

Data Collection and Preprocessing Phase

Date	5 December 2024
Team ID	XXXXXX
Project Title	COVID – 19 Analysis using Tableau
Maximum Marks	3 Marks

Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
COVID-19 in India	Missing values in testing and vaccination columns	Moderate	Imputed missing values using forward fill or averages to maintain time-series consistency.
	Discrepancies in cumulative case counts across dates	High	Verified with official sources and corrected anomalies using updated records.
	Duplicate entries for specific dates and states	Moderate	Removed duplicate rows after verifying correctness of state and date combinations.
	Incorrect date formats causing sorting issues	Low	Standardized Date column to DateTime format for proper filtering and analysis.

	Inconsistent state names (e.g., abbreviations vs full names)	Moderate	Standardized state names using a mapping table to unify naming conventions.
COVID-19 Explained Through Visualizations	Missing data for certain regions and time periods	Moderate	Used interpolation to estimate missing values where trends were consistent.
	Unclear column labels in raw data	Low	Renamed columns for better clarity (e.g., Total Tests -> Cumulative Tests).
	Mixed data types in numerical columns (e.g., string with numbers)	High	Converted all numerical columns to proper integer/float format after removing non-numeric characters.
	Misaligned data for state-wise testing and vaccination figures	High	Re-aligned columns to ensure correct mapping of testing and vaccination data per state.