**EDA Questions**

-- Total number of sales

SELECT COUNT(\*) FROM sales

A screenshot of a computer

Description automatically generated

-- Top 5 customers (in terms of sales revenue)

SELECT s.customer\_id as customer, SUM(round(sales::numeric,2)) as sales

FROM sales s LEFT JOIN customers c USING (customer\_id)

GROUP BY s.customer\_id

ORDER BY SUM(sales::numeric) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

-- Top 5 customers (in terms of # of sales transactions)

SELECT s.customer\_id as customer, COUNT(sales) as no\_of\_sales

FROM sales s LEFT JOIN customers c USING (customer\_id)

GROUP BY s.customer\_id

ORDER BY COUNT(sales) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

-- Top 5 products (in terms of sales revenue)

SELECT p.description as product, SUM(sales::numeric) as sales

FROM SALES s JOIN PRODUCTS p USING (stock\_code)

GROUP BY p.description

ORDER BY SUM(sales::numeric) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

-- Top 5 products (in terms of # of sales transactions)

SELECT p.description as product, COUNT(sales) as no\_of\_sales

FROM SALES s JOIN PRODUCTS p USING (stock\_code)

GROUP BY p.description

ORDER BY COUNT(sales) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

-- What region, state and city had the highest revenue? (Top 5 for state and city)

SELECT region, SUM(sales::numeric) as sales

FROM sales s LEFT JOIN customers c USING (customer\_id)

GROUP BY region

HAVING region IS NOT NULL  --Filtering out guests

ORDER BY SUM(sales) DESC;

Filtering guest customers

A screenshot of a computer

Description automatically generated

Without filtering guest customers, we can see guest customers account for a large number of sales. Guest customers region shows as [null] since we don’t have their information.

A screenshot of a computer

Description automatically generated

SELECT order\_state as state, SUM(sales::numeric) as sales

FROM sales s LEFT JOIN customers c USING (customer\_id)

GROUP BY order\_state

HAVING order\_state IS NOT NULL  --Filtering out guests

ORDER BY SUM(sales) DESC

LIMIT 5;

Filtering guest customers

A screenshot of a computer

Description automatically generated

SELECT order\_city as city, order\_state as state, SUM(round(sales::numeric,2)) as sales

FROM sales s LEFT JOIN customers c USING (customer\_id)

GROUP BY order\_city, order\_state

HAVING order\_city IS NOT NULL  --Filtering out guests

ORDER BY SUM(sales) DESC

LIMIT 5;

Filtering out guest customers

A screenshot of a computer

Description automatically generated

-- Which products got more returns?

SELECT DISTINCT(p.description),

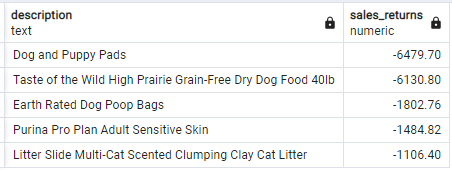
SUM(ROUND(sales::numeric,2)) OVER (PARTITION BY p.description) as sales\_returns

FROM SALES s JOIN PRODUCTS p USING (stock\_code)

WHERE invoice\_no = 'return'

ORDER BY sales\_returns

LIMIT 5;



-- Which days of the month generate more sales?

SELECT day, SUM(sales::numeric) as sales

FROM sales

GROUP BY day

ORDER BY SUM(sales) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

-- Which days of the month generate less sales?

SELECT day, SUM(sales::numeric) as sales

FROM sales

GROUP BY day

ORDER BY SUM(sales)

LIMIT 5;

A screenshot of a computer

Description automatically generated

A graph with blue bars

Description automatically generated

A graph with blue lines

Description automatically generated

It seems the days with higher sales are usually in the first half of the month.

-- Which days of the week generate more sales?

SELECT day\_of\_week, SUM(sales::numeric) as sales

FROM sales

GROUP BY day\_of\_week

ORDER BY SUM(sales) DESC

A screenshot of a computer

Description automatically generated

We can see weekdays generate more sales than weekends.

-- Which months generate more sales?

SELECT month, SUM(sales::numeric) as sales

FROM sales

GROUP BY month

ORDER BY SUM(sales) DESC

A screenshot of a calendar

Description automatically generated

We can see Nov/Dec (holiday season) as the months with more sales.

-- Top 5 highest shipping cost products?

SELECT description as product, weight, shipping\_cost\_1000\_mile

FROM products

ORDER BY shipping\_cost\_1000\_mile DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

-- Correlation between product weight and shipping cost?

SELECT round(corr(weight, shipping\_cost\_1000\_mile)::numeric,2) as weight\_shipping\_corr

FROM products

A screen shot of a website

Description automatically generated

As expected, there is a high correlation between the product’s weight and shipping cost.