# **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------------------------------------**