Great! Now that you’ve calculated the ride duration, we can proceed with \*\*Step 2\*\*, which involves aggregating and summarizing the data. This will help you identify trends and insights regarding the ride duration, station usage, and bike types.

### Step 2: Aggregating Data (Ride Statistics and Frequency Analysis)

Here’s the SQL code to analyze the aggregated data, including:

- Calculating the \*\*average\*\*, \*\*minimum\*\*, \*\*maximum\*\*, and \*\*median\*\* ride durations.

- Identifying the most frequently used stations (both start and end stations).

```sql

-- Step 2: Aggregated analysis on ride duration and station usage

WITH ride\_data AS (

-- Step 1: Calculate ride duration (reusing the previous calculation)

SELECT

rideable\_type,

started\_at,

ended\_at,

start\_station\_id,

end\_station\_id,

TIMESTAMP\_DIFF(ended\_at, started\_at, MINUTE) AS duration\_minutes

FROM

`alpine-proton-431618-q7.cyclist\_portfolio.filtered\_combined\_table`

WHERE

started\_at IS NOT NULL

AND ended\_at IS NOT NULL

)

-- Aggregating the data

SELECT

rideable\_type,

start\_station\_id,

end\_station\_id,

COUNT(\*) AS total\_rides, -- Total number of rides per start and end station pair

AVG(duration\_minutes) AS avg\_duration, -- Average ride duration for each combination

MIN(duration\_minutes) AS min\_duration, -- Shortest ride duration for each combination

MAX(duration\_minutes) AS max\_duration, -- Longest ride duration for each combination

PERCENTILE\_CONT(duration\_minutes, 0.5) OVER () AS median\_duration -- Median ride duration

FROM

ride\_data

GROUP BY

rideable\_type,

start\_station\_id,

end\_station\_id

ORDER BY

total\_rides DESC; -- Order by most frequent start/end station combinations

```

### Explanation of Step 2:

- \*\*Aggregating Ride Data\*\*:

- \*\*`COUNT(\*)`\*\*: This counts the total number of rides between a specific start station and end station for each bike type.

- \*\*`AVG(duration\_minutes)`\*\*: This calculates the average ride duration for each station pair and bike type.

- \*\*`MIN(duration\_minutes)` and `MAX(duration\_minutes)`\*\*: These calculate the shortest and longest ride durations for each combination of bike type and station pair.

- \*\*`PERCENTILE\_CONT(duration\_minutes, 0.5) OVER ()`\*\*: This calculates the median ride duration across all rows. This will give you insights into the typical ride duration (50th percentile).

- \*\*Grouping and Ordering\*\*:

- The query groups by `rideable\_type`, `start\_station\_id`, and `end\_station\_id`, so you can see the statistics for each station combination and bike type.

- Results are ordered by `total\_rides` in descending order, so the most frequently used start/end stations appear at the top.

### Example Output

The result of this query will provide the following columns:

- \*\*rideable\_type\*\*: The type of bike used (e.g., standard bike, electric bike, etc.).

- \*\*start\_station\_id\*\*: The station ID where the ride started.

- \*\*end\_station\_id\*\*: The station ID where the ride ended.

- \*\*total\_rides\*\*: The total number of rides that occurred between this start and end station pair.

- \*\*avg\_duration\*\*: The average ride duration for this station pair.

- \*\*min\_duration\*\*: The shortest ride duration for this station pair.

- \*\*max\_duration\*\*: The longest ride duration for this station pair.

- \*\*median\_duration\*\*: The median ride duration for this station pair.

Next Steps

1. \*\*Interpreting Results\*\*:

- Once you run this query, you can analyze trends such as:

- Which stations have the most rides.

- Which types of bikes are used most frequently.

- Whether certain bike types (e.g., electric bikes) are associated with longer or shorter ride durations.

- Whether certain station pairs are consistently busy.

2. \*\*Additional Analysis\*\*:

- You can further break down the analysis by \*\*hour of the day\*\*, \*\*day of the week\*\*, or \*\*season\*\* to identify patterns in bike usage.

- If you have membership data (e.g., casual vs. annual members), you could also add this information to the query to compare how casual riders and annual members differ in their ride behavior.

Let me know if you'd like to dive deeper into any specific analysis or if you need further help interpreting the results!