Report for Summer Internship Cistup IIsc 2023

(Name : Sajid Javid)

Important Libraries and Packages Used:

- !pip install geopandas # ! --> Google Collab
- !pip install folium
- import pandas as pd
- import numpy as np
- import matplotlib.pyplot as plt
- import time
- import geopandas as gpd
- import seaborn as sns
- import folium
- from folium import plugins
- from shapely.geometry import Point, Polygon
- from multiprocessing import Pool
- from scipy.spatial import Voronoi, voronoi_plot_2d

Question no 1 (part 1)

- **Maximum duration of the trip (in minutes): 518.00**
- **Minimum duration of the trip (in minutes): 1.00**
- **Total number of trips corresponding to the minimum duration: 89**
- **Percentage of total circular trips: 2.95%**
- **Total runtime: 0.04 seconds**

```
process_bike_data(dataset_path)
```

```
Maximum duration of the trip (in minutes): 518.00
Minimum duration of the trip (in minutes): 1.00
Total number of trips corresponding to the minimum duration: 89
Percentage of total circular trips: 2.95%
Total runtime: 0.04 seconds
```

Question no 1 (part 2)

Total feasible pairs of trips: 51951

Total runtime: 41.68 seconds

```
        runtime = end_time - start_time
        print(f"Total runtime: {runtime:.2f} second

    find_feasible_pairs(dataset_path)
```

```
Total feasible pairs of trips: 51951
Total runtime: 41.68 seconds
```

Question no 1 (part 3)

Code has been implemented and the results could not be displayed because of the python new version.

Question no 2 (part 1)

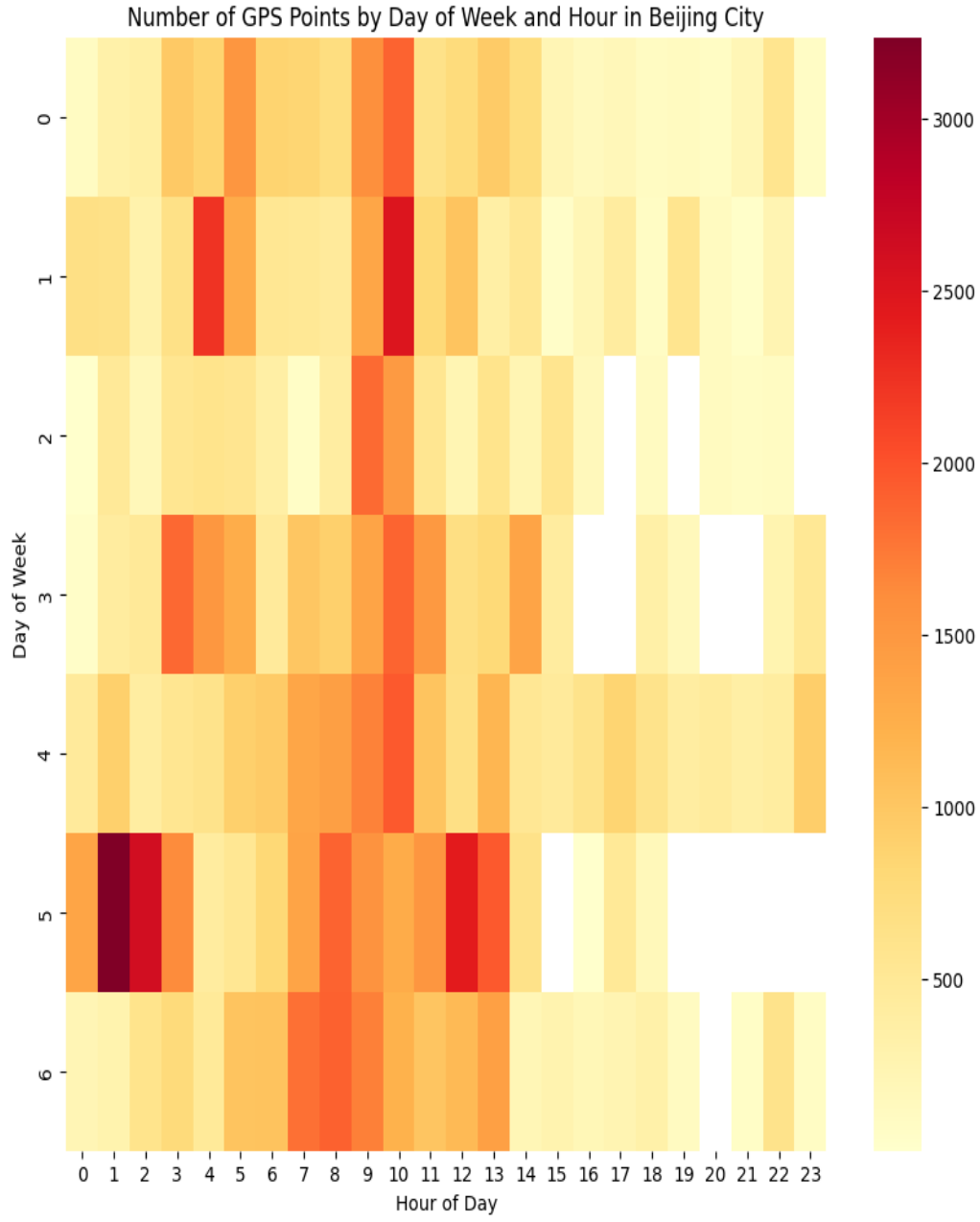
individual_id	total_distance
1.0	75.079503
2.0	10.369296
3.0	47.014150
4.0	171.644505
5.0	105.945486

```
   individual_id  total_distance
0            1.0       75.079503
1            2.0       10.369296
2            3.0       47.014150
3            4.0      171.644505
4            5.0      105.945486
```

63°F

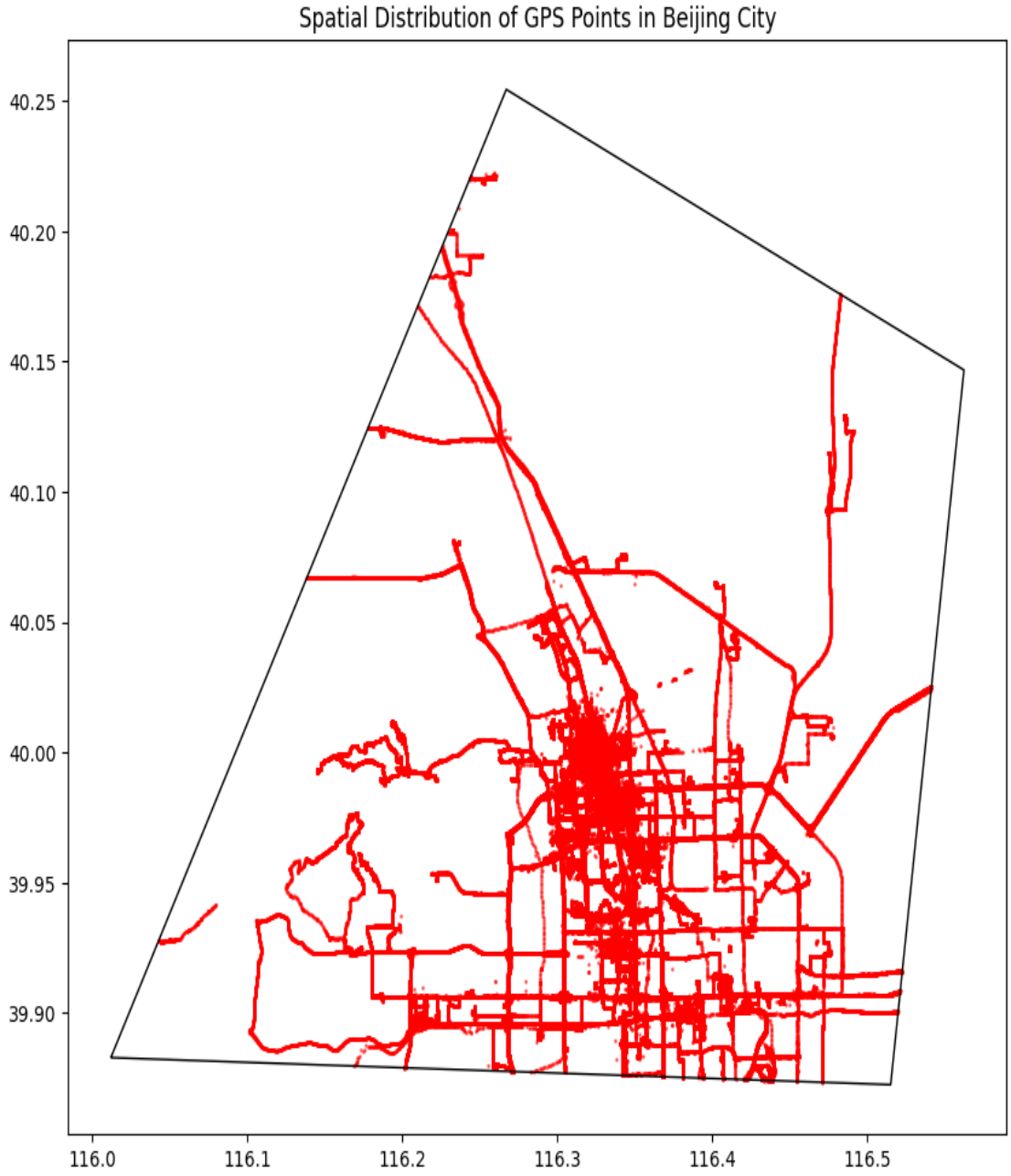
Question no 2 (Part 2 - (i)) - Spatio-Temporal Anaylsis

Temporal Analysis : Temporal data suggests that people in Beijing city tend to move around more during weekdays and rush hours, and that the city is less active during weekends and at night. There are several analyses that can be driven to gain insights into the underlying pattern.
For example, by analyzing this temporal data, we can identify activity hotspots, or areas where people are most active during certain times of the day or week. This information can be useful for identifying areas of high demand for public services, retail businesses, or other amenities.



<u>Spatio-Analysis:</u> There is more rush of people (GPS Points) at the centre of Beijing City in the range of latitudes (39.95 - 40) and longitudes (116.3 - 116.5).

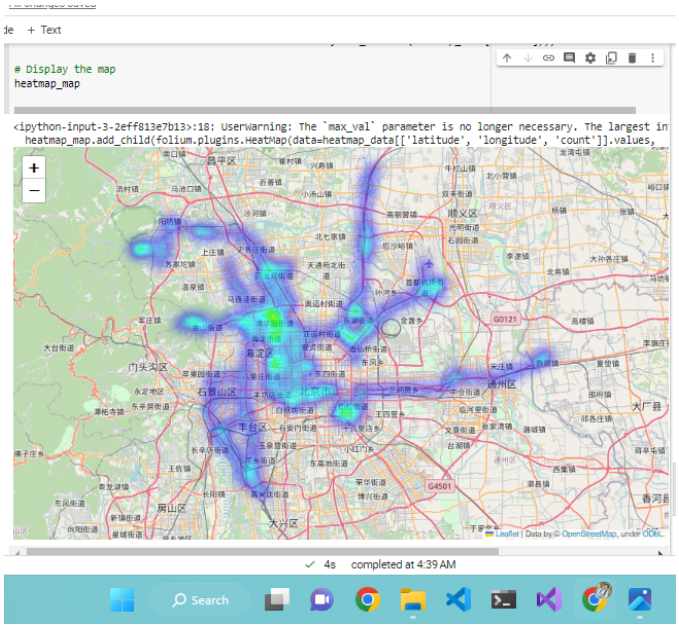
This spatial data can be used to identify areas of high population density, to analyze patterns of land use, and to plan new infrastructure projects such as roads, public transportation, or parks.



Question no 2 (Part 2 - (ii)) - Other Visualisations

(1) - Heatmap : The resulting heatmap shows where GPS points are concentrated and can be used to identify areas of high activity or density. Depending on the purpose of our analysis, we may then take different actions based on this information.

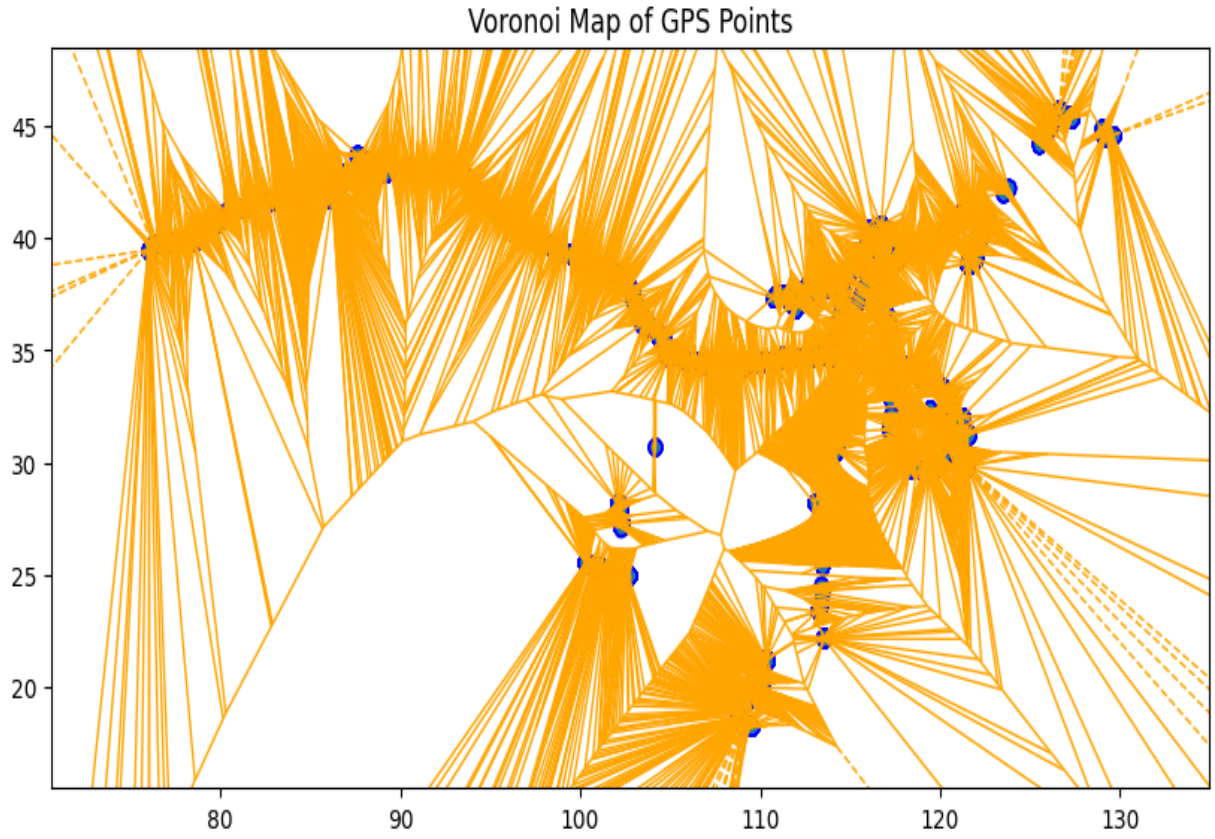
For example, if we are analyzing GPS data for a transportation system such as a bus network, we may use the heatmap to identify areas where buses are frequently delayed due to heavy traffic. This information can be used to prioritize investments in transportation infrastructure in those areas or to adjust bus schedules to account for the increased travel time.



Question no 2 (Part 2 - (ii)) - Other Visualisations

(2) - Voronoi Map of GPS Points :

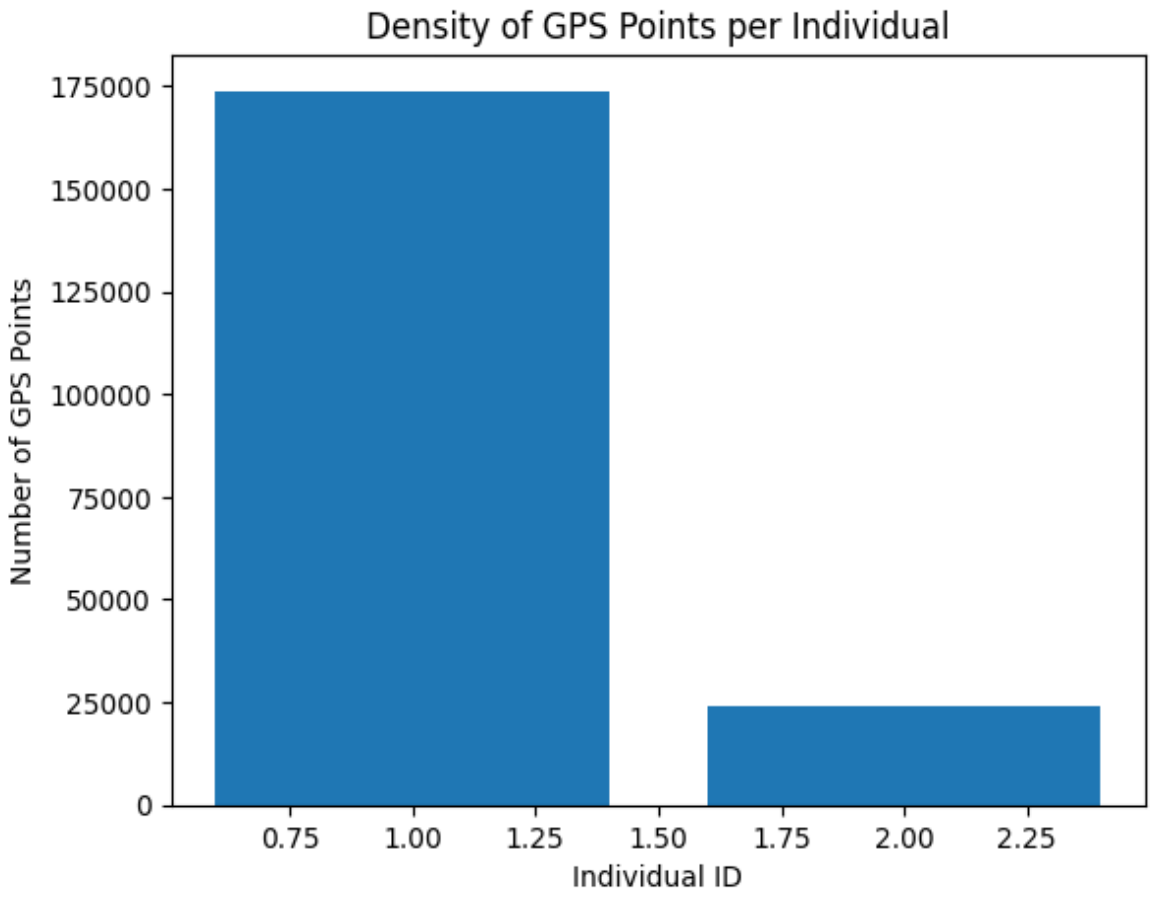
This spatial analysis that can be used to divide a geographic area into regions based on the proximity of GPS points where each GPS point is represented by small blue circle assigned to a region that encompasses all points that are closer to that point than to any other point in the dataset.
Using this Voronoi map of GPS points in Beijing city, we can gain insights into the spatial distribution of activity and mobility across the city.
For example we can identify activity centers(blue circles) , or areas where there are clusters(yellow clusters) of GPS points. These areas can be further analyzed to understand the types of activities that are taking place and to identify opportunities for urban design or other interventions.



Question no 2 (Part 2 - (ii)) - Other Visualisations

(3) - Density of GPS Points per Individual:

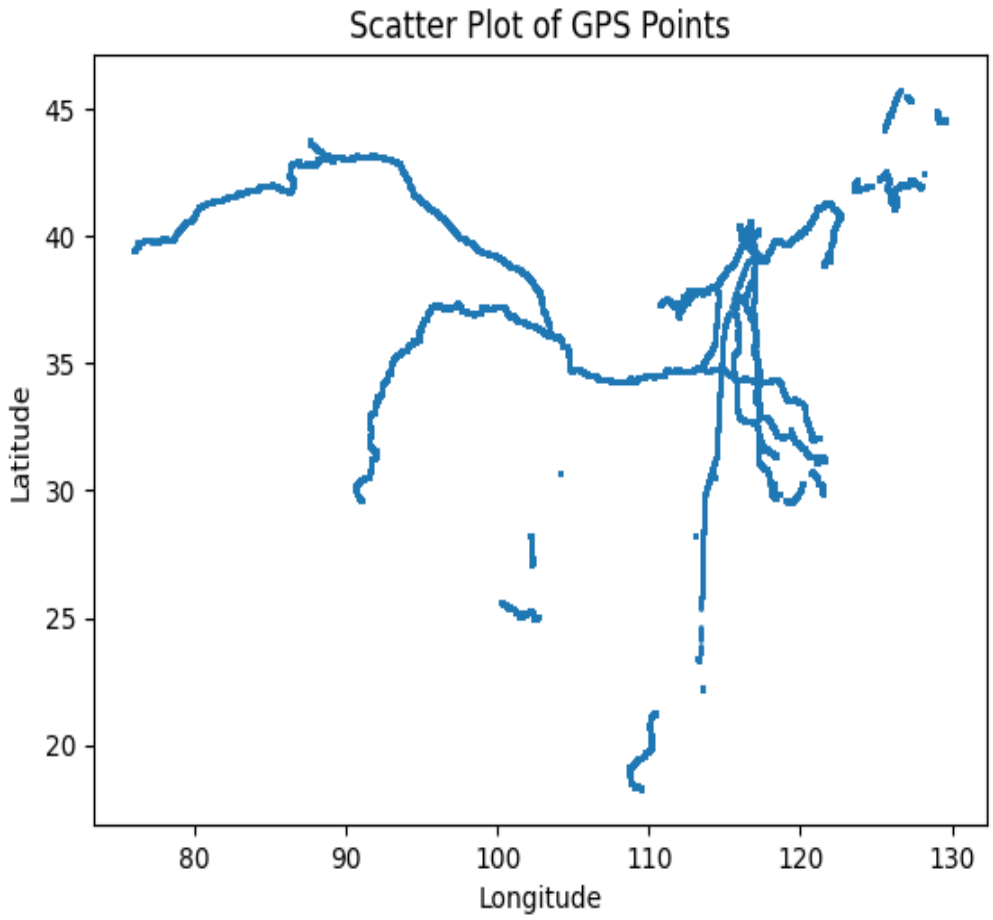
Analyzing the density of GPS points per individual provides insights into individual mobility patterns, activity spaces, and other aspects of human behavior. By analyzing the density of GPS points for different individuals in a dataset, we can identify patterns and trends that can be used to inform urban design, transportation planning, and other applications. Here we can see individual with id = 1 having more activity than the rest of the other individuals.
By analyzing the density of GPS points for different individuals, we can identify their activity spaces, or the areas where they spend the most time. This information can be used to design interventions that improve access to services and amenities.



Question no 2 (Part 2 - (ii)) - Other Visualisations

(4) - Scatter Plot of GPS Points:

A scatter plot of GPS points has provided valuable insights into the spatial distribution of points and patterns in the data. Here some potential insights we could gain from analyzing this scatter plot of GPS points are that at lat ~ 40 and long ~ 20 individuals tend to spend more time or these are areas with high levels of activity. This could indicate areas of interest or hotspots. This could be useful for identifying popular destinations by looking at the sequence of GPS points over time like we have analyzed movement pattern in this scatter.



Question no 2 (Part 2 - (ii)) - Other Visualisations

(5) - Bar Chart of GPS Points per Individual:

We can clearly see the individuals who have an unusually high or low number of GPS points compared to the rest of the group. This could indicate that they spend more or less time in certain areas or that their GPS device was malfunctioning.
We can also compare the number of GPS points for each individual to see who has the most and least. This could be useful in understanding individual behaviors and activities.