

DATA COLLECTION AND PREPROCESSING PHASE

Date	18 March 2025
Team ID	LTVIP2025TMID26712
Project Title	Heritage Treasures: An In-Depth Analysis of UNESCO World Heritage Sites in Tableau
Maximum Marks	10 Marks

Data Exploration and Preprocessing Template

Section	Description
Data Overview	To effectively analyze UNESCO World Heritage Sites within Tableau, a comprehensive data overview is essential, focusing on key dimensions and measures that reveal insightful patterns. Geospatial analysis forms a cornerstone, utilizing latitude and longitude to map site distributions, colour-coding by category (Cultural, Natural, Mixed), and sizing points by area to represent scale. Temporal trends emerge through line charts depicting inscription rates over time, highlighting shifts in site recognition. Categorical breakdowns, using bar charts and tree maps, quantify sites per country, continent, and UNESCO region, while word clouds and stacked bar charts illuminate the prevalence and interplay of UNESCO criteria.
Data Cleaning	The "Criteria" field often requires splitting and parsing due to its multi-value nature, enabling effective analysis of site classification. Year inscribed data must be checked for accuracy and consistency, while area measurements should be standardized to a single unit. Text fields like "Short Description" may need cleaning to remove extraneous characters and ensure readability. Duplicate records must be identified and eliminated to prevent skewed analysis. Finally, data type verification is essential, ensuring that numerical fields are recognized as numbers, date fields as dates, and geographical fields as geographical data.
Data Transformation	Transforming UNESCO World Heritage Sites data in Tableau involves manipulating the raw dataset to create meaningful insights and visualizations. A crucial transformation is parsing the "Criteria" field, which often contains multiple values separated by delimiters. Splitting this field into individual criteria allows for detailed analysis of the reasons for site inscription. Creating calculated fields is essential; for instance, deriving a "Decade Inscribed" field from the "Year Inscribed" column enables trend analysis over time.
Data Type Conversion	Data type conversion involves verifying and adjusting the inherent nature of each field to match its content. For example, the "Year Inscribed" column, initially often imported as a string, must be converted to an integer to facilitate numerical analysis and time-based visualizations. Latitude and longitude coordinates, which may be treated as strings, need to be converted to decimal numbers for accurate geospatial mapping. The "Area (ha)" field should also be verified as a numerical field. The "Criteria" field, if split into individual values, should remain as strings,

	as they represent categorical information.
Column Splitting and Merging	The "Criteria" column, often containing multiple criteria separated by delimiters (like commas or semicolons), necessitates splitting. This process involves dividing the single column into multiple columns, each representing an individual criterion. This transformation allows for detailed analysis of the frequency and co-occurrence of different criteria. Conversely, column merging can be used to create combined fields that enhance visualization and analysis. For example, merging "Latitude" and "Longitude" columns into a single "Location" field can simplify map creation.
Data Modeling	Creating calculated fields within the data model, such as decade from year or combined location fields, can enhance analysis. Hierarchical dimensions, like continent to country, facilitate drill-down explorations. Relationships between tables should be clearly defined, enabling Tableau to perform accurate aggregations and joins. This well-defined data model ensures efficient querying, accurate visualization, and insightful analysis of the UNESCO World Heritage Sites dataset
Save Processed Data	saving the processed data is crucial for future analysis and collaboration. Tableau offers several options for saving this refined dataset. For optimal performance and portability, exporting the processed data as a <code>.hyper</code> file is recommended. This format, optimized for Tableau's data engine, ensures fast query speeds and efficient data handling. Alternatively, saving the data as a <code>.csv</code> file allows for broader compatibility with other data analysis tools and platforms. If the transformations involved calculated fields or complex data modeling, saving the data source as a <code>.tds</code> or <code>.tdsx</code>