Determinants of Asthma in New York City

Authors: Tawney Lott, Chris Garvin, Youssef Sultan, Tyler Watts and Danielle Chen (Team 28)

Introduction

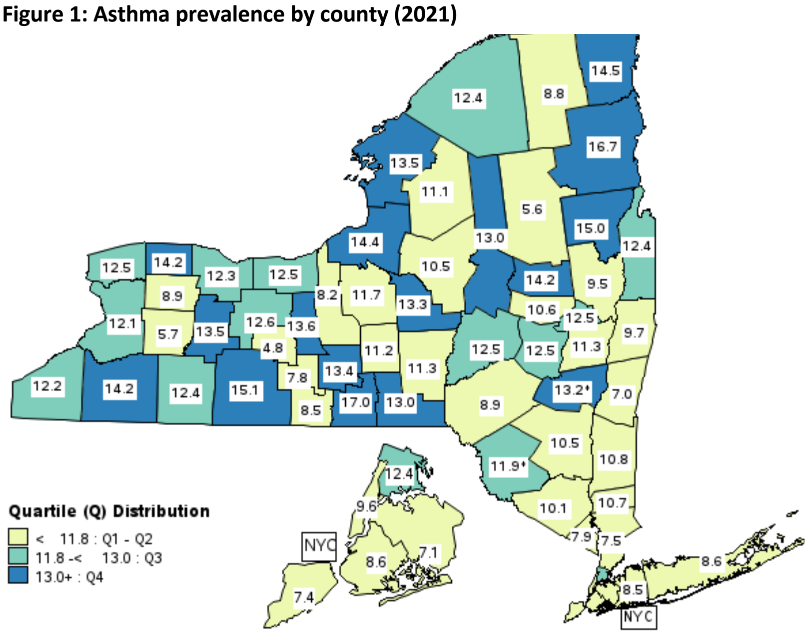
An estimated 14% of New York City's population has asthma. While New York City (NYC) has lower overall asthma prevalence among the adult population relative to the state of New York, the city has higher scores across most indicators for asthma-related hospitalizations and deaths.[1]

Against this backdrop, this analysis seeks to understand the causes of the asthma burden in New York City. We believe there are many behavioral and environmental factors that can be investigated and modified to reduce the asthma burden. This analysis draws on classification modeling to predict city residents with an asthma diagnosis and interpret the most significant predictors of asthma. We combine this with data on financial costs per person with asthma to estimate the potential cost savings from implementing asthma-targeted interventions. In doing so, we aim to provide health policymakers and intervention agencies with insights that can be used to reduce the financial and healthcare burden of asthma in New York City.

Background and existing research

The estimated number of New Yorkers diagnosed with asthma translates to an estimated $1.3 billion annual cost.[2] While Figure 1 to the right shows that NYC has a lower asthma burden relative to many counties across the state, we see higher prevalence in areas like Bronx County relative to Manhattan. This is demonstrated further in the Exploratory Data Analysis section below.

Figure 1: Asthma prevalence by county (2021)



Several risk factors have been linked to asthma. These include a family history of asthma among immediate relatives, as well as lifestyle and behavioral factors such as smoking or exposure to secondhand smoke and being overweight. Environmental factors such as airborne allergens and air quality can also precipitate asthma.[3]

---

[1] Department of Health. 2022. Available: https://webbi1.health.ny.gov/SASStoredProcess/guest?_program=%2FEBI%2FPHIG%2Fapps%2Fasthma_dashboard%2Fad_dashboard&p=ch&cos=60.

[2] Ibid.

[3] MayoClinic. 2023. Asthma. Available: https://www.mayoclinic.org/diseases-conditions/asthma/symptoms-causes/syc-20369653.

Existing literature has also shown that there are many direct and indirect costs associated with asthma. Direct costs include the cost of doctors, social support, drugs, hospital treatment, disposable equipment. Indirect costs include the loss of productivity or work by patient and loss of productivity by patient's family and friends. There also exist intangible costs such as grief, fear, pain, and unhappiness.[4]

Research by McKinsey & Company has found that, for every $1 invested into improving health, a potential return of $2 to $4 of economic returns can be realized.[5] However, it requires substantial changes in healthcare delivery, and functioning community infrastructure, particularly in high poverty neighborhoods. Drawing on this research, we will use a mix of EDA and modeling to narrow in on which factors to target to formulate a targeted and impactful act that will not only reduce cases of asthma but also improve the overall health of NYC communities. Our estimated dollar impact will take into consideration the reduction of workforce and school days lost, healthcare and mortality due to asthma. This approach is described further in the sections that follow.

Data

Community Health Survey

The analysis draws on several data sources to provide a holistic view of asthma in NYC, the first of which is the NYC Community Health Survey (CHS) data.[6] CHS is an annual telephonic survey conducted by the Bureau of Epidemiology Services at the Department of Health and Mental Hygiene. It is a cross-sectional survey with a random sample of approximately 10,000 NYC adults aged 18 and older and includes residents from all five of NYC's boroughs (Manhattan, Brooklyn, Queens, Bronx, and Staten Island).

As part of the analysis, we combined the data from the 2018 and 2019 implementations of the survey, which resulted in approximately 17,500 observations. While the survey itself collects data on more than 100 attributes, we narrowed these down to those which appeared most appropriate to the context of this challenge.[7] One challenge in narrowing down these attributes is that not all data is collected in each year of the survey. For instance, while the 2018 iteration included data on dwelling (indoor allergens) and exposure to secondhand smoke, these questions were not asked in the 2019 iteration of the survey. Ultimately, we narrowed down on a list of 26 attributes, which are shown in the Appendix. Our target variable for this part of the analysis is whether the respondent has ever been diagnosed with asthma.

Several data cleaning steps were necessary to prepare the data for analysis, the first of which was relabeling the values for each column and observation. That is, while the labels were encoded with a numerical value, these needed to be correctly labeled with their corresponding categorical value prior to analysis.[8] This is because the numerical labels would have implied an ordinal relationship between the values, which in fact does not exist.

We also needed to impute missing values; in this case, we used a different approach depending on the extent of missing values in the column and the data type of the column. While we tested imputing missing

---

[4] Barnes PJ, Jonsson B, Klim JB. The costs of asthma. Eur Respir J. 1996 Apr;9(4):636-42. doi: 10.1183/09031936.96.09040636. PMID: 8726924.

[5] McKinsey & Company. 2020. Prioritizing Health: A prescription for Prosperity. Available: https://www.mckinsey.com/industries/healthcare/our-insights/prioritizing-health-a-prescription-for-prosperity.

[6] NYC Community Health Survey. Available: https://www.nyc.gov/site/doh/data/data-sets/community-health-survey.page.

[7] For instance, while the survey collects data on sexual behaviors, these were excluded from this analysis.

[8] For example, values of 1,2,3…n in marital status were relabeled as 'married', 'divorce', etc.

BMI values using linear regression, we ultimately abandoned this technique as we were not able to identify a high enough quality model. In the end, we found which attributes were most correlated with BMI (US born, exercise and alcohol consumption) and then replaced the missing value with group means based on those attributes. We took a similar approach for fruit and vegetable consumption, replacing null values with the full column means.

We also investigated imputing values for our binary features using random forest modeling, although we only achieved 75% accuracy for variables such as birth sex and exercise. Ultimately, for both binary and categorical features, we dropped the remaining observations with null values as we were reluctant to introduce too much noise to the dataset. The resulting dataset after the above cleaning techniques was then used for exploratory data analysis and predicting asthma diagnosis.

New York State Asthma Control Program

The second dataset in this analysis comes from the New York State Asthma Control Program dashboard, which tracks asthma data at the state, county and zip code levels and includes 40 asthma-related indicators.[9] We focused on major socio-demographic factors, asthma prevalence, emergency department visits, hospital discharges, mortality, asthma health care utilizations and costs to examine annual correlations and trends which can be used after filtering and aggregating. Additionally, some of this data is available at the borough level (although aggregated in 3-year periods instead of our preferred yearly view) which we have joined to our air quality data to explore additional correlations between air quality and asthma prevalence at lower levels of the population.

NYC Air Quality

The third data set is data on air quality for NYC, [10] which is focused on ozone, nitrogen dioxide and fine particulate matter levels over the last decade down to the county level. Literature and research have found that exposure to these items can adversely impact respiratory and worsen the severity of asthma conditions.[11] This dataset, once filtered and aggregated, was joined to the NYS Asthma Control Program data to draw correlations and model relationships related to hospitalizations, emergency department visits, or other important factors at the city and borough levels.[12]

Exploratory data analysis

Community Health Survey: Individual lifestyle attributes and asthma diagnosis

We started with exploratory data analysis to understand the relationship between our input variables and the output variable. This analysis revealed several important findings that needed to be considered as part of our modeling process, the first of which is the class imbalance of our output variable, being an asthma diagnosis. Given only 14% of our sample has had an asthma diagnosis, we needed to consider strategies to address class imbalance, which is discussed further in the Methods section below.

---

[9] New York State Department of Health. Asthma. Available: https://www.health.ny.gov/statistics/ny_asthma/.

[10] NYC Open Data. Air Quality. Available: https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r/data.

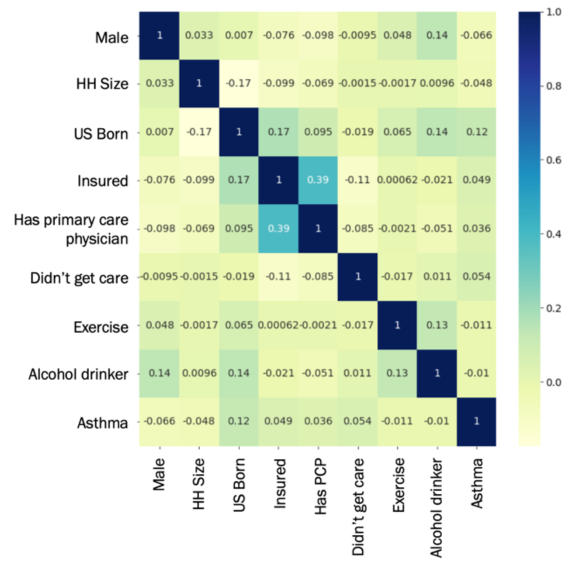
[11] Department of Health. FPM Questions and Answers. Available: https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm.

[12] Ideally, we would be able to join all our datasets together; however, the CHS survey data does not include enough granular information to tag the individuals to their corresponding physical locations.

In terms of relationships between our input variables, we found weak to moderate correlation between these variables, which provided an early signal that multicollinearity would likely not be a major concern for modeling. The correlation matrix between selected variables is shown in the figure to the right.

Findings from the EDA also support insights from the existing literature. As noted in the Background section above, research has shown a relationship between asthma diagnosis and lifestyle factors such as being overweight and smoking cigarettes. We have also seen these bear out in the analysis; Figure 9 in the Appendix shows higher reported asthma rates among current and former smokers as well as those who consume more than one soda a day on average.

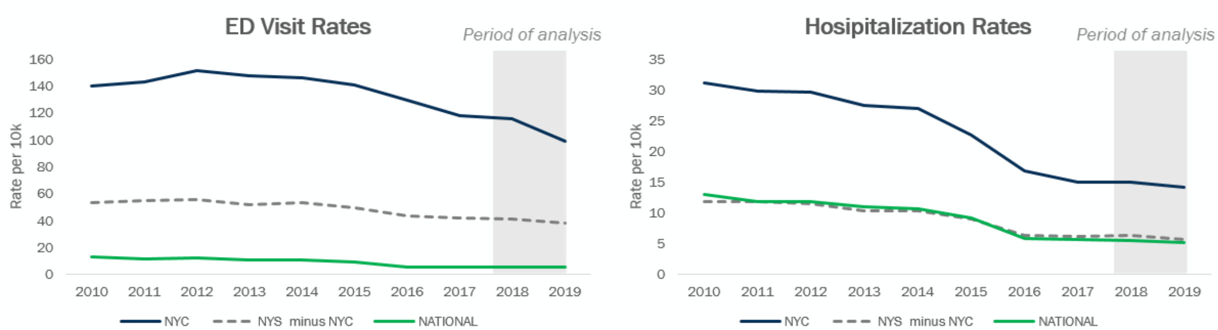
Figure 2: Correlation matrix



Department of Health Asthma Control Program: Locational Disparities in Asthma Severity

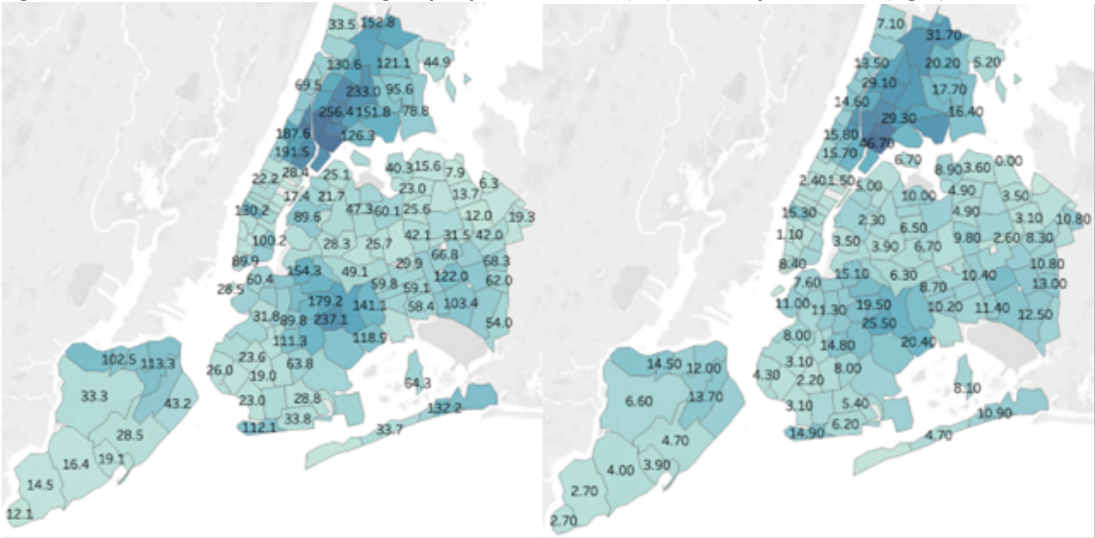
From our exploratory data analysis, we see that NYC has consistently over two times the rate of both asthma-related emergency department and hospitalization rates compared to both the NY State (when omitting NYC) and the national rates. This highlights the importance of tackling the city's asthma burden.

Figure 3: Asthma related emergency department visits and hospitalizations over time



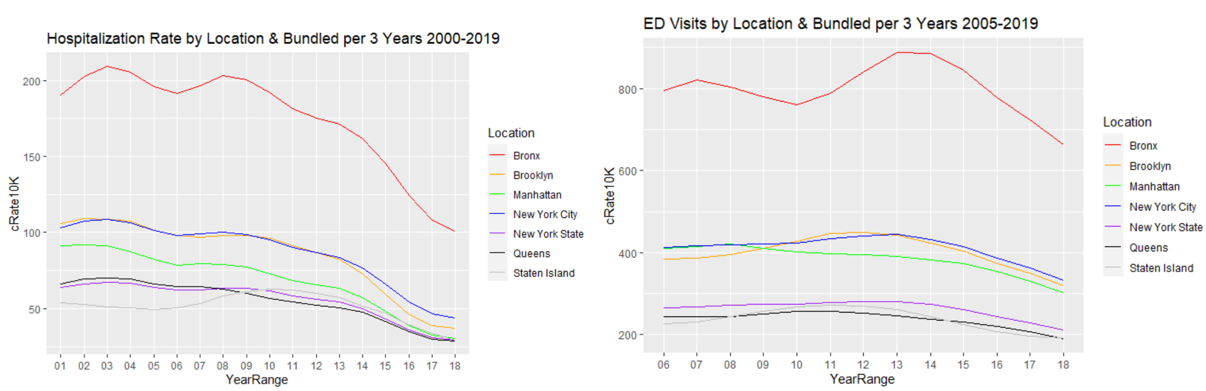
However, it is also important to understand these trends at the borough level. Across NYC's five boroughs, there are locational disparities in the asthma burden and severity, as shown in the figures below. The darker shading in the upper left-hand corner of each figure indicates that asthma-related emergency department visits and hospitalizations are highest in areas of Bronx County, particularly in the Mott Haven area. Although to a lesser extent, this is also the case in Brownsville, which is in Kings County (Brooklyn), shown in the center of each figure.

Figure 4: Rate of asthma-related emergency department visits (left) and hospitalizations (right) 2018-2020



While these figures provide a recent snapshot of asthma-related emergency department visits and hospitalizations by zip code, it is also useful to view these trends over time for each of the boroughs. The figures below show these averages overtime for each borough, as well as the city and state average for comparison. The Bronx has consistently had the highest rate of hospitalizations as well as emergency department visits per 10,000 people from 2000-2019, which is unsurprising given the findings above. While there has been a downward trend when it comes to both emergency department visits and hospitalizations in all boroughs, NYC, as well as New York state, hospitalizations in Bronx (and up until recently) Brooklyn, have been higher than the average for both the city and the state.

Figure 5: Borough Timespan Breakdown- Asthma related emergency department visits and hospitalizations



These are also the lowest income areas in NYC, as indicated by the lighter shades in the corresponding geographic areas shown in Figure 10 in the Appendix. Mott Haven has the lowest median income at $21,800, which is nearly 75% lower than the city average median of $83,000. Similarly, Brownsville had a median income of $31,700, which is 63% lower than the city average median. These early findings suggest that lower income areas of the Bronx and Brooklyn would provide good geographic targets for asthma-related interventions.

Methods

Predicting asthma diagnoses

Given our focus on understanding the most significant predictors of an asthma diagnosis, our primary model is logistic regression. Logistic regression works by wrapping the linear function in the logit link function, thereby scaling the linear output to be a probability between 0 and 1. However, due to imbalance in our target class, we explored a couple of approaches to improve overall prediction, the first of which is random oversampling of the positive class and adjusting the decision threshold.[13]

While logistic regression is our preferred model due to its interpretability, we also use random forest to establish a baseline on what a “good” score looks like for this dataset. Random forest is an ensemble technique that builds off the concept of decision trees. Random forests use bagging (random sampling with replacement) and feature randomness (each tree includes a random subset of features) to create individual trees. These trees are combined to form a “forest” of uncorrelated trees whose prediction as a unit will be more accurate than that of any of the individual trees.

For each of the above techniques, we looped through hyperparameter values to identify the optimal values for each model and dataset. To ensure robustness of our approach, we also use Monte Carlo Cross Validation (MCCV) with 50-100 iterations. MCCV randomly splits the training data into two groups, using a different percent split in each iteration. The model is then fitted on the train data and the test error is found using the fitted model on the test set. After all iterations, we then find the average test error.[14]

Given our focus on identifying those with asthma to understand significant predictors, our primary scoring metric is recall (sensitivity), which is the ability of the model to correctly identify the positive observations. As a result, we are willing to trade off some of the predictiveness of the negative class to identify as much of our positive class as possible.

Understanding the relationship between asthma and air quality

To investigate the relationship between air quality measures of FPM and NO2 levels and asthma-related hospitalizations, we used linear regression. In this case, our target variable was asthma-related hospitalizations which was taken from the New York State Asthma Control Program data. Our input variables were the FPM and NO2 levels, taken from the Air Quality data.

---

[13] Due to the relatively small sample size, we have opted for oversampling the positive class, as opposed to undersampling the negative class, to avoid removing some of the variance in this class.

[14] Rebecca Patro. 2021. Cross-Validation: K Fold vs Monte Carlo. Available: https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b.

Estimating Dollar Impact

To estimate the business impact of the project we plan to leverage prior legislative acts and their economic impacts as a proxy. It is also worth noting that the benefits of heathier living and mortality reduction will only continue to grow over time as programs take their full effect. An example listed below is the Clean Act of 1990[15] which is an emissions control program that aims to reduce air pollution. In the example to the right, we calculate the value from the mitigated costs of loss of labor, medical costs, to mortality. The average costs are taken from various research linked below[16] and we multiply the number of prevented cases to the average costs to get total dollar amount mitigated, which in 2020 was $33 Billion. To evaluate what that means to NYC we multiplied that $33 Billion by 2.55%[17] to get an estimated dollar impact of $844 Million saved just from one act.

Figure 6: Estimating Dollar Impact to NYC (Clean Act of 1990 Example)

$$
\begin{aligned}
&\mathit{Business\ Impact\ of\ Project\ (in\ \$\ value)} \\
&\quad = Labor_{workdays} * Cases + Labor_{schooldays} * Cases + Medical_{asthma} \\
&\quad * Cases + Medical_{heart} * Cases + Medical_{bronchitis} * Cases \\
&\quad + Medical_{ERvisits} * Costs + Mortality_{all} * Cases
\end{aligned}
$$

Economic Impact of the Clean Act of 1990 in 2020

Type	Variable	# Prevented (Cases)	Avg Cost	Total $ Mitigated
Labor	Lost Workdays	17,000,000	$ 340[1]	$ 5,777,280,000
Labor	School Loss Days	5,400,000	$ 65[2]	$ 352,836,000
Medical	Asthma Exacerbation	2,400,000	$ 9,223[3]	$ 22,135,200,000
Medical	Heart Disease - Acute Myocardial Infarction	200,000	$ 11,664[4]	$ 2,332,800,000
Medical	Chronic Bronchitis	75,000	$ 1,876[5]	$ 140,700,000
Medical	Emergency Room Visits	120,000	$ 1,082[6]	$ 129,840,000
Mortality	Mortality - ozone	7100	$ 9,351.00[7]	$ 66,392,100
Mortality	Adult Mortality - particles	230,000	$ 9,351.00[7]	$ 2,150,730,000
Mortality	Infant Mortality - particles	280	$ 9,351.00[7]	$ 2,618,280
			Total:	**$ 33,088,396,380**
			Total (NYC Adjusted ~2.55% total US Pop):	**$ 843,754,108**

Results

Predicting asthma diagnoses

In predicting NYC residents with asthma, we tested random forest and logistic regression both with and without oversampling of our positive class.[18] The results of each test regarding their recall scores are captured in Table 1 on the following page.

---

[15] Clean Air Act Case Study https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-second-prospective-study.

[16] US Department of Labor. 2022. News Release: employer costs for employee compensation - December 2022. Available: https://www.bls.gov/news.release/pdf/ecec.pdf. Box Pure Air. 2022. Who Pays the Price when Students Miss School? Available: https://www.boxpureair.com/blog/who-pays-the-price-when-students-miss-school. Ivanova et al. 2012. Effect of asthma exacerbations on health care costs among asthmatic patients with moderate and severe persistent asthma. Available: https://www.jacionline.org/article/S0091-6749(12)00127-3/fulltext#:~:text=Patients%20with%20exacerbations%20had%20significantly,0001. Nicholson et al. 2016. Patient-level costs of major cardiovascular conditions: a review of the international literature. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5036826/#:~:text=Across%20the%20114%20studies%20included,%2413%2C501%20for%20percutaneous%20coronary%20intervention. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5036826/#:~:text=Across%20the%20114%20studies%20included,%2413%2C501%20for%20percutaneous%20coronary%20intervention. Miravitlles et al. 2003. Costs of chronic bronchitis and COPD: a 1-year follow-up study. Available: https://pubmed.ncbi.nlm.nih.gov/12628879/#:~:text=Results%3A%20The%20global%20mean%20direct,of%20mild%20COPD%20(%241%2C484). Latifa deGraft-Johnson. 2022. How much does an er visit cost in 2022? With and without insurance. Available: https://www.khealth.com/learn/healthcare/er-visit-cost/. Empathy. 2023. The Cost of Dying. Available: https://www.empathy.com/costofdying.

[17] NYC makes up about 2.55% of the US Population.

[18] While we also investigated whether shifting the decision threshold would improve prediction, we ultimately did not pursue this strategy. While the approach did improve prediction on the positive class, it significantly reduced negative class prediction. As a result, we determined that the costs of this approach outweighed the benefits.

Table 1: Model recall results

Split	Random Forest	Logistic regression	Oversampled RF	Oversampled LR[19]
Train	0.0	0.03	0.62	0.62
Test	0.0	0.03	0.60	0.62

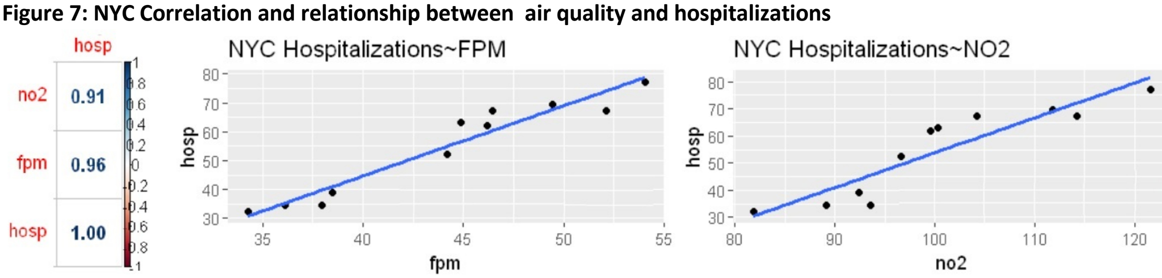
The first takeaway from the above results is that oversampling of the positive class was an effective strategy in addressing class imbalance and improving overall predictiveness of the models. For both random forest and logistic regression, the recall score without oversampling was less than 5% of our positive class. However, after oversampling, we were able to push the average recall score of our logistic regression model up to 0.62, indicating that the model was able to correctly identify 62% of residents with asthma (average specificity of 66%). Furthermore, the delta between the train and test set results indicates only minor overfitting.

Finally, we evaluated the goodness of fit of our logistic regression model for predicting asthma by examining the Pearson and Deviance residuals, as well as the Chi-squared test. The goodness-of-fit tests indicated that the logistic regression model fits the data well (p-values > 0.05). As a result of the above findings, we believe that this is an acceptable model.

Understanding the relationship between asthma and air quality

While we aimed to investigate the relationship between air quality and asthma-related hospitalizations and emergency room visits, we encountered a few challenges in this regard. Although this dataset covers a wide range of time, the smallest period for joining was year, which provided very limited observations. Upon testing the regressions to predict hospitalizations (with FPM, NO2, or both), each appeared to exhibit problematic features. The fine particulate matter regression residuals exhibited non-constant variance, the nitrogen dioxide regression exhibited auto-correlation as shown by the Durbin-Watson test, and the regression using both predictors exhibited multicollinearity as shown by its variance inflation factors. However, we do find a strong correlation between both NO2, FPM, and hospitalization rates, as shown in the figures below. This along with industry research indicates that these factors are connected to asthma hospitalization rates. While this area warrants further research in future, managing air quality factors will likely continue to be important in controlling overall asthma prevalence.

Figure 7: NYC Correlation and relationship between air quality and hospitalizations



Findings and conclusion

Our analysis into the determinants of asthma in NYC is novel for its use of multiple datasets to understand both the micro and macro factors associated with the city's asthma burden. The Results section above demonstrated that we were able to produce a well-performing logistic regression model to identify those

[19] Average score across 100 folds of MCCV.

with asthma, and that there is also a strong correlation between asthma and air quality. As a result, we combine the interpretation of model coefficients with findings from the hospitalization and air quality data sets to provide insights for policy makers.

The table below depicts the probability of having asthma based on some of the strongest and relevant predictors from our logistic regression model.

Table 2: Selected logistic regression coefficients

Coefficient	Estimate	Std. Error	z value	Pr(>|z|)
Poor general health	1.3398045	0.06499447	20.61413	2.05E-94
Hispanic	0.31706476	0.05028198	6.305733	2.87E-10
Didn’t get care	0.31614643	0.03846021	8.220091	2.03E-16
> 1 Soda per day	0.18566141	0.05764591	3.220721	0.0012787
Delay pay rent	0.18502281	0.03778504	4.896722	9.74E-07
Male sex at birth	-0.3517569	0.02733757	-12.8672	6.89E-38
Never smoked	-0.080309	0.04025316	-1.9951	0.0460322

Based on our logistic regression model, we find that individuals with poor general health are most likely to have asthma. Among racial groups, Hispanic residents are most likely to have asthma while males are less likely to have asthma compared to females. There also appears to be a disparity in asthma burden by income group based on two of the coefficients above; respondents who did not get medical care in the past 12 months because they could not afford it as well as those who have delayed paying rent in the past 12 months are more likely to have had an asthma diagnosis. Additionally, the full model output in the Appendix shows that those who are above the food poverty line are less likely to have asthma. This is further evidenced by the fact that the EDA section above demonstrated that residents in neighborhoods in the Bronx and Brooklyn (which also relates to some of the lowest income areas) are more likely to have more severe asthma-related medical issues. Taken together, these findings suggest that lower income groups, particularly women and those of Hispanic backgrounds, should be prioritized by healthcare agencies to address the asthma burden.

With this understanding of at-risk populations, we can create policy recommendations that can help reduce the asthma burden by providing support to these groups, while also delivering interventions to prevent or reduce the likelihood of developing asthma among the city’s population in general. Aggregating together all the analysis conducted, we propose a new targeted act which we have titled the Asthma Prevention and Community Health Improvement Act (APACHI). This act would advocate for adding asthma education and awareness training to existing physical education curricula schools and in places of work. To address the environmental factors that have an impact on asthma and other respiratory illnesses, the act would also include building codes that require air filtration systems for all schools as well as in places of work, starting in lower income at-risk areas in the Bronx and Brooklyn. Furthermore, the act would promote continued regulatory focus on reduction of fine particulate matter and nitrogen dioxide as shown to be strongly correlated with asthma-related hospitalizations in NYC.

We estimate that APACHI can mitigate $174 million per year which will continue to compound over the years, as shown in Figure 8 on the following page. As this act is not in effect, we took the number of cases of asthma related costs in NYS[20] (and adjusted it for NYC which is 48% of the population) and calculated the estimated percentage in reduction of cases. Figure 7 is a high-level calculation of estimated value. The

[20] NY Health Asthma Statistics https://www.health.ny.gov/prevention/prevention_agenda/asthma.htm.

2% reduction leverages the Clean Act of 1990's case[21] looking at their year over year (YoY) increase of cases mitigated from 2010 to 2020 which came to be around 4% YoY. We take a conservative approach of estimating that APACHI can be half as effective as Clean Act, making it a 2% reduction in asthma related cases for NYC.

Figure 8: Estimating the dollar impact of APACHI

Type	Variable (Asthma Related)	# Cases (NYS)	NYC Adjusted	% Reduction from APACHI	Avg Cost	Total $ Mitigated
Labor	Lost Work Days	14,200,000	6,060,766	2%	$ 340	$ 96,514,560
Labor	School Loss Days	14,400,000	6,146,129	2%	$ 65	$ 18,817,920
Medical	Hospitalization	444,000	189,506	2%	$ 5,000	$ 44,400,000
Medical	Emergency Room Visits	1,700,000	725,585	2%	$ 400	$ 13,600,000
Mortality	Mortality	3,600	1,537	2%	$ 9,351	$ 673,272
					Total:	**$ 174,005,752**

$$\begin{aligned} Dollar\ Impact\ of\ APACHI &= \sim\$174Million \\ &= \left(Labor_{work} * Cases\ +\ Labor_{school} * Cases\ + Medical_{hosp} * Cases + Medical_{ER} * Cases \right. \\ &\quad \left. + Mortality_{all} * Cases\right) * 2\%\ Reduction \end{aligned}$$

NYC is a micro example but given that this proposal gets taken into consideration and passed this can serve as a framework of analysis by using analytics to build a data back story to inform policy makers for other metropolitan areas. APACHI will be the first step into that direction where it focuses on early prevention and education on asthma while lifting our lower income communities in NYC.

Potential Further Analysis and Business Impacts

As we see from our analysis, our model prediction quality (measured by recall score) is 0.62 which we believe to be suitable. However, given more time, we would be interested in analyzing interaction terms to see how that would affect our model fit and quality. In a general sense, adding interaction terms to our model would allow us to see how synergies between multiple predictors could affect the response variable. Additionally, with more time we would be able to file the needed paperwork to gain location data on the survey results so we could join all the datasets together to draw stronger conclusions on the relationships between asthma, air quality, and NYC respondents.

Furthermore, while we have recommended a policy intervention given the findings of this analysis, given more time we also would want to investigate the costs and potential drawbacks of implementing such a program. This would allow for a more comprehensive understanding of the financial impact of APACHI.

[21] Clean Air Act Case Study https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-second-prospective-study.

References

American Thoracic Society. 2018. Asthma Costs the U.S. Economy More than $80 Billion Per Year. Available: https://www.thoracic.org/about/newsroom/press-releases/journal/2018/asthma-costs-the-us-economy-more-than-80-billion-per-year.php.

Barnes PJ, Jonsson B, Klim JB. The costs of asthma. Eur Respir J. 1996 Apr;9(4):636-42. doi: 10.1183/09031936.96.09040636. PMID: 8726924.

Box Pure Air. 2022. Who Pays the Price when Students Miss School? Available: https://www.boxpureair.com/blog/who-pays-the-price-when-students-miss-school.

Department of Health. 2022. Asthma Dashboard. Available: https://webbi1.health.ny.gov/SASStoredProcess/guest?_program=%2FEBI%2FPHIG%2Fapps%2Fasthma_dashboard%2Fad_dashboard&p=ch&cos=60.

Department of Health. FPM Questions and Answers. Available: https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm

Empathy. 2023. The Cost of Dying. Available: https://www.empathy.com/costofdying.

Ivanova et al. 2012. Effect of asthma exacerbations on health care costs among asthmatic patients with moderate and severe persistent asthma. Available: https://www.jacionline.org/article/S0091-6749(12)00127-3/fulltext#:~:text=Patients%20with%20exacerbations%20had%20significantly,0001.

Latifa deGraft-Johnson. 2022. How much does an er visit cost in 2022? With and without insurance. Available: https://www.khealth.com/learn/healthcare/er-visit-cost/.

Miravitlles et al. 2003. Costs of chronic bronchitis and COPD: a 1-year follow-up study. Available: https://pubmed.ncbi.nlm.nih.gov/12628879/#:~:text=Results%3A%20The%20global%20mean%20direct,of%20mild%20COPD%20(%241%2C484).

MayoClinic. 2023. Asthma. Available: https://www.mayoclinic.org/diseases-conditions/asthma/symptoms-causes/syc-20369653.

McKinsey & Company. 2020. Prioritizing Health: A Prescription for Prosperity. Available: https://www.mckinsey.com/industries/healthcare/our-insights/prioritizing-health-a-prescription-for-prosperity#section-header-5.

New York Department of Health. 2023. Enrolment by county. Available: https://www.health.ny.gov/health_care/medicaid/enrollment/docs/by_resident_co/2019/2019-07-01.pdf.

Nicholson et al. 2016. Patient-level costs of major cardiovascular conditions: a review of the international literature. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5036826/#:~:text=Across%20the%20114%20studies%20included,%2413%2C501%20for%20percutaneous%20coronary%20intervention.

NYC Community Health Survey. Available: https://www.nyc.gov/site/doh/data/data-sets/community-health-survey.page.

NYC Open Data. Air Quality. Available: https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r/data.

The Prevalence and Cost of Asthma in New York State. 2014. Available: https://www.osc.state.ny.us/files/reports/special-topics/pdf/health-asthma-2014.pdf.

Queensland Government 2023. Available: https://www.qld.gov.au/environment/management/monitoring/air/air-pollution/pollutants/.

Rebecca Patro. 2021. Cross-Validation: K Fold vs Monte Carlo. Available: https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b.

United States Environmental Protection Agency. 2022. NO2. Available: https://www.epa.gov/no2-pollution/basic-information-about-no2#What%20is%20NO2.

United States Environmental Protection Agency. 2022. FPM. Available: https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#effects.

US Census Bureau. American Community Survey. Available: https://data.census.gov/table?q=median+income&g=040XX00US36$8600000.

US Department of Labor. 2022. News Release: EMPLOYER COSTS FOR EMPLOYEE COMPENSATION - DECEMBER 2022. Available: https://www.bls.gov/news.release/pdf/ecec.pdf.

US Environmental Protection Agency. Benefits and Costs of the Clean Air Act 1990-2020, the Second Prospective Study. Available: https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-second-prospective-study.

Appendix

Figure 9: Asthma rates by behavioral attributes

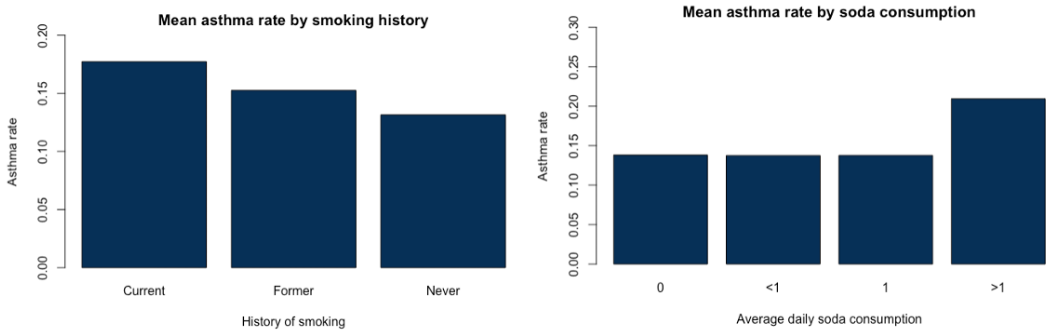


Figure 10: Median income in the past 12 months (in 2021 inflation-adjusted dollars)

Table 3: Logistic regression model

Coefficient	Estimate	Std.Error	z value	Pr(>|z|)
agegroup25-44 years	-0.2404	0.0549	-4.3806	0.000012
agegroup45-64 years	-0.3894	0.0588	-6.6273	0.000000
agegroup65+ years	-0.5812	0.0659	-8.8161	0.000000
avgsodaperday>1	0.1857	0.0576	3.2207	0.001279
avgsodaperday0	0.1147	0.0306	3.7504	0.000177
avgsodaperday1	-0.1156	0.0576	-2.0059	0.044863
birthsex_male	-0.3518	0.0273	-12.8672	0.000000
bmi	0.0260	0.0021	12.5748	0.000000
cyclingNever	-0.1960	0.0474	-4.1316	0.000036
cyclingSeveral times a month	-0.1238	0.0585	-2.1181	0.034166
delaypayrent	0.1850	0.0378	4.8967	0.000001
didntgetcare	0.3161	0.0385	8.2201	0.000000
drinker	0.0596	0.0276	2.1623	0.030597
educationHighschool	-0.1561	0.0378	-4.1256	0.000037
employment_status Not in labor force	0.0966	0.0336	2.8756	0.004032
employment_status Unemployed	0.1116	0.0527	2.1189	0.034100
exercise	0.0940	0.0297	3.1626	0.001564
generalhealthFair	0.8576	0.0484	17.7185	0.000000
generalhealthGood	0.4888	0.0425	11.5019	0.000000
generalhealthPoor	1.3398	0.0650	20.6141	0.000000
generalhealthVery good	0.3546	0.0421	8.4176	0.000000
helpneighbors Strongly disagree	0.4143	0.0636	6.5108	0.000000
hhsize	-0.0613	0.0113	-5.4197	0.000000
imputed_povertygroup100-<200% FPL	-0.1090	0.0383	-2.8440	0.004456
imputed_povertygroup200-<400% FPL	-0.1271	0.0413	-3.0769	0.002092
imputed_povertygroup400-<600% FPL	-0.0866	0.0435	-1.9893	0.046667
insuredgateway	0.3654	0.0520	7.0331	0.000000
maritalstatusMarried	-0.1131	0.0426	-2.6542	0.007950
maritalstatusWidowed	-0.1422	0.0549	-2.5926	0.009526
newraceHispanic	0.3171	0.0503	6.3057	0.000000
newraceOther Non-Hispanic	0.3273	0.0843	3.8842	0.000103
nutrition_fruit	-0.0376	0.0112	-3.3445	0.000824
nutrition_veg	0.0295	0.0101	2.9232	0.003465
pcp	0.2637	0.0427	6.1807	0.000000
smokerNever	-0.0803	0.0403	-1.9951	0.046032
usborn	0.8010	0.0305	26.2995	0.000000