

Data Collection and Preprocessing Phase

Date	15 March 2025
Team ID	SWTID1749896042
Project Title	Unemployed insurance beneficiary forecasting
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report summarizes data quality issues identified in the selected source, including severity levels and resolution plans. This ensures systematic identification and rectification of data discrepancies to maintain data integrity and improve model performance.

Data Source	Data Quality Issue	Severity	Resolution Plan
Data.World Unemployment Insurance Beneficiaries Dataset	Missing values in the 'Number of Beneficiaries' and 'Benefit Amount Paid' columns for some counties and years.	Moderate	Use mean/median imputation or forward-fill methods for missing numerical values to maintain time-series consistency.
Data.World Unemployment Insurance Beneficiaries Dataset	Inconsistent formatting of county names	Low	Apply string normalization techniques such as trimming spaces, converting to title case, and standardizing names across the dataset.