

BRIGHTLEARN

Practical exercise (Google BigQuery)

Dataset: Retail Sales Dataset

QUESTIONS

1. WHERE Clause

Q1. Filter all transactions that occurred in the year 2023.

Expected output: All columns

The screenshot shows the Google BigQuery Studio interface. The left sidebar has sections for Pipelines and integration, Data transfers, Dataform, Scheduled queries, Scheduling, Governance, and Sharing (Analytics Hub). The main area shows an 'Untitled query' with the following SQL code:

```
1 SELECT * FROM `retail-479317.Sales_data` LIMIT 1000;
2 -- 1. WHERE Clause Q1. Filter all transactions that occurred in the year 2023.
3 SELECT *
4 FROM `retail-479317.Sales_data`
5 WHERE EXTRACT(YEAR FROM DATE) = 2023;
```

A green checkmark indicates 'Query completed'. Below the code is a 'Using on-demand processing quota' message. The 'Results' tab is selected in the navigation bar. The results table has columns: Row, Transaction ID, Date, Customer ID, Gender, Age, and Product Category. The data shows three rows of transactions from 2023:

Row	Transaction ID	Date	Customer ID	Gender	Age	Product Category
1	191	2023-10-18	CUST191	Male	64	Beauty
2	204	2023-09-28	CUST204	Male	39	Beauty
3	230	2023-04-23	CUST230	Male	54	Beauty

2. Filtering + Conditions

Q2. Display all transactions where the Total Amount is more than the average Total Amount

of the entire dataset.

Expected output: All columns

The screenshot shows the Google BigQuery Studio interface. The left sidebar has sections for Pipelines and integration, Data transfers, Dataform, Scheduled queries, Scheduling, Governance, and Sharing (Analytics Hub). The main area shows an 'Untitled query' with the following SQL code:

```
9 SELECT
10   *
11   FROM
12     `retail-479317`.Sales_data
13   WHERE
14     `Total Amount` > (
15       | SELECT
```

A green checkmark indicates 'Query completed'. Below the code is a 'Using on-demand processing quota' message. The 'Results' tab is selected in the navigation bar. The results table has columns: Row, Transaction ID, Date, Customer ID, Gender, Age, and Product Category. The data shows three rows of transactions from January 2023:

Row	Transaction ID	Date	Customer ID	Gender	Age	Product Category
1	21	2023-01-14	CUST021	Female	50	Beauty
2	28	2023-04-23	CUST028	Female	43	Beauty
3	128	2023-07-05	CUST128	Male	25	Beauty

3. Aggregate Functions

Q3. Calculate the total revenue (sum of Total Amount).

Expected output: Total_Revenue

The screenshot shows the BigQuery Studio interface. The left sidebar has sections for Overview, Studio (selected), Pipelines and integration, Data transfers, Dataform, Scheduled queries, Scheduling, Governance, Sharing (Analytics Hub), Partner Centre, Settings (selected), and Release notes. The main area shows an Untitled query with the following SQL code:

```
11 FROM `retail-479317`.Sales.data
12 WHERE `Total Amount` > (
13   SELECT
14     AVG(`Total Amount`)
15   FROM
16     `retail-479317`.Sales.data
17 );
18
19 -- 3. Aggregate Functions Q3. Calculate the total revenue (sum of Total Amount). Expected output: Total_Revenue
20 SELECT SUM(`Total Amount`) AS Total_Revenue
21 FROM `retail-479317`.Sales.data;
```

The status bar at the bottom indicates "Query completed". The results pane shows a single row with "Total_Revenue" and the value "456000".

4. DISTINCT

Q4. Display all distinct Product Categories in the dataset.

Expected output: Product_Category

The screenshot shows the BigQuery Studio interface. The left sidebar has sections for Overview, Studio (selected), Pipelines and integration, Data transfers, Dataform, Scheduled queries, Scheduling, Governance, Sharing (Analytics Hub), Partner Centre, Settings (selected), and Release notes. The main area shows an Untitled query with the following SQL code:

```
23 FROM `retail-479317`.Sales.data`;
24
25 -- 4. DISTINCT Q4. Display all distinct Product Categories in the dataset.
26 SELECT DISTINCT `Product Category`
27 FROM `retail-479317`.Sales.data;
```

The status bar at the bottom indicates "Query completed". The results pane shows three rows with "Product Category": "Beauty", "Clothing", and "Electronics".

5. GROUP BY

Q5. For each Product Category, calculate the total quantity sold.

Expected output: Product_Category, Total_Quantity

```

27   FROM `retail-479317.Sales.data`;
28   -- GROUP BY Q5. For each Product Category, calculate the total quantity sold.
29
30   SELECT
31   'Product Category',
32   SUM(Quantity) AS Total_Quantity
33   FROM `retail-479317.Sales.data`
34   GROUP BY 'Product Category';
35

```

Query completed

Product Category	Total_Quantity
Beauty	771
Clothing	894
Electronics	849

6. CASE Statement

Q6. Create a column called Age_Group that classifies customers as ‘Youth’ (<30), ‘Adult’

(30–59), and ‘Senior’ (60+).

Expected output: Customer_ID, Age, Age_Group

```

37   SELECT
38   'Customer ID',
39   Age,
40   CASE
41   WHEN Age < 30 THEN 'Youth'
42   WHEN Age BETWEEN 30 AND 59 THEN 'Adult'
43   ELSE 'Senior'
44   END AS Age_Group
45   FROM `retail-479317.Sales.data`;
46

```

Query completed

Customer ID	Age	Age_Group
CUST191	64	Senior
CUST204	39	Adult
CUST230	54	Adult
CUST232	43	Adult

7. Conditional Aggregation

Q7. For each Gender, count how many high-value transactions occurred (where Total Amount > 500).

Expected output: Gender, High_Value_Transactions

The screenshot shows the BigQuery web interface. The left sidebar has sections for Overview, Studio, Pipelines and integration, Governance, and Release notes. The Studio section is active. The main area shows an 'Untitled query' with the following SQL code:

```
45 FROM `retail-479317.Sales.data`;
46
47 -- 7.. Conditional Aggregation Q7. For each Gender, count how many high-value transactions occurred (where Total Amount > 500).
48 SELECT
49   Gender,
50   COUNTIF(`Total Amount` > 500) AS High_Value_Transactions
51   FROM `retail-479317.Sales.data`
52   GROUP BY Gender;
53
54
```

A green checkmark indicates 'Query completed'. Below the code, a button says 'Using on-demand processing quota'. The 'Results' tab is selected in the 'Query results' header. The results table shows:

Row	Gender	High_Value_Trans...
1	Male	144
2	Female	155

At the bottom, there are navigation icons and a footer with 'Results per page: 50 ▾ 1 - 2 of 2'.

8. HAVING Clause

Q8. For each Product Category, show only those categories where the total revenue exceeds 5,000.

Expected output: Product_Category, Total_Revenue

The screenshot shows the BigQuery web interface. The left sidebar has sections for Overview, Studio, Pipelines and integration (Data transfers, Dataform, Scheduled queries, Scheduling), Governance (Sharing (Analytics Hub)), Partner Centre, Settings (Preview selected), and Release notes. The main area shows an 'Untitled query' with the following SQL code:

```
52 GROUP BY Gender;
53
54 -- 8. HAVING Clause Q8. For each Product Category, show only those categories where the total revenue exceeds 5,000
55 SELECT
56   `Product Category`,
57   SUM(`Total Amount`) AS Total_Revenue
58 FROM `retail-479317.Sales_data`
59 GROUP BY `Product Category`
60 HAVING SUM(`Total Amount`) > 5000;
```

A green checkmark indicates 'Query completed'. Below the code, a button says 'Using on-demand processing quota'. The 'Results' tab is selected under 'Query results'. The results table has columns 'Row', 'Product Category', and 'Total_Revenue'. The data is as follows:

Row	Product Category	Total_Revenue
1	Beauty	143515
2	Clothing	155580
3	Electronics	156905

At the bottom, there are buttons for 'Save results' and 'Open in', and a footer with 'Results per page: 50' and navigation icons.

9. Calculated Fields

Q9. Display a new column called Unit_Cost_Category that labels a transaction as:

- 'Cheap' if Price per Unit < 50
 - 'Moderate' if Price per Unit between 50 and 200
 - 'Expensive' if Price per Unit > 200

Expected output: Transaction_ID, Price_per_Unit, Unit_Cost_Category

The screenshot shows the Google BigQuery web interface. On the left, there's a sidebar with navigation links like Overview, Studio, Pipelines and integration, Data transfers, Dataform, Scheduled queries, Scheduling, Governance, and Sharing (Analytics Hub). The main area is titled 'Untitled query' and contains the following SQL code:

```

62 -- 9. Calculated Fields Q9. Display a new column called Unit_Cost_Category that labels a transaction as:
63 SELECT
64   'Transaction ID',
65   'Price per Unit',
66   CASE
67     WHEN 'Price per Unit' < 50 THEN 'Cheap'
68     WHEN 'Price per Unit' BETWEEN 50 AND 200 THEN 'Moderate'
69     ELSE 'Expensive'
70   END AS Unit_Cost_Category
71 FROM `retail-479317.Sales.data`;

```

A green checkmark indicates 'Query completed'. Below the code, a message says 'Using on-demand processing quota'. The 'Results' tab is selected in the 'Query results' section. The table has columns: Row, Transaction ID, Price per Unit, and Unit_Cost_Category. The data is as follows:

Row	Transaction ID	Price per Unit	Unit_Cost_Category
1	191	25	Cheap
2	204	25	Cheap
3	230	25	Cheap
4	232	25	Cheap

At the bottom, it says 'Results per page: 50 1 – 50 of 1000'.

10. Combining WHERE + CASE

Q10. Display all transactions from customers aged 40 or older and add a column

Spending_Level showing 'High' if Total Amount > 1000, otherwise 'Low'.

Expected output: Customer_ID, Age, Total_Amount, Spending_Level

The screenshot shows the Google BigQuery web interface. The sidebar and navigation links are identical to the previous screenshot. The main area is titled 'Untitled query' and contains the following SQL code:

```

73 -- 10. Combining WHERE + CASE Q10. Display all transactions from customers
74 SELECT
75   'Customer ID',
76   'Age',
77   'Total Amount',
78   CASE
79     WHEN 'Total Amount' > 1000 THEN 'High'
80     ELSE 'Low'
81   END AS Spending_Level
82 FROM `retail-479317.Sales.data`
83 WHERE Age >= 40;

```

A green checkmark indicates 'Query completed'. Below the code, a message says 'Using on-demand processing quota'. The 'Results' tab is selected in the 'Query results' section. The table has columns: Row, Customer ID, Age, Total Amount, and Spending_Level. The data is as follows:

Row	Customer ID	Age	Total Amount	Spending_Level
1	CUST191	64	25	Low
2	CUST230	54	25	Low
3	CUST232	43	25	Low

At the bottom, it says 'Results per page: 50 1 – 50 of 558'.