GRAMENER CASE STUDY

LOAN DATA - RISK ANALYSIS

Team Members

Praval Rastogi

Deepsi Shrivastava

Prakhar Shukla

Lakshmipathi Kakarla

Analysis Overview

- Business Objective

- Data Exploration

- Data Cleaning and Manipulation

- Data Analysis

- Conclusion

Business Objective

Business Objective

- Identify the **driving factors** to identify new borrowers that would likely default on their loans

Data Exploration

Data Exploration

- o 39,717 loan requests issued through the time period 2007 to 2011

- o Contains borrower's credit history and loan information

- o Unordered Categorical Variables (Nominal)
    - o home_ownership - Home Ownership Status
    - o verification_status - Income Source Verification Status
    - o purpose - Loan Category
    - o term - Loan Term(Number of Payments)
    - o loan_status - Current Status of the Loan

- o Ordered Categorical Variables (Ordinal)
    - o grade - Loan Grade
    - o emp_length - Employment Length

Data Preparation and Processing

Data Preparation and Processing

- o Removed columns that had missing values for all observations(E.g. tot_coll_amt, open_acc_6m etc.)

- o Removed columns that had identical data for all observations (E.g. pymnt_plan, policy_code etc.)

- o Removed columns that had unique values for all observations as analysis done on the collection (E.g. id, url etc)

- o Removed columns that had bad quality data (i.e. zip code etc.)

- o Converted continuous variables to range of values to enhance interpretation of results (E.g. annual_inc)

- o Derived metrics(month, year, quarter) based date columns

Assumptions

Employment Length(emp_length)

- o Considered missing values as 0 (years), <1 year as 0 , and 10+ years as 10 (years).

Loan Status(loan_status)

- o Considered observations with loan status as either 'FULLY PAID' or 'CHARGED OFF'.

- o Ignored 'CURRENT' loan status in our analysis

Annual Income(annual_inc)

- o Considered up to 98th percentile of annual income to remove outliers

Debt To Income(dti)

- o Considered up to 98th percentile of debt to income to remove outliers

Data Analysis

The analysis divided into three parts to identify driving factors for defaulters:

- o Univariate Analysis
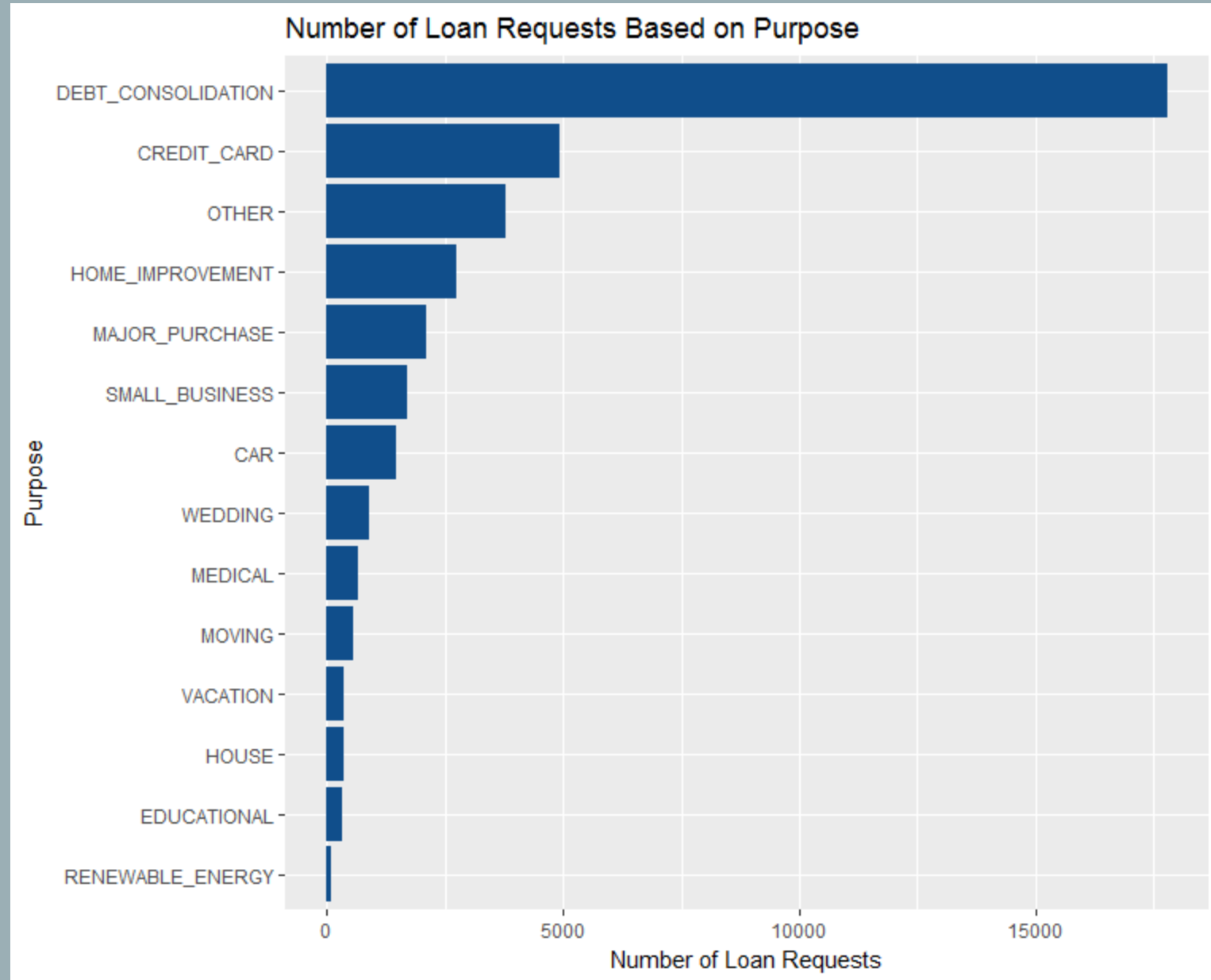
- o Segmented Univariate Analysis

- o Bivariate Analysis

Univariate Analysis - Annual Income

o Univariate Analysis on Annual Income indicates that, majority of people, whose income ranges from 20,000 to 1,00,000 are opting for loans


Income Range Of Borrowers

Univariate Analysis - Loan Purpose

o Univariate Analysis on Loan Purpose indicates that, majority of loans requested for Debit Consolidation, followed by Credit Card.
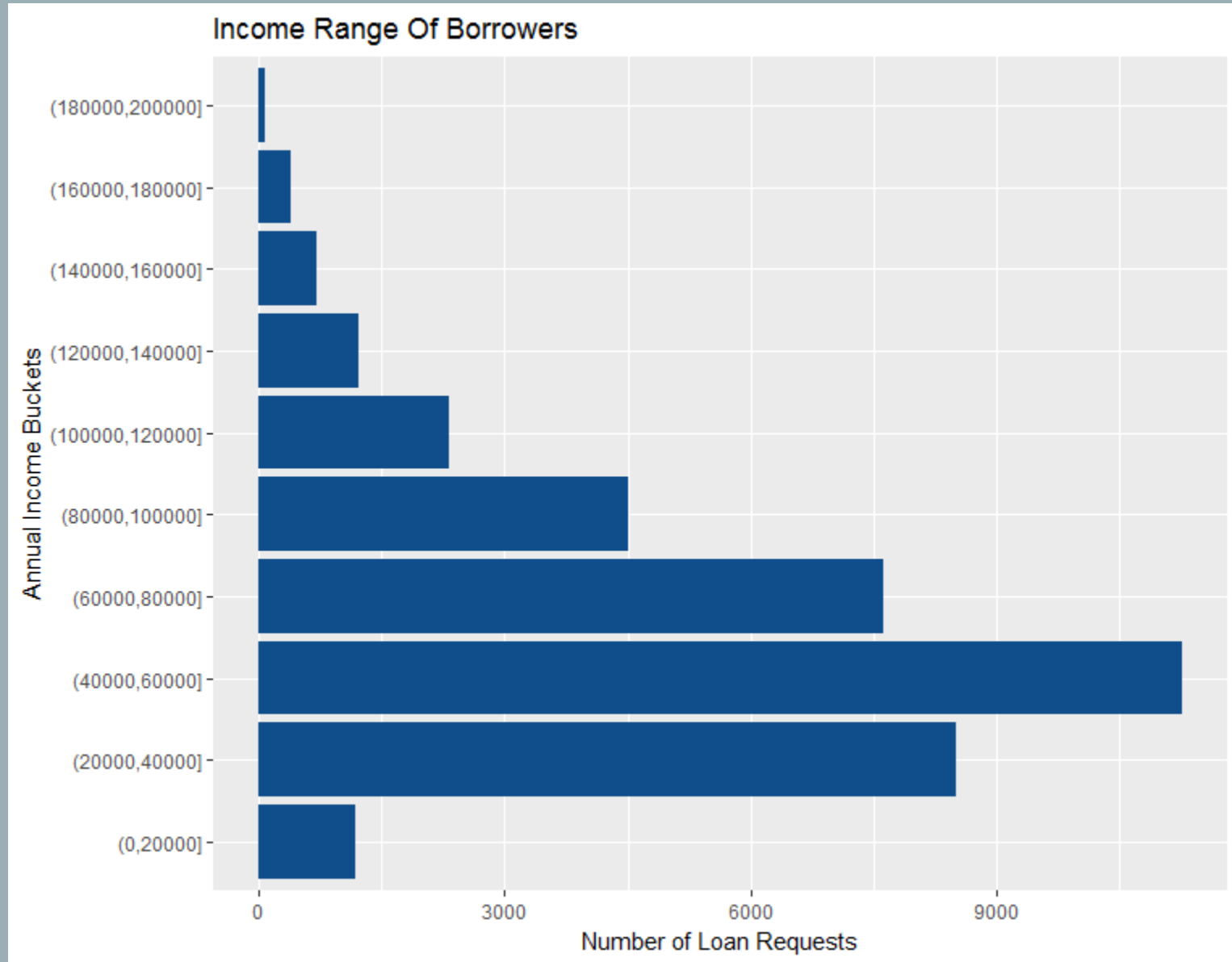

Number of Loan Requests Based on Purpose

Univariate Analysis - Loan Issued Date

o Univariate Analysis on **Loan Issued Year/Quarter** indicates that, number of loan requests processed doubled every year.

o And consistently, majority of the loan requests processed in Q4 (holiday season)
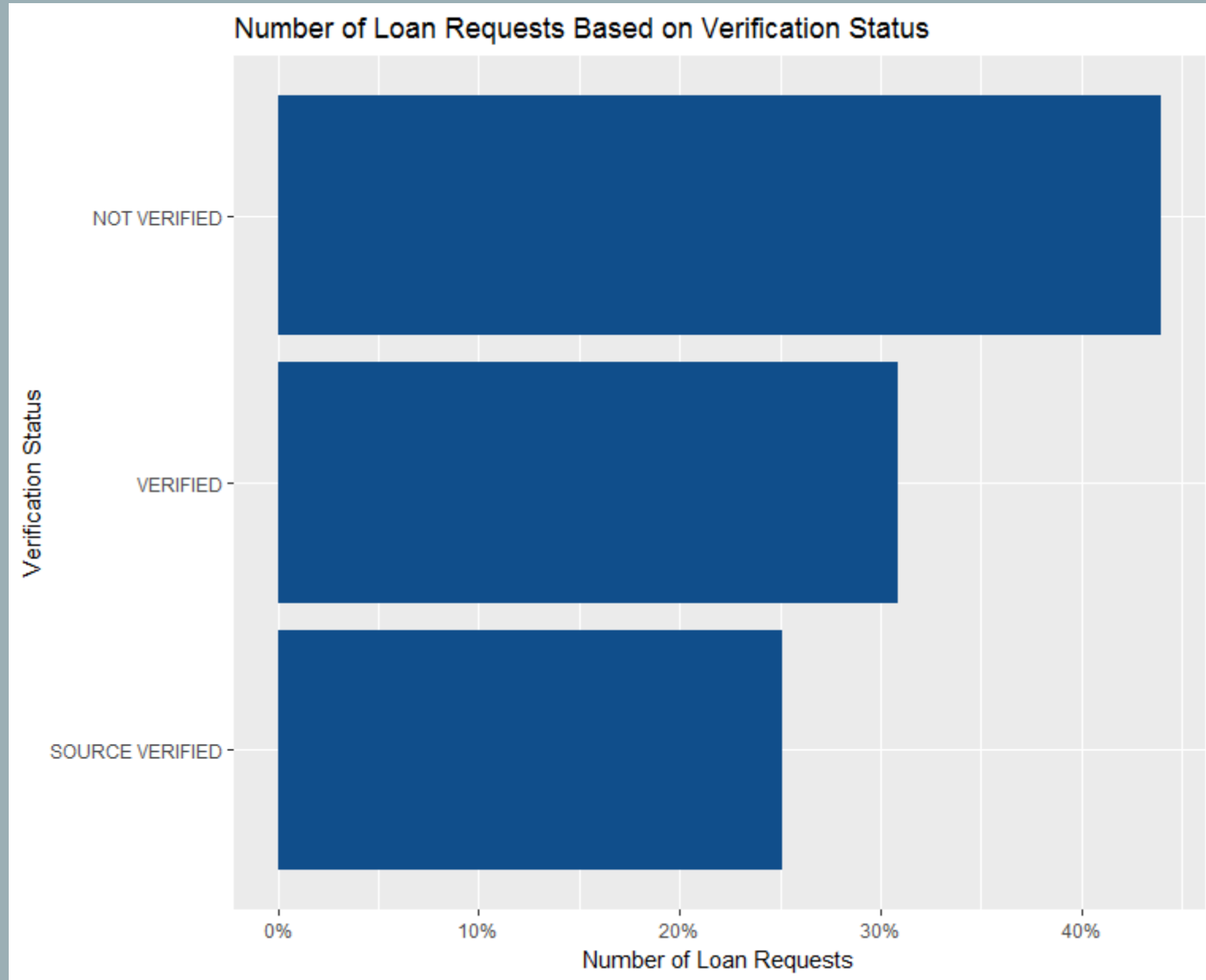

Number of Loans Issued by Year and Quarter
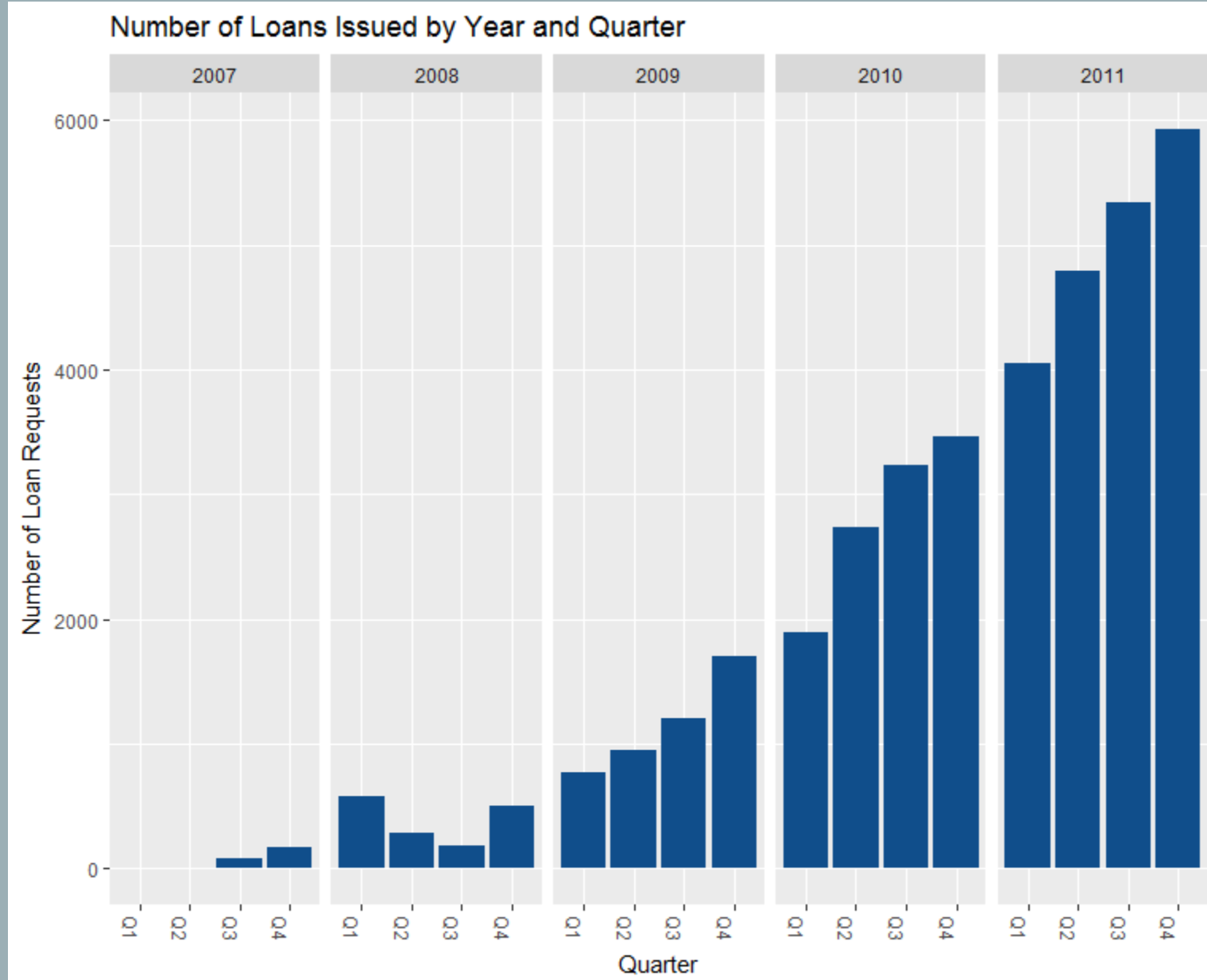
Univariate Analysis - Loan Grade

o   Univariate Analysis on **Loan Grade** indicates that, 75% of loan requests issues with loan grade 'A', 'B' and 'C'.
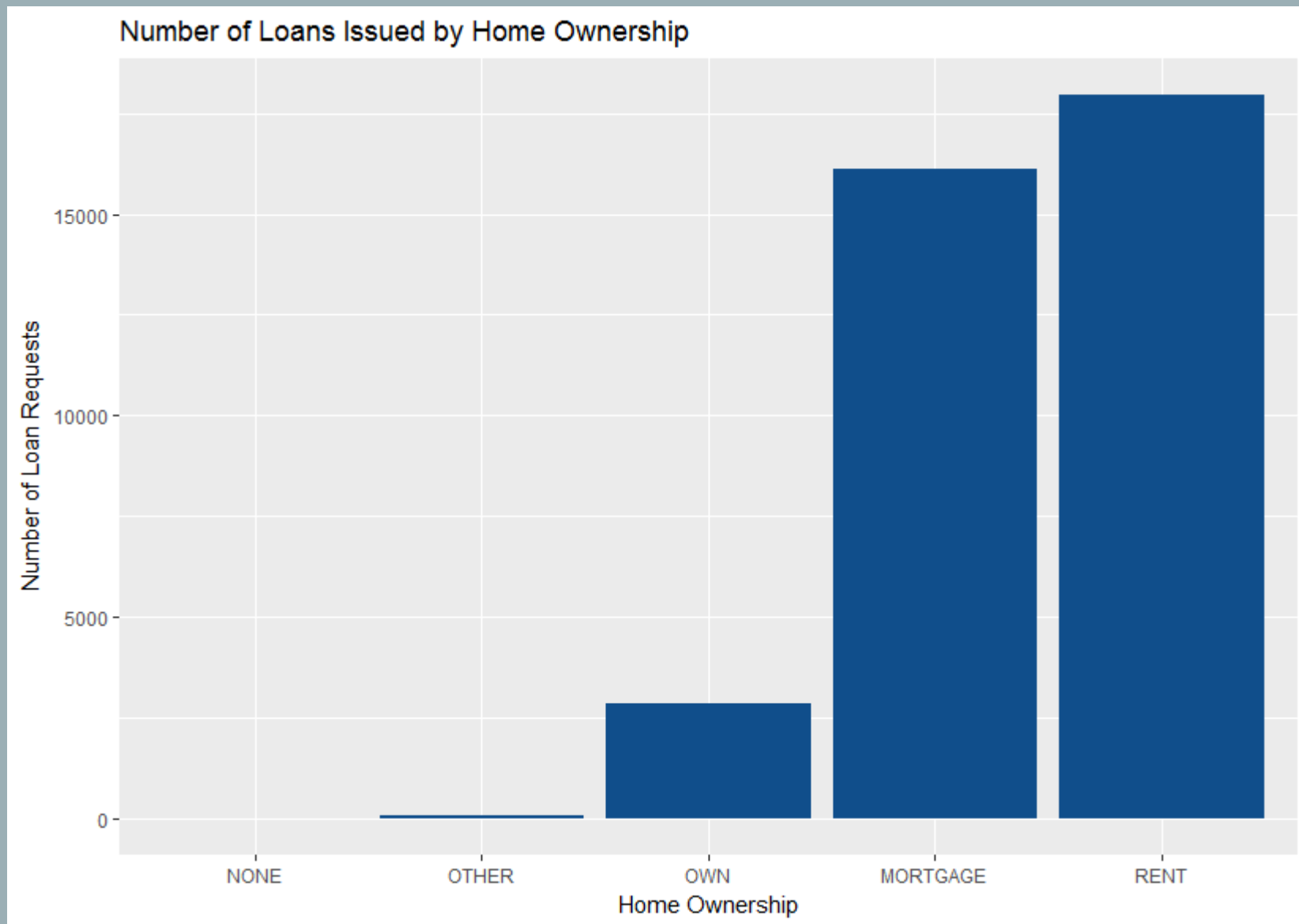

Percentage of Loans Issued by Grade
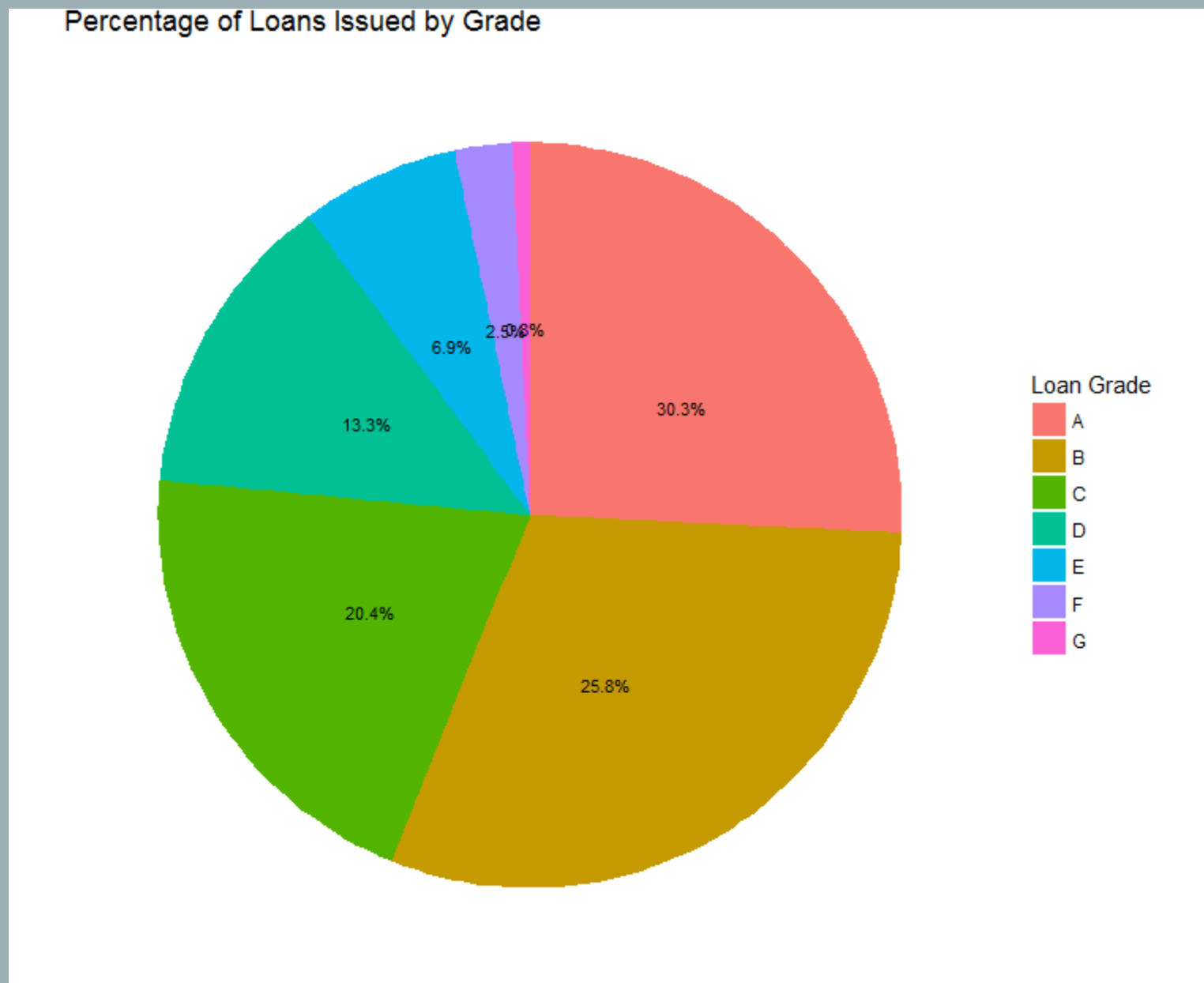
Univariate Analysis - Income Source Verification Status

o Univariate Analysis on Income Source Verification Status indicates that, around 45% of loan requests were issued even though income source was not verified.


Number of Loan Requests Based on Verification Status

Univariate Analysis - Home Ownership

o Univariate Analysis on Home Ownership indicates that, people with own home took significantly lesser number of loans compare to people with mortgage and rent.


Number of Loans Issued by Home Ownership

Univariate Analysis - Loan Status

o Univariate Analysis on Loan Status indicates that, most(around 85%) of the investors were able to receive their funds (with interest).
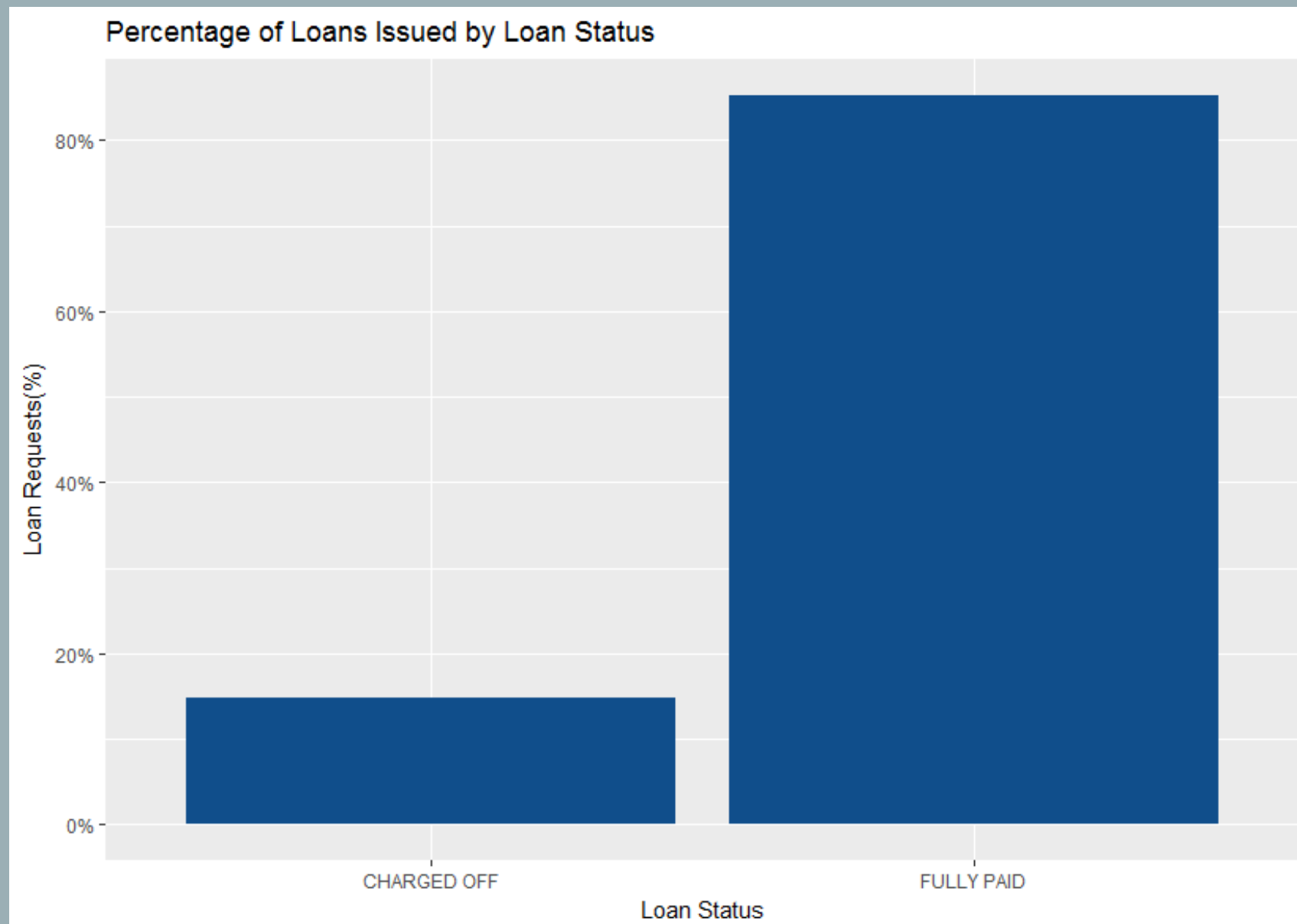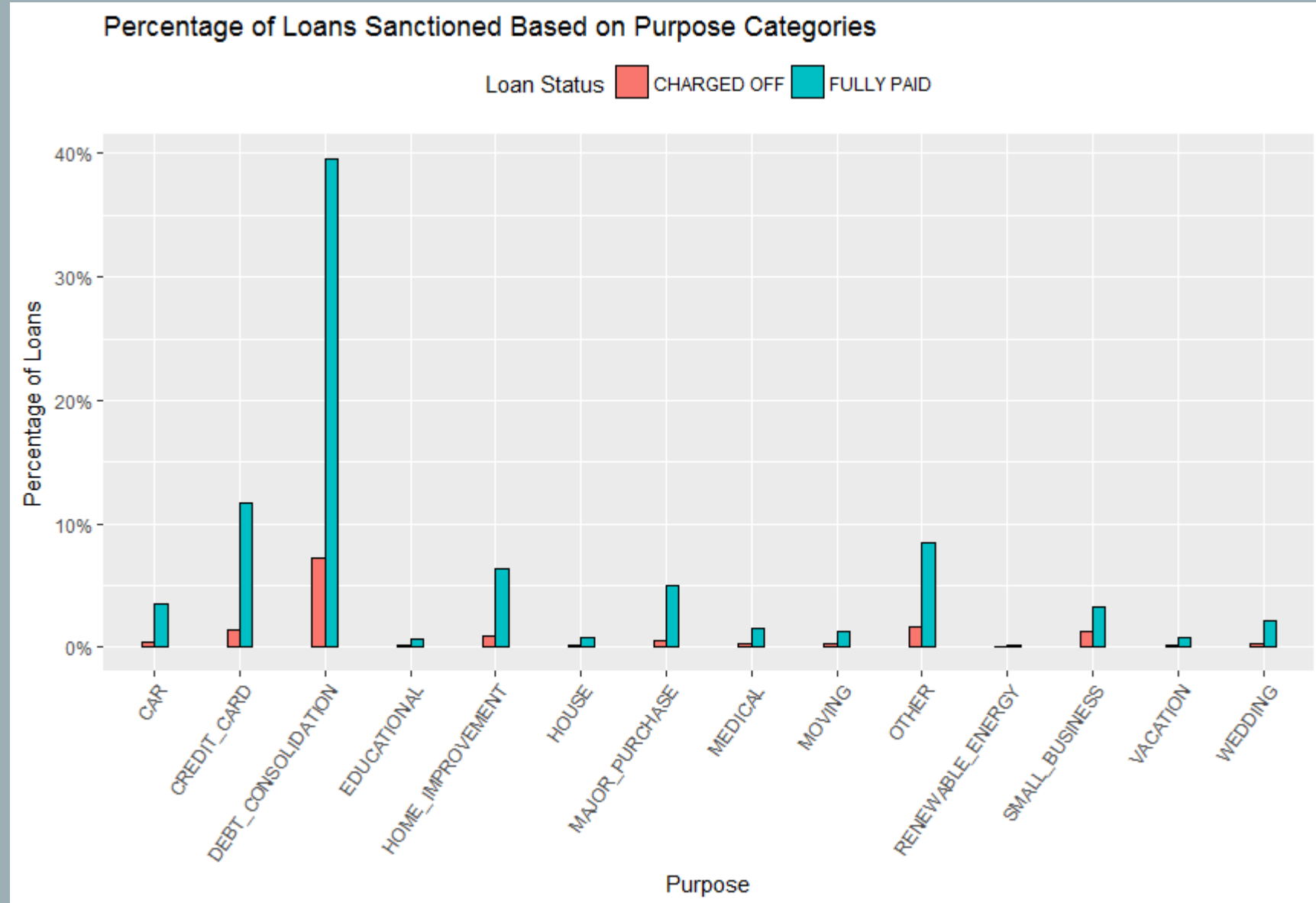
Univariate Analysis - Loan Status

o Based on Univariate Analysis on **Loan Amount (for Charged Off),** we observed that, majority of loan defaulters are having loan amount multiples of 5000$.
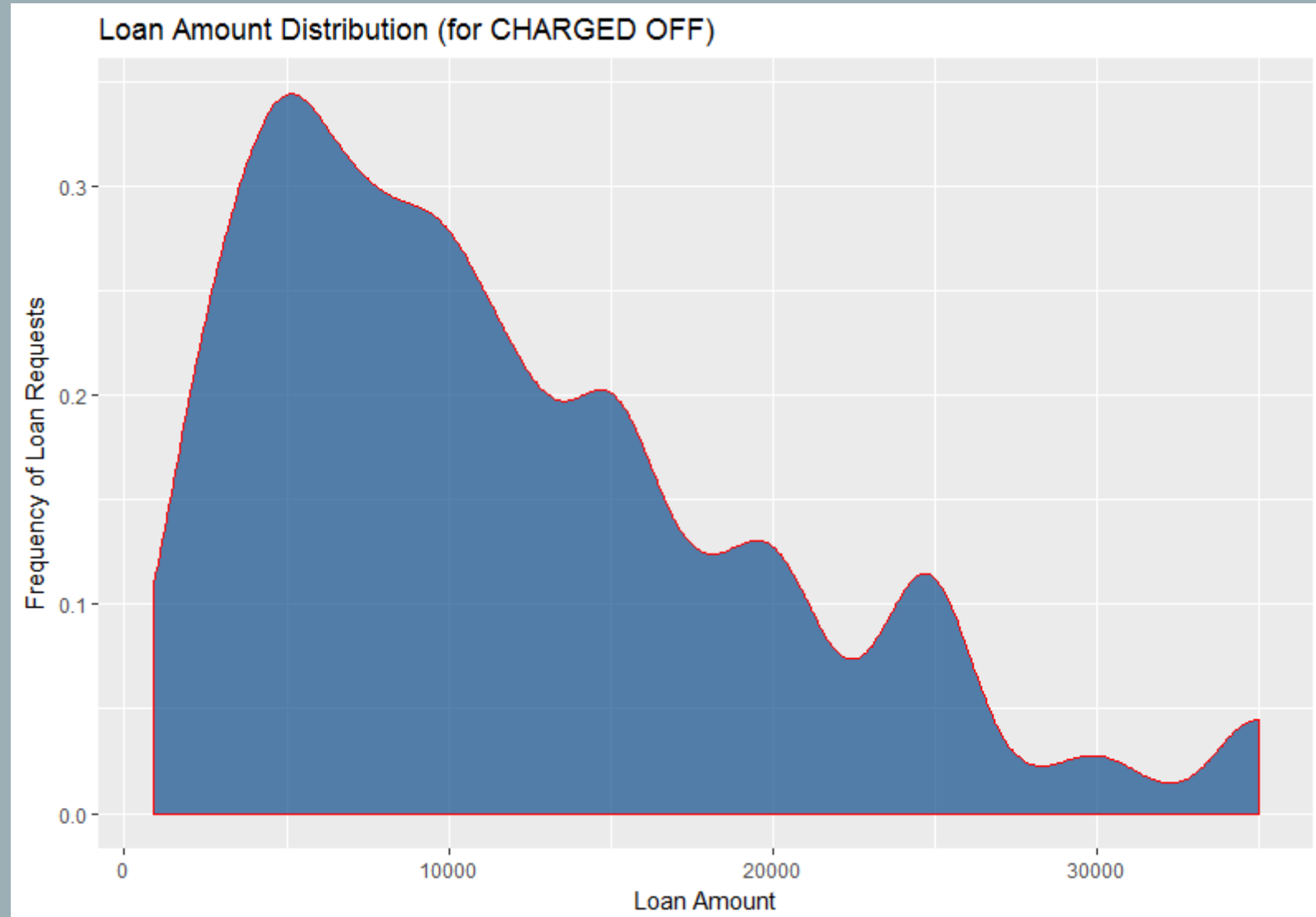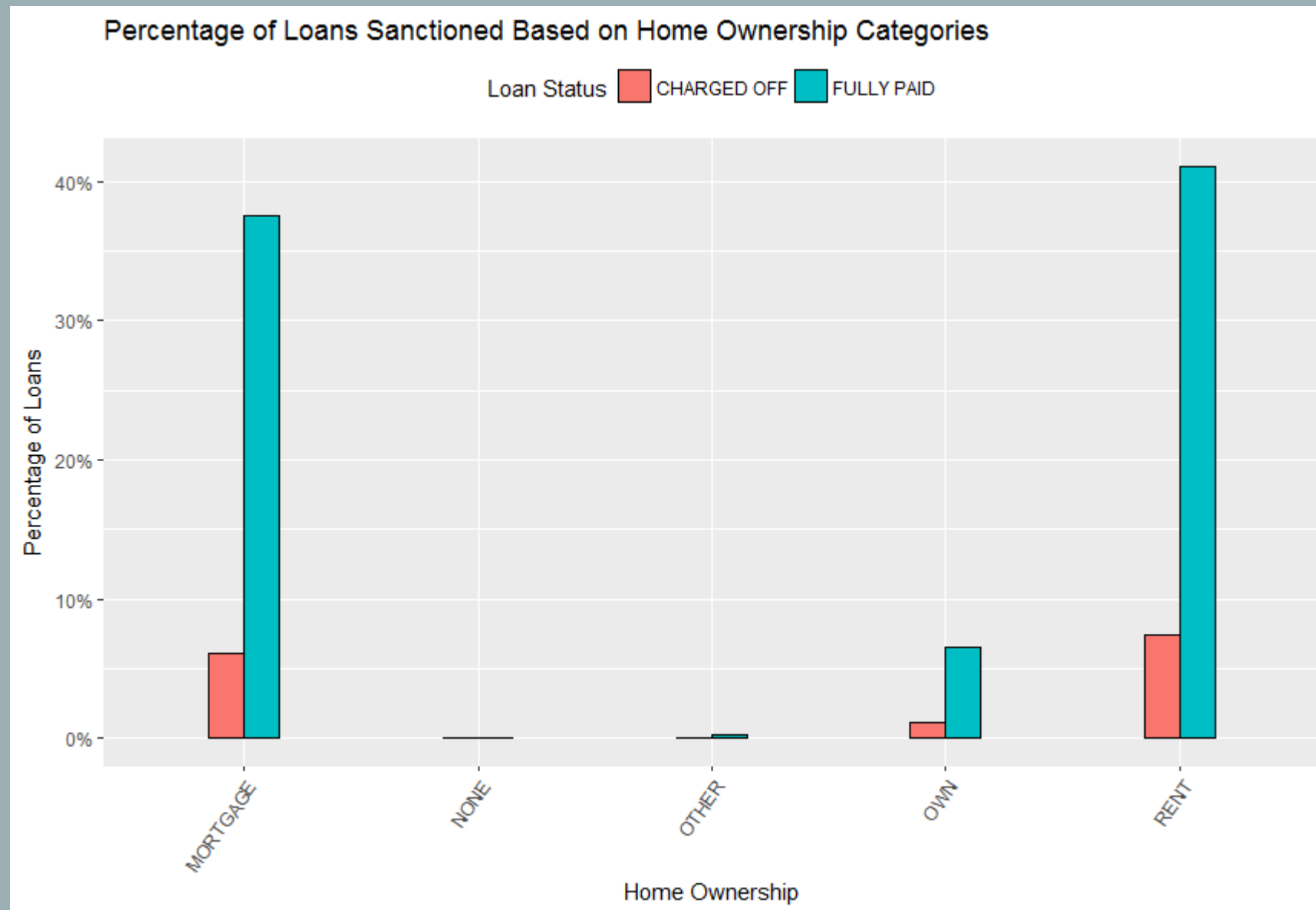
Segmented Univariate Analysis - Loan Purpose

o Based on Segmented Univariate Analysis on Loan Purpose categories, following loan purposes identified as highest risk factors

- o DEBT_CONSOLIDATION
- o CREDIT_CARD
- o HOME_IMPROVEMENT
- o SMALL_BUSINESS
- o OTHER

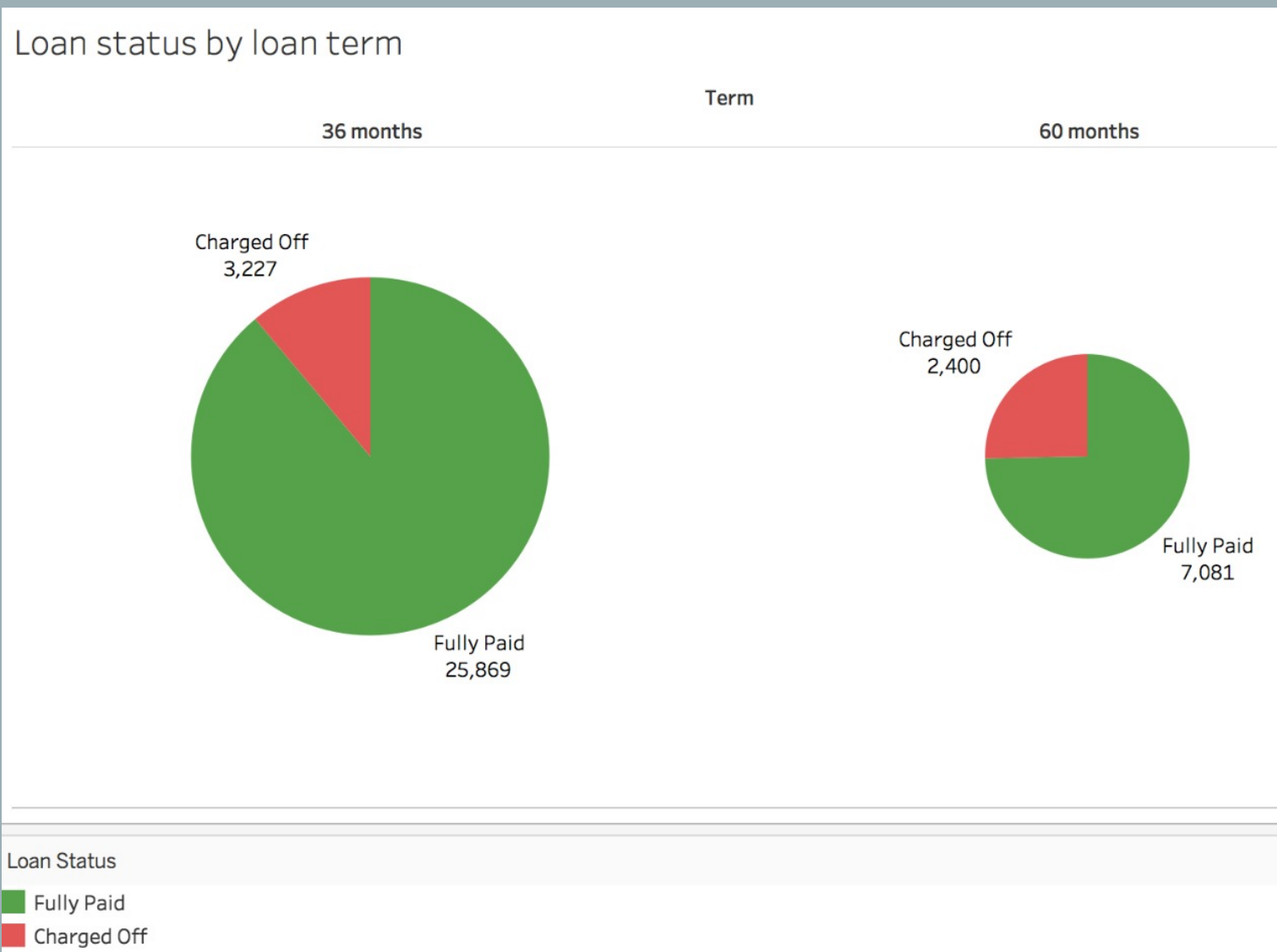Percentage of Loans Sanctioned Based on Purpose Categories

Segmented Univariate Analysis - Home Ownership

○ Based on Segmented Univariate Analysis on Home Ownership categories, borrower's without own home are prone to default on their loan compare to others.


Percentage of Loans Sanctioned Based on Home Ownership Categories

Segmented Univariate Analysis - Loan Term

o Based on Segmented Univariate Analysis on **Loan Term**, loan's having term as 60 months have highest risk of being default compare to 36 months.
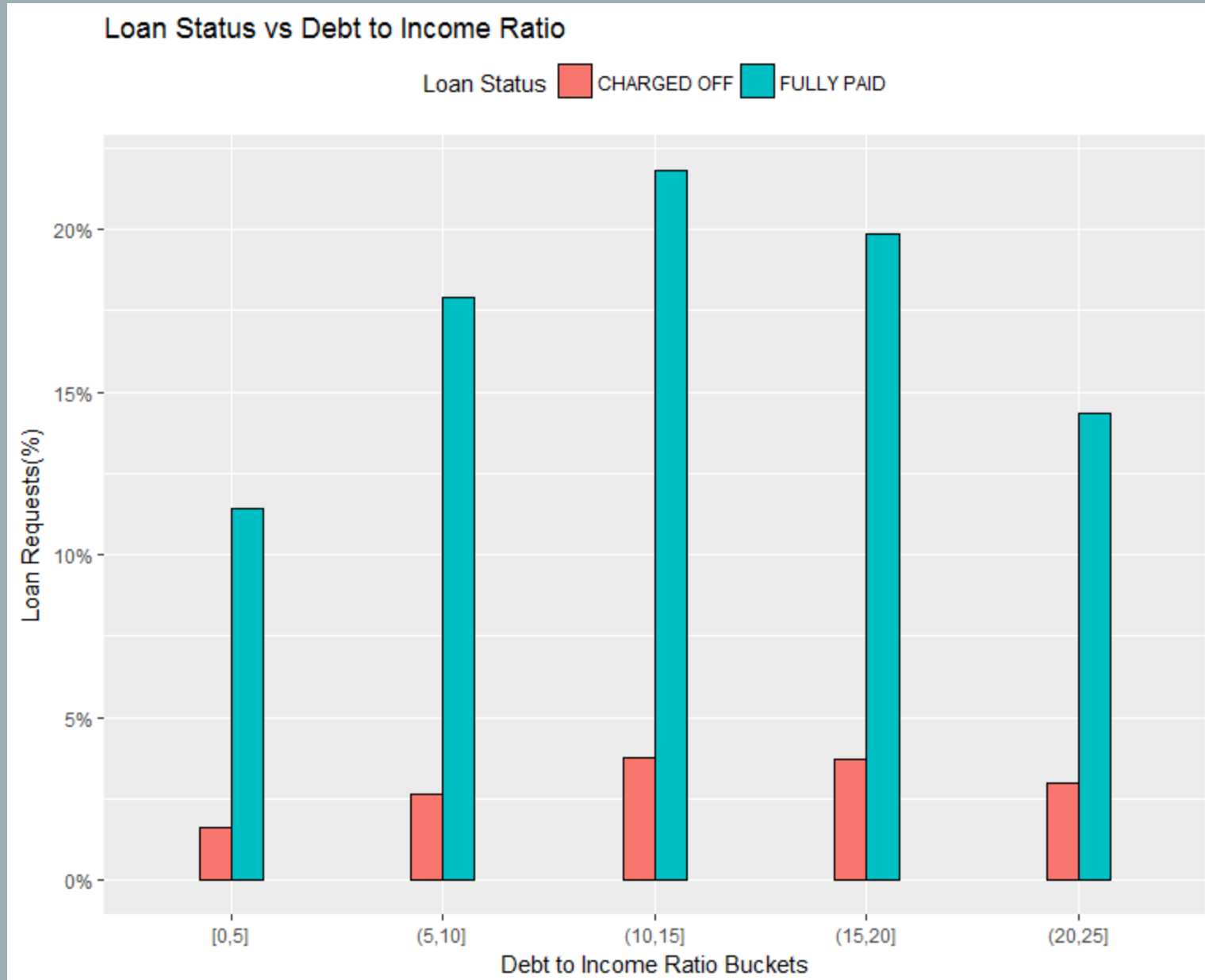
Bivariate Analysis

o Bivariate Analysis on **Loan Grade** and **Interest Rate** indicates that, number of loan requests consistently decreased as loan grade increases.

o And interest rate increases along with loan grade.
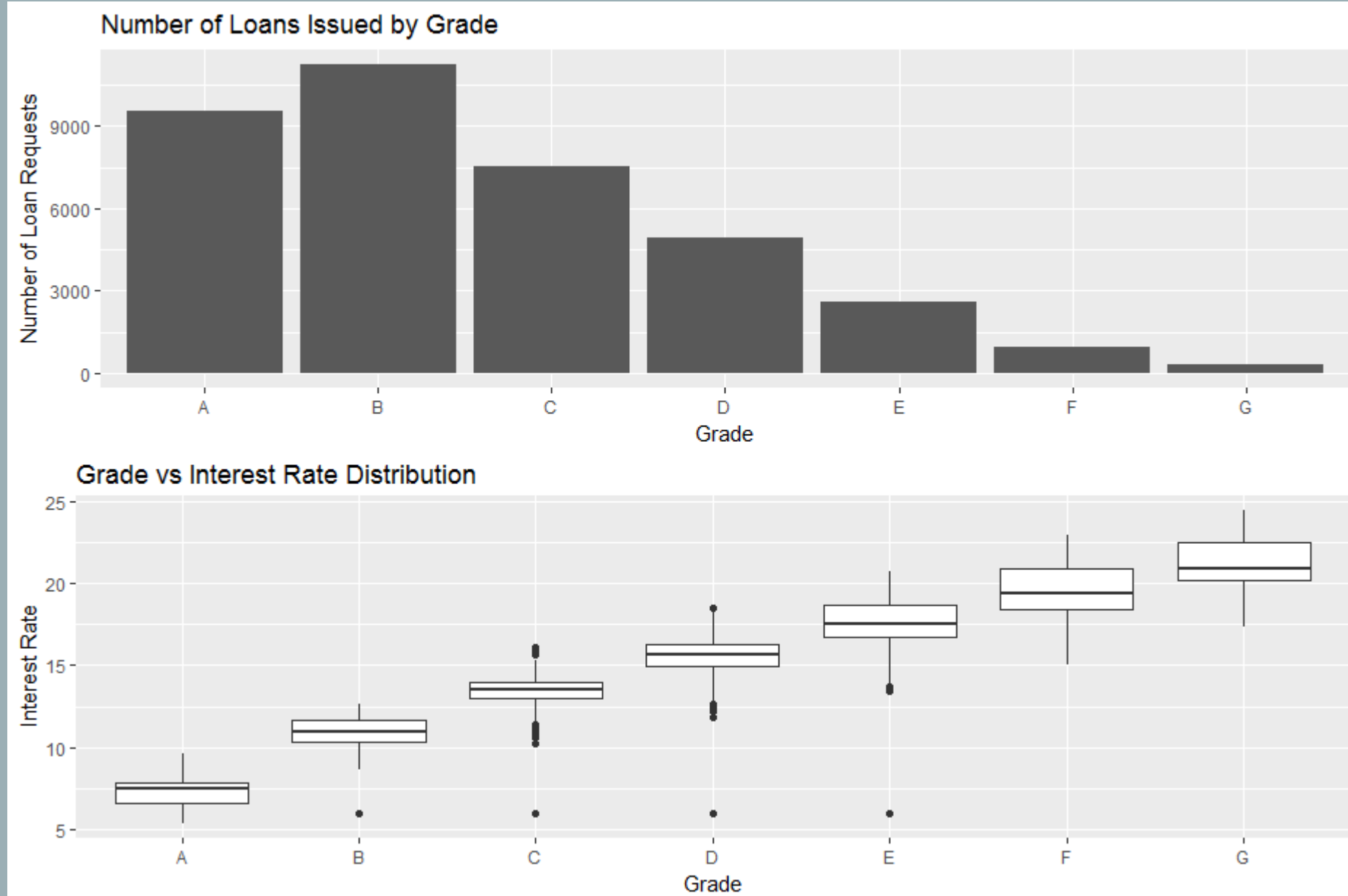
Bivariate Analysis

o Based on Bivariate Analysis on Debt To Income(DTI) and Loan Status, we observed a steady increase in the charged off loans ratio as DTI increases compare to fully paid loans.

Loan Status vs Debt to Income Ratio

Bivariate Analysis

o Based on Bivariate Analysis on Annual Income Buckets and Loan Status, we observed that, borrowers having annual income ranges from 20,000$ to 1,00,000$ are likely to be defaulters compare to other buckets.

Percentage of Loans Sanctioned Based on Annual Income Buckets
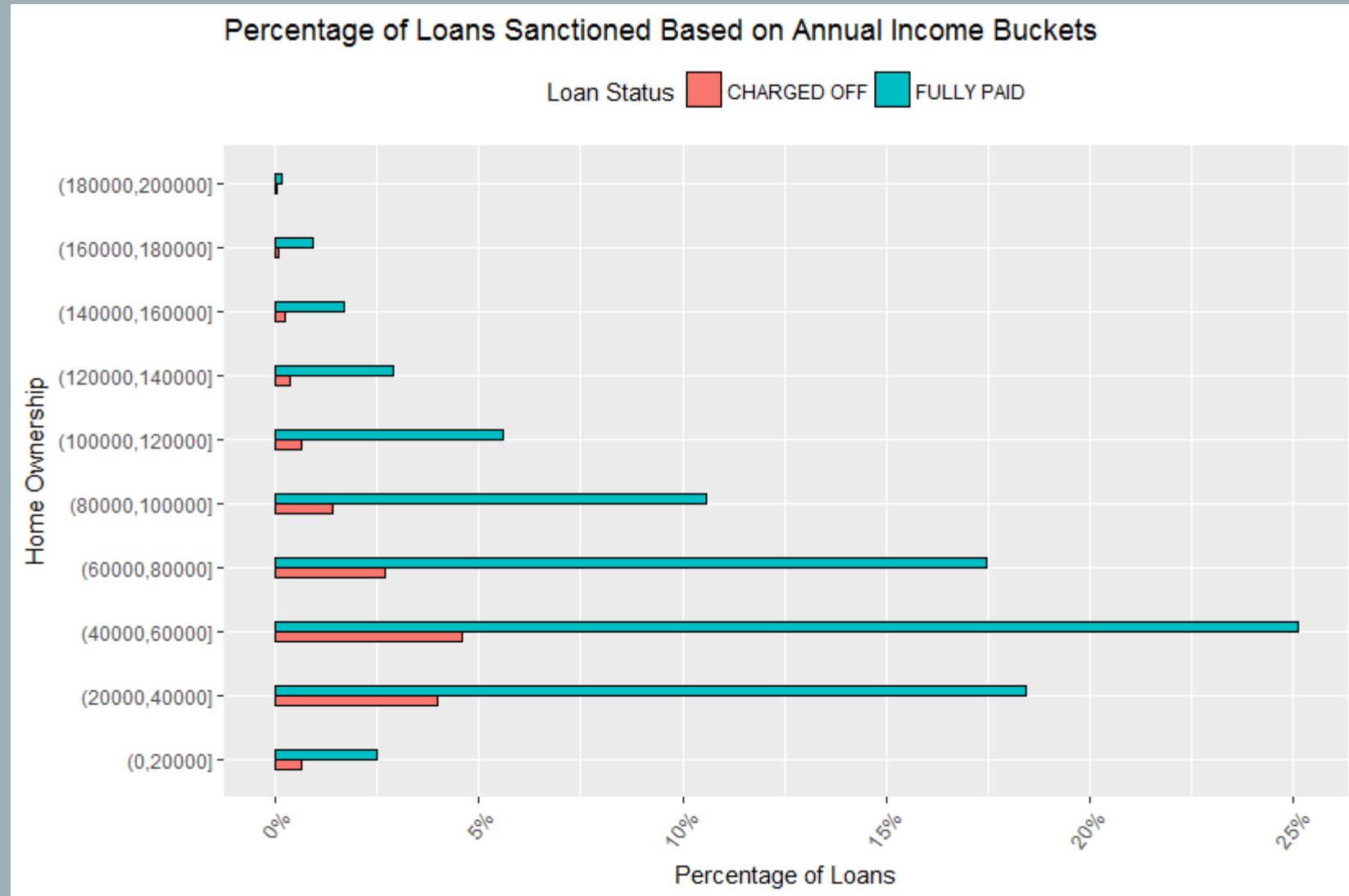
Conclusion

Borrower's with following factors are highly prone to be defaulters

- Home Ownership
    - Mortgage
    - Rent
- Loan Term selected as 60 months
- Loan Purpose
    - DEBT_CONSOLIDATION
    - CREDIT_CARD
    - HOME_IMPROVEMENT
    - SMALL_BUSINESS
    - OTHER
- Annual Income falling in range from 20,000$ to 1,00,000$
- Higher Debt to Income(DTI) ratio