Walmart Sales Analysis Using SQL

By Akhilesh Maurya

Problem Statement

The objective of this project is to perform a comprehensive analysis of Walmart's sales data to uncover actionable insights and trends that can drive improved business strategies and decision-making processes. This analysis will focus on multiple dimensions of the data to provide a holistic understanding of Walmart's sales performance, customer behavior, product trends, and operational efficiency.

Dataset Overview

1. Invoice ID
2. Branch
3. City
4. Customer Type
5. Gender
6. Product Line
7. Unit Price
8. Quantity
9. Tax 5%
10. Total
11. Date
12. Time
13. Payment
14. COGS (Cost of Goods Sold)
15. Gross Margin Percentage
16. Gross Income
17. Rating

Dataset Overview

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

Invoice_ID	Branch	City	Customer_type	Gender	Product_line	Unit_price	Quantity	Tax_pct	Total	Date	Time	Payment	cogs	gross_margin_percentage	gross_income	Rating
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2019-01-05	13:08:00	Ewallet	522.83	4.761904762	26.1415	9.1
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	2019-03-08	10:29:00	Cash	76.4	4.761904762	3.82	9.6
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2019-03-03	13:23:00	Credit card	324.31	4.761904762	16.2155	7.4
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	2019-01-27	20:33:00	Ewallet	465.76	4.761904762	23.288	8.4
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2019-02-08	10:37:00	Ewallet	604.17	4.761904762	30.2085	5.3
699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	2019-03-25	18:30:00	Ewallet	597.73	4.761904762	29.8865	4.1
355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2019-02-25	14:36:00	Ewallet	413.04	4.761904762	20.652	5.8
315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2019-02-24	11:38:00	Ewallet	735.6	4.761904762	36.78	8
665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	2019-01-10	17:15:00	Credit card	72.52	4.761904762	3.626	7.2
692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2019-02-20	13:27:00	Credit card	164.52	4.761904762	8.226	5.9
351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	2019-02-06	18:07:00	Ewallet	57.92	4.761904762	2.896	4.5
529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	2019-03-09	17:03:00	Cash	102.04	4.761904762	5.102	6.8
365-64-0515	A	Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375	246.4875	2019-02-12	10:25:00	Ewallet	234.75	4.761904762	11.7375	7.1
252-56-2699	A	Yangon	Normal	Male	Food and beverages	43.19	10	21.595	453.495	2019-02-07	16:48:00	Ewallet	431.9	4.761904762	21.595	8.2
829-34-3910	A	Yangon	Normal	Female	Health and beauty	71.38	10	35.69	749.49	2019-03-29	19:21:00	Cash	713.8	4.761904762	35.69	5.7
299-46-1805	B	Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	2019-01-15	16:19:00	Cash	562.32	4.761904762	28.116	4.5
656-95-9349	A	Yangon	Member	Female	Health and beauty	68.93	7	24.1255	506.6355	2019-03-11	11:03:00	Credit card	482.51	4.761904762	24.1255	4.6
765-26-6951	A	Yangon	Normal	Male	Sports and travel	72.61	6	21.783	457.443	2019-01-01	10:39:00	Credit card	435.66	4.761904762	21.783	6.9
329-62-1586	A	Yangon	Normal	Male	Food and beverages	54.67	3	8.2005	172.2105	2019-01-21	18:00:00	Credit card	164.01	4.761904762	8.2005	8.6
310-50-2348	B	Mandalay	Normal	Female	Home and lifestyle	40.3	2	4.03	84.63	2019-03-11	15:30:00	Ewallet	80.6	4.761904762	4.03	4.4

Data Cleaning & features Engineering

-- Add the time_of_day new column

```
alter table walmart add column time_of_day varchar(20);


update walmart
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);
```

time_of_day
Afternoon
Morning
Afternoon
Evening
Morning
Evening
Afternoon
Morning
Evening
Afternoon
Evening
Evening
Morning
Evening
Evening
Evening
Morning

-- add day_name column

```
29      -- add day_name column
30
31  •   select date,dayname(date) from walmart;
32
33  •   alter table walmart add column day_name varchar(20);
34
35  •   update walmart
36      set day_name = dayname(date);
37
```

day_name
Saturday
Friday
Sunday
Sunday
Friday
Monday
Monday
Sunday
Thursday
Wednesday
Wednesday
Saturday
Tuesday
Thursday
Friday
Tuesday
Monday
Tuesday
Monday

-- add month_name column

```
-- add month_name column

SELECT
    date,
    MONTHNAME(date)
FROM walmart;

ALTER TABLE walmart ADD COLUMN month_name VARCHAR(10);

UPDATE walmart
SET month_name = MONTHNAME(date);
```

month_name
January
March
March
January
February
March
February
February
January
February
February
March
February
February

Analysis

Q1.How many unique product lines does the data have?

```
51  -- ------------------------------------------------------------
52  -- Q1.How many unique product lines does the data have?
53
54  select count(distinct product_line) from walmart;
55
56  -- answer. 6
```

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

count(distinct product_line)
6

Q2.What is the most common payment method?

```
58  -- Q2.What is the most common payment method?
59
60  select Payment,count(payment) as payment_type from walmart
61  group by Payment
62  order by payment_type desc;
63
64  -- answer. Ewallet
65
```

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

Payment	payment_type
Ewallet	345
Cash	344
Credit card	311

Q3.What is the most selling product line?

```sql
-- Q3.What is the most selling product line?

SELECT
    SUM(quantity) as qty,
    product_line
FROM walmart
GROUP BY product_line
ORDER BY qty DESC;
```

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

qty	product_line
971	Electronic accessories
952	Food and beverages
920	Sports and travel
911	Home and lifestyle
902	Fashion accessories
854	Health and beauty

Q4.What is the total revenue by month?

```sql
-- Q4.What is the total revenue by month?

SELECT
    month_name AS month,
    SUM(total) AS total_revenue
FROM walmart
GROUP BY month_name
ORDER BY total_revenue desc;
```

Result Grid | Filter Rows: | Export:

month	total_revenue
January	116291.86800000005
March	109455.50700000004
February	97219.37399999997

Q5.What month had the largest COGS?

```sql
-- Q5.What month had the largest COGS?

SELECT
    month_name AS month,
    SUM(cogs) AS cogs
FROM walmart
GROUP BY month_name
ORDER BY cogs desc;
```

Result Grid | Filter Rows: | Export:

month	cogs
January	110754.16000000002
March	104243.33999999997
February	92589.88

Q6.What product line had the largest revenue?

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

Result Grid | Filter Rows: | Export:

product_line	total_revenue
Food and beverages	56144.844000000005
Sports and travel	55122.826499999996
Electronic accessories	54337.531500000005
Fashion accessories	54305.895
Home and lifestyle	53861.91300000001
Health and beauty	49193.739000000016

Q7.What is the city with the largest revenue?

```
109
110     -- Q7.What is the city with the largest revenue?
111
112 •   SELECT
113         branch,
114         city,
115         SUM(total) AS total_revenue
116     FROM walmart
117     GROUP BY city, branch
118     ORDER BY total_revenue;
```

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

branch	city	total_revenue
B	Mandalay	106197.67199999996
A	Yangon	106200.3705000001
C	Naypyitaw	110568.70649999994

Q8.What product line had the largest VAT?

```
122     -- Q8.What product line had the largest VAT?
123
124 •   SELECT
125         product_line,
126         AVG(tax_pct) as avg_tax
127     FROM walmart
128     GROUP BY product_line
129     ORDER BY avg_tax DESC;
```

Result Grid | Filter Rows: | Export: | W

product_line	avg_tax
Home and lifestyle	16.03033125000001
Sports and travel	15.812629518072285
Health and beauty	15.411572368421048
Food and beverages	15.365310344827583
Electronic accessories	15.22059705882354
Fashion accessories	14.528061797752809

Q9.Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales.

```
134 •    SELECT
135          AVG(quantity) AS avg_qnty
136      FROM walmart;
137
138 •    SELECT
139          product_line,
140          CASE
141              WHEN AVG(quantity) > 6 THEN "Good"
142              ELSE "Bad"
143          END AS remark
144      FROM walmart
145      GROUP BY product_line;
```

Result Grid | Filter Rows: | Export: | Wrap

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

Q10.Which branch sold more products than average product sold?

```
151  SELECT
152      branch,
153      SUM(quantity) AS qnty
154  FROM walmart
155  GROUP BY branch
156  HAVING SUM(quantity) > (SELECT AVG(quantity) FROM walmart);
```

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

branch	qnty
A	1859
C	1831
B	1820

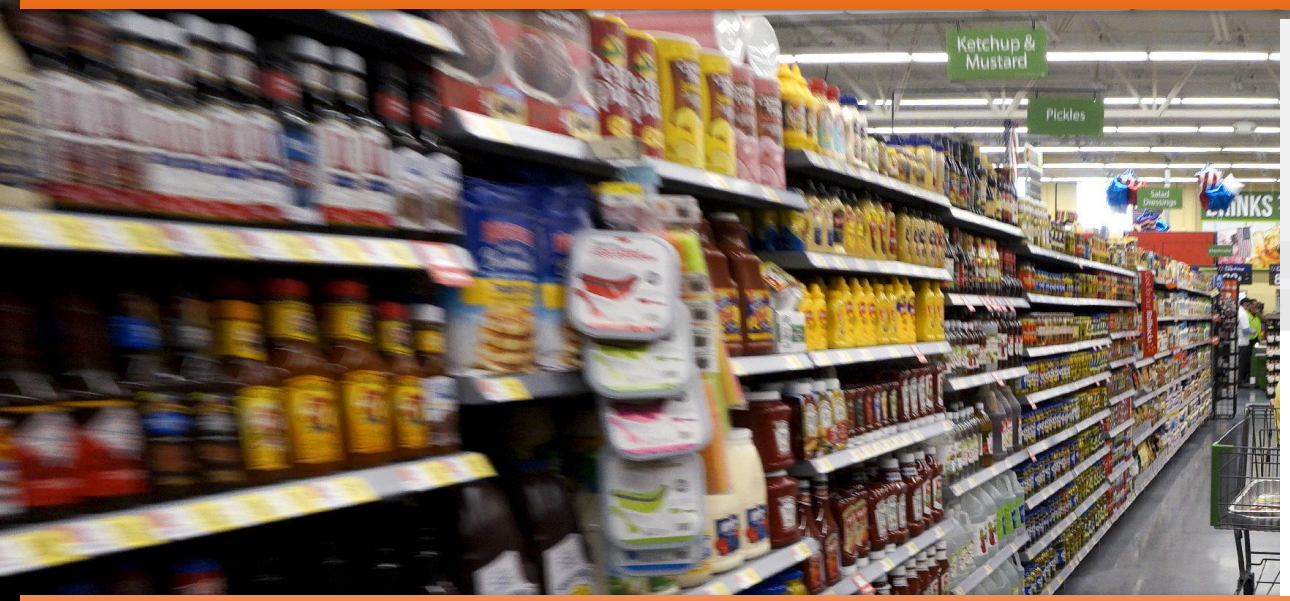
Q11.What is the most common product line by gender?

```
161  SELECT
162      gender,
163      product_line,
164      COUNT(gender) AS total_cnt
165  FROM walmart
166  GROUP BY gender, product_line
167  ORDER BY total_cnt DESC;
```

Result Grid | Filter Rows: | Export:

gender	product_line	total_cnt
Female	Fashion accessories	96
Female	Food and beverages	90
Male	Health and beauty	88
Female	Sports and travel	88
Male	Electronic accessories	86
Female	Electronic accessories	84
Male	Food and beverages	84

Q12.What is the average rating of each product line?




```
173 •   SELECT
174         ROUND(AVG(rating), 2) as avg_rating,
175         product_line
176     FROM walmart
177     GROUP BY product_line
178     ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows: | Export: | Wra

avg_rating	product_line
7.11	Food and beverages
7.03	Fashion accessories
7	Health and beauty
6.92	Electronic accessories
6.92	Sports and travel
6.84	Home and lifestyle

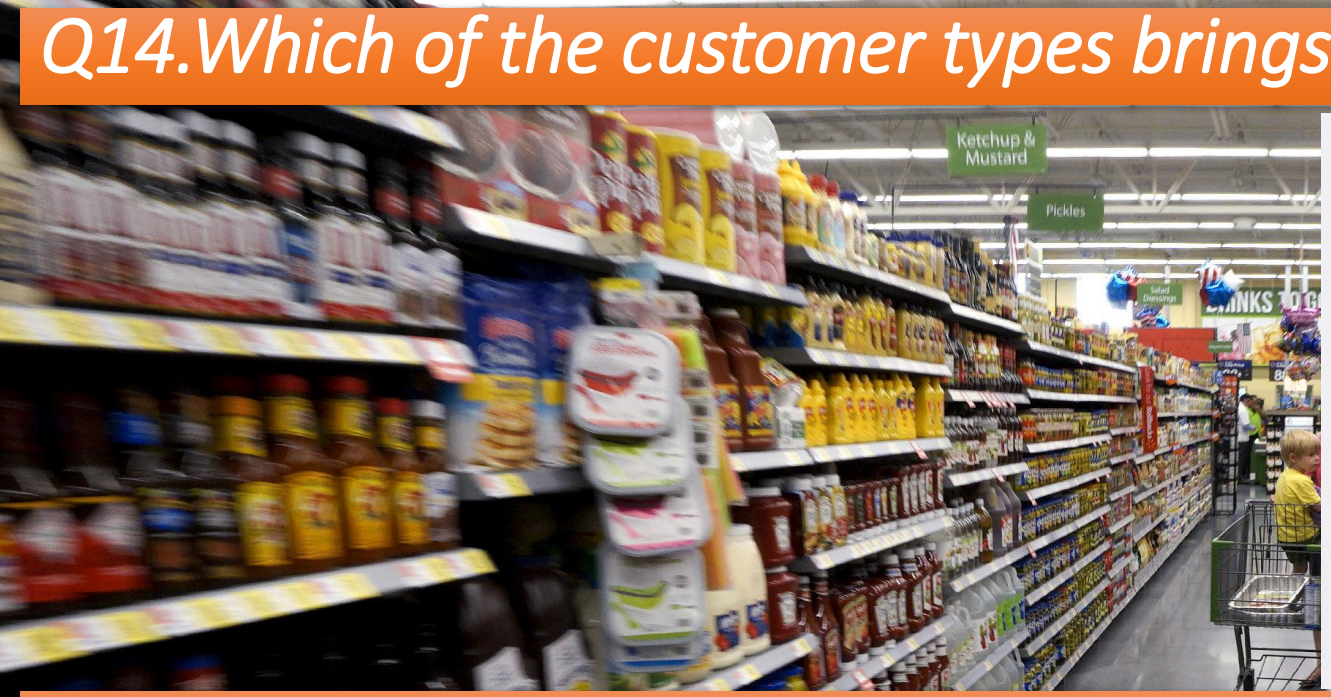
Q13.Number of sales made in each time of the day per weekday.

```
185 •   SELECT
186         time_of_day,
187         COUNT(*) AS total_sales
188     FROM walmart
189     WHERE day_name = "Sunday"
190     GROUP BY time_of_day
191     ORDER BY total_sales DESC;
```

Result Grid | Filter Rows: | Export:

time_of_day	total_sales
Evening	58
Afternoon	53
Morning	22

Q14.Which of the customer types brings the most revenue?




```
196
197 •   SELECT
198         customer_type,
199         SUM(total) AS total_revenue
200     FROM walmart
201     GROUP BY customer_type
202     ORDER BY total_revenue;
```

Result Grid | Filter Rows: | Export:

customer_type	total_revenue
Normal	158743.30500000005
Member	164223.44400000002

Q15.Which city has the largest tax percent/ VAT (Value Added Tax)?

```
208 •   SELECT
209         city,
210         ROUND(AVG(tax_pct), 2) AS avg_tax_pct
211     FROM walmart
212     GROUP BY city
213     ORDER BY avg_tax_pct DESC;
```

Result Grid | Filter Rows: | Export: | Wrap C

city	avg_tax_pct
Naypyitaw	16.05
Mandalay	15.23
Yangon	14.87

Q16.How many unique payment methods does the data have?

```
239
240 •   SELECT
241         DISTINCT payment
242     FROM walmart;
```

Result Grid | Filter Rows:

payment
Ewallet
Cash
Credit card

Q17.Which customer type buys the most?

```
249
250 •   SELECT
251         customer_type,
252         COUNT(*)
253     FROM walmart
254     GROUP BY customer_type;
```

Result Grid | Filter Rows:

customer_type	COUNT(*)
Member	501
Normal	499

Q18.Which time of the day do customers give most ratings?

```
283 •   SELECT
284         time_of_day,
285         AVG(rating) AS avg_rating
286     FROM walmart
287     GROUP BY time_of_day
288     ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows:

time_of_day	avg_rating
Afternoon	7.031299734748012
Morning	6.960732984293193
Evening	6.926851851851853

Q19.Which day of the week has the best avg ratings?

```
305 •   SELECT
306         day_name,
307         AVG(rating) AS avg_rating
308     FROM walmart
309     GROUP BY day_name
310     ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows:

day_name	avg_rating
Monday	7.153599999999999
Friday	7.076258992805756
Sunday	7.011278195488723
Tuesday	7.003164556962025
Saturday	6.901829268292688
Thursday	6.88985507246377
Wednesday	6.805594405594405