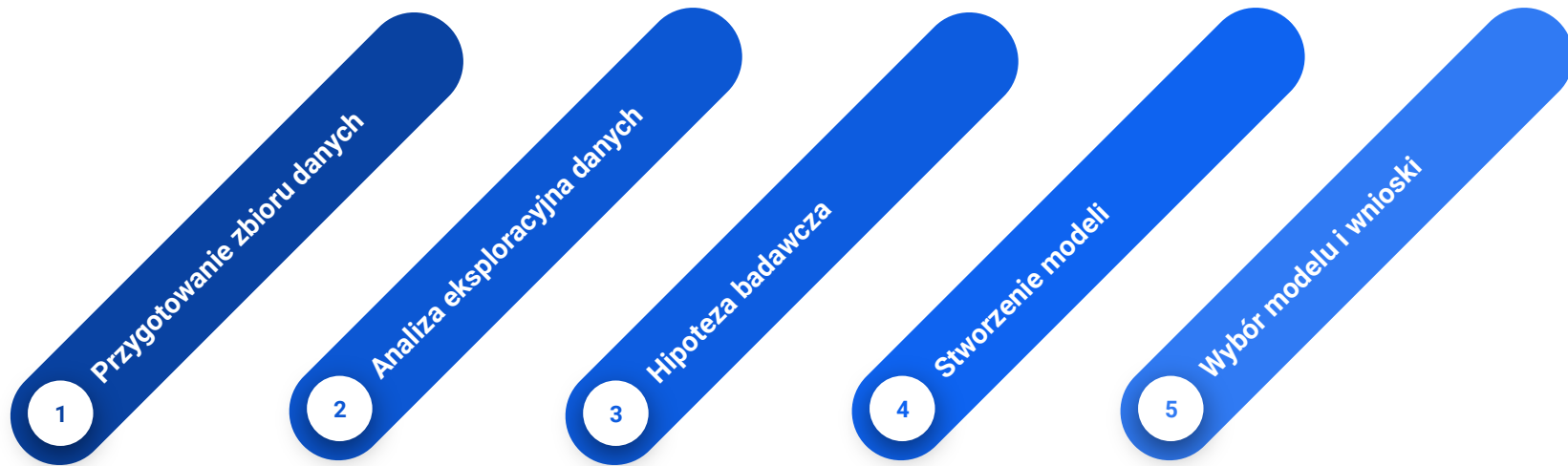




Analiza danych o warunkach pogodowych w Londynie

Daria Danieluk | Weronika Zbierowska

Etapy



Zbiór danych

Dzienne pomiary warunków
pogodowych ze stacji
meteorologicznej w pobliżu lotniska
Heathrow w Londynie
w latach 1979-2020

dane udostępnione przez:
European Climate Assessment (ECA)



<https://www.kaggle.com/datasets/emmanuelwerr/london-weather-data>

<https://www.ecad.eu/dailydata/index.php>

Dzienne pomiary poziomu
rezerwuarów wodnych w Londynie
w latach 1989-2023

dane udostępnione przez:
Environment Agency



<https://data.london.gov.uk/dataset/london-reservoir-levels>

<https://www.gov.uk/government/collections/water-situation-reports-for-england>

Ramki danych

	date	cloud_cover	sunshine	global_radiation	max_temp	mean_temp	min_temp	precipitation	pressure	snow_depth
0	19790101	2.0	7.0	52.0	2.3	-4.1	-7.5	0.4	101900.0	9.0
1	19790102	6.0	1.7	27.0	1.6	-2.6	-7.5	0.0	102530.0	8.0
2	19790103	5.0	0.0	13.0	1.3	-2.8	-7.2	0.0	102050.0	4.0
3	19790104	8.0	0.0	13.0	-0.3	-2.6	-6.5	0.0	100840.0	2.0
4	19790105	6.0	2.0	29.0	5.6	-0.8	-1.4	0.0	102250.0	1.0

	date	month	year	lower_lee_group	lower_thames_group
0	01-Jan-89	Jan	1989	83	76
1	02-Jan-89	Jan	1989	83	75
2	03-Jan-89	Jan	1989	83	75
3	04-Jan-89	Jan	1989	82	75
4	05-Jan-89	Jan	1989	82	75

Przygotowanie ramek danych

	date	month	year	lower_lee_group	lower_thames_group
0	01-Jan-89	Jan	1989	83	76
1	02-Jan-89	Jan	1989	83	75
2	03-Jan-89	Jan	1989	83	75
3	04-Jan-89	Jan	1989	82	75
4	05-Jan-89	Jan	1989	82	75



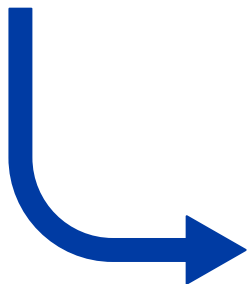
	parsed_date	lower_thames_group
0	19890101	76
1	19890102	75
2	19890103	75
3	19890104	75
4	19890105	75

df_reservoir - stan początkowy

df_reservoir - po przygotowaniu

Zbiór danych po połączeniu ramek

	date	cloud_cover	sunshine	global_radiation	max_temp	mean_temp	min_temp	precipitation	pressure	snow_depth	lower_thames_group
0	19890101	7.0	0.0	13.0	8.2	7.4	6.6	0.0	103870.0	0.0	76.0
1	19890102	6.0	0.0	13.0	9.5	7.1	6.0	0.0	103550.0	0.0	75.0
2	19890103	8.0	0.0	13.0	11.2	6.2	3.0	0.4	102950.0	0.0	75.0
3	19890104	5.0	3.8	39.0	11.1	8.0	4.8	0.0	102530.0	0.0	75.0
4	19890105	5.0	0.0	13.0	10.7	6.2	1.2	3.7	102340.0	0.0	75.0

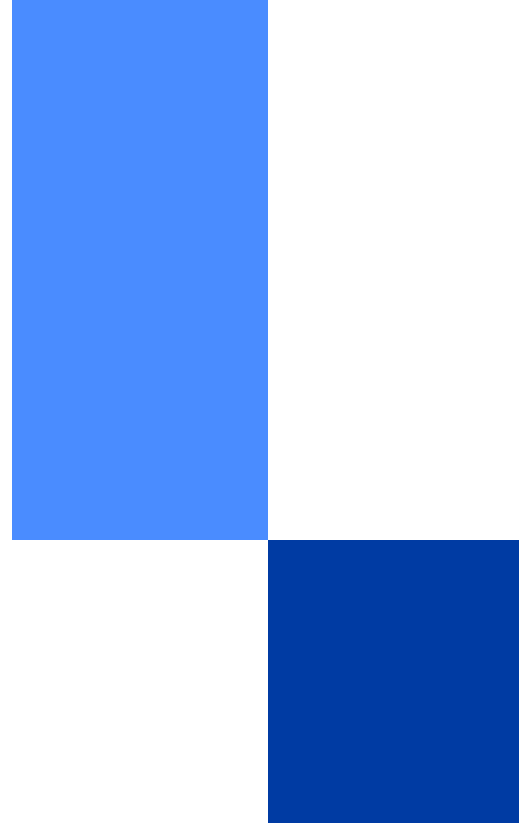


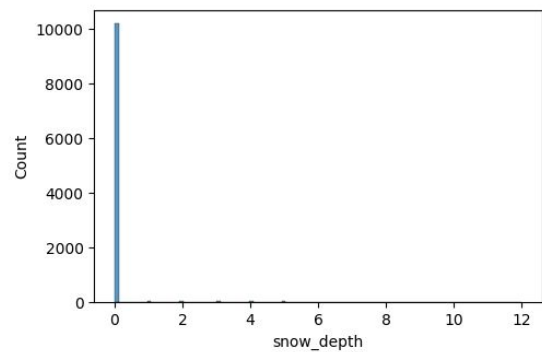
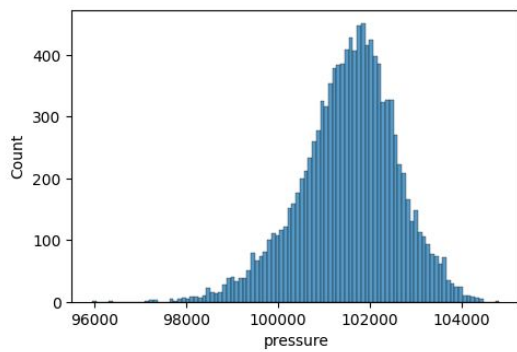
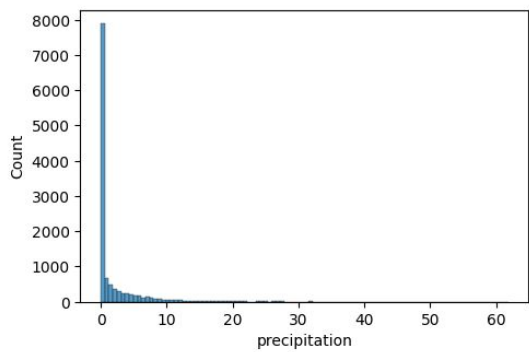
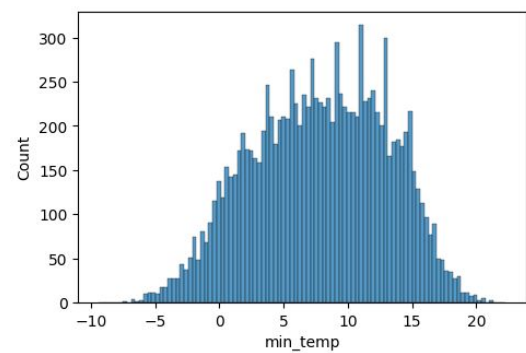
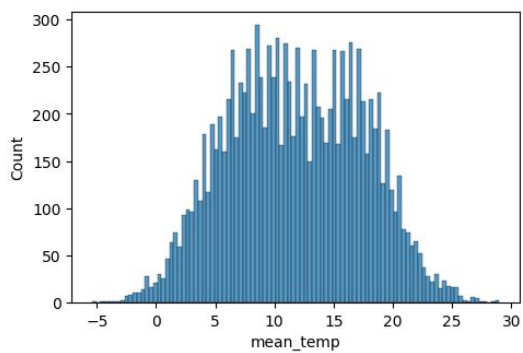
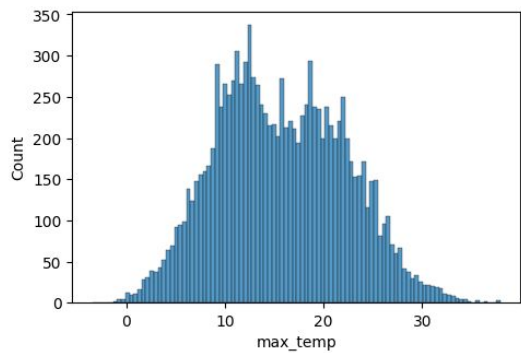
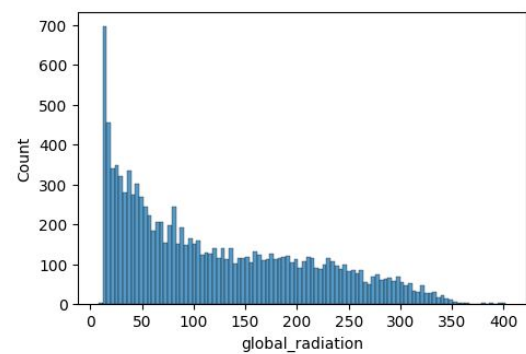
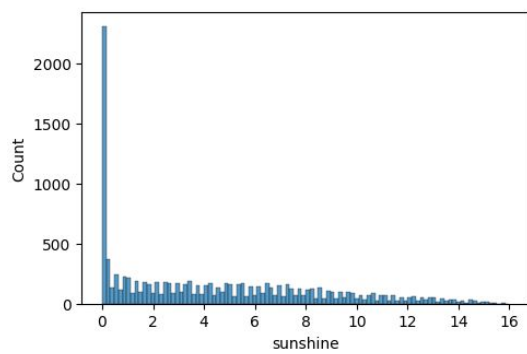
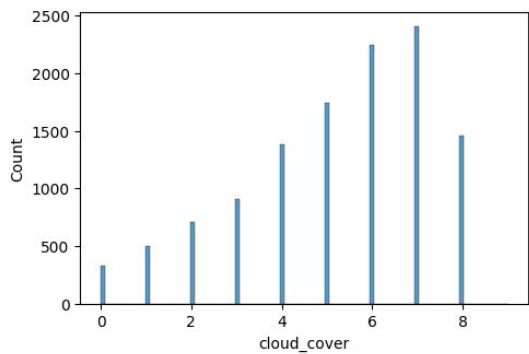
year	month
1989	1
1989	1
1989	1
1989	1
1989	1

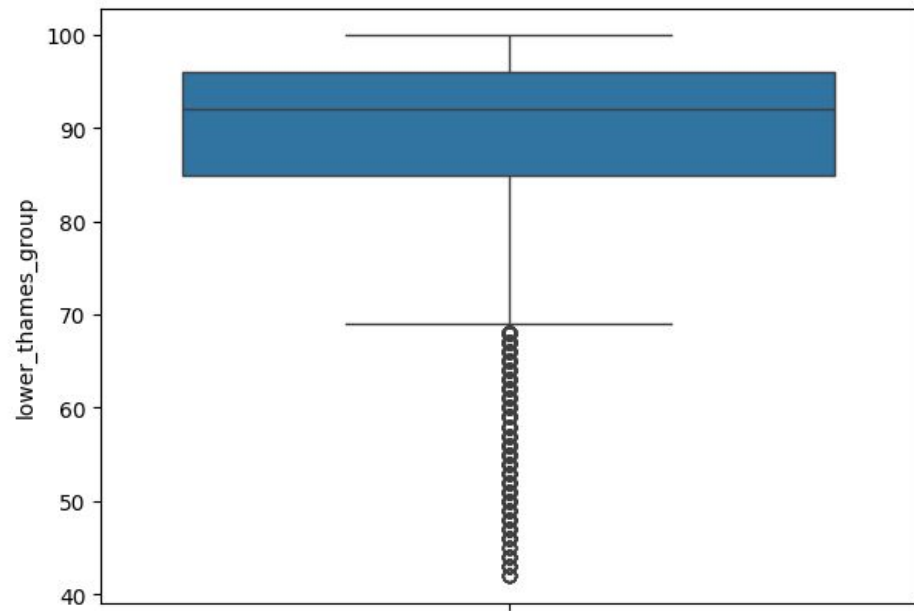
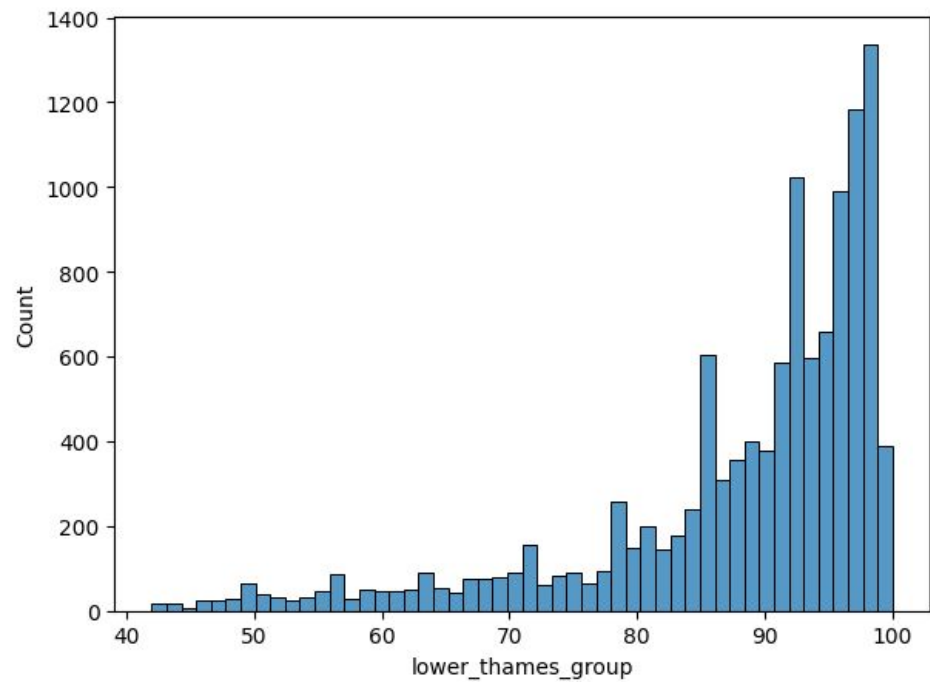
12 atrybutów

11688 obserwacji

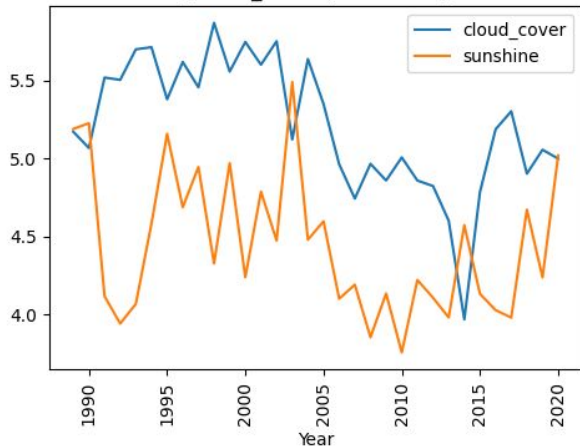
Analiza zmienności



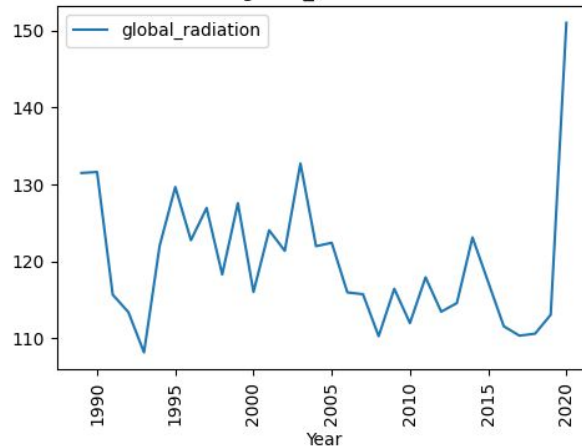




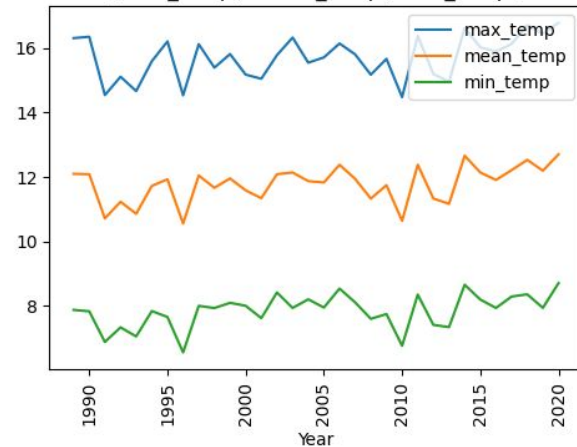
['cloud_cover', 'sunshine']



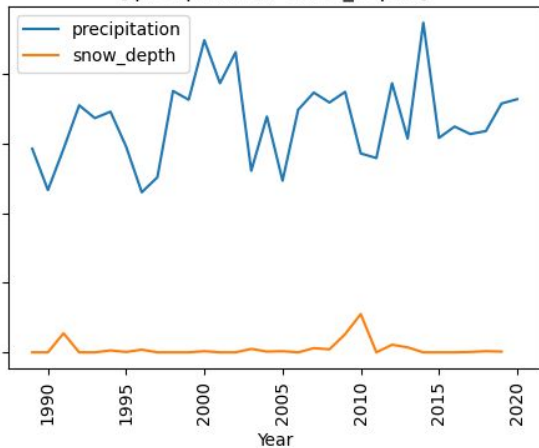
global_radiation



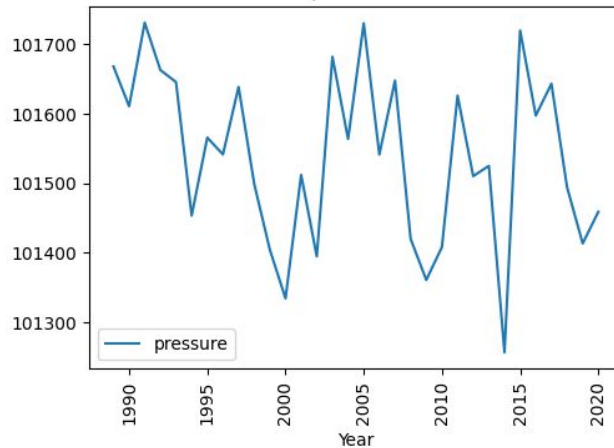
['max_temp', 'mean_temp', 'min_temp']



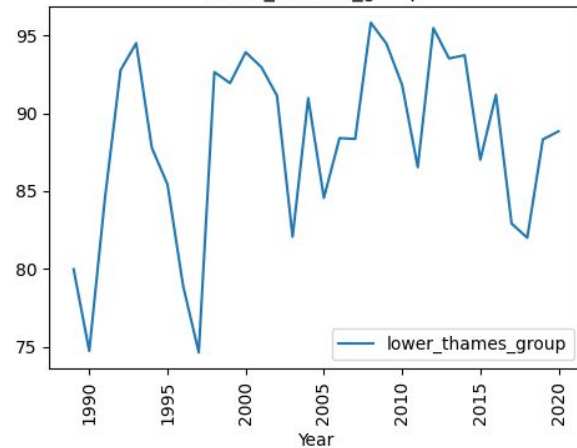
['precipitation', 'snow_depth']

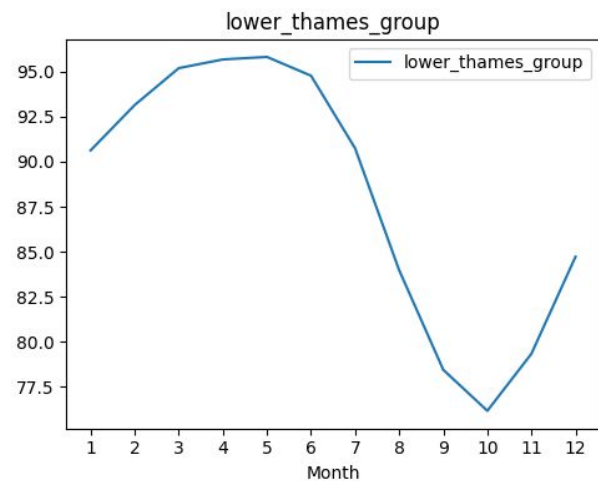
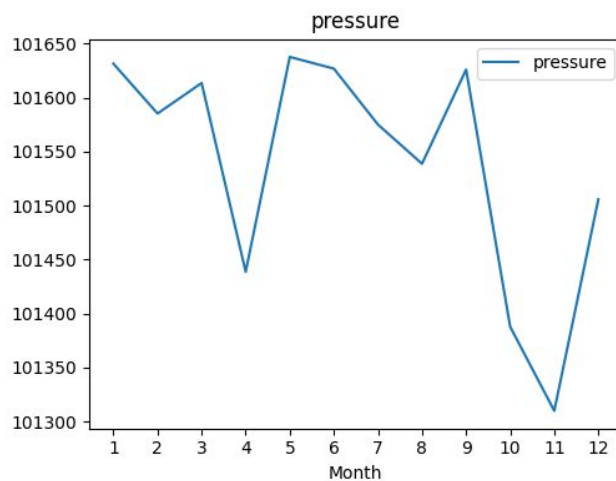
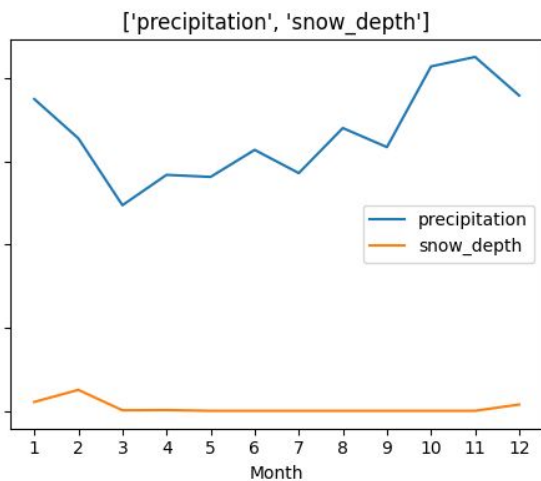
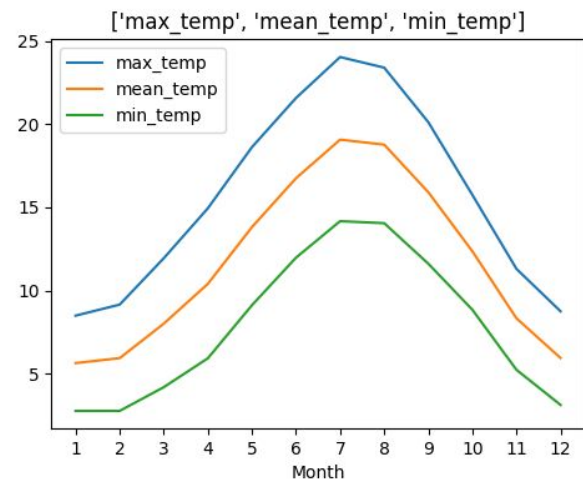
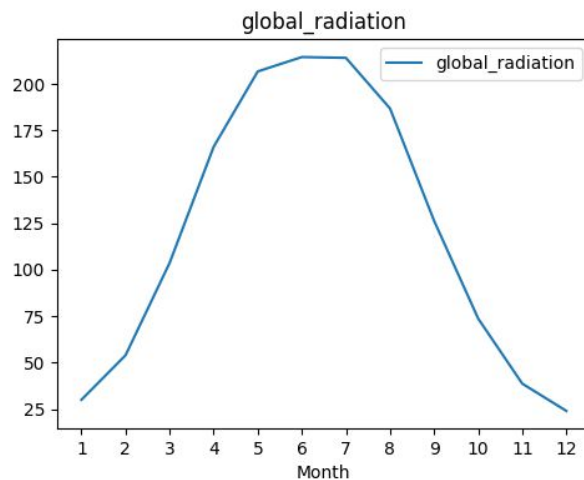
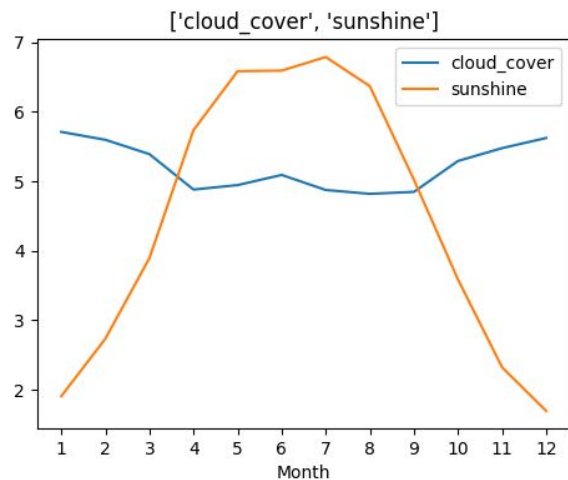


pressure



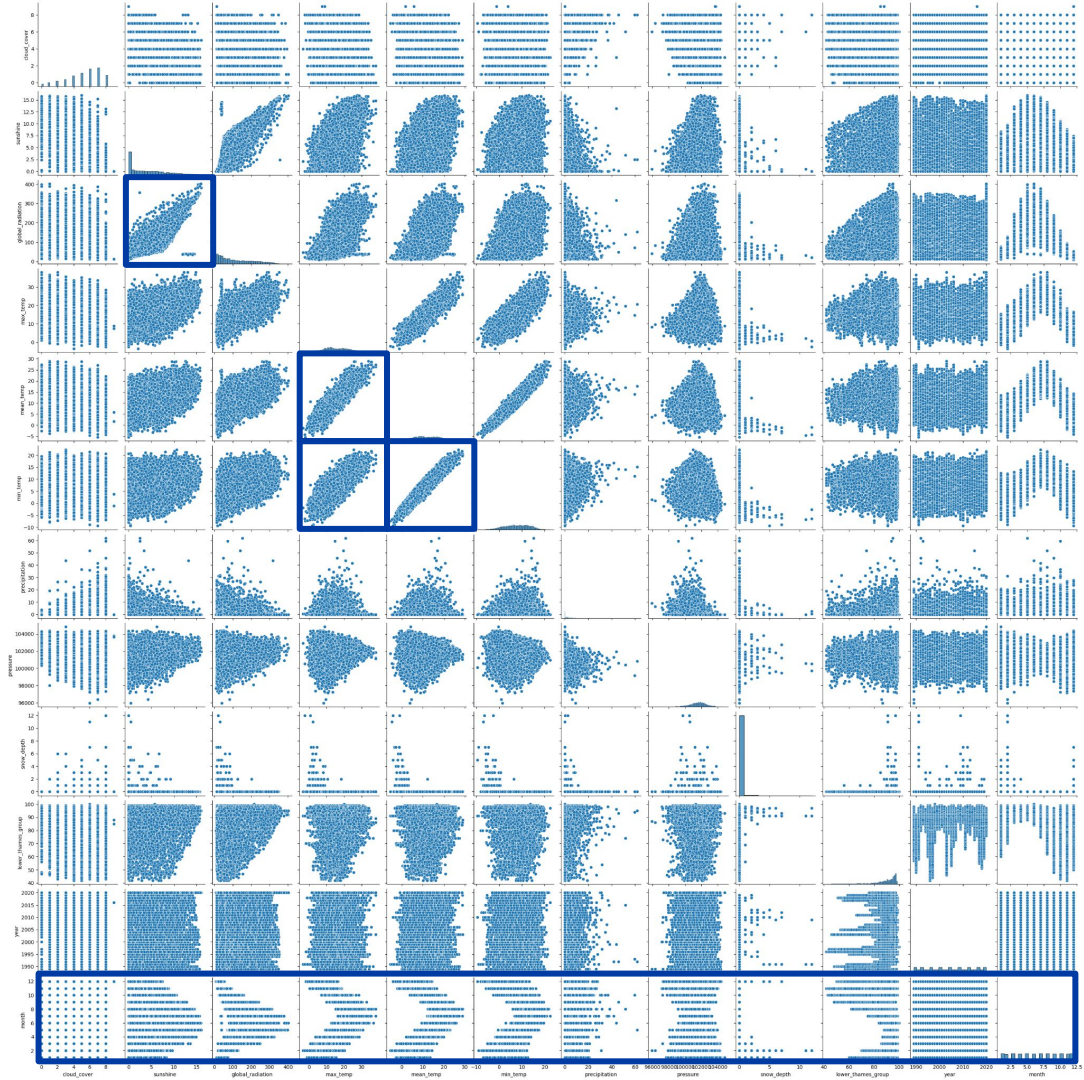
lower_thames_group

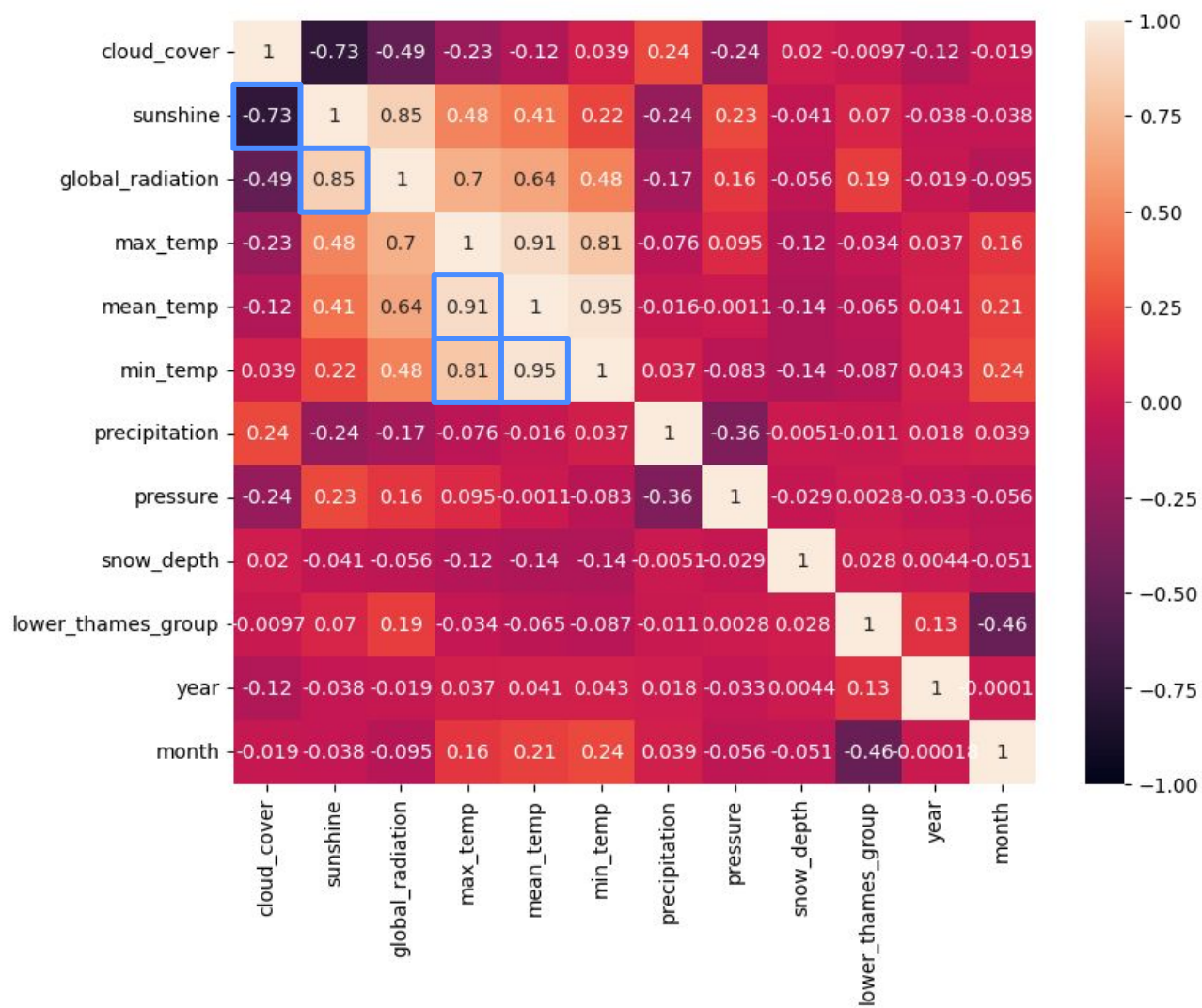




Analiza zależności między atributami







Braki w danych

Braki w 'snow_depth':

- 2020 - brak wszystkich danych
- 2006-2011 - brak ok. 1/3 danych
- 2005, 2015 - pojedyncze braki

```
Year 2009
2      3
4      23
5      31
6      30
7      31
8      31
9      30
10     1
12     1
Name: month
```

```
cloud_cover      6
sunshine         0
global_radiation 19
max_temp         6
mean_temp        36
min_temp         2
precipitation     6
pressure          3
snow_depth       1441
lower_thames_group 13
year             0
month            0
```

Hipoteza:

Jaki wpływ na procent użytkowej pojemności wody w rezerwuarach 'Grupy Dolnej Tamizy' mają warunki pogodowe w Londynie i okres ich pomiarów?

Przygotowanie zbioru danych

Imputacja danych

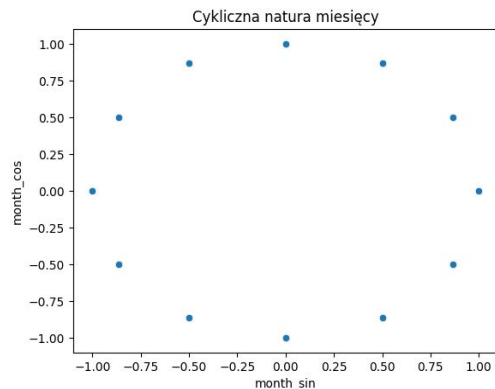
- **mediana:**
'cloud_cover'
- **średnia:**
'global_radiation',
'max_temp',
'mean_temp',
'min_temp',
'precipitation',
'pressure',
'snow_depth'
- **usunięcie:**
'lower_thames_group'

Utworzenie nowych cech

- month_sin
- month_cos
- log_sunshine
- log_global_radiation
- log_precipitation
- log_snow_depth

Podział zbioru na uczący i testowy

Standaryzacja

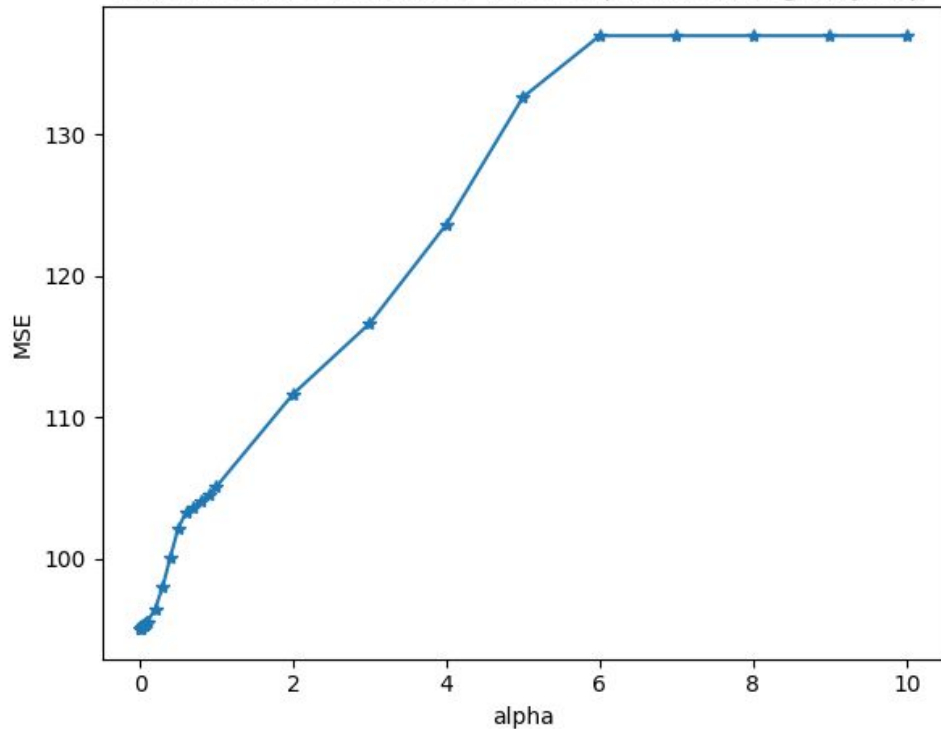


Utworzone modele

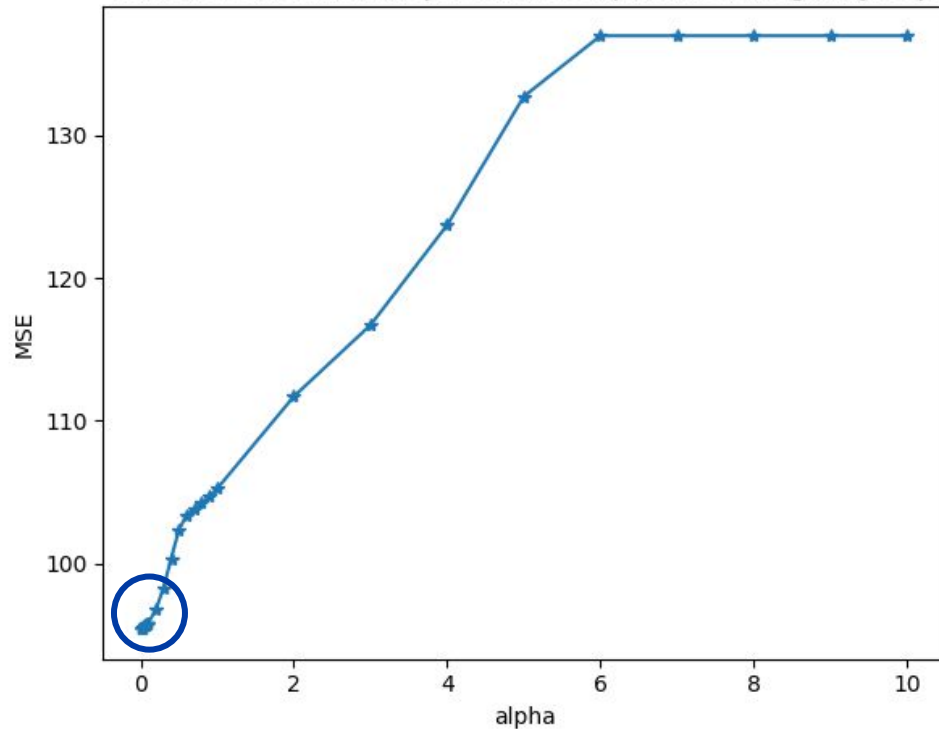
Z oryginalnymi cechami		
regresja wieloraka bez regularyzacji	regresja grzbietowa (L2)	regresja Lasso (L1)
Z oryginalnymi oraz nowo utworzonymi cechami		
regresja wieloraka bez regularyzacji	regresja grzbietowa (L2)	regresja Lasso (L1)
Z automatycznym doбором cech: regresja krokowa w przód		
regresja wieloraka bez regularyzacji	regresja grzbietowa (L2)	regresja Lasso (L1)
Z automatycznym doбором cech: regresja krokowa w tył		
regresja wieloraka bez regularyzacji	regresja grzbietowa (L2)	regresja Lasso (L1)

Dobór współczynnika regularyzacji

Zależność MSE uczenia od wartości parametru regularyzacji

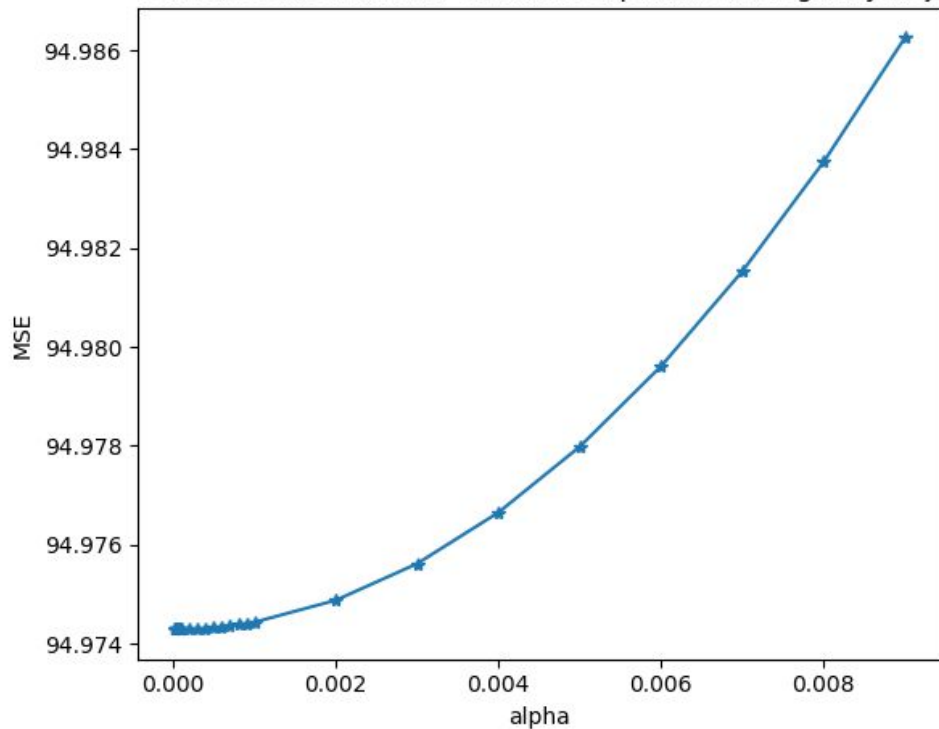


Zależność MSE walidacji od wartości parametru regularyzacji

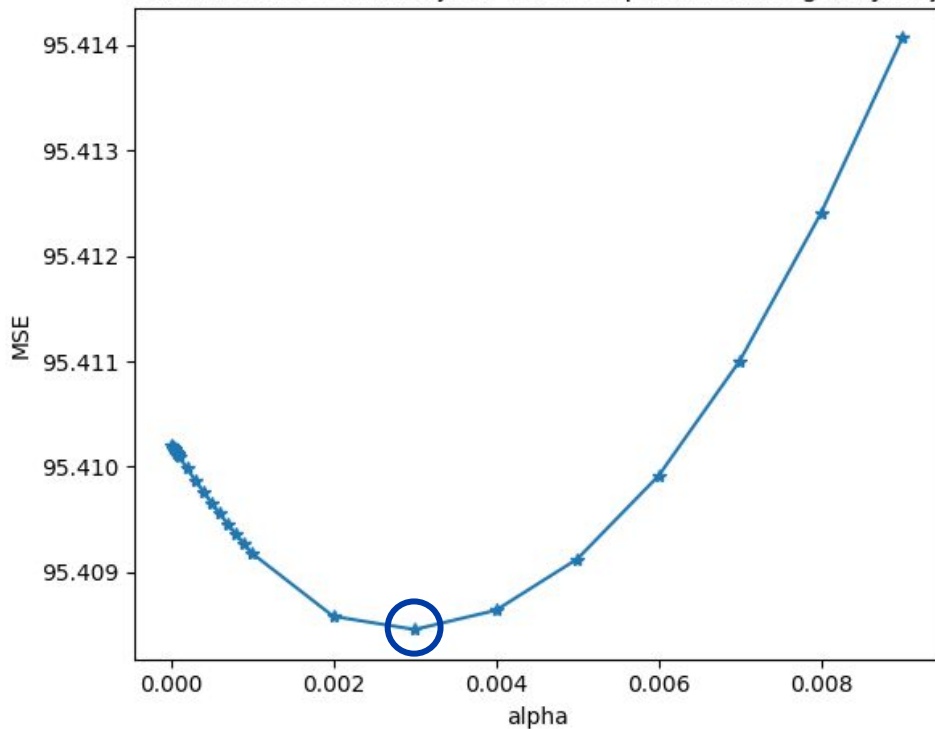


Dobór współczynnika regularyzacji

Zależność MSE uczenia od wartości parametru regularyzacji



Zależność MSE walidacji od wartości parametru regularyzacji



Cechy wybrane przez regresję krokową w przód

- 'month_sin'
- 'month_cos'
- 'year',
- 'log_global_radiation'
- 'global_radiation'
- **'max_temp'**
- 'month'
- 'log_sunshine'
- 'sunshine'
- 'cloud_cover'
- 'precipitation'
- 'min_temp'
- 'mean_temp'

Wybór 13 z 17 cech

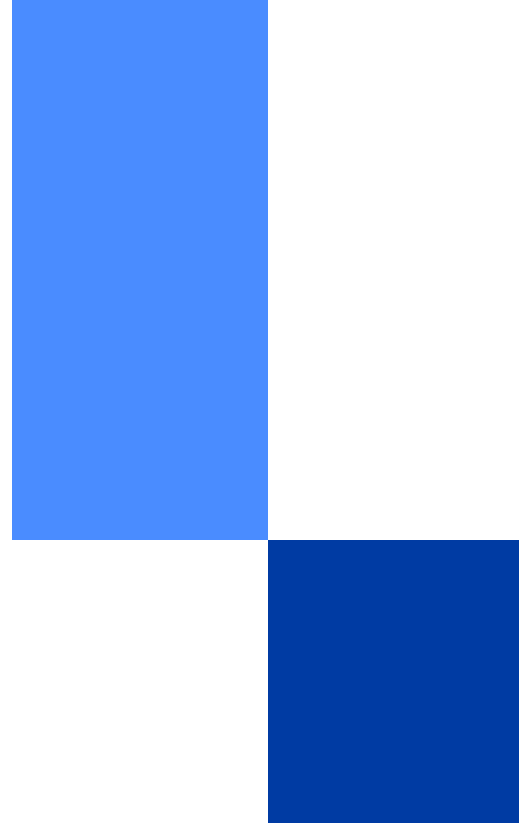
Cechy wybrane przez regresję krokową w tył

- 'cloud_cover'
- 'sunshine'
- 'global_radiation'
- 'mean_temp'
- 'min_temp'
- 'precipitation'
- **'snow_depth'**
- 'year'
- 'month'
- 'month_sin'
- 'month_cos'
- 'log_sunshine'
- 'log_global_radiation'
- **'log_snow_depth'**

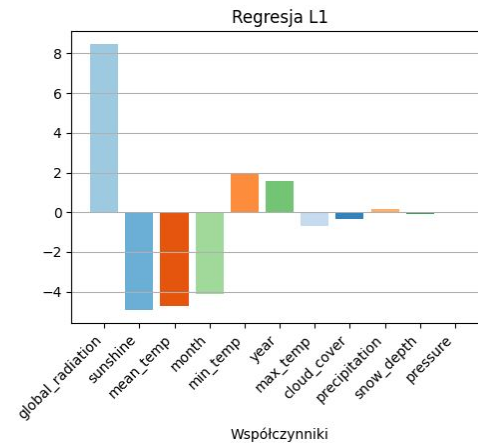
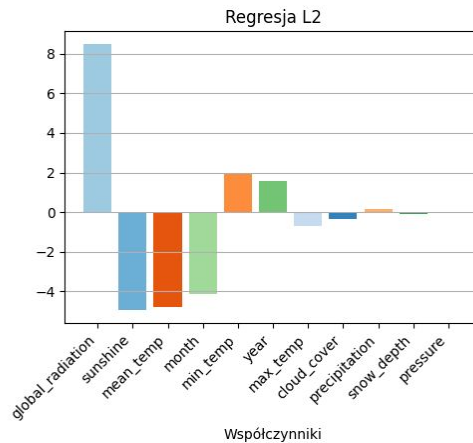
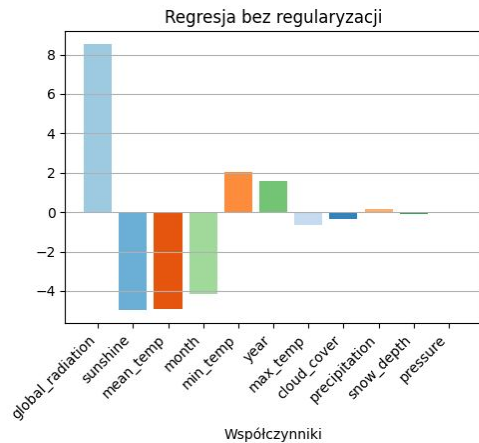
Wybór 14 z 17 cech

'pressure', 'log_precipitation' - cechy odrzucone przez obydwie metody

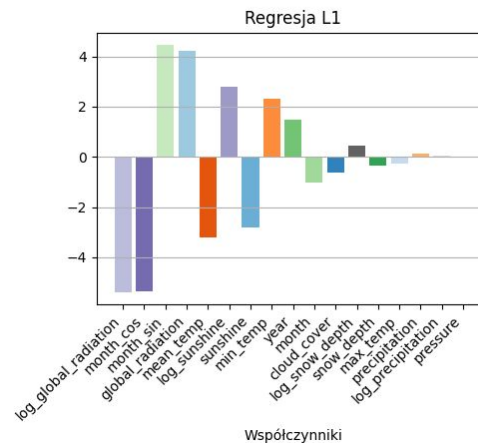
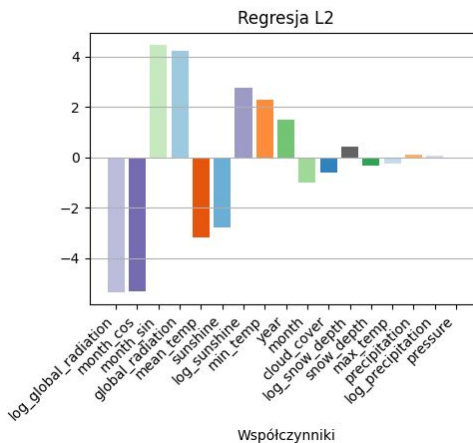
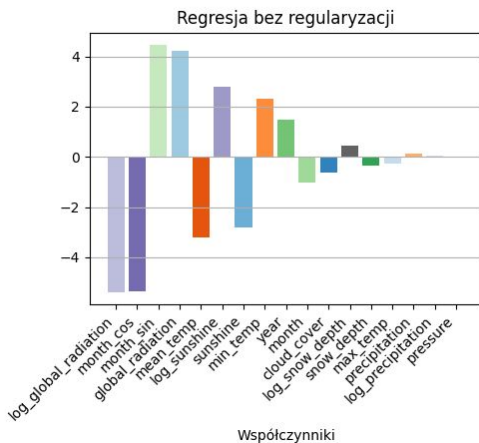
Porównanie modeli



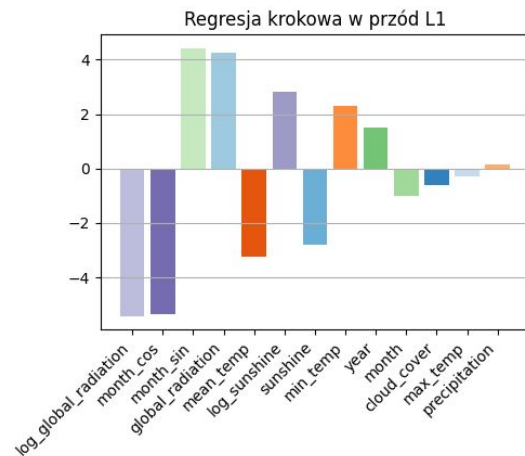
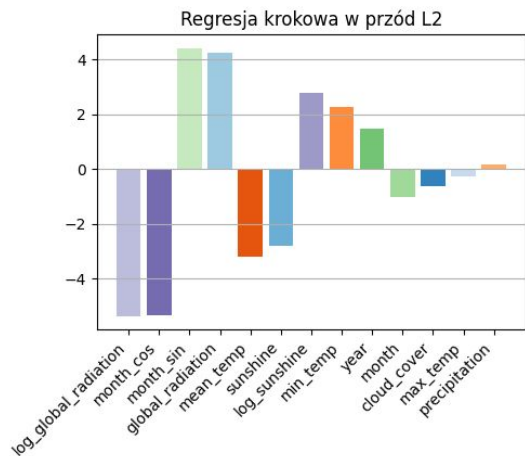
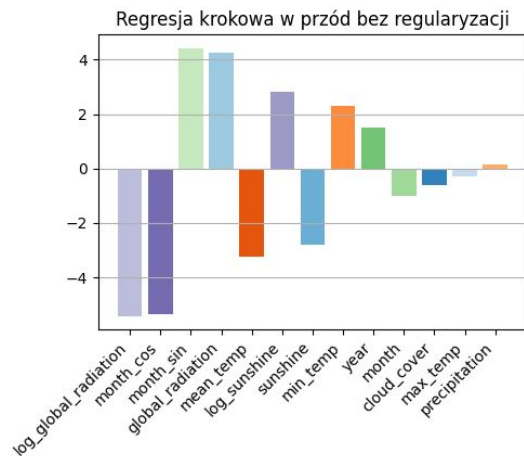
Porównanie modeli regresji z oryginalnymi cechami



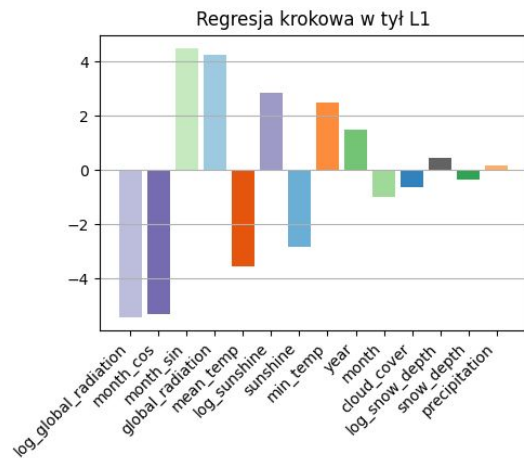
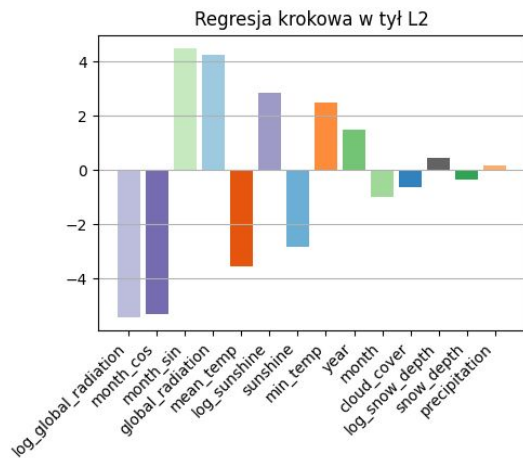
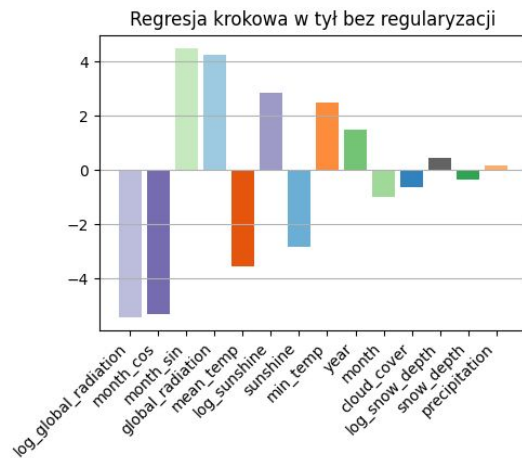
Porównanie modeli regresji z nowymi cechami



Porównanie modeli regresji z cechami wybranymi metodą krokową w przód



Porównanie modeli regresji z cechami wybranymi metodą krokową w tył

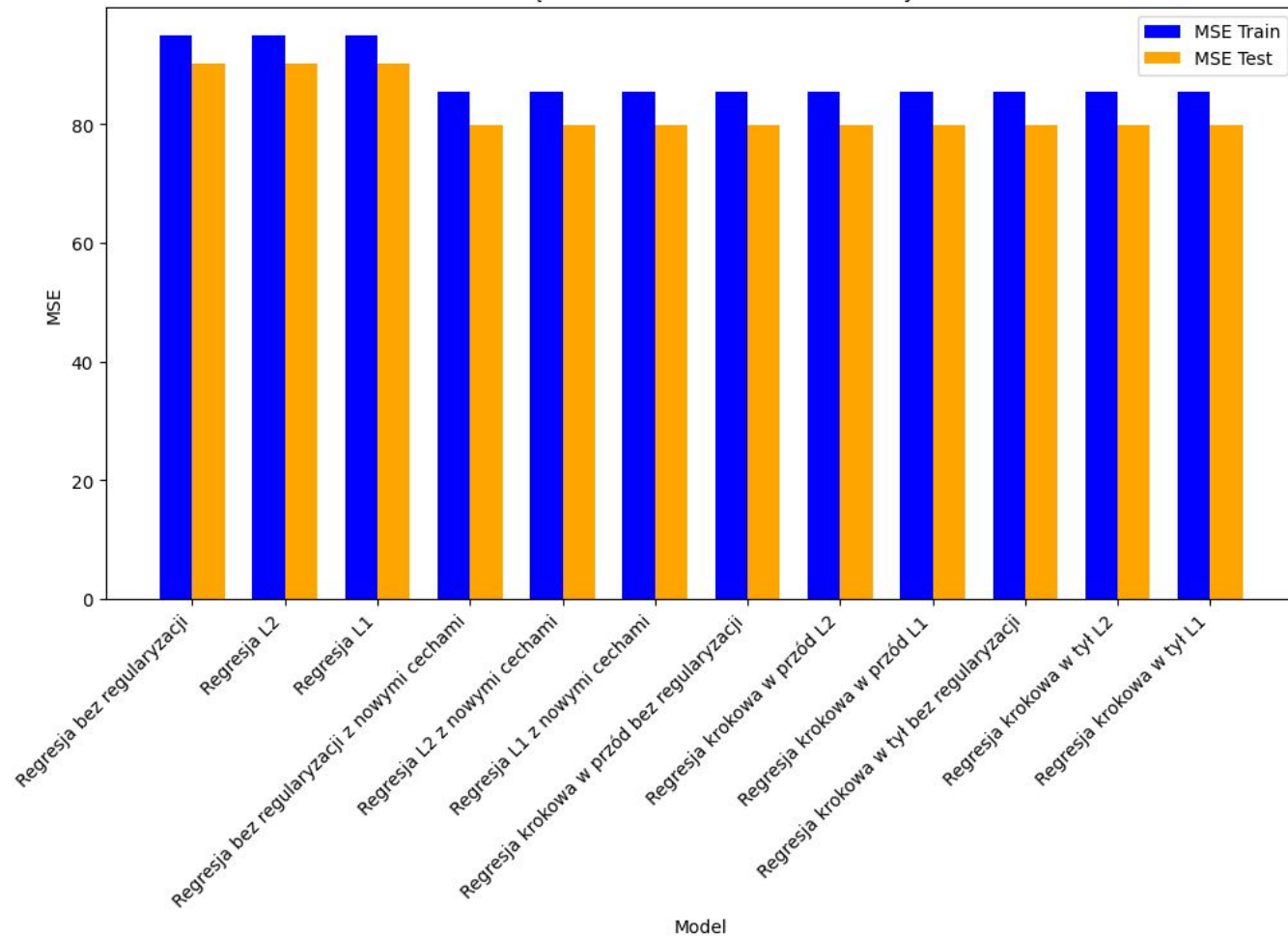


Współczynniki

Współczynniki

Współczynniki

Porównanie błędów uczenia i testowania dla różnych modeli



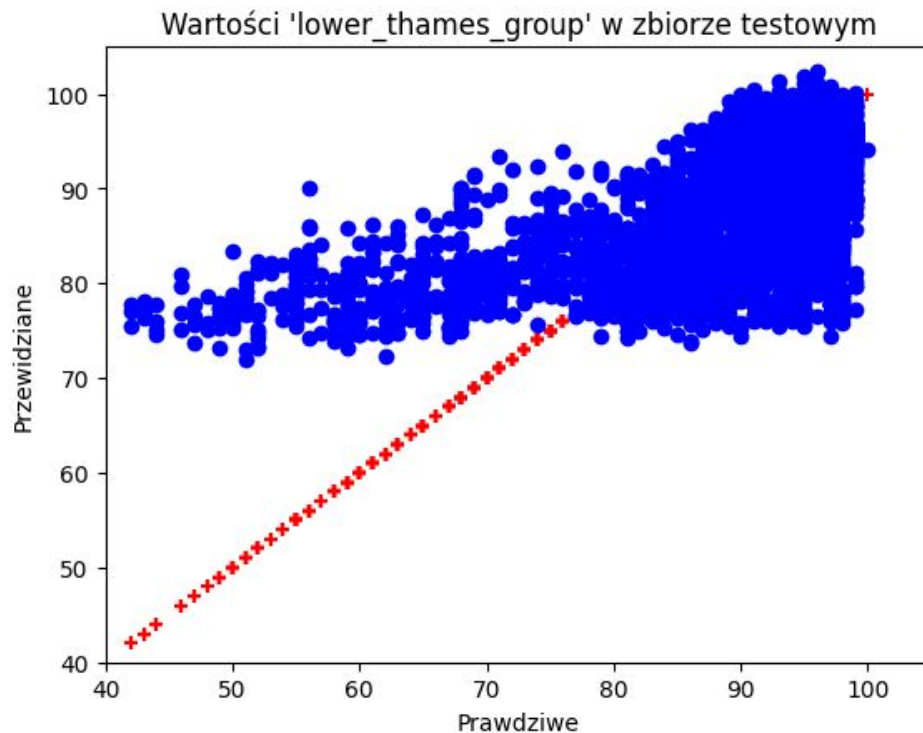
Model	MSE Train	MSE Test
Regresja bez regularyzacji	94.9743	90.2442
Regresja L2	94.9747	90.2238
Regresja L1	94.9756	90.21
Regresja bez regularyzacji z nowymi cechami	85.4422	79.9383
Regresja L2 z nowymi cechami	85.4423	79.9399
Regresja L1 z nowymi cechami	85.4422	79.9383
Regresja krokowa w przód bez regularyzacji	85.4687	79.9445
Regresja krokowa w przód L2	85.4687	79.9465
Regresja krokowa w przód L1	85.4687	79.9445
Regresja krokowa w tył bez regularyzacji	85.4504	79.9669
Regresja krokowa w tył L2	85.4504	79.9669
Regresja krokowa w tył L1	85.4504	79.9669

Wybór modelu

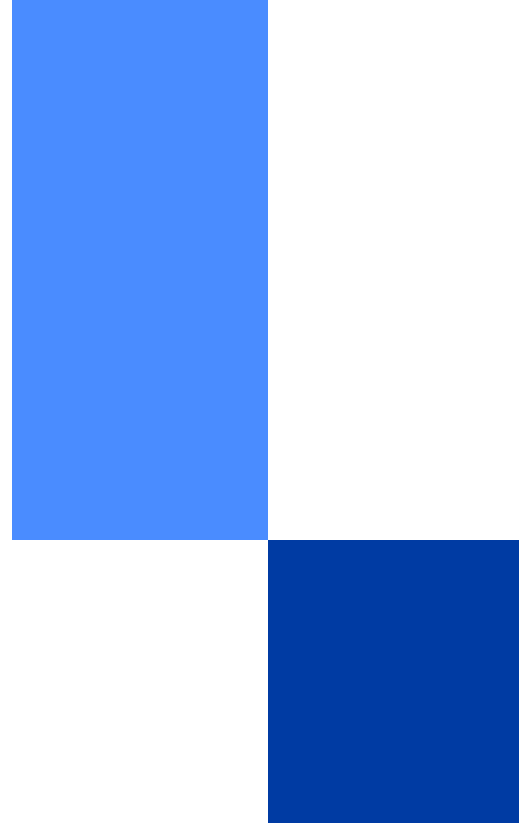
regresja bez regularyzacji
z nowymi cechami

MSE train: 85.442231

MSE test: 79.938269



Wnioski



**Dziękujemy
za uwagę**