

```
create database walmart_analysis;  
use walmart_analysis;
```

```
create table walmart(  
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(10) not null,  
Unit_price    decimal(10,2) not null,  
Quantity int not null,  
Tax_5_per float(6,4) not null,  
Total decimal(10,2) not null,  
full_Date date not null,  
full_Time time not null,  
Payment varchar(20) not null,  
cogs decimal(10,2) not null,  
gross_margin_percentage decimal(10,9),  
gross_income float(11,9),  
Rating float(2,1) not null,  
Sub_category varchar(150)  
);
```

```
LOAD DATA INFILE 'walmart.csv'  
INTO TABLE walmart  
CHARACTER SET utf8mb4  
FIELDS TERMINATED BY ','  
ENCLOSED BY '"'  
IGNORE 1 LINES;
```

**\*\*\*\*\* 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.**

**alter table walmart**

**add column time\_of\_day varchar(30);**

**update walmart**

**set time\_of\_day=**

**CASE**

**when `full\_Time` between "00:00:00" And "12:00:00" then "Morning"**

**when `full\_Time` between "12:01:00" And "16:00:00" then "Afternoon"**

**else "Evening"**

**END;**

**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.**

**alter table walmart**

**add column day\_name varchar(30);**

**update walmart**

**set day\_name=dayname(full\_Date);**

**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.**

**Alter Table walmart**

**add column month\_name varchar(30);**

```
update walmart
set month_name=
CASE
When month(full_Date) = 1 then 'January'
When month(full_Date) = 2 then 'Feburary'
When month(full_Date) = 3 then 'March'
When month(full_Date) =4 then 'April'
When month(full_Date) =5 then 'May'
When month(full_Date) =6 then 'June'
When month(full_Date) =7 then 'July'
When month(full_Date) =8 then 'August'
When month(full_Date) =9 then 'September'
When month(full_Date) =10 then 'October'
When month(full_Date) =11 then 'November'
Else 'December'
End;
```

**\*\*\*\*\* Business Questions To Answer \*\*\*\*\***

**# A-GENERIC QUESTION**

**1-Find The Total Sum Of Revenue in this data.**

```
select round(sum(Total),2) as 'Total Revenue' from Walmart;
```

**2. Find The Total Product Lines In this Data.**

```
select count(distinct(Sub_category)) as 'Total Productline' from walmart;
```

**3-How many unique cities does the data have?**

```
select count(distinct(City))as 'Unique City' from walmart;
```

**4-Find The Total Branches Available In This data.**

```
select count(distinct(Branch)) as 'Total Branch' from walmart;
```

## **#B-PRODUCT QUESTION**

**1-What is the most common payment method?**

```
select Payment,count(Payment) as 'Payment Method Total' from walmart  
group by Payment  
order by `Payment Method Total` desc;
```

**2-What is the most selling product line?**

```
select Sub_category,count(Sub_Category) as 'Most Selling Productline' from walmart  
group by Sub_category  
order by `Most Selling Productline` desc;
```

**3-What is the most common product line by gender?**

```
select Sub_Category,Gender,count(Sub_Category)as 'Gender Common Product Line' from  
walmart  
group by Sub_category,Gender  
order by `Sub_category` desc;
```

**4-What product line had the largest revenue?**

```
select Sub_Category,round(sum(Total),1) as 'Product Line With Revenue' from walmart  
group by Sub_category  
order by `Product Line With Revenue` desc;
```

**5-What product line had the largest VAT?**

```
select Sub_category,round(max(Tax_5_Per),1)as 'Product Line With Max Tax' from walmart
```

**group by Sub\_category**  
**order by `Product Line With Max Tax` desc;**

**6-What is the total revenue by month?**

**select month\_name,round(sum(Total),2) as 'Month Revenue' from walmart**  
**group by month\_name**  
**order by `Month Revenue` desc;**

**7-What month had the largest COGS?**

**select month\_name,round(max(cogs),1)as 'Month With Max COGS' from walmart**  
**group by month\_name**  
**order by `Month With Max COGS` desc;**

**8-What is the city with the largest revenue?**

**select city,round(sum(Total),2) as 'City With Revenue' from walmart**  
**group by city**  
**order by `City With Revenue` desc;**

**9-Which branch sold more products than average product sold?**

**select Branch,SUM(Quantity)as 'Avg Product Sold' from walmart**  
**group by Branch**  
**having SUM(Quantity)>(select avg(Quantity) from walmart)**  
**order by `Avg Product Sold` desc;**

**10-What is the average rating of each product line?**

**select Sub\_category,round(avg(Rating),2) as 'Average Rating' from walmart**  
**group by Sub\_category**  
**order by `Average Rating` desc;**

**#C-SALES QUESTION**

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

```
select day_name,time_of_day,count(Total)as 'Sales Made' from walmart  
group by time_of_day,day_name  
order by `day_name`;
```

```
select time_of_day,count(Total)as 'Sales Made In Quantity' from walmart  
where day_name='Saturday'  
group by time_of_day  
order by `Sales Made In Quantity` desc;
```

**2-Which of the customer types brings the most revenue?**

```
select Customer_type,round(sum(Total),2) as 'Cust With Most Revenue' from walmart  
group by Customer_type  
order by `Cust With Most Revenue` desc;
```

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

```
select City,round(avg(Tax_5_per),2) as 'City With Max Tax' from walmart  
group by City  
order by `City With Max Tax` desc;
```

**4-Which customer type pays the most in VAT?**

```
select Customer_type,round(avg(Tax_5_per),2) as 'Customer Type With Most Tax' from  
walmart  
group by Customer_type  
order by `Customer Type With Most Tax` desc;
```

**#D-CUSTOMER QUESTION**

**1-How many unique customer types does the data have?**

**select distinct(Customer\_type) from walmart;**

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

**select distinct(Payment) from walmart;**

**3-What is the most common customer type?**

**4-Which customer type buys the most?**

**select customer\_type,count(Total)as 'Cust With Most Buys' from walmart  
group by `customer\_type`  
order by `Cust With Most Buys` desc;**

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

**select gender,count(gender) as 'Cust Gender Count' from walmart  
group by gender  
order by `Cust Gender Count` desc;**

**6-What is the gender distribution per branch?**

**select Branch,gender,count(gender) as 'Cust Gender Count' from walmart  
group by Branch,gender  
order by `Branch`;**

**7-Which time of the day do customers give most ratings?**

**select time\_of\_day,round(avg(Rating),2)as 'Day Time Rating' from walmart  
group by time\_of\_day  
order by `Day Time Rating` desc;**

**8-Which time of the day do customers give most ratings per branch?**

**select time\_of\_day,Branch,round(avg(Rating),2)as 'Day Time Rating' from walmart**

```
group by time_of_day,Branch  
order by `Branch`;
```

**9-Which day of the week has the best avg ratings?**

```
select day_name,round(avg(Rating),2)as 'Best Avg_Rating' from walmart  
group by day_name  
order by `Best Avg_Rating` desc;
```

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

```
select day_name,Branch,round(avg(Rating),2)as 'Avg_Rating' from walmart  
group by day_name,Branch  
order by day_name;
```