Dharma Devi 25th April HW

Write a SQL query to categorize the flights based on their distance. The categories are defined as follows:

* Distance between 0 and 500: Category 1
* Distance between 501 and 1000: Category 2
* Distance between 1001 and 2000: Category 3
* Distance between 2001 and 3000: Category 4
* Distance between 3001 and 4000: Category 5
* Distance between 4001 and 5000: Category 6

Add it to above pyspark code.

SELECT

flight\_id,

flight\_num,

flight\_date,

distance,

CASE

WHEN distance BETWEEN 0 AND 500 THEN 'Category 1'

WHEN distance BETWEEN 501 AND 1000 THEN 'Category 2'

WHEN distance BETWEEN 1001 AND 2000 THEN 'Category 3'

WHEN distance BETWEEN 2001 AND 3000 THEN 'Category 4'

WHEN distance BETWEEN 3001 AND 4000 THEN 'Category 5'

WHEN distance BETWEEN 4001 AND 5000 THEN 'Category 6'

ELSE 'Unknown Category' -- This handles cases where distance exceeds 5000

END AS distance\_category

FROM flight\_data;

**Local Development Setup**



