NAME :ANJALI A S

ANJALI A S-G18SQL&POWER BI

ASSIGNMENT 1

Q1.What is Power Query , and what role does it play in Power BI?

Power Query is a data connection and transformation tool developed by Microsoft . I t is used in Excel and Power BI . Power query helps to bring data from different places ,clean it up and get it ready for analysis .

In Power BI ,Power query plays crucial role in data preparation step like connect to many different data sources , remove unwanted rows and columns , change data types ,split or merge columns ,merging multiple tables ,grouping data by categories ,creating calculated columns etc .

Q2.How do you connect to different data sources using Power Query ?

In Power Query, connect to different data sources by using the "Get Data" option, which allows select from various categories such as files, databases, and online services. After selecting the desired source, we can import the data and utilize the Power Query Editor to perform necessary transformations like filtering and merging. Finally, we can load the cleaned data into Excel or Power BI for analysis and reporting.

Q3.Can you explain the process of data transformation in Power Query ?

Power Query involves several key steps that help us clean and prepare our data for analysis.

First, load our data into the Power Query Editor, where we can view and manipulate it .Next , we can perform various transformations such as removing unnecessary columns, filtering rows, changing data types, and merging or appending tables. Power Query provides is a user-friendly interface with options to apply these transformations without needing to write code.

Q4.What are some common data transformation operations you can perform on a Power Query ?

1)Filtering rows and columns : Remove unwanted rows based on specific criteria like the values

2)Changing Data Types: Convert data types of columns .

3)Removing columns : Remove unnecessary columns to simplify your dataset and focus on relevant information.

4)Merging Queries :Combine datas from multiple tables based on a common column.

5)Appending Queries :combining two or more tables with similar structures into a single table.

6)Grouping Data : Grouping data in Power Query allows us to summarize and aggregate our data based on one or more columns.

7)Pivoting and Unpivoting :Reshape our data by pivoting columns into rows or unpivoting rows into columns, which helps in organizing data for analysis.

8)Conditional columns : Making columns based on some conditions applied to the exsisting columns.

9)Sorting Data : Organize data in ascending or descending order based on one or more columns.

Q5)How do you handle missing or duplicate data in Power Query ?

Handling missing data: Handling missing data means finding and fixing gaps in our information. In Power Query, we can fill in missing values, remove rows with empty spots, or use nearby values to fill in the blanks. This helps make sure our data is complete and reliable.

Handling duplicate data : Handling duplicate data means finding and dealing repeated entries in the data . In Power Query, we can remove duplicates to keep only unique items or group them to summarize the data.

Q6) Explain the difference between the appending and merging queries in power query ?

Appending queries means putting two or more tables on top of each other to make one bigger table with more rows. This is helpful when we have different datasets that have the same columns and want to combine them into one.

Merging queries involves joining two or more tables based on a common column to create a new table that combines columns from both tables. This is similar to a SQL join operation.

Q7)What are the conditional columns , and how do you create them in Power Query ?

Conditional columns in Power Query allow us to create new columns based on specific conditions or rules applied to existing columns.

In the Add Column tab, select Conditional Column, then enter the new column name. Next, choose the column to base your conditions on, set the conditions, and finally click OK.

Q8)Describe the purpose of index columns in Power Query ?

Row identification : easily find or reference specific rows in our data, making it simpler to track changes or analyze specific entries.

Sorting and Ordering : After filtering or rearranging our data, wecan use the index to restore the original order if needed.

Joining and Merging :When combining data from multiple sources, we can use the index to ensure that the correct rows match up.

Tracking Changes : If we update our data, the index allows us to see what has changed by comparing the index numbers.

Q9)How can you create a calendar table using a Power Query ?

Open the Power Query Editor and creating a blank query. Within the Advanced Editor, input code that defines a start and end date for our calendar range. This code then generates a list of all dates within that period and transforms it into a table with a "Date" column. Following this, the code adds various calculated columns, automatically extracting useful date parts like the year, month name, day of the week, and quarter. This resulting calendar table acts as a central, organized source of date information that we can connect to our other data, it simplify time-based analysis .

Q10)What are the best practices for optimizing data loading and transformation in Power Query ?

Filter Early and Often**:** Apply filters as early in our query steps as possible to reduce the amount of data Power Query needs to load and process and it also reduces time .

Select Only Necessary Columns**:** Remove any columns we don't need right at the source. Less number of columns mean less data to load into memory and process through each transformation step.

Optimize Data Types: Ensure columns have the correct data types. Incorrect types can lead to unexpected errors. We can change the datatype using "Change Type."

Avoid Row-by-Row Operations: Power Query works best with column-level operations. Try to avoid operations row by row using complex conditional column within "Add Custom Column”.

Optimize Data Source Connection: Ensure your connection to the data source is efficient and reliable. Slow network connections or poorly performing data sources will naturally impact loading times.