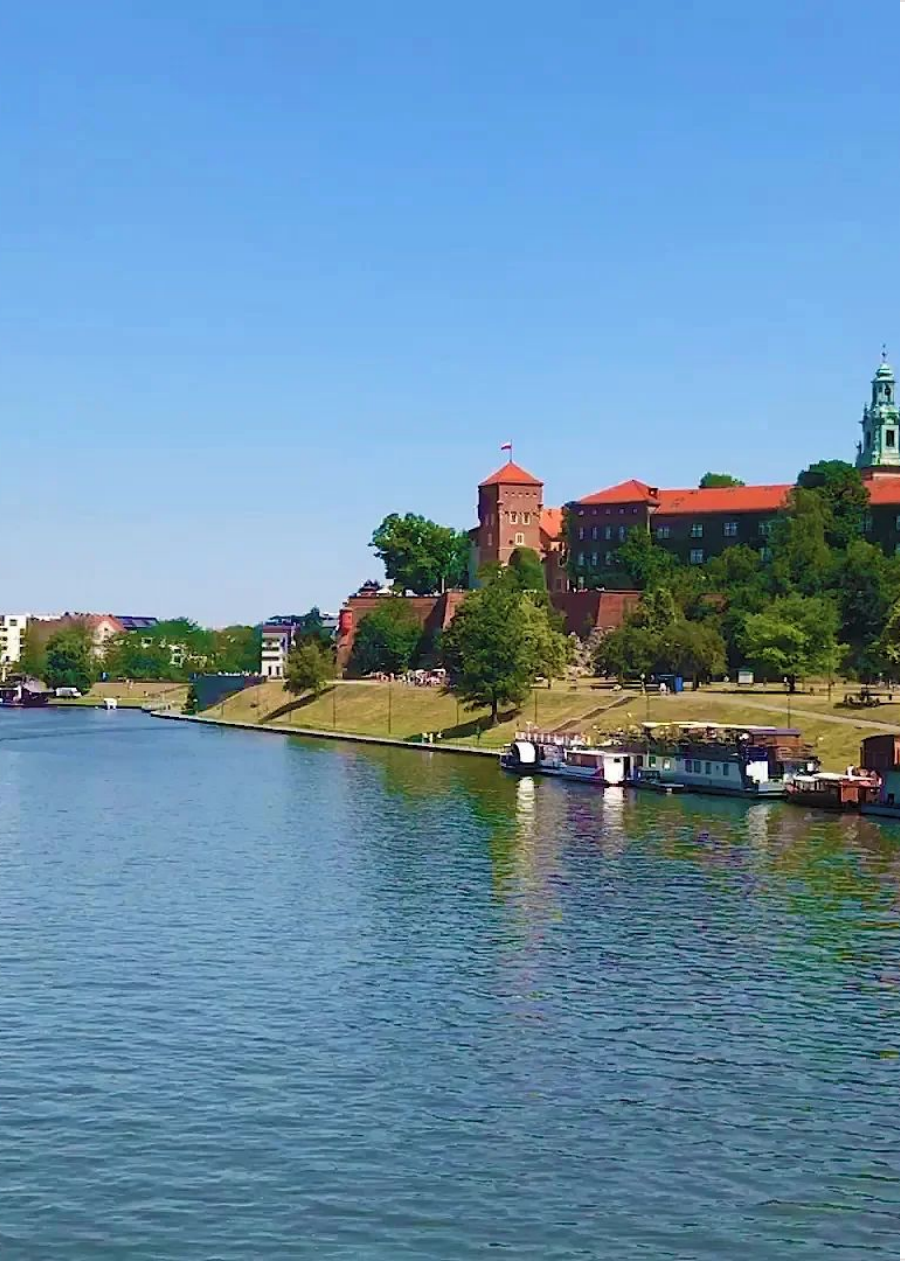


Przewidywanie poziomu zanieczyszczenia powietrza w Krakowie i okolicach na podstawie danych z czujników Airly i sytuacji pogodowej.

Milena Biernacka, Paweł Świder, Szymon Musiał

Cel projektu

1 Analiza jakości powietrza

Projekt ma na celu analizę jakości powietrza w Krakowie i okolicach na różnych skalach czasowych.

2 Wizualizacja danych

Dane te będą poddane statystycznej analizie i wizualizacji w celu zidentyfikowania trendów i wzorców zmian zanieczyszczenia w analizowanych obszarach.

3 Analiza wpływu czynników zewnętrznych

Projekt będzie badał również potencjalne wpływy czynników zewnętrznych, takich jak warunki pogodowe, na jakość powietrza w tych obszarach.

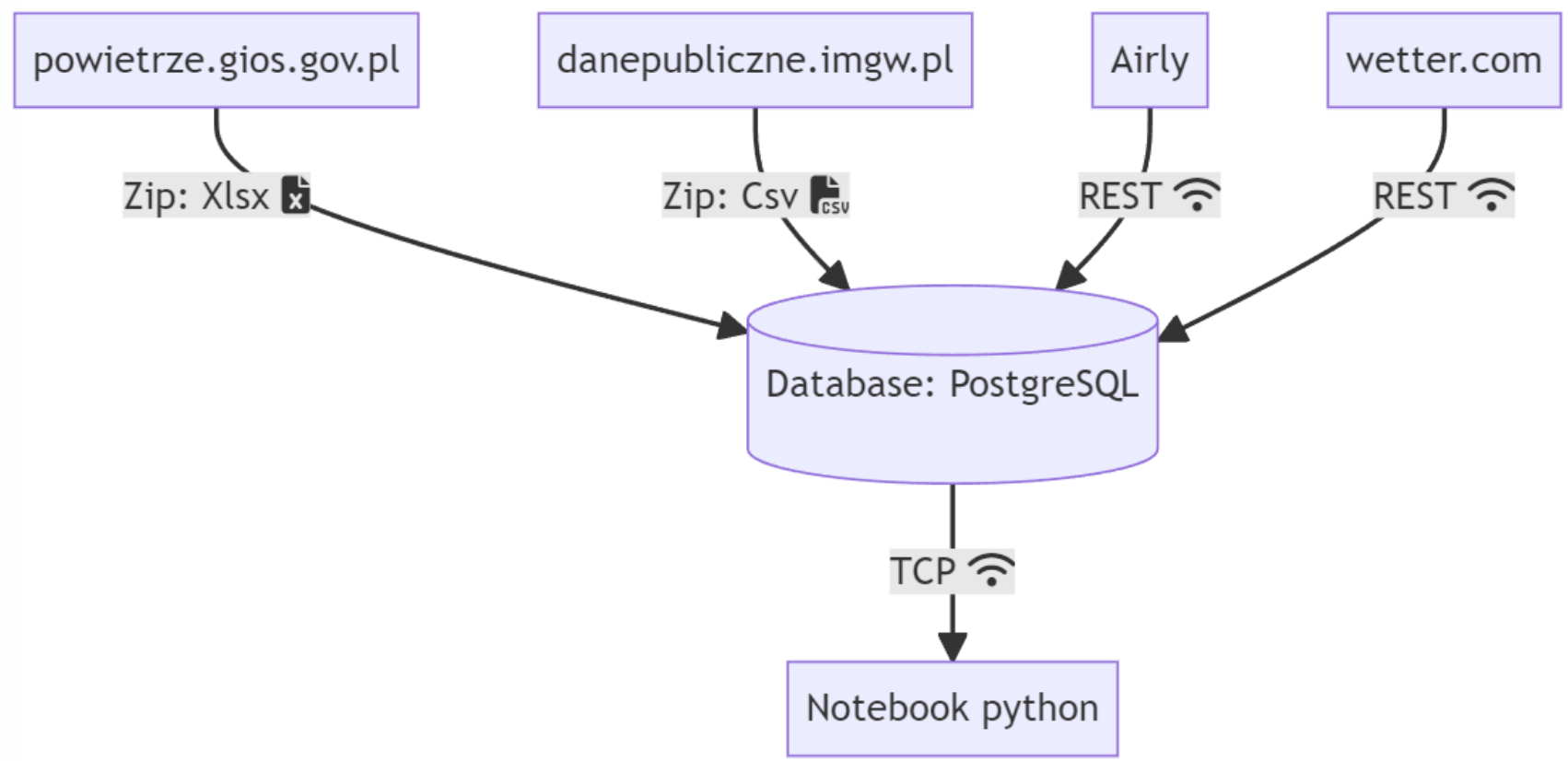
4 Opracowanie modelu predykcyjnego

Jednym z celów jest opracowanie modelu predykcyjnego poziomu zanieczyszczeń dla gmin i okolic Krakowa w różnych horyzontach czasowych.

Dwa podejścia do przetwarzania danych

Przetwarzanie danych historycznych z GIOS i IMGW

Używanie API Airly do pobierania aktualnych danych i prognoz pogodowych



Architektura oprogramowania użyta do eksploracji

Dane były dodawane asynchronicznie do bazy danych

Notebook Pythona komunikował się z serwerem bazy danych, pobierając przygotowane dane i wykorzystując je do analizy i przetwarzania.

Serwisy procesujące dane napisane w języku C#

Serwisy procesujące dane zostały napisane w języku C# z wykorzystaniem .Net 7.0.

Jak działa nasz program zbierania danych

Źródła danych

Cztery różne źródła danych

Konfiguracja

Odczytanie konfiguracji z pliku JSON lub zmiennych środowiskowych

Baza danych

Utworzenie kontekstu bazy danych i dedykowane repozytorium dla każdego programu

Przetwarzanie danych

Tworzenie repozytoriów i przetwarzanie danych wejściowych

Zapis do bazy danych

Dodawanie przetworzonych rekordów do bazy danych

Analiza zanieczyszczeń powietrza w Krakowie

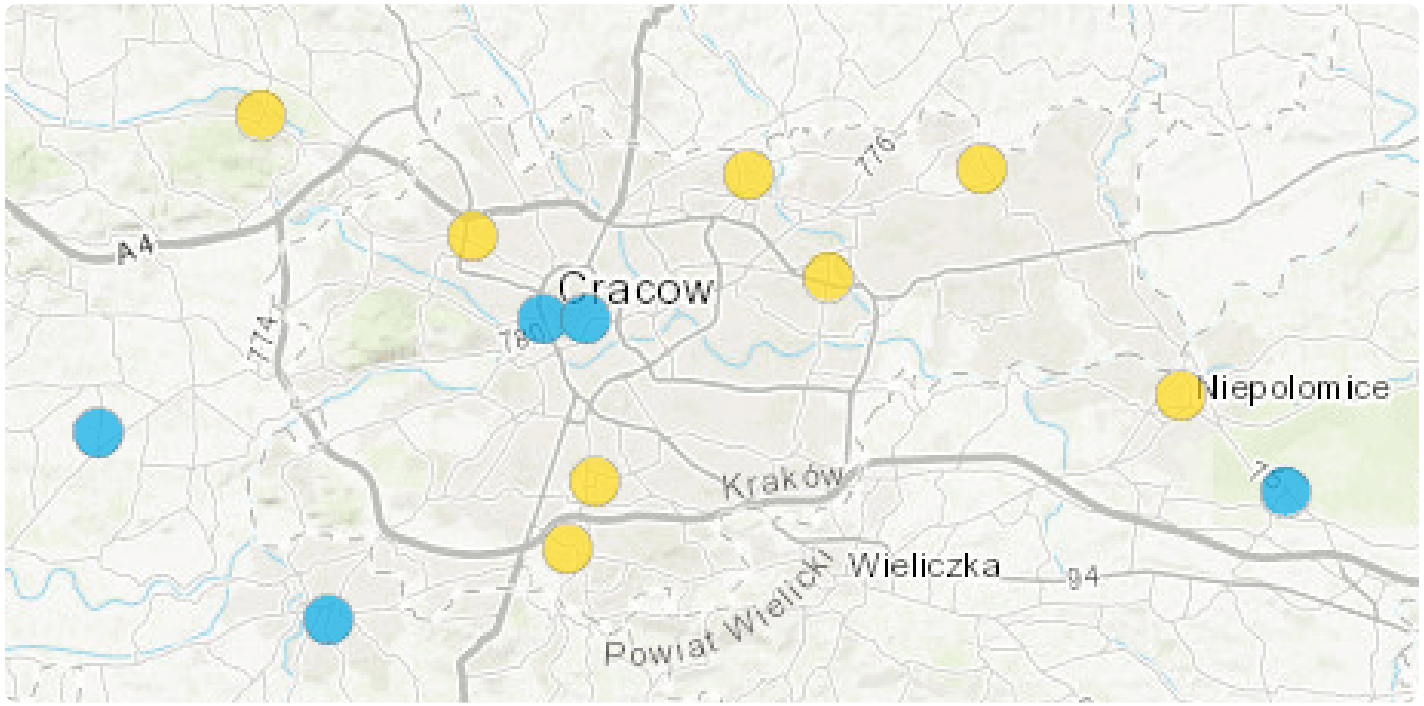
Stacje pomiarowe

Analiza skupiła się na stacjach pomiarowych wybranych w taki sposób, aby uwzględnić wpływ aglomeracji na centrum miasta i umożliwić dokładną analizę tego zjawiska.

Różnorodność zanieczyszczeń

Wybrane stacje charakteryzowały się różnym zakresem dostępnych mierzonych zanieczyszczeń.

Wybrane stacje



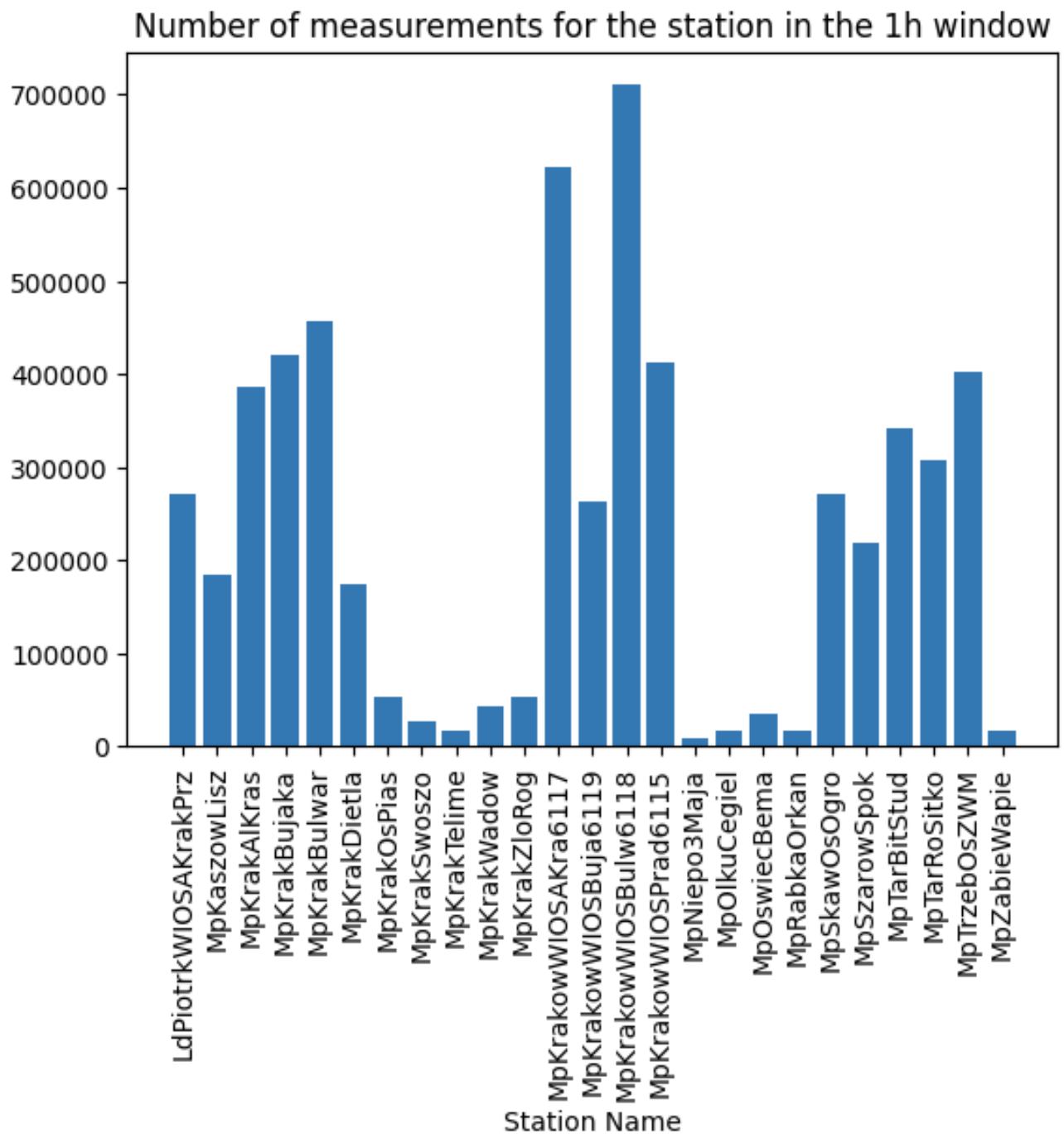
Serwisy

powietrze.gios.gov.pl

danepubliczne.imgw.pl

Dostępne dane o jakości powietrza ze strony powietrze.gios.gov.pl.

> Nie wszystkie stacje pomiarowe mają pomiary dla zanieczyszczeń takich jak PM10, PM2.5, O3.

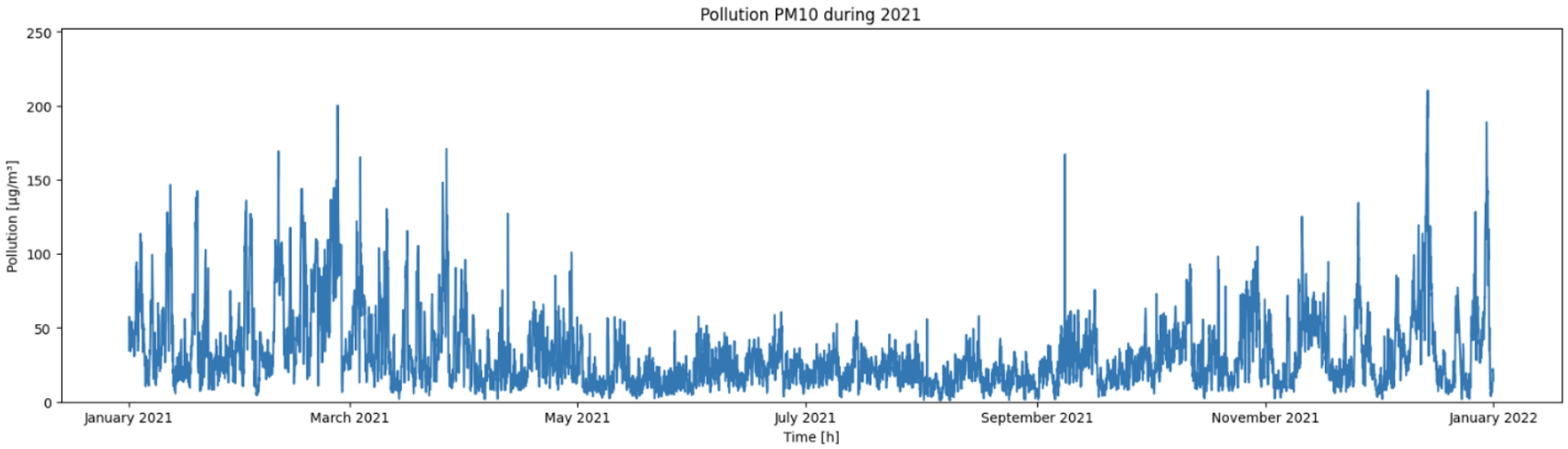


Dostępne dane meteorologiczne

Serwis danepubliczne.imgw.pl zapewnia różnorodne dane

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

Próby analizy



Miesiące zimowe

Duży poziom zanieczyszczeń oraz duże różnice tego poziomu.

Miesiące letnie

Niski i dużo bardziej stabilny poziom zanieczyszczeń.

Zakres czasu raportowanych danych

Stacje Monitorujące

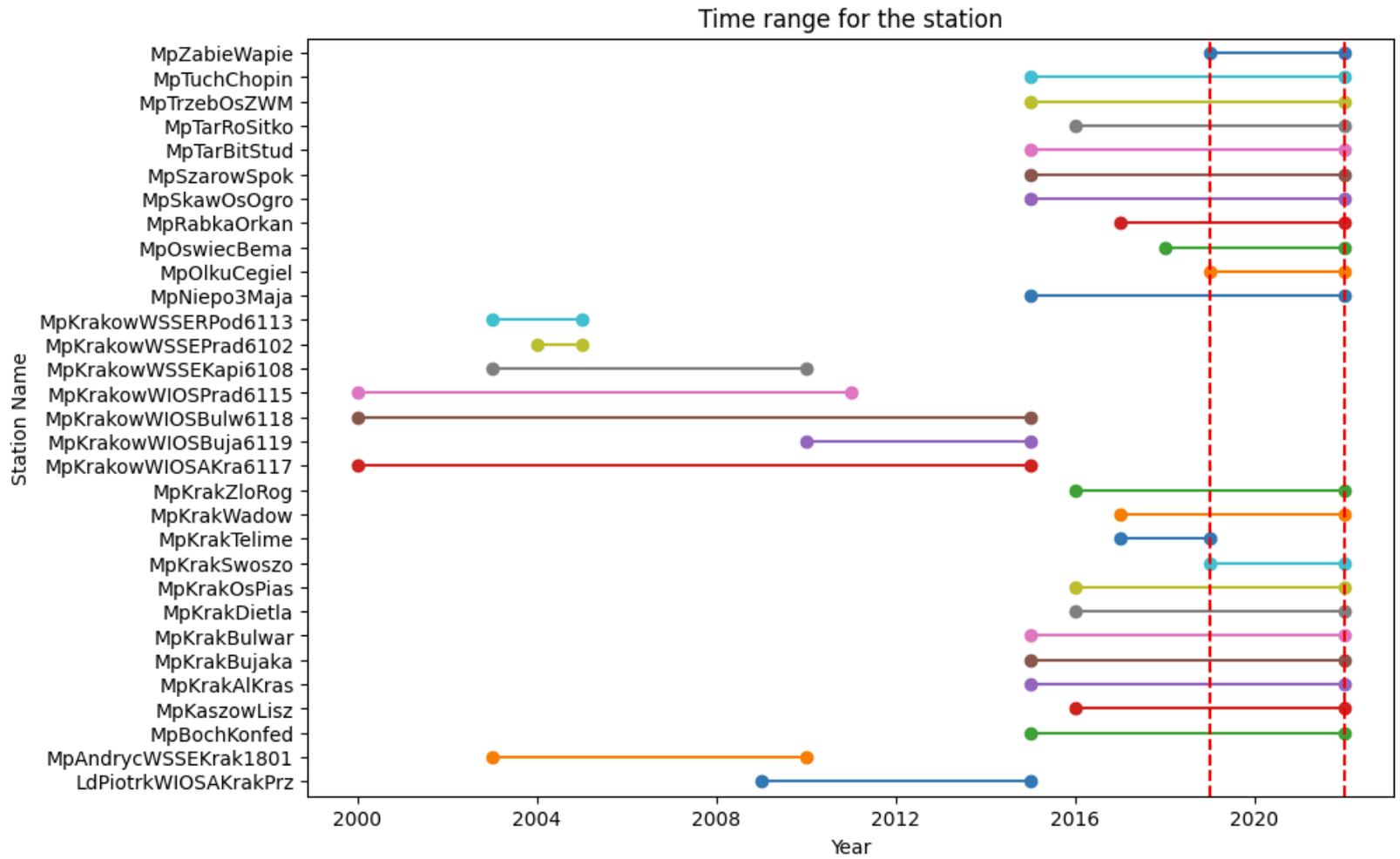
Region Małopolski

Okres

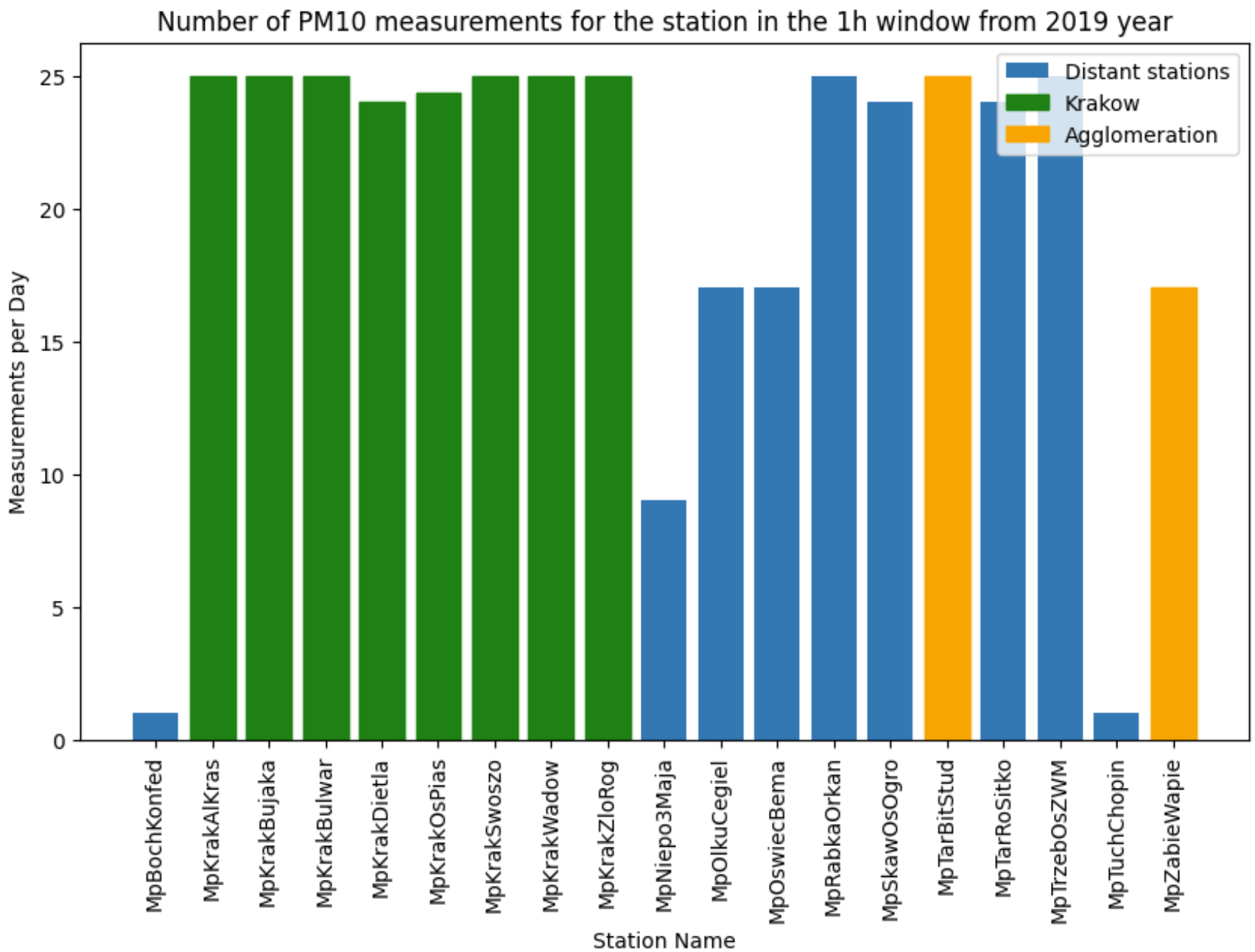
2000 - 2020

Braki Danych

Znacząca liczba, szczególnie w poprzedniej dekadzie



Wybrane stacje 2019+, PM10



Stacje oznaczone kolorem zielonym reprezentują stacje w Krakowie

Stacje znajdujące się w aglomeracji zostały oznaczone kolorem żółtym.

Natomiast pozostałe stacje z terenu Małopolski zostały oznaczone kolorem niebieskim.

Ze względu na brak dostępnych danych meteorologicznych w serwisie powietrze.gios.gov.pl, skorzystano z serwisu danepubliczne.imgw.pl

Wpływ zanieczyszczeń atmosferycznych na jakość powietrza jest ściśle powiązany z warunkami pogodowymi

W przypadku tego źródła w Krakowie raportowane są sytuacje meteorologiczne z dwóch stacji

Analiza danych

1 Braki w danych

Niedostateczna liczba danych ograniczyła możliwość pełnej oceny i wnioskowania na temat badanych zależności, co utrudniło osiągnięcie zamierzonych celów projektu.

2 Poszukiwanie alternatywnych źródeł danych

Zwrócono się z prośbą o dostęp do baz danych serwisu Airly. Niestety, w odpowiedzi otrzymano umowę o zbyt restrykcyjnych warunkach

3 Ostateczny wybór: Api Airly

Zdecydowano się więc ograniczyć pobieranie danych jedynie poprzez API serwisu Airly.

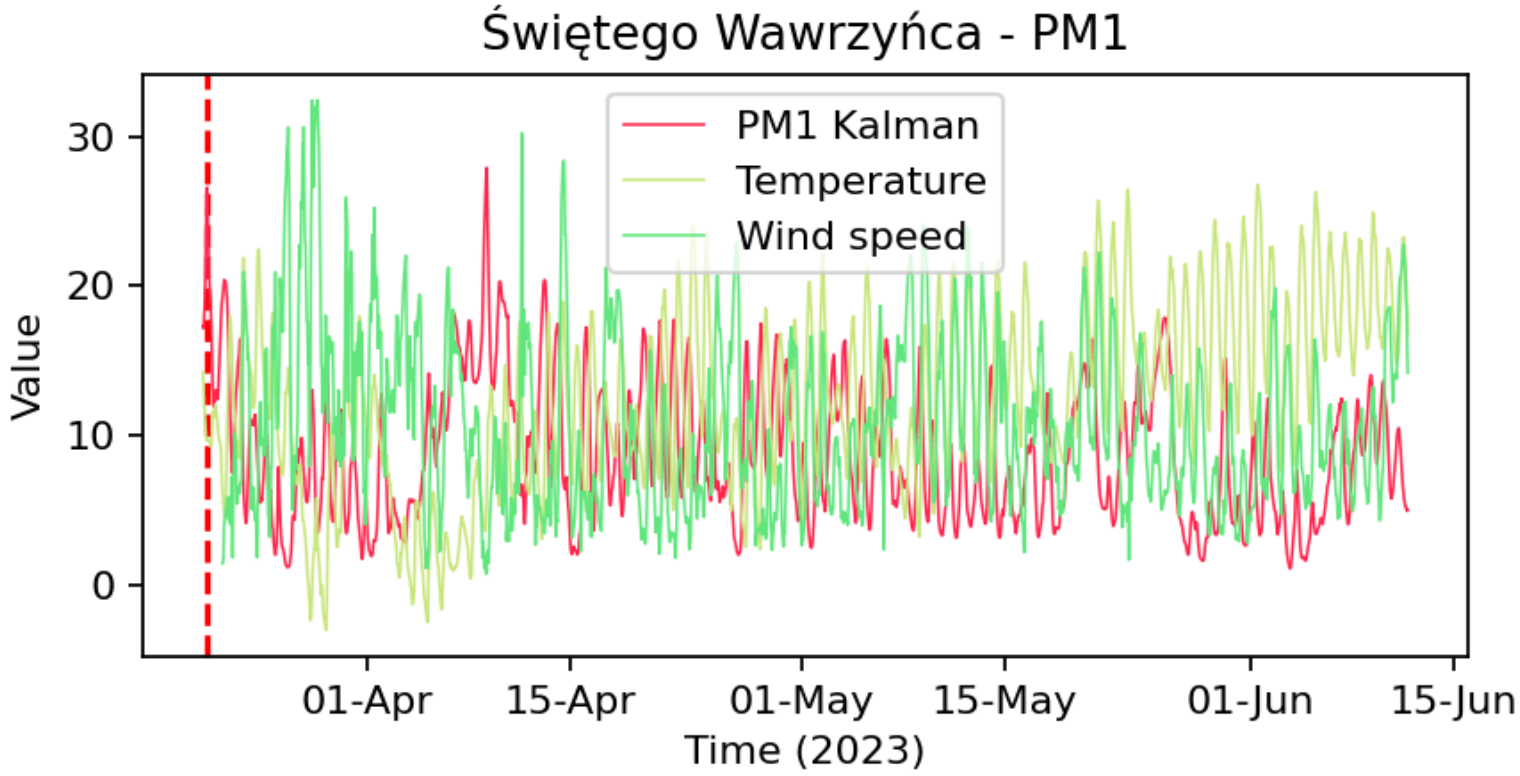
Ostateczne wybrane źródło danych: Api Airly

Wybrane stacje Airly oraz ich pomiary

- **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
- **Zabierzów, Krakowska -** NO2, O3, PM1, PM10, PM2.5
- **Wieliczka, Kuczkiewicza -** PM1, PM10, PM2.5
- **Węgrzce -** PM1, PM10, PM2.5
- **Prusy, Miętowa -** PM1, PM10, PM2.5
- **Niepołomice, Szkolna -** PM1, PM10, PM2.5
- **Mogilany, Rynek -** 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

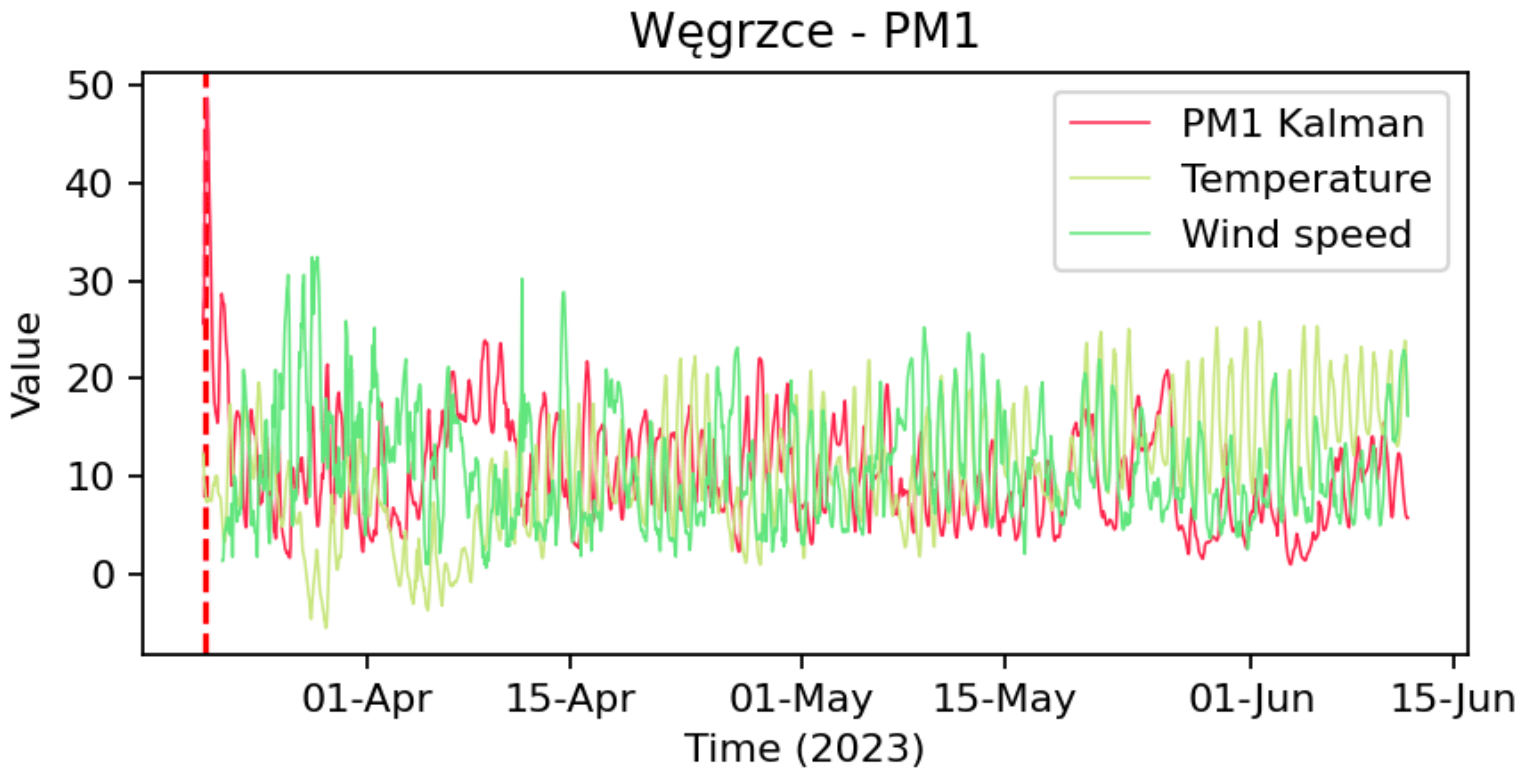
> Stacje oferują różne pomiary zanieczyszczeń (NO2, O3, PM1, PM10, PM25, CO, SO2) oraz dane pogodowe (temperaturę, ciśnienie, wilgotność). Odczyty są odczytami godzinnymi. Pierwszy odczyt pochodzi z dnia 20.03.2023 o godzinie 16:00:00.

Analiza danych z Airly



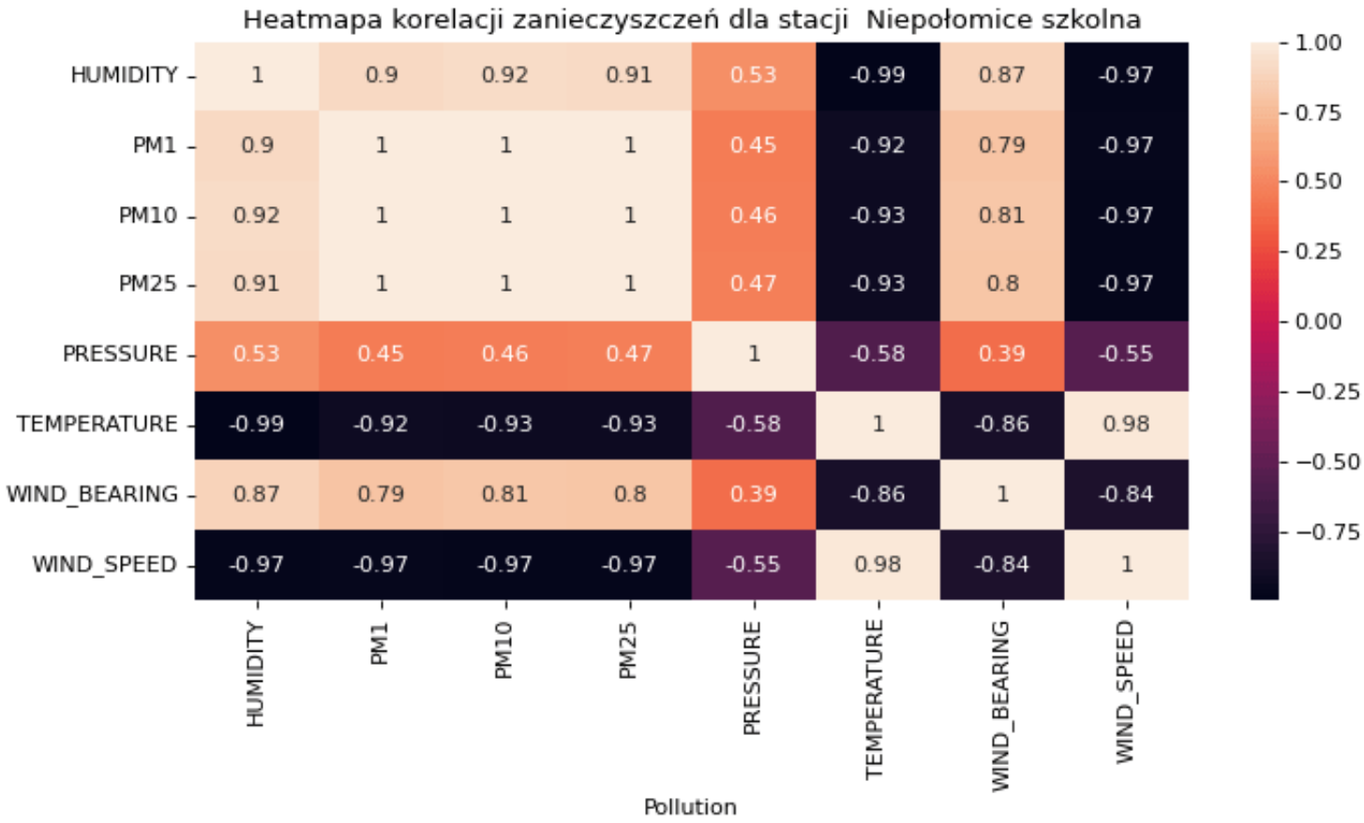
Wzrost temperatury wiąże się ze spadkiem poziomu zanieczyszczenia

Analiza danych z Airly



Większa prędkość wiatru skutkuje redukcją poziomu zanieczyszczenia

Korelacje



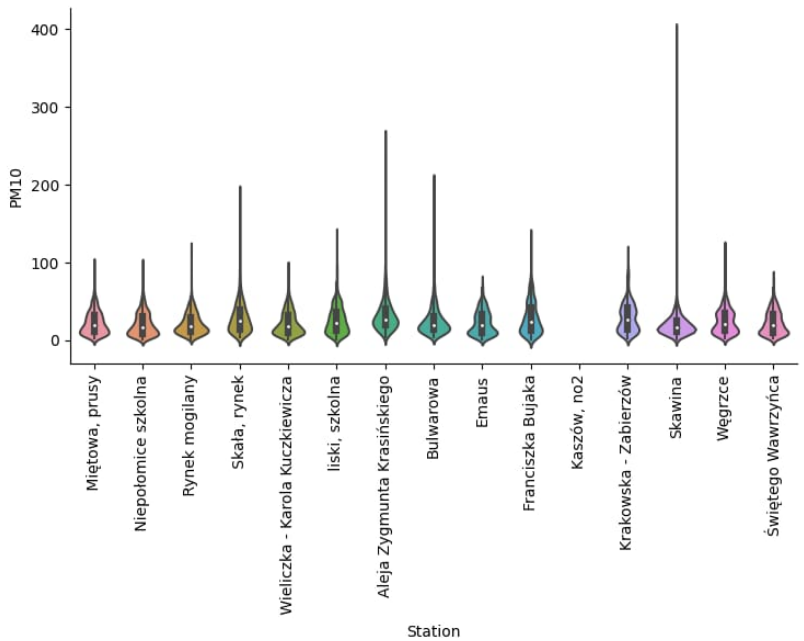
1 **Bliska -1 korelacja prędkości wiatru i temperatury z zanieczyszczeniami**

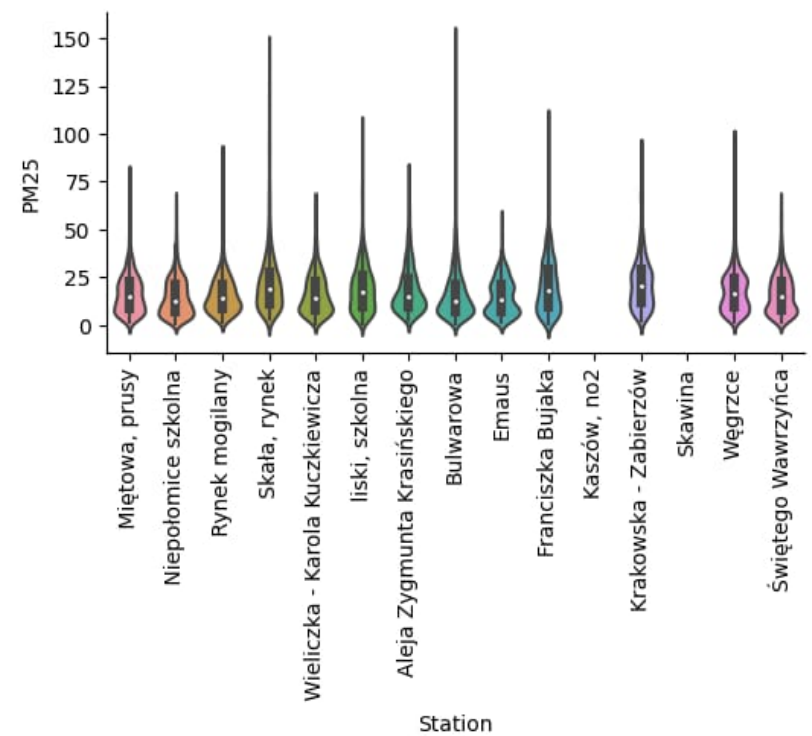
2 **Im większa wilgotność tym więcej zanieczyszczeń**

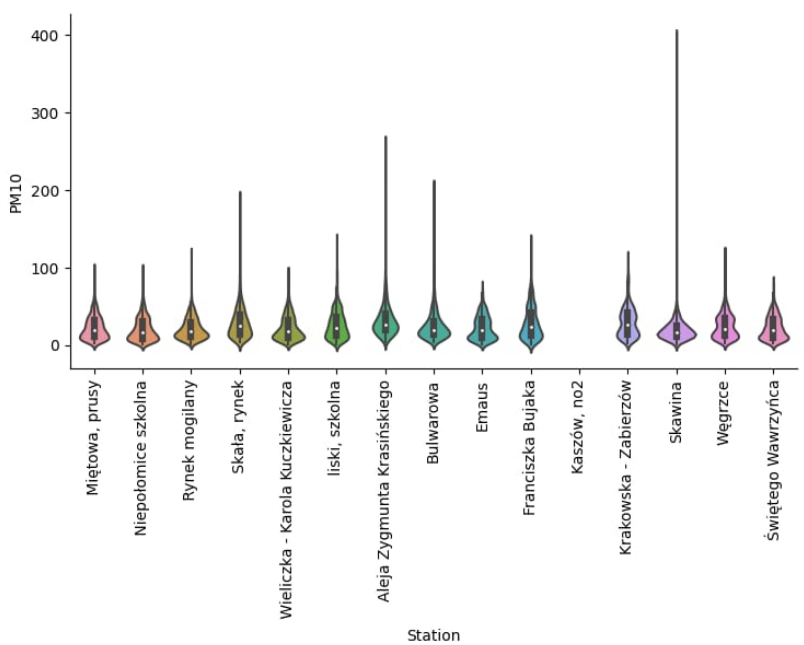
3 **Korelacja ciśnienia z zanieczyszczeniami przyjmuje różne poziomy**

4 **Dla stacji Skała, Rynek korelacje nie są tak dominujące jak dla innych stacji**

Wykresy skrzypcowe





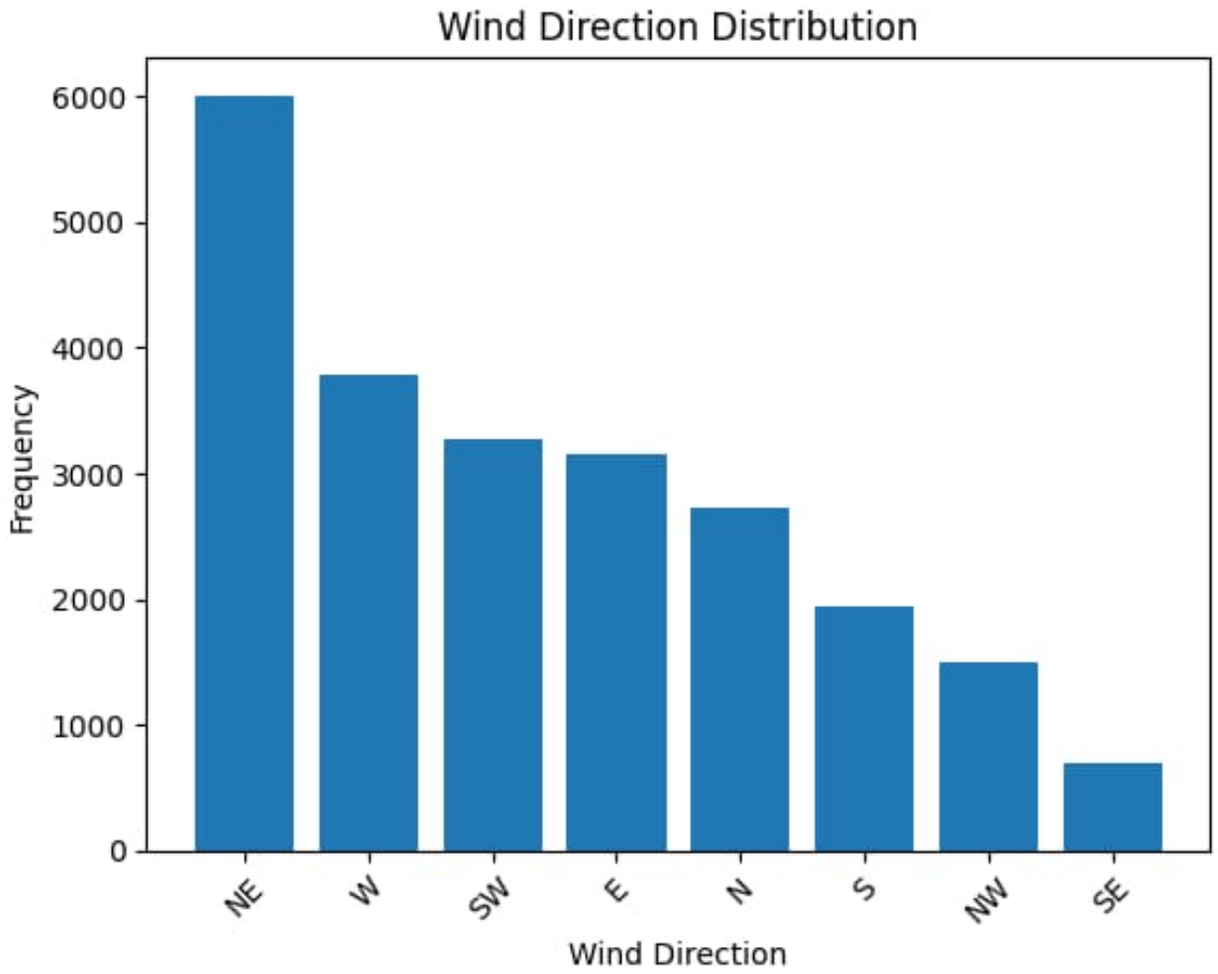


1 **Większe średnie zanieczyszczenia dla stacji:**

- Zabierzów, F. Bujaka, Skawina, Al. Z. Krasińskiego, Skała dla pyłów PM
- Kaszów dla O3
- Al. Z. Krasińskiego dla NO2

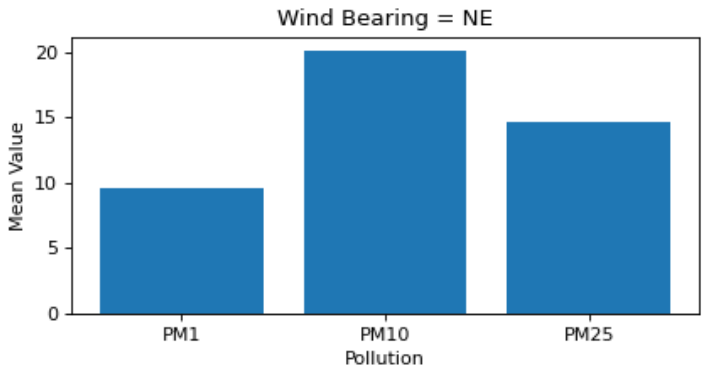
2 **Podobieństwo wykresów PM1 oraz P2.5**

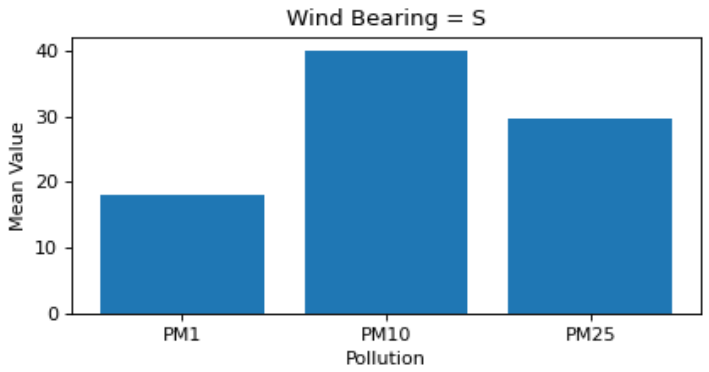
Dystrybucja wiatru



Najczęściej występującym kierunkiem wiatru był wiatr północno - wschodni.

Średnie PM dla kierunków wiatru



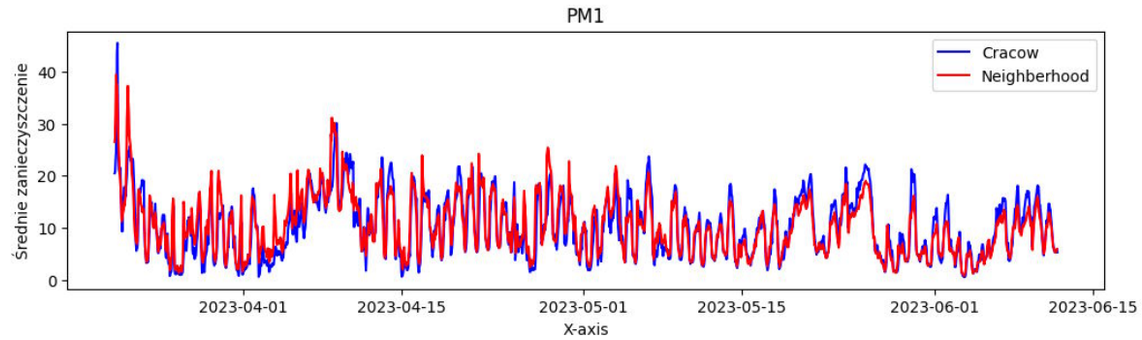


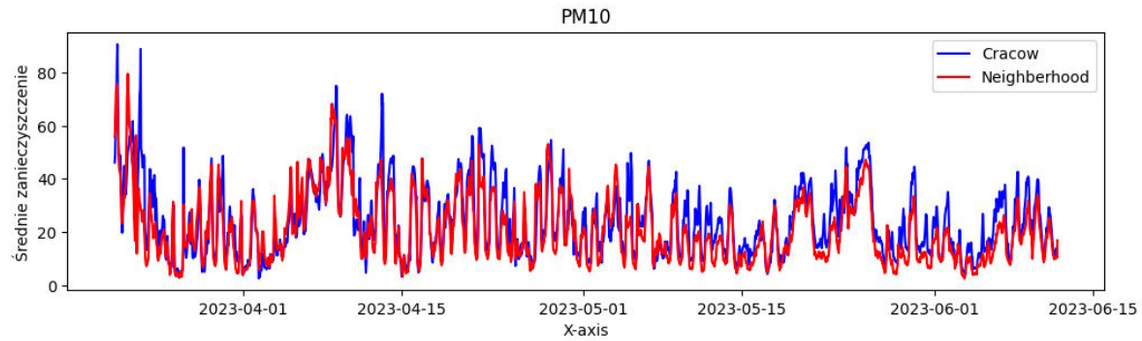
Kierunki wiatru z **południa** i **południowego-zachodu** mają wyższe średnie wartości zanieczyszczeń.

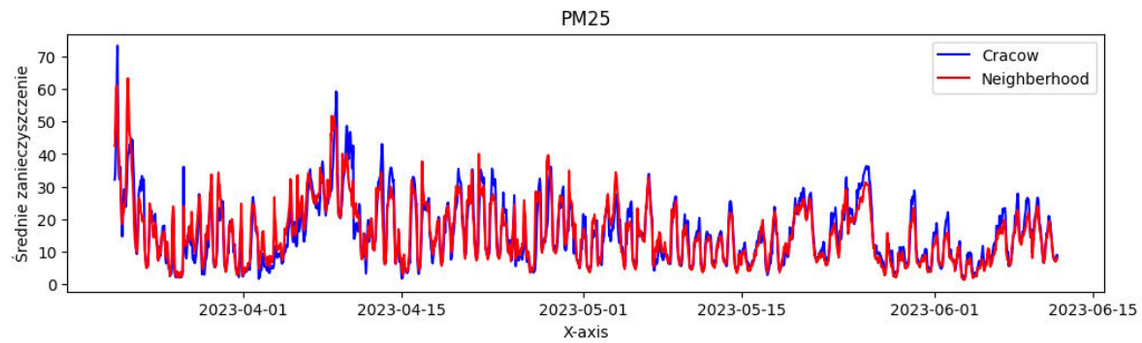
Podział na Kraków oraz okolice

Korelacje średnich pomiarów zanieczyszczeń w Krakowie i okolicach:

1. **PM1 - 90%**
2. **PM2.5 - 91%**
3. **PM10 - 88%**

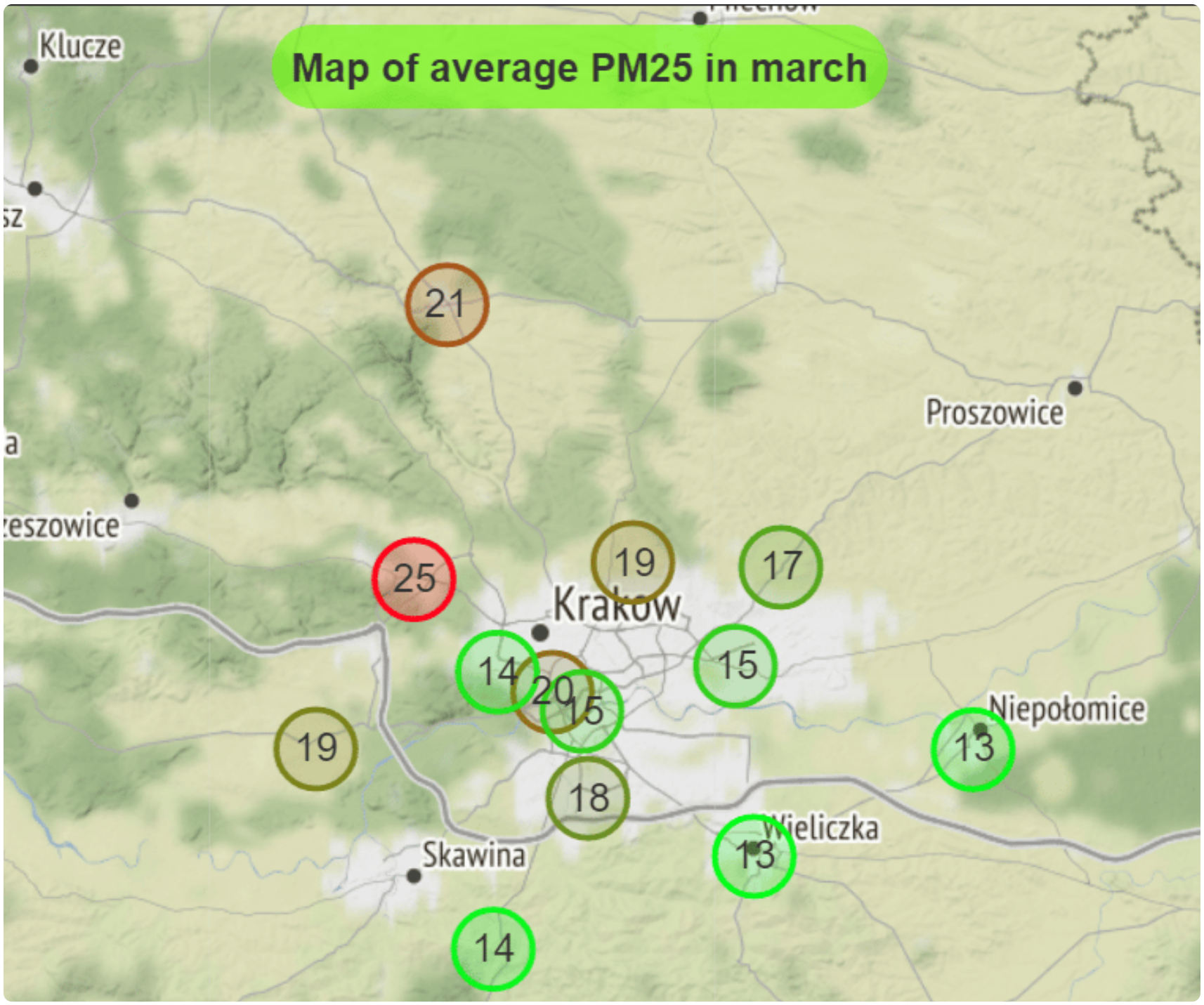






Zanieczyszczenie w Krakowie było większe niż w "Obwarzanku" w przeważającej części badanego okresu.

Mapy średniego miesięcznego zanieczyszczenia według stacji



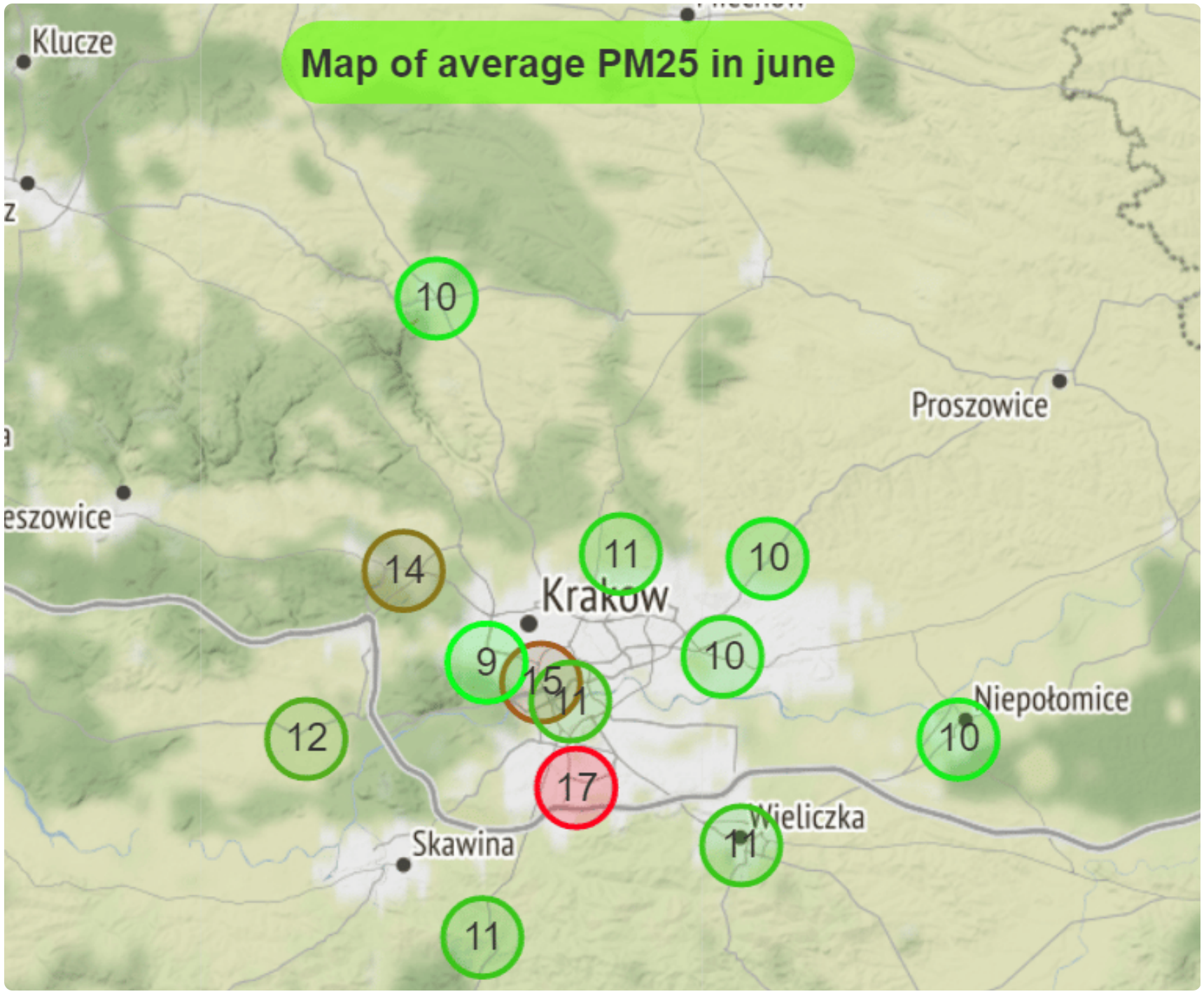
Średnie poziomy zanieczyszczeń

Dla wybranych stacji na przestrzeni marzec-czerwiec.

Wizualizacja wartości na mapach

Aby można było dostrzec czy stacja znajduje się w mieście czy w "Obwarzanku".

Mapy średniego miesięcznego zanieczyszczenia według stacji



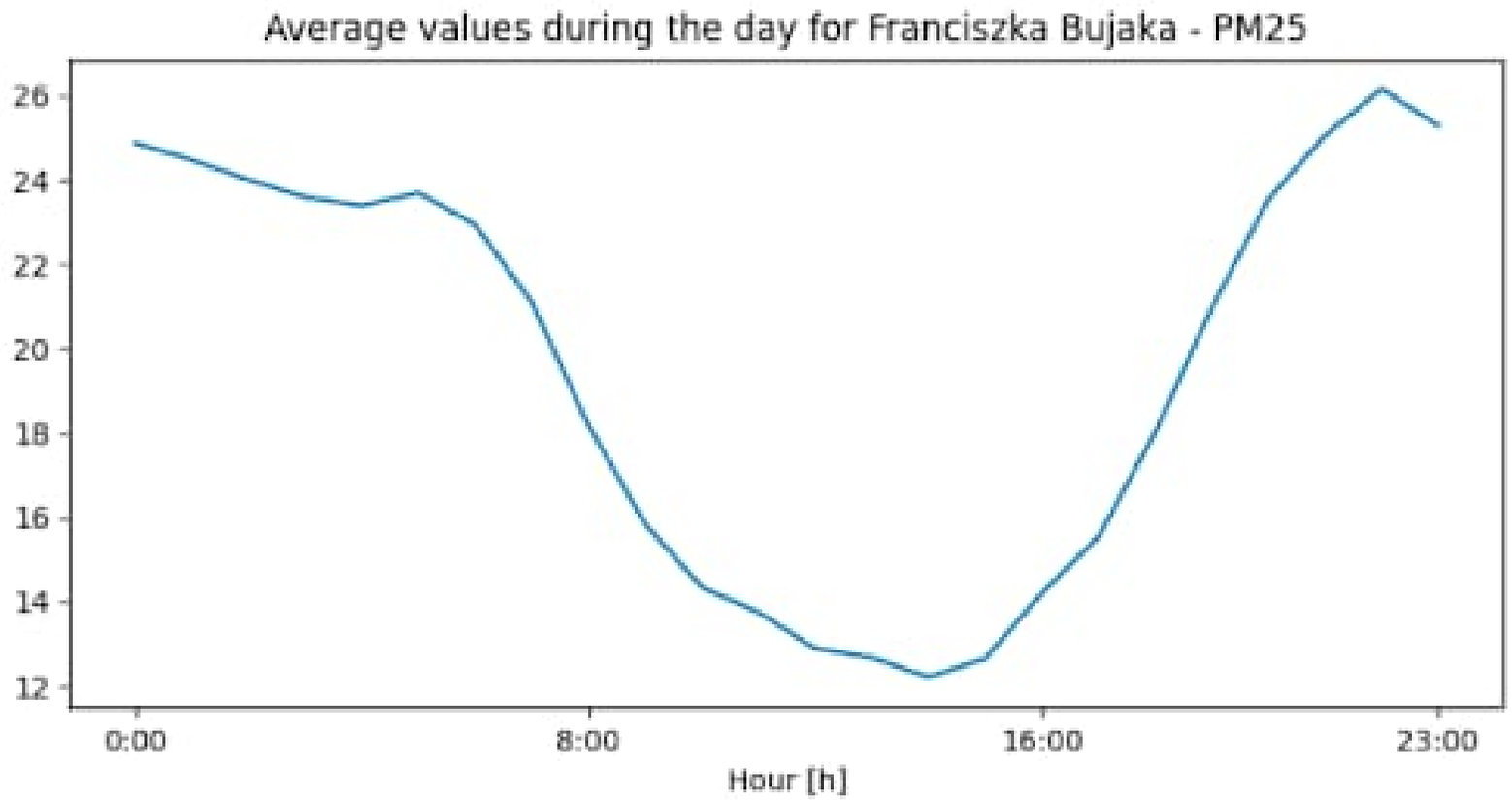
Wzrost w kwietniu

Wzrost zanieczyszczeń w przypadku wszystkich stacji względem marca.

Czerwiec z najniższymi poziomami

Miesiąc o najniższych poziomach zanieczyszczeń spośród badanych miesięcy.

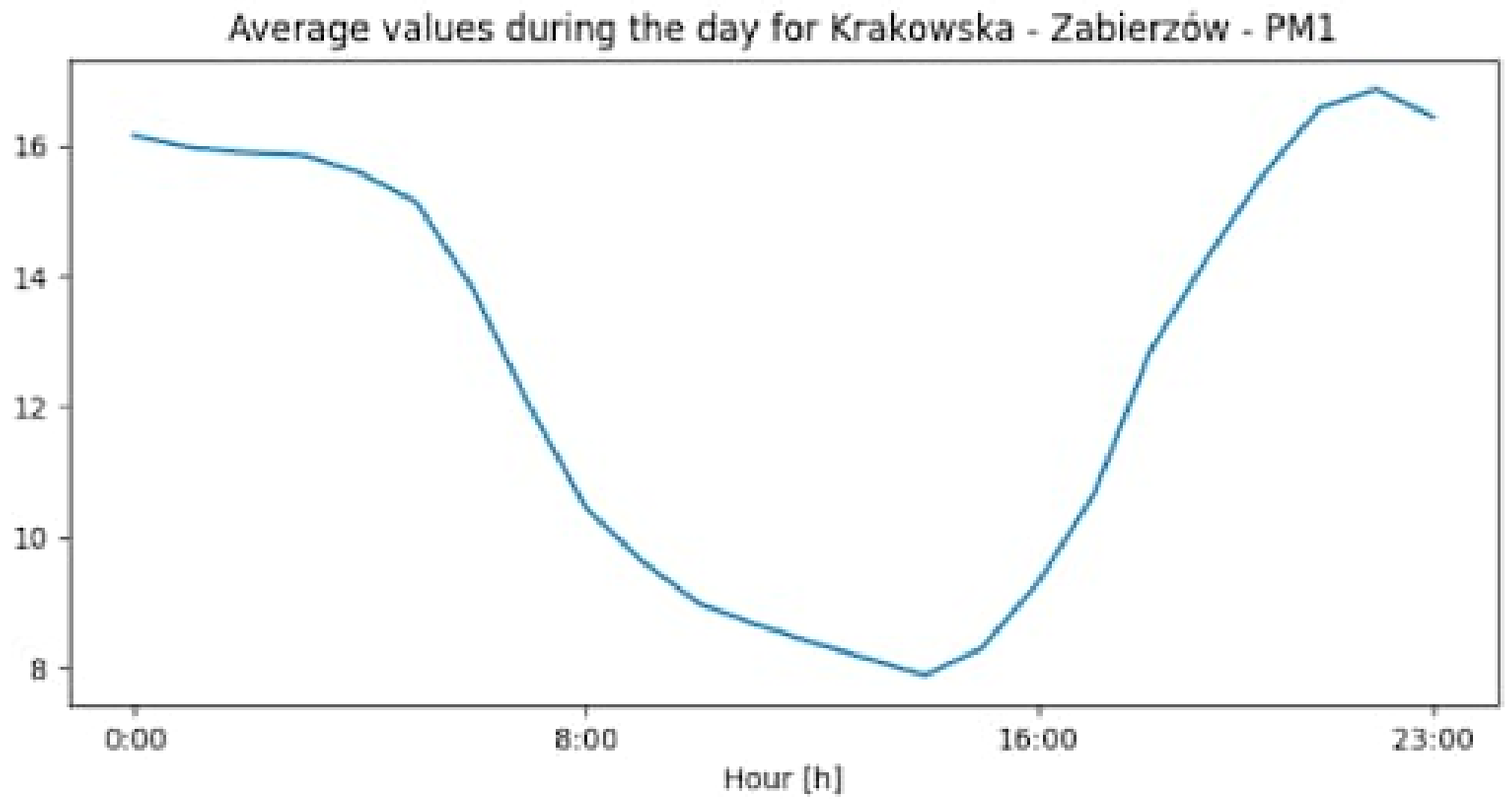
Średnie zanieczyszczenie w ciągu doby



Cykliczność

Największe wartości zanieczyszczenia występują nocą i rano.

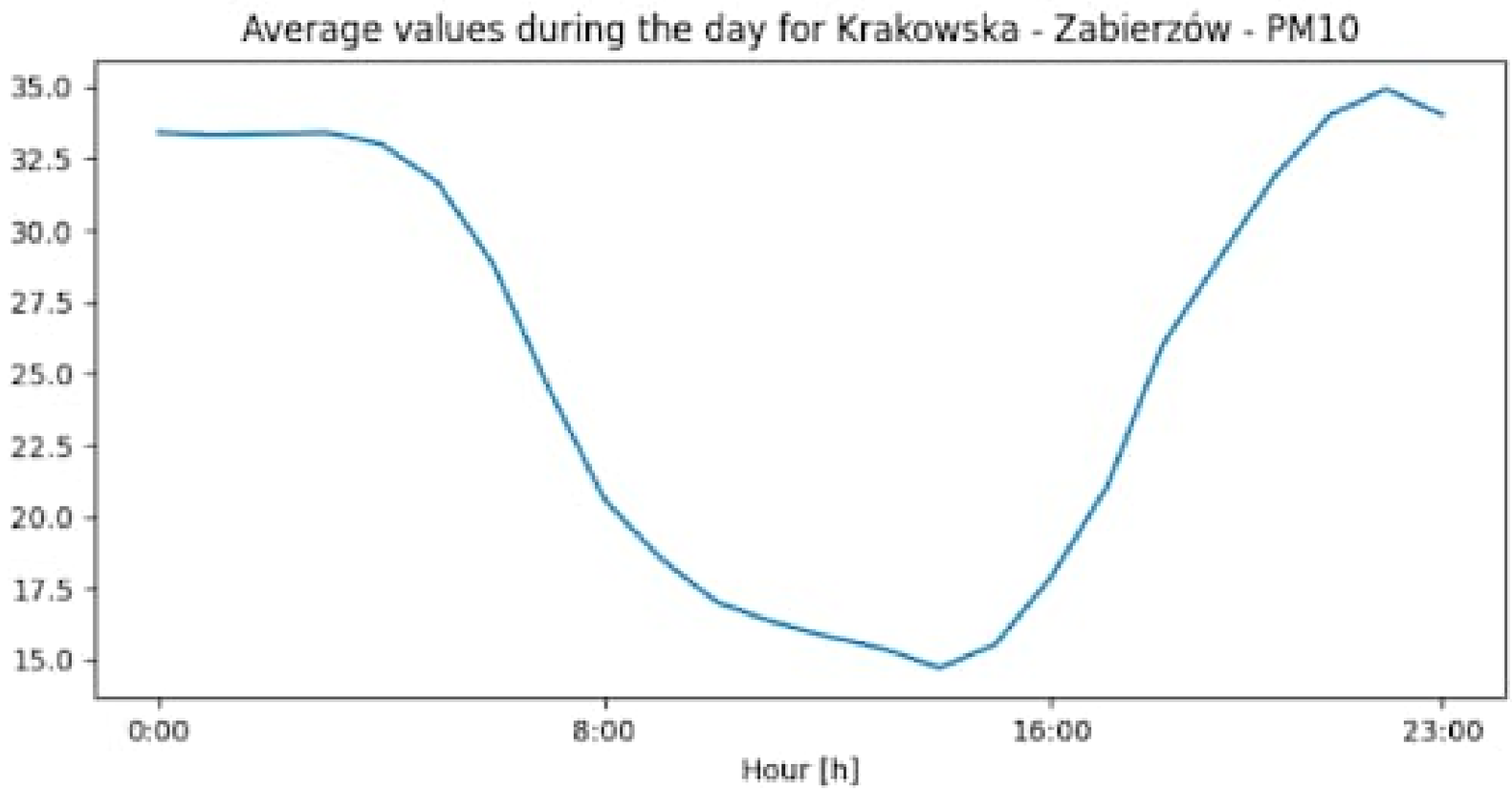
Średnie zanieczyszczenie w ciągu doby



Mniejszy ruch samochodowy rano

Wczesne godziny poranne charakteryzują się mniejszym ruchem samochodowym, co może przyczynić się do zmniejszenia emisji spalin i zanieczyszczeń powietrza.

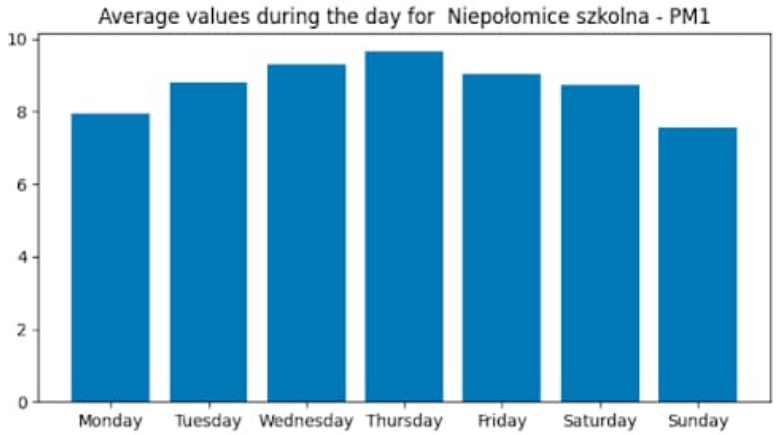
Średnie zanieczyszczenie w ciągu doby

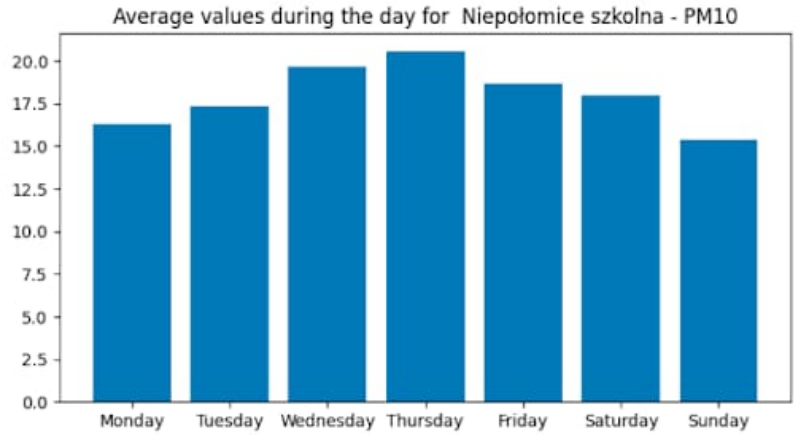


Wzrost zanieczyszczenia po południu

Poziomy zanieczyszczeń ponownie zaczynają wzrastać około godziny 15, kiedy to zwiększa się ruch samochodowy.

Zanieczyszczenie powietrza w ciągu tygodnia





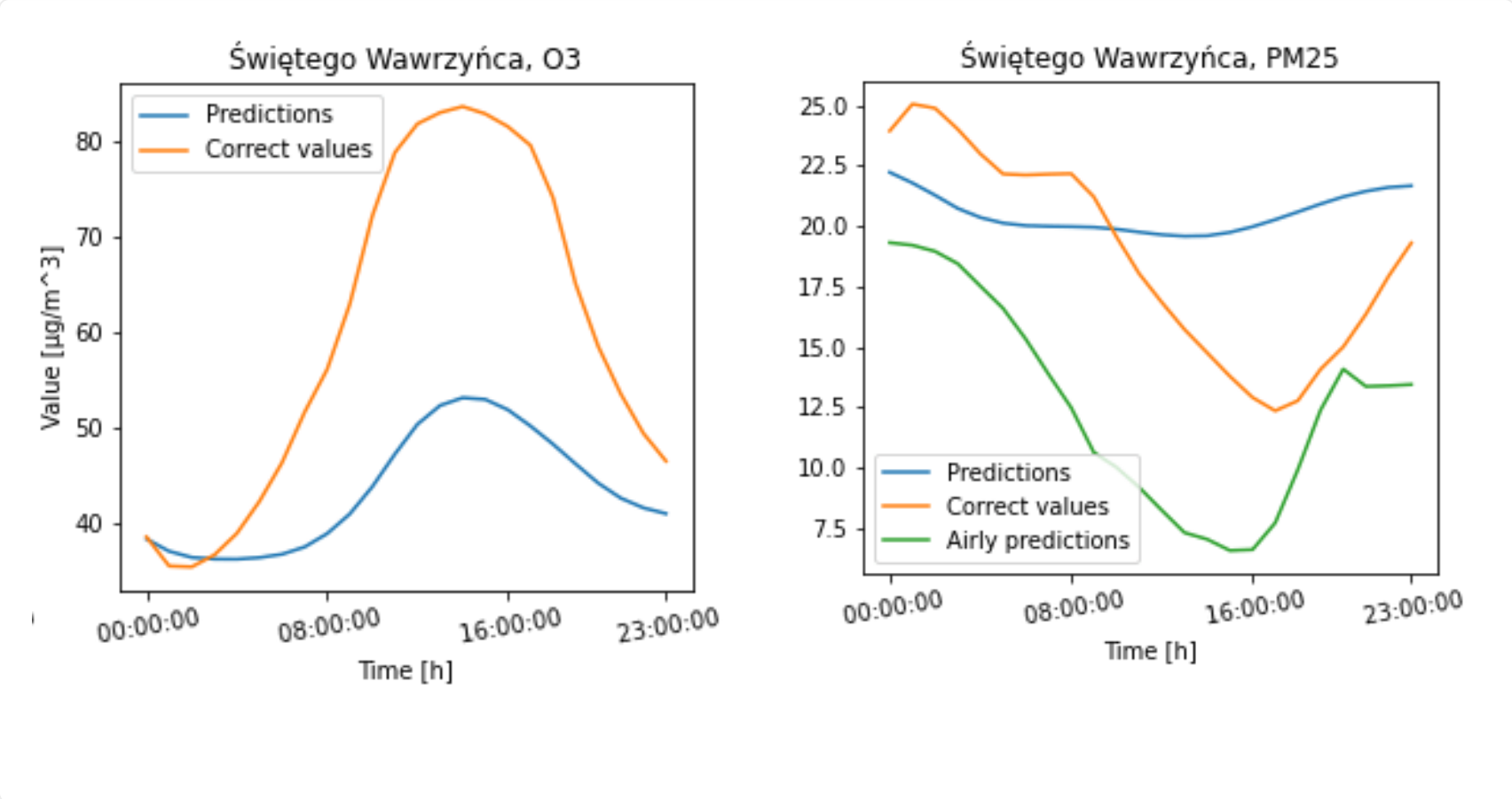
Spadek zanieczyszczenia podczas weekendu

W ciągu tygodnia, poziom zanieczyszczeń jest zazwyczaj wyższy, ponieważ jest więcej ruchu ulicznego i działalności przemysłowej.

Największe zanieczyszczenie w środę i czwartek

Największe zanieczyszczenie występuje w połowie tygodnia, szczególnie w środę i czwartek.

Predykcja modelu oraz Airly



Nasze predykcje są dokładniejsze, jeśli wykonujemy je dla krótszego czasu.

Zmieniające się warunki pogodowe utrudniają predykcje

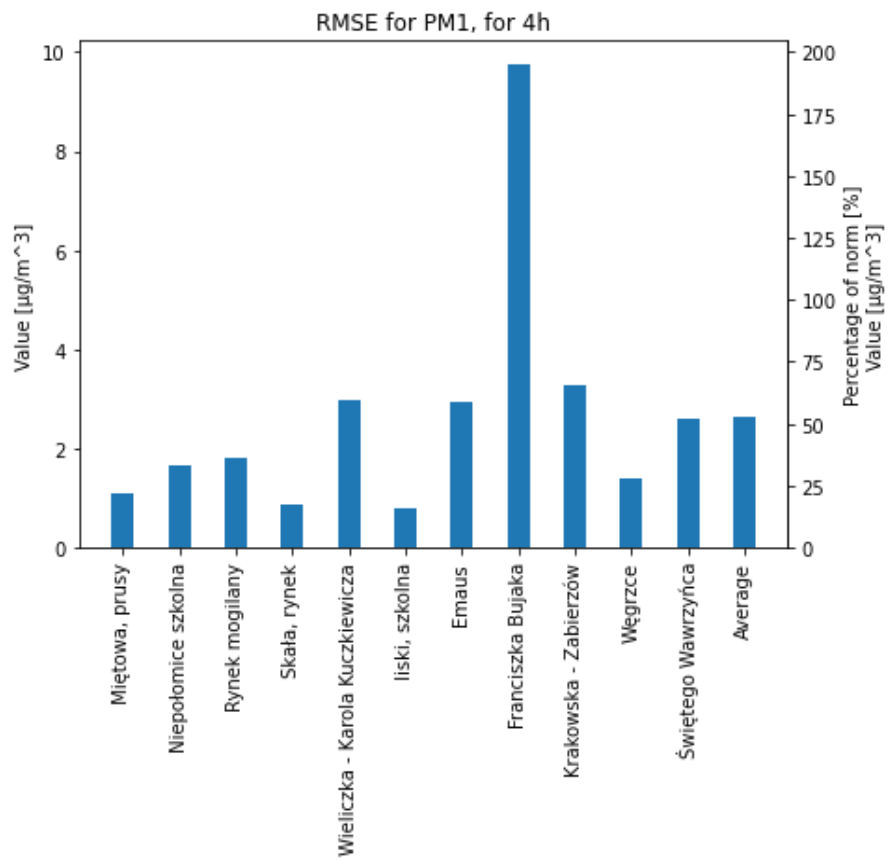
Błędy w predykcjach

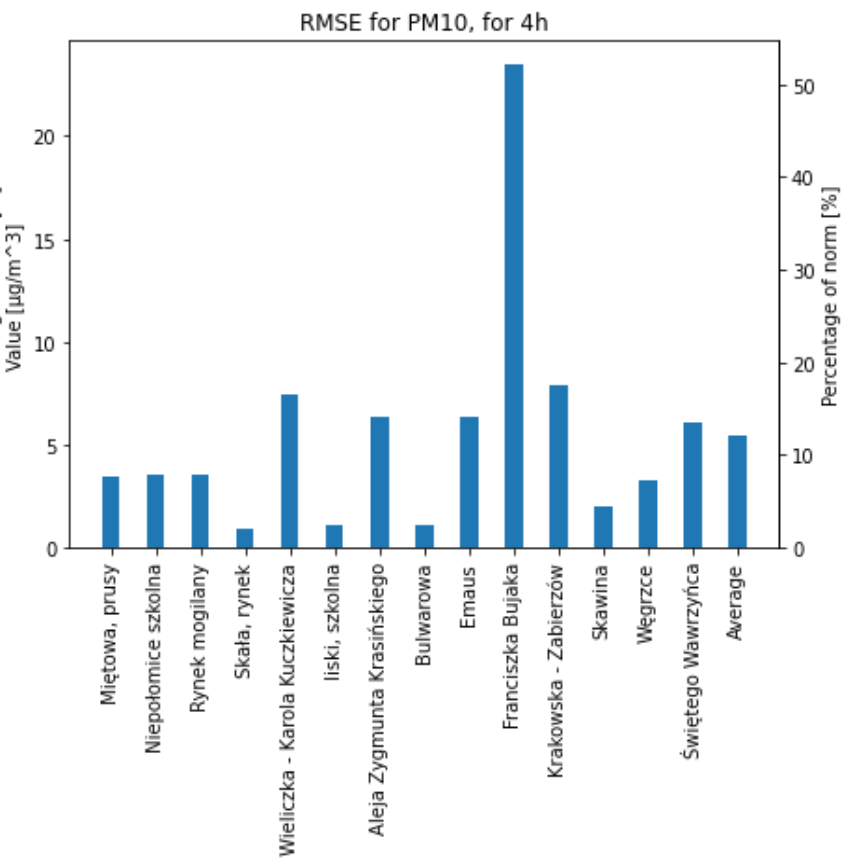
Błędy pojawiają się zazwyczaj w momencie, w którym poziom zanieczyszczenia gwałtownie wzrósł (np. Święty Wawrzyniec, O3). Inną przyczyną błędów jest początkowy błąd (np. Święty Wawrzyniec, PM25).

Zmieniające się warunki pogodowe utrudniają predykcje

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

Analiza błędu predykcji





1 Stacja Franciszka Bujaka

Obecny model zwraca o wiele gorsze wyniki niż dla pozostałych stacji (w szczególności dla 4 godzinnych pomiarów - dla 24 błąd zmniejszył się).

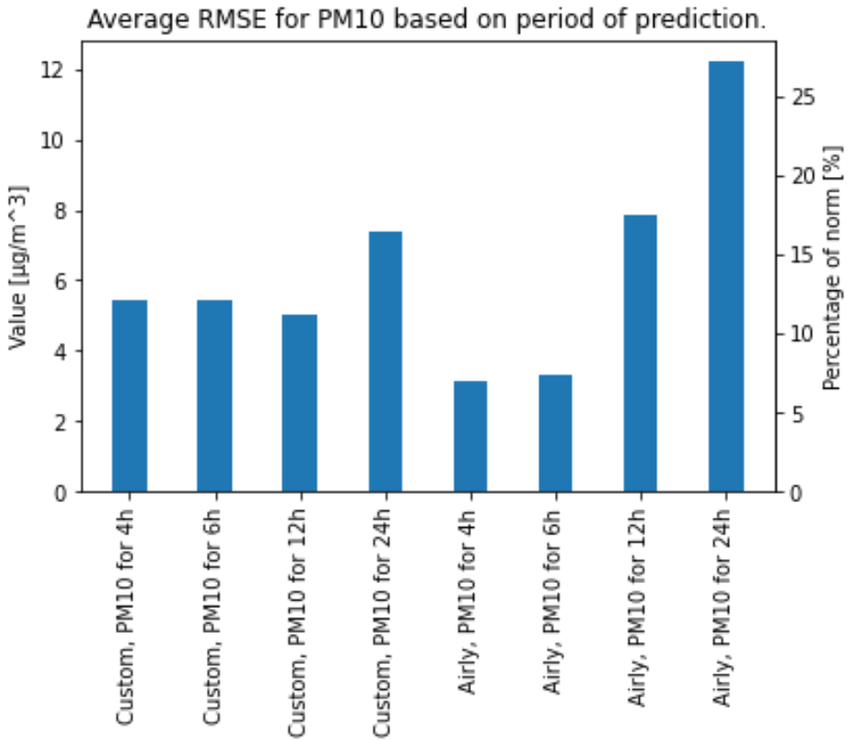
2 Pozostałe stacje

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.

3 Błędy w porównaniu do norm

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.

Porównanie błędów z Airly



Błędy modelu

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.

Porównanie z Airly

Osiągamy gorsze wyniki dla prognoz krótkoterminowych, a lepsze dla długoterminowych w porównaniu z Airly.

Podsumowanie

- Przeprowadzono analizę zanieczyszczeń w okolicach Krakowa dla różnych okresów
- Odkryto trendy dobowe, tygodniowe oraz roczne
- Zaobserwowano, że warunki atmosferyczne, takie jak temperatura, prędkość wiatru i wilgotność, mają istotny wpływ na poziomy zanieczyszczeń
- Stworzono model predykcji zanieczyszczeń korzystający w archiwalnych danych oraz prognozy pogody przewidujący poziom zanieczyszczeń następnego dnia