

Hlatsjwayo Jonas

Data science assessment 3

05 April 2024

I have provided two documents for my assessment:

- A google doc that you are currently reading. It contains the code I wrote for each query, along with explanations. The code is numbered to match the task numbers.
- An excel worksheet with the results for each of my queries.

To review my work, please check the excel sheet for the actual results, and refer to this google doc for the code and explanations.

Nb! The database I created is called 'company_sales' and the table is 'sales' as instructed.

1. Retrieve all columns from the 'sales' table.

```
SELECT * FROM company_sales.sales;
```

This query simply retrieves all the columns and rows from the 'sales' table, allowing us to view the complete dataset.

2. Find the total sales and total profit generated.

```
SELECT SUM(sales) AS total_sales, SUM(profit) AS total_profit  
FROM company_sales.sales;
```

This query calculates the total sales and total profit by summing the 'sales' and 'profit' columns, respectively. The SUM() aggregate function is used to add up all the values, and the aliases 'total_sales' and 'total_profit' are used to provide meaningful column names.

3. Calculate the average sales and average profit.

```
SELECT AVG(sales) AS avg_sales, AVG(profit) AS avg_profit  
FROM company_sales.sales;
```

This query calculates the average sales and average profit by using the AVG() aggregate function on the 'sales' and 'profit' columns, respectively. The aliases 'avg_sales' and 'avg_profit' are used to provide meaningful column names.

4. Identify the top 3 regions with the highest total sales.

```
SELECT region, SUM(sales) AS total_sales  
FROM company_sales.sales  
GROUP BY region  
ORDER BY total_sales DESC  
LIMIT 3;
```

This query first groups the data by the 'region' column using the GROUP BY clause. It then calculates the total sales for each region by summing the 'sales' column. The results are ordered in descending order by the 'total_sales' column, and the LIMIT 3 clause ensures that only the top 3 regions are returned.

5. Determine the product with the highest profit margin.

```
SELECT product, (profit / sales) AS profit_margin  
FROM company_sales.sales  
ORDER BY profit_margin DESC  
LIMIT 1;
```

This query calculates the profit margin for each product by dividing the 'profit' column by the 'sales' column. The results are then ordered in descending order by the 'profit_margin' column, and the LIMIT 1 clause returns the product with the highest profit margin.

6. Find the total sales and profit for each product.

```
SELECT product, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM company_sales.sales
GROUP BY product;
```

This query groups the data by the 'product' column and calculates the total sales and total profit for each product using the SUM() aggregate function. The GROUP BY clause ensures that the results are aggregated at the product level.

7. Calculate the average profit for each region.

```
SELECT region, AVG(profit) AS avg_profit
FROM company_sales.sales
GROUP BY region;
```

This query groups the data by the 'region' column and calculates the average profit for each region using the AVG() aggregate function. The GROUP BY clause ensures that the results are aggregated at the region level.

8. Identify the date with the highest sales.

```
SELECT date, SUM(sales) AS total_sales
FROM company_sales.sales
GROUP BY date
ORDER BY total_sales DESC
LIMIT 1;
```

This query groups the data by the 'date' column and calculates the total sales for each date using the SUM() aggregate function. The results are then ordered in descending order by the 'total_sales' column, and the LIMIT 1 clause returns the date with the highest total sales.

9. Determine the product with the highest sales in the East region.

```
SELECT product, SUM(sales) AS total_sales
FROM company_sales.sales
WHERE region = 'East'
GROUP BY product
ORDER BY total_sales DESC
LIMIT 1;
```

This query first filters the data to only include rows where the 'region' is 'East' using the WHERE clause. It then groups the data by the 'product' column and calculates the total sales for each product using the SUM() aggregate function. The results are ordered in descending order by the 'total_sales' column, and the LIMIT 1 clause returns the product with the highest sales in the East region.

10. Identify the top 5 dates with the highest sales.

```
SELECT date, SUM(sales) AS total_sales
FROM company_sales.sales
GROUP BY date
ORDER BY total_sales DESC
LIMIT 5;
```

This query groups the data by the 'date' column and calculates the total sales for each date using the SUM() aggregate function. The results are then ordered in descending order by the 'total_sales' column, and the LIMIT 5 clause returns the top 5 dates with the highest total sales.

The entire code

```
SELECT * FROM company_sales.sales;

SELECT SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM company_sales.sales;

SELECT AVG(sales) AS avg_sales, AVG(profit) AS avg_profit
FROM company_sales.sales;

SELECT region, SUM(sales) AS total_sales
FROM company_sales.sales
GROUP BY region
ORDER BY total_sales DESC
LIMIT 3;

SELECT product, (profit / sales) AS profit_margin
FROM company_sales.sales
ORDER BY profit_margin DESC
LIMIT 1;

SELECT product, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM company_sales.sales
GROUP BY product;

SELECT region, AVG(profit) AS avg_profit
FROM company_sales.sales
GROUP BY region;
```

```
SELECT date, SUM(sales) AS total_sales
FROM company_sales.sales
GROUP BY date
ORDER BY total_sales DESC
LIMIT 1;

SELECT product, SUM(sales) AS total_sales
FROM company_sales.sales
WHERE region = 'East'
GROUP BY product
ORDER BY total_sales DESC
LIMIT 1;

SELECT date, SUM(sales) AS total_sales
FROM company_sales.sales
GROUP BY date
ORDER BY total_sales DESC
LIMIT 5;
```

Overall Analysis Report

The sales data analysis provides valuable insights into the company's performance. The total sales generated across all products and regions is R166,846.10, with a total profit of R75,523.49. On average, each sale generates R556.15 in sales and R251.74 in profit.

The top 3 regions by total sales are the West, East, and South, indicating these regions are the company's strongest markets. The product with the highest profit margin is likely the company's most efficient offering, while the product with the highest sales in the East region could be a focus area for growth. Additionally, the analysis identifies the date with the highest sales and the top 5 dates with the highest sales, which could help the company plan for peak demand periods.