Walmart Sales Data Analysis

Introduction

This project aims to analyze the sales data of Walmart, one of the largest retail companies in the world, using SQL. The data contains information about sales transactions from three Walmart branches in different cities, including details like product lines, customer types, payment methods, and more. The project will use various SQL queries and data manipulation techniques to uncover valuable insights from this data and provide recommendations for enhancing and optimizing sales strategies.

Analysis Objectives

The main objectives of this project are:

- To perform an analysis on the data to gain insights into different product lines, determine the top-performing product lines, and identify areas for improvement in other product lines.
- To address the inquiry regarding the sales trends of the product. The outcomes of this analysis can assist in evaluating the efficiency of each applied sales strategy in the business and determining necessary modifications to increase sales.
- To identify various customer segments, understand purchasing trends, and evaluate the profitability associated with each of these customer segments.

Generic Questions

1.How many unique cities does the data have?

```
select distinct(city) as unique_cities
from sales;
```

	Unique_cities
▶	Yangon
	Naypyitaw
	Mandalay

2.In which city is each branch?

```
select distinct(city), branch
from sales;
```

	city	branch
▶	Yangon	A
	Naypyitaw	C
	Mandalay	B

Product Questions

1. How many unique product lines does the data have?

```
select count(distinct(product_line)) as
no_of_product_line
from sales;
```

	no_of_product_line
▶	6

2. What is the most common payment method?

```
select payment, count(payment) as cnt
from sales
group by 1
order by 2 desc;
```

	payment	cnt
▶	Cash	344
	Ewallet	342
	Credit card	309

3. What is the most selling product line?

```
select product_line, sum(quantity) as
no_of_item_sold
from sales
group by 1
order by 2 desc;
```

product_line	no_of_item_sold
Electronic accessories	961
Food and beverages	952
Home and lifestyle	911
Sports and travel	902
Fashion accessories	902
Health and beauty	844

4. What is the total revenue by month?

```
select month_name, sum(unit_price*quantity) as
total_revenue
from sales
group by 1
order by 2 desc;
```

month_name	total_revenue
January	110754.16
March	103683.00
February	91168.93

5. What month had the largest COGS?

```
select month_name, sum(cogs) as total_cogs
from sales
group by 1
order by 2 desc;
```

month_name	total_cogs
January	110759
March	103695
February	91177

6. What product line had the largest revenue?

```
select product_line, sum(unit_price*quantity) as
total_revenue
from sales
group by 1
order by 2 desc;
```

product_line	total_revenue
Food and beverages	53471.28
Fashion accessories	51719.90
Sports and travel	51367.74
Home and lifestyle	51297.06
Electronic accessories	51222.13
Health and beauty	46527.98

7. What is the city with the largest revenue?

```
select city, sum(unit_price*quantity) as total_revenue
from sales
group by 1
order by 2 desc;
```

city	total_revenue
Naypyitaw	105229.31
Yangon	100820.01
Mandalay	99556.77

8. What product line had the largest VAT?

```
select product_line, round(avg(5/100 * cogs),2) as VAT
from sales
group by 1
order by 2 desc;
```

product_line	VAT
Home and lifestyle	16.03
Sports and travel	15.76
Health and beauty	15.41
Food and beverages	15.37
Electronic accessories	15.16
Fashion accessories	14.53

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

```
with cte1 as
(
select product_line,round(avg((5/100 * cogs) + cogs),2) as avg_sales_product ,
(select round(avg((5/100 * cogs) + cogs),2) from sales) as average_sales
from sales
group by 1
order by 2 desc
)
select *,
case when avg_sales_product > average_sales then "Good" else "Bad" end as review
from cte1;
```

product_line	avg_sales_product	average_sales	review
Home and lifestyle	336.65	322.53	Good
Sports and travel	330.92	322.53	Good
Health and beauty	323.56	322.53	Good
Food and beverages	322.72	322.53	Good
Electronic accessories	318.26	322.53	Bad
Fashion accessories	305.13	322.53	Bad

10.Which branch sold more products than average product sold?

```
select branch, sum(quantity) as quantity_sold
from sales
group by 1
having sum(quantity)> (select avg(quantity) from sales)
order by 2 desc;
```

branch	quantity_sold
A	1849
C	1828
B	1795

11. What is the most common product line by gender?

```
select product_line, gender, count(gender) as total_cnt
from sales
group by 1,2
order by 3 desc;
```

product_line	gender	total_cnt
Fashion accessories	Female	96
Food and beverages	Female	90
Health and beauty	Male	88
Sports and travel	Female	86
Electronic accessories	Male	86
Food and beverages	Male	84
Electronic accessories	Female	83
Fashion accessories	Male	82
Home and lifestyle	Male	81
Home and lifestyle	Female	79
Sports and travel	Male	77
Health and beauty	Female	63

12. What is the average rating of each product line?

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

product_line	avg_rating
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	6.98
Electronic accessories	6.91
Sports and travel	6.86
Home and lifestyle	6.84

Sales Questions

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

```
select day_name,time_of_day , count(*) as total_sales
from sales
where day_name  not in ("saturday", "sunday")
group by day_name,time_of_day
order by 1,3 desc;
```

day_name	time_of_day	total_sales
Friday	Afternoon	58
Friday	Evening	51
Friday	Morning	29
Monday	Evening	56
Monday	Afternoon	48
Monday	Morning	20
Thursday	Evening	56
Thursday	Afternoon	49
Thursday	Morning	33
Tuesday	Evening	69
Tuesday	Afternoon	53
Tuesday	Morning	36
Wednesday	Afternoon	61
Wednesday	Evening	58
Wednesday	Morning	22

2.Which of the customer types brings the most revenue?

```
select customer_type, round(sum(total),2) as total_rev
from sales
group by customer_type
order by total_rev desc;
```

customer_type	total_rev
Member	163625.10
Normal	157261.29

3.Which city has the largest tax percent/ VAT (Value Added Tax)?

```
select city, round(avg(tax_pct),2) as avg_tax_pcnt
from sales
group by 1
order by 2 desc;
```

city	avg_tax_pcnt
Naypyitaw	16.09
Mandalay	15.13
Yangon	14.87

4.Which customer type pays the most in VAT?

```
select customer_type, round(avg(5/100 * cogs),2) as VAT
from sales
group by 1
order by 2 desc;
```

customer_type	VAT
Member	15.62
Normal	15.10

Customer Questions

1.How many unique customer types does the data have?

```
select distinct(customer_type) as customer_type,
(select count(distinct(customer_type) ) from sales) as
total_count
from sales
group by 1;
```

customer_type	total_count
Normal	2
Member	2

2.How many unique payment methods does the data have?

```
select distinct(payment) as customer_type,
(select count(distinct(payment) ) from sales) as
total_count
from sales
group by 1;
```

customer_type	total_count
Credit card	3
Ewallet	3
Cash	3

3.What is the most common customer type?

```
select customer_type, count(customer_type) as cnt
from sales
group by 1
order by 2 desc
limit 1;
```

	customer_type	cnt
▶	Member	499

4.Which customer type buys the most?

```
select customer_type, count(*) as cnt
from sales
group by 1
order by 2 desc
limit 1;
```

	customer_type	cnt
▶	Member	499

5.What is the gender of most of the customers?

```
select gender, count(*) as gender_cnt
from sales
group by 1
order by 2 desc;
```

gender	gender_cnt
Male	498
Female	497

6.What is the gender distribution per branch?

```
select branch,gender, count(*) as gender_cnt
from sales
group by 1,2
order by 1,3 desc;
```

branch	gender	gender_cnt
A	Male	179
A	Female	160
B	Male	169
B	Female	160
C	Female	177
C	Male	150

7.Which time of the day do customers give most ratings?

```
select time_of_day, count(rating) as total_rating_cnt
from sales
group by 1
order by 2 desc;
```

time_of_day	total_rating_cnt
Evening	429
Afternoon	376
Morning	190

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

```
select branch, time_of_day, count(rating) as
total_rating_cnt
from sales
group by 1,2
order by 1,3 desc;
```

	branch	time_of_day	total_rating_c
▶	A	Evening	140
	A	Afternoon	126
	A	Morning	73
	B	Evening	147
	B	Afternoon	124
	B	Morning	58
	C	Evening	142
	C	Afternoon	126
	C	Morning	59

9.Which day fo the week has the best avg ratings?

```
select day_name, round(avg(rating),2) as avg_rating
from sales
group by 1
order by 2 desc;
```

	day_name	avg_rating
▶	Monday	7.13
	Friday	7.06
	Tuesday	7
	Sunday	6.99
	Saturday	6.9
	Thursday	6.89
	Wednesday	6.76

10.Which day of the week has the best average ratings per branch?

```
select branch, day_name, avg_rating
from(
select branch, day_name, round(avg(rating),2) as
avg_rating,
rank() over(partition by branch order by  avg(rating) desc )
as rnk
from sales
group by 1,2
order by 1,3 desc) a
where rnk=1
```

branch	day_name	avg_rating
A	Friday	7.31
B	Monday	7.27
C	Saturday	7.23