

# SQL PROJECT

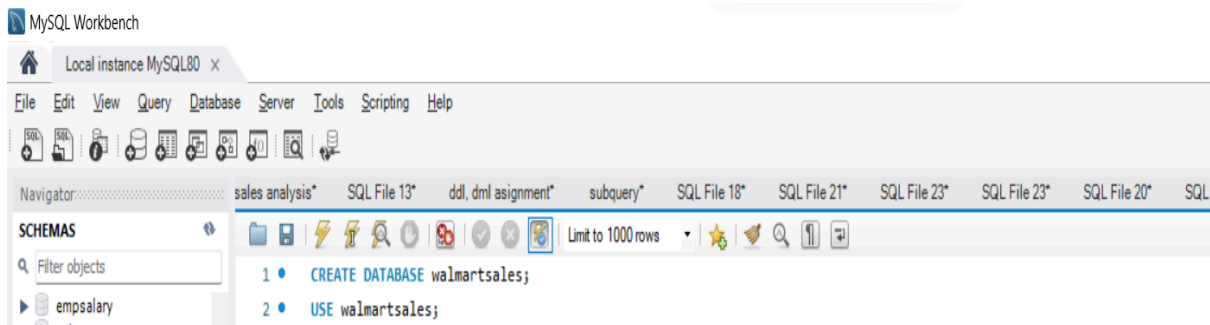
## WALMART SALES ANALYSIS

**AIM:** 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.

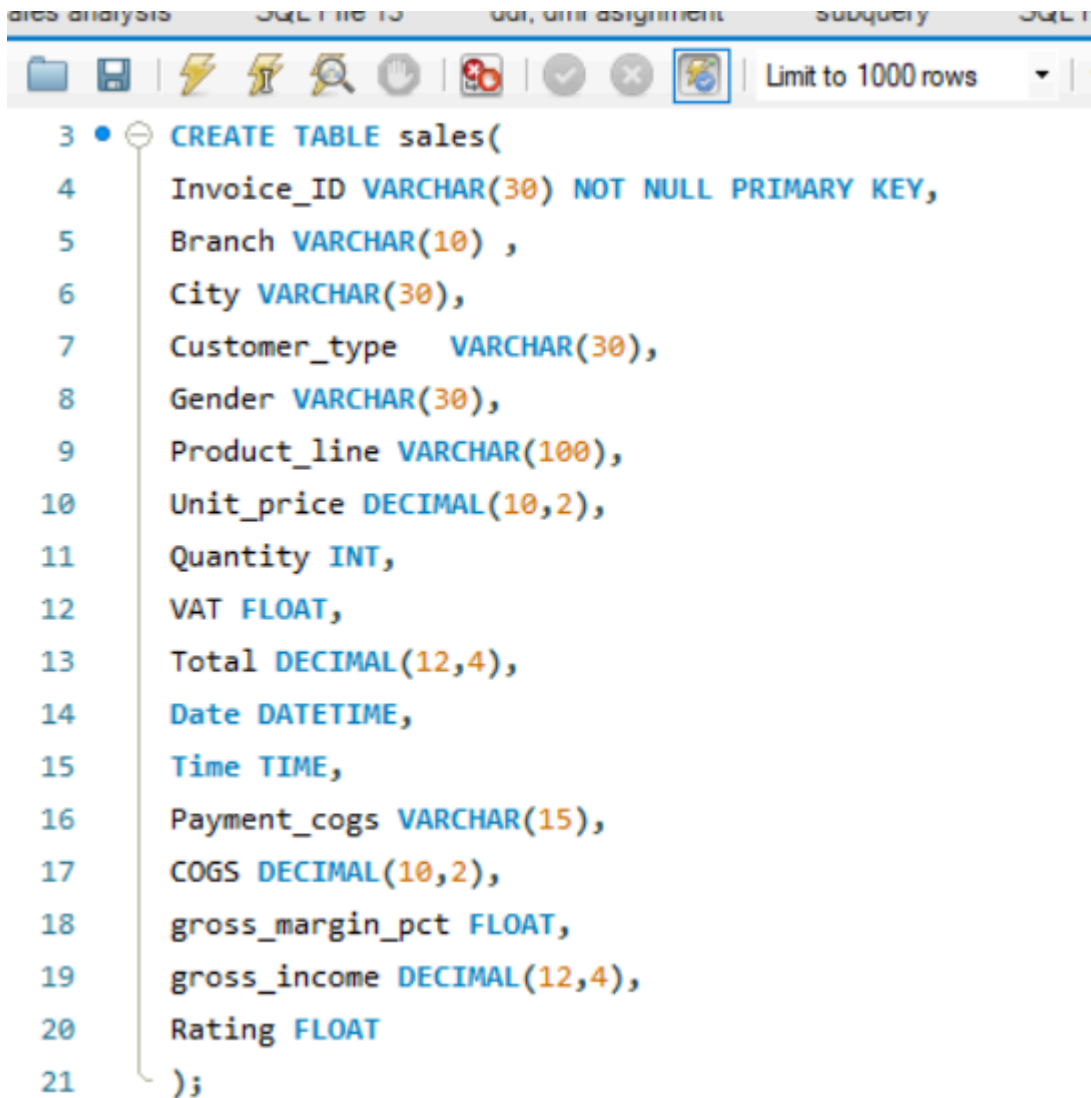
### ABOUT:

This dataset contains sales transactions from three different branches of Walmart, respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows.

### CREATING DATABASE:



### CREATE TABLE AND INSERT COLUMNS:



```
3 • CREATE TABLE sales(  
4     Invoice_ID VARCHAR(30) NOT NULL PRIMARY KEY,  
5     Branch VARCHAR(10) ,  
6     City VARCHAR(30),  
7     Customer_type VARCHAR(30),  
8     Gender VARCHAR(30),  
9     Product_line VARCHAR(100),  
10    Unit_price DECIMAL(10,2),  
11    Quantity INT,  
12    VAT FLOAT,  
13    Total DECIMAL(12,4),  
14    Date DATETIME,  
15    Time TIME,  
16    Payment_cogs VARCHAR(15),  
17    COGS DECIMAL(10,2),  
18    gross_margin_pct FLOAT,  
19    gross_income DECIMAL(12,4),  
20    Rating FLOAT  
21 );
```

## APPROACH:

### FEATURE ENGINEERING:

1. Add a new column named **time\_of\_day** to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.

```

22
23 -----FEATURE ENGINEERING-----
24 #Add a new column named time_of_day to give insight of sales in the Morning, Afternoon and Evening.
25 #This will help answer the question on which part of the day most sales are made
26 SELECT
27     time,
28     (CASE
29         WHEN `time` BETWEEN "00:00:00" AND "12:00:00" THEN "Morning"
30         WHEN `time` BETWEEN "12:01:00" AND "16:00:00" THEN "Afternoon"
31         ELSE "Evening"
32     END) AS time_of_day
33 FROM sales;
34
35

```

time	time_of_day
19:44:00	Evening
12:36:00	Afternoon
17:52:00	Evening
18:02:00	Evening
12:22:00	Afternoon
15:10:00	Afternoon
11:26:00	Morning
15:01:00	Afternoon
11:36:00	Morning

Result 1 x

2.

2. Add a new column named `day_name` that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.

### 3. #ADD NEW COLUMN DAY\_NAME

```

SELECT date,
       DAYNAME(date) AS day_name
FROM sales;
ALTER TABLE sales ADD COLUMN day_name VARCHAR(10);

```

## EXPLORATORY DATA ANALYSIS (EDA):

- 1.) How many unique cities does the data have?

```

67 -----EXPLORATORY DATA ANALYSIS-----
68
69 #1 How many unique cities does the data have?
70 • SELECT distinct city from sales;
71
72 #2 In which city is each branch?

```

city
Yangon
Naypyitaw
Mandalay

2.) In which city is each branch?

```
70 • SELECT distinct city from sales;
```

```
71
```

```
72 #2 In which city is each branch?
```

```
73 • SELECT distinct city,branch from sales;
```

```
74
```

```
75 #3 How many unique product lines does the data have?
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
city	branch		
Yangon	A		
Naypyitaw	C		
Mandalay	B		

3.) How many unique product lines does the data have?

```
74
```

```
75 #3 How many unique product lines does the data have?
```





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

```
77
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
COUNT(DISTINCT product_line)			
6			





4.) what is the most selling product line?

```
78
79 #4 what is the most selling product line?
80 • SELECT product_line,
81 COUNT(product_line) AS CNT FROM sales
82 GROUP BY product_line
83 ORDER BY CNT DESC;
84
```

Result Grid     Filter Rows: <input type="text"/>   Export:    Wrap Cell Content: 		
	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

5.) What is the total revenue by month?



```
84
85 #5 What is the total revenue by month?
86 • SELECT month_name AS MONTH, SUM(TOTAL) AS total_revenue
87 FROM sales
88 GROUP BY month_name
89 ORDER BY total_revenue DESC;
90
```

Result Grid     Filter Rows: <input type="text"/>   Export:    Wrap Cell Content: 		
	MONTH	total_revenue
▶	January	116291.8680
	March	108867.1500
	February	95727.3765

- 6.) Create a stored procedure that retrieves all the details of a sale based on the provided invoice ID.
- 7.) Create a stored procedure that takes two parameters, branch\_name and gender, and returns a summary of sales transactions for the specified branch and gender.

8.) What product line had the largest revenue?

```
117 #8-- What product line had the largest revenue?
118 • SELECT
119     product_line,
120     SUM(total) as total_revenue
121 FROM sales
122 GROUP BY product_line
123 ORDER BY total_revenue DESC;
124
125 #9-- What is the city with the largest revenue?
```

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

	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

9.) What is the city with the largest revenue?

```
125 #9-- What is the city with the largest revenue?
126 • SELECT
127     branch,
128     city,
129     SUM(total) AS total_revenue
130 FROM sales
131 GROUP BY city, branch
132 ORDER BY total_revenue;
133
134 #10 Which branch sold more products than average product so
```

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

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

10.) Which branch sold more products than average product sold?

```
134 #10 Which branch sold more products than average product sold?
135 • SELECT
136     branch,
137     SUM(quantity) AS qty
138 FROM sales
139 GROUP BY branch
140 HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);
141
142 #11 What is the most common product line by gender
```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	branch	qty			
▶	A	1849			
	C	1828			
	B	1795			

11.) What is the most common product line by gender

```
142      #11 What is the most common product line by gender
143 •    SELECT
144         gender,
145         product_line,
146         COUNT(gender) AS total_cnt
147     FROM sales
148     GROUP BY gender, product_line
149     ORDER BY total_cnt DESC;
150
151      #12 How many unique customer types does the data have?
```

Result Grid			
		Filter Rows:	
		Export:	
		Wrap Cell Content:	
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	
Male	Sports and travel	77	

12.) How many unique customer types does the data have?

```
150
151      #12 How many unique customer types does the data have?
152 •    SELECT
153         DISTINCT customer_type
154     FROM sales;
155
156      #13 How many unique payment methods does the data have?
157 •    SELECT
158         DISTINCT payment_cogs
159     FROM sales;
160
```

Result Grid	
Filter Rows:	
Export:	
Wrap Cell Content:	
customer_type	
Normal	
Member	



13.) How many unique payment methods does the data have?

```
154 FROM sales;
155
156 #13 How many unique payment methods does the data have?
157 • SELECT
158     DISTINCT payment_cogs
159 FROM sales;
```

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

payment_cogs
Credit card
Ewallet
Cash

14.) Which customer type buys the most?

```
160
161 #14 Which customer type buys the most?
162 • SELECT
163     customer_type,
164     COUNT(*)
165 FROM sales
166 GROUP BY customer_type;
```

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

customer_type	COUNT(*)
Normal	496
Member	499

15.) What is the gender of most of the customers?

```
170 #15 What is the gender of most of the customers?
171 • SELECT
172     gender,
173     COUNT(*) as gender_cnt
174 FROM sales
175 GROUP BY gender
176 ORDER BY gender_cnt DESC;
177
178 #16 Which time of the day do customers give most ratings?
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
gender	gender_cnt		
Male	498		
Female	497		

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

```
178 #16 Which time of the day do customers give most ratings?
179 • SELECT
180     time_of_day,
181     AVG(rating) AS avg_rating
182 FROM sales
183 GROUP BY time_of_day
184 ORDER BY avg_rating DESC;
185
186 #17 Number of sales made in each time of the day per weekday
187 • SELECT
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
time_of_day	avg_rating		
Afternoon	7.02340		
Morning	6.94474		
Evening	6.90536		

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

```
186 #17 Number of sales made in each time of the day per weekday
187 • SELECT
188     time_of_day,
189     COUNT(*) AS total_sales
190 FROM sales
191 WHERE day_name = "Sunday"
192 GROUP BY time_of_day
193 ORDER BY total_sales DESC;
194
195 #18 Which of the customer types brings the most revenue?
196 • SELECT
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	time_of_day	total_sales		
▶	Evening	58		
	Afternoon	52		
	Morning	22		

18.) Which of the customer types brings the most revenue

```
195 #18 Which of the customer types brings the most revenue?
196 • SELECT
197     customer_type,
198     SUM(total) AS total_revenue
199 FROM sales
200 GROUP BY customer_type
201 ORDER BY total_revenue;
202
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	customer_type	total_revenue		
▶	Normal	157261.2930		
	Member	163625.1015		



19.)

20.)

21.)