-- Delete useless Columns

ALTER TABLE [dbo].[DataCoSupplyChain]

DROP COLUMN [Customer\_Password], -- Useless values

[Customer\_Email], -- Useless values

[Order\_Item\_Cardprod\_Id], --duplicated with Product card ID

[Product\_Description], -- Useless values

[Order\_Profit\_Per\_Order], --duplicated with benifit per order

[Product\_Image], -- Useless values

[Product\_Status], -- Useless values

[Order\_Zipcode], -- Useless values

[Order\_Customer\_Id], --duplicated with customer ID

[Order\_Item\_Total], --duplicated with Sales per Customer

[Order\_Item\_Product\_Price], --duplicated with Product Price

[Category\_Id] --duplicated with Product Category ID

 -- Checking missing data in some necessary columns shouldn't be null

SELECT \* FROM [dbo].[DataCoSupplyChain] WHERE

[Sales\_per\_customer] IS NULL OR

[Delivery\_Status] IS NULL OR

[Late\_delivery\_risk] IS NULL OR

[Category\_Id] IS NULL OR

[Category\_Name] IS NULL OR

[Customer\_Id] IS NULL OR

[Department\_Id] IS NULL OR

[Department\_Name] IS NULL OR

[Order\_Id] IS NULL OR

[Order\_Item\_Discount] IS NULL OR

[Order\_Item\_Discount\_Rate] IS NULL OR

[Order\_Item\_Id] IS NULL OR

[Sales] IS NULL OR

[Order\_Item\_Total] IS NULL OR

[Product\_Category\_Id] IS NULL OR

[Product\_Name] IS NULL OR

[Product\_Price] IS NULL

-- Checking Dublication data

SELECT [Order\_Id],[Order\_Item\_Id], COUNT(\*) AS DuplicateCount

FROM [dbo].[DataCoSupplyChain]

GROUP BY [Order\_Id],[Order\_Item\_Id]

HAVING COUNT(\*) > 1;

-- Change Payment type 'payment' to Other

UPDATE [dbo].[DataCoSupplyChain]

SET [Type]= case

when [Type]='PAYMENT'then 'OTHER'

else [Type]

end;

-- Get Total Sales , total orders , Total Quanntaty , total profit

-- total Discount , AVG Sales , total Customers

SELECT CONCAT(ROUND((SUM([Sales])) / 1000000, 2), ' Million')

AS Total\_Sales ,

COUNT(DISTINCT [Order\_Id]) AS Total\_Orders ,

Sum([Order\_Item\_Quantity]) AS Total\_quantity ,

CONCAT(ROUND((SUM([Benefit\_per\_order])) / 1000000, 2), ' Million')

AS Total\_Prorfit ,

CONCAT(ROUND((SUM([Order\_Item\_Discount])) / 1000000, 2), ' Million')

AS Total\_Discount ,

ROUND ((SUM([Sales]) / count (DISTINCT [Customer\_Id]) ), 2)

AS AVG\_Sales\_per\_customer ,

COUNT(DISTINCT [Customer\_Id]) AS Count\_Of\_Customers

FROM [dbo].[DataCoSupplyChain]

--top 10 total sales , orders and customers per country

SELECT top 10[Order\_Country],

CONCAT(ROUND((SUM([Sales])) / 1000, 2), ' K') AS Total\_Sales ,

COUNT(DISTINCT [Order\_Id]) AS Total\_Orders , COUNT( DISTINCT [Customer\_Id]) AS Customer\_num

FROM [dbo].[DataCoSupplyChain]

GROUP BY [Order\_Country]

ORDER BY ROUND((SUM([Sales])) / 1000, 2) DESC;

--total sales , orders and customers per market

SELECT [Market],

CONCAT(ROUND((SUM([Sales])) / 1000, 2), ' K') AS Total\_Sales ,

COUNT(DISTINCT [Order\_Id]) AS Total\_Orders , COUNT( DISTINCT [Customer\_Id]) AS Customer\_num

FROM [dbo].[DataCoSupplyChain]

GROUP BY [Market]

ORDER BY ROUND((SUM([Sales])) / 1000, 2) DESC;

--total sales , orders and customers per Order Region

SELECT [Order\_Region],

CONCAT(ROUND((SUM([Sales])) / 1000, 2), ' K') AS Total\_Sales ,

COUNT(DISTINCT [Order\_Id]) AS Total\_Orders , COUNT(DISTINCT [Customer\_Id]) AS Customer\_num

FROM [dbo].[DataCoSupplyChain]

GROUP BY [Order\_Region]

ORDER BY ROUND((SUM([Sales])) / 1000, 2) DESC;

--top 10 total sales and orders per CAT

Select top 10 [Category\_Name],

CONCAT(ROUND((SUM([Sales])) / 1000, 0), ' K') AS Total\_Sales ,

COUNT(DISTINCT [Order\_Id]) AS Total\_Orders

FROM [dbo].[DataCoSupplyChain]

group by [Category\_Name]

ORDER BY ROUND((SUM([Sales])) / 1000, 2) DESC;

--top 10 total sales and orders per Order Status

Select [Order\_Status],

CONCAT(ROUND((SUM([Sales])) / 1000, 0), ' K') AS Total\_Sales ,

COUNT(DISTINCT [Order\_Id]) AS Total\_Orders

FROM [dbo].[DataCoSupplyChain]

group by [Order\_Status]

ORDER BY ROUND((SUM([Sales])) / 1000, 2) DESC;

-- most common payment method that Customers used to pay

SELECT [Type], count (\*)

as Number\_of\_using ,

CONCAT(CAST(COUNT(\*) \* 100.0 / SUM(COUNT(\*)) OVER () AS DECIMAL(5, 0)), ' %')

AS Percentage

FROM [dbo].[DataCoSupplyChain]

group by [Type]

order by Number\_of\_using desc

-- most common status of delivery

SELECT [Delivery\_Status], COUNT(\*)

AS Count\_of\_Status,

CONCAT(CAST(COUNT(\*) \* 100.0 / SUM(COUNT(\*)) OVER () AS DECIMAL(5, 0)), ' %')

AS Percentage

FROM [dbo].[DataCoSupplyChain]

GROUP BY [Delivery\_Status]

ORDER BY Count\_of\_Status DESC;

-- most common Shipping Mode

SELECT [Shipping\_Mode], COUNT(\*) AS Count\_of\_mode,

CONCAT(CAST(COUNT(\*) \* 100.0 / SUM(COUNT(\*)) OVER () AS DECIMAL(5, 0)), ' %')

AS Percentage

FROM [dbo].[DataCoSupplyChain]

GROUP BY [Shipping\_Mode]

ORDER BY Count\_of\_mode DESC;

-- add new column contain year from Shipping date column

-- get Sales and total orders per Year

ALTER TABLE [dbo].[DataCoSupplyChain]

ADD year\_column INT;

UPDATE [dbo].[DataCoSupplyChain]

SET [year\_column] = YEAR ([shipping\_date\_DateOrders]) ;

SELECT [year\_column],

CONCAT(ROUND((SUM([Sales])) / 1000000, 2), ' Milion')

AS Total\_Sales\_Per\_year , COUNT(\*) as Total\_Orders\_Per\_year

FROM [dbo].[DataCoSupplyChain]

GROUP BY [year\_column]

ORDER BY ROUND((SUM([Sales])) / 1000, 2) DESC;