

Wrocław, 29.11.2024

Projekt

Metody analizy i eksploracji danych

Eksploracyjna analiza danych

**„Badanie mikroklimatów na podstawie danych
meteorologicznych i topograficznych”**

Autorzy: Bartosz Palmer 260346

Nataniel Jargiło 252888

Termin zajęć: Wtorek 17:05

Prowadzący: dr inż. Agata Migalska

1. Identyfikacja zmiennych

W projekcie zostały wykorzystane dwa zbiory danych. Pierwszy zawierające dane synoptyczne pochodzi z IMGW i obejmuje miesięczne obserwacje dla poszczególnych stacji meteorologicznych w Polsce, co składa się na 708 rekordów, zawierających 14 zmiennych (z 59, które zawierał zbiór danych). Drugi zawiera numeryczny model terenu Polski w obszarze 5x5km, którego środkiem są współrzędne danej stacji pomiarowej. Zestaw ten pochodzi z serwisu geoportal.

Poniżej przedstawiona została identyfikacja wybranych zmiennych z danych synoptycznych:

- Kod_stacji
 - Znaczenie: Kod stacji meteorologicznej, w której zostały zebrane dane.
 - Typ: Numeryczna dyskretna
 - Kodowanie: int64
- Nazwa_stacji
 - Znaczenie: Nazwa stacji meteorologicznej, z której pochodzą dane.
 - Typ: Tekstowa
 - Kodowanie: UTF-8
- Miesiac
 - Znaczenie: Miesiąc, w którym zostały zebrane dane.
 - Typ: Numeryczna dyskretna
 - Kodowanie: int64
- Absolutna_temperatura_maksymalna
 - Znaczenie: Najwyższa zanotowana temperatura w danym miesiącu w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Srednia_temperatura_maksymalna
 - Znaczenie: Średnia wartość maksymalnej temperatury dla danego miesiąca w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Absolutna_temperatura_minimalna
 - Znaczenie: Najniższa zanotowana temperatura w danym miesiącu w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Srednia_temperatura_minimalna
 - Znaczenie: Średnia wartość minimalnej temperatury dla danego miesiąca w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Srednia_temperatura_miesieczna
 - Znaczenie: Średnia miesięczna temperatura z pomiarów danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64

- Minimalna_temperatura_przy_gruncie
 - Znaczenie: Najniższa zanotowana temperatura przy gruncie w danym miesiącu w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Miesieczna_suma_opadow
 - Znaczenie: Łączna ilość opadów w danym miesiącu w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Maksymalna_dobowa_suma_opadow
 - Znaczenie: Najwyższa ilość opadów zanotowana jednego dnia w danym miesiącu w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Miesieczna_suma_uslonecznienia
 - Znaczenie: Łączna liczba godzin słonecznych w danym miesiącu w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Liczba_dni_z_opadem_deszczu
 - Znaczenie: Liczba dni z opadem deszczu w danym miesiącu w danej stacji.
 - Typ: Numeryczna dyskretna
 - Kodowanie: float64
- Srednia_miesieczna_predkosc_wiatru
 - Znaczenie: Średnia miesięczna prędkość wiatru zmierzona w danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Srednie_miesieczne_cisnienie_na_poziomie_stacji
 - Znaczenie: Średnie miesięczne ciśnienie atmosferyczne na poziomie danej stacji meteorologicznej.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Srednia_miesieczna_wilgotnosc_wzgledna
 - Znaczenie: Średnia miesięczna wilgotność względna w rejonie danej stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64

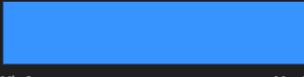
Kolejnymi wykorzystanymi danymi są rastry pochodzące z serwisu geoportal, które stanowią numeryczny model terenu (NMT) Polski. Rastry te są reprezentowane przez jedną zmienną ilościową o charakterze ciągłym, odpowiadającą wysokości terenu w danym punkcie. W ramach projektu wykorzystano dane z obszaru 5x5 km o rozdzielczości 500x500 px, co oznacza, że każda wartość na siatce reprezentuje średnią

wysokość terenu w obszarze o boku 10 m, skupionym wokół stacji pomiarowej. Na podstawie tych danych, dla każdej stacji pomiarowej utworzono następujące zmienne:

- Srednia_wysokosc
 - Znaczenie: Średnia wysokość dla danego rastra (danej stacji)
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Min_wysokosc
 - Znaczenie: Minimalna wysokość w okolicach stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Max_wysokosc
 - Maksymalna wysokość w okolicach stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64
- Procent_plaski
 - Znaczenie: Procent płaskiego terenu, wokół stacji.
 - Typ: Numeryczna ciągła
 - Kodowanie: float64

2. Identyfikacja braków i błędów w danych

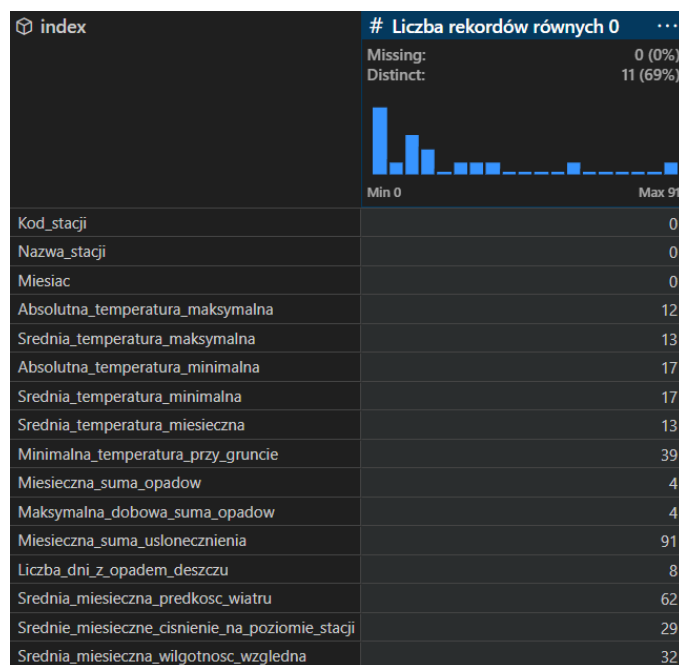
Na Rys. 1. analiza zestawu danych meteorologicznych wykazała brak brakujących rekordów, co oznacza, że dla każdej zmiennej posiadamy pełne dane. Jest to korzystna sytuacja, ponieważ brak luk w danych ułatwia dalszą analizę i ogranicza konieczność uzupełniania brakujących wartości.

index		# Liczba brakujących rekordów	...
		Missing:	0 (0%)
		Distinct:	1 (6%)
			
		Min	0
		Max	0
Kod_stacji			0
Nazwa_stacji			0
Miesiac			0
Absolutna_temperatura_maksymalna			0
Srednia_temperatura_maksymalna			0
Absolutna_temperatura_minimalna			0
Srednia_temperatura_minimalna			0
Srednia_temperatura_miesieczna			0
Minimalna_temperatura_przy_gruncie			0
Miesieczna_suma_opadow			0
Maksymalna_dobowa_suma_opadow			0
Miesieczna_suma_uslonecznienia			0
Liczba_dni_z_opadem_deszczu			0
Srednia_miesieczna_predkosc_wiatru			0
Srednie_miesieczne_cisnienie_na_poziomie_stacji			0
Srednia_miesieczna_wilgotnosc_wzgledna			0

Rys. 1. Tabela brakujących rekordów w danych IMGW

Dodatkowo, na Rys 2. przedstawione zostało występowanie wartości równych zero dla każdej zmiennej, co może sugerować nietypowe sytuacje meteorologiczne lub błędy w danych. Poniżej omówione zostały zmienne, dla których liczba rekordów równych zero jest znacząca:

- `Srednia_temperatura_minimalna` i `Srednia_temperatura_miesieczna`: kolejno 17 i 13 wartości zerowych. Może to wynikać z błędu pomiarowego lub być rezultatem specyficznych warunków klimatycznych, w których temperatura przez cały miesiąc oscylowała wokół zera (okresy zimowe). Wnioski te dotyczą wszystkich zmiennych temperatur.
- `Miesieczna_suma_uslonecznienia`: 91 wartości zerowych, może to świadczyć o brakujących pomiarach w niektórych miesiącach, dla niektórych stacji pomiarowych.
- `Srednie_miesieczne_cisnienie_na_poziomie_stacji`: 29 wartości zerowych, które są brakami pomiarowymi.
- `Liczba_dni_z_opadem_deszczu`: 8 wartości zerowych, co wskazuje na brak dni deszczowych w danym miesiącu na niektórych stacjach. Jest to możliwe w okresach suszy i jest zgodne z rzeczywistością meteorologiczną.
- `Srednia_miesieczna_predkosc_wiatru` oraz `Srednia_miesieczna_wilgotnosc_wzgledna`: kolejno 62 i 32 wartości zerowe. Są to najprawdopodobniej braki pomiarowe, ponieważ takie warunki są raczej niemożliwe.



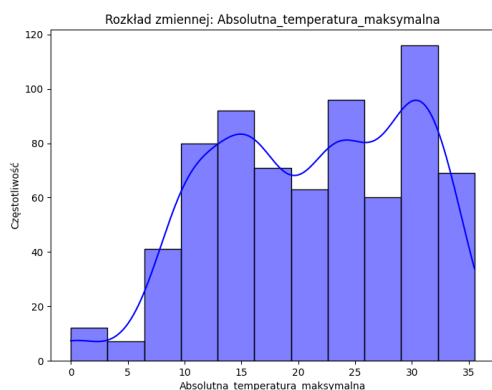
Rys. 2. Tabela rekordów równych 0 w danych IMGW

3. Analiza rozkładów zmiennych

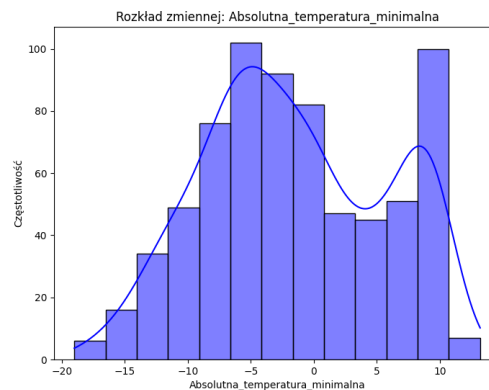
Poniżej przedstawiono wykresy analizy rozkładu zmiennych meteorologicznych dla danych z IMGW. Wykresy rozkładów przedstawiają summaryczne wartości dla wszystkich stacji i miesięcy, bez wyszczególnienia konkretnych lokalizacji czy czasu. Dzięki temu możliwe jest wyznaczenie ogólnych wzorców pogodowych oraz identyfikacja zakresów i częstości występowania poszczególnych wartości.

Rozkład przedstawiony na Rys. 3. jest lewo skośny, z większością wartości w przedziale 10-30 C.

Rozkład z Rys. 4. jest lewo skośny, z większością w zakresie -15 – 5 C.



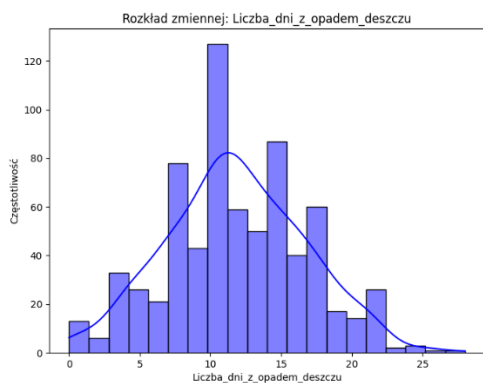
Rys. 3. Rozkład zmiennej abs. temp. maks.



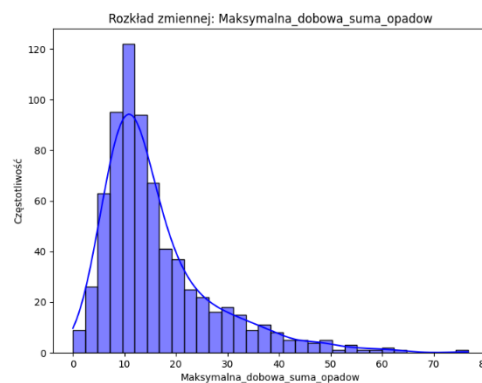
Rys. 4. Rozkład zmiennej abs. temp. min.

Rozkład na Rys. 5. jest zbliżony do normalnego, z najczęstszymi wartościami wokół 10 dni.

Rozkład na Rys. 6. jest prawo skośny, z większością wartości poniżej 20 mm.



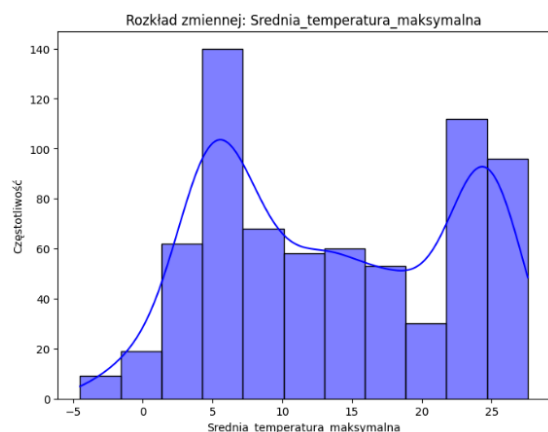
Rys. 5. Rozkład zmiennej l. dni z opad. deszcz.



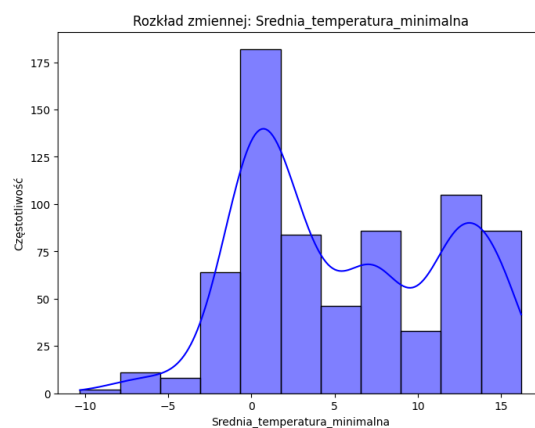
Rys. 6. Rozkład zmiennej maks. dob. suma opadów

Rozkład z Rys. 7. jest dwumodalny ze szczytami w okolicach 5 oraz 25 C.

Rozkład z Rys. 8. jest lewo skośny z najwyższą częstotliwością w okolicach 0 C.



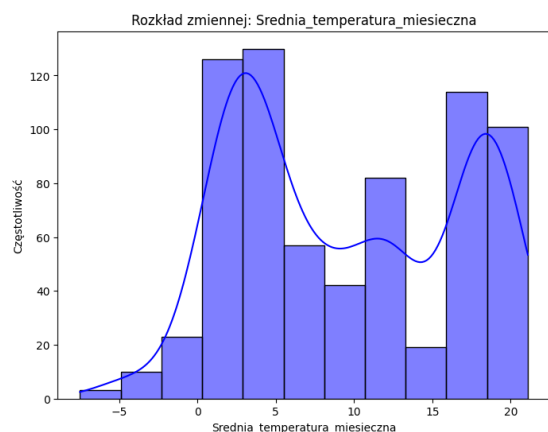
Rys. 7. Rozkład zmiennej średnia temp maks.



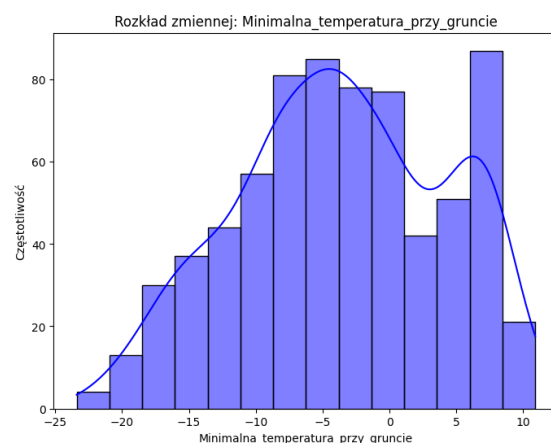
Rys. 8. Rozkład zmiennej średnia temp min.

Rozkład z Rys. 9. jest dwumodalny ze szczytami w okolicach 0 i 15 C.

Rozkład z Rys. 10. jest lewo skośny z największą liczbą obserwacji w przedziale -10 – 5 C.



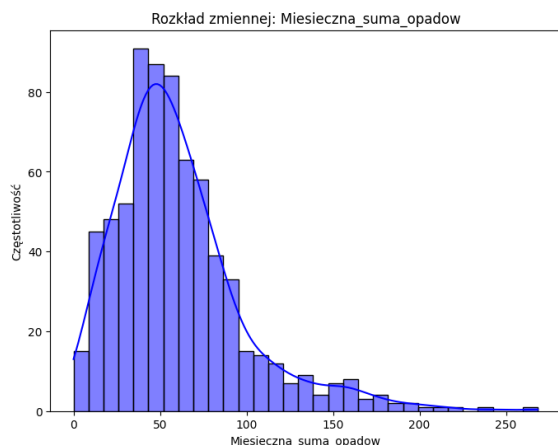
Rys. 9. Rozkład zmiennej średnia temp. mies.



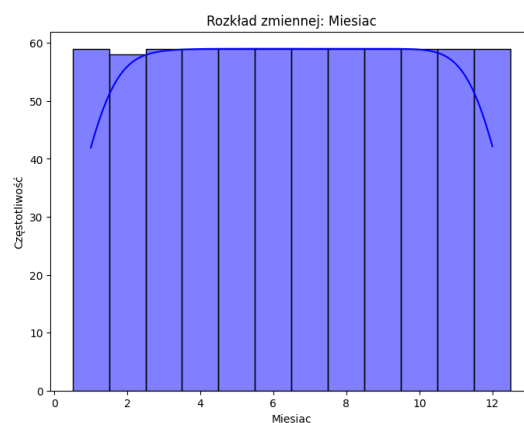
Rys. 10. Rozkład zmiennej min. temp. przy grun.

Rozkład z Rys. 11. jest mocno prawo skośny, z największą liczbą obserwacji w przedziale 20 – 60 mm.

Rozkład z Rys. 12. jest równomierny.



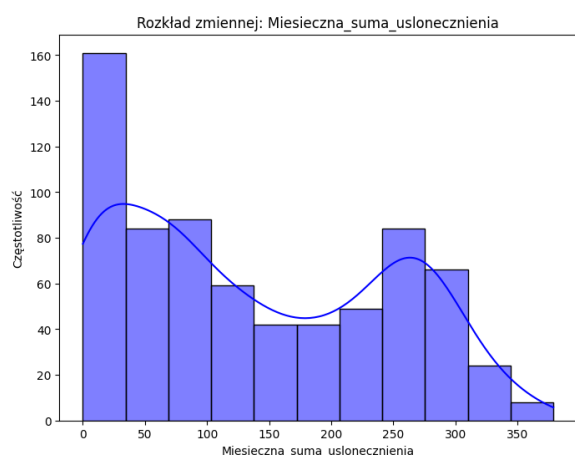
Rys. 11. Rozkład zmiennej mies. suma opadów



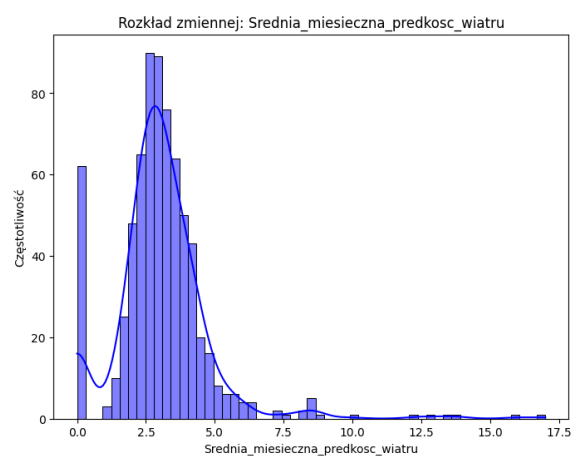
Rys. 12. Rozkład zmiennej miesiąc

Rozkład z Rys. 13. jest dwumodalny ze szczytami ok. 0 – 50 godz. oraz ok. 250 godz.

Rozkład z Rys. 14. jest prawoskośny z większością wartości w zakresie 2 – 4 m/s oraz widocznym szczytem dla 0 m/s.



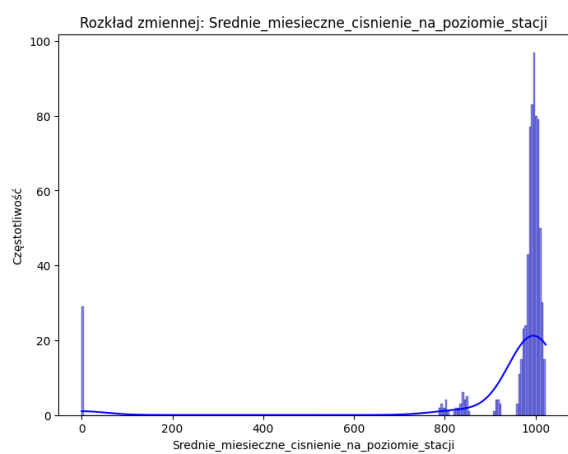
Rys. 13. Rozkład zmiennej mies. suma usłonecz.



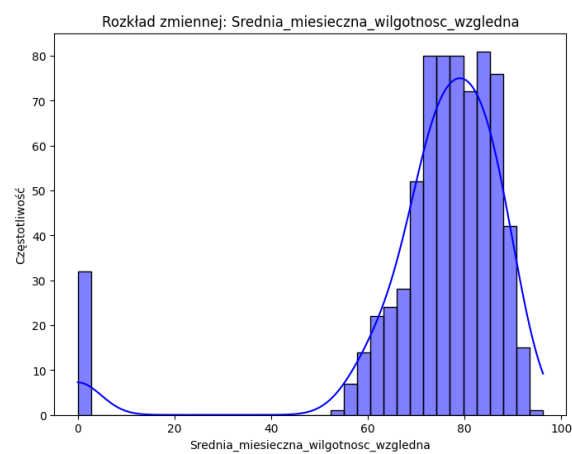
Rys. 14. Rozkład zmiennej śred. mies. pręđ. wiatru

Rozkład z Rys. 15. jest lewo skośny z większością wartości ok. 1000 hPa, oraz małym szczytem dla 0 hPa, który reprezentuje braki w danych.

Rozkład z Rys. 16. jest lewoskośny, z większością wartości w zakresie 60 – 90 %, oraz wartościami odstającymi dla 0 %.



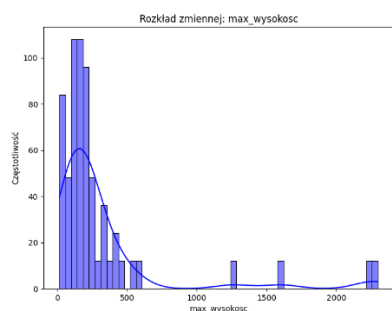
Rys. 15. Rozkład zmiennej śred. mies. ciśnienie



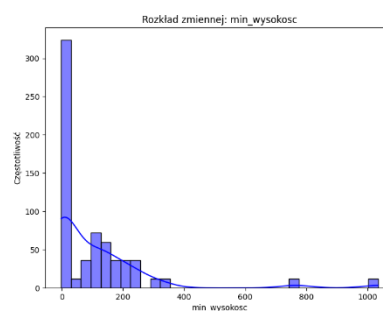
Rys. 16. Rozkład zmiennej śred. mies. wilgot. wzgl.

Rozkład z Rys. 17. jest mocno prawo skośny z większością wartości ok. 0 – 400 m.

Rozkład z Rys. 18. jest mocno prawo skośny z większością wartości równych 0 m.



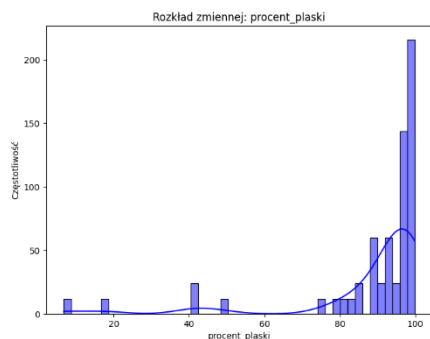
Rys. 17. Rozkład zmiennej max_wysokosc



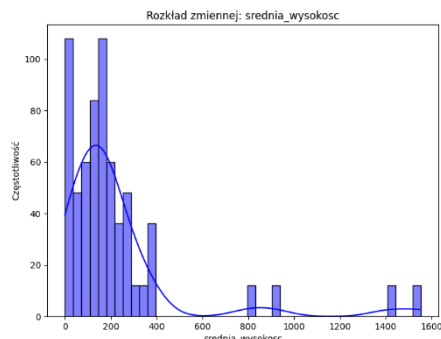
Rys. 18. Rozkład zmiennej min_wysokosc

Rozkład z Rys. 19. jest mocno lewo skośny z większością wartości w zakresie 95-90%.

Rozkład z Rys. 20. jest mocno prawo skośny z większością wartości w zakresie 0 – 300 m.

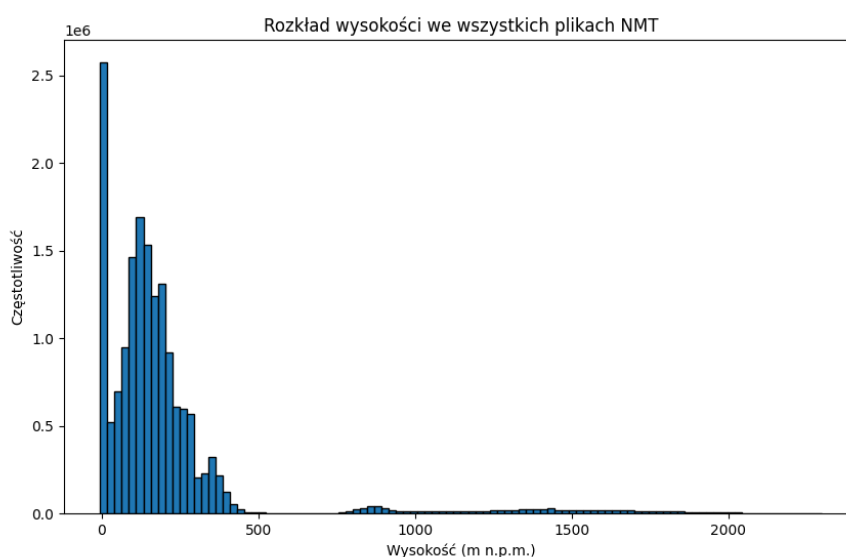


Rys. 17. Rozkład zmiennej max_wysokosc



Rys. 18. Rozkład zmiennej min_wysokosc

Rozkład wysokości z Rys. 21. jest mocno prawo skośny, z największą liczbą obserwacji w przedziale 0 – 500 m. Widoczny jest również mocny szczyt dla wartości 0, co odpowiada pomiarom z miejscowości nadmorskich, gdzie obszar morza jest również brany pod uwagę.

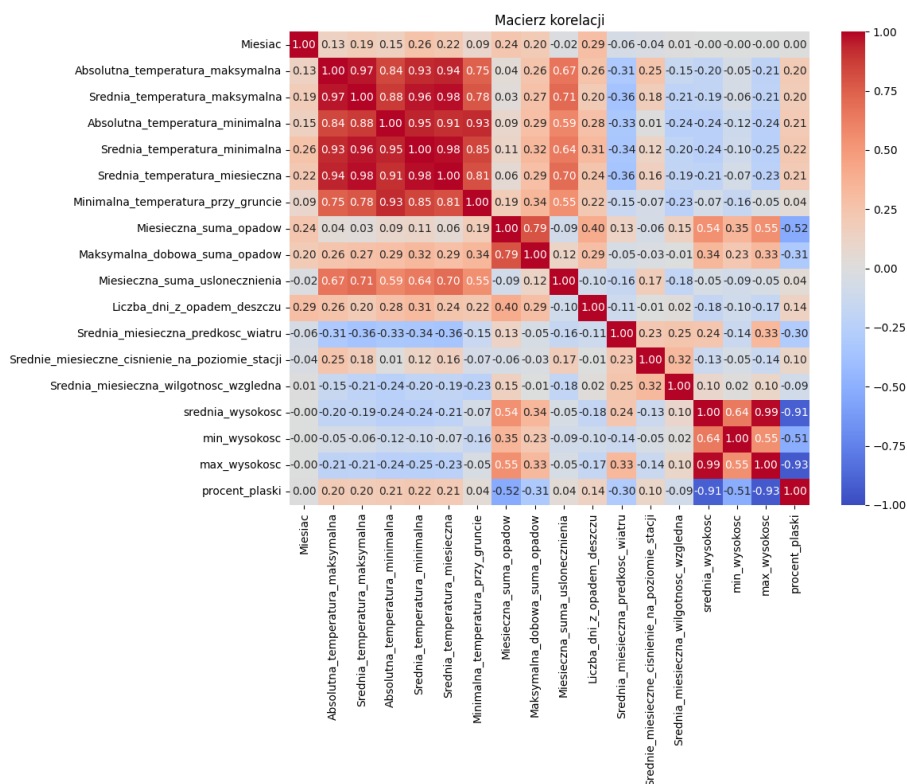


Rys. 21. Rozkład wysokości we wszystkich plikach NMT

4. Analiza relacji między zmiennymi

Przeprowadzona została analiza relacji między zmiennymi meteorologicznymi w celu zrozumienia wzajemnych powiązań między temperaturą, opadami, usłonecznieniem, wilgotnością i innymi czynnikami klimatycznymi. W wyniku tej analizy powstała macierz korelacji widoczna na Rys. 22.

- Występuje wysoka dodatnia korelacja między wszystkimi zmiennymi dotyczącymi temperatury, wynosząca od 0,75 do 0,98.
- Ustónecznienie wykazuje wysoką korelację z temperaturami, w przedziale od 0,55 do 0,71.
- Miesięczna suma opadów jest silnie skorelowana z maksymalną dobową sumą opadów (0,79).
- Miesięczna suma opadów ma umiarkowaną korelację z liczbą dni z opadem deszczu (0,4).
- Średnia miesięczna wilgotność względna wykazuje umiarkowaną korelację ze średnim miesięcznym ciśnieniem na poziomie stacji (0,32).
- Średnia miesięczna prędkość wiatru wykazuje umiarkowaną ujemną korelację z temperaturami, w zakresie od -0,36 do -0,15.
- Zmienne dotyczące temperatur wykazują słabą ujemną korelację ze średnią miesięczną wilgotnością względną, od -0,24 do -0,15.
- Wszystkie zmienne wysokościowe wykazują między sobą duże korelacje zarówno dodatnie jak i ujemne.
- Miesięczna suma opadów ma umiarkowaną dodatnią korelację (0,54-0,55) ze średnią oraz maksymalną wysokością oraz umiarkowaną ujemną korelację (-0,52) z procentem płaskiego terenu.
- Procent płaski wykazują niską korelację dodatnią ze zmiennymi pogodowymi, od 0.2 do 0.22.
- Średnia i maksymalna wysokość wykazują niską ujemną korelację ze zmiennymi temperaturowymi, od -0.25 do -0.19.



Rys. 22. Macierz wariancji dla połączonego zestawu danych