**Data Description**

The project utilizes three primary datasets:

**Dataset “NYCPilot1PM.csv”:** This file contains sensor readings from various locations across NYC, likely including attributes like timestamps, sensor ID, location coordinates (latitude, longitude), and pollutant concentrations (PM2.5, potentially other pollutants).

**Dataset “nyc1.csv”:** This dataset contains taxi mobility data such as pickup and drop off location, fare amount, trip distance, passengers trip amount etc.

**Dataset “nycpolygon.geojson”:** This data represents the street network of NYC. The specific format may vary (e.g., shapefile, GeoJSON), but it should contain information about street segments, including their geometries (linestrings) and potentially additional attributes like street names or road types.

## Conclusion and Future Work

This project has successfully developed a route planning system that integrates air quality data (PM2.5) into the pathfinding process. By utilizing exploratory data analysis, spatial techniques, and a novel weighting scheme, the system can recommend routes that minimize pollution exposure while still considering factors like distance. The evaluation using stratified sampling demonstrated the effectiveness of the system in providing accurate route recommendations based on sampled data, offering a balance between computational efficiency and accuracy.

There are several exciting avenues for future work to further enhance this system. First, applying the system to a different dataset, such as the NYC taxi mobility data, would allow for a broader evaluation of its performance and generalizability. Second, incorporating additional factors like traffic congestion or elevation changes into the weighting scheme could provide even more comprehensive route recommendations. Furthermore, exploring advanced metrics like PM2.5 dose reduction rate could quantify the system's potential health benefits. Also, based on passenger data we can determine how much people are affect due to polluted way, or we can indicate using shortest path in context of pollution, how much they are safe as compared to less polluted way. Finally, developing a user-friendly interface for this system would make it accessible to a broader audience, allowing individuals to make informed decisions about their routes while considering both travel time and air quality. By pursuing these avenues of future work, this project has the potential to significantly contribute to the development of practical and health-conscious route planning tools.