Walmart Sales Data Analysis

Primary KPI's-

1) How many unique cities does data have?
2) In which city is each branch?

KPI's- 1. PRODUCT ANALYSIS-

1) How many unique product lines does the data have?
2) What is the most common payment method?
3) What are the most selling product lines?
4) What is the total revenue by month?
5) What month had the largest COGS?
6) What product line had the largest revenue?
7) What is the city with the largest revenue?
8) What product lines had the largest VAT?
9) Fetch each product lines and add a column to those product lines showing "Good" and "Bad". Good if it's greater than Average Sales?
10) Which branch sold more product than Average product sold?
11) What is the most common product line by Gender?
12) What is the Average rating of each product lines?

2. SALES ANALYSIS-

1) Number of Sales made in each Time of Day per Weekday?
2) Which customer type brings the most revenue?
3) Which city has the largest tax/VAT percent (Value add tax)?
4) Which customer type pays the most in VAT?

3.CUSTOMER ANALYSIS-

1) How many Unique Customer type does the data have?
2) How many unique payment methods does the data have?
3) What is the most common customer type?
4) Which customer type buys the most?
5) What is the gender of most of the customers?
6) What is the gender distribution per branch?
7) Which time of day do customers give most Ratings?
8) Which time of day do customers give the most Ratings per Branch?
9) Which day of the week has the best AVG Ratings?
10) Which day of the week has the best Average Ratings per Branch?

Walmart Sales Data Analysis (SQL Queries)

-- Create database ---

```sql
CREATE DATABASE IF NOT EXISTS Walmart_sales_BA;
```

-- Create table ---

```sql
CREATE TABLE IF NOT EXISTS Walmart_Data
    (`Invoice Id` VARCHAR(30) 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 5%` FLOAT NOT NULL,
    Total DECIMAL(12 , 4 ) NOT NULL,
    `Date` DATE NOT NULL,
    `Time` TIME NOT NULL,
    Payment VARCHAR(15) NOT NULL,
    Cogs DECIMAL(10 , 2 ) NOT NULL,
    `Gross Margin %` FLOAT,
    `Gross Income` DECIMAL(12 , 4 ),
    rating FLOAT);
```

--- Data file (.CSV) Import from EXCEL --------

-- Data cleaning & Processing ---

```sql
SELECT * FROM walmart_data;
```

-- Add Column “Time of day” ---

```sql
ALTER TABLE walmart_data
ADD COLUMN `Time Of Day` VARCHAR(20);
```

-- Update Values in Time of day Column ----

```sql
UPDATE walmart_data
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);
```

-- Add column “Day name” ----

```
ALTER TABLE walmart_data
ADD COLUMN `Day Name` VARCHAR(10);
```

--- Update Values in Day Name column ----

```
UPDATE walmart_data
SET `Day name` = DAYNAME(date);
```

-- Add Month Name column –

```
ALTER TABLE walmart_data
ADD COLUMN `Month Name` VARCHAR(10);
```

--- Update values in Month Name Column ----

```
UPDATE Walmart_data
SET `Month Name` = MONTHNAME(date);
```

---------- Generic Questions ----------

-- How many unique cities does the data have?

```
SELECT DISTINCT city
FROM walmart_data;
```

city
Yangon
Naypyitaw
Mandalay

-- In which city is each branch?

```
SELECT DISTINCT city, Branch
FROM walmart_data;
```

city	Branch
Yangon	A
Naypyitaw	C
Mandalay	B

---------- Product Analysis ----------

-- What are the Product lines does data have?

```
SELECT DISTINCT `Product line`
FROM walmart_data;
```

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

-- What is the most common Payment method Customer used?

```
SELECT Payment, COUNT(Payment) AS `Number of Payment`
FROM walmart_data
GROUP BY Payment
ORDER BY `Number of Payment` DESC;
```

Payment	Number of Payment
Ewallet	345
Cash	344
Credit card	311

-- What are the most selling product line?

```
SELECT `Product line`, COUNT(Quantity) AS Quantity
FROM walmart_data
GROUP BY `Product line`
ORDER BY Quantity DESC;
```

Product line	Quantity
Fashion accessories	178
Food and beverages	174
Electronic accessories	170
Sports and travel	166
Home and lifestyle	160
Health and beauty	152

-- What is the total revenue by month?

```
SELECT `Month Name`, round(SUM(Total), 2) AS `Total Revenue`
FROM walmart_data
GROUP BY `Month Name`
ORDER BY `Total Revenue` DESC;
```

Month Name	Total Revenue
January	116291.87
March	109455.51
February	97219.37

-- What month had the largest COGS?

```
SELECT `Month Name`, SUM(Cogs) AS Cogs
FROM walmart_data
GROUP BY `Month Name`
ORDER BY Cogs DESC;
```

Month Name	Cogs
January	110754.16
March	104243.34
February	92589.88

-- What product line had the largest revenue?

```
SELECT `Product line`, SUM(Total) AS Revenue
FROM walmart_data
GROUP BY `Product line`
ORDER BY Revenue DESC;
```

Product line	Revenue
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

-- What is the city with the largest revenue?

```
SELECT City, Branch, SUM(Total) AS Revenue
FROM walmart_data
GROUP BY City , Branch
ORDER BY Revenue DESC;
```

City	Branch	Revenue
Naypyitaw	C	110568.7065
Yangon	A	106200.3705
Mandalay	B	106197.6720

-- What product line had the largest VAT?

```
SELECT `Product line`, AVG(`Tax 5%`) AS VAT
FROM walmart_data
GROUP BY `Product line`
ORDER BY VAT DESC;
```

Product line	VAT
Home and lifestyle	16.030331237986683
Sports and travel	15.812629552131677
Health and beauty	15.411572383030466
Food and beverages	15.365310291449228
Electronic accessories	15.220597051872927
Fashion accessories	14.528061806485894

-- 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) AS avg_qnty FROM walmart_data)
        THEN 'Good'
        ELSE 'Bad'
    END AS Remarks
FROM walmart_data
GROUP BY `Product line`;
```

Product Line	Remarks
Food and beverages	Bad
Health and beauty	Good
Sports and travel	Good
Fashion accessories	Bad
Home and lifestyle	Good
Electronic accessories	Good

-- Which branch sold more products than the average product sold?

```sql
SELECT Branch, SUM(Quantity) AS Sold_product
FROM walmart_data
GROUP BY Branch
HAVING Sold_product > (SELECT AVG(Quantity) FROM walmart_data);
```

Branch	Sold_product
A	1859
C	1831
B	1820

-- What is the most common product line by gender

```sql
SELECT Gender, `Product line`, COUNT(Quantity) AS Sold
FROM walmart_data
GROUP BY Gender , `Product line`
ORDER BY Sold DESC;
```

Gender	Product line	Sold
Female	Fashion accessories	96
Female	Food and beverages	90
Male	Health and beauty	88
Female	Sports and travel	88
Male	Electronic accessories	86
Male	Food and beverages	84
Female	Electronic accessories	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

-- What is the average rating of each product line

```sql
SELECT `Product line`, ROUND(AVG(rating), 2) AS avg_rating
FROM walmart_data
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	7
Sports and travel	6.92
Electronic accessories	6.92
Home and lifestyle	6.84

---------- Sales Analysis ----------

-- Number of Sales made in each Time of Day per Weekday?

```
SELECT `Day Name`, `Time Of Day`, COUNT(*) AS total_sales
FROM walmart_data
WHERE `Day Name` = 'Sunday'
GROUP BY `Time Of Day` , `Day Name`
ORDER BY total_sales DESC;
```

Day Name	Time Of Day	total_sales
Sunday	Evening	58
Sunday	Afternoon	53
Sunday	Morning	22

-- Evenings experience most sales and the stores are

-- filled during the evening hours

-- Which customer type brings the most revenue?

```
SELECT `Customer type`, SUM(total) AS total_revenue
FROM walmart_data
GROUP BY `Customer type`
ORDER BY total_revenue DESC;
```

Customer type	total_revenue
Member	164223.4440
Normal	158743.3050

-- Which city has the largest tax/VAT percent (Value add tax)?

```
SELECT city, ROUND(AVG(`Tax 5%`), 2) AS avg_tax_pct
FROM walmart_data
GROUP BY city
ORDER BY avg_tax_pct DESC;
```

city	avg_tax_pct
Naypyitaw	16.05
Mandalay	15.23
Yangon	14.87

-- Which customer type pays the most in VAT?

```
SELECT `Customer type`,
Round(AVG(`Tax 5%`), 2) AS total_tax
FROM walmart_data
GROUP BY `Customer type`
ORDER BY total_tax DESC;
```

Customer type	total_tax
Member	15.61
Normal	15.15

---------- Customer Analysis ----------

-- How many Unique Customer type does the data have?

```
SELECT DISTINCT `Customer type`
FROM walmart_data;
```

Customer type
Normal
Member

-- How many unique payment methods does the data have?

```
SELECT DISTINCT payment
FROM walmart_data;
```

payment
Credit card
Ewallet
Cash

-- What is the most common customer type?

```
SELECT `Customer type`, COUNT(*) AS Purchase
FROM walmart_data
GROUP BY `Customer type`
ORDER BY Purchase DESC;
```

Customer type	Purchase
Member	501
Normal	499

-- Which customer type buys the most?

```
SELECT `Customer type`, COUNT(*) Bought_qnty
FROM walmart_data
GROUP BY `Customer type`
ORDER BY Bought_qnty DESC;
```

Customer type	Bought_qnty
Member	501
Normal	499

-- What is the gender of most of the customers?

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

gender	gender_cnt
Female	501
Male	499

-- What is the gender distribution per branch?

```
SELECT Gender, Branch, COUNT(*) AS Gender_Distribution
FROM walmart_data
WHERE branch = 'A'
GROUP BY Branch , Gender
ORDER BY Gender_Distribution DESC;
```

-- (Branch "A")--

Gender	Branch	Gender_Distribution
Male	A	179
Female	A	161

```sql
SELECT Gender, Branch, COUNT(*) AS Gender_Distribution
FROM walmart_data
WHERE branch = 'B'
GROUP BY Branch , Gender
ORDER BY Gender_Distribution DESC;
```

-- (Branch "B")--

Gender	Branch	Gender_Distribution
Male	B	170
Female	B	162

```sql
SELECT Gender, Branch, COUNT(*) AS Gender_Distribution
FROM walmart_data
WHERE branch = 'C'
GROUP BY Branch, Gender
ORDER BY Gender_Distribution DESC;
```

-- (Branch "C")--

Gender	Branch	Gender_Distribution
Female	C	178
Male	C	150

-- Gender per branch is more or less same hence, I don't think has

-- an effect of the sales per branch and other factors.

-- Which time of day do customers give most Ratings?

```sql
SELECT `Time Of Day`, ROUND(AVG(rating), 2) AS Avg_rating
FROM walmart_data
GROUP BY `Time Of Day`
ORDER BY Avg_rating DESC;
```

Time Of Day	Avg_rating
Afternoon	7.03
Morning	6.96
Evening	6.93

-- Looks like time of the day does not really affect the rating, its

-- more or less same rating each time of the day.

-- Which time of day do customers give most Ratings per Branch?

-- (Branch "A") --

```sql
SELECT Branch, `Time Of Day`, ROUND(AVG(rating), 2) AS Avg_rating
FROM walmart_data
WHERE Branch = 'A'
GROUP BY Branch , `Time Of Day`
ORDER BY avg_rating DESC;
```

Branch	Time Of Day	Avg_rating
A	Afternoon	7.19
A	Morning	7.01
A	Evening	6.89

-- (Branch "B") –

```sql
SELECT Branch, `Time Of Day`, ROUND(AVG(rating), 2) AS Avg_rating
FROM walmart_data
WHERE Branch = 'B'
GROUP BY Branch , `Time Of Day`
ORDER BY avg_rating DESC;
```

Branch	Time Of Day	Avg_rating
B	Morning	6.89
B	Afternoon	6.84
B	Evening	6.77

-- (Branch "C")--

```sql
SELECT Branch, `Time Of Day`, ROUND(AVG(rating), 2) AS Avg_rating
FROM walmart_data
WHERE Branch = 'C'
GROUP BY Branch , `Time Of Day`
ORDER BY avg_rating DESC;
```

Branch	Time Of Day	Avg_rating
C	Evening	7.12
C	Afternoon	7.07
C	Morning	6.97

-- Branch A and C are doing well in ratings, branch B needs to do a

-- little more to get better ratings.

-- Which day of the week has the best AVG Ratings?

```sql
SELECT `Day Name`, ROUND(AVG(rating), 2) AS avg_rating
FROM walmart_data
GROUP BY `Day Name`
ORDER BY avg_rating DESC;
```

Day Name	avg_rating
Monday	7.15
Friday	7.08
Sunday	7.01
Tuesday	7
Saturday	6.9
Thursday	6.89
Wednesday	6.81

-- Monday, Friday and Sunday are the top best days for good ratings

-- why is that the case, how many sales are made on these days?

-- Which day of the week has the best Average Ratings per Branch?

```sql
SELECT `Day Name`, ROUND(AVG(Rating) , 2) as Ratings
FROM walmart_data
WHERE Branch = "A"
GROUP BY `Day Name`
ORDER BY Ratings DESC;
```

Day Name	Ratings
Friday	7.31
Monday	7.1
Sunday	7.08
Tuesday	7.06
Thursday	6.96
Wednesday	6.92
Saturday	6.75

---------- **END** ----------