|  |  |
| --- | --- |
| QUESTION 1 | |
| …………… | The following questions were addressed using the dataset including bicycles that are stored at fixed docking stations throughout the city. |
| PART 1 | Write a function that removes all trips of duration 0 minutes and prints the following values on the console. The function remove\_zero\_duration\_trips(df) is defined to address the necessary issues. The combined format of df['started\_at'] and df['ended\_at'] is %d-%m-%Y%H:%M, which the datetime library helped to extract the date and time from.  Maximum duration of the trip (in minutes):518 minutes  Minimum duration of the trip (in minutes):1 minutes  Total number of trips corresponding to the minimum duration:89  Percentage of total circular trips: 2.4610455803116356  Total runtime for the function: 0.03 seconds |
| PART 2 | Filter the original dataset to include only the trips starting between 06:00 AM and 06:00 PM. Find the total number of feasible pairs of trips. After filtering by the required format (df['started\_at'].dt.hour >6) & (df['started\_at'].dt.hour < 18) the number of rows we obtained after filtering are 4680. As a result of comparing each value in the "started\_at" column of the "df\_duplicate" DataFrame to each value in the "ended\_at" column of the "df\_filtered" DataFrame, df\_duplicate is formed from df\_filtered. Another mask to exclude rows in the df\_duplicate dataframe where the corresponding start\_lat and start\_lng values in the df\_duplicate dataframe match the end\_lat and end\_lng values in the df\_filtered dataframe and the started\_at time is greater than or equal to the ended\_at time in the df\_filtered dataframe.  The items in mask with True indices are returned by np.argwhere(mask). These indexes provide access to the entries in the df\_duplicate and df\_filtered tables that fulfil the criteria specified in the mask  Total Number of matching pairs: 41782  Total runtime for the function: 0.22 seconds |
| paRT 3 | Filter the original dataset to include only the first 100 trips (i.e., trip id 1 to 100). In the report, mention the number of unique depots used to serve these trips Number of unique depots: 98 After dropping the duplicate files through depots\_df.drop\_duplicates(subset=['start\_lat', 'start\_lng', 'end\_lat', 'end\_lng'], inplace=True)  Maximum shortest path length: 4393.912000000001 meters  Minimum shortest path length: 7.19 meters  Total runtime: 3922.05 seconds  Distance between path is given by:  <https://docs.google.com/document/d/1FdRyFyFAKadm5lIpbM27yMHYoZivJ6UTeH4HYkPB7lc/edit>  For comparing each dataset individually they are compared by using a locus as Washington, DC, USA  (38.889248, -77.050636) and then calculating the shortest path length between start and end points for each row in the depots\_df DataFrame.  Total runtime for the function: 47.93 seconds  Maximum distance: 6501.599999999999 meters  Minimum distance: 99.73 meters  Distance between path is given by:  <https://docs.google.com/document/d/1VnGrfWYxNJLOSHrQogkI1pDN3ySGytJhhYBOS2qxTn0/edit> |