Friedrich-Alexander-Universität
Technische Fakultät

Correlation between air pollution and number of vehicles with combustion motors in Nordrhein-Westfalen.

Joaquin Ayzanoa

2024-01-17

Agenda

01 Main question

02 Datasources

03 Datapipeline

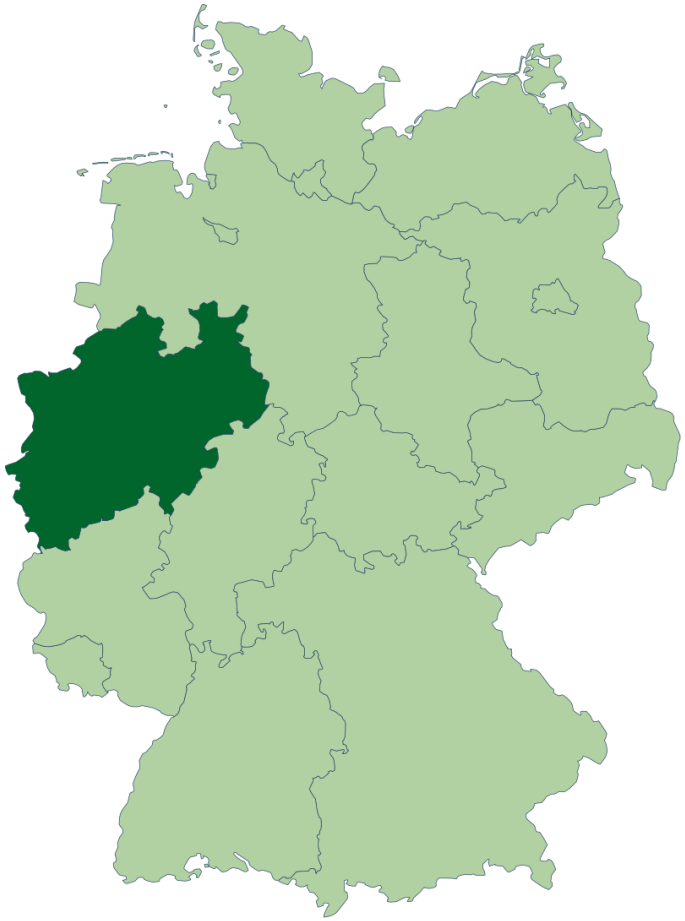
04 Findings

05 Conclusions

Air Pollution

Main question

1.Is the number of combustion motor vehicles the main factor in air pollution in Nordrhein-Westfalen?



Datasources

Datasource1: Stock of motor vehicles by vehicle type in Nordrhein-Westfalen

- Metadata URL: https://mobilithek.info/offers/-4132669826481765343
- Data URL: https://www.landesdatenbank.nrw.de/ldbnrwws/downloader/00/tables/46251-02iz_00.csv
- Data Type: CSV
- Stock of motor vehicles by motor vehicle type of cities: Köln, Münster, Detmold, Arnsberg, Düsseldorf

Datasource2: Annual parameters of air pollutants in Nordrhein-Westfalen

- Metadata
  URL: https://www.opengeodata.nrw.de/produkte/umwelt_klima/luftqualitaet/luqs/eu_jahreskenngroessen
- Data
  URL: https://www.opengeodata.nrw.de/produkte/umwelt_klima/luftqualitaet/luqs/eu_jahreskenngroessen/LUQS-EU-Kenngroessen-2022.xlsx
- Data Type: xlsx
- Annual parameters of air pollutants in Nordrhein-Westfalen for 2022: Nitrogen dioxide, fine dust (PM10), fine dust (PM2.5), sulfur dioxide, benzene, lead, arsenic, cadmium, nickel, benzopyrene

Datapipeline

Datasourse1 is a csv file and the information extraction process was as follows:

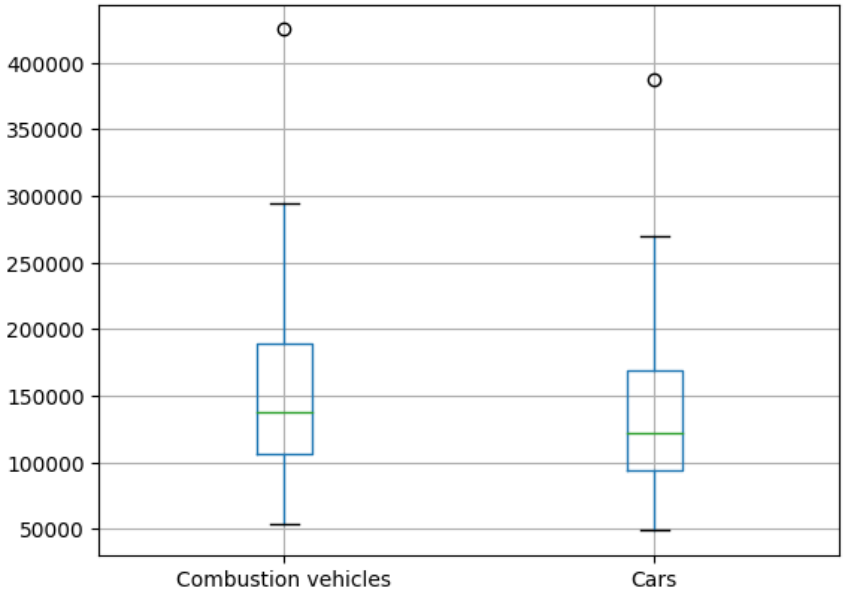
1. The csv was read through the pandas library using 'pandas.read_csv'. Since it was not necessary to read all the rows of the file, the 'skiprows' and 'nrows' parameters were added with which you can read from a row 'x' to a row ‘y’.
2. Then, the dataframe column names were renamed to easily understandable ones.
3. Next, the missing information that has been represented as '-' in the csv file was removed.
4. Finally, A sql database was created with the designated database name using 'sql.create_engine', the variable types of each column were set, and the table was saved with the designated name.

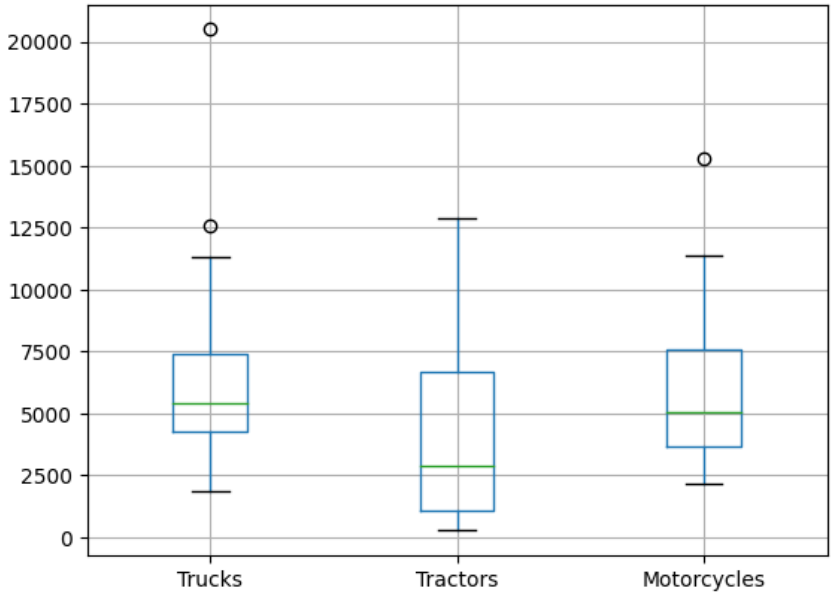
Datasourse2 is a xlsx file and the information extraction process was as follows:

1. The xlsx file was captured through the pandas library using 'pandas.ExcelFile(url). Then, the sheet ('EU-Jahreskenngrößen 2022') of the xlsx book, where the table you want to read is located, was selected. Subsequently, the rows and columns of interest in the dataframe were selected.
2. Then, the dataframe column names were renamed to easily understandable ones.
3. Next, The strings '-', '--', 'nan' were replaced with np.nan value.
4. Finally, a sql database was created with the designated database name using 'sql.create_engine', the variable types of each column were set, and the table was saved with the designated name.

Findings

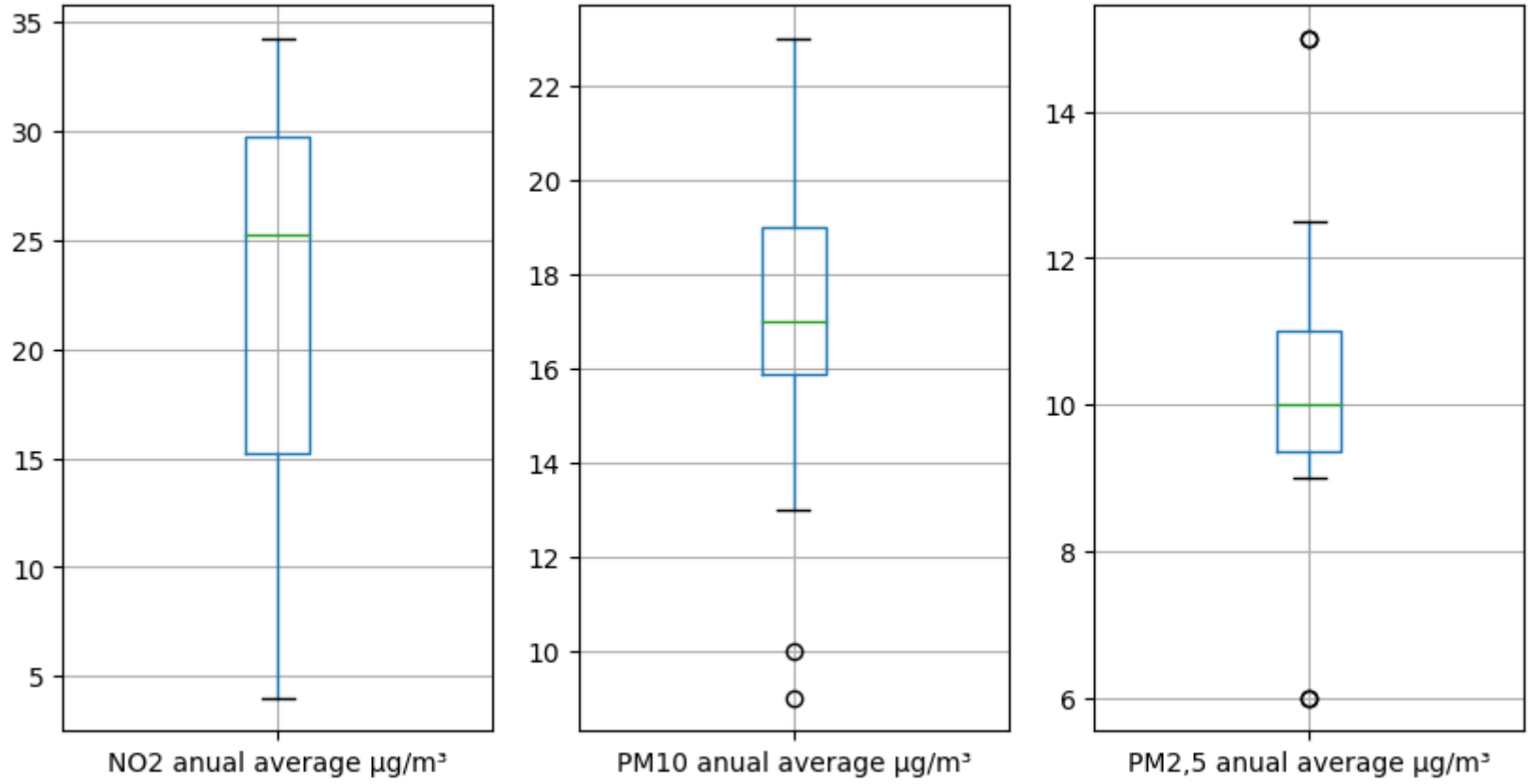
Number of vehicles





Findings

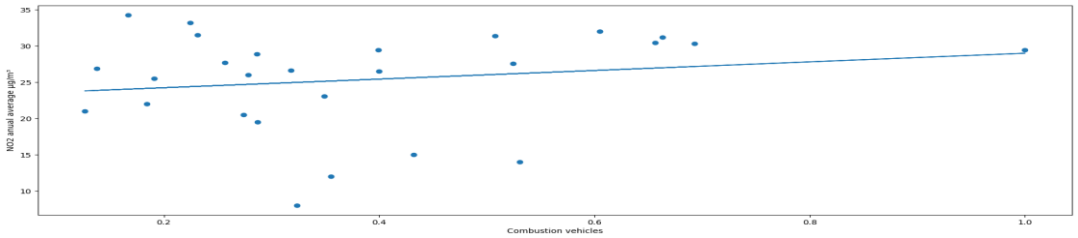
Pollutans

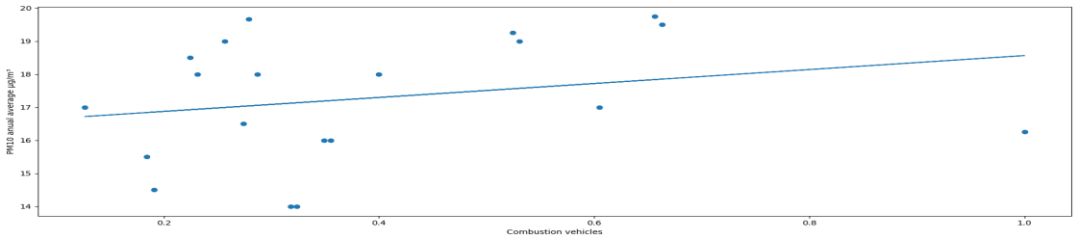


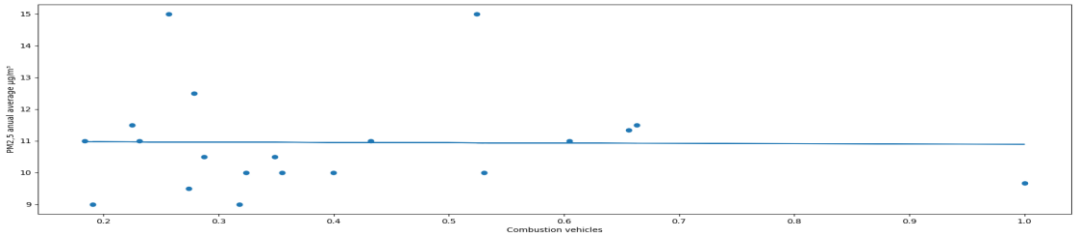
Findings

Pollutans

	Combustion vehicles	NO2 anual average µg/m³	PM10 anual average µg/m³	PM2,5 anual average µg/m³
Combustion vehicles	1.000000	0.243302	0.241226	-0.012554
NO2 anual average µg/m³	0.243302	1.000000	0.416301	0.328342
PM10 anual average µg/m³	0.241226	0.416301	1.000000	0.647095
PM2,5 anual average µg/m³	-0.012554	0.328342	0.647095	1.000000







Conclusions

1. The analysis show that there is a weak correlation between NO2 pollutant and the quantity of combustion vehicles acording to the correlation matrix, a weak linear aproximation could be possible just to see how the amount of combustion vehicles if affecting the quantity of NO2 pollutant in the air. This correlation between this variables is positive, which means that the more quantity of combustion vehicles, the more of NO2 pollutant in the air. The case of PM10 is similar to the case of NO2.
2. The analysis show that there is no correlation between PM2.5 pollutant and the quantity of combustion vehicles acording to the correlation matrix, the slope of the linear aproximation is almost zero, which means there is no correlation between variables.

Thank you!