Data Collection and Preprocessing Phase

Date	18 March 2024
Team ID	LTVIP2025TMID26683
Project Title	Heritage Treasures: An In-Depth Analysis of UNESCO World Heritage Sites in Tableau
Maximum Marks	10 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	**Description**
Data Overview	This is the authoritative source for information on World Heritage Sites. The core data is typically structured, making it suitable for analysis in Tableau.The data for this project is primarily obtained from UNESCO, and contains a wealth of geographical, categorical, and temporal data. This data can be enriched by supplemental data from other sources. By understanding the data's characteristics, you can create a robust and insightful Tableau dashboard.
Data Cleaning	Handle missing values, duplicates, and correct errors as well as datatype for each category/field.
Data Transformation	This process involves reshaping, aggregating, and enriching the data to derive meaningful insights. Initially, the dataset may contain raw information about UNESCO World Heritage Sites, including site names, categories, countries, and years inscribed. Data transformation begins with aggregating relevant measures, such as counting the number of sites per country or calculating the time elapsed since inscription.
Data Type Conversion	The dataset, containing information about various UNESCO World Heritage Sites, often comes with fields that are initially assigned incorrect data types, hindering proper analysis. For instance, the "Year Inscribed" field, which should be treated as a numerical or date field, might be imported as a string. To rectify this, a conversion to an integer or date data type is necessary, enabling time-based analyses and calculations.
Column Splitting and Merging	column splitting and merging are essential data transformation techniques to refine and restructure the dataset for enhanced analysis. Column splitting is particularly useful when a single column contains multiple pieces of information that need to be separated. For example,

	if a "Location" column combines city, region, and country, splitting this column into individual "City," "Region," and "Country" columns allows for more granular filtering and analysis.
Data Modeling	This process involves creating relationships between different tables, defining hierarchies, and establishing appropriate data types. The core data model revolves around a central table containing information about each UNESCO World Heritage Site, including its name, category, country, and year inscribed.
Save Processed Data	After meticulously cleaning, transforming, and modeling the UNESCO World Heritage Sites data within Tableau, the next crucial step is to save the processed data for efficient future use and sharing. This involves exporting the refined dataset in a format that preserves the transformations and ensures compatibility with Tableau's visualization capabilities.