STATISTICS CASE STUDY
(Loan Default)

SUBMISSION

Group Name : Quantitative-Soothsayers

1. Chayan Chatterjee
2. Devanshi Kulshreshtha
3. Hemanth Ponnada
4. Himanshu Srivastava

Loan Default Analysis

-- Exploratory Data Analysis to understand how the customer attributes & loan attributes influence tendency of loan defaults

Background

- **CFC** is a consumer finance company specialised in lending various types of loans to urban customers. The company approves or disapproves the applicant's loan based on their profile and loan's characteristics.
- Some customers pay their loan and some default on it. The company wants to know which type of customers are likely to default or pay their loan by studying past data trends

Objective

- The objective is to identify such applicant's profile who are often defaulting on their loans **plus** Identification of such loan types which attribute to more defaults

Benefits

- Minimization in risk of losing the money while approving or disapproving a loan

Problem solving methodology

Business & Data Understanding → **Data Preprocessing** → **EDA & Visualization** → **Conclusion & Recommendations**

Business & Data Understanding

- **CFC** is a consumer finance company specialised in lending various types of loans to urban customers. The company approves or disapproves the applicant's loan based on their profile and loan's characteristics
- Some customers pay their loan and some default on it. The company wants to know which type of customers are likely to default or pay their loan by studying past data trends

Data Preprocessing

- Data was sourced from UpGrad & cleaned –
  - Columns with 100% NA values were removed from data
  - Outliers were removed using Q1+1.5*IQR & Q1-1.5*IQR method
  - Missing Values were imputed & strings manipulations done
- New metrics were derived using methods like –
  - Type-driven
  - Business-driven
  - Data-driven

EDA & Visualization

- EDA performed using techniques like –
  - Univariate Analysis
  - Segmented Univariate Analysis
  - Bivariate Analysis
- Visualization was done by plotting various categorical & continuous variables
  - Histograms
  - Boxplots
  - Bar Charts
  - Proportion Stacked Bar Charts

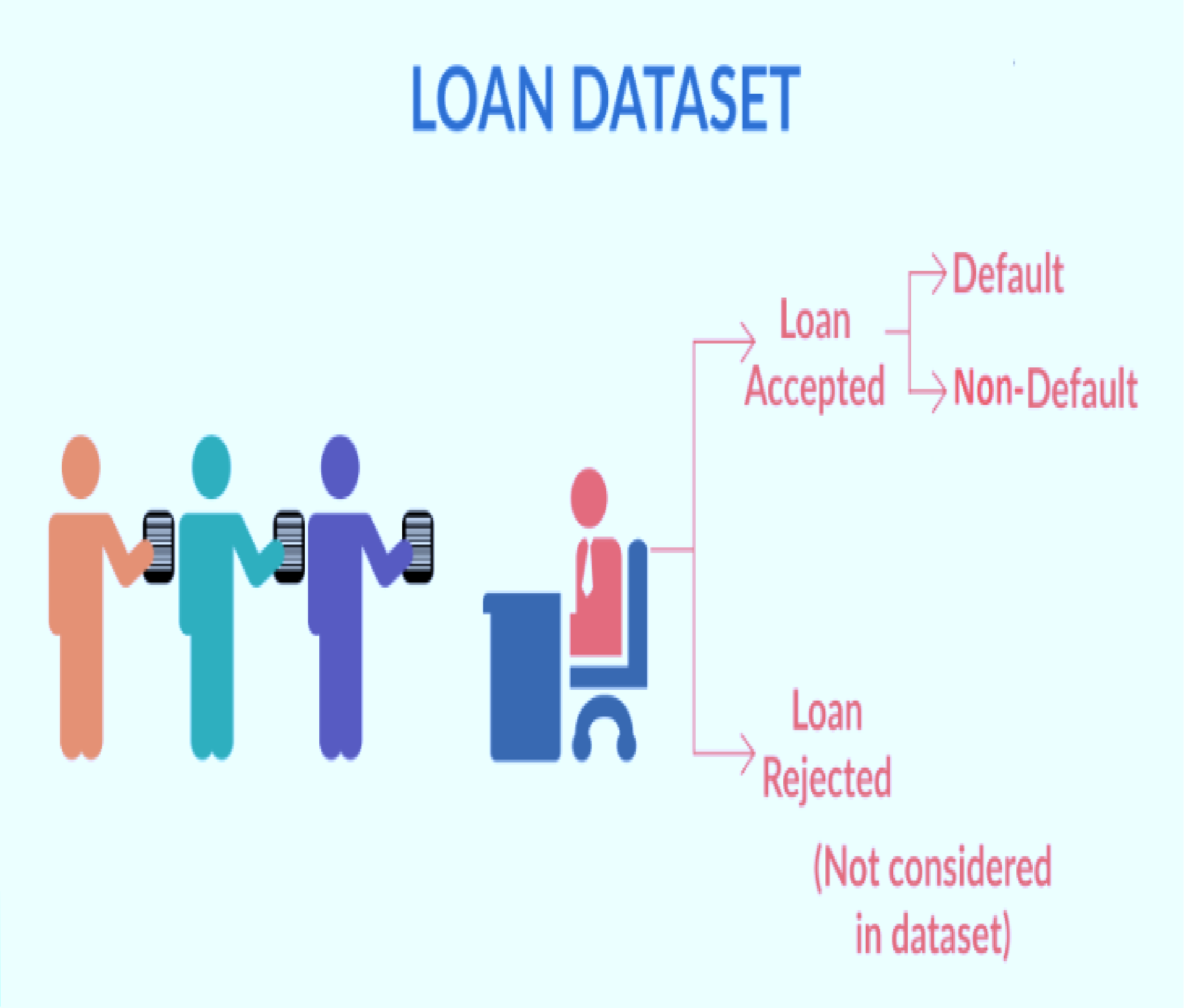
Conclusion & Recommendations

- Based on the insights extracted from various analysis and visualizations, conclusion were made about the driving factors behind loan defaults
- Recommendations were given to the loan sanctioning department of the company in order to minimize the risk while approving & disapproving the loans in future

Quick Summary

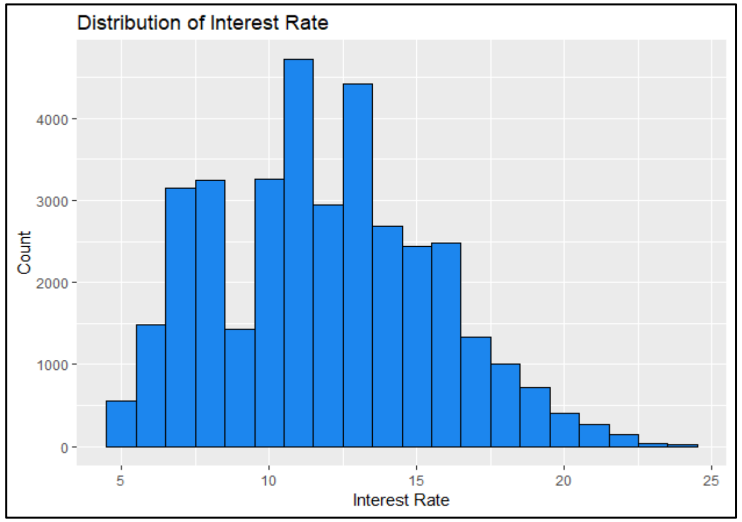
Key Take-Away --

- Total No. of Approved Loan Requests (2007-2011) **: ~40k**
- All the loan request **IDs** are **unique**
- Overall Loan Default **: ~14%**
- Number of Columns where all the values were **NA : 54**
- Average Income of the customers **:** ~**61K**
- Average loan amount **: ~10.5K**
- Percentage of loan having 3 Years term **: ~75%**
- Percentage of the loans taken for debt consolidation purpose **:** ~**50%**

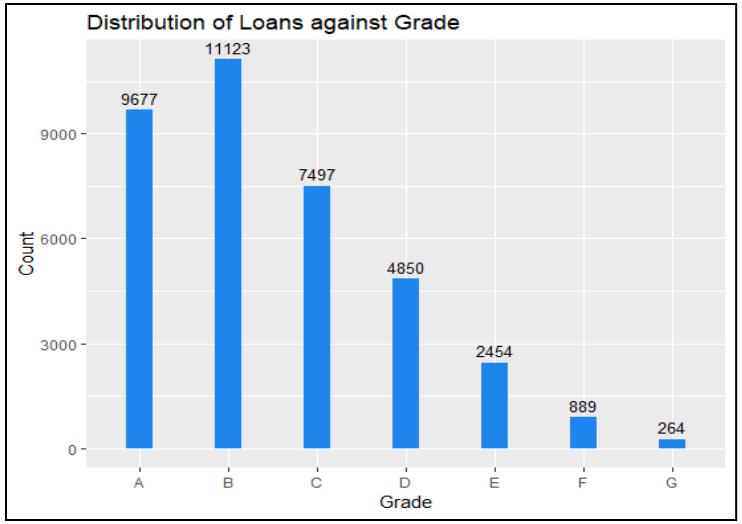


- ✓ Dataset Name : **loan.csv**
- ✓ Programming done in : **R**

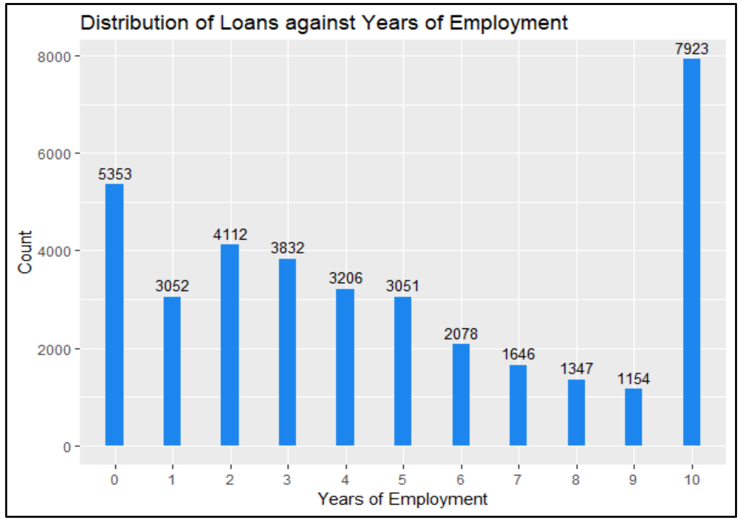
Univariate Analysis



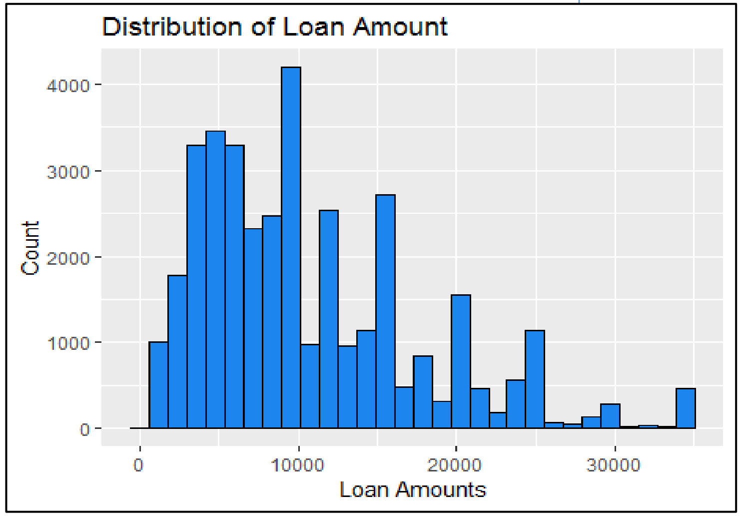
- The Interest rates for all the approved loans lies between a range of 5% & 25%
- Most of the loans were approved between an interest rate of 10% & 15%



- Most number of loans were approved under B grade segment
- Least number of loans were approved under G grade segment

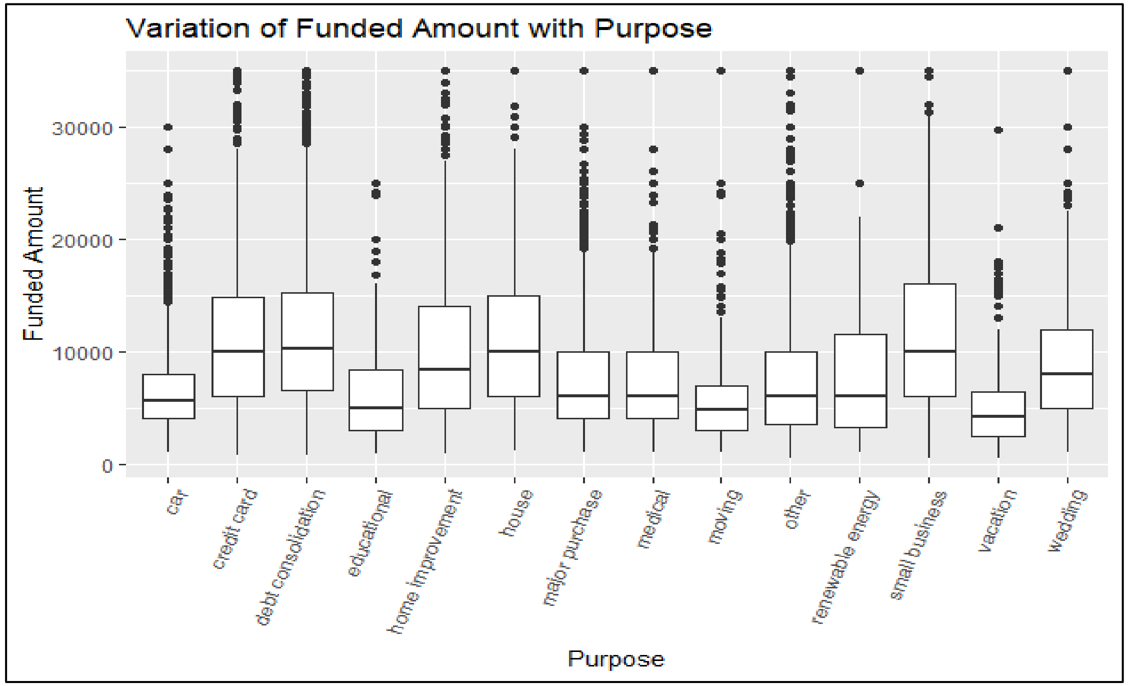


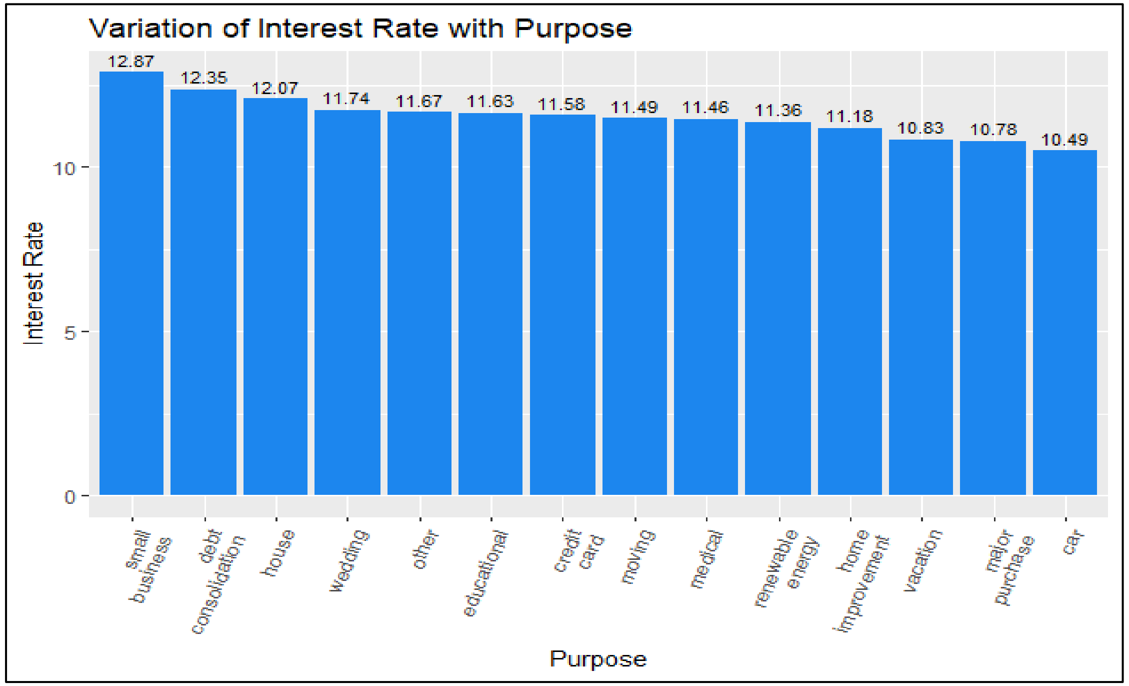
- There is no trend in years of employment and the number of loans given out.
- People with 10+ years of experience took the most number of loans, while those with less than 1 year of experience were second



- 50% of the loans were issued with an amount less than 15k dollars

Segmented Univariate Analysis (1/2)

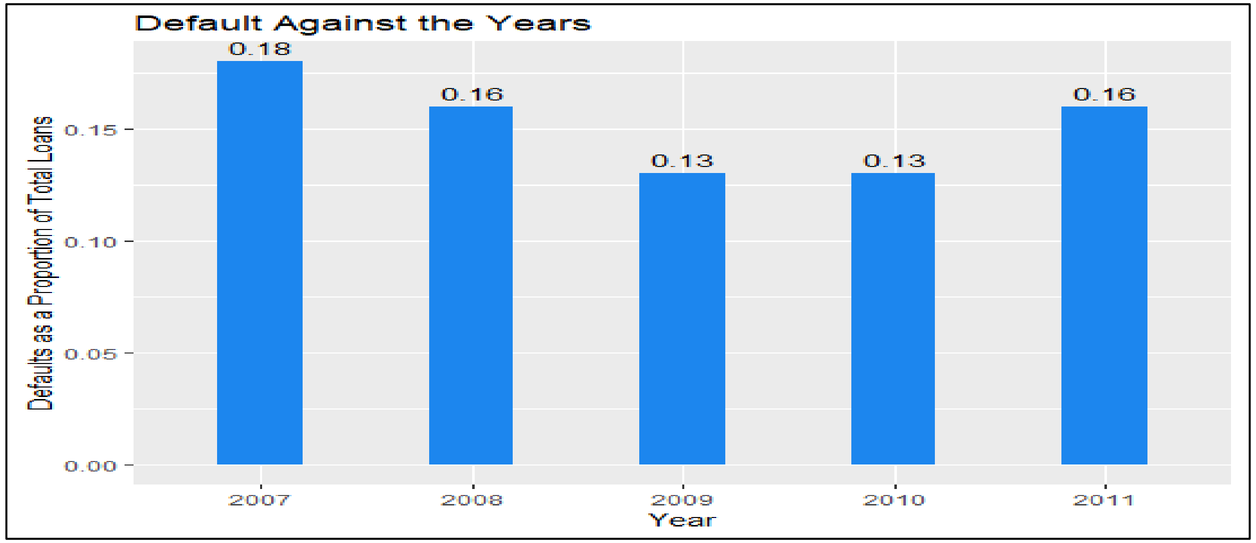




- Loan amount was the highest for loans taken for credit card, debt consolidation and small businesses.
- Vacation saw the least amount of loan.

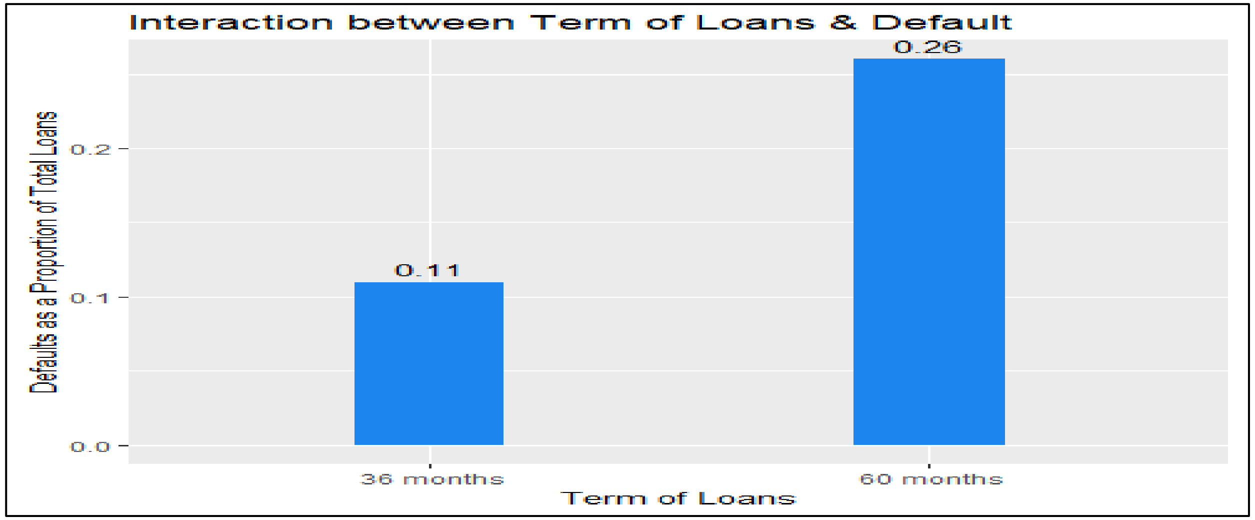
- Loans taken out for small businesses attracted the most interest rate, followed by debt consolidation and house.
- Car loans and major purchases saw the least interest rate.

Segmented Univariate (2/2)(Time Period Analysis)

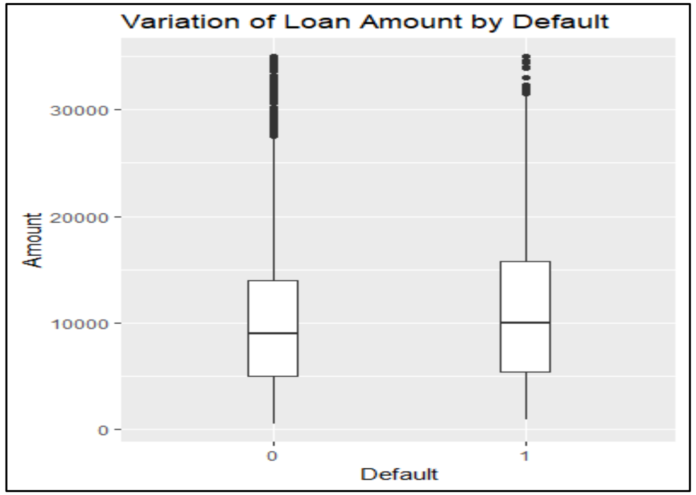


- A higher default rate (18%) was observed in 2007 followed by a dip in subsequent years 2008,2009 and 2010
- In 2011, again it spiked to 16%

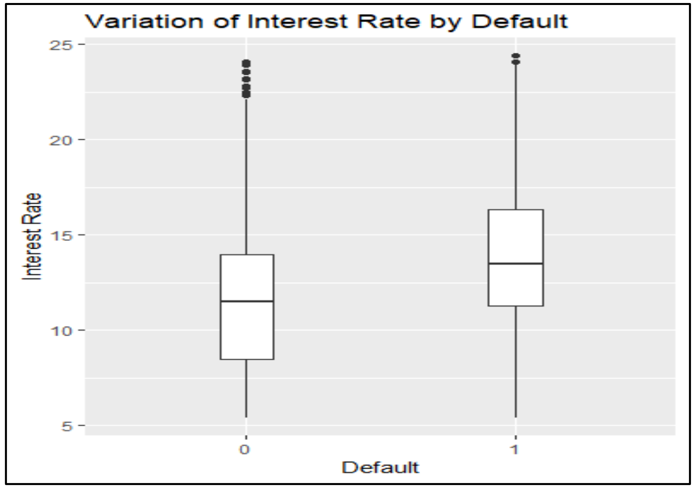
- Higher terms loans (60months) are nearly twice as likely to default as compared Lower terms (36 months)



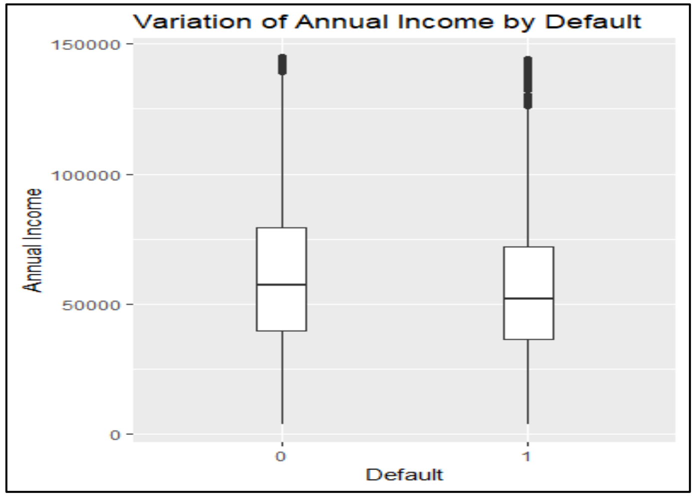
Bivariate Analysis (1/3) – Drivers of Loan Default



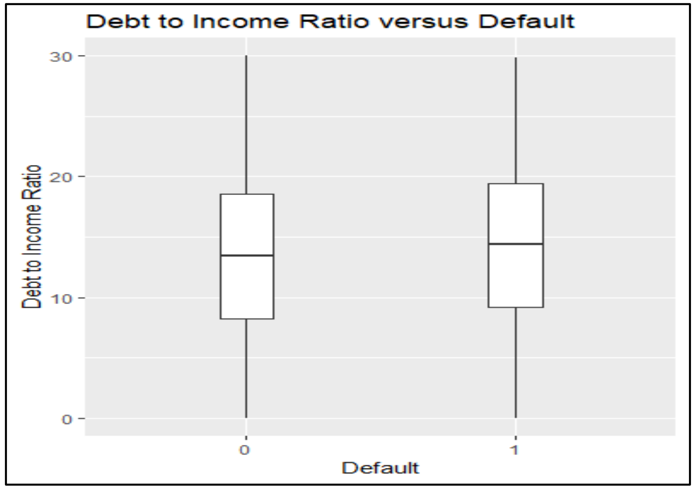
- Higher Proportion of loan defaults observed with higher median value of loan amount issued



- Loan Defaulters seem to have a higher median interest rate as compared to Non-defaulters

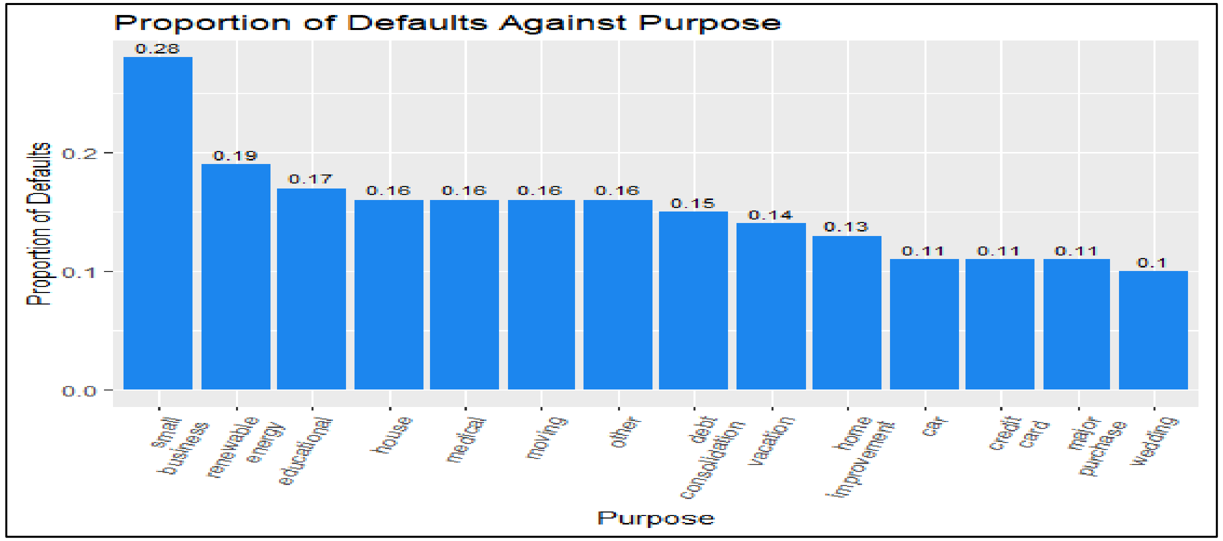


- Lower Median Annual Income drives loan defaults

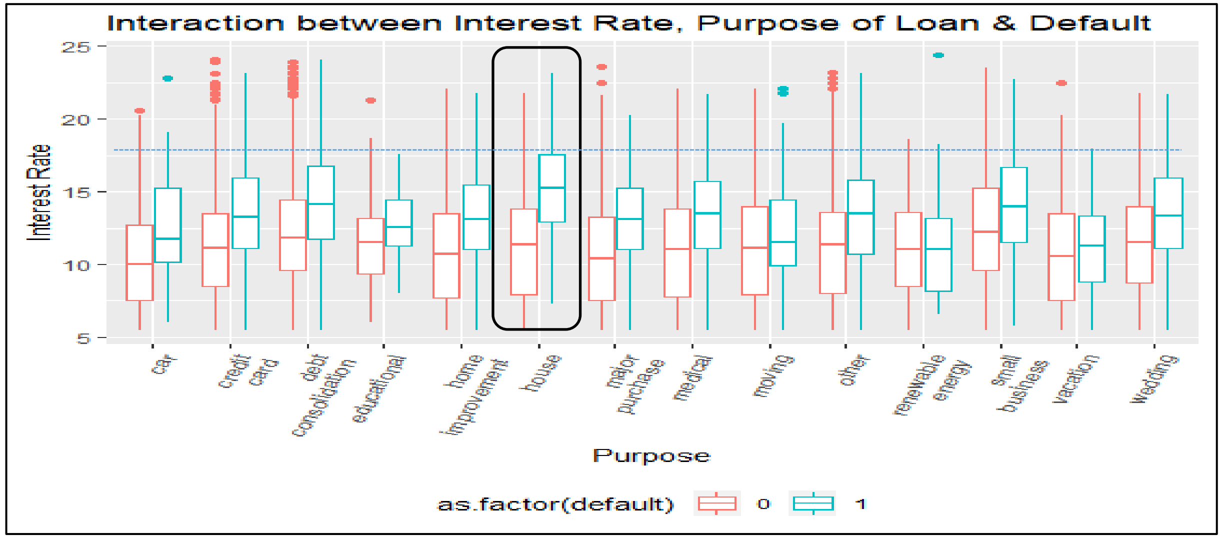


- Median of Debt to income ratio is higher amongst defaulters as compared to non-defaulters

Bivariate Analysis (2/3) – Drivers of Loan Default

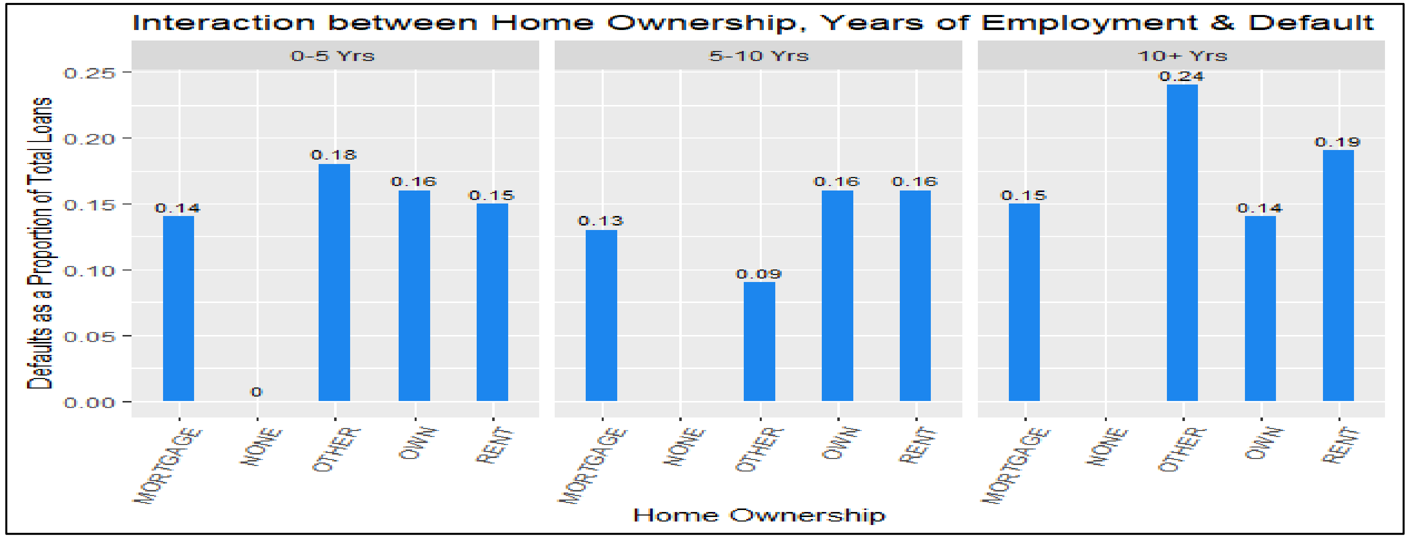


- **Small businesses** are most likely to default followed by **renewal energy**
- Loans taken for **wedding** purpose have the least likeliness of defaulting

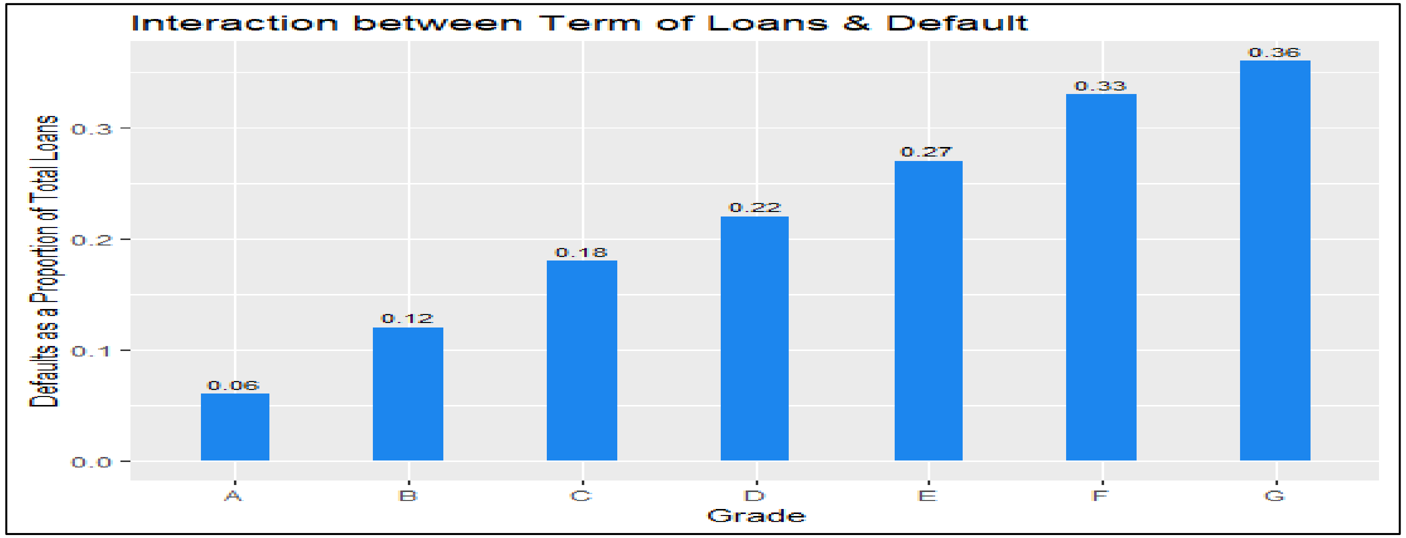


- Median **Interest rate** is higher for defaulters irrespective of the purpose for which the loan was taken
- Difference between Interest rate of defaulters and non-defaulters is highest for those who are taking the loan for a house

Bivariate Analysis (3/3) – Drivers of Loan Default



- Customers with 10+ Years of employment have highest proportion of defaults amongst all the Home Ownership categories except those who have a house of their own



- Proportion of defaults is maximum for **G** grade segment customers
- Those who fall under **A** grade segment defaults the least

Conclusion & Recommendations

Conclusion :

The key drivers of loan defaults [as per past trend (2007-2011)] were :

- Loan Amount
- Purpose of Loan
- Loan Interest Rates
- Year of Employment
- Debt to Income Ratio, Term of loan & Grade Segment
- Revolving Utilization

Recommendations :

The Consumer Finance Company while dispersing loans should take into account the following :

- Loan amount is lower than 10000$
- Interest rate is in the range 10-12
- Customers should have an income higher than 50000.
- Loans should be of a lower term (more 3 year loans than 5 year loans)
- The Debt to Income Ratio of customers should be checked to give out more loans to customers with a favourable ratio.
- Loan Grades are a good predictor of default, and should be used more often while giving out loans.