WALMART SALES DATA ANALYSIS WITH SQL

PRESENTED BY : POOJA SUHAG

ABOUT

This project aims to explore Walmart sales data to identify top-performing branches and products, analyze sales trends, and understand customer behavior. By studying these aspects, we seek to uncover insights that can help improve and optimize sales strategies.

PURPOSE OF THE PROJECT

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

ANALYSIS LIST

Product Analysis

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

Sales Analysis

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

Customer Analysis

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

METHODOLOGY USED

Data Wrangling

The initial step involves inspecting the data to identify and address any NULL or missing values. Data replacement methods are employed to ensure completeness. The process includes:

1.Building a Database : Creating tables and inserting the data.
2. Handling NULL Values : During table creation, each field is set to NOT NULL, ensuring that null values are filtered out. This guarantees that the database is free of null values.

Feature Engineering

This step involves generating new columns from the existing data to derive additional insights:

1. Adding Time of Day Column : A new column named `time_of_day` is created to categorize sales into Morning, Afternoon, and Evening. This helps analyze which part of the day experiences the highest sales.
2. Adding Day Name Column : A new column named `day_name` is added to capture the day of the week (Mon, Tue, Wed, Thur, Fri) for each transaction. This assists in determining which day of the week each branch is busiest.
3. Adding Month Name Column : A new column named `month_name` is included to denote the month of the year (Jan, Feb, Mar) for each transaction. This helps identify which month records the highest sales and profit.

Exploratory Data Analysis (EDA): Exploratory data analysis is done to answer the listed questions and aims of this project.

QUESTIONS

Generic Question

1. How many unique cities does the data have?
2. In which city is each branch?

Product

1. How many unique product lines does the data have?
2. What is the most common payment method?
3. What is the most selling product line?
4. What is the total revenue by month?
5. What month had the largest COGS?
6. What product line had the largest revenue?
7. What is the city with the largest revenue?
8. What product line had the largest VAT?
9. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales
10. Which branch sold more products than average product sold?
11. What is the most common product line by gender?
12. What is the average rating of each product line?

QUESTIONS

Sales

1. Number of sales made in each time of the day per weekday
2. Which of the customer types brings the most revenue?
3. Which city has the largest tax percent/ VAT (Value Added Tax)?
4. Which customer type pays the most in VAT?

Customer

1. How many unique customer types does the data have?
2. How many unique payment methods does the data have?
3. What is the most common customer type?
4. Which customer type buys the most?
5. What is the gender of most of the customers?
6. What is the gender distribution per branch?
7. Which time of the day do customers give most ratings?
8. Which time of the day do customers give most ratings per branch?
9. Which day fo the week has the best avg ratings?
10. Which day of the week has the best average ratings per branch?

GENERIC

Query 1. How many unique cities does the data have?

```
SELECT DISTINCT
    city
FROM
    sales;
```

city
Yangon
Naypyitaw
Mandalay

Insight : Yangon, Naypyitaw and Mandalay are the 3 unique cities..

Query 2. How many unique branch are there?

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

branch
A
C
B

Insight : There are 3 unique branches (A, B, C)..

PRODUCT

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

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

	COUNT(DISTINCT product_line)
▶	6

Insight : There are 6 unique product lines..

Query 2. What is the most common payment method?

```
SELECT
    payment_method, COUNT(payment_method) AS cnt
FROM
    sales
GROUP BY payment_method
ORDER BY cnt DESC;
```

	payment_method	cnt
▶	Cash	344
	Ewallet	342
	Credit card	309

Insight : Cash is the most common Payment method ..

Query 3. 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	cnt
Fashion accessories	178
Food and beverages	174
Electronic accessories	169
Sports and travel	163
Home and lifestyle	160
Health and beauty	151

Insight : Fashion accessories is the most common product line..

Query 4.What is the total revenue by month?

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

month	cogs
January	110759
March	103695
February	91177

Insight : January month has the highest revenue of 110759Rs. followed by March 103695Rs. and Feburary 91177Rs ..

Query 5. What product line had the largest revenue?

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

product_line	total_revenue
Food and beverages	56144.8440
Fashion accessories	54305.8950
Sports and travel	53936.1270
Home and lifestyle	53861.9130
Electronic accessories	53783.2365
Health and beauty	48854.3790

Insight : Food and beverages has the largest revenue of @56,144,8440..

Query 6. What is the city with the largest revenue?

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

branch	city	total_revenue
C	Naypyitaw	110490.7755
A	Yangon	105861.0105
B	Mandalay	104534.6085

Insight : Naypyitaw is the city with highest revenue of @110490.7755..

Query 7.What product line has the largest vat?

```
SELECT
    product_line, AVG(VAT) AS avg_tax
FROM
    sales
GROUP BY product_line
ORDER BY avg_tax DESC;
```

product_line	avg_tax
Home and lifestyle	16.03033124
Sports and travel	15.75697549
Health and beauty	15.40661591
Food and beverages	15.36531029
Electronic accessories	15.15447632
Fashion accessories	14.52806181

Insight : Home and Lifestyle has the largest VAT..

Query 8. Which branch sold more products than average product sold?

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

branch	qty
A	1849
C	1828
B	1795

Insight : Branch A sold most Products ..

Query 9. What is the most common product line by gender?

```
SELECT
    gender, product_line, COUNT(gender) AS total_cnt
FROM
    sales
GROUP BY gender , product_line
ORDER BY total_cnt DESC;
```

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

Insight : Female(Fashion Accessories) ..

Query 10.What is the average rating of each product line?

```
SELECT
    ROUND(AVG(rating), 2) AS avg_rating, product_line
FROM
    sales
GROUP BY product_line
ORDER BY avg_rating DESC;
```

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

Insight : Avg Rating of top product_line is 7.11 ..

Query 11.What is the city with the largest revenue?

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

branch	city	total_revenue
C	Naypyitaw	110490.7755
A	Yangon	105861.0105
B	Mandalay	104534.6085

Insight : Naypyitaw is the city with largest revenue followed by Yangon and Mandalay..

SALES

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

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

time_of_day	total_sales
Evening	132

Insight : Evening sales are 132 ..

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

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

customer_type	total_revenue
Member	163625.1015
Normal	157261.2930

Insight : Member brings the most revenue of @ 163625.10 ..

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

```
SELECT
    city, AVG(VAT) AS VAT
FROM
    sales
GROUP BY city
ORDER BY VAT DESC;
```

city	VAT
Naypyitaw	16.09010850
Mandalay	15.13020824
Yangon	14.87020798

Insight : Naypyitaw is the city with largest tax ..

Query 4.Which customer type pays the most in VAT?

```
SELECT
    customer_type, AVG(VAT) AS VAT
FROM
    sales
GROUP BY customer_type
ORDER BY VAT DESC;
```

Result Grid | Filter Rows:

customer_type	VAT
Member	15.61457214
Normal	15.09805040

Insight : Member pays the most VAT..

CUSTOMER

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

```
SELECT DISTINCT
    customer_type
FROM
    sales;
```

customer_type
Normal
Member

Insight : Normal and Member are the two unique customer types..

Query 2.What is the most common customer type?

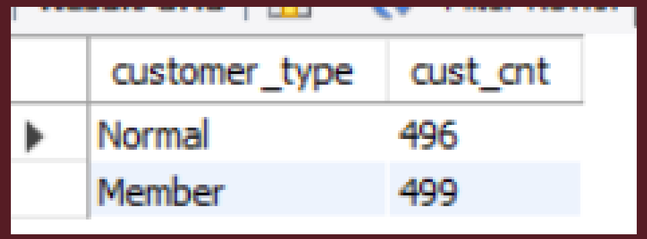
```
SELECT
    customer_type, COUNT(*) AS count
FROM
    sales
GROUP BY customer_type
ORDER BY count DESC;
```

customer_type	count
Member	499
Normal	496

Insight : Member is the most common customer type with a count of 499 ..

Query 3.Which customer type buys the most?

```
SELECT
    customer_type, COUNT(*) AS cust_cnt
FROM
    sales
GROUP BY customer_type;
```

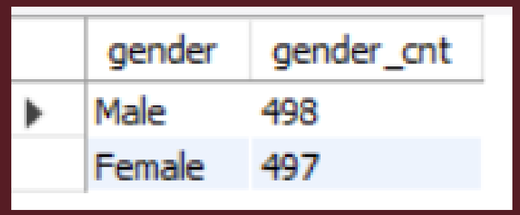


customer_type	cust_cnt
Normal	496
Member	499

Insight : Normal buys the most ..

Query 4.What is the gender of most of the customers?

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



gender	gender_cnt
Male	498
Female	497

Insight : Male ..

Query 5.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;
```

gender	gender_cnt
Male	179
Female	160

Insight : Male -179 and Female- 160 ..

Query 6.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
Evening	6.95749

Insight : Evening ..

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

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

time_of_day	avg_rating
Evening	7.01829

Insight : Evening is the time customers give most ratings ..

Query 8.Which day fo 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	7.13065
Friday	7.05507
Tuesday	7.00316
Sunday	6.98864
Saturday	6.90183
Thursday	6.88986
Wednesday	6.76028

Insight : Monday is the day with best avg_ratings around 7.19065 ..

Query 9.Which day of the week has the best average ratings per branch?

```sql
SELECT
    day_name, COUNT(day_name) total_sales
FROM
    sales
WHERE
    branch = 'C'
GROUP BY day_name
ORDER BY total_sales DESC;
```

day_name	total_sales
Tuesday	54
Saturday	54
Wednesday	50
Thursday	48
Sunday	46
Monday	38
Friday	37

Insight : Tuesday is the day with best ratings ..

KEY INSIGHTS

- Most common payment_method : Cash @344
- Highest selling product_line : Fashion Accessories @ 178
- Month with highest revenue : January @110759
- Product_line with highest revenue : Food and beverages @56,144,8440
- City with highest revenue : Naypyitaw @ 110490.7755
- Avg_Rating of top product_line : 7.11 (Food and beverages)
- Time_of_day with most sales : Evening

THANK YOU