Walmart Sales Analysis

SQL Project

Table Description

Sales

Column Name	Data Type	Description
invoice_id	VARCHAR(30)	Invoice of the sales made
branch	VARCHAR(5)	Branch at which sales were made
city	VARCHAR(30)	The location of the branch
customer_type	VARCHAR(30)	The type of the customer
gender	VARCHAR(10)	Gender of the customer making purchase
product_line	VARCHAR(100)	Product line of the product sold
unit_price	DECIMAL(10, 2)	The price of each product
quantity	INT	The amount of the product sold
VAT	FLOAT(6, 4)	The amount of tax on the purchase
total	DECIMAL(10, 2)	The total cost of the purchase
date	DATE	The date on which the purchase was made
time	TIMESTAMP	The time at which the purchase was made
payment_method	DECIMAL(10, 2)	The total amount paid
cogs	DECIMAL(10, 2)	Cost Of Goods sold
gross_margin_percentage	FLOAT(11, 9)	Gross margin percentage
gross_income	DECIMAL(10, 2)	Gross Income
rating	FLOAT(2, 1)	Rating

Generic Question

What is the average rating of each product line

```
select productline, round(avg(rating),2) as avg_rating
from sales
group by productline
order by avg_rating  desc;
```

productline	avg_rating
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	7
Electronic accessories	6.92
Sports and travel	6.92
Home and lifestyle	6.84

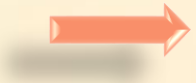
How many unique cities does the data have

```
SELECT DISTINCT
    (city)
FROM
    sales;
```

city
Yangon
Naypyitaw
Mandalay

In which city is each branch

```
SELECT DISTINCT
    branch
FROM
    sales;
```



branch
A
C
B

Product

How many unique product lines does the data have

```
SELECT
    COUNT(DISTINCT productline)
FROM
    sales;
```

	COUNT(DISTINCT productline)
▶	6

What is the most common payment method

```
SELECT
    payment, COUNT(payment) AS Cnt
FROM
    sales
GROUP BY payment
ORDER BY Cnt DESC;
```

	payment	Cnt
▶	Ewallet	345
	Cash	344
	Credit card	311

What is the most selling product line

```
SELECT
    productline, COUNT(productline) AS Most_Selling
FROM
    sales
GROUP BY productline
ORDER BY Most_Selling DESC;
```

productline	Most_Selling
Fashion accessories	178
Food and beverages	174
Electronic accessories	170
Sports and travel	166
Home and lifestyle	160
Health and beauty	152

What is the total revenue by month

```
SELECT
    month_name AS month, SUM(total)AS total_revenue
FROM
    sales
GROUP BY Month
ORDER BY total_revenue DESC;
```

month	total_revenue
January	116291.86800000005
March	109455.50700000004
February	97219.37399999997

What month had the largest COGS

```sql
SELECT
    month_name AS Month, SUM(cogs) AS Largest_cogs
FROM
    sales
GROUP BY Month
ORDER BY Largest_cogs DESC;
```

Month	Largest_cogs
January	110754.16000000002
March	104243.33999999997
February	92589.88

What productline has the largest revenue

```sql
SELECT
    productline, SUM(total) AS Largest_revenue
FROM
    sales
GROUP BY productline
ORDER BY Largest_revenue DESC;
```

productline	Largest_revenue
Food and beverages	56144.844000000005
Sports and travel	55122.826499999996
Electronic accessories	54337.531500000005
Fashion accessories	54305.895
Home and lifestyle	53861.91300000001
Health and beauty	49193.739000000016

What is the city with the largest revenue

```
SELECT
    branch, city, SUM(total) AS Largest_revenue
FROM
    sales
GROUP BY branch , city
ORDER BY Largest_revenue DESC;
```

branch	city	Largest_revenue
C	Naypyitaw	110568.70649999994
A	Yangon	106200.3705000001
B	Mandalay	106197.67199999996

What productline has the largest VAT

```
SELECT
    productline, AVG(Tax) AS avg_Tax
FROM
    sales
GROUP BY productline
ORDER BY avg_tax ;
```

productline	avg_Tax
Fashion accessories	14.528061797752809
Electronic accessories	15.22059705882354
Food and beverages	15.365310344827583
Health and beauty	15.411572368421048
Sports and travel	15.812629518072285
Home and lifestyle	16.03033125000001

Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales

```sql
WITH overall_avg AS (
    SELECT AVG(Quantity) AS avg_qty
    FROM sales
)
SELECT productline,
    CASE
        WHEN AVG(Quantity) > (SELECT avg_qty FROM overall_avg) THEN 'Good'
        ELSE 'Bad'
    END AS remark
FROM sales
GROUP BY productline;
```

	productline	remark
▶	Health and beauty	Good
	Electronic accessories	Good
	Home and lifestyle	Good
	Sports and travel	Good
	Food and beverages	Bad
	Fashion accessories	Bad

Which branch sold more products than average product sold

```sql
SELECT
    branch, SUM(quantity) AS quty
FROM
    sales
GROUP BY branch
HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);
```

	branch	quty
▶	A	1859
	C	1831
	B	1820

What is the most common product line by gender

```sql
SELECT
    gender, productline, COUNT(gender) AS cnt
FROM
    sales
GROUP BY gender , productline
ORDER BY cnt DESC
```

	gender	productline	cnt
▶	Female	Fashion accessories	96
	Female	Food and beverages	90
	Male	Health and beauty	88
	Female	Sports and travel	88
	Male	Electronic accessories	86
	Female	Electronic accessories	84
	Male	Food and beverages	84
	Male	Fashion accessories	82
	Male	Home and lifestyle	81
	Female	Home and lifestyle	79
	Male	Sports and travel	78
	Female	Health and beauty	64

What is the average rating of each product line

```
select productline, round(avg(rating) as avg_rating
from sales
group by  productline
order by avg_rating desc;
```

	productline	avg_rating
▶	Food and beverages	7.11
	Fashion accessories	7.03
	Health and beauty	7
	Electronic accessories	6.92
	Sports and travel	6.92
	Home and lifestyle	6.84

Sales

Which of the customer types brings the most revenue

```
SELECT
    customertype,
    SUM(total) AS total_revenue
FROM sales
GROUP BY customertype
ORDER BY total_revenue;
```

	Customertype	total_revenue
▶	Normal	158743.30500000005
	Member	164223.44400000002

Number of sales made in each time of the day per weekday

```
SELECT
    time_of_day,
    COUNT(*) AS total_sales
FROM sales
WHERE day_name = "Monday"
GROUP BY time_of_day
ORDER BY total_sales DESC;
```

	time_of_day	total_sales
▶	Evening	56
	Afternoon	48
	Morning	21

Evenings experience most sales, the stores are filled during the evening hours

Which customer type pays the most in VAT

```sql
SELECT
    customertype, SUM(tax) AS total_vat
FROM
    sales
GROUP BY customertype
ORDER BY total_vat DESC
LIMIT 1;
```

	customertype	total_vat
▶	Member	7820.164000000002

Which city has the largest tax percent/ VAT (**Value Added Tax**)

```sql
SELECT city, tax
FROM sales
ORDER BY tax DESC
LIMIT 1;
```

	city	tax
▶	Naypyitaw	49.65

Customer

How many unique customer types does the data have

```
SELECT DISTINCT
    (customertype)
FROM
    sales
```

	customertype
▶	Member
	Normal

How many unique payment methods does the data have

```
SELECT DISTINCT
    (payment)
FROM
    sales
```

	payment
▶	Ewallet
	Cash
	Credit card

What is the most common customer type

```sql
SELECT
    customertype, COUNT(*) AS count
FROM
    sales
GROUP BY customertype
ORDER BY count DESC
```

	customertype	count
▶	Member	501
	Normal	499

Which customer type buys the most

```sql
SELECT
    customertype,
    SUM(quantity) AS total_quantity
FROM
    sales
GROUP BY
    customertype
ORDER BY
    total_quantity DESC
LIMIT 1;
```

	customertype	total_quantity
▶	Member	2785

What is the gender of most of the customers

```sql
SELECT
    gender, COUNT(*) AS gender_cnt
FROM
    sales
GROUP BY gender
ORDER BY gender_cnt DESC;
```

	gender	gender_cnt
▶	Female	501
	Male	499

What is the gender distribution per branch

```sql
SELECT
    gender, COUNT(*) AS gender_cnt
FROM
    sales
WHERE
    branch = 'A'
GROUP BY gender
ORDER BY gender_cnt DESC
```

For branch – ‘A’

	gender	gender_cnt
▶	Male	179
	Female	161

```sql
SELECT
    gender, COUNT(*) AS gender_cnt
FROM
    sales
WHERE
    branch = 'B'
GROUP BY gender
ORDER BY gender_cnt DESC
```

For branch – ‘B’

	gender	gender_cnt
▶	Male	170
	Female	162

For branch – ‘C’

```sql
SELECT
    gender, COUNT(*) AS gender_cnt
FROM
    sales
WHERE
    branch = 'C'
GROUP BY gender
ORDER BY gender_cnt DESC
```

	gender	gender_cnt
▶	Female	178
	Male	150

Gender per branch is more or less the same hence, I don't think has an effect of the sales per branch and other factors.

Which time of the day do customers give most ratings

```sql
SELECT
    time_of_day, AVG(rating) AS avg_rating
FROM
    sales
GROUP BY time_of_day
ORDER BY avg_rating DESC
```

time_of_day	avg_rating
Afternoon	7.031299734748012
Morning	6.960732984293193
Evening	6.926851851851853

Looks like time of the day does not really affect the rating, its more or less the same rating each time of the day

Which time of the day do customers give most ratings per branch

For branch – ‘A’

```sql
SELECT
    time_of_day, branch, AVG(rating) AS avg_rating
FROM
    sales
WHERE
    branch = 'A'
GROUP BY time_of_day
ORDER BY avg_rating DESC
```

	time_of_day	branch	avg_rating
▶	Afternoon	A	7.188888888888891
	Morning	A	7.005479452054794
	Evening	A	6.893617021276596

For branch – ‘B’

```sql
SELECT
    time_of_day, branch, AVG(rating) AS avg_rating
FROM
    sales
WHERE
    branch = 'B'
GROUP BY time_of_day
ORDER BY avg_rating DESC
```

	time_of_day	branch	avg_rating
▶	Morning	B	6.891525423728813
	Afternoon	B	6.836799999999998
	Evening	B	6.7729729729729735

```sql
SELECT
    time_of_day, branch, AVG(rating) AS avg_rating
FROM
    sales
WHERE
    branch = 'C'
GROUP BY time_of_day
ORDER BY avg_rating DESC
```

time_of_day	branch	avg_rating
Evening	C	7.118881118881118
Afternoon	C	7.066666666666664
Morning	C	6.974576271186442

Branch A and C are doing well in ratings, branch B needs to do a little more to get better ratings.

Which day of the week has the best avg ratings

```sql
SELECT
    day_name, AVG(rating) AS avg_rating
FROM
    sales
GROUP BY day_name
ORDER BY avg_rating DESC;
```

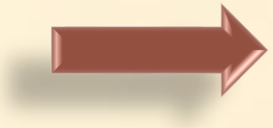
	day_name	avg_rating
▶	Monday	7.153599999999999
	Friday	7.076258992805756
	Sunday	7.011278195488723
	Tuesday	7.003164556962025
	Saturday	6.901829268292688
	Thursday	6.88985507246377
	Wednesday	6.805594405594405

Monday, Tuesday and Friday are the top best days for good ratings

Which day of the week has the best average ratings per branch

For branch –’A’

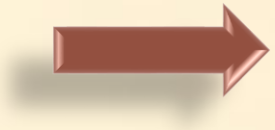
```sql
SELECT
    day_name, branch,
    AVG(rating) AS avg_rating
FROM
    sales
WHERE
    branch = 'A'
GROUP BY
    day_name, branch
ORDER BY
    avg_rating DESC;
```



	day_name	branch	avg_rating
▶	Friday	A	7.3119999999999985
	Monday	A	7.097916666666666
	Sunday	A	7.078846153846157
	Tuesday	A	7.0588235294117645
	Thursday	A	6.958695652173914
	Wednesday	A	6.916279069767441
	Saturday	A	6.745999999999998

Which day of the week has the best average ratings per branch

```sql
SELECT
    day_name, branch,
    AVG(rating) AS avg_rating
FROM
    sales
WHERE
    branch = 'B'
GROUP BY
    day_name, branch
ORDER BY
    avg_rating DESC;
```



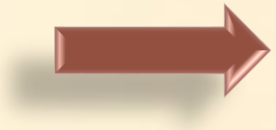
For branch –’B’

	day_name	branch	avg_rating
▶	Monday	B	7.335897435897434
	Tuesday	B	7.001886792452827
	Sunday	B	6.888571428571429
	Thursday	B	6.752272727272726
	Saturday	B	6.7366666666666655
	Friday	B	6.694117647058826
	Wednesday	B	6.451999999999999

Which day of the week has the best average ratings per branch

For branch –’C’

```sql
SELECT
    day_name, branch,
    AVG(rating) AS avg_rating
FROM
    sales
WHERE
    branch = 'C'
GROUP BY
    day_name, branch
ORDER BY
    avg_rating DESC;
```



	day_name	branch	avg_rating
▶	Monday	B	7.335897435897434
	Tuesday	B	7.001886792452827
	Sunday	B	6.888571428571429
	Thursday	B	6.752272727272726
	Saturday	B	6.7366666666666655
	Friday	B	6.694117647058826
	Wednesday	B	6.451999999999999

Thank You

Nidhi Dewangan