Analysis on how COVID-19 has impacted the usage of transportation and how the latter has affected air quality.

Arian Khademi, Ashvat Ranadive, Malhar Vivek Padir, Manit Ankit Gosalia

Friday, November 5, 2021

Problem Description and Research Question

How did COVID-19 impact the air quality in the United Kingdom, with respect to the most prominent modes of transportation.

While we were surfing the internet, we stumbled across an article on the National Geographic website[i] regarding the impact of vehicles on the already deteriorating environment. This got us thinking about whether or not the lockdown affected the environment in any way. We decided to analyze the impact of COVID-19 on transportation which may have affected air quality. Our reason to analyze the air quality in particular leads back to not only our passion for the environment but also because it is our responsibility to ensure the world is a better place for future generations. This can be done by taking small steps one at a time, and making sure that we have clean air is one of them. At the same time, we love hiking, and each weekend the four of us find time to hike at Crothers Woods, and we love that feeling of breathing fresh air at the end of our hike. We hope that the next generation can experience this feeling too, thus we decided to analyze this further with respect to the pandemic.

2020 was a devastating year for everyone, but at the very least, it had a positive impact on the environment, or did it? The UK, specifically, was hit severely with COVID-19 as there were two lockdowns in 2020, the first being from March 23rd to June 15th, and the second being from November 5th to December 2nd. Due to the lockdowns, the use of both public and private transportation had reduced, while there was an increase in cyclists. Over the last few years, there has been a growing concern among environmentalists regarding how our lifestyles have negatively impacted air quality. We will be focusing specifically on how our usage of different means of transport has contributed to air pollution. Nitrogen dioxide, ozone, and particulate matter (smoke, soot, dust, pollen, liquid droplets, etc.) are the three impurities we will be focusing on in our research.

Dataset Description

The datasets are all from the government of the United Kingdom, and their formats are '.csv'. The transportation dataset contains observations for each day in the year 2020 and the percentage usage of different modes of transport compared to the previous month's first week. This dataset contains data from multiple government datasets. The data for road traffic that will be used is calculated from DfT's Quarterly Road Traffic National Statistics series, which is used to estimate traffic change. The daily road traffic estimates are a suitable indication of traffic change. The data provided is indexed to the first week of February and each comparison is to the same day of that week. Regular traffic can vary by $\pm 20\%$. The methodology to calculate data for other modes of transportation is the same. The other three datasets represent the monthly mean concentrations of ozone, nitrogen dioxide in $\mu g/m^3$, and particulate matter for 2020.

Computational Plan

Since we are working with multiple datasets, we will first convert the ".csv" files to a list of lists to more easily perform data manipulation in python. We will open the csv files using the csv.reader() function. Once we have these files as lists it makes it easier to filter the datasets and remove null values. For the purpose of our project, we need

[i] https://www.nationalgeographic.com/environment/article/environmental-impact

to group together variables - “National Rail”, “Transport for London Bus”, “Bus (excl. London)”, and “Transport for London Tube”. This group of variables will be treated as public transport. Variable “All motor vehicles” will represent the usage of private transportation. There will be a total of three variables in the new data frame that we use: public vehicles, private vehicles, and cycling. For this dataset we want to calculate the average percentage usage of each mode of transport per month as the dataset for the air index uses months as a measure of time. Thus, we will use loops in python to create two parallel lists: a list of the average usage of the mode of transport per month and a list of months. When it comes to data visualization and future prediction based on limited data, we will use the plotly library. In this project, we aim to use plotly for statistical purposes as we want to predict how the trend in air purity index would be based on the usage of transportation assuming COVID-19 did not exist and then compare it to the actual data. The histograms will be used to represent the visualizations of each mode of transportation and the average transportation per month, both over the same time period. These will be two separate sets of histograms. For data on the air quality index, we will use a scatter plot as we want to fit models of regression and try to predict what the air quality would be like assuming the UK never went into lockdown, thus finding the true impact of COVID-19 on the air quality. This will be done using ML Regression tools on plotly. We will use Ordinary Least Square (OLS) with plotly and the functions like plotly.express.scatter() to actually show the scatterplots with regression lines, and to compare the prediction with the actual data we will make use of more rigorous methods. These will make use of “scikit-learn” to use functions like LinearRegression.predict() to predict the outcome for the unseen data. This alongside the OLS functions will help perform the visualisation on the predicted data. These functions in the plotly, OLS and scikit-learn libraries help perform calculations, determine regression coefficients and plot the data in scatter plots and we will study and implement them to our data. ML-based regression models only allow us to do this much, however, we can then compare the regression model with the actual data scatterplot to derive whether COVID-19 had an impact on the air index or we would have had the same outcome if the pandemic had never happened.

References

“Concentrations of Nitrogen Dioxide.” GOV.UK, National Statistics Concentrations of Nitrogen Dioxide, 29 Apr. 2021,
https://www.gov.uk.glasswall-icap.com/government/statistics/air-quality-statistics/ntrogen-dioxide.

“Concentrations of Ozone.” GOV.UK, National Statistics Concentrations of Ozone, 29 Apr. 2021,
https://www.gov.uk/government/statistics/air-quality-statistics/concentrations-of-ozone.

“Concentrations of Particulate Matter (PM10 and PM2.5).” GOV.UK, National Statistics Concentrations of Particulate Matter, 29 Apr. 2021,
https://www.gov.uk/government/statistics/air-quality-statistics/concentrations-of-particulate-matter-pm10-and-pm25.

“Covid-19 Domestic Transport Data: Methodology Note.” GOV.UK, 29 Apr. 2021,
https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic/covid-19-transport-data-methodology-note.

“Getting Started with Plotly.” Plotly,
https://plotly.com/python/getting-started/.

“Transport Use during the Coronavirus (COVID-19) Pandemic.” GOV.UK, 3 Nov. 2021,
https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic.