Walmart Sales Data Analysis

Create database

```
CREATE DATABASE IF NOT EXISTS walmartSales;
```

Create table

```
CREATE TABLE IF NOT EXISTS sales(
    invoice_id VARCHAR(30) NOT NULL PRIMARY KEY,
    branch VARCHAR(5) NOT NULL,
    city VARCHAR(30) 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_pct FLOAT(6,4) NOT NULL,
    total DECIMAL(12, 4) NOT NULL,
    date DATETIME NOT NULL,
    time TIME NOT NULL,
    payment VARCHAR(15) NOT NULL,
    cogs DECIMAL(10,2) NOT NULL,
    gross_margin_pct FLOAT(11,9),
    gross_income DECIMAL(12, 4),
    rating FLOAT(2, 1)
);
```

ADD the **time_of_day** column

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

invoice_id	branch	city	customer_type	gender	product_line	unit_price	quantity	tax_pct	total	date	time	payment	cogs	gross_margin_pct	gross_income	rating	time_of_day
101-17-6199	A	Yangon	Normal	Male	Food and beverages	45.79	7	16.0265	336.5565	2019-03-13 00:00:00	19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	*NULL*
101-81-4070	C	Naypyitaw	Member	Female	Health and beauty	62.82	2	6.2820	131.9220	2019-01-17 00:00:00	12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	*NULL*

UPDATE time_of_day column

```
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)

);
```

invoice_id	branch	city	customer_type	gender	product_line	unit_price	quantity	tax_pct	total	date	time	payment	cogs	gross_margin_pct	gross_income	rating	time_of_day
101-17-6199	A	Yangon	Normal	Male	Food and beverages	45.79	7	16.0265	336.5565	2019-03-13 00:00:00	19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	evening
101-81-4070	C	Naypyitaw	Member	Female	Health and beauty	62.82	2	6.2820	131.9220	2019-01-17 00:00:00	12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	afternoon
102-06-2002	C	Naypyitaw	Member	Male	Sports and travel	25.25	5	6.3125	132.5625	2019-03-20 00:00:00	17:52:00	Cash	126.25	4.761904716	6.3125	6.1	evening
102-77-2261	C	Naypyitaw	Member	Male	Health and beauty	65.31	7	22.8585	480.0285	2019-03-05 00:00:00	18:02:00	Credit card	457.17	4.761904716	22.8585	4.2	evening
105-10-6182	A	Yangon	Member	Male	Fashion accessories	21.48	2	2.1480	45.1080	2019-02-27 00:00:00	12:22:00	Ewallet	42.96	4.761904716	2.1480	6.6	afternoon
105-31-1824	A	Yangon	Member	Male	Sports and travel	69.52	7	24.3320	510.9720	2019-02-01 00:00:00	15:10:00	Credit card	486.64	4.761904716	24.3320	8.5	afternoon
106-35-6779	A	Yangon	Member	Male	Home and lifestyle	44.34	2	4.4340	93.1140	2019-03-27 00:00:00	11:26:00	Cash	88.68	4.761904716	4.4340	5.8	morning
109-28-2512	B	Mandalay	Member	Female	Fashion accessories	97.61	6	29.2830	614.9430	2019-01-07 00:00:00	15:01:00	Ewallet	585.66	4.761904716	29.2830	9.9	afternoon
109-86-4363	B	Mandalay	Member	Female	Sports and travel	60.08	7	21.0280	441.5880	2019-02-14 00:00:00	11:36:00	Credit card	420.56	4.761904716	21.0280	4.5	morning

ADD the **day_name** column

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

ender	product_line	unit_price	quantity	tax_pct	total	date	time	payment	cogs	gross_margin_pct	gross_income	rating	time_of_day	day_name
lale	Food and beverages	45.79	7	16.0265	336.5565	2019-03-13 00:00:00	19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	evening	NULL
emale	Health and beauty	62.82	2	6.2820	131.9220	2019-01-17 00:00:00	12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	afternoon	NULL
lale	Sports and travel	25.25	5	6.3125	132.5625	2019-03-20 00:00:00	17:52:00	Cash	126.25	4.761904716	6.3125	6.1	evening	NULL
lale	Health and beauty	65.31	7	22.8585	480.0285	2019-03-05 00:00:00	18:02:00	Credit card	457.17	4.761904716	22.8585	4.2	evening	NULL
lale	Fashion accessories	21.48	2	2.1480	45.1080	2019-02-27 00:00:00	12:22:00	Ewallet	42.96	4.761904716	2.1480	6.6	afternoon	NULL
lale	Sports and travel	69.52	7	24.3320	510.9720	2019-02-01 00:00:00	15:10:00	Credit card	486.64	4.761904716	24.3320	8.5	afternoon	NULL
lale	Home and lifestyle	44.34	2	4.4340	93.1140	2019-03-27 00:00:00	11:26:00	Cash	88.68	4.761904716	4.4340	5.8	morning	NULL
emale	Fashion accessories	97.61	6	29.2830	614.9430	2019-01-07 00:00:00	15:01:00	Ewallet	585.66	4.761904716	29.2830	9.9	afternoon	NULL

UPDATE day_name column

```
UPDATE sales
SET day_name = DAYNAME(date);
```

	unit_price	quantity	tax_pct	total	date	time	payment	cogs	gross_margin_pct	gross_income	rating	time_of_day	day_name
es	45.79	7	16.0265	336.5565	2019-03-13 00:00:00	19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	evening	Wednesday
	62.82	2	6.2820	131.9220	2019-01-17 00:00:00	12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	afternoon	Thursday
	25.25	5	6.3125	132.5625	2019-03-20 00:00:00	17:52:00	Cash	126.25	4.761904716	6.3125	6.1	evening	Wednesday
	65.31	7	22.8585	480.0285	2019-03-05 00:00:00	18:02:00	Credit card	457.17	4.761904716	22.8585	4.2	evening	Tuesday
es	21.48	2	2.1480	45.1080	2019-02-27 00:00:00	12:22:00	Ewallet	42.96	4.761904716	2.1480	6.6	afternoon	Wednesday
	69.52	7	24.3320	510.9720	2019-02-01 00:00:00	15:10:00	Credit card	486.64	4.761904716	24.3320	8.5	afternoon	Friday
	44.34	2	4.4340	93.1140	2019-03-27 00:00:00	11:26:00	Cash	88.68	4.761904716	4.4340	5.8	morning	Wednesday
es	97.61	6	29.2830	614.9430	2019-01-07 00:00:00	15:01:00	Ewallet	585.66	4.761904716	29.2830	9.9	afternoon	Monday
	60.08	7	21.0280	441.5880	2019-02-14 00:00:00	11:36:00	Credit card	420.56	4.761904716	21.0280	4.5	morning	Thursday
es	39.43	6	11.8290	248.4090	2019-03-25 00:00:00	20:18:00	Credit card	236.58	4.761904716	11.8290	9.4	evening	Monday

ADD the **month_name** column

```
1 ALTER TABLE sales add COLUMN month_name VARCHAR(10);
2
```

UPDATE **month_name** column

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

SELECT * FROM sales ;

ct	total	date	time	payment	cogs	gross_margin_pct	gross_income	rating	time_of_day	day_name	month_name
265	336.5565	2019-03-13 00:00:00	19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	evening	Wednesday	March
320	131.9220	2019-01-17 00:00:00	12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	afternoon	Thursday	January
125	132.5625	2019-03-20 00:00:00	17:52:00	Cash	126.25	4.761904716	6.3125	6.1	evening	Wednesday	March
585	480.0285	2019-03-05 00:00:00	18:02:00	Credit card	457.17	4.761904716	22.8585	4.2	evening	Tuesday	March
480	45.1080	2019-02-27 00:00:00	12:22:00	Ewallet	42.96	4.761904716	2.1480	6.6	afternoon	Wednesday	February
320	510.9720	2019-02-01 00:00:00	15:10:00	Credit card	486.64	4.761904716	24.3320	8.5	afternoon	Friday	February
340	93.1140	2019-03-27 00:00:00	11:26:00	Cash	88.68	4.761904716	4.4340	5.8	morning	Wednesday	March

Business Questions To Answer

.....Generic Question.....

1. How many unique cities does the data have?

```
SELECT DISTINCT city
FROM sales ;
```

city
Yangon
Naypyitaw
Mandalay

2. In which city is each branch?

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

branch	city
A	Yangon
C	Naypyitaw
B	Mandalay

.....Product.....

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

```
SELECT DISTINCT product_line
FROM sales ;
```

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

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

COUNT(DISTINCT product_line)
6

2. What is the most common payment method?

```
SELECT
   Payment,
   COUNT(Payment) as cnt
FROM sales
GROUP BY Payment
ORDER BY cnt DESC;
```

Payment	cnt ▾ 1
Ewallet	345
Cash	344
Credit card	311

3. What is the most selling product line?

```
SELECT product_line, COUNT(product_line) as cnt
FROM sales
GROUP BY product_line
ORDER BY cnt DESC;
```

product_line	cnt 1
Fashion accessories	178
Food and beverages	174
Electronic accessories	170
Sports and travel	166
Home and lifestyle	160
Health and beauty	152

4. What is the total revenue by month?

```
SELECT month_name  AS month, SUM(total) as total_revenue
FROM sales
GROUP BY month_name
ORDER BY  total_revenue DESC;
```

month	total_revenue 1
January	116291.8680
March	109455.5070
February	97219.3740

5. What month had the largest COGS?

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

month	cogs 1
January	110754.16
March	104243.34
February	92589.88

6. What 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 1
Food and beverages	56144.8440
Sports and travel	55122.8265
Electronic accessories	54337.5315
Fashion accessories	54305.8950
Home and lifestyle	53861.9130
Health and beauty	49193.7390

7. What is the city with the largest revenue?

```
SELECT city, SUM(total) AS total_revenue
FROM sales
GROUP BY city
ORDER BY total_revenue DESC;
```

city	total_revenue 1
Naypyitaw	110568.7065
Yangon	106200.3705
Mandalay	106197.6720

8. What product line had the largest VAT?

```
SELECT product_line, AVG(tax_pct) AS avg_tax
FROM sales
GROUP BY product_line
ORDER BY avg_tax DESC;
```

product_line	avg_tax 1
Home and lifestyle	16.03033124
Sports and travel	15.81262955
Health and beauty	15.41157238
Food and beverages	15.36531029
Electronic accessories	15.22059705
Fashion accessories	14.52806181

9. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales
10. 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	1859
B	1820
C	1831

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

```
SELECT gender ,product_line ,COUNT(gender) AS total_count
FROM sales
GROUP BY gender,product_line
ORDER BY  total_count DESC;
```

gender	product_line	total_count ▼ 1
Female	Fashion accessories	96
Female	Food and beverages	90
Female	Sports and travel	88
Male	Health and beauty	88
Male	Electronic accessories	86
Female	Electronic accessories	84
Male	Food and beverages	84
Male	Fashion accessories	82
Male	Home and lifestyle	81
Female	Home and lifestyle	79
Male	Sports and travel	78
Female	Health and beauty	64

12.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 ▾ 1
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	7.00
Electronic accessories	6.92
Sports and travel	6.91
Home and lifestyle	6.84

.....Sales.....

....

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

```
SELECT time_of_day,COUNT(total) AS total_sales
FROM sales
GROUP BY time_of_day;
```

time_of_day	total_sales
afternoon	377
evening	432
morning	191

2. 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 1
Member	164223.4440
Normal	158743.3050

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

```
SELECT city,AVG(tax_pct) AS VAT
FROM sales
GROUP BY city
ORDER BY VAT DESC;
```

city	VAT 1
Naypyitaw	16.05236732
Mandalay	15.23202414
Yangon	14.87400148

4. Which customer type pays the most in VAT?

```
SELECT customer_type,AVG(tax_pct) AS VAT
FROM sales
GROUP BY customer_type
ORDER BY VAT DESC;
```

customer_type	VAT 1
Member	15.60910977
Normal	15.14870742

……..Customer……..

1. How many unique customer types does the data have?

```
SELECT DISTINCT customer_type
FROM sales;
```

customer_type
Normal
Member

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

```
SELECT DISTINCT payment
FROM sales;
```

payment
Credit card
Ewallet
Cash

3. What is the most common customer type?

```
SELECT Customer_type, COUNT(*) AS Total_Count
FROM sales
GROUP BY Customer_type
ORDER BY Total_Count DESC
LIMIT 1;
```

Customer_type	Total_Count
Member	501

4. Which customer type buys the most?

```
SELECT Customer_type, COUNT(*) AS Total_Count
FROM sales
GROUP BY Customer_type
ORDER BY Total_Count DESC
LIMIT 1;
```

Customer_type	Total_Count
Member	501

5. What is the gender of most of the customers?

```
SELECT gender, COUNT(*) AS gender_cnt
FROM sales
GROUP BY  gender
ORDER BY gender_cnt DESC
LIMIT 1;
```

gender	gender_cnt
Female	501

6. What is the gender distribution per branch?

```
SELECT gender, COUNT(*) AS gender_cnt
FROM sales
WHERE branch = "A"
GROUP BY  gender
ORDER BY gender_cnt DESC;
```

A

gender	gender_cnt ▾ 1
Male	179
Female	161

B

gender	gender_cnt ▾ 1
Male	170
Female	162

C

gender	gender_cnt ▾ 1
Female	178
Male	150

7. 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
LIMIT 1;
```

time_of_day	avg_rating
afternoon	7.03103

8. 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
LIMIT 1;
```

A

time_of_day	avg_rating
afternoon	7.18889

B

time_of_day	avg_rating
morning	6.88983

C

time_of_day	avg_rating
evening	7.11818

9. Which day fo 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
LIMIT 1;
```

day_name	avg_rating
Monday	7.15280

10. Which day of the week has the best average ratings per branch?

```
SELECT day_name, branch, AVG(rating) AS avg_rating
FROM sales
GROUP BY day_name , branch
ORDER BY avg_rating DESC
LIMIT 3;
```

day_name	branch	avg_rating ▾ 1
Monday	B	7.33333
Friday	A	7.31200
Friday	C	7.27632