GRAMENER CASE STUDY

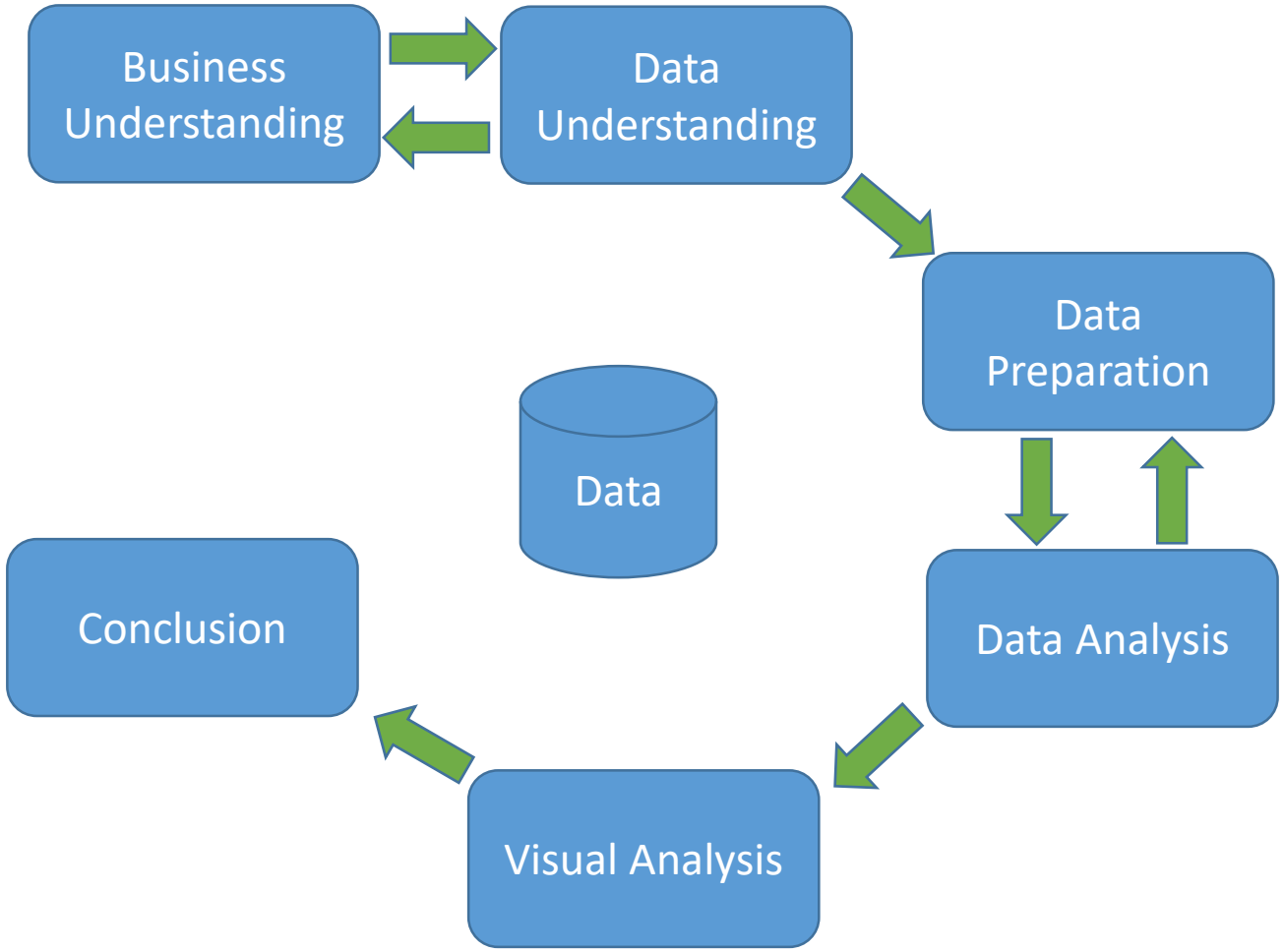
SUBMISSION

1. Shagun Agarwal
2. Deepak Singh
3. Ravi M
4. Abhimanyu Khandare

Abstract

- BACKGROUND AND PROBLEM STATEMENT:
  - A consumer finance company specialises in lending various types of loans to urban customers. When the company receives a loan application, they have to make a decision for loan approval based on the applicant's profile. Using EDA we need to understand how consumer attributes and loan attributes influence the tendency of default i.e. the factors/variables which are strong indicators of default.
- APPROACH:
  - We have adopted multi pronged approach involving business-driven, type-driven and data-driven metrics. We have performed univariate and bivariate analysis with the help of robust visualization techniques in R. We were able to recognize patterns using histograms, dodged/stacked bar chart helping us to identify the driver variables.
- CONCLUSION:
  - Per our analysis, the **six** driver variables for loan default are as below:
    - Annual Income
    - Home Ownership
    - Grade
    - Sub-grade
    - Purpose
    - DTI

Problem solving methodology

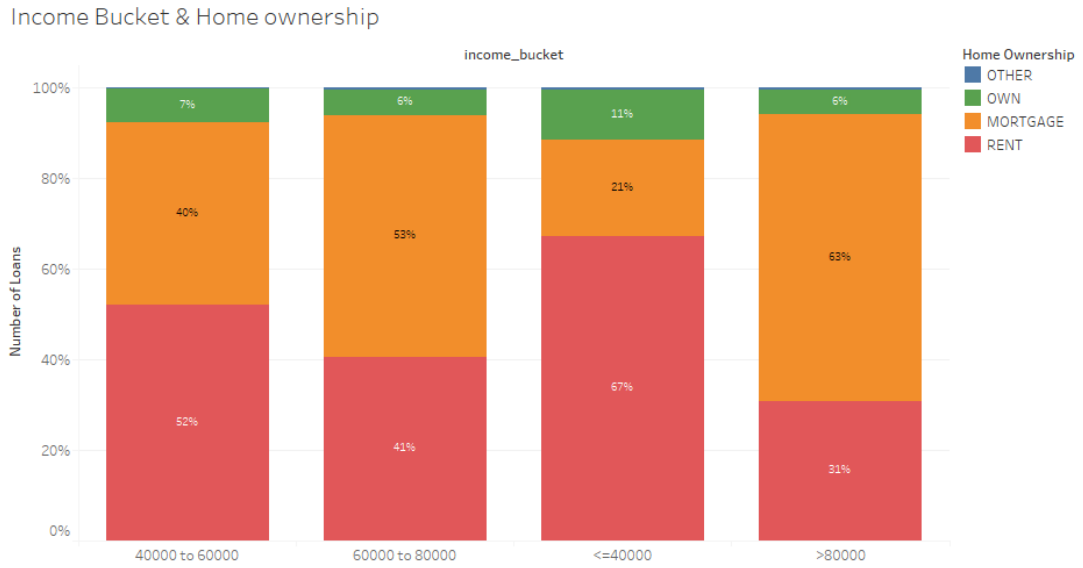


Data understanding, cleaning and manipulation

- Data driven cleaning is followed in the following sequence
  - Identified columns with NAs and deleted from the data set such as tot_hi_cred_limit, num_sats etc
  - Columns with no significant patterns in the data were identified and deleted such as pymnt_plan, url etc
  - Checked for duplication of rows
  - Deleted columns where more than 90% of rows were found to be blank such as next_pymnt_d
- Business driven cleaning is followed in the following sequence
  - Columns were renamed to be more intuitive and self explanatory
- Type driven cleaning is followed in the following sequence
  - Non numeric characters such as %, ‘months’ were removed to ensure consistency in data type
  - Variables considered as categorical were converted to factors
- Data manipulation
  - Derived columns were created to further enable segmented analysis such as issued_month, issued_year
  - Annual income was categorized to perform segmented analysis
  - Home ownership was recategorized to identify non owned houses (Rent + Mortgaged)

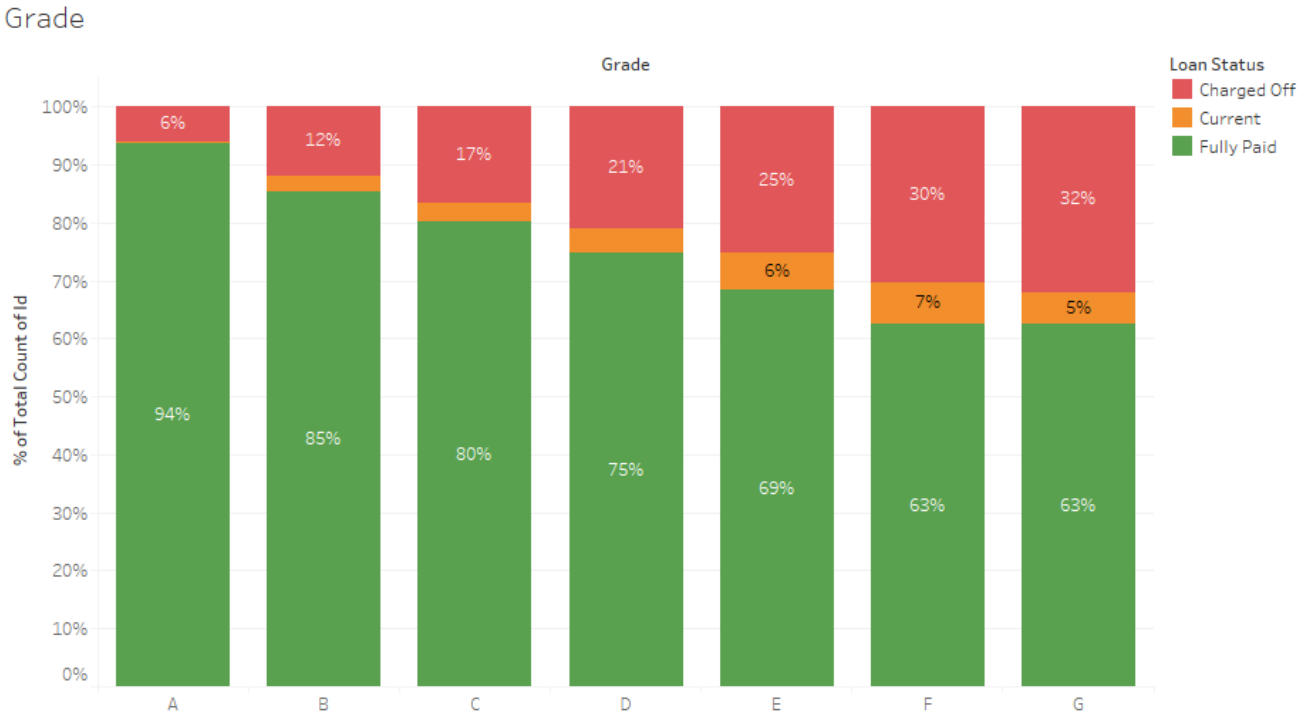
Data Analysis

- **Annual Income & Home ownership**
  - 18% of loan defaulters have annual income <= 40,000
    - 67% of loan defaulters in the above category have rented homes



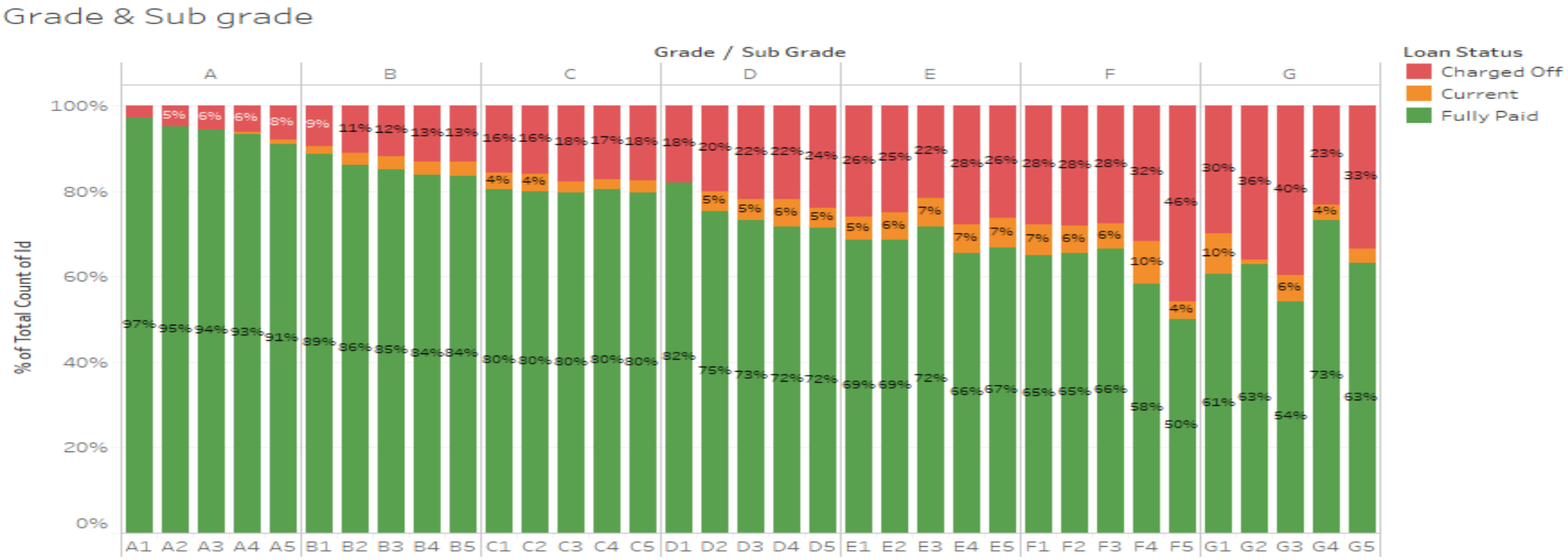
Data Analysis

- **Grade**
  - As the grade moves from A to G, the percentage of defaulting increases from 6% (in A) to 32% (in G)



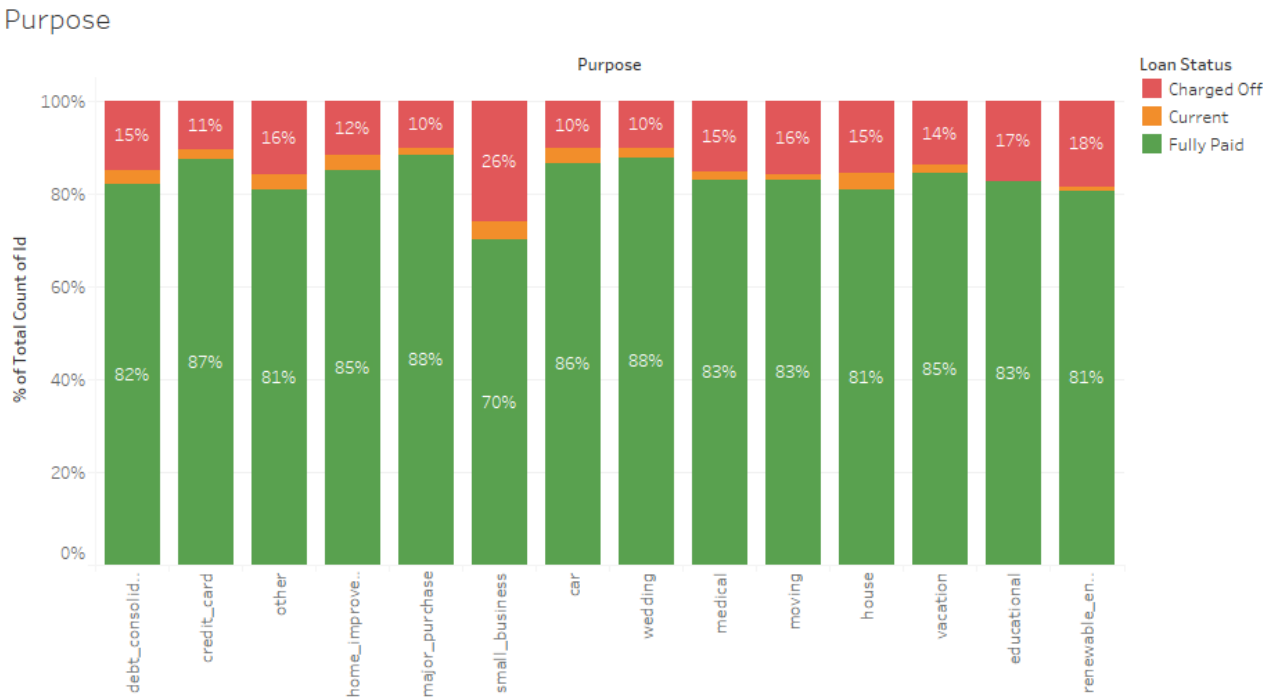
Data Analysis

- **Sub-grade**
  - Similar pattern is observed when we move from Sub-grade A1 to G5
  - F5 has highest default rate – 46%
  - Sub-grades in G have higher default rate in sub-grades in A
    - Observed one outlier – default rate of F5(46%) > G1(30%)



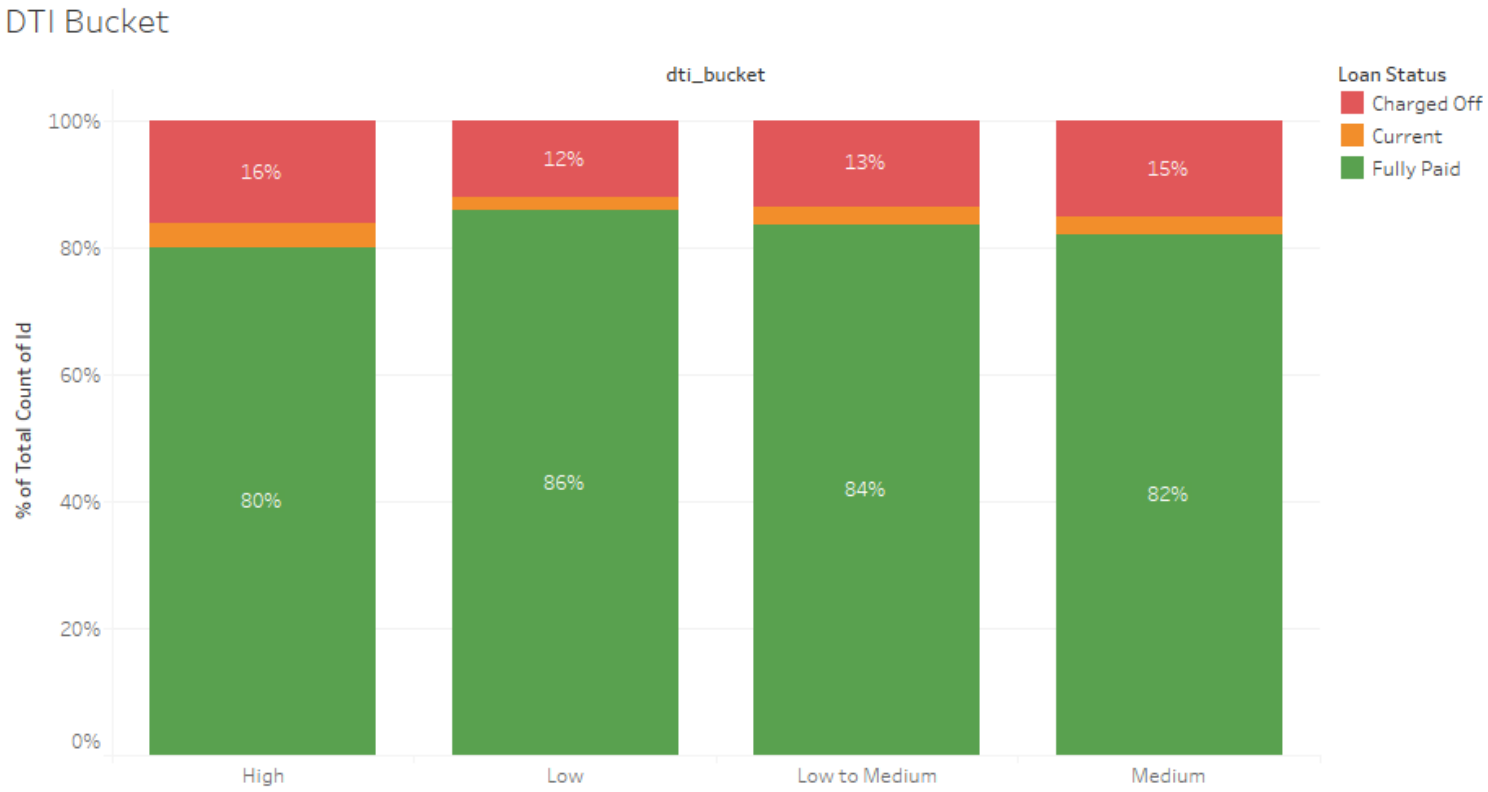
Data Analysis

- **Purpose**
  - Small business has the highest default rate of 26%
  - Other category has 16% default rate



Data Analysis

- **DTI Bucket**
  - The chances of loan default increases marginally as the DTI increases.
  - Charged off percentage for Low DTI – 12% and High DTI – 16%



Conclusion & Recommendation

- Per our analysis, the **six** driver variables for loan default are as below:
  - Annual Income
  - Home Ownership
  - Grade
  - Sub-grade
  - Purpose
  - DTI

- Recommendations to reduce loan default rate
  - Extra verification for applicants with annual income <= 40,000 and having rented home
  - Increase tenure for applicants in higher sub-grades and in Grade G
  - High interest rate for small business and other category as ‘purpose’
  - Increase tenure for applicants with high DTI