Eksploracja danych 2022/2023

Przewidywanie poziomu zanieczyszczenia powietrza w Krakowie i okolicach na podstawie danych z czujników Airly i sytuacji pogodowej

Milena Biernacka

Paweł Świder

Szymon Musiał

1. Wstęp **2**
Źródła 2
2. Minutki **3**
3. Proces pobierania danych **4**
4. Próba analizy na przestrzeni lat **9**
Wstępna analiza danych z pojedynczego czujnika w roku 2021 16
5. Dane **19**
6. Analizy **20**
6.1. Poziomy zanieczyszczeń, wiatr, temperatura 20
6.2. Korelacje 27
6.3. Wykresy skrzypcowe pyłów PM1, PM2.5 i PM10 31
6.4. Wiatr 33
6.5. Średnie zanieczyszczenie w podziale na Kraków i okolice 35
6.6. Mapy średniego miesięcznego zanieczyszczenia według stacji 37
6.7. Średnie zanieczyszczenie w ciągu doby i tygodnia 39
7. Predykcja **41**
7.1 Opis modelu 41
7.2 Wyniki predykcji 41
7.3 Analiza błędu predykcji 44
7.4 Porównanie błędów z Airly 47
8. Podsumowanie **49**
9. Spis rysunków i tabel **50**

1. Wstęp

Celem niniejszego projektu jest analiza jakości powietrza w Krakowie i pobliskich miejscowościach na różnych skalach czasowych, w celu zrozumienia zmian zanieczyszczenia w ciągu dnia, tygodnia czy na przestrzeni kilku miesięcy.

W ramach projektu, dane dotyczące jakości powietrza, zebrane z różnych stacji pomiarowych w gminach i okolicach Krakowa, będą analizowane. Będą one poddane statystycznej analizie oraz wizualizacji w celu zidentyfikowania trendów i wzorców zmian zanieczyszczenia w analizowanych obszarach. Będzie to obejmować analizę zmian w stężeniu pyłów zawieszonych PM10, PM2.5 i PM10 oraz innych szkodliwych substancji.

Projekt będzie badał potencjalne wpływy czynników zewnętrznych, takich jak warunki pogodowe, na jakość powietrza w tych obszarach. Analiza zostanie przeprowadzona w kontekście obwarzanka, czyli obszarów gminnych i okolic Krakowa. Będzie to obejmować analizę porównawczą poziomów zanieczyszczeń w zależności od warunków pogodowych, takich jak temperatura czy kierunek wiatru.

Dodatkowo jednym z głównych celów tego projektu jest opracowanie modelu predykcyjnego poziomu zanieczyszczeń dla gmin i okolic Krakowa w różnych horyzontach czasowych. Model ten będzie oparty na analizie historycznych danych dotyczących jakości powietrza, warunków pogodowych oraz innych istotnych czynników. Celem budowy tego modelu jest umożliwienie prognozowania poziomu zanieczyszczeń na przyszłe okresy i weryfikacja jego dokładności w porównaniu do prognoz udostępnianych przez serwis Airly.

Źródła

Dane zanieczyszczeń oraz prognozy pogodowe są pobierane z wybranych stacji w centrum Krakowa, oraz jego aglomeracji za okres od marca do czerwca 2023 roku z wykorzystaniem API serwisu Airly https://developer.airly.org/en/api . Pomiary obejmują informacje dotyczące dnia poprzedniego oraz prognoz na kolejny dzień. Ponadto wykorzystano również prognozy z serwisu wettercom https://rapidapi.com/wettercom-wettercom-default/api/forecast9/

Dodatkowo podjęto próbę analizy danych o zanieczyszczeniach na przestrzeni lat z analogicznego obszaru z serwisu https://powietrze.gios.gov.pl w połączeniu z danymi meteorologicznymi z serwisu https://danepubliczne.imgw.pl

2. Minutki

Repozytorium: https://github.com/szymon-musial/data-mining

Minutka 1 - 06.03.2023 - wziąć po 2 stacje z każdego kierunku świata (np. Skała, Wielka Wieś, Skawina - 8 stacji będzie ok). W tabelce dodać kierunki świata. Podpisać rysunek (i kółka). Korzystamy jednak ze strony rządowej (możemy zrezygnować z Airly). Pogodowe dane powinny być z punktów pomiarowych (nie z samego Krakowa). Krótki horyzont czasowy. Predykcje na danych bieżących, a trendy itp. na historycznych.
Na następne zajęcia:

- zebrać dane
- zaproponować bazę danych
- patrzymy na rozklady w trybie ciągłym

Minutka 2 - 20.03.2023 :

- pobierać bieżące dane z Airly i je analizować

Minutka 3 - 03.04.2023r rozkłady temperatury, kierunki, prędkość wiatru. Rozkłady porobić, korelacja np. wiatru z zanieczyszczeniem. Skąd najczęściej wieje wiatr.

Minutka 4 - 17.04.2023 - zweryfikować skoki zanieczyszczeń, wnioski na nowych danych, porównanie modelu z predykcjami z Airly, powtórzyć na uzupełnionych danych, można spróbować skrócić czas predykcji

Minutka 5 - 09.05.2023 - predykcja z krótką historią, obecną zostawić, żeby mieć porównanie. “Naprawić wiatr”. Przy zmianach wiatru współczynnik dodatkowy.

Minutka 6 - 22.05.2023 - komentarze, analiza.

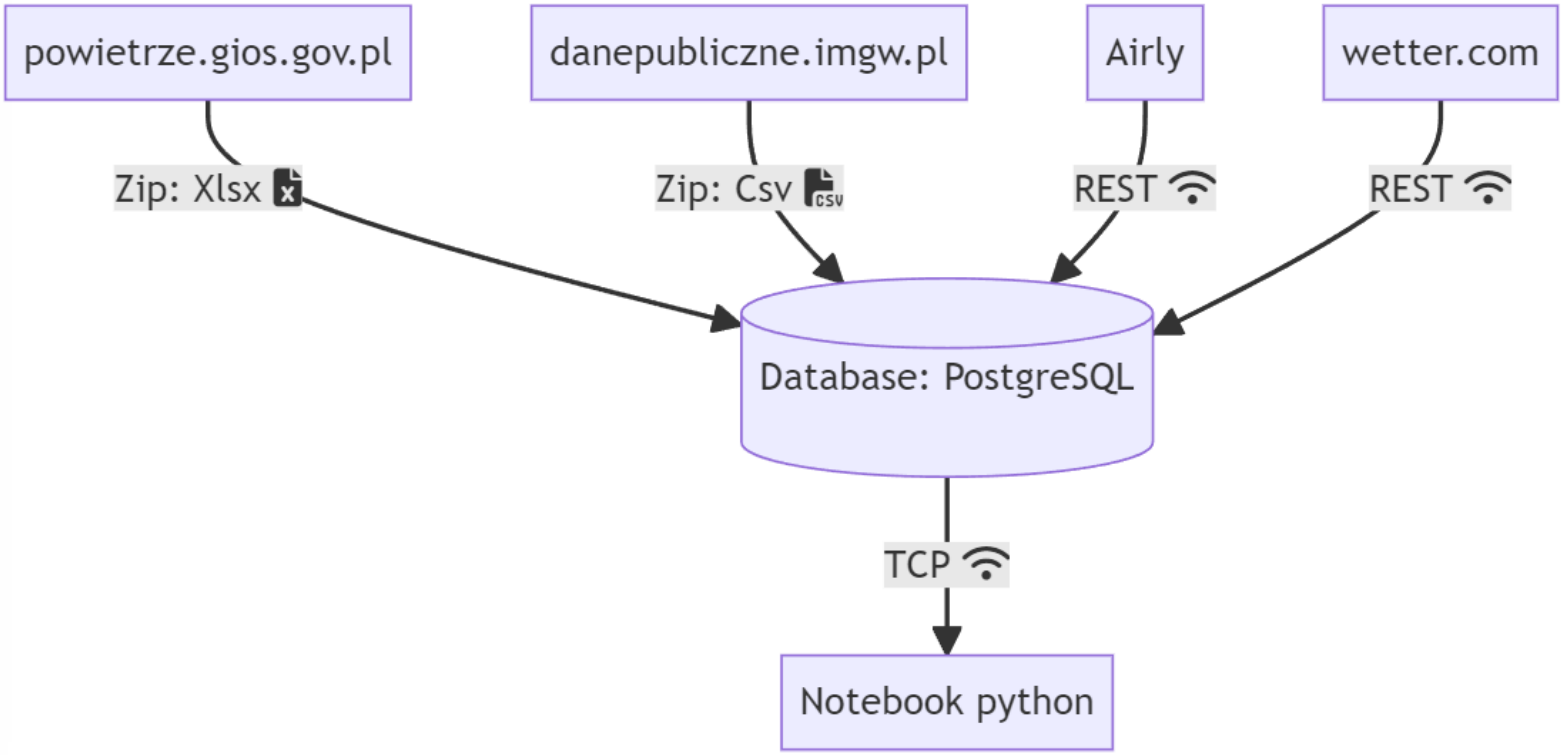
Minutka 7 - 12.06.2023 - numery, tytuły, cytacja wykresów/tabel itp., pozyskanie danych. podsumowanie, koncepcja, cel. Prezentacja na następne zajęcia. (ok 10 minut)

3. Proces pobierania danych

W projekcie wykorzystano dwa podejścia

- Przetwarzanie danych historycznych. Dane historyczne zostały zebrane z serwisów GIOS i IMGW,
- Przetwarzanie z API serwisu Airly. Pobieranie aktualnych danych i prognoz pogodowych.

Całość procesu, obejmującego zbieranie, przetwarzanie i analizę danych, została zobrazowana na schemacie w rysunku 3.1.



Rys 3.1. Architektura oprogramowania użyta do eksploracji

W przedstawionej architekturze, na rysunku 1.1, zidentyfikowana została warstwa danych, której reprezentantem jest baza danych PostgreSQL.Dane były dodawane do bazy danych asynchronicznie. Notebook Pythona komunikował się z serwerem bazy danych, pobierając przygotowane dane i wykorzystując je do analizy i przetwarzania.

Serwisy procesujące dane zostały napisane w języku C# z wykorzystaniem .Net 7.0. Przedstawiono ogólny szkielet programów w poniższym fragmencie 1.

```
using airly_data_fetch;
using Microsoft.Extensions.Configuration;
Console.WriteLine(DateTime.Now);
var builder = new ConfigurationBuilder().AddUserSecrets<Program>();
var configurationRoot = builder.Build();
var connectionString = configurationRoot.GetSection("connectionString").Value;
var apiKey = configurationRoot.GetSection("apiKey").Value;
if (connectionString is null) {
	connectionString = Environment.GetEnvironmentVariable("connectionString");
}
if (apiKey is null) {
	apiKey = Environment.GetEnvironmentVariable("apiKey");
}
ArgumentNullException.ThrowIfNull(apiKey);
if (connectionString is not null) {
	AppDbContext.connectionString = connectionString;
}
var appDbContext = new AppDbContext();
var arilyApiRepository = new ArilyApiRepository(apiKey);
var appRepository = new AppRepostioty(appDbContext);
foreach (var friendlyStationName in Urls.GetFriendlyStationName()) {
	Console.WriteLine($"Processing ... {friendlyStationName.Value}");
	var request = await arilyApiRepository.GetMeasurementsForInstallationId(friendlyStationName);
	if(request is null)  {
	continue;
	}
	var responseModel = await arilyApiRepository.ProcessResponse(request);
	appRepository.InsertEntity(responseModel, friendlyStationName);
	Console.WriteLine($"Finished {friendlyStationName.Value}");
}
```

Fragment 1. Fragment kodu punktu wejścia programu pobierającego dane z api Airly

Stworzono cztery programy obsługujące pięć różnych źródeł danych, które są zaimplementowane w analogiczny sposób jak przedstawiony fragment kodu 1.

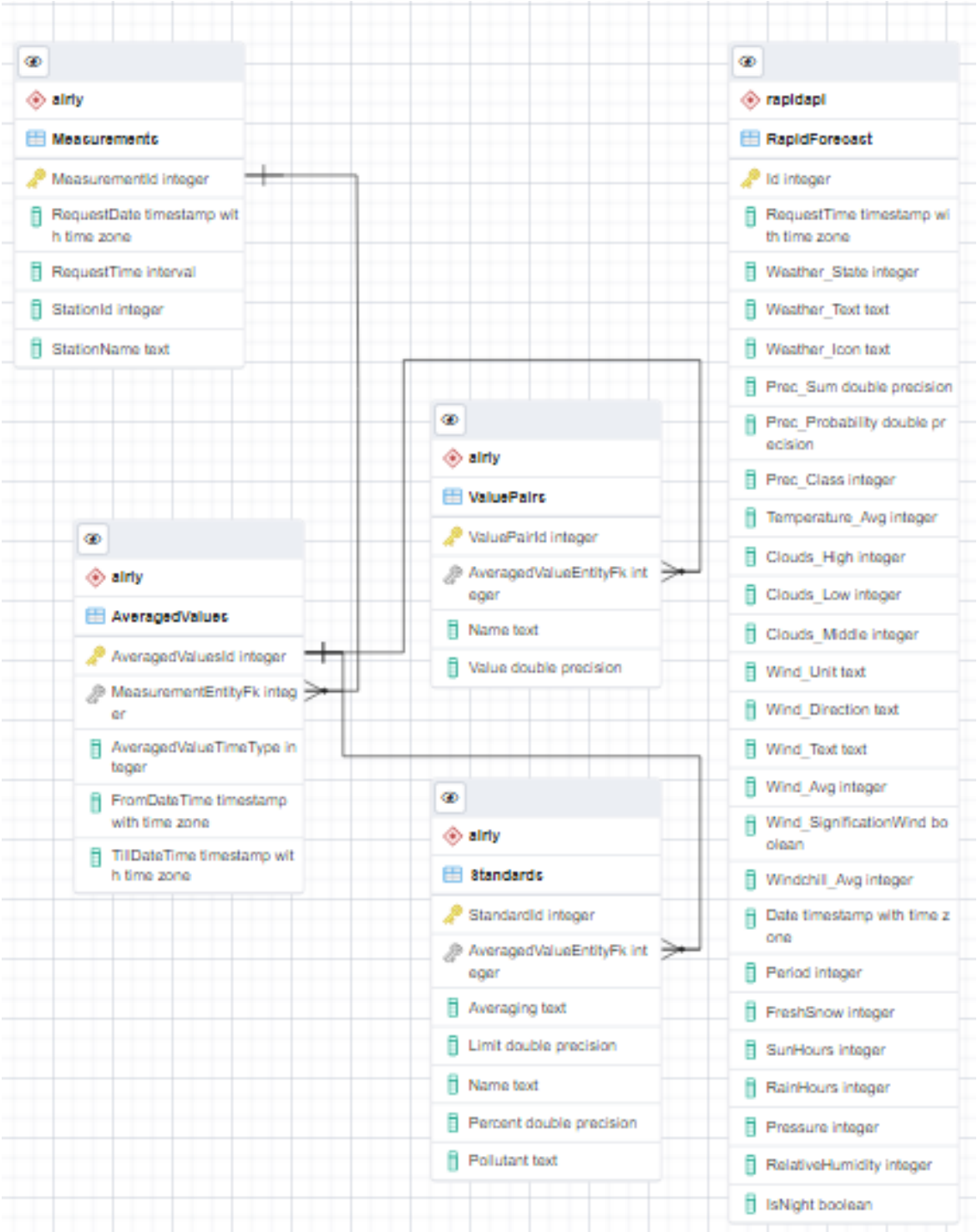
Na początku każdego programu następuje odczytanie konfiguracji z pliku JSON podczas procesu rozwoju aplikacji lub zmiennych środowiskowych w przypadku wdrożenia aplikacji na własny klaster Kubernetes.

Następnie, korzystając z biblioteki ORM Entity Framework, tworzona jest instancja kontekstu bazy danych, która jest przekazywana do dedykowanego repozytorium dla każdego programu.

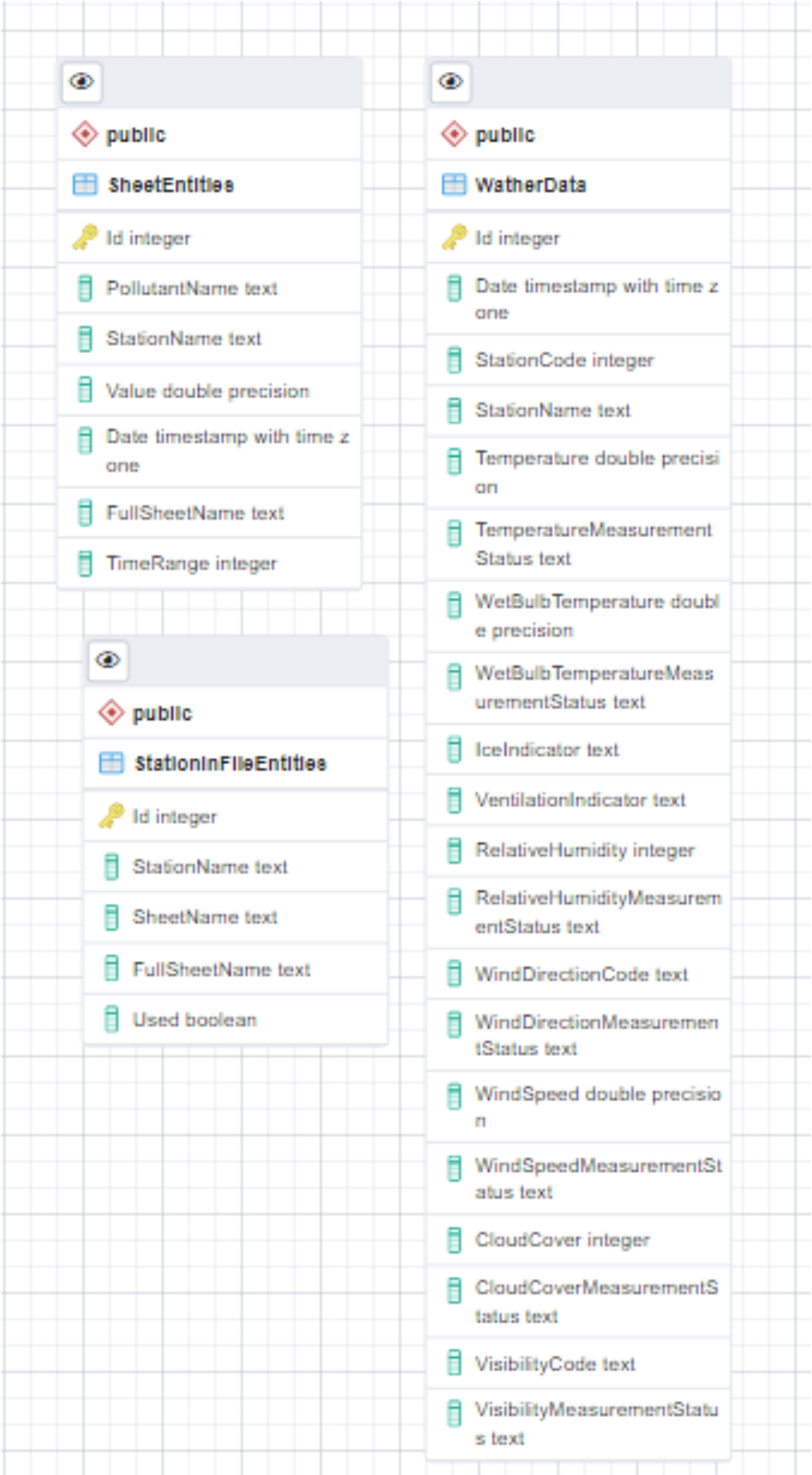
Kolejno, w zależności od źródła danych, tworzone są odpowiednie repozytoria, które przetwarzają dane wejściowe:

- Dla źródła powietrze.gios.gov.pl, obsługiwane są pliki ZIP i XLSX.
- Dla źródła danepubliczne.imgw.pl, obsługiwane są pliki ZIP i CSV.
- Dla źródła airly, wykorzystywane są zapytania REST.
- Dla źródła wetter.com za pośrednictwem RapidAPI, również wykorzystywane są zapytania REST.

Następnie, w pętli, przetworzone rekordy są dodawane do bazy danych, która została stworzona przy użyciu podejścia "Code First" w Entity Framework. Migracje schematu bazy danych są automatycznie przeprowadzane, a struktura bazy danych jest prezentowana na rysunkach 3.2 i 3.3.



Rys 3.2. Schemat ERD wykorzystanej bazy



Rys 3.3. Schemat ERD porzuconych danych wykorzystanej bazy

Serwisy obsługujące przetwarzanie plików zostały uruchomione jednorazowo, natomiast programy odpowiedzialne za pobieranie danych online za pośrednictwem interfejsu REST zostały wdrożone na klaster Kubernetes.

W klastrze Kubernetes została skonfigurowana usługa typu CRON, która planuje i uruchamia te programy w określonym harmonogramie czasowym. W przypadku tego projektu, programy są uruchamiane dwukrotnie w ciągu dnia. Jest to związane z limitem wywołań narzuconym przez serwisy, z których pobierane są dane, szczególnie Airly.

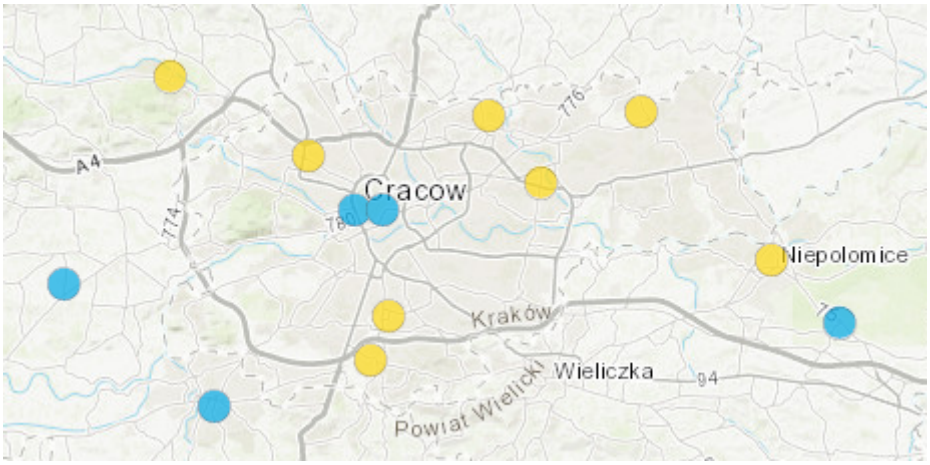
Dodatkowo, na klastrze Kubernetes znajduje się również baza danych, która służy do przechowywania zgromadzonych danych. Dzięki temu dane są trwale przechowywane i dostępne dla innych usług i aplikacji w klastrze.

4. Próba analizy na przestrzeni lat

Zgodnie z modelem iteracyjnym CRISP-DM, podjęto próby analizy zanieczyszczeń powietrza na przestrzeni lat 2000-2023, z uwzględnieniem trzech miesięcy przed aktualną datą.

Do analizy wykorzystano serwisy powietrze.gios.gov.pl, które dostarczały dane dotyczące zanieczyszczeń z kilku punktów pomiarowych w Krakowie, oraz serwisy danepubliczne.imgw.pl, które zawierały historyczne dane meteorologiczne.

W trakcie analizy zanieczyszczeń skupiono się na wybranych stacjach znajdujących się na rysunku 4.1. Wybór stacji był dokonywany w taki sposób, aby uwzględnić wpływ aglomeracji na centrum miasta i umożliwić analizę tego zjawiska.



Rys. 4.1. Wybrane stacje pokazane kolorem żółtym oraz niebieskim do analizy z serwisu powietrze.gios.gov.pl

Wybrane stacje charakteryzowały się różnym zakresem dostępnych mierzonych zanieczyszczeń. Nie wszystkie stacje posiadały pomiary dla wszystkich składników, takich jak PM10, O3 itp. Przedstawiono fragment dostępnych pomiarów w tabeli 4.1. Dane były pogrupowane w okna czasowe o szerokości 1 godziny i/lub 24 godzin, umożliwiając analizę dla różnych granulacji.

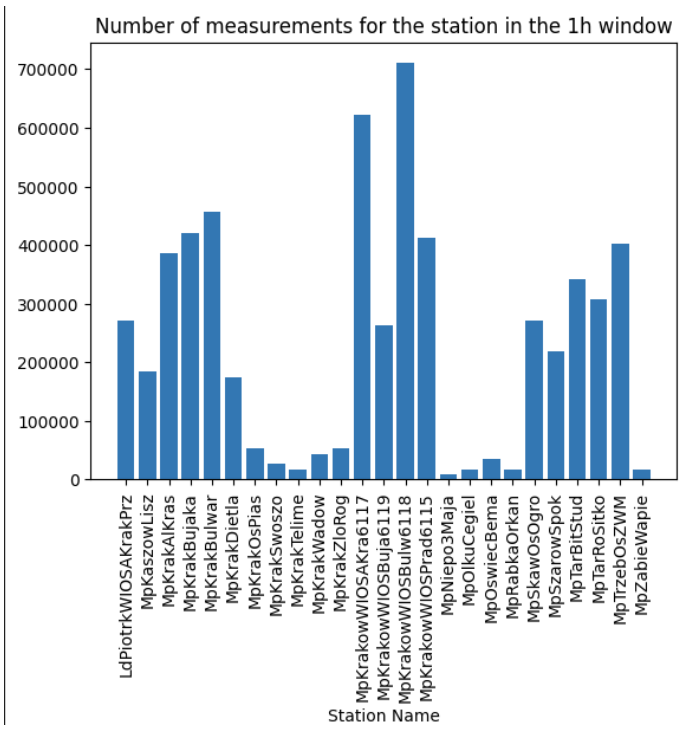
Tabela 4.1. Wybrane stacje do analizy z serwisu powietrze.gios.gov.pl wraz z dostępnymi pomiarami zanieczyszczenia

Nazwa stacji	Kod stacji/krajowy	Zanieczyszczenie
Kaszów , Bory	MpKaszowLisz	tlenek azotu dwutlenek azotu tlenki azotu ozon
Skawina , os. Ogrody	MpSkawOsOgro	tlenek azotu dwutlenek azotu tlenki azotu pył zawieszony PM10 dwutlenek siarki benzen
Zabierzów , Wapienna	MpZabieWapie	pył zawieszony PM10 benzo(a)piren w PM10
Kraków, ul. Złoty Róg	MpKrakZloRog	pył zawieszony PM10 benzo(a)piren w PM10
Kraków , Lusińska	MpKrakSwoszo	pył zawieszony PM10 benzo(a)piren w PM10
Kraków, ul. Bujaka	MpKrakBujaka	arsen w PM10 benzo(a)antracen w PM10 benzo(a)piren w PM10 benzo(b)fluoranten w PM10 kadm w PM10 dibenzo(a,h)antracen w PM10 indeno(1,2,3-cd)piren w PM10 nikiel w PM10 tlenek azotu dwutlenek azotu tlenki azotu ozon ołów w PM10 pył zawieszony PM10 pył zawieszony PM2.5 dwutlenek siarki benzo(j)fluoranten w PM10 benzen
Kraków , al. Krasińskiego	MpKrakAlKras	tlenek węgla tlenek azotu dwutlenek azotu tlenki azotu pył zawieszony PM10 pył zawieszony PM2.5 benzen

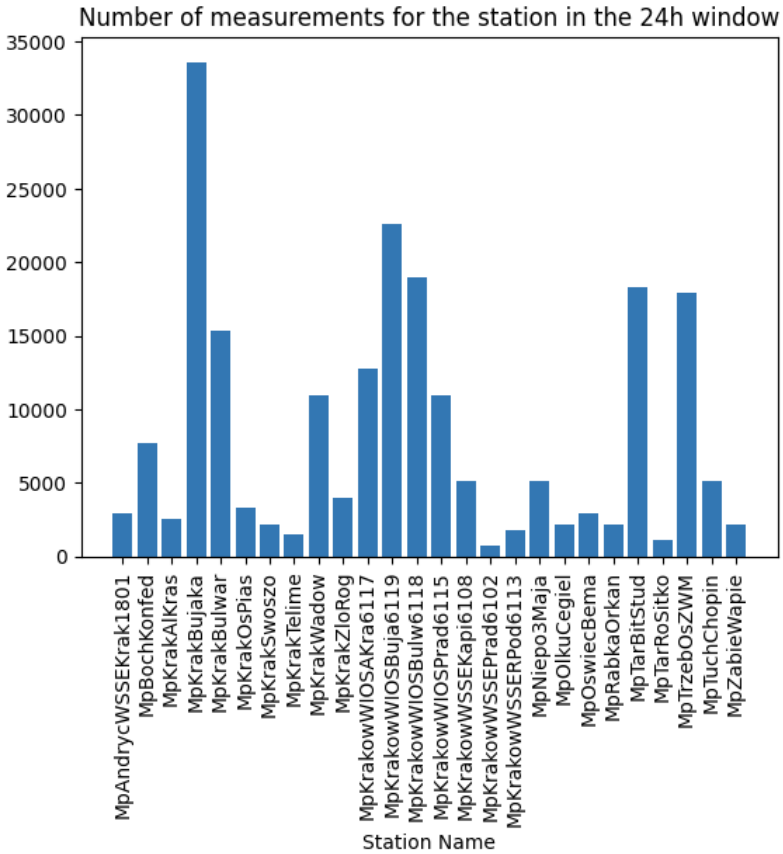
Niepołomice, ul. 3 Maja	MpNiepo3Maja	pył zawieszony PM10 benzo(a)piren w PM10
Szarów, ul. Spokojna	MpSzarowSpok	tlenek azotu dwutlenek azotu tlenki azotu ozon
Wadowice, ul. Bałysa	MpWadowiBalyMOB	pył zawieszony PM10
Oświęcim, ul. J. Bema	MpOswiecBema	pył zawieszony PM10 benzo(a)piren w PM10 benzen etylobenzen o-ksylen toluen
Olkusz, ul. Cegielniana	MpOlkuCegiel	pył zawieszony PM10 benzo(a)piren w PM10

Po przetworzeniu danych z archiwum i sprawdzeniu logów w celu upewnienia się, że nie wystąpiły błędy konwersji, podjęto próby uzupełnienia brakujących danych. Oprócz użycia średniej wartości, zdecydowano się także skorzystać z interfejsu API Airly w celu uzupełnienia danych. Jednakże, ze względu na różnice w stacjach pomiędzy obiema usługami, wyniki nie były zadowalające.

Dane z serwisu powietrze.gios.gov.pl zawierały uśrednione wartości dla różnych okien czasowych, jak pokazano na rysunkach 4.2 i 4.3. Do analizy wybrano najmniejsze dostępne okno czasowe, czyli 1 godzinę.



Rys. 4.2. Histogram dostępnych pomiarów dla stacji o granulacji 24h



Rys. 4.3. Histogram dostępnych pomiarów dla stacji o granulacji 1h

Serwis danepubliczne.imgw.pl dostarcza różnorodne dane meteorologiczne.Te dane są dostępne dla poszczególnych stacji pomiarowych w określonych punktach czasowych. Dostępne dane ze wspomnianego serwisu są przedstawione poniżej:

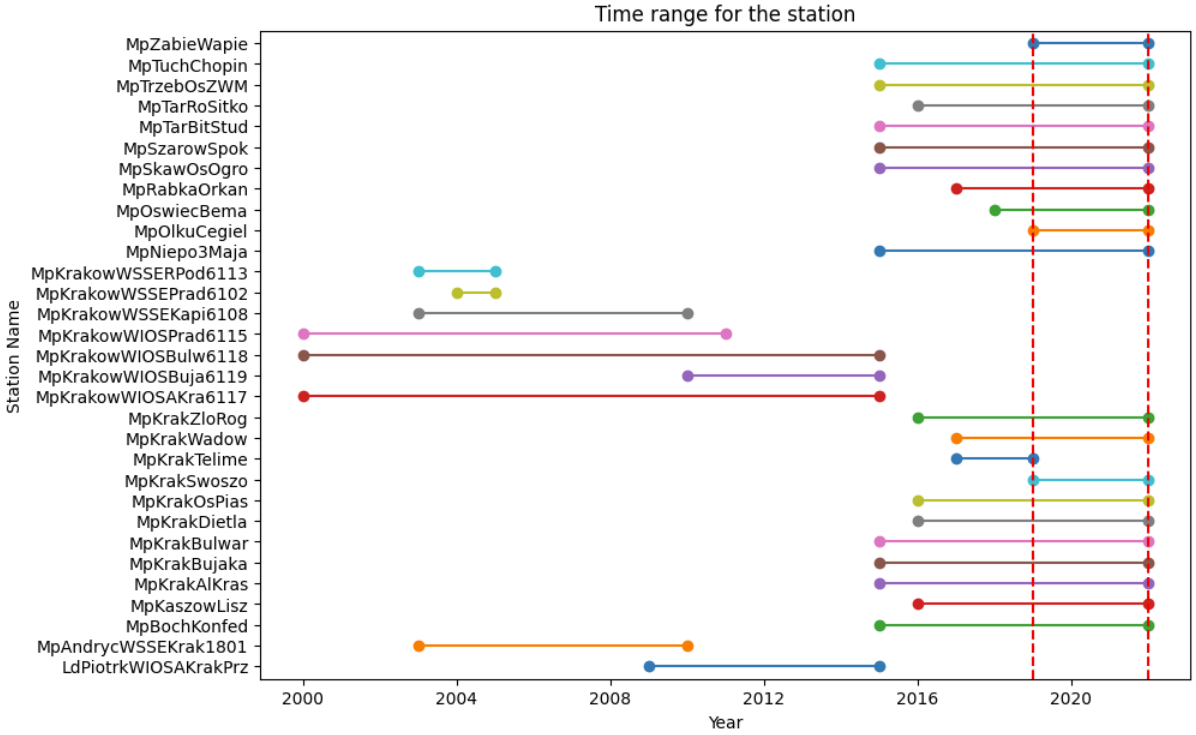
- Stacja
- Data i godzina
- Temperatura
- Wilgotność
- Wskaźnik lodu / wentylacji
- Kod kierunku wiatru oraz prędkość
- Zachmurzenie (oktany), widzialność

Ustawa antysmogowa, która weszła w życie 1 września 2019 roku w Krakowie, wprowadziła zakaz stosowania węgla i drewna w kotłach, piecach i kominkach. Jej celem jest ograniczenie emisji zanieczyszczeń powietrza, szczególnie w okresie sezonu grzewczego.

W kontekście analizy wpływu ustawy na stan zanieczyszczeń planowano ponadto wziąć pod uwagę dane dotyczące cen energii elektrycznej udostępnionych przez Urząd Regulacji Energetyki (URE).Te dane mogą być istotne w kontekście oceny ewentualnej zmiany preferencji energetycznych gospodarstw domowych po wprowadzeniu ustawy antysmogowej.

Podczas prób analizy danych zgromadzonych w ramach projektu, napotkano znaczące braki w szeregach czasowych, które miały negatywny wpływ na możliwość przeprowadzenia analizy. Wybrane stacje monitorujące zanieczyszczenia powietrza od 2000 roku często posiadały niepełne dane, co utrudniło porównanie i analizę wpływu zanieczyszczeń na przestrzeni lat.

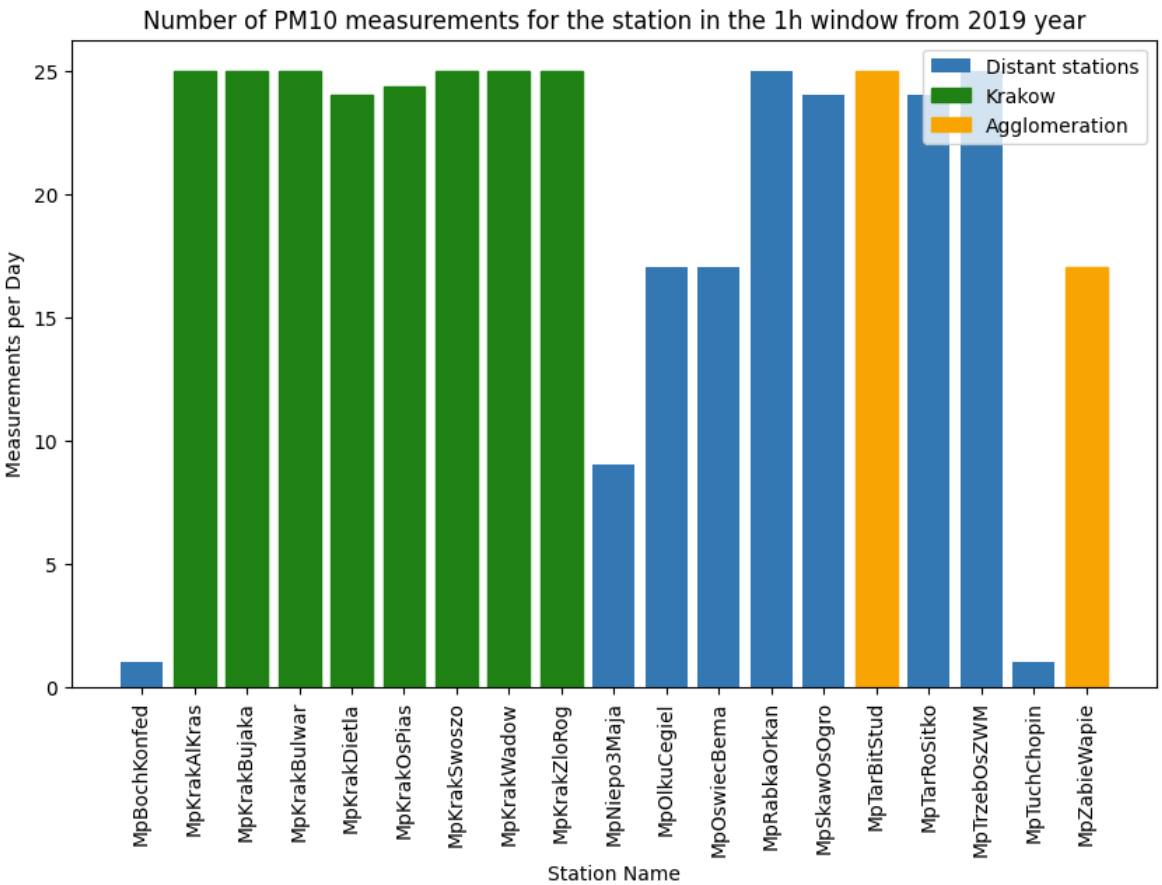
Rysunki 4.4 i 4.5 ukazują zakres brakujących danych w poszczególnych stacjach. Niedostateczna ilość danych uniemożliwia dokładne porównanie trendów zanieczyszczeń oraz wnioskowanie na temat ich zmian w czasie.



Rys 4.4. Okres, w którym stacje raportowały zanieczyszczenia

Na rysunku 4.4. przedstawiono wykres okresem czasu, w którym to poszczególne stacje monitorujące zanieczyszczenia raportowały wyniki. Na osi y znajdują się stacje z małopolski, natomiast na osi x przedstawione są lata. Wykres ten obrazuje, że występuje znaczna liczba braków danych szczególnie w poprzedniej dekadzie, które utrudniają zakładaną analizę.

Na rysunku 4.4. zaznaczono czerwonymi liniami zakres od 2019 roku do początku 2023, w którym to jest względnie dużo stacji. Podczas analizy dostępnych danych dla stacji dla wybranego zanieczyszczenia PM10, zbudowano histogram przedstawiony na rysunku 4.5 dla zaznaczonego okresu czasu.



Rys 4.5. Histogram ilości pomiarów dla stacji w okresie sty 2019 - sty 2023 dla zanieczyszczenia PM10

Stacje na rysunku 4.5 oznaczone kolorem zielonym znajdują się w Krakowie. Natomiast, stacje znajdujące się w aglomeracji zostały oznaczone kolorem żółtym, a pozostałe stacje z terenu Małopolski zostały oznaczone kolorem niebieskim.

Ten kolorowy podział umożliwia wizualne zobrazowanie rozmieszczenia stacji pomiarowych. Widać tutaj niedostateczną ilość reprezentatywnych stacji z okolic miasta.

Wpływ zanieczyszczeń atmosferycznych oraz warunków pogodowych na jakość powietrza jest powiązany. Z uwagi na brak dostępnych danych meteorologicznych w serwisie powietrze.gios.gov.pl, skorzystano z danepubliczne.imgw.pl w celu uzupełnienia informacji meteorologicznych do analiz. W przypadku tego źródła w krakowie raportowane są sytuacje meteorologiczne z dwóch stacji:

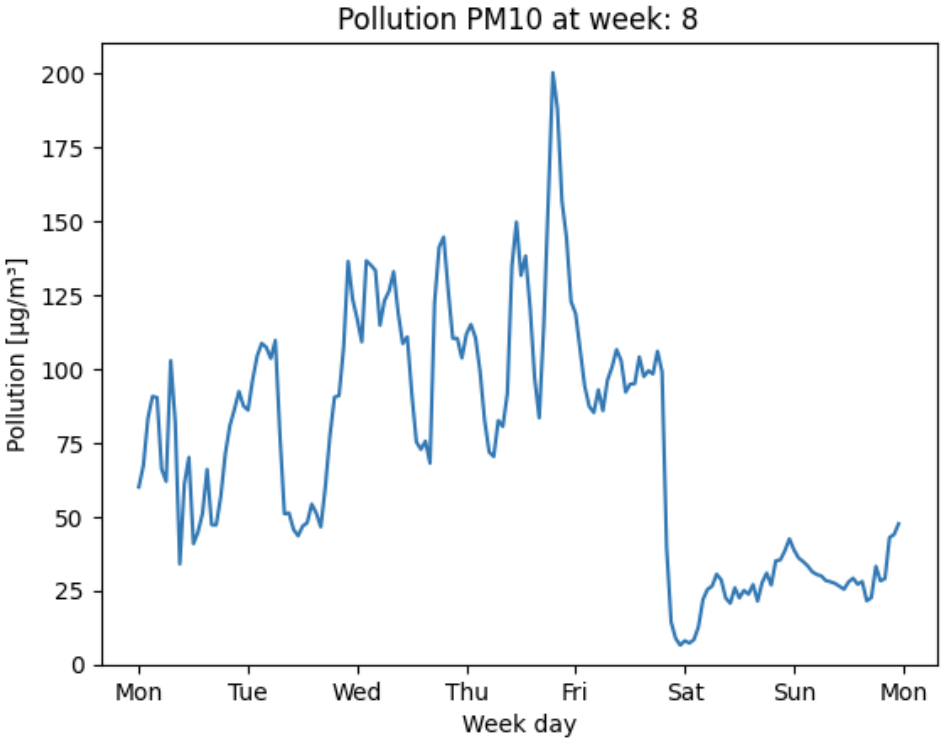
- Kraków balice (lotnisko),
- Kraków obserwatorium (ul. Kopernika 27).

W związku z tymi brakami w szeregach czasowych, podjęto decyzję o zrezygnowaniu z dalszej analizy dla wybranych źródeł danych w ramach. Niedostateczna liczba danych ograniczyła możliwość pełnej oceny i wnioskowania na temat badanych zależności, co utrudniło osiągnięcie zamierzonych celów projektu.

Z tego powodu podjęto próby wyboru innego źródła danych, zwrócono się z prośbą o dostęp do baz danych serwisu Airly, jednak w odpowiedzi otrzymano umowę o zbyt restrykcyjnych warunkach. W związku z tym, zdecydowano się ograniczyć pobieranie danych jedynie poprzez API serwisu Airly. W następnym rozdziale dokumentacji zostaną przedstawione wyniki analizy oparte na danych uzyskanych wspomnianym sposobem.

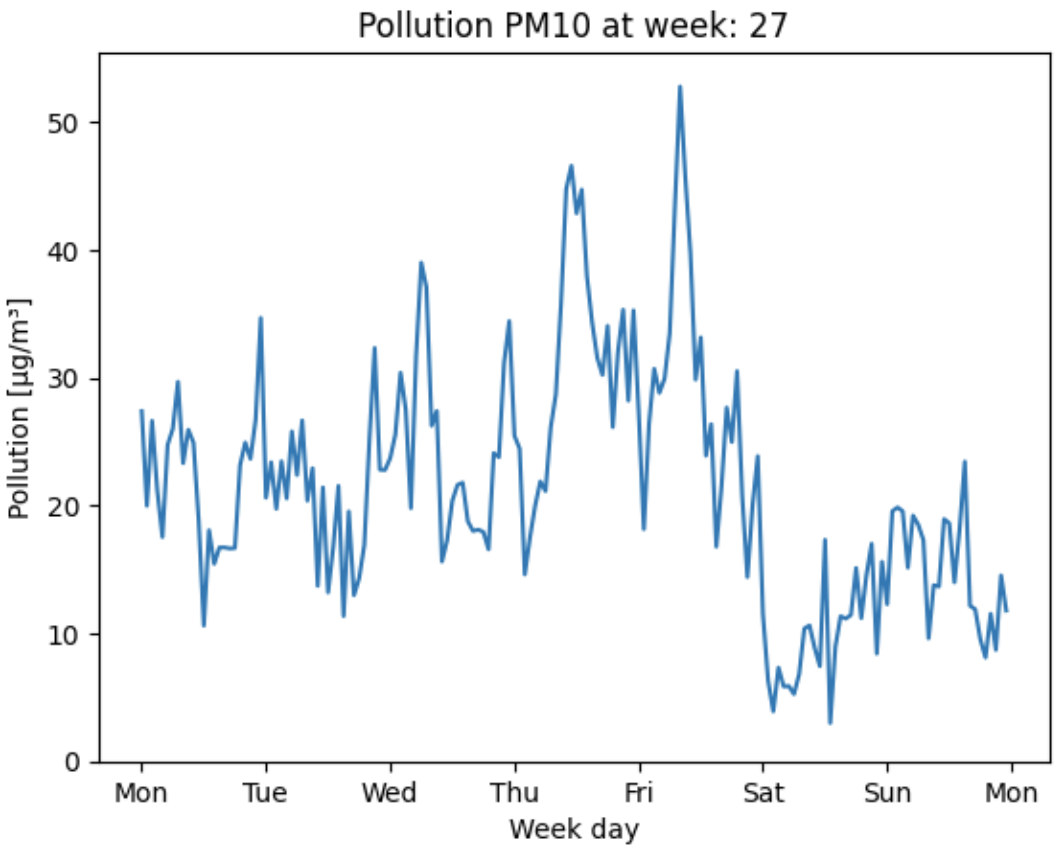
Wstępna analiza danych z pojedynczego czujnika w roku 2021

Przeanalizowana została stacja MpKrakBujaka w roku 2021 i zanieczyszczenie PM10. Dane zawierają odczyty z częstotliwością 1 godziny. Na rysunkach 4.6. - 4.9. przedstawione zostały zmiany poziomu zanieczyszczenia w różnych okresach.



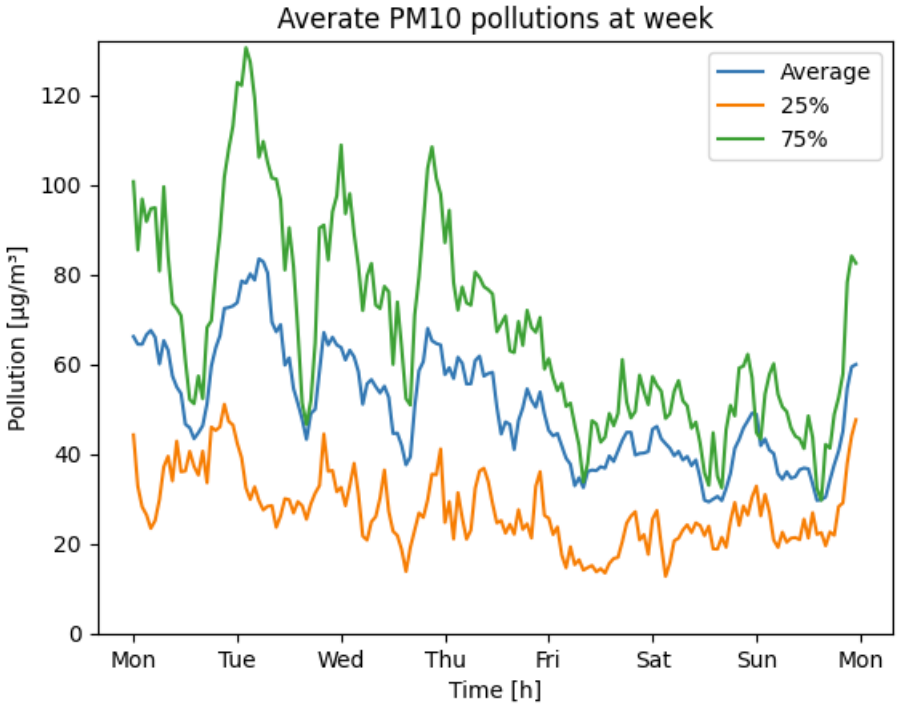
Rys. 4.6. Zmiana poziomu zanieczyszczenia PM10 w 8. tygodniu 2021 roku dla stacji Kraków, ul. Bujaka

Na rysunku 4.6. zostało przedstawione zanieczyszczenie w zimie (8 tydzień roku), a na rysunku 4.7. w lecie (27 tydzień roku). W tygodniu zimowym poziomy zanieczyszczeń osiągały znacznie większe wartości niż w tygodniu letnim. W 8. tygodniu możemy zauważyć duże skoki zanieczyszczenia oraz gwałtowny spadek pod koniec tygodnia. Zmniejszenie poziomu zanieczyszczenia wraz z rozpoczęciem weekendu jest również widoczne w tygodniu letnim.



Rys. 4.7. Zmiana poziomu zanieczyszczenia PM10 w 27. tygodniu 2021 roku dla stacji Kraków, ul. Bujaka

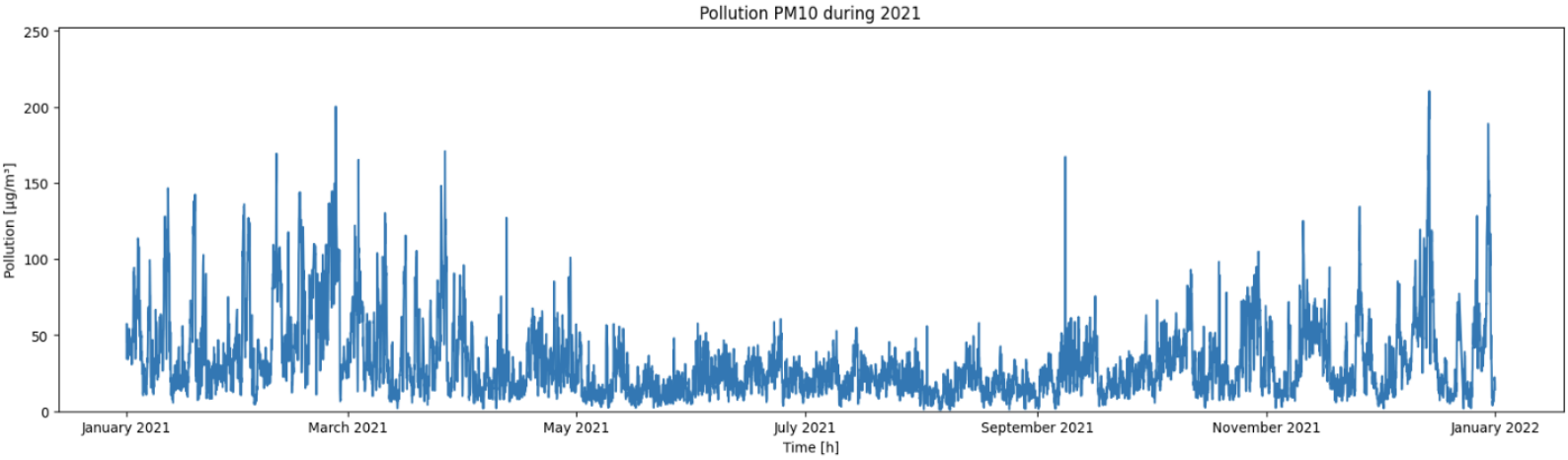
Rysunek 4.8. przedstawia średni poziom zanieczyszczenia w tygodniach zimowych oraz 25. oraz 75. percentyl. Dla 75. percentyla możemy zaobserwować duże skoki zanieczyszczenia. Ciekawy jest również gwałtowny skok występujący pod koniec tygodnia.



Rys. 4.8. Średni poziom zanieczyszczenia PM10 w tygodniu zimowym 2021 roku dla stacji Kraków, ul. Bujaka wraz z percentylami 25% i 75%

Z rysunku 4.9. można odczytać, że poziom zanieczyszczenia w ciągu roku wyglądał następująco:

- w miesiącach zimowych mamy duży poziom zanieczyszczeń oraz duże różnice tego poziomu
- w miesiącach letnich mamy niski i dużo bardziej stabilny poziom zanieczyszczeń



Rys. 4.9. Zmiana poziomu zanieczyszczenia PM10 w 2021 roku dla stacji Kraków, ul. Bujaka

5. Dane

Tabela 5.1. przedstawia wybrane 15 stacji z centrum Krakowa oraz okolicznych gmin wraz z ich dostępnymi pomiarami. Stacje oferują różne pomiary zanieczyszczeń (NO2, O3, PM1, PM10, PM25, CO, SO2) oraz dane pogodowe (temperaturę, ciśnienie, wilgotność). Odczyty są odczytami godzinnymi. Pierwszy odczyt pochodzi z dnia 20.03.2023 o godzinie 16:00:00.

Tabela 5.1. Wybrane stacje do analizy z serwisu Airly

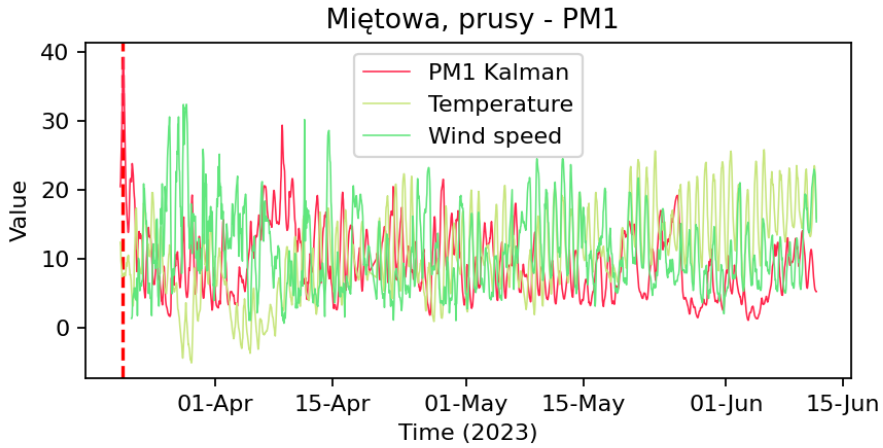
Nazwa stacji	Dostępne pomiary zanieczyszczeń
Franciszka Bujaka	NO2, O3, PM1, PM10, PM2.5
Świętego Wawrzyńca	NO2, O3, PM1, PM10, PM2.5
Liszki, Szkolna	PM1, PM10, PM2.5
Emaus	NO2, O3, PM1, PM10, PM2.5
Krakowska - Zabierzów	NO2, O3, PM1, PM10, PM2.5
Wieliczka - Karola Kuczkiewicza	PM1, PM10, PM2.5
Węgrzce	PM1, PM10, PM2.5
Miętowa, Prusy	PM1, PM10, PM2.5
Niepołomice, Szkolna	PM1, PM10, PM2.5
Rynek Mogilany	PM1, PM10, PM2.5
Skała, Rynek	PM1, PM10, PM2.5
Bulwarowa	CO, NO2, PM10, PM2.5, SO2
Aleja Zygmunta Krasińskiego	CO, NO2, PM10, PM2.5
Skawina	NO2, PM10, SO2
Kaszów	NO2, O3

6. Analizy

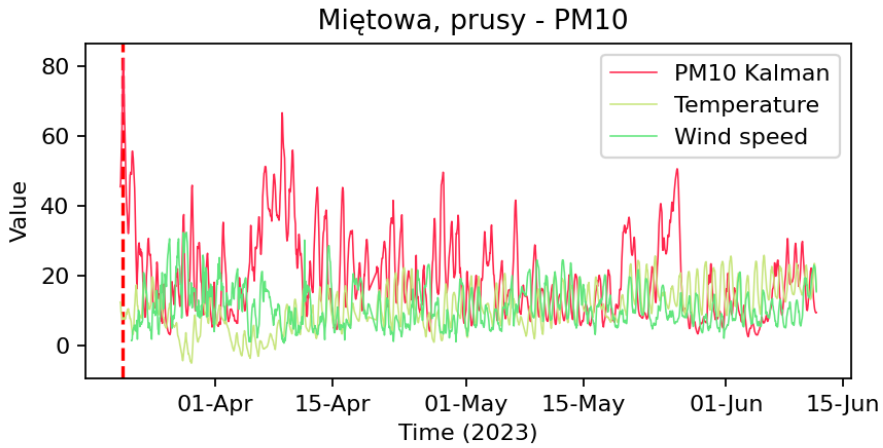
Na wykresach 6.1. - 6.19.przedstawiono zmiany poziomów temperatury, zanieczyszczenia oraz prędkości wiatru w analizowanych stacjach pomiarowych na przestrzeni całego analizowanego okresu. Każdy z wykresów przedstawia jednocześnie te trzy wartości (lub dwie w przypadku braku danych o temperaturze).

W celu uzyskania wygładzonych danych, oryginalne wartości poziomów zanieczyszczeń zostały poddane procesowi filtrowania za pomocą filtra Kalmana. Filtr Kalmana jest popularną techniką stosowaną w analizie sygnałów, która pozwala na estymację wartości na podstawie kombinacji obserwacji i prognoz. W przypadku tego projektu filtr Kalmana został zastosowany w celu usunięcia zakłóceń i szumów z oryginalnych danych, co pozwoliło uzyskać bardziej precyzyjne i stabilne wartości poziomów zanieczyszczeń.

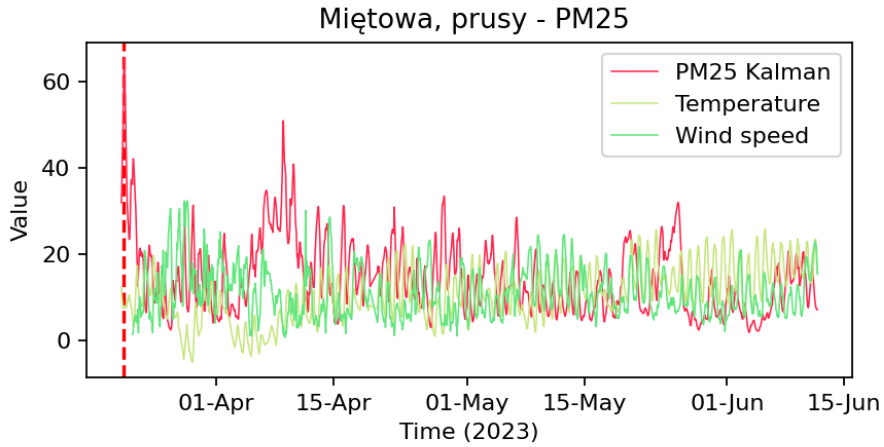
6.1. Poziomy zanieczyszczeń, wiatr, temperatura



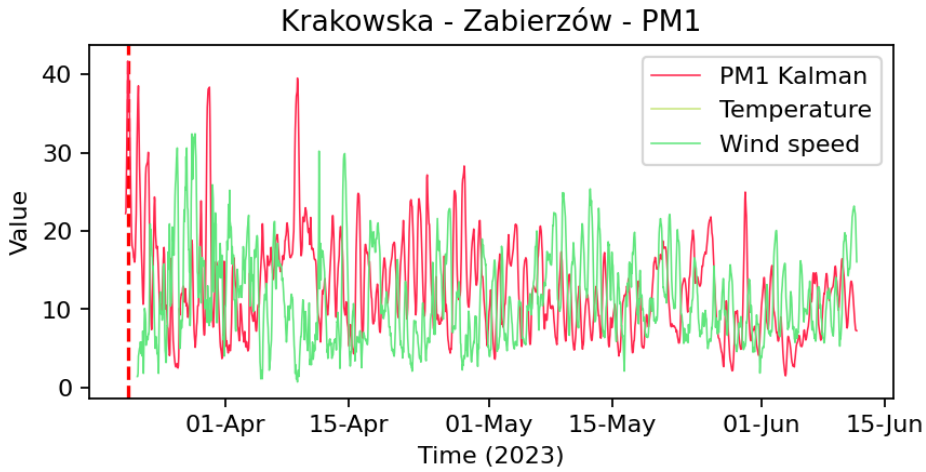
Rys. 6.1.1. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Miętowa, Prusy



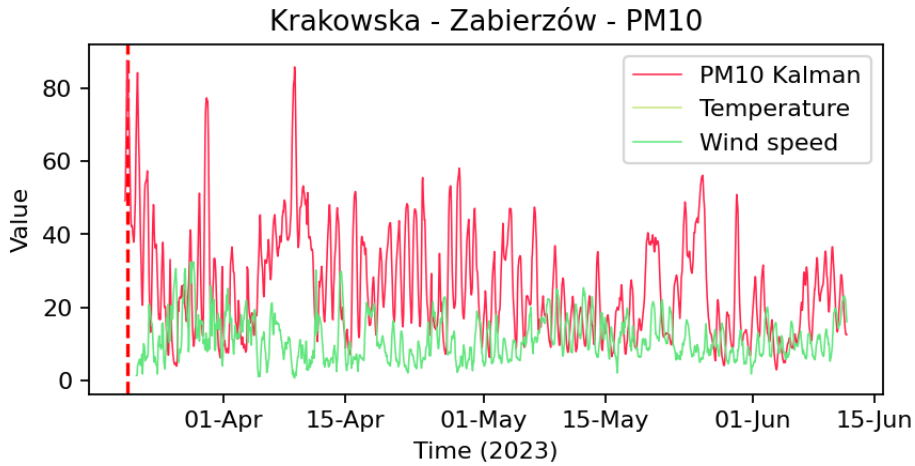
Rys. 6.1.2. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Miętowa, Prusy



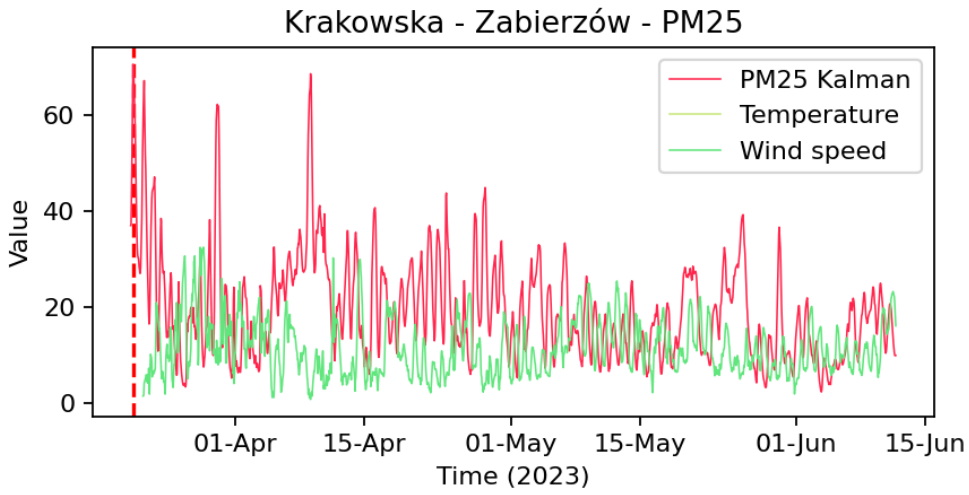
Rys. 6.1.3. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Miętowa, Prusy



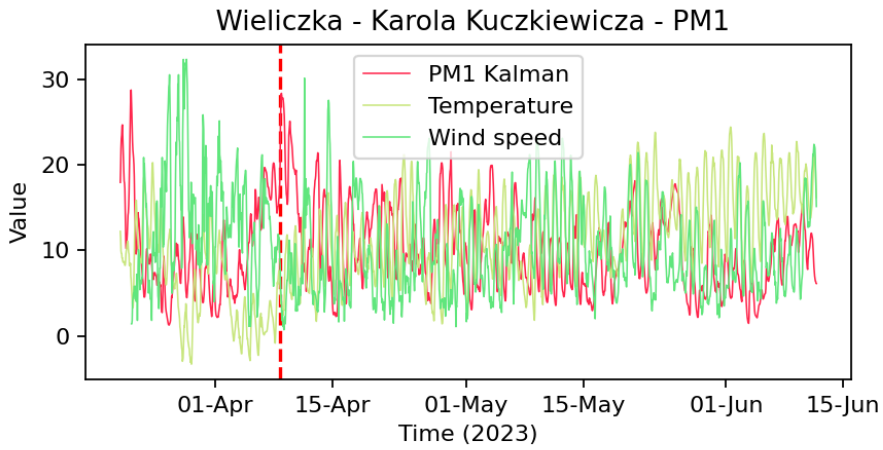
Rys. 6.1.4. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Zabierzów



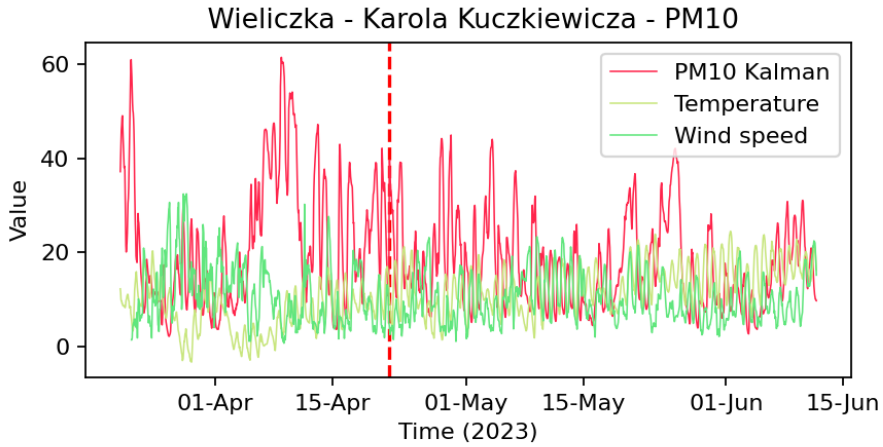
Rys. 6.1.5. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Zabierzów



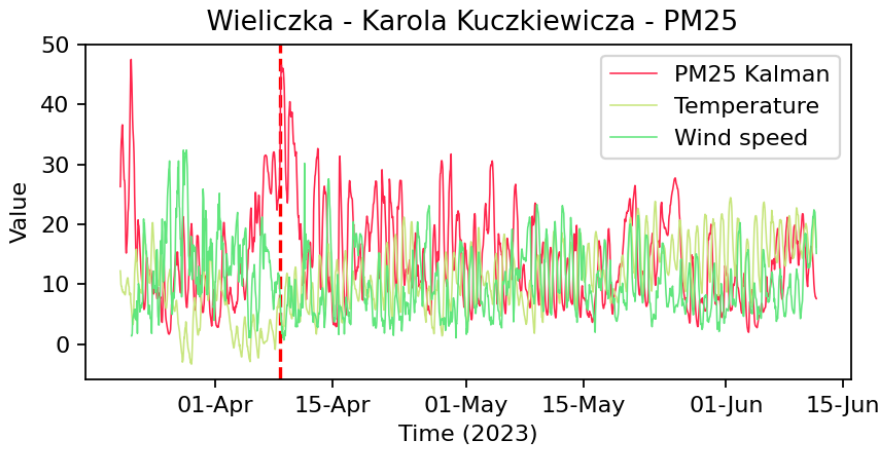
Rys. 6.1.6. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Zabierzów



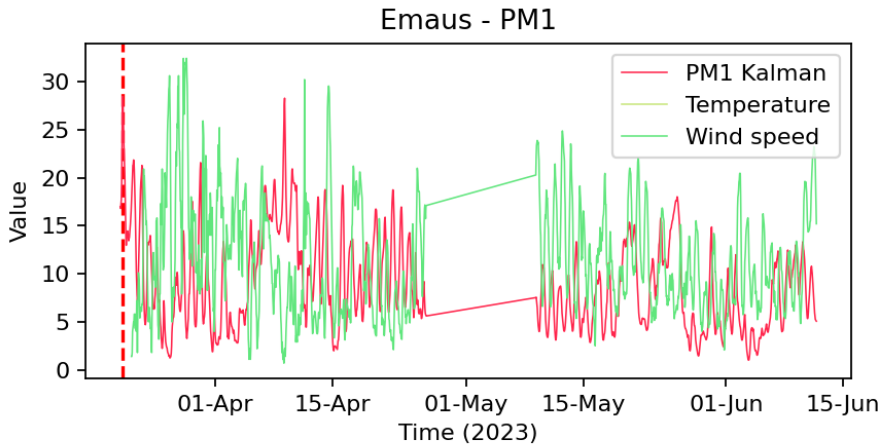
Rys. 6.1.7. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Wieliczka, Kuczkiewicza



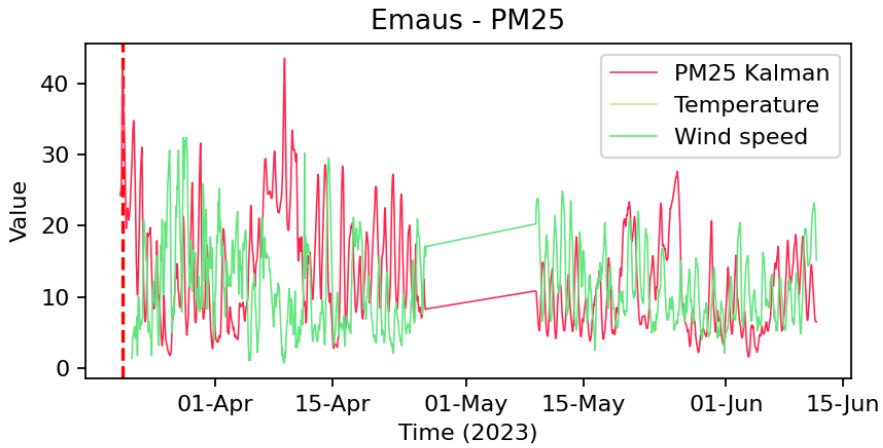
Rys. 6.1.8. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Wieliczka, Kuczkiewicza



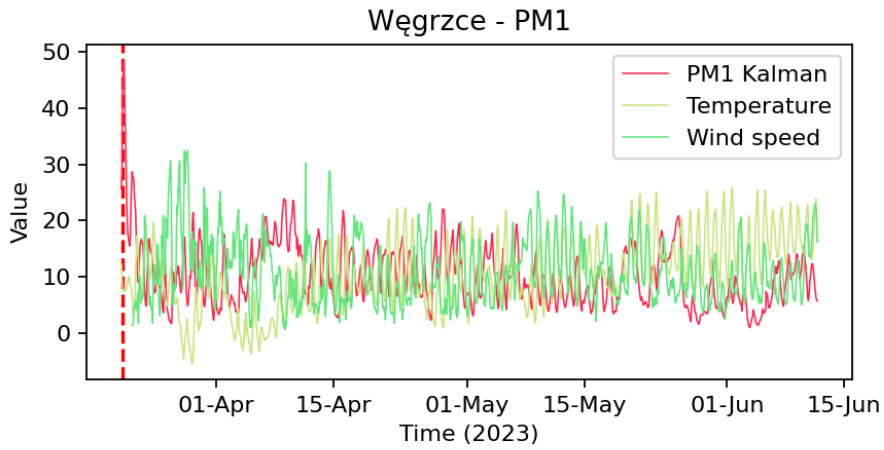
Rys. 6.1.9. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Wieliczka, Kuczkiewicza



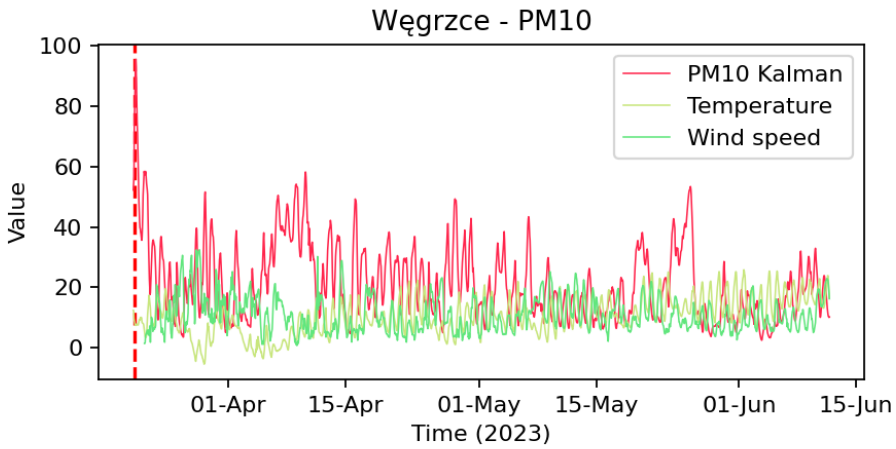
Rys. 6.1.10. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru dla stacji Emaus



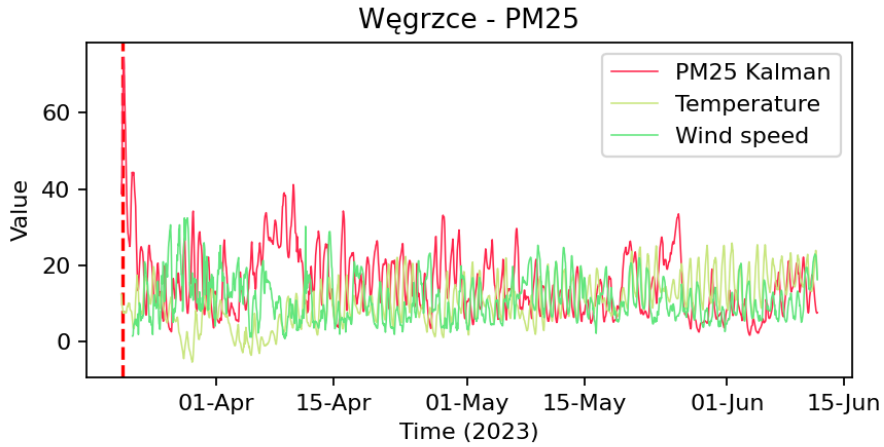
Rys. 6.1.11. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru dla stacji Emaus



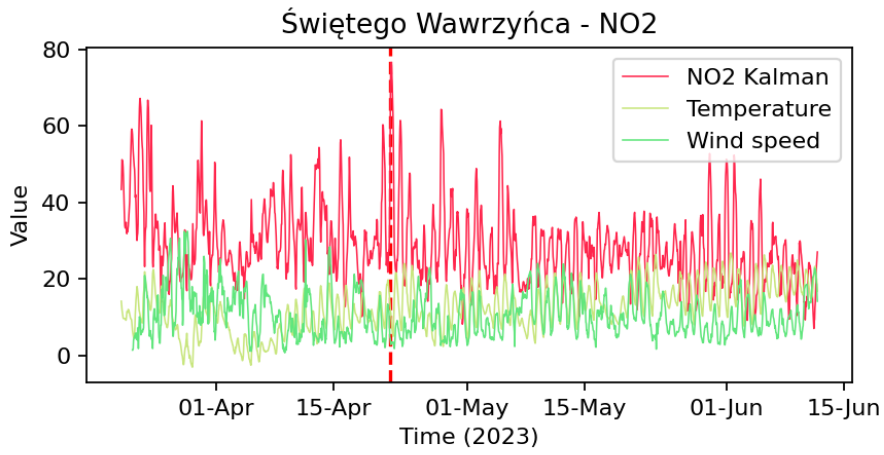
Rys. 6.1.12. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Węgrzce



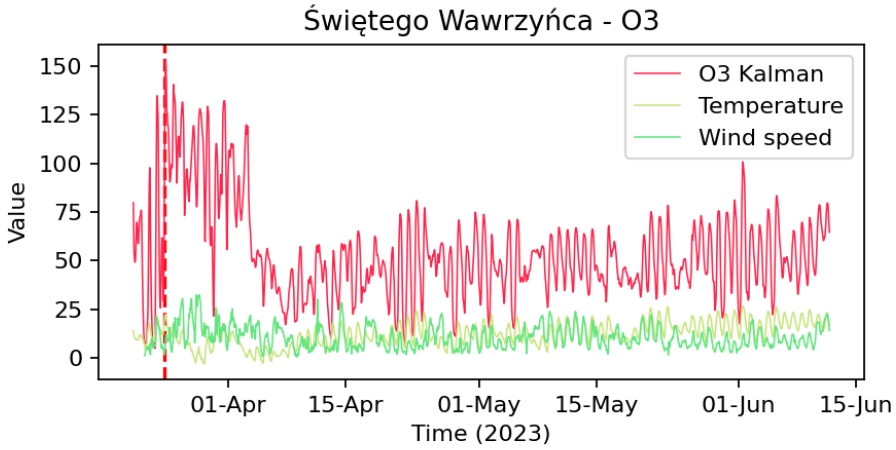
Rys. 6.1.13. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Węgrzce



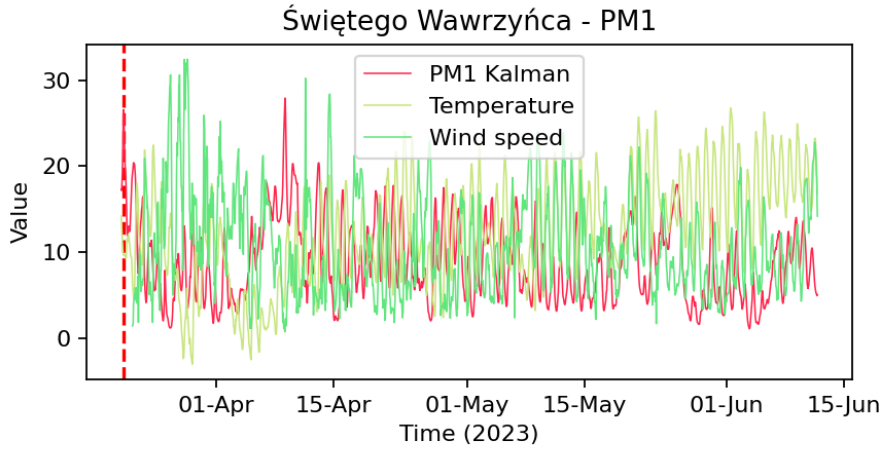
Rys. 6.1.14. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Węgrzce



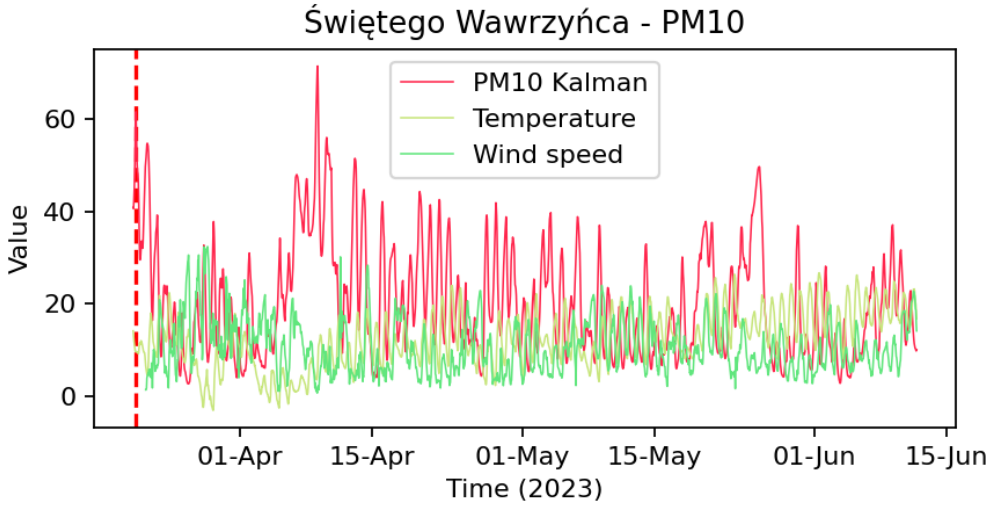
Rys. 6.1.15. Poziom zanieczyszczenia NO2 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca



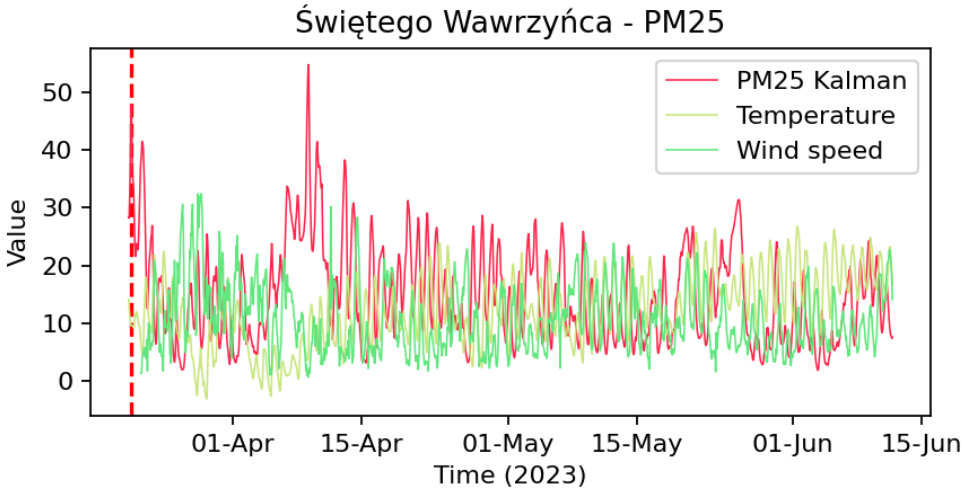
Rys. 6.1.16. Poziom zanieczyszczenia O3 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca



Rys. 6.1.17. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca



Rys. 6.1.18. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca



Rys. 6.1.19. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca

Na wykresach można zauważyć brak pomiarów dla pewnych okresów. Wiąże się to z tym, że niektóre stacje nie udostępniały tego konkretnego pomiaru meteorologicznego. W przypadku niektórych stacji zaobserwowano też kilkudniową przerwę w dostawie pomiarów w okolicy początku maja.

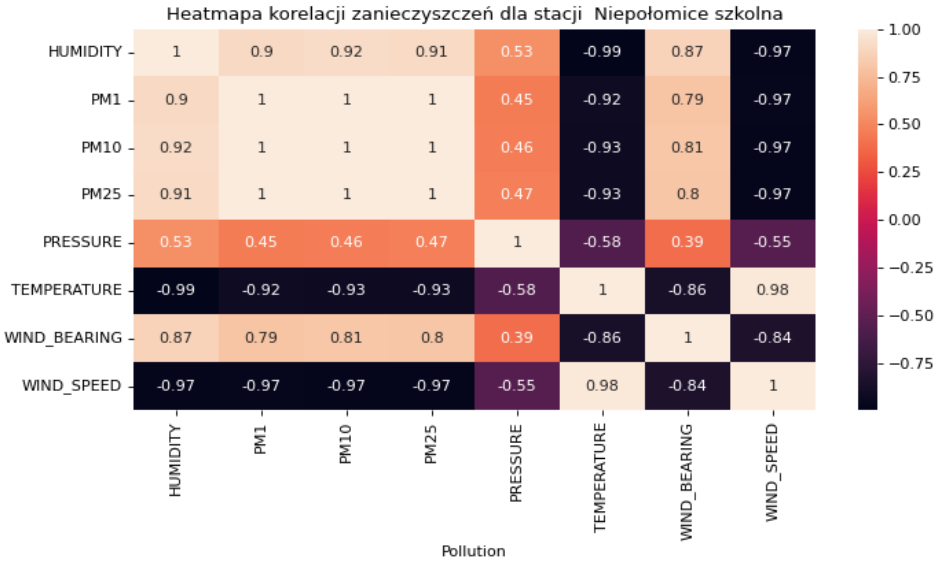
Analizując zmiany poziomu wartości zanieczyszczeń oraz zmian temperatury na powyższych wykresach można wnioskować, że kiedy podnosi się temperatura, to spada poziom zanieczyszczeń. Można zakładać, że jest to związane z zapotrzebowaniem na ogrzewanie. Spadek temperatury często prowadzi do większego zużycia energii i większej produkcji ciepła. Przy porównywaniu prędkości powietrza z zanieczyszczeniami zauważamy, że wyższe prędkości skutkują zmniejszeniem zanieczyszczeń.

Temperatura ma również wpływ na warunki atmosferyczne, takie jak prędkość wiatru i opady. Silniejszy wiatr może pomóc w rozpraszaniu zanieczyszczeń, zmniejszając ich stężenie w powietrzu, zwłaszcza w przypadku lokalnych źródeł emisji, takich jak fabryki, elektrownie czy ruch samochodowy. W rezultacie, obszary o wyższych prędkościach wiatru mogą doświadczać niższych poziomów zanieczyszczeń. Warto też zauważyć, że niższe temperatury mogą powodować stabilność atmosferyczną, co prowadzi do zatrzymywania zanieczyszczeń w niższych warstwach atmosfery, czyli wzrostu poziomu zanieczyszczeń.

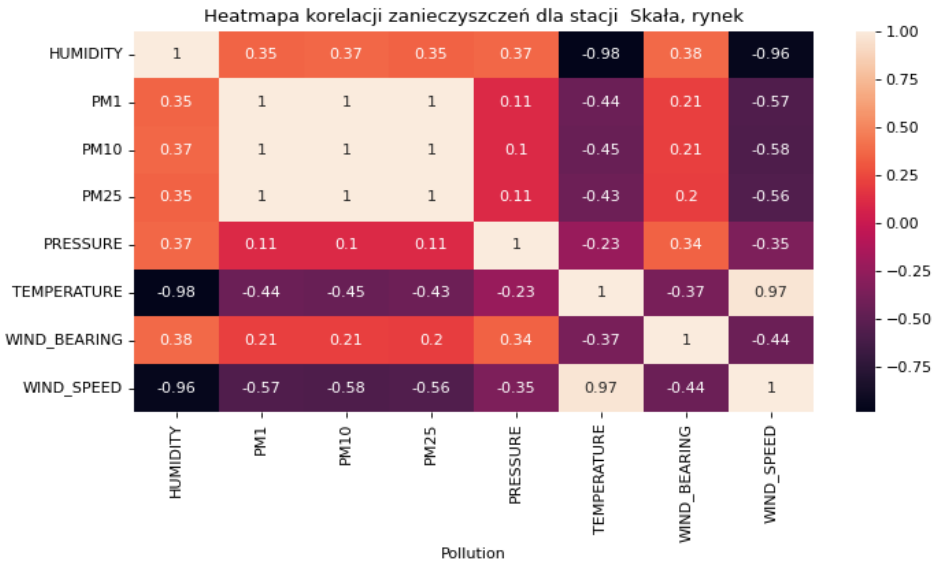
Na wykresach można również odnotować kilka okresów nagłych skoków poziomu zanieczyszczeń. Występowały one m.in. około 10 kwietnia, 25 maja oraz pod koniec marca.

6.2. Korelacje

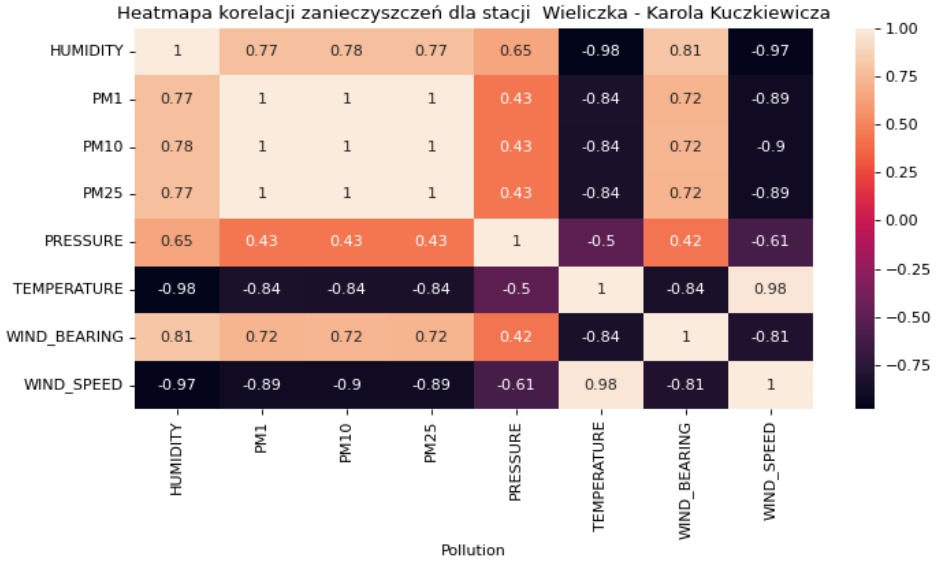
Rysunki 6.2.1. - 6.2.6. przedstawiają heatmapy korelacji pomiarów dla wybranych stacji.



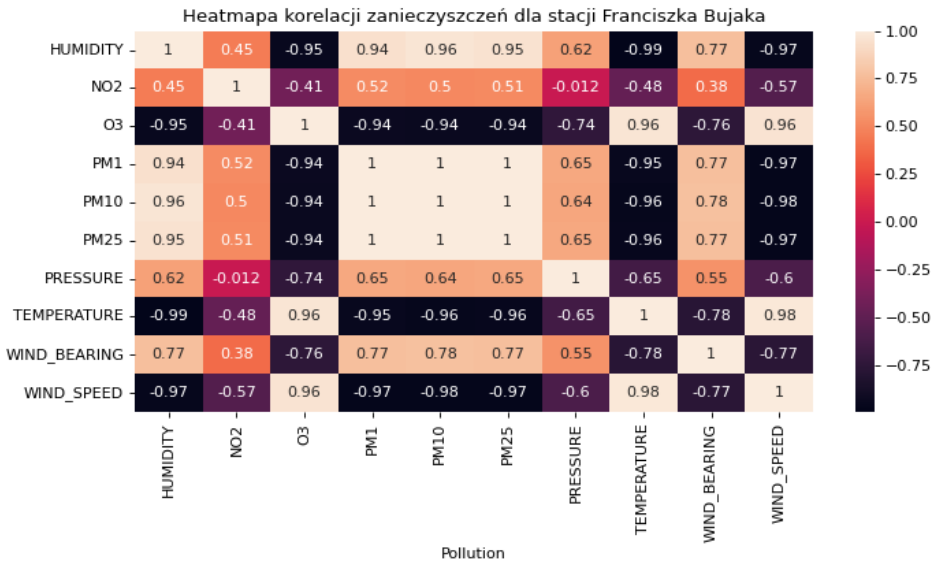
Rys. 6.2.1. Korelacje pomiarów stacji Niepołomice, Szkolna



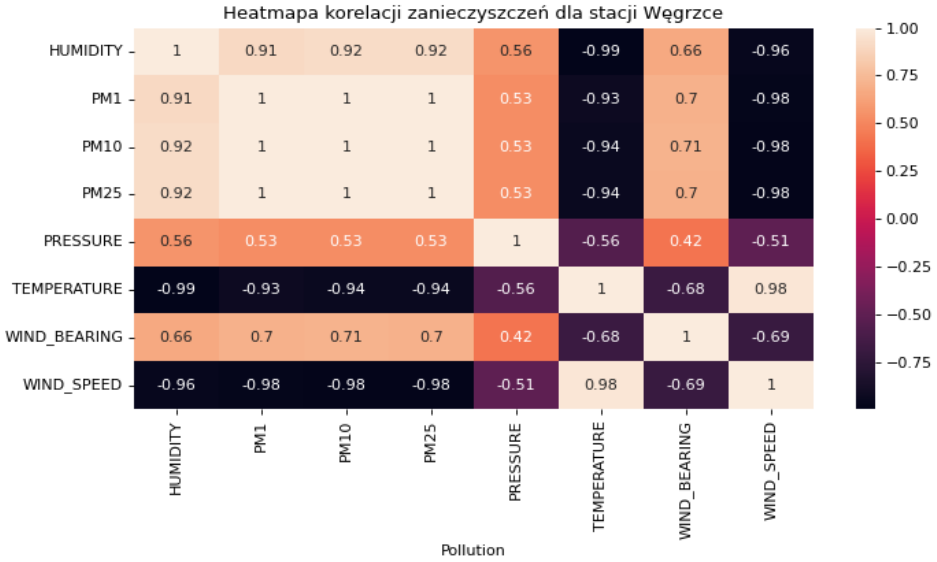
Rys. 6.2.2. Korelacje pomiarów stacji Skała, Rynek



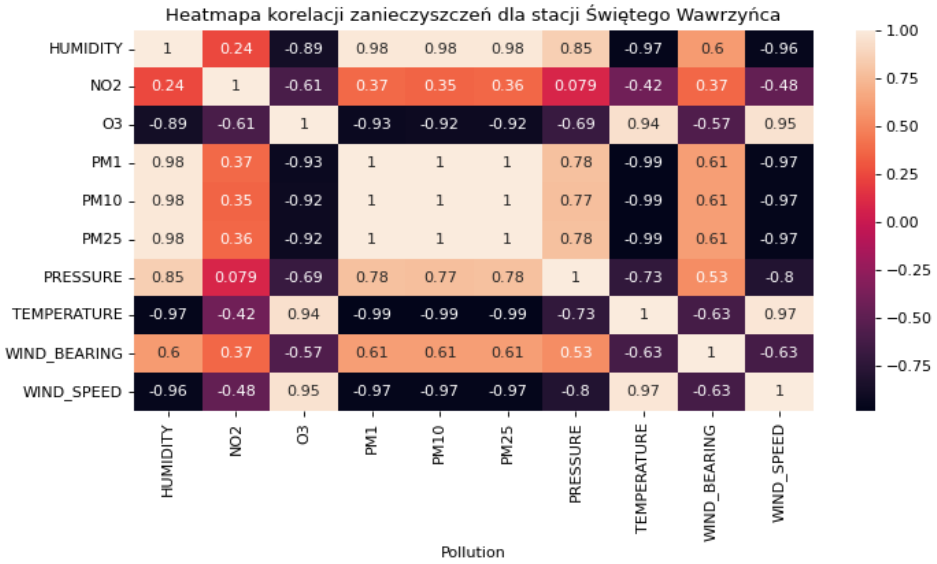
Rys. 6.2.3. Korelacje pomiarów stacji Wieliczka, Kuczkiewicza



Rys. 6.2.4. Korelacje pomiarów stacji Franciszka Bujaka



Rys. 6.2.5. Korelacje pomiarów stacji Węgrzce

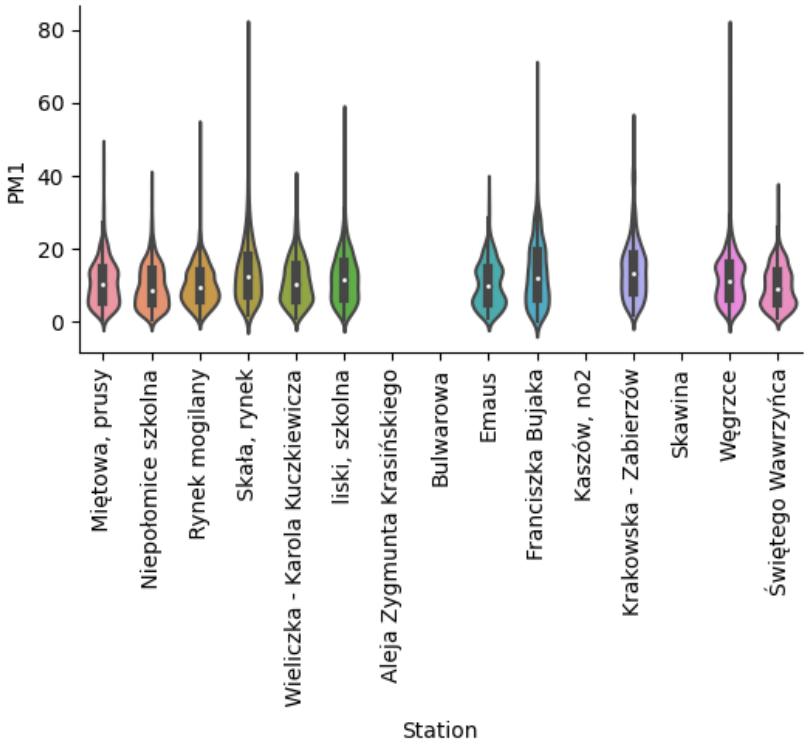


Rys. 6.2.6. Korelacje pomiarów stacji Świętego Wawrzyńca

Analizując heatmapy korelacji można zauważyć, że:

- Bliska -1 korelacja prędkości wiatru z zanieczyszczeniami oraz temperatury z zanieczyszczeniami, co potwierdza analizę wykresów
- Wysoka dodatnia korelacja wilgotności z pyłami PM
- Korelacja ciśnienia z zanieczyszczeniami przyjmuje różne poziomy, np. dla stacji Skała, Rynek jest to 10%, a dla stacji Św. Wawrzyńca 78%
- Dla stacji Skała, Rynek korelacje nie są tak mocne jak dla innych stacji
- Bliska -1 korelacja O3 z innymi pomiarami
- Słaba korelacja SO2 z innymi pomiarami

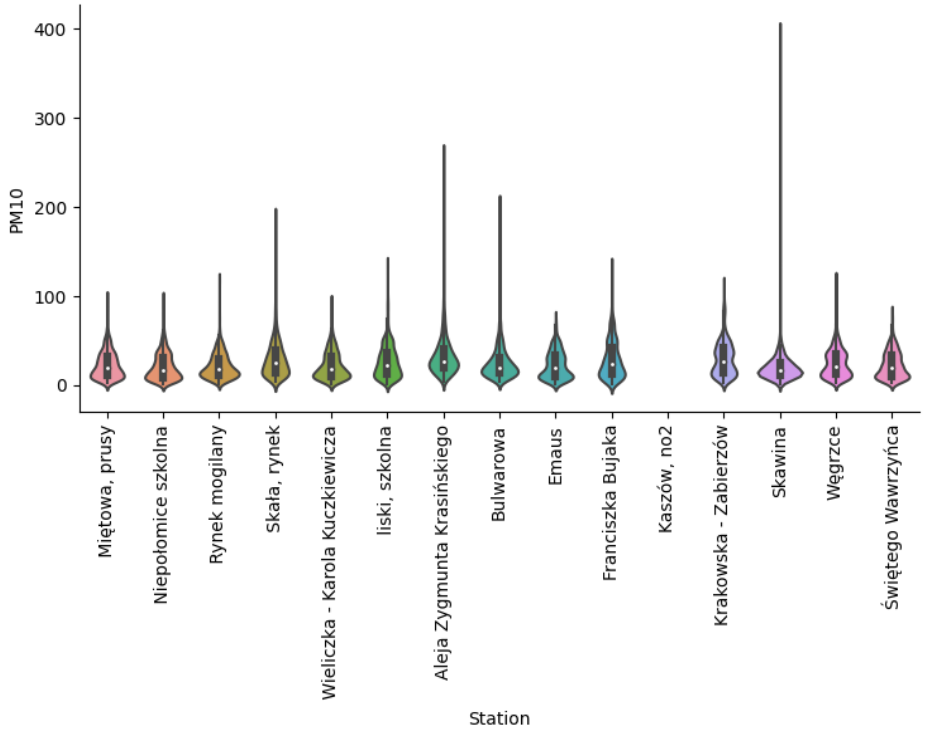
6.3. Wykresy skrzypcowe pyłów PM1, PM2.5 i PM10



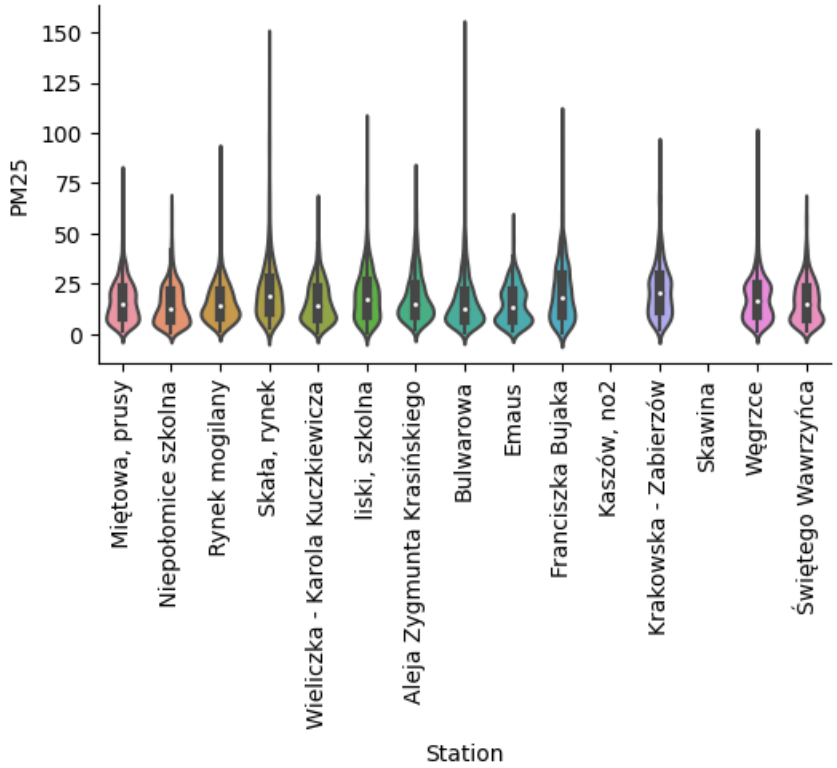
Rys. 6.3.1. Wykresy skrzypcowe pomiarów PM1

Na powyższych wykresach da się zauważyć, że niektóre stacje osiągają większe średnie wartości dla wszystkich zanieczyszczeń. Są to stacje np. Krakowska - Zabierzów, Franciszka Bujaka, Skawina, Aleja Zygmunta Krasińskiego czy Skała, Rynek dla pomiarów PM1, PM2.5, PM10. Aleja Zygmunta Krasińskiego osiąga największe wartości dla NO2, a Kaszów dla O3.

Wykresy skrzypcowe przedstawiają rozkłady wartości zanieczyszczeń PM1, PM2.5, PM10 dla każdej ze stacji. Można zauważyć, że wykresy dla PM1 i PM2.5 dla odpowiadających stacji są do siebie podobne, bardziej smukłe. Z kolei wykresy dla PM10 mają przeważnie większe prawdopodobieństwo dla mniejszych wartości zanieczyszczeń.



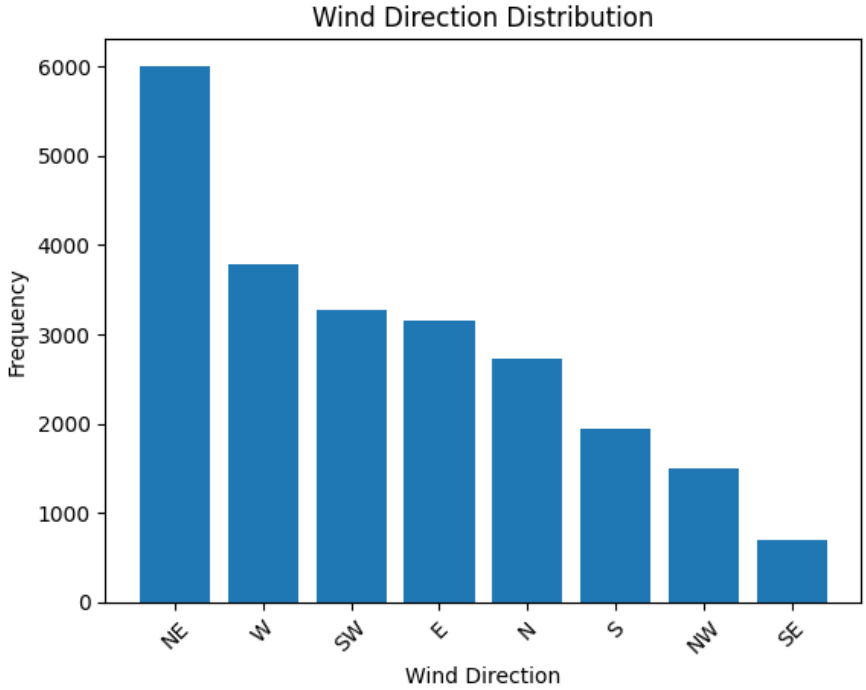
Rys. 6.3.2. Wykresy skrzypcowe pomiarów PM10



Rys. 6.3.3. Wykresy skrzypcowe pomiarów PM2.5

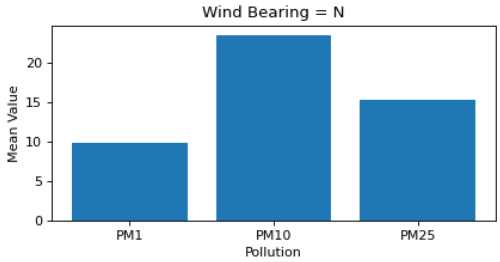
6.4. Wiatr

Częstość kierunków wiatru:

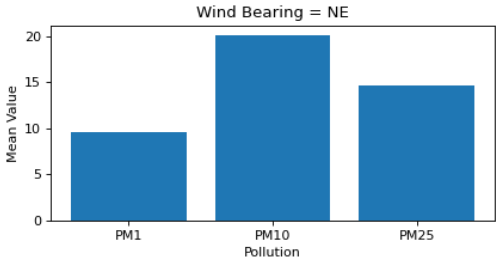


Rys. 6.4.1. Dystrybucja kierunków wiatru

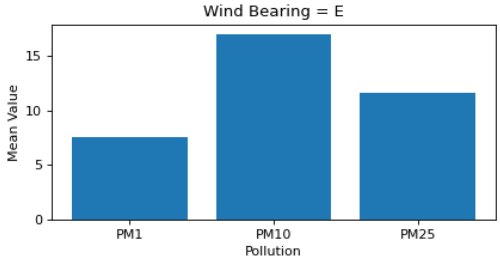
Na aktualny moment najczęściej występował wiatry północno-wschodni.



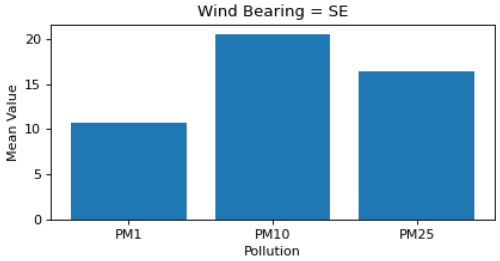
Rys. 6.4.2. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru północnego



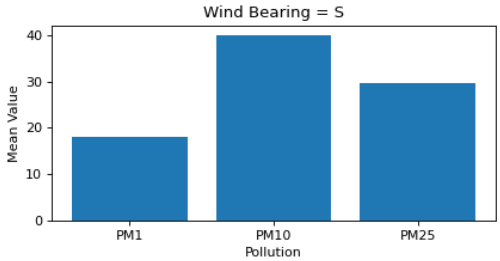
Rys. 6.4.3. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru północno-wschodniego



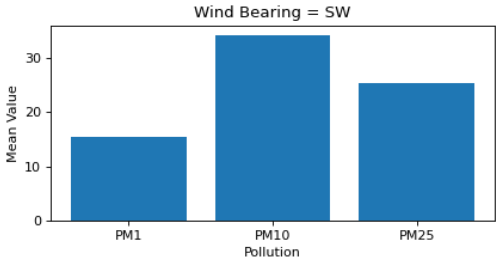
Rys. 6.4.4. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru wschodniego



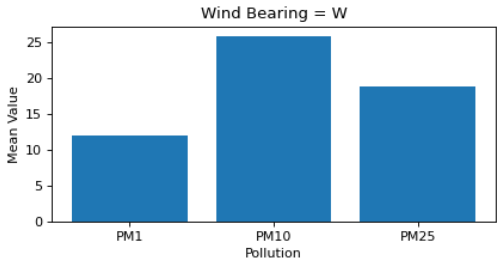
Rys. 6.4.5. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla południowo-wschodniego



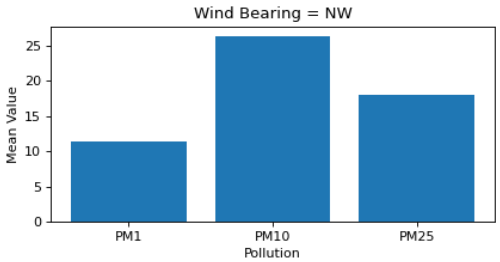
Rys. 6.4.6. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru południowego



Rys. 6.4.7. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru południowo-zachodniego



Rys. 6.4.8. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru zachodniego



Rys. 6.4.9. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru północno-zachodniego

Jeśli chodzi o wysokość zanieczyszczeń w zależności od kierunku wiatru, w przypadku kierunków południowego i południowo-zachodniego średnie dla wszystkich zanieczyszczeń są wyższe niż dla pozostałych kierunków. Kierunek wiatru jest o tyle istotny, że wiatr może transportować zanieczyszczenia z jednego obszaru do drugiego. Na przykład, jeśli wiatr wieje z obszaru o dużej koncentracji zanieczyszczeń do obszaru o niższej koncentracji, może dojść do wzrostu poziomu zanieczyszczeń w obszarze docelowym.

6.5. Średnie zanieczyszczenie w podziale na Kraków i okolice

W tabelach 6.5.1. - 6.5.3. zostały przedstawione korelacje pomiędzy średnimi zanieczyszczeniami pyłów PM1, PM2.5 i PM10 w podziale na stacje znajdujące się w mieście oraz te z okolicznych gmin. Są to średnie z całego analizowanego okresu.

Tabela 6.5.1. Korelacja średniego zanieczyszczenia PM1 w Krakowie i okolicach

PM1	Kraków	Okolice
Kraków	1	0.90286
Okolice	0.90286	1

Tabela 6.5.2. Korelacja średniego zanieczyszczenia PM10 w Krakowie i okolicach

PM10	Kraków	Okolice
Kraków	1	0.876703
Okolice	0.876703	1

Tabela 6.5.3. Korelacja średniego zanieczyszczenia PM2.5 w Krakowie i okolicach

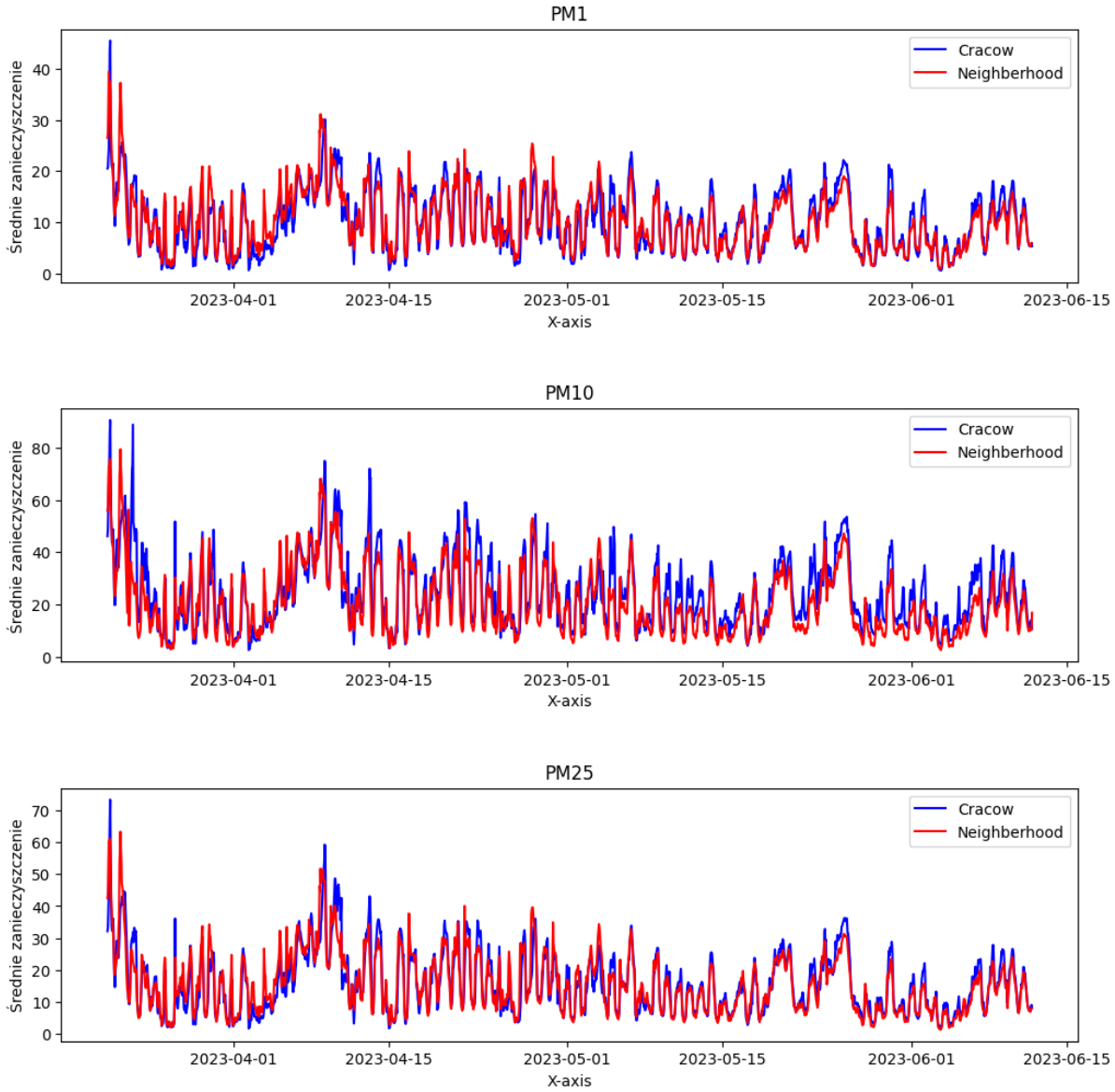
PM2.5	Kraków	Okolice
Kraków	1	0.910903
Okolice	0.910903	1

Współczynnik korelacji wskazuje na silną dodatnią korelację pomiędzy badanymi obszarami pod względem wysokości zanieczyszczeń. W przypadku wszystkich zanieczyszczeń typu PM osiągnęły one poziom około 90%. Zatem centrum miasta oraz okoliczne gminy wpływają na siebie nawzajem.

Interpretacja tego wyniku może wskazywać na kilka możliwych scenariuszy. Przede wszystkim, wysoka korelacja może sugerować, że główne źródła zanieczyszczenia znajdują się w centrum miasta i wpływają na jakość powietrza w sąsiednich gminach. Na przykład, jeśli w centrum miasta znajdują się fabryki, czy duży ruch samochodowy, to emisje z tych źródeł mogą się rozprzestrzeniać na okoliczne obszary, co powoduje wysokie stężenie zanieczyszczeń również w tych gminach.

Z drugiej strony, istnieje również możliwość, że korelacja ta wynika z podobnych czynników atmosferycznych i geograficznych, które wpływają na obie lokalizacje. Na przykład, wiatr czy ukształtowanie terenu mogą sprzyjać rozprzestrzenianiu się zanieczyszczeń z centrum miasta na okoliczne gminy.

Rysunek 6.5.1. składa się z trzech wykresów przedstawiającą zmiany uśrednionych wartości pyłów zawieszonych PM w podziale na te pochodzące z miasta Krakowa oraz okolic podczas całego badanego okresu.

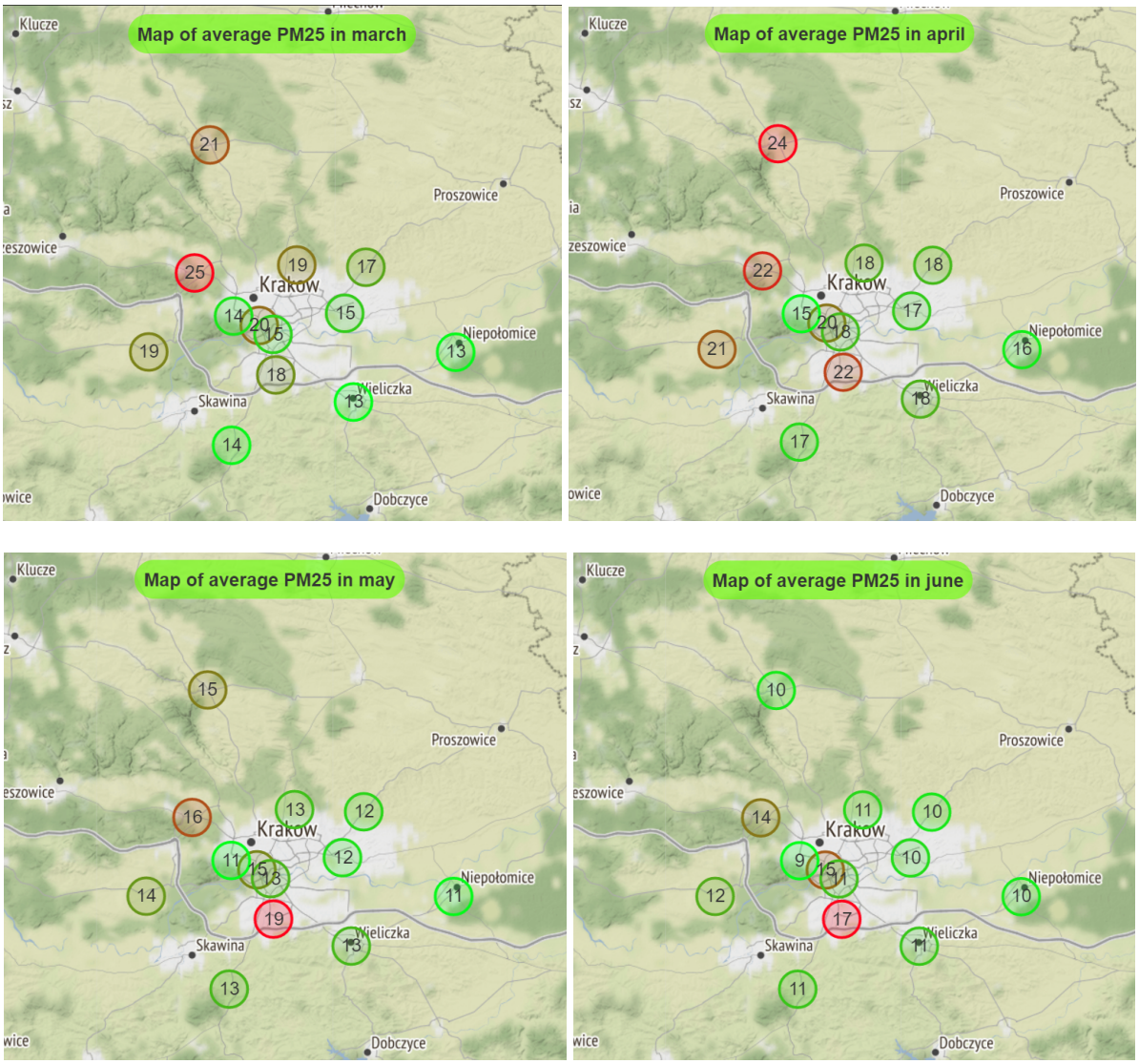


Rys. 6.5.1. Zmiana średniego poziomu zanieczyszczeń PM1, PM2.5, PM10 w podziale na Kraków i okolice

Analizując wykresy poziomów średnich zanieczyszczeń, można zauważyć, że przez większą część badanego okresu to poziom zanieczyszczeń z miasta Krakowa był większy. Oczywiście, były też momenty, w których to stacje znajdujące się w “Obwarzanku” notowały wyższe niż w centrum miasta zanieczyszczenia. Taką sytuację można zaobserwować do końcówki pierwszego tygodnia kwietnia.

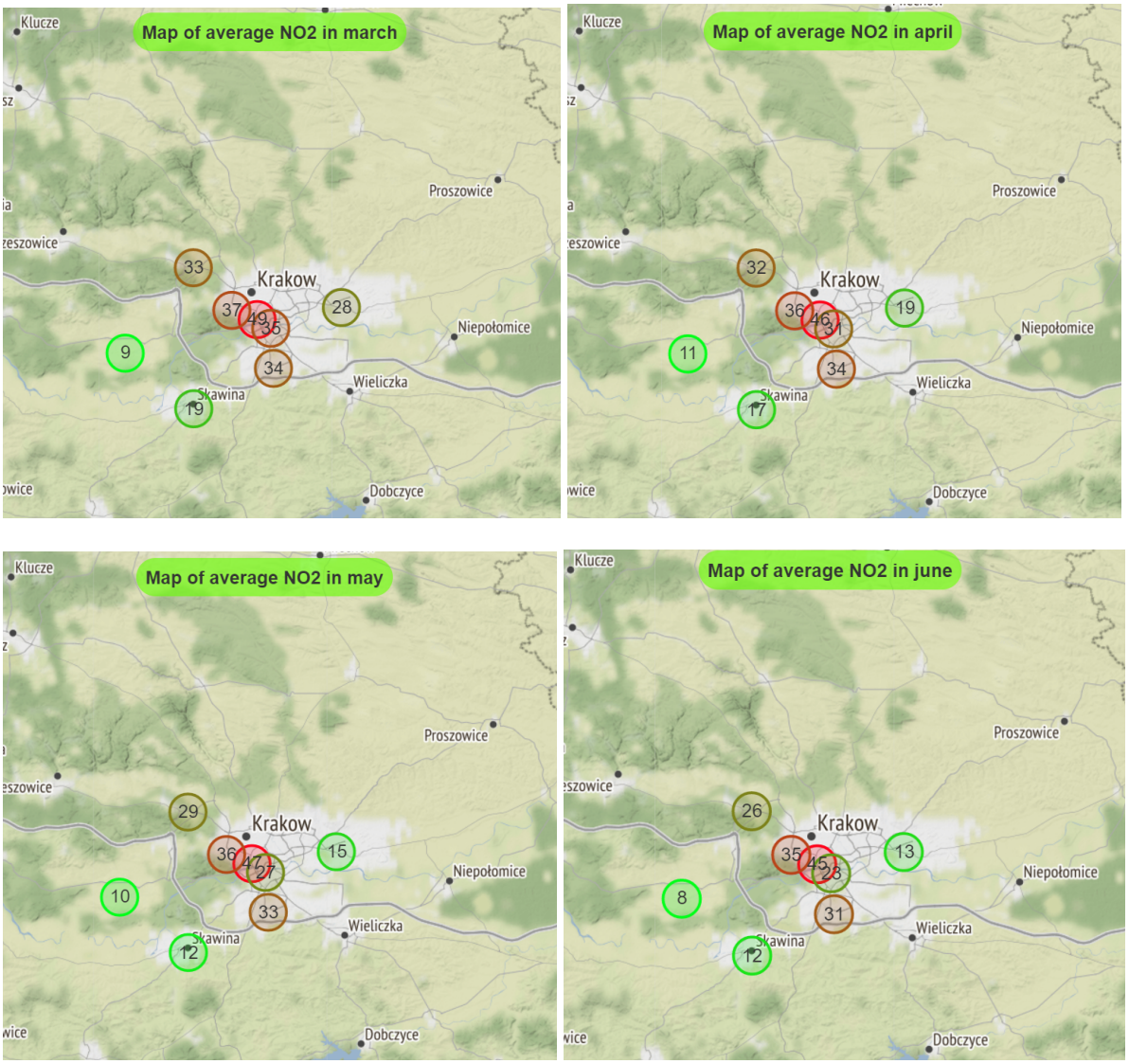
6.6. Mapy średniego miesięcznego zanieczyszczenia według stacji

Rysunek 6.6.1. oraz 6.6.2. przedstawiają średnie poziomu wybranych zanieczyszczeń dla stacji na przestrzeni marzec-czerwiec. Wartości te zostały zwizualizowane na mapach, aby można było dostrzec czy stacja znajduje się w mieście czy w “Obwarzanku”.



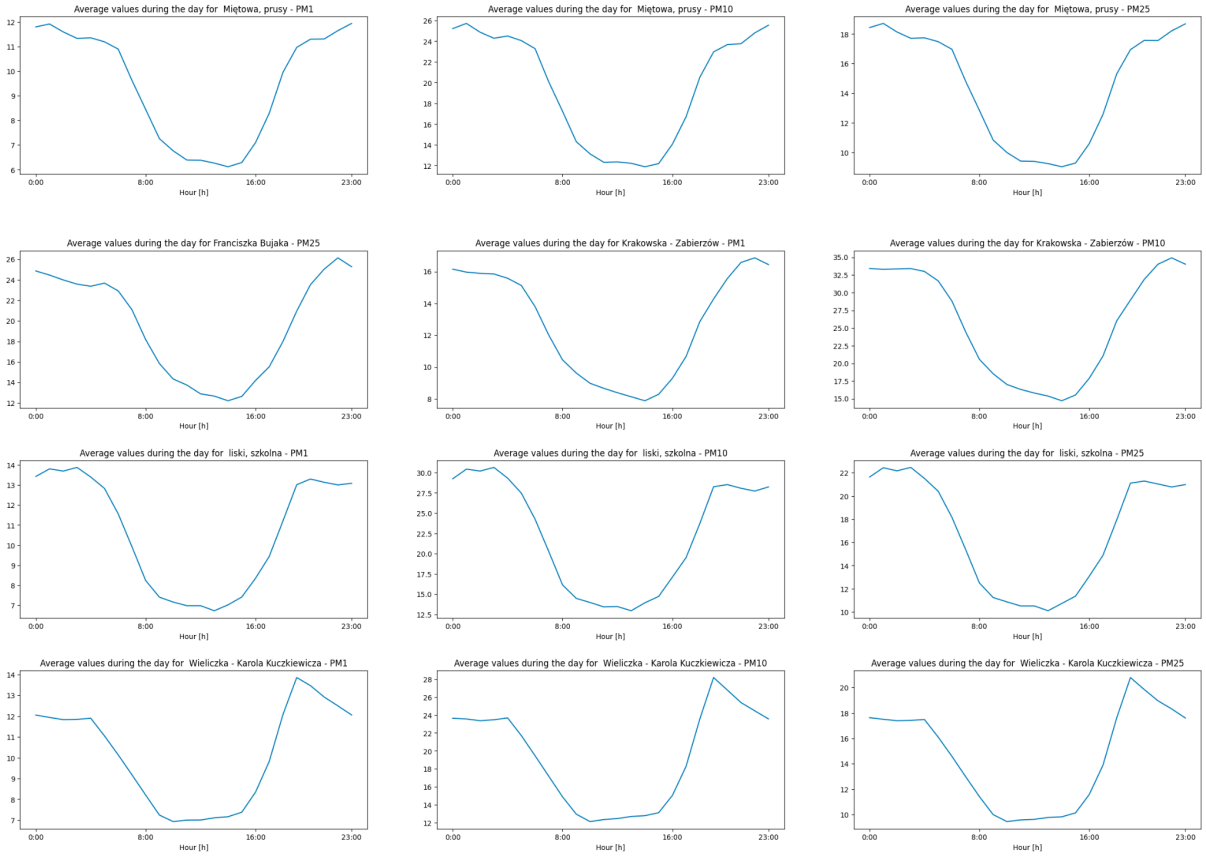
Rys. 6.6.1. Mapy średniego zanieczyszczenia PM2.5 na przestrzeni marzec-czerwiec według stacji

Dla każdego zanieczyszczenia zmiany wyglądały dość podobnie. W kwietniu nastąpił wzrost zanieczyszczeń w przypadku wszystkich stacji względem marca. Później te średnie zaczęły maleć i czerwiec jest miesiącem o najniższych poziomach zanieczyszczeń spośród badanych 4 miesięcy.



Rys. 6.6.2. Mapy średniego zanieczyszczenia NO2 na przestrzeni marzec-czerwiec według stacji

6.7. Średnie zanieczyszczenie w ciągu doby i tygodnia

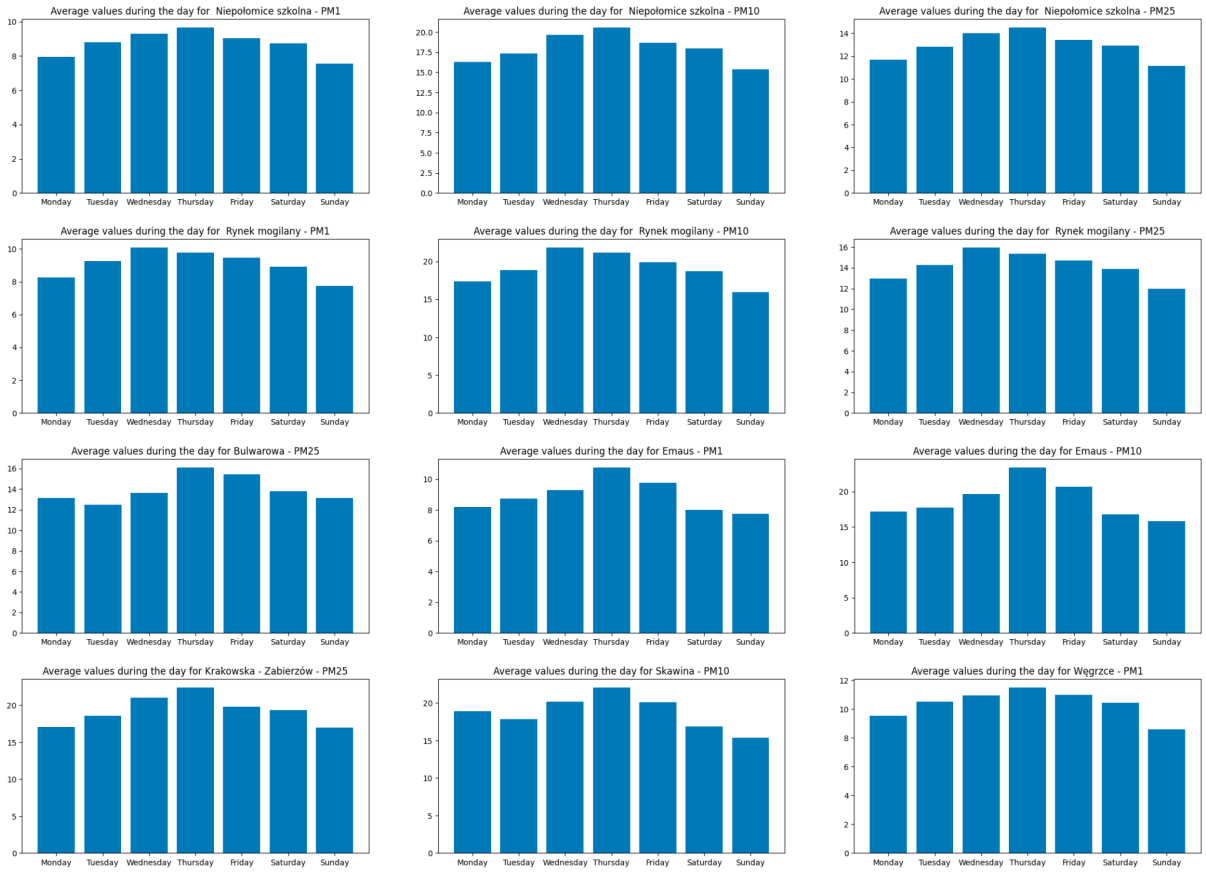


Rys. 6.7.1. Zmiana średniego zanieczyszczenia pyłów PM1, PM2.5, PM10 na przestrzeni doby dla wybranych stacji

Opisując średnie zanieczyszczenie na przestrzeni 24 godzin, można zauważyć, że ten wzorzec jest podobny dla każdej stacji pomiarowej i każdego rodzaju zanieczyszczenia. Największe wartości występują w godzinach nocnych, zaczynając od wieczora i utrzymując się do wczesnych godzin porannych.

W okolicy godziny 5 rano obserwuje się początek spadku poziomu zanieczyszczeń. Może to być związane z kilkoma czynnikami. Po pierwsze, wczesne godziny poranne często charakteryzują się mniejszym ruchem samochodowym, zwłaszcza w porównaniu do godzin szczytu wieczornego, co może przyczynić się do zmniejszenia emisji spalin i zanieczyszczeń powietrza. Ponadto, warunki atmosferyczne, takie jak niskie temperatury, mogą sprzyjać zatrzymywaniu się cząstek zanieczyszczeń bliżej powierzchni ziemi.

Po godzinie 5 rano zanieczyszczenia zaczynają maleć, a poziomy powietrza stają się stopniowo lepsze. Ten trend utrzymuje się przez większość godzin porannych i przedpołudniowych, aż do około godziny 15, kiedy to poziom zanieczyszczeń ponownie zaczyna wzrastać. Może to być związane ze zwiększającym się ruchem samochodowym wraz z końcem dnia pracy i rozpoczynającymi się godzinami szczytu. Dodatkowo działalność przemysłowa czy energetyczna, która może być większa w popołudniowych godzinach, może również przyczynić się do wzrostu zanieczyszczeń.



Rys. 6.7.2. Zmiana średniego zanieczyszczenia pyłów PM1, PM2.5, PM10 na przestrzeni tygodnia dla wybranych stacji

Na postawie wykresów średnich zanieczyszczeń na przestrzeni dni tygodnia, zauważyć można, że poziomy zanieczyszczeń maleją wraz z początkiem weekendu, czyli w piątek, a zaczynają rosnąć na początku tygodnia. Zazwyczaj najwyższe poziomy zanieczyszczeń osiągane są w środę lub czwartek.

Analizując wykresy średnich zanieczyszczeń na przestrzeni dni tygodnia, można zauważyć pewne powtarzające się wzorce. Malejące poziomy zanieczyszczeń w piątek i wzrost na początku tygodnia sugerują związek między dniem tygodnia a jakością powietrza.

Najprawdopodobniej powodem spadku zanieczyszczeń w piątek jest zmniejszona aktywność przemysłowa i mniejszy ruch komunikacyjny związane z nadchodzącym weekendem. Często w piątki wieczorem i weekendy obserwuje się mniejszą aktywność przemysłową, zwłaszcza w niektórych branżach, co przekłada się na mniejszą emisję zanieczyszczeń. Ponadto, mniejszy ruch samochodowy w czasie weekendu również wpływa na obniżenie emisji spalin.

Z kolei wzrost poziomów zanieczyszczeń na początku tygodnia może być spowodowany wzrostem ruchu komunikacyjnego i aktywności przemysłowej po weekendzie. Po weekendzie wielu ludzi powraca do pracy, a fabryki i zakłady rozpoczynają swoją działalność, co prowadzi do wzrostu emisji zanieczyszczeń. Dodatkowo, zjawisko takie jak tzw. "efekt poniedziałkowy" może wpływać na wzrost poziomów zanieczyszczeń, gdy ludzie wracają do codziennych aktywności po weekendzie.

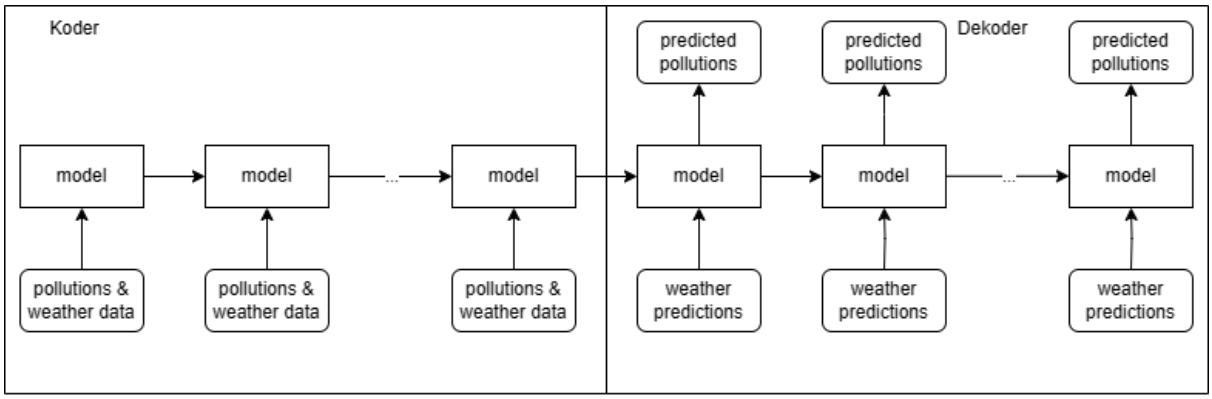
Najwyższe poziomy zanieczyszczeń często obserwuje się w środę lub czwartek. Może to wynikać z kumulacji zanieczyszczeń w atmosferze przez kilka dni. Emisje z różnych źródeł w poprzednich dniach tygodnia mogą się kumulować i osiągać szczyt w połowie tygodnia.

7. Predykcja

7.1 Opis modelu

Do predykcji modelu użyliśmy sieci rekurencyjnych sieci neuronowych oraz komórek LSTM.

Architektura modelu prezentuje się następująco:



Rys. 7.1.1. Architektura modelu predykcyjnego

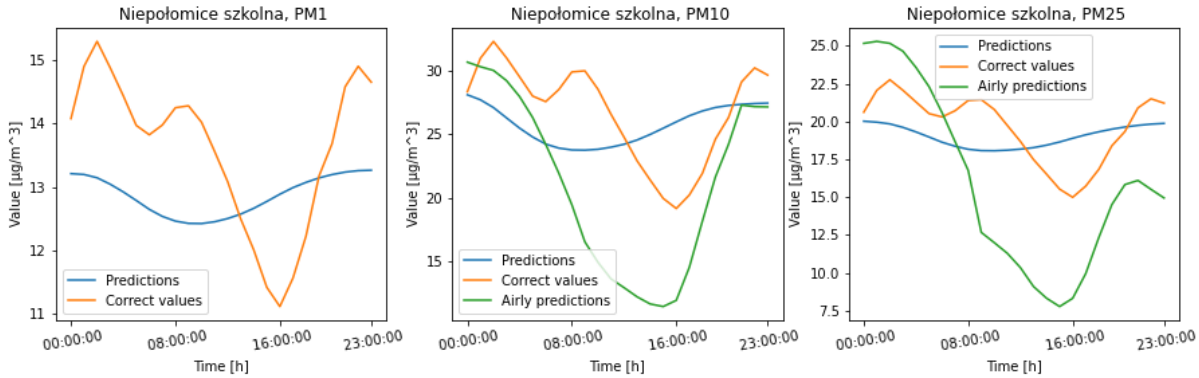
Model składa się z dwóch elementów:

- kodera - który przyjmuje znane wartości zanieczyszczeń oraz pogody z poprzednich godzin, które wykorzysta potem do predykcji nowych wartości
- dekodera - który zwraca przewidywane poziomy zanieczyszczeń dla poszczególnych godzin, dekoder w naszym przypadku przyjmuje dodatkowo przewidywane wartości pogodowe w celu lepszego dostosowania się do danych.

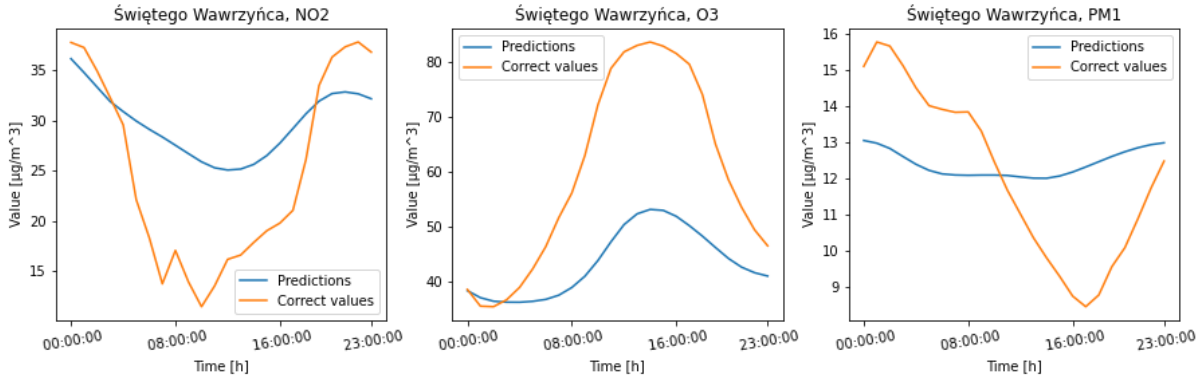
Umożliwienie dodania przewidywanych wartości pogodowych polepsza działanie modelu, gdyż model nie musi samemu martwić się o predykcję pogody, tylko może skorzystać z bardziej wiarygodnych źródeł. Pozwala to mu skupić się na prognozowaniu zanieczyszczeń powietrza.

7.2 Wyniki predykcji

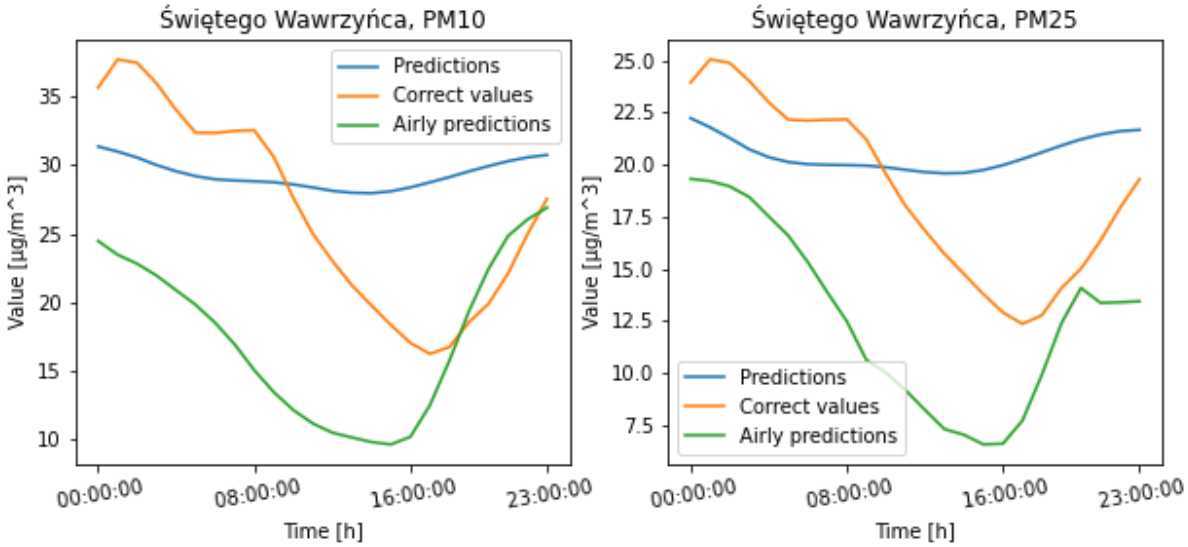
Na poniższych wykresach znajdują się wykresy zanieczyszczeń wraz z predykcją naszego modelu oraz Airly dla zanieczyszczeń PM10 oraz PM25.



Rys. 7.2.1. Zmiany poziomu zanieczyszczeń dla stacji Niepołomice szkolna dla wartości PM1, PM10, PM2.5 wraz z przewidzianymi wartościami oraz wartościami przewidywanymi przez Airly dla PM10 oraz PM2.5



Rys. 7.2.2. Zmiany poziomu zanieczyszczeń dla stacji Świętego Wawrzyńca dla zanieczyszczeń: NO2, O3 oraz PM1 wraz z przewidzianymi wartościami



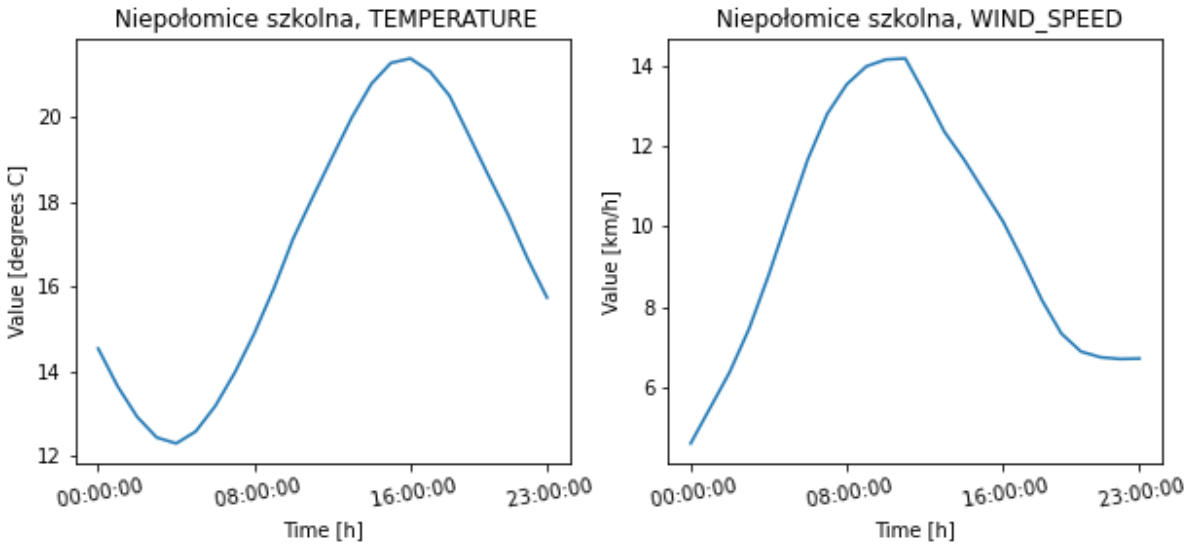
Rys. 7.2.3. Zmiany poziomu zanieczyszczeń dla stacji Świętego Wawrzyńca dla zanieczyszczeń PM10, PM2.5 wraz z przewidzianymi wartościami oraz wartościami przewidywanymi przez Airly

Z rysunków 7.2.1-3 widać, że zarówno nasz model jak i Airly popełnia zauważalne błędy. Błędy pojawiają się zazwyczaj w momencie, w którym poziom zanieczyszczenia

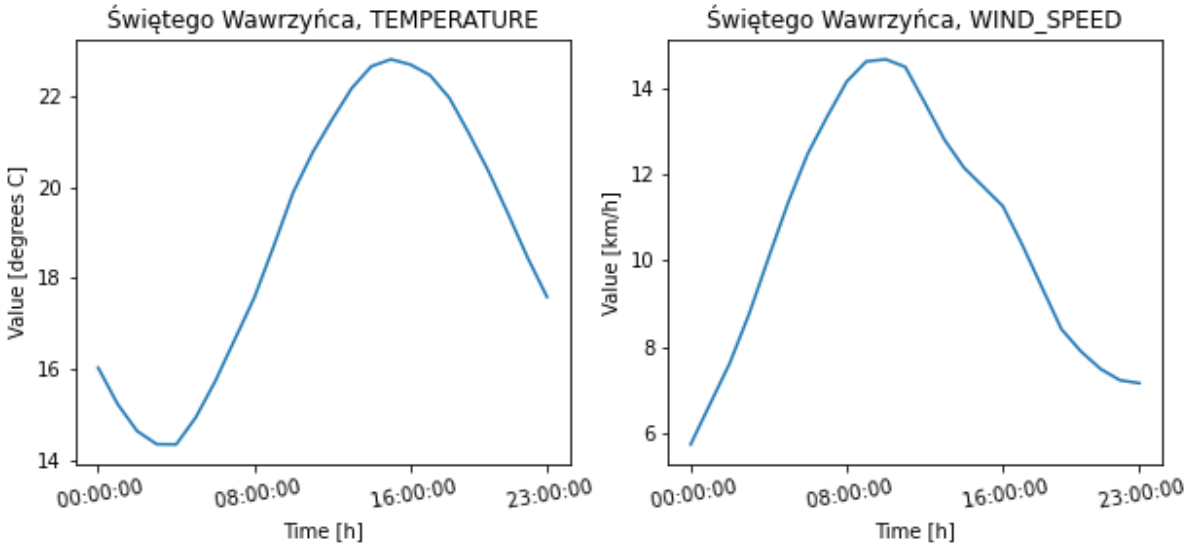
gwałtownie wzrósł (jak dla Świętego Wawrzyńca, O3). Inną przyczyną jest początkowy błąd jak dla (Świętego Wawrzyńca, PM2.5).

Dla większości przypadków błąd nie różni się znacząco od poprawnych wartości. Kolejną obserwacją jest fakt, że nasze predykcje są dokładniejsze jeśli wykonujemy je dla krótszego czasu.

Głównymi powodami utrudniającymi predykcje zanieczyszczeń są zmieniające się warunki pogodowe jak wiatr mające znaczący wpływ na poziom zanieczyszczeń.



Rys. 7.2.4. Zmiana temperatury oraz prędkości wiatru dla stacji Niepołomice Szkolna



Rys. 7.2.5. Zmiana temperatury oraz prędkości wiatru dla stacji Niepołomice Szkolna

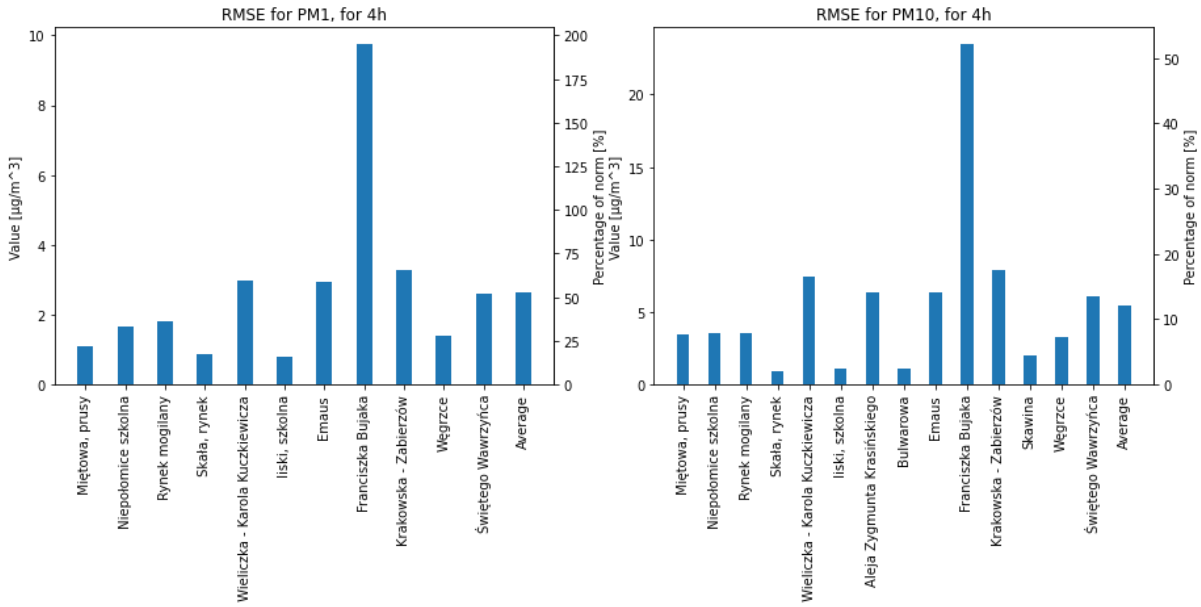
Porównując wartości temperatury dla stacji Świętego Wawrzyńca, możemy zauważyć, że wzrost temperatury spowodował spadek wartości zanieczyszczeń dla pyłów zawieszonych w powietrzu.

W tym przypadku również temperatura oraz prędkość wiatru jest podobna dla powyższych stacji (choć dla stacji w centrum Krakowa temperatura jest o 2 stopnie wyższa z powodu ulokowania w centrum miasta).

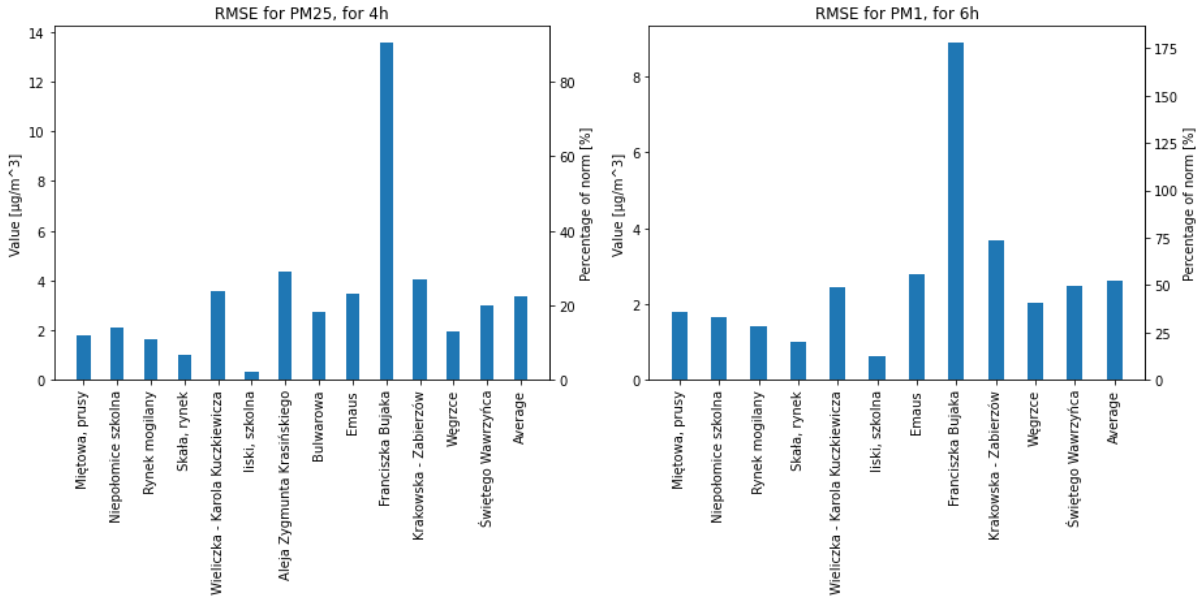
Ponieważ warunki pogodowe mają istotny wpływ na poziom zanieczyszczeń, dlatego powinny być one uwzględniane w modelu.

7.3 Analiza błędu predykcji

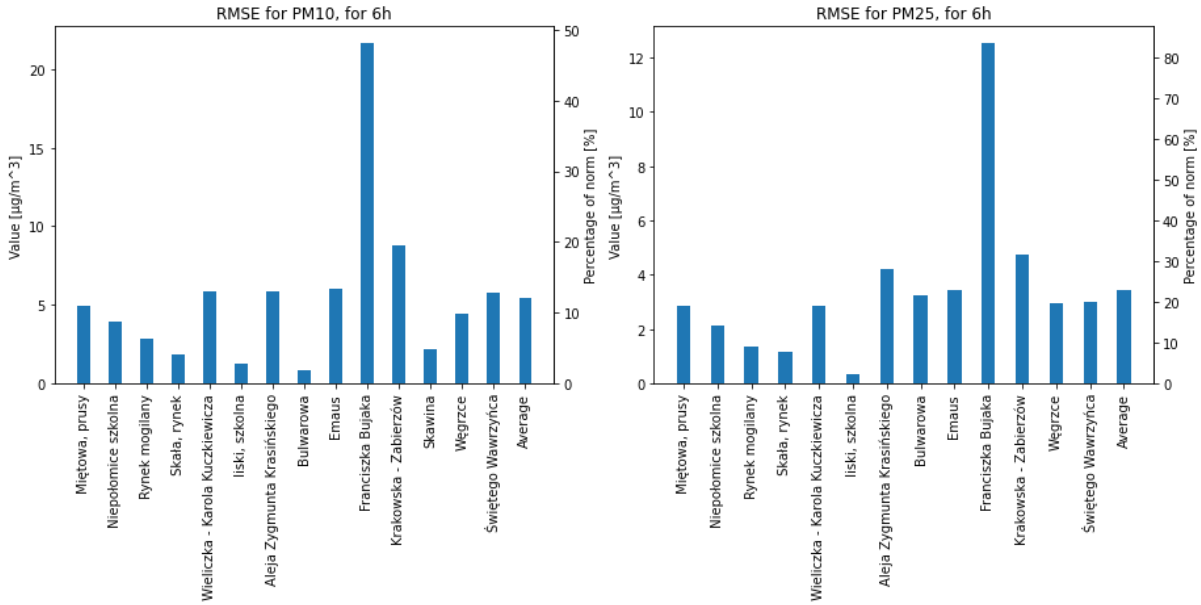
Na poniższych wykresach znajdują się wartości błędu średniokwadratowego dla badanych stacji:



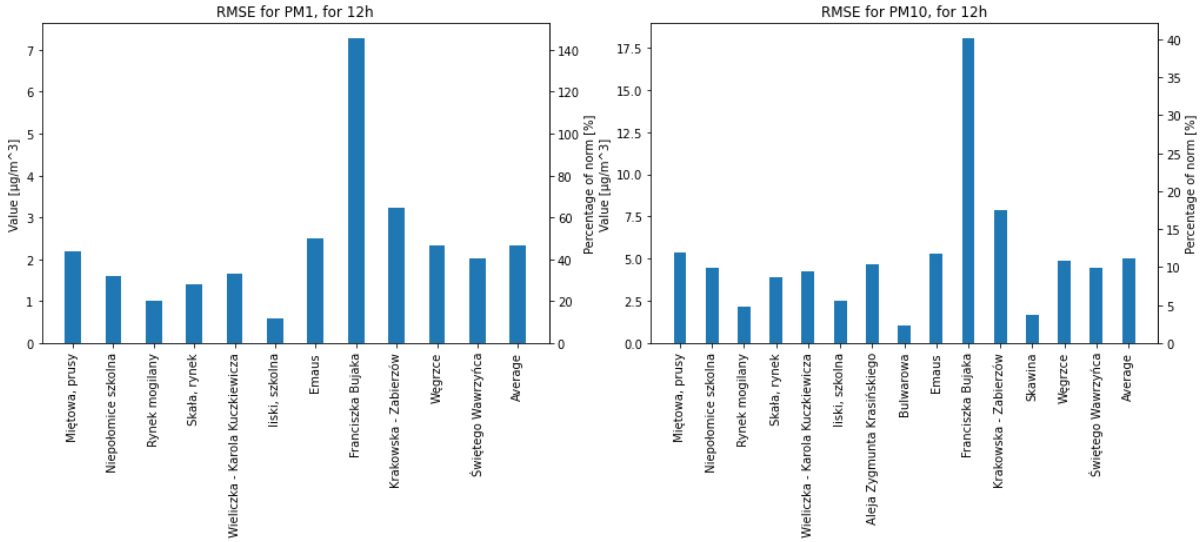
Rys. 7.3.1. Błąd średniokwadratowy modelu dla zanieczyszczeń PM25 oraz PM1 dla wybranych stacji dla 4 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń



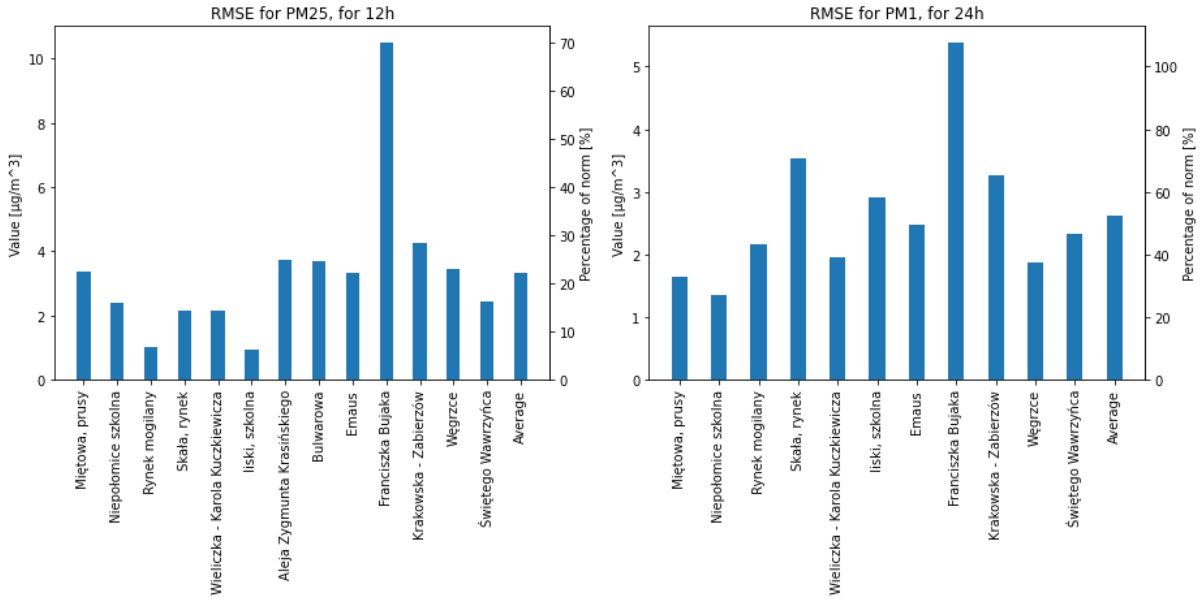
Rys. 7.3.2. Błąd średniokwadratowy modelu dla zanieczyszczeń PM25 dla 4 godzinnej predykcji oraz PM1 dla 6 godzinnej predykcji dla wybranych stacji wraz z porównaniem z normą zanieczyszczeń



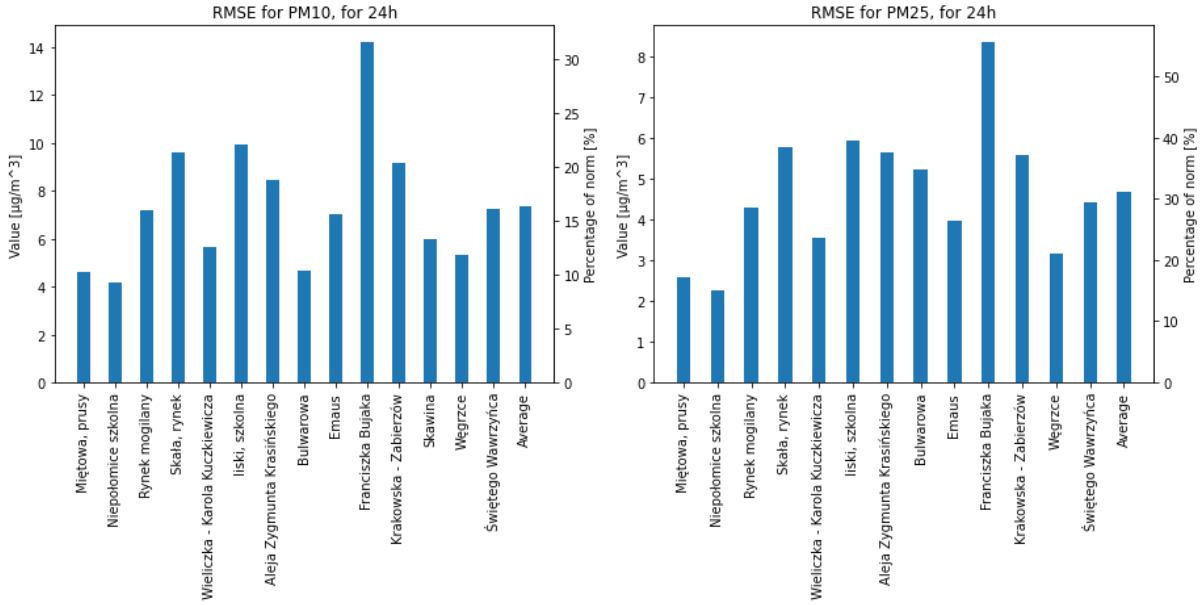
Rys. 7.3.3. Błąd średniokwadratowy modelu dla zanieczyszczeń PM10 oraz PM25 dla wybranych stacji dla 6 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń



Rys. 7.3.4. Błąd średniokwadratowy modelu dla zanieczyszczeń PM1 oraz PM10 dla wybranych stacji dla 12 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń



Rys. 7.3.5. Błąd średniokwadratowy modelu dla zanieczyszczeń PM25 dla 12 godzinnej predykcji oraz PM1 dla 24 godzinnej predykcji dla wybranych stacji wraz z porównaniem z normą zanieczyszczeń



Rys. 7.3.6 Błąd średniokwadratowy modelu dla zanieczyszczeń PM10 oraz PM24 dla wybranych stacji dla 24 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń

Możemy zaobserwować, że dla stacji Franciszka Bujaka obecny model zwraca o wiele gorsze wyniki niż dla pozostałych stacji (w szczególności dla 4-godzinnych pomiarów - dla 24 błąd zmniejszył się).

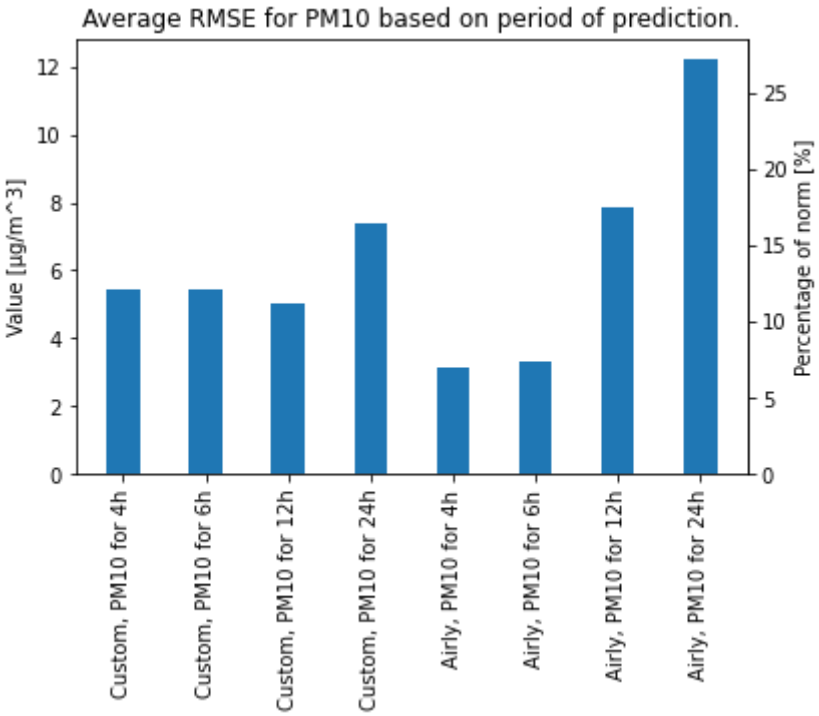
Dla pozostałych stacji błąd jest podobny i przykładowo dla 4 godzinnych predykcji dla PM10 waha się od kilku do kilkunastu procent, aby wzrosnąć do 10-25% dla 24-godzinnej predykcji.

Porównując błędy modelu w porównaniu do dopuszczalnej normy dla danego zanieczyszczenia możemy zauważyć że największe błędy obserwujemy dla PM1 i sięgają one kilkudziesięciu procent - jest to częściowo spowodowane tym że norma jest bardziej

restrykcyjna oraz zanieczyszczenie to osiąga mniejsze wartości - z tego powodu taki sam błąd bezwzględny jest dużo większym błędem względnym.

7.4 Porównanie błędów z Airly

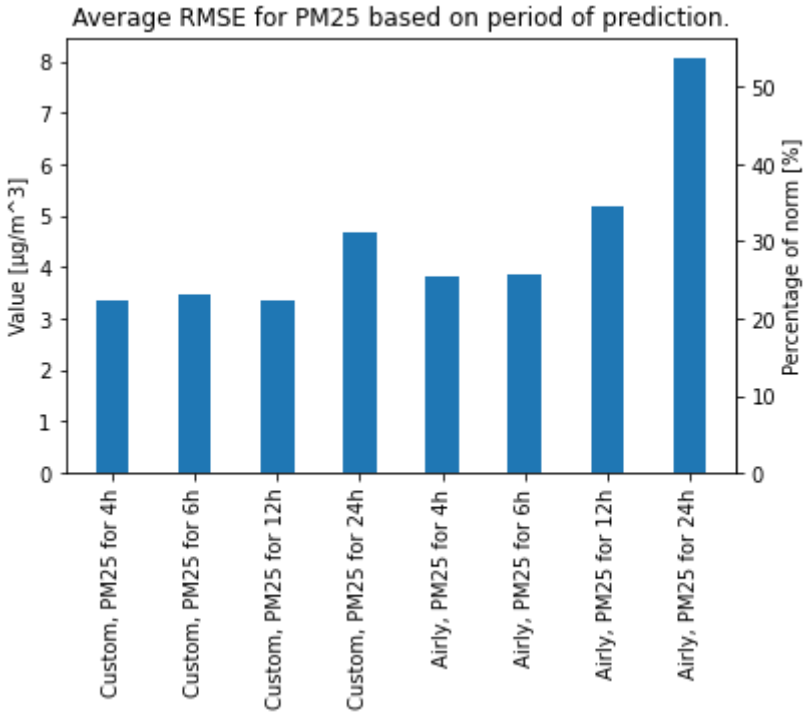
Na poniższych wykresach znajduje się porównanie błędów pomiaru dla różnych czasów predykcji wraz z porównaniem z błędami popełnianymi przez Airly dla analogicznego okresu czasu.



Rys. 7.4.1. Porównanie błędów Airly oraz modelu dla różnych czasów predykcji dla zanieczyszczenia PM10

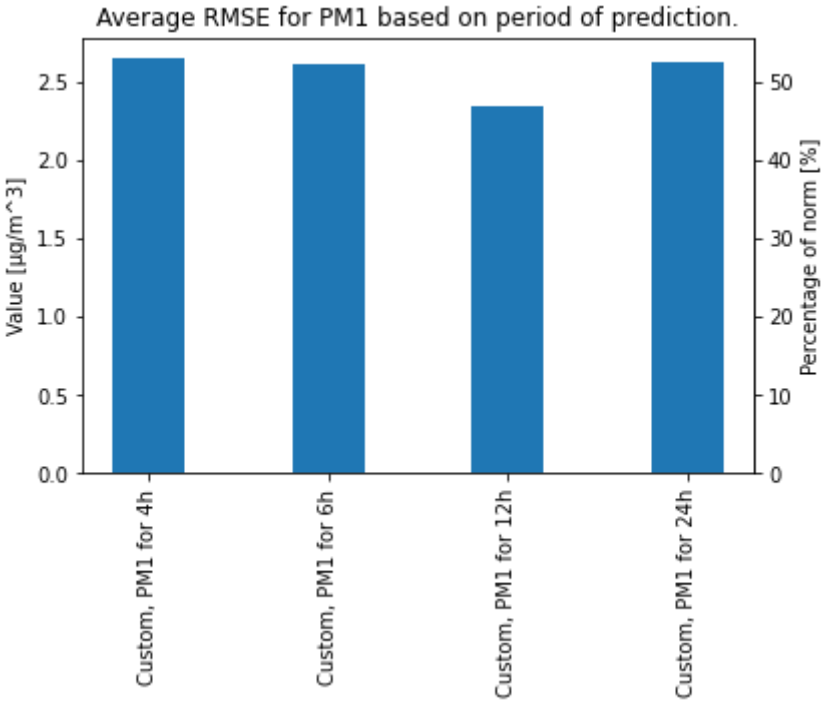
Możemy zauważyć, że nasz model zaczyna zwracać większe błędy dopiero dla prognozy 24-godzinnej, może to być spowodowane wysokim błędem dla stacji Franciszka Bujaka, który następnie malał. Dla Airly widzimy, że im większy czas predykcji tym większy błąd otrzymujemy.

Porównując nasz model z Airly możemy dojść do wniosku, że osiągamy gorsze wyniki dla prognoz krótkoterminowych, a lepsze dla długoterminowych.



Rys. 7.4.2. Porównanie błędów Airly oraz modelu dla różnych czasów predykcji dla zanieczyszczenia PM25

Dla PM25 nasz model zwraca wartości nieco lepsze od Airly dla każdego okna czasowego. Największa różnica jest dla prognoz długoterminowych.



Rys. 7.4.3. Porównanie modelu dla różnych czasów predykcji dla zanieczyszczenia PM25

Dla PM1 nie mamy znaczących różnic pomiędzy różnymi oknami czasowymi. Jest to spowodowane prawdopodobnie faktem, że dla 4-godzinnej predykcji stacja Franciszka Bujaka ma błąd wynoszący 200% normy, a dla 24 wartość ta spada do 100%.

8. Podsumowanie

Pierwsze dane pochodzące ze stron rządowych (powietrze.gios.gov.pl oraz danepubliczne.imgw.pl) okazały się niewystarczające, po pierwsze nie obejmowały wystarczającej liczny stacji w okolicach Krakowa oraz nie zawierały wystarczającej ilości danych meteorologicznych. Mimo wszystko podjęto próbę analizy długoterminowej. Do reszty analiz skorzystaliśmy z danych oferowanych przez platformę Airly z której na bieżąco pobieraliśmy wszystkie potrzebne dane (pogodowe oraz poziomów zanieczyszczeń). W celu polepszenia predykcji modelu zaczęliśmy pobierać prognozowane dane pogodowe z serwisu wetter.com.

Patrząc na dane całoroczne dotyczące zanieczyszczeń możemy dojść do wniosku, że największy poziom zanieczyszczeń jest w zimie. W tym okresie obserwujemy wiele krótkich skoków zanieczyszczeń, dla których wartości przekraczają normę nawet kilkudziesięciokrotnie. Średni poziom zanieczyszczeń w zimie jest również dużo większy niż w lecie. W miesiącach letnich nie obserwujemy również skoków zanieczyszczeń.

Warunki atmosferyczne odgrywają istotną rolę w determinowaniu poziomu zanieczyszczeń. Wyższe temperatury często korelują z niższymi poziomami zanieczyszczeń, co może wynikać z mniejszego zapotrzebowania na ogrzewanie. Wyższe prędkości wiatru również przyczyniają się do zmniejszenia zanieczyszczeń poprzez rozpraszanie ich w powietrzu. Zauważono również, że dla niektórych kierunków wiatru średnie poziomy zanieczyszczeń bywają wyższe niż dla innych. Jednocześnie wyższa wilgotność może przyczynić się do redukcji zanieczyszczeń.

Porównując Kraków oraz przygraniczne miejscowości możemy dość do wniosku, że nie ma dużych stałych różnic w poziomie pyłów zawieszonych w powietrzu. W poszczególnych miesiącach można było zauważyć obszary o większym zanieczyszczeniu niż pozostałe (jak na Rys. 6.6.1), jednak w kolejnych miesiącach zanieczyszczenie nie różniło się od pozostałych. Dla NO2 możemy zauważyć rzecz odwrotną - w centrum Krakowa poziom zanieczyszczenia jest nawet kilkukrotnie razy większy niż w miejscowościach przygranicznych.

Analizując zmiany poziomów zanieczyszczenia na przestrzeni różnych okresów, zaobserwowano istnienie trendu dobowego i tygodniowego. Na przestrzeni dobry zanieczyszczenie osiąga wyższe wartości w godzinach wieczornych i nocnych, a mniejsze w ciągu dnia. Związane jest to najpewniej z okresami większego i mniejszego ruchu samochodowego, a także bardziej intensywną działalnością przemysłową w godzinach popołudniowych. Dodatkowo, czynniki atmosferyczne również mają swój wpływ na poziomy zanieczyszczeń w ciągu doby.

Na przestrzeni tygodnia poziom zanieczyszczeń zaczyna spadać w piątek, po czym wzrasta od poniedziałku. Fakt ten jest skutkiem zmniejszonej aktywności przemysłowej i mniejszemu ruchowi komunikacyjnemu podczas weekendu oraz powrotem do pracy w poniedziałek.

Stworzony w ramach projektu model predykcji zanieczyszczeń pozwala na predykcję do 24 godzin w przód na podstawie wcześniejszych pomiarów oraz prognozowanej pogody (wiatr, temperatura). Model osiąga błędy zbliżone do Airly będącego konkurencyjnym modelem, udostępniający bezpłatnie swoje predykcje. Model osiąga lepsze wyniki dla krótszych okresów predykcyjnych a nagłe i nieprzewidziane zmiany pogody mogą powodować większe błędy.

9. Spis rysunków i tabel

9.1. Rysunki

Rys 3.1. Architektura oprogramowania użyta do eksploracji 5

Rys 3.2. Schemat ERD wykorzystanej bazy 7

Rys 3.3. Schemat ERD porzuconych danych wykorzystanej bazy 8

Rys. 4.1. Wybrane stacje pokazane kolorem żółtym oraz niebieskim do analizy z serwisu powietrze.gios.gov.pl 10

Rys. 4.2. Histogram dostępnych pomiarów dla stacji o granulacji 24h 13

Rys. 4.3. Histogram dostępnych pomiarów dla stacji o granulacji 1h 13

Rys 4.4. Okres, w którym stacje raportowały zanieczyszczenia 15

Rys 4.5. Histogram ilości pomiarów dla stacji w okresie sty 2019 - sty 2023 dla zanieczyszczenia PM10 15

Rys. 4.6. Zmiana poziomu zanieczyszczenia PM10 w 8. tygodniu 2021 roku dla stacji Kraków, ul. Bujaka 16

Rys. 4.7. Zmiana poziomu zanieczyszczenia PM10 w 27. tygodniu 2021 roku dla stacji Kraków, ul. Bujaka 17

Rys. 4.8. Średni poziom zanieczyszczenia PM10 w tygodniu zimowym 2021 roku dla stacji Kraków, ul. Bujaka wraz z percentylami 25% i 75% 18

Rys. 4.9. Zmiana poziomu zanieczyszczenia PM10 w 2021 roku dla stacji Kraków, ul. Bujaka 18

Rys. 6.1.1. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Miętowa, Prusy 20

Rys. 6.1.2. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Miętowa, Prusy 20

Rys. 6.1.3. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Miętowa, Prusy 21

Rys. 6.1.4. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Zabierzów 21

Rys. 6.1.5. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Zabierzów 21

Rys. 6.1.6. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Zabierzów 22

Rys. 6.1.7. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Wieliczka, 22

Rys. 6.1.8. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Wieliczka, 22

Rys. 6.1.9. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Wieliczka, Kuczkiewicza 23
Rys. 6.1.10. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru dla stacji Emaus 23
Rys. 6.1.11. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru dla stacji Emaus 23
Rys. 6.1.12. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Węgrzce 24
Rys. 6.1.13. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Węgrzce 24
Rys. 6.1.14. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Węgrzce 24
Rys. 6.1.15. Poziom zanieczyszczenia NO2 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca 25
Rys. 6.1.16. Poziom zanieczyszczenia O3 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca 25
Rys. 6.1.17. Poziom zanieczyszczenia PM1 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca 25
Rys. 6.1.18. Poziom zanieczyszczenia PM10 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca 26
Rys. 6.1.19. Poziom zanieczyszczenia PM2.5 Kalman, prędkość wiatru, temperatura dla stacji Św. Wawrzyńca 26
Rys. 6.2.1. Korelacje pomiarów stacji Niepołomice, Szkolna 27
Rys. 6.2.2. Korelacje pomiarów stacji Skała, Rynek 28
Rys. 6.2.3. Korelacje pomiarów stacji Wieliczka, Kuczkiewicza 28
Rys. 6.2.4. Korelacje pomiarów stacji Franciszka Bujaka 29
Rys. 6.2.5. Korelacje pomiarów stacji Węgrzce 29
Rys. 6.2.6. Korelacje pomiarów stacji Świętego Wawrzyńca 30
Rys. 6.3.1. Wykresy skrzypcowe pomiarów PM1 31
Rys. 6.3.2. Wykresy skrzypcowe pomiarów PM10 32
Rys. 6.3.3. Wykresy skrzypcowe pomiarów PM2.5 32
Rys. 6.4.1. Dystrybucja kierunków wiatru 33
Rys. 6.4.2. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru północnego 33
Rys. 6.4.3. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru północno-wschodniego 33
Rys. 6.4.4. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru wschodniego 34
Rys. 6.4.5. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla południowo-wschodniego 34

Rys. 6.4.6. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru południowego 34
Rys. 6.4.7. Śr ednie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru południowo-zachodniego 34
Rys. 6.4.8. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru zachodniego 34
Rys. 6.4.9. Średnie zanieczyszczenia PM1, PM10, PM2.5 dla wiatru północno-zachodniego 34
Rys. 6.5.1. Zmiana średniego poziomu zanieczyszczeń PM1, PM2.5, PM10 w podziale na Kraków i okolice 36
Rys. 6.6.1. Mapy średniego zanieczyszczenia PM2.5 na przestrzeni marzec -czerwiec według stacji 37
Rys. 6.6.2. Mapy średniego zanieczyszczenia NO2 na przestrzeni marzec -czerwiec według stacji 38
Rys. 6.7.1. Zmiana średniego zanieczyszczenia pyłów PM1, PM2.5, PM10 na przestrzeni doby dla wybranych stacji 39
Rys. 6.7.2. Zmiana średniego zanieczyszczenia pyłów PM1, PM2.5, PM10 na przestrzeni tygodnia dla wybranych stacji 40
Rys. 7.1.1. Architektura modelu predykcyjnego 41
Rys. 7.2.1. Zmiany poziomu zanieczyszczeń dla stacji Niepołomice szkolna dla wartości PM1, PM10, PM2.5 wraz z przewidzianymi wartościami oraz wartościami przewidywanymi przez Airly dla PM10 oraz PM2.5 42
Rys. 7.2.2. Zmiany poziomu zanieczyszczeń dla stacji Świętego Wawrzyńca dla zanieczyszczeń: NO2, O3 oraz PM1 wraz z przewidzianymi wartościami 42
Rys. 7.2.3. Zmiany poziomu zanieczyszczeń dla stacji Świętego Wawrzyńca dla zanieczyszczeń PM10, PM2.5 wraz z przewidzianymi wartościami oraz wartościami przewidywanymi przez Airly 42
Rys. 7.2.4. Zmiana temperatury oraz prędkości wiatru dla stacji Niepołomice Szkolna 43
Rys. 7.2.5. Zmiana temperatury oraz prędkości wiatru dla stacji Niepołomice Szkolna 43
Rys. 7.3.1. Błąd średniokwadratowy modelu dla zanieczyszczeń PM25 oraz PM1 dla wybranych stacji dla 4 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń 44
Rys. 7.3.2. Błąd średniokwadratowy modelu dla zanieczyszczeń PM25 dla 4 godzinnej predykcji oraz PM1 dla 6 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń 44
Rys. 7.3.3. Błąd średniokwadratowy modelu dla zanieczyszczeń PM10 oraz PM25 dla wybranych stacji dla 6 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń 45
Rys. 7.3.4. Błąd średniokwadratowy modelu dla zanieczyszczeń PM1 oraz PM10 dla wybranych stacji dla 12 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń 45
Rys. 7.3.5. Błąd średniokwadratowy modelu dla zanieczyszczeń PM25 dla 12 godzinnej predykcji oraz PM1 dla godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń 46

Rys. 7.3.6 Błąd średniokwadratowy modelu dla zanieczyszczeń PM10 oraz PM24 dla wybranych stacji dla 24 godzinnej predykcji wraz z porównaniem z normą zanieczyszczeń 46
Rys. 7.4.1. Porównanie błędów Airly oraz modelu dla różnych czasów predykcji dla zanieczyszczenia PM10 47
Rys. 7.4.2. Porównanie błędów Airly oraz modelu dla różnych czasów predykcji dla zanieczyszczenia PM25 48
Rys. 7 .4.3. Porównanie modelu dla różnych czasów predykcji dla zanieczyszczenia PM25 48

9.2. Tabele

Tabela 4.1. Wybrane stacje do analizy z serwisu powietrze.gios.gov.pl wraz z dostępnymi pomiarami zanieczyszczenia 11

Tabela 5.1. Wybrane stacje do analizy z serwisu Airly 19

Tabela 6.5.1. Korelacja średniego zanieczyszczenia PM1 w Krakowie i okolicach 35

Tabela 6.5.2. Korelacja średniego zanieczyszczenia PM10 w Krakowie i okolicach 35

Tabela 6.5.3. Korelacja średniego zanieczyszczenia PM2.5 w Krakowie i okolicach 35

Fragment 1. Fragment kodu punktu wejścia programu pobierającego dane z api Airly 5