

Data Collection and Preprocessing Phase

Date	1 July 2024
Team ID	739854
Project Title	Software Salary Prediction
Maximum Marks	2 Marks

Data Quality Report

The data quality report for software salary prediction highlights completeness, consistency, and accuracy of the dataset. Missing values were found in >5% of records and handled via imputation. Categorical variables showed consistent encoding, and numerical features were within expected ranges with minimal outliers. Data was checked for duplicates, resulting in a 2% removal. Overall, the dataset is reliable for analysis, with preprocessing steps ensuring it meets the requirements for building predictive models.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle	Missing values in the "Rating, Name"	Low	Drop the null values as there are so less this won't affect the model.
Kaggle	Categorical data in the dataset	Moderate	Encoding has to be done in the data

