Data Collection and Preprocessing Phase

Date	5 December 2024
Team ID	Xxxxxx
Project Title	COVID – 19 Analysis using Tableau
Maximum Marks	10 Marks

Section	Description
Data Overview	**Dataset 1**: COVID-19 in India • Provides data on COVID-19 cases, deaths, recoveries, testing, and vaccinations across India. • Includes demographic, regional, and time-series data. **Dataset 2**: COVID-19 Explained Through Visualizations • Focuses on visual explanations and additional context for COVID-19 trends globally and in India. • Features detailed testing, recovery, and vaccination metrics.
Data Cleaning	• **Missing Values**: 1. Identified null entries in testing and vaccination columns. 2. Imputed missing values using forward fill or averages for time-series consistency. • **Duplicates**: Checked and removed duplicate entries based on unique identifiers like State and Date.

	• **Error Correction**: Verified discrepancies in cumulative case counts and corrected anomalies by cross-referencing official reports.
Data Transformation	• Filter out irrelevant columns, such as unrelated regional data. • Sort data chronologically by Date for trend analysis. • Pivot State data to compare metrics like testing rates and cases side by side. • Created calculated columns for metrics like Test Positivity Rate and Recovery Rate.
Data Type Conversion	• Converted Date to DateTime format for accurate filtering and aggregation. • Ensured numerical columns like Confirmed Cases and Deaths are in integer format.
Column Splitting and Merging	• **Splitting**: Split State-Test column (if present) into separate State and Test Count columns for clarity. • **Merging**: Combined First Dose and Second Dose data to create a Total Vaccination column.
Data Modeling	Created measures for: • Weekly totals (e.g., weekly cases, recoveries, deaths). • Percentages (e.g., percentage of vaccinated population).
Save Processed Data	• Documented data cleaning steps for reproducibility.