# EDA CASE STUDY

Objectives

- Identification of risky loan applicants who are labelled as "defaulters".

- Understand the "driving factors"(or driver variables) behind loan default.

- The company can utilize this knowledge for its portfolio and risk assessment.

Problem Solving Methodology

- Understanding the dataset, i.e. data sourcing

- Data cleaning (Imputing the missing values)

- Data visualization using graphs and charts

- Univariate, segmented and bivariate analysis
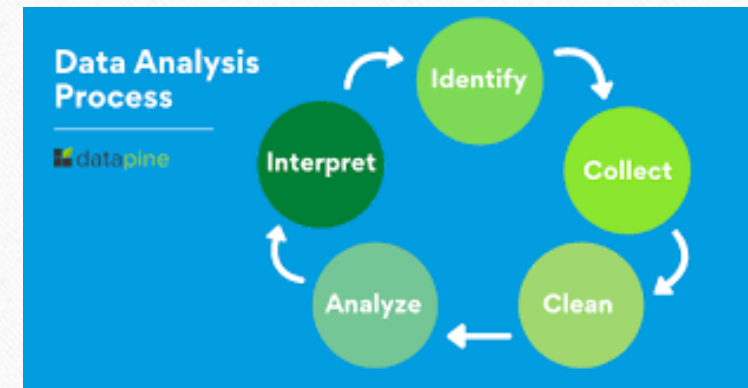
Steps for data cleaning

- **Delete columns**: Delete unnecessary columns.

- **Remove outliers**: Remove high and low values that would disproportionately affect the results of your analysis.

- **Missing values:** Treat missing values with appropriate approach.

- **Duplicate data:** Remove identical rows, remove rows where some columns are identical.

- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.

Data Visualisation

- The essence of the whole project is to analyze how consumer attributes and loan attributes are influencing the tendency of defaulting.

- We performed data cleaning and preparation on the Loan dataset.

- During univariate analysis we have created histograms and bar charts to check out the distribution of all the driver variables and performed the Multivariate analysis to understand how different variables interact with each other in the analysis.

Loan Grade Analysis

- As the grade of loan goes from A to G, default rate increases.

- Higher interest rates have higher tendency to default the loan.

- Hence, we can decide accordingly which grades to work on.
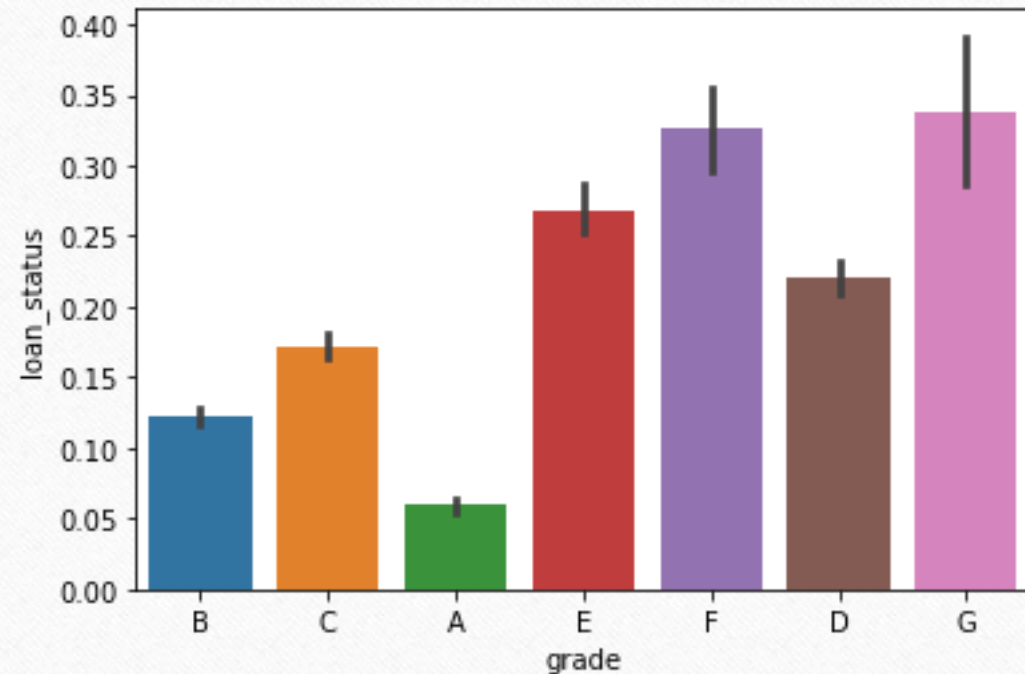
Loan Purpose Analysis
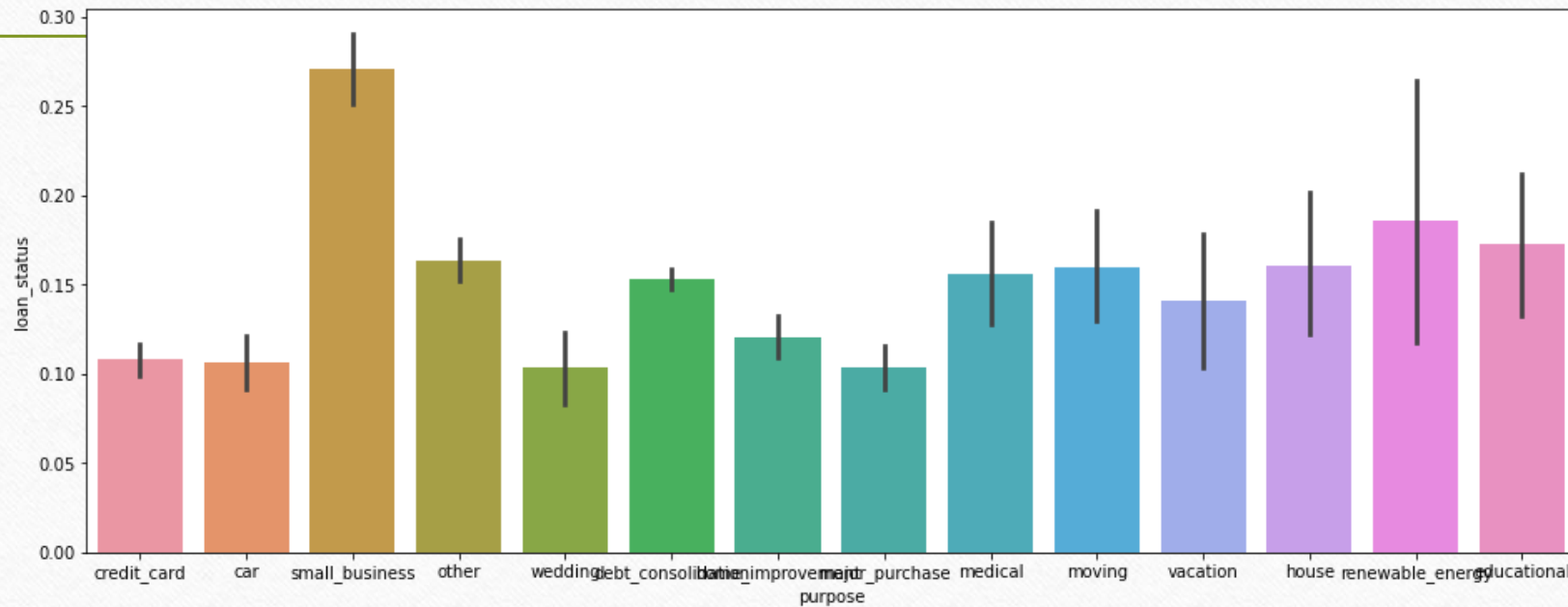
- People with small businesses default the most for the loan closely followed by renewable energy sources and educational institutions.

Loan Amount Analysis

- Funded amount is left skewed. Most of the loan amount given is below 7 lakhs.

- Probability of people with loan amount greater than 7 lakhs tends to default is more than the people with less than 7 lakhs.
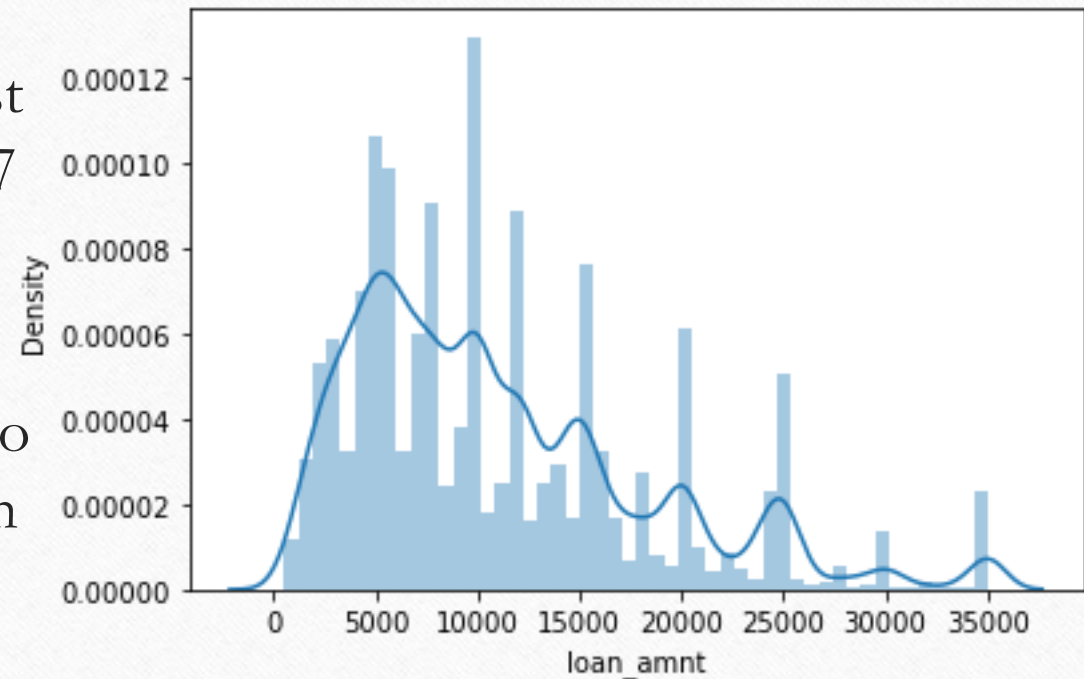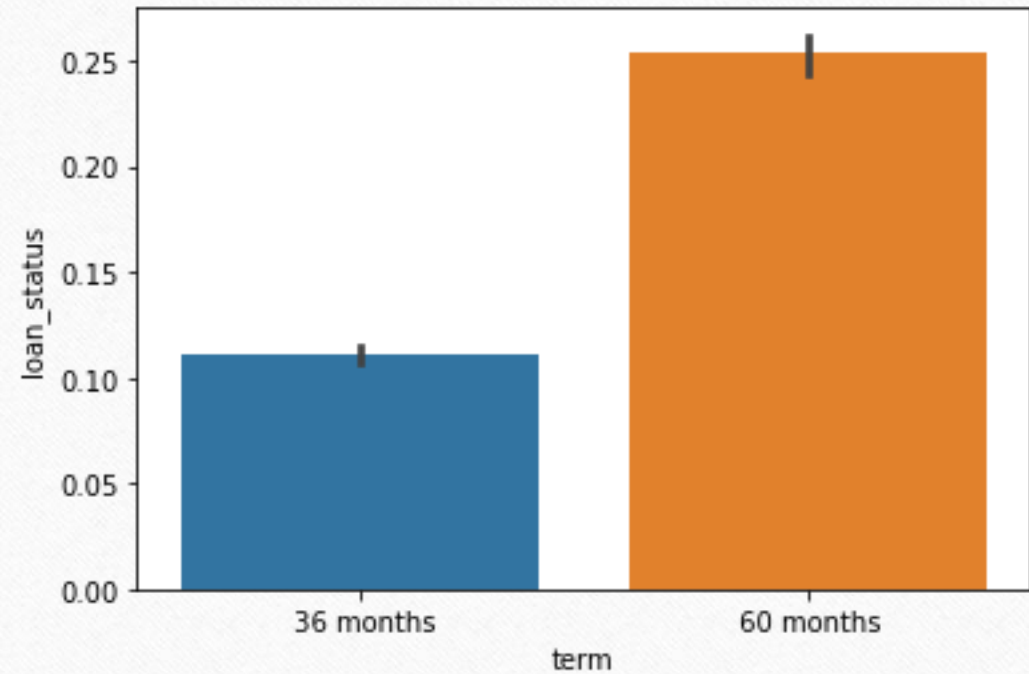
Loan Purpose Analysis

- Most of the loans are granted were for debt consolidation.

- Applicants having higher income tends to default the loan.

- Very less amount of loan is granted for home renovation and credit card.
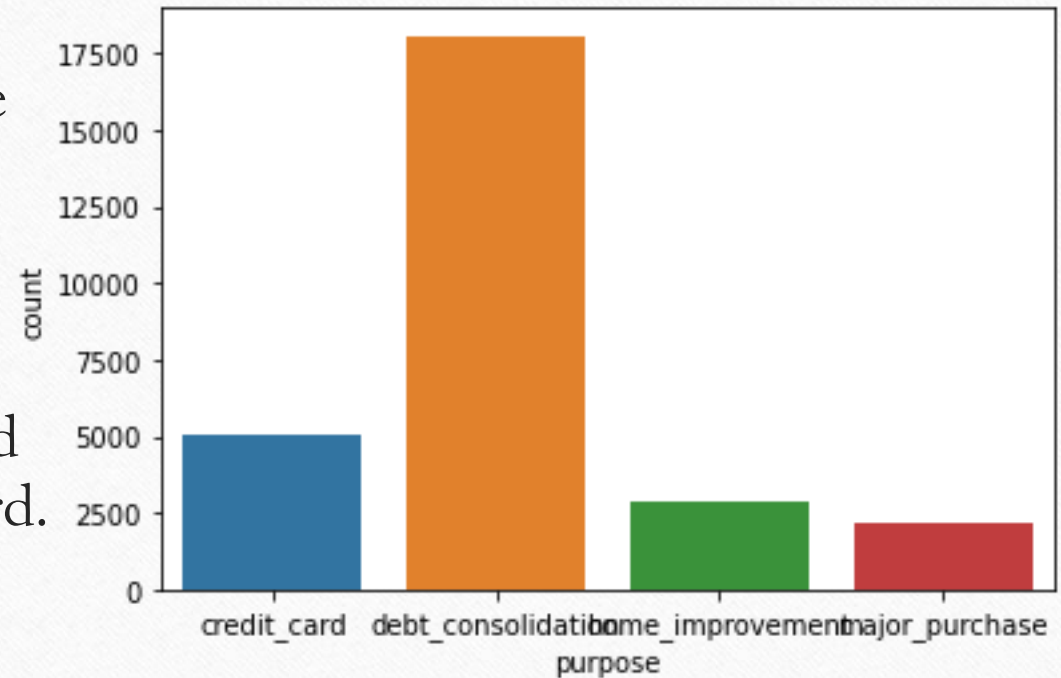
Loan Status vs Loan Term

- Most of the loans granted were of 60 months.

- 60 months loans default more than 36 months loans.

Conclusion

- Low grade loans have high tendency to default.

- Grading system is working as expected.

- Loans having higher interest rate have more defaulters.

- When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default.

- Extra scrutiny must be done for the applicants belonging to small business, as tendency to default is high.

THANK YOU