

Planning Document

1. Understanding the Data

- Load and inspect the data.
- Check data types.
- Identify missing values.
- Find duplicates.

2. Basic Statistics

- Review summary stats (e.g., `describe()`, `value_counts()`).
- Look for zero values and constant values.

3. Variable Distribution

- **Numerical Features:** Use histograms and KDE plots.
- **Categorical Features:** Use bar plots and check for imbalance.

4. Correlations

- Check numerical correlations (e.g., `corr()`).
- Explore categorical correlations.

5. Feature Interactions

- Use scatter plots to analyze interactions.

6. Target Variable Analysis

- Analyze the target variable's distribution for class imbalance and decide if resampling or weighting is needed.
- Examine relationships with predictors using scatter plots and box plots.