AUGUST 16, 2022

Superstore Sales

SQL PROJECT

DANIELOKOLI448@GMAIL.COM
OKOLI NNAMDI NNAEMEKA
09067260556

Professional Background

With over three years of versatile experience in both office and freelance settings, I have honed my skills as a data analyst. My impactful tenure at Guinness Nigeria was marked by the optimization of supply chain operations through strategic data analysis, resulting in a notable reduction in processing time.

As a National Youth Service Corps (NYSC) graduate, I served as a Data Entry personnel, utilizing Microsoft Excel to navigate extensive datasets and presenting insights through clear Excel tables and PowerPoint presentations. Transitioning to the role of IT Assistant at Delta Broadcasting Service (DBS) Warri, I further strengthened my data management skills over six months.

My commitment to continuous learning is evident in my proficiency in SQL, MS Excel, and Python. Additionally, I bring expertise in visualization tools such as Tableau and Power BI, with ongoing efforts to enhance my skills in R language. Practical applications include the creation of interactive dashboards using Power BI and experimentation with R language in controlled environments.

Notably, my analytical approach focuses on tangible outcomes and quantifiable achievements. At Guinness Nigeria, my data-driven strategies had a direct and positive impact on supply chain efficiency. Eager to contribute this expertise to a remote data analyst role, I am poised to deliver results through a combination of technical proficiency and strategic problem-solving.

Table of Contents

Introduction

The **"Superstore Sales"** dataset is a comprehensive and versatile collection of data that provides valuable insights into sales, customer behaviour, and product performance. This dataset offers a rich resource for in-depth analysis.
Containing information from diverse regions and segments, the dataset enables exploration of trends, patterns, and correlations in **sales and customer preferences.**

The Business problem is to improve on the profits made by the Superstore.

I used the dataset Superstore_sale to analyze and answer this problem.

I applied SQL commands to analyse data: JOIN, ORDER BY, AS, WHERE, AND, OR, SUM(), COUNT(), GROUP BY, HAVING.

Also, I used Root Cause Analysis to understand the problem and ask right questions.

As a result, I have found out crucial insights of provided data sets, prepared visualisations, and report for my team.

Root Cause Analysis

The business problem is that we need to improve the total amount of profits made from sales by the Superstore.

In order to better understand the problem, I need to analyze the existing database of sales in the store. I will also present some crucial numbers and visualization of the datasets. So, I decided to ask some questions.

- How many sales do we currently have recorded in our database?
- What timeframe does this dataset cover?
- How much profit was made in the recorded dataset?
- How much was gotten from sales in the database?
- How many sales was made each year and what was the equivalent profit made?
- What shipping mode was used the most?

Also, I decided to apply Root Cause Analysis to the problem to figure out the underlying issues in order to identify appropriate solutions.

1. Why doesn't the Superstore make enough profits from sales?
   Ans. They make losses from multiple sales.

2. Why does the Superstore make so many losses from sales?
   Ans. These losses are due to the giving discounts of over 0.2% on sales to customers.

3. Why were high discount rates given to customers?
   Ans. High discount were given to customers from the central region of the country.

4. Why do customers from the central region of the country get high discount rates?
   Ans. These customers are given higher discount rates when they purchase office supplies from the Superstore.

5. Why do customers from the central region of the country get high discount rates when they buy office supplies?
   Ans. They get higher discount rates when they buy binders office supplies.

Insights from the Analysis

I made use of a relational databases Superstore_sale to answer the problem.

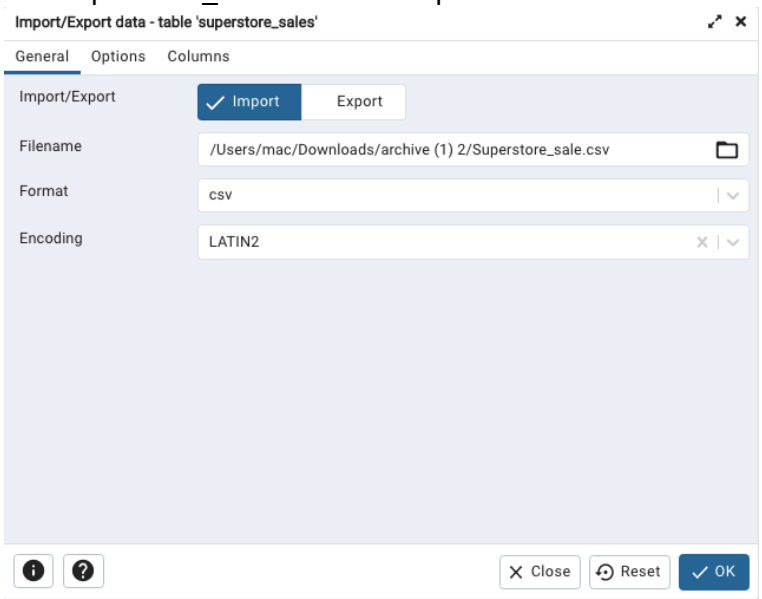
POSTGRESQl Database Management System was used to find out main insights.

Superstore_sale includes such data:
- Row ID
- Order ID
- Order Date
- Ship Date
- Ship Mode
- Customer ID
- Customer Name
- Segment
- Country
- City
- State
- Postal Code
- Region
- Product ID
- Category
- Sub-Category
- Product Name
- Sales
- Quantity
- Discount
- Profit

A Superstore_sale table was created using Postgresql.

```
4  CREATE TABLE Superstore_sales(
5      Row_ID SERIAL,
6      Order_ID VARCHAR(30),
7      Order_Date VARCHAR(30),
8      Ship_Date VARCHAR(30),
9      Ship_Mode TEXT,
10     Customer_ID VARCHAR(20),
11     Customer_Name TEXT,
12     Segment TEXT,
13     Country TEXT,
14     City TEXT,
15     State TEXT,
16     Postal_Code INT,
17     Region TEXT,
18     Product_ID VARCHAR(30),
19     Category TEXT,
20     Sub_Category VARCHAR(30),
21     Product_Name VARCHAR(500),
22     Sales REAL,
23     Quantity INT,
24     Discount REAL,
25     Profit REAL
26 );
```

The Superstore_sale was then imported into the created table with the "LATIN2" Encoding



I Used ALTER() to change the datatype of the order_date and ship_date columns.

```
33 ALTER TABLE Superstore_sales
34 ALTER COLUMN order_date TYPE DATE
35 USING to_date(order_date, 'DD-MM-YYYY');
36
37 ALTER TABLE Superstore_sales
38 ALTER COLUMN Ship_Date TYPE DATE
39 USING to_date(Ship_Date, 'DD-MM-YYYY');
```

SELECT statement was used to fetch data from a database.

```
29  SELECT * FROM Superstore_sales;
```

To find the recorded number of crimes with the COUNT() Function, I used a command like this:

```
31  SELECT COUNT(Row_ID) FROM Superstore_sales;
```

To get the timeframe of the dataset, I used the MIN() and MAX():

```
41  SELECT MAX(order_date)
42  FROM Superstore_sales;
43
44  SELECT MIN(order_date)
45  FROM Superstore_sales;
```

To know the sum of the profits and sales, I used SUM():

```
64  SELECT SUM(profit)
65  FROM Superstore_sales;
66
67  SELECT SUM(sales)
68  FROM Superstore_sales;
```

To know the total sales and profit of each year, I used the COUNT() and SUM() function:

```
48  SELECT COUNT(order_date), SUM(profit)
49  FROM Superstore_sales
50  WHERE order_date BETWEEN '2011-01-04' AND '2011-12-31';
51
52  SELECT COUNT(order_date), SUM(profit)
53  FROM Superstore_sales
54  WHERE order_date BETWEEN '2012-01-01' AND '2012-12-31';
55
56  SELECT COUNT(order_date), SUM(profit)
57  FROM Superstore_sales
58  WHERE order_date BETWEEN '2013-01-01' AND '2013-12-31';
59
60  SELECT COUNT(order_date), SUM(profit)
61  FROM Superstore_sales
62  WHERE order_date BETWEEN '2014-01-01' AND '2014-12-31';
```

I used the GROUP BY function to find the shipping mode that was used the most.

```
70  SELECT Ship_Mode, COUNT(*)
71  FROM Superstore_sales
72  GROUP BY ship_mode
73  ORDER BY COUNT(*) DESC;
```

I used the GROUP BY function and HAVING clause to find the state with over 100 purchases from the superstore.

```
80  SELECT state, COUNT(*)
81  FROM Superstore_sales
82  GROUP BY state
83  HAVING COUNT(*) > 100
84  ORDER BY COUNT(*) DESC;
```

I used the GROUP BY function to find the total number of customers from all regions. Also, I used the AS command to give an alias to COUNT(*). The ORDER BY query was used to organize the data:

```
94  SELECT region, COUNT(*) AS Customers, SUM(profit)
95  FROM Superstore_sales
96  WHERE discount > 0.2
97  AND profit < 0
98  GROUP BY region
99  ORDER BY COUNT(*) DESC;
```

To know the top 20 losses from sales i wrote this query. I used FETCH to get the first 20 outputs.

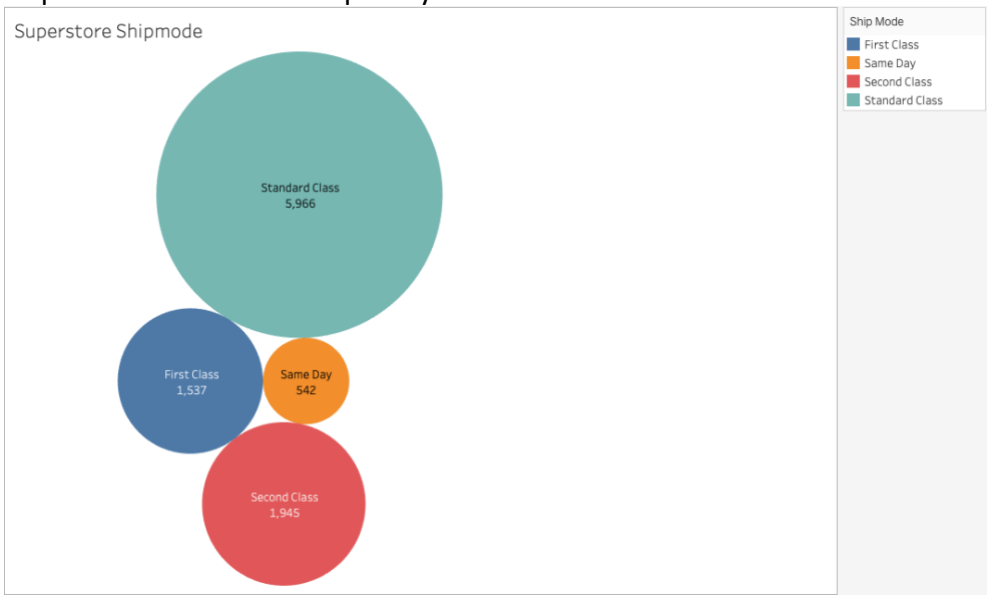
```
108  SELECT Row_ID, Region, Sales, Discount, Profit
109  FROM Superstore_sales
110  WHERE discount > 0.2
111  AND profit < 0
112  ORDER BY Profit
113  FETCH FIRST 20 ROW ONLY;
```

This shows the top 100 discounts given to customers and the category of what they purchased.

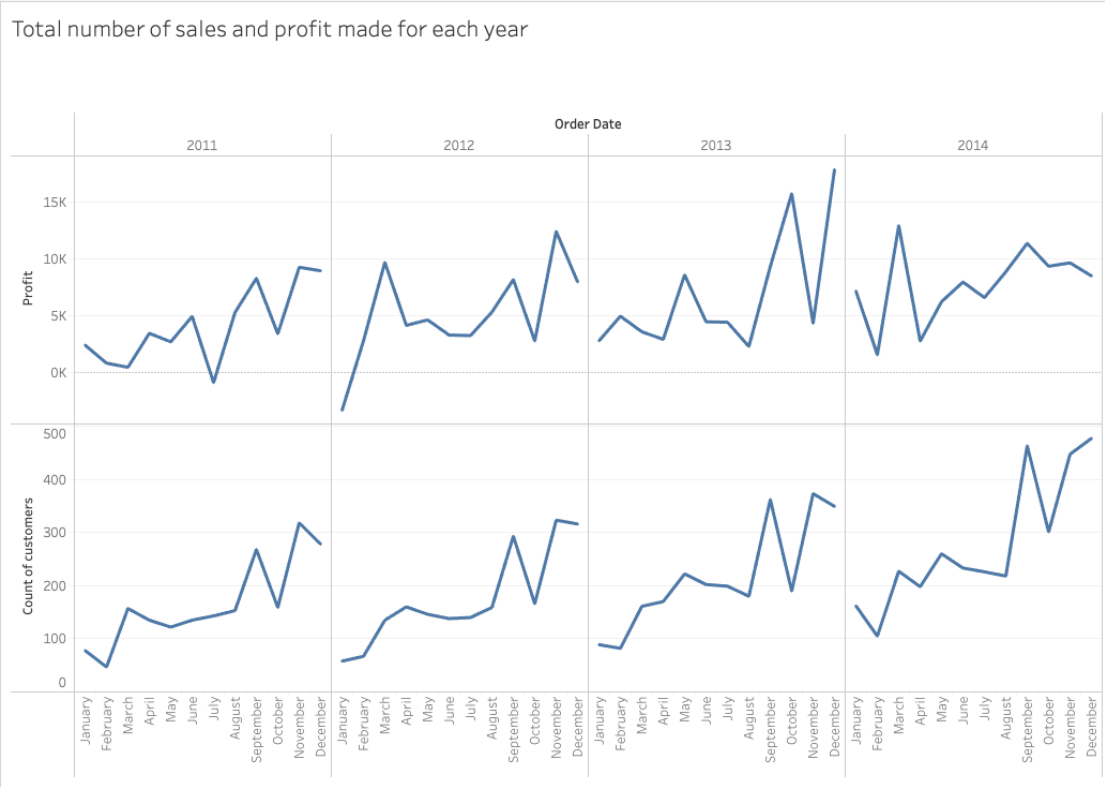
```
134  SELECT Row_ID, Region, Category, sub_category, Sales, Discount
135  FROM Superstore_sales
136  WHERE discount > 0.2
137  ORDER BY discount DESC
138  LIMIT 100;
```

Tableaus was used as a very powerful tool for data analysis and Visualization for better understanding of the data.

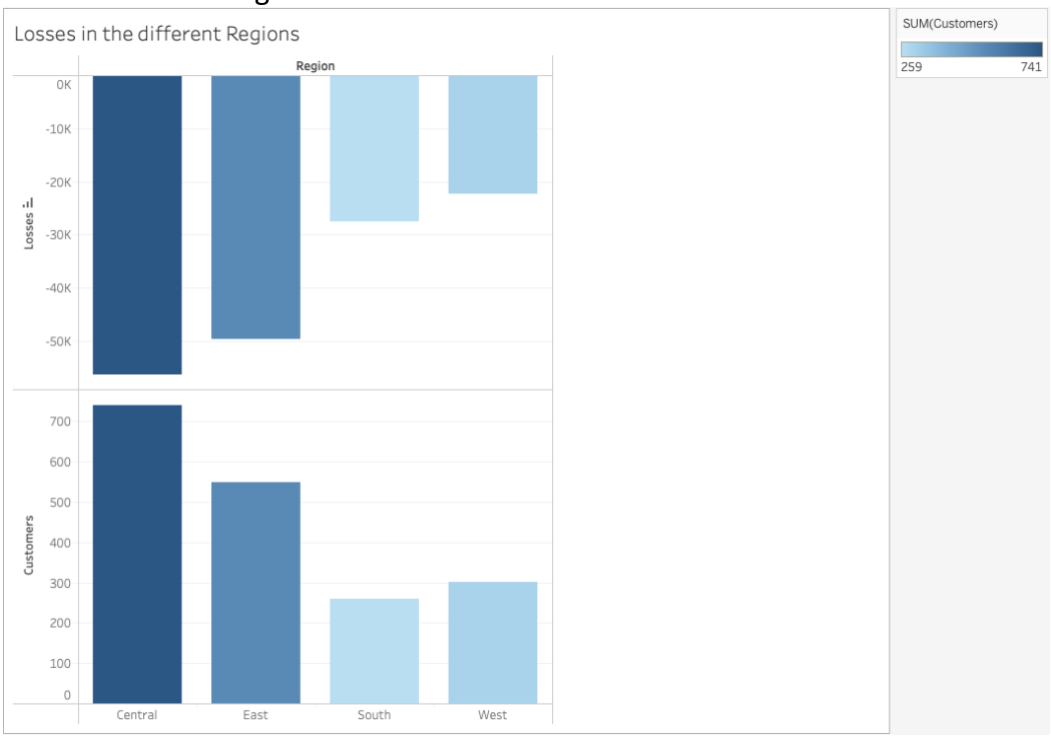
This Tableau clearly shows the different shipping modes of the Superstore. Standard class ship mode is the most frequently used.



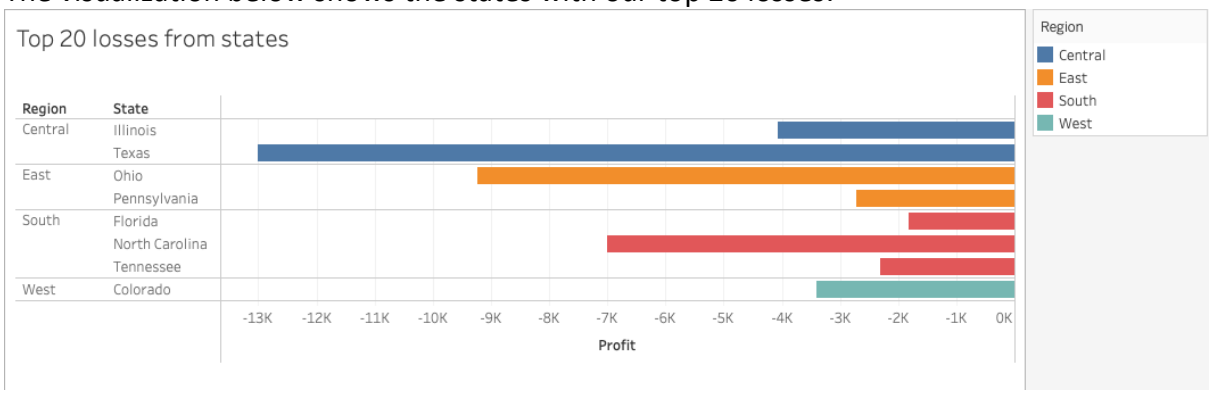
This time series visualization shows the progression and regression of the total number of sales and profits made each year in our dataset. We can see the most profit was made December 2013 while the greatest loss was recorded in January 2012.



This visualization shows the total number of customers with a discount of over 0.2 and the losses from each region.



The visualization below shows the states with our top 20 losses.



Findings and Recommendations

Here are the results of the data set analysis:

- The total number of recorded sales we have in the database is 9990
- The dataset covers over 4 years of sales by the Superstore between 04/01/2011 and 31/12/2014
- 286,197.47 was the total profit made from sales from the dataset.
- 2,295,821.47 was the total sales amount from the dataset.
- Standard class shipping mode was used the most.

Record of total number of sales and profit made each year.

Number of sales	Total profit	Year
3317	93356.41	1014
2579	81702.73	2013
2101	61594.40	2012
1993	49544.00	2011

We can see there was a continuous increase in both number of sales and profits made each year.

Total number of times each ship mode was used.

ship_mode	count
Standard Class	5966
Second Class	1945
First Class	1537
Same Day	542

'Same day' was used the least by customers.

Status of each crime from arrests to continued investigation.

status_desc	number_of_victims
Invest Cont	667644
Adult Other	89864
Adult Arrest	72652
Juv Arrest	2710
Juv Other	1446
UNK	4

The count of customers with a discount rate of over 0.2% from each region and the total loss from each region

region	customers	sum
Central	741	-56314.848
East	549	-49561.3
West	301	-22392.26
South	259	-27504.838

Thus, we can see the Superstore is hit with great loss when these high discounts are given to customers.

This is a list of the top 100 sales made to customers with a discount rate of over 0.2%

row_id	region	category	sub_category	sales	discount
3085	Central	Office Supplies	Binders	252.784	0.8
3436	Central	Office Supplies	Binders	15.08	0.8
2005	Central	Office Supplies	Binders	11.364	0.8
2923	Central	Office Supplies	Binders	3.882	0.8
3260	Central	Office Supplies	Binders	3.136	0.8
3394	Central	Office Supplies	Binders	6.928	0.8
1758	Central	Office Supplies	Binders	41.568	0.8
1996	Central	Office Supplies	Appliances	294.62	0.8
2503	Central	Office Supplies	Binders	9.36	0.8
2883	Central	Office Supplies	Binders	8.85	0.8
3194	Central	Office Supplies	Binders	3.648	0.8
3243	Central	Office Supplies	Binders	13.468	0.8
3308	Central	Office Supplies	Binders	2.286	0.8
3379	Central	Office Supplies	Binders	13.776	0.8
1515	Central	Office Supplies	Binders	6.744	0.8
1676	Central	Office Supplies	Binders	6.588	0.8

1858	Central	Office Supplies	Binders	34.24	0.8
1973	Central	Office Supplies	Binders	5.104	0.8
2174	Central	Office Supplies	Binders	9.264	0.8
2502	Central	Office Supplies	Binders	2.304	0.8
2688	Central	Office Supplies	Binders	3.98	0.8
2847	Central	Office Supplies	Binders	762.594	0.8
3161	Central	Office Supplies	Binders	5.18	0.8
3193	Central	Office Supplies	Binders	6.37	0.8
3212	Central	Office Supplies	Binders	11.228	0.8
3233	Central	Office Supplies	Binders	2.896	0.8
1272	Central	Office Supplies	Appliances	48.792	0.8
3295	Central	Office Supplies	Binders	30.96	0.8
3325	Central	Office Supplies	Binders	896.99	0.8
3359	Central	Office Supplies	Binders	3.208	0.8
1326	Central	Office Supplies	Binders	17.46	0.8
1514	Central	Office Supplies	Binders	21.38	0.8
988	Central	Office Supplies	Binders	1.112	0.8
1568	Central	Office Supplies	Binders	8.784	0.8
1771	Central	Office Supplies	Binders	12.222	0.8
1821	Central	Office Supplies	Appliances	92.064	0.8
1108	Central	Office Supplies	Binders	10.78	0.8
1944	Central	Office Supplies	Binders	3.24	0.8
2138	Central	Office Supplies	Binders	3.656	0.8

2168	Central	Office Supplies	Binders	6.286	0.8
2347	Central	Office Supplies	Binders	1.272	0.8
2429	Central	Office Supplies	Binders	8.272	0.8
2508	Central	Office Supplies	Appliances	32.784	0.8
2528	Central	Office Supplies	Binders	9.648	0.8
2769	Central	Office Supplies	Appliances	32.192	0.8
2827	Central	Office Supplies	Binders	2.724	0.8
3088	Central	Office Supplies	Appliances	12.992	0.8
3142	Central	Office Supplies	Binders	210.392	0.8
3186	Central	Office Supplies	Binders	26.046	0.8
470	Central	Office Supplies	Binders	4.788	0.8
102	Central	Office Supplies	Binders	1.788	0.8
677	Central	Office Supplies	Appliances	2.688	0.8
680	Central	Office Supplies	Binders	182.994	0.8
3232	Central	Office Supplies	Binders	26.046	0.8
537	Central	Office Supplies	Binders	42.616	0.8
550	Central	Office Supplies	Binders	11.364	0.8
1104	Central	Office Supplies	Binders	2.934	0.8
1273	Central	Office Supplies	Binders	44.848	0.8
902	Central	Office Supplies	Binders	18.336	0.8
918	Central	Office Supplies	Binders	8.568	0.8
3326	Central	Office Supplies	Binders	1.234	0.8
3349	Central	Office Supplies	Binders	16.03	0.8

1174	Central	Office Supplies	Binders	5.176	0.8
1285	Central	Office Supplies	Binders	2.772	0.8
1333	Central	Office Supplies	Binders	1.08	0.8
1334	Central	Office Supplies	Appliances	7.96	0.8
204	Central	Office Supplies	Appliances	66.284	0.8
622	Central	Office Supplies	Binders	8.69	0.8
1548	Central	Office Supplies	Binders	23.912	0.8
1551	Central	Office Supplies	Appliances	9.324	0.8
15	Central	Office Supplies	Appliances	68.81	0.8
1113	Central	Office Supplies	Binders	1.192	0.8
1819	Central	Office Supplies	Appliances	75.6	0.8
837	Central	Office Supplies	Binders	51.184	0.8
742	Central	Office Supplies	Binders	3.54	0.8
1045	Central	Office Supplies	Appliances	58.464	0.8
281	Central	Office Supplies	Binders	2.08	0.8
262	Central	Office Supplies	Appliances	1.624	0.8
76	Central	Office Supplies	Binders	1.248	0.8
2107	Central	Office Supplies	Binders	0.876	0.8
16	Central	Office Supplies	Binders	2.544	0.8
2160	Central	Office Supplies	Binders	8.808	0.8
1200	Central	Office Supplies	Binders	1088.792	0.8
2207	Central	Office Supplies	Binders	29.932	0.8
170	Central	Office Supplies	Appliances	177.98	0.8

2428	Central	Office Supplies	Binders	1.344	0.8
663	Central	Office Supplies	Binders	2.724	0.8
175	Central	Office Supplies	Appliances	52.448	0.8
2527	Central	Office Supplies	Binders	6.9	0.8
177	Central	Office Supplies	Appliances	97.264	0.8
379	Central	Office Supplies	Appliances	8.652	0.8
394	Central	Office Supplies	Binders	12.462	0.8
381	Central	Office Supplies	Binders	12.176	0.8
659	Central	Office Supplies	Binders	2.308	0.8
522	Central	Office Supplies	Binders	14.112	0.8
1103	Central	Office Supplies	Binders	2.694	0.8
3097	Central	Office Supplies	Appliances	5.768	0.8
3135	Central	Office Supplies	Binders	30.56	0.8
3172	Central	Office Supplies	Binders	3.168	0.8
3507	Central	Office Supplies	Binders	3.392	0.8

Conclusively, we can understand from this finding that with over0.2 percent discount where all from the central region, with a discount of 0.8%, a sales category of office supplies and a majority sub-category of Binders. Thus, this set of issues are the main reasons for loss in sales.

Conclusion

I have analysed the dataset Superstore_sale to help improve the total amount of profits made from sales by the Superstore.
Hence, I asked some general questions to get an insight of the database. Some of the insights I got were: There was an increase in profits and sales moving into ever new year in the database, a profit of 286,197.47 was made from sales within 4 years and the standard class shipping mode was the most used shipping mode.

However, there are some crucial points that we need to count and try to use to improve the total amount of profits made from sales.

To begin with, i found out that considerable losses were made in sales within the 4 years recorded. This was due to the that many customers were given high discount rates on theirs purchases from the store. I suggest better consideration should be taking when given discount rates of over 0.2% to customers.

Furthermore, I noticed that a lot of this discount rates over 0.2% were given to customers from the central region of the country (Texas and Illinois). Surprisingly, all discount rates of 0.8% which was the highest rate recorded were only given to these customers from the central region. This is an issue that needs to be looked over and resolved.
Therefore, a reduction in discount rates for customers from the central region of country should be made to improve on the profits and decreases losses made by the Superstore.

Moreover, I observed that high discount rates are given to customers from the central region of the country who purchase Binders from the office supply category. Most discount rates of 0.8% were also given to customers who buy binders while all sales with this discount rate were given for office supplies.

In conclusion, a reduction in discount rates to customers from the central region of the country who buy office supplies needs to be made. A general reduction in all discount rates is highly advised to reduce the losses and in turn provide an increase in profits made by the superstore.