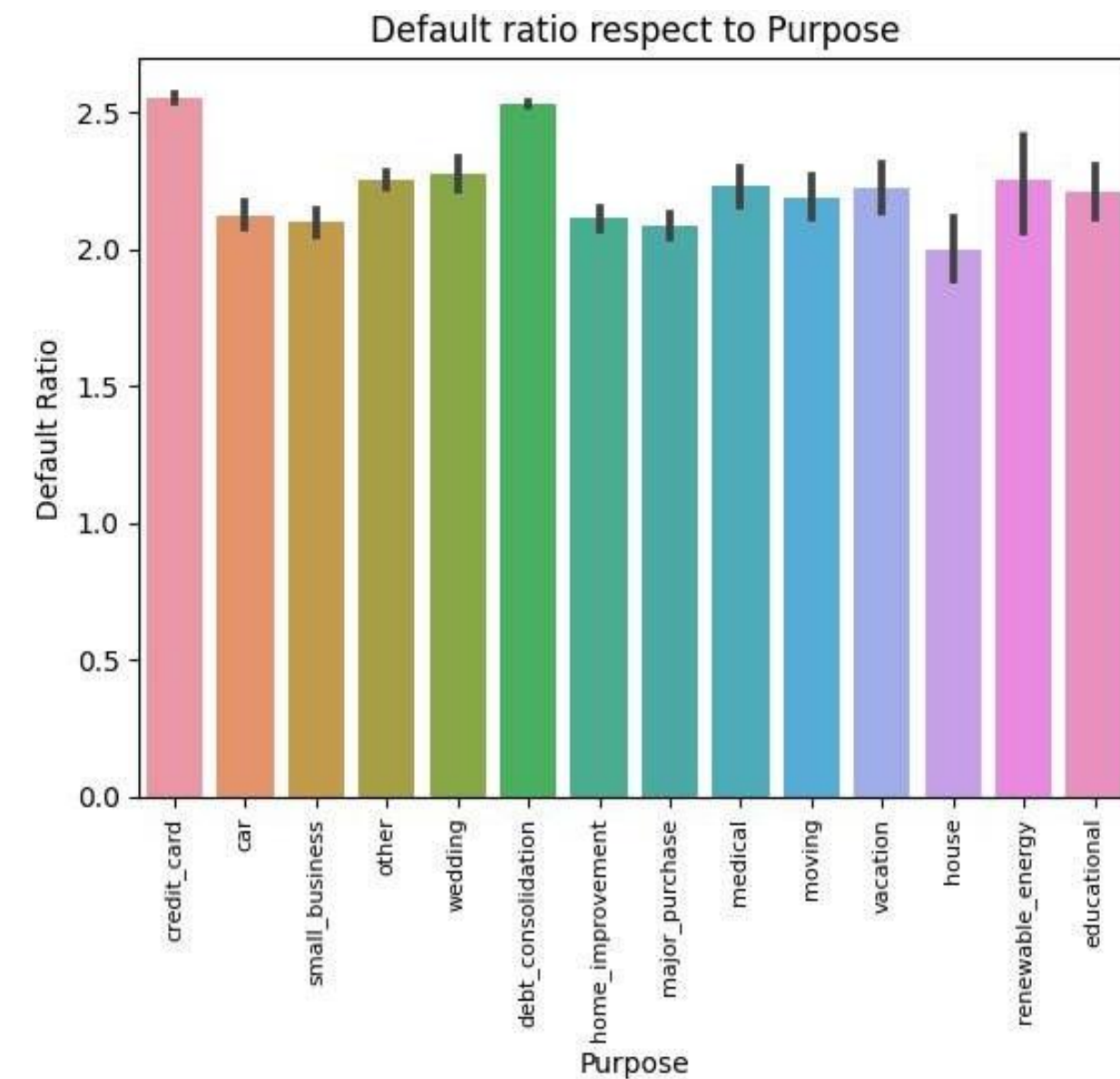


As the Grade of the people increases, the default ratio also increases. Both are directly proportional in this case

Default ratio with respect to purpose

```
▶ sns.barplot(x='purpose', y='dti_log', data=df1_filtered)
plt.title('Default ratio respect to Purpose')
plt.xlabel('Purpose')
plt.ylabel('Default Ratio')
plt.xticks(rotation=45)
plt.tick_params(axis='x', labels=8)

plt.show()
```

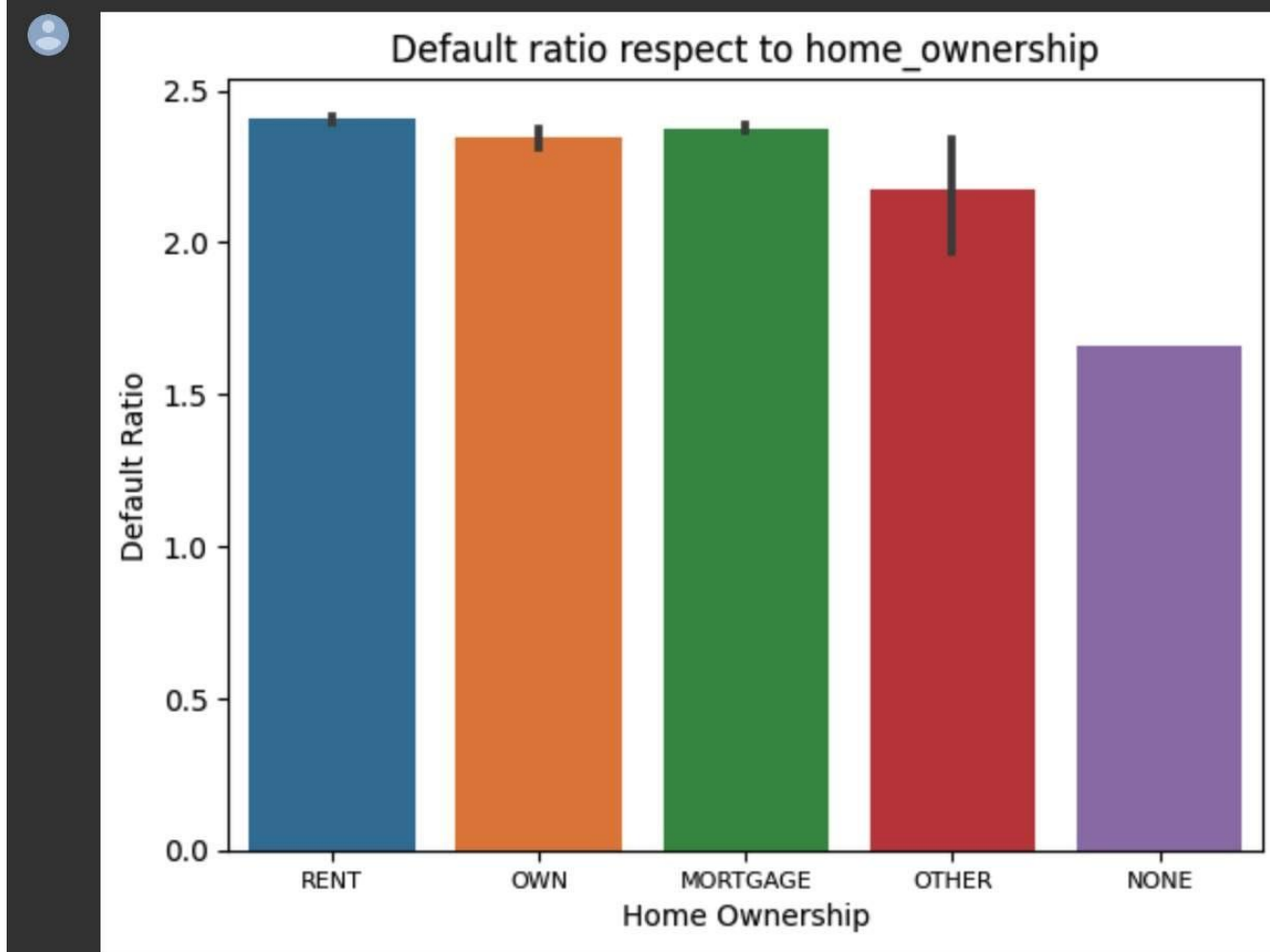


The default ratio is highest for Credit cards, The default ratio of credit card users can vary depending on various factors such as the specific credit card provider, the target market, creditworthiness of the users, economic conditions, and risk management practices.

Default ratio with respect to Home ownership

```
sns.barplot(x='home_ownership', y='dti_log', data=df1_filtered)
plt.title('Default ratio respect to home_ownership')
plt.xlabel('Home Ownership')
plt.ylabel('Default Ratio')
#plt.xticks(rotation=45)
plt.tick_params(axis='x', labels=8)

plt.show()
```



There are various factors that effects the default ratio of rent people.

1. Financial Instability
2. Affordability Challenges
3. Lack of Home Equity:
4. Housing Market Conditions

Box plot of interest rate and loan status

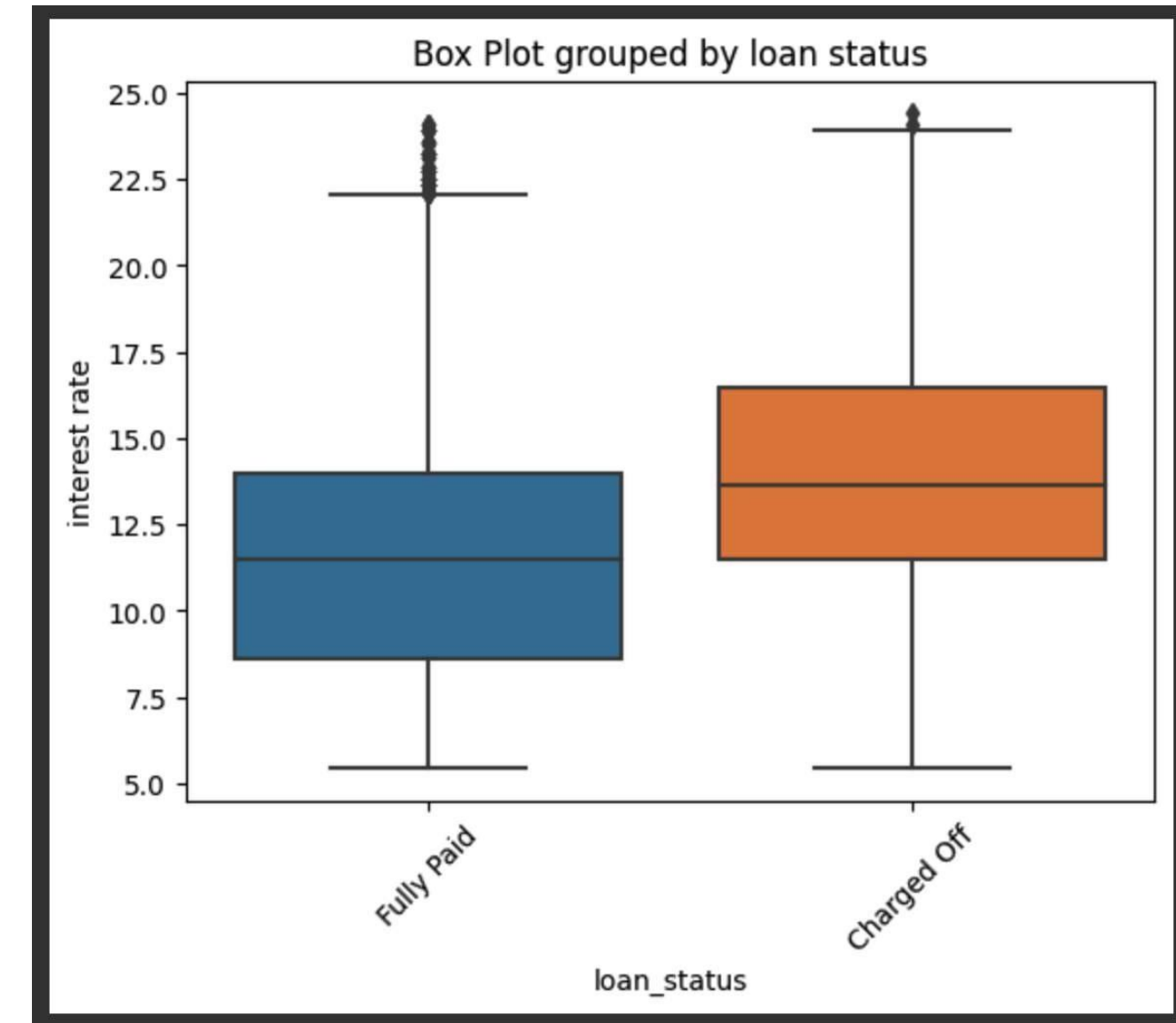
```
[ ] #df3['int_rate'] = df3['int_rate'].str.rstrip('%').astype(float)

# Create a box plot
sns.boxplot(x='loan_status', y='int_rate', data=df3)

# Customize the plot
plt.xlabel('loan_status')
plt.ylabel('interest rate')
plt.title('Box Plot grouped by loan status')

# Rotate x-axis labels for better readability
plt.xticks(rotation=45)

# Display the plot
plt.show()
```



Charged off interest rates are more as compare to fully paid interest rates

Thank You