

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/1FdRyFyFAKadm5IpbM27yMHYoZivJ6UTeH4HYkPB7Ic/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/1VnGrfWYxNJLOSHrQogkl1pDN3ySGytJhhYBOS2qxTn0/edit>