Gold fact sales – Exploratory data analysis

Database exploration is the process of understanding the database structure, identifying key tables, columns, and relationships, examining data types and constraints, reviewing sample data, and preparing the dataset for tasks such as reporting, analysis, or data cleaning.

```
-- Explore all the objects in the database
select * from INFORMATION_SCHEMA.TABLES

-- what schema exits in the database?
Select distinct TABLE_SCHEMA
from INFORMATION_SCHEMA.TABLES;

-- Explore all the columns in the database
select * from INFORMATION_SCHEMA.COLUMNS
where TABLE_NAME = 'gold.fact_sales'
```

Dimensions Exploration is the process of analyzing, understanding, and profiling dimension tables within a data warehouse or reporting database.
This process helps in:

- Identifying **key attributes**
- Assessing **cardinality** (number of unique values)
- Evaluating **data quality**
- Determining if **hierarchies** are defined
- Understanding how **dimensions are related to fact tables** through joins or foreign keys

It ensures that dimensions are reliable, well-structured, and ready for accurate reporting and analysis.

```
-- Explore all the tables
select top 3 * from [dbo].[gold.fact_sales]
select top 3 * from [dbo].[gold.report_customers]
select top 3 * from  [dbo].[gold.report_products]
select top 3 * from  [dbo].[gold.dim_customers]
select top 3 * from   [dbo].[gold.dim_products]

-- How many countries are involved?
select distinct [country]
from [dbo].[gold.dim_customers]                    -- 6 countries+ 1 n/a
```

-- Explore the category of the data, which is the major division as per our data

-- Level 1 -- How many categories??

```
Select distinct [category]
from [dbo].[gold.dim_products]                -- 4 categories + 1 n/a
```

-- Level 2 -- How many sub-categories??

```
select distinct [subcategory]
from [dbo].[gold.dim_products]                -- 36 Sub-categories + 1 n/a
```

-- Level 3 -- How many Products??

```
select distinct [product_name]
from [dbo].[gold.dim_products]                -- 295 Products
```

-- checking all together

```
select distinct [category],
                [subcategory],
                [product_name]
from [dbo].[gold.dim_products]
```

Dates Exploration is the process of time-based profiling to understand trends, perform growth or loss analysis, and support time-based reporting.
It involves:

- Identifying the **earliest and latest dates** to define the time boundaries of the data
- Analysing **temporal patterns**, such as seasonality, trends, and gaps
- Supporting **date-based slicing and dicing** in reports and dashboards

-- Find the date of first and last order

```
select min([order_date]) as First_Order,
       Max([order_date]) as Last_order
from [dbo].[gold.fact_sales]
```

-- How many years of data is avalible?

```
select datediff(Year, Min([order_date]), Max([order_date])) as Years
from [dbo].[gold.fact_sales]
```

-- Check the customer age range

```
select min([birthdate]) Oldest_customer,
       Max([birthdate]) as Youngest_customer
```

```
from [dbo].[gold.dim_customers]
```

Measures Exploration is the process of analysing and understanding quantitative metrics used to track business performance.
It involves:

- Identifying **what can be measured and analysed**
- Understanding the **correct aggregation method** for each measure (e.g., sum, average, count)
- Ensuring **data consistency** across reports and systems
- Maintaining **accuracy in reporting** and decision-making

This process helps ensure that key performance indicators (KPIs) and other metrics reflect the true state of the business.

```
-- Write all the important business metrics
  select 'Total_sales' as Measure_name,
        sum([sales_amount]) as Total_sales
  from [dbo].[gold.fact_sales]
    union all
    select 'Average_selling_Price' as Measure_name,
    avg([price]) as Avg_selling_price
     from [dbo].[gold.fact_sales]
    union all
    select 'Total_Qunatity' as Measure_name,
    sum([quantity]) as Total_Qunatity
     from [dbo].[gold.fact_sales]
    union all
    select 'Total_Customers' as Measure_name,
    count( distinct [customer_id] ) as Total_Customers
    from [dbo].[gold.dim_customers]
    union all
    select 'Total_Products' as Measure_name,
    count ( distinct [product_id] ) as Total_Products
     from [dbo].[gold.dim_products]
      union all
```

```
select 'Total_Categories' as Measure_name,
 count (distinct [category_id] ) as Total_Categories
   from [dbo].[gold.dim_products]
     union all
select 'Total_Sub_categories' as Measure_name,
 count(distinct [subcategory])  as Total_sub_categories
  from [dbo].[gold.dim_products]
```

-- write all the key Sales- metrics in the business

```
select 'Total_sales' as Measure_name,
sum([sales_amount]) as Measure_value
from [dbo].[gold.fact_sales]
union all
select 'Total_Orders' as Measure_name,
count(distinct [order_number]) as Total_Orders
from [dbo].[gold.fact_sales]
Union all
select 'Total_Qunatity' as Measure_name,
sum([quantity]) as Total_Qunatity_sold
from [dbo].[gold.fact_sales]
union all
 select 'Average_order_value' as Measure_name,
 ( sum([sales_amount]) / count(distinct [order_number])) as AOV
from [dbo].[gold.fact_sales]
union all
 select 'Average_selling_price' as Measure_name,
(sum([sales_amount])/sum([quantity]))
 from [dbo].[gold.fact_sales]
```

Magnitude Measures Exploration involves analyzing calculations that help understand the scale and size of the business.

These measures are typically summed, counted, or averaged to provide insights into overall volume, value, or frequency within the data.

```
-- Find the total customers by countries
  select [country], count(distinct [customer_id]) as Total_customers
        from [dbo].[gold.dim_customers]
        group by [country]
-- Find the total customers by gender
    select [gender], count(distinct [customer_id]) as Total_customers
        from [dbo].[gold.dim_customers]
        Group by [gender]
-- Find the total products by category
    select [category],count( distinct [product_id]) as Total_products
        from [dbo].[gold.dim_products]
        group  by [category]
-- Find the average costs in each category
    select [category], Avg([cost]) as Average_cost
        from [dbo].[gold.dim_products]
        group  by [category]
-- Find total revenue generated by each category
    select c.[category], sum(s.[sales_amount]) as Toal_revenue
        from [dbo].[gold.dim_products] c Left join [dbo].[gold.fact_sales] s
        on c.[product_key] = s.[product_key]
        Group by c.[category]
-- What is the total value generated by each customer?
    select c.[customer_id], c.[first_name],sum(s.[sales_amount]) as Total_revenue_by_customer
        from [dbo].[gold.dim_customers] c left join [dbo].[gold.fact_sales] s
        on c.[customer_key] = s.[customer_key]
        Group by c.[customer_id],c.[first_name]
-- What is the disturbution of items sold across countries?
    select c.[country], sum(s.[quantity]) as Total_quantity_across_Countries
        from [dbo].[gold.dim_customers] c left join [dbo].[gold.fact_sales] s
```

```
on c.[customer_key] = s.[customer_key]

Group by c.[country]

Order by sum(s.[quantity]) desc
```

Ranking Analysis Exploration involves calculating and ordering data points based on specific metrics to identify top or bottom performers.
This helps highlight leaders, laggards, and relative positions within the dataset for better decision-making and insights.

```
-- What 5 products generate highest revenue
 select top 5 p.[product_id], p.[product_name], sum(s.[sales_amount]) as Sales_amount

from [dbo].[gold.dim_products] p left join [dbo].[gold.fact_sales] s

on p.[product_key] = s.[product_key]

group by p.[product_id], p.[product_name]

order by sum(s.[sales_amount]) desc

-- What 5 products generate lowest revenue
 select top 5 p.[product_id], p.[product_name], sum(isnull(s.[sales_amount],0)) as
Sales_amount

from [dbo].[gold.dim_products] p right join [dbo].[gold.fact_sales] s

on p.[product_key] = s.[product_key]

group by p.[product_id], p.[product_name]

order by sum(isnull(s.[sales_amount],0)) asc
```

----------EOD----------