Gramener Case Study

Group Name:

1. Naveen Masali
2. Nikhil Tiwari
3. Sagar Sheth
4. Rahul Varaganti

Objectives

XYZ Bank deals in giving personal Financial Loans of all sorts.

Business Objectives:
1. Analyse the Loan data set of 4 years of current, Fully paid and Defaulted Loans.
2. Analyse the data set and hypnotize the multiple cause of Defaults .
3. Possible Solution to the Issue.

Problem analysis

Understanding Data:
This data has information about 4 years of loans that have been issued by the bank. They are mainly of three types *Fully Paid off, Current and Defaulted*. The data set contains all sorts of information about customer like annual income, employee length, purpose of loan, no. of bankruptcies etc.

On high level following are stages in the workflow:
- Gather data for the analysis.
- Clean and Format the data for readability.
- Extrapolate the Derived Metrics.
- Filter out the outliers based on Annual Income.
- Filter Data Set on Loan Status.
- Perform Univariate and Bivariate analysis on the Data.
- Plot the results of Analysis and Hypnotize the Root of the Issue.

Problem Analysis

Assumptions:

1. As current loans have not yet been completed, we decided to ignore them during analysis.
2. Outliers (Based on Annual Income) have not been considered during analysis.

Tools for the workflow:

- RStudio is used for data cleansing, modelling, analysis and plotting.

- Tableau is also used for plotting the results.

Plots: Annual Income vs Grade

- The mean income as well as the range of the incomes is almost same across all the grades.
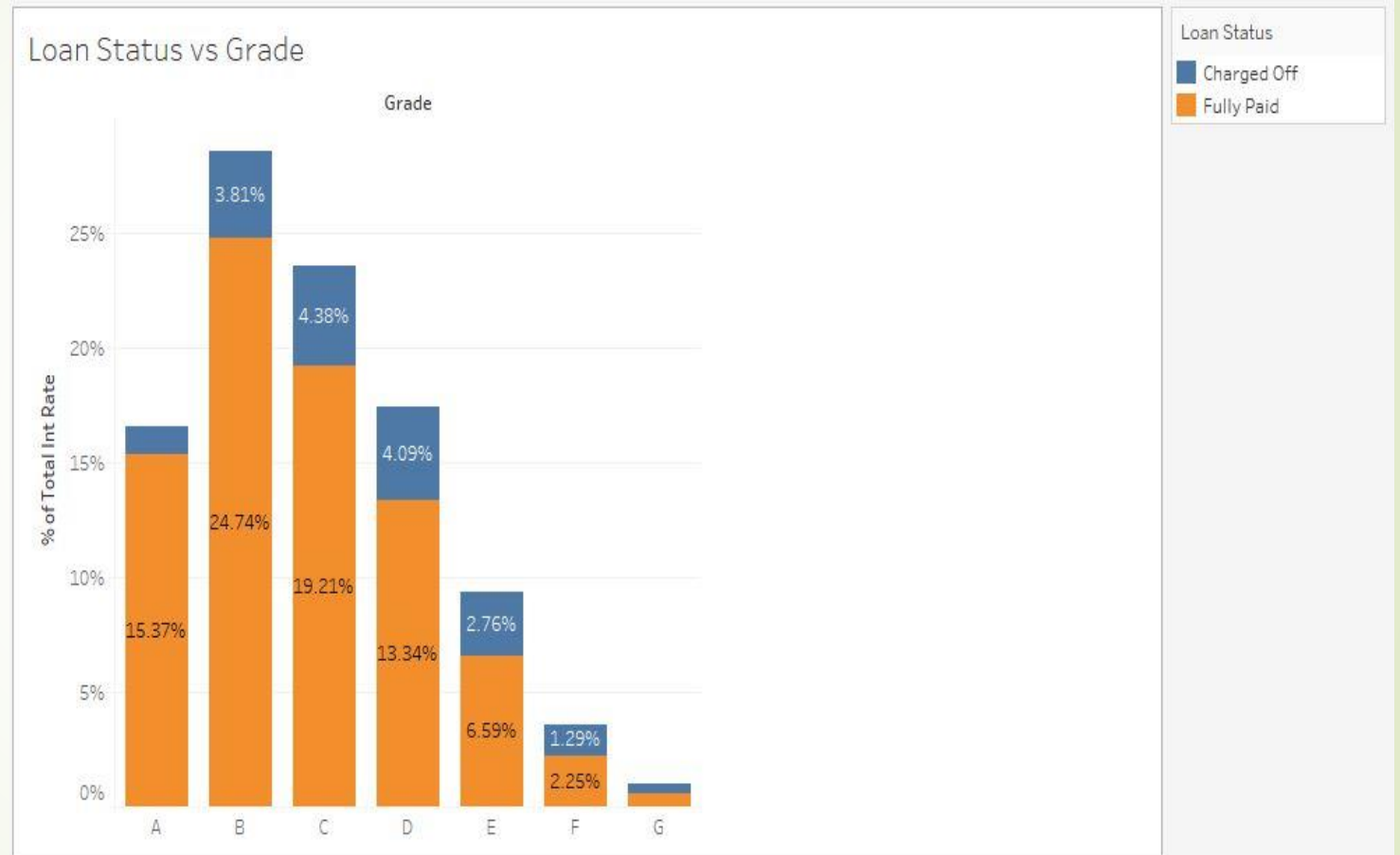
- It is a little more for grades E,F and G.

Annual Income vs Grade
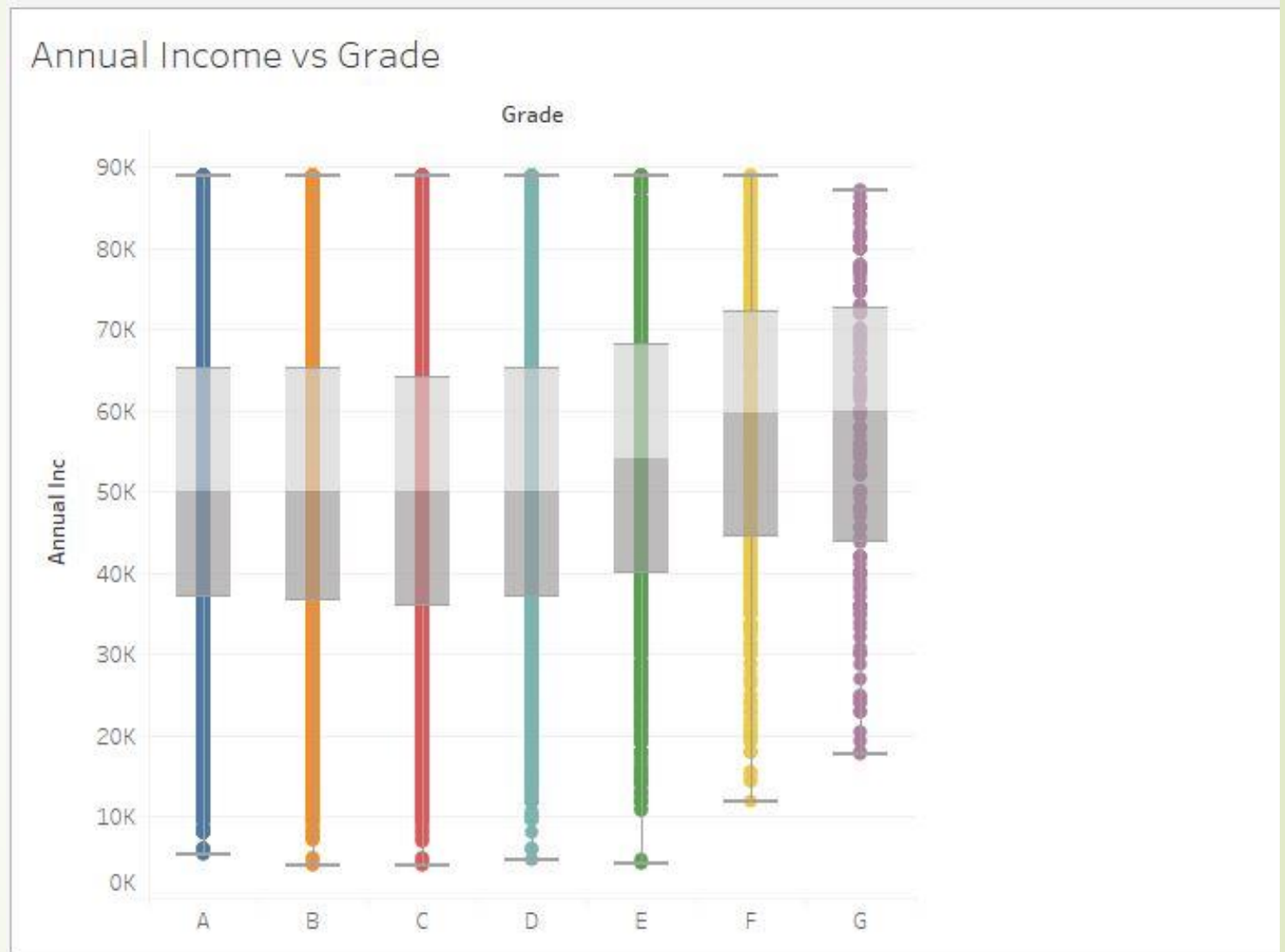
Loan amount vs Loan Status/Grade

➡ The graph is right skewed, that shows the grades from D to G follow a decreasing trend for both Default and Fully paid.
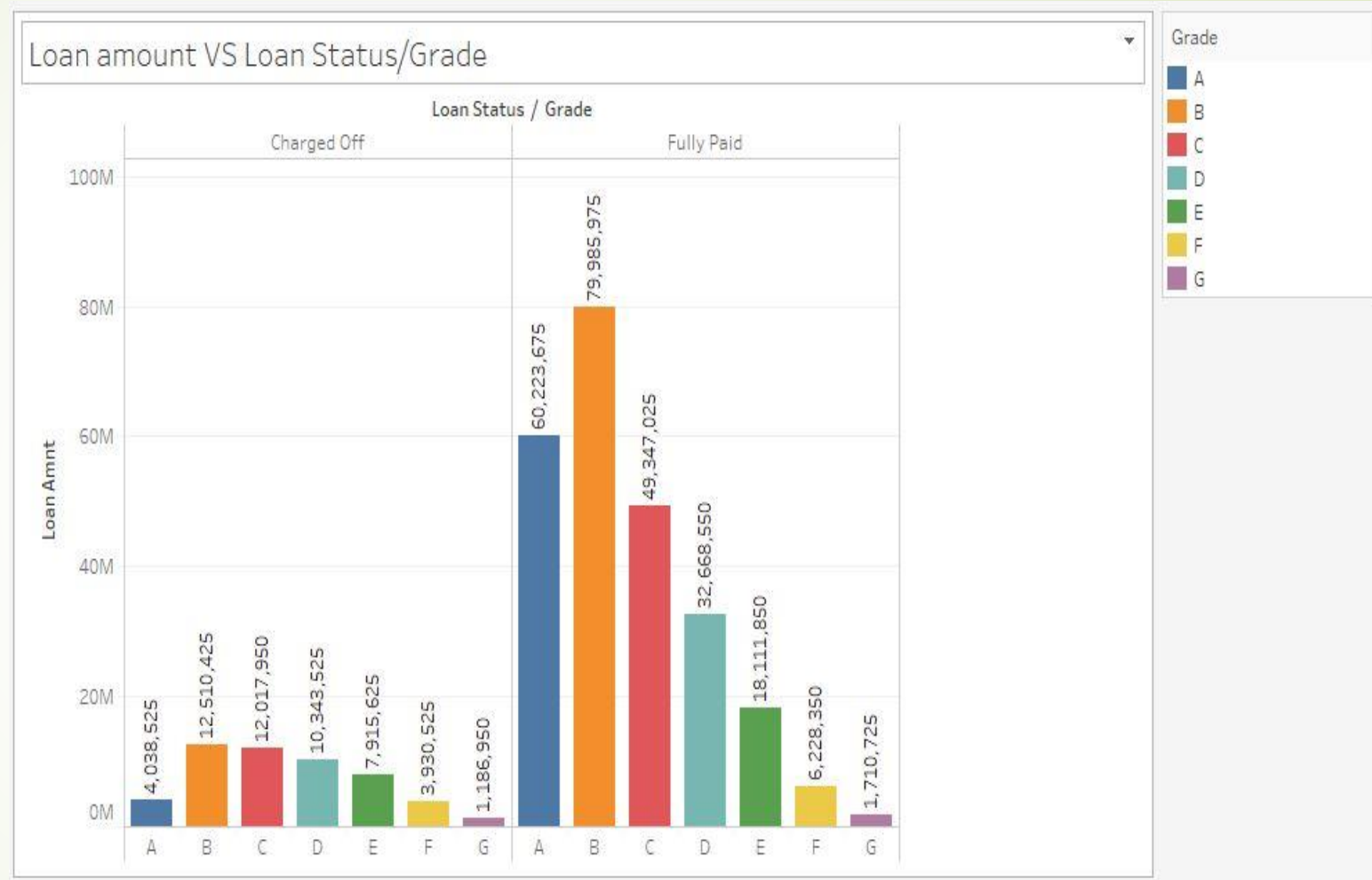
Loan Status vs Grade

- The graph clearly shows in percentages for the contribution in both charged off and Fully paid.

- The percentage of contribution gap is closing from grades C to G.

Loan Status vs Grade

Grade

Loan Status
- Charged Off
- Fully Paid

% of Total Int Rate

A: 15.37%
B: 24.74%, 3.81%
C: 19.21%, 4.38%
D: 13.34%, 4.09%
E: 6.59%, 2.76%
F: 2.25%, 1.29%
G

Loan Status vs Grade

- The graph clearly signifies the closing gap between the Charged off and Fully paid statuses across all the grades.

- Hence, particularly for grades D,E,F and G , it is better if loans are lent with caution.

Funded amount vs Grade

- This graph shows that the amount of defaulters are more in the grades from E to G.

- The pattern is very similar where the grades from D to G are less performing grades for the bank.
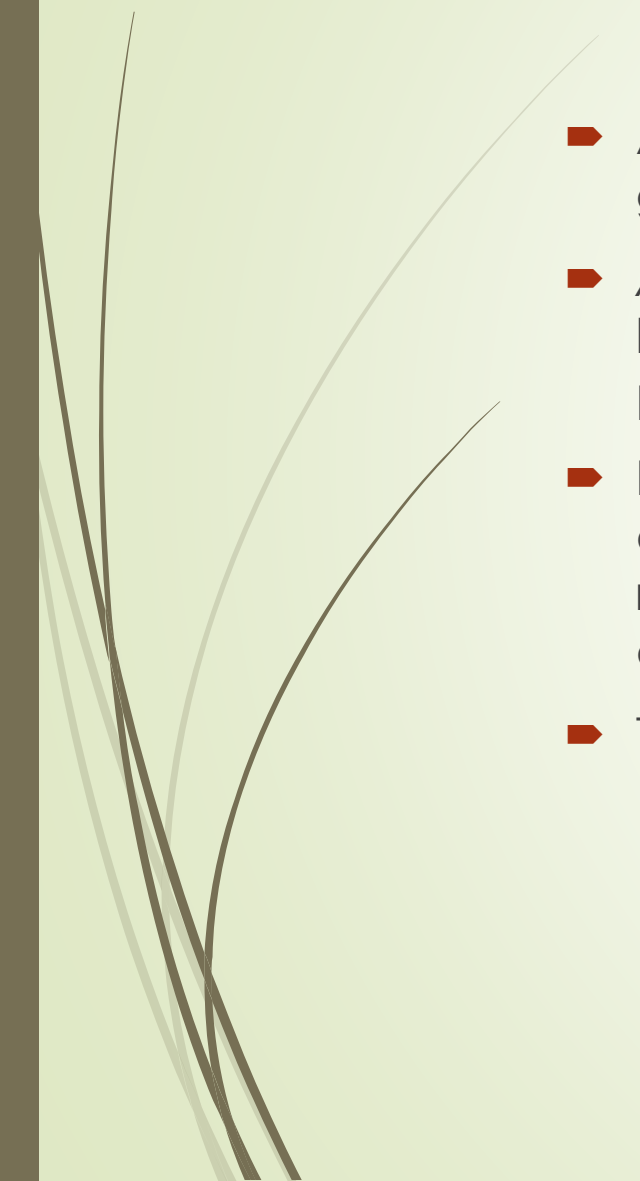

Funded amount vs Grade

Loan amount vs Verification

- We can see here that the number of defaults for source verified and verified are high in number.

- We need to investigate whether the documents have been correctly identified and verified before the loan is lent.

Loan Amount vs Verification Status/Loan Status

Loan Status
- Charged Off
- Fully Paid

Conclusion

- After examining annual income and grades, we have found that all the grades have almost similar distribution.

- *Annual Income, interest rate* and *verification* have bearing on whether an loan will become fully paid or become charged off. Hence, these parameters need to be critically monitored.

- In the current disbursal, from grade D to grade G, the percentage of defaults are more. Grades also play an important role in loan disbursal and repayment and can be reference from current data to make future decisions.

- Target based/ last minute loan disbursals should be avoided.

Recommendation

- Customer should be classified clearly based on their income, frequency of application, purpose of use, loan duration and intended use based on this interest rate and disbursed amount should be calculated.

- The process of income verification should be more accurate as it is an important parameter to control defaults.