Rapido Assignment

Part 1. Exploratory data analysis

Perform an exploratory data analysis on the given dataset and share your findings.

Part 2. Metric calculation

What is the average duration between the 1st trip and the 2nd trip of customers? Note: Consider only the customers who have done 2 or more trips.

Part 3. Model building

Build a model to predict trip_fare using travel_distance and travel_time. Measure the accuracy of the model and use the model to predict trip_fare for a trip with travel_distance of 3.5 kms and travel_time of 15 minutes.

Part 4. Top Hex clusters

Top 5 pairs of hex (resolution=8) clusters where most of the trips happened? You can refer to the library listed below to get hexid for a given latitude and longitude.

- https://github.com/uber/h3-py

Expected output

Rank	Hex pair (source_hexid, destination_hexid)	Total trips
1	883c8e4159fffff to 88754e6499fffff	34
2	883db66b55fffff to 883c8e4159fffff	28

Dataset

Serial No.	Column names	Description
1	trip_id	Unique identifier for customer
2	customer_id	Unique identifier for customer
3	timestamp	Time stamp of the trip in Epoch format
4	pick_lat	Latitude of the pickup location
5	pick_lng	Longitude of the pickup location
6	drop_lat	Latitude of the drop location
7	drop_lng	Longitude of the drop location
8	travel_distance	Distance of trip measured in KMs
9	travel_time	Duration of the trip measured in Minutes
10	trip_fare	Trip fare calculated in Rupees