**Exploratory Data Analysis**

**Business Questions To Answer**

Generic Question

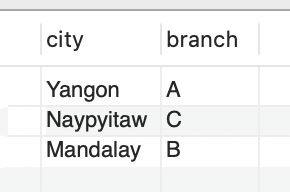
1. How many unique cities does the data have?

SELECT COUNT(DISTINCT city) FROM sales;

1. In which city is each branch?

SELECT DISTINCT city,

branch FROM sales;

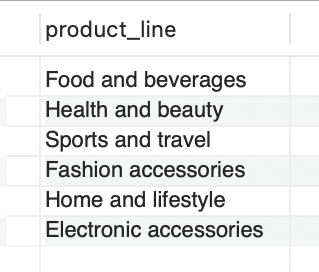


Product

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

SELECT DISTINCT product\_line

FROM sales;



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

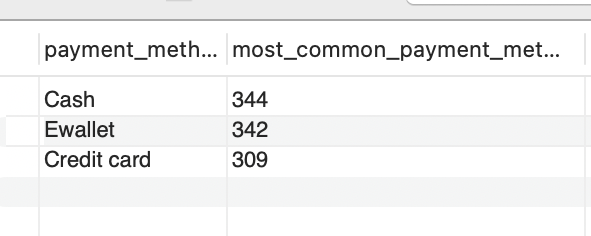
SELECT payment\_method,

COUNT(payment\_method) AS most\_common\_payment\_method

FROM sales

GROUP BY payment\_method

ORDER BY COUNT(payment\_method) DESC;



1. **What is the most selling product line?**

SELECT product\_line,

SUM(quantity) AS product\_line\_sales

FROM sales

GROUP BY product\_line

ORDER BY SUM(quantity) DESC;



1. **What is the total revenue by month?**

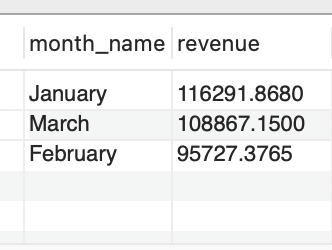
SELECT month\_name,

SUM(total) AS revenue

FROM sales

GROUP BY month\_name

ORDER BY SUM(total) DESC;



1. **What month had the highest COGS?**

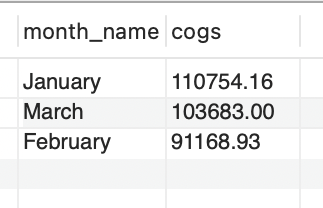
SELECT month\_name,

SUM(cogs) AS cogs

FROM sales

GROUP BY month\_name

ORDER BY SUM(cogs) DESC;



1. **What product line had the highest revenue?**

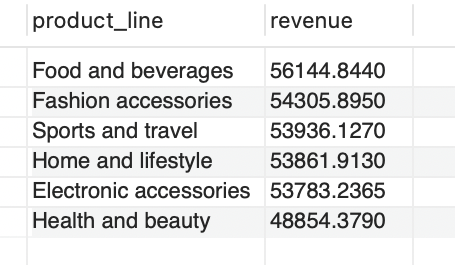
SELECT product\_line,

SUM(total) AS revenue

FROM sales

GROUP BY product\_line

ORDER BY SUM(total) DESC;



1. **What is the city with the highest revenue?**

SELECT city,

SUM(total) AS revenue

FROM sales

GROUP BY city

ORDER BY SUM(total) DESC;



1. **What product line had the highest VAT?**

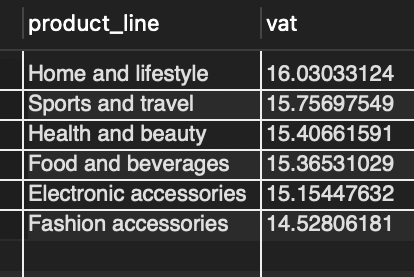
SELECT product\_line,

AVG(vat) AS vat

FROM sales

GROUP BY product\_line

ORDER BY AVG(vat) DESC;



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

SELECT

product\_line,

ROUND(AVG(total),2) AS avg\_sales,

(CASE

WHEN AVG(total) > (SELECT AVG(total) FROM sales) THEN "Good"

ELSE "Bad"

END)

AS Remarks

FROM sales

GROUP BY product\_line;



1. **Which branch sold more products than average product sold?**

SELECT

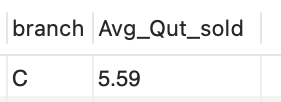
branch,

ROUND(AVG(quantity),2) AS Avg\_Qut\_sold

FROM sales

GROUP BY branch

HAVING AVG(quantity) > (SELECT AVG(quantity) FROM sales);



1. **What is the most common product line by gender?**

SELECT product\_line,

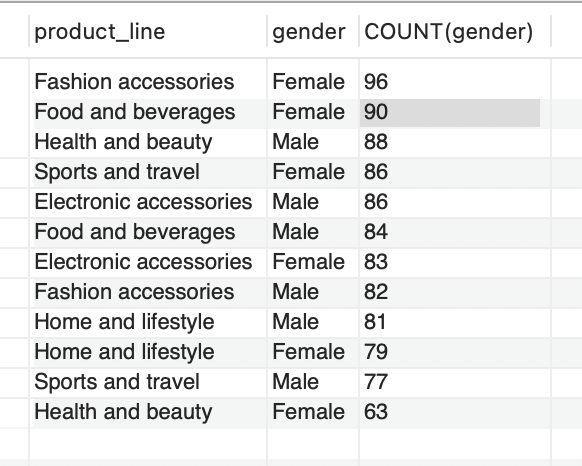
gender,

COUNT(gender)

FROM sales

GROUP BY product\_line, gender

ORDER BY COUNT(gender) DESC;



1. **What is the average rating of each product line?**

SELECT product\_line,

ROUND(AVG(rating),2) AS Rating

FROM sales

GROUP BY product\_line



1. **What product line has highest profit?**

SELECT product\_line,

ROUND(SUM(gross\_income),2) AS Profit

FROM sales

GROUP BY product\_line

ORDER BY profit DESC;



**Sales**

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

SELECT

time\_of\_day,

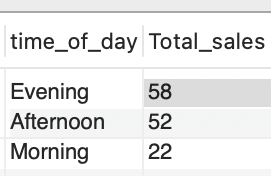
COUNT(\*) AS Total\_sales

FROM sales

WHERE day\_name = "Sunday"

GROUP BY time\_of\_day

ORDER BY Total\_sales DESC;



**Note – By changing day\_name we can get total sales count for different days.**

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

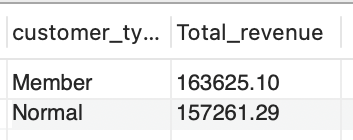
SELECT customer\_type,

SUM(total) AS Total\_revenue

FROM sales

GROUP BY customer\_type

ORDER BY Total\_revenue DESC;



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

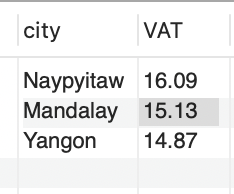
SELECT city,

ROUND(AVG(VAT),2) AS VAT

FROM sales

GROUP BY city

ORDER BY VAT DESC;



1. **Which customer type pays the most in VAT?**

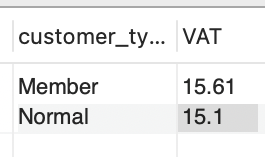
SELECT customer\_type,

ROUND(AVG(VAT),2) AS VAT

FROM sales

GROUP BY customer\_type

ORDER BY VAT DESC;

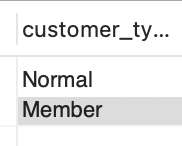


### Customer

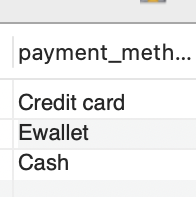
1. **How many unique customer types does the data have?**

SELECT DISTINCT customer\_type

FROM sales;



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



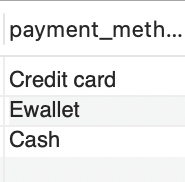
1. **What is the most common customer type?**

SELECT customer\_type,

COUNT(\*) AS COUNT

FROM sales

GROUP BY customer\_type;



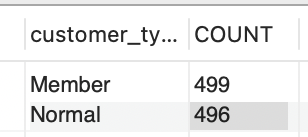
1. **Which customer type buys the most?**

SELECT customer\_type,

COUNT(\*) AS COUNT

FROM sales

GROUP BY customer\_type;



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

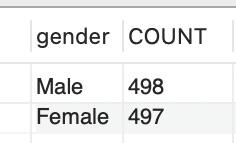
SELECT gender,

COUNT(\*) AS COUNT

FROM sales

GROUP BY gender

ORDER BY COUNT DESC;



1. **What is the gender distribution per branch?**

SELECT branch,

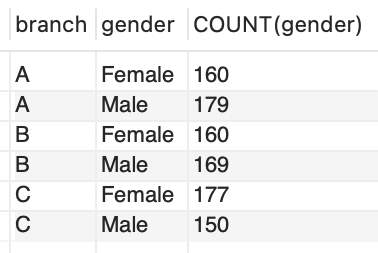
gender,

COUNT(gender)

FROM sales

Group BY branch, gender

ORDER by branch;



1. **What type of customer gives better ratings?**

SELECT customer\_type,

ROUND(AVG(rating),2) AS rating

FROM sales

GROUP BY customer\_type

