```sql
CREATE database amazon;
use amazon

create table amazon_sales (
invoice_id varchar(30) primary key not null,
branch varchar(5) not null,
city varchar(30) not null,
customer_type varchar(30) not null,
gender varchar(10) 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(10,2) not null,
date date not null,
time time not null,
payment varchar(20) not null,
cogs decimal(10,2) not null,
gross_margin_percentage float not null,
gross_income decimal(10,2) not null,
rating decimal(3,1) not null,
);

-- Add new columns
ALTER TABLE amazon_sales
ADD COLUMN timeofday VARCHAR(15),
ADD COLUMN dayname VARCHAR(15),
ADD COLUMN monthname VARCHAR(15);

-- Update timeofday
UPDATE amazon_sales
SET timeofday =
  CASE
    when time BETWEEN '05:00:00' AND '11:59:59' THEN 'Morning'
    WHEN time BETWEEN '12:00:00' AND '16:59:59' THEN 'Afternoon'
    WHEN time BETWEEN '17:00:00' AND '20:59:59' THEN 'Evening'
    ELSE 'Night'
  END;

-- Update dayname and monthname
UPDATE amazon_sales
SET dayname = DAYNAME(date),
    monthname = MONTHNAME(date);

-- Exploratory Data Analysis (EDA) — Business Questions with SQL

-- 1.What is the count of distinct cities in the dataset?
SELECT COUNT(DISTINCT city) AS distinct_cities
FROM amazon_sales;

-- 2.For each branch, what is the corresponding city?
SELECT DISTINCT branch, city
FROM amazon_sales
ORDER BY branch;

-- 3.What is the count of distinct product lines in the dataset?
SELECT COUNT(DISTINCT product_line) AS total_product_lines
FROM amazon_sales;
```

```sql
-- 4.Which payment method occurs most frequently?
SELECT payment, COUNT(*) AS count
FROM amazon_sales
GROUP BY payment
ORDER BY count DESC;

-- 5.Which product line has the highest sales?
SELECT product_line, SUM(quantity) AS total_sales
FROM amazon_sales
GROUP BY product_line
ORDER BY total_sales DESC;

-- 6.How much revenue is generated each month?
SELECT monthname AS month, SUM(total) AS monthly_revenue
FROM amazon_sales
GROUP BY monthname
ORDER BY monthly_revenue desc;

-- 7.In which month did the cost of goods sold reach its peak?
SELECT monthname, SUM(cogs) AS total_cogs
FROM amazon_sales
GROUP BY monthname
ORDER BY total_cogs DESC;

-- 8.Which product line generated the highest revenue?
SELECT product_line, SUM(total) AS total_revenue
FROM amazon_sales
GROUP BY product_line
ORDER BY total_revenue DESC;

-- 9.In which city was the highest revenue recorded?
SELECT city, SUM(total) AS total_revenue
FROM amazon_sales
GROUP BY city
ORDER BY total_revenue DESC;

-- 10.Which product line incurred the highest Value Added Tax?
SELECT product_line, SUM(tax_5) AS total_tax
FROM amazon_sales
GROUP BY product_line
ORDER BY total_tax DESC;

-- 11.For each product line, add a column indicating "Good" if its
sales are above average, otherwise "Bad."
SELECT
    product_line,
    SUM(total) AS total_sales,
    CASE
        WHEN SUM(total) > (SELECT AVG(total) FROM amazon_sales) THEN
'Good'
        ELSE 'Bad'
    END AS performance
FROM amazon_sales
GROUP BY product_line;

-- 12.Identify the branch that exceeded the average number of products
sold.
```

```sql
SELECT
    branch,
    SUM(quantity) AS total_products_sold
FROM amazon_sales
GROUP BY branch
HAVING total_products_sold > (SELECT AVG(quantity) FROM amazon_sales);


-- 13.Which product line is most frequently associated with each
gender?
SELECT gender, product_line, COUNT(*) AS count
FROM amazon_sales
GROUP BY gender, product_line
HAVING COUNT(*) = (
  SELECT MAX(cnt)
  FROM (
      SELECT gender AS g, COUNT(*) AS cnt
      FROM amazon_sales AS a2
      WHERE a2.gender = amazon_sales.gender
      GROUP BY product_line
  ) AS subquery
);


-- 14.Calculate the average rating for each product line.
SELECT product_line, ROUND(AVG(rating),2) AS avg_rating
FROM amazon_sales
GROUP BY product_line
ORDER BY avg_rating DESC;


-- 15.Count the sales occurrences for each time of day on every
weekday.
SELECT dayname, timeofday, COUNT(*) AS total_sales
FROM amazon_sales
GROUP BY dayname, timeofday
ORDER BY dayname, total_sales DESC;


-- 16.Identify the customer type contributing the highest revenue.
SELECT customer_type, SUM(total) AS total_revenue
FROM amazon_sales
GROUP BY customer_type
ORDER BY total_revenue DESC;


-- 17.Determine the city with the highest VAT percentage.
SELECT city, ROUND(AVG(tax_5),4) AS avg_tax
FROM amazon_sales
GROUP BY city
ORDER BY avg_tax DESC;


-- 18.Identify the customer type with the highest VAT payments.
SELECT customer_type, ROUND(SUM(tax_5),4) AS total_tax
FROM amazon_sales
GROUP BY customer_type
ORDER BY total_tax DESC;


-- 19.What is the count of distinct customer types in the dataset?
SELECT COUNT(DISTINCT customer_type) AS total_customer_types
FROM amazon_sales;


-- 20.What is the count of distinct payment methods in the dataset?
```

```sql
SELECT COUNT(DISTINCT payment) AS total_payment_methods
FROM amazon_sales;

-- 21.Which customer type occurs most frequently?
SELECT customer_type, COUNT(*) AS count
FROM amazon_sales
GROUP BY customer_type
ORDER BY count DESC;

-- 22.Identify the customer type with the highest purchase frequency.
SELECT customer_type, SUM(quantity) AS total_items
FROM amazon_sales
GROUP BY customer_type
ORDER BY total_items DESC;

-- 23.Determine the predominant gender among customers.
SELECT gender, COUNT(*) AS total_customers
FROM amazon_sales
GROUP BY gender
ORDER BY total_customers DESC;

-- 24.Examine the distribution of genders within each branch.
SELECT branch, gender, COUNT(*) AS count
FROM amazon_sales
GROUP BY branch, gender
ORDER BY branch;

-- 25.Identify the time of day when customers provide the most ratings.
SELECT timeofday, COUNT(rating) AS rating_count
FROM amazon_sales
GROUP BY timeofday
ORDER BY rating_count DESC;

-- 26.Determine the time of day with the highest customer ratings for
each branch.
SELECT branch, timeofday, ROUND(AVG(rating),2) AS average_rating
FROM amazon_sales
GROUP BY branch, timeofday
ORDER BY branch, average_rating DESC;

-- 27.Identify the day of the week with the highest average ratings.
SELECT dayname, ROUND(AVG(rating),2) AS avg_rating
FROM amazon_sales
GROUP BY dayname
ORDER BY avg_rating DESC;

-- 28.Determine the day of the week with the highest average ratings
for each branch.
SELECT branch, dayname, ROUND(AVG(rating),2) AS average_rating
FROM amazon_sales
GROUP BY branch, dayname
ORDER BY branch, average_rating DESC;

-- Highest Revenue Gender:
SELECT gender ,sum(total) as total_revenue
FROM amazon_sales
GROUP BY gender
ORDER BY total_revenue DESC;
```

```sql
-- sales of males and females
SELECT gender,SUM(quantity) AS total_units_sold
FROM amazon_sales
GROUP BY gender
ORDER BY total_units_sold DESC;

-- Distribution Of Members Based On Gender
SELECT gender,COUNT(*) AS member_count
FROM amazon_sales
WHERE customer_type = 'Member'
GROUP BY gender
ORDER BY member_count DESC;

-- Peak Sales Day of week
SELECT dayname,sum(total) as sales
FROM amazon_sales
GROUP BY dayname
ORDER BY sales DESC;
```

-- Key Findings from Amazon Sales Dataset

Product Analysis

-- Highest Sales Product Line: Electronic accessories - 971units sold7

-- Highest Revenue Product Line: Food and beverages 56144.96

-- Lowest Sales Product Line: Health and Beauty - 854 units sold

-- Lowest Revenue Product Line: Health and Beauty — 49193.84

Customer Analysis

-- Most Predominant Gender: Female - 501

-- Most Predominant Customer Type: Member -501

-- Highest Revenue Gender: Female-167883.26

-- Highest Revenue Customer Type: Member - 164223.81

-- Most Popular Product Line (Male): Health and Beauty - 88

-- Most Popular Product Line (Female): Fashion Accessories - 96

-- Distribution Of Members Based On Gender: Female-261 & Male-240

-- Sales Male: 2641 units

-- Sales Female: 2869 units

Sales Analysis:
-- Month With Highest Revenue: January-116292.11

-- City & Branch With Highest Revenue: city:Naypyitaw branch:C -
110568.86

```
-- Month With Lowest Revenue: February  97219.58

-- City & Branch With Lowest Revenue: city:Mandalay branch:B -
106198.00

-- Peak Sales Time Of Day: Afternoon

-- Peak Sales Day of week: Saturday     56120.86
```