SALES SUPERSTORE ANALYSIS

Prepared By: Andre Ahmad Al Farisi

Created at : Januari 5, 2024

Table Of Contents

Abstrak

A sales data set was found on Kaggle which is considered to be the sales data of a superstore. The dataset contains sales data of 4922 different orders, 1326 of which have been shipped, 3 product categories, 1849 product names, 529 cities, 1 country and 49 regions. This analysis is a descriptive analysis which aims to find insight with extensive exposure, exploring sales trends and customer behavior, which will be developed through a list of questions that will describe the purpose of the analysis.

Keywords : Analysis Descriptive, Sales, Kaggle, Superstore

Introduction

The dataset contains information about sales from a superstore, including details such as order ID, transaction date, customer name, customer ID, customer type, city, country, product category, product subcategory, product name, country, city, region and sales. In general, this dataset contains a combination of product and customer data that is tied into the sales system. The dataset contains data with 9800 rows and 18 columns.

Objective

Objective / Goal Analysis

The objective of the analysis is to find insights from datasets such as sales trends, best-selling products, certain seasonalities that influence sales, understand customer profiles, purchasing habits, and evaluate sales in various regions, so that the resulting insights can provide a complete and broad description for decision making business.

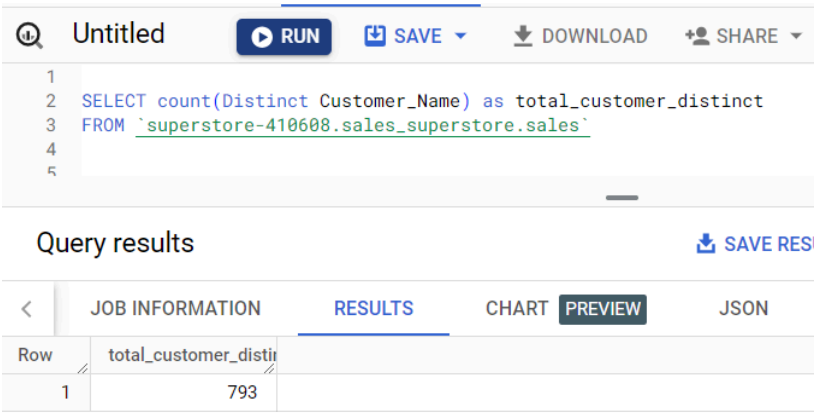
Analysis

- Analysis steps

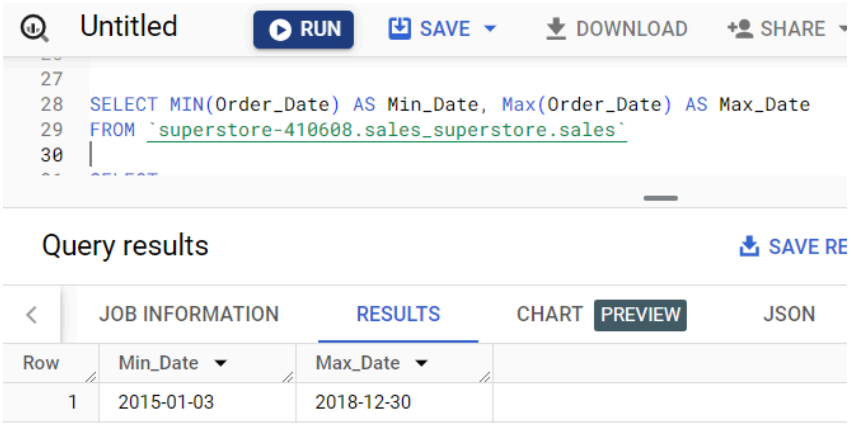
1. Eksploratory Data Analysis (EDA) Using BigQuery (SQL Language)

Identify descriptive statistics, distributions, and general trends in datasets.

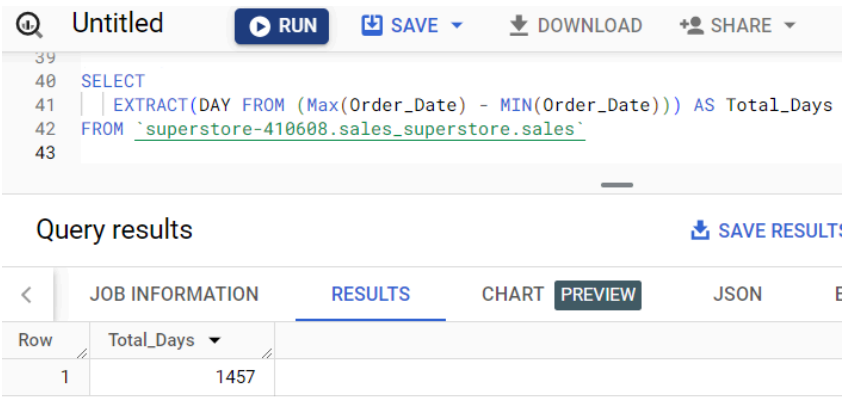
➔ Number of distinct customer names



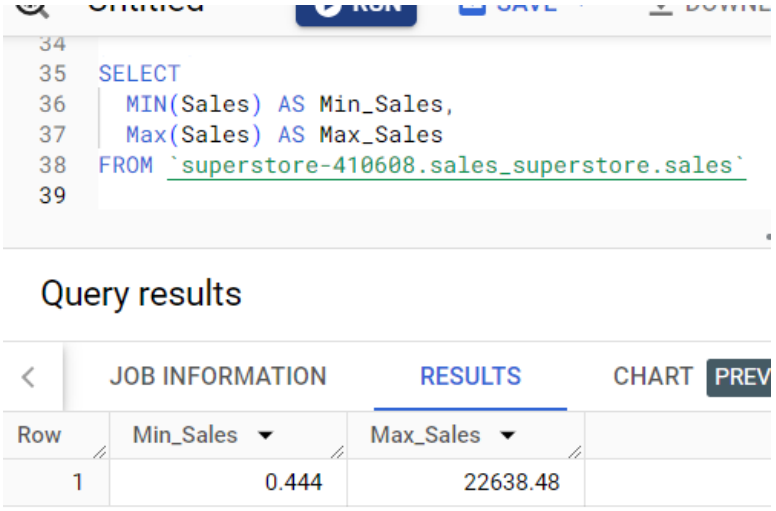
➔ First order_date and Final Order_date



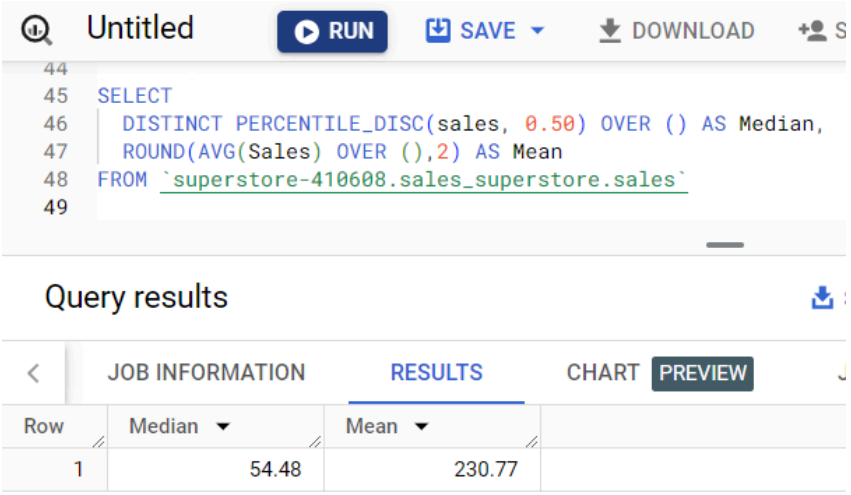
➔ Dataset range



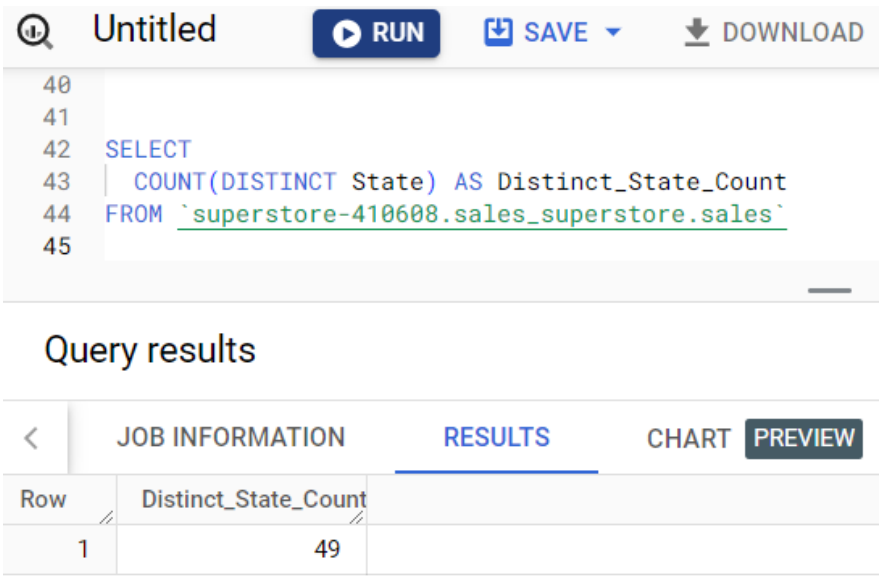
➔ Sales range



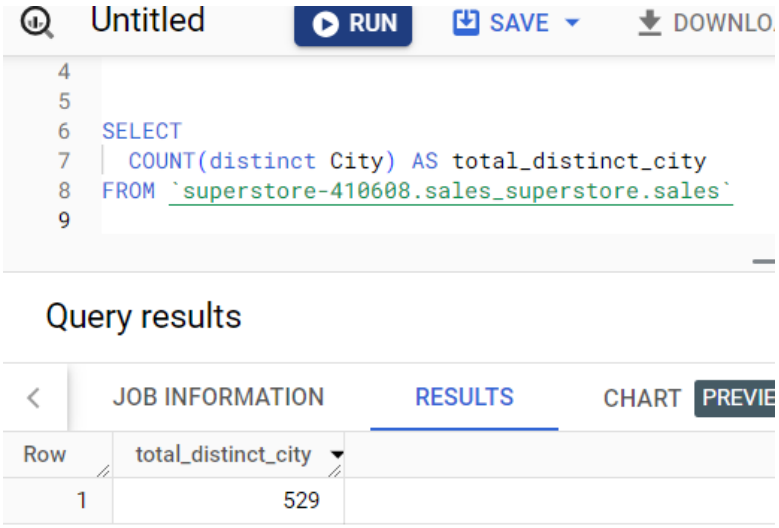
➔ Sales distribution (Mean, Median Sales)



➔ Number of distinct state



➔ Number of distinct city



➔ Number of distinct segment

```
10
11  SELECT
12    COUNT (DISTINCT Segment) AS total_distinct_segment
13  FROM `superstore-410608.sales_superstore.sales`
14
15
```

Query results

JOB INFORMATION | RESULTS | CHART | PREVIEW

Row	total_distinct_segme
1	3

➔ Number of distinct product name

Untitled RUN SAVE DOWNLOAD

```
14
15
16  SELECT
17    COUNT(DISTINCT Product_Name) AS total_product_distinct
18  FROM `superstore-410608.sales_superstore.sales`
19
```

Query results

JOB INFORMATION | RESULTS | CHART | PREVIEW

Row	total_product_distinc
1	1849

➔ Number of distinct product kategory

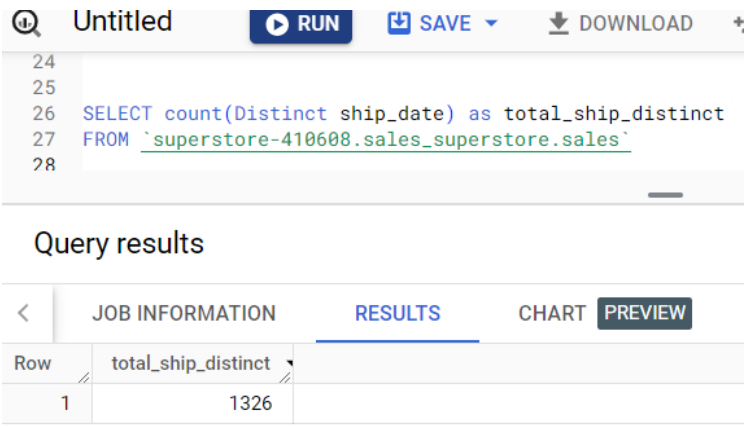
```
19
20
21  SELECT
22    COUNT(DISTINCT Category) AS total_category_distinct
23  FROM `superstore-410608.sales_superstore.sales`
24
```

Query results

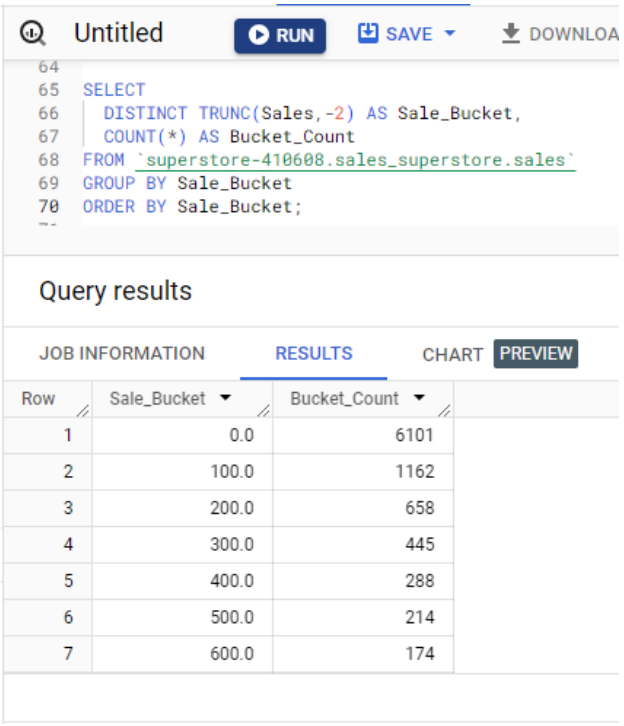
JOB INFORMATION | RESULTS | CHART | PREVIEW

Row	total_category_distin
1	3

➔ Number of distinct ship mode



➔ Sales Distribution (histogram sales)



2. Analysis Sales Superstore

A more granular analysis of the sales data of the global superstore using SQL. The analysis aimed to answer the following questions about the dataset:

❖ What was the most profitable region, state, and city in the dataset?

Analysis_Superstore_Sales

RUN

```
2
3  SELECT
4    DISTINCT Region,
5    ROUND(SUM(sales), 2) as total_sales_reg
6  FROM `superstore-410608.sales_superstore.sales`
7  GROUP BY Region
8  ORDER BY total_sales_reg DESC
9    -- West is Region with most profitable
10
```

Query results

JOB INFORMATION | RESULTS | CHART PREVIEW

Row	Region	total_sales_reg
1	West	710219.68
2	East	669518.73
3	Central	492646.91
4	South	389151.46

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE | MO

```
11 SELECT
12   DISTINCT State,
13   ROUND(SUM(Sales) OVER(PARTITION BY State), 2) AS sales_state_west_Reg,
14   ROUND(SUM(Sales) OVER (),2) AS West_Region_Total,
15   ROUND((SUM(Sales) OVER (PARTITION BY State))/(SUM(Sales) OVER ()),2) AS West_Region_State_Pct
16 FROM `superstore-410608.sales_superstore.sales`
17 WHERE Region = 'West'
18 ORDER BY sales_state_west_Reg DESC LIMIT 3
19   -- California was the most profitable state in the western region with a sales total of $446,306.46,
```

Query results

SAVE R

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRA

Row	State	sales_state_west_Re	West_Region_Total	West_Region_State_I
1	California	446306.46	710219.68	0.63
2	Washington	135206.85	710219.68	0.19
3	Arizona	35272.66	710219.68	0.05

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE | MORE

```
22 SELECT
23   DISTINCT City,
24   State,
25   ROUND (SUM(Sales) OVER (PARTITION BY City, State), 2) AS Sales_Cal_city_totals,
26   ROUND (SUM(Sales) OVER (), 2) AS Cal_State_Total,
27   ROUND((SUM(Sales) OVER (PARTITION BY City, State) / SUM(Sales) OVER ()), 2) AS Cal_State_Region_pct
28 FROM `superstore-410608.sales_superstore.sales`
29 WHERE State = 'California'
30 ORDER BY Sales_Cal_city_totals DESC LIMIT 3
31   -- Los Angeles was most profitable in the California state with a sales $ 173,420.18
32   -- and persentase 39% from total sales in the california city
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	City	State	Sales_Cal_city_totals	Cal_State_Total	Cal_State_Region_pc
1	Los Angeles	California	173420.18	446306.46	0.39
2	San Francisco	California	109041.12	446306.46	0.24
3	San Diego	California	47521.03	446306.46	0.11

❖ What was the most profitable category and sub category?

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE

```sql
-- based on category
SELECT
  DISTINCT Category,
  ROUND(SUM(Sales) OVER (PARTITION BY Category), 2) AS total_sales_cat,
  ROUND(SUM(Sales) OVER (), 2) AS total_sales,
  ROUND((SUM(Sales) OVER (PARTITION BY Category)/ SUM(Sales) OVER ()), 2) AS Contribute_sales
FROM `superstore-410608.sales_superstore.sales`
ORDER BY total_sales_cat DESC LIMIT 3
  -- Technology was most profitable in the category with total sales $ 827,455.00
  -- And Contribute sales is 37%
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECU

Row	Category	total_sales_cat	total_sales	Contribute_sales
1	Technology	827455.87	2261536.78	0.37
2	Furniture	728658.58	2261536.78	0.32
3	Office Supplies	705422.33	2261536.78	0.31

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE

```sql
-- Product name in the technology category was most profitable
SELECT
  DISTINCT Product_Name,
  ROUND(SUM(Sales) OVER (PARTITION BY Product_Name), 2) AS tot_sales_tech_cat,
  ROUND(SUM(Sales) OVER (), 2) AS tot_sales_Cat,
  ROUND((SUM(Sales) OVER (PARTITION BY Product_Name) / SUM(Sales) OVER ()), 2) AS Contribute_Sales
FROM `superstore-410608.sales_superstore.sales`
WHERE Category = 'Technology'
ORDER BY tot_sales_tech_cat DESC LIMIT 3
  -- Canon image Class 2200 Advanced Copier product was most profitable in the technology category w
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION G

Row	Product_Name	tot_sales_tech_cat	tot_sales_Cat	Contribute_Sales
1	Canon imageCLASS 2200 Adva...	61599.82	827455.87	0.07
2	Cisco TelePresence System EX...	22638.48	827455.87	0.03
3	Hewlett Packard LaserJet 331...	18839.69	827455.87	0.02

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE

```sql
-- Based on Sub category
SELECT
  DISTINCT Sub_Category,
  ROUND(SUM(Sales) OVER (PARTITION BY Sub_Category), 2) AS tot_sales_sub_cat,
  ROUND(SUM(Sales) OVER (), 2) AS tot_sales,
  ROUND((SUM(Sales) OVER (PARTITION BY Sub_Category) / SUM(Sales) OVER ()), 2) AS Contribute_Sales
FROM `superstore-410608.sales_superstore.sales`
ORDER BY tot_sales_sub_cat DESC LIMIT 3
  -- Sub Category was most profitable is Phones with total sales is $ 327,782.45
  -- With contribute sales 14%
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION

Row	Sub_Category	tot_sales_sub_cat	tot_sales	Contribute_Sales
1	Phones	327782.45	2261536.78	0.14
2	Chairs	322822.73	2261536.78	0.14
3	Storage	219343.39	2261536.78	0.1

Analysis_Superstore_Sales RUN SAVE QUERY SHARE SCHEDULE

```
-- Product name in the Sub category phone was most profitable
SELECT
  DISTINCT Product_Name,
  ROUND(SUM(Sales) OVER (PARTITION BY Product_Name), 2) AS tot_sales_sub_cat_phone,
  ROUND(SUM(Sales) OVER (), 2) AS tot_sales_sub_cat,
  ROUND((SUM(Sales) OVER (PARTITION BY Product_Name)/SUM(Sales) OVER ()), 2) AS Contribute_sales
FROM `superstore-410608.sales_superstore.sales`
WHERE Sub_Category = 'Phones'
ORDER BY tot_sales_sub_cat_phone DESC LIMIT 3
-- Samsung Galaxy Mega 6.3 was most profitable in the sub category phone with total sales is &
```

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART | PREVIEW | JSON | EXECUTION DETAILS | EXECUTIO

Row	Product_Name	tot_sales_sub_cat_ph	tot_sales_sub_cat	Contribute_sales
1	Samsung Galaxy Mega 6.3	13943.67	327782.45	0.04
2	Apple iPhone 5	12996.6	327782.45	0.04
3	Wilson Electronics DB Pro Sign...	8878.4	327782.45	0.03

Analysis_Superstore_Sales RUN SAVE QUERY SHARE SCHEDULE

```
-- 3. What is the most profitable in the product

SELECT
  DISTINCT Product_Name,
  ROUND(SUM(Sales) OVER (PARTITION BY Product_Name), 2) AS tot_sales_product,
  ROUND(SUM(Sales) OVER (), 2) AS tot_sales,
  ROUND((SUM(Sales) OVER (PARTITION BY Product_Name)/SUM(Sales) OVER ()), 2) AS Contribute_sales
FROM `superstore-410608.sales_superstore.sales`
ORDER BY tot_sales_product DESC LIMIT 3
-- As a whole was most profitable in the product name is canno image CLASS 2200 Advanced with sal
```

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART | PREVIEW | JSON | EXECUTION DETAILS | EXECUTION

Row	Product_Name	tot_sales_product	tot_sales	Contribute_sales
1	Canon imageCLASS 2200 Adva...	61599.82	2261536.78	0.03
2	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	27453.38	2261536.78	0.01
3	Cisco TelePresence System EX...	22638.48	2261536.78	0.01

❖ What is the most profitable in the product ?

Analysis_Superstore_Sales RUN SAVE QUERY SHARE SCHEDULE

```
SELECT
  DISTINCT Product_Name,
  ROUND(SUM(Sales) OVER (PARTITION BY Product_Name), 2) AS tot_sales_product,
  ROUND(SUM(Sales) OVER (), 2) AS tot_sales,
  ROUND((SUM(Sales) OVER (PARTITION BY Product_Name)/SUM(Sales) OVER ()), 2) AS Contribute_sales
FROM `superstore-410608.sales_superstore.sales`
ORDER BY tot_sales_product DESC LIMIT 3
-- As a whole was most profitable in the product name is canno image CLASS 2200 Advanced with s
```

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART | PREVIEW | JSON | EXECUTION DETAILS | EXECUTIC

Row	Product_Name	tot_sales_product	tot_sales	Contribute_sales
1	Canon imageCLASS 2200 Adva...	61599.82	2261536.78	0.03
2	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	27453.38	2261536.78	0.01
3	Cisco TelePresence System EX...	22638.48	2261536.78	0.01

❖ What is the most profitable based on ship mode ?

a. The ship mode which the most used by customers

Analysis_Superstore_Sales RUN SAVE QUERY SHARE SCHEDULE

```
93  -- the shipmode most used by customers
94  SELECT
95    DISTINCT Ship_Mode,
96    COUNT (*) OVER (PARTITION BY Ship_Mode) AS count_ship_mode,
97    COUNT (*) OVER () AS total_count_ship,
98    ROUND ((COUNT(*) OVER (PARTITION BY Ship_Mode) / COUNT(*) OVER ()), 2) AS Contribute_sales_shipmode
99  FROM `superstore-410608.sales_superstore.sales`
100 ORDER BY count_ship_mode DESC LIMIT 3
```

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAI

Row	Ship_Mode	count_ship_mode	total_count_ship	Contribute_sales_shi
1	Standard Class	5859	9800	0.6
2	Second Class	1902	9800	0.19
3	First Class	1501	9800	0.15

a. What Shipmode was the most profitable

Analysis_Superstore_Sales RUN SAVE QUERY SHARE SCHEDULE

```
104 -- Shipmode which the most profitable
105 SELECT
106   DISTINCT Ship_Mode,
107   ROUND(SUM(Sales) OVER (PARTITION BY Ship_Mode), 2) AS Tot_sales_shipmode,
108   ROUND(SUM(Sales) OVER (), 2) AS tot_sales,
109   ROUND((SUM(Sales) OVER (PARTITION BY Ship_Mode)/SUM(Sales) OVER()), 2) AS Contribute_sales_shipmode
110 FROM `superstore-410608.sales_superstore.sales`
111 ORDER BY Tot_sales_shipmode DESC LIMIT 3
112   -- Shipmode which the most profitable is Standard Class with sales $ 1,340,831.78
```

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAF

Row	Ship_Mode	Tot_sales_shipmode	tot_sales	Contribute_sales_shi
1	Standard Class	1340831.31	2261536.78	0.59
2	Second Class	449914.18	2261536.78	0.2
3	First Class	345572.26	2261536.78	0.15

❖ What the difference in days between the order date and the ship date affect sales?

Analysis_Superstore_Sales RUN SAVE QUERY

```
116 -- Korelasi between Difference days with sales
117 SELECT
118   ROUND(CORR(Diff_Day, Sales), 2) AS Coefisien_korelasi
119 FROM
120   (
121   SELECT
122     Order_date,
123     Ship_date,
124     DATE_DIFF(Ship_date, Order_date, DAY) AS Diff_Day,
125     Sales
126   FROM `superstore-410608.sales_superstore.sales`
127   )
128   -- There is no correlation between the amount of time betw
      coeffiienct value -0.01
```

Query results

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON

Row	Coefisien_korelasi
1	-0.01

❖ What was the most profitable year ?

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE | M

```sql
SELECT
  DISTINCT sub.Year,
  ROUND(SUM(Sub.Sales) OVER(PARTITION BY sub.Year), 2) AS tot_sales_year,
  ROUND(SUM(Sub.Sales) OVER(), 2) AS total_sales,
  ROUND((SUM(Sub.Sales) OVER (PARTITION BY sub.Year)/SUM(Sub.Sales) OVER ()), 2) AS Contribute_Sales
FROM (
    SELECT
      EXTRACT(YEAR FROM Order_date) AS Year,
      Sales
    FROM `superstore-410608.sales_superstore.sales`
    ) AS Sub
ORDER BY tot_sales_year DESC LIMIT 3
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GR

Row	Year	tot_sales_year	total_sales	Contribute_Sales
1	2018	722052.02	2261536.78	0.32
2	2017	600192.55	2261536.78	0.27
3	2015	479856.21	2261536.78	0.21

Analysis_Superstore_Sales

RUN | SAVE QUERY | SHARE | SCHEDULE | MORE

```sql
-- 7. What is the sales trend from year to year?

CREATE TEMP TABLE t1 AS
SELECT
  sub1.Year,
  sub1.Year_Sales_Total,
  COALESCE(ROUND(sub1.Year_Sales_Total - LAG(sub1.Year_sales_total,1) OVER (ORDER BY sub1.Year),2),0)
  AS YoY_Total_Sales_Change,
  COALESCE(ROUND((sub1.Year_Sales_Total - LAG(sub1.Year_sales_total,1) OVER (ORDER BY sub1.Year))
  /(LAG(sub1.Year_Sales_Total,1) OVER (ORDER BY sub1.Year)),2),0) AS YoY_Total_Sales_Pct_Change
FROM
  (SELECT
      DISTINCT sub.Year,
      ROUND(SUM(sub.Sales) OVER (PARTITION BY sub.Year),2) AS Year_Sales_Total,
      ROUND(SUM(sub.sales) OVER (),2) AS Total_Sales,
      ROUND((SUM(sub.Sales) OVER (PARTITION BY sub.Year))/( SUM(sub.sales) OVER ()),2) AS Sales_Pct
  FROM
    (SELECT
      EXTRACT(YEAR FROM Order_Date) AS Year,
      Sales
    FROM `superstore-410608.sales_superstore.sales`
    ) AS sub
    ORDER BY Year_Sales_Total DESC
  ) AS sub1
ORDER BY sub1.Year;
```

Query results

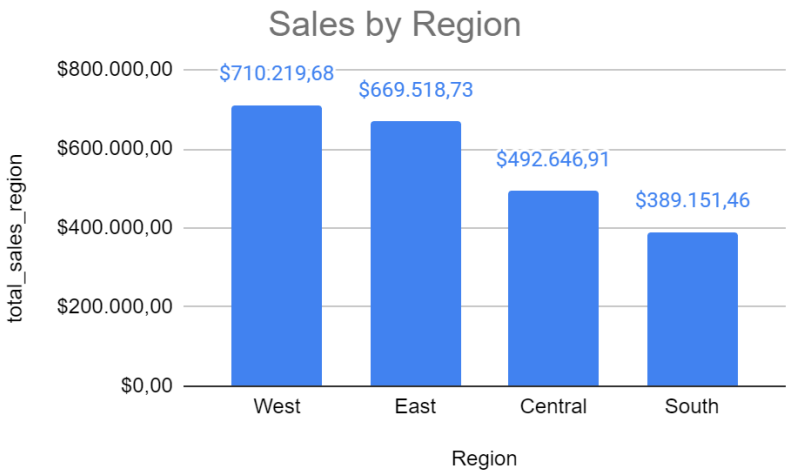
JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS

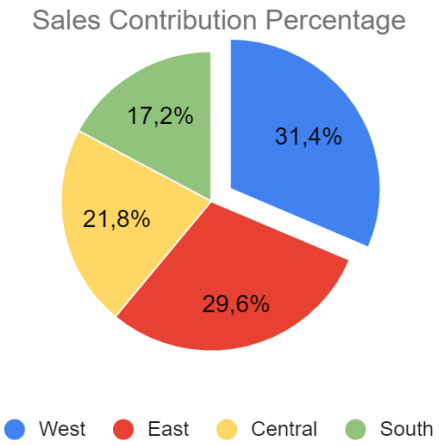
Row	Year	Year_Sales_Total	YoY_Total_Sales_Change	YoY_Total_Sales_Pct_Cha
1	2015	479856.21	0.0	0.0
2	2016	459436.01	-20420.2	-0.04
3	2017	600192.55	140756.54	0.31
4	2018	722052.02	121859.47	0.2

Conclusion

The objective of the analysis can be explained through various questions

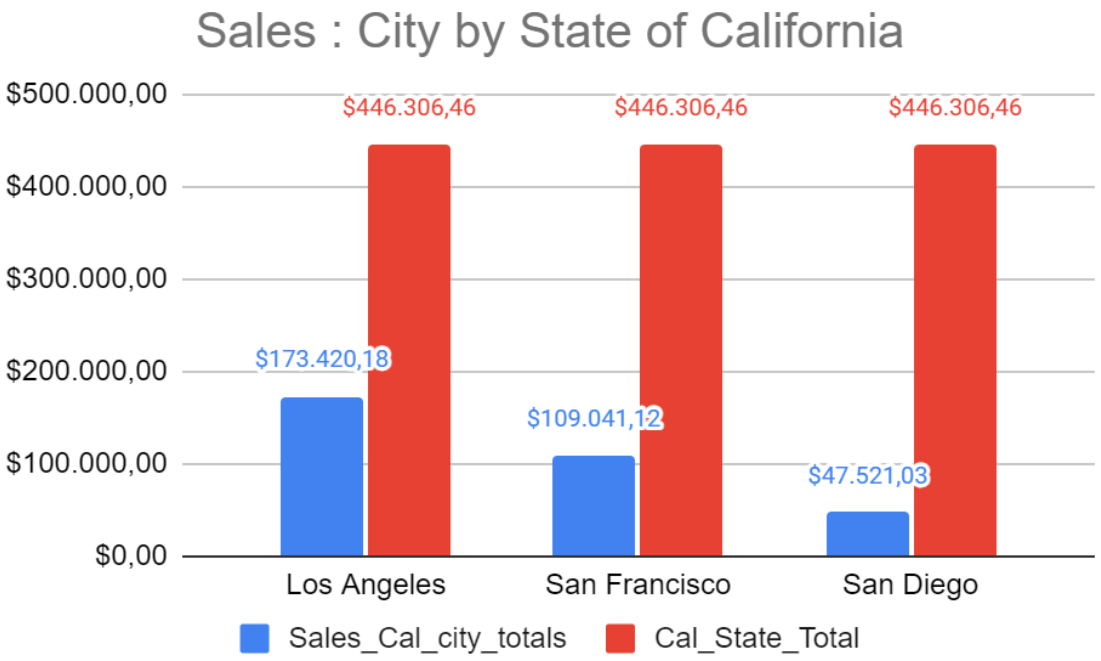
1. What was the most profitable region, state, and city in the dataset ?

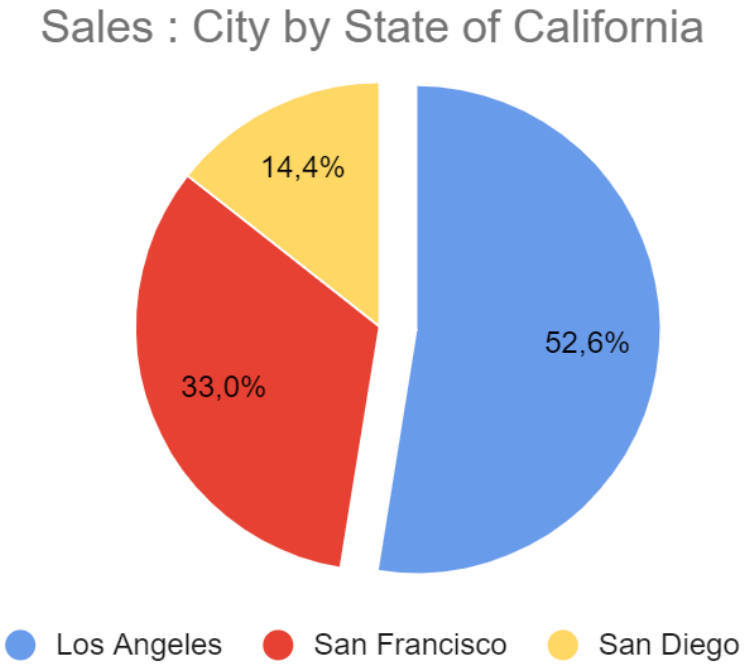




The most profitable sales results when viewed from the regional division are Western. Western contributed the most sales proceeds amounting to $ 710.219,68. These results will become the center of attention for sales performance, so that we can dig deeper into sales mechanisms, customer habits, and maintain and even improve sales quality.

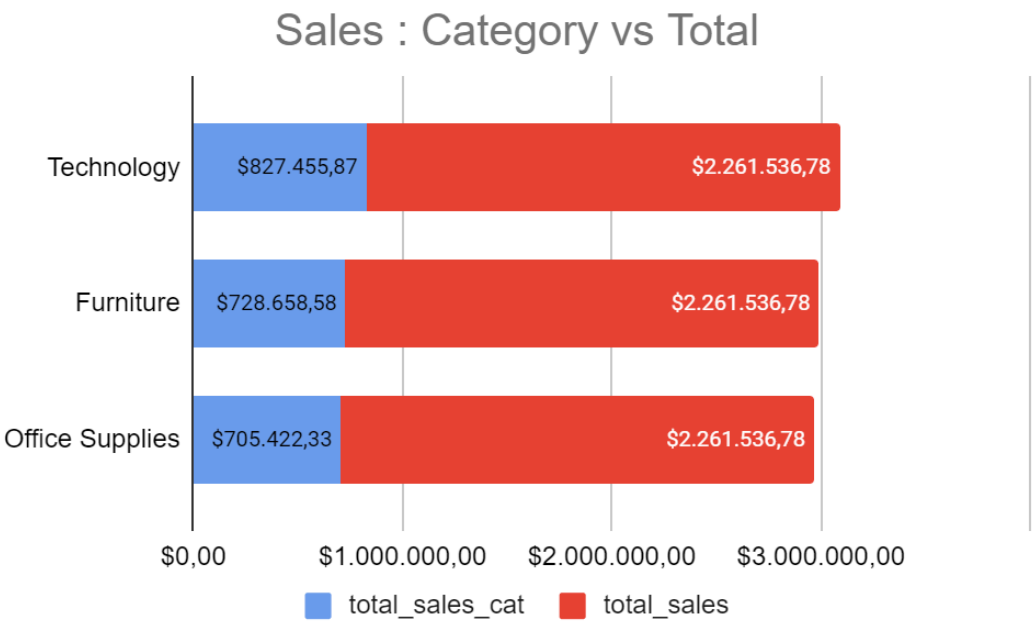
Then specifically about the profitable region, which State in the western dominates sales more is California

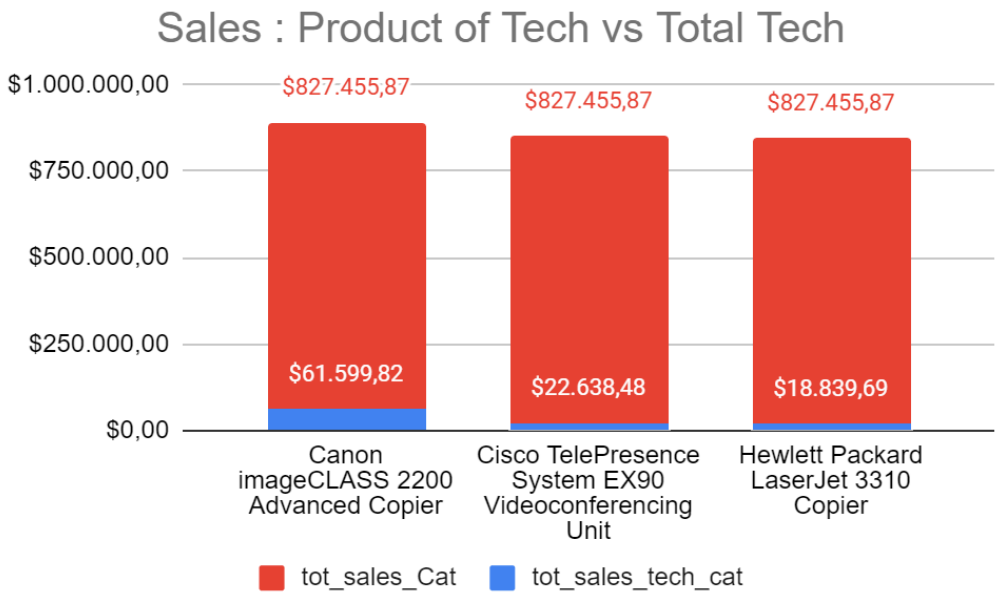




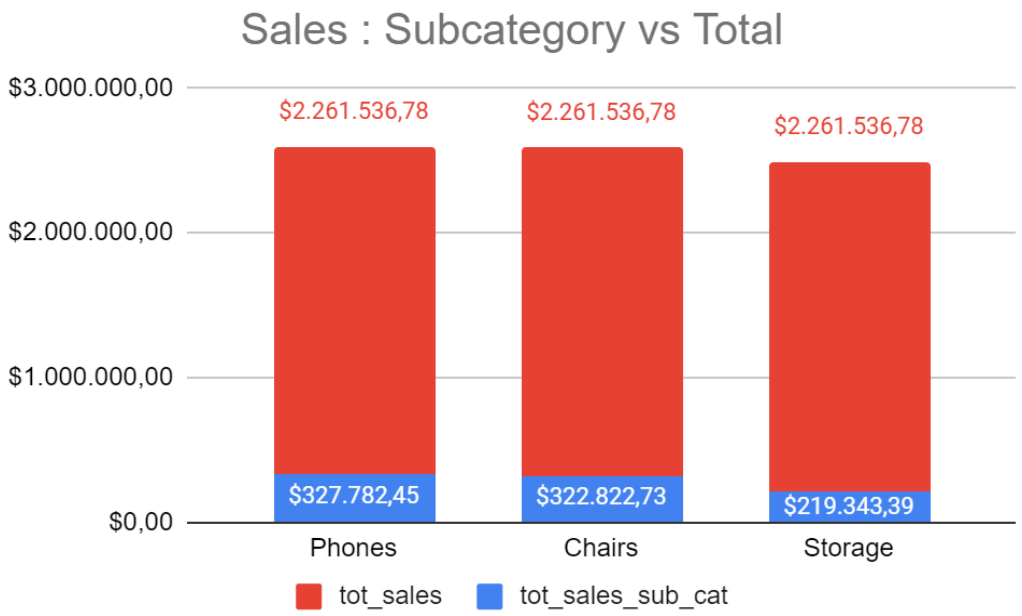
In the region of western was the state most profitable is California with sales amount of $ 446.306,46, and contribution of sales in state of california is 52,6%

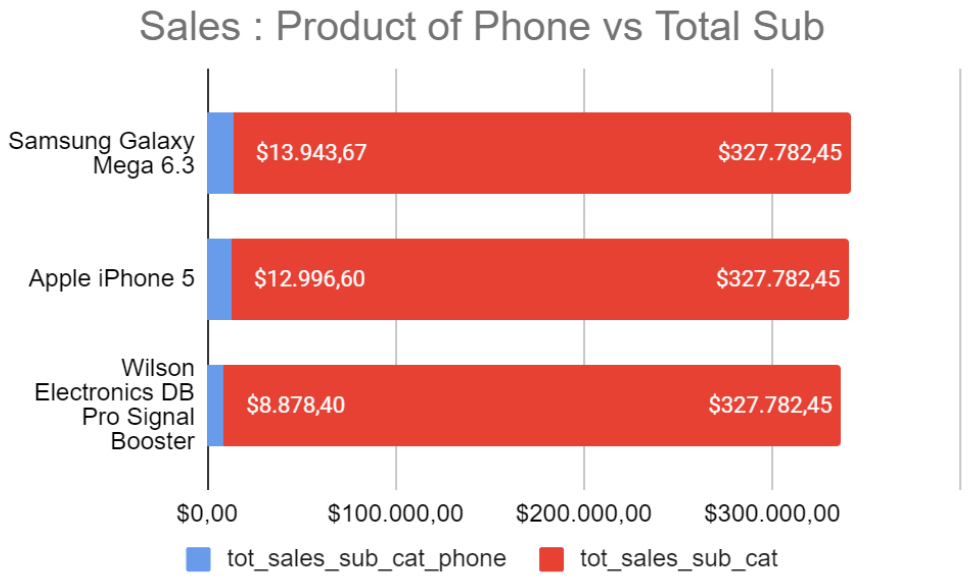
2. What was most profitable based on category and subcategory ?





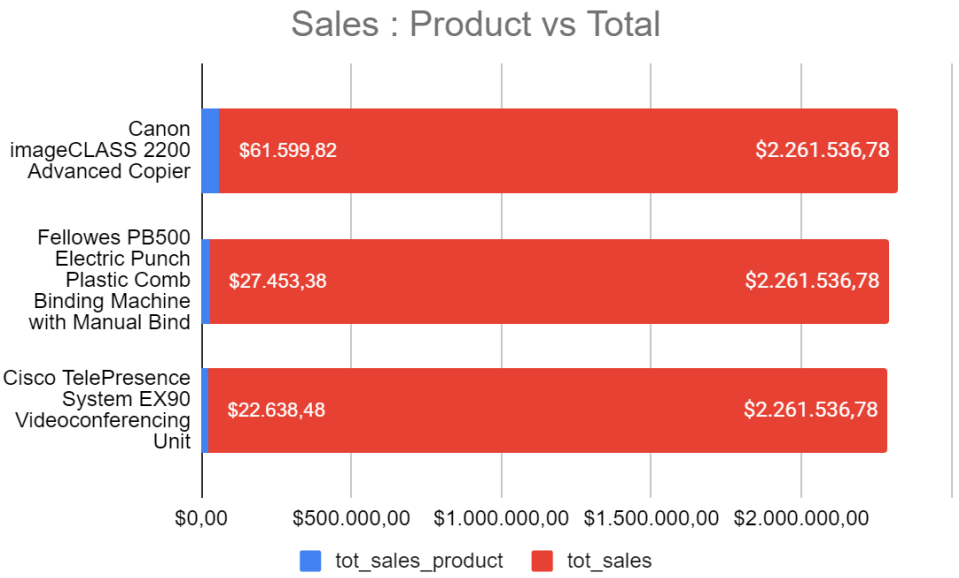
Technology was most profitable in the category with total sales $ 827,455.00 and the most profitable product name in the technology category is “ Canon image Class 2200 Advance Copier ” amount of $ 61.599,82





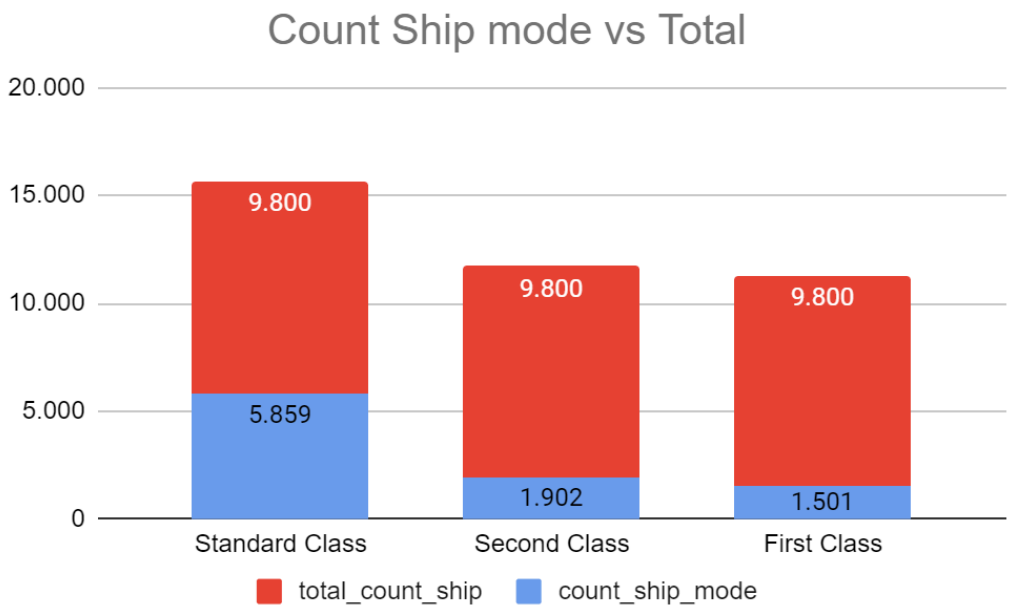
The biggest sales contributor in the sub category is phones with sales of $327.785,45 and the best selling product from the phones subcategory is Samsung Galaxy Mega 6.3 with $12.996,60. From the information above, it shows that products from categories or subcategories that contain elements of technology are the biggest contributors to sales and are the most popular.

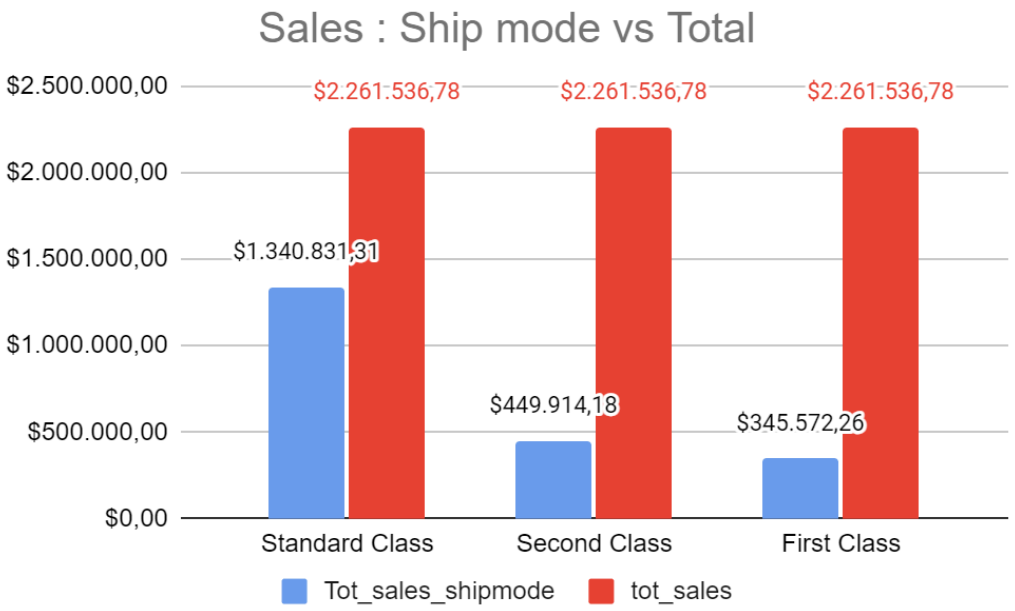
3. What is the most profitable based on product ?



Overall, the product with the most sales is the Canon Image Class 2200 Advanced Copier product for $ 61.599,82 which is also included in the technology category.

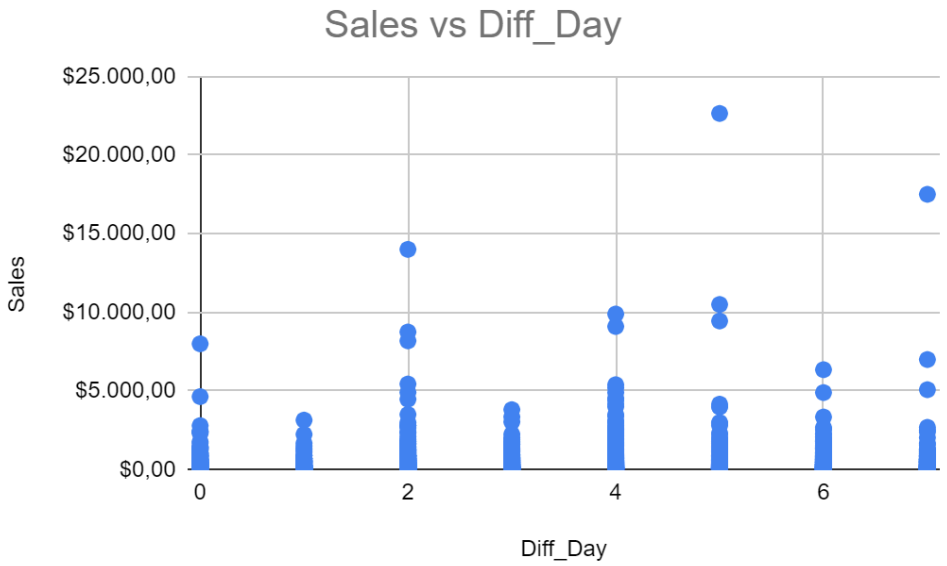
4. What is the most profitable based on ship mode ?





The most widely used shipping mode is standard class with count is 5,859 and the most profitable is Standard Class with sales $ 1.340.831,78. From the data above, it shows that customers prefer the standard class delivery mode, this may indicate customers' financial conditions, customer habits, or good service that is satisfactory for customers.

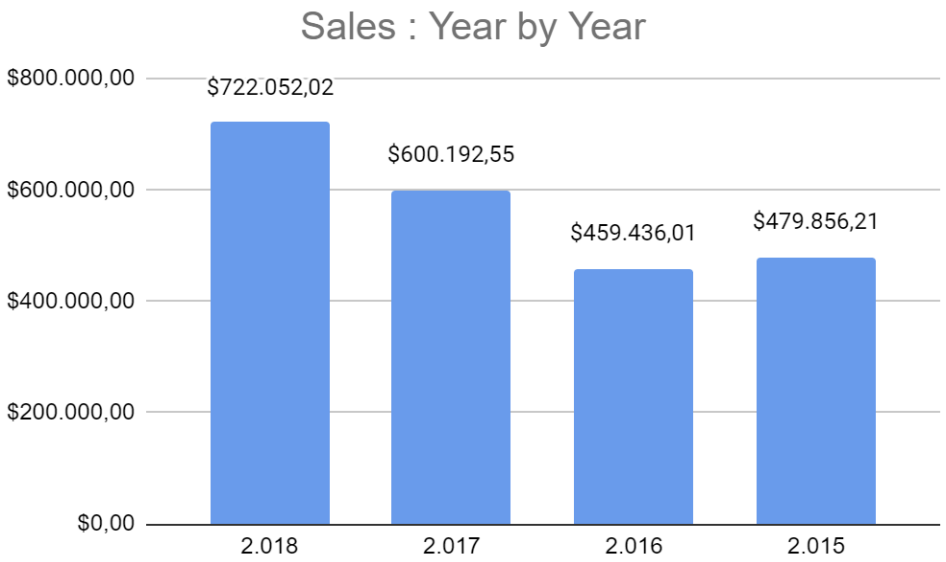
5. What the difference in days between the order date and the ship date affect sales?



There is no correlation between the length of time between order date and delivery date. Shown in the correlation coefficient value -0.01.

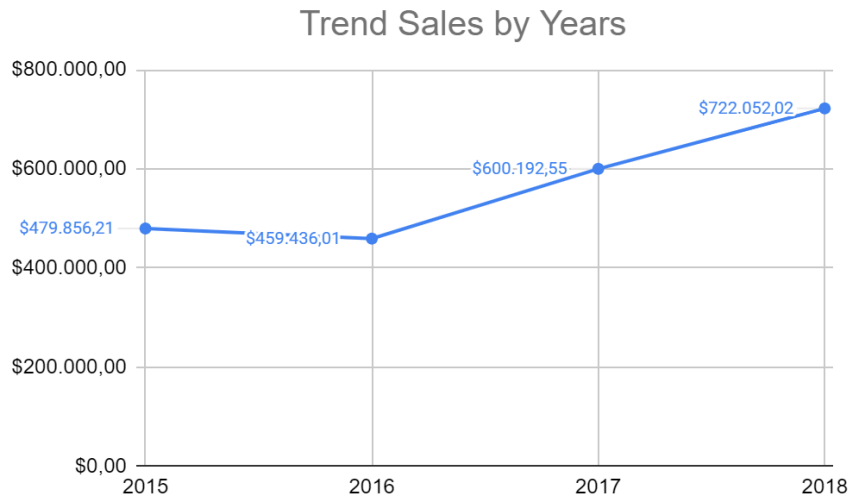
From this question, investigate whether the difference between the order date and the shipping date can influence sales, whether the greater the difference in days between order and shipping, the sales decrease or vice versa. It turns out that the relationship between these two variables does not have a positive correlation, the correlation value shows -0.001. This means that the difference in days does not affect sales.

6. What is the most profitable year during the order period ?

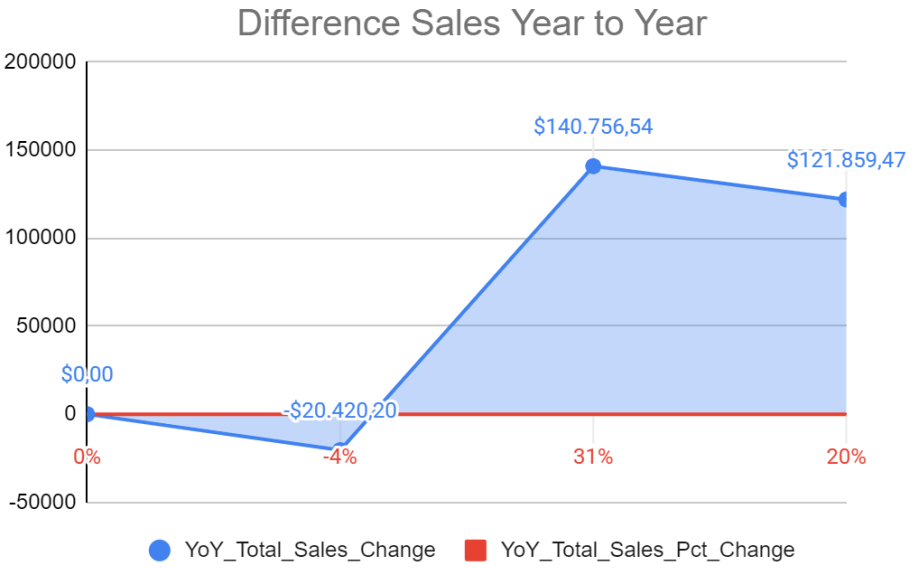


The most profitable year was 2018 with the most sales of $722.052,02, from this information you can find out more about the causes and effects that provided the most profitable sales in that year.

7. What is the sales trend from year to year ?



Sales trends from year to year are unstable, there was a decline in sales from 2015 to 2016, then increased again the following year. The peak increase at the end of the period was in 2018 with sales of $722.052,02. 2015 to 2016 became the center of attention as material for evaluation and understanding of field conditions in that year.



The trend of the difference in sales from year to year decreased from 2015 to 2016 amounting to -$-$20,420.20 with a percentage decrease of -4%, then in the following year sales increased by $140,756.54 with a percentage increase from the previous 31%, in the final year sales increased not too significant, not bigger than the previous $121,859.47 with a percentage of 20%

Appendix

- To access the data repository click Here
- To acces the spreadsheet chart click Her