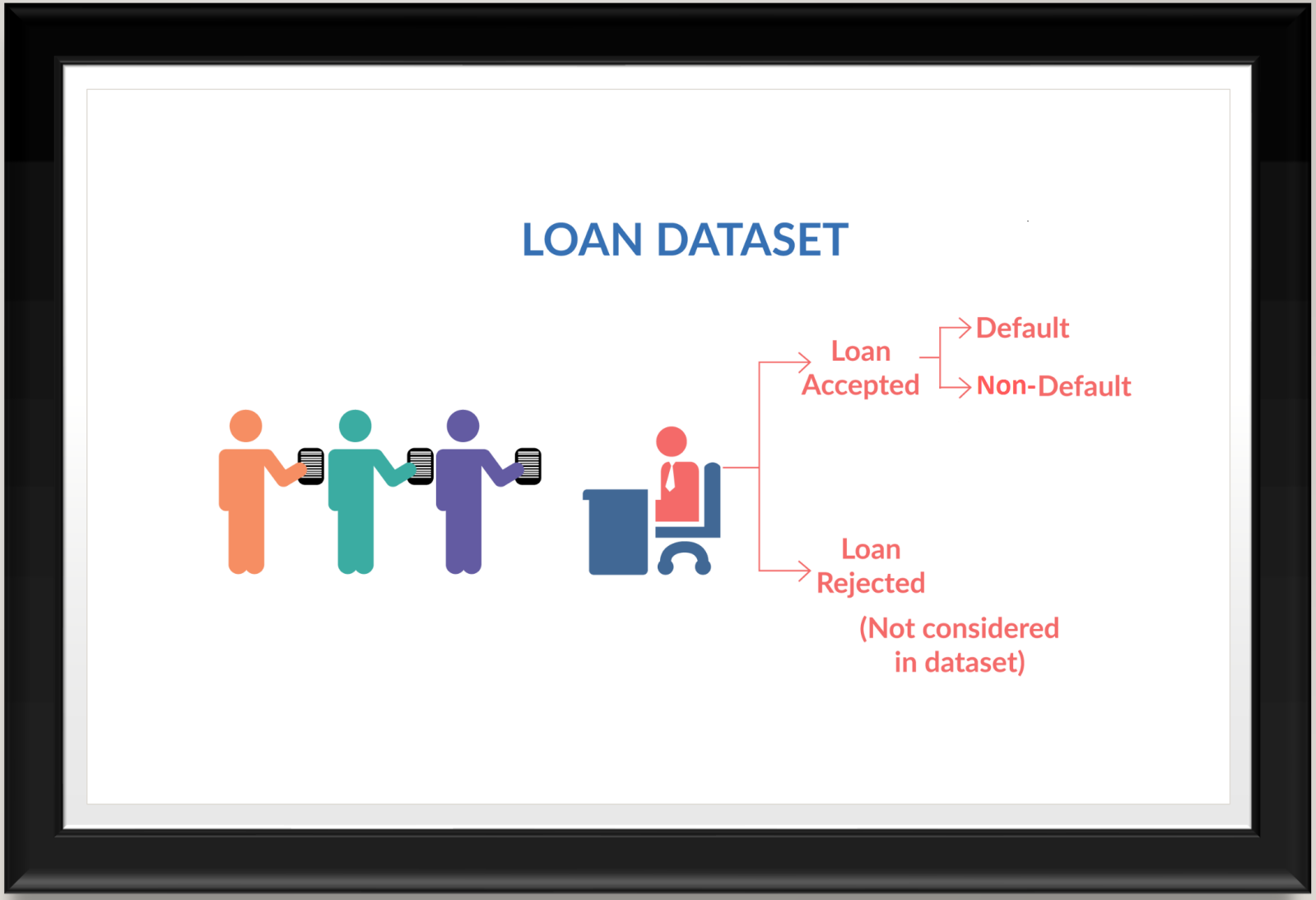
LENDING CLUB CASE STUDY

PROGRAM: EPGPML- C49

GROUP: ANKUSH GARG

PULKIT GANGWAR



OBJECTIVE

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

If one can identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

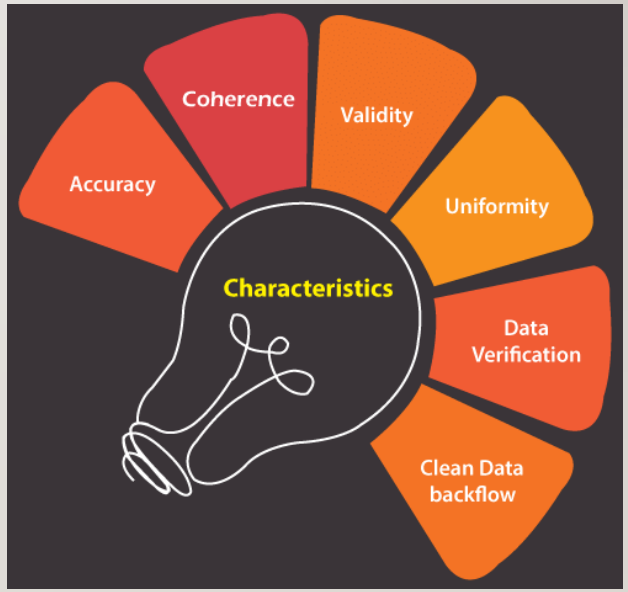
DATA UNDERSTANDING & CLEANING

DATA UNDERSTANDING

- Checked the shape of data
- Checked Data types and null values
- Checked Statistical Information for numerical data e.g. Mean, Median, 25%, 75% etc)
- Checked the number of unique values
- Checked the number of categorical and numerical columns

DATA CLEANING

- Removed currently in progress
- Removed columns with 90% or greater null values
- Removed variables not present during the loan processing.
- Removed redundant variables.



UNIVARIATE / BIVARIATE ANALYSIS

UNIVARIATE ANALYSIS

- Converted most numeric columns to binned categorical columns for ease in analysis
- Removed outliers from numerical columns
- Used bar charts, Pie plots, box plots, histograms etc to perform analysis.

BIVARIATE ANALYSIS

- Created a function to plot bar chart for categorial column and target variable
- Used bar charts, pie plots, etc to perform analysis

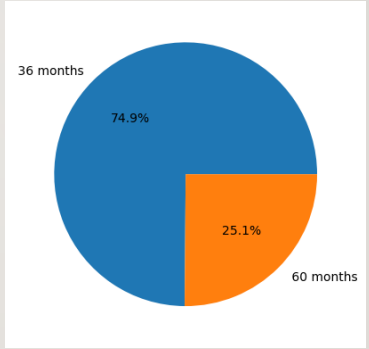
DRIVER VARIABLES

Based on the analysis the most influential variables on loan status are :

1. TERM

Univariate analysis

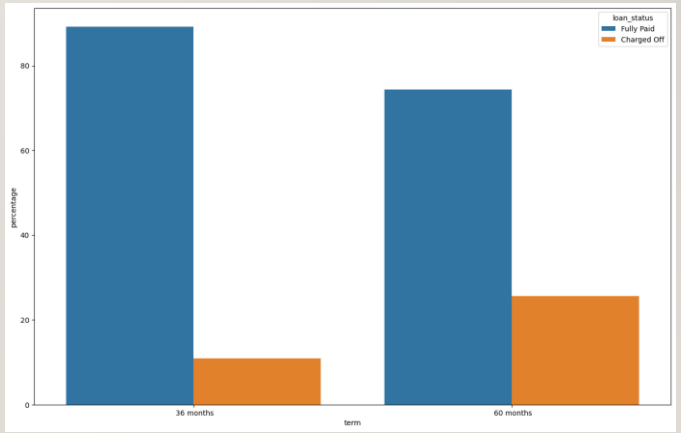
- Most of the loans were given for 36 months term.



Term – Univariate Analysis

Bivariate analysis

- 60 months term have more defaults as compared to 36 months term



Term – Bivariate Analysis

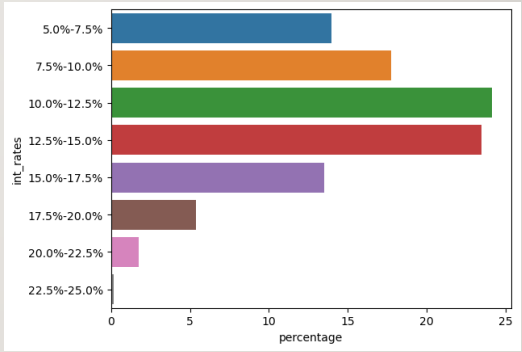
DRIVER VARIABLES

Based on the analysis the most influential variables on loan status are :

2. int_rate

Univariate analysis

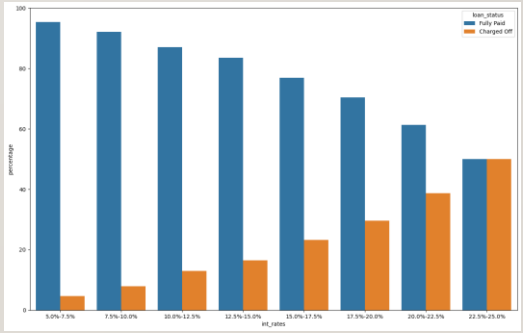
- Most int_rate offered on the loan are in between 10% to 12.5%



Int_rate – Univariate Analysis

Bivariate analysis

- As the interest rate is increasing the default rate is also increasing.



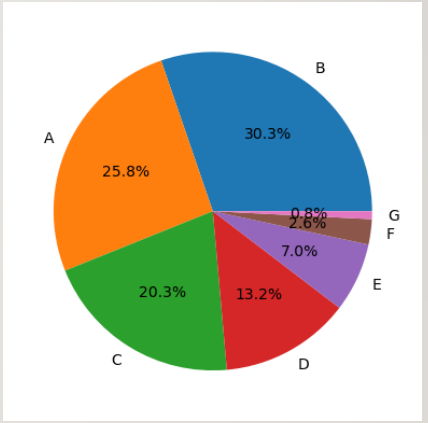
Int_rate – Bivariate Analysis

DRIVER VARIABLES

3. grade

Univariate analysis

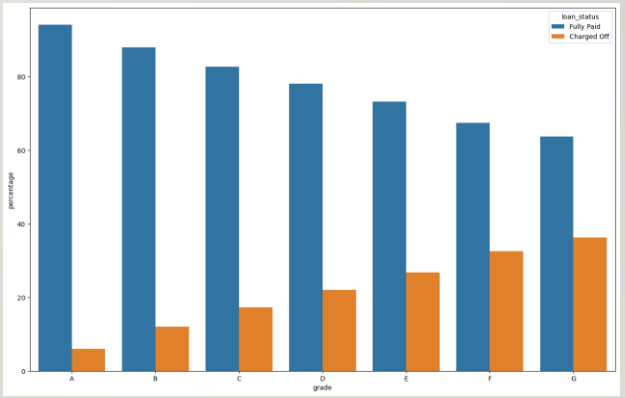
- Most loans are of grade B followed by A and C



grade – Univariate Analysis

Bivariate analysis

- Grade G and F have more defaults as compared to other grades.
- As the grades are increasing, defaults are also increasing with it.



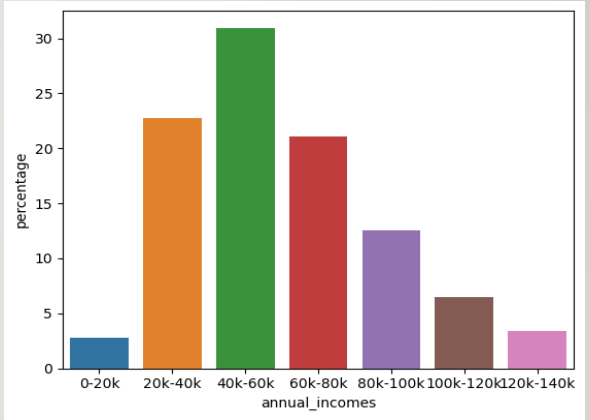
grade – Bivariate Analysis

DRIVER VARIABLES

4. annual_income

Univariate analysis

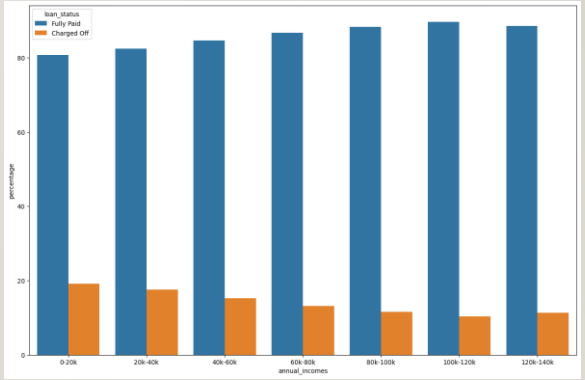
- Most borrowers with loans have annual income in between 40000 to 60000



annual_income – Univariate Analysis

Bivariate analysis

- Borrowers with annual incomes less than 40,000 have slightly more defaults. As the annual income is increasing the default rate is decreasing.



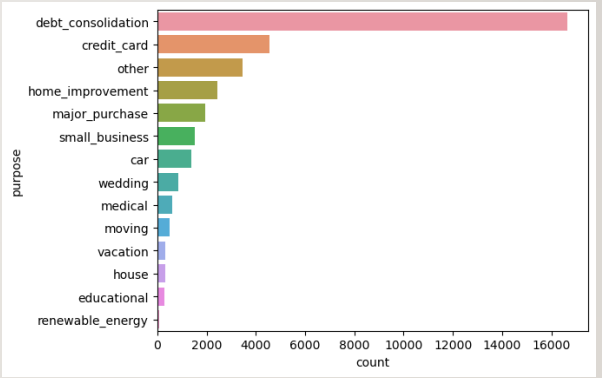
annual_income – Bivariate Analysis

DRIVER VARIABLES

5. purpose

Univariate analysis

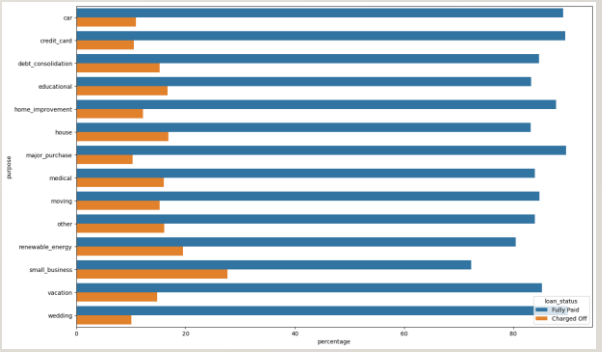
- About 47% borrowers have taken loan for debt consolidation.



purpose – Univariate Analysis

Bivariate analysis

- purpose category with value small business have higher percentage of defaults.



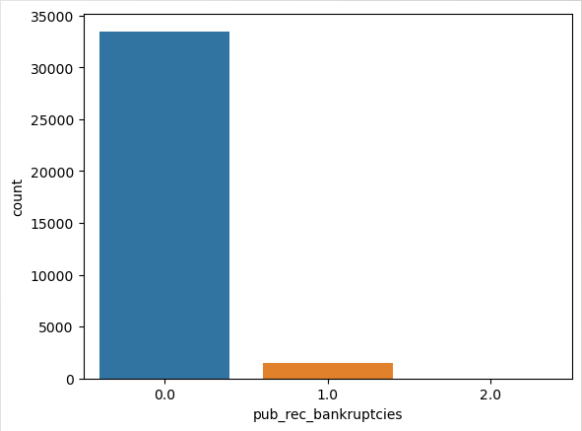
purpose – Bivariate Analysis

DRIVER VARIABLES

6. pub_rec_bankruptcies

Univariate analysis

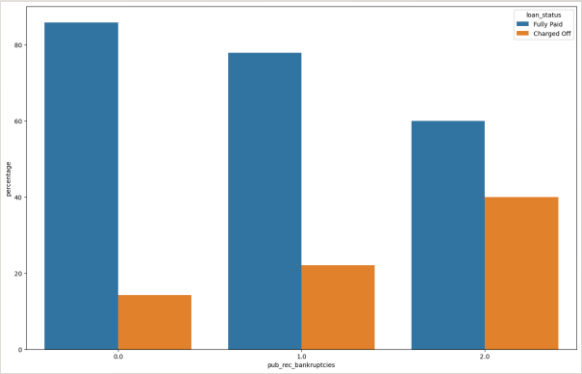
- About 95% borrowers have no bankruptcies and 4% borrowers have a single bankruptcy.



pub_rec_bankruptcies – Univariate Analysis

Bivariate analysis

- borrowers with 2 bankruptcies have more defaults.



pub_rec_bankruptcies – Bivariate Analysis

RECOMMENDATIONS

1. Term - Loans for term 60 months are more prone to default as compared to 36 months. Our suggestion is to convince borrowers to opt for 36 months term.

2. pub_rec_bankruptcies - Borrowers having 1 or more bankruptcies are prone to defaults. Our suggestion is to avoid approving loans for borrowers having 1 or more bankruptcies

3. loan_amnt - Borrower with loan amount higher than 30000 are prone to defaults. Our suggestion is to reduce the number of loans for loan amount greater than 30000

4. int_rate - Higher interest rate tends to more loans being charged off. Loans with interest rate higher than 15% are more prone to defaults. Our suggestion is to approve loans with low interest rate

5. annual_income - Borrower having low annual income (<40k) are more prone to defaults. Our suggestion is to reduce the number of loans for borrowers having annual income less than 40000

6. purpose - Borrowers taking loans for small business or renewable energy are more prone to defaults. Our suggestion is to reduce the number of loans for these purposes.

7. grade - As the grade increases defaults also increases. Grade higher or equal to D are more prone to defaults. Our suggestion is to reduce the loans for grade higher or equal to D