Data Engineering (LTAT.02.007) project proposal
Group 19 - Toomas Kangur, Uku Kangur

Research problem description

Air quality is a critical factor influencing public health worldwide. Prolonged exposure to poor air can lead to serious health issues such as strokes, ischemic heart disease, lung cancer, pneumonia, and more. Understanding the factors that affect air quality is therefore essential for mitigating these risks.

In our project, we aim to explore how different **meteorological features,** such as temperature, wind speed, and direction, can impact the **air quality** in Tallinn. We focus on how these conditions influence the movement and concentration of pollutants in the air. For example, strong winds tend to disperse pollutants, improving air quality, while stagnant air can trap them near the ground, leading to increased pollution levels. Temperature inversions, where warm air traps cooler air near the surface, can prevent pollutants from dispersing. Higher temperatures can also accelerate chemical reactions in the atmosphere, contributing to the formation of ground-level ozone. In our project, we assess whether these phenomena also occur in Tallinn and to what extent, by analysing their correlations and temporal patterns.

Datasets

Dataset 1 - meteorological features (temperature, wind)
Our meteorological features come from the Estonian Ilmateenistus portal. We will be using the sub-dataset from Tallinn-Harku meteorological station as we can get the air pollution numbers from the same region. They provide the temperature (average, minimum, and maximum) and wind (speed and direction) data on an hourly basis from 2004 January to 2024 June. The data is given in .xlsx format.
link to the data source: https://www.ilmateenistus.ee/kliima/ajaloolised-ilmaandmed/

Dataset 2 - air pollution measurements (PM10, SO2, NO2, O3)
Origin of the air pollution data is Eesti Keskkonnauuringute Keskus. We requested this dataset directly from the institution. Daily data covers the period 01.01.2001 to 31.12.2020. The dataset includes the following air content measurements: PM2.5-10 ($\mu g/m^3$), PM2.5 ($\mu g/m^3$), PM10 ($\mu g/m^3$), SO2 ($\mu g/m^3$), NO2 ($\mu g/m^3$), and O3 ($\mu g/m^3$).
link to the data source: https://www.klab.ee/seired-ja-uuringud/ohuseire/

Main research questions

Research question 1: Does temperature influence air pollution levels in Tallinn, and if so, how?

Research question 2: Does wind speed influence air pollution levels in Tallinn, and if so, how?

Research question 3: On which days of the week, months, and seasons do meteorological features have the greatest and least impact on air pollution in Tallinn?