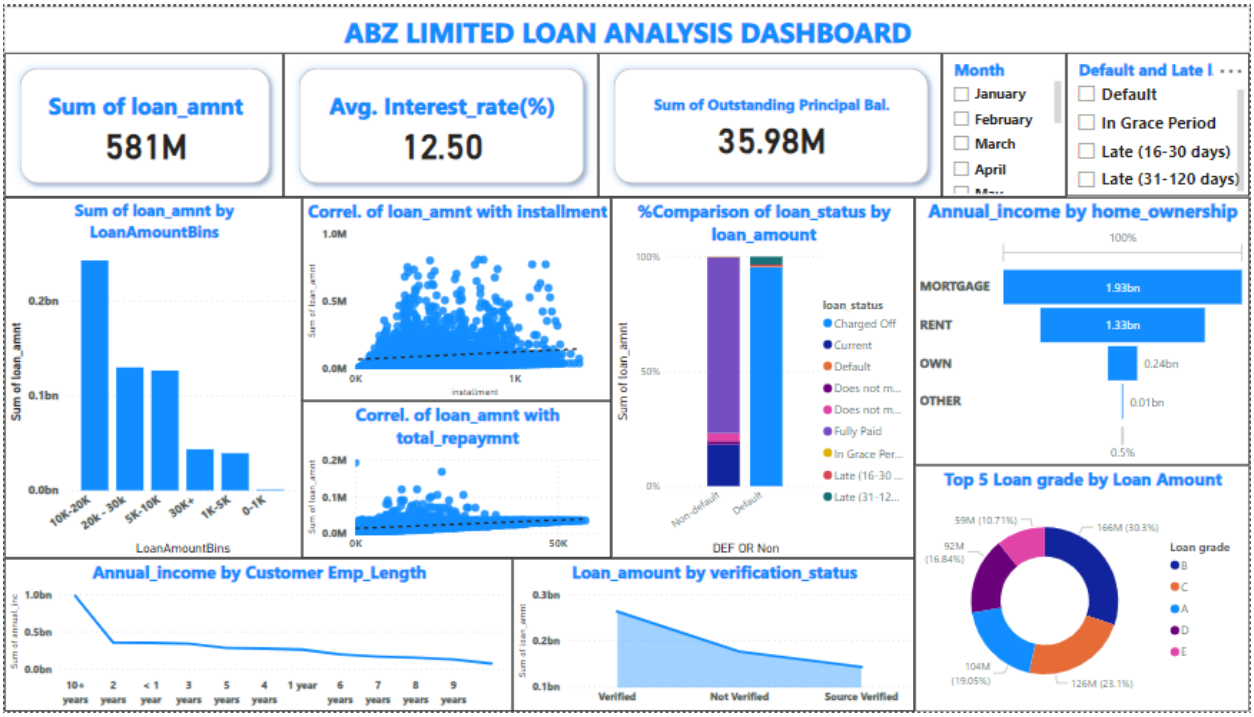
ABC Financial Loan Analysis – A Case Study

1. **A dashboard that communicates my findings**:



2. **Brief overview of the data analysis:**

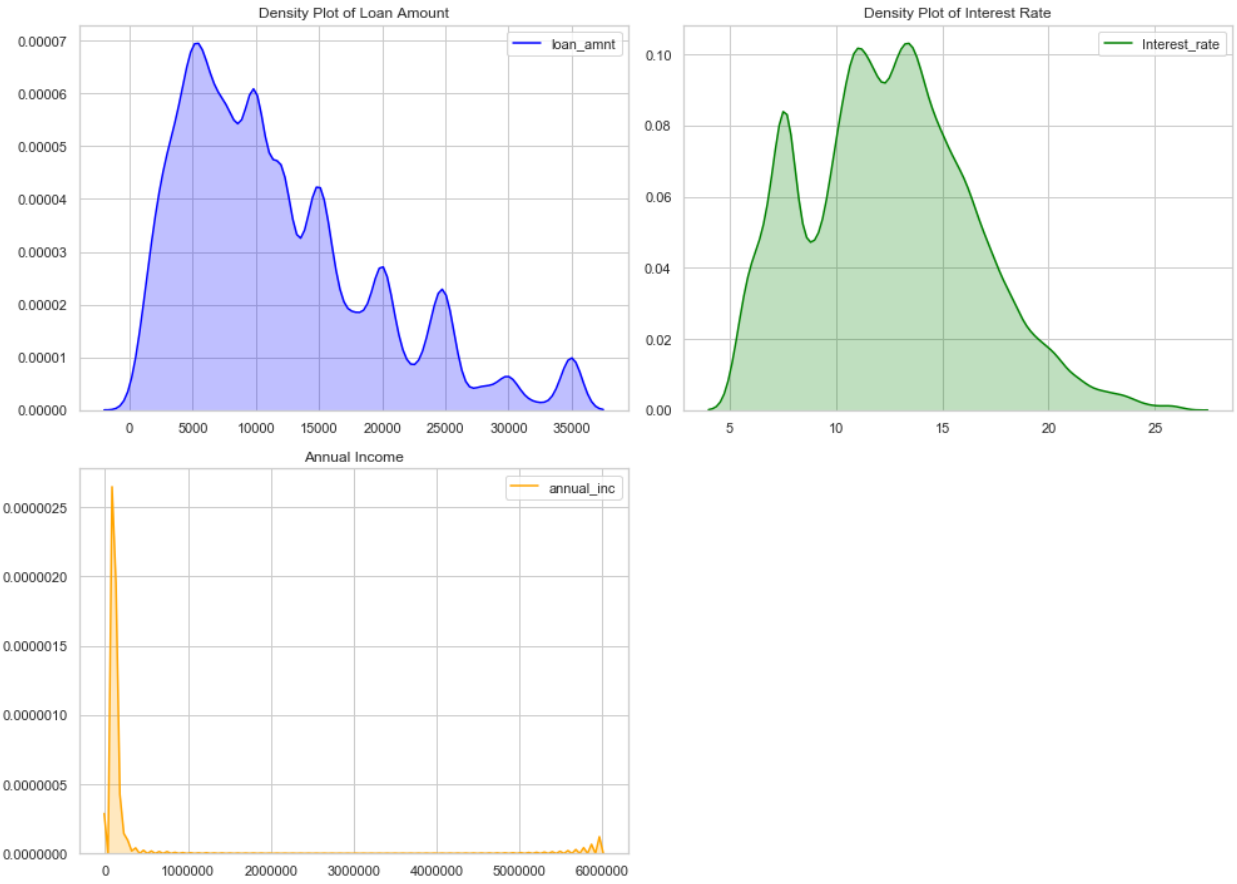
The data was first imported into excel for initial analysis, appropriate data cleaning was done like removing duplicates, checking for misspelt words, dealing with null data, setting the right data type, where the loan_status column was recoded into default and non-default with "Default, In grace, both late categorical data" set as default and others as non-default. And another column to replace non-default with 1 and default with 0, to make the loan_status column numerical data for the statistical analysis sake.

Then the data was imported into python jupyter notebook to carry some statistical analysis on the data, such as:

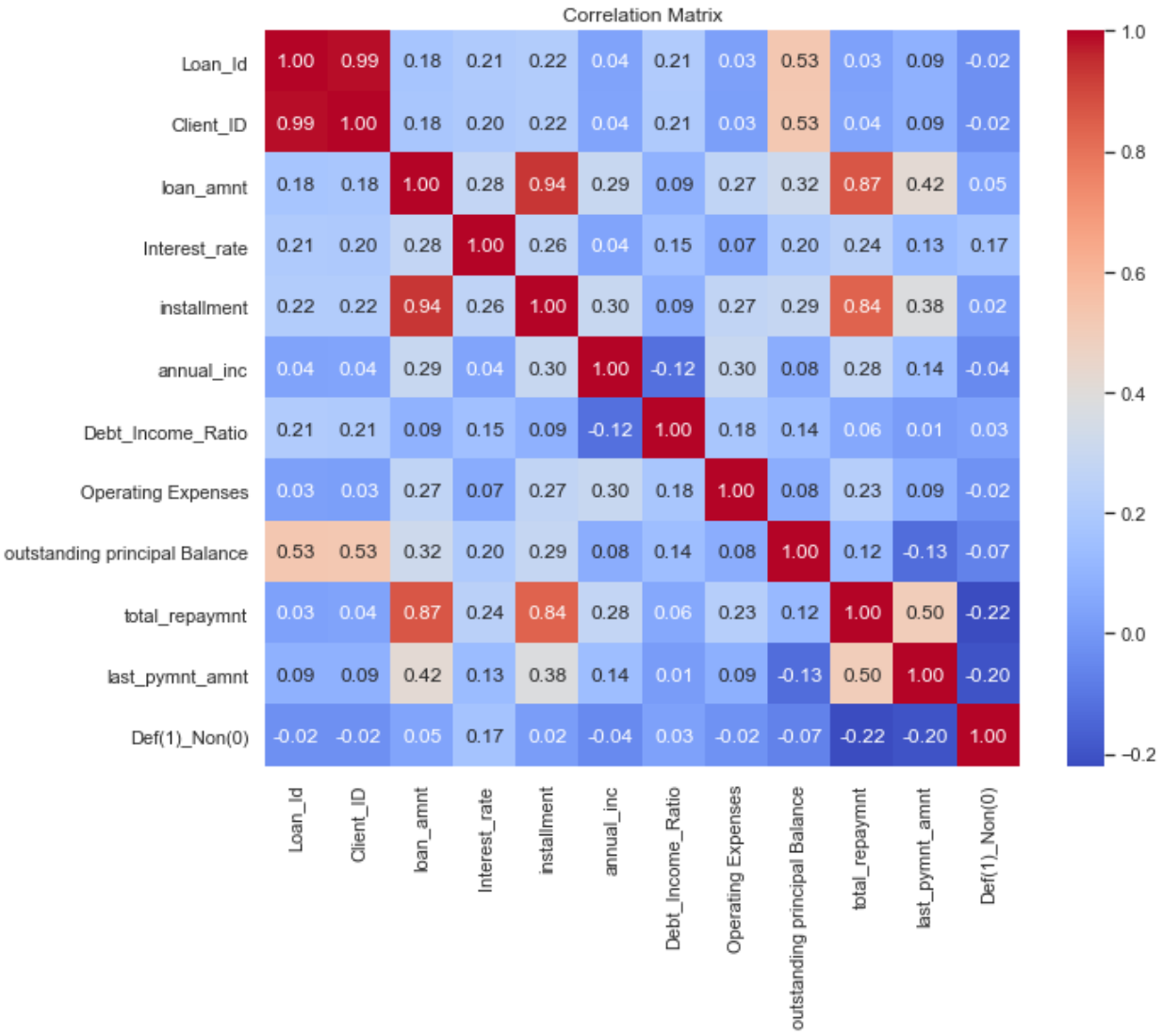
1. Descriptive statistics
2. Distribution analysis and
3. Correlation analysis.

Descriptive statistics revealed the summary of the entire data where the measure of central tendency such as mean, median and mode were identified and the measure of central variation such a range, standard variation and variance were also determined.

Distribution analysis showed the dispersion of the numerical data and gave an idea about the skewness of the data, all of these skewed to the right showing that 25%(Q1) and 50%(Median) of the population are close to each other.



Correlation analysis was then carried out before visualizing to det few insights into the data, the image below shows the correlation matrix that preformed the dashboard.



Among all, loan_amount shows high correlation with installment payment and total_repayment, likewise installment shows high correlation with total_repayment.

There was a challenge in checking the correlation of categorical data with numerical data at the moment, hence powerbi was used to create the visuals to check for these.

Key finding and Insights:

Note: Defaulters are limited to observations with “Default”, “Late (16-30 days)”, “Late (31-120 days)”, “Does not meet the credit policy. Status:Charged Off” and “Charged off” while others were treated as Non-default

1. The top 5 Loan defaulters are found decreasingly in Grade C, followed by B, then D, then E and F. Garde C being the highest
2. Also, the 5 Loan non-defaulters are found increasingly in Grade E, B, A, C and B. Grade B being the highest.

3. The highest amount of loan by defaulters were disbursed to them in December, which is 3 times the highest among other months.
4. Exciting to know that the loan amount disbursement both for defaulters and non-defaulters increases per month.
5. The highest count of loaners collects between the range of 10k-20k, non-defaulters have equal count for range “10k-20k” and”5k-10k”, while defaulters have more “10k-20k” than ”5k-10k”.
6. There is a consistent trend and pattern in employment length and loan amount for both cases of loan status.
7. There is a very strong correlation between loan amount and installment and also loan amount and total repayment.
8. Verification status has a consistent trend for both cases of loan status
9. Debt consolidation is the primary purpose for loan collection across the loan status

Recommended Cause and Strategies

1. Since Grade B loaners have the track record of paying loans back in time, they could be used to orient the Grade C loaners on strategies to always repay back in time.
2. Loans disbursement should be critically analyzed in December before releasing as it maybe that most people collect the loan to shop at discounted price or just enjoy the new year period.
3. The team should be prepared to disburse more loan than each previous month as it was evident that loans applications and disbursement increases by month.