SQL Project

GitHub: Link to Project

Walmart Sales Data Analysis

Created by

Ankush Verma

Thanks to for Reference:

Code With Prince

Index

Link to Different Pages in the Projects

- Data Exploration
- Creating a Database and Table
- Feature Engineering & Add Columns
- Business Questions
- Thank You

GitHub: Link to Project

Data Exploration

Open up Walmart Sale Data, explore different columns, and check for inconsistent data, if found then clean the data file with no Null Value.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Invoice ID	Branch	City	Customer	Gender	Product li	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross mar	gross inco	Rating
2	750-67-84	A	Yangon	Member	Female	Health an	74.69	7	26.1415	548.9715	1/5/2019	13:08:00	Ewallet	522.83	4.761905	26.1415	9.1
3	226-31-30	C	Naypyitav	Normal	Female	Electronic	15.28	5	3.82	80.22	3/8/2019	10:29:00	Cash	76.4	4.761905	3.82	9.6
4	631-41-31	A	Yangon	Normal	Male	Home and	46.33	7	16.2155	340.5255	3/3/2019	13:23:00	Credit car	324.31	4.761905	16.2155	7.4
5	123-19-11	A	Yangon	Member	Male	Health an	58.22	8	23.288	489.048	1/27/2019	20:33:00	Ewallet	465.76	4.761905	23.288	8.4
6	373-73-79	A	Yangon	Normal	Male	Sports and	86.31	7	30.2085	634.3785	2/8/2019	10:37:00	Ewallet	604.17	4.761905	30.2085	5.3
7	699-14-30	C	Naypyitav	Normal	Male	Electronic	85.39	7	29.8865	627.6165	3/25/2019	18:30:00	Ewallet	597.73	4.761905	29.8865	4.1
8	355-53-59	A	Yangon	Member	Female	Electronic	68.84	6	20.652	433.692	2/25/2019	14:36:00	Ewallet	413.04	4.761905	20.652	5.8
9	315-22-56	C	Naypyitav	Normal	Female	Home and	73.56	10	36.78	772.38	2/24/2019	11:38:00	Ewallet	735.6	4.761905	36.78	8
10	665-32-91	A	Yangon	Member	Female	Health an	36.26	2	3.626	76.146	1/10/2019	17:15:00	Credit car	72.52	4.761905	3.626	7.2
11	692-92-55	B	Mandalay	Member	Female	Food and	54.84	3	8.226	172.746	2/20/2019	13:27:00	Credit car	164.52	4.761905	8.226	5.9
12	351-62-08	B	Mandalay	Member	Female	Fashion ac	14.48	4	2.896	60.816	2/6/2019	18:07:00	Ewallet	57.92	4.761905	2.896	4.5
13	529-56-39	B	Mandalay	Member	Male	Electronic	25.51	4	5.102	107.142	3/9/2019	17:03:00	Cash	102.04	4.761905	5.102	6.8
14	365-64-05	A	Yangon	Normal	Female	Electronic	46.95	5	11.7375	246.4875	2/12/2019	10:25:00	Ewallet	234.75	4.761905	11.7375	7.1
15	252-56-26	A	Yangon	Normal	Male	Food and	43.19	10	21.595	453.495	2/7/2019	16:48:00	Ewallet	431.9	4.761905	21.595	8.2
16	829-34-39	A	Yangon	Normal	Female	Health an	71.38	10	35.69	749.49	3/29/2019	19:21:00	Cash	713.8	4.761905	35.69	5.7
17	299-46-18	B	Mandalay	Member	Female	Sports and	93.72	6	28.116	590.436	1/15/2019	16:19:00	Cash	562.32	4.761905	28.116	4.5
18	656-95-93	A	Yangon	Member	Female	Health an	68.93	7	24.1255	506.6355	3/11/2019	11:03:00	Credit car	482.51	4.761905	24.1255	4.6
19	765-26-69	A	Yangon	Normal	Male	Sports and	72.61	6	21.783	457.443	1/1/2019	10:39:00	Credit car	435.66	4.761905	21.783	6.9
20	329-62-15	A	Yangon	Normal	Male	Food and	54.67	3	8.2005	172.2105	1/21/2019	18:00:00	Credit car	164.01	4.761905	8.2005	8.6
21	319-50-33	B	Mandalay	Normal	Female	Home and	40.3	2	4.03	84.63	3/11/2019	15:30:00	Ewallet	80.6	4.761905	4.03	4.4
22	300-71-46	C	Naypyitav	Member	Male	Electronic	86.04	5	21.51	451.71	2/25/2019	11:24:00	Ewallet	430.2	4.761905	21.51	4.8

Creating Database & Table

Open up MySQL Database and create a new connection, after creating the Database and Data Table as per CSV data columns with datatype and constraints

```sql
CREATE DATABASE walmartsaledata;

USE walmartsaledata;

CREATE TABLE sale (
    invoice_id VARCHAR(30) PRIMARY KEY,
    branch VARCHAR(5) NOT NULL,
    city VARCHAR(50) 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_prct FLOAT(6, 4) NOT NULL,
    TOTAL DECIMAL(10, 4) NOT NULL,
    date DATETIME NOT NULL,
    time TIME NOT NULL,
    payment VARCHAR(50) NOT NULL,
    cogs DECIMAL(10, 2) NOT NULL,
    gross_margin_prct FLOAT(11, 9),
    gross_income DECIMAL(12, 9),
    rating FLOAT(2, 1)
);
```

Feature Engineering

- Introduce a "time_of_day" column to provide insights into sales during Morning, Afternoon, and Evening periods. This addition will aid in determining the peak sales time of the day.

- Incorporate a "day_name" column that captures the days of the week for each transaction (Mon, Tue, Wed, Thur, Fri). This inclusion facilitates an analysis of each branch's busiest days, answering questions about weekly patterns.

- Integrate a "month_name" column, extracting the months of the year for each transaction (Jan, Feb, Mar). This modification assists in identifying the months with the highest sales and profit, contributing to a comprehensive analysis of yearly trends.

Adding Column

-- Add the time_of_day column


```sql
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 sale;

ALTER TABLE sale
ADD COLUMN time_of_day VARCHAR(20);
```

Adding Column

-- Add the time_of_day column

```
UPDATE sale 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, time_of_day FROM sale;
```

Adding Column

-- Add day_name column


```sql
SELECT date, DAYNAME(date) FROM sale;

ALTER TABLE sale
ADD COLUMN day_name VARCHAR(20);

UPDATE sale
SET day_name = DAYNAME(date);

SELECT invoice_id, day_name FROM sale;
```

Adding Column

-- Add month_name column


```sql
SELECT date, MONTHNAME(date) FROM sale;

ALTER TABLE sale
ADD COLUMN month_name VARCHAR(20);

UPDATE sale
SET month_name = MONTHNAME(date);

SELECT invoice_id, month_name FROM sale;
```

Business Question

#Q1. -- How many unique cities does the data have?

```
5 •   SELECT DISTINCT city FROM sale;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

city
Yangon
Naypyitaw
Mandalay

Business Question

#Q2. -- In which city is each branch?

```sql
SELECT DISTINCT city, branch
FROM sale
ORDER BY branch;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

city	branch
Yangon	A
Mandalay	B
Naypyitaw	C

Business Question

#Q3. -- How many unique product lines does the data have?

```sql
SELECT DISTINCT product_line
FROM sale;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

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

Business Question

#Q4 -- What is the most selling product line

```sql
SELECT product_line, SUM(quantity) AS total_sale
FROM sale
GROUP BY product_line
ORDER BY total_sale DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

product_line	total_sale
Electronic accessories	961
Food and beverages	952
Home and lifestyle	911
Sports and travel	902
Fashion accessories	902

Business Question

#Q5 -- What is the most selling product line

```sql
SELECT
    month_name AS month,
    SUM(total) AS revenue
FROM sale
GROUP BY month_name
ORDER BY revenue;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

month_name	revenue
February	95727.3765
March	108867.1500
January	116291.8680

Business Question

#Q6. -- What month had the largest COGS?

```sql
SELECT
    month_name AS month,
    SUM(cogs) AS total_cogs
FROM sale
GROUP BY month
ORDER BY total_cogs;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

month	total_cogs
February	91168.93
March	103683.00
January	110754.16

Business Question

#Q7. -- What product line had the largest revenue?

```sql
SELECT product_line, SUM(total) as revenue
FROM sale
GROUP BY product_line
ORDER BY revenue DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

product_line	revenue
Food and beverages	56144.8440
Fashion accessories	54305.8950
Sports and travel	53936.1270
Home and lifestyle	53861.9130
Electronic accessories	53783.2365

Business Question

#Q8. -- What is the city with the largest revenue?

```sql
SELECT city, SUM(total) as revenue
FROM sale
GROUP BY city
ORDER BY revenue;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

city	revenue
Mandalay	104534.6085
Yangon	105861.0105
Naypyitaw	110490.7755

Business Question

#Q9. -- What product line had the largest VAT?

```sql
SELECT product_line, SUM(tax_prct) as tax
FROM sale
GROUP BY product_line
ORDER BY tax;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

product_line	tax
Health and beauty	2326.3990
Electronic accessories	2561.1065
Home and lifestyle	2564.8530
Sports and travel	2568.3870
Fashion accessories	2585.9950

Business Question

#Q10. -- Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales

```sql
SELECT
    product_line,
        CASE WHEN AVG(quantity) > (SELECT AVG(quantity) FROM sale) THEN 'Good'
        ELSE 'Bad'
        END as product_remark
FROM sale
GROUP BY product_line;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

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

Business Question

#Q11. -- Which branch sold more products than average product sold?

```sql
SELECT branch, SUM(quantity) as total_sale
FROM sale
GROUP BY branch
HAVING total_sale > (SELECT AVG(quantity) FROM sale);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

branch	total_sale
A	1849
C	1828
B	1795

Business Question

#Q12. -- What is the most common product line by gender

```sql
SELECT gender, product_line, COUNT(gender) as total_count
FROM sale
GROUP BY gender, product_line
ORDER BY total_count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	product_line	total_count
Female	Fashion accessories	96
Female	Food and beverages	90
Male	Health and beauty	88
Female	Sports and travel	86
Male	Electronic accessories	86

Business Question

#Q13. -- What is the average rating of each product line

```sql
SELECT
    product_line as product,
    ROUND(AVG(rating), 2) as avg_rating
FROM sale
GROUP BY product_line
ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

product	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

Business Question

#Q14. -- How many unique customer types does the data have?

```
SELECT DISTINCT customer_type
FROM sale;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_type
Normal
Member

Business Question

#Q15. -- How many unique payment methods does the data have?

```
SELECT DISTINCT payment
FROM sale
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

payment
Credit card
Ewallet
Cash

Business Question

#Q16. -- What is the most common customer type?

```sql
SELECT
    customer_type,
    COUNT(*) as count
FROM sale
GROUP BY customer_type
ORDER BY count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_type	count
Member	499
Normal	496

Business Question

#Q17. -- Which customer type buys the most?

```sql
SELECT customer_type, COUNT(*)
FROM sale
GROUP BY customer_type
ORDER BY COUNT(*) DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_type	COUNT(*)
Member	499
Normal	496

Business Question

#Q18. -- What is the gender of most of the customers?

```sql
SELECT gender, COUNT(*) as gender_count
FROM sale
GROUP BY gender
ORDER BY gender_count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	gender_count
Male	498
Female	497

Business Question

#Q19. -- What is the gender distribution per branch?

```sql
SELECT gender, branch, COUNT(*) as gender_dist_count
FROM sale
GROUP BY gender, branch
ORDER BY branch;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	branch	gender_dist_count
Male	A	179
Female	A	160
Female	B	160
Male	B	169
Female	C	177

Business Question

#Q20. -- Which time of the day do customers give most ratings?

```sql
SELECT
    time_of_day,
    ROUND(AVG(rating),2) as avg_rating
FROM sale
GROUP BY time_of_day
ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

time_of_day	avg_rating
Afternoon	7.02
Morning	6.94
Evening	6.91

Business Question

#Q21. Which time of the day do customers give most ratings per branch?

```sql
SELECT branch, time_of_day, ROUND(AVG(rating),2) AS avg_rating
FROM sale
GROUP BY branch, time_of_day
ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

branch	time_of_day	avg_rating
A	Afternoon	7.19
C	Evening	7.1
C	Afternoon	7.07
A	Morning	7.01
C	Morning	6.97

Business Question

#Q22. -- Which day of the week has the best avg ratings?

```sql
SELECT day_name, ROUND(AVG(rating),2) as avg_rating
FROM sale
GROUP BY day_name
ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

day_name	avg_rating
Monday	7.13
Friday	7.06
Tuesday	7
Sunday	6.99
Saturday	6.9

Business Question

#23. -- Which day of the week has the best average ratings per branch?

```sql
SELECT
    branch,
    day_name,
    ROUND(AVG(rating),2) as avg_rating
FROM sale
GROUP BY branch, day_name
ORDER BY avg_rating DESC
;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

branch	day_name	avg_rating
A	Friday	7.31
B	Monday	7.27
C	Saturday	7.23
C	Friday	7.21
A	Monday	7.1

Business Question

#Q24. -- Number of sales made in each time of the day per weekday

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

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

time_of_day	total_sales
Evening	58
Afternoon	52
Morning	22

Business Question

#Q25. -- Which of the customer types brings the most revenue?

```sql
SELECT customer_type, SUM(total) as revenue
FROM sale
GROUP BY customer_type
ORDER BY revenue DESC
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_type	revenue
Member	163625.1015
Normal	157261.2930

Business Question

#Q26. -- Which city has the largest tax/VAT percent?

```sql
SELECT city, ROUND(AVG(tax_prct),2) as tax
FROM sale
GROUP BY city
ORDER BY tax DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

city	tax
Naypyitaw	16.09
Mandalay	15.13
Yangon	14.87

Business Question

#Q27. -- Which customer type pays the most in VAT?

```sql
SELECT customer_type, ROUND(AVG(tax_prct),2) as total_tax
FROM sale
GROUP BY customer_type
ORDER BY total_tax DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_type	total_tax
Member	15.61
Normal	15.1

Follow Ankush Verma for More

Thank You

CONTACT

GitHub

ankushverma2807@gmail.com

+91 99969 02603