Steps :-

**Part 1. Exploratory data analysis**

Performed an exploratory data analysis on the given dataset (ds.xlsx) and shared my findings.

**Part 2. Metric calculation**

Calculated average duration between the 1st trip and the 2nd trip of customers. Note: Considered only the customers who have done 2 or more trips.

**Part 3. Model building**

Built a model to predict trip\_fare using travel\_distance and travel\_time. Measured the accuracy of the model and used 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**

Identified top 5 pairs of hex (resolution=8) clusters where most of the trips happened. (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 |