Walmart Sales

SQL Project

Objective of the project

This project aims to explore the Walmart Sales data to understand top performing branches and products, sales trends of different products and customer behaviour. The aim is to study how sales strategies can be improved and optimized.

The major aim of this project is to gain insight into the sales data of Walmart to understand the different factors that affect the sales of different branches.

Handling Null-values in the dataset

There are no null values in the database as while creating the table, I have set NOT NULL for each field, hence the null values are filtered out.

```
CREATE TABLE sales (
    invoice_id VARCHAR(30) NOT NULL PRIMARY KEY,
    branch VARCHAR(5) NOT NULL,
    city VARCHAR(30) NOT NULL,
    customer_type VARCHAR(30) NOT NULL,
    gender VARCHAR(30) NOT NULL,
    product_line VARCHAR(100) NOT NULL,
    unit_price DECIMAL(10,2) NOT NULL,
    quantity INT NOT NULL,
    tax_pct FLOAT(6) NOT NULL,
    total DECIMAL(12, 4) NOT NULL,
    date TIMESTAMP NOT NULL,
    time TIME NOT NULL,
    payment VARCHAR(15) NOT NULL,
    cogs DECIMAL(10,2) NOT NULL,
    gross_margin_pct FLOAT(9),
    gross_income DECIMAL(12, 4),
    rating FLOAT(2));
```

Feature Engineering

Adding a new column named time_of_day to give insight of sales in Morning, Afternoon and Evening

```sql
-- Step 1
SELECT TIME, CASE
                    WHEN TIME BETWEEN '00:00:00' AND '12:00:00' THEN 'Morning'
                    WHEN TIME BETWEEN '12:01:00' AND '16:00:00' THEN 'Afternoon'
                    ELSE 'Evening'
             END AS TIME_OF_DAY
FROM SALES;


-- Step 2
ALTER TABLE SALES ADD COLUMN TIME_OF_DAY VARCHAR(20);


-- Step 3
UPDATE SALES
SET TIME_OF_DAY = (CASE
                        WHEN TIME BETWEEN '00:00:00' AND '12:00:00' THEN 'Morning'
                        WHEN TIME BETWEEN '12:01:00' AND '16:00:00' THEN 'Afternoon'
                        ELSE 'Evening'
                   END);
```

Feature Engineering

Adding a new column named day_name that contains the extracted days of the week on which the given transaction took place (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday)

```sql
-- Step 1
SELECT date, TO_CHAR(date, 'Day') AS DAY_NAME
FROM SALES;


-- Step 2
ALTER TABLE SALES ADD COLUMN DAY_NAME VARCHAR(10);


-- Step 3
UPDATE SALES
SET DAY_NAME = TO_CHAR(date, 'Day');
```

Feature Engineering

Adding a new column named month_name that contains the extracted months of the year on which the given transaction took place (January, February, March....)

```sql
-- Step 1
SELECT date, TO_CHAR(date, 'Month') AS MONTH_NAME
FROM SALES;


-- Step 2
ALTER TABLE SALES ADD COLUMN MONTH_NAME VARCHAR(10);


-- Step 3
UPDATE SALES
SET MONTH_NAME = TO_CHAR(date, 'Month');
```

Product Related Questions

Analysis on the data to understand the different product lines, best performing products lines and the product lines that need to be improved.

How many unique product lines does the data have?

```
SELECT DISTINCT PRODUCT_LINE
FROM SALES;
```

product_line character varying (100)
Fashion accessories
Health and beauty
Electronic accessories
Food and beverages
Sports and travel
Home and lifestyle

What is the most common payment method?

```
SELECT PAYMENT, COUNT(PAYMENT) AS CNT
FROM SALES
GROUP BY PAYMENT
ORDER BY CNT DESC;
```

payment character varying (15)	cnt bigint
Ewallet	345
Cash	344
Credit card	311

What is the most selling product line?

```sql
SELECT PRODUCT_LINE, COUNT(PRODUCT_LINE) AS CNT
FROM SALES
GROUP BY PRODUCT_LINE
ORDER BY CNT DESC;
```

product_line character varying (100)	cnt bigint
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?

```sql
SELECT MONTH_NAME AS MONTH, SUM(TOTAL) AS TOTAL_REVENUE
FROM SALES
GROUP BY MONTH_NAME
ORDER BY TOTAL_REVENUE DESC;
```

month character varying (10)	total_revenue numeric
January	116291.8680
March	109455.5070
February	97219.3740

What month had the highest COGS?

```sql
SELECT MONTH_NAME AS MONTH, SUM(COGS) AS COGS
FROM SALES
GROUP BY MONTH_NAME
ORDER BY COGS DESC
LIMIT 1;
```

month character varying (10)	cogs numeric
January	110754.16

What product line had the highest revenue?

```sql
SELECT PRODUCT_LINE, SUM(TOTAL) AS REVENUE
FROM SALES
GROUP BY PRODUCT_LINE
ORDER BY REVENUE DESC
LIMIT 1;
```

product_line character varying (100)	revenue numeric
Food and beverages	56144.8440

What is the city with the highest revenue?

```sql
SELECT CITY, SUM(TOTAL) AS TOTAL_REVENUE
FROM SALES
GROUP BY CITY
ORDER BY TOTAL_REVENUE DESC
LIMIT 1;
```

city character varying (30)	total_revenue numeric
Naypyitaw	110568.7065

What product line had the highest VAT?

```sql
SELECT PRODUCT_LINE, sum(TAX_PCT) AS VAT
FROM SALES
GROUP BY PRODUCT_LINE
ORDER BY VAT DESC
LIMIT 1;
```

product_line character varying (100)	vat real
Food and beverages	2673.5632

Which branch sold more products than average product sold?

```sql
SELECT BRANCH, SUM(QUANTITY) AS QUANTITY
FROM SALES
GROUP BY BRANCH
HAVING SUM(QUANTITY) > (SELECT AVG(QUANTITY) FROM SALES);
```

branch character varying (5)	quantity bigint
A	1859
C	1831
B	1820

What is the most common product line by gender?

```
SELECT PRODUCT_LINE, COUNT(GENDER) AS GENDER
FROM SALES
GROUP BY PRODUCT_LINE
ORDER BY GENDER DESC;
```

product_line character varying (100)	gender bigint
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 average rating of each product line?

```
SELECT PRODUCT_LINE, AVG(RATING) AS AVG_RATING
FROM SALES
GROUP BY PRODUCT_LINE
ORDER BY AVG_RATING DESC;
```

product_line character varying (100)	avg_rating double precision
Food and beverages	7.11321838970842
Fashion accessories	7.0292134660013605
Health and beauty	7.003289457998778
Electronic accessories	6.924705881230971
Sports and travel	6.916265062538974
Home and lifestyle	6.8375

Sales Related Questions

This analysis aims to answer the question of the sales trends of product. The result of this can help to measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

Number of sales made in each time of the day

```
SELECT TIME_OF_DAY, COUNT(*) AS NUMBER_OF_SALES
FROM SALES
GROUP BY TIME_OF_DAY
ORDER BY NUMBER_OF_SALES DESC;
```

time_of_day character varying (20)	number_of_sales bigint
Evening	432
Afternoon	377
Morning	191

Which of the customer type brings the most revenue?

```sql
SELECT CUSTOMER_TYPE, SUM(TOTAL) AS REVENUE
FROM SALES
GROUP BY CUSTOMER_TYPE
ORDER BY REVENUE DESC
LIMIT 1;
```

customer_type character varying (30)	revenue numeric
Member	164223.4440

Which city has the highest tax percent/ VAT?

```sql
SELECT CITY, SUM(TAX_PCT) as TAX
FROM SALES
GROUP BY CITY
ORDER BY TAX DESC
LIMIT 1;
```

city character varying (30)	tax real
Naypyitaw	5265.175

Which customer type pays the most in VAT?

```sql
SELECT CUSTOMER_TYPE, SUM(TAX_PCT) AS TAX
FROM SALES
GROUP BY CUSTOMER_TYPE
ORDER BY TAX DESC
LIMIT 1;
```

customer_type character varying (30)	tax real
Member	7820.1597

Customer Related Questions

This analysis aims to uncover the different customer segments, purchase trends and the profitability of each customer segment.

Which customer type buys the most?

```sql
SELECT CUSTOMER_TYPE, COUNT(*)
FROM SALES
GROUP BY CUSTOMER_TYPE
ORDER BY COUNT(*) DESC;
```

customer_type character varying (30)	count bigint
Member	501
Normal	499

What is the gender distribution per branch?

```sql
SELECT GENDER, COUNT(*)
FROM SALES
WHERE BRANCH = 'B'
GROUP BY GENDER
ORDER BY COUNT(*) DESC;
```

gender character varying (30)	count bigint
Male	170
Female	162

```sql
SELECT GENDER, COUNT(*)
FROM SALES
WHERE BRANCH = 'A'
GROUP BY GENDER
ORDER BY COUNT(*) DESC;
```

gender character varying (30)	count bigint
Male	179
Female	161

```sql
SELECT GENDER, COUNT(*)
FROM SALES
WHERE BRANCH = 'C'
GROUP BY GENDER
ORDER BY COUNT(*) DESC;
```

gender character varying (30)	count bigint
Female	178
Male	150

Which time of the day do customers give most ratings?

```sql
SELECT TIME_OF_DAY, COUNT(RATING) AS RATINGS
FROM SALES
GROUP BY TIME_OF_DAY
ORDER BY RATINGS DESC;
```

time_of_day character varying (20)	ratings bigint
Evening	432
Afternoon	377
Morning	191

Which time of the day do customers give most ratings?

```sql
SELECT DAY_NAME, AVG(RATING) AS AVERAGE_RATINGS
FROM SALES
WHERE BRANCH = 'B'
GROUP BY DAY_NAME
ORDER BY AVERAGE_RATINGS DESC
LIMIT 1;
```

day_name character varying (10)	average_ratings double precision
Monday	7.335897384545742

```sql
SELECT DAY_NAME, AVG(RATING) AS AVERAGE_RATINGS
FROM SALES
WHERE BRANCH = 'A'
GROUP BY DAY_NAME
ORDER BY AVERAGE_RATINGS DESC
LIMIT 1;
```

day_name character varying (10)	average_ratings double precision
Friday	7.311999988555908

```sql
SELECT DAY_NAME, AVG(RATING) AS AVERAGE_RATINGS
FROM SALES
WHERE BRANCH = 'C'
GROUP BY DAY_NAME
ORDER BY AVERAGE_RATINGS DESC
LIMIT 1;
```

day_name character varying (10)	average_ratings double precision
Friday	7.278947328266344

Thank You!