EDA ON DATASET OF SUPERMARKET USING MYSQL

Step 1: Firstly, use MySQL

Step 2: Create database

Step 3: Import dataset

Step 4: Analyze data

In analyzation we found that dataset includes following columns

- Invoice Id:- Computer generated sales slip
- Branch :- Branch of supermarket identified by A, B, C
- City:- Location of supermarket
- Customer type:- type of customer
    1. member:- customer using membership card
    2. Normal:- customer without membership card
- Gender :- gender of customer
- Product Line:- Gender item categories groups:

1.electric
2. health & beauty
3. Home & lifestyle
4. fashion
5. Sports &travel
6. food & beverages.

- Unit point :- Price of product in $
- Quantity :- Number of products purchased by customer.
- Tax 5%:- 5% tax of customer buying
- Total:- Total price of product including taxes.
- Date:- Date of purchase
- Time:- Time of purchase
- Payment:- Mode of payment
    1. Cash
    2. Ewallet
    3. Credit card
- Cogs:- Cost of goods sold.
- Gross margin percentage:- Gross margin percentage
- Gross income:- income
- Rating:- Customer satisfaction rating on overall shopping experience

Step 5: Solve queries related to dataset and visualized it.

-- 1. Display the first 5 rows from the dataset.

select * from Supermarket limit 5;

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross marg percentag
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08	Ewallet	522.83	4.7619047
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	3/8/2019	10:29	Cash	76.4	4.7619047
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23	Credit card	324.31	4.7619047
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.7619047
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	604.17	4.7619047

-- 2. Display the last 5 rows from the dataset.

select * from Supermarket 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 ma percenta
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08	Ewallet	522.83	4.761904
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	3/8/2019	10:29	Cash	76.4	4.761904
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23	Credit card	324.31	4.761904
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761904
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	604.17	4.761904

-- 3. Display random 5 rows from the dataset.

select * from Supermarket order by rand() 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
228-96-1411	C	Naypyitaw	Member	Female	Food and beverages	98.7	8	39.48	829.08	3/4/2019	20:39	Cash	789.6	4.761904762
243-47-2663	C	Naypyitaw	Member	Male	Electronic accessories	18.77	6	5.631	118.251	1/28/2019	16:43	Credit card	112.62	4.761904762
743-04-1105	B	Mandalay	Member	Male	Health and beauty	97.22	9	43.749	918.729	3/30/2019	14:43	Ewallet	874.98	4.761904762
590-83-4591	B	Mandalay	Member	Male	Electronic accessories	87.45	6	26.235	550.935	2/17/2019	14:40	Credit card	524.7	4.761904762
423-57-2993	B	Mandalay	Normal	Male	Sports and travel	93.39	6	28.017	588.357	3/27/2019	19:18	Ewallet	560.34	4.761904762

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

```
select count(`gross income`) as count,
min(`gross income`) as minimum,
max(`gross income`) as maximum,
avg(`gross income`) as average,
std(`gross income`) as Standard_deviation
from Supermarket;
```

count	minimum	maximum	average	Standard_deviation
1000	0.5085	49.65	15.379369000000002	11.702969603922716

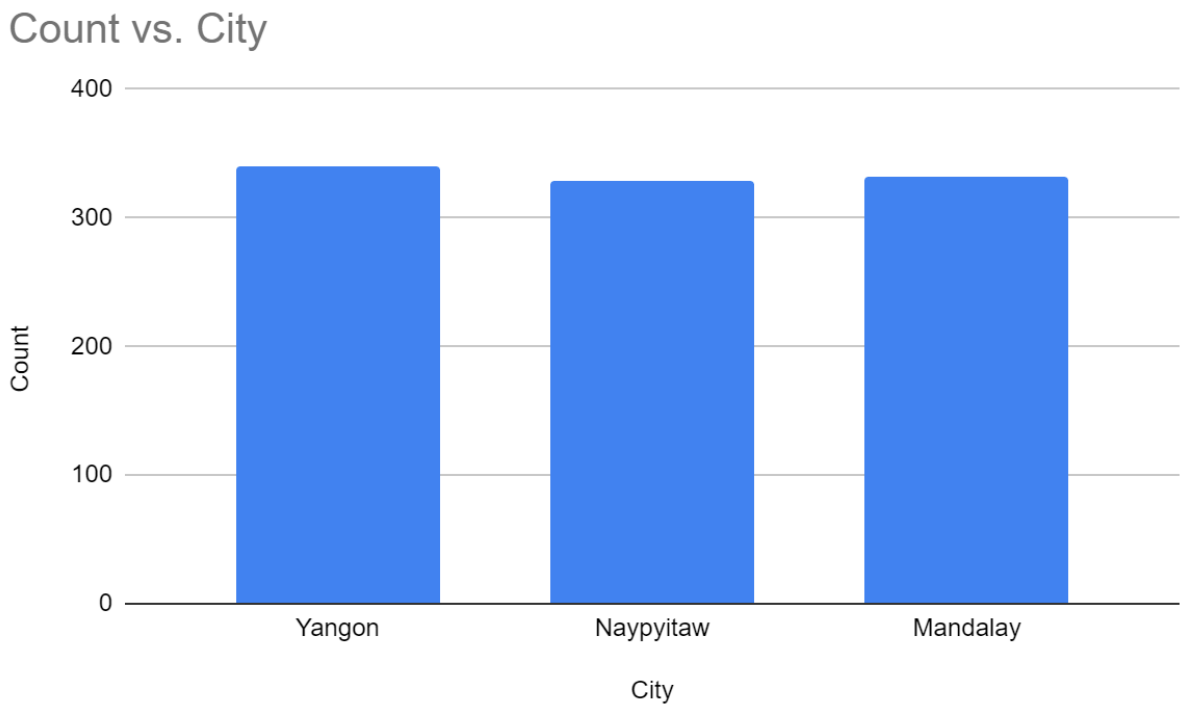
-- 5. Find the number of missing values.

```
select count(*) as null_value from supermarket where branch is null
```

-- 6. Count the number of occurrences of each city.

select city, count(city) as no_of_cities from supermarket group by city;

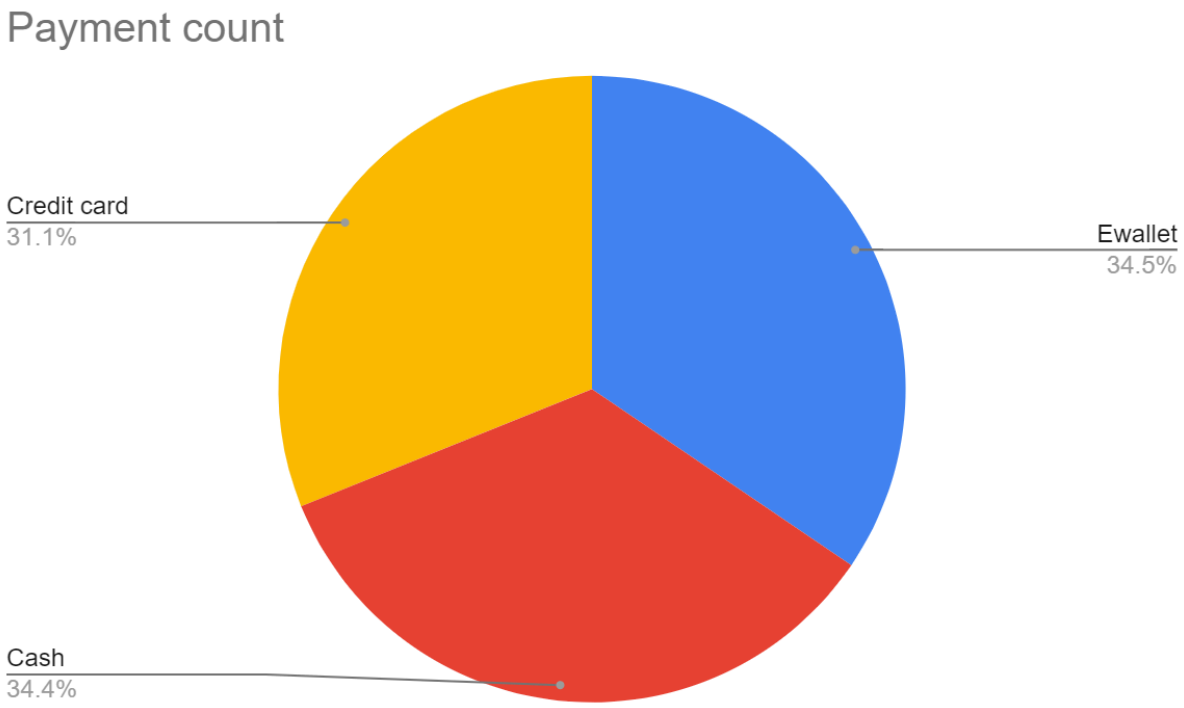
city	no_of_cities
Yangon	340
Mandalay	332
Naypyitaw	328



-- 7. Find the most frequently used payment method.

```
select  payment, count(*) as Payment_count
from supermarket
group by payment
order by count(*) desc;
```

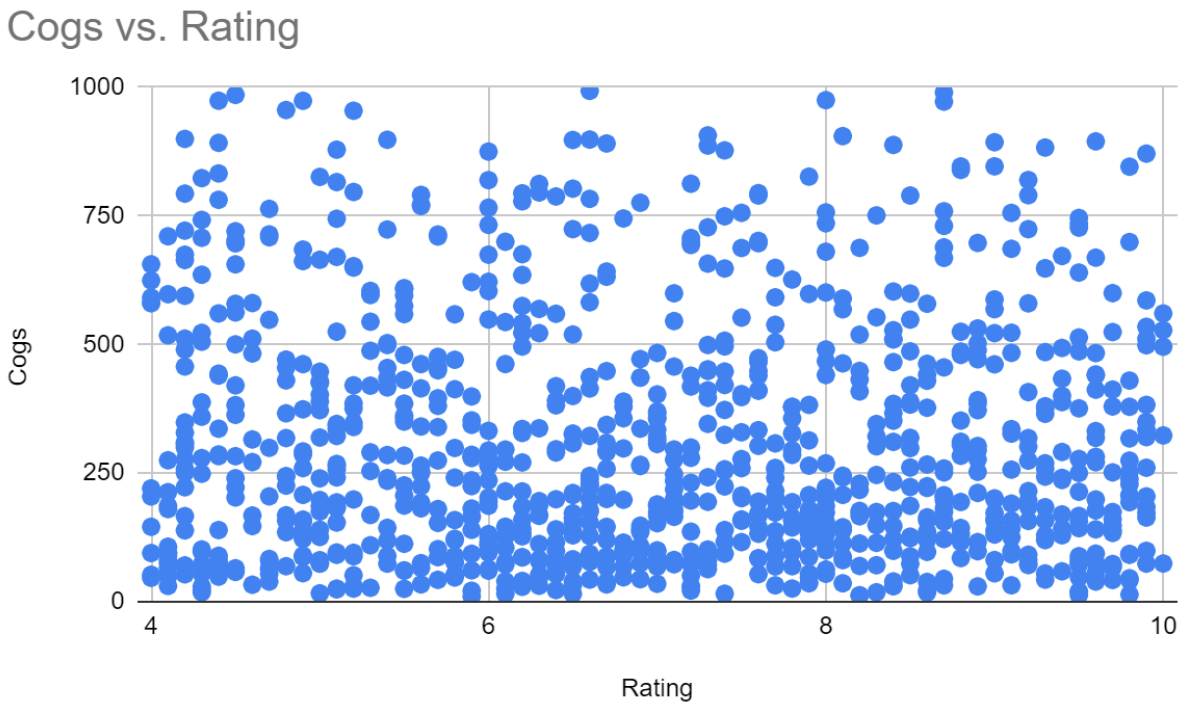
payment	Payment_count
Ewallet	345
Cash	344
Credit card	311



-- 8. Does The Cost of Goods Sold Affect The Ratings That The Customers Provide?

select Rating, cogs from supermarket;

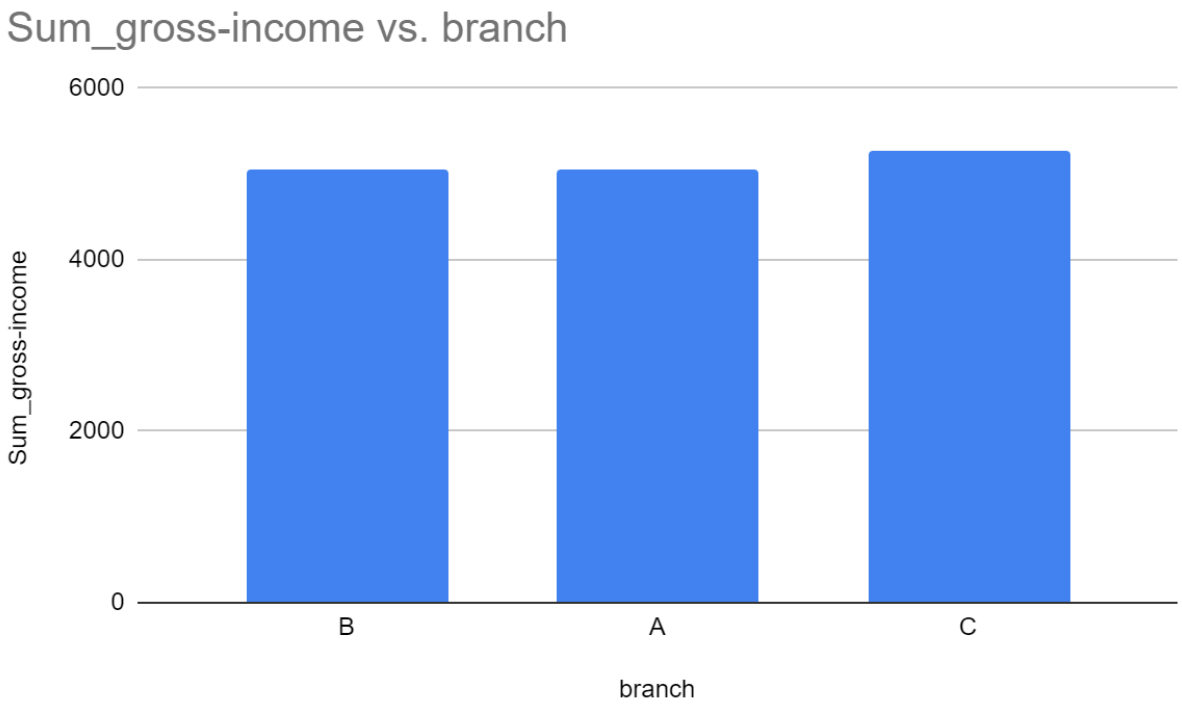
	Rating	cogs
	9.1	522.83
	9.6	76.4
	7.4	324.31
	8.4	465.76
	5.3	604.17
	4.1	597.73
	5.8	413.04
	8	735.6
	7.2	72.52
	5.9	164.52
	4.5	57.92



-- 9. Find the most profitable branch as per gross income.

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

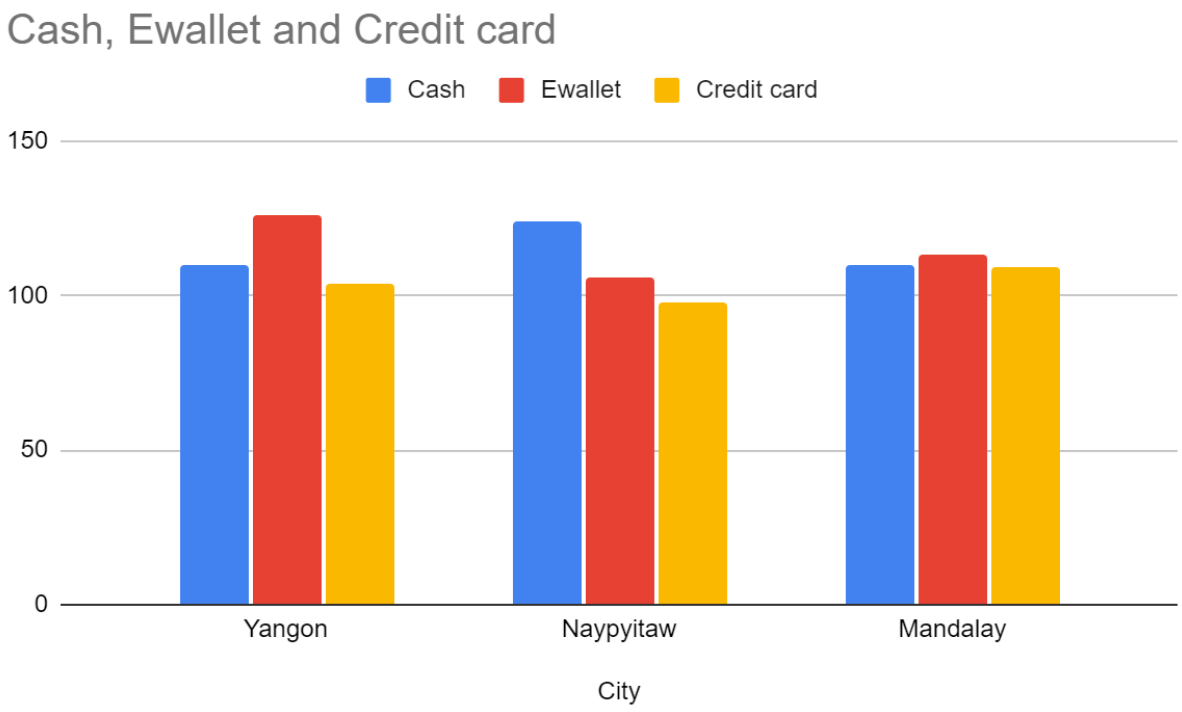
branch	sum_gross_income
B	5057.03
A	5057.16
C	5265.18



-- 10. Find the most used payment method city-wise.

```
select city,
sum(case when payment = "cash" then 1 else 0 end) as "cash",
sum(case when payment = "Ewallet" then 1 else 0 end) as "Ewallet",
sum(case when payment = "credit card" then 1 else 0 end) as "Credit_casrd"
from supermarket group by city;
```

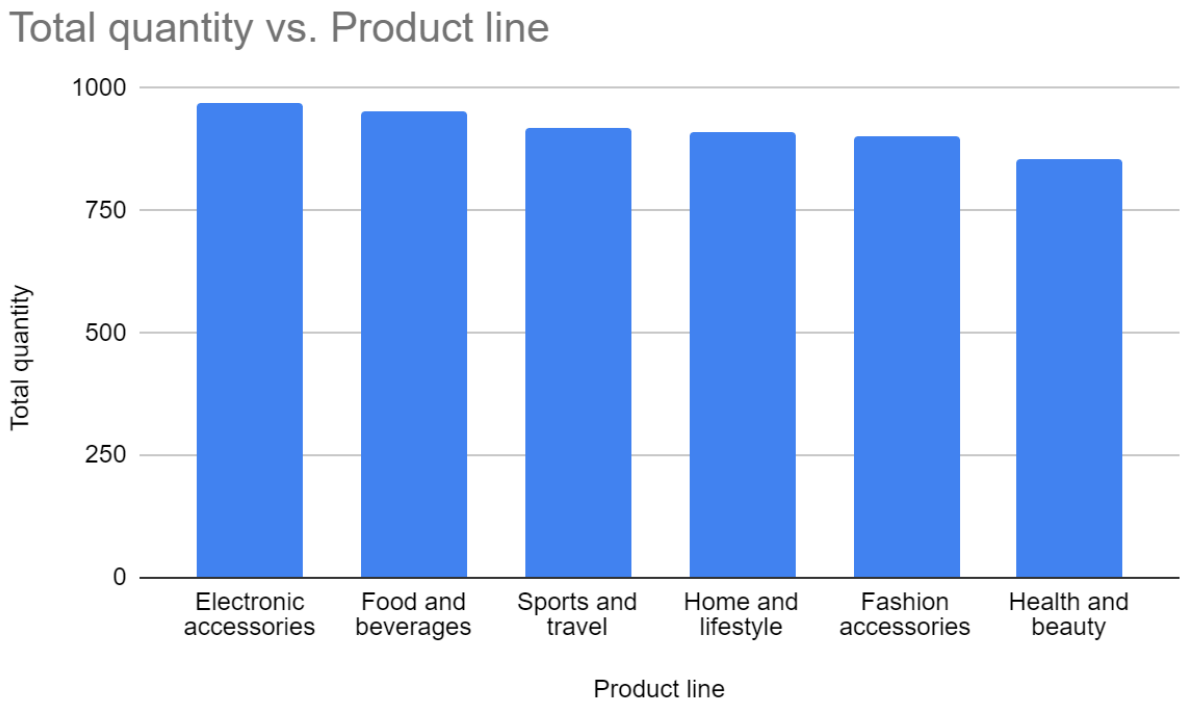
city	cash	Ewallet	Credit_casrd
Yangon	110	126	104
Naypyitaw	124	106	98
Mandalay	110	113	109



-- 11. Find the product line purchased in the highest quantity.

select `Product line`, sum(quantity) as Total_quantity from supermarket
group by `Product line` order by Total_quantity desc;

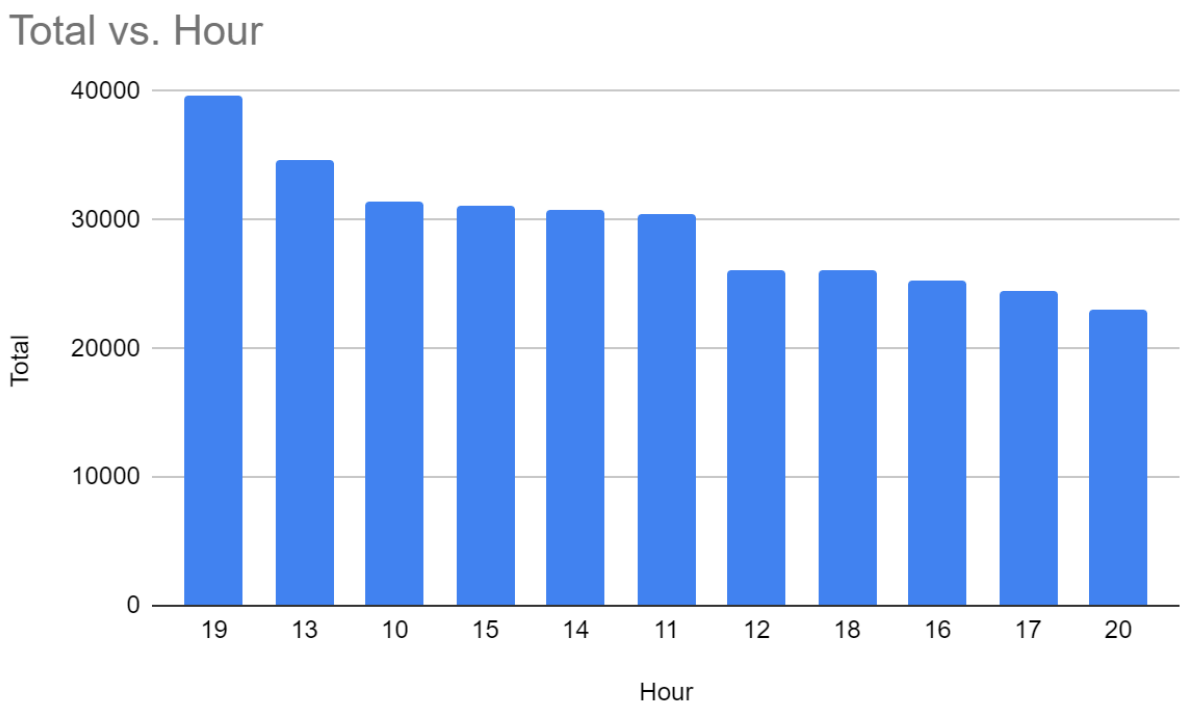
Product line	Total_quantity
Electronic accessories	971
Food and beverages	952
Sports and travel	920
Home and lifestyle	911
Fashion accessories	902
Health and beauty	854



-- 12. Find the time at which sales are highest.

select hour(Time) as hour, sum(Total) as total from supermarket
group by hour order by total desc;

hour	total
19	39699.51300000002
13	34723.22700000001
10	31421.48100000001
15	31179.508499999996
14	30828.399
11	30377.329499999996
12	26065.882499999996
18	26030.339999999986
16	25226.323499999995
17	24445.218
20	22969.527000000002



-- 13. Which gender spends more on average?

select Gender, avg(`gross income`) from supermarket group by Gender;

	Gender	avg(`gross income`)
▶	Female	15.956936127744514
	Male	14.7994869739479