**Matthew Latondresse**  **9/24/2023**

**Project Subject Area**: The project aims to analyze and visualize data related to environmental factors affecting air quality in urban areas. By combining data from different sources, we will explore the relationships between weather conditions, pollution levels, and public health. This project has practical implications for understanding air quality's impact on urban populations and can assist in making informed decisions regarding public health and policy interventions.

**Data Sources**:

* **Flat File (CSV/Excel)**:
  + Link or Flat File: Air Quality Data
  + This dataset contains hourly air quality measurements (e.g., PM2.5, NO2, CO) from various monitoring stations across a city. It also includes geographical coordinates and timestamps.
* **API (OpenWeatherMap API)**:
  + Link: [OpenWeatherMap API](https://openweathermap.org/api)
  + We will use the OpenWeatherMap API to fetch real-time weather data for the same city where air quality data was collected. This will provide information on temperature, humidity, wind speed, and other meteorological factors.
* **Website (Wikipedia):**
  + Link: [Air Quality Index](https://en.wikipedia.org/wiki/Air_quality_index)
  + This website contains tables with information on pollution sources, emissions data, and public health statistics for the same urban area. The data includes the number of cases related to respiratory illnesses and known pollution sources in the city.

**Relationships**: The relationships between these data sources are as follows:

* **Air Quality Data (Flat File) and Weather Data (API):**
  + We can link the two datasets based on the timestamp and geographical coordinates of the air quality monitoring stations. This will allow us to analyze how weather conditions (e.g., temperature, wind speed) influence air quality.
* **Air Quality Data (Flat File) and Pollution Sources/Health Data (Website):**
  + We will create a relationship by matching the geographical locations of monitoring stations with known pollution sources in the city. This will help us assess the impact of specific pollution sources on air quality and public health.
* **Weather Data (API) and Pollution Sources/Health Data (Website):**
  + By using the geographical coordinates of the urban area, we can cross-reference weather conditions with public health statistics. This will enable us to investigate correlations between weather patterns, pollution sources, and health outcomes.

**Project Approach and Challenges**: To accomplish the project's milestones, we will:

* Collect and preprocess data from the three sources, ensuring data consistency and compatibility.
* Merge the datasets based on common attributes, such as timestamps and geographical coordinates.
* Perform exploratory data analysis to identify patterns and correlations.
* Create visualizations to communicate findings effectively.

**Ethical Implications**: We must ensure data privacy and handle health-related information responsibly. Ethical considerations include anonymizing health data, obtaining necessary permissions, and maintaining transparency in our analysis and reporting.

**Challenges**:

* Data quality issues and missing values may require data cleaning.
* Ensuring the ethical use of health data.
* Managing data from diverse sources and formats.
* Interpretation challenges due to complex relationships between weather, pollution, and health.

Throughout the project, ethical considerations played a crucial role. Handling personal and environmental data requires a commitment to privacy and responsible data usage. Ensuring the anonymization of sensitive information and abiding by relevant data protection laws is paramount. Additionally, transparency in data cleansing methods and acknowledgment of potential biases in the dataset is crucial for maintaining ethical standards.

The project highlighted the importance of accurate and reliable data for making informed decisions, particularly in domains like air quality monitoring. Cleansing and merging data effectively contribute to the accuracy of analyses and subsequent visualizations, enabling stakeholders to derive meaningful insights.

In conclusion, this project served as a practical application of data cleaning, merging, and visualization techniques. Beyond the technical aspects, the project underscored the ethical responsibilities associated with handling diverse datasets, emphasizing the need for transparency, privacy, and an awareness of potential biases.