

WPŁYW GLOBALNEGO OCIEPLENIA NA KLIMAT,

CZYLI BADANIE ZMIAN TEMPERATURY, ZASIĘGU LODU MORSKIEGO
ORAZ POKRYWY ŚNIEŻNEJ

AUTORZY:

Alicja Dąbrowska (260349)
Anna Pawlik (260339)



ŹRÓDŁA DANYCH:

National Centers for Environmental Information:

<https://www.ncdc.noaa.gov/snow-and-ice/extent/>

Narodowa Agencja Aeronautyki i Przestrzeni Kosmicznej Goddard Institute for Space Studies (NASA-GISS):

<https://www.fao.org/faostat/en/#data/ET>

JĘZYK PROGRAMOWANIA:

Python 3.10

BIBLIOTEKI:

statsmodels, matplotlib, numpy, pandas, seaborn, sklearn, pmdarima

SPIS TREŚCI

1.	WSTĘP.....	2
1.1.	CEL.....	2
1.2.	DANE I NARZĘDZIA.....	2
2.	ZMIANA TEMPERATURY NA PRZESTRZENI LAT.....	2
2.1.	ŽRÓDŁO DANYCH.....	2
2.2.	PRZYGOTOWANIE I WSTĘPNA ANALIZA DANYCH.....	2
2.3.	MODEL.....	7
2.4.	ANALIZA WYNIKÓW.....	10
2.5.	WNIOSKI.....	10
3.	ZMIANA ZASIĘGU LODU MORSKIEGO.....	10
3.1.	ŽRÓDŁO DANYCH.....	10
3.2.	PRZYGOTOWANIE I WSTĘPNA ANALIZA DANYCH.....	10
3.3.	MODEL.....	12
3.4.	ANALIZA WYNIKÓW.....	16
3.5.	WNIOSKI.....	16
4.	ZMIANA ZASIĘGU POKRYWY ŚNIEŻNEJ.....	16
4.1.	ŽRÓDŁO DANYCH.....	16
4.2.	PRZYGOTOWANIE I WSTĘPNA ANALIZA DANYCH.....	16
4.3.	MODEL.....	19
4.4.	ANALIZA WYNIKÓW.....	25
4.5.	WNIOSKI.....	25
5.	ZAKOŃCZENIE I WNIOSKI.....	25

1. WSTĘP

CEL

Celem projektu było bliższe przyjrzenie się skutkom wzrostu średniej temperatury, który obserwowany jest w ostatnich latach. Zjawisko to nazywane jest globalnym ociepleniem. W projekcie ujęte zostały trzy czynniki: temperatura, zasięg lodu morskiego oraz zasięg pokrywy śnieżnej.

DANE I NARZĘDZIA

Dane temperaturowe, na których zostały oparte analizy opierają się na publicznie dostępnych danych GISTEMP, danych dotyczących globalnej zmiany temperatury powierzchni, rozpowszechnianych przez Narodową Agencję Aeronautyki i Przestrzeni Kosmicznej Goddard Institute for Space Studies (NASA-GISS). Dane dotyczące pokrywy śnieżnej opierają się na danych Rutgers University Global Snow Laboratory (GSL), natomiast dane nawiązujące do zasięgu lodu morskiego pochodzą z National Snow and Ice Data Center (NSIDC). W projekcie do analiz zdecydowano się użyć języka Python i dostępnych w nim bibliotek, jako że jest to język programowania powszechnie używany do analizy danych i zawiera wiele bibliotek ją umożliwiających.

2. ZMIANA TEMPERATURY NA PRZESTRZENI LAT

ŹRÓDŁO DANYCH: <https://www.fao.org/faostat/en/#data/ET>

FAOSTAT (dane temperaturowe) rozpowszechnia statystyki zmian średniej temperatury powierzchni według krajów, z corocznymi aktualizacjami. Dane pochodzą z lat 1961-2021. Dane oparte są na publicznie dostępnych danych GISTEMP, danych dotyczących globalnej zmiany temperatury powierzchni, rozpowszechnianych przez Narodową Agencję Aeronautyki i Przestrzeni Kosmicznej Goddard Institute for Space Studies (NASA-GISS).

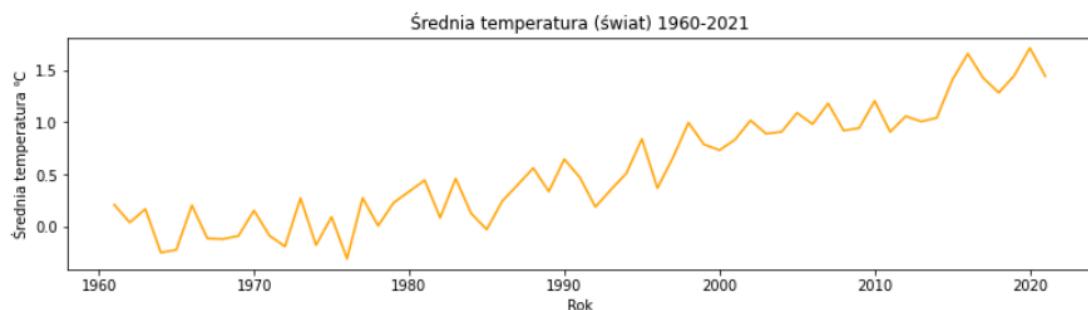
PRZYGOTOWANIE I WSTĘPNA ANALIZA DANYCH

1. Sprawdzenie kompletności danych, czy dane nie zawierają braków.
2. Usunięcie zbędnych kolumn i wizualizacja danych.
3. Sprawdzenie podstawowych statystyk - średnia, odchylenie standardowe, minimalna i maksymalna wartość oraz kwartyle Q1, Q2, Q3.
4. Zbadanie korelacji między upływem lat a średnią temperaturą.
5. Czynności te powtórzone zarówno jak i dla świata, jak i również badanych kontynentów.

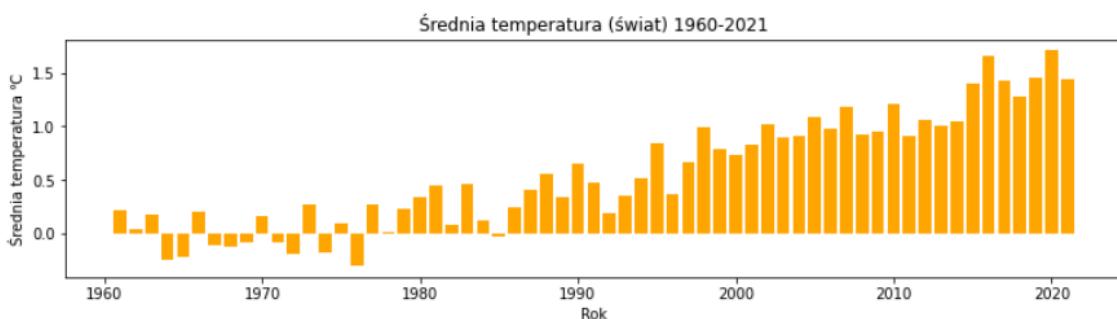
Zadane zostały dwa pytania:

- Czy na świecie temperatura rzeczywiście rośnie?
- Czy na każdym kontynencie zmiana jest tak samoauważalna?

Na początek zajęto się analizą na poziomie świata, wartości temperatury podawane w skali rocznej. Dane wyświetcone na dwa sposoby - wykresem liniowym oraz słupkowym.

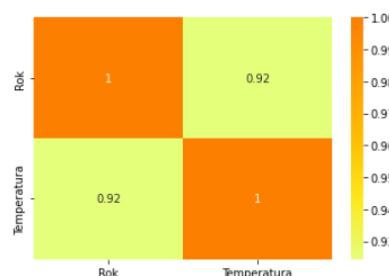


Wyk. 1.: Dane średniej temperatury na świecie w latach 1960-2021, wykres liniowy



Wyk. 2.: Dane średniej temperatury na świecie w latach 1960-2021, wykres słupkowy

Już tutaj widać znaczną zmianę średniej temperatury na świecie oraz jej tendencję wzrostową. Przyjrzano się też korelacji między temperaturą a rokiem. Okazała się ona bardzo silna i dodatnia, co oznacza, że wraz ze wzrostem jednej zmiennej, druga zmienna analogicznie rośnie. Zauważono to na poniższej macierzy korelacji.

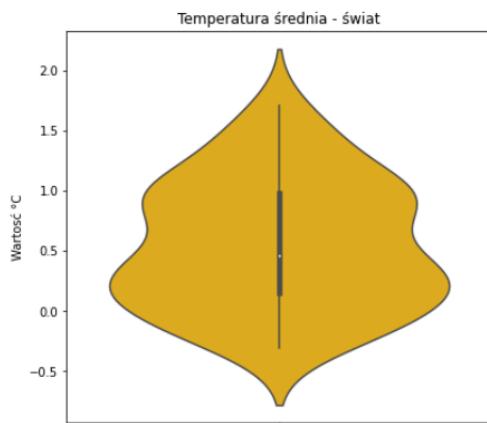


Rys. 1.: Macierz korelacji ukazująca korelację między upływającymi latami a średnią temperaturą na świecie

Sprawdzone zostały też podstawowe statystyki. Te podstawowe zależności zostały ukazane również na wykresie wiolinowym.

średnia	mean	0.555967
odchylenie standardowe	std	0.534132
wartość minimalna	min	-0.310000
kwartyl Q1	25%	0.153000
kwartyl Q2	50%	0.459000
kwartyl Q3	75%	0.982000
wartość maksymalna	max	1.713000

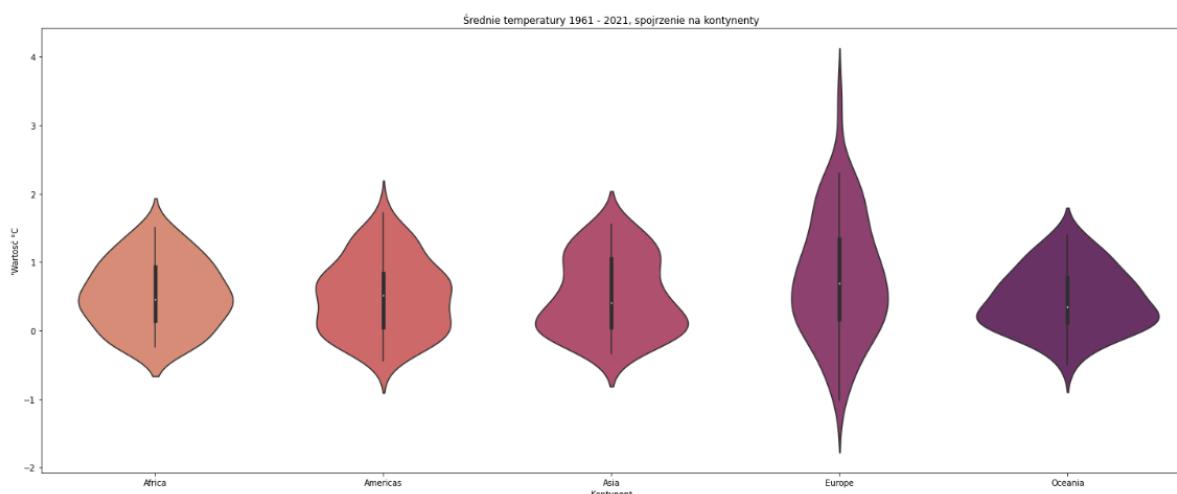
Rys. 2.: Podstawowe statystyki dotyczące średniej temperatury na świecie.



Wyk. 3 Wykres violinowy dotyczący średniej temperatury na świecie

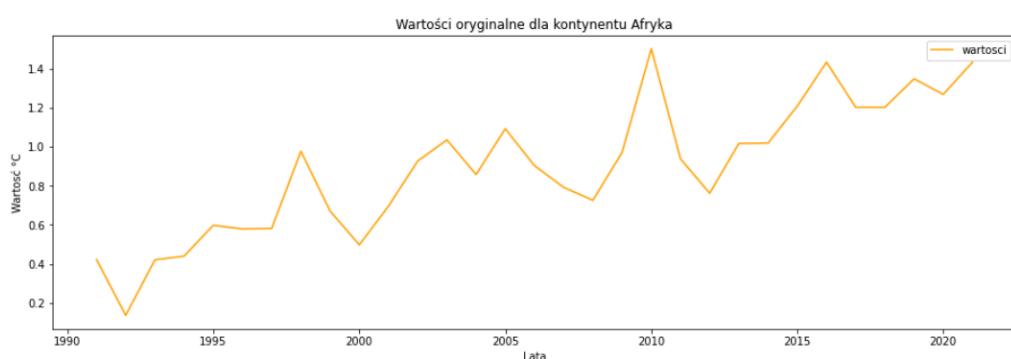
Kolejnym krokiem było przyjrzenie się każdemu kontynentowi z osobna, wykonane zostały te same kroki, przez które przeszliśmy analizując świat.

Wykres violinowy - kontynenty.



Wyk. 4.: Wykres violinowy dotyczący poszczególnych kontynentów

Afryka

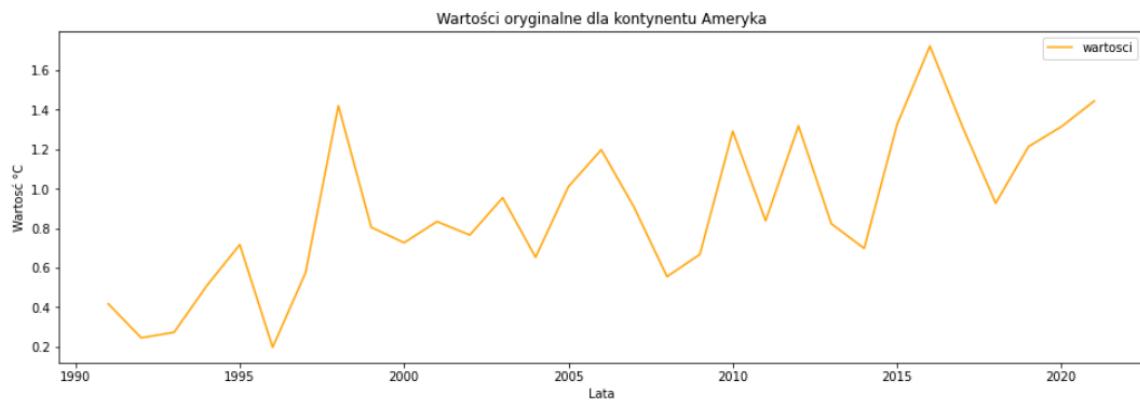


Wyk. 5.: Dane dotyczące średniej temperatury w Afryce w latach 1960-2021, wykres liniowy

średnia	mean	0.527443
odchylenie standardowe	std	0.487853
wartość minimalna	min	-0.238000
kwartyl Q1	25%	0.137000
kwartyl Q2	50%	0.455000
kwartyl Q3	75%	0.927000
wartość maksymalna	max	1.502000

Rys. 3.: Podstawowe statystyki dotyczące średniej temperatury w Afryce.

AMERYKA

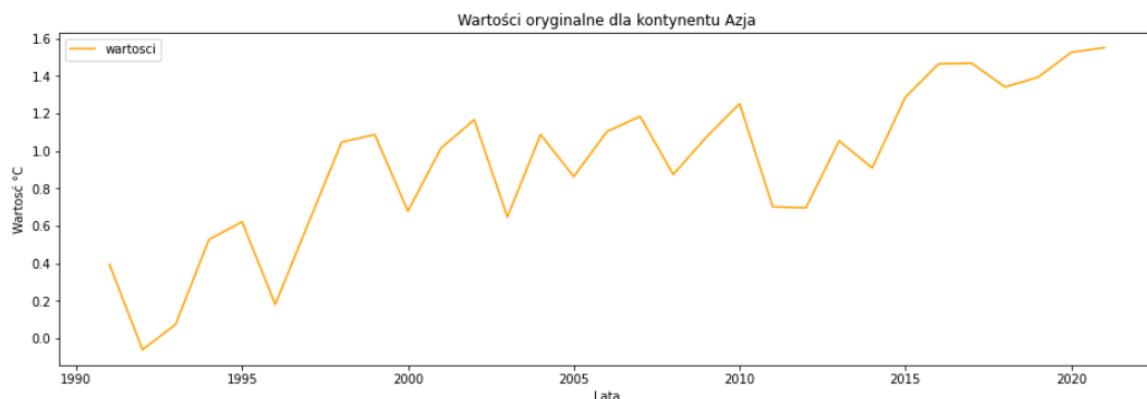


Wyk. 6.: Dane dotyczące średniej temperatury w Ameryce w latach 1960-2021, wykres liniowy

średnia	mean	0.501098
odchylenie standardowe	std	0.532237
wartość minimalna	min	-0.444000
kwartyl Q1	25%	0.051000
kwartyl Q2	50%	0.510000
kwartyl Q3	75%	0.834000
wartość maksymalna	max	1.723000

Rys. 4.: Podstawowe statystyki dotyczące średniej temperatury w Ameryce.

AZJA

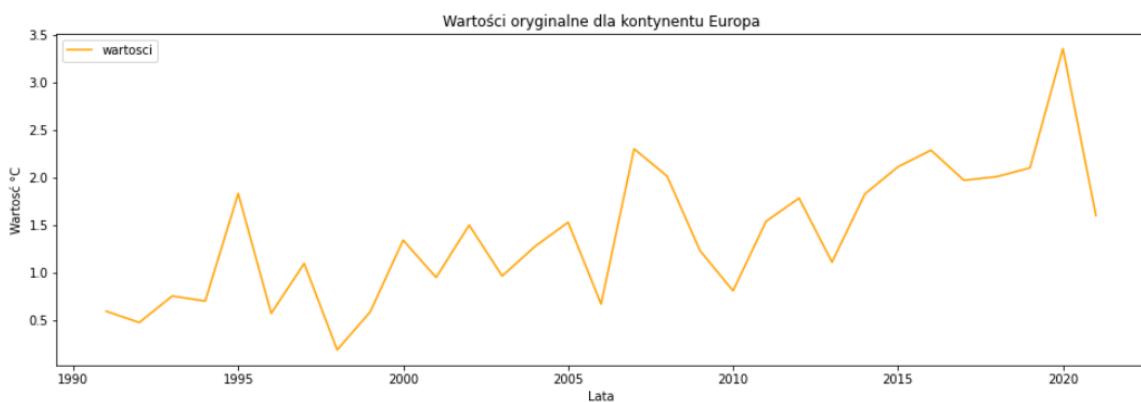


Wyk. 7.: Dane dotyczące średniej temperatury w Azji w latach 1960-2021, wykres liniowy

średnia	mean	0.509639
odchylenie standardowe	std	0.549537
wartość minimalna	min	-0.335000
kwartyl Q1	25%	0.050000
kwartyl Q2	50%	0.411000
kwartyl Q3	75%	1.047000
wartość maksymalna	max	1.552000

Rys. 5.: Podstawowe statystyki dotyczące średniej temperatury w Azji.

EUROPA

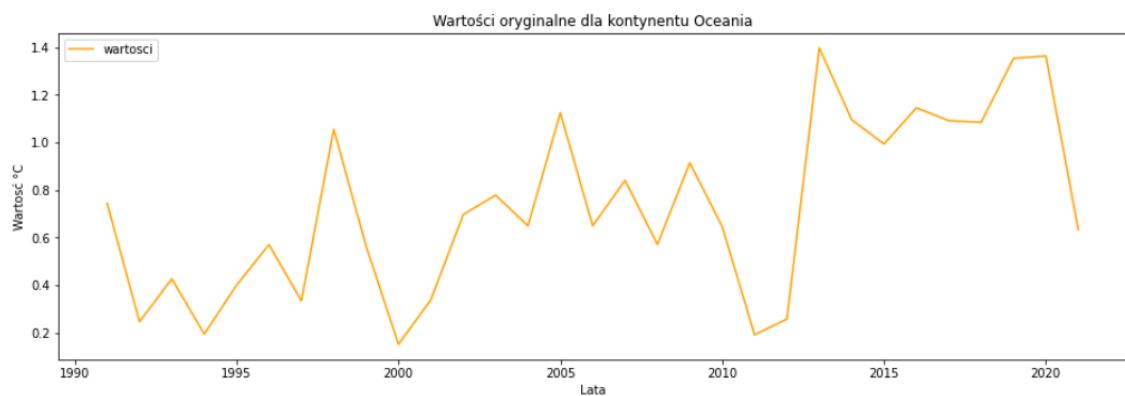


Wyk. 8.: Dane dotyczące średniej temperatury w Europie w latach 1960-2021, wykres liniowy

średnia	mean	0.783574
odchylenie standardowe	std	0.876356
wartość minimalna	min	-1.012000
kwartyl Q1	25%	0.172000
kwartyl Q2	50%	0.696000
kwartyl Q3	75%	1.337000
wartość maksymalna	max	3.353000

Rys. 6.: Podstawowe statystyki dotyczące średniej temperatury w Europie.

AUSTRALIA I OCEANIA



Wyk. 9.: Dane dotyczące średniej temperatury w Australii i Oceanii w latach 1960-2021, wykres liniowy

	mean	0.448393
średnia	std	0.450483
odchylenie standardowe	min	-0.505000
wartość minimalna	25%	0.118000
kwartyl Q1	50%	0.350000
kwartyl Q2	75%	0.779000
kwartyl Q3	max	1.398000
wartość maksymalna		

Rys. 7.: Podstawowe statystyki dotyczące średniej temperatury w Australii i Oceanii.

MODEL

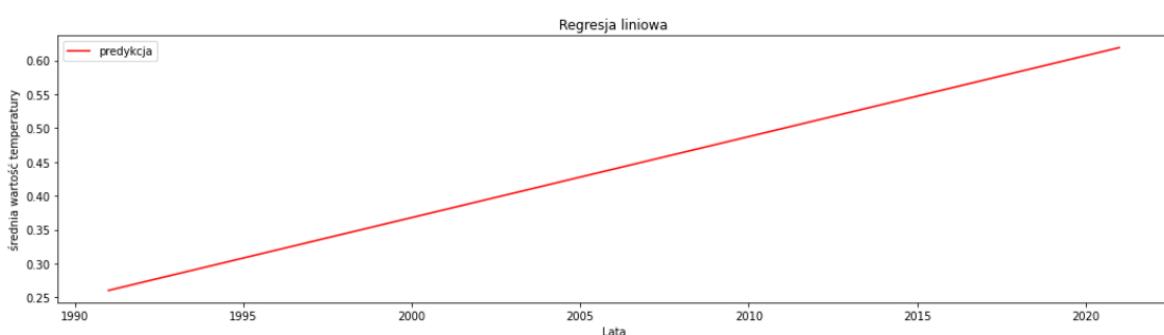
Wybór modelu

Przy analizie średniej temperatury postanowiono skorzystać z modelu regresji, sprawdzono zarówno modele regresji liniowej jak i wielomianowej o współczynnikach 2-20. Dla każdego modelu policzony został błąd średniej kwadratowej, na podstawie którego dla każdego kontynentu jak i świata osobno został wybrany najlepszy model.

Zdecydowano się na użycie regresji jako modelu z kilku powodów. Regresja liniowa charakteryzuje się przede wszystkim **prostotą, interpretowalnością** – dzięki prostym wzorom bardzo łatwo wyjaśnić, jak dana cecha wpływa na wynik modelu oraz **szynkością** – nawet przy dużej liczbie danych dla tych prostych algorytmów wyniki dostajemy prawie od razu. Jednak świat nie składa się tylko z liniowych zależności, dlatego przyjrzano się także regresji wielomianowej.

Model 1. - Świat

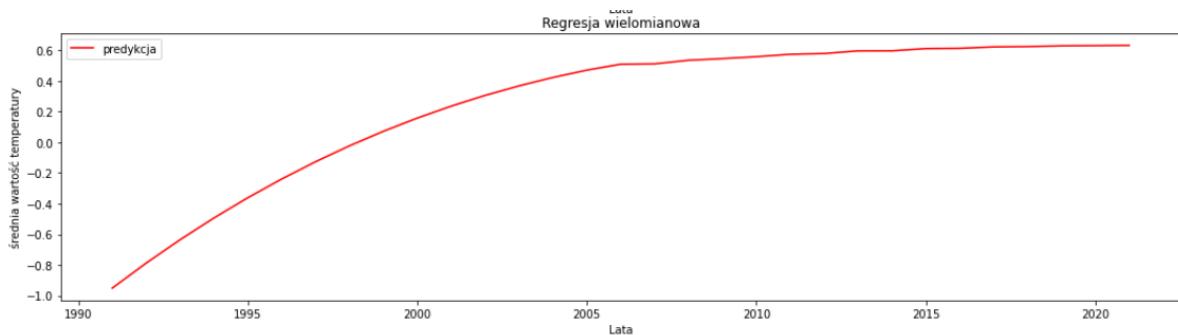
Jak się okazało dla danych na poziomie światowym - najlepszym z sprawdzanych modeli jest **regresja liniowa**, całkiem dobra była również regresja wielomianowa trzeciego stopnia.



Wyk. 10.: Predykcja średniej temperatury świata wykonana na danych testowych, regresja liniowa

współczynniki: $a = 0.01682603$, $b = -33.11805347$

model: $y = 0.01682603 * x - 33.11805347$



Wyk. 11.: Predykcja średniej temperatury świata wykonana na danych testowych, regresja wielomianowa trzeciego stopnia

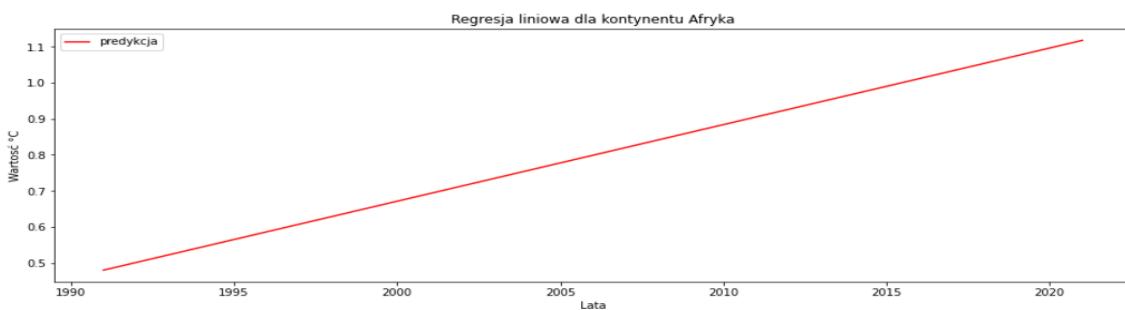
model:

$$y = 384749.76687714 - 581.852555 * x + 0.293292628 * (x^2) - 4.92768173 * 10^{-5} * (x^3)$$

Przyjrzyjmy się każdemu z kontynentów z osobna.

Model 2. - Afryka

Tutaj również najlepszym modelem spośród sprawdzanych okazała się **regresja liniowa**.



Wyk. 12.: Predykcja średniej temperatury Afryki wykonana na danych testowych, regresja liniowa

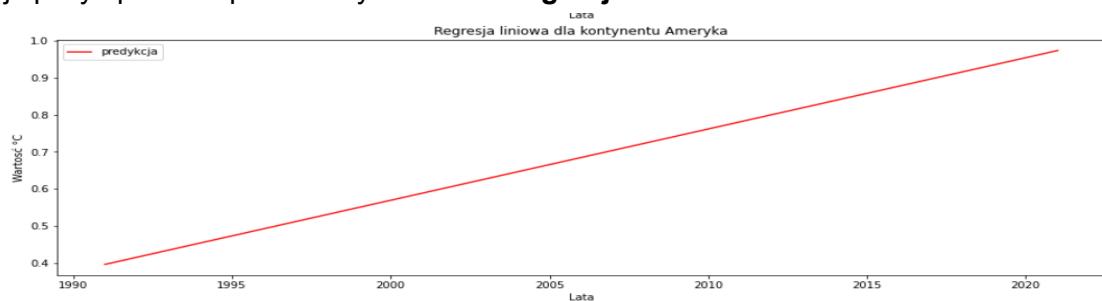
współczynniki: $a = 0.02126007$, $b = -41.84872851$

model: $y = 0.02126007 * x - 41.84872851$

Ze sprawdzonych modeli regresji wielomianowej najlepszy okazał się ten stopnia trzeciego.

Model 3. - Ameryka

Najlepszy spośród sprawdzonych modeli: **regresja liniowa**.



Wyk. 13.: Predykcja średniej temperatury Ameryki wykonana na danych testowych, regresja liniowa

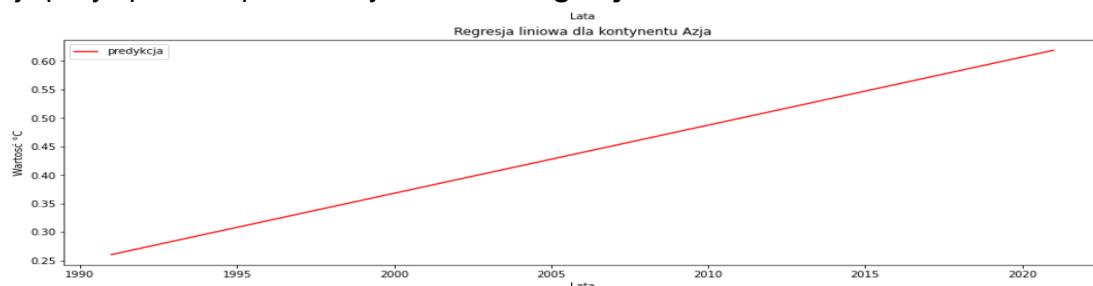
współczynniki: $a = 0.01927164$, $b = -37.97431524$

model: $y = 0.01927164 * x - 37.97431524$

Najlepszy model regresji wielomianowej: stopnia trzeciego.

Model 4. - Azja

Najlepszy spośród sprawdzonych modeli: **regresja liniowa.**



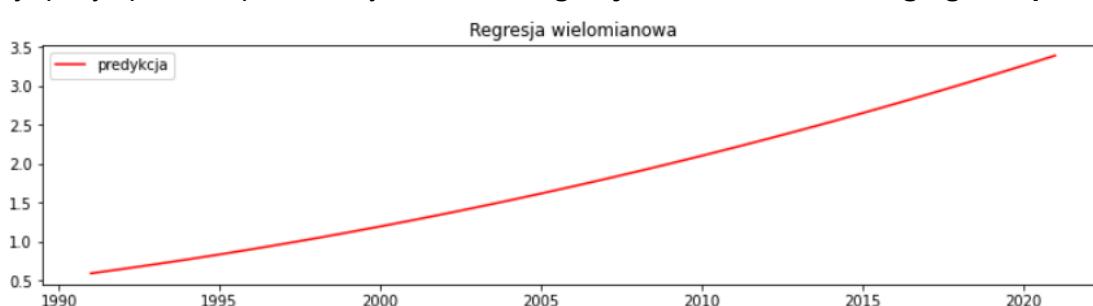
współczynniki: $a = 0.01196085$, $b = -23.55378339$

model: $y = 0.01196085 \cdot x - 23.55378339$

Najlepszy model regresji wielomianowej: stopnia trzeciego.

Model 5. - Europa

Najlepszy spośród sprawdzonych modeli: **regresja wielomianowa drugiego stopnia.**

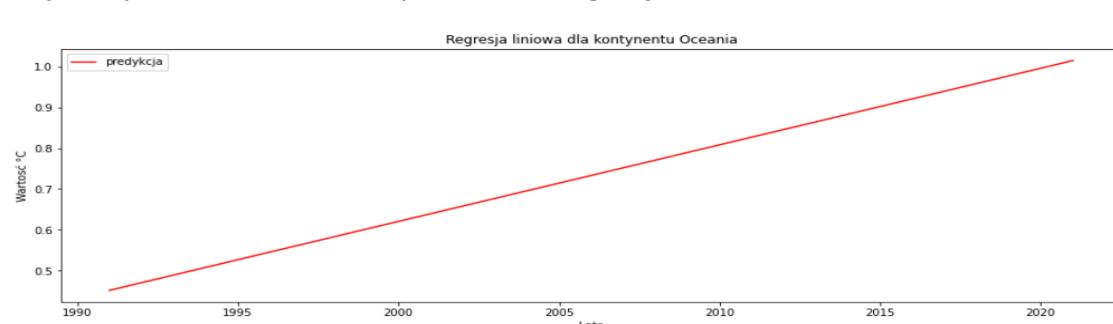


współczynniki: $a = -4.93585418$, $b = 1.25352574 \cdot 10^{-3}$, $c = 4858.79791416$

model: $y = -4.93585418 \cdot x + 1.25352574 \cdot 10^{-3} \cdot (x^2) + 4858.79791416$

Model 6. - Australia i Oceania

Najlepszy spośród sprawdzonych modeli: **regresja liniowa.**



Wyk. 16.: Wykres regresji liniowej dla kontynentu Oceanii. Oś pionowa: Wartość °C (0.5 do 1.0). Oś pozioma: Lata (1990 do 2020). Linia czerwona przedstawia predykcję.

współczynniki: $a = 0.018760855$, $b = -36.90105006$

model: $y = 0.01876085 \cdot x - 36.90105006$

Najlepszy model regresji wielomianowej: stopnia trzeciego.

ANALIZA WYNIKÓW

Każdy z modeli oceniany był miarą błędu średniej kwadratowej. Każdy model charakteryzuje się w miarę małym błędem, dlatego możemy uznać każdy z nich za wartościowy.

Po dokonaniu analiz otrzymanych dostrzega się trend wzrostowy średniej temperatury na całym świecie. Średnia temperatura wzrasta głównie liniowo. Okazało się, że najszybszy wzrost odnotowuje się w Europie, która znacznie wyróżnia się na tle innych kontynentów. Co ciekawe to właśnie w Europie 'rozstrzał' temperatury średniej jest największy.

WNIOSKI

Na podstawie uzyskanych wyników można stwierdzić, że średnia temperatura rośnie. Należy przewidywać dalszy wzrost średniej temperatury na świecie, jak i na każdym kontynencie z osobna. Najprawdopodobniej to właśnie w Europie średnie temperatury będą rosnąć najszybciej, co więcej będą one sięgać najwyższych wartości.

3. ZMIANA ZASIĘGU LODU MORSKIEGO

ŹRÓDŁO DANYCH: <https://www.ncdc.noaa.gov/snow-and-ice/extent/>

Zasięg lodu morskiego (value) dostarczany przez National Snow and Ice Data Center (NSIDC) w latach 1979-2022. Jednostką pomiaru są miliony *km²*.

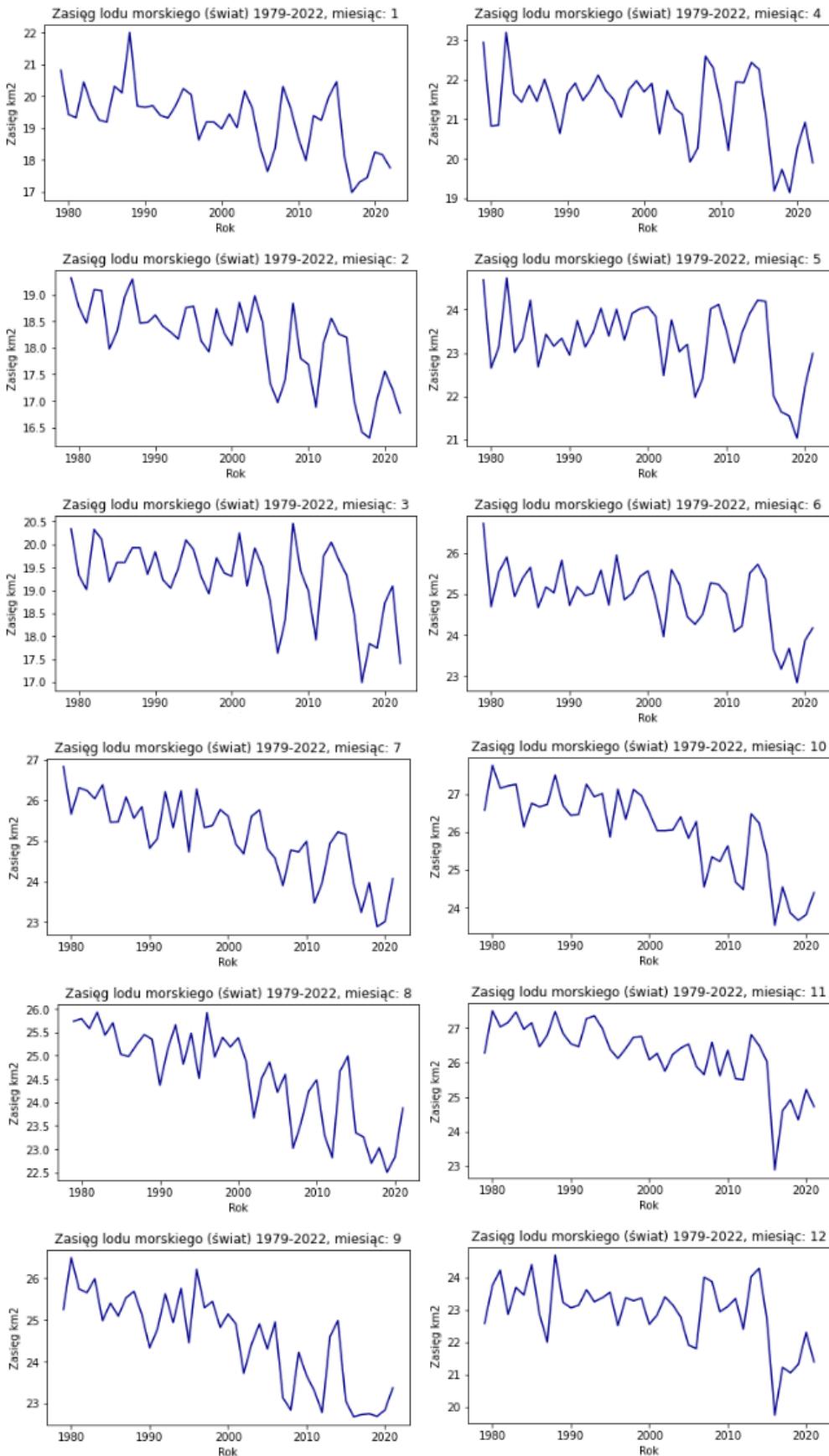
PRZYGOTOWANIE I WSTĘPNA ANALIZA DANYCH

W celu odpowiedniego przygotowania i wstępnej analizy danych zostały dokonane następujące kroki:

1. Wczytanie danych oraz sprawdzenie, czy dane nie zawierają braków.
2. Uzupełnienie znalezionych braków za pomocą średniej.
3. Wizualizacja danych dla każdego miesiąca.
4. Utworzenie nowej kolumny poprzez złączenie kolumn roku i miesiąca w celu wyświetlenia danych w formie szeregu czasowego.
5. Usunięcie zbędnych kolumn oraz określenie typu danych nowej kolumny jako datetime
6. Wizualizacja danych w formie szeregu czasowego.

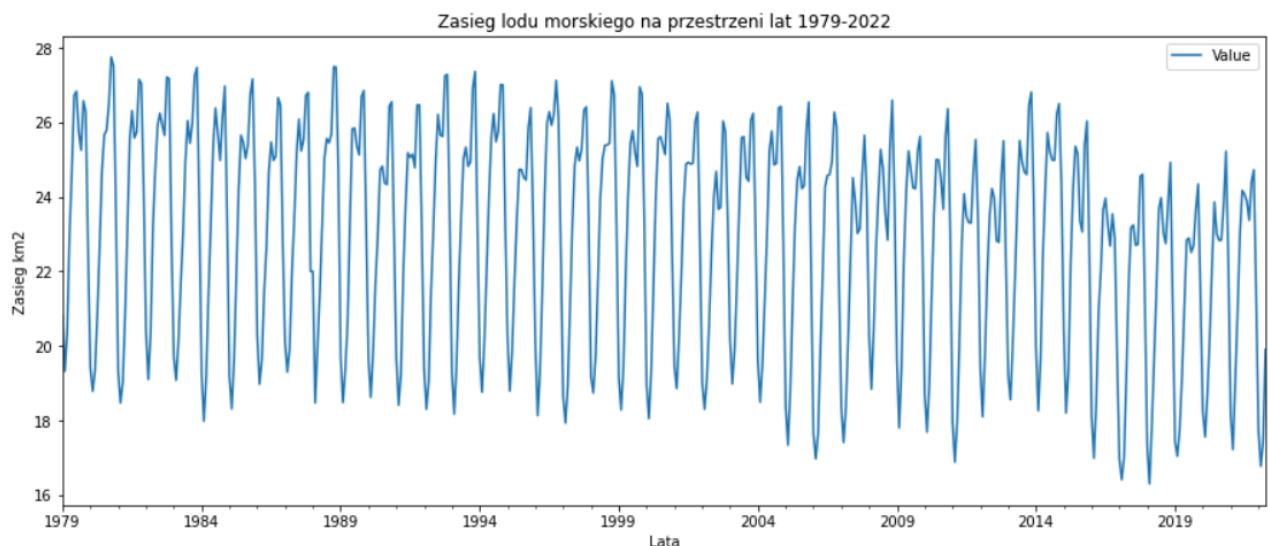
Na podstawie wczytanych danych można zbadać, czy zasięg lodu morskiego zmniejsza się wraz z upływem lat.

Na początek wykonano wizualizację danych dla każdego miesiąca:



Rys.8.: Na rysunku przedstawiono wykresy przedstawiające zasięg lodu morskiego dla każdego miesiąca od stycznia do grudnia w latach 1979-2022.

Po wizualizacji danych dla każdego miesiąca, dane zostały przekształcone w celu wizualizacji zasięgu lodu morskiego w skali rocznej w postaci szeregu czasowego.



Wyk.17.: Wykres przedstawia zasięgu lodu morskiego w latach 1979-2022

MODEL

Wybór modelu

Po wyświetleniu danych na wykresie, można zauważyc pewne rozróżnialne wzorce. Wyświetlony szereg czasowy cechuje się wyraźnym wzorcem sezonowości, a także ogólną tendencją spadkową. Jedną z najczęstszych metod w dokonywaniu prognoz dla szeregów czasowych jest model ARIMA (Autoregressive Integrated Moving Average- czyli zintegrowany model autoregresyjny ze średnią ruchomą). Problem z ARIMA polega na tym, że nie obsługuje danych sezonowych, czyli szeregów czasowy z powtarzającym się cyklem. Rozszerzeniem ARIMA, które wprost obsługuje jednowymiarowe dane w postaci szeregów czasowych z komponentem sezonowym to **model SARIMA** lub sezonowa ARIMA (Seasonal Autoregressive Integrated Moving Average). SARIMA dodaje trzy nowe hiperparametry do określenia autoregresji (AR), różnicowania (I) i średniej ruchomej (MA) dla składnika sezonowego szeregu, a także dodatkowy parametr dla okresu sezonowości. Ma on postać **SARIMA (p,d,q) (P,D,Q)s** i może być zapisany wzorem:

$$\phi_p(L)\tilde{\phi}_P(L^s)\Delta^d\Delta_s^Dy_t = A(T) + \theta_q(L)\tilde{\theta}_Q(L^s)\varsigma_t$$

gdzie:

- (p,d,q) - to niesezonowa część modelu,
- (P,D,Q)s - to sezonowa część modelu,

- p reprezentuje auto-regresywną część modelu,
- d reprezentuje różnicującą część modelu,
- q reprezentuje część modelu ze średnią ruchomą,
- s to okresowość szeregu czasowego (4 dla okresów kwartalnych, 12 dla okresów rocznych itd.)
- P, D, Q analogicznie przyjmują te same definicje co ich małe odpowiedniki , ale są stosowane do składnika sezonowego szeregu czasowego.

Wybór parametrów modelu

Aby dopasować dane z szeregu czasowego do modelu SARIMA, pierwszym krokiem jest znalezienie wartości ARIMA(p,d,q) (P,D,Q)s. W tym celu dokonana została optymalizacja parametrów metodą “grid search”, aby iteracyjnie badać różne kombinacje parametrów. Po zbadaniu kombinacji parametrów i dopasowaniu ich do modelu SARIMA wybrany został optymalny zestaw, o najniższej wartości AIC (Akaike Information Criterion). AIC mierzy, jak dobrze model pasuje do danych, biorąc pod uwagę ogólną złożoność modelu.

Dopasowanie modelu do danych

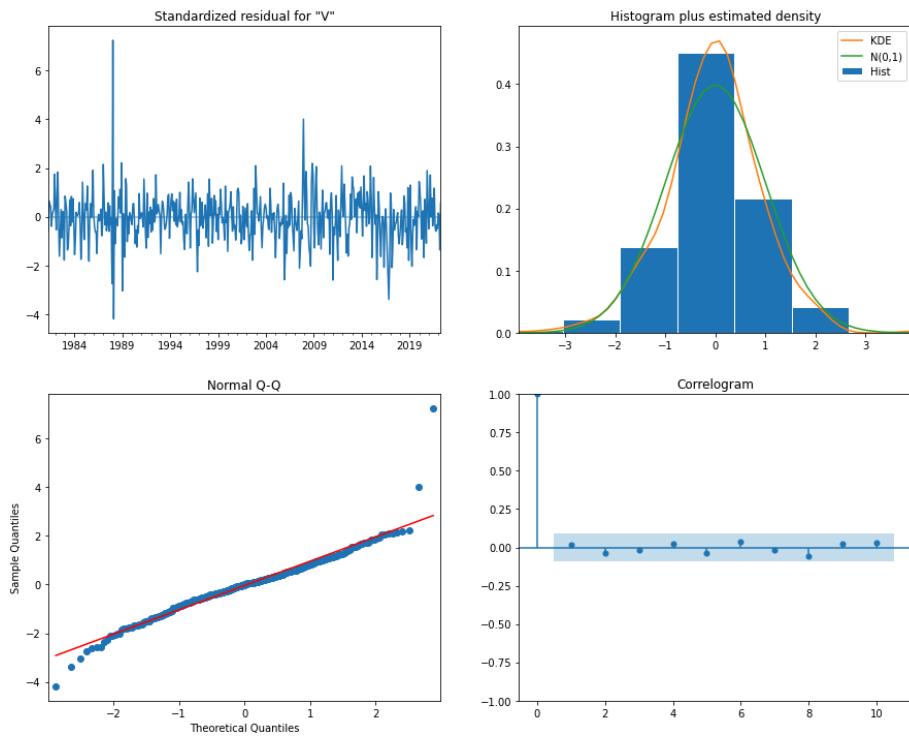
Po dobraniu optymalnych parametrów i dopasowaniu modelu do danych, wyświetlono wyniki modelu SARIMA dla szeregu czasowego.

SARIMAX Results						
Dep. Variable:		Value	No. Observations:	520		
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 12)	Log Likelihood	-327.680			
Date:	Wed, 08 Jun 2022	AIC	665.360			
Time:	16:45:59	BIC	686.363			
Sample:	01-01-1979 - 04-01-2022	HQIC	673.606			
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.7469	0.027	27.528	0.000	0.694	0.800
ma.L1	-1.0151	0.011	-92.179	0.000	-1.037	-0.994
ar.S.L12	-4.906e-06	0.011	-0.000	1.000	-0.022	0.022
ma.S.L12	-0.8934	0.025	-36.021	0.000	-0.942	-0.845
sigma2	0.2080	0.009	23.178	0.000	0.190	0.226
Ljung-Box (L1) (Q):	0.19	Jarque-Bera (JB):	834.92			
Prob(Q):	0.67	Prob(JB):	0.00			
Heteroskedasticity (H):	0.95	Skew:	0.53			
Prob(H) (two-sided):	0.73	Kurtosis:	9.29			

Rys.9.: Rysunek przedstawia wyniki dopasowania modelu SARIMA do danych

Kolumna coef pokazuje wagę (tj. znaczenie) każdej cechy oraz wpływ każdej z nich na szereg czasowy. Kolumna P>|z| zawiera informacje o znaczeniu poszczególnych wag funkcji. Tutaj każda waga poza trzecią ma wartość p mniejszą lub zbliżoną do 0,05

Diagnostyka modelu SARIMA

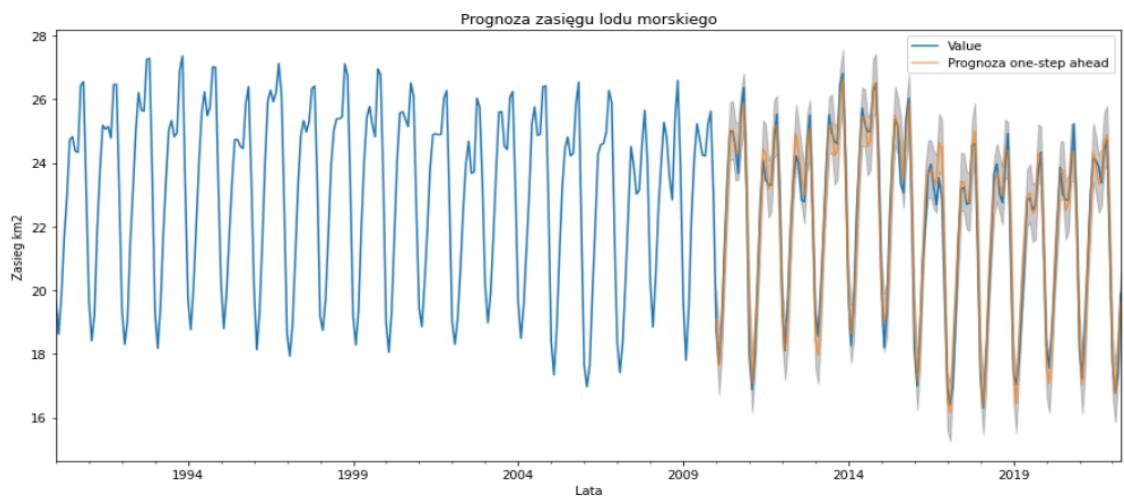


Wyk. 18.: Wykres jest złożony z 4 podwykresów, na podstawie których możliwa jest diagnostyka modelu.

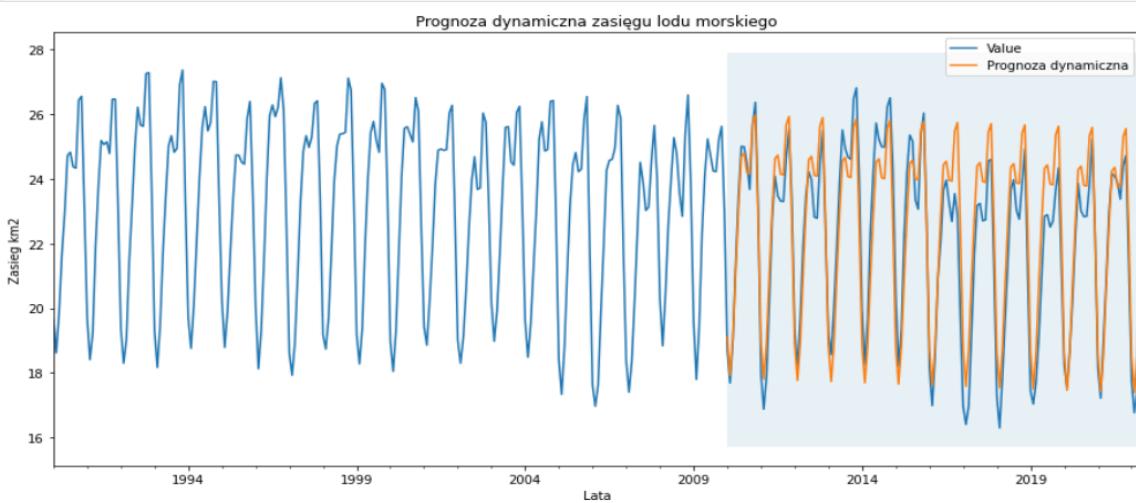
Na prawym górnym diagramie widać, że czerwona linia KDE podąża za linią $N(0,1)$ (gdzie $N(0,1)$ jest notacją standardową dla rozkładu normalnego ze średnią 0 i odchyleniem standardowym 1). Oznacza to, że rezyduły są normalnie rozmieszczone. Wykres QQ w lewym dolnym rogu pokazuje, że uporządkowany rozkład pozostałości (niebieskie kropki) jest zgodny z liniowym przebiegiem próbek ze standardowego rozkładu normalnego o wartości $N(0,1)$. Jest to również wyraźna oznaka normalnego rozmieszczenia rezyduów. Pozostałości w czasie (wykres w lewym górnym rogu) nie wykazują oczywistej sezonowości i wydają się być białym szumem. Potwierdza to wykres autokorelacji w prawym dolnym rogu, który pokazuje, że rezyduły szeregów czasowych wykazują niewielką korelację z opóźnionymi wersjami samych siebie.

Prognoza

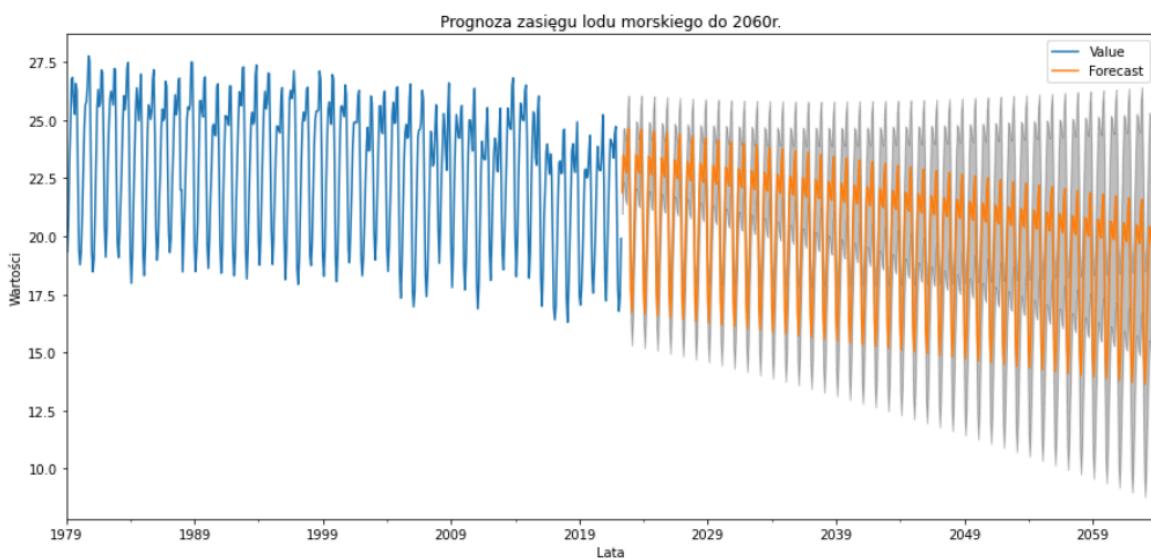
Otrzymaliśmy model, który pozwala nam tworzyć prognozy. Najpierw dokonane zostało porównanie przewidywanych wartości z rzeczywistymi wartościami szeregów czasowych za pomocą prognozy o jeden krok do przodu (one-step ahead forecast) i prognozy dynamicznej (dynamic forecast). Na końcu wykonana została prognoza przyszłych wartości do roku 2060.



Wyk. 19.: Wykres przedstawia prognozę o jeden krok do przodu zasięgu lodu morskiego



Wyk. 20.: Wykres przedstawia prognozę dynamiczną zasięgu lodu morskiego



Wyk. 21.: Wykres przedstawia prognozę zasięgu lodu morskiego do roku 2060.

ANALIZA WYNIKÓW

Z diagnostyki modelu wynika, że zapewnia on zadowalające dopasowanie. Dzięki temu możliwe jest zrozumienie danego szeregu czasowego i prognoza przyszłych wartości. Po dokonaniu dwóch rodzajów prognozy, obliczony został błąd średniokwadratowy dla każdej z nich. Jego wartość w prognozowaniu dynamicznym wynosi 0.97. Jest to wyższy wynik od poprzedniej wartości MSE z prognozowania one-step ahead, który wyniósł 0.22, co jest dosyć niską jest wartością. Błąd MSE równy 0 oznaczałby, że estymator przewiduje obserwacje parametru z całkowitą dokładnością, co nie byłoby możliwe, zatem lepszą reprezentację rzeczywistej zdolności przewidywania można uzyskać dzięki dynamicznym prognozom. Z prognozy do 2060 r. można zauważyc, że w miarę coraz dalszego przewidywania przyszłych wartości generowane przez nasz model przedziały ufności są coraz większe, ponieważ im dalszych dokonujemy prognoz, tym mniejsza jest nasza ufność w prognozowane wartości.

WNIOSKI

Z dokonanej analizy wynika, że zarówno prognozy o krok do przodu, jak i prognozy dynamiczne potwierdzają słuszność tego modelu dla szeregu czasowego. Zatem bazując na dokonanych przewidywaniach możemy dostrzec, że zasięg lodu morskiego będzie nadal malał w stałym tempie.

4. ZMIANA ZASIĘGU POKRYWY ŚNIEŻNEJ

ŽRÓDŁO DANYCH: <https://www.ncdc.noaa.gov/snow-and-ice/extent/>

Zakres pokrywy śnieżnej dostarczany przez Rutgers University Global Snow Laboratory (GSL) dostępny w latach 1967-2022 dla półkuli północnej, Eurazji, Ameryki Północnej oraz Grenlandii. Jednostką pomiaru są miliony km^2 .

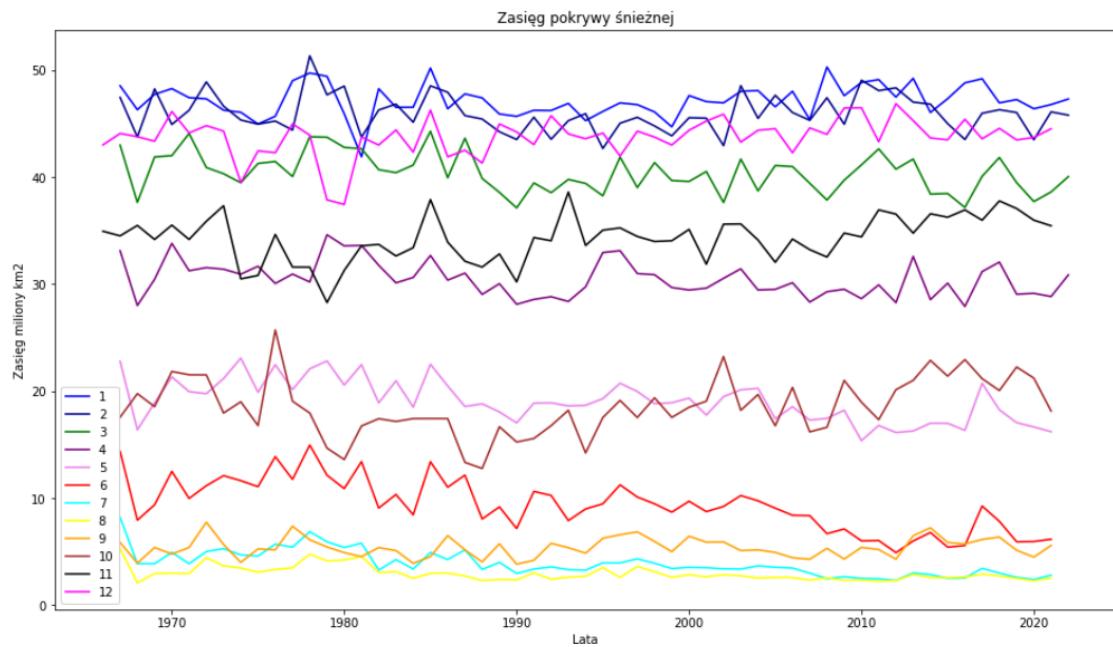
PRZYGOTOWANIE I WSTĘPNA ANALIZA DANYCH

W celu odpowiedniego przygotowania i wstępnej analizy danych zostały dokonane następujące kroki:

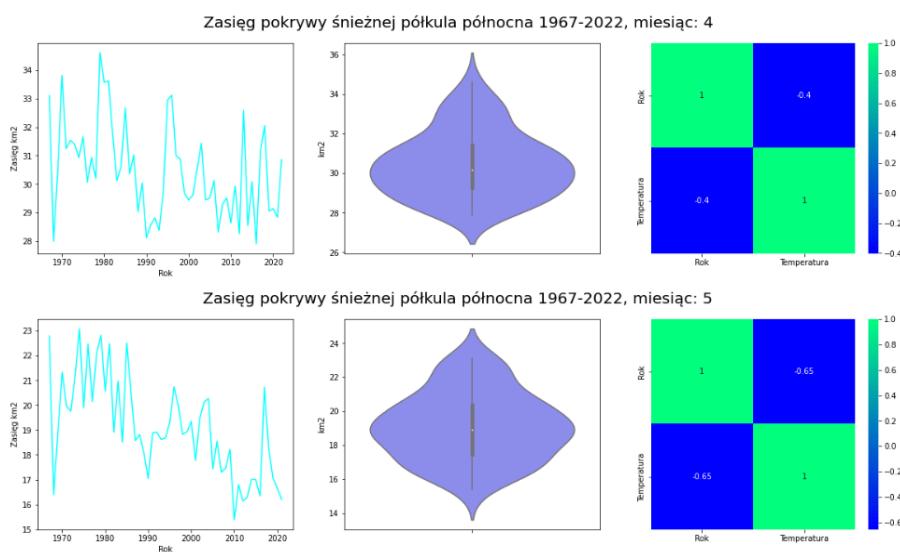
1. Wczytanie danych oraz sprawdzenie, czy dane nie zawierają braków.
2. Uzupełnienie znalezionych braków za pomocą średniej z danej kolumny.
3. Wizualizacja danych dla każdego miesiąca.

4. Utworzenie nowej kolumny poprzez złączenie kolumn roku i miesiąca w celu wyświetlenia danych w formie szeregu czasowego.
5. Usunięcie zbędnych kolumn oraz określenie typu danych nowej kolumny jako datetime
6. Wizualizacja danych w formie szeregu czasowego.
7. Sprawdzenie stacjonarności i trendu w szeregu czasowym

Najpierw wyświetcone zostały dane dla każdego miesiąca na jednym wykresie, a następnie wyświetlono także dane dla każdego miesiąca na osobnych wykresach wraz z macierzą korelacji oraz wykresem wiolinowym dającym informację o podstawowych statystykach.



Wyk. 22.: Zmiana zasięgu pokrywy śnieżnej w milionach km² dla każdego miesiąca

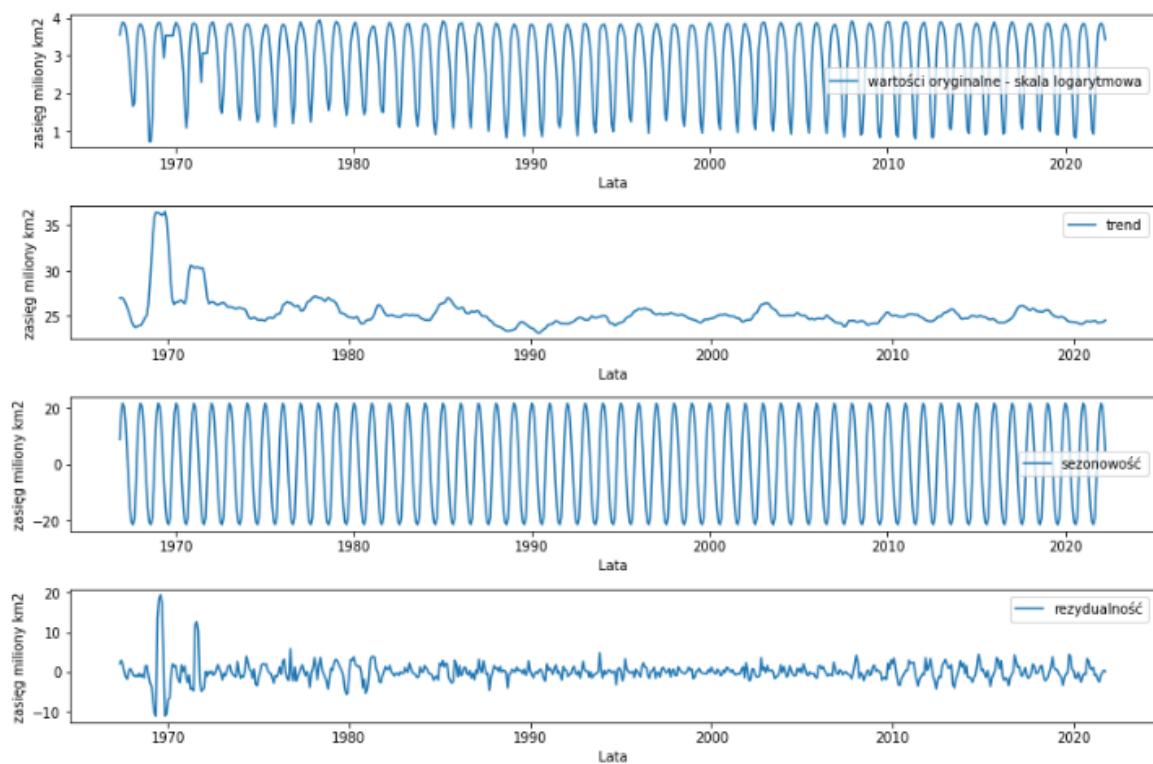


Wyk. 23.: Fragment wykresów dotyczących poszczególnych miesięcy, zmiana zasięgu pokrywy śnieżnej w milionach km² dla miesiąca kwietnia i maja

Na podstawie uzyskanych macierzy korelacji zauważono pewną zależność - zasięg pokrywy śnieżnej w miesiącach uznawanych za cieplejsze (np. kwiecień, czy maj) maleje wraz z upływem czasu, jednak w miesiącach uznawanych za zimniejsze (np. styczeń) zasięg śniegu rośnie lub rośnie nieznacznie.

Sprawdzanie stacjonarności omawianego szeregu czasowego:

W większości modeli szeregów czasowych badane dane powinny pochodzić z szeregu stacjonarnego. Szereg czasowy uznaje się za stały, jeżeli jego właściwości statystyczne, takie jak średnia, wariancja i kowariancja, pozostają stałe w czasie. Formalne sposoby weryfikacji to wykreślanie danych, przeprowadzenie analizy wizualnej i zastosowanie testu statystycznego. Wyświetlone zostały podstawowe charakterystyki szeregu czasowego:



Wyk. 24.: Zasięg pokrywy śnieżnej, własności szeregu czasowego

Wykres pokazuje, że dane są zarówno związane z trendami, jak i sezonowością. To znaczy, że szereg nie jest stacjonarny.

Test statystyczny:

Aby potwierdzić naszą obserwację wizualną na powyższym wykresie, używamy testu hipotezy Dickey-Fullera.

Hipoteza zerowa: Szereg nie jest stacjonarny.

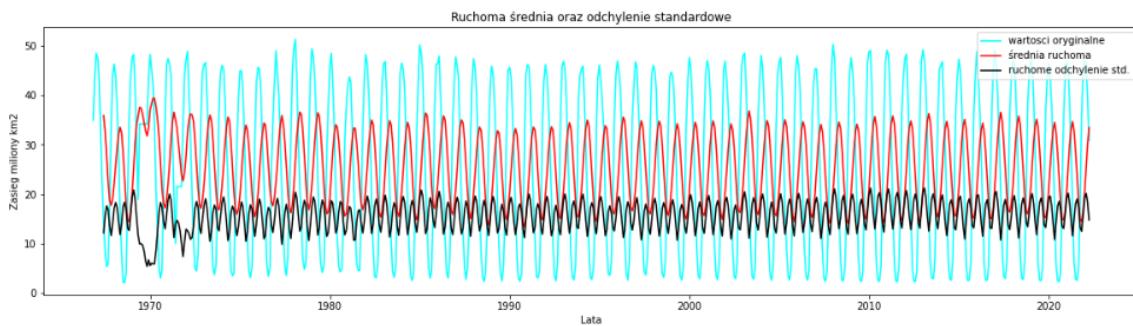
Alternatywna hipoteza: Szereg jest stacjonarny.

```
Wynik testu statystycznego adfuller:  
Test Statistics           -4.894458  
p-value                  0.000036  
Lags Used                20.000000  
Number of observation used 645.000000  
Critical Value (1%)      -3.440529  
Critical Value (5%)      -2.866031  
Critical Value (10%)     -2.569162  
dtype: float64
```

Wyk. 25.: Wyniki testu statystycznego - Dickey-Fullera

Z wartością $p = 0.000036$, mniejszą niż 0.05, można odrzucić hipotezę zerową, która mówi o tym, że szereg nie jest stacjonarny. Oznacza to, że można przyjąć, że szereg jest stacjonarny.

Wyświetlono także ruchomą średnią oraz odchylenie standardowe dla omawianego szeregu czasowego.



Wyk. 26.: Ruchoma średnia oraz odchylenie standardowe

MODEL

Wybór modelu

Dla naszych danych zastosowane zostaną dwa modele: **model SARIMA** oraz **Auto SARIMA**. Model SARIMA jest rozszerzeniem modelu ARIMA (Auto-Regressive Integrated Moving Average, co oznacza zintegrowany model autoregresyjny ze średnią ruchomą), który jest w rzeczywistości klasą modeli, która "wyjaśnia" dany szereg czasowy na podstawie własnych przeszłych wartości. ARIMA jest stosowany dla szeregów "niesezonowych", które mają wzorce i nie są przypadkowymi białymi szumami. Ze względu na sezonowość danego szeregu czasowego należy użyć więc modelu, który uwzględnia sezonowość, czyli modelu SARIMA (Seasonal Auto-Regressive Integrated Moving Average). Dodatkowo szereg musi być stacjonarny, co dla rozważanego przez nas szeregu zostało

zbadane i potwierdzone. Dla danego szeregu czasowego także użyty zostanie model **Auto SARIMA**. Zaletą korzystania z tego modelu jest fakt, że po wstępnym przetworzeniu danych możemy pominąć kolejne etapy i dopasować bezpośrednio do naszego modelu. Model sam znajdzie optymalne parametry, korzystając z wartości AIC (Akaike Information Criterion) i BIC (Bayesian Information Criterion) generowane przez różne kombinacje wartości p,q i d.

Model przyjmuje postać **SARIMA (p,d,q) (P,D,Q)s** i może być zapisany wzorem:

$$\phi_p(L)\tilde{\phi}_P(L^s)\Delta^d\Delta_s^Dy_t = A(T) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta_t$$

gdzie:

- (p,d,q) - to niesezonowa część modelu,
- (P,D,Q)s - to sezonowa część modelu,
- p reprezentuje auto-regresywną część modelu,
- d reprezentuje różnicującą część modelu,
- q reprezentuje część modelu ze średnią ruchomą,
- s to okresowość szeregu czasowego (4 dla okresów kwartalnych, 12 dla okresów rocznych itd.)
- P, D, Q analogicznie przyjmują te same definicje co ich małe odpowiedniki , ale są stosowane do składnika sezonowego szeregu czasowego.

MODEL AUTO SARIMA

Dobór parametrów modelu i dopasowanie do danych:

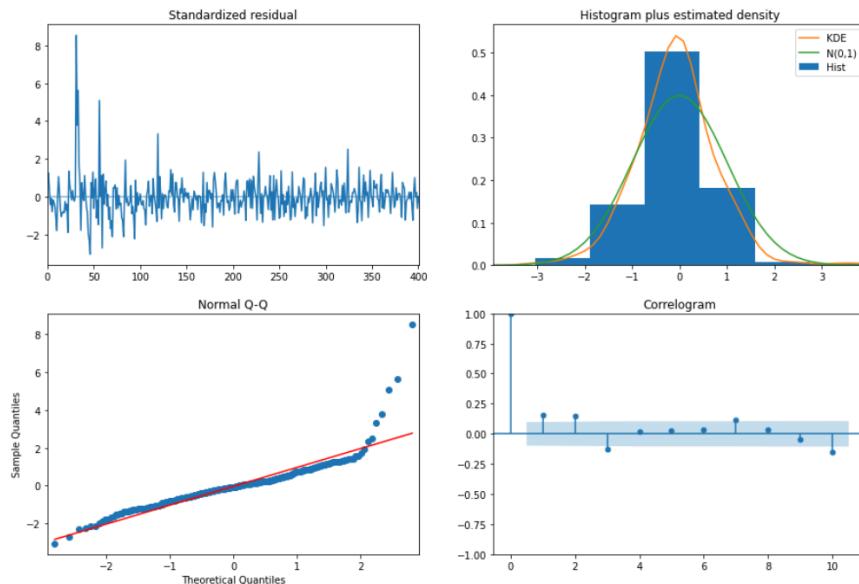
Dobór parametrów w modelu Auto SARIMA odbywa się automatycznie. Model sam znajduje optymalne parametry. Po dobraniu odpowiednich przez model parametrów i dopasowaniu modelu do danych uzyskane zostały wyniki:

SARIMAX Results						
Dep. Variable:	y	No. Observations:	402			
Model:	SARIMAX(4, 0, 5)	Log Likelihood	-1002.591			
Date:	Thu, 09 Jun 2022	AIC	2025.181			
Time:	18:25:24	BIC	2065.146			
Sample:	0	HQIC	2041.005			
	- 402					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	1.7779	0.109	16.318	0.000	1.564	1.991
ar.L2	-0.1258	0.297	-0.423	0.672	-0.709	0.457
ar.L3	-1.6056	0.297	-5.401	0.000	-2.188	-1.023
ar.L4	0.9535	0.109	8.765	0.000	0.740	1.167
ma.L1	-1.0153	0.102	-9.910	0.000	-1.216	-0.815
ma.L2	-1.0545	0.204	-5.162	0.000	-1.455	-0.654
ma.L3	1.5355	0.109	14.127	0.000	1.322	1.749
ma.L4	0.1474	0.063	2.344	0.019	0.024	0.271
ma.L5	-0.5792	0.056	-10.307	0.000	-0.689	-0.469
sigma2	8.2573	0.289	28.557	0.000	7.691	8.824
Ljung-Box (L1) (Q):	10.03	Jarque-Bera (JB):		5217.69		
Prob(Q):	0.00	Prob(JB):		0.00		
Heteroskedasticity (H):	0.22	Skew:		2.39		
Prob(H) (two-sided):	0.00	Kurtosis:		19.99		

Rys.10.: Rysunek przedstawia tabelę wyników dopasowania modelu SARIMA do danych

Można zauważyć, że w kolumnie $P>|z|$, która zawiera informacje o znaczeniu poszczególnych wag funkcji, jedna z wag ma wartość p znacznie większą niż 0,05.

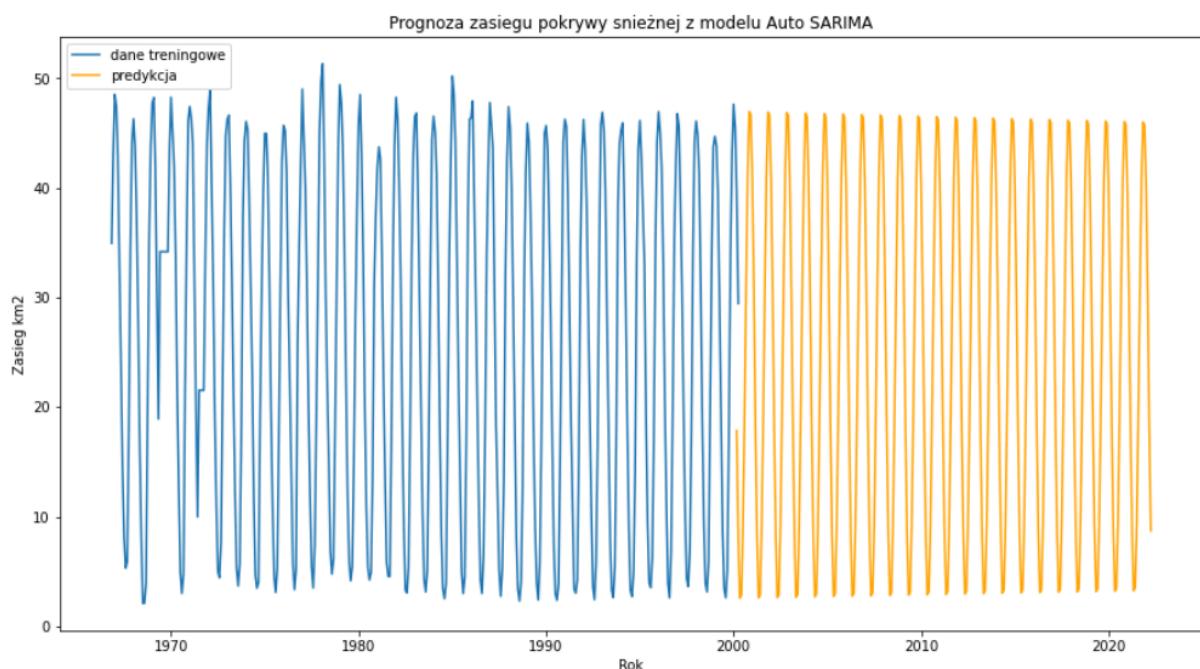
Diagnostyka modelu Auto SARIMA:



Wyk.27.: Wykres przedstawia 4 podwykresy umożliwiające diagnostykę modelu Auto SARIMA.

Z analizy wykresów widać, że rezyduły rozmiieszczone są normalnie, co zapewnia zadowalające dopasowanie modelu do danych.

Prognoza:



Wyk.28.: Wykres przedstawia prognozę zasięgu pokrywy śnieżnej z modelu Auto SARIMA.

MODEL SARIMA

Dobór parametrów modelu i dopasowanie modelu do danych:

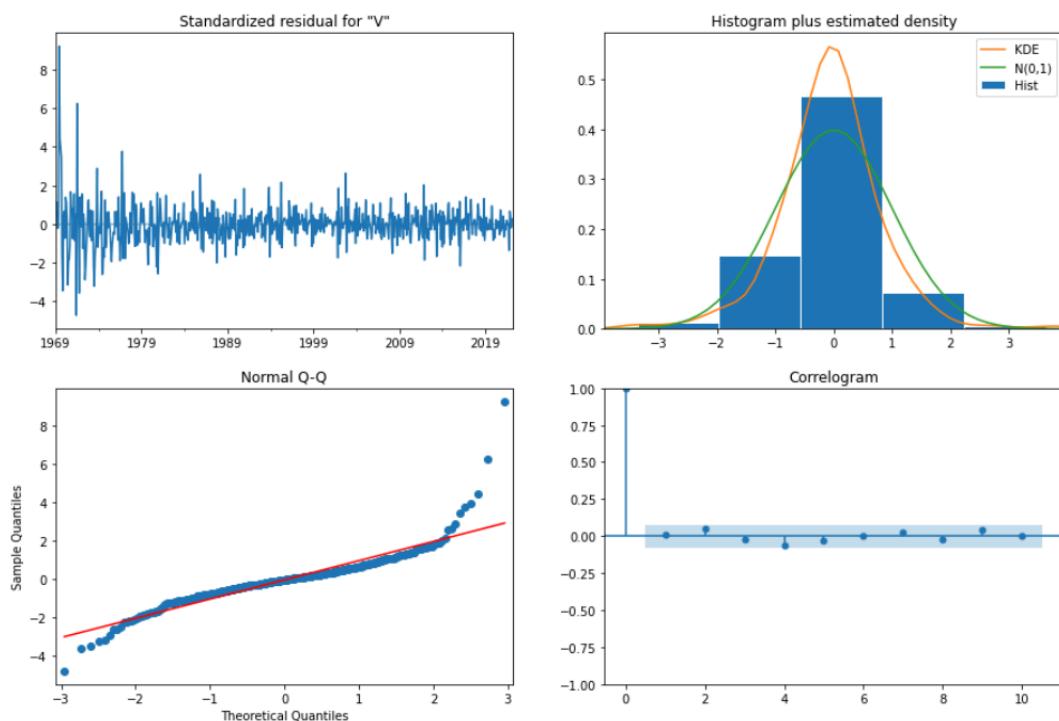
Dobór parametrów modelu SARIMA odbywa się tak samo jak to zostało opisane przy zmianie zasięgu lodu morskiego, czyli za pomocą metody grid search, gdzie optymalne parametry minimalizowały wartość AIC. Po dobraniu odpowiednich parametrów dopasowujemy model do danych i wyświetlamy wynik:

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.6390	0.020	31.601	0.000	0.599	0.679
ma.L1	0.1465	0.028	5.302	0.000	0.092	0.201
ar.S.L12	-0.1704	0.022	-7.867	0.000	-0.213	-0.128
ma.S.L12	-0.7908	0.030	-26.547	0.000	-0.849	-0.732
sigma2	4.2215	0.124	33.994	0.000	3.978	4.465

Rys.10.: Tabela wynikowa modelu SARIMA

Można ponownie zauważyc, że w kolumnie P>|z|, która zawiera informacje o znaczeniu poszczególnych wag funkcji, każda waga ma wartość p mniejszą lub zbliżoną do 0,05

Diagnostyka modelu SARIMA:



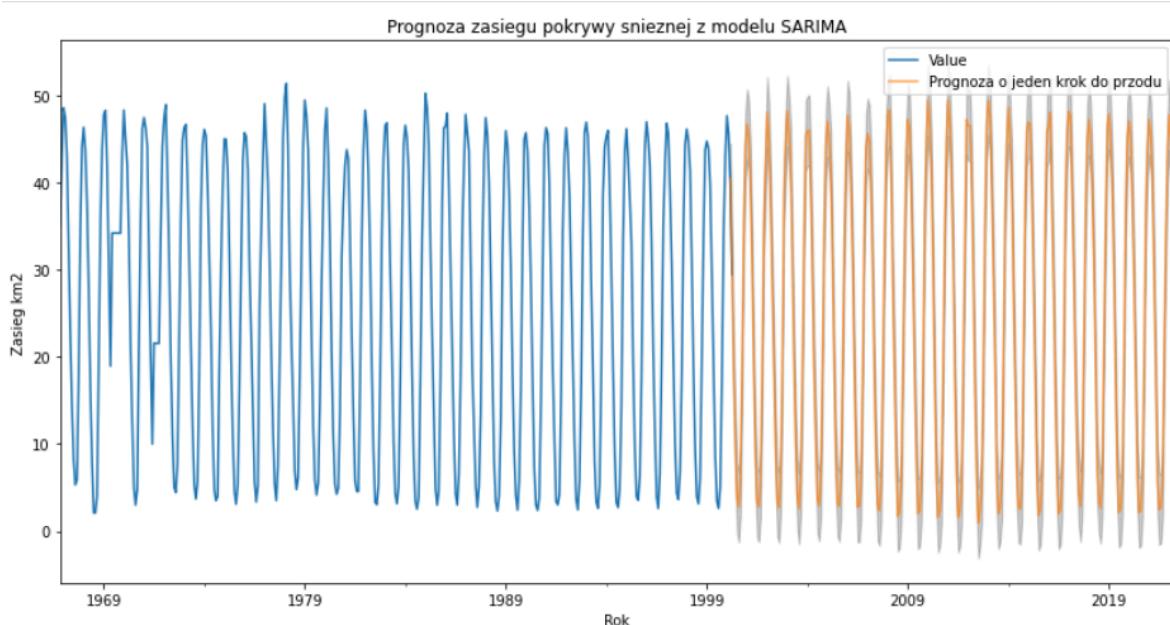
Wyk.22.: Wykres przedstawia 4 podwykresy, których analiza umożliwia diagnostykę modelu

Ponownie jak w poprzednim przykładzie na prawym górnym wykresie widać, że czerwona linia KDE podąża za linią $N(0,1)$. Wykres QQ pokazuje, że uporządkowany rozkład

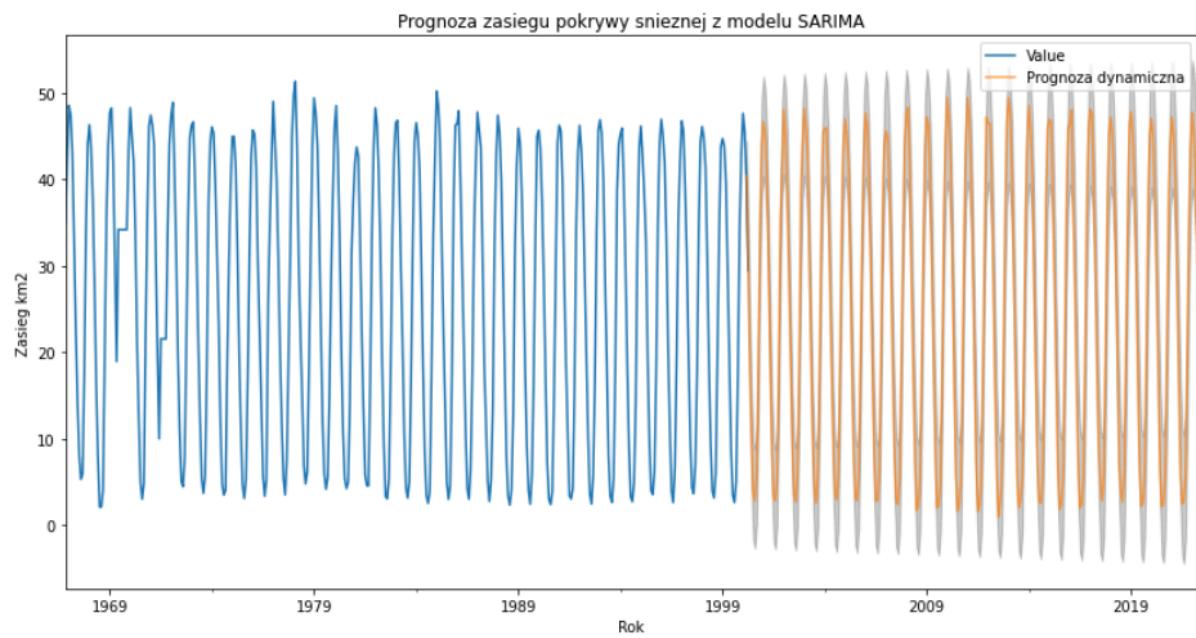
pozostałości (niebieskie kropki) jest zgodny z liniowym przebiegiem próbek ze standardowego rozkładu normalnego o wartości N (0,1). Zatem oba wykresy potwierdzają normalne rozmieszczenie rezyduów. Wykres autokorelacji pokazuje, że rezyduły szeregów czasowych wykazują niewielką korelację z opóźnionymi wersjami samych siebie, co jest potwierdzeniem wyniku na lewym górnym wykresie.

Prognoza:

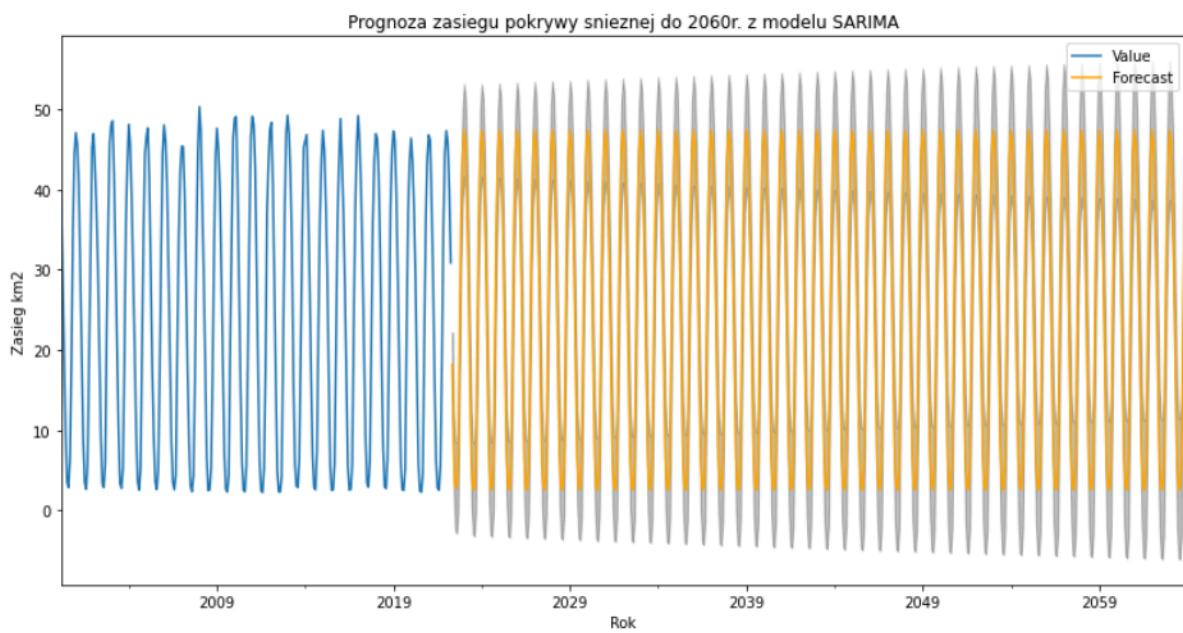
Po dopasowaniu modelu SARIMA do danych dokonana została prognoza "o jeden krok do przodu" (ang. one-step ahead forecast) oraz prognoza dynamiczna (ang. dynamic forecast), a na końcu zrobiona została prognoza przyszłych wartości do 2060 r.



Wyk.29.: Wykres przedstawia prognozę "o jeden krok do przodu" zasięgu pokrywy śnieżnej z modelu SARIMA.



Wyk.30.: Wykres przedstawia prognozę dynamiczną zasięgu pokrywy śnieżnej z modelu SARIMA.



Wyk.28.: Wykres przedstawia prognozę do 2060r. zasięgu pokrywy śnieżnej z modelu SARIMA.

ANALIZA WYNIKÓW

Z diagnostyki modeli wynika, że oba modele zapewniają zadowalające dopasowanie, jednak dopasowanie modelu SARIMA jest lepsze, co widać po wartości wag funkcji w kolumnie $P>|z|$. Po policzeniu błędu średniokwadratowego dla prognozy w modelu Auto SARIMA okazało się, że jego wartość, równa 254.57, jest dosyć wysoka, co było przyczyną zmiany modelu na SARIMA. Po wykonaniu prognozy "o jeden krok do przodu" oraz dynamicznej w modelu SARIMA, okazało się, że ich błędy średniokwadratowe dają odpowiednio wynik 1.76 i 3.43. Wynika z tego, że prognozy z modelu SARIMA dają dokładniejsze wyniki, jednak lepszą reprezentację rzeczywistej zdolności przewidywania można uzyskać dzięki dynamicznym prognozom.

WNIOSKI

Z dokonanej analizy można wysnuć wniosek, że zasięg pokrywy śnieżnej maleje w bardzo wolnym tempie. Jest to minimalny spadek, widoczny głównie w cieplejszych miesiącach, zatem na modelu obejmującym cały rok zależność spadkowa nie jest mocno zauważalna.

5. ZAKOŃCZENIE I WNIOSKI

W niniejszej pracy zbadane zostały 3 czynniki, których analiza pozwoliła na określenie wpływu, jaki wywiera na naszą planetę zjawisko globalnego ocieplenia, czyli ogólnego wzrostu średniej temperatury. Pierwszym czynnikiem, który został omówiony była temperatura na świecie oraz na poszczególnych kontynentach. Badając jej zmianę na przestrzeni lat, można było zauważać jej stały wzrost na powierzchni Ziemi, co potwierdza zjawisko globalnego ocieplenia. Następnie dokonana została analiza zasięgu lodu morskiego na planecie, co potwierdziło, iż lód morski maleje w stałym tempie. Zjawisko to jest potwierdzeniem negatywnego wpływu wzrostu temperatury na Ziemię. Ostatnim badanym czynnikiem była zmiana zasięgu pokrywy śnieżnej. Analiza wykazała znacznie mniejszy spadek tego czynnika w porównaniu do lodu morskiego, jednak wciąż można zauważać zmniejszenie zasięgu pokrywy śnieżnej na Ziemi, co również wywołane jest negatywnymi skutkami globalnego ocieplenia. Z dokonanej analizy można wysnuć wniosek, że globalne ocieplenie ma realny wpływ na naszą planetę i wraz z ciągłym wzrostem temperatury, staje się ono coraz większym zagrożeniem. Należy zatem zapobiegać negatywnym skutkom tego procesu.