

WALMART



SALES REPORT



By: Modita

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

Analysis

- **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 modifications 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.

Approach Used

- Data Wrangling

This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace, missing or NULL values.

- Feature Engineering

This will help use generate some new columns from existing ones.

- Exploratory Data Analysis (EDA)

This is done to answer the listed questions and aims of this project.

Example of Feature Engineering

```
-- Add the time_of_day column
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;

ALTER TABLE sales ADD COLUMN time_of_day VARCHAR(20);

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
);

select invoice_id,date,time, time_of_day from sales;
```

invoice_id	date	time	time_of_day
101-17-6199	2019-03-13 00:00:00	19:44:00	Evening
101-81-4070	2019-01-17 00:00:00	12:36:00	Afternoon
102-06-2002	2019-03-20 00:00:00	17:52:00	Evening
102-77-2261	2019-03-05 00:00:00	18:02:00	Evening
105-10-6182	2019-02-27 00:00:00	12:22:00	Afternoon
105-31-1824	2019-02-01 00:00:00	15:10:00	Afternoon
106-35-6779	2019-03-27 00:00:00	11:26:00	Morning
109-28-2512	2019-01-07 00:00:00	15:01:00	Afternoon
109-86-4363	2019-02-14 00:00:00	11:36:00	Morning
110-05-6330	2019-03-25 00:00:00	20:18:00	Evening
110-48-7033	2019-01-29 00:00:00	14:12:00	Afternoon
114-35-5271	2019-02-07 00:00:00	15:06:00	Afternoon
115-38-7388	2019-03-30 00:00:00	12:51:00	Afternoon

Q1 : How many unique cities does the data have?

```
SELECT  
    COUNT(DISTINCT city) AS no_of_unique_cities  
FROM  
    sales;
```

no_of_unique_cities
3

Q2 : In which city is each branch?

```
SELECT DISTINCT  
    branch, city  
FROM  
    sales;
```

branch	city
A	Yangon
C	Naypyitaw
B	Mandalay

Q3: What month had the largest COGS?

```
SELECT  
    month_name, SUM(cogs) AS cogs_by_month  
FROM  
    sales  
GROUP BY month_name  
ORDER BY cogs_by_month DESC  
LIMIT 1;
```

month_name	cogs_by_month
January	110754.16

Q4 : Which of the customer types brings the most revenue?

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

customer_type	revenue
Member	163625.1015
Normal	157261.2930

⇒ Customers who are 'Members' brings the most revenue

Q5 :What is the most common payment method?

```
SELECT  
    payment_method, COUNT(payment_method) AS number_  
FROM  
    sales  
GROUP BY payment_method  
ORDER BY number_ DESC  
LIMIT 1;
```

	payment_method	number_
	Cash	344

Q7: Which branch sold more products than average product sold?

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

branch	total_quantity
A	1849
C	1828
B	1795

Q8: Number of sales made in each time of the day per weekday

```
SELECT  
    day_name, time_of_day, COUNT(*) AS number  
FROM  
    sales  
GROUP BY day_name , time_of_day  
ORDER BY day_name , time_of_day;
```

day_name	time_of_day	number
Friday	Afternoon	58
Friday	Evening	51
Friday	Morning	29
Monday	Afternoon	48
Monday	Evening	56
Monday	Morning	20
Saturday	Afternoon	55
Saturday	Evening	81
Saturday	Morning	28
Sunday	Afternoon	52
Sunday	Evening	58
Sunday	Morning	22
Thursday	Afternoon	49
Thursday	Evening	56
Thursday	Morning	33
Tuesday	Afternoon	53
Tuesday	Evening	69
Tuesday	Morning	36
Wednesday	Afternoon	61
Wednesday	Evening	58
Wednesday	Morning	22

Q9: What is the gender distribution per branch?

```
SELECT  
    branch, gender, COUNT(gender) AS number  
FROM  
    sales  
GROUP BY branch , gender  
ORDER BY branch;
```

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

Q10: Which time of the day do customers give most ratings?

```
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	7.02340
Morning	6.94474
Evening	6.90536

⇒ Customers give the most ratings in Afternoon and the least in Evening

Q11: Which time of the day do customers give most ratings per branch?

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

branch	time_of_day	avg_rating
A	Afternoon	7.18889
A	Morning	7.00548
A	Evening	6.87143
B	Morning	6.83793
B	Afternoon	6.81129
B	Evening	6.75102
C	Evening	7.09859
C	Afternoon	7.06667
C	Morning	6.97458

Q12: Which day of the week has the best average ratings?

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

day_name	rating
Monday	7.13065
Friday	7.05507
Tuesday	7.00316
Sunday	6.98864
Saturday	6.90183
Thursday	6.88986
Wednesday	6.76028

⇒ Monday has the best average ratings and Wednesday the least

Some more Insights

- Cash payments are predominant despite the rise of digital payment methods, emphasizing the need to cater to diverse payment preferences.
- Fashion accessories emerge as the top-selling product line, indicating strong consumer interest and demand in this category.
- January records the highest sales volume, while February experiences a decline, suggesting potential seasonal factors or opportunities for targeted promotions.
- Food and beverages contribute the most to revenue, highlighting the importance of enhancing offerings within this category to drive sales and improve customer satisfaction.

Thank You!



moditagarg.99@yahoo.com

