Air Pollution in Brno

PV251 Project, Filip Gregora

I chose this topic because, during this semester, I was searching for a new apartment to rent. When choosing an apartment there are many properties based on which we will rank the apartment. These factors include price, size, distance to work or school, proximity to leisure activities, and air quality.

I believe air quality is particularly important because we spend most of our time in our homes. However, information about air quality is often difficult to access. This inspired me to create a visualization that presents air quality levels across Brno.

Data sources

I used the dataset titled *Kvalita ovzduší - data z mobilních stanic (Air quality on site measurements)*, which is accessible at:
https://data.brno.cz/datasets/c0d62b5fe8474c22ad74721b6fbb92dd_0/explore
This dataset contains measurements done by a mobile air quality monitoring station collected over one year. Since Brno has only one station, measurements were conducted at 12 different locations. Each location contains approximately 55 measurements, but these measurements were not taken simultaneously.

To better analyze the data, I grouped the measurements by season. The seasons were defined as follows: winter ends on March 21, spring ends on June 21, summer ends on September 21, and autumn ends on December 21. Due to the limited number of measurement locations, I computed values for every possible location in Brno using interpolation.

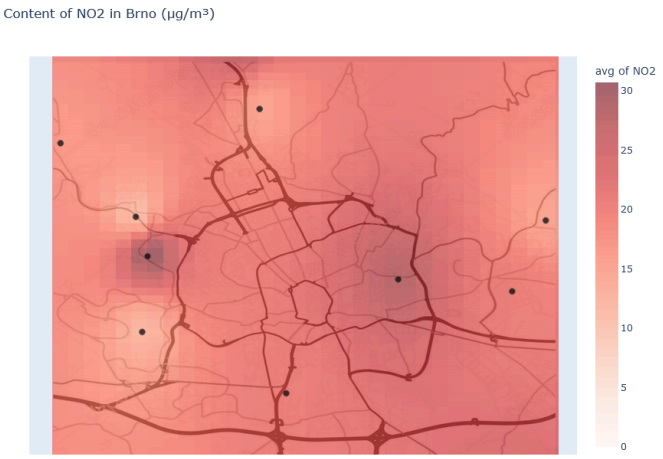
$$val_{X,Y} = \frac{1}{\sum_{i=1}^{n} (X-X_i)^2+(Y-Y_i)^2} \cdot \sum_{i=1}^{n} \frac{val_i}{(X-X_i)^2+(Y-Y_i)^2}$$

Where n is the number of places where measurements have been taken, X_i is X coordinate at place i, Y_i is Y coordinate at place i and val_i is value at place i.

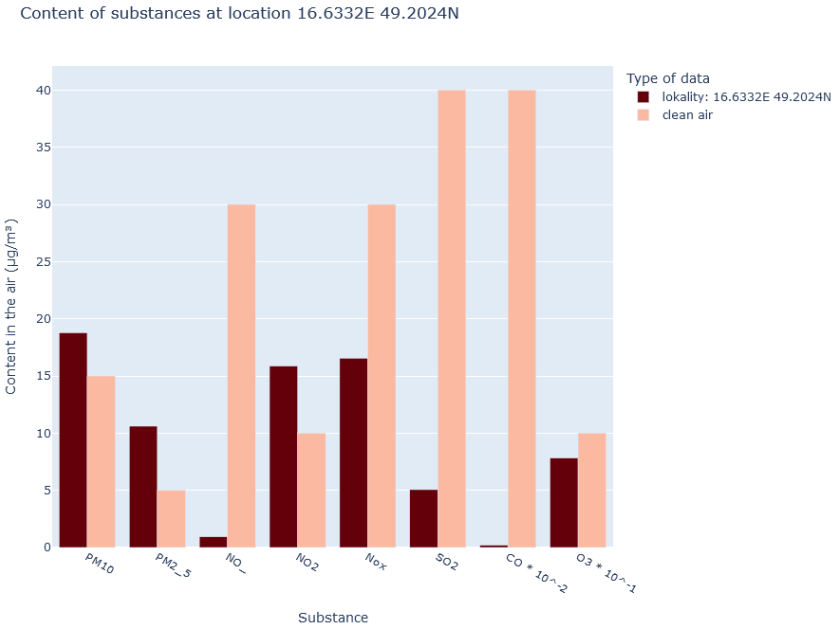
Each point's value is calculated based on data from all mobile station locations, weighted by the square of the distance. This approach assumes that air pollution spreads in a circular pattern in a 2D space.

Design

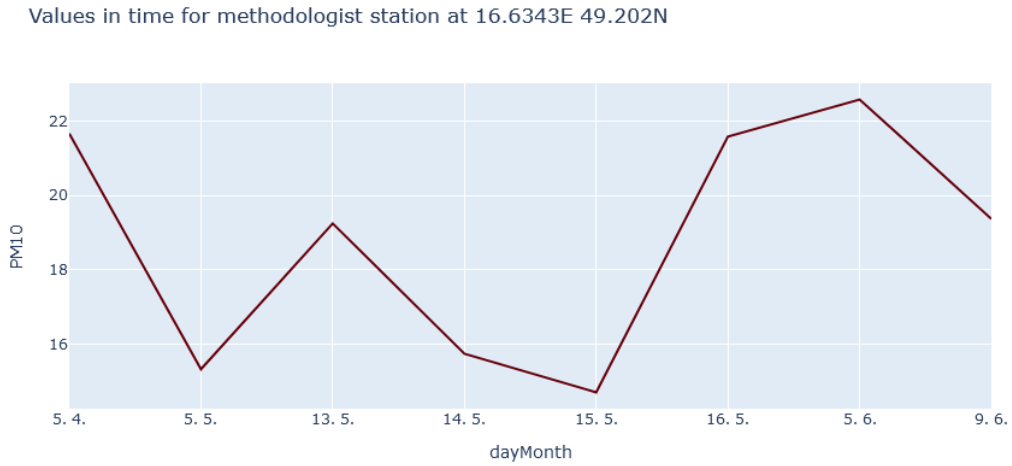
First presented plot is the density histogram. The histogram visualizes pollution levels across Brno using a grid of evenly spaced points. Each point's pollution value is precomputed. Users can select a specific pollutant, season, and whether to fix the range for easier comparisons across seasons. A map of Brno below the histogram provides context, with black dots indicating measurement locations. This plot shows the pollution level in Brno, so we can easily compare two different places and identify worse locations for living.



Second presented plot is the bar plot. The bar plot displays the levels of all measured pollutants at a selected location for the chosen season. The bars are compared against WHO standards for clean air, helping users quickly identify if the location meets health guidelines. This plot is showing us all pollution in a selected place, so we can easily compare if the location fulfills health limits.

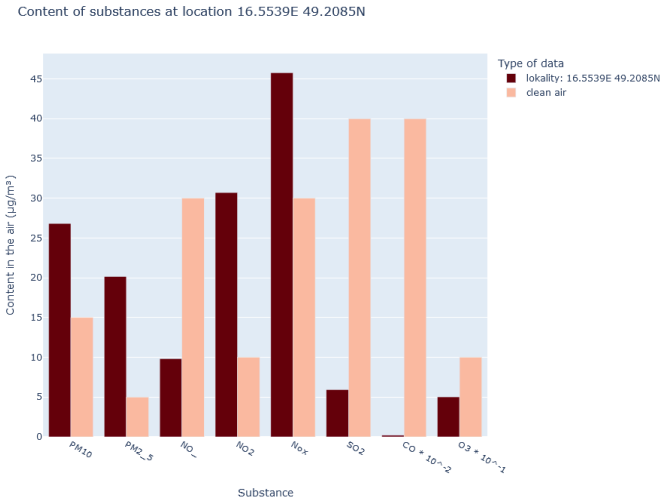


Last presented plot is a line plot. This line plot shows values of a selected substance in the chosen season for closest places, where measurements have been taken. This plot is showing us how trustworthy values are from the closest place. The closest place has the biggest impact on computed value.



Interesting observations

One of the most interesting observations was that some locations in Brno did not meet WHO air quality standards. For example, areas closer to industrial zones or closer to roads tended to have higher pollution levels. This highlights the importance of considering air quality when choosing a home.



Technologies

This visualization was created using Python, with Jupyter Notebook for development and Plotly for interactive visualizations. Additional Python libraries, such as pandas for data manipulation and numpy for calculations, were used extensively. To run the visualization, users need to install Python, set up a virtual environment, and install the required dependencies.

Lessons learned

The most challenging aspect of this project was learning to navigate the Plotly environment. While it offers a wide range of features for creating interactive visualizations, its complexity made it difficult to get started.

The hardest subtask for me was to implement the image of a map of Brno into visualization and compute all coordinates for density histogram.

On the other hand, this project provided valuable insights into data visualization, enhanced my proficiency with the Jupyter environment, and improved my skills in using pandas for data processing.