**Walmart sales data**

**Purpose of the object :**

The major aim of this project is to gain insight into the sales data of store to understand the different factors that affect sales of the different branches.

**About Data**

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

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 solf 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)

**Analysis List**

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 modifications are needed to gain more sales.

**3. Customer Analysis**

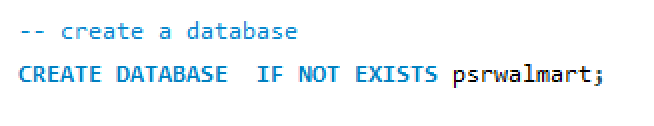
This analysis aims to uncover the different customers segments, purchase trends and the profitability of each customer segment.

1. **Data Wrangling:**

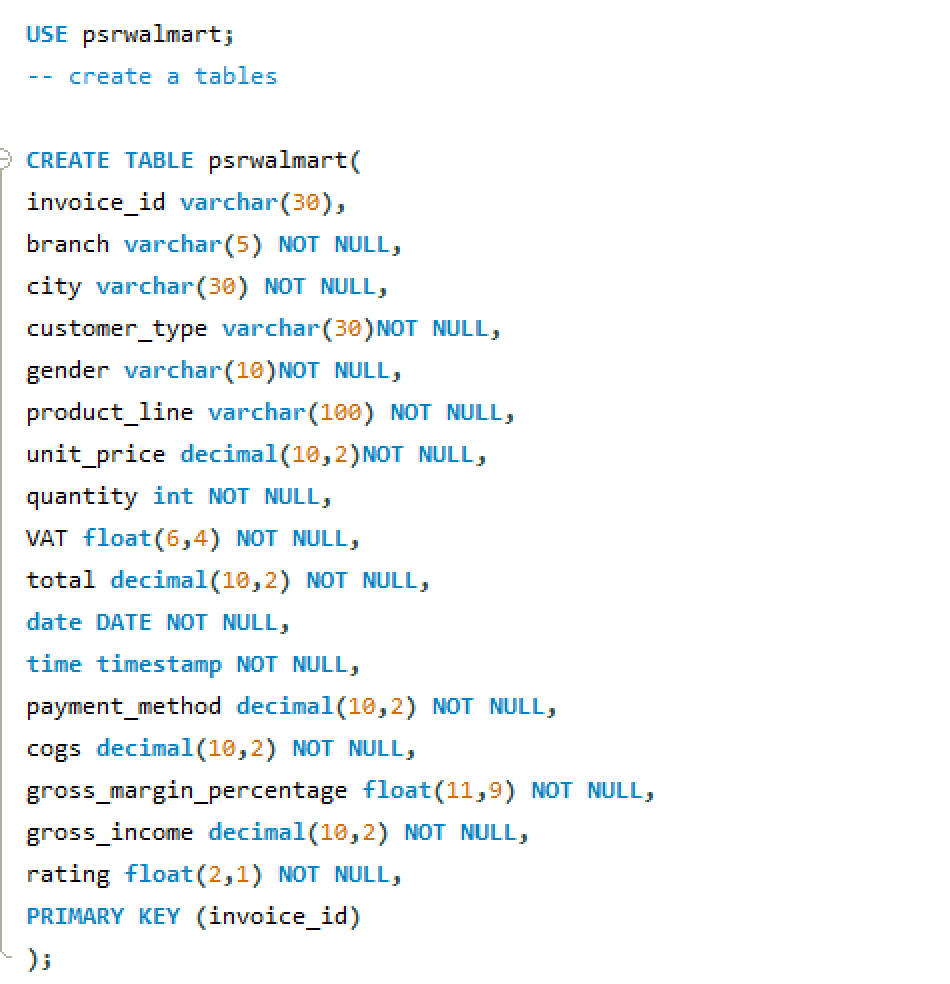
This is the first step where inspection of data is done to make sure NULL values and

missing values are detected and data replacement methods are used to replace, missing or NULL values.

* 1. Build a database

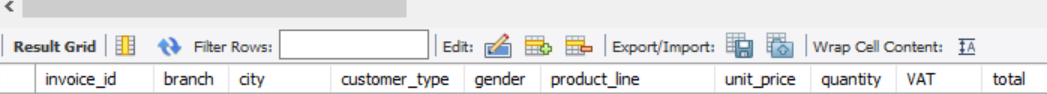


* 1. Create table and insert the data.



**Import Data from CSV**

After creating a database, it’s time to import data from a CSV file. For this select import and a new window will open Select the path from the local machine.

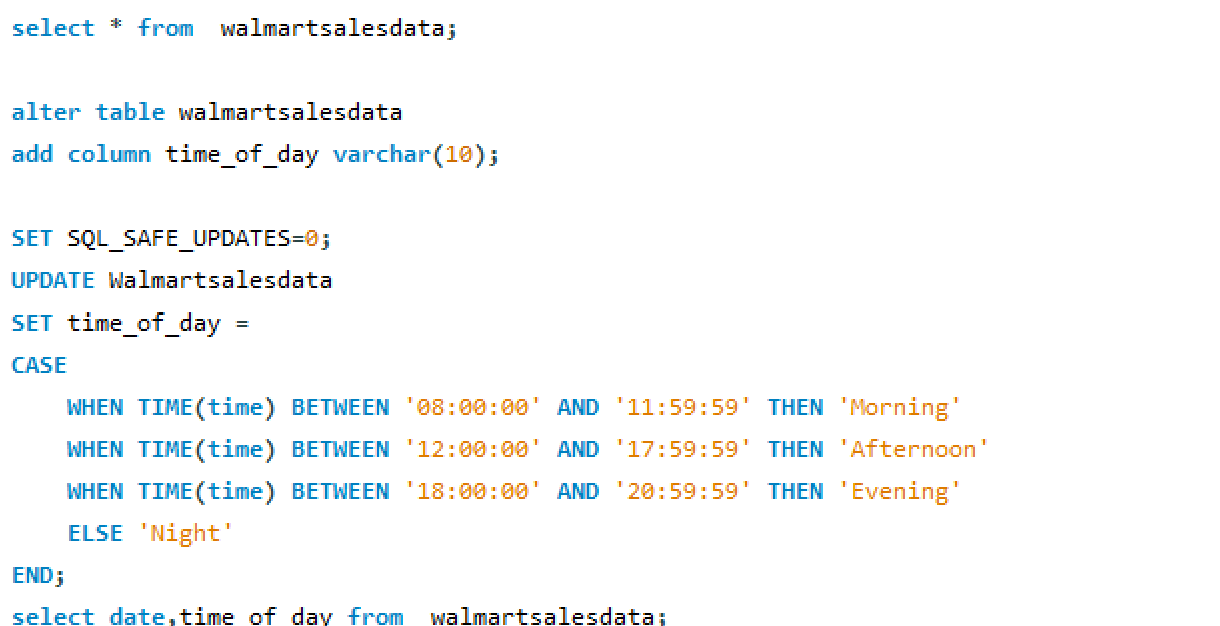


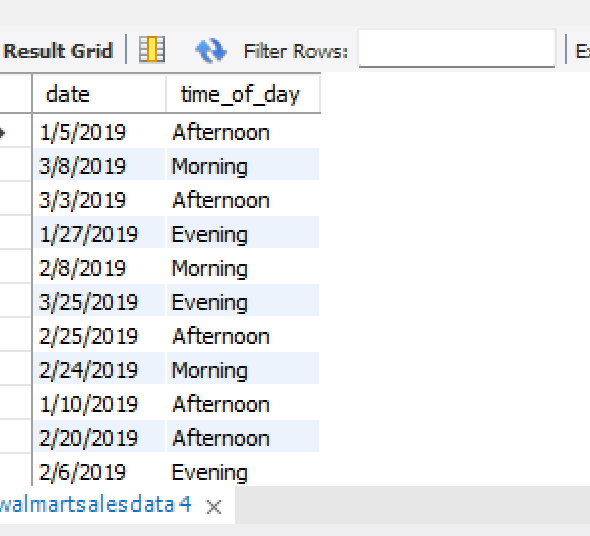
* 1. Select columns with null values in them. There are no null values in our database as in creating the tables, we set NOT NULL for each field, hence null values are filtered out.

1. **Feature Engineering:**

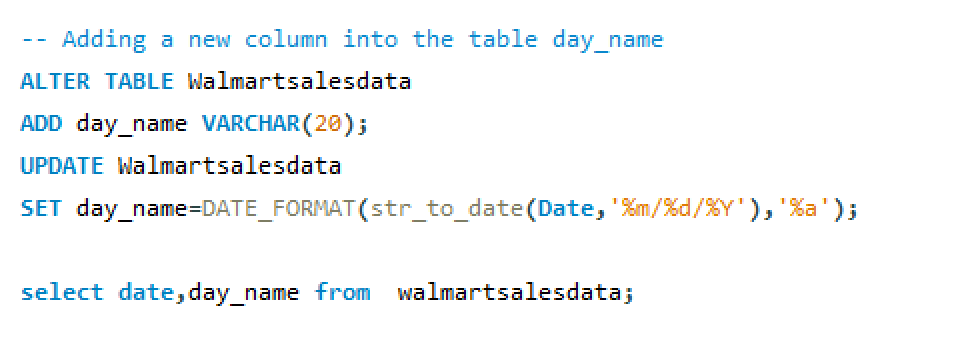
This will help use generate some new columns from existing ones.

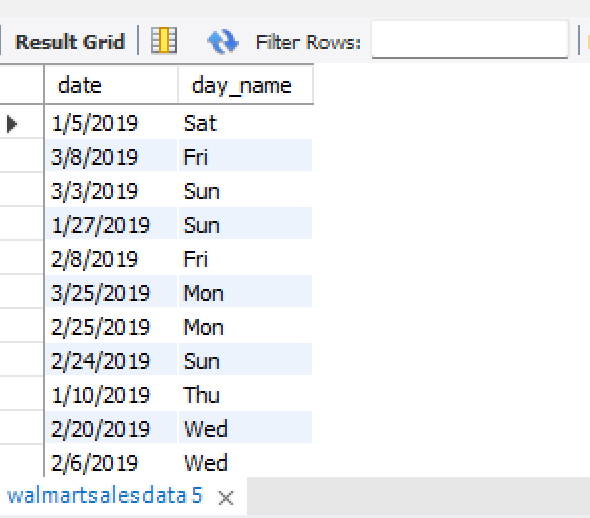
1. Add a new column named time\_of\_day to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.



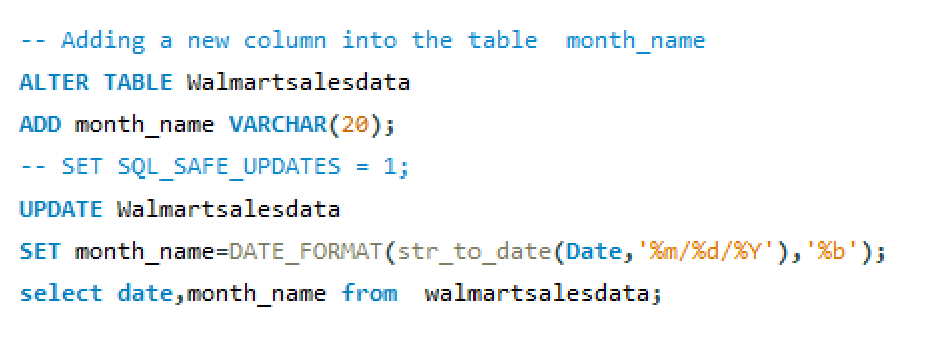


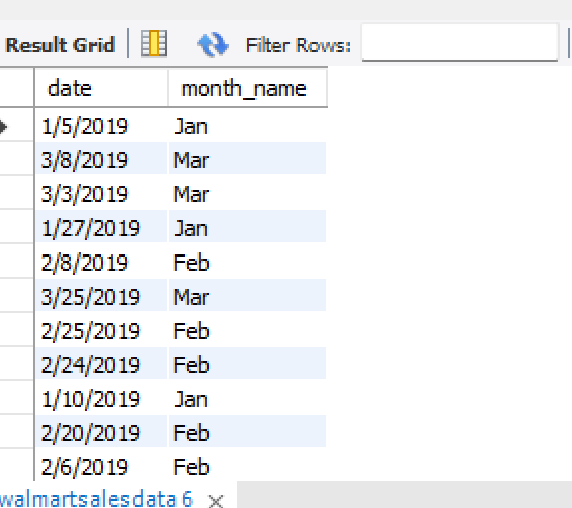
1. Add a new column named day\_name that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.





1. 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.





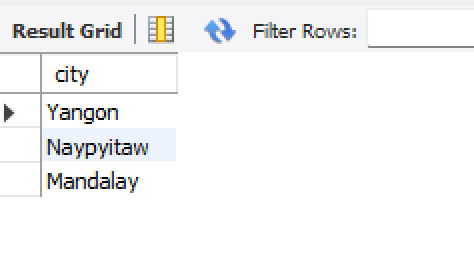
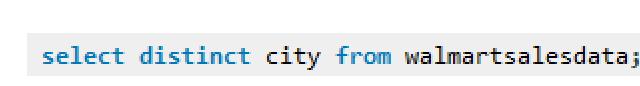
**3. Exploratory Data Analysis (EDA)**

Conducting exploratory data analysis is essential to address the project's listed questions and objectives.

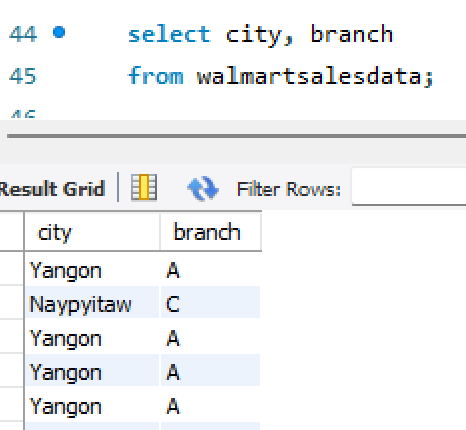
**Business Questions To Answer**

**Generic Question**

1. How many unique cities does the data have?

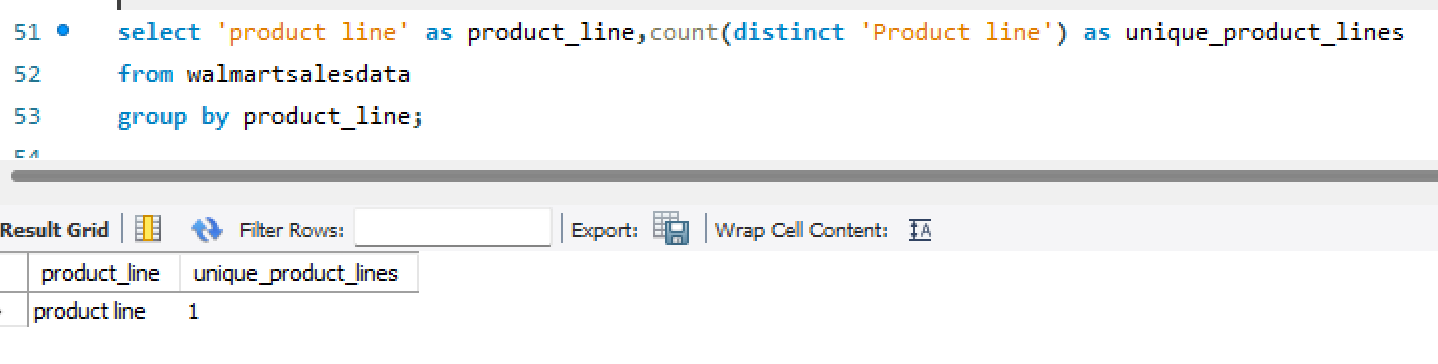


1. In which city is each branch?

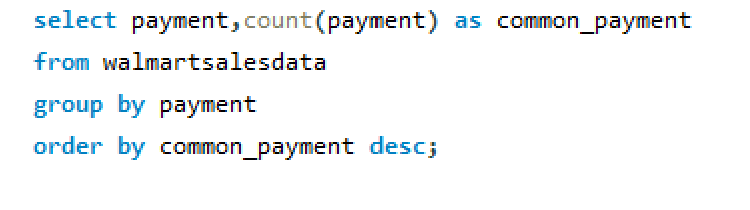


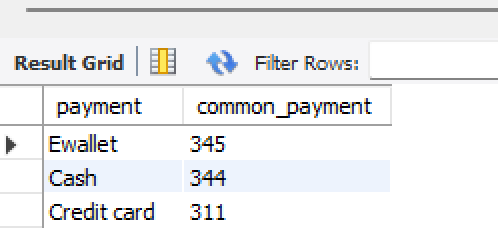
**Product**

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

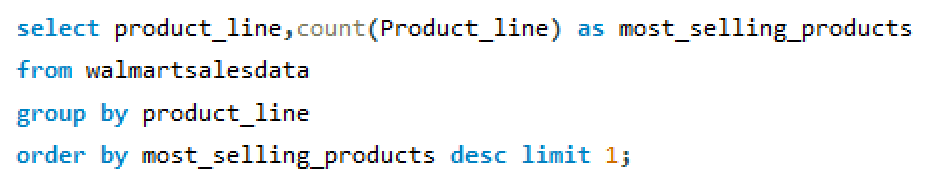


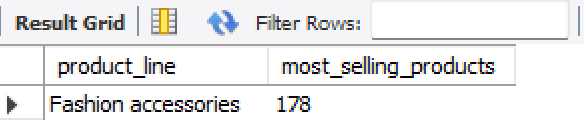
1. What is the most common payment method?



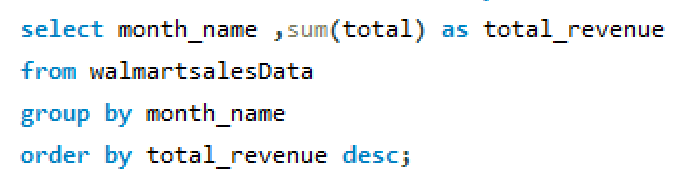


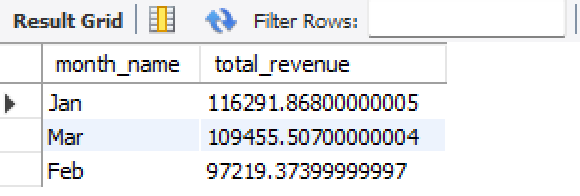
1. What is the most selling product line?



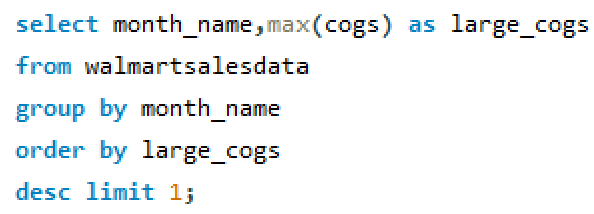


1. What is the total revenue by month?



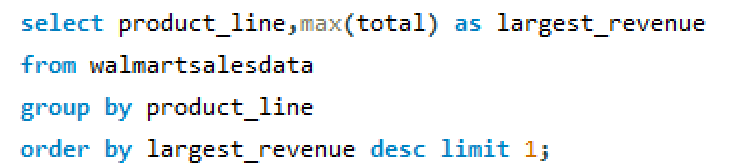


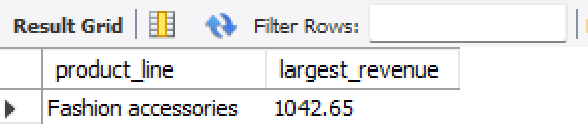
1. What month had the largest COGS?



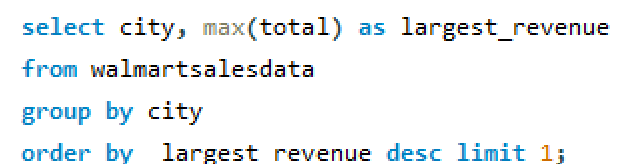


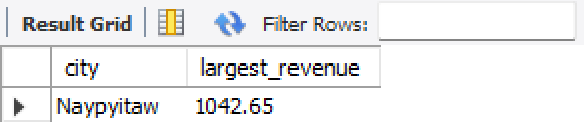
1. What product line had the largest revenue?



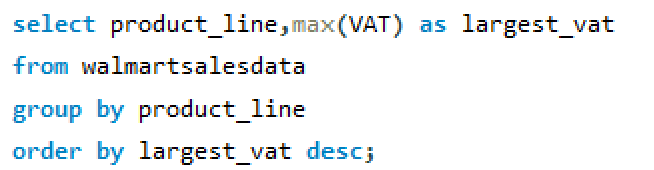


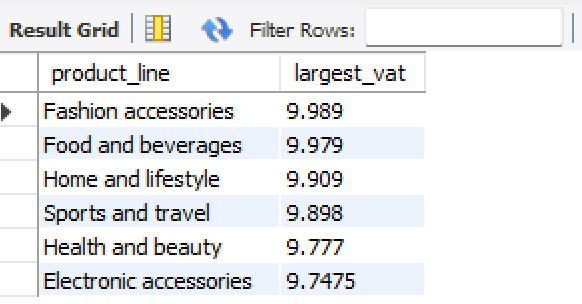
7. What is the city with the largest revenue?



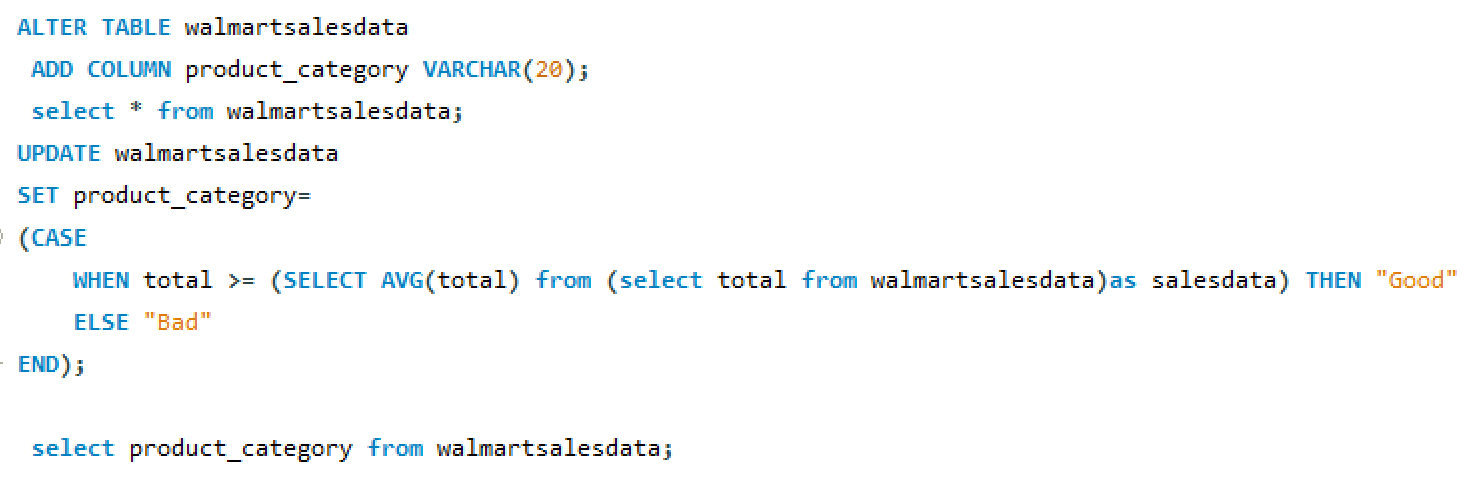


1. What product line had the largest VAT?



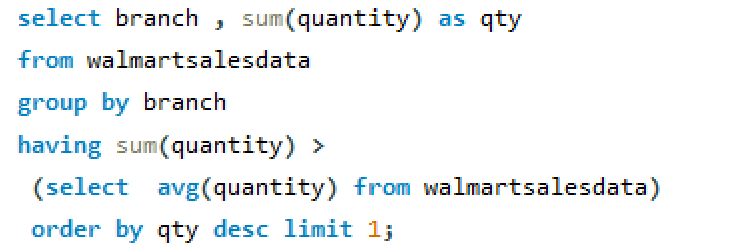


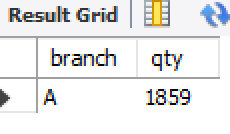
9. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales



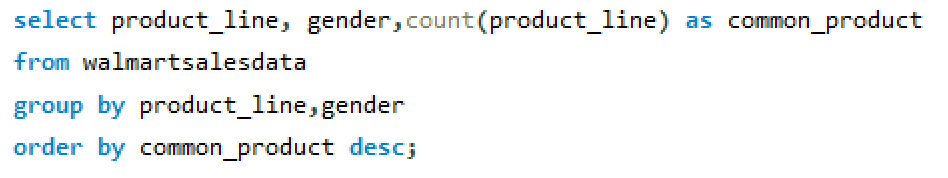


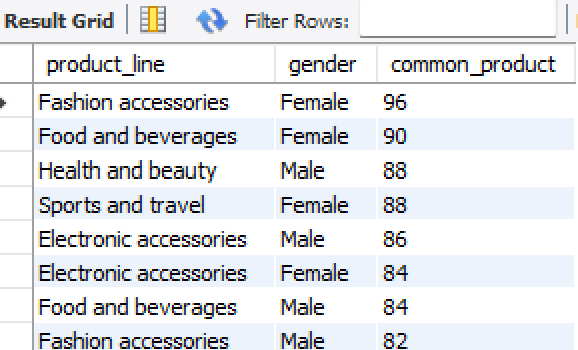
10. Which branch sold more products than average product sold?



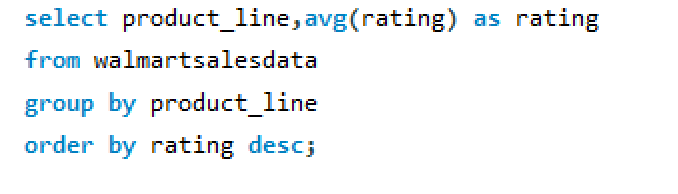


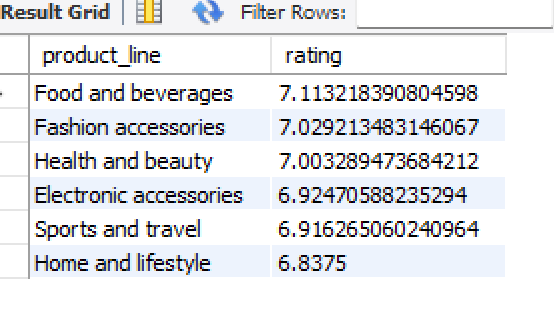
11. What is the most common product line by gender?





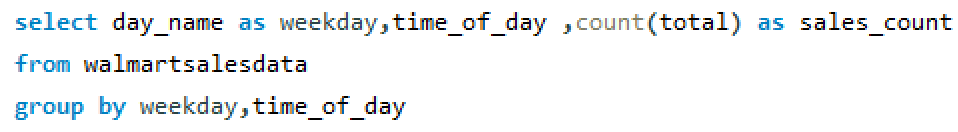
12. What is the average rating of each product line?

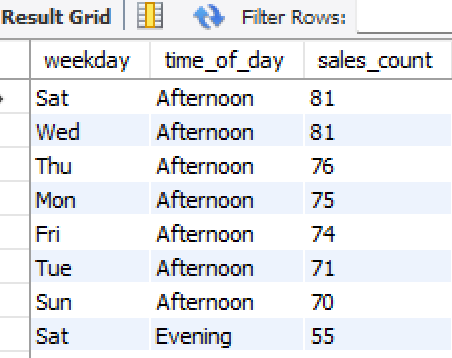




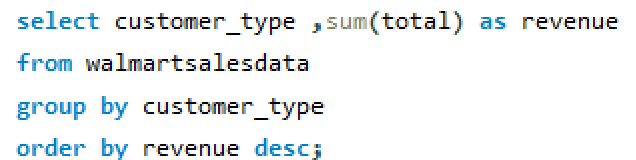
**Sales**

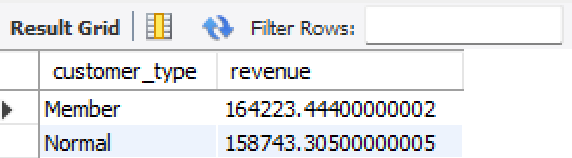
1. Number of sales made in each time of the day per weekday



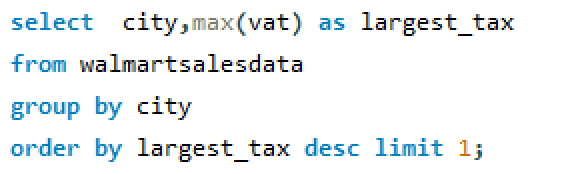


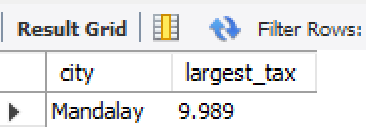
1. Which of the customer types brings the most revenue?



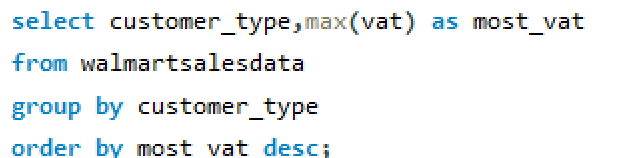


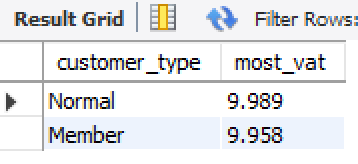
1. Which city has the largest tax percent/ VAT (Value Added Tax)?





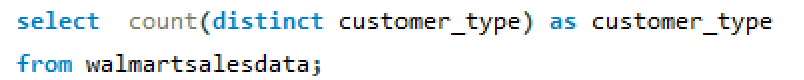
1. Which customer type pays the most in VAT?

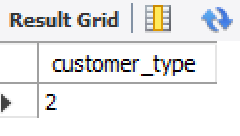




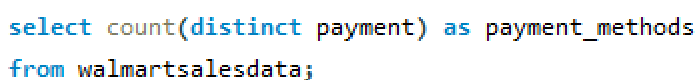
**Customer**

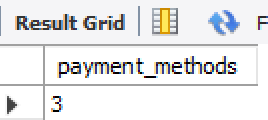
1. How many unique customer types does the data have?



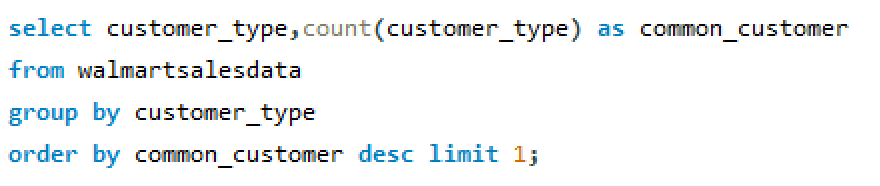


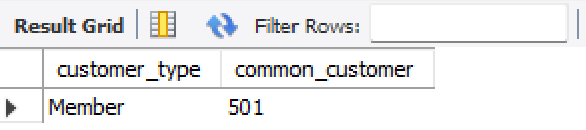
1. How many unique payment methods does the data have?



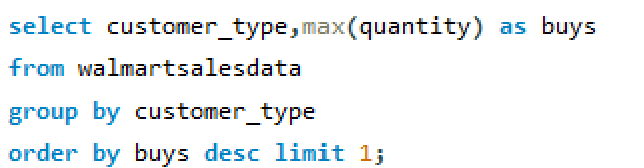


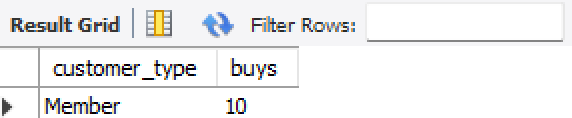
1. What is the most common customer type?



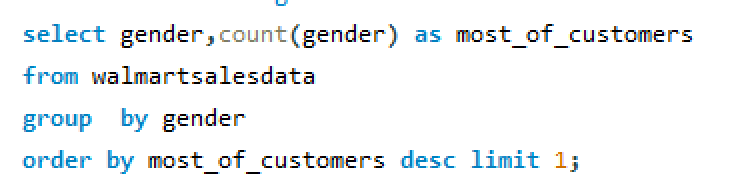


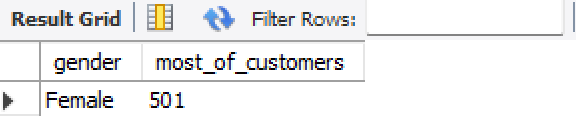
1. Which customer type buys the most?



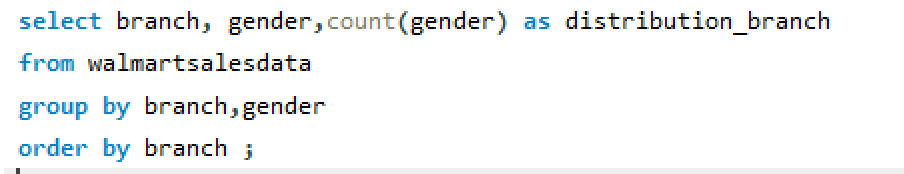


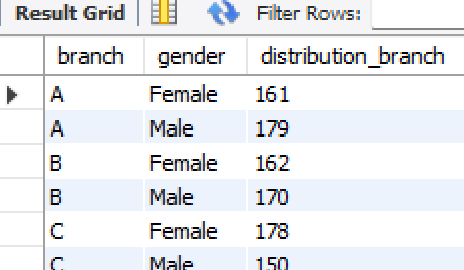
1. What is the gender of most of the customers?



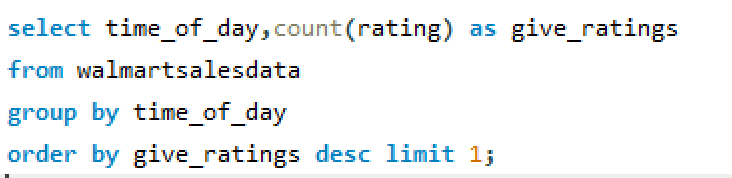


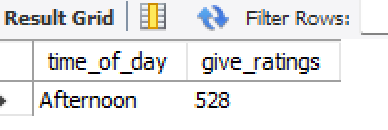
1. What is the gender distribution per branch?



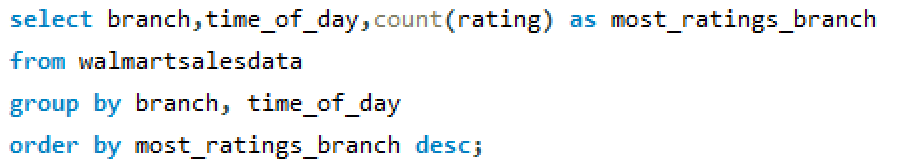


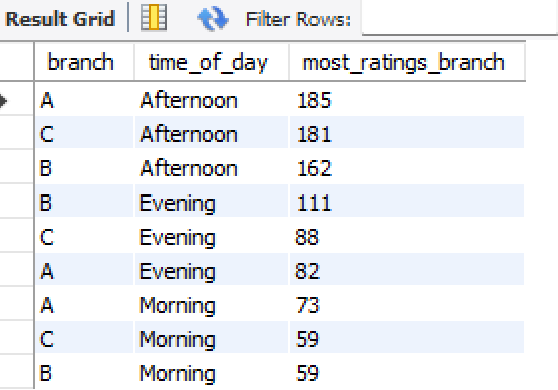
1. Which time of the day do customers give most ratings?



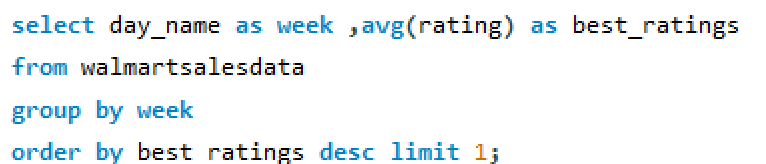


1. Which time of the day do customers give most ratings per branch?



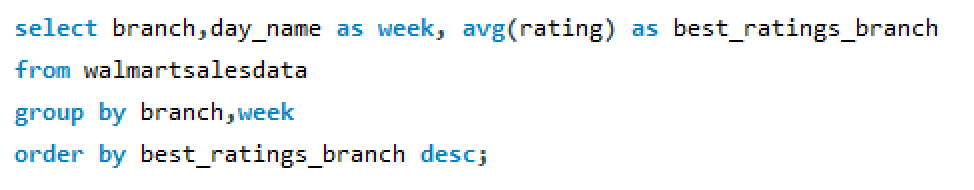


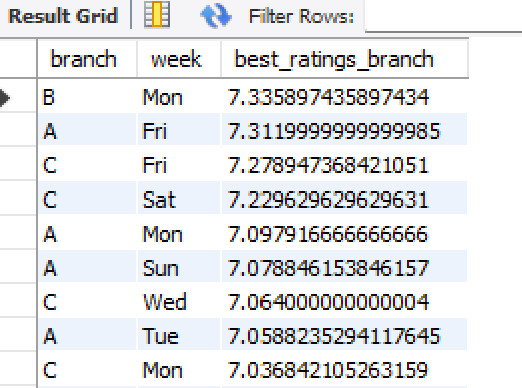
1. Which day of the week has the best avg ratings?





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





Conclusions:

Data analysis is a continuous process, even though this research has yielded insightful information. There's always room for more research and strategy improvement as long as Walmart keeps gathering sales information. The enhancement of Walmart's sales forecasting and operational optimization will be built upon by this initiative in the future.