



ASSIGNMENT TITLE

DATA CLEANING, TRANSFORMATION AND MODELLING USING POWER BI

Submitted By : Abhilasha Pareek

Course : Data Analytics With AI – September batch Live

Institute : PW Skills

Date : 12 January 2026

Email Id: abhilashapareek8974@gmail.com

Assignment Description

This assignment focuses on the concepts of **Data Cleaning, Data Transformation, and Data Modelling using Power BI**, which are critical stages in the data analytics and business intelligence lifecycle. The objective of this assignment is to understand how raw and unstructured data is prepared, refined, and organized before it is used for meaningful analysis and reporting.

The assignment explains the importance of data quality, accuracy, and consistency while working with real-world datasets. It also demonstrates how Power BI tools such as Power Query and Data Model View help in cleaning data, handling missing and duplicate values, and creating efficient data models. These practices ensure reliable dashboards and support effective decision-making.

Overall, this assignment provides a strong foundation in data preparation techniques that are widely used in professional data analytics and reporting environments.

Question 1: What is Data Loading in Power BI and why is it important?

Answer:

Data Loading in Power BI refers to the process of importing data from various sources such as Excel files, CSV files, databases, web sources, and cloud platforms into Power BI Desktop. This step makes the data available for further processing, analysis, and visualization.

Data loading is important because it acts as the **starting point of data analysis**. Without loading data, it is not possible to perform data cleaning, transformation, modeling, or reporting. Proper data loading ensures that the dataset is correctly structured and accessible for analysis.

In real-world business scenarios, organizations often work with data coming from multiple systems. Data loading allows analysts to combine these datasets into a single analytical platform, improving data visibility and consistency.

Question 2: Explain the difference between “Load” and “Transform Data” options in Power BI.

Answer:

The **Load** option in Power BI directly imports data into the Power BI data model without allowing any modifications. This option is suitable when the dataset is already clean, accurate, and properly structured.

The **Transform Data** option opens the Power Query Editor, where data can be cleaned and modified before loading. Using this option, analysts can remove duplicates, handle missing values, change data types, rename columns, and create new calculated columns.

Difference Summary:

Load	Transform Data
Loads data as-is	Allows data cleaning and transformation
Faster but limited	Slightly slower but flexible
No preprocessing	Full preprocessing possible

In practical applications, Transform Data is preferred because it improves data quality and ensures reliable analytical results.

Question 3: What is a Fact Table and a Dimension Table? Explain with examples.

Answer:

A **Fact Table** contains numerical or measurable data that is used for analysis, such as sales amount, confirmed cases, deaths, or revenue. Fact tables usually contain foreign keys that link to dimension tables.

A **Dimension Table** contains descriptive information that provides context to the data in the fact table, such as country, date, customer name, or product category.

Example:

- **Fact Table:** Confirmed Cases, Deaths, Recovered Cases
- **Dimension Table:** Country, State, Date, Vaccination Status

In Power BI, separating data into fact and dimension tables improves performance and makes the data model easier to understand and maintain.

Question 4: Why is Star Schema preferred over Snowflake Schema in Power BI?

Answer:

Star Schema is preferred in Power BI because it offers a **simple and efficient data model**. In a Star Schema, the fact table is directly connected to multiple dimension tables, forming a star-like structure.

Compared to Snowflake Schema, Star Schema:

- Improves query performance
- Reduces model complexity
- Is easier to understand and maintain
- Works efficiently with Power BI's DAX engine

Due to these advantages, Star Schema is commonly used in business intelligence solutions built using Power BI.

Question 5: Identify and remove duplicate records.

Answer:

Steps:

1. Open **Transform Data**
2. Select relevant columns
3. Click **Remove Rows** → **Remove Duplicates**
4. Apply the changes

Explanation:

Duplicate records can lead to incorrect calculations and misleading reports. Removing duplicates improves data accuracy and ensures that each record represents a unique observation.

Question 6: Handle missing or null values in the dataset.

Answer:

Steps:

1. Select the column containing null values
2. Choose **Replace Values**
3. Replace nulls with appropriate values such as "Unknown" or "Not Available"

Explanation:

Handling missing values improves data completeness and prevents errors during analysis. It also ensures smoother report generation and better visualization results.

Question 7: Create a calculated column (e.g., Recovery Rate).

Answer:

Formula:

$\text{Recovery Rate} = (\text{Recovered} / \text{Confirmed}) * 100$

Explanation:

Calculated columns help in deriving meaningful insights from existing data. Recovery Rate is a useful metric for comparing performance across different regions or time periods.

Question 8: Create a summarized table using Group By.

Answer:

Steps:

1. Open **Power Query**
2. Select relevant columns
3. Click **Group By**
4. Apply aggregation (Sum / Count)

Explanation:

Summarized tables help reduce data volume and make analysis more efficient. Grouping data improves clarity and supports comparative analysis.

Conclusion:

This assignment provided a comprehensive understanding of data cleaning, transformation, and modeling using Power BI. By combining theoretical concepts with practical implementation, the importance of data quality and structured data models was clearly demonstrated. Clean and well-organized data plays a crucial role in producing accurate insights, reliable dashboards, and effective business decisions. Overall, this assignment strengthened practical skills required for real-world data analytics and business intelligence applications.