--1) DATA CLEANING

--Step 1: To understand the data I am working with, I had to view all.

Select \* from SuperStore;

--Step 2: I removed an unnecessary column.

Alter table SuperStore

Drop column Row\_ID;

--Step 3: Checked for duplicate values.

Select count (distinct order\_id) as single,

Count (order\_id) as duplicate

From SuperStore;

--Step 4: I deleted duplicate values

With delete\_duplicate as (

Select

\* ,

ROW\_NUMBER() over (partition by order\_id order by order\_id) as Number

From SuperStore)

Delete from delete\_duplicate

Where Number > 1;

--2) Find the Total Avg Revenue per Customer and Total Avg Revenue per Product

Select top 1

(Select round (sum(sales)/ count(distinct Customer\_ID), 2)

From SuperStore) as Avg\_Rev\_Customer,

(Select round (sum(sales)/ count(distinct Product\_ID), 2)

From SuperStore) as Avg\_Rev\_Product

From SuperStore;

--3) find the Products that generated revenue below the Average Revenue per Product in 2017

Select

product\_name, category, region, (sales) as Revenue

From

SuperStore

Where

(Sales) < (select sum(sales)/ count(distinct Product\_ID)

From SuperStore)

And

DATEPART (year, ship\_date) = ‘2017’

Order by Revenue asc;

--3b) How many of such products fall under the three product categories?

Alter View Products\_below\_AvgRev as

Select Product\_name, Category, Region

From

(Select

product\_name, category, region, (sales) as Revenue

From

SuperStore

Where

(Sales\*quantity) < (select sum(sales)/ count(distinct Product\_ID)

From SuperStore)

And

DATEPART (year, ship\_date) = '2017'

) AS TAB;

Update Products\_below\_AvgRev

Set Category = 'Office\_Supplies'

Where Category = 'Office supplies';

Select \* from Products\_below\_AvgRev

PIVOT

(Count (Product\_name) for category in (Furniture, Office\_Supplies, Technology)) as PBA\_pivot;

--NB: After saving my result, I had to re-update my view and change "Office\_Supplies" to "Office Supplies" as it originally was. This is because, whatever change you make in the View also affects the Source table, and I do not want to mess up the category column in the Source table.

--4) Calculate the Y-O-Y growth rate from 2014 - 2017

With Y\_O\_Y (YOY\_2014, YOY\_2015, YOY\_2016, YOY\_2017) AS (

Select

(Select sum(sales)

From SuperStore

Where DATEPART (year, ship\_date) = '2014'),

(Select sum(sales)

From SuperStore

Where DATEPART (year, ship\_date) = '2015'),

(Select sum(sales)

From SuperStore

Where DATEPART (year, ship\_date) = '2016'),

(Select sum(sales)

From SuperStore

Where DATEPART (year, ship\_date) = '2017')

From SuperStore)

Select TOP 1

Concat(round(((YOY\_2015/YOY\_2014) - 1), 5) \* 100, '%') AS YOY\_2015,

Concat(round(((YOY\_2016/YOY\_2015) - 1), 5) \* 100, '%') AS YOY\_2016,

Concat(round(((YOY\_2017/YOY\_2016) - 1), 5) \* 100, '%') AS YOY\_2017

From Y\_O\_Y;

--5) What is the Total Cost of goods sold (COGS) compared to the Revenue in each region from 2014 - 2018?

Alter view storesales as

Select DATEPART (year, ship\_date) as year, region,

(Sales) - profit) as COGS

From superstore;

Create view storesales\_rev as

Select DATEPART (year, ship\_date) as year, region,

(Sales) as Revenue

From superstore;

With COGS\_cte as (

Select \* from StoreSales

PIVOT

(Sum(COGS) for region in (East, West, South, Central)) as COGS\_pivot),

Rev\_cte as (

Select \* from StoreSales\_rev

PIVOT

(Sum(revenue) for region in (East, West, South, Central)) as Rev\_pivot)

Select COGS\_cte.year,

Rev\_cte.East as Revenue\_East, COGS\_cte.East as COGS\_East,

Rev\_cte.West as Revenue\_West, COGS\_cte.West as COGS\_West,

Rev\_cte.South as Revenue\_South, COGS\_cte.South as COGS\_South,

Rev\_cte.Central as Revenue\_Central, COGS\_cte.Central as COGS\_Central

From COGS\_cte

Inner join Rev\_cte on COGS\_cte.year = Rev\_cte.year

Group by COGS\_cte.year, COGS\_cte.East, Rev\_cte.East, COGS\_cte.West, Rev\_cte.West, COGS\_cte.South, Rev\_cte.South, COGS\_cte.Central, Rev\_cte.Central;

--6) Return a table that shows profit by state and region.

Select State, Region, round(Profit, 5) as profit,

Sum(Profit) over (partition by Region) as Regional\_Profit

From SuperStore

Where DATEPART (year, ship\_date) = '2018';

7) What is the total number of orders we receive on each day of the week for each year? (From 2014-2017)

With Order2014 as (

Select

DATENAME(weekday, Order\_date) as Day\_of\_Week,

Count(Order\_ID) as Orders14

From

SuperStore

Where

datepart(year, Order\_Date) = 2014

Group by

DATENAME(weekday, Order\_date)),

Order2015 as (

Select

DATENAME(weekday, Order\_date) as Day\_of\_Week,

Count(Order\_ID) as Orders15

From

SuperStore

Where

Datepart(year, Order\_Date) = 2015

Group by

DATENAME(weekday, Order\_date)),

Order2016 as (

Select

DATENAME(weekday, Order\_date) as Day\_of\_Week,

Count(Order\_ID) as Orders16

From

SuperStore

Where

Datepart(year, Order\_Date) = 2016

Group by

DATENAME(weekday, Order\_date)),

Order2017 as (

Select

DATENAME(weekday, Order\_date) as Day\_of\_Week,

Count(Order\_ID) as Orders17

From

SuperStore

Where

Datepart(year, Order\_Date) = 2017

Group by

DATENAME(weekday, Order\_date))

Select a.day\_of\_week, a.orders14, b.orders15, c.orders16, d.orders17

From

Order2014 a, Order2015 b, Order2016 c, Order2017 d

Where a.Day\_of\_Week = b.Day\_of\_Week

And a.Day\_of\_Week = c.Day\_of\_Week

And a.Day\_of\_Week = d.Day\_of\_Week;