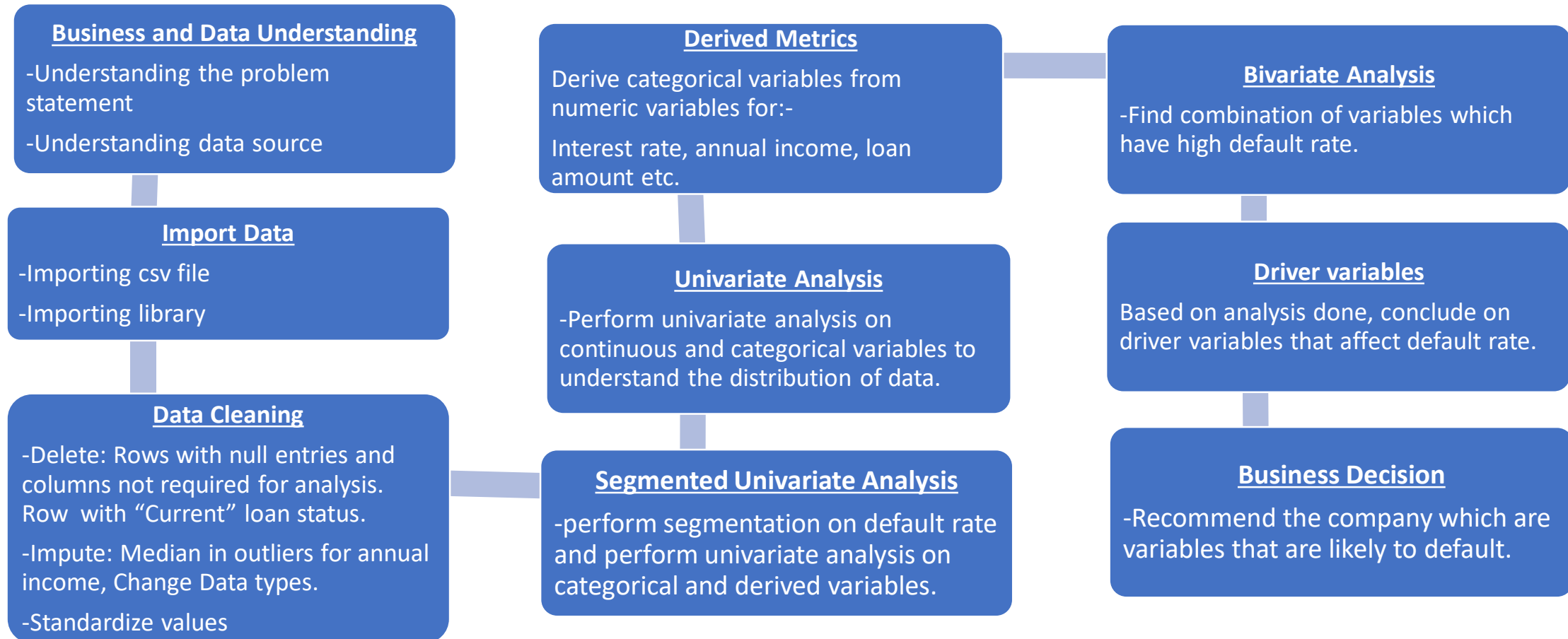# LENDING CLUB CASE STUDY

SUBMISSION

Name:

Amar Parsekar (Group-Facilitator)

Snehal Gupta

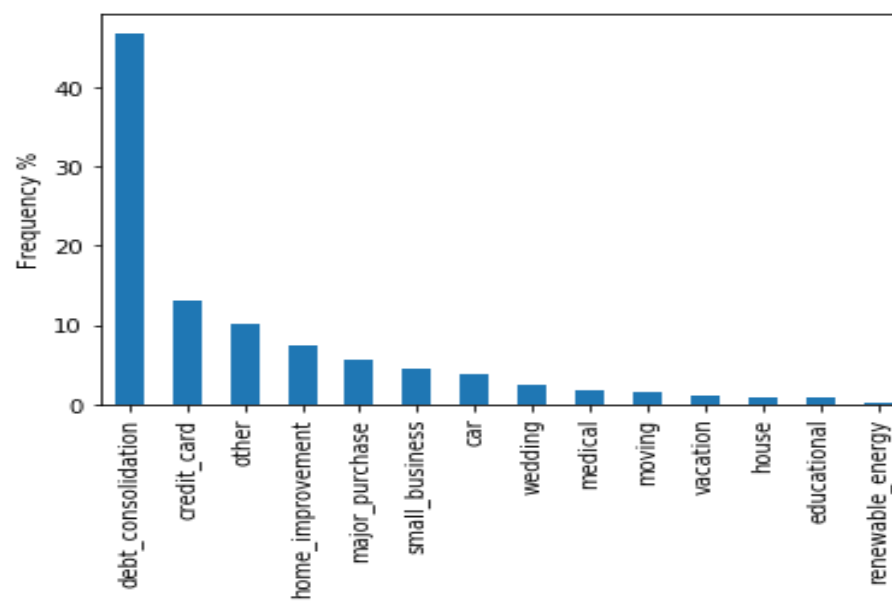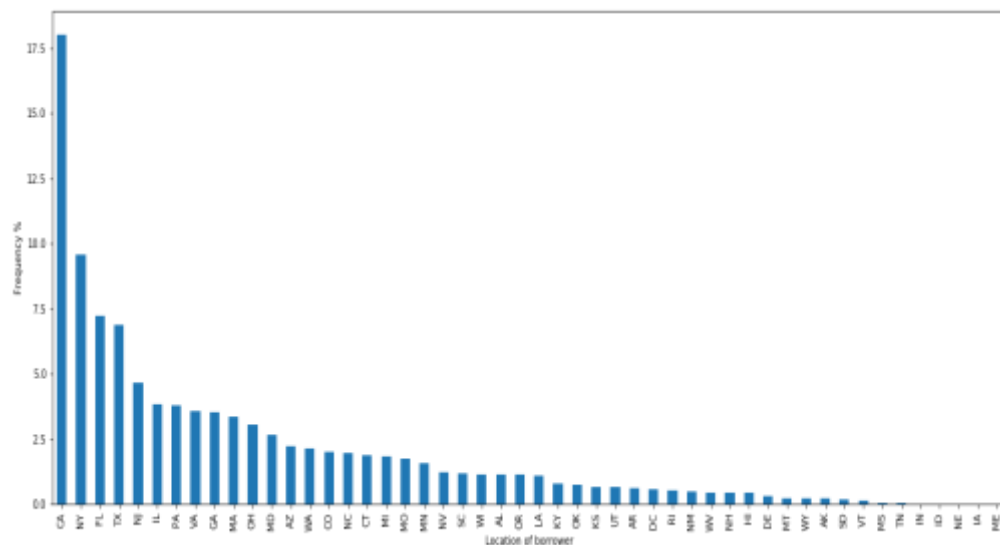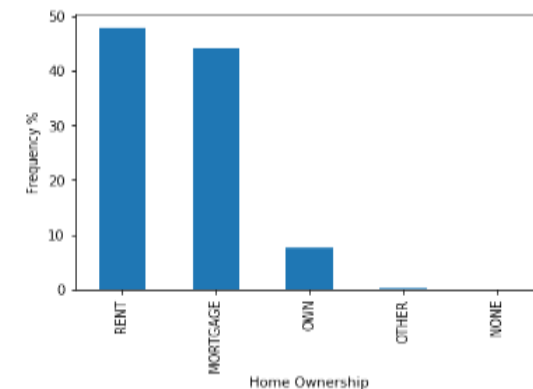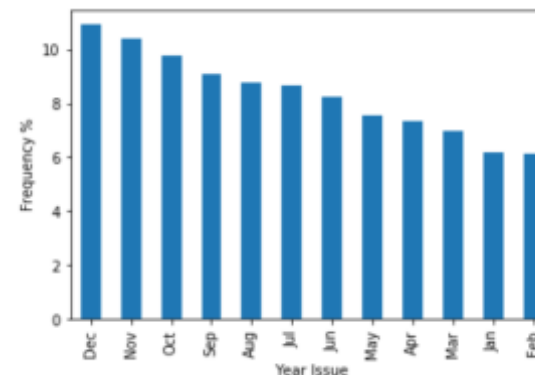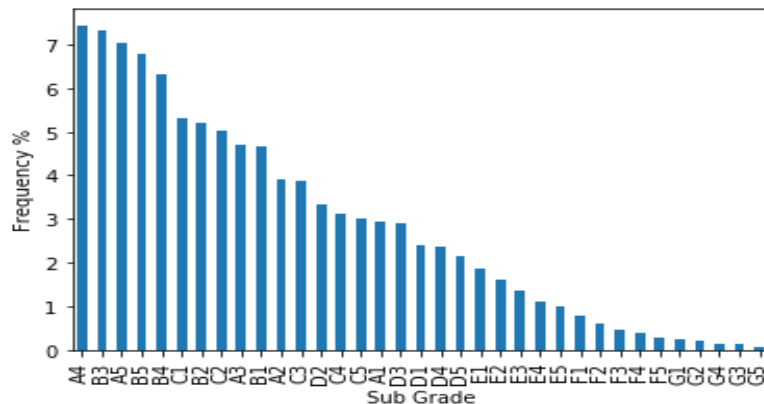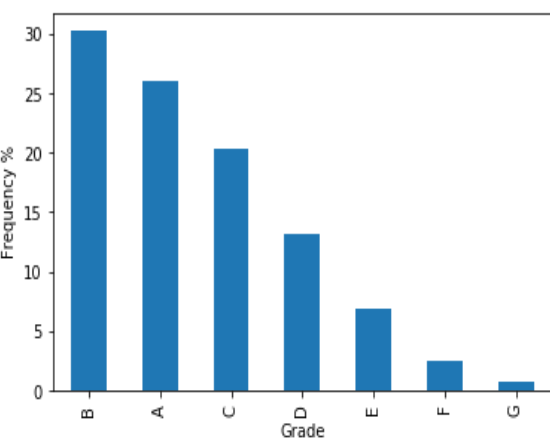# Problem Statement

➢ **Business Understanding:** 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.

• 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. **defaulted** on the loan

• **Loan rejected**: The company had rejected the loan

➢ **Business Objectives**

Identify driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Flow Chart

Business and Data Understanding

-Understanding the problem statement

-Understanding data source

Import Data

-Importing csv file

-Importing library

Data Cleaning

-Delete: Rows with null entries and columns not required for analysis. Row with "Current" loan status.

-Impute: Median in outliers for annual income, Change Data types.

-Standardize values

Derived Metrics

Derive categorical variables from numeric variables for:-

Interest rate, annual income, loan amount etc.

Univariate Analysis

-Perform univariate analysis on continuous and categorical variables to understand the distribution of data.

Segmented Univariate Analysis

-perform segmentation on default rate and perform univariate analysis on categorical and derived variables.

Bivariate Analysis

-Find combination of variables which have high default rate.

Driver variables

Based on analysis done, conclude on driver variables that affect default rate.

Business Decision

-Recommend the company which are variables that are likely to default.

Univariate Analysis-Categorical Variables

Observations:
- Majority applicants belong to CA state
- Majority of applicants have taken loan for debt consolidation
- Majority of applicants belong to grade B
- Majority of applicants belong to sub_grade A4.
- More than 40% applicants have rented home or mortgage
- Majority loan were approved at the end of year i.e. Dec.

Univariate Analysis-Continuous Variables
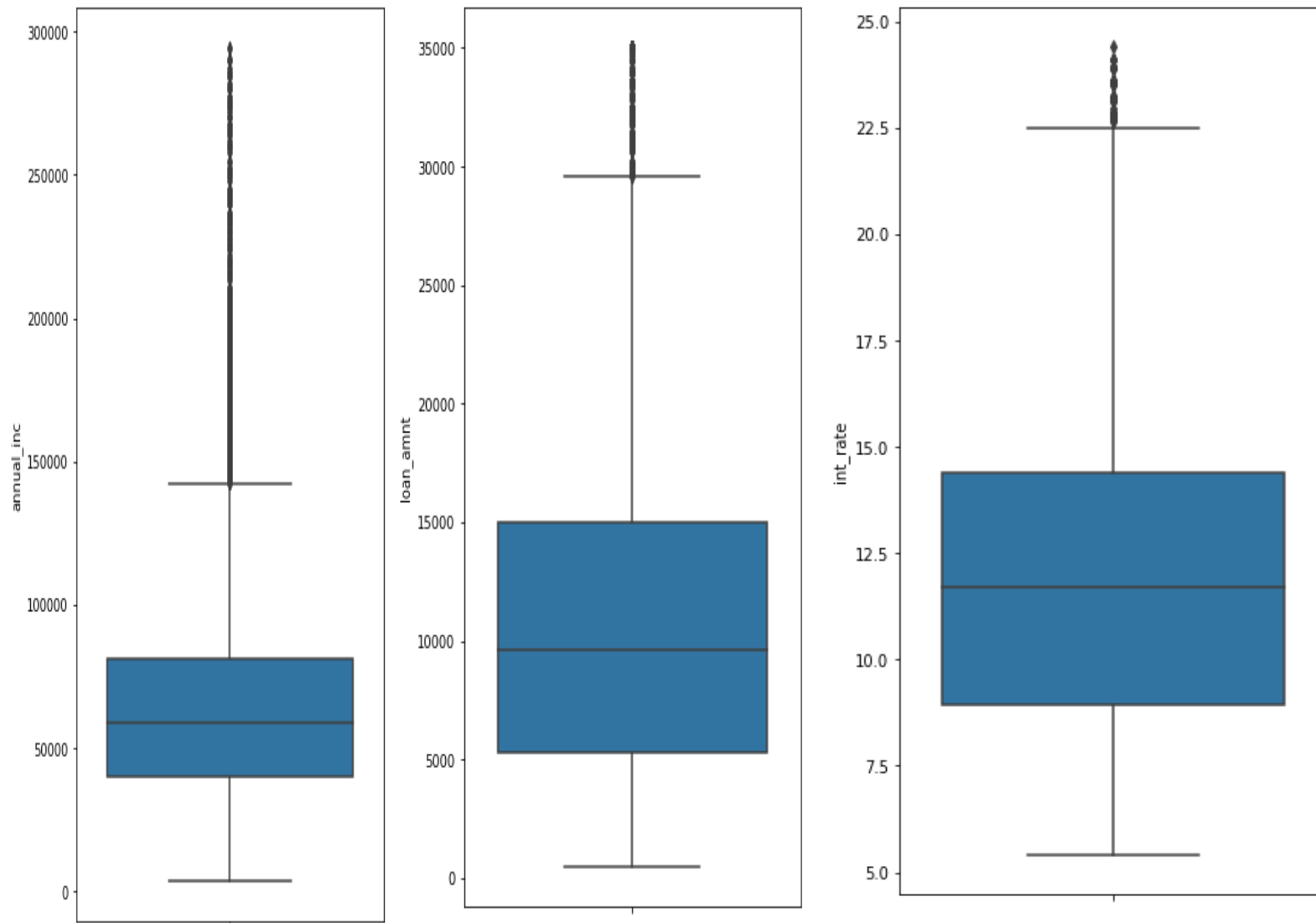
Observations:
- Majority of applicants have taken loan at 10 to 15% interest rate.
- Most of the loan amounts are distributed between 8000 to 20000 USD.
- Most of the applicants earns between 40000 to 90000 USD annually.
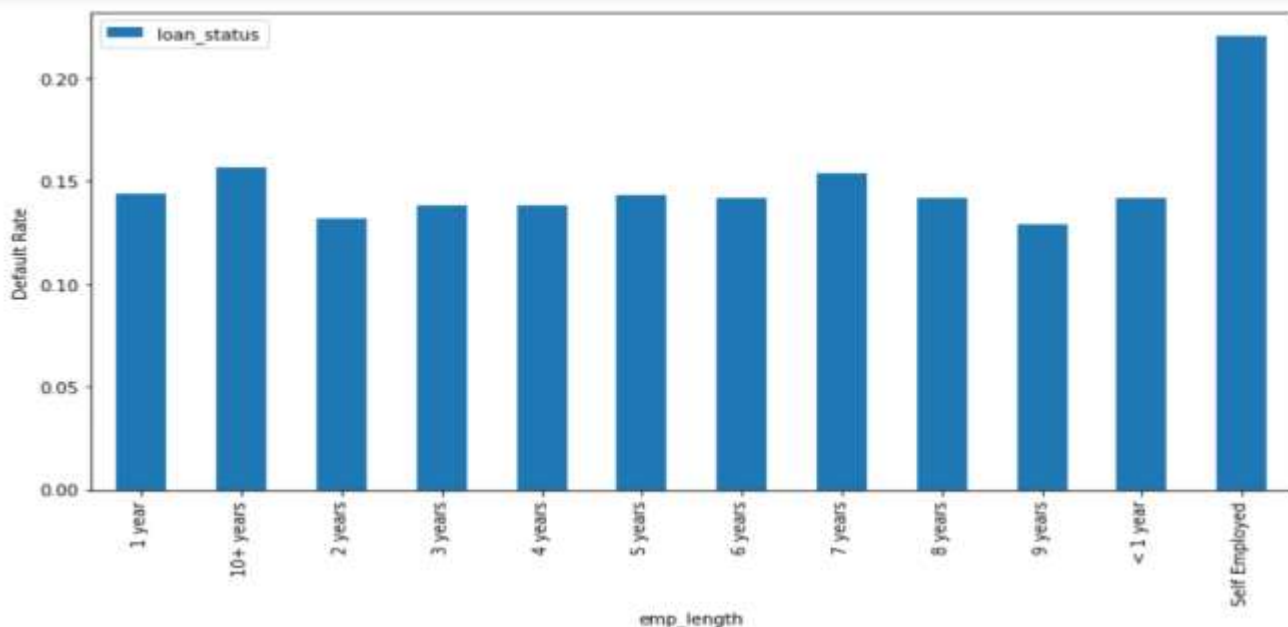
Univariate Segmented Analysis

Strategy: We performed univariate segmented analysis on demographic and loan characteristics.

- Segmentation is done on default rate.

- Default Rate= $\dfrac{\text{Number of defaults(Charged-Off)}}{\text{Number of applicants}}$
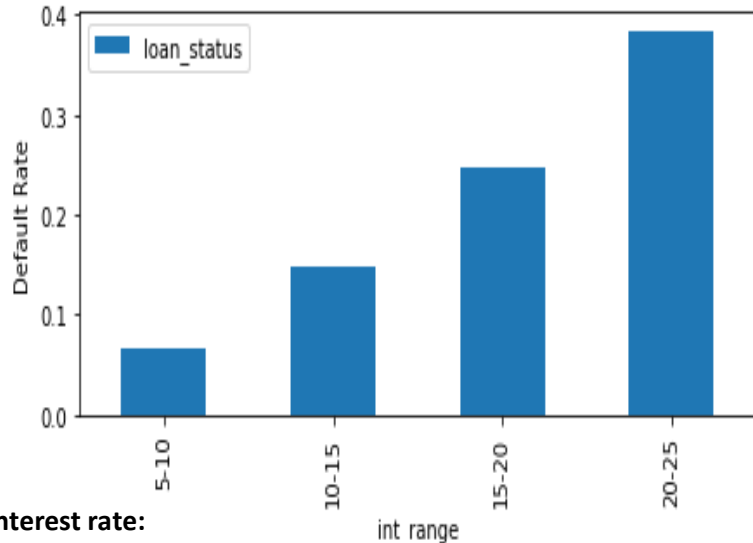
Plots:

Default Rate for Employee length

Employee length:
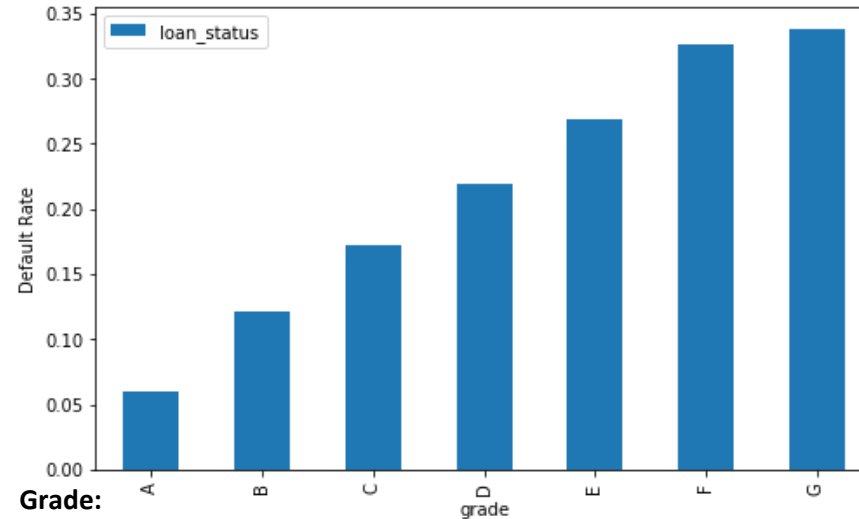Applicants who self employed are more likely to default.
Assumption: For employees whose employment type is n/a is assumed as self employed
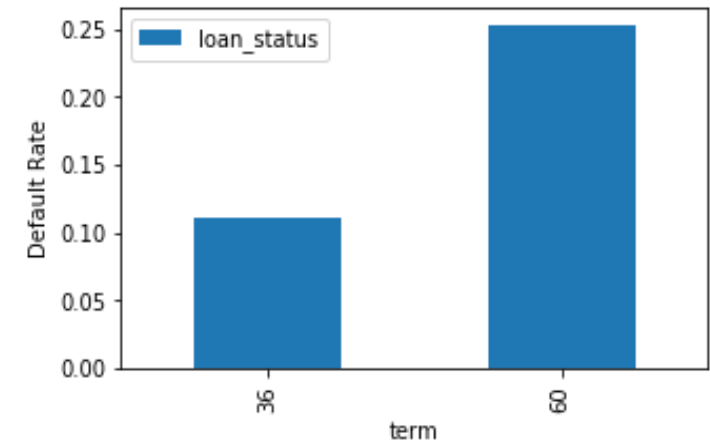
Default Rate for Interest Rate range

Interest rate:
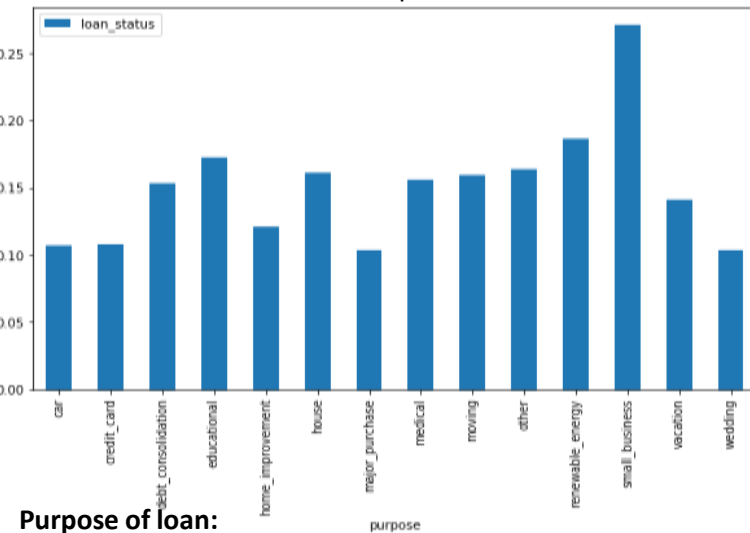As interest rate increases default rate increases.

Default Rate for Grade

Grade:
Grade G applicants are more likely to default. Default Rate increases from grade A to G.
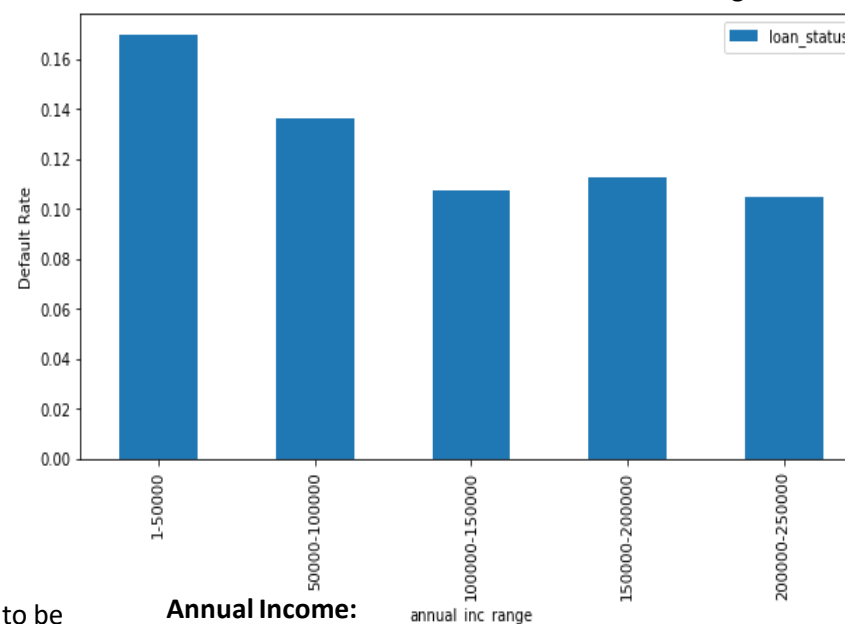
Default Rate for Term

Term:
Default rate is high if loan is taken for longer term.

Default Rate for Purpose

Purpose of loan:
Applicants who have taken loan for small business are more likely to be defaulters.

Default Rate for Annual Income Range

Annual Income:
Applicants who have annual income less than 50000 are more likely to default.

Bivariate Analysis

Strategy: We analysed various combinations of two variables against default rate and obtained top 3 combinations based on Number of defaulters greater than 1500 for the combination.

Assumption: Number of defaulters are considered because default rate is 1 even when there is only 1 defaulter applicant for the combination. Eg. Suppose Grade D3 and term 36mon has default rate 1. It is possible that there is only 1 applicant with this combination but default rate is still high hence number of defaulters is considered to extract top 3 combinations.

Plots:
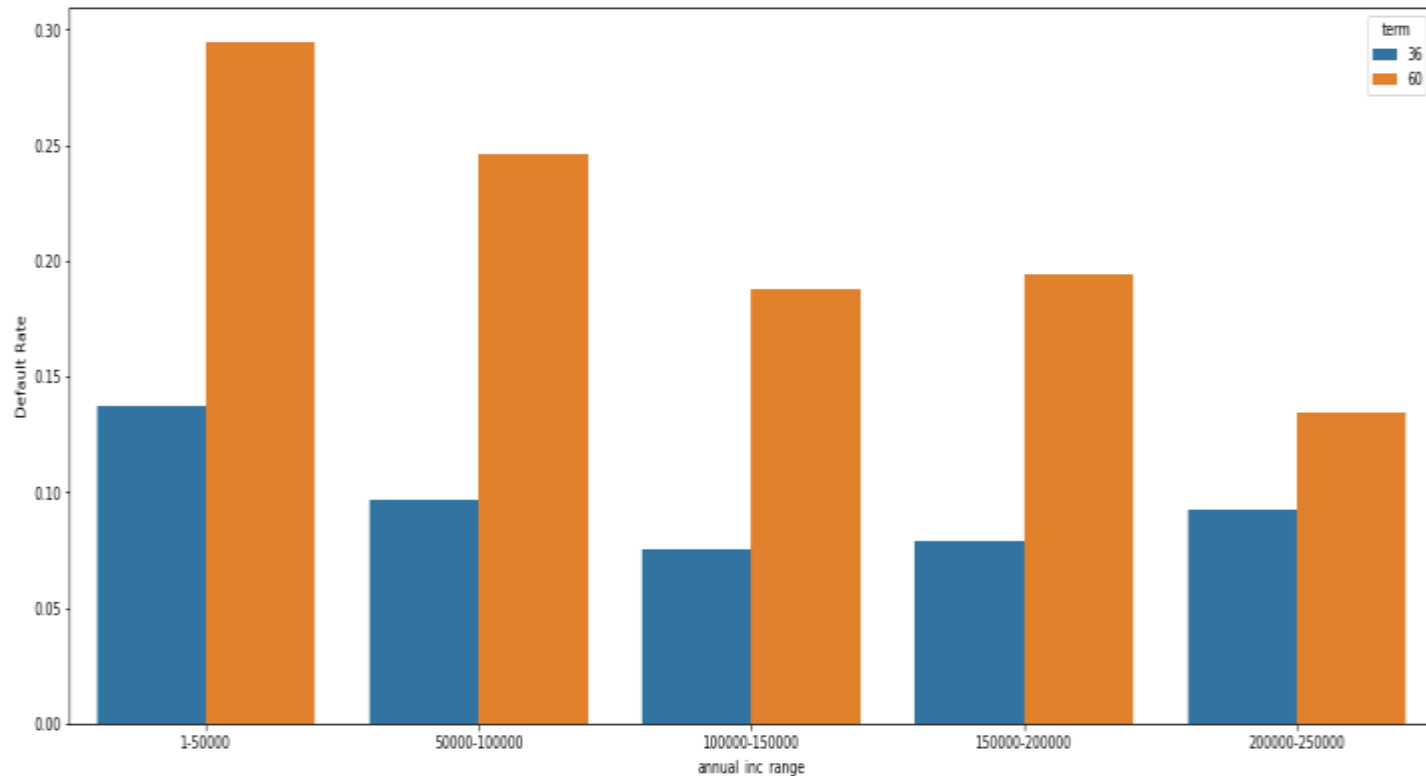
Default Rate for Home ownership vs annual income
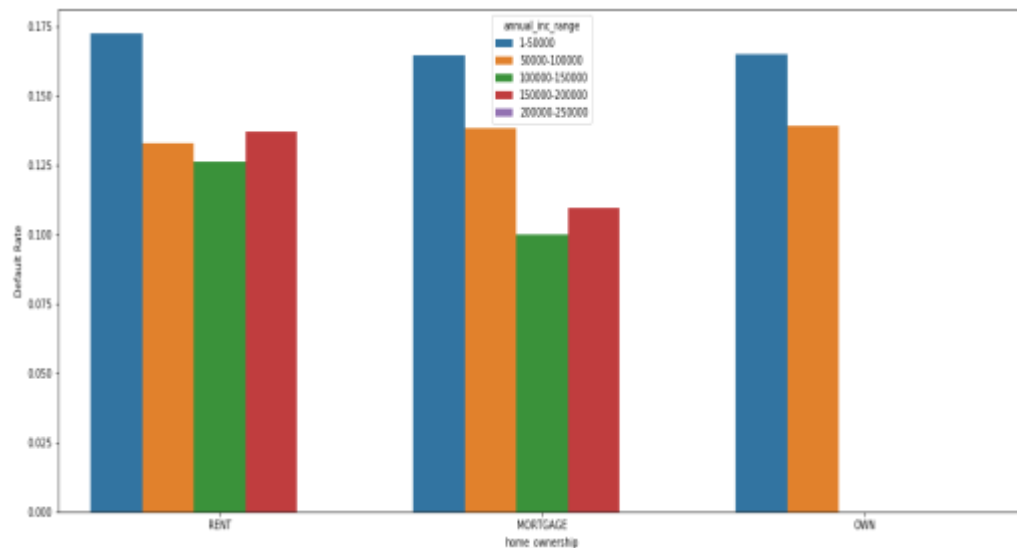
Top 10 Defaulters for Home ownership vs annual income

home_ownership	annual_inc_range	Default Rate	No. of Defaulters
RENT	1-50000	0.173	1669
MORTGAGE	50000-100000	0.139	1289
RENT	50000-100000	0.133	999
MORTGAGE	1-50000	0.165	669
MORTGAGE	100000-150000	0.100	276
OWN	1-50000	0.165	268
OWN	50000-100000	0.139	148
RENT	100000-150000	0.126	128
MORTGAGE	150000-200000	0.110	69
RENT	150000-200000	0.137	31

Observations:
- Applicant with low income and rented home or mortgage are more likely to default.
- Annual income range is less than 50000 and default rate is around 0.173
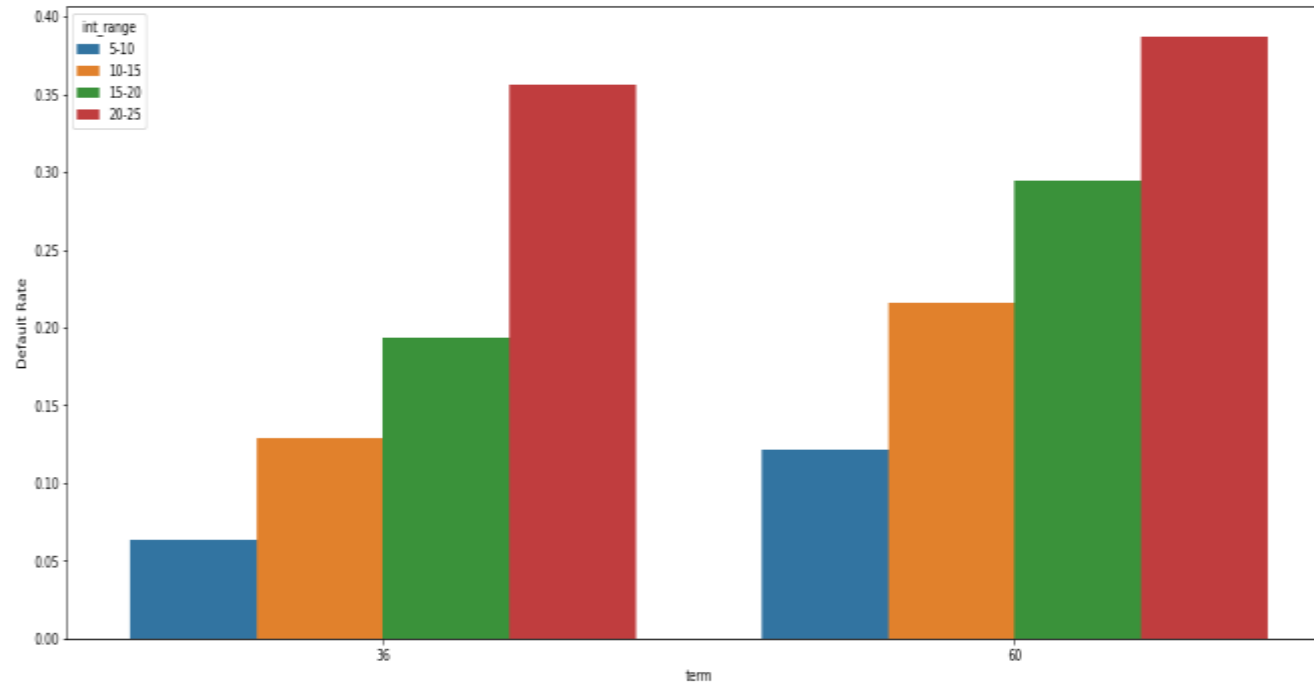
Default for Rate Annual Income vs Term

Top 10 Defaulters for Annual Inc Range vs Term

annual_inc_range	term	Default Rate	No. of Defaulters
1-50000	36	0.137	1673
50000-100000	36	0.096	1267
50000-100000	60	0.246	1179
1-50000	60	0.295	940
100000-150000	36	0.075	215
100000-150000	60	0.188	214
150000-200000	60	0.194	52
150000-200000	36	0.079	51
200000-250000	36	0.092	18
200000-250000	60	0.134	11

Observations:
- Applicant with low income and loan taken for longer term are likely to default.
- Default rate is around 0.137 .
- Applicant who have taken loan for 60mon terms across all income brackets are likely to default compare to loan taken for shorter duration.

Default Rate for term vs interest rate

Observations:
- Applicant who have taken loan at high interest rate and for long term are likely to default.
- Interest rate is around 20-25% , term is 60mon and default rate is around 0.129

Top 10 Defaulters for Term vs Interest

term	int_range	Default Rate	No. of Defaulters
36	10-15	0.129	1828
60	15-20	0.294	1157
60	10-15	0.216	879
36	5-10	0.064	736
36	15-20	0.193	637
60	20-25	0.387	270
60	5-10	0.121	94
36	20-25	0.356	26

Driver Variables & Recommendations

Driver Variables:-

Univariate & Segmented Univariate analysis- Purpose, Grade, Interest rate, Loan amount, Annual Income, Employee length, Term, Home ownership

Bivariate Analysis-Term vs Annual Income, Home ownership vs Annual Income, Term vs Interest rate range

Recommendations:

Lending club should take care while approving loan for applicants with any of the below variables as there are chances of default:-

- **Grade G** applicants are more likely to default. Default Rate increases from grade A to G.

- Applicants who have taken loan for **small business** are more likely to be defaulters.

- Applicants who have low annual income (**less than 50000**) are more likely to default.

- Applicants who **self employed or employment type is not specified** are more likely to default.(**Assumption**: For employees whose employment type is n/a is assumed as self employed)

- As interest rate increases default rate increases. Company needs to be careful while considering loan application at **20% and above interest rate**.

- Default rate is high if loan is taken for longer term. Company needs to be careful while considering loan application for **60 months**.

- Applicant with low income (**less than 50000**) and **rented** home or **mortgage** are more likely to default.

- Applicant with low income (**less than 50000**) and loan taken for longer term (**60mon**) are likely to default.

- Applicant who have taken loan at high interest rate (**20% & above**) and for long term(**60mon**) are likely to default.