**Assignment Problems: Day 3**

The following problems require access to the file named ‘Dataset\_Day3.csv’, provided with this assignment.

About the data,

* key - a unique identifier for each trip
* fare\_amount - the cost of each trip in usd
* pickup\_datetime - date and time when the meter was engaged
* passenger\_count - the number of passengers in the vehicle (driver entered value)
* pickup\_longitude - the longitude where the meter was engaged
* pickup\_latitude - the latitude where the meter was engaged
* dropoff\_longitude - the longitude where the meter was disengaged
* dropoff\_latitude - the latitude where the meter was disengaged

Problems to solve,

1. Use the location coordinates to find the distance for each trip. Create a new column – ‘*distance’* and store its value there. Use the following formula:
2. Find all the ‘*key*’ values for which the attributes: *fare\_amount* & *passenger\_count & distance* are outliers. **Remove all rows with outliers**.
3. Show the scatterplot between *distance* & *fare\_amount.* Is there any relationship that you can identify?