Analysis of PM2.5 Pollution and Its Impact on Public Health in NYC (2009-2022)

Team Members:
Shubhi Vyas
Srishti RV Singh

Contents covered

Introduction

Problem statement:

Air pollution, particularly PM2.5, poses a **severe public health risk**, causing respiratory and cardiovascular diseases. Despite a 78% reduction in U.S. pollutants since 1970, **NYC remains heavily impacted**, with significant racial and neighborhood disparities. Our project aims to examines PM2.5 trends (2009–2022) and their **health impacts & neighborhood disparities** to highlight the need for targeted interventions.

Motive:

Despite reductions in overall air pollution, PM2.5 continues to **significantly impact public health**, particularly in densely populated cities like NYC. Vulnerable neighborhoods, often home to marginalized communities, face persistent disparities in air quality. This study seeks to **analyze PM2.5 trends**, link exposure to **health outcomes**, and generate **actionable insights** to inform policies addressing environmental justice and public **health improvement**.

Data Source

1. **Air Quality Dataset** (Geographical type, Air Quality indicators & measurements & Time period)

2. **Datasets on PM2.5 levels and hospitalizations** (Geographical type & Annual estimated cases)
   a) Asthma emergency hospitalizations
   b) Cardiovascular hospitalizations = CV_hospitalizations
   c) Respiratory hospitalizations = resp_hospitalizations

3. **Tools: Python**
   a) pandas, numpy: data manipulation
   b) matplotlib, seaborn: visualization
   c) Statsmodels: regression modelling

Research Goals & Hypotheses

Goals:

1. Assess air quality disparities across neighborhoods
2. Evaluate the trend in PM2.5 levels
3. Link PM2.5 levels to hospitalization rates(asthma, other respiratory illness & cardio-vascular diseases)

Hypotheses:

1. Neighborhoods with more people of color(population density) have worse air quality.
2. PM2.5 levels have decreased over time from 2009 to 2022.
3. Higher PM2.5 exposure leads to increased hospitalizations.

Data Cleaning and Preparation

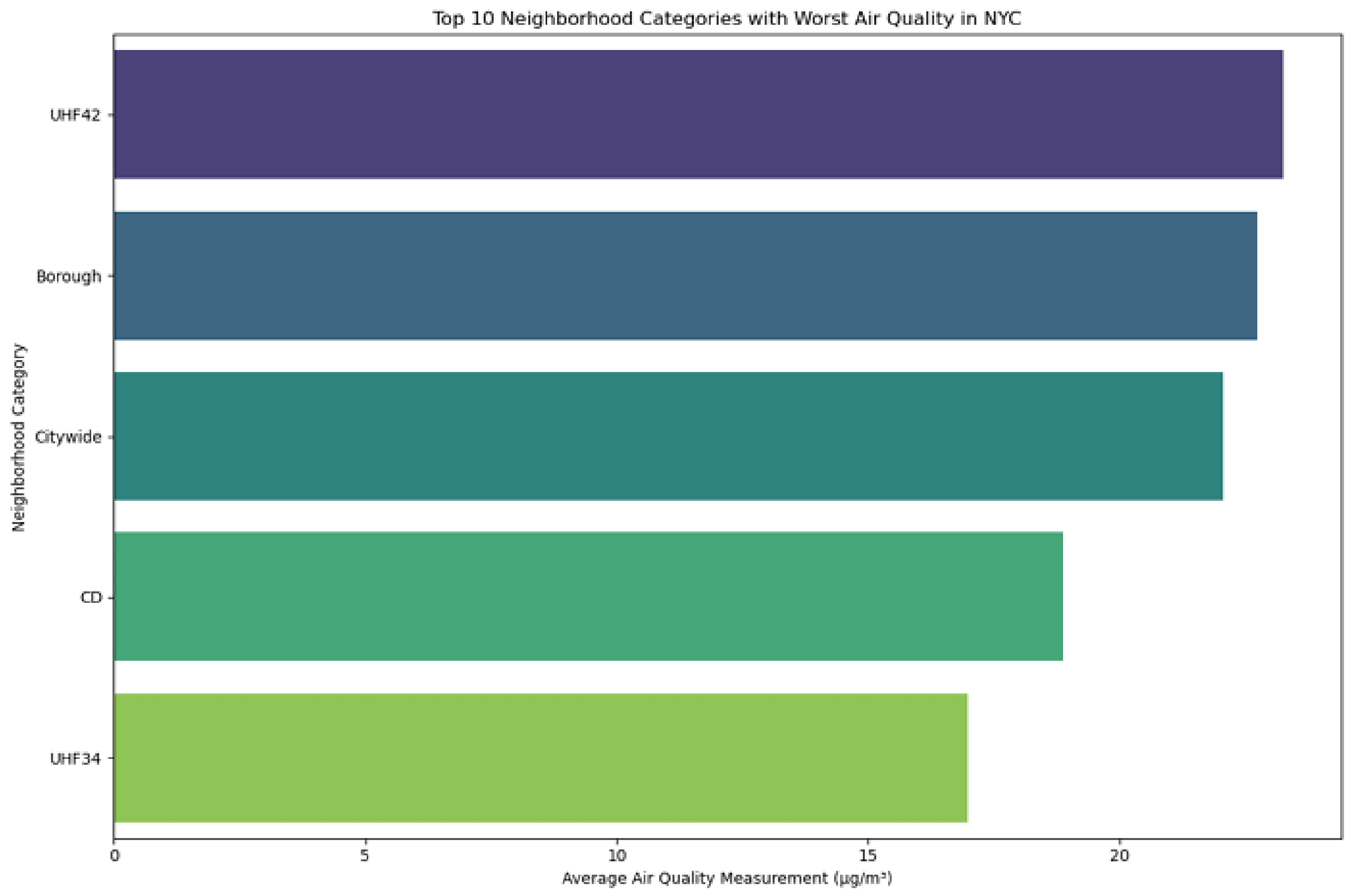
Air Quality Data Preparation

- Addressed missing values with **null value treatment based on geographic type**.
- Analyzed data for 2009–2022, excluding 2006–2008 due to incomplete **seasonality patterns**.
- Identified **PM2.5 as the primary pollutant** for analysis.
- Checked **data skewness** for distribution insights.

Asthma, Respiratory, and Cardiovascular Data Processing

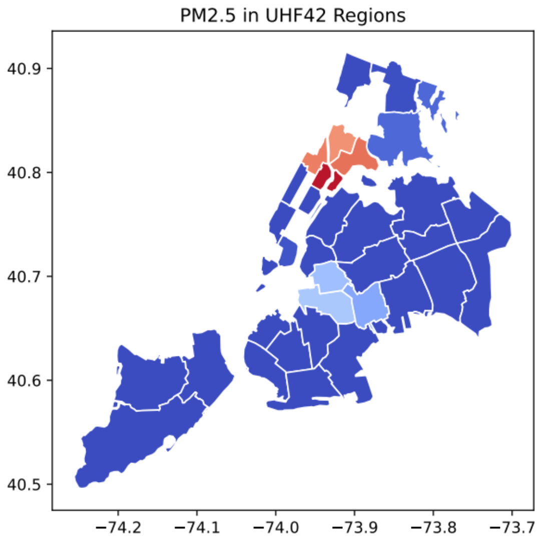
- **Integrated data** from three different sources.
- **Merged datasets** using geographic type (Geo Place Name) and year as common keys.
- **Extracted year** from date fields for **consistency**.
- Removed rows with **null values** across all three datasets.
- **Transposed** and merged data using pandas for a seamless analysis framework.

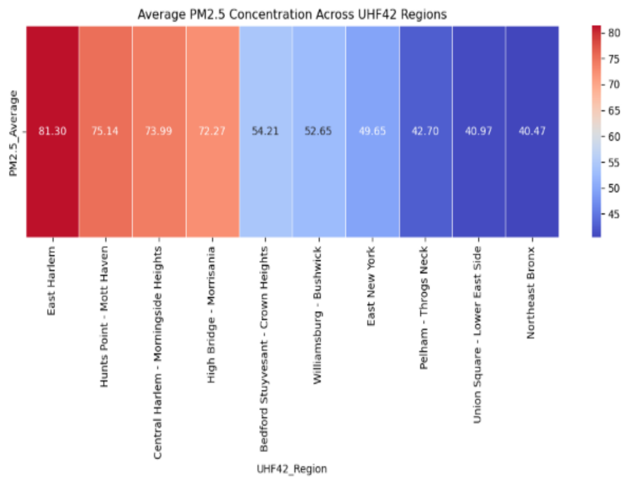
Results (H1 - Disparities in Air Quality)

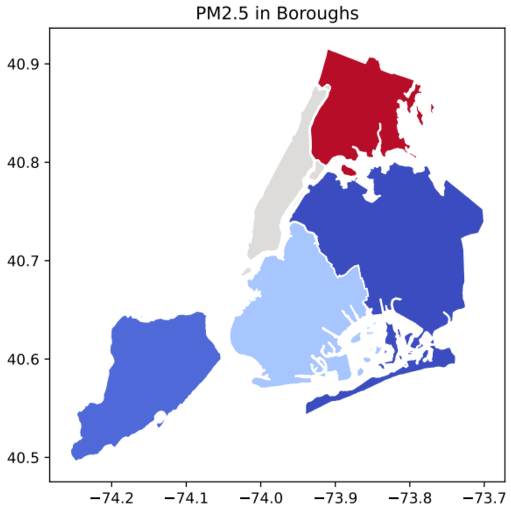


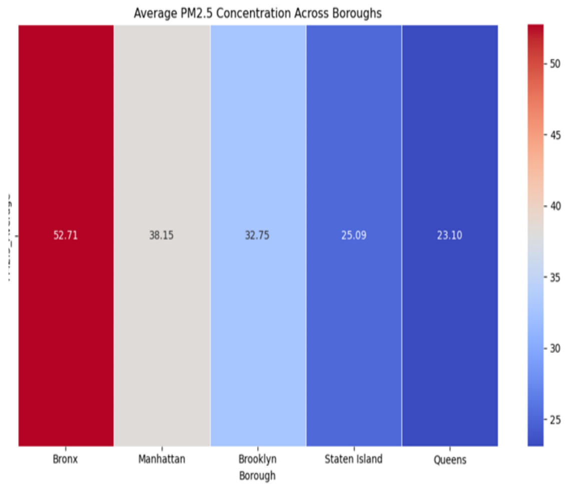
- Geospatial Analysis
- **Citywide** as **benchmark**
- Area of focus: **UHF42** and **Borough**
- UHF42 – United Hospital Fund, consisting of 42 city neighborhoods. Boundaries are based on the zip codes
- Borough – administrative units

Results (H1 - Disparities in Air Quality)



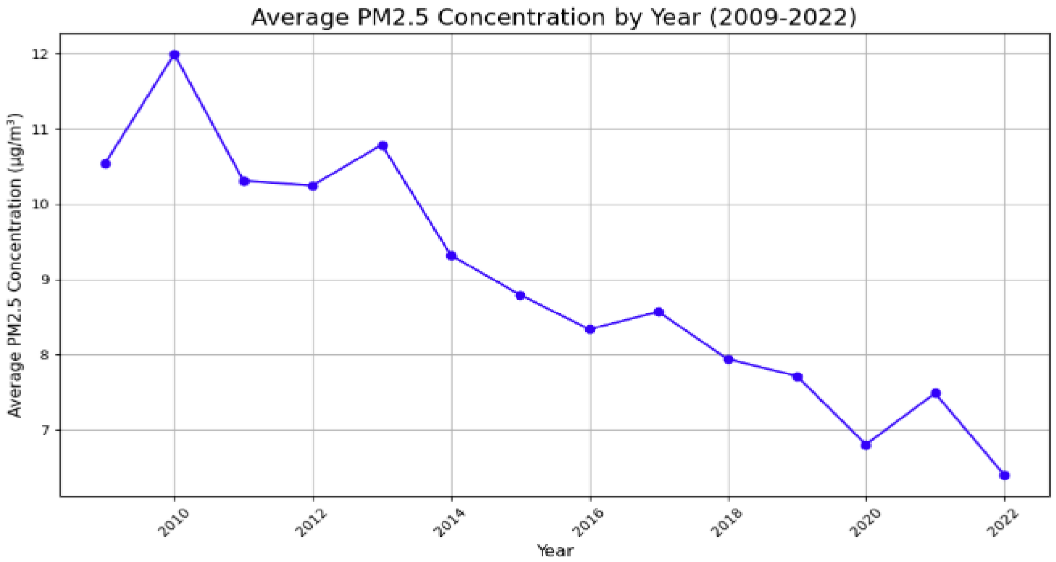


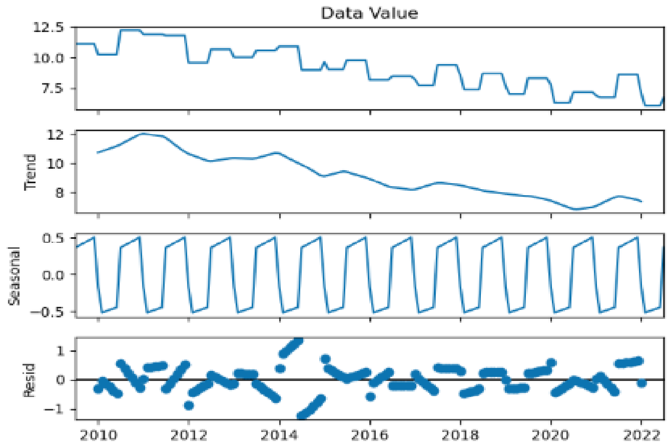




- Disparities in PM2.5 concentrations:
  - Higher levels in neighborhoods with people of color.
  - **Bronx**(Borough) and **East Harlem**(UHF42) among the worst-affected areas.
- These neighborhoods are historically **underserved and have higher minority populations** – like Hispanic and Blacks (US Census Data, 2020)
- Goldstein (1972) and Sicard *et al.* (2023) - **disadvantaged communities**, especially those **located near industrial or high-traffic zones**, bear the brunt of air pollution, leading to **chronic health issues**

Results (H2 - Time-Series Analysis)





- **General decline** in PM2.5 from **2009 to 2022**: indicating improvements in air quality over the years.
- While **seasonal patterns** are evident in PM2.5 as well through the years.
- **Residuals** suggest some unexplained variations due to **short-term events** (such as pollution spikes or specific events like wildfires or industrial accidents).

Results (H2 - Time-Series Analysis)

```
ADF Statistic: -0.614285084118284
p-value: 0.8677214026498994
The time series is non-stationary.
```

ADF Test 1 - Results

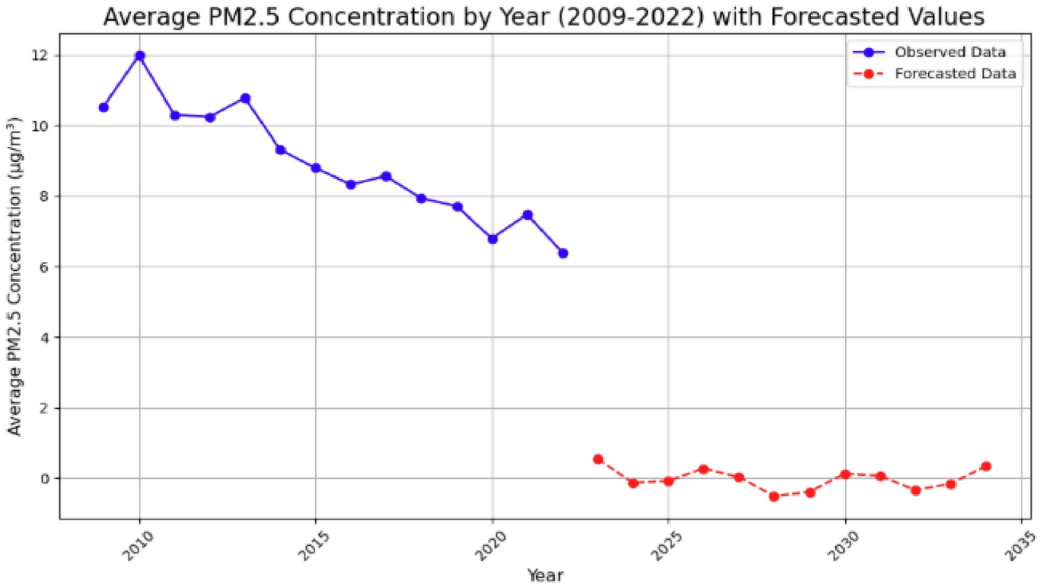
```
ADF Statistic: -8.593892429770564
p-value: 7.163528322744075e-14
The differenced time series is stationary.
```

ADF Test 2 - Results

Summary of ARIMA Model				
Parameter	Coefficient	Std. Error	z-value	p-value
AR(1)	0.6986	0.250	2.800	0.005
AR(2)	-0.7447	0.291	-2.560	0.010
MA(1)	-1.7351	0.195	-8.887	0.000
MA(2)	1.5390	0.278	5.527	0.000

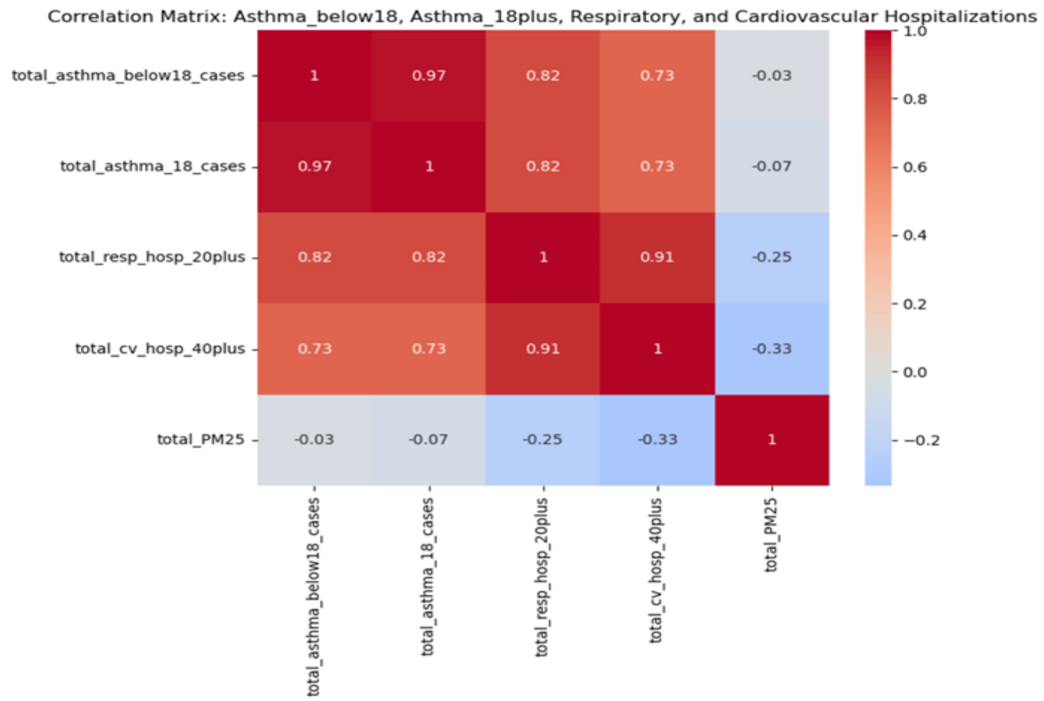
Approach:

- Stationary & Non-Stationary: ADF Test
  - Our data was non-stationary - First differencing to make data stationary
- ACF & PACF used to understand the structure of time series- **ARIMA Model(6,1,4)**
  - Forecasted the data



Results (H3 - Health Impacts)

Regression 1: Asthma Cases under 18				
Variable	Coefficient	Std Error	z-value	p-value
Constant	35.542	10.561	3.365	0.001
PM2.5	-0.6475	1.097	-0.590	0.555
Regression 2: Asthma Cases above 18				
Variable	Coefficient	Std Error	z-value	p-value
Constant	21.5462	5.312	4.056	0.000
PM2.5	-0.7720	0.552	-1.399	0.162
Regression 3: Asthma Cases Total				
Variable	Coefficient	Std Error	z-value	p-value
Constant	57.088	10.144	5.627	0.000
PM2.5	-1.420	1.054	-1.347	0.178
Regression 4: Cardiovascular cases (40+)				
Variable	Coefficient	Std Error	z-value	p-value
Constant	14.456	0.813	17.784	0.000
PM2.5	-0.114	0.084	-1.353	0.176
Regression 5: Other Respiratory cases (20+)				
Variable	Coefficient	Std Error	z-value	p-value
Constant	9.981	0.888	11.244	0.000
PM2.5	0.062	0.092	0.670	0.503



- **Correlation** between PM2.5 and hospitalizations
- Linear regressions – **results are non-significant**
- Need to incorporate additional variables

Discussion

- **Racial and Neighborhood Disparities**: Findings reveal that neighborhoods with higher populations of people of color face worse air quality (e.g., Kinney et al., 2002; Simkovich et al., 2019).
- **Declining Trends with Persistent Hotspots**: While PM2.5 levels have decreased citywide, persistent pollution hotspots, such as East Harlem and the Bronx
- **Health Impacts of PM2.5**: The association between PM2.5 exposure and hospitalizations aligns with studies linking fine particulate pollution to respiratory and cardiovascular illnesses (Manisalidis et al., 2020; Sicard et al., 2023). An aim should be made to focus on reducing particulate pollution.
- **Policy Implications**: The results highlight the need for targeted policies to reduce pollution in vulnerable areas and mitigate health risks, indicating the need for localized interventions and targeted interventions are necessary to address pollution hotspots
- **Broader Research Support**: Literature emphasizes how urbanization and industrial activity exacerbate pollution, reinforcing the importance of data-driven strategies to manage air quality (Dockery, 2012; Chen et al., 2024).

Conclusion, Limitation & Future Directions

Conclusion

- **Disparities exist** in the neighborhoods with a **higher population of people of color (H1)**
- PM2.5 trends have **reduced** over time from 2009 to 2022 **(H2)**
- **Insignificant results** to determine rising cases of respiratory illness and cardiovascular diseases due to PM2.5 **(H3)**

Limitations

- The analysis excluded **socio-economic factors** and other environmental variables that might influence hospitalization rates.
- Limited scope for **seasonal or short-term event analysis**, which may affect PM2.5 trends and health outcomes.

Future Directions

- Incorporate **additional variables**, such as income, occupation, and access to healthcare, for a more nuanced analysis.
- Conduct **seasonal studies** to identify periods of heightened risk and tailor interventions.
- Develop **predictive models** using real-time monitoring to proactively manage air quality and reduce health risks.

References

Chen, F., Zhang, W., Mfarrej, M. F. B., Saleem, M. H., Khan, K. A., Ma, J., Raposo, A. & Han, H. (2024). Breathing in danger: Understanding the multifaceted impact of air pollution on health impacts. Ecotoxicology and Environmental Safety, 280(116532). https://doi.org/10.1016/j.ecoenv.2024.116532

Cifuentes, L., Borja-Aburto, V. H., Gouveia, N., Thurston, G., & Davis, D. L. (2001). Assessing the Health Benefits of Urban Air Pollution Reductions Associated with Climate Change Mitigation (2000-2020): Santiago, São Paulo, México City, and New York City. Environmental Health Perspectives, 109, 419–425. https://doi.org/10.2307/3434790

Dockery, D. W. (2009). Health effects of particulate air pollution. Annals of epidemiology, 19(4), 257–263. https://doi.org/10.1016/j.annepidem.2009.01.018

Goldstein, I. F. (1972). Interaction of Air Pollution and Weather in Their Effects on Health. HSMHA Health Reports, 87(1), 50–55. https://doi.org/10.2307/4594426

Kinney, P. L., Chillrud, S. N., Ramstrom, S., Ross, J., & Spengler, J. D. (2002). Exposures to Multiple Air Toxics in New York City. Environmental Health Perspectives, 110, 539–546. http://www.jstor.org/stable/3455248

Manisalidis, I., Stavropoulou, E., Stavropoulos & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontier Public Health, Sec. Environmental Health and Exposome*, *8*. https://doi.org/10.3389/fpubh.2020.00014

Sicard, P., Agathokleous, E., Anenberg, S. C., Marco, A. D., Paoletti, E., Calatayud, V. (2023). Trends in urban air pollution over the last two decades: A global perspective. Science of The Total Environment, 858(2). https://doi.org/10.1016/j.scitotenv.2022.160064

Simkovich, S.M., Goodman, D., Roa, C., Crocker, M.E., Gianella, G.E., Kirenga, B.J., Wise, R.A., Checkley, W. (2019) The health and social implications of household air pollution and respiratory diseases. *NPJ Primary Care Respiratory Med Medicine*, *29*, 12

United States Environmental Protection Agency: (2024, April 30). *Progress Cleaning the Air and Improving People's Health*. https://www.epa.gov/clean-air-act-overview/progress-cleaning-air-and-improving-peoples-health

US Census Bureau: Population Division. (2020). 2020 Census Detailed Demographic and Housing Characteristics File A (DDHC-A). https://www.nyc.gov/site/planning/planning-level/nyc-population/2020-census.page

Appendix

1. New York City Department of Health, Environment & Health Data Portal. Health impacts of air pollution data. Asthma emergency department visits due to PM2.5. Accessed at https://a816-dohbesp.nyc.gov/IndicatorPublic/data-explorer/health-impacts-of-air-pollution/?id=2117 on 12/01/2024.
2. New York City Department of Health, Environment & Health Data Portal. Health impacts of air pollution data. Respiratory hospitalizations due to PM2.5 (age 20+). Accessed at https://a816-dohbesp.nyc.gov/IndicatorPublic/data-explorer/health-impacts-of-air-pollution/?id=2120 on 12/01/2024.
3. New York City Department of Health, Environment & Health Data Portal. Health impacts of air pollution data. Cardiovascular hospitalizations due to PM2.5 (age 40+). Accessed at https://a816-dohbesp.nyc.gov/IndicatorPublic/data-explorer/health-impacts-of-air-pollution/?id=2117 on 12/01/2024
4. City of New York. (2020, November 10). Dataset [Air Quality]. NYC OpenData. https://catalog.data.gov/dataset/air-quality
5. Organization for the NYC Environmental Health Services team at DOHMH. (2020). NYC Geography [UHF34]. Github. https://github.com/nycehs/NYC_geography
6. Organization for the NYC Environmental Health Services team at DOHMH. (2020). NYC Geography [UHF42]. Github. https://github.com/nycehs/NYC_geography
7. Veltman, N. (2018). Data [nyc-neighborhoods.geo.json]. Github. https://github.com/veltman/snd3/blob/master/data/nyc-neighborhoods.geo.json?short_path=730a833

Thank You!