



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.

Column	Description	Data Type
:-----	:-----	:-----
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(10)
product_line	Product line of the product sold	VARCHAR(100)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT
VAT	The amount of tax on the purchase	FLOAT(6, 4)
total	The total cost of the purchase	DECIMAL(10, 2)
date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIMESTAMP
payment_method	The total amount paid	DECIMAL(10, 2)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_percentage	Gross margin percentage	FLOAT(11, 9)
gross_income	Gross Income	DECIMAL(10, 2)
rating	Rating	FLOAT(2, 1)

PROBLEM STATEMENT

1. Product Analysis

>Conduct analysis on the data to understand the different product lines, the products lines performing best and the product lines that need to be improved.

2. Sales Analysis

>This analysis aims to answer the question of the sales trends of product. The result of this can help use measure the effectiveness of each sales strategy the business applies and what modificatoins are needed to gain more sales.

3. Customer Analysis

>This analysis aims to uncover the different customers segments, purchase trends and the profitability of each customer segment.

Approach Used

1. Data Wrangling: This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace, missing or NULL values.

- > *1. Build a database*
- > *2. Create table and insert the data.*

2. Feature Engineering: This will help use generate some new columns from existing ones.

- > *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.*
- > *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 a new column named `month_name` that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.*

Business Questions To Answer

#Generic Question

1. How many unique cities does the data have?
2. Find The Total Branches Available In This data.
3. Find The Total Sum Of Revenue..?

#Product Question

1. How many unique product lines does the data have?
2. What is the most common payment method?
3. What is the most selling product line?
4. What is the most common product line by gender?
5. What product line had the largest revenue?
6. What product line had the largest VAT?
7. What is the total revenue by month?
8. What month had the largest COGS?
5. What is the city with the largest revenue?
8. Which branch sold more products than average product sold?
9. What is the average rating of each product line?

#Sales Question

1. Number of sales made in each time of the day per weekday
2. Which of the customer types brings the most revenue?
3. Which city has the largest tax percent/ VAT (**Value Added Tax**)?
4. Which customer type pays the most in VAT?

#Customer Question

1. How many unique customer types 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 the day do customers give most ratings?
8. Which time of the day do customers give 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?