Data Collection and Preprocessing Phase

Date	18 March 2025
Team ID	LTVIP2025TMID26708
Project Title	Measuring the Pulse of Prosperity: An index of Economic Freedom Analysis
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	The analysis assesses the quality and consistency of the data used in the index. This involves examining the sources of data, the potential for bias, and the accuracy of the measurements. • Consideration is given to how well the data captures the realities of diverse economic environments.
Data Cleaning	When analysing the Index of Economic Freedom, "data cleaning" refers to the essential process of ensuring the accuracy, consistency, and reliability of the data used to construct the index. This is a crucial step in producing a valid and meaningful assessment of economic freedom. Economic data from various countries may have gaps or missing values. Data cleaning involves determining the appropriate methods for handling these gaps, such as imputation or exclusion.
Data Transformation	When considering "data transformation" within the context of the Index of Economic Freedom, it's about how raw data is processed and converted into the index's standardized scores. This process is crucial for creating a consistent and comparable measure of economic freedom across diverse economies. Raw data from various sources often uses different units and scales. Transformation standardizes this data, typically onto a common scale (often 0 to 100), allowing for direct comparisons. This process ensures that factors like tax rates, trade tariffs, and property rights assessments are all measured on a uniform basis.
Data Type Conversion	When discussing "data type conversion" in the context of the Index of Economic Freedom, we're focusing on how raw data, which comes in various forms, is transformed into a standardized numerical format suitable for inclusion in the index. The index draws data from a wide range of sources, including: • Government statistics (e.g., tax rates, government spending).

	This data comes in various formats: • Percentages (e.g., tax burden). **Standardization Requirement:** • To create a meaningful index, all these diverse data types must be converted into a common numerical scale.
Column Splitting and Merging	When analysing the Index of Economic Freedom, the concepts of "column splitting and merging" relate to how the raw data is organized and processed to create the final index scores. **Column Splitting:** • **Disaggregation of Data:** o The Index of Economic Freedom uses numerous data points to assess each country. Sometimes, raw data might be provided in a single column that contains multiple pieces of information. **Column Merging:** • **Aggregation of Data:** o Conversely, "column merging" involves combining data from multiple columns to create a composite score or a broader metric. o For example, after individual scores are calculated for "property rights," "judicial effectiveness," and "government integrity," these scores might be merged to create an overall "rule of law" score.
Data Modeling	When considering "data modelling" within the context of the Index of Economic Freedom, it's essential to understand how the raw economic data is structured and organized to create a meaningful and usable index. **Conceptual Modelling:** • This involves defining the core concepts of economic freedom and identifying the relevant factors to measure it. **Logical Modelling:** • This involves defining the relationships between the different data elements. **Physical Modelling:** • This involves determining how the data will be stored and accessed.
Save Processed Data	When considering the "saved data process" of the Index of Economic Freedom analysis project, it's crucial to understand how the collected, cleaned, and transformed data is stored, managed, and made accessible. **Data Storage:** Databases: Relational databases or other database systems to store structured data.