Exploring the Environmental Impact: Can staying at home enhance Air Quality

Joseph Pulikeyil, Raghu Ram Sattanapalle & Ariane Uwase
Northeastern University, Boston, MA, USA

Abstract

This study investigates the impact of COVID-19 lockdown measures on air quality in Massachusetts, using the Air Quality Index (AQI) and the Social Distancing Index (SDI) as proxies for human mobility changes. The results reveal a weak to moderate correlation between energy demand and AQI, and a weak correlation between AQI and COVID-19 cases across various counties. Although lockdown measures had some impact on air quality, the changes were not significant, and AQI values remained within the good to moderate range. The study concludes that the relationship between AQI, COVID-19 cases, SDI, and energy demand is complex and varies by location and time, suggesting the need for further research incorporating additional factors such as population density and healthcare access.

Introduction

Air quality, a critical determinant of public health, is often quantified using the Air Quality Index (AQI). The AQI is a composite measure that incorporates various pollutants, such as particulate matter, ozone, sulfur dioxide, and nitrogen dioxide, providing a comprehensive overview of air quality [1].

The COVID-19 pandemic and subsequent lockdown measures led to significant changes in human behavior and reduced vehicle use. These changes have potential implications for air quality, and studies have reported improvements in air quality during lockdown periods [2,3]. In this study, we utilize the Social Distancing Index (SDI) derived from the study by Gae et al. (2020) [12], as a proxy for vehicle use, which captures changes in human mobility patterns during the pandemic. The SDI incorporates various metrics related to daily trips and distances traveled, related to vehicle use. By examining changes in the SDI, we can investigate the relationship between changes in human mobility and air quality. This approach allows us to understand the environmental impacts of pandemic-induced behavioral changes and their connection to air quality and public health.

In this project, we aim to investigate the COVID-19 pandemic's impact on the AQI in Massachusetts, a state that experienced various degrees of lockdown measures over the pandemic. The analysis is based on a dataset of daily AQI values from 2010 to 2023 for different counties in Massachusetts. The data is analyzed in relation to the number of COVID-19 cases reported in these counties. The analysis is divided into different time periods, including before, during, and after the lockdown, to understand the changes in AQI values over time and their correlation with the number of COVID-19 cases. To quantify the changes in human mobility and activity levels during the pandemic, we employ the Social Distancing Index (SDI). The SDI is a score-based index designed to measure the extent to which residents and visitors in a geographical area are practicing social distancing. The SDI assigns a score between 0 and 100 to each area, with 0 indicating no social distancing and 100 indicating perfect social distancing compared to benchmark days before the COVID-19 outbreak [4].

The research questions that this project seeks to answer are:

1. How did the pandemic impact AQI across different counties in Massachusetts?
2. Did AQI show significant variations before, during, and after lockdown periods?
3. Is there a correlation between COVID-19 case numbers and AQI?
4. How does social distancing, as measured by the SDI, relate to AQI changes during the pandemic?

The findings of this study could provide valuable insights into the environmental impacts of large-scale behavioral changes, such as those induced by the COVID-19 pandemic. Furthermore, understanding the relationship between human activity, air quality, and respiratory health could inform public health policies and strategies for sustainable development and pollution control [5].

This study is significant as it combines the analysis of air quality, disease spread, and human behavior in the context of a global pandemic. It provides a comprehensive understanding of the impacts of the COVID-19 pandemic on air quality in Massachusetts, contributing to the broader understanding of the environmental impacts of human behavioral changes on a large scale.

Data Sources and Methods

This study leveraged data from multiple sources and implemented a series of processing steps to ensure the appropriateness of the data for analysis.

Data Acquisition and Processing

1. **COVID-19 Data:** Data was sourced from the Massachusetts government and USAFacts websites [6, 7, 8]. The dataset, detailing daily COVID-19 cases by county, was loaded into a Python environment, date-formatted, and sorted by 'County Name' and 'date'. Daily new COVID-19 cases were computed by differentiating the cumulative case numbers. Negative values due to data corrections were replaced with zero, ensuring the dataset contained only non-negative values. Regarding the specific lockdown dates in Massachusetts, we relied on two definitions, The **lockdown period** in Massachusetts started on March 24, 2020, when Governor Charlie Baker declared a state of emergency, giving the administration more flexibility to respond to the Coronavirus outbreak [9] The first phase of reopening started on May 18, 2020, with certain sectors being permitted to operate [10]. This period can be considered the strict lockdown period. A **broader definition** that includes the period of restrictions. Once again this started on March 24, 2020, and lasted until August 1, 2021, when the state of emergency officially terminated [10]
2. **AQI Data:** The Environmental Protection Agency (EPA) Air Data - Multiyear Tile Plot tool [11] provided daily AQI values for all counties from 2010 to 2023. The CSV data was loaded into Python, and a unified dataframe was created by concatenating individual county data.
3. **Social Distancing Index (SDI) Data:** SDI was computed based on the study by Gao et al. [12], scoring each geographical area's adherence to social distancing practices from 0 to 100. The data, sourced from the Harvard Dataverse [13], was loaded and refined to include only Massachusetts counties. The dataframe was reshaped to a long format with 'date' and 'SDI' as the variable and value columns, respectively. The Social Distancing Index (SDI) is designed to capture the extent to which residents and visitors in an area are practicing social distancing, and it does so by examining changes in mobility patterns. Specifically, it incorporates variables related to the percentage of residents staying at home, the number of work and non-work trips made daily, and the average distances travelled per person. Thus, while the SDI does not directly measure vehicle use, it can serve as a useful proxy
4. **Energy Demand Data:** The data for this report was obtained from the ISO New England website. The website provides information on load and demand in the New England region. The

code provided is used to process and consolidate the data from multiple files. The files named "2011_smd_hourly.xls" to "2023_smd_hourly.xlsx" contain hourly load and demand data for each year. The code reads each file and determines the appropriate sheet to extract the data from based on the year. For years up to and including 2015, the sheet name "NEMASSBOST" is used, while for subsequent years, the sheet name "NEMA" is used. After reading the data from each file, it is stored in separate dataframe named "df_year" (e.g., df_2011, df_2012, etc.), where "year" corresponds to the respective year. Next, the code concatenates the separate dataframes into a single dataframe named "energydf" using the pd.concat() function. It also adds columns to represent the day of the year and the week number for each entry in the "Date" column. A new dataframe named "new_energydf" is created by selecting only the "Date" and "DEMAND_VALUE" columns from "energydf". The "DAY" and "WEEK" columns are duplicated and populated with the respective values from the "Date" column. Then, the data is grouped by "Date," "DAY," and "WEEK" and the demand values are summed for each group. The resulting "new_energydf" dataframe contains consolidated information on the date, demand value, day of the year, and week number, which can be further analyzed and used for the report.

Correlation Between AQI and COVID-19

Time series graphs were produced for each county, depicting 7-day rolling averages of AQI values and COVID-19 cases. This facilitated a preliminary visual representation of potential correlations. The pandas library's in-built correlation method calculated the Pearson correlation between COVID-19 cases and AQI values. A simple linear regression model, treating AQI as the predictor and COVID-19 cases as the dependent variable, was used to assess the assumptions of linear regression. Further, Spearman and Kendall correlation coefficients captured the monotonic relationships between the variables. Temporal and seasonal impacts on these correlations were investigated by computing correlations for different time periods or seasons at a county level. Lastly, mean AQI values were calculated for different lockdown periods for each county to discern the potential impacts of lockdown on air quality.

Correlation Between AQI and SDI

A consolidated dataframe was created, combining COVID-19 case numbers and SDI values. Pearson, Spearman, and Kendall correlations were calculated to examine the relationship between COVID-19 cases and SDI. Correlation analysis was conducted for each county individually, using Pearson for linear relationships and Spearman/Kendall for monotonic relationships. Temporal SDI data was plotted for each county, along with heatmaps showing key COVID-19 restriction dates. AQI and SDI data were plotted side-by-side to identify patterns, and the correlation between AQI and SDI was calculated for each county.

In summary, this study employed a comprehensive data analysis approach, yielding a deeper understanding of the relationship between SDI, AQI, and COVID-19 cases across different counties over time.

Analysis

AQI vs COVID-19 Cases

The first step in this analysis involved understanding the variation of AQI values in these counties during the pre-pandemic and pandemic periods. For the pre-pandemic analysis, the time-period from 2010-2019 was considered, and for the pandemic analysis, the period from 2020-2023 was utilized.

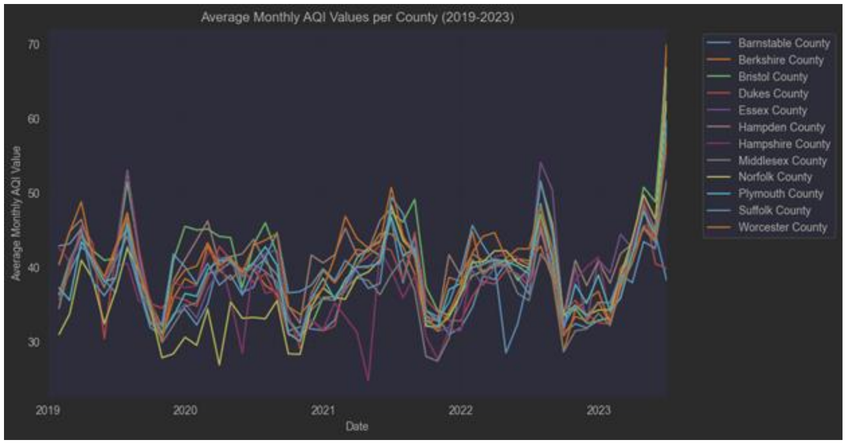

Figure 1: Average Monthly AQI Values per County (2019-2023)

This plot displays the average monthly AQI values for each county from 2019 to 2023. It allows for a visual comparison of air quality across different counties and over time.

Next, to understand the relationship between COVID-19 and AQI, a correlation analysis was performed. The overall correlation between the number of COVID-19 cases and AQI values was found to be approximately *0.007*, indicating a negligible correlation. This suggests that there is not a linear relationship between these two variables at an aggregate level.

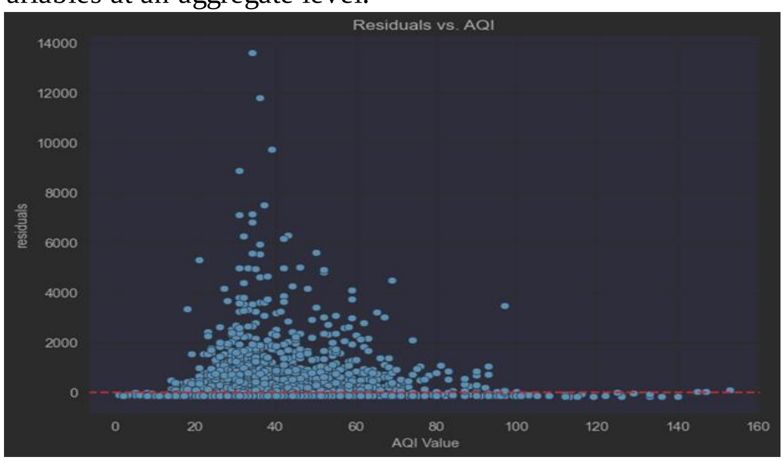

Figure 2: Residuals vs. AQI

The plot shows residuals from a linear regression model with AQI and COVID-19 cases. There is no clear pattern or trend in the residuals, indicating no strong linear relationship between AQI and COVID-19 cases. Improved air quality during the pandemic does not seem to directly affect the number of COVID-19 cases.

SDI vs AQI

To better understand the correlation between AQI and COVID-19 lockdown, we used the social distance index (SDI) as our proxy for Massachusetts residents staying in during the pandemic which would indicate a decrease in activities like driving. Figure 3 highlights how the SDI varies across time in different counties. The SDI measures the extent to which residents of a given area are practicing social distancing, with a higher number indicating greater adherence to social distancing measures.

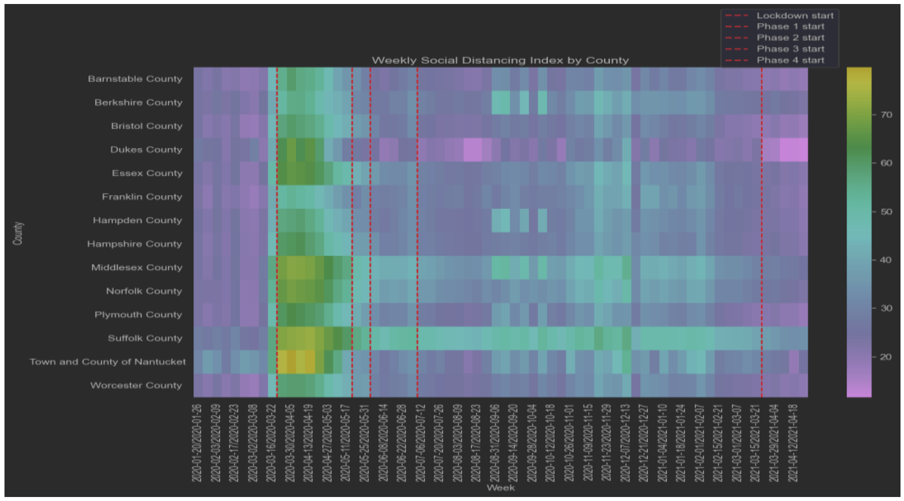

Figure 3: Heatmap of weekly social distancing index by county

During the lockdown in Massachusetts starting on March 24, 2020, counties showed high SDI values, indicating significant adoption of social distancing measures. Reopening began on May 18, 2020, with Phase 1, resulting in decreased SDI values across all counties. Variations in SDI could be due to factors like population density or adherence levels. Phase 2 in June 2020 allowed more businesses to reopen, leading to further SDI decreases. Suffolk County consistently had the highest SDI, reflecting sustained adherence to social distancing, while Dukes County had consistently low SDIs. By the start of Phase 4 on March 22, 2021, SDI values reached their lowest levels. All industries were permitted to open by May 29, 2021, with most restrictions lifted, and the state of emergency ended on June 15, 2021. The decline in SDI values reflects societal adjustment to the "new normal," balancing economic activities and public health measures.

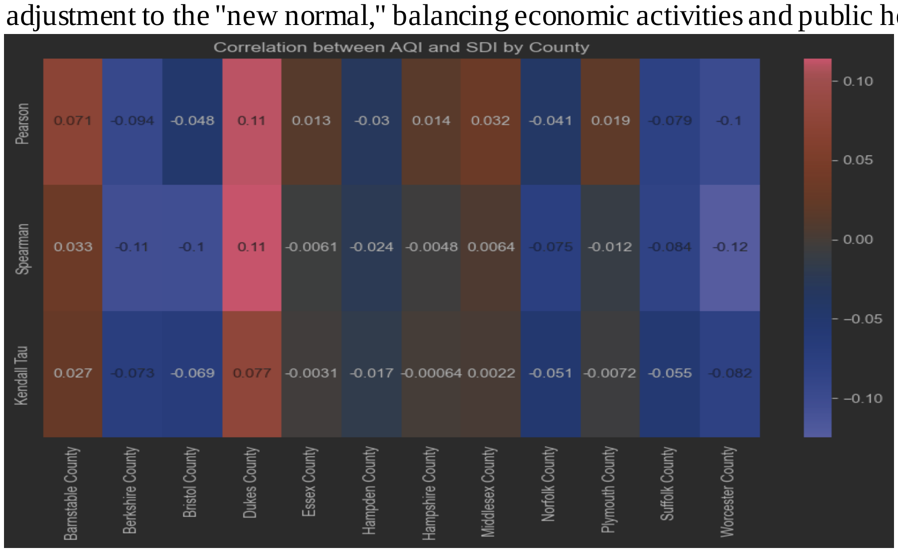

Figure 4: Pearson, Spearman, and Kendall Tau correlation by county for AQI and SDI

A correlation analysis was performed between the Air Quality Index (AQI Value) and the Social Distancing Index (SDI) for each county in Massachusetts. We performed the analysis using three different methods: Pearson, Spearman, and Kendall Tau (Figure 5). Pearson correlation assumes a linear relationship between variables and that the variables are normally distributed. In contrast, Spearman and Kendall Tau correlations do not assume a linear relationship and can handle ordinal data. Considering that SDI is a measure of social distancing which influences mobility and by extension, vehicle use and emissions, which in turn affects the Air Quality Index (AQI) we see that,

1. **Barnstable, Dukes, Essex, Hampshire, Middlesex, and Plymouth Counties**: These counties

display a small positive correlation: higher SDI (more social distancing) is linked to slightly higher AQI (worse air quality). Other factors like weather or natural pollution sources may contribute, implying reduced vehicle use from social distancing may not be the main factor impacting air quality.

2. **Berkshire, Bristol, Hampden, Norfolk, Suffolk, and Worcester Counties**: These counties show a slight negative correlation. This suggests that higher SDI (more social distancing, less travel) is associated with slightly lower AQI (better air quality). This could suggest that reduced mobility due to social distancing had a minor impact on improving air quality in these counties.

Overall, the correlations are quite weak, suggesting that while social distancing may have some effect on air quality, there are many other factors at play. In addition, because SDI is a complex measure that takes into account several aspects of social distancing, not all of which are directly related to vehicle emissions (for example, the number of people staying at home does not necessarily equate to fewer vehicle trips if those who do go out make multiple trips), it may not be a perfect indicator of changes in emissions.

Energy Demand and AQI

This report aims to go beyond mere data description and focuses on creating a model to examine the relationship between long-term trends in energy usage and its potential impact on the Air Quality Index (AQI). The study primarily focuses on analyzing energy demand in Northeast Massachusetts, specifically Suffolk, Essex, and Middlesex counties, spanning the years 2011 to 2023.

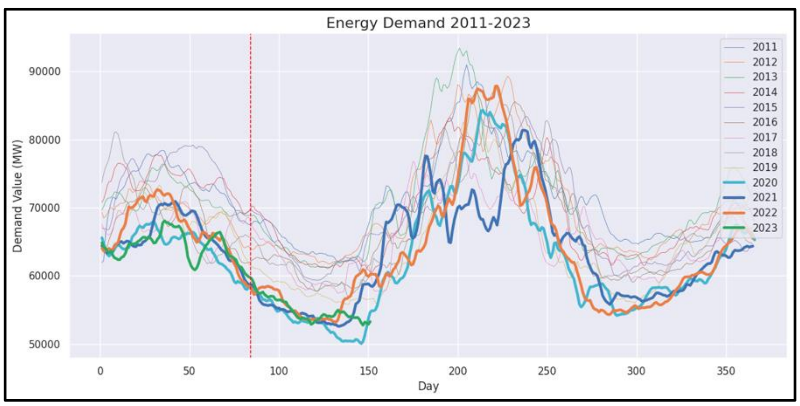

Figure 5: Energy demand per year for Northeast Massachusetts

A notable observation from the chart depicting energy demand is the decline in energy demand during the initial months of the year in post-COVID years. However, despite this decline, the peak energy demand in July remains consistent with previous years. This pattern suggests a potential shift in energy consumption behavior following the COVID-19 pandemic.

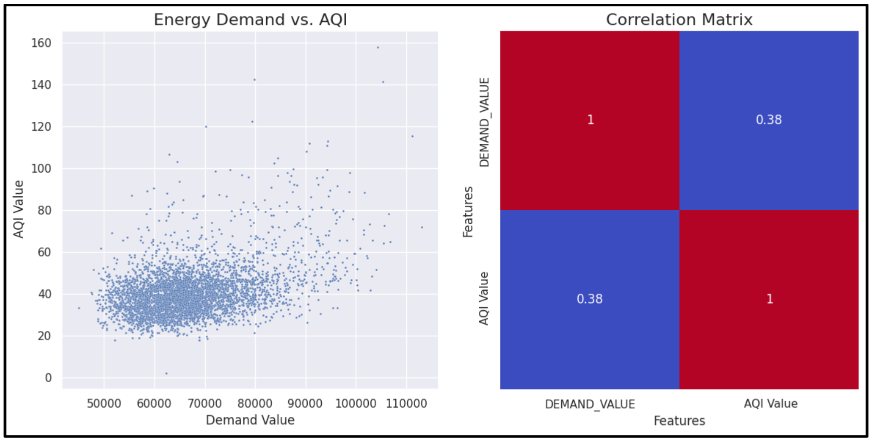

Figure 6: Scatter Plot and Correlation matrix for Energy Demand vs AQI

To assess the correlation between energy demand and AQI, a scatter plot was generated, plotting energy demand against AQI. The scatter plot reveals a correlation coefficient of 0.38, indicating a weak to moderate correlation between these variables. A correlation coefficient of 0.38 can be considered significant given the complex nature of factors influencing air quality. While the correlation is not particularly strong, it suggests a discernible association between energy demand and AQI. However, it is important to note that other variables and factors may contribute to fluctuations in air quality, which should be considered in future analyses. Further research is required to establish causality and identify additional factors influencing air quality.

Conclusions

In conclusion, the results suggest the lockdown measures might have impacted the air quality, as indicated by the changes in average AQI values. However, the changes are not significant, and the AQI values are still within the good to moderate range throughout all periods. Other factors not included in this analysis could also influence the AQI values. It is also important to note that these are average values, and there could be variations within each period. For us to understand if there was a correlation between AQI and COVID-19, we used SDI as a proxy.

However, coefficient correlations between the three are still remarkably close to zero which shows that there is little correlation between each other. The lockdown measures had some impact on air quality, but the changes are not significant, and AQI values remained within the good to moderate range. The relationship between AQI, COVID-19 cases, SDI, and energy demand is intricate and varies by location and time. Further studies should consider incorporating additional factors such as population density, healthcare access, public health measures and various other socioeconomic factors for better understanding of the relationships between these variables.

Author Contributions

Raghu and Joseph worked on data collection and analysis. Using python to inspect the following relationships; AQI vs Covid-19 vs SDI and Energy demand vs AQI. Ariane wrote and finalized the report. Each member took a turn in presenting the project during the two days science project fair.

References

1. Environmental Protection Agency (EPA). (2021). Air Quality Index (AQI) Basics. Retrieved from https://www.airnow.gov/aqi/aqi-basics/
2. Bao, R., & Zhang, A. (2020). Does lockdown reduce air pollution? Evidence from 44 cities in northern China. Science of The Total Environment, 731, 139052.
3. Collivignarelli, M. C., Abbà, A., Bertanza, G., Pedrazzani, R., Ricciardi, P., & Miino, M. C. (2020). Lockdown for CoViD-2019 in Milan: What are the effects on air quality? Science of The Total Environment, 732, 139280.
4. Zhang, F., Wang, Q., & Liu, Z. (2020). Social Distancing Index: A Data-Driven Metric for Quantifying the Effectiveness of Social Distancing. Scientific Reports, 10, 21861.
5. Chen, K., Wang, M., Huang, C., Kinney, P. L., & Anastas, P. T. (2020). Air pollution reduction and mortality benefit during the COVID-19 outbreak in China. The Lancet Planetary Health, 4(6), e210-e212.
6. Commonwealth of Massachusetts. (2023). COVID-19 Response Reporting. Retrieved from https://www.mass.gov/info-details/covid-19-response-reporting
7. Commonwealth of Massachusetts. (2023). Archive of COVID-19 Cases in Massachusetts. Retrieved from https://www.mass.gov/info-details/archive-of-covid-19-cases-in-massachusetts
8. USAFacts. (2023). Massachusetts Coronavirus Map and Case Count. Retrieved from https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/massachusetts/
9. Commonwealth of Massachusetts. (2023). Reopening Massachusetts. Retrieved from https://www.mass.gov/info-details/reopening-massachusetts.
10. Commonwealth of Massachusetts. (2023). COVID-19 State of Emergency. Retrieved from https://www.mass.gov/info-details/covid-19-state-of-emergency.
11. Environmental Protection Agency (EPA). 2023. Air Data - Multiyear Tile Plot. Available at: https://www.epa.gov/outdoor-air-quality-data/air-data-multiyear-tile-plot (Accessed: 17 June 2023).
12. Gao, S., Rao, J., Kang, Y., Liang, Y., & Kruse, J. (2020). Mapping county-level mobility pattern changes in the United States in response to COVID-19. Scientific Reports, 10(1), 1-11. Available at: https://www.nature.com/articles/s41598-020-77751-2#Sec7
13. Harvard Dataverse. (n.d.). Social Distancing Index Data. Available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZAKKCE