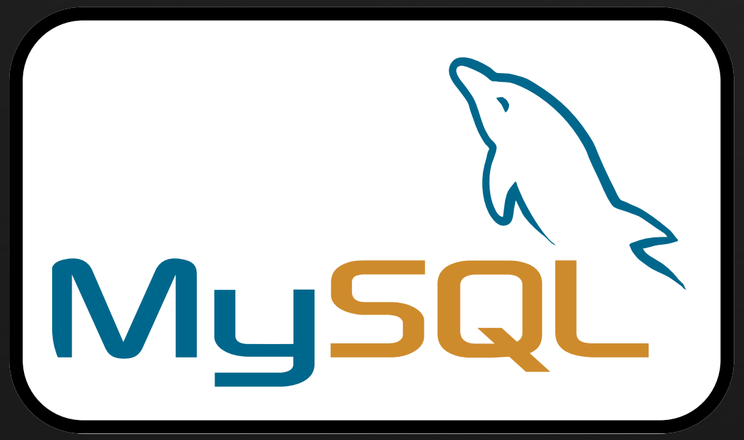
SQL

DATA ANALYSIS PROJECT



– Ardhendu Bhusan Panda| Data Analyst

OBJECTIVE

The objective of this project is to explore the Walmart Sales data to understand top performing products, sales trend of different products and customer behavior.

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

DATASET

This dataset contains sales transactions from three different branches of Walmart, located in Mandalay, Yangon and Naypyitaw. The data contains 15 columns and 1000 rows

Column	Description	Data Type
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(20)
product_line	Product line of the product sold	VARCHAR(100)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The quantity of the product sold	INT
sales	sales of the products sold	DECIMAL(10, 2)
tax_pct	The amount of tax on the purchase	FLOAT(10, 4)
total	The total cost of the purchase	DECIMAL(12, 4)
date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIMESTAMP
payment_method	The mode of payment	VARCHAR(15)
rating	Rating	DECIMAL(2, 1)

ANALYSIS

PRODUCTS

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

SALES

To understand sales trends of product to measure the effectiveness of sales strategy the business applies and modifications needed to gain more sales.

CUSTOMERS

To understand the different customers segments, purchase trends and the profitability of each customer segment.

```sql
-- day_name
ALTER TABLE sales ADD COLUMN day_name VARCHAR(10);

UPDATE sales
SET day_name = dayname(date);



-- month_name
ALTER TABLE sales ADD COLUMN month_name
VARCHAR(15);
UPDATE sales
SET month_name = monthname(date);
```

```sql
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
);
```

```sql
-- How many unique cities does the data have?
SELECT
    DISTINCT city
FROM sales;
```

	city
▶	Yangon
	Naypyitaw
	Mandalay

```sql
-- In which city is each branch?

SELECT
    DISTINCT city,
    branch
FROM sales;
```

city	branch
Yangon	A
Naypyitaw	C
Mandalay	B

```sql
-- How many unique product lines does the data have?
SELECT
    COUNT(DISTINCT product_line)
FROM sales;
```

COUNT(DISTINCT product_line)
6

```sql
-- What is the most common payment method?
SELECT
    payment_method,
    COUNT(payment_method) AS count
FROM sales
GROUP BY payment_method
ORDER BY count DESC;
```

payment_method	count
Cash	344
Ewallet	342
Credit card	309

```sql
-- What is the most selling product line?
SELECT
    product_line,
    COUNT(product_line) AS count
FROM sales
GROUP BY product_line
ORDER BY count DESC;
```

product_line	count
Fashion accessories	178
Food and beverages	174
Electronic accessories	169
Sports and travel	163
Home and lifestyle	160
Health and beauty	151

```sql
-- What is the total revenue by month?
SELECT
    month_name,
    SUM(total) AS total_revenue
FROM sales
GROUP BY month_name
ORDER BY total_revenue DESC;
```

month_name	total_revenue
January	116291.8680
March	108867.1500
February	95727.3765

```sql
-- What month had the largest sales?
SELECT
    month_name,
    SUM(sales) AS sales
FROM sales
GROUP BY month_name
ORDER BY sales DESC;
```

month_name	sales
January	110754.16
March	103683.00
February	91168.93

```sql
-- Which 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

```sql
-- 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

```sql
-- What product line had the largest tax?
SELECT
    product_line,
    AVG(tax_pct) 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

```sql
-- Fetch each product line and add a column to those product line showing "Good", "Bad".
-- Good if its greater than average sales
SELECT
    product_line,
    CASE
        WHEN AVG(quantity) > (SELECT AVG(quantity) FROM sales) THEN "Good"
        ELSE "Bad"
    END AS remark
FROM sales
GROUP BY product_line;
```

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

```sql
-- 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

```sql
-- 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	20

```sql
-- What is the most common product line by gender?
SELECT
    gender,
    product_line,
    COUNT(gender) as cnt
FROM sales
GROUP BY gender, product_line
ORDER BY cnt DESC;
```

gender	product_line	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
Male	Sports and travel	77
Female	Health and beauty	63

```sql
-- 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

```sql
-- Which of the customer types brings the most revenue?
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

```sql
-- Which customer type pays the most tax?
SELECT
    customer_type,
    ROUND(AVG(tax_pct), 2) AS avg_tax
FROM sales
GROUP BY customer_type
ORDER BY avg_tax DESC;
```

customer_type	avg_tax
Member	15.61
Normal	15.1

```sql
-- How many unique customer types does the data have?
SELECT
    COUNT(DISTINCT customer_type)
FROM sales;
```

COUNT(DISTINCT customer_type)
2

```sql
-- How many unique payment methods does the data have?
SELECT
    COUNT(DISTINCT payment_method)
FROM sales;
```

COUNT(DISTINCT payment_method)
3

```sql
-- What is the most common customer type?
SELECT
    customer_type,
    COUNT(customer_type) AS count
FROM sales
GROUP BY customer_type
ORDER BY count DESC;
```

```sql
-- Which customer type buys the most?
SELECT
    customer_type,
    COUNT(*) AS count
FROM sales
GROUP BY customer_type
ORDER BY count DESC;
```

customer_type	count
Member	499
Normal	496

```sql
-- What is the gender of most of the customers?
SELECT
    gender,
    COUNT(gender) as cnt
FROM sales
GROUP BY gender
ORDER BY cnt DESC;
```

gender	cnt
Male	498
Female	497

```sql
-- What is the gender distribution per branch?
SELECT
    gender,
    COUNT(gender) as cnt
FROM sales
WHERE branch = "C"
GROUP BY gender
ORDER BY cnt DESC;
```

gender	cnt
Female	177
Male	150

```sql
-- 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

```sql
-- Which time of the day do customers give most ratings per branch?
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
Afternoon	7.18889
Morning	7.00548
Evening	6.87143

```sql
-- Which day of the week has the best avg ratings?
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

```sql
-- Which day of the week has the best average ratings per branch?
SELECT
    day_name,
    AVG(rating) AS avg_rating
FROM sales
WHERE branch = "C"
GROUP BY day_name
ORDER BY avg_rating DESC;
```

day_name	avg_rating
Saturday	7.22963
Friday	7.20541
Wednesday	7.06400
Monday	7.03684
Sunday	7.02826
Tuesday	6.95185
Thursday	6.95000

THANK YOU