

# EDA on Superstore



---

## **Introduction:**

The *Superstore Sales Analysis Project* uses SQL and data visualization to gain insights from sales data. Focusing on measures like sales, product categories, payment methods, and customer attributes. The project aims to identify trends that can improve business decisions. By analyzing sales patterns across cities, customer types, and time, we get to know top-performing products. Visualizations like bar charts, pie charts, and scatter plots provide a clear picture of performance, helping stakeholders make data-driven decisions to optimize sales and enhance customer satisfaction.

## **Process:**

- ❖ I have used Mysql Workbench for performing analysis.
- ❖ Below steps will give overview of project:
- ❖ Process:
  1. Create database named Superstore
  2. Import dataset i.e(Superstore sales data) into it
  3. Next i have analyzed the data:

(There are 17 columns and 1000 rows each one contains following info)

- 1]Invoice ID:

Computer generated sales slip -[ Invoice Identification]

- 2]Branch:

Branches of superstore -[ A,B,C]

- 3]City:

Location of superstore - [Yangon,Naypyidaw,Mandalay]

- 
- **4]Customer type:**  
**[1]Member-having member card,2) Normal-without member card)**
  - **5]Gender:**  
**Gender of customer-[Male,Female]**
  - **6]Product line:**  
**Categories of items -**  
**[Health & beauty, Electronic Accessories, Home and lifestyle,**  
**Sports and Travel, Food and Beverages, Fashion Accessories]**
  - **7]Unit price:**  
**Price of product (each item)**
  - **8]Quantity:**  
**Number of product purchased by customer**
  - **9]Tax 5%:**  
**Tax on customer buying**
  - **10]Total:**  
**Total price including tax**
  - **11]Date:**  
**Purchased date-jan 2019 to mar 2019**
  - **12]Time:**  
**Purchase time(24 hr clock wise)**
  - **13]Payment:**

---

**Payment method used by customer- [Cash,E Wallet,Credit Card**

- **14]COGS:**

**Cost of Goods Sold**

- **15]Gross Margin Percentage:**

**Gross(bulk) margin percentage**

- **16]Gross Income:**

**Income**

- **17]Rating:**

**Customer satisfaction rating on shopping experience(scale of - 10)**

#### **4. Final Analysis**

# Final Analysis:

(Using Mysql )

Q. 1] Display the first 5 rows from the dataset.

**select \* from table\_superstore limit 5;**

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2019-01-05	13:08	Ewallet	522.83	4.761904762	26.1415	9.1
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	2019-03-08	10:29	Cash	76.4	4.761904762	3.82	9.6
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2019-03-03	13:23	Credit card	324.31	4.761904762	16.2155	7.4
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	2019-01-27	20:33	Ewallet	465.76	4.761904762	23.288	8.4
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2019-02-08	10:37	Ewallet	604.17	4.761904762	30.2085	5.3

**Analysis:** It will give you 5 rows.

Q. 2] Display the last 5 rows from the dataset.

**select \* from table\_superstore order by 'Invoice ID' desc limit 5 ;**

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2019-01-05	13:08	Ewallet	522.83	4.761904762	26.1415	9.1
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	2019-03-08	10:29	Cash	76.4	4.761904762	3.82	9.6
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2019-03-03	13:23	Credit card	324.31	4.761904762	16.2155	7.4
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	2019-01-27	20:33	Ewallet	465.76	4.761904762	23.288	8.4
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2019-02-08	10:37	Ewallet	604.17	4.761904762	30.2085	5.3

**Analysis:** Firstly sorted data using column 'Invoice ID' in descending order and limit 5 is given so that it will give last 5 records.

Q 3] Display random 5 rows from the dataset.

**select \* from table\_superstore order by rand() limit 5;**

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
796-12-2025	C	Naypyitaw	Normal	Male	Fashion accessories	62.12	10	31.06	652.26	2019-02-11	16:19	Cash	621.2	4.761904762	31.06	5.9
131-15-8856	C	Naypyitaw	Member	Female	Food and beverages	72.52	8	29.008	609.168	2019-03-30	19:26	Credit card	580.16	4.761904762	29.008	4
325-90-8763	C	Naypyitaw	Member	Female	Electronic accessories	46.57	10	23.285	488.985	2019-01-27	13:58	Cash	465.7	4.761904762	23.285	7.6
227-50-3718	A	Yangon	Normal	Male	Health and beauty	14.62	5	3.655	76.755	2019-03-04	12:23	Cash	73.1	4.761904762	3.655	4.4
592-46-1692	C	Naypyitaw	Member	Female	Food and beverages	36.77	7	12.8695	270.2595	2019-01-11	20:10	Cash	257.39	4.761904762	12.8695	7.4

**Analysis:** As rand() function is used to get random values so it will give you random 5 rows.

---

**Q.4] Display count, min, max, avg, and std values for a column(gross income) in the dataset.**

Select count(gross income), min(gross income), round(max(gross income),1), round(avg(gross income),1), round(std(gross income),1) from table\_superstore ;

	count(`gross income`)	min(`gross income`)	round(max(`gross income`),1)	round(avg(`gross income`),1)	round(std(`gross income`),1)
►	1000	0.5085	49.6	15.4	11.7

**Analysis:** By using aggregate functions i calculated count is 1000 for rows, Minimum is 0.5 ,Maximum of gross income is 49.6, Average is 15.4 & Standard Deviation of gross income is 11.7

**Q.5] Find the number of missing values.**

select count(\*) from table\_superstore where branch is null;

select count(\*) from table\_superstore where `Invoice Id` is null ;

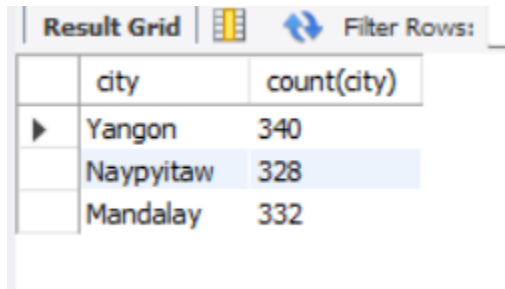
	count(*)
►	0

**Analysis:** As our data is cleaned so no null values present there.

---

**Q.6] Count the number of occurrences of each city.**

**select city, count(city) from table\_superstore group by city ;**



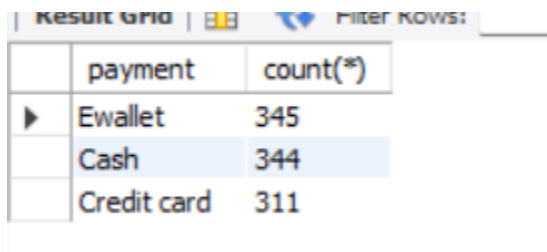
The screenshot shows a 'Result Grid' window with a 'Filter Rows' button. The grid contains three rows of data. The first row is highlighted with a mouse cursor. The columns are 'city' and 'count(city)'.

	city	count(city)
▶	Yangon	340
	Naypyitaw	328
	Mandalay	332

**Analysis:** By using group by function occurrences of cities have been calculated. Where Yangon has more number.

**Q.7] Find the most frequently used payment method.**

**select payment, count(\*) from table\_superstore group by payment order by count(\*) desc limit 1;**



The screenshot shows a 'Result Grid' window with a 'Filter Rows' button. The grid contains three rows of data. The first row is highlighted with a mouse cursor. The columns are 'payment' and 'count(\*)'.

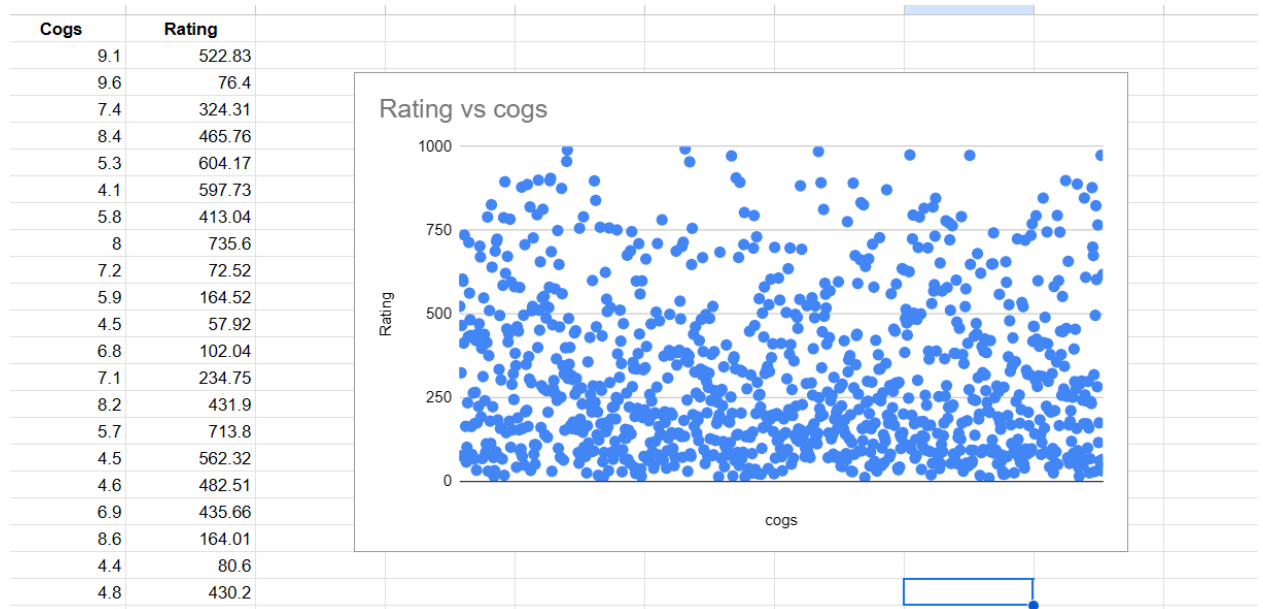
	payment	count(*)
▶	Ewallet	345
	Cash	344
	Credit card	311

**Analysis:** Using group by function count of frequently used payment method is calculated and by sorting them in descending order gives result that Ewallet is the method.

---

**Q.8] Does The Cost of Goods Sold Affect The Ratings That The Customers Provide?**

**select cogs,Rating from table\_superstore ;**



**Analysis:** By plotting cogs vs rating scatter plot we can see that cogs wont get affected by rating.



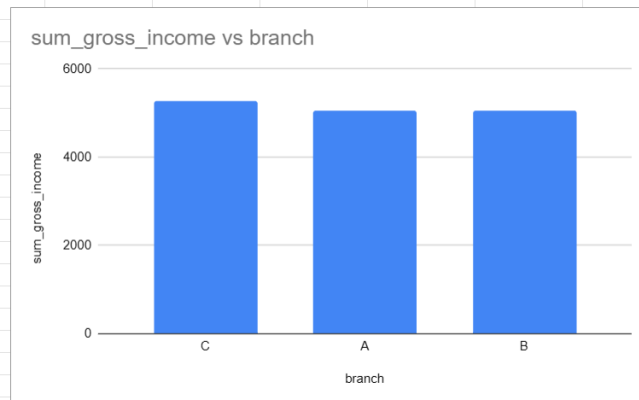
---

**Q.9] Find the most profitable branch as per gross income.**

```
select branch,round(sum(`gross income`),2)sum_gross_income from  
table_superstore group by branch order by sum_gross_income desc;
```

	branch	sum_gross_income
▶	C	5265.18
	A	5057.16
	B	5057.03

Branch	sum_gross_income
C	5265.18
A	5057.16
B	5057.03



**Analysis:** In vertical bar graph we can see that branch C has most profit than A and B.

**Q.10] Find the most used payment method city-wise.**

```
[select payment,city,count(payment) from table_superstore group by  
payment,city ;] or
```

```
select city,
```

```
sum(case when payment= "cash" then 1 else 0 end)Cash,
```

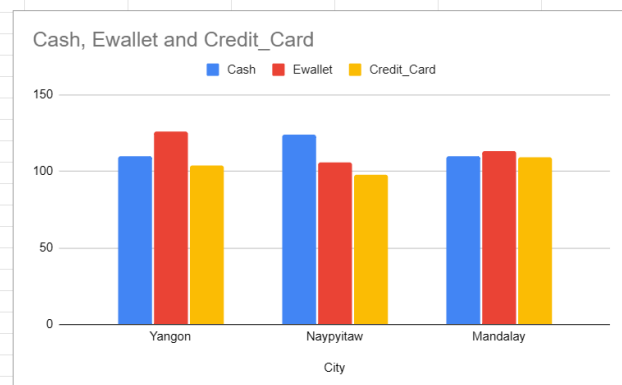
```
sum(case when payment="Ewallet" then 1 else 0 end)Ewallet,
```

```
sum(case when payment="Credit Card" then 1 else 0 end)CreditCard
```

```
from table_superstore group by city ;
```

	city	Cash	Ewallet	CreditCard
►	Yangon	110	126	104
	Naypyitaw	124	106	98
	Mandalay	110	113	109

City	Cash	Ewallet	Credit_Card
Yangon	110	126	104
Naypyitaw	124	106	98
Mandalay	110	113	109



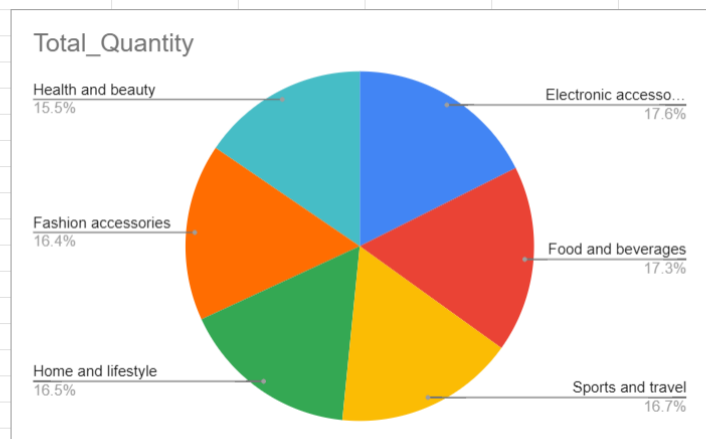
**Analysis:** Here we can see that in Yangon city Ewallet is mostly used and in Naypyitaw Cash method is used and in Mandalay Ewallet is there.

**Q. 11] Find the product line purchased in the highest quantity.**

**select `product line`,sum(quantity)total\_quantity from table\_superstore group by `product line` order by total\_quantity desc ;**

product line	total_quantity
Electronic accessories	971
Food and beverages	952
Sports and travel	920
Home and lifestyle	911
Fashion accessories	902
Health and beauty	854

Product_Line	Total_Quantity
Electronic accessories	971
Food and beverages	952
Sports and travel	920
Home and lifestyle	911
Fashion accessories	902
Health and beauty	854



**Analysis:** Electronic Accessories has highest quantity of purchasing & it is calculated by using group by of product line and sorting of quantity.

---

**Q.12] Display the daily sales by day of the week.**

**UPDATE table\_superstore**

**SET date = CASE**

**WHEN date LIKE '%/%/%' THEN STR\_TO\_DATE(date, '%m/%d/%Y') --**  
**Change to MM/DD/YYYY**

**WHEN date LIKE '%-%-%' THEN STR\_TO\_DATE(date, '%Y-%m-%d') -- Keep**  
**this for YYYY-MM-DD**

**ELSE NULL END**

**WHERE date IS NOT NULL; -- Update only if date is not NULL**

**SET SQL\_SAFE\_UPDATES = 1;**

**Select        dayname(date),dayofweek(date),round(sum(Total),2)        from**  
**table\_superstore group by dayname(date),dayofweek(date);**

	dayname(date)	dayofweek(date)	round(sum(Total),2)
►	Saturday	7	56120.81
	Friday	6	43926.34
	Sunday	1	44457.89
	Monday	2	37899.08
	Thursday	5	45349.25
	Wednesday	4	43731.14
	Tuesday	3	51482.25

**Analysis:** As we can see that on saturday the superstore has most sales or purchasing total which is (56120.81).

---

**Q. 13] Find the month with the highest sales.**

```
select    monthname(date)    name    ,round(sum(total),2)total    from  
table_superstore group by name order by total desc;
```

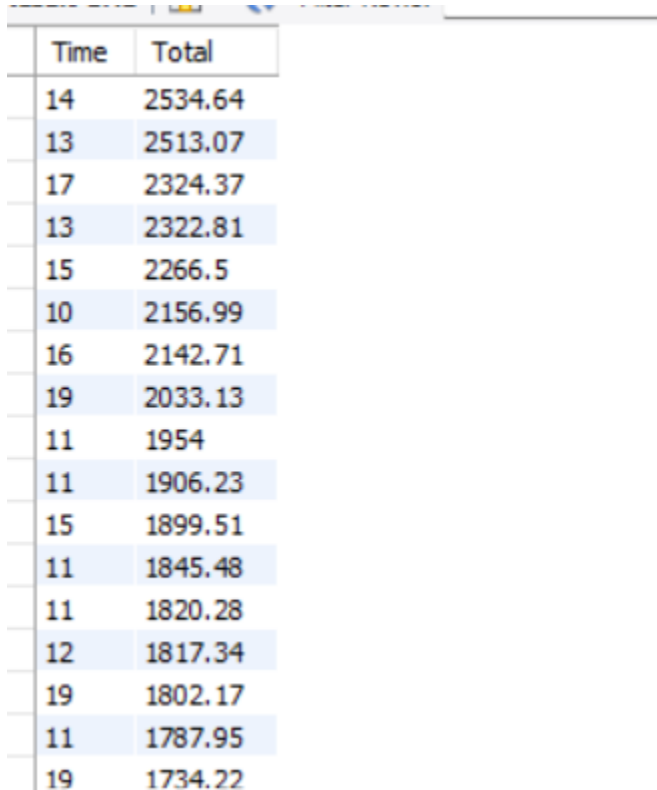
name	total
January	116291.87
March	109455.51
February	97219.37

**Analysis:** In january total is more that is (116291.87).So we can analyze that in january there is highest level of purchasing is there.

---

**Q.14] Find the time at which sales are highest.**

```
select hour(Time)Time,round(sum(Total),2)Total from table_superstore  
group by Time order by Total desc;
```



Time	Total
14	2534.64
13	2513.07
17	2324.37
13	2322.81
15	2266.5
10	2156.99
16	2142.71
19	2033.13
11	1954
11	1906.23
15	1899.51
11	1845.48
11	1820.28
12	1817.34
19	1802.17
11	1787.95
19	1734.22

**Analysis:** By grouping time and sorting total we can see the total per hour of superstore. So it makes easy to analyze the most profitable hour of superstor.

---

**Q. 15] Which gender spends more on average?**

```
select gender,round(avg(`Total`),2)Total from table_superstore group by gender;
```

gender	Total
Female	335.1
Male	310.79

**Analysis:** Female customer rate is more than the male for purchasing in superstore.

## CONCLUSION:

Based on the analysis, it is concluded that the superstore has a strong performance in January, achieving the highest sales figures of 116,291.87 .Yangon is the most frequented city, with E-wallet emerging as the preferred payment method, while Cash dominates in Naypyitaw. The analysis shows that Electronic Accessories are the top-selling product category, and Branch C generates the most profit than A and B. Notably, sales peak on Saturdays, suggesting a trend that could inform marketing strategies. Additionally, the higher purchasing rate among female customers indicates a potential target for future promotions. Overall, these insights provide valuable information for optimizing effective performance.