Purushottam Kumar Mandal 0187CS221157 CSE-3

Objective- Study and document the different phases of a data analytics project (Data Collection, Cleaning, Processing, Analysis, and Visualization).

Phases of a Data Analytics Project A data analytics project typically involves several interconnected phases: Data Collection, Cleaning, Processing, Anaysis, and Visualization. Each phase plays a critical role in transforming raw data into actionable insights. Below is a detailed study of these phases:

1. **Data Collection :-** This phase involves gathering data from various sources to address specific questions or objectives. Steps: Identify the questions or objectives the data needs to answer. Choose appropriate data collection methods, such as surveys, interviews, observations, or secondary sources like government reports. Determine the required amount of data and the sampling method (random, systematic, stratified). Ensure the data source is trustworthy and reliable. Tools: Online forms, APIs, databases, sensors, and manual collection methods.
2. **Data Cleaning :-** Data cleaning ensures the accuracy, consistency, and reliability of the dataset by removing errors and irrelevant information. Steps: Detect and correct inaccuracies such as missing values or duplicates. Fix structural errors (e.g., inconsistent formatting). Filter outliers and irrelevant observations. Automate repetitive cleaning tasks using scripts or tools. Importance: Clean data improves analysis accuracy, efficiency, and reliability. Tools: Python libraries (Pandas), R scripts, Excel.
3. **Data Processing : -** This phase transforms raw data into a usable format for analysis. Steps: Input cleaned data into systems like CRMs or databases. Apply algorithms to process the data based on its intended use (e.g., customer insights, trend analysis). Translate processed data into readable formats like tables or graphs. Importance: Enables structured interpretation of raw data for further analysis. Tools: SQL databases, cloud platforms (AWS), machine learning frameworks.
4. **Data Analysis :-** Data analysis involves applying statistical or computational techniques to extract insights from processed data. Types of Analysis: Descriptive Analysis: Summarizes historical data. Diagnostic Analysis: Explains why certain events occurred. Predictive Analysis: Forecasts future trends. Prescriptive Analysis: Suggests actions based on predictions. Process: Perform exploratory analysis to understand patterns in the dataset. Validate findings through statistical tests or modeling techniques. Tools: Python (NumPy, SciPy), R, Tableau, Power BI.
5. **Visualization :-** Visualization presents insights in a graphical format for easier interpretation by stakeholders. Methods: Use charts (bar graphs, pie charts), dashboards, and infographics to summarize findings visually. Focus on clarity and relevance to ensure

stakeholders can derive actionable insights quickly. Importance: Enhances decision- making by simplifying complex datasets into digestible formats. Tools: Tableau, Power BI, Matplotlib (Python).





Purushottam Kumar Mandal 0187CS221157 CSE-3



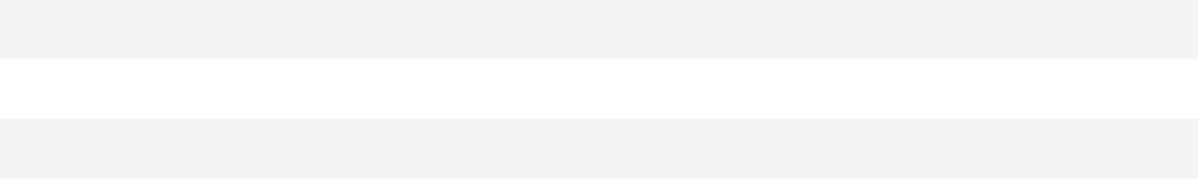
















Purushottam Kumar Mandal 0187CS221157 CSE-3

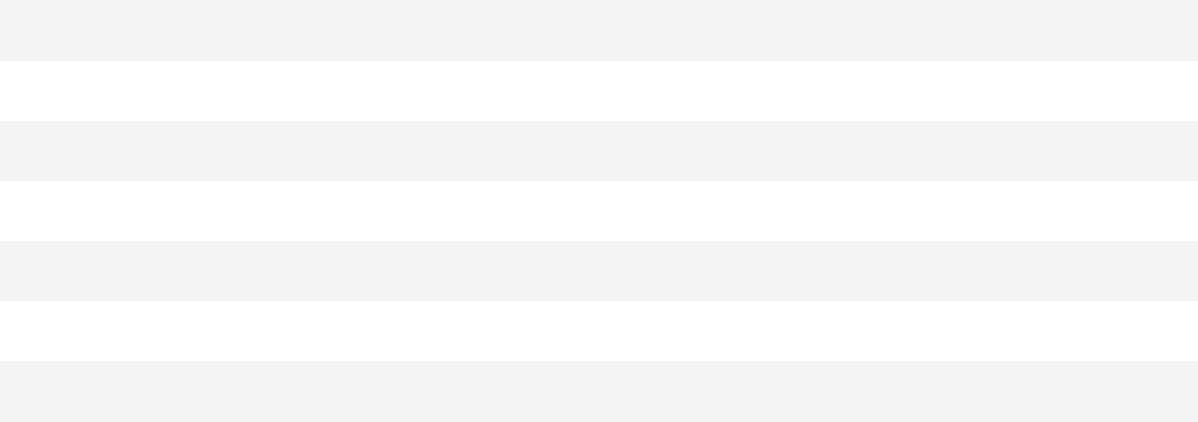












Purushottam Kumar Mandal 0187CS221157 CSE-3







Purushottam Kumar Mandal 0187CS221157 CSE-3









Purushottam Kumar Mandal 0187CS221157 CSE-3











Purushottam Kumar Mandal 0187CS221157 CSE-3













Purushottam Kumar Mandal 0187CS221157 CSE-3









Purushottam Kumar Mandal 0187CS221157 CSE-3



