

Data collection and preprocessing phase

Date	15 June 2024
Team ID	740677
Project Name	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 Quality Report :

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	'Rating', 'Company Name', 'Job Title', 'Salary', 'Salaries Reported', 'Location', 'Employment Status', 'Job Roles'	Moderate	Implement rigorous data validation processes and conduct thorough feature engineering to enhance predictive accuracy for software salary prediction.
Kaggle Dataset	Categorical data in	Moderate	Encoding has to be

	the dataset		done in the data
--	-------------	--	------------------