Impact of PM2.5 Concentration in atmosphere on Respiratory Hospitalizations in New York

By- Utkarsh Bansal
(23546954)

Datasets Used

NYC Open Data (Air-Quality Data):-

- **Description**:- This dataset provides detailed measurements of PM2.5 levels, across various Locations in New York and the data is in CSV format.

- **Structure and Quality**:- The data is disaggregated by UHF-42, Time Period and pollutant type, enabling granular analysis of air quality trends over time.

- **Licensing**:- The dataset is publicly accessible for research and analysis

Datasets Used

NYC Environmental Public Health Data Explorer (Hospitalisations Data):-

- • **Description**:- This dataset provides data on number of hospitalizations caused due to air pollution, including PM2.5 exposure, in New York City across various neighbourhoods.

- • Structure and Quality:- The data includes detailed breakdowns by UHF-42, demographic groups, and time periods, supporting an in-depth examination of air pollution's effects on public health.

- **Licensing**:- The dataset is publicly accessible for research and analysis.

Pre-Processing

Data Cleaning and Preparation

• In the Air Quality Dataset Missing values and irrelevant columns were removed. Only PM2.5 data was retained, and the cleaned dataset was stored in a database.

• Hospitalization Dataset was already cleaned; directly stored in the database.
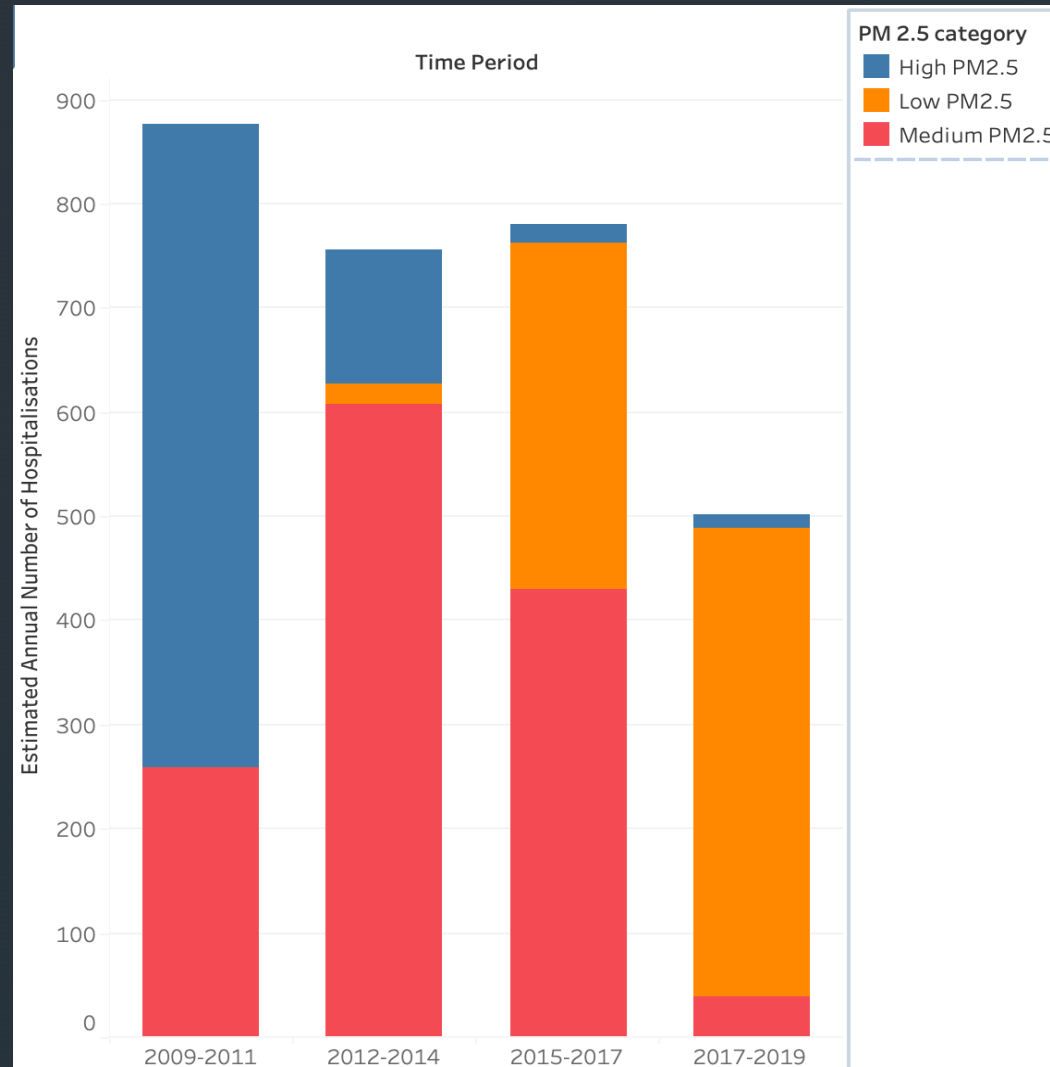
Aggregation and Joining

• The air quality data (yearly) was aggregated into multi-year time periods (e.g., 2009–2011) to match the hospitalization dataset. PM2.5 averages were calculated for each UHF-42 region..

• The datasets were joined using SQL, with UHF-42 regions and time periods as keys.
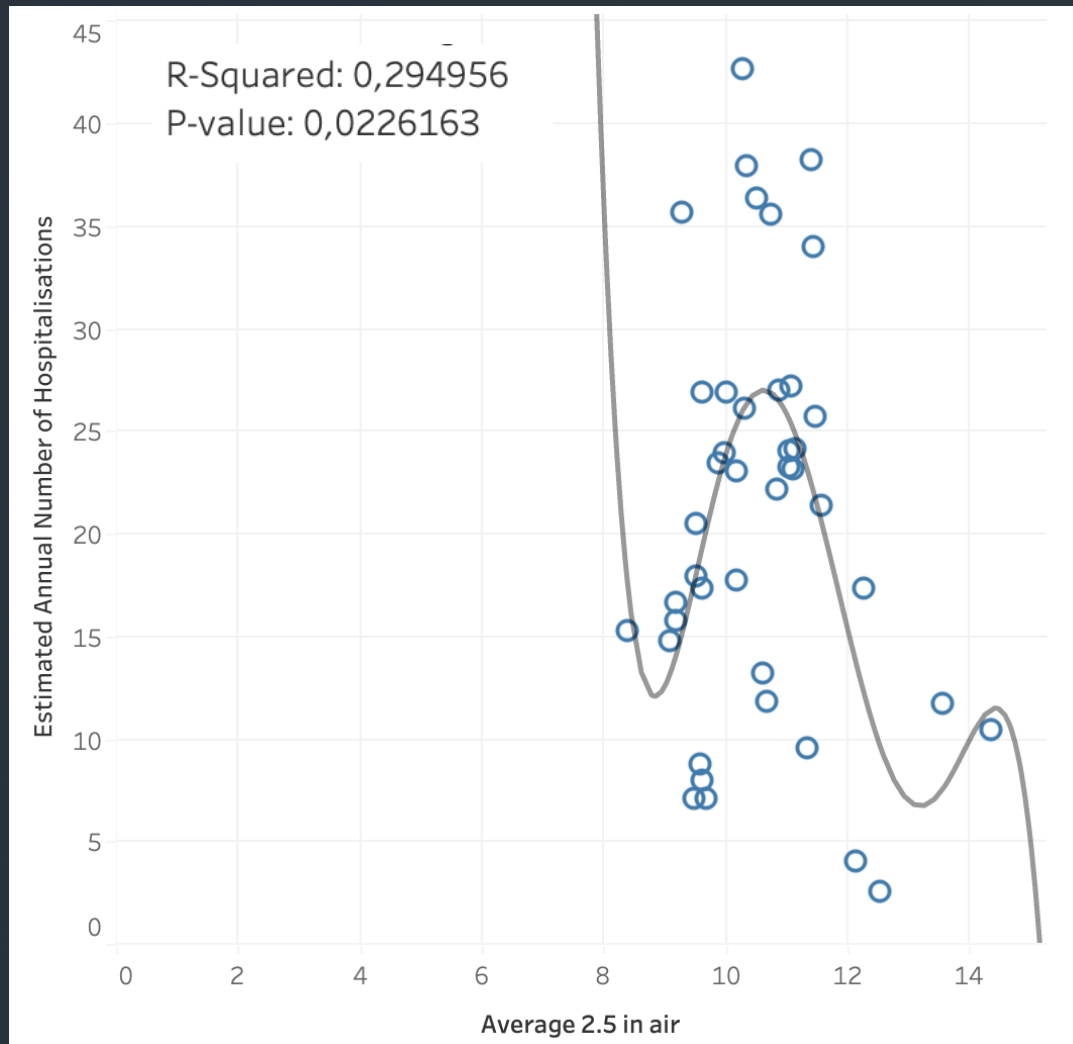
Results
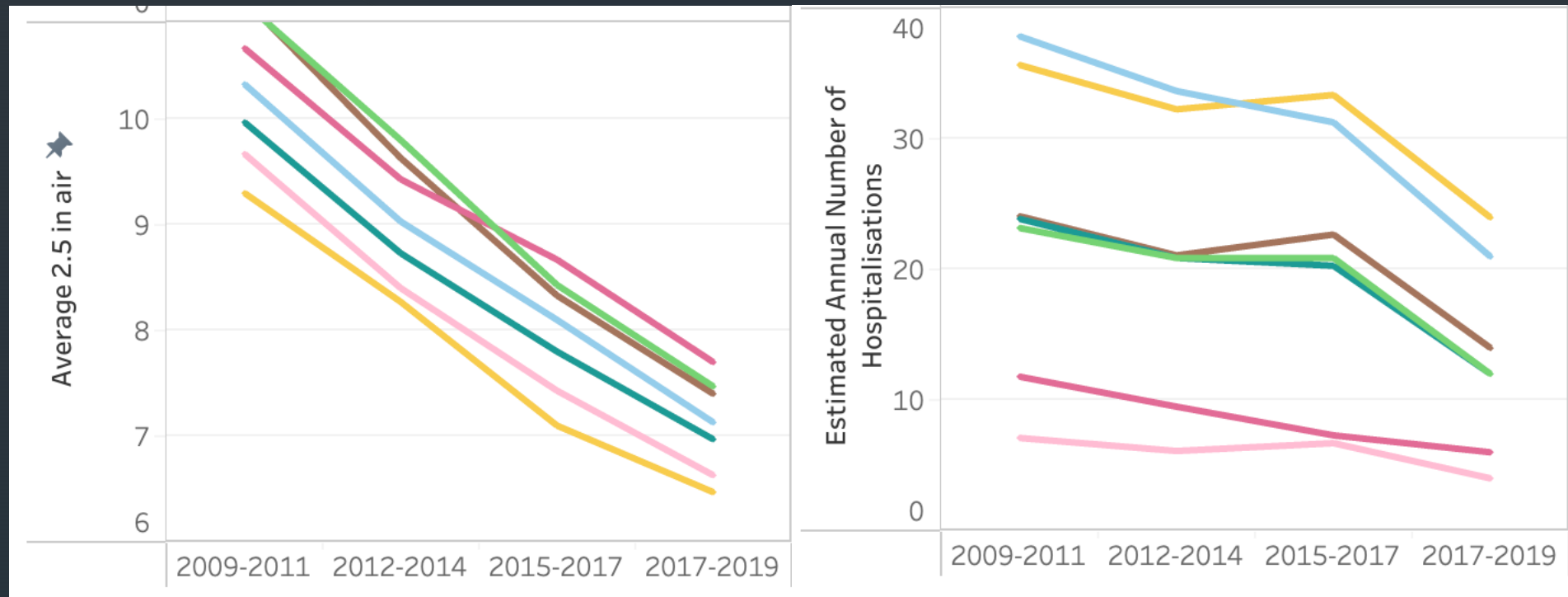
Results

Results

Results

Conclusion

The analysis demonstrates a clear link between PM2.5 concentrations and respiratory hospitalizations in New York City. While reductions in PM2.5 levels over time have led to decreased hospitalizations, the relationship is only partially explained by air pollution (R^2 = 0.29) suggesting that other factors also play a substantial role on the number of hospitalisations in an area so, further analysis is required to investigate additional variables like age distribution , pre-existing health conditions and availability of medical facilities and so on . Clustering analysis highlights the need to prioritize neighborhoods in Cluster 3, where high pollution levels and low hospitalizations indicate potential gaps in healthcare access. These findings emphasize the importance of addressing both environmental and systemic factors to improve respiratory health outcomes citywide.