Anish Gupta

anish.gupta@rutgers.edu

Number of Pages: 19

Air Quality Trends in the Neighborhoods of New York City for People Suffering from ASTHMA

**Abstract**

In New York City (NYC) thousands of patients visit the Emergency room, are hospitalized, and even face untimely death due to pollution related asthma, heart and lung problems. In this study I specifically look at a pollutant called PM2.5 and its impact on asthma. PM2.5 capable of getting deep inside a person’s lungs as well as into the bloodstream. These particles are the biggest threat to people’s health and contributes to 2000 deaths and 5150 ER visits per year in NYC. The segment of population most at risk is older adults, children, and people with preexisting conditions. This study will help stakeholders such as healthcare providers, first responders, patients, policy makers, city planners as it will help them identify high risk areas and vulnerable populations.

The main objective of this study is to provide stakeholders with the information to make crucial decisions related to identifying high risks areas in NYC with high levels of PM2.5, as well as areas with high rates of PM2.5 related asthma hospitalizations and deaths and then subsequently mitigating exposure to the pollutants thereby addressing the healthcare crisis and improving air quality.

This study has used spatial analysis and clustering using K-means clustering. The key findings of the study showed that Manhattan has the greatest levels of PM2.5 along with highest amounts of traffic. Overall, all 5 boroughs show decreasing levels of PM2.5. Results also showed that Bronx, with the second highest levels of PM2.5 has the highest amounts of emergency room visits for both children and adults as well as the highest amounts of death for adults. The Bronx has the lowest median household income.

**Keywords**

**PM2.5** - are small particles in the air that are <= 2.5 micrometers and are so small that they can be inhaled and cause serious health problems.

**K-means clustering** - is an unsupervised learning algorithm that groups data points into a specified number of clusters. The algorithm iteratively assigns data points to clusters based on their similarity to the cluster centers.

**Spatial Analysis –** According toArcGIS: “is the process of using analytical techniques to find relationships, discover patterns, and solve problems with geographic data.”

**Community District** – Neighborhood boundaries in New York City. There are 59 community districts in the city.

**t-map -** is an actively maintained open-source [R](https://stackoverflow.com/tags/r/info)-library for drawing thematic maps

**Tables**

Table 1 – National Air Quality Standards

**Figures**

Fig 1 – PM2.5 size comparison

Fig 2 – Species ratio by cluster

Fig 3 – PM2.5 concentration

Fig 4 - PM2.5 levels by Community Districts (CD)

Fig 5 – Traffic by CD

Fig 6 – PM2.5 by Borough Over Time

Fig 7 - PM2.5 Asthma ER visits per 100,000 Children

Fig 8 - PM2.5 Asthma ER visits per 100,000 Adults

Fig 9 - PM2.5 deaths per 100,000 Adults

Fig 10 – Median Household Income 2016

Fig 11 – Population by Race and Borough for 2016

Fig 12 – Scree Plot

Fig 13 – Household Income VS PM2.5 by Cluster

Fig 14 - Clusters from K-means

**Main Body**

**Introduction**

The Environmental Protection Agency defines air pollution as “any visible or invisible particle or gas found in the air that is not part of the natural composition of air.” Air pollution can come from many different places, some of which are created by humans and some of which are natural. Air pollution negatively impacts people’s health. One such health issue caused by air pollution is asthma. Studies show that air pollution can make asthma symptoms worse.

The system I wish to design will track air quality trends of the pollutant Particulate Matter (PM2.5) that impact people who suffer from asthma and other respiratory illnesses in the New York City area. The system will track how the air quality changes over time and will answer specific questions raised by the stakeholders of the system.

Some of the stakeholders of this system are healthcare providers, first responders, and people who suffer from respiratory illnesses.

**Research Questions**

What are the characteristics and trends in neighborhoods with poor air quality?

How do income levels affect the healthcare response to health issues related to poor air quality?

**Background and Literature Review**

**The Basics of PM2.5**

According to the article “Particulate Matter (PM) Basics”, from EPA’s website particulate matter (PM) as a mixture of solid particles and liquid droplets found in the air. These particles come in many sizes and can be harmful to human health. Some PM is emitted directly from sources like construction sites and fires, while other PM forms in the atmosphere. The government regulates PM pollution to protect public health. People can reduce their exposure to PM by checking air quality alerts and taking precautions when PM levels are high.

A diagram of a human hair

Description automatically generated

Fig 1

Particle pollution includes PM10 and PM2.5

PM10 are inhalable particles, with diameter <= 10 micrometers

PM2.5 are fine inhalable particles, with diameter <= 2.5 micrometers

Some of these particles are so small that they can be inhaled and cause serious health problems. Some of the particles < 10 micrometers are capable of getting deep inside a person’s lungs and also into the bloodstream. The particles known as PM2.5 are the biggest threat to people’s health.

**Health risk posed by PM2.5**

The article “Inhalable Particulate Matter and Health (PM2.5 and PM10)” defines PM2.5 and PM10 just like the previous article. Exposure to PM, especially PM2.5, poses significant health risks, including respiratory problems like aggravated asthma, chronic obstructive pulmonary disease (COPD), cardiovascular diseases like increased risk of heart attack, stroke, cancer such as potential link to lung cancer and premature death like shortened lifespan due to respiratory and cardiovascular complications.

Vulnerable Populations include children, older adults, individuals with pre-existing respiratory or cardiovascular conditions, and pregnant women are particularly vulnerable to the health effects of PM.

The article “The Public Health Impacts of PM2.5 from Traffic Air Pollution” discusses about exposure to traffic-related PM2.5 has been linked to a range of adverse health outcomes, including respiratory problems such as bronchitis, pneumonia, cardiovascular diseases like elevated chances of heart attack, stroke, cancer like the potential association with lung cancer and other types of cancer, premature death such as reduced lifespan due to respiratory and cardiovascular complications, cognitive decline such as studies suggest PM2.5 exposure might contribute to cognitive impairment and dementia, and birth defects like the potential link to low birth weight and other prenatal complications.

Research estimates that PM2.5 exposure contributes to millions of premature deaths globally each year, with a significant proportion attributable to traffic emissions. In the United States alone, traffic-related PM2.5 is estimated to cause thousands of premature deaths annually, resulting in substantial economic costs due to healthcare expenditures and lost productivity.

**How PM2.5 can be monitored and controlled?**

PM Regulations and Standards include US Environmental Protection Agency and the California Air Resources Board (CARB). The EPA sets national ambient air quality standards (NAAQS) for PM2.5 and PM10 to protect public health. CARB implements stricter standards for PM2.5 and PM10 in California, aiming to achieve cleaner air quality.

Strategies for Reducing PM Exposure include reduce PM emissions such as implement stricter emissions controls for vehicles and industries, promote cleaner fuels and energy sources, and individual actions such as limit exposure during high PM periods, utilize air purifiers, and stay informed about air quality forecasts.

A screenshot of a computer

Description automatically generated

Table 1

EPA regulates these inhalable particles. However, they do not regulate anything larger than 10 micrometers. EPA’s guidance help state and local governments to meet the agency’s national air quality standards. People can protect themselves from air pollution by using air quality alert systems. They can use AirNow that will tell them how healthy and safe their air is.

The article “A framework to spatially cluster air pollution monitoring sites in US based on the PM2.5 composition” by Austin et al mention that PM2.5 particles are a major component of air pollution and can have a negative impact on human health. They also note that the composition of PM2.5 particles can vary depending on the location.

It discusses how the sites can be grouped based on the composition of the PM2.5 particles. The authors used a k-means clustering algorithm to group the sites. They identified 31 clusters in total. The clusters were then validated and characterized based on various factors. The authors found that the clustering method was robust and produced meaningful results.

A map of the united states with different colored dots

Description automatically generated

Fig 2

Data was used from the U.S. Environmental Protection Agency (EPA) to perform the clustering. The data includes measurements of PM2.5 mass and various concentrations of PM2.5 components such as sulfate, nitrate, ammonium, organic carbon, and elemental carbon.

Many methods were used to validate the clusters and these included silhouette analysis and gap statistics. The clusters were characterized based on PM2.5 components and other factors like the location of the monitoring sites. Other important details of this research include – the authors utilized the Positive Matrix Factorization (PMF) model to determine the sources of PM2.5 in each cluster, the authors found that the clusters were associated with different types of air pollution sources, such as motor vehicles, coal-fired power plants, and secondary organic aerosols.

In the article “Spatial Enablement to Support Environmental, Demographic, Socioeconomics, and Health Data Integration and Analysis for Big Cities: A Case Study With Asthma Hospitalizations in New York City”, the authors explored a method using spatial clustering analysis using ArcGIS grouping analysis tool to predict asthma hospitalization rates. They used New York City data as a case study. They found a connection between asthma and several factors, including air pollution, poverty, and obesity. The authors used a geographically weighted regression model to evaluate the relationship between asthma hospitalization rates and the difference between various environmental and socioeconomic factors.

A map of the united states

Description automatically generated

Fig 3

They found that air pollution, poverty, and obesity were all significantly associated with asthma hospitalizations. They also found that the relationship between these factors and asthma hospitalizations varied across different geographic areas of the city.

The authors suggest that their findings can be used to develop more targeted interventions to reduce asthma hospitalizations in cities. They also suggest that spatial analysis can be a valuable tool for studying the complex relationships between environmental and socioeconomic factors and health outcomes.

**Challenges and Future Directions for traffic related to PM2.5**

The article “The Public Health Impacts of PM2.5 from Traffic Air Pollution” also specifies that despite the existing knowledge and available strategies, numerous challenges remain in effectively reducing traffic related PM2.5 and mitigating its health impacts. Some key challenges include policy implementation like ensuring effective implementation of emissions regulations and investing in infrastructure, addressing legacy infrastructure like transitioning away from older, high-polluting vehicles and infrastructure that contribute significantly to PM2.5 emissions, and addressing global disparities like low and middle income countries often deal with greater challenges in implementing effective PM2.5 reduction strategies due to limited resources and capacity.

Future research is needed to develop more comprehensive air quality models like refine the understanding of the complex interactions between traffic emissions and other sources of PM2.5, investigate long-term health effects such as conduct longitudinal studies to better understand the long-term health consequences of PM2.5 exposure, and evaluating the effectiveness of interventions such as implementing robust monitoring and evaluation frameworks to assess the impact of policies and strategies aimed at reducing traffic related PM2.5.

**Data Source**

I am using the Air\_Quality dataset and it contains the variables with the pollutant name, quantity, neighborhood, year, time period, etc. The Air\_Quality dataset was retrieved from Data.gov. I joined this dataset with the Total\_Population\_by\_Race\_Ethnicity\_NYC, Median\_Income\_NYC, and Total\_Population datasets. I got these three datasets from the Citizens’ Committee for Children of New York webpage.

**Methods and Results**

The tools used in this study are as follows –

* ggplot2 to generate graphs to analyze the data
* Spatial analysis to study PM2.5 levels by community districts in NYC.
* K-means clustering to take a look at the different clusters of community districts by PM2.5 levels, population, income, demographics.

PM2.5 levels by Community Districts

A map of a city

Description automatically generated

Fig 4

Looking at PM2.5 levels in different community districts, CDs with the highest PM2.5 levels are in Manhattan followed by some CDs in Brooklyn, Queens and Bronx.

Traffic-Density Annual Vehicle Million Miles Traveled Per km2

A map of a city

Description automatically generated

Fig 5

The above map shows the Traffic-Density Annual Vehicle Million Miles Traveled Per km2

According to the NYC.gov EH data portal, traffic is a major source of PM2.5. So I looked at the traffic information. Unfortunately the latest information I have is from 2016. It does show that Manhattan CDs have the highest levels of traffic.

A graph of different colored lines

Description automatically generated

Fig 6

The good news is PM2.5 levels has been constantly decreasing in all boroughs in New York City. Manhattan is the brough with the highest level of PM2.5.

A graph of a patient's emergency room

Description automatically generated

Fig 7

PM2.5 attributable asthma emergency room visits for children has been decreasing in every borough. The borough with the highest number of PM2.5 attributable asthma related emergency room visits for children is Bronx.

A graph of emergency room visits per 100, 000 adults

Description automatically generated

Fig 8

PM2.5 attributable asthma emergency room visits for adults has been decreasing in every borough. The borough with the highest number of PM2.5 attributable asthma related emergency room visits for adults is Bronx.

Overall, the number of asthma related emergency room visits is greater for children than adults.

A graph of different colored bars

Description automatically generated

Fig 9

PM2.5-Attributable Deaths per 100,000 Adults has been decreasing in each borough. The two boroughs with the highest number of PM2.5 attributable deaths per 100,000 adults is Bronx and Staten Island

A graph of a number of household income

Description automatically generated

Fig 10

We can see from the above bar graph that the borough with the lowest median household income for 2016 is Bronx. Therefor the borough with the highest PM2.5 attributable hospital visits is also the borough with the lowest income.

A graph of different colored lines

Description automatically generated

Fig 11

The borough with the lowest income is Bronx and has high population of people that are Hispanic or Latino.

Scree Plot

A graph of a number of clusters

Description automatically generated

Fig 12

For K-means clustering a scree plot was created. It looks like 4 is a good number for clusters.

A graph of a graph showing a number of people

Description automatically generated with medium confidence

Fig 13

The K-mean cluster plot shows that most of the community districts with higher level of PM2.5 are in Manhattan which is the second highest borough in terms of income. However it is important to note that the community districts in Manhattan that have lower levels of PM2.5 have the highest income. Lower income CDs in Manhattan have high levels of PM2.5.

Clusters from K-means

**A map of a city

Description automatically generated**

Fig 14

Plotting the clusters from K-means, shows that CDs with similar characteristics in terms of income, ethnicity and PM2.5 levels are grouped together in similar nearby neighborhoods. Manhattan in cluster 2 has higher PM2.5 levels.

**Key Findings in Context of Research Questions**

Here are the key findings in terms of the research question.

What are the characteristics and trends in neighborhoods with poor air quality, and/or high levels of hospitalizations and deaths in terms of PM2.5?

Some neighborhoods like the ones in Manhattan have some of the highest levels of PM2.5 and have some of the highest levels of traffic in New York City. Manhattan also has the highest average income. Bronx has the lowest average income, but the highest rates of emergency room visits although the PM2.5 level in Bronx is lower than Manhattan. Bronx also has a high percentage of Hispanic and African American population. Children visit the emergency room for PM2.5 related asthma in these neighborhoods at a much higher rate than adults. My most interesting finding is that Staten Island which has a high average income and low PM2.5 levels has the most PM2.5 attributable deaths.

How do income levels affect the healthcare response to health issues related to poor air quality?

It is clear that income levels and demographics play a major role in terms of access to healthcare and disease management. Although Manhattan community districts have a high level of PM2.5, their rates of hospitalization and deaths related to PM2.5 is much lower than community districts in boroughs with lower PM2.5. This clearly shows that low income equates to less access to quality healthcare. This calls for better health insurance, better clinics, and hospitals in poorer neighborhoods.

**Summary**

These results suggest that having higher levels of pollutants in the air does equate to greater amounts of critical illness or death at least in certain areas like Bronx. In addition, in Bronx which is a low-income area a with large Hispanic and African American population has some of the highest levels of critical illness or death. We can see from the study that higher levels of PM2.5 in the air has a direct correlation with the amount of traffic as we have seen in the case of Manhattan. The areas with the greater amounts of pollutants like Manhattan are the same areas with the highest median household income. This suggests that having better paying jobs and higher income allows people to access better healthcare and remedy health problems more effectively. They have better health insurance, better clinics, hospitals etc. Some of the implications of this study include better access to healthcare and more clinics in poorer neighborhoods can improve the situation with PM2.5 related asthma. Reduced drug prices for lower income families will make drugs like inhalers more affordable for people suffering from asthma. This improved level of access to better healthcare will reduce the number of PM2.5 related hospital visits and deaths for people living in poorer neighborhoods.

**Future Research**

Further research is needed for us to understand why Staten Island has the highest death rate attributable to PM2.5. We need to perform a deeper analysis of the area. We can take a deeper dive and look at income, demographics, and pollution levels at the census tract or even block group levels to gain a better understanding of why Staten Island with a high average income has the highest death rate among the boroughs in New York City.

**References**

Pala, D., Pagán, J., Parimbelli, E., Rocca, M. T., Bellazzi, R., & Casella, V. (2019). Spatial enablement to support environmental, demographic, socioeconomics, and Health Data Integration and analysis for big cities: A case study with asthma hospitalizations in New York City. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6499162/

(2021). Retrieved from https://a816-dohbesp.nyc.gov/IndicatorPublic/data-stories/traffic-and-air-pollution/

Austin, E., Coull, B. A., Zanobetti, A., & Koutrakis, P. (2013). Retrieved from https://www.sciencedirect.com/science/article/pii/S0160412013001219

(N.d.). Retrieved from https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health#:~:text=Emissions%20from%20combustion%20of%20gasoline,a%20significant%20proportion%20of%20PM10

(2023). Retrieved from <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>

United States Environmental Protection Agency. (2023). Retrieved from https://www.epa.gov/outdoor-air-quality-data

(2023). Retrieved from https://www.epa.gov/air-trends

(2022). Retrieved from https://www.cdc.gov/nceh/tracking/topics/AirQuality.htm#anchor\_1606247930783.

(2023). Retrieved from https://catalog.data.gov/dataset/air-quality

(2021). Retrieved from https://data.cccnewyork.org/data/