```
-- 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)
);

-- Data cleaning
SELECT
      *
FROM sales;


-- Add the time_of_day column
SELECT
      time,
      (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) AS time_of_day
FROM sales;


ALTER TABLE sales ADD COLUMN time_of_day VARCHAR(20);

-- For this to work turn off safe mode for update
-- Edit > Preferences > SQL Edito > scroll down and toggle safe mode
-- Reconnect to MySQL: Query > Reconnect to server
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
);


-- Add day_name column
SELECT
      date,
      DAYNAME(date)
FROM sales;

ALTER TABLE sales ADD COLUMN day_name VARCHAR(10);

UPDATE sales
SET day_name = DAYNAME(date);


-- Add month_name column
SELECT
      date,
      MONTHNAME(date)
FROM sales;

ALTER TABLE sales ADD COLUMN month_name VARCHAR(10);

UPDATE sales
SET month_name = MONTHNAME(date);

-- --------------------------------------------------------------------
-- ---------------------------- Generic ------------------------------
-- --------------------------------------------------------------------
-- How many unique cities does the data have?
SELECT
      DISTINCT city
FROM sales;

-- In which city is each branch?
SELECT
      DISTINCT city,
    branch
FROM sales;

-- --------------------------------------------------------------------
-- ---------------------------- Product -------------------------------
-- --------------------------------------------------------------------

-- How many unique product lines does the data have?
SELECT
      DISTINCT product_line
FROM sales;


-- What is the most selling product line
SELECT
```

```
      SUM(quantity) as qty,
    product_line
FROM sales
GROUP BY product_line
ORDER BY qty DESC;

-- What is the most selling product line
SELECT
      SUM(quantity) as qty,
    product_line
FROM sales
GROUP BY product_line
ORDER BY qty DESC;

-- 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;


-- What month had the largest COGS?
SELECT
      month_name AS month,
      SUM(cogs) AS cogs
FROM sales
GROUP BY month_name
ORDER BY cogs;


-- 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;

-- What is the city with the largest revenue?
SELECT
      branch,
      city,
      SUM(total) AS total_revenue
FROM sales
GROUP BY city, branch
ORDER BY total_revenue;


-- 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;


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

SELECT
      AVG(quantity) AS avg_qnty
FROM sales;

SELECT
      product_line,
      CASE
                  WHEN AVG(quantity) > 6 THEN "Good"
        ELSE "Bad"
    END AS remark
FROM sales
GROUP BY product_line;


-- Which branch sold more products than average product sold?
SELECT
      branch,
    SUM(quantity) AS qnty
FROM sales
GROUP BY branch
HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);


-- What is the most common product line by gender
SELECT
      gender,
    product_line,
    COUNT(gender) AS total_cnt
FROM sales
GROUP BY gender, product_line
ORDER BY total_cnt DESC;

-- What is the average rating of each product line
SELECT
      ROUND(AVG(rating), 2) as avg_rating,
    product_line
FROM sales
GROUP BY product_line
ORDER BY avg_rating DESC;

-- --------------------------------------------------------------------
-- --------------------------------------------------------------------

-- --------------------------------------------------------------------
-- -------------------------- Customers -------------------------------
-- --------------------------------------------------------------------
```

```sql
-- How many unique customer types does the data have?
SELECT
      DISTINCT customer_type
FROM sales;

-- How many unique payment methods does the data have?
SELECT
      DISTINCT payment
FROM sales;


-- What is the most common customer type?
SELECT
      customer_type,
      count(*) as count
FROM sales
GROUP BY customer_type
ORDER BY count DESC;

-- Which customer type buys the most?
SELECT
      customer_type,
    COUNT(*)
FROM sales
GROUP BY customer_type;


-- 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;

-- What is the gender distribution per branch?
SELECT
      gender,
      COUNT(*) as gender_cnt
FROM sales
WHERE branch = "C"
GROUP BY gender
ORDER BY gender_cnt DESC;
-- Gender per branch is more or less the same hence, I don't think has
-- an effect of the sales per branch and other factors.

-- 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;
```

```
-- Looks like time of the day does not really affect the rating, its
-- more or less the same rating each time of the day.alter


-- 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;
-- Branch A and C are doing well in ratings, branch B needs to do a
-- little more to get better ratings.


-- 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;
-- Mon, Tue and Friday 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?
SELECT
      day_name,
      COUNT(day_name) total_sales
FROM sales
WHERE branch = "C"
GROUP BY day_name
ORDER BY total_sales DESC;


-- --------------------------------------------------------------------
-- --------------------------------------------------------------------

-- --------------------------------------------------------------------
-- ---------------------------- Sales ---------------------------------
-- --------------------------------------------------------------------

-- Number of sales made in each time of the day per weekday
SELECT
      time_of_day,
      COUNT(*) AS total_sales
FROM sales
WHERE day_name = "Sunday"
GROUP BY time_of_day
ORDER BY total_sales DESC;
-- Evenings experience most sales, the stores are
-- filled during the evening hours

-- 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;

-- Which city has the largest tax/VAT percent?
SELECT
      city,
    ROUND(AVG(tax_pct), 2) AS avg_tax_pct
FROM sales
GROUP BY city
ORDER BY avg_tax_pct DESC;

-- Which customer type pays the most in VAT?
SELECT
      customer_type,
      AVG(tax_pct) AS total_tax
FROM sales
GROUP BY customer_type
ORDER BY total_tax;
```