***What Contributes to Asthma Disparity***

***in New York City?***

Team 45 Final Report

Hevander Da Costa

Veronica Venture

Yohanna Georgis

Sina Hashemi

*August 2021*

1. **INTRODUCTION: 4**

1.1 BACKGROUND [4](#_c8ajtxus52nu)

1.2 PURPOSE OF REPORT [4](#_c8ajtxus52nu)

1.3 PURPOSE OF REPORT [4](#_c8ajtxus52nu)

1.4 OUR MODEL [4](#_c8ajtxus52nu)  
 1.5 IMPLICATIONS 4

1. [**DATASET DESCRIPTION & CLEANING PROCESS . 5**](#_p5x7k92o4k5s) 2.1 OUTDOOR AIR QUALITY 5  
    2.2 INDOOR AIR QUALITY6  
    2.3 ASTHMA6  
    2.4 SOCIAL DETERMINANTS OF HEALTH7
2. **EXPLORATORY DATA ANALYSIS**[**.**](#_p5x7k92o4k5s) **8** 3.1 OUTDOOR AIR QUALITY 8  
    3.2 INDOOR AIR QUALITY14  
    3.3 SOCIAL DETERMINANTS OF HEALTH16
3. **STATISTICAL ANALYSIS & PREDICTIVE MODELING**[**.**](#_p5x7k92o4k5s) **22** 3.1 OUTDOOR AIR QUALITY REGRESSION 22  
    3.2 OUTDOOR AIR QUALITY CHI SQUARED27  
    3.3 INDOOR AIR QUALITY REGRESSION28
4. [**DASHBOARD DESCRIPTION 3**](#_z8yt4w1aupei)**0**
5. [**CONCLUSION 3**](#_ohvpqclwxtow)**1**
6. [**RECOMMENDATIONS 3**](#_uaom9dnpsjm)**2**
7. [**LINKS AND REFERENCES**](#_jnbba4r97qnf) **33**

**LIST OF FIGURES**

Figure 1 Changes in Air Quality in NYC Over Time8

Figure 2 Sulfur dioxide from 2009-20149

Figure 3 Locations with Highest Asthma Rates9

Figure 4 Pollutant Levels in Neighborhoods with High Asthma Rates10

Figure 5 Fine Particulate Matter levels in the Five Neighborhoods with Highest Asthma Rates11

Figure 6 Pollution levels within highest asthma boroughs vs. areas with the highest pollution12

Figure 7 Five neighborhoods with the highest emergency department visits for asthma13

Figure 8 Top five neighborhoods with the most complaints of each type14

Figure 9 Poverty Levels15

Figure 10 Poverty in high asthma areas neighborhoods16

Figure 11 Top asthma regions and high Poverty16

Figure 12 Highest rates of households with no health care insurance17

Figure 13 Race/Ethnicity distribution18

Figure 14 Poverty and demographics18

Figure 15 Matrix of scatter plots Pollutants vs. Emergency visits21  
Figure 16 QQ plots for each pollutant 22

Figure 17 Log QQ plots for each pollutant23

Figure 19 Log Matrix of scatter plots Pollutants vs. Emergency visits27

Figure 20 Dashboard Description29

**LIST OF TABLES**

Table 1 Average of the mean yearly pollutant levels……………………………………………. 10

Table 2 Race/Ethnicity Graphs in NYC (Occurence of Poverty).................................................. 19

Table 3 Race/Ethnicity Graphs in Top Asthma Regions 19  
Table 4 Intra-Race Poverty by Region 20  
Table 5 Regression model Combinations 23  
Table 6 Best Regression Model 25  
Table 7 Bivariate frequency table 26  
Table 8 Regression model using remaining variables 28

**1. INTRODUCTION**

#### **1.1 BACKGROUND**

Asthma is widely recognized as a growing public health concern in urban areas. Although the exact cause of asthma is unknown, this is primarily believed to result from residents in urban areas having higher exposure to air pollutants and a variety of allergens compared to those in rural and suburban areas ([1a](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5523938/#R7), [2a](https://www.publichealth.columbia.edu/research/columbia-center-childrens-environmental-health/asthma)). New York City in particular has relatively high asthma rates in comparison to the national average. This high asthma burden does not affect all neighborhoods in the city equally, with certain neighborhoods having significantly higher asthma rates than others. Some community members and experts have suggested that this is due to higher environmental pollutant exposure in Black and Brown and low income communities, primarily due to discrimination in infrastructural and environmental policy making. These communities are often the target for busy roadways, toxic waste facilities, and a variety of industrial and commercial sites.

##### **1.2 PURPOSE OF REPORT**

The purpose of our project was to analyze New York City data on asthma contributors and Social Determinants of Health to uncover what potentially drives asthma disparity in this city. Although there are a wide variety of potential asthma contributors, for this project we focused on indoor and outdoor air quality because they are widely believed to be the main contributors. We also decided to examine social determinants of health, which are non-medical factors that influence health and well-being, because they could potentially play a role in asthma development. For the purpose of our project, we focused on race, income and access to health care as social determinants of health.

#### **1.3 OUR MODEL** Our model had two distinct parts. In the first part, we did exploratory data and regression analysis to examine the relationship that asthma contributors and social determinants of health had on asthma rates in the city. For the second part of our model, we made an interactive dashboard that includes a map of New York City that highlights neighborhoods at-risk for asthma.

#### **1.4 IMPLICATION & IMPORTANCE**

The ultimate purpose of this project is to inform change in the modifiable factors that contribute to asthma rates in NYC. The findings of our project can provide insight into which combinations of air quality, infrastructure, and social determinants of health by neighborhood conduce the exacerbation of asthma, guiding future investment and planning. Our data is applicable to a variety of industries that can improve confounding factors to asthma occurrence such as technology companies with a focus on air quality, healthcare and public health institutions, and policy makers.

**2. DATASET DESCRIPTIONS AND CLEANING PROCESS**

We used four distinct groups of datasets separated by categories. The categories are outdoor air quality, indoor air quality, asthma rates, and social determinants of health. The data for indoor and outdoor air quality and asthma rates was divided by regions called UHF- 42 Neighborhoods. These are the 42 United Hospital Fund Neighborhoods that have boundaries based on zip codes. These neighborhood boundaries were outlined by the health department in the 1980s and are meant to closely match community districts ([3a](https://a816-dohbesp.nyc.gov/IndicatorPublic/Closerlook/geographies/index.html)). The neighborhoods for datasets related to social determinants of health are divided by community districts. New York City has 59 community districts with several regions overlapping multiple UHF neighborhoods.

#### **2.1 OUTDOOR AIR QUALITY**

Dataset Summaries:

NYC Open Data has several datasets related to air pollutants. The datasets used in this project contain information about the following air pollutants: fine particulate matter, nitrogen dioxide, ozone, and sulfur. The data was originally stored in three different tables separated by different types of air pollutants:

* [Ozone](https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=2027,719b87,122,Summarize) : This dataset includes information about the air pollutant ozone which is measured in parts per billion. The measurements represent the yearly mean amount of ozone in each of the 42 UHF neighborhoods from 2009 to 2019.
* [Fine Particulate Matter and Nitrogen Dioxide](https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r) : This dataset includes information about fine particulate matter measured in micrograms per cubic meter, and nitrogen dioxide measured in parts per billion. The measurements represent the yearly mean amount of each pollutant in each of the 42 UHF neighborhoods from 2009 to 2018.
* [Sulfur](https://data.cityofnewyork.us/Environment/Air-Quality/fyf4-hrcu) : This dataset includes information about sulfur. Sulfur is measured in parts per billion. The measurements represent the yearly mean of sulfur in each of the 42 UHF neighborhoods for years 2008 to 2015.

Cleaning Process:

The cleaning process for these datasets involved filtering the data to only include information about the 42 UHF neighborhoods. The data was originally broken down by different types of neighborhoods such as community district, borough and UHF 34 neighborhoods. The Pandas library in Python was used to create new datasets that only included information for the desired neighborhoods. Another part of the process involved transforming the datasets from a ‘narrow’ format with data about each pollutant stacked on top of each other to a ‘wide’ format where there is a separate column with the mean of each pollutant for each neighborhood.

**2.2 INDOOR AIR QUALITY**

Dataset Summaries:

The source of the data set is NYC Open Data. The original data set contained location data comprising the borough to address level and also coordinate location. Each row contained an indoor air quality related complaint per address. Variables related to the complaint such as date, descriptor and complaint status are also included.

* [DOHMH Indoor Environmental Complaints](https://data.cityofnewyork.us/Health/DOHMH-Indoor-Environmental-Complaints/9jgj-bmct/data) : This data is largely received via complaints to 311; Each record represents a single complaint; This data can be used to help determine if the DOHMH has received an environmental complaint from a particular address; The listing of a particular address does not indicate that a condition or violation was found, only that a complaint was made to DOHMH. The data set spans 11 years from July 2010 to July 2021
* [Listing Complaints](https://github.com/paulvaldes/group_45_ds4a-project/blob/main/listing_complaints.csv) : This data set is derived from the DOHMH Indoor Environmental Complaints dataset. The data contains counts of complaints per year for each neighborhood. This data set spanned 11 years. Most address information besides borough and neighborhood was stripped.

Cleaning Process:

1. The DOHMH Indoor Environment Complaints data set was first stripped of extraneous variables relating to location and court district. A dictionary was created to match and filter the zipcodes for those only associated with UHF 42 neighborhoods. It was finally queried to form specified datasets that only contained one complaint type. The individual datasets were saved as csv files.
2. The aforementioned data sets were then used to derive the Listing Complaints data set. First the “Date\_Recevied” variable was converted to datetime and changed to only display year per row. Then the data was grouped by “UHF\_42” , “Incident\_Address\_Borough”, “Date\_Received”. Another column “{complaint type}\_ count” was created to store the counts per year for every neighborhood. The datasets were saved individually as csv files. The datasets were then outer joined to display the complaint types as individual columns along with UHF\_42 neighborhood, borough, and date.

#### **2.3 ASTHMA**

Dataset Summaries:

The Environment & Health Data Portal is a tool run by the NYC Department of Health and Mental Hygiene. Its purpose is to share data and analyze trends on how the environment affects health. For our project we have selected and cleaned a variety of datasets all pertaining to asthma burden in NYC. We have datasets in three categories: Emergency Department Visits, Hospitalizations, and Prevalence.

* [Emergency Department Visits](https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=2380,4466a0,11,Summarize) : This dataset includes information on the number and rate of adults that have visited the emergency room for a reason asthma-related by each UHF 34 neighborhood. The data represents years 2003-2017.
* [Hospitalizations](https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=2382,4466a0,11,Summarize) : This dataset includes information on the number and rate of asthma-related hospitalizations among adults in each UHF 42 neighborhood. This data represents years 2005-2016.
* [Prevalence](https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=18,4466a0,11,Summarize) : This dataset includes information on adults who were medically diagnosed with asthma and symptoms in the past 12 months. This dataset represents years 2003-2017.

Cleaning Process:

These datasets were cleaned by first removing any columns or rows with metadata. The columns were also renamed to be more precise, commas were removed from all the numerical values, and the data in each column was changed to the appropriate data type.

#### **2.4 SOCIAL DETERMINANTS OF HEALTH (SDOH)**

Dataset Summaries:

Citizens’ Committee for Children of New York maintains a comprehensive [compilation](https://data.cccnewyork.org/data/download#0,4,8,10,13/99) of data on various dimensions of well-being across New York City’s 59 community districts. This tool aims to provide a better understanding of where risks to well-being may concentrate. In this project, we selected and cleaned a variety of datasets pertaining to [SDOH that contribute to asthma](https://www.healthify.us/healthify-insights/addressing-sdoh-needs-care-coordination) in urban areas. The three main variables we assessed fall into the SDOH categories of economic stability, social context, and the health care system, respectively:

* Income Levels & [Poverty](https://data.cccnewyork.org/data/download#0,4,8,10,13/99): This dataset includes geographical information (by Fips code and neighborhood) on the income of households in New York City that fall into each category of the Federal Poverty Level (FPL) in the USA.
* Below 100% FPL = Below $25,750
* 100 to 199% FPL = $25,750 to $51,499
* Below 200% FPL = Below $51,500
* 200 to 399% FPL = $51,500 to $102,999
* 400% FPL and above = $100,300 and above
* [Poverty & Ethnicity/Race](https://data.cccnewyork.org/data/download#0,4,8,10,13/99,1423): This dataset includes geographical information (by Fips code and neighborhood) on the income of households in New York City that fall into each category of the Federal Poverty Level (FPL) in the USA divided by ethnic groups: Asian, White, Black, and Latino.
* [Health Care Insurance](https://data.cccnewyork.org/data/download#0,4,8,10,13/0,86): This dataset includes geographical information (by Fips code and neighborhood) on the households insured for healthcare in New York City.

Cleaning Process:

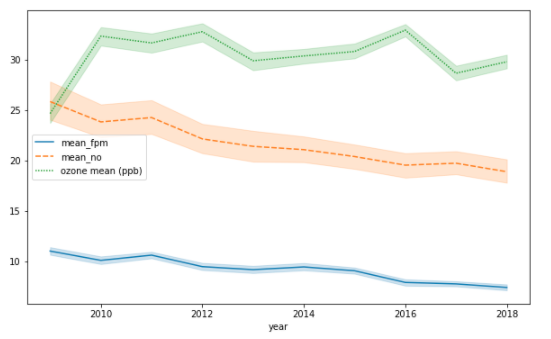
The datasets were cleaned by first removing the rows that had percent dataforms. Then, all rows that included total counts for each of the 5 NYC boroughs were removed to prevent double-counting.

**EXPLORATORY DATA ANALYSIS**

**OUTDOOR AIR QUALITY**Background on Effects of Outdoor Air Quality  
The pollutants included in the air quality report are ground-level ozone, fine particulate matter, nitrogen dioxide, and sulfur. Ground-level ozone is created when nitrogen oxides and volatile organic compounds react in sunlight ([5a](https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics) ). High levels of ozone usually occur in summer months and can cause asthma attacks ([5b](https://www.epa.gov/ground-level-ozone-pollution/health-effects-ozone-pollution)). Nitrogen dioxide is an air pollutant formed when fossil fuels are burned at high temperatures. It is known to cause a greater risk for asthma attacks and an increased likelihood of emergency department visits and hospitalizations ([6c](https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/nitrogen-dioxide) ). Fine particulate matter refers to a mixture of tiny solids (2.5 micrometers or smaller) and liquid droplets in the air ([5c](https://www.epa.gov/pm-pollution/particulate-matter-pm-basics)). These tiny particles are hazardous and can get into human lungs causing increased risk of asthma ([5d](https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm) ). Sulfur dioxide is a gaseous air pollutant created when fuel containing sulfur is burned. Sulfur dioxide also converts to sulfates, in the atmosphere, which are a large component of fine particulate pollution ([6a](https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/sulfur-dioxide) ).

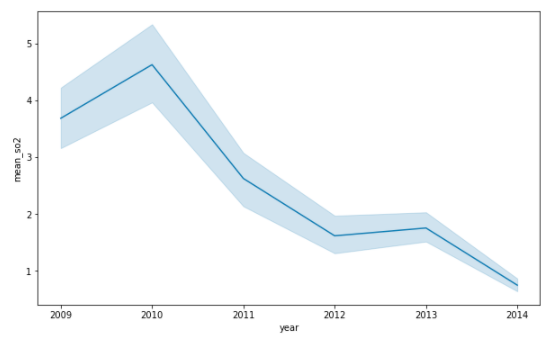
**Changes in Air Quality in NYC Over Time**

The line plot below shows the changes in mean fine particulate matter (fpm\_mean), mean nitrogen dioxide (mean\_no) and mean ground ozone ( ozone mean (ppb)).

****

The line plot shows the changes in air pollutants from 2009 to 2018. We can see that nitrogen dioxide and fine particulate matter levels (indicated by the orange and blue lines respectively) have declined over the years. Ground ozone levels have varied over time; there is not a clear trend in the changes of ozone levels.

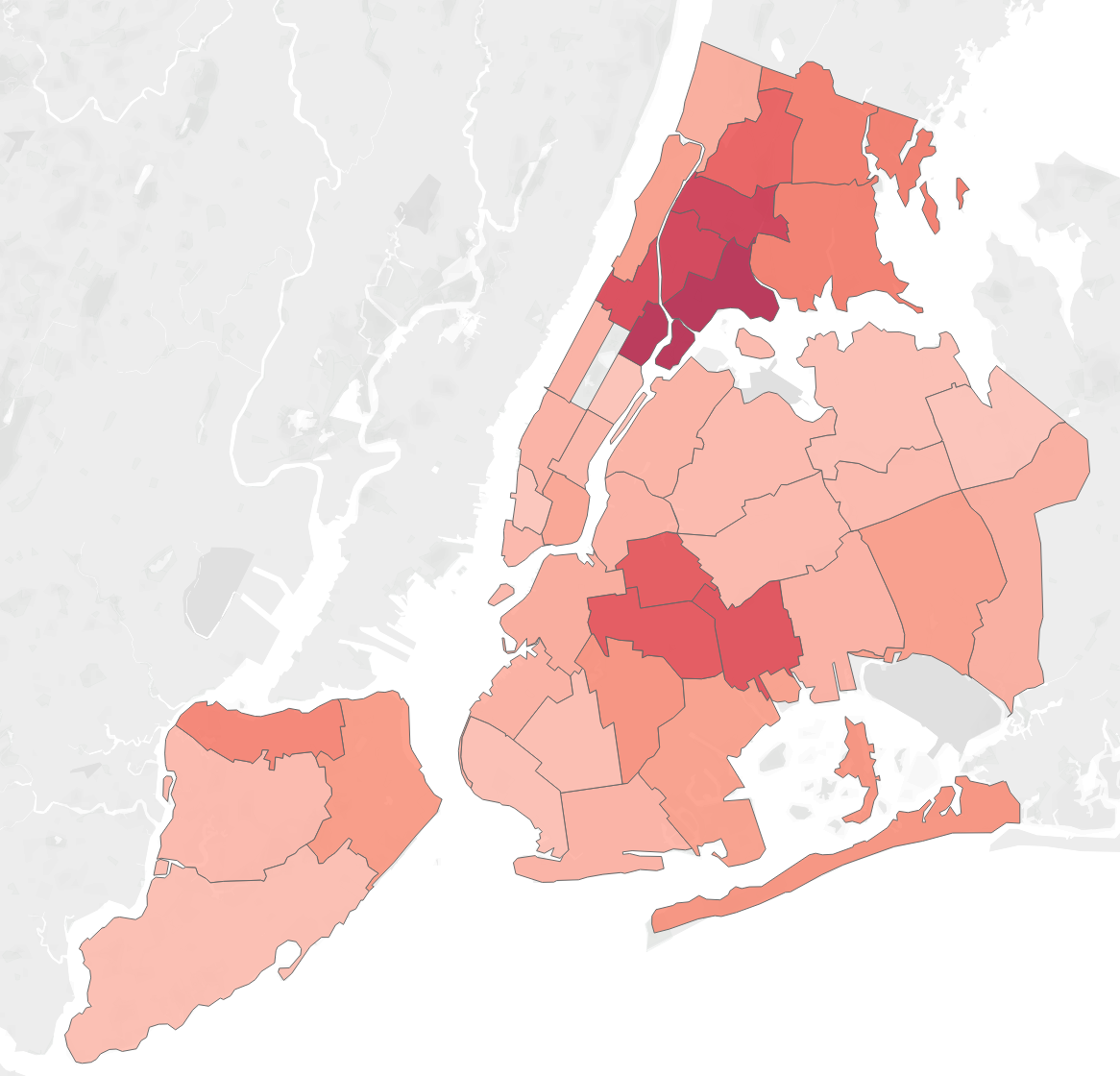
The line plot below shows changes in sulfur dioxide from 2009-2014.



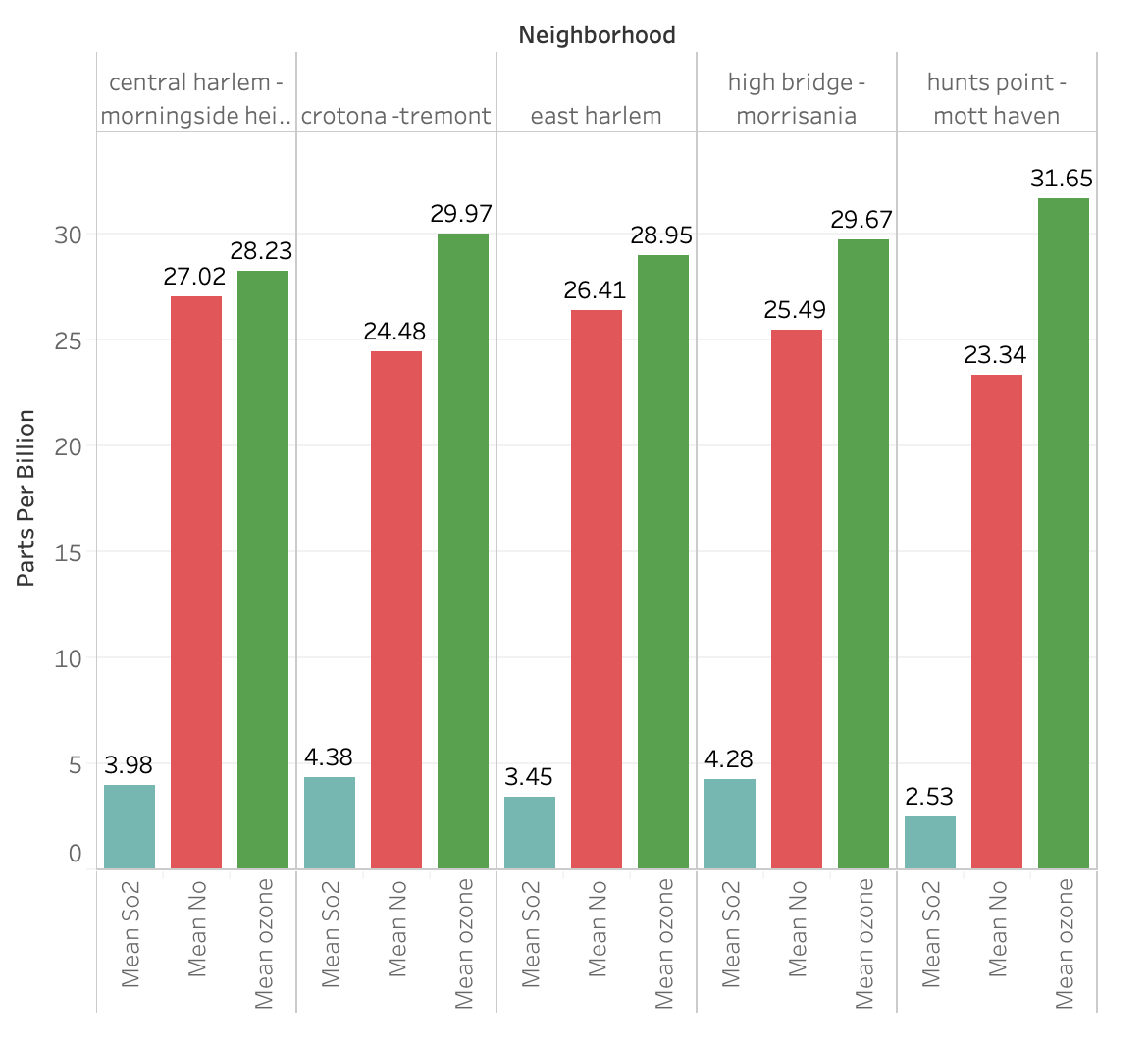
We can see that sulfur dioxide levels have been decreasing since 2010, where they peak, with a slight increase in 2013.

**Locations with Highest Asthma Rates**

The five neighborhoods with the highest asthma rates are Hunts Point - Mott Haven, East Harlem, Crotona - Tremont, High Bridge - Morrisania, and Central Park - Morningside Heights.

****

Pollutant Levels in Neighborhoods with High Asthma Rates

****

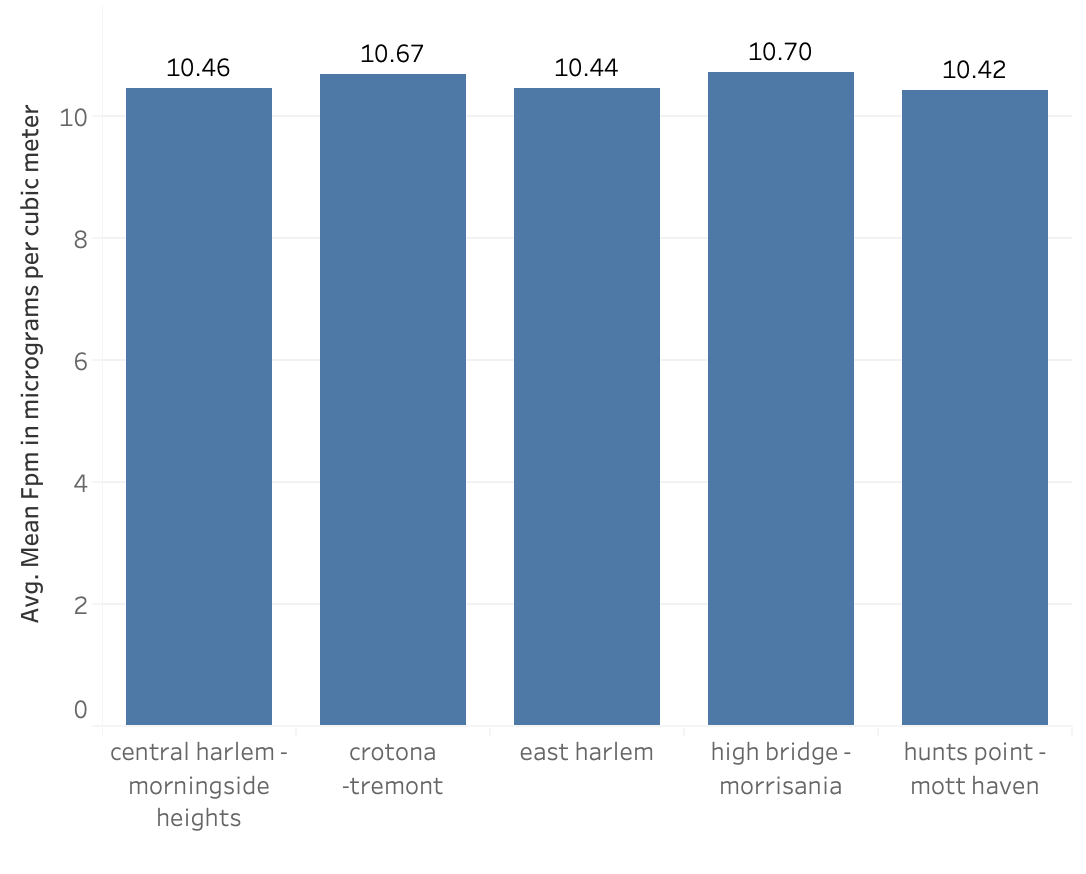
**How do these levels compare with the average yearly pollution rates for all neighborhoods?**

The table shows the average of the mean yearly pollutant levels for all neighborhoods. Measurements are in parts per billion.

| Mean Sulfur | 2.51 |
| --- | --- |
| Mean Nitrogen Dioxide | 23.08 |
| Mean Ozone | 30.26 |

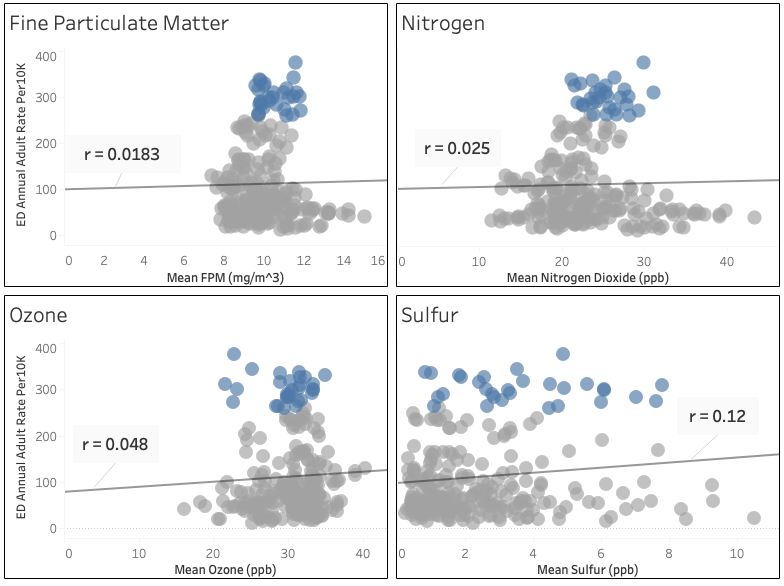
Each of the five neighborhoods has higher-than-average levels of Sulfur and Nitrogen Dioxide. Four of the neighborhoods have a lower than average level of ozone.

**Fine Particulate Matter Levels in the Five Neighborhoods with Highest Asthma Rates**

****

The average of the mean yearly fine particulate matter for all neighborhoods is 9.97 micrograms per cubic meter. The five neighborhoods included here all have a greater average of mean yearly fine particulate matter levels.

**How do the pollution levels in areas with high asthma rates compare to the areas with the highest pollution?**

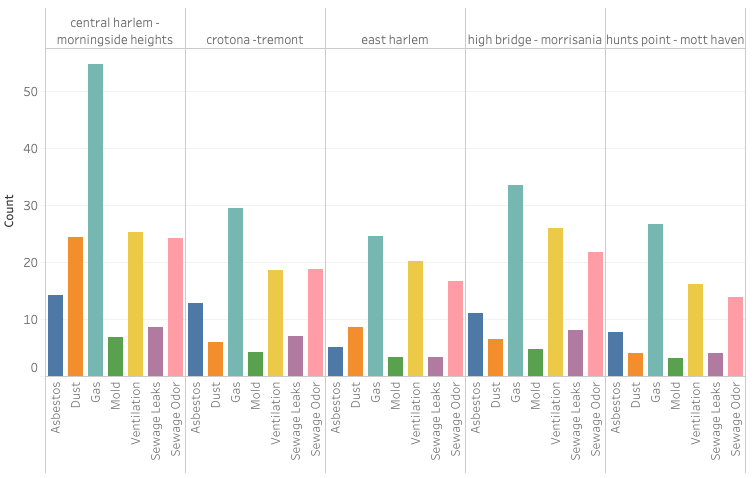


In the scatter plots above, the blue points represent the five locations with the highest asthma rates in New York City: Hunts Point - Mott Haven, East Harlem, Crotona - Tremont, High Bridge - Morrisania, and Central Park - Morningside Heights. We can see that there is a very low correlation between each of the pollutants, and emergency department visits. In the scatter plot for the pollutant Nitrogen Dioxide, the locations with the highest pollutant levels have the lowest asthma rates. The scatter plot for Sulfur shows a wide spread of pollution levels for areas with high and low asthma rates. The scatter plot for fine particulate matter shows a cluster of points mostly between 8 and 12 micrograms per cubic meter, corresponding to a wide range of asthma rates. The scatter plot for Ozone shows a similar trend with the majority of points between 25 and 35 parts per billion corresponding to a wide range of asthma rates. It is clear from these scatter plots that there are other factors besides air pollution that contribute to high emergency department visits for asthma.

#### **Indoor Air Quality**

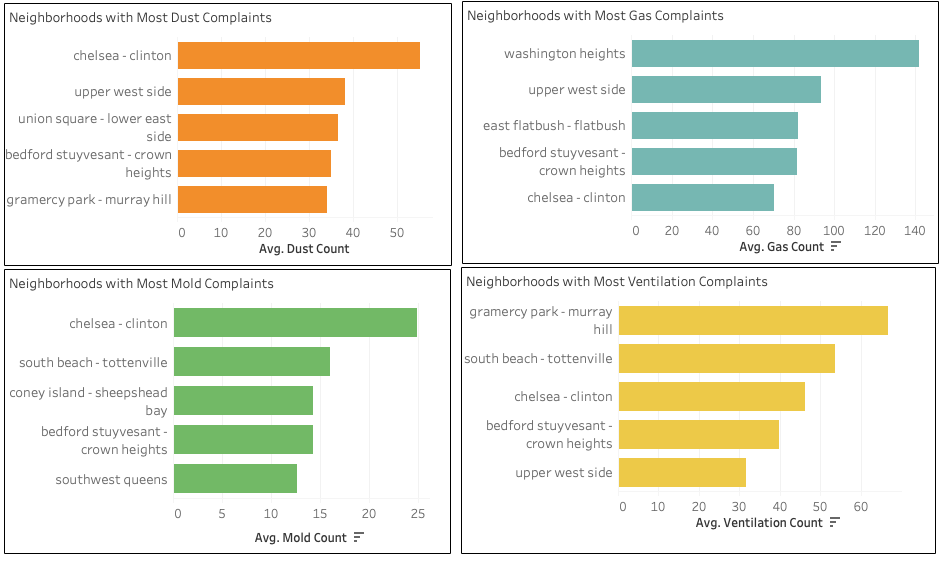
Indoor air quality, which includes ventilation, exposure to gasses, and exposure to dust and other allergens, is a contributing factor to asthma risk. In a study on the impact of ventilation on asthma conducted by the National Center for Biotechnology Information ([2b](https://www.ncbi.nlm.nih.gov/books/NBK224478/)), it was shown that ventilation could have an impact on concentrations of some indoor air pollutants associated with asthma such as dust mites, mold and bacteria. The study defines ‘ventilation’ as the flow of outside air indoors and vice versa. Poor ventilation in areas with high air pollution could cause increased pollution indoors which in turn exacerbates asthma risks. Exposure to gasses indoors can also increase risk of asthma. Fuel-burning appliances in the home such as fireplaces, space heaters, and gas furnaces produce nitrogen dioxide which can trigger asthma ([7a](https://www.webmd.com/children/guide/asthma)).

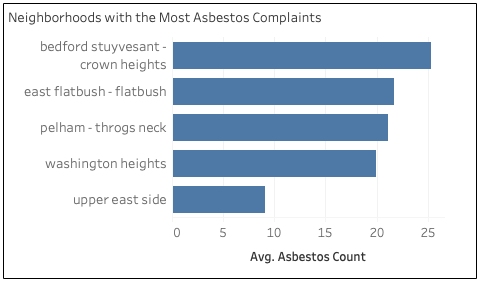
The bar charts below represent the five neighborhoods with the highest emergency department visits for asthma and show the average number of complaints for each of the seven indoor air factors: Asbestos, Dust, Gas, Mold, Ventilation, Sewage Leaks, and Sewage Odor. We can see that gas complaints have the highest count followed by ventilation.

****

In the bar charts below, we can see the top five neighborhoods with the most complaints of each type. The top five neighborhoods with the highest asthma rates are not among the neighborhoods with the highest complaints. We can see that the neighborhood with the highest average amount of gas complaints is Washington Heights with an average of 140 complaint calls per year. This amount is much higher than the average for Central Harlem which has one of the highest asthma rates in New York City. Looking at ventilation complaints, we can see that the highest number of complaints come from Gramercy Park - Murray Hill, with about 65 complaints per year. This is much higher than the complaints in the top asthma regions.

These results demonstrate a low correlation between complaint calls for indoor air quality and emergency department visits for asthma.

****

****

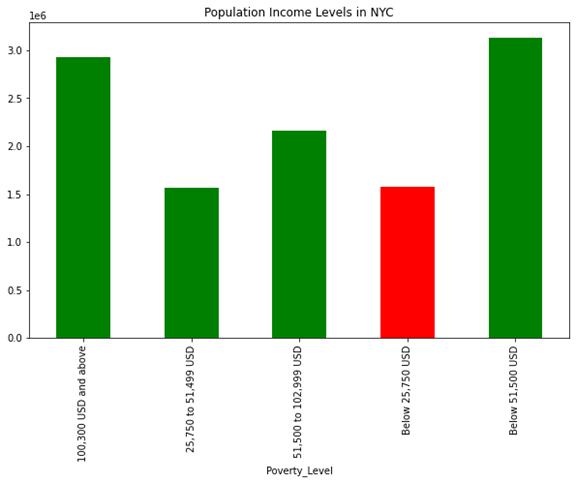
**SDOH IN NEIGHBORHOODS WITH HIGH ASTHMA RATES**

Social determinants of health are any non-medical factors that may influence differences in population health outcomes. Factors we analyzed amongst New York City’s population cover the economic stability, physical environment, health care system, and social context categories of SDOH. The data was categorized by community districts across NYC and by community districts that composed the top 5 UHF neighborhoods (where the ED asthma visit data originates) with the highest rates of asthma admissions in emergency departments, based on Fips code.

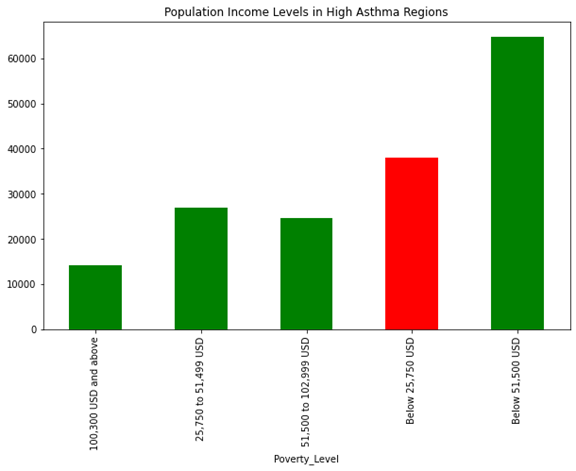
Poverty Levels - Economic Stability

The graph shows the mean number of households in New York City that fall into each category of the Federal Poverty Level (FPL) in the USA.

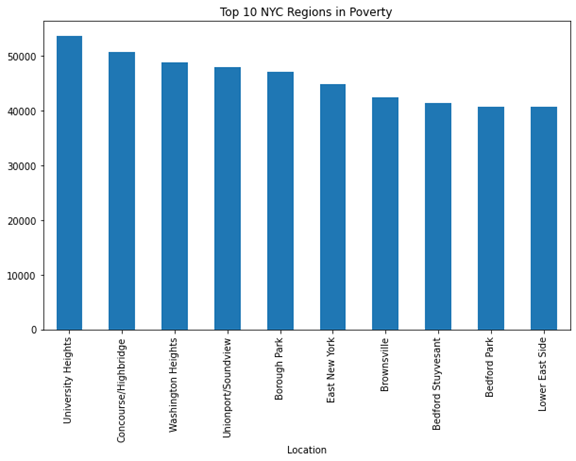
* Below 100% FPL = Below $25,750
* 100 to 199% FPL = $25,750 to $51,499
* Below 200% FPL = Below $51,500
* 200 to 399% FPL = $51,500 to $102,999
* 400% FPL and above = $100,300 and above



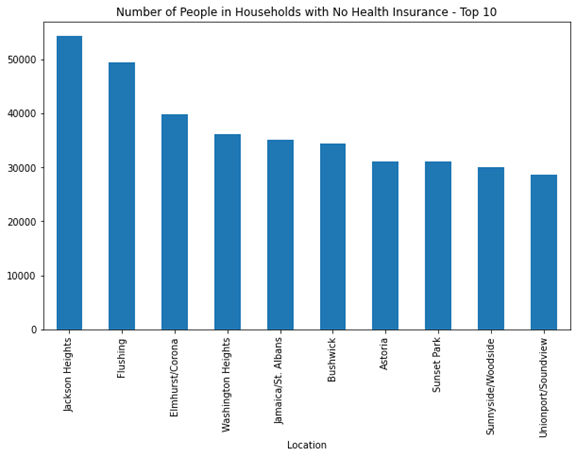
The data shows that the number of people that live with an income 400% FPL and above ($100,300 USD annual income and above) is the predominant group in NYC. The number of people that live 400% or more above the FPL are nearly twice the number of people living in poverty (income of below $25,750 USD annual income), giving some insight into the wealth distribution in NYC. It should be noted that the livable income for one adult (not including households with any dependable members) in NYC is [$43,540 USD](https://livingwage.mit.edu/counties/36061) annual income.

Poverty Levels in Neighborhoods with High Asthma Rates

The wealth distribution in the neighborhoods with the highest asthma rates do not have a similar trend to the city as a whole. People living below the FPL make up the predominant group in the regions of NYC with the highest asthma rates.

****Taking a further look below, two of the top asthma regions (University Heights and Highbridge) are the top two regions in NYC with the highest number of people that have income below the FPL.

Health Insurance - Health Care System

****

The figure above shows the top 10 neighborhoods in NYC with the highest rates of households with no health care insurance. None of the regions with the highest asthma rates in NYC are within the top 10 uninsured regions in NYC. Originally, we had hypothesized different possibilities for the explanation behind how being uninsured for healthcare and rates of emergency department visits may be correlated. On one hand, we expected that areas with the highest rates of asthma, as measured by ED visits for asthma, would have the highest insured populations. This was based on the assumption that being uninsured would act as a barrier to accessing health care for those that have asthma. Alternatively we hypothesized that it is possible that people who go to the ED for asthma are people who are uninsured for a primary care provider and may only seek medical attention in dire circumstances.

While examining the literature, we found a study conducted in 2017 by [Zhou and colleagues](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5754025/) that determined that people who are not insured for healthcare do not use the ED substantially more than those who are insured or those who used Medicaid. On this basis, we suggest that being insured for healthcare is not a SDOH that has a significant contribution to emergency department visits for asthma.

**Race/Ethnicity - Social Context**

|  |  |
| --- | --- |

This pair of graphs demonstrates the population distribution of NYC by the mean population from the year 2007 to 2014 divided into racial/ethnic groups. The largest racial/ethnic group in NYC as a whole is the White population, whereas, the largest racial/ethnic group in high asthma regions are Latinos followed by the Black population.

|  |  |
| --- | --- |

Examining NYC’s population distribution experiencing poverty by ethnic group, it can be seen that Latinos are still the predominant racial group in poverty, followed by the Black population. Poverty is defined as living with an income below the Federal Poverty Line of $25750 USD annual income. The population distribution of people in poverty by race/ethnicity in the top asthma regions do not follow the general trend of NYC as a whole. While Latino and Black racial groups represent the highest populations in poverty in regions with high asthma, showing a similar trend to NYC as a whole, the White and Asian populations in poverty in high asthma regions are significantly lower in proportion to Black and Latino populations in high asthma regions compared to NYC as a whole. Latinos are the predominant group living in poverty followed by the Black population. These findings do not coincide with findings in the literature that shows Black and American Indian/Alaska Native populations having the highest asthma rates in the USA. However, [rates vary significantly among subgroups of latinos](https://www.lung.org/research/trends-in-lung-disease/asthma-trends-brief/current-demographics), with the highest asthma occurrence in any ethnic-subgroup being in 14% of Puerto Ricans across the USA from years 2016-2018. Taking a further look below, the higher rates of asthma ED visits are broken down based on the numbers from the graphs above and consolidated into the last table.

| Race/Ethnicity Graphs in NYC (Occurence of Poverty) | |
| --- | --- |
| Pop. by Ethnicity | Pop. by Ethnicity in Poverty |
| Other 2.200543e+05  Asian 1.095875e+06  Black 1.878944e+06  Latino 2.389002e+06  White 2.747519e+06 | Asian 199639.375001  White 325659.625003  Black 404223.500001  Latino 658471.250003 |
| % of Pop. in Poverty by Ethnicity   * Asian: 199639.375001 / 1.095875e+06 = 18.22% * White: 325659.625003 / 2.747519e+06 = 11.85% * Black: 404223.500001 / 1.878944e+06 = 21.51% * Latino: 658471.250003 / 2.389002e+06 = 27.56% | |

| Race/Ethnicity Graphs in Top Asthma Regions (Occurence of Poverty w Possible Correlation to Asthma) | |
| --- | --- |
| Pop. by Ethnicity | Pop. by Ethnicity in Poverty |
| Other 1777.079545  Asian 2440.647727  White 5811.852273  Black 36309.102273  Latino 59606.102273 | Asian 731.403509  White 1047.828125  Black 12000.640625  Latino 23995.578125 |
| % of Pop. in Poverty by Ethnicity   * Asian: 731.403509 / 2440.647727 = 29.97% * White: 1047.828125 / 5811.852273 = 18.03% * Black: 12000.640625 / 36309.102273 = 33.05% * Latino: 23995.578125 / 59606.102273 = 40.26% | |

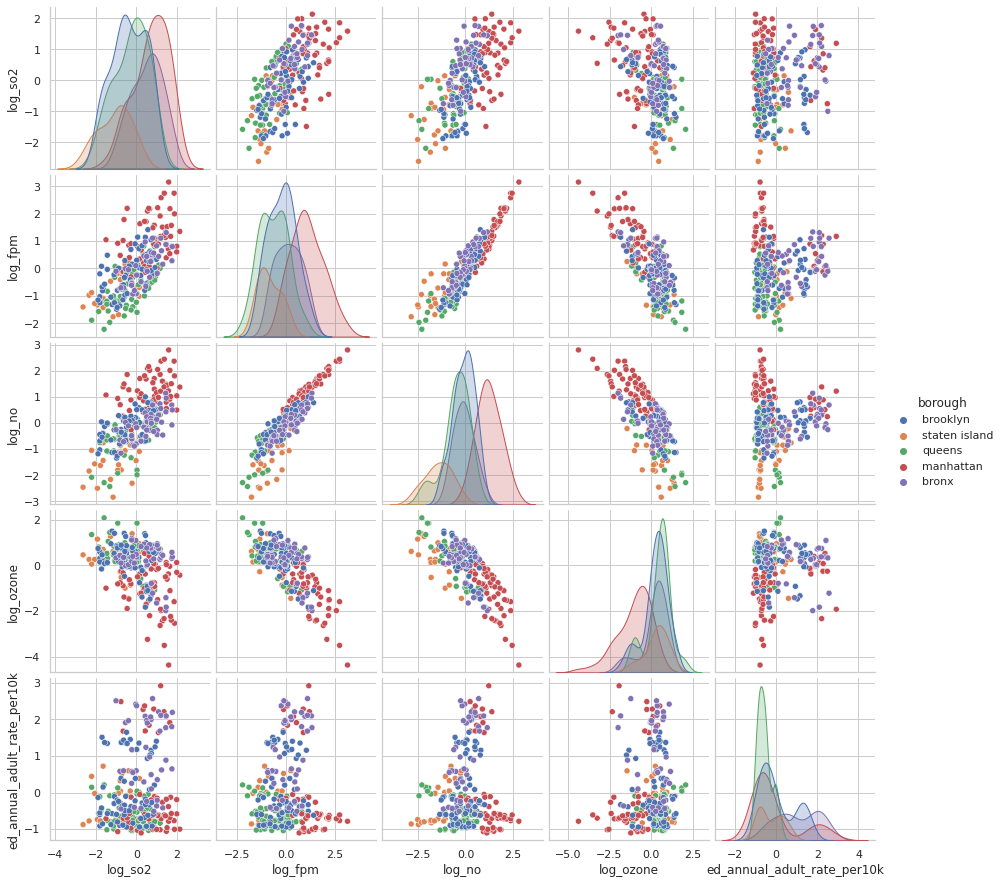
| Intra-Race Poverty by Region | | |
| --- | --- | --- |
| **Race** | **All of NYC** | **Top Asthma Regions** |
| Asian | 18.22% | 29.97% |
| White | 11.85% | 18.03% |
| Black | 21.51% | 33.05% |
| Latino | 27.56% | 40.26% |

This final table demonstrates the rate of poverty within each race divided geographically. The highest intra-race poverty percentage occurs in Latinos, with 27.56% of all Latinos in NYC living in poverty. In the top asthma regions of NYC, it can be seen that 40.46% of all Latinos and nearly 1 out of every 3 people of a Black racial background in those regions live in poverty. These results lead us to believe that race and income can be correlated with asthma exacerbation.

**STATISTICAL ANALYSIS AND PREDICTIVE MODELING**

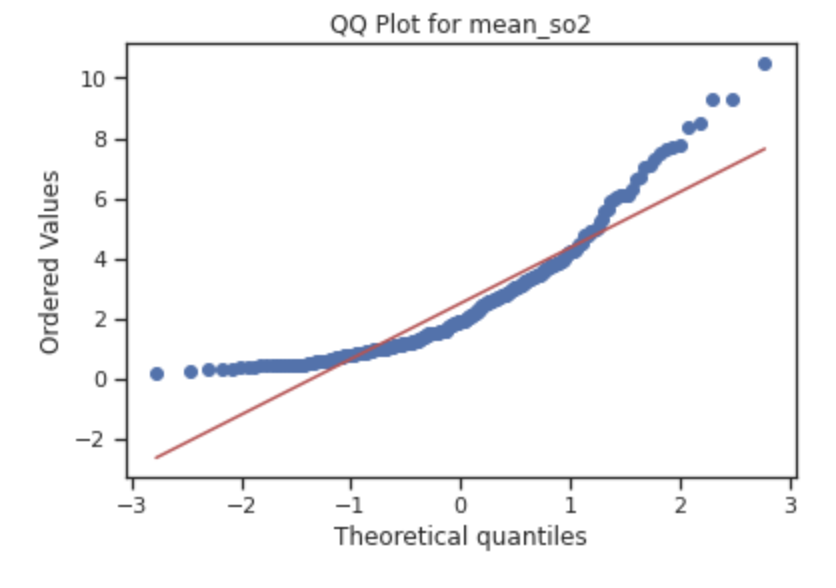
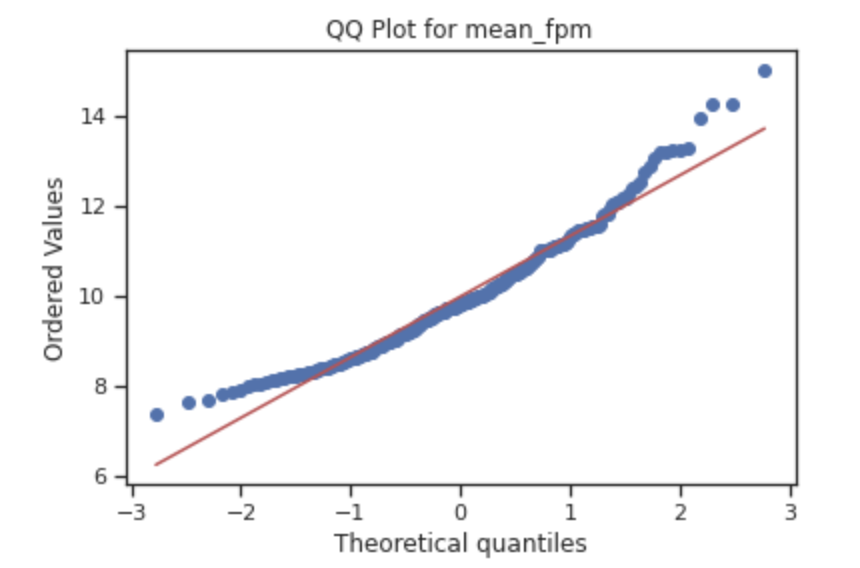
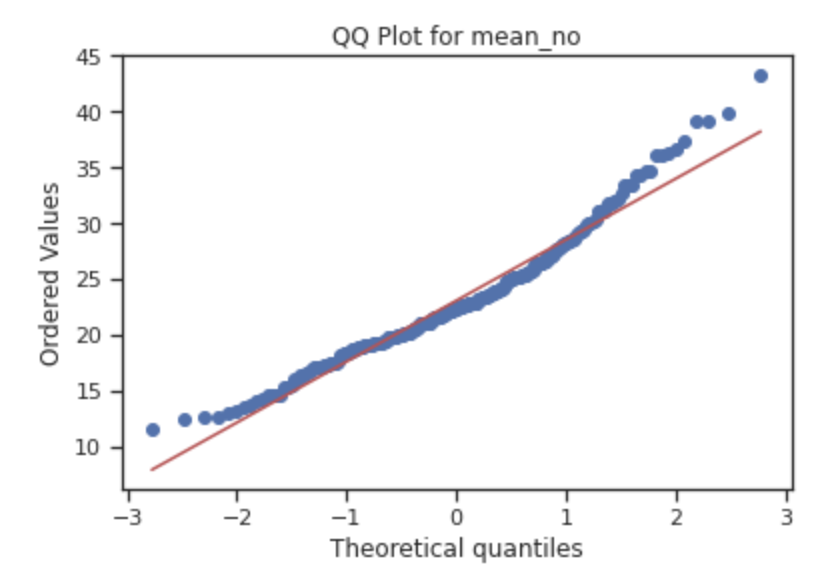
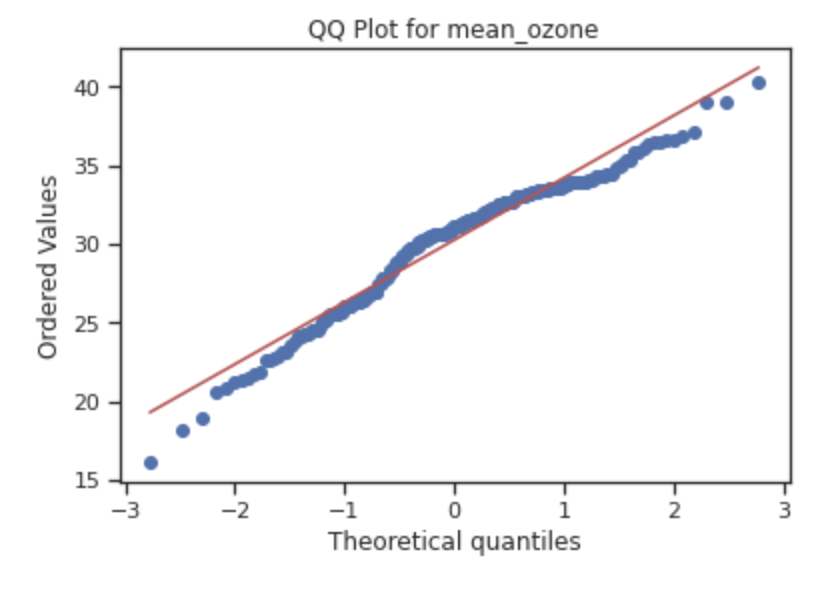
#### **OUTDOOR AIR QUALITY**

In this section, we will outline the requirements for linear regression and the process for creating a predictive model using outdoor air quality factors as predictors. The four outdoor air pollutants used in this regression analysis are fine particulate matter, nitrogen dioxide, ozone, and sulfur. The matrix of scatter plots below shows the relationship between each of the pollutants (after they were log transformed and standardized) and the rate of emergency department visits for asthma (standardized).

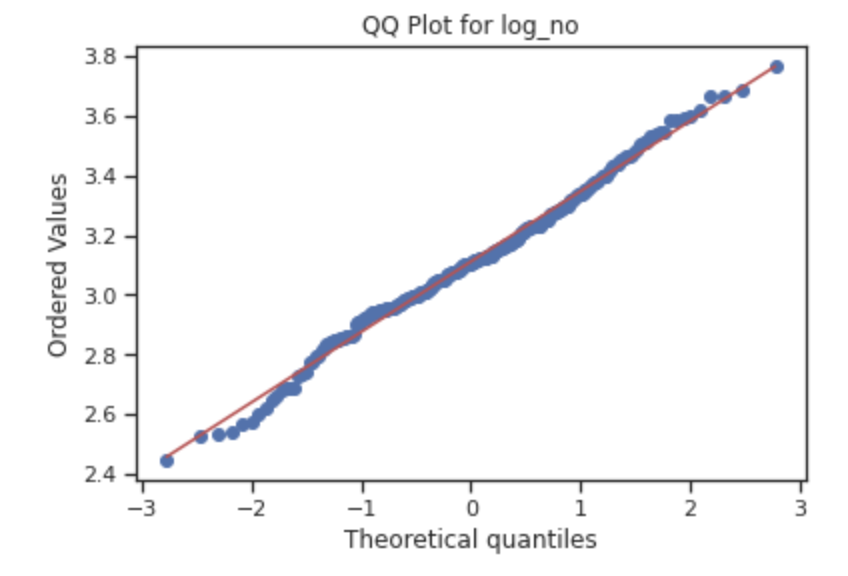
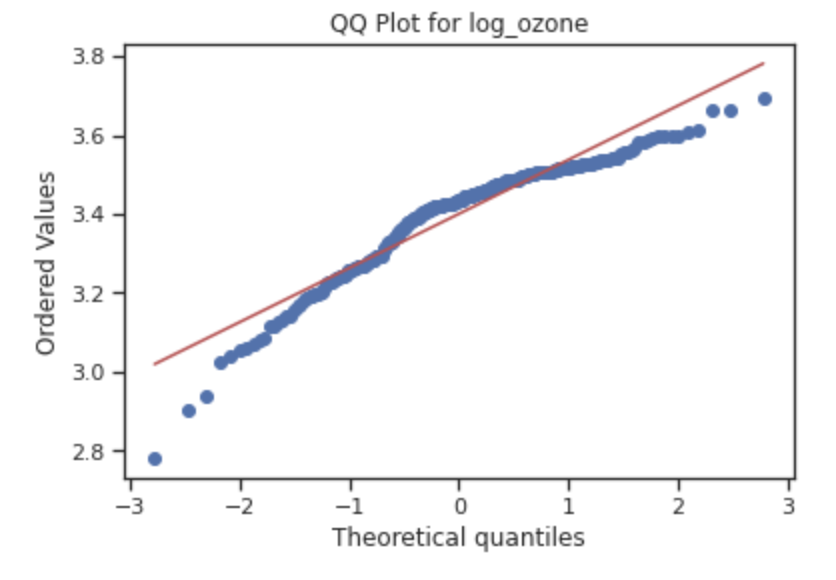


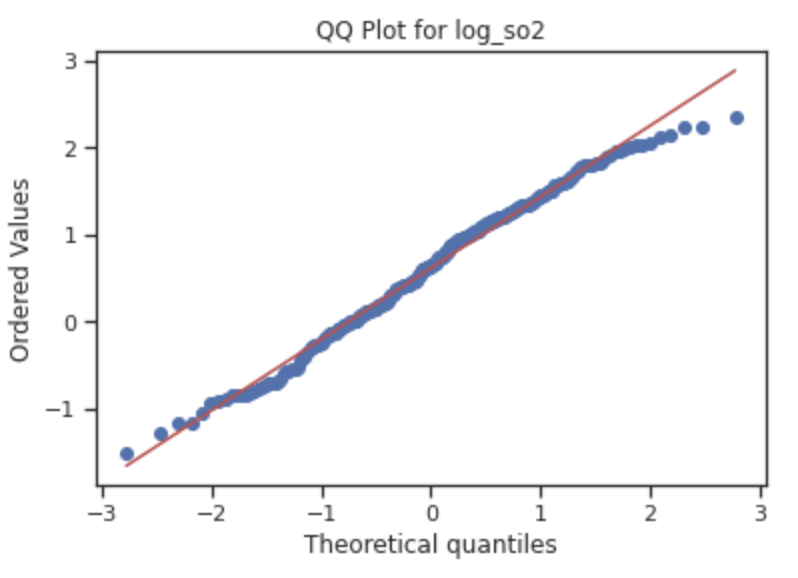
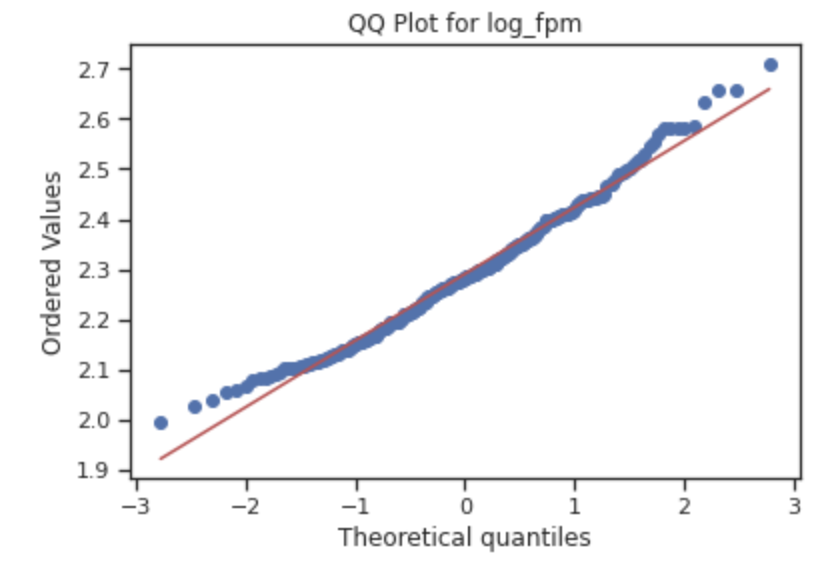
**Testing Requirements for Linear Regression with Outdoor Air Quality:**

1. Linear relationship: Is there an apparent linear relationship between the predictors and the target variable? The scatter plots show that there is not a strong linear relationship between any of the predictor variables and the target variable of emergency department visits. The scatter plots may not tell the whole story therefore it is worth comparing models to determine if there are any predictors that demonstrate a linear relationship with the target variable.
2. Normal distribution of each variable: We used QQ plots to check for normal distribution of each variable.



The QQ plots show that the data for mean\_fpm, mean\_no, and mean\_so2 are right skewed and the data for mean\_ozone are left skewed. The variables were transformed using a log transformation to try to create distributions that more closely resemble a normal distribution. The following QQ plots represent the data after a log transformation. It is clear that the data are closer to a normal distribution after the transformation.





1. The third assumption of linear regression is that there is no autocorrelation. This means that when considering each predictor variable independently, one observation is not dependent on the previous observation. Each observation represents the yearly average of a certain air pollutant in a particular neighborhood and is measured independently of the previous year.
2. No multicollinearity: This means that the predictor variables are not highly correlated with each other. We can see that mean\_fpm and mean\_no are highly correlated with each other, and therefore only one of them should be included in the model.
3. Homoscedasticity: this means that the variance of the observations is the same for all the data. We can see that is not the case with these data sets since the variability of the dependent variable tends to increase as the independent variable increases.

**Regression Process:**

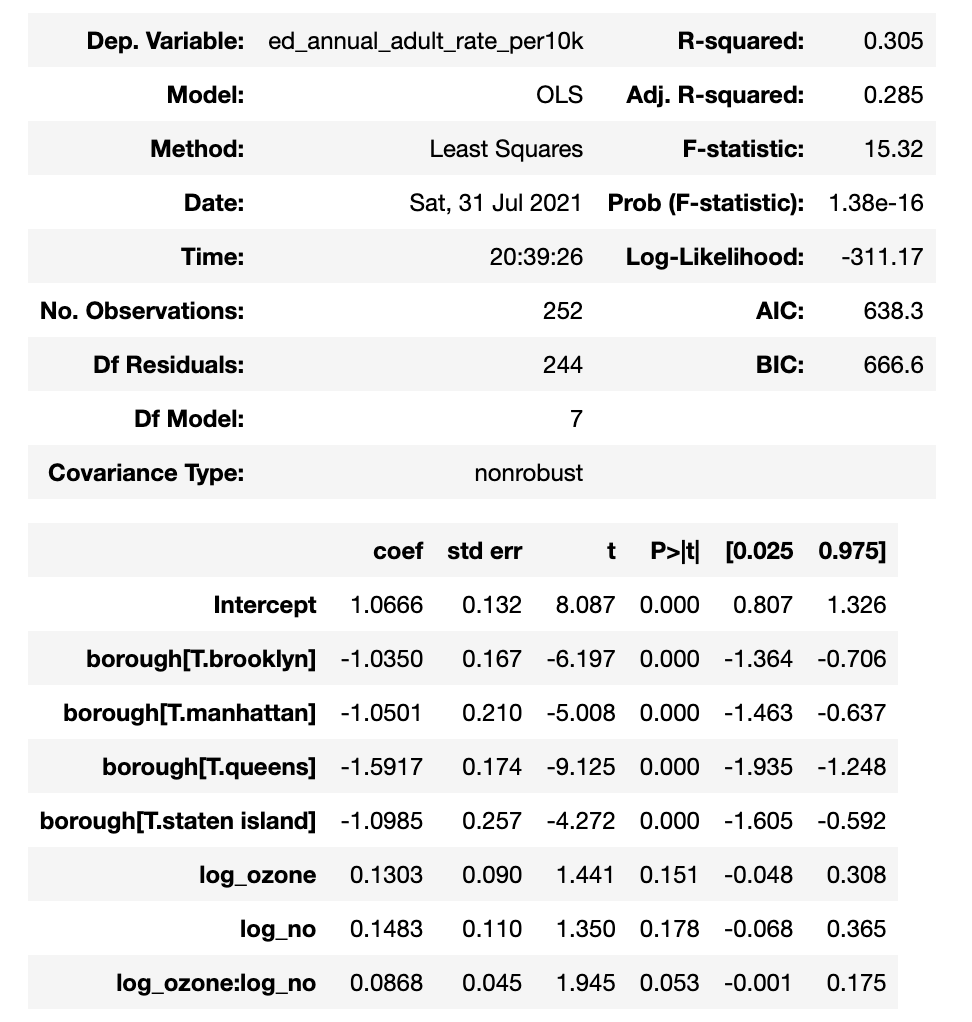
After checking the assumptions of linear regression, the data was transformed using a log transformation and standardized. Several regression models were tested using ordinary least squares regression to determine the best model. The first few models tested related annual emergency department visits for asthma to each of the four air pollutants. These models showed r-squared values close to 0.0 and high AIC values over 700.

The next step was to include the categorical variable ‘borough’ into the model. Introducing this categorical variable raised the r-square value significantly and decreased the AIC value. Various models were tested using different combinations of variables. The table summarizes the results.

| Predictor Variables in Model | r-squared | AIC | Significant Variables |
| --- | --- | --- | --- |
| log\_no | 0.005 | 716.8 | none |
| log\_ozone | 0.003 | 717.3 | none |
| log\_so2 | 0.022 | 712.6 | log\_so2 |
| log\_fpm | 0.002 | 717.7 | none |
| borough | 0.278 | 642.2 | Brooklyn Manhattan  Queens Staten island |
| borough, log\_ozone, log\_no, log\_ozone x log\_no | 0.305 | 638.3 | Brooklyn  Manhattan  Queens  Staten island |

The model that had the highest r-squared and the lowest AIC is the model that includes the categorical variable ‘borough’, log-transformed mean\_ozone, log-transformed mean\_no, and the interaction of log\_ozone and log\_no. This model indicates that asthma rates have some dependence on borough. There might be some underlying causes for that which could relate to social determinants of health such as income levels, number of people below the federal poverty line, and race.

The regression summary for the model is shown below:



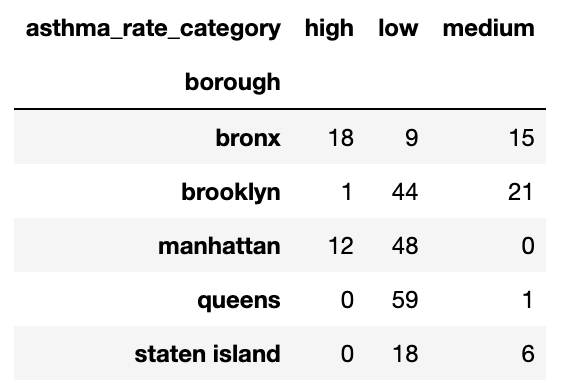
In the summary report, each borough is compared to The Bronx. The coefficient for each of the four boroughs are negative which means that Brooklyn, Manhattan, Queens and Staten Island have lower asthma rates than The Bronx.

The variables log\_ozone, log\_no, and the interaction of log\_ozone and log\_no have p-values greater than 0.05 which means that they are not significant.

The results of the regression analysis indicate that the air pollution factors of mean yearly fine particulate matter, nitrogen dioxide, sulfur and ozone are not significant predictors for asthma rates in New York City. Even though research shows that these air pollutants are closely linked to high asthma rates, our analysis did not reflect this conclusion. One of the possible reasons why our data does not show similar results to other studies of air quality and asthma rates is that the original datasets regarding air quality that are used in our study were aggregated to indicate the yearly mean levels of air pollutants. Our findings could be missing important information about daily air pollution rates that factor into high asthma rates.

**CHI SQUARE ANALYSIS**

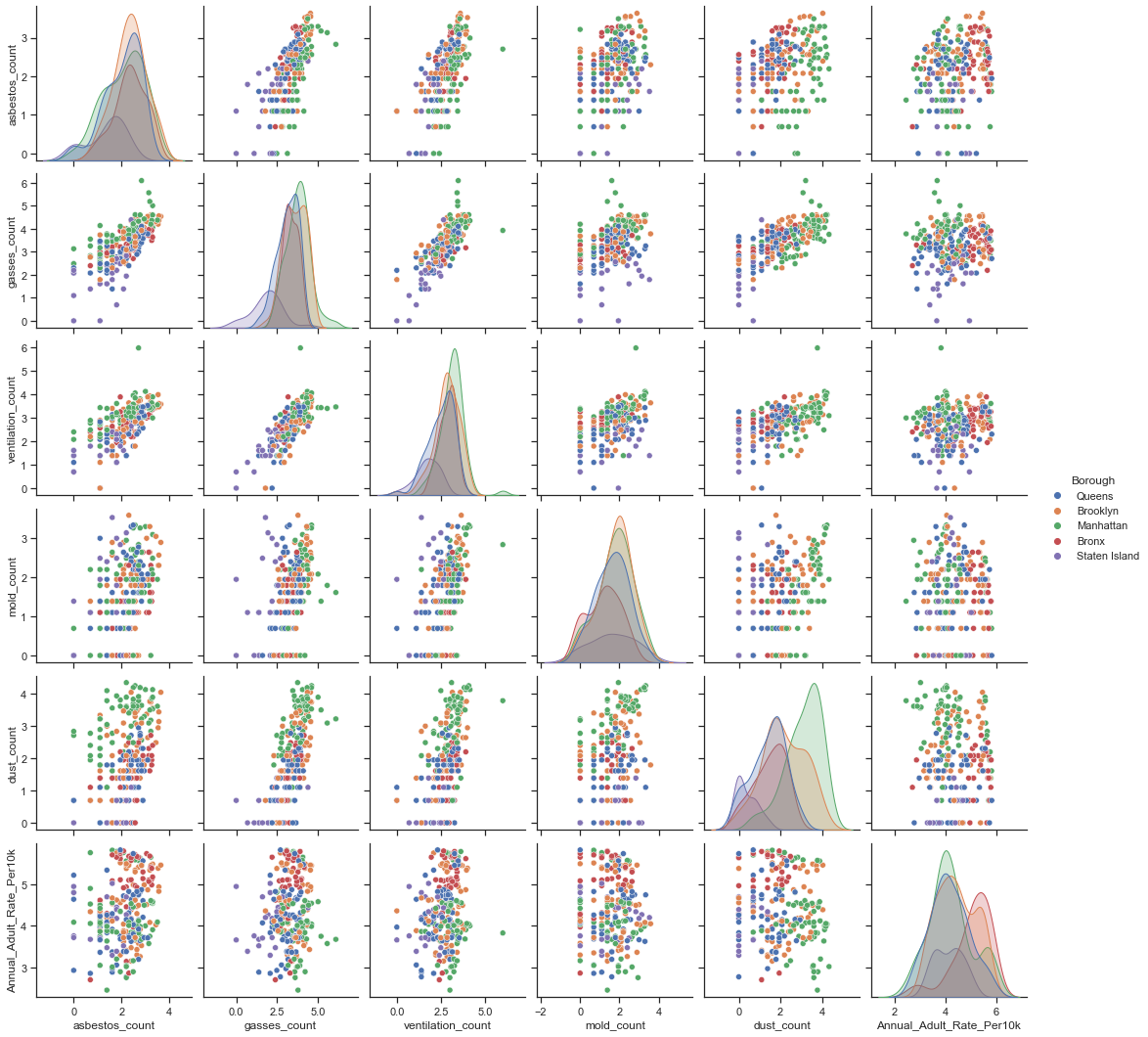
The results of the regression analysis indicate that borough is a significant predictor of asthma rates. For further investigation, a chi squared analysis was performed to test the null hypothesis that asthma rates in each borough are equal. In order to perform the chi squared test, asthma rates needed to be turned into a categorical variable. The asthma rates were transformed to a categorical variable indicating low (rate between 0 and 126), medium (rate from 126 to 252) and high (rate 252 or higher). Then a bivariate frequency table was created to show the number of low medium and high asthma rates in each borough.



From the table, it is clear that The Bronx has the most regions with high asthma rates, and Queens has the most regions with low asthma rates. The Chi Square analysis returned a p-value of . This value is significant, so we can reject the null hypothesis that asthma rates in each borough are the same. This result is consistent with the significance of the boroughs in the regression analysis.

#### **INDOOR AIR QUALITY**

#### The scatter plots below show the relationship between each of the log-transformed and standardized indoor air quality factors and the standardized emergency department visit rate for adults.



In constructing the regression model we initially included all 7 complaint types: mold, dust, ventilation, asbestos, sewage odor, and sewage leaks. The dataset counting their respective counts spanned from 2010 to 2021. The asthma data set spanned from 2002 to 2017 with its most consistent data ending at 2016. Because of this only the rows pertaining to mutual years could be extracted: years 2010, 2011, 2012, 2013, 2014, 2015, and 2016. This significantly reduced the size of our data set. This was particularly true for sewage odor and sewage leak data, which only recently began to be recorded (around 2017). Consequently, the two aforementioned complaint types were not included in the model.

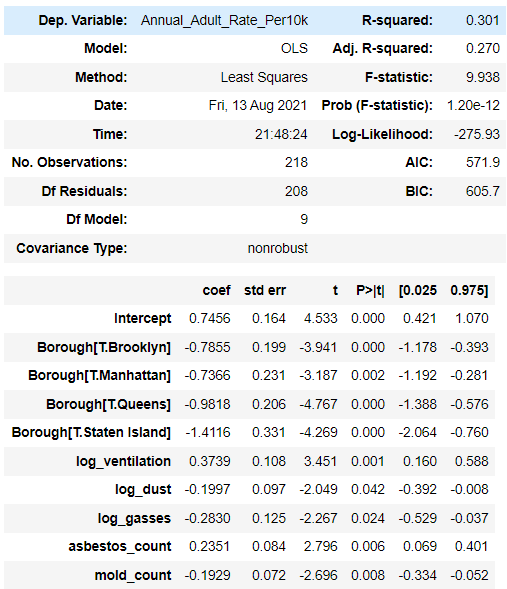
The variable geography- indicating UHF-42 neighborhood was also not included as it caused over fitting. Moreover, this regression model is about predicting the effect of indoor air quality; including geography would be too granular of a variable to speak generally on indoor air quality. Categorical data in the form of ‘borough’ was included as a way to view how the indoor air quality varied among boroughs. Another caveat to mention is that the data set for this model did not include 2015 data as there were methodological changes in asthma data collection in that year.

The remaining data was used in a regression model, the target variable being Adult\_Asthma\_Rate\_Per10K and the predictors being as follows: Mold, Asbestos, Dust, ventilation, gasses. At first, no transformations were performed on this data; the r-squared value of 0.21 was achieved. After performing a log- transformation on all variables the r-squared

was 0.245. Finally a standardization was performed on the data. A mix of log-standardized

and standardized variables were used as predictors for asthma rate. This process achieved the highest r-squared value of 0.301.

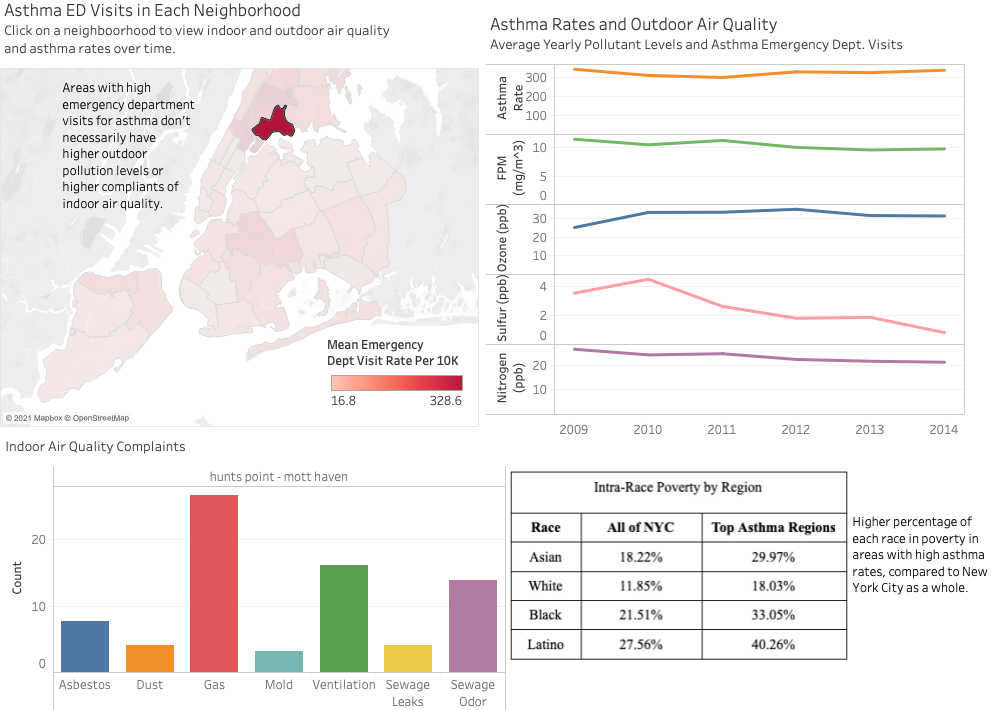
The regression summary for the model is shown below:

****

All the variables used were significant, having a p-value of less than 0.05. The results show the log value of ventilation counts and asbestos counts as being the strongest factors in a positive correlation between asthma rates and complaint types. The results also show that log\_dust, log\_gasses and mold\_count show a negative coefficient meaning that the correlation with asthma rates is negative. The coefficient for each of the four boroughs shown is a comparison to the borough The Bronx. We can see that asthma rates are lower in each of the four boroughs compared to The Bronx since all the coefficients are negative.

#### **DASHBOARD DESCRIPTION**

[**https://tinyurl.com/team45-dashboard**](https://tinyurl.com/team45-dashboard)



The dashboard displays a map of New York City divided by UHF Neighborhoods and colored based on the rate of emergency department visits for asthma per 10,000. A darker red color indicates a higher rate. The dashboard also includes a timeline showing the change in air pollution and change in asthma rates over time. Furthermore, the dashboard includes a bar chart demonstrating the counts of indoor air quality complaints in each neighborhood. The dashboard is interactive and allows the user to click a neighborhood on the map to filter the timeline and barcharts for that region. The functionality of the dashboard allows the user to compare indoor and outdoor air quality for all the UHF neighborhoods in New York City.

#### **CONCLUSION**

Through our investigation, we found evidence that there is disparity in the rates of emergency department visits for asthma in different regions throughout New York City. The highest rate of emergency department visits for asthma occurs in East Harlem with a rate of 328.6 per 10 thousand people. The lowest rate of emergency department visits for asthma occurs in Greenwich Village - SoHo with a rate of 16.8 per 10 thousand people.We found a low correlation between outdoor air pollution and the rate of emergency department visits for asthma. Certain areas with high pollution levels such as Gramercy Park, and Chelsea actually have very low rates of emergency department visits for asthma. We did find that there was a distinction in rates of emergency department visits for asthma in each borough with The Bronx having the highest rates and Queens having the lowest rates according to our regression model.

The reason our data analysis does not align with the results of other similar studies could be because of the limitations of our data. The original datasets for outdoor air quality were aggregated by average yearly pollution levels, so we are missing information about daily pollution levels and the amount of time a region had elevated levels of pollution. Furthermore, the variable we are using to represent asthma prevalence is emergency department visits and this does not encompass adults living with asthma who did not have severe enough cases to go to the hospital.Through our analysis of various Social Determinants of Health, we found that race and income may contribute to the exacerbation of asthma in urban areas such as NYC. In the community districts of NYC where the highest emergency department visits for asthma occurred, the predominant population demographic was Latino people living under the federal poverty line. This population demographic differed from that of NYC in general where the predominant population demographic was White people living 400% and more above the federal poverty line. Limitations to our findings may be due to a number of reasons including the nature of the ED visits for asthma (eg, severe or simple cases), a lack of precision with the overlap of community districts and UHF neighborhoods, and limitations of the data, being only from NYC.

#### **RECOMMENDATIONS**

Further research could be done into the connection between pollution levels and asthma rates by gathering data on daily pollution levels to get a more granular view of the changes in air pollution over time. Suggestions for further research also include looking into housing quality including air ventilation, allergens, and pests such as mice and roaches. Another area of investigation could be to look into areas with high amounts of construction and road density. The asthma data used in this study pertained to adults specifically. Studies referenced to support the correlative effect of indoor air quality on asthma development showed an acute relationship with children. Expanding our study to include the rates of asthma among persons under eighteen would presumably increase the predictive power of the regression models. Moreover, a ratio of complaint counts in each neighborhood to population may yield a greater correlation than simply using complaint counts. Using a ratio would remove the effect of larger neighborhoods having more counts simply due to population. Therefore, neighborhoods with a small population but relatively high complaints would show more prominently in our data. Based on our findings of elevated asthma rates in certain neighborhoods, we suggest policy makers employ asthma care initiatives to ease the burden in these areas. Some options for initiatives to support residents living with asthma are education programs to teach about causes and treatment methods, providing free or low cost asthma medication, and ensuring there are an adequate number of health clinics in highly populated areas

#### 

#### **LINKS AND REFERENCES**

**Dashboard:** [**https://tinyurl.com/team45-dashboard**](https://tinyurl.com/team45-dashboard)

**Github with Python Notebooks:** [**https://github.com/vlventure/Team45\_python\_notebooks**](https://github.com/vlventure/Team45_python_notebooks)

**Research Sources:**

1. Columbia Center for Children’s Environmental Health
   1. <https://www.publichealth.columbia.edu/research/columbia-center-childrens-environmental-health/asthma>
2. National Center for Biotechnology Information
   1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5523938/#R7>
   2. <https://www.ncbi.nlm.nih.gov/books/NBK224478/>
   3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5754025/>
3. NYC Dept of Health
   1. <https://a816-dohbesp.nyc.gov/IndicatorPublic/Closerlook/geographies/index.html>
4. Healthify
   1. <https://www.healthify.us/healthify-insights/addressing-sdoh-needs-care-coordination>
5. Environmental Protection Agency
   1. <https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics>
   2. <https://www.epa.gov/ground-level-ozone-pollution/health-effects-ozone-pollution>
   3. <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>
   4. <https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm>
6. American Lung Association
   1. [https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/sulfur-dioxid](https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/sulfur-dioxide)
   2. <https://www.lung.org/research/trends-in-lung-disease/asthma-trends-brief/current-demographics>
   3. <https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/nitrogen-dioxide>
7. Web MD
   1. <https://www.webmd.com/children/guide/asthma>
8. Living Wage Calculator
   1. <https://livingwage.mit.edu/counties/36061>

**Dataset Sources:**

1. Environmental Health and Data Portal Ozone
   1. <https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=2027,719b87,122,Summarize>
2. Air Quality NYC Open Data
   1. <https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r>
   2. <https://data.cityofnewyork.us/Environment/Air-Quality/fyf4-hrcu>
3. DOHMH Indoor Environmental Complaints NYC Open Data
   1. <https://data.cityofnewyork.us/Health/DOHMH-Indoor-Environmental-Complaints/9jgj-bmct/data>
4. Environmental Health and Data Portal Asthma ED Visits
   1. <https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=2380,4466a0,11,Summarize>
5. Environmental Health and Data Portal Asthma Hospitalizations
   1. <https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=2382,4466a0,11,Summarize>
6. Environmental Health and Data Portal Asthma within Past 12 Months
   1. <https://a816-dohbesp.nyc.gov/IndicatorPublic/VisualizationData.aspx?id=18,4466a0,11,Summarize>
7. Citizens’ Committee for Children of New York
   1. <https://data.cccnewyork.org/data/download#0,4,8,10,13/99>