# Lending Case Study

EXPLORATORY DATA ANALYSIS

# Problem Statement

You work for a **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

# Objective

Use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

# Constraint

When a person applies for a loan, there are two types of decisions that could be taken by the company:

• **Loan accepted**: If the company approves the loan, there are 3 possible scenarios described below:

• **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate) • **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

• **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

• **Loan rejected:** The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the • loan was rejected, there is here is no transactional history of those applicants with the company and so this data is not available with the company.

# Data Summary

Loan.csv file contains 39717 rows and 111 columns.

Columns contains header.

Columns available with all null values.

# Data Cleaning

There were no excel header and footer, and data is properly aligned with header.

No duplicate rows available.

Out of 39717, **loan_status** column contains 1140 rows of ‘Current’ status, 32950 rows of ‘Fully Paid’ status and 5627 rows of ‘Charged off’ status.

There were 55 columns which is having all the rows values as null/blank and doesn’t participate in analysis has been removed.

‘url’ and ‘member_id’ is unique in nature and has been deleted.

‘desc’ and ‘title’ text/description’ values and doesn’t participate has been dropped.

8 columns whose values were 1, and is uniqueness in nature has been dropped.

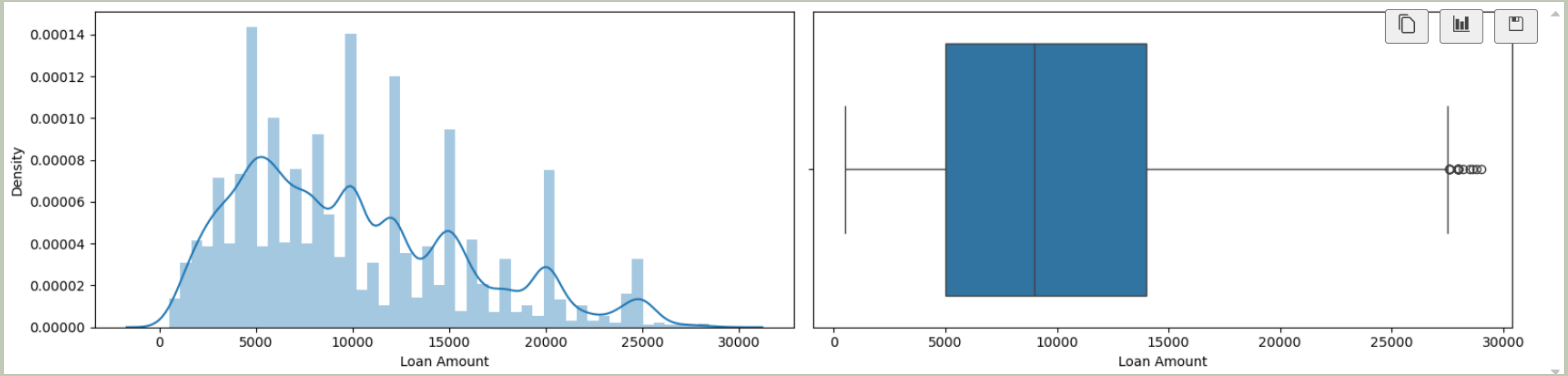
# Data Conversion

- ‘int_rate’ has been converted from string to int. Additional ‘%’ has been trimmed regex method.
- Column ‘loan_funded_amnt’ and ‘funded_amnt’ converted to float type.
- issue_d has been converted to yyyy-mm-dd format.
- Creating a derived columns for ‘issue_year’ and ‘issue_month ’ from ‘issue_d’ which will be using for further analysis.
- Loan_amnt, annual_inc, int_rate, dti has been bucketed and created new columns.
- Additional string value has been trimmed from ‘term’ column and has been converted to int data types.
- Removed Outliers data for 'loan_amnt', 'funded_amnt', ‘funded_amnt_inv','int_rate', 'installment', ‘annual_inc’.

# Loan Amount

**Observations:**

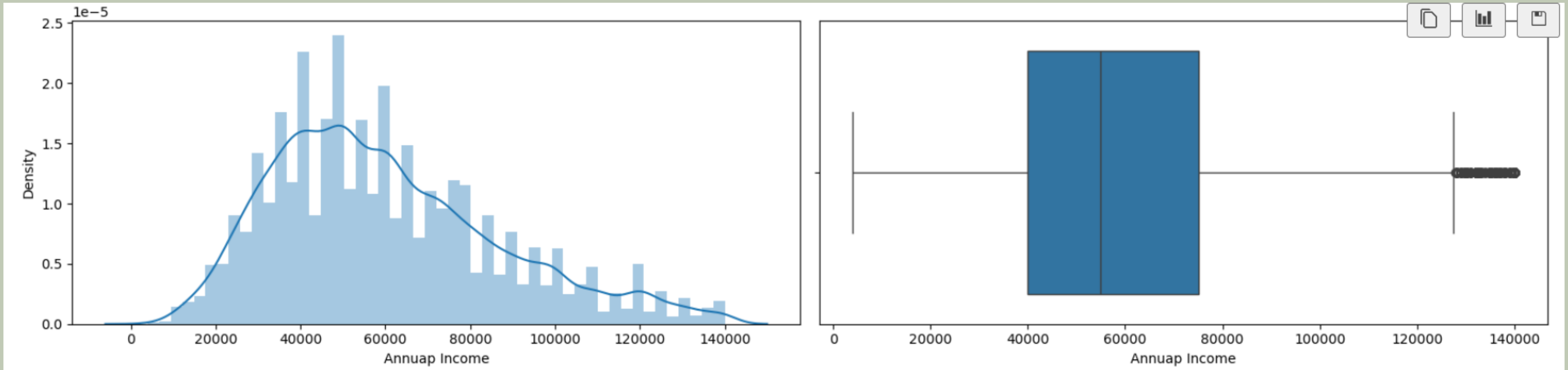
- Most of the loan amount lies between 5k and 14k.



# Annual Income

## Observations:

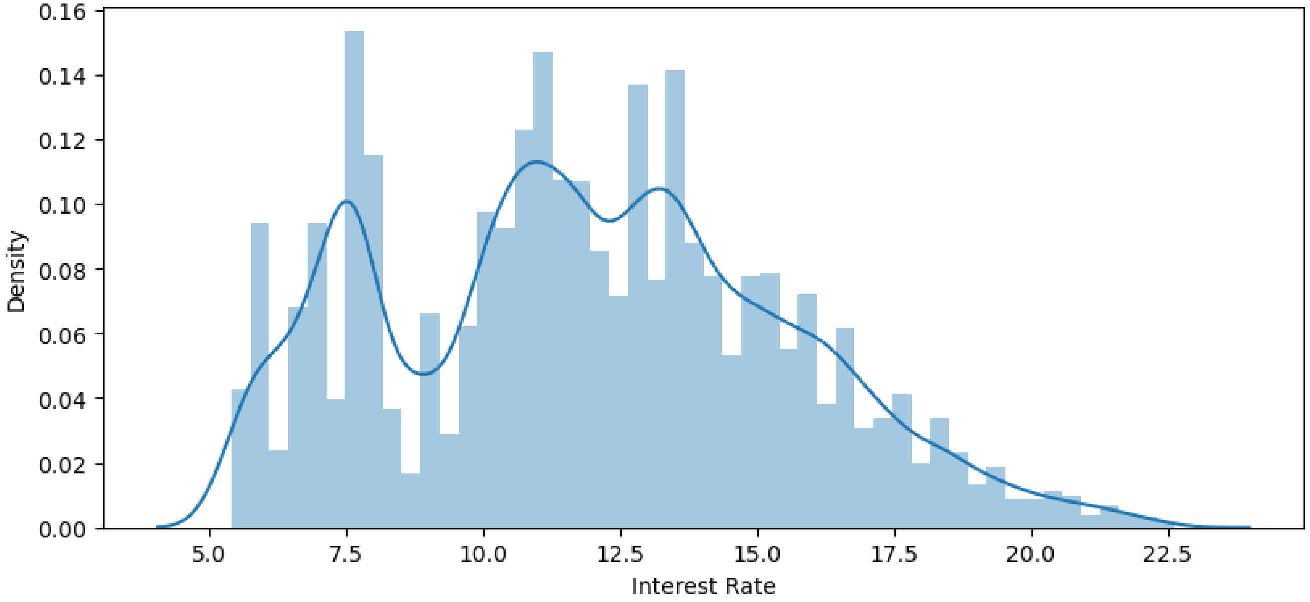
The Annual income of most if applicants lies between 40k-75k.

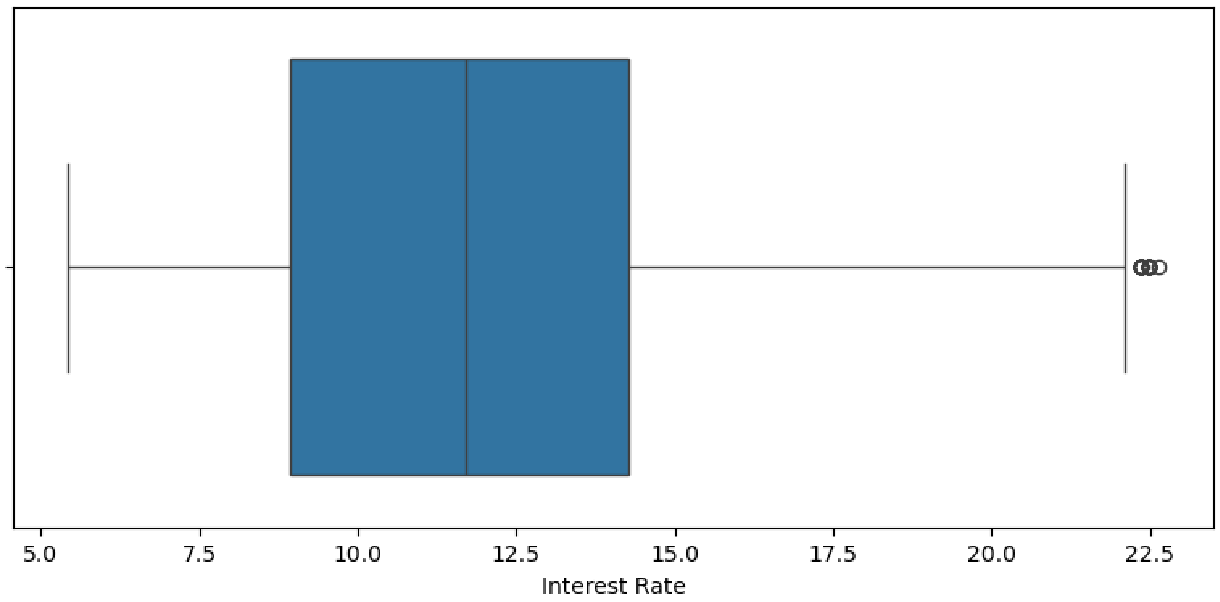


# Interest Rate

**Observations:**

Most of the applicant's rate of interest is between in the range of 8%-14%.





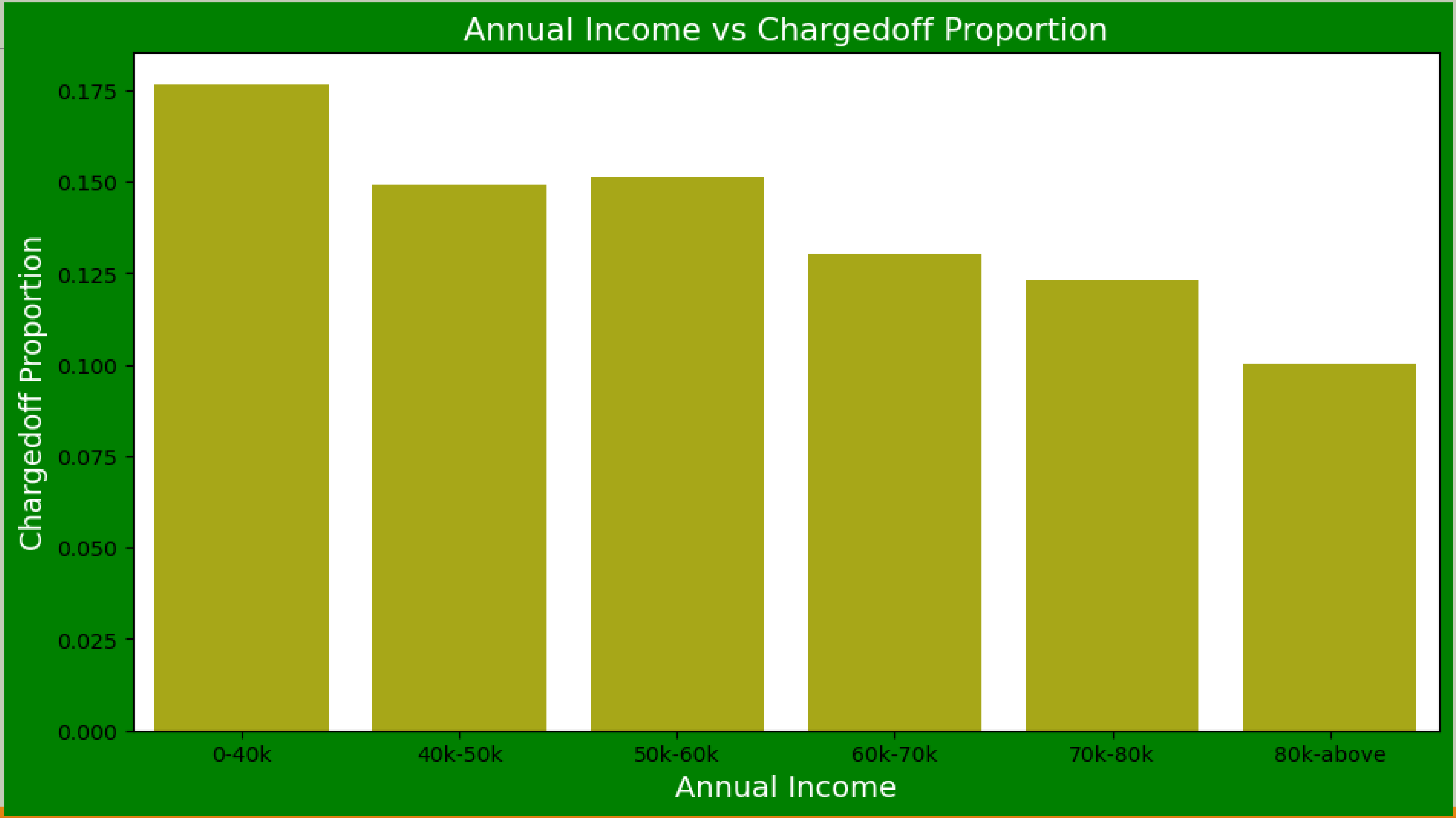
# Annual Income vs Charged Off

**Observations:**

Income range of 0-40k has more chances of charged off.

Income range above 80k has less chances of charged off.

Annual income is inversely proportional with the charged off proportion.

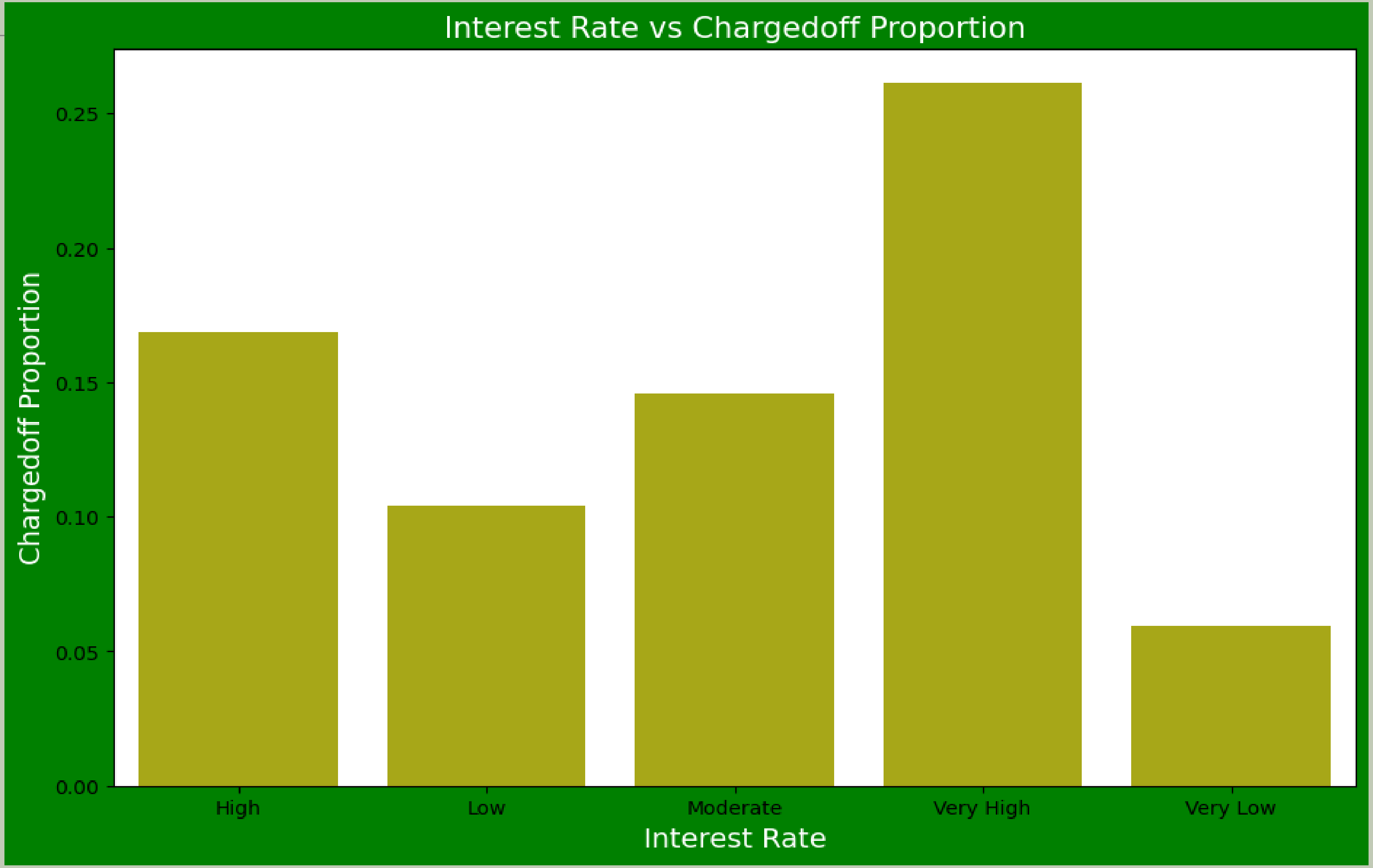


# Interest Rate vs Charged Off Proportion

**Observations:**

Very high interest rate has more chances of being charged off.

Very low interest rate has less chances of being charged off.



# Conclusion

Income range between 0-20K has high chances of charged off.

Those who are not owning the home is having high chances of loan defaulter.

Interest rate more than 16% has good chances of charged off as compared to other category interest rates.

High DTI value having high risk of defaults.