ANKKIT KUMAR GUPPTA

DATA ANALYST | PROMPT ENGINEER

DUPLICATE

VS

REFERENCE

In Power BI's Power Query, **DUPLICATE** & **REFERENCE** are two options for creating new queries based on an existing query, but they serve different purposes

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

DUPLICATE

- **Functionality:** Creates an exact copy of the original query, including all the applied steps.
- **Independence:** The new query is independent of the original. Any changes made to the duplicate query will not affect the original query, and vice versa.
- **Use Case:** Use Duplicate when you want to make significant changes to a query without altering the original query. This is helpful when experimenting with different transformations or creating variations of the same data.

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

EXAMPLE 1: USING DUPLICATE

Goal:

Create two different reports, one for Top 10 Products by Sales and another for Sales by Region, without affecting the original data.

1. Duplicate the "SalesData" table:

- Right-click on the "SalesData" query in Power Query & select Duplicate.
- A new query named "SalesData (2)" is created, which is an exact copy of the original "SalesData".

2. Modify "SalesData (2)" to show Top 10 Products by Sales:

- In "SalesData (2)", remove columns that are not needed, such as Region.
- Sort the table by SalesAmount in descending order.
- Keep only the top 10 rows.

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

EXAMPLE 1: USING DUPLICATE

3. Create another Duplicate for Sales by Region:

- Right-click on the "SalesData (2)" and select Duplicate again.
- Rename it to "SalesByRegion".
- Remove columns that are not needed, such as Product.
- Group by Region to summarize the SalesAmount by each region.

Result:

- "SalesData (2)" shows the Top 10 Products by Sales.
- "SalesByRegion" shows the total SalesAmount for each Region.
- Both queries are independent; changes in "SalesData" will not affect these duplicates, and modifications in one duplicate will not impact the other.

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

REFERENCE

- **Functionality:** Creates a new query that references the original query as its source. It does not copy the steps but rather points to the result of the original query.

- **Dependence:** The reference query is dependent on the original query. If you make changes to the original query, those changes will affect the reference query as well.

- **Use Case:** Use Reference when you need to build a new query that starts from the same point as an existing query but might require additional transformations. This is useful for maintaining a single data source with different views or calculations.

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

EXAMPLE 2: USING REFERENCE

Goal:

Analyze Monthly Sales Trends and Sales Growth by Region based on the same underlying data, ensuring that any changes to the data preparation in the original query are reflected in both analyses.

1. Reference the "SalesData" table:

- Right-click on the "SalesData" query and select Reference.
- A new query named "SalesData (2)" is created. It references the results of the original "SalesData" query.

2. Modify "SalesData (2)" for Monthly Sales Trends:

- Rename the query to "MonthlySalesTrends".
- Add a new column to extract the Month from the Date.
- Group by Month and calculate the sum of SalesAmount for each month.

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

EXAMPLE 2: USING REFERENCE

3. Create another Reference for Sales Growth by Region:

- Right-click on the "SalesData" query again and select Reference.
- Rename the query to "SalesGrowthByRegion".
- Group by Region and calculate the percentage growth in SalesAmount compared to the previous period.

Result:

- "MonthlySalesTrends" shows the total SalesAmount by Month.
- "SalesGrowthByRegion" shows the Sales Growth by Region.
- Both queries reflect any changes made to the "SalesData" query since they are referencing it. If you, for instance, filter out certain products in the original "SalesData" query, those changes will be reflected in both referenced queries.

ANKKIT KUMAR GUPPTA
DATA ANALYST | PROMPT ENGINEER

SUMMARY

DUPLICATE allows you to create separate, independent analyses, useful when you need different transformations or views of the data.

REFERENCE is ideal when you need to maintain consistency across different analyses based on the same underlying data and want changes to the base query to propagate automatically.