

Analiza zależności gruźlicy i indeksów giełdowych od pogody

Damian Kąkol
Jakub Kępka

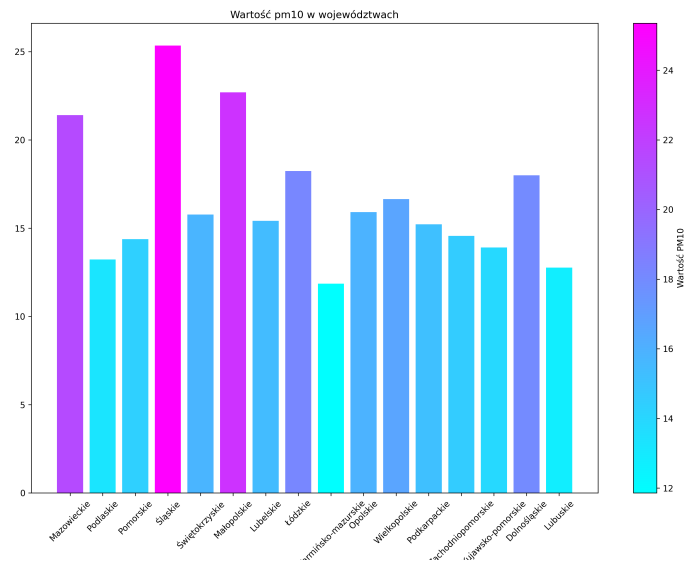
June 2024

1 Analiza struktury zanieczyszczenia powietrza w polskich województwach

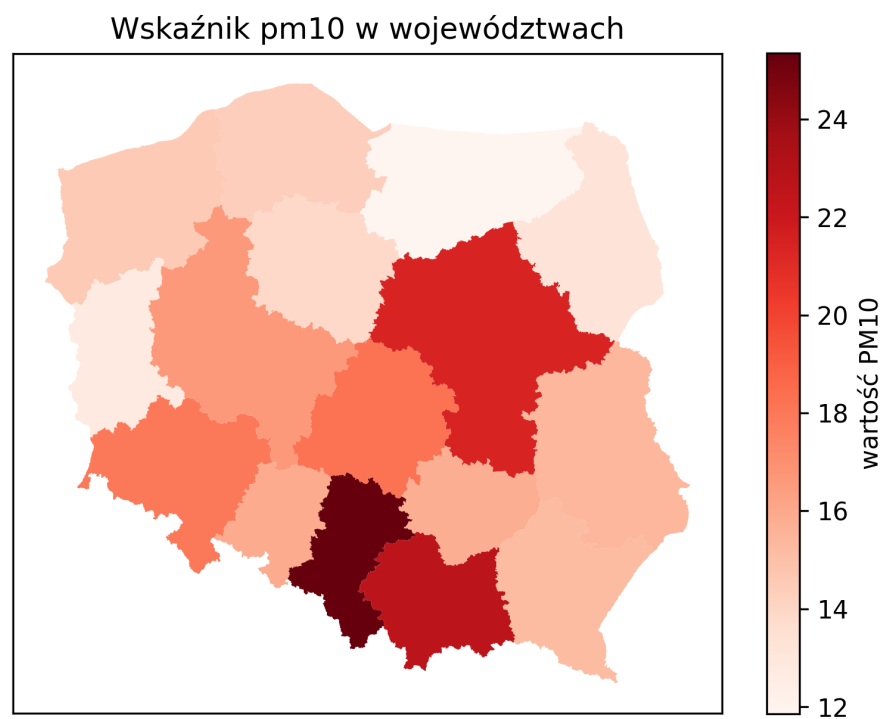
1.1 Średnia roczna wartość wskaźnika pm10 w województwach

PM10 to mieszanina zawieszonych w powietrzu cząsteczek, których średnica nie przekracza 10 mikrometrów. Jest szkodliwy z uwagi na zawartość takich elementów jak benzopireny, furany, dioksyny – jednym słowem, rakotwórcze metale ciężkie. Norma średniego, dobowego stężenia tego pyłu wynosi według WHO 50 mikrogramów na metr sześcienny, a roczna 20 mikrogramów na metr sześcienny. Warto jednak wiedzieć, że informację o przekroczonych normach ogłasza się, gdy dobowe stężenie PM10 wyniesie 200 mikrogramów na metr sześcienny – widać więc doskonale jak często wszyscy myślimy, że powietrze jest w porządku, podczas gdy jest ono bardzo zanieczyszczone, jednak nie został jeszcze osiągnięty poziom alarmowy.¹

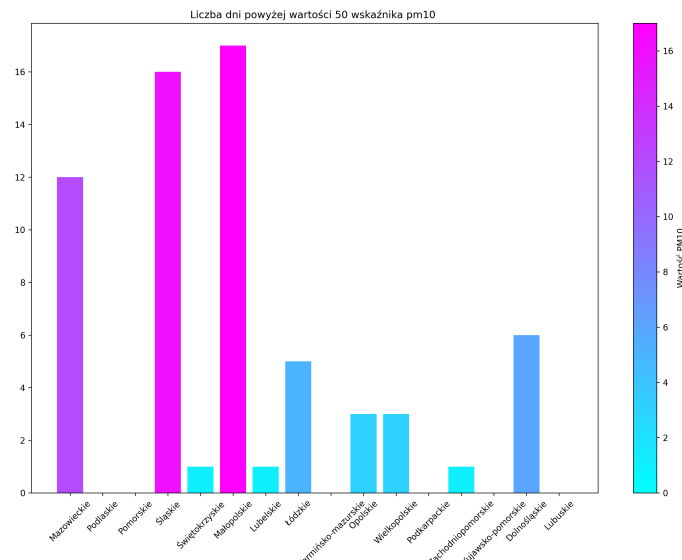
¹<https://airly.org/pl/pyl-zawieszony-czym-jest-pm10-a-czym-pm2-5-aerозole-atmosferyczne/>



Rysunek 1: Średnia roczna wartość wskaźnika pm10 w województwach



Rysunek 2: Średnia roczna wartość wskaźnika pm10 w województwach nałożona na mapę

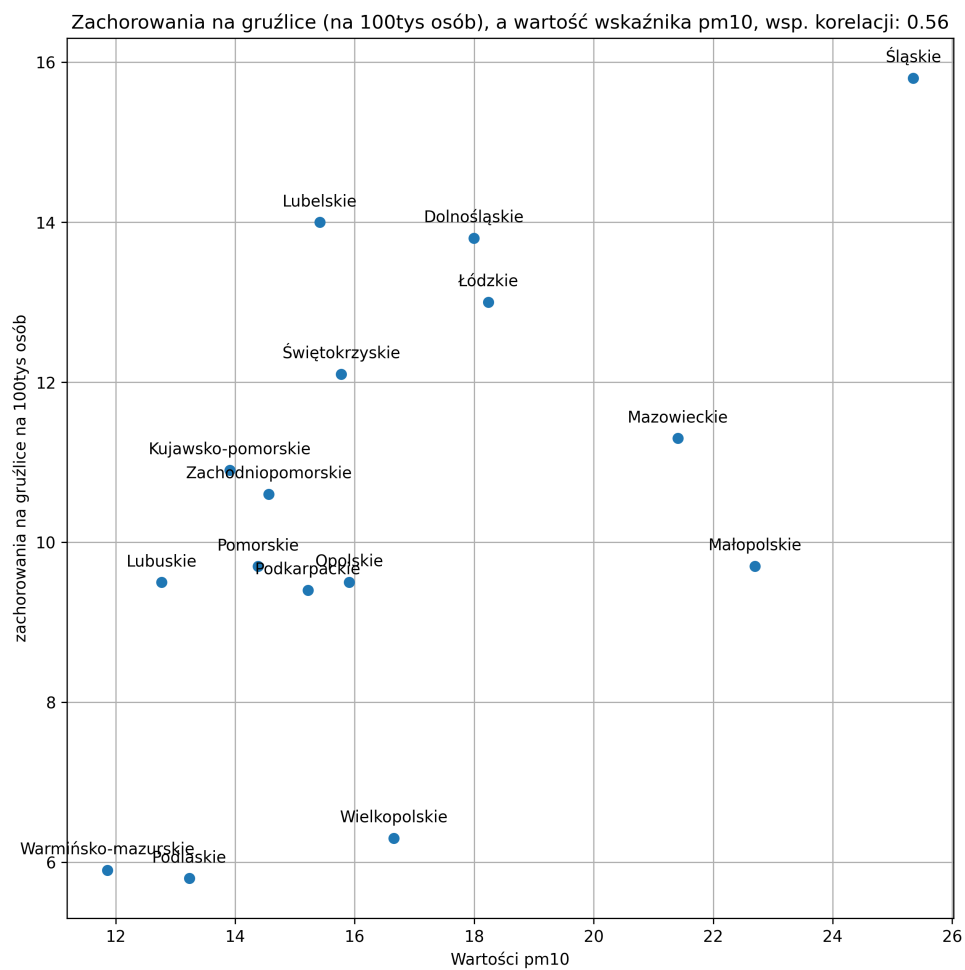


Rysunek 3: Liczba dni z przekroczoną wartością 50 wskaźnika pm10(norma do 35dni)

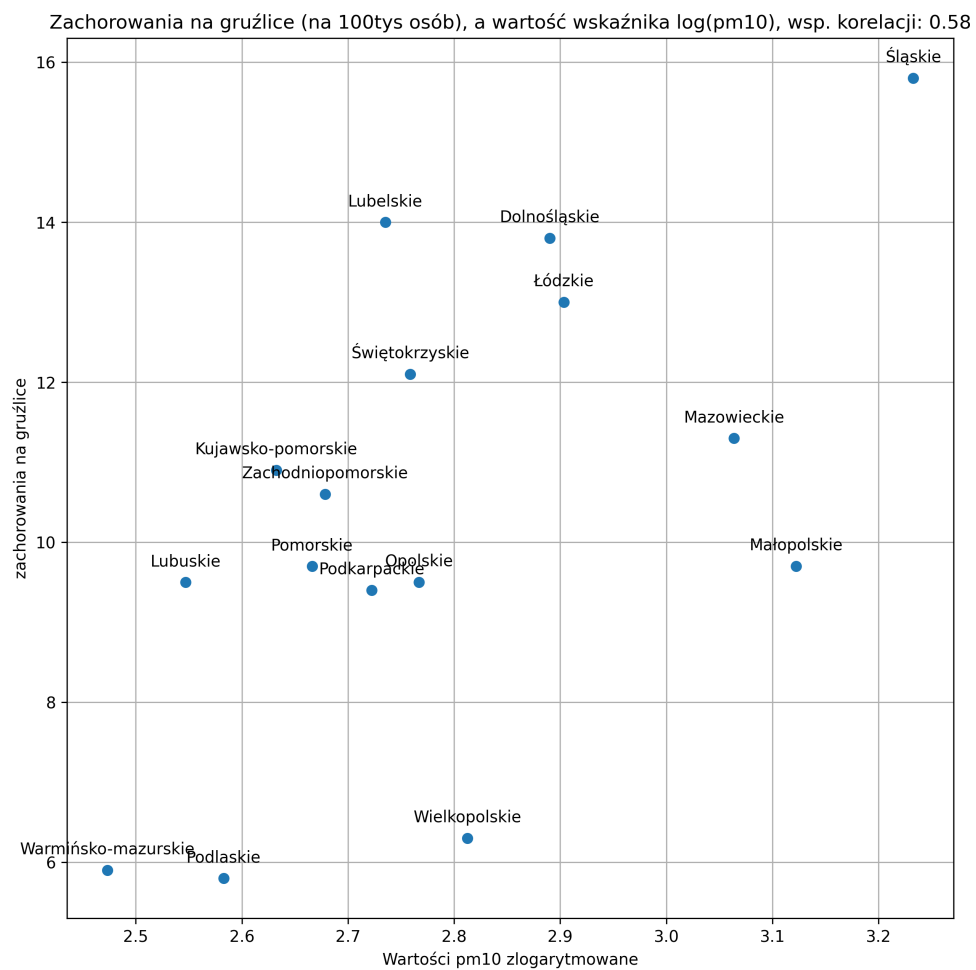
1.2 Analiza korelacji z liczbą zachorowań na gruźlicę

"Udowodniony jest wpływ ekspozycji na pył zawieszony na ryzyko infekcji u ludzi (Ghio 2014). Cząstki pyłu mogą osadzać się na nabłonku nosa, gardła, krtani, tchawicy, oskrzeli i proksymalnej części płuc. Górnictwo, rolnictwo, dym tytoniowy, burze piaskowe, emisje związane komunikacją są źródłem cząstek, z którymi wiąże się zwiększone ryzyko wystąpienia infekcji dróg oddechowych. Narażenie na te czynniki może ponadto skutkować zwiększonym ryzykiem wystąpienia gruźlicy i atypowych infekcji"²

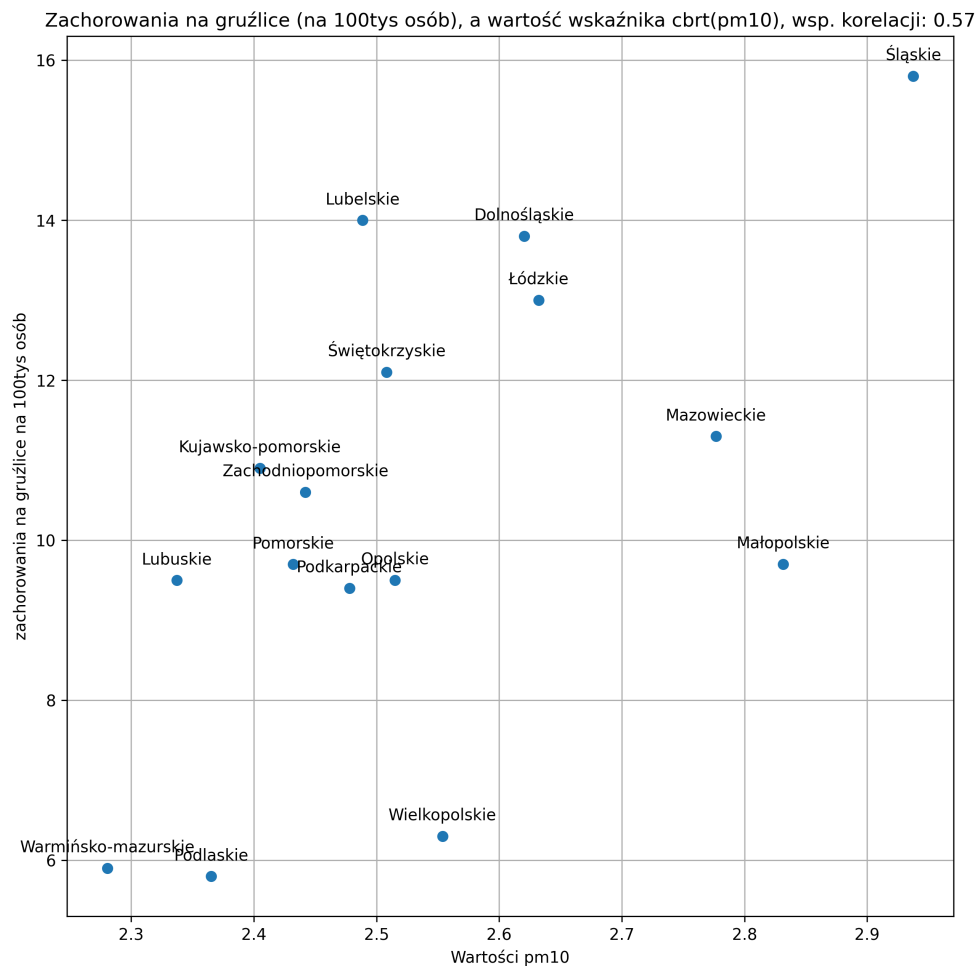
²WPŁYW ZANIECZYSZCZENIA POWIETRZA NA CHOROBY UKŁADU ODDECHOWEGO - <https://bibliotekanauki.pl/articles/1034398.pdf>



Rysunek 4: Średnia roczna wartość wskaźnika pm10 w województwach, a zachorowania na gruźlicę



Rysunek 5: Zlogarytmowana średnia roczna wartość wskaźnika $\text{pm}10$ w województwach, a zachorowania na gruźlicę



Rysunek 6: spierwiastkowana średnia roczna wartość wskaźnika pm10 w województwach, a zachorowania na gruźlicę

1.3 Wnioski

- Gruźlica nie jest powszechną chorobą, do tego jest zakaźna więc niewielkie zdarzenia w ciągu roku mają znaczący wpływ na dane
- Korzystanie z ogólniejszych statystyk np. chorób układu oddechowego umożliwiłoby precy-

zyjniejszą analizę

- Specyfika api open meteo utrudnia dostarczenie danych z większych obszarów
- Gdyby przełożyć uzyskaną zależność na inne choroby układu oddechowego to należałoby uznać, że jakość powietrza ma kluczowy wpływ na zdrowie naszych płuc

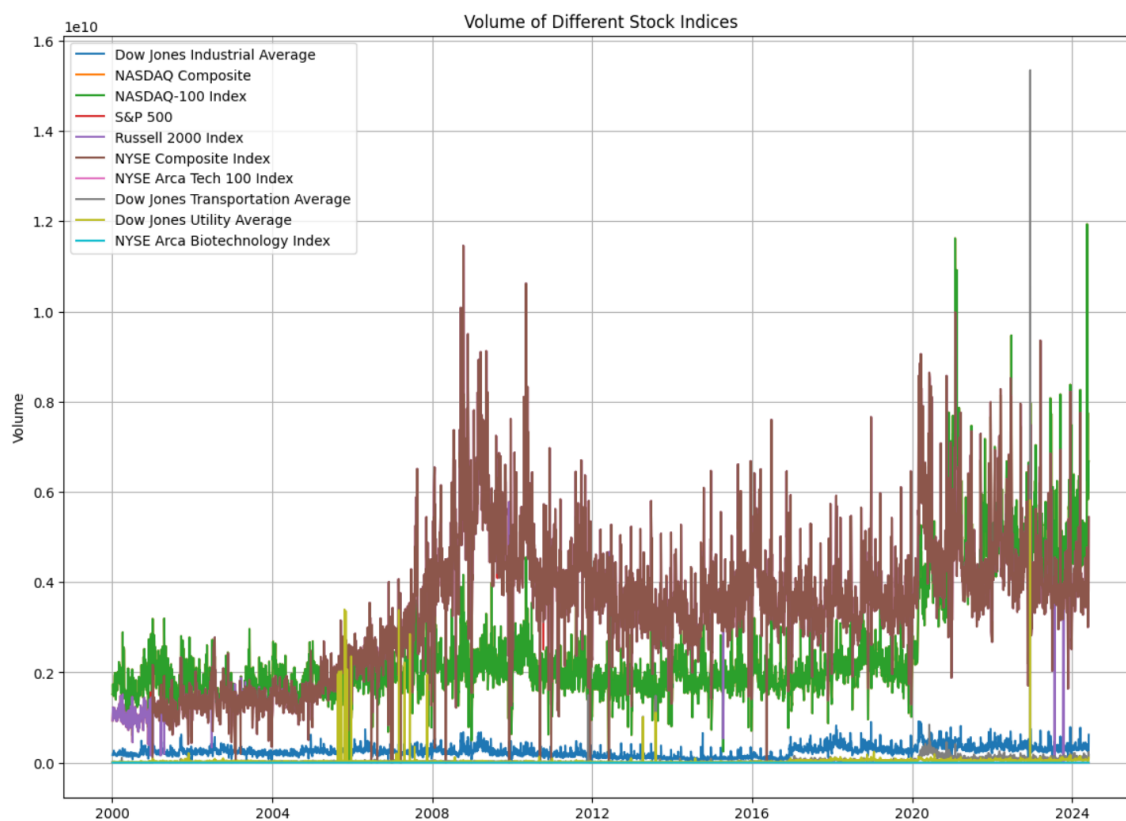
2 Analiza zależności indeksów giełdowych i danych pogodowych

2.1 Wskaźniki giełdowe

Wybrałem 10 różnych indeksów giełdowych z giełdy Nowojorskiej. Konkretnie interesował mnie parametr wolumenu, ale uwzględniałem też najmniejsze i największe ceny z dnia. Wszystkie parametry uwzględniały dane z danego pojedynczego dnia. Wszystkie dane pobrałem z biblioteki python yahoofinance. Rozpatrywany przedział czasowy to 2000-01-01 do 2024-06-01. Oto krótka charakterystyka każdego z nich:

- **Dow Jones Industrial Average (^DJI)**: Jest to indeks, który śledzi akcje 30 największych i najbardziej znanych firm w Stanach Zjednoczonych, reprezentujących różne sektory przemysłowe. Jest powszechnie używany jako barometr zdrowia gospodarki USA.
- **NASDAQ Composite (^IXIC)**: Indeks, który obejmuje wszystkie akcje notowane na giełdzie NASDAQ, zarówno technologiczne, jak i nie-technologiczne, oferujący wgląd w ogólny klimat na rynku akcji, zwłaszcza wśród firm technologicznych.
- **NASDAQ-100 Index (^NDX)**: Składa się z 100 największych spółek notowanych na giełdzie NASDAQ pod względem kapitalizacji rynkowej, głównie firm technologicznych i innowacyjnych, co czyni go wskaźnikiem trendów w sektorze technologicznym.
- **S&P 500 (^GSPC)**: To indeks, który obejmuje 500 największych firm notowanych na giełdzie amerykańskiej, reprezentujących różne sektory gospodarki. Jest jednym z najczęściej używanych wskaźników dla amerykańskiego rynku akcji.
- **Russell 2000 Index (^RUT)**: Skupia się na 2000 mniejszych firmach notowanych na giełdzie amerykańskiej, co czyni go miernikiem dla wydajności mniejszych kapitalizacyjnie spółek.
- **NYSE Composite Index (^NYA)**: Jest to indeks, który obejmuje wszystkie spółki notowane na New York Stock Exchange (NYSE), oferujący szeroki wgląd w ogólny stan rynku akcji na NYSE.

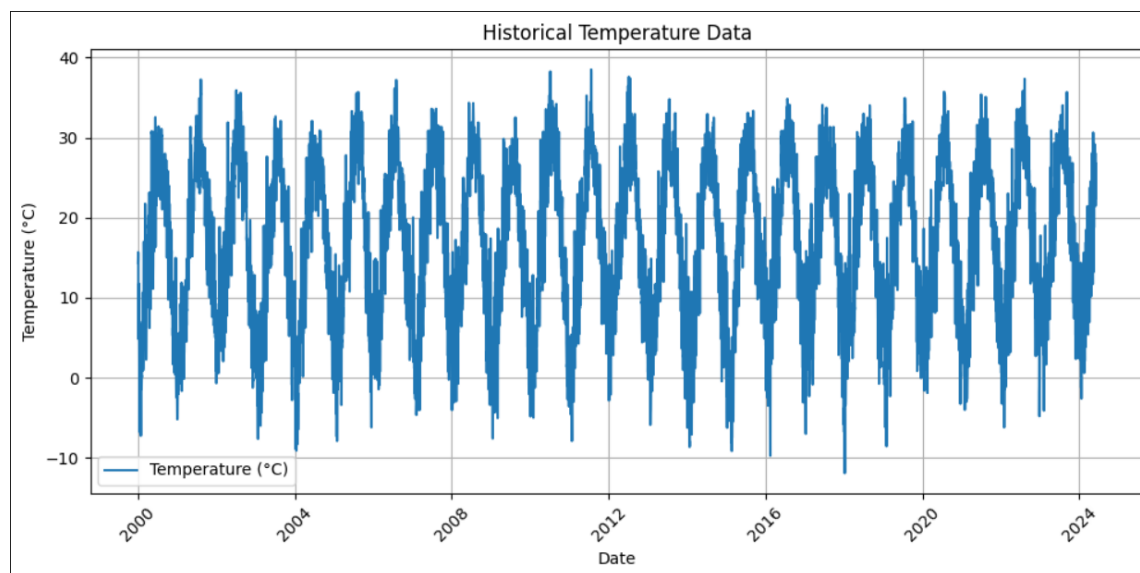
- **NYSE Arca Tech 100 Index (^NYFANG):** Koncentruje się na 100 największych technologicznych spółkach notowanych na NYSE Arca, co czyni go miarą dla wydajności wiodących technologicznych firm na tej giełdzie.
- **Dow Jones Transportation Average (^DJT):** Indeks, który śledzi wydajność 20 kluczowych firm transportowych w USA, dostarczając informacji o zdrowiu sektora transportowego gospodarki.
- **Dow Jones Utility Average (^DJU):** Zawiera akcje 15 głównych przedsiębiorstw publicznych z sektora usług publicznych, takich jak energia, gaz, woda, co czyni go miernikiem dla wydajności firm świadczących usługi publiczne.
- **NYSE Arca Biotechnology Index (^BTK):** Indeks skupiający się na firmach biotechnologicznych notowanych na NYSE Arca, co czyni go wskaźnikiem trendów w sektorze biotechnologicznym na tej giełdzie.



Rysunek 7: Dane historyczne wolumeny sprzedaży akcji dla każdego indeksu giełdowego

2.2 Dane pogodowe

Dane pogodowe wybrałem ze strony open-meteo. Parametry które mnie interesowały to temperatura maksymalna, minimalna średnia oraz odczuwalna (również tutaj maksymalna, minimalna i średnia), opady deszczu oraz wiatr. Zakres czasowy z jakiego brałem dane to 2000-01-01 do 2024-06-01.



Rysunek 8: Dane pogodowe historyczne od 2000 roku

2.3 Dlaczego takie dane?

Model człowieka racjonalnego, działającego jak komputer został podważony (czy też zostaje podważany). Wyobrażenie takie okazuje się coraz bardziej fałszywe. Wiemy teraz że człowiek jest głęboko zależny od środowiska, okoliczności, emocji, ludzi otaczających go czy też nawet stanu jego żołądka. Wydaje się również że pogoda ma również wpływ na nasze zachowanie. Z tą powstaje pytanie o zależności stanu indeksów giełdowych (ceny akcji oraz ilości transakcji) od pogody. Jedno z badań wyszukanych przeze mnie badań potwierdza tezę o korelacji, jednak zauważa że ta korelację się zmniejszała wraz z momentem rozpowszechnienia się powszechnie klimatyzacji.³

2.4 Analiza korelacji danych pogodowych i stanu indeksów giełdowych

Pierwszym z narzędzi jakich użyłem do wyznaczania zależności jest korelacja Pearsona. Okazała się ona jednak bez skuteczna, dała wyniki na poziomie bliskim zera dla wszystkich indeksów i dla różnych parametrów pogodowych i giełdowych. Największą korelację jaką znalazłem to była korelacja na poziomie 0.25 między 4 pierwszymi indeksami, a prędkością wiatru.

Kolejno korzystałem z modelu regresji liniowej, co myślę że w sposób przewidywalny dało wynik bezwartościowy. Większość wyników wahało się blisko wyniku: Mean Squared Error: $1.0594516060826866e+18$ R-squared: 0.013276888050757885

³Stock Returns, weather, and air conditioning - <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219439>

Użyłem modelu RandomForest który okazał się równie nie użyteczny. Tak samo z modelem SVM.

2.5 Wnioski

Z moich obliczeń, modeli których wykorzystywałem i danych na które się powoływałem wynika że nie istnieje żadna korelacja między pogodą (temperaturą, wiatrem, opadami deszczu i śniegu), a indeksami z giełdy Nowojorskiej (wolumen, cena otwarcia, zamknięcia, największa i najmniejsza cena z dnia). Wydaje się że średnia temperatura i wszystkie "normalne" które wykorzystywałem w obliczeniach mogą zakłócać znaczenie dni o większym odchyleniu od normy, które to już mogą mieć wpływ na działania inwestorów na giełdzie (jak w przedstawionym wyżej badaniu oraz wielu innych). Należy więc podejść do tego badania z inną metodologią i patrzeć bardziej lokalnie, wyłapując dni z dużym odchyleniem.