Assessing the Impact of ULEZ on Air Quality in London



Group-3

Data-1

CFG-Summer-2023

# 

# Introduction

## Research question

The main goal of this project was to analyse and assess the impact of the Ultra Low Emission Zone (ULEZ) on air quality in London. We wanted to know **if the introduction of low emission zones had reduced the levels of Nitrogen Dioxide (NO2) in London’s air.**



**“The ULEZ has contributed to almost 50%  
reduction in toxic Nitrogen Dioxide pollution in central London.”**

**- Greater London Authority**

We expected to see that:

* **the level of NO2 particulate matter was higher in London boroughs before the introduction of ULEZ**
* **the level of NO2 particulate matter in a London borough decreased following the introduction of ULEZ**

We considered the possibility that the longer a borough had adopted ULEZ the better the air quality i.e., NO2 levels and time elapsed would have a relationship. We acknowledged that additional factors including changing public attitudes, advancements in cleaner car technology, the cost of living crisis and COVID19 would also have an effect on this hypothesis. We felt that these factors were beyond the scope of this study but would be an interesting consideration for a future project.

We chose weather data as our control variable in this project. We understood that weather and air quality were closely linked and therefore felt it was appropriate to include it.

## Background

The Ultra Low Emission Zone (ULEZ) is an area in London, UK, where an emissions standard based charge is applied to non-compliant vehicles. The scheme was conceived to address the high levels of Nitrogen dioxide (NO2) in London's air. The first phase was introduced in April 2019 and encompassed the existing Central London Congestion Charge zone (CCZ). The area was expanded in October 2021 to cover the Inner London area inside the North Circular and South Circular roads. A proposed expansion is due to come into effect on 29th August 2023 and will include all 32 London boroughs.

Proponents of the scheme, namely the London Mayor's office, claim the data shows NO2 pollution has reduced by nearly 50% in Central London and that pollution has reduced five times faster in London than elsewhere in the UK between 2016 and 2020 ([london.gov.uk](https://www.london.gov.uk/programmes-strategies/environment-and-climate-change/pollution-and-air-quality/ultra-low-emission-zone-ulez-london)). A study published in 2021 by researchers from Imperial College London suggests that the data shows a much more modest change to NO2 levels in London's air ([Liang Ma et al 2021 Environ. Res. Lett. 16 124001](https://iopscience.iop.org/article/10.1088/1748-9326/ac30c1)).

## Project aims

In order to make the project achievable:

* **we focused on NO2 levels as a measure of air quality. The larger the particulate number the worse the air quality**
* **we analysed how NO2 levels have changed over time highlighting key dates such as the introduction of new zones**
* **we tried predictive modelling to show how NO2 levels will change in the future**

Audience

Our project can appeal to a diverse audience with varying expertise and interests. This is a brief overview of our target audience and their potential motivations for reading our report:

**General Public and Community:**

including local residents, environmental activists, parents, and individuals who care about sustainability. Our report can help raise awareness about the ULEZ's effects on air quality, providing insights that people can use to make informed decisions about transportation choices and advocacy efforts. This audience might not have a deep technical understanding of data science concepts.

**Policy Makers and Government Officials:**

Government bodies, local councils, and policymakers are key stakeholders interested in the effectiveness of measures like the ULEZ. Our findings can influence decisions related to environmental policies, public transportation, and urban planning.

**Transportation and Urban Planning Professionals:**

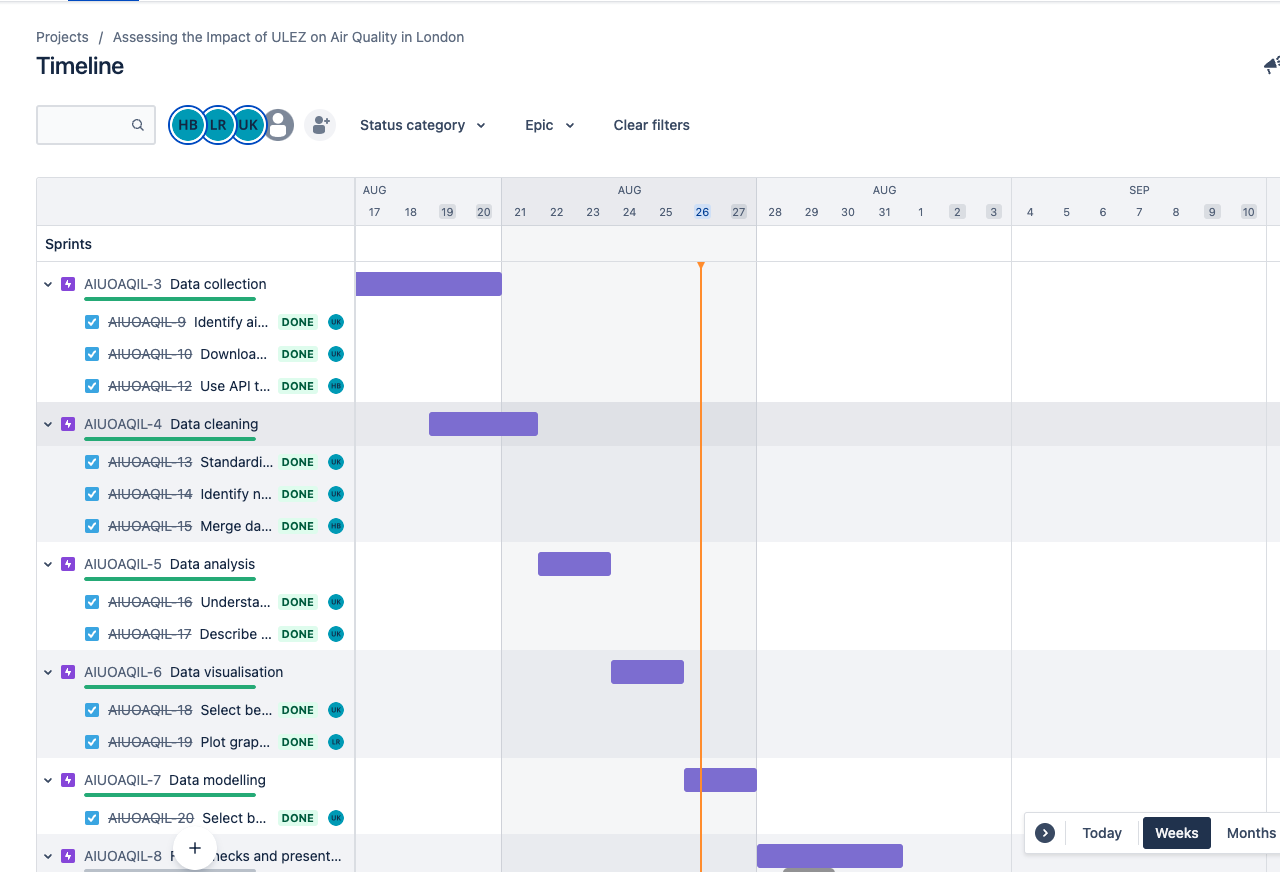
Professionals working in transportation, urban planning, and environmental fields would be interested in understanding how the ULEZ has impacted air quality and whether the goals of the initiative are being met.

**Data Scientists and Researchers:**

Data scientists and researchers might be interested in your methodology, techniques, and results. The project could contribute to the broader field of air quality analysis and urban planning.

# Specifications and Design

We began planning our project in Trello to share ideas and resources. We quickly identified that data collection and preprocessing was going to take the majority of our time so we visualised a timeline in Jira starting with the data collection tasks and assigning them to team members:



Each task within the sprint set out what was required and team members could keep each other up-to-date by adding notes and comments.

Most of the code was written in Colab as this was the easiest way to share notebooks between team members. We stored files in a shared Drive folder so we could all mount the same resources and check the code. All of our code was written in Python. We had considered creating an SQL database to store and organise the data however we felt it would be more time efficient to work with one language and one interface. We created a GitHub repository in which we stored clean versions of our files.

# 

# Implementation and Execution

## Data collection

Our data collection process consisted of three main parts: **1. identifying air monitoring sites** within the ULEZ boundaries; **2. using an API** to call relevant data **3. installing a library** to gather historical weather data.

**Sources:**

* [Local Government website:](https://apps.london.gov.uk/air-quality/) An interactive map of London showing ULEZ boundaries and air monitoring sites. We used this in the discovery phase to get a sense of the geography and what metrics we could use. We decided to focus on the Regulatory Air Quality Monitoring sites as they were more extensive, most of them recorded NO2, and most had data for our timeframe. We stored this data in a .tsv file.
* [London Air:](https://www.londonair.org.uk/LondonAir/Default.aspx) website of the London Air Quality Network (LAQN). LAQN was set up in 1993 to coordinate air pollution monitoring. The website is maintained by the [Environmental Research Group](https://www.imperial.ac.uk/school-public-health/environmental-research-group/) at Imperial College London. The monitoring is owned and funded by local authorities, Business Improvement Districts (BIDs), TfL and Defra. We used [London Air’s API](https://www.londonair.org.uk/Londonair/API/) to gather pollution data for the sites we had identified from the interactive map. We added this data to a .csv file.
* [Meteostat:](https://dev.meteostat.net/) a free source of historical weather data. We used the Meteostat Python library to obtain daily average weather information based on the latitude and longitudes of each pollution monitoring site. We captured this data in a .csv file.

**Considerations:**

We wanted to work with a large dataset and we needed it to be open, reliable and reflective of our timeframe. This was relatively straightforward once we understood who the key stakeholders were. We had to discuss how we were going to gather the site information evaluate how useful the site would be for your project e.g., when did the site open, was it open for the full duration of our time period, did the site record NO2, was it clear which phase of ULEZ they were in.

When it came to using an API we needed to be unrestricted with the ability to make unlimited calls as we were trying to learn how to pull the data.

There were several options for historical weather data but we decided to use Meteostat as it was unrestricted and the library option for Python was appealing. It also provided daily averages which worked well with our air pollution data which was also daily.

**Challenges:**

We knew that the topic was complex and that our background knowledge was limited. We also knew that our ambition to gather data from multiple sources and merge it would be a big challenge. Admittedly the data collection phase was time consuming and we had to do a lot of reading to understand if we were collecting the correct information however it was enjoyable. In the collection process we encountered monitoring sites with incomplete data and missing values. We tackled these issues in the data preprocessing phase.

## Data preprocessing

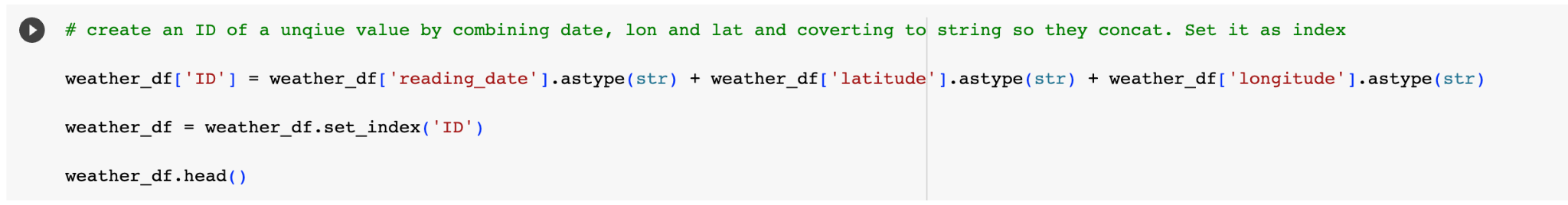
We cleaned at each stage of the gathering process to make sure that what we had extracted was going to be usable and mergeable. It also allowed different members of the team to work on the collection and processing tasks simultaneously, which saved time.

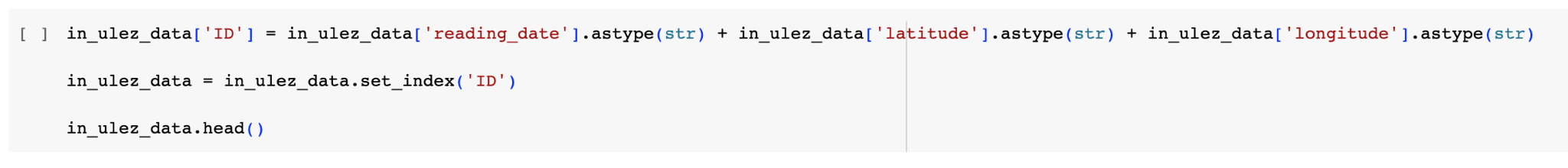
We imported our data into Jupyter notebooks to read the data. We used Pandas to convert the data into DataFrames. We used techniques associated with merging, standardisation, normalisation, de-duplication and handling missing/NaN values.

Here are some examples:

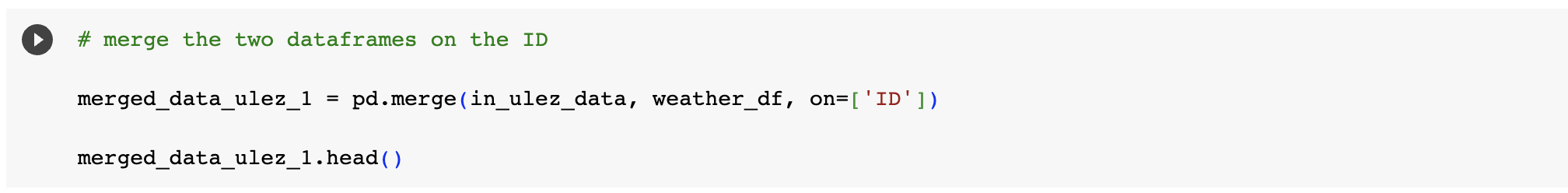
Merging

We needed to create a primary key in our datasets before merging. We had multiple data points for each site so site name or code wasn’t enough. We needed to isolate what was unique about the site (latitude and longitude) and combine it with the date field (reading date):



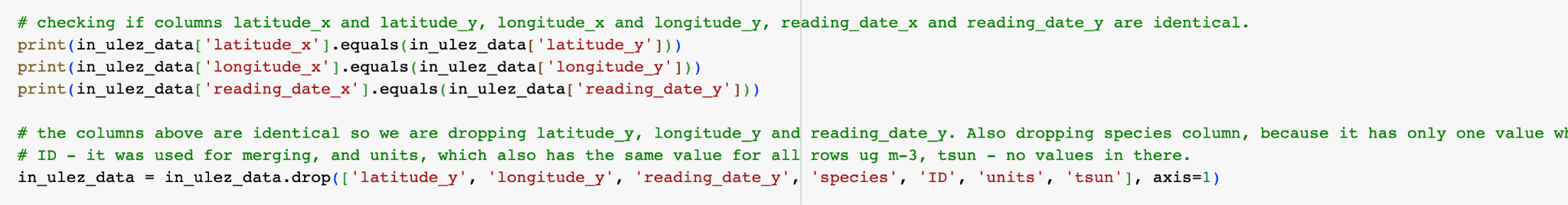


It was then possible to use the Pandas merge method to combine the data into a single DataFrame:



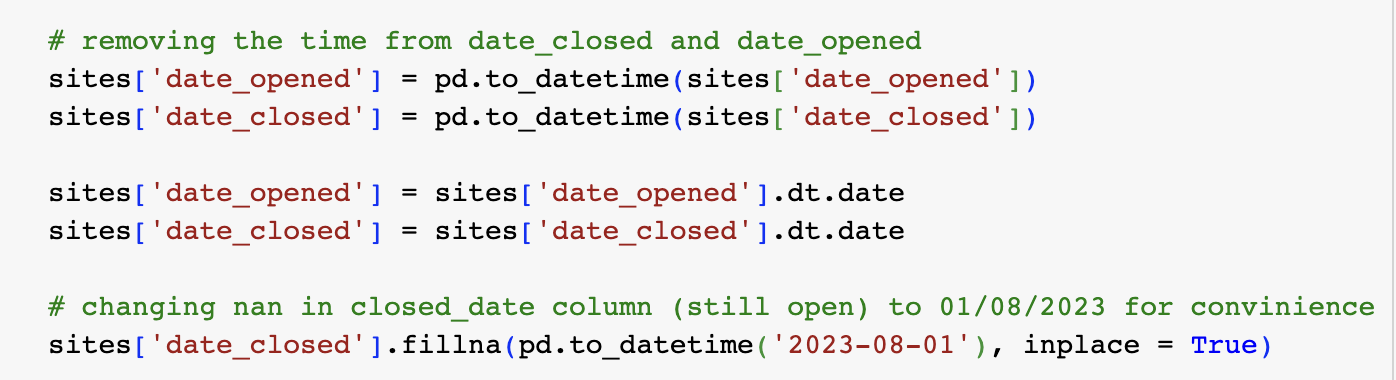
De-duplication

Once we merged we cleaned the data further to remove superfluous columns:



Standardisation

The date was an important variable in our data. In order to make it uniform and readable we used the Pandas datetime function:



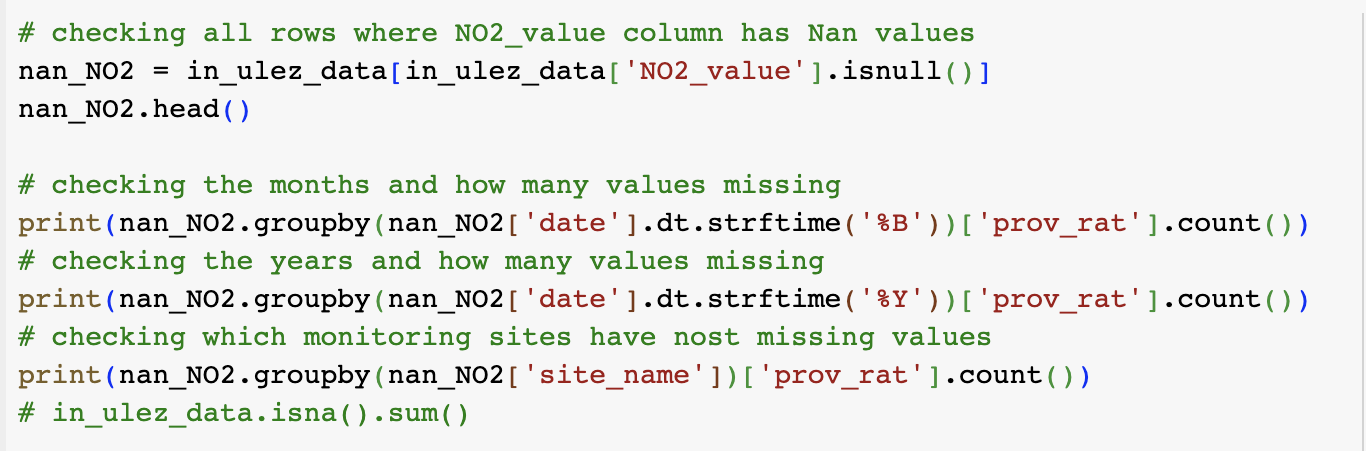
Normalisation

In order for us to be able to query the data easily we needed to ensure that our column names and cases were consistent. We ended up with an Initial ULEZ dataset and an Expanded ULEZ dataset and it was imperative that there was normalisation across the data:



Missing values

We discussed different options for dealing with NaN values in our data. Our data contained daily averages and filling missing days would have required us to spend a significant amount of time on working out monthly averages or taking data from previous or future years and carefully filling the gaps. We felt our data was comprehensive enough for us to drop rows with NaN values and allowed us to move on to visualisation:



**Considerations:**

Amongst our considerations for data preprocessing and cleaning were: How will we merge data, how difficult will it be to extract the relevant sites, how will we make a decision on which data to drop.

**Challenges:**

Time was a limiting factor as we would have liked to have explored how we could fill null NO2 values with historical averages. This would have required us to explore additional data sources and gather more historical data.

## Data analysis and visualisation

We knew that data visualisation was going to be important in the storytelling of our analysis for this project. We did extensive research to find the most appropriate tools for analysis and which visualisations would help convey the results to our audience.

In our discovery phase we found maps to be an excellent tool for providing context to our understanding of ULEZ so we wanted to incorporate our own into this project. We found an excellent [tutorial](https://towardsdatascience.com/lets-make-a-map-using-geopandas-pandas-and-matplotlib-to-make-a-chloropleth-map-dddc31c1983d) with an explainer on how to use Geopandas, Pandas and Matplotlib.

Forecasting

Our data has the historical dates the NO2 readings were captured which is perfect for time series analysis. We experimented with different models and they produced unexpected results. So we chose to use the Prophet package to produce the forecast using monthly averages. Prophet is an open source software released by Facebooks’s data science team.

## **Challenges:**

In the initial stages we observed that each future year was showing a repetitive trend, see NO2 static yearly pattern.PNG. To increase the reliability of the model we add in UK holiday dates and adjust the model to account for the seasonality and the output fit into our hypothesis that the level of NO2 particulate matter in a London borough will decrease following the introduction of ULEZ.

Our findings summary towards the end of this report goes into more detail about what the visualisations show.

**Considerations:**

We wanted to make sure that we applied the best analysis for our data types. For example, we wanted to visualise the continuous data (NO2, temperature) against our time series so line charts proved to be particularly useful. We really enjoyed this part of the process as at this point we were starting to test our research question and formulate conclusions.

**Limitations:**

Our lack of academic knowledge in this field meant that we could not confidently check if the data was behaving as it should. Learning about and trying to understand the syntax of new plots was often challenging, however we got better with practice and discovered new visualisations that we had not considered before this project.

## Experimenting with modelling

We really wanted to explore modelling in our project and use predictive analysis to make future predictions about the success of ULEZ on tackling air pollution. We experimented with a few models but found linear regression to be the best for our dataset and our level of understanding.

We played around with Random Forest Regression, Gradient Boosting Regression, Support Vector Regression, ElasticNet Regression.

**Considerations:**

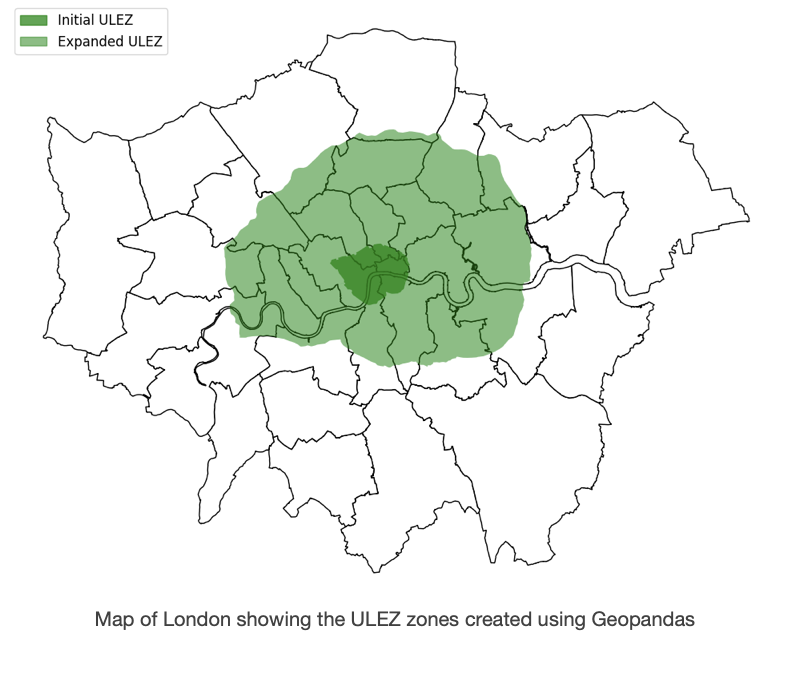
We wanted to demonstrate our interest in modelling and show how this project could be developed given more time. We felt linear regression was a good starting point.

**Challenges:**

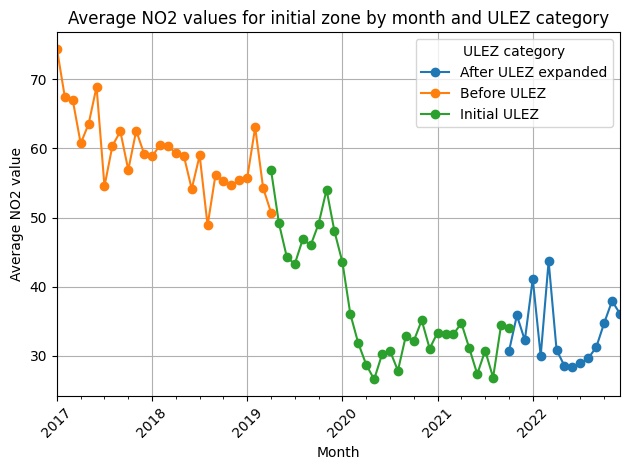
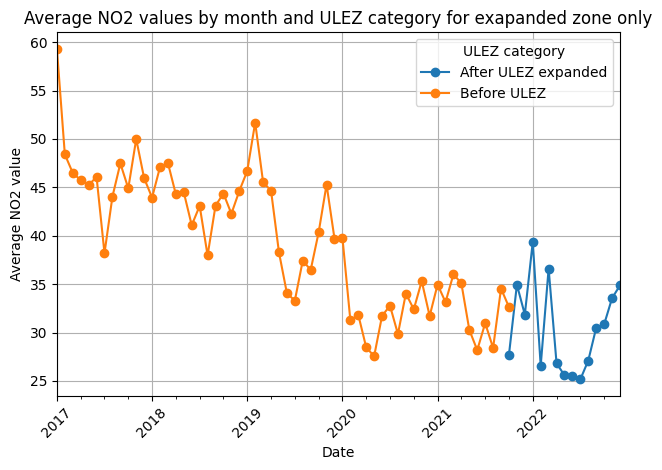
Our knowledge of predictive modelling was limited and we did not have enough time to deep dive into the subject. We did however enjoy experimenting and would like to see if we could develop this project in the future.

## Summary of findings

This project set out to find **if the introduction of low emission zones had reduced the levels of Nitrogen Dioxide (NO2) in London’s air**.

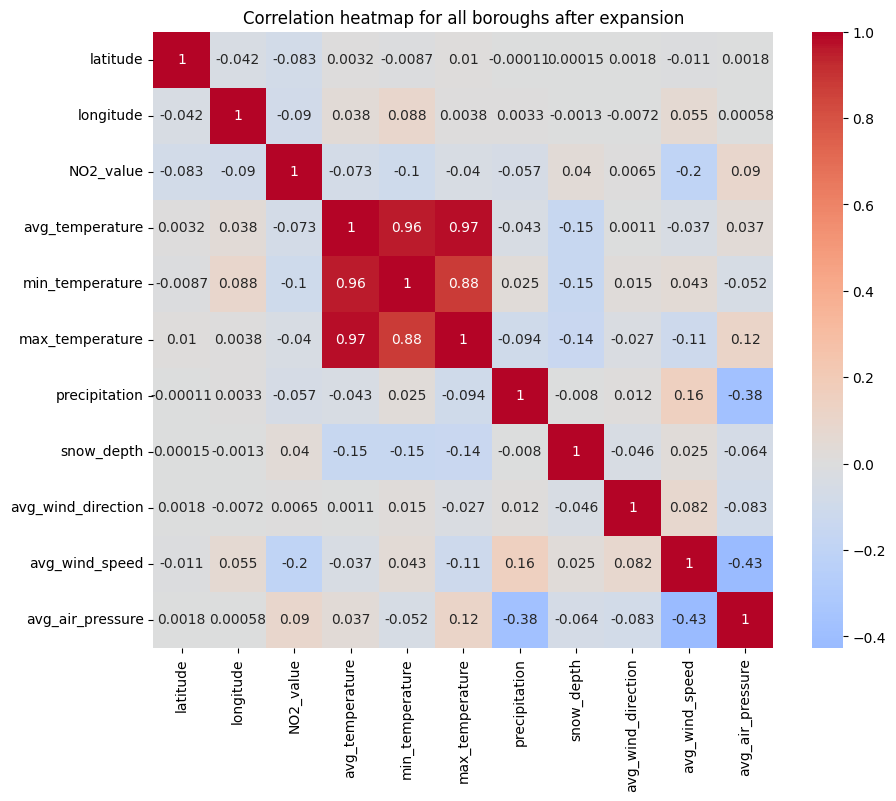


Our visualisations showed that NO2 particulate matter did reduce from 2017 to 2023 in the boroughs that ULEZ was in effect. We can also see a decline in line with the expansion timings. This suggests that **the introduction of ULEZ has produced a positive impact on reducing NO2 levels**:



There is a general decline in average monthly NO2 when ULEZ zones are introduced. We can see a steep reduction in the first half of 2019 and again around the early part of 2020. This is when COVID-19 lockdown measures came into effect meaning that the number of vehicles on the road would have been significantly reduced. Even so the plots show a positive improvement in NO2 levels post covid 21/22 onwards after the ULEZ zone has been expanded.

We included weather data in our project because our background reading suggested that higher temperatures contributed to higher pollution levels. Our findings did not support this explicitly however using daily averages may have had something to do with this. Further research and a better understanding of climate factors may reveal additional metrics that need to be considered when exploring this correlation.



Throughout this project we acknowledged that we were taking a simplified approach to this topic. We would have liked to include traffic data as we felt that this would have given a fuller picture of the impact of ULEZ. In particular, we understood that ULEZ probably impacted the number of cars in the zones which in turn impacted the levels of traffic pollution. In addition, car type and car fuel are likely to be other factors. Adding this data to the project would be a great next step.

## Conclusion

This was a challenging project that tested all of the knowledge and skills we had amassed over the past four months. We enjoyed taking a subject that we were interested in and digging deep into the analysis. At times we thought we may have chosen a topic that was too complex but once we started to visualise the data we could see that our research question was sound and the methods we had used were solid. It was interesting to explore Python libraries and stretch ourselves with the visualisations. Using Geopandas was a highlight and definitely something we want to keep exploring in the future.