SUPERMARKET SALES ANALYSIS

Using snowflake,excel,powerbi

Fathima R

OBJECTIVE :

To analyze,

- Sales and profit
- Product and Category performance
- Customer Segmentation

STEP 1: EXTRACT

Dataset: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final

Shape of the dataset: 9800 observations and 20 variables

Variables Description:

- Order ID => Unique Order ID for each Customer.
- Order Date => Order Date of the product.
- Ship Date => Shipping Date of the Product.
- Ship Mode=> Shipping Mode specified by the Customer.
- Customer ID => Unique ID to identify each Customer.
- Customer Name => Name of the Customer.
- Segment => The segment where the Customer belongs.
- Country => Country of residence of the Customer.
- City => City of residence of of the Customer.
- State => State of residence of the Customer.
- Postal Code => Postal Code of every Customer.
- Region => Region where the Customer belong.
- Product ID => Unique ID of the Product.
- Category => Category of the product ordered.
- Sub-Category => Sub-Category of the product ordered.
- Product Name => Name of the Product
- Sales => Sales of the Product.
- Quantity => Quantity of the Product.
- Discount => Discount provided.
- Profit => Profit/Loss incurred.

STEP 2:LOAD

Load the data into excel and change the datatype of certain columns like

- Ordered date and shipped date to date datatype
- Sales and Profit column to currency format
- Discount to percentage format

Snowflake:

- In snowflake create warehouse, database and the stage
- Load the cleaned csv format in the internal stage
- Create table in the database by using the following query

```
create or replace TABLE PRACTICE.PUBLIC.SUPERSTORE (
    ORDER_ID VARCHAR(40),
    ORDER_DATE DATE,
    SHIP_DATE DATE,
    SHIP_MODE VARCHAR(40),
    CUSTOMER_ID VARCHAR(40),
    CUSTOMER_NAME VARCHAR(40),
    SEGMENT VARCHAR(40),
    COUNTRY VARCHAR(40),
    CITY VARCHAR(40),
    STATE VARCHAR(40),
    POSTAL_CODE VARCHAR(40),
    REGION VARCHAR(40),
    PRODUCT_ID VARCHAR(40),
    CATEGORY VARCHAR(40),
    SUB_CATEGORY VARCHAR(40),
    PRODUCT_NAME VARCHAR(40),
    SALES FLOAT,
    QUANTITY FLOAT,
    DISCOUNT FLOAT,
    PROFIT FLOAT
);
```

- Load the data in the internal stage into the table named superstore
- To check whether the data is loaded properly,

```
select * from superstore limit 5;
```

Output for the following query will be

	ORDER_ID	ORDER_DATE	SHIP_DATE	SHIP_MODE	CUSTOMER_ID	CUSTOMER_NAME	SEGMENT ···
1	CA-2016-152156	2017-11-08	2017-11-11	Second Class	CG-12520	Claire Gute	Consumer
2	US-2015-108966	2016-10-11	2016-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer
3	CA-2014-115812	2015-06-09	2015-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer
4	CA-2014-115812	2015-06-09	2015-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer
5	CA-2014-115812	2015-06-09	2015-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer

PRODUCT AND CATEGORY PERFORMANCE:

1.Find the most popular category

```
select category,count(*) as frequency  from superstore group by category order by count(*) desc;
```

	CATEGORY	···	FREQUENCY
1	Office Supplies		3,429
2	Technology		890
3	Furniture		765

From this we can conclude office sales is the most popular category.

2.Find the most popular subcategory in each category.

```
select category,sub_category,frequency from (select category,sub_category,count(*) as frequency ,row_number() over (partition by category
order by category)as rank from superstore group by category,sub_category) where rank=1 order by frequency;
```

Using row_number we can give the rank to each category for all subcategories

Row_number function:

It takes two parameters

partition by: it is used in grouping the rows based on particular column

Order by: we can order the partitioned groups based on particular column

Output:

	CATEGORY	SUB_CATEGORY	FREQUENCY
1	Technology	Phones	530
2	Office Supplies	Storage	407
3	Furniture	Bookcases	53

Phones are the most popular(sold) sub category.

3.Find which is the most profitable category,sub category,product.

Most profitable category:

```
select category,sum(profit) as profit from superstore group by category order by profit desc;
```

	CATEGORY	PROFIT
1	Technology	101,310.1
2	Office Supplies	59,626.64
3	Furniture	8,811.9

Technology is the most profitable one.

we can infer the most sold category is not the most profitable.

Most profitable sub category:

```
select category,sub_category,sprofit from
select category,sub_category,sum(profit) as sprofit,
row_number() over (partition by category order by category)as rank

from superstore
group by category,sub_category)
where rank=1
order by sprofit desc limit 2;
```

Output:

	CATEGORY	SUB_CATEGORY	SPROFIT
1	Technology	Phones	30,236.42
2	Office Supplies	Storage	11,288.93

Most profitable and most sold ones are the phones.

4.Find the product which is more profitable and least profitable.

```
select product_name,profit from superstore order by profit desc limit 1;
```

	PRODUCT_NAME	PROFIT
1	Canon imageCLASS 2200 Advanced Copier	8,399.98

Canon imageCLASS 2200 Advanced Copier is the most profitable product.

```
select product_name,profit from superstore order by profit limit 1;
```

	PRODUCT_NAME	PROFIT
1	Ibico EPK-21 Electric Binding System	-2,929.48

Ibico epk-15 electric binding system is the least profitable one.

5.Find the product which is most sold.

```
select product_name,count(*) as frequency
from superstore
group by product_name
order by frequency desc limit 2;
```

	PRODUCT_NAME	FREQUENCY
1	Staple envelope	47
2	Staples	46

6.Find the trend for each category.

```
select ord_year,Category, profit
from (select ord_year,Category, sum(profit) as profit,row_number()
over (partition by ord_year order by Category) as rank
from superstore  group by ord_year,Category)
where rank=1  order by ord_year;
```

	ORD_YEAR	CATEGORY	PROFIT
1	2015	Furniture	1,271.23
2	2016	Furniture	4,667.74
3	2017	Furniture	1,443.4
4	2018	Furniture	1,292.33
5	2019	Furniture	137.2

We can see we did not make good profit as year progress for the category furniture.

	ORD_YEAR	CATEGORY	PROFIT
1	2015	Technology	13,158.75
2	2016	Technology	23,658.93
3	2017	Technology	27,177.64
4	2018	Technology	37,289.37
5	2019	Technology	25.41

Technology category has an appreciable increase in profit over years except for the year 2019.

NOTE: year 2019 the company did not have good sales and profit due to covid outbreak.

7.Analyze the trend for most profitable product over years.

```
select ord_year,product_name,sum(profit) as profit
from superstore
group by ord_year,product_name
having product_name='Canon imageCLASS 2200 Advanced Copier'
order by ord_year;
```

	ORD_YEAR	PRODUCT_NAME	PROFIT
1	2017	Canon imageCLASS 2200 Advanced Copier	9,519.98
2	2018	Canon imageCLASS 2200 Advanced Copier	15,679.96

It shows a spiral trend.

Note: The company can introduce this product to gain more profit.

8.Analyze the trend for the less profitable product .

```
select ord_year,product_name,sum(profit) as profit
from superstore
group by ord_year,product_name
having product_name='Ibico EPK-21 Electric Binding System'
order by ord_year;
```

	ORD_YEAR	PRODUCT_NAME	PROFIT
1	2015	Ibico EPK-21 Electric Binding System	4,630.48
2	2017	Ibico EPK-21 Electric Binding System	1,644.29
3	2018	Ibico EPK-21 Electric Binding System	-2,929.48

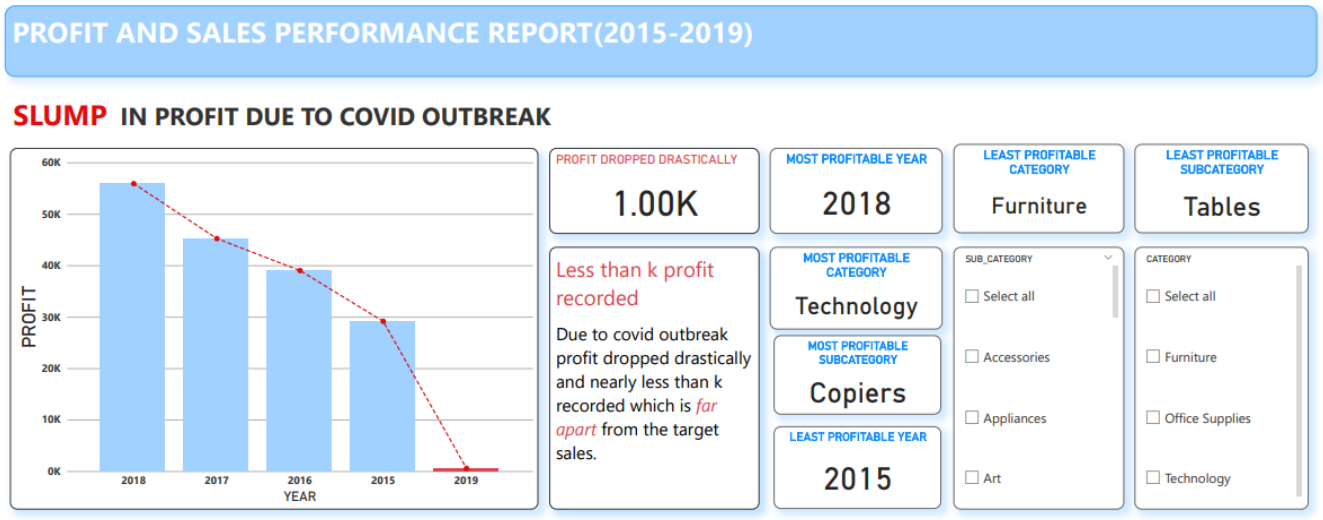
It follows a downline trend.

Note: The company can eliminate or find a way to avoid the loss.

POWER BI REPORT FOR PROFIT ANALYSIS:

Load the data from the snowflake account to power bi software directly.

- Connecting Power BI to Snowflake:
- Open Power BI Desktop.
- Click on "Get Data" and select "Snowflake" from the list of available data sources.
- Enter your Snowflake account information, such as your account URL, warehouse, database, and authentication credentials.
- Power BI will establish a connection to your Snowflake data, and you can then select the tables/views you want to use.



CUSTOMER SEGMENTATION:

To segment the customers based on region.

```
select region,sum(profit) as profit from superstore group by region order by profit desc;
```

	REGION	PROFIT
1	West	58,203.01
2	East	51,984.94
3	Central	32,829.35
4	South	26,731.34

The West region is more profitable in the United States.

Trend analysis for West region:

```
select ord_year,region,profit
from (select ord_year,region,sum(profit) as profit
,row_number() over(partition by ord_year order by region)
as rank from superstore
group by region,ord_year)
where rank=4 order by ord_year;
```

	ORD_YEAR	REGION	PROFIT
1	2015	West	10,645.23
2	2016	West	10,859.5
3	2017	West	12,223.81
4	2018	West	24,162.91
5	2019	West	311.56

It shows a spiral trend.

Find the most profitable state in the West region.

```
select state,sum(profit) as profit
from superstore where region='West'
group by state
order by profit
desc limit 3;
```

	STATE	PROFIT
1	California	35,299.79
2	Washington	17,239.7
3	Nevada	2,843.84

California got the most profit compared to other states.

Trend analysis for the state California:

```
1  select ord_year,state,profit ,region
2  from (select ord_year,state,region,sum(profit)
3  as profit,row_number()
4  over(partition by ord_year order by ord_year)
5  as rank from superstore where region='West'
6  group by region,state,ord_year)
7  where rank=1 order by ord_year;
```

	ORD_YEAR	STATE	PROFIT	REGION
1	2015	California	7,273.3	West
2	2016	California	6,650.01	West
3	2017	California	8,513.64	West
4	2018	California	12,585.78	West
5	2019	California	277.06	West

It shows a positive trend except for the year 2019.

Most profitable city in state california.

```
1  select city,sum(profit) as profit
2  from superstore where state='California'
3  group by city
4  order by profit desc limit 3;
```

	CITY	PROFIT
1	Los Angeles	15,417.09
2	San Francisco	8,181.57
3	San Diego	3,265.82

Los Angeles got the most profit compared to other states.

Most two profitable products in the city Los Angeles and San Francisco.

```
5   select  product_name,city,sum(profit) as profit
6   from superstore where city='Los Angeles'
7   group by product_name,city
8   order by profit desc limit 2;
```

	PRODUCT_NAME	CITY	PROFIT
1	Hewlett Packard LaserJet 3310 Copier	Los Angeles	1,343.97
2	Canon Imageclass D680 Copier / Fax	Los Angeles	1,049.99

```
5   select  product_name,city,sum(profit) as profit
6   from superstore where city='San Francisco'
7   group by product_name,city
8   order by profit desc limit 2;
```

	PRODUCT_NAME	CITY	PROFIT
1	Sharp AL-1530CS Digital Copier	San Francisco	434.99
2	Hoover Upright Vacuum With Dirt Cup	San Francisco	419.82

Top 5 customers:

```
select customer_id,state,city,count(*)
as  number_of_orders
from superstore
group by customer_id,state,city
order by number_of_orders
desc limit 5;
```

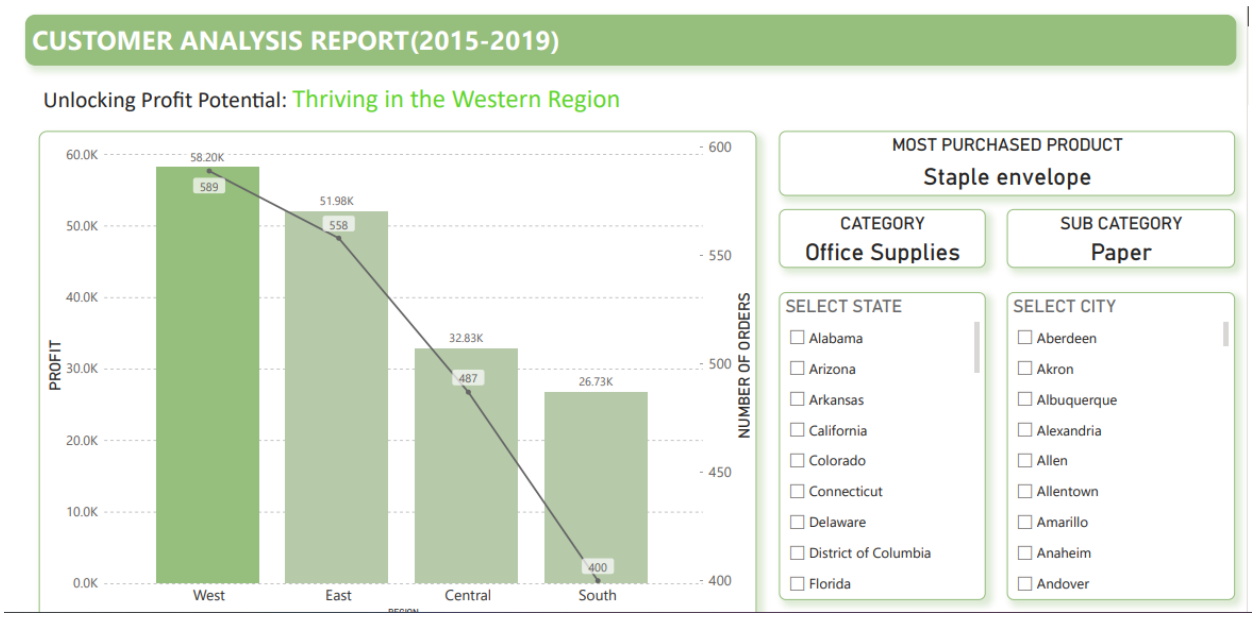
	CUSTOMER_ID	STATE	CITY	NUMBER_OF_ORDERS
1	XP-21865	California	Los Angeles	7
2	AI-10855	California	Los Angeles	7
3	RL-19615	Illinois	Chicago	7
4	SU-20665	Illinois	Chicago	7
5	NP-18325	New York	New York City	6

Products that are mostly brought together by customers.

```
select  a.product_name as product1,
b.product_name as product2,
sum(a.profit+b.profit) as total_profit
from superstore as a
join superstore as b
on a.customer_id=b.customer_id
and a.product_id<b.product_id
group by a.product_name,b.product_name
order by total_profit desc limit 5;
```

	PRODUCT1	PRODUCT2	TOTAL_PROFIT
1	Staple envelope	Canon imageCLASS 2200 Advanced Copier	18,522.97
2	Acco Perma 4000 Stacking Storage Drawers	Canon imageCLASS 2200 Advanced Copier	15,129.7
3	Enermax Acrylux Wireless Keyboard	Canon imageCLASS 2200 Advanced Copier	8,584.24
4	Xerox 1881	Canon imageCLASS 2200 Advanced Copier	8,428.84
5	Xerox 1983	Canon imageCLASS 2200 Advanced Copier	8,411.7

POWER BI REPORT FOR CUSTOMER SEGMENTATION:


CUSTOMER ANALYSIS REPORT(2015-2019)
Unlocking Profit Potential: Thriving in the Western Region
PROFIT
NUMBER OF ORDERS
60.0K
50.0K
40.0K
30.0K
20.0K
10.0K
0.0K
600
550
500
450
400
58.20K
589
51.98K
558
32.83K
487
26.73K
400
West
East
Central
South
MOST PURCHASED PRODUCT
Staple envelope
CATEGORY
Office Supplies
SUB CATEGORY
Paper
SELECT STATE
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
SELECT CITY
Aberdeen
Akron
Albuquerque
Alexandria
Allen
Allentown
Amarillo
Anaheim
Andover