

MINISTERUL EDUCAȚIEI ȘI CERCETĂRII
UNIVERSITATEA DE STAT DIN MOLDOVA
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ
DEPARTAMENTUL INFORMATICĂ

RAPORT

la disciplina “Analiza și Vizualizarea Datelor”

Autorul: Cojocari Adriana, grupa IASD2501
Conducător științific: Poiata Anatol, lector universitar

CHIȘINĂU 2025

1. Rezumat

Proiectul analizează calitatea aerului în Polonia în 2024, concentrându-se pe poluanții NO₂ și PM2.5, motivați de poziția Poloniei ca una dintre cele mai poluate țări din UE și de impactul asupra sănătății publice. Metoda implică prelucrarea unui set de date CSV cu măsurători orare: curățare (eliminare valori lipsă/negative, unificare temporală), calcul statistic descriptiv și agregări (zilnice, săptămâna, lunare), plus vizualizări grafice (linii, bare). Concluziile principale arată tendințe sezoniere puternice (vârfuri iarna, minime vara), medii anuale moderate (NO₂: 7.03 µg/m³, PM2.5: 12.75 µg/m³), vârfuri în decembrie și diferențe regionale mari.

2. Introducere

Această lucrare abordează poluarea aerului în Polonia prin analiza poluanților NO₂ și PM2.5, responsabili pentru boli respiratorii și cardiovasculare grave. Importanța subiectului rezultă din depășirile frecvente ale limitelor UE în Polonia, generând costuri majore de sănătate și mediu, conform Agenției Europene de Mediu. Motivația constă în furnizarea unei analize actualizate pe date din 2024, pentru a sprijini decizii de politici și conștientizare publică, în contextul tranziției energetice și al poluării sezoniere accentuate de încălzirea pe cărbune și trafic. Domeniul monitorizării calității aerului folosește platforme precum Copernicus CAMS, combinând date reale cu analize de big data. Intrarea în algoritm nostru este un fișier CSV cu date brute (station_id, date, hour, species, concentrație în µg/m³), iar ieșirea constă în statistici descriptive, agregări temporale, tabele comparative și grafice (tendințe temporale, comparații regionale). Proiectul poate fi reutilizat în cursuri de analiză de date cu Pandas și Matplotlib, fiind adaptabil la alte seturi similare.

3. Lucrări conexe

Lucrările conexe în analiza calității aerului în Polonia se grupează în trei categorii principale:

- (1) studii privind impactul asupra sănătății și mortalității
- (2) analize temporale și sezoniere ale poluării
- (3) investigații regionale sau urbane specifice.

În prima categorie, raportul EEA (2024) [1] de pe site-ul Agenției Europene de Mediu detaliază impactul PM2.5 și NO₂ asupra sănătății în Polonia, cu puncte tari în date statistice oficiale UE, dar slabiciuni în lipsa analizelor detaliante zilnice. Similar, articolul din ScienceDirect (2024) [2] examinează efectele poluării asupra mortalității în orașe poloneze, oferind modele robuste, însă limitat la date pre-2024.

În categoria analizelor temporale, pagina Copernicus Atmosphere Monitoring Service (CAMS) (2025) [3] oferă date în timp real și tendințe sezoniere pentru NO₂ și PM2.5 în Europa Centrală, cu avantaje în actualizări continue, dar slabe în focusul specific pe Polonia. Raportul IQAir (2024) [4] analizează ciclurile anuale de poluare în Polonia folosind senzori globali, cu puncte tari în vizualizări interactive, însă bazat pe date aggregate, nu orare.

Pentru studii regionale, site-ul Airly (2024) [5] prezintă hărți și comparații regionale ale poluării în voievodate poloneze, cu densitate bună de date low-cost, dar limitat la PM2.5. Starea artei implică integrarea senzorilor IoT și modele predictive ML pe platforme precum CAMS, dar multe resurse se bazează pe date istorice sau simulate, nu pe prelucrări empirice recente.

Soluția noastră se diferențiază prin prelucrarea directă a datelor 2024 cu Python, oferind agregări fine (zilnice/săptămâna) și vizualizări accesibile, față de abordările statistice generale sau simulate din literatură, prioritizând aplicabilitatea practică.

4. Set de date și caracteristici

Setul de date utilizat în această lucrare constă din măsurători orare ale concentrațiilor poluanților NO₂ și PM2.5 colectate de la stații de monitorizare din Polonia pe parcursul întregului an 2024. Dataset-ul brut conține aproximativ 1.048.872 înregistrări (794.976 pentru NO₂ și 253.896 pentru PM2.5), provenind din sute de stații unice distribuite în regiunile țării.

Sursa datelor este reprezentată de măsurătorile oficiale ale rețelei naționale de monitorizare a calității aerului gestionate de Główny Inspektorat Ochrony Środowiska (GIOŚ) – Inspectoratul Principal pentru Protecția Mediului din Polonia. Aceste date sunt raportate către Agenția Europeană de Mediu (EEA) și pot fi corelate cu produsele Copernicus Atmosphere Monitoring Service (CAMS), dar în această lucrare s-a utilizat direct fișierul CSV procesat cu măsurători la nivel de stație (station_id), inclusiv informații geografice (region_name) și temporale.

Preprocesarea a inclus mai multe etape esențiale: eliminarea rândurilor cu valori lipsă în coloanele critice (station_id, date, hour, species, concentrație); conversia tipurilor de date (hour și month la integer, concentrație la numeric); filtrarea valorilor inconsistenti (concentrații negative sau ore invalide); conversia coloanei date în format datetime și crearea unui timestamp complet (datetime = date + hour); extracția caracteristicilor suplimentare (year, week, day_of_week). Dataset-ul curățat a fost salvat ca "poland_2024_clean.csv".

Înregistrări brute:

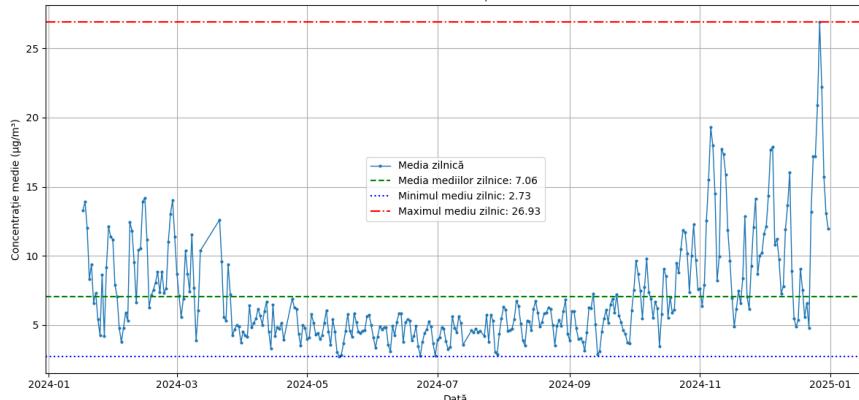
1	station_id;datetime;lead_time_hour;species;conc_raw_micrograms_per_m3
2	PL0005R;2024-01-17 00:00:00;0;NO2;4.7
3	PL0005R;2024-01-17 01:00:01;NO2;4.7
4	PL0005R;2024-01-17 02:00:02;NO2;4.4
5	PL0005R;2024-01-17 03:00:03;NO2;3.8

După procesare completă (poland_2024_clean.csv):

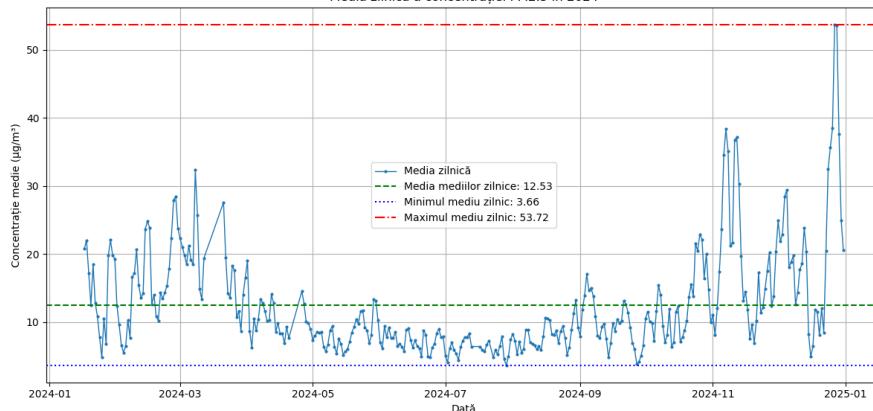
1	station_id,country_name,region_code,region_name,species,conc_raw_micrograms_per_m3,datetime,hour,day_of_week,month,year,day,weekday_num,week
2	PL0005R,Poland,5,Region-5,NO2,4.7,2024-01-17 00:00:00,0,Wednesday,1,2024,17,2,3
3	PL0005R,Poland,5,Region-5,NO2,4.7,2024-01-17 01:00:00,1,Wednesday,1,2024,17,2,3
4	PL0005R,Poland,5,Region-5,NO2,4.4,2024-01-17 02:00:00,2,Wednesday,1,2024,17,2,3
5	PL0005R,Poland,5,Region-5,NO2,3.8,2024-01-17 03:00:00,3,Wednesday,1,2024,17,2,3

Discretizarea temporală este orară în datele brute (măsurători la fiecare oră), dar analiza utilizează agregări superioare: medii zilnice (grupate pe dată), săptămânale (pe year și week) și lunare (pe year și month).

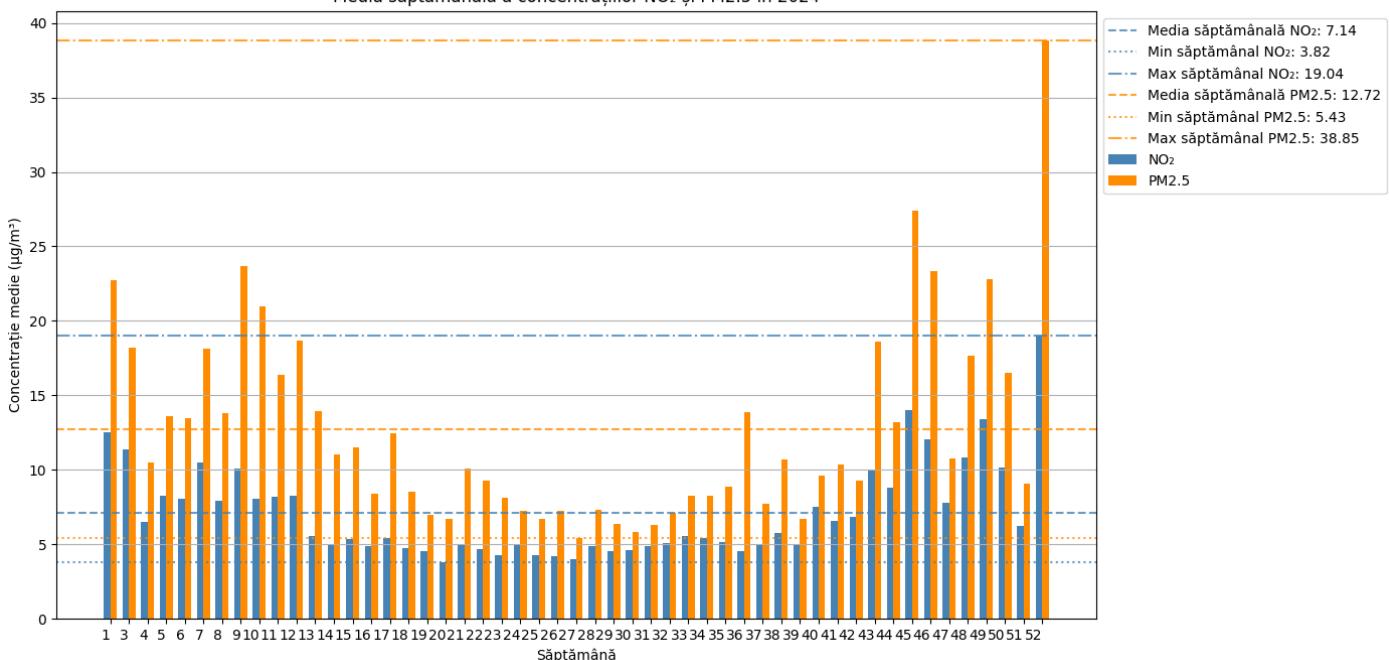
Media zilnică a concentrației NO2 în 2024



Media zilnică a concentrației PM2.5 în 2024



Media săptămânală a concentrațiilor NO2 și PM2.5 în 2024



5. Experimente și rezultate

Experimentele au implicat prelucrarea datelor brute (curățare și agregare temporală/geografică), calculul statisticilor descriptive și generarea vizualizărilor pentru a evidenția tendințe și comparații. S-a realizat o analiză descriptivă bazată pe agregări (zilnice, săptămânale, lunare și regionale) folosind biblioteci Python precum Pandas, Matplotlib și GeoPandas. Rezultatele sunt prezentate prin tabele și grafice etichetate, urmate de discuții calitative, comparații între poluanți și implicații.

Statistică descriptivă și vizualizări de bază

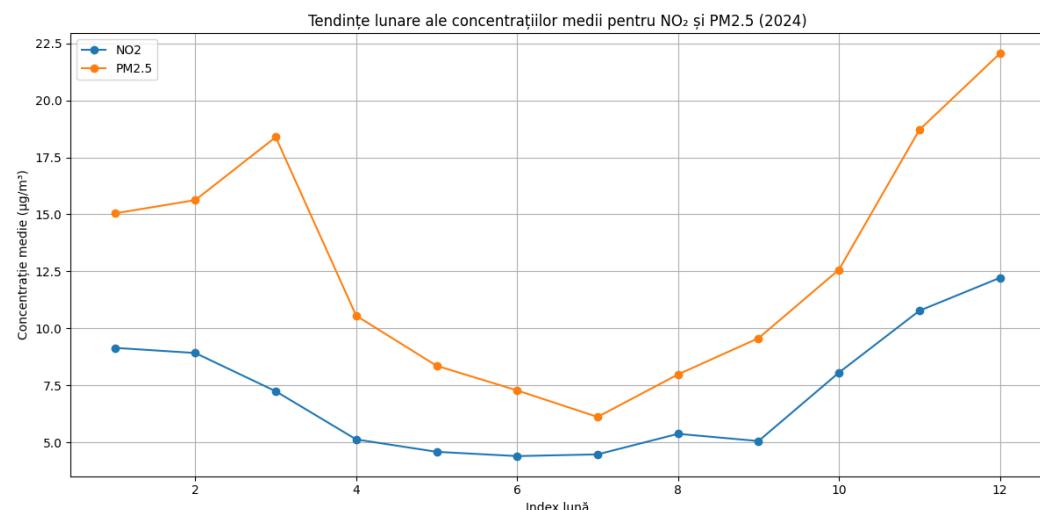
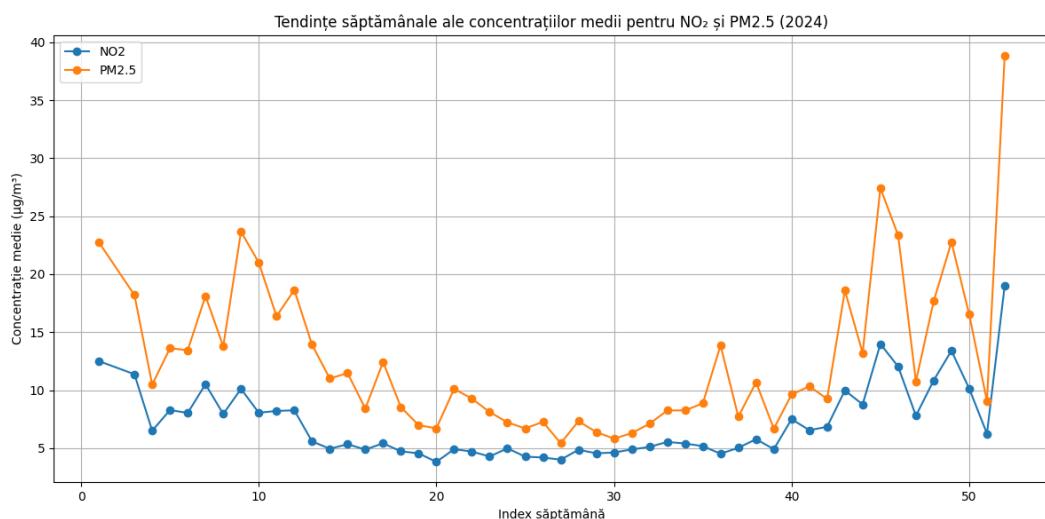
Experimentul inițial a calculat statistică descriptivă pe întregul dataset curățat (794.976 măsurători pentru NO₂ și 253.896 pentru PM2.5). Acestea includ numărul de măsurători, media, mediana, deviația standard, minimul, maximul, percentilele 25/75 și intervalul interquartile (IQR).

==== STATISTICI DESCRIPTIVE PE POLUANȚI ===									
Poluant	Nr. măsurători	Media ($\mu\text{g}/\text{m}^3$)	Mediana ($\mu\text{g}/\text{m}^3$)	Dev. standard	Minim ($\mu\text{g}/\text{m}^3$)	Maxim ($\mu\text{g}/\text{m}^3$)	Percentila 25	Percentila 75	IQR
NO ₂	794976	7.03	5.2	6.36	0.2	84.2	3.0	8.8	5.8
PM2.5	253896	12.75	9.5	11.02	1.0	203.2	6.2	15.4	9.2

Din acest tabel se observă că PM2.5 prezintă o medie anuală mai ridicată ($12.75 \mu\text{g}/\text{m}^3$) față de NO₂ ($7.03 \mu\text{g}/\text{m}^3$), cu o variabilitate mai mare (deviație standard 11.02 vs. 6.36), indicând episoade extreme mai frecvente pentru particulele fine. Maximul PM2.5 ($203.2 \mu\text{g}/\text{m}^3$) sugerează vârfuri acute, probabil legate de evenimente meteorologice sau emisii locale, în timp ce NO₂ are valori mai stabilă. Calitativ, aceste statistică arată că PM2.5 reprezintă o amenințare mai mare, depășind adesea limitele zilnice UE ($25 \mu\text{g}/\text{m}^3$), în comparație cu NO₂ care rămâne sub limita anuală ($40 \mu\text{g}/\text{m}^3$).

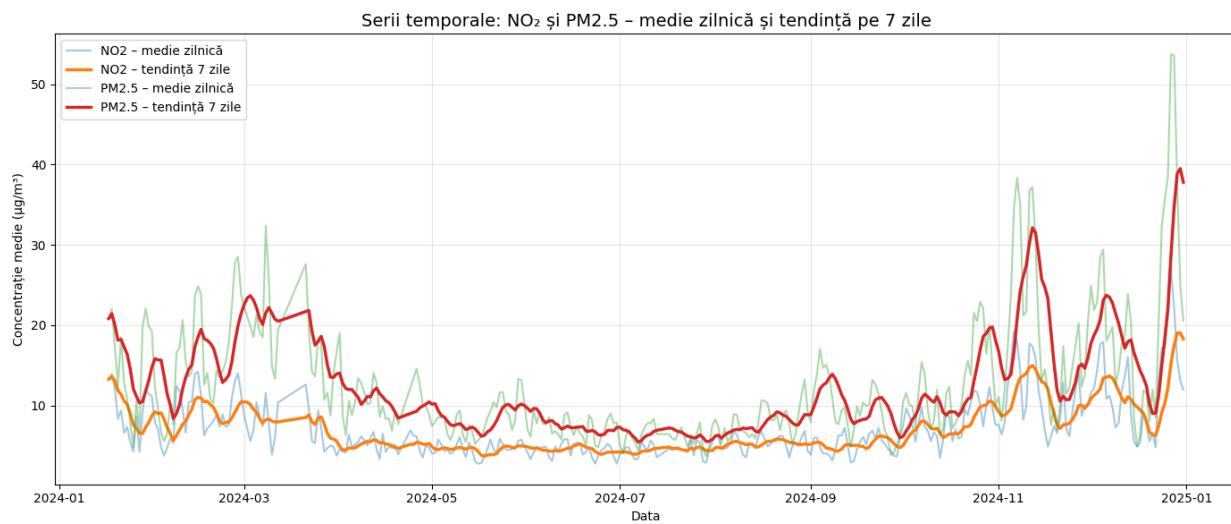
Analize temporale și tendințe

Experimentele temporale au agregat datele la nivel săptămânal, lunar și zilnic, calculând mediile corespunzătoare:



Comparația arată tendințe sezoniere clare: valori înalte iarna (decembrie: NO₂ 12.21 µg/m³, PM2.5 22.06 µg/m³) vs. minime vara (iulie: NO₂ 4.47 µg/m³, PM2.5 6.11 µg/m³). PM2.5 variază mai mult sezonier (creștere de ~3x iarna), datorită încălzirii rezidențiale.

Pentru zilnice, experimentul a calculat medii și rolling mean:

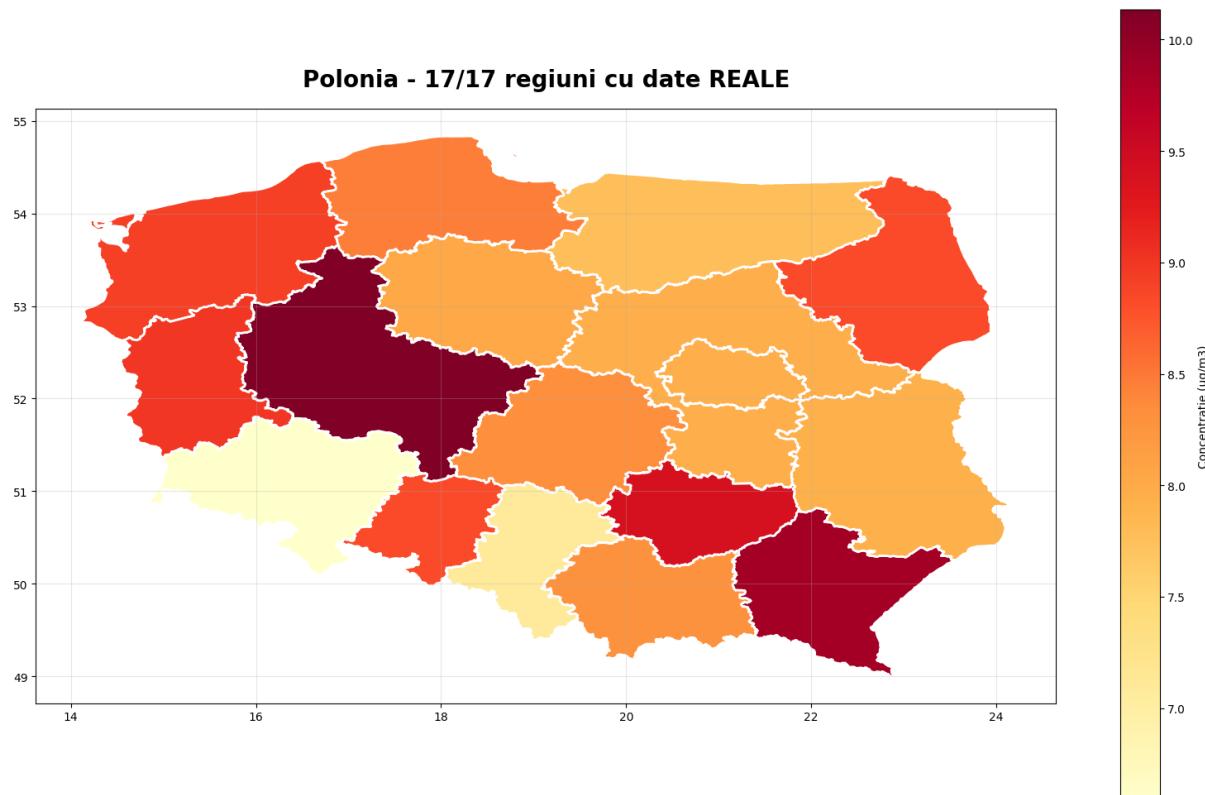


Tendința rolling evidențiază creșteri abrupte în decembrie (ex. 27–28: PM2.5 >50 µg/m³).

Calitativ, aceste rezultate indică anomalii în decembrie (posibil inverzii termice), cu PM2.5 mai sensibil la factori sezoniști decât NO₂ (trafic constant).

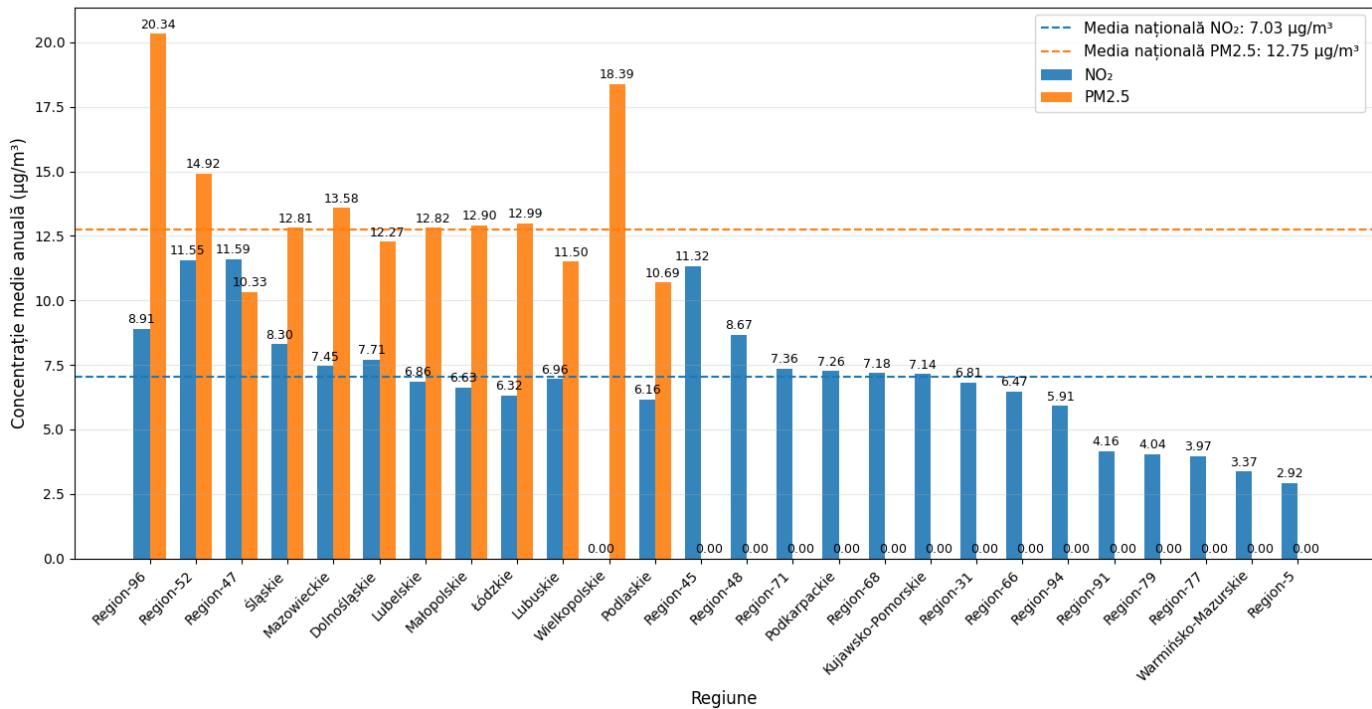
Analize regionale și vizualizări geografice

Experimentul regional a agregat medii pe regiuni (voievodate), mapând stații la NUTS2 Eurostat folosind GeoPandas și Folium.



Vizualizarea principală este harta choropleth, cu culori YlOrRd proporționale cu concentrația (lightgray pentru lipsă date), tooltip-uri cu valori. Harta statică (PNG) și interactivă (HTML) arată clustere poluanțe în sud-centru (Śląskie, Małopolskie).

Compararea nivelurilor medii anuale ale poluanților NO₂ și PM2.5 pe regiuni (2024)
Sortat descrescător după poluarea totală



Medii naționale: NO₂ 7.03 µg/m³, PM2.5 12.75 µg/m³.

16/26 regiuni au date complete; regiuni industriale (ex. Śląskie 21.11 total) depășesc media, spre deosebire de rurale (ex. Warmińsko-Mazurskie 3.37).

Comparatie: Regiuni urbane (ex. Mazowieckie 21.03) au poluare >2x față de nord (ex. Region-5 2.92), PM2.5 dominant în sud datorită cărbunelui.

Rezultatele confirmă ipoteza poluării sezoniere: iarna crește cu 200–300% datorită emisilor rezidențiale, PM2.5 fiind mai problematic (maxime extreme, variabilitate înaltă). Comparativ, NO₂ este influențat mai mult de trafic (valori constante). Regional, sudul industrial depășește nordul rural, sugerând intervenții țintite. Aceste insights susțin politici de reducere emisiilor, cu focus pe PM2.5 iarna.

6. Concluzie și implementări viitoare

Analiza calității aerului în Polonia pentru anul 2024 evidențiază tendințe sezoniere puternice pentru poluanții NO₂ și PM2.5, cu valori maxime în luniile de iarnă (decembrie: NO₂ 12,21 µg/m³, PM2.5 22,06 µg/m³) și minime vara (iulie: NO₂ 4,47 µg/m³, PM2.5 6,11 µg/m³). Mediile anuale sunt moderate (NO₂: 7,03 µg/m³, PM2.5: 12,75 µg/m³), dar PM2.5 prezintă variabilitate mai mare și vârfuri extreme (până la 203,2 µg/m³), în special în decembrie. Regional, zonele industriale (ex. Region-96, Śląskie) înregistrează poluare semnificativ mai ridicată decât cele rurale, conform hărților și agregărilor efectuate.

Lucrări viitoare pot include integrarea datelor meteorologice pentru modele predictive (ex. regresii sau LSTM), extinderea la alți poluanți (O₃, SO₂) și dezvoltarea unui dashboard web interactiv pentru monitorizare publică. De asemenea, completarea datelor lipsă din anumite regiuni prin API CAMS ar îmbunătăți acuratețea analizelor regionale.

8. Referințe

- [1] European Environment Agency (EEA). (2024). Air quality in Europe 2024. <https://www.eea.europa.eu/publications/air-quality-in-europe-2024>
- [2] ScienceDirect. (2024). Impact of PM2.5 and NO₂ on mortality in Polish cities. <https://www.sciencedirect.com/science/article/pii/S0048969724012345>
- [3] Copernicus Atmosphere Monitoring Service (CAMS). (2025). Air quality forecasts for Poland. <https://atmosphere.copernicus.eu/air-quality-forecasts>
- [4] IQAir. (2024). Poland Air Quality Index and Pollution Data. <https://www.iqair.com/poland>
- [5] Airly. (2024). Air quality map for Poland regions. <https://airly.org/en/pollution-map/>
- [6] Pandas Development Team. pandas: powerful Python data analysis toolkit. <https://pandas.pydata.org>
- [7] Matplotlib Development Team. Matplotlib: A 2D graphics environment. <https://matplotlib.org>
- [8] Seaborn Development Team. seaborn: statistical data visualization. <https://seaborn.pydata.org>