

# Prognozowanie Inflacji w Polsce

Szymon Budziak, Dominik Ciołczyk, Karol Wyřbkiewicz

|  |           |
|--|-----------|
| <b>1. Zrozumienie uwarunkowań.....</b>   | <b>3</b>  |
| 1.1. Cel główny.....   | 3         |
| 1.2. Cel biznesowy.....  | 3         |
| 1.3. Cele poboczne.....  | 3         |
| <b>2. Analiza literatury.....</b>  | <b>3</b>  |
| 2.1. Model Phillipsa i Jego Ewolucja.....  | 3         |
| 2.2. Prognozy Inflacji z Wykorzystaniem Krzywej Phillipsa.....                               | 4         |
| 2.2.1. Zmiana w Procesie Inflacyjnym i Relacjach Predykcyjnych.....                          | 4         |
| 2.2.2. Model ADL-u jako Prototyp Prognozy Inflacji.....                                      | 4         |
| 2.2.3. Modele Jednowymiarowe a Model UC-SV.....  | 4         |
| 2.2.4. Sukcesy Krzywej Phillipsa.....  | 4         |
| 2.2.5. Kombinacja Prognoz.....   | 4         |
| 2.2.6. Zmienne Inne niż Aktywność Gospodarcza.....   | 4         |
| 2.2.7. Wahania Inflacji a Sukces Prognoz.....  | 5         |
| 2.2.8. Density Forecasting.....  | 5         |
| 2.2.9. Prognozowanie Gęstościowe.....  | 5         |
| 2.3. Wnioski.....  | 5         |
| <b>3. Zrozumienie danych.....</b>  | <b>5</b>  |
| 3.1. Gromadzenie danych wejściowych.....   | 5         |
| 3.2. Kalendarium ważnych wydarzeń gospodarczych i ekonomicznych.....                         | 7         |
| 3.3. Opis danych.....  | 12        |
| 3.4. Wstępny przegląd danych.....  | 16        |
| 3.5. Weryfikacja jakości danych.....   | 20        |
| <b>4. Przygotowanie danych.....</b>  | <b>21</b> |
| 4.1. Funkcje oraz kroki w ekstrakcji danych.....   | 21        |
| 4.2. Czyszczenie danych.....   | 24        |
| 4.3. Feature selection.....  | 27        |
| <b>5. Przewidywanie inflacji.....</b>  | <b>29</b> |
| 5.1. Przewidywanie inflacji za pomocą modeli z biblioteki Darts na jednej kolumnie HICP..... | 29        |
| 5.1.1. Naive Seasonal.....   | 29        |
| 5.1.2. ARIMA.....  | 30        |
| 5.1.3. XGBoost.....  | 31        |
| 5.1.4. Prophet.....  | 33        |
| 5.1.5. Wnioski.....  | 33        |
| 5.2. Przewidywanie inflacji na kilku kolumnach wybranych w procesie Feature Selection.....   | 35        |
| 5.2.1. Feature Selection.....  | 35        |
| 5.2.2. Linear Regression.....  | 37        |

|   |    |
|---|----|
| 5.2.3. Ridge Regression.....                        | 38 |
| 5.2.4. Polynomial Regression degree 3.....          | 39 |
| 5.2.5. Minutes po spotkaniu 10.05.....              | 40 |
| 5.2.6. LSTM.....                                    | 40 |
| 6. Prognoza inflacji w Statistica oraz RStudio..... | 47 |
| 6.1. Regresja wieloraka.....                        | 47 |
| 6.2. ARIMA w Statistice.....                        | 49 |
| 6.3. ARIMA i regresja liniowa w RStudio.....        | 49 |
| 6.4. ARIMA z transformacją danych.....              | 50 |
| 7. Wnioski.....                                     | 52 |
| 8. Dalsze badania.....                              | 53 |

# 1. Zrozumienie uwarunkowań

## 1.1. Cel główny

Zrozumienie i prognozowanie inflacji w Polsce, z uwzględnieniem wpływu wydarzeń zewnętrznych na jej przebieg oraz różnych danych makroekonomicznych.

## 1.2. Cel biznesowy

Stworzenie wiarygodnego modelu prognozującego inflację, który może być używany przez instytucje finansowe, inwestorów i decydentów politycznych do podejmowania świadomych decyzji.

## 1.3. Cele poboczne

1. Analiza historycznych danych dotyczących inflacji w Polsce.
2. Badanie wpływu wydarzeń zewnętrznych na inflację.
3. Opracowanie metodologii do przewidywania inflacji.
4. Testowanie i walidacja modelu prognozującego.
5. Prezentacja wyników w przystępnej formie dla różnych grup odbiorców.

# 2. Analiza literatury

Inflacja to jedno z kluczowych zjawisk makroekonomicznych, które od lat fascynuje ekonomistów, analityków i decydentów. Zrozumienie jej dynamiki i przewidywanie jej przyszłych zmian mają fundamentalne znaczenie dla polityki pieniężnej, fiskalnej oraz podejmowania decyzji inwestycyjnych.

## 2.1. Model Phillipsa i Jego Ewolucja

Podstawową koncepcją teoretyczną przewidywania inflacji jest krzywa Phillipsa. Od lat 60. XX wieku przeszła ona znaczną ewolucję. Pierwotnie sugerowała ujemną korelację między bezrobociem a inflacją, jednak późniejsze badania wykazały, że ta relacja jest bardziej złożona i zmienna w czasie.

**Klasyczna krzywa Phillipsa:** Związek między bezrobociem a inflacją cenową.

**Krzywa Phillipsa z oczekiwaniami adaptacyjnymi:** Uwzględnia wpływ oczekiwań inflacyjnych.

**Krzywa Phillipsa z NAIRU:** Wprowadza pojęcie "Naturalnej Stopy Bezrobocia" (Non-Accelerating Inflation Rate of Unemployment).

## 2.2. Prognozy Inflacji z Wykorzystaniem Krzywej Phillipsa

W badaniach przeprowadzonych przez naukowców z Harvardu zauważono, że skuteczność prognoz inflacji z wykorzystaniem krzywej Phillipsa jest zróżnicowana i zależy od okresu próby oraz wybranych zmiennych aktywności gospodarczej.

### 2.2.1. Zmiana w Procesie Inflacyjnym i Relacjach Predykcyjnych

Badania wykazały znaczącą zmienność czasową w procesie inflacyjnym, relacjach predykcyjnych i prognozach opartych na krzywej Phillipsa. W zależności od okresu prób różni autorzy osiągalni różne wnioski dotyczące prognozowania inflacji przy użyciu krzywej Phillipsa.

### 2.2.2. Model ADL-u jako Prototyp Prognozy Inflacji

Modele typu ADL-u (Autoregressive Distributed Lag with unemployment gap) stanowiły podstawę dla porównania prognoz z krzywej Phillipsa z modelem UC-SV (Unobserved Components-Stochastic Volatility). Wyniki sugerują, że modele te radzą sobie stosunkowo dobrze w okresach, w których inne modele aktywności również są skuteczne.

### 2.2.3. Modele Jednowymiarowe a Model UC-SV

Nie znaleziono jednego modelu jednowymiarowego, który przewyższałby model UC-SV. Spośród rozważanych modeli jednowymiarowych (ARIMA), tylko nieliczne w niektórych podokresach przewyższały model UC-SV.

### 2.2.4. Sukcesy Krzywej Phillipsa

Chociaż modele z krzywej Phillipsa wykazywały znaczne ulepszenia w prognozach w latach 70. i 80., nie zawsze udaje się im przewyższyć model UC-SV. Niemniej jednak, w okresach recesji i wysokiego bezrobocia, modele te przewidywały inflację stosunkowo dobrze.

### 2.2.5. Kombinacja Prognoz

Kombinacja prognoz (Forecast Combining) zwykle poprawia dokładność prognoz z krzywej Phillipsa, ale nie zawsze przewyższa model UC-SV. Kombinacja prognoz działa lepiej przed rokiem 1993, podczas gdy po tym okresie poprawy nie są duże.

### 2.2.6. Zmienne Inne niż Aktywność Gospodarcza

Prognozy oparte na zmiennych innych niż aktywność gospodarcza nie sprawdzają się dobrze, szczególnie w okresie po 1992 roku. Wykorzystanie alternatywnych wskaźników rdzenia inflacji (np. mediana, CPI trim mean) nie przynosi znaczącej poprawy.

#### 2.2.7. Wahania Inflacji a Sukces Prognoz

Prognozy z krzywej Phillipsa wydają się być bardziej skuteczne, gdy inflacja jest zmienna, ale związek ten jest ograniczony. Najlepsze wyniki osiągane są w sytuacjach, gdy luka bezrobocia jest skrajna.

#### 2.2.8. Density Forecasting

Badacze przyjęli modele QRF, będące rozszerzeniem metod Random Forests, aby uchwycić szeroki zakres nieliniowych dynamik inflacji. Metoda ta okazała się być skuteczniejsza od tradycyjnych modeli liniowych w niektórych scenariuszach, szczególnie przy prognozowaniu inflacji bazowej w strefie euro. Siła tego modelu leży w jego zdolności do uwzględniania szerokiego zestawu potencjalnych determinant inflacji, takich jak oczekiwania inflacyjne, naciski kosztowe, rzeczywista aktywność i zmienne finansowe.

#### 2.2.9. Prognozowanie Gęstościowe

Zaawansowane techniki prognozowania obecnie często używają prognoz gęstościowych zamiast tradycyjnych prognoz punktowych. Te prognozy gęstościowe pomagają ocenić prawdopodobieństwo różnych potencjalnych przyszłych scenariuszy, co jest kluczowe dla podejmowania świadomych decyzji politycznych.

### 2.3. Wnioski

Analiza literatury pokazuje, że prognozowanie inflacji jest zadaniem trudnym i niejednoznacznym. Modele oparte na krzywej Phillipsa mogą okazać się użyteczne w niektórych warunkach, szczególnie gdy występują znaczne odchylenia bezrobocia od NAIRU. Modele univariate, takie jak ARIMA, są proste do wdrożenia i często dają dobre wyniki w stabilnych warunkach ekonomicznych. Mogą być mniej skuteczne w zmieniających się warunkach, ponieważ opierają się tylko na danych historycznych dotyczących samej inflacji. Modele wielowymiarowe, takie jak VAR, mogą uwzględniać więcej zmiennych ekonomicznych, co może poprawić dokładność prognoz. Wymagają więcej danych i są bardziej złożone do wdrożenia, ale mogą lepiej uwzględniać złożone interakcje między różnymi czynnikami ekonomicznymi.

W obliczu tych wyzwań opracowanie skutecznego modelu predykcji inflacji wymaga kombinacji podejść oraz uważnego doboru zmiennych.

## 3. Zrozumienie danych

### 3.1. Gromadzenie danych wejściowych

Przy gromadzeniu danych korzystaliśmy z następujących źródeł:

- <https://data.ecb.europa.eu/search-results?searchTerm=Inflation> - miesięczne wskaźniki HICP dla Polski, Niemiec oraz strefy Euro
- <https://nbp.pl/statystyka-i-sprawozdawczosc/inflacja-bazowa/> - miesięczna inflacja bazowa dla Polski
- <https://stat.gov.pl/wskazniki-makroekonomiczne/> -miesięczne wskaźniki makroekonomiczne dla Polski
- <https://tradingeconomics.com/poland/consumer-price-index-cpi> - wskaźniki CPI dla Polski
- <https://www-genesis.destatis.de/genesis/online?operation=abruftabelleBearbeiten&levelindex=1&levelid=1709382443109&auswahloperation=abruftabelleAuspraegungAuswahlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&code=61111-0002&auswahltext=&werteabruf=Value+retrieval#abreadcrumb> - miesięczne wskaźniki CPI dla Niemiec
- <https://fred.stlouisfed.org/series/CPHPTT01EZM659N> - miesięczne wskaźniki CPI dla strefy Euro

### 3.2. Kalendarium ważnych wydarzeń gospodarczych i ekonomicznych

- 2019-01-17: PE przegłosował stanowisko w sprawie powiązania funduszy z UE z praworządnością
- 2019-01-19: protesty przeciw nowelizacji kodeksu pracy
- 2019-01-23: KE: 215 mln euro z budżetu UE na Baltic Pipe
- 2019-02-02: kolejny marsz „żółtych kamizelek” we Francji
- 2019-02-06: Niemcy obiecują NATO dalsze zwiększanie nakładów na obronność
- 2019-02-07: parlament Ukrainy wniósł do konstytucji zapis o dążeniu do UE i NATO
- 2019-02-08: Ambasadorowie krajów UE dali zielone światło dla dyrektywy dot. Nord Stream 2
- 2019-02-16: Merkel ostrzega w Monachium przed izolowaniem Rosji
- 2019-02-17: Trump apeluje do krajów UE o przyjęcie ponad 800 bojowników IS
- 2019-02-26: PSL zapowiada złożenie projektu ustawy podnoszącej wynagrodzenie nauczycieli o 1000 zł
- 2019-03-01: MliR: 200 mln zł z UE dla firm na innowacje dla seniorów i niepełnosprawnych
- 2019-03-02: Morawiecki: Wspólna Polityka Rolna UE musi być kontynuowana i rozwijana
- 2019-03-12: Właściciele małych sklepów apelują o wycofanie ograniczenia handlu w niedzielę
- 2019-03-15: Zapowiedzi wzrastających cen paliwa
- 2019-03-19: PFR i dwie zagraniczne firmy przejmują DCT Gdańsk
- 2019-03-29: Brytyjska Izba Gmin odrzuciła projekt umowy w sprawie brexitu
- 2019-03-31: Czaputova wygrała wybory prezydenckie na Słowacji
- 2019-04-03: Manifestacja rolników
- 2019-04-06: Francja: "żółte kamizelki" protestowały po raz 21.
- 2019-04-12: Sejm za nowelą zakładającą dalsze uszczelnienie systemu VAT
- 2019-04-13: Francja: starcia "żółtych kamizelek" z policją w Tuluzie
- 2019-04-23: Manifestacja nauczycieli przed MEN w Warszawie
- 2019-05-09: Korea Północna wystrzeliła niezidentyfikowane pociski
- 2019-05-14: ZNP chce od czerwca podwyżek o 9,6 proc. dla strajkujących nauczycieli
- 2019-05-15: Dane o PKB w I kw. prognozują wzrost ok. 4,5 proc. w całym 2019 r.
- 2019-05-19: Zapowiedź bezpłatnego pobytu przy dziecku w szpitalu od lipca
- 2019-05-26: Wybory do PE
- 2019-05-28: PKO BP podnosi prognozę wzrostu PKB roku 2019 z 4,2 proc. do 4,6 proc.
- 2019-06-01: Trump krytykuje May w sprawie brexitu i popiera Johnsona
- 2019-06-03: wizyta państwowa Donalda Trumpa w Wielkiej Brytanii
- 2019-07-01: Uchwalenie ustawy dotyczącej 500+
- 2019-08-28: MF: w 2020 r. wzrost akcyzy o 3 proc. na alkohol i tytoń
- 2019-09-12: 3,5 mld zł z funduszy norweskich przeznaczone dla Polski
- 2019-09-17: Gowin: dla małych firm chcemy wzrost pensji minimalnej zneutralizować przez obniżenie ZUS

- 2019-09-29: Demonstracje przeciwko aresztowaniu katalońskich separatystów w Hiszpanii
- 2019-10-03: Kurs złotego wobec euro wzrósł po ogłoszeniu wyroku TSUE ws. kredytów indeksowanych we frankach szwajcarskich
- 2019-10-04: UE przeznaczy 880 mln euro na modernizację polskich kolei
- 2019-10-11: NCBR: ok. 200 mln zł na innowacyjne nawozy dla rolnictwa
- 2019-10-13: Wybory parlamentarne 2019
- 2019-10-18: Protesty w Hiszpanii przeciwko wyrokowi Sądu Najwyższego wobec regionalnych separatystów
- 2019-11-04: Szczyt szefów parlamentów w Budapeszcie
- 2019-11-07: Ewakuacja 1600 migrantów z improwizowanych obozów pod Paryżem
- 2019-11-11: Polska dołącza do amerykańskiego programu ruchu bezwizowego
- 2019-11-14: Rząd przyjął projekt, przewidujący 10-proc. wzrost akcyzy na alkohole i wyroby tytoniowe
- 2019-11-16: Czechy: ponad 250 tys. osób demonstrowało w Pradze przeciwko premierowi Babiszowi
- 2019-12-05: NATO potwierdza zobowiązanie do kolektywnej obrony i wzrost wydatków wojskowych
- 2020-01-01: Mikrorachunki w podatkach
- 2020-01-01: Wzrost minimalnej płacy do 2600 zł
- 2020-01-13: Hiszpania: wielotysięczne manifestacje na rzecz jedności kraju
- 2020-01-21: Przeciętne wynagrodzenie po raz pierwszy przebiło poziom 5,5 tys. zł
- 2020-01-22: Trzy ofiary śmiertelne nowego koronawirusa w Chinach
- 2020-01-28: Chiny: już ponad 100 ofiar śmiertelnych nowego koronawirusa
- 2020-01-31: W Rosji wykryto dwa pierwsze przypadki koronawirusa
- 2020-02-02: Filipiny: pierwsza śmiertelna ofiara koronawirusa poza Chinami
- 2020-02-05: Chiny: 491 ofiar śmiertelnych koronawirusa
- 2020-02-07: Chiny: 636 ofiar śmiertelnych koronawirusa
- 2020-02-09: Chiny: liczba ofiar śmiertelnych koronawirusa wzrosła do 811
- 2020-02-15: Zelenski: wojna na Ukrainie to wojna w Europie (aneksja Krymu)
- 2020-02-16: Liczba ofiar śmiertelnych koronawirusa w Chinach wzrosła do 1665
- 2020-02-19: Liczba ofiar śmiertelnych koronawirusa w Chinach przekroczyła 2 tys.
- 2020-02-22: Włochy: już 61 zarażonych koronawirusem, 2 osoby nie żyją
- 2020-02-24: UE przeznacza 230 mln euro na walkę z koronawirusem na świecie
- 2020-03-05: Ponad 93 tys. przypadków koronawirusa na świecie; 3,2 tys. zgonów
- 2020-03-07: 233 osoby zmarły we Włoszech z powodu koronawirusa
- 2020-03-09: 16 potwierdzonych przypadków koronawirusa w Polsce
- 2020-03-09: Ponad 1000 przypadków koronawirusa w Niemczech
- 2020-03-10: Włochy: szturm na sklepy w kraju
- 2020-03-11: Zamknięte placówki oświatowe, uczelnie odwołują zajęcia; instytucje kultury zawieszają funkcjonowanie
- 2020-03-12: Czechy: wprowadzono na 30 dni stan wyjątkowy w związku z koronawirusem
- 2020-03-12: W Polsce 51 przypadków zakażenia koronawirusem, w tym jeden zgon
- 2020-03-15: Liczba potwierdzonych przypadków zakażenia koronawirusem wzrosła do 123
- 2020-03-16: KE zapowiada tymczasowe ograniczenia dotyczące podróży do UE



- 2020-03-17: Włochy: już 2503 zmarłych zakażonych koronawirusem; w ciągu doby 345 zgonów
- 2020-03-26: Hiszpania: w ciągu doby 655 nowych zgonów na Covid-19
- 2020-03-26: Koalicja Obywatelska chce wprowadzenia w Polsce stanu klęski żywiołowej
- 2020-03-27: Macedonia Północna przystąpiła do NATO
- 2020-03-28: Sejm uchwalił specustawę o wsparciu dla firm w związku z epidemią
- 2020-03-30: Hiszpania: w ciągu doby liczba zgonów na Covid-19 wzrosła o 812
- 2020-03-31: Premier: zakładamy scenariusz, że polska gospodarka będzie zamknięta na dłuższy czas
- 2020-04-01: Blisko 900 osób zmarło ubiegłej doby w Hiszpanii z powodu koronawirusa
- 2020-04-06: Premier: spodziewamy się, że szczyt zachorowań jest przed nami - w maju, czerwcu
- 2020-04-08: Premier: proponujemy tarczę finansową; kierujemy co najmniej 100 mld zł do firm
- 2020-04-12: W. Brytania: 10 612 zmarłych z powodu koronawirusa
- 2020-04-15: Państwa G20 uzgodniły zawieszenie spłaty długów najbiedniejszych krajów
- 2020-04-25: Liczba ofiar śmiertelnych koronawirusa na świecie przekroczyła 200 tys.
- 2020-04-27: Ponad 3 mln zakażeń Covid-19, szybki rozwój epidemii w Europie Wschodniej
- 2020-04-27: Orlen kupił ponad 333 mln akcji Energi
- 2020-05-04: 3,5 mln zakażonych koronawirusem na całym świecie
- 2020-05-11: Francja wychodzi z kwarantanny po ośmiu tygodniach:
- 2020-05-13: Ponad 8,27 mld zł trafiło do przedsiębiorców dzięki tarczy antykrzysowej
- 2020-05-21: Ponad 5 mln przypadków koronawirusa na świecie
- 2020-05-23: Hiszpania: samochodowy protest w przeciw rządowi w Madrycie
- 2020-06-08: Ponad 7 mln zakażeń koronawirusem na świecie
- 2020-09-23: UOKiK nałożył ponad 60 mln zł kar na BNP Paribas, Millennium i Santander Bank
- 2020-09-29: Ponad milion ofiar śmiertelnych koronawirusa na świecie
- 2020-10-20: Prezydent podpisał ustawę o Funduszu Medycznym
- 2020-10-26: Polska uzyskała 660 mln euro w unijnym programie Horyzont 2020
- 2020-10-27: Premier: wszystkie przedsiębiorstwa dotknięte obostrzeniami zostaną objęte programem wsparcia
- 2020-10-31: Lockdown w Anglii
- 2020-11-05: Spadek PKB w strefie euro na poziomie 7,8 proc.
- 2020-11-08: Joe Biden wybrany na prezydenta USA
- 2020-11-29: Sklepy w galeriach i parkach handlowych otwarte dla kupujących
- 2020-12-05: Demonstracje we Francji
- 2020-12-08: W Wielkiej Brytanii rozpoczynają się szczepienia przeciwko koronawirusowi
- 2020-12-14: UOKiK nałożył 723 miliony złotych kary na Biedronkę
- 2021-01-14: Rusza Tarcza Finansowa 2.0 Polskiego Funduszu Rozwoju
- 2021-01-23: Około 20 tysięcy lokali gastronomicznych wznowiło działalność po 18 stycznia

- 2021-01-27: Ponad 100 mln zakażeń koronawirusem na całym świecie
- 2021-01-28: Odmrożony handel i kultura
- 2021-01-30: Na świecie podano już 87 milionów dawek szczepionki
- 2021-02-05: Otwarte hotele, kina i teatry
- 2021-03-03: Ambasadorowie UE uzgodnili przedłużenie sankcji za podważanie integralności terytorialnej Ukrainy
- 2021-03-12: Premier zapowiada pomoc dla firm – kolejne 4,5 mld zł wsparcia
- 2021-03-24: Najwięcej nowych przypadków zakażenia koronawirusem od początku pandemii
- 2021-04-23: Dodatkowy zasiłek opiekuńczy wydłużony do 9 maja
- 2021-05-04: Centra handlowe w pełni otwarte
- 2021-05-15: Prawo i Sprawiedliwość zaprezentowało założenia Polskiego Ładu
- 2021-05-23: Nadzwyczajny szczyt Rady Europejskiej
- 2021-05-28: Ukraina zamyka przestrzeń powietrzną dla białoruskich samolotów
- 2021-06-02: Dodatkowy zasiłek opiekuńczy wydłużony do 25 czerwca
- 2021-08-31: Stan wyjątkowy na granicy z Białorusią
- 2021-09-14: Minister Czarnek zapowiada duże podwyżki dla nauczycieli
- 2021-10-14: Rekord 4. fali koronawirusa
- 2021-10-17: Protesty przeciwko przepustkom sanitarnym we Francji
- 2021-11-23: Siłowe próby przekroczenia granicy przez Białorusinów
- 2021-11-30: Rząd obniża akcyzę na paliwa i energię
- 2021-12-08: Olaf Scholz nowym kanclerzem Niemiec
- 2021-12-14: Zamieszki w Niemczech wywołane przez koronasceptyków
- 2021-12-21: Władimir Putin oskarża Zachód o napięcia w Europie
- 2022-01-31: Dania zniosła wszystkie restrykcje
- 2022-02-13: Szef MSWiA: przygotowujemy się na napływ uchodźców z Ukrainy
- 2022-02-14: Rząd brytyjski zapowiada nowy pakiet pomocy wojskowej dla Ukrainy
- 2022-02-18: Putin: wspólnie z Białorusią zapewnimy sobie bezpieczeństwo
- 2022-02-24: Atak Rosji na Ukrainę
- 2022-03-03: Prezydent Zełenski chce bezpośrednio negocjować z Władimirem Putinem
- 2022-03-03: Do Polski przybyło ponad 575 tys. uchodźców z Ukrainy
- 2022-03-09: ONZ: z Ukrainy uciekło już 2,1-2,2 mln uchodźców
- 2022-03-12: Z Ukrainy do Polski wjechało około 1,6 mln uchodźców
- 2022-03-16: Prezydent Ukrainy nawołuje do wstrzymania handlu z Rosją
- 2022-03-24: Koniec obowiązku noszenia maseczek, izolacji i kwarantanny
- 2022-03-28: Atak rakietowy na Łuck przeprowadzony z terytorium Białorusi
- 2022-04-06: Rosyjskie wojska wycofały się spod Kijowa, starają się zająć Donbas
- 2022-04-07: Ukraina apeluje do krajów NATO o zwiększenie dostaw broni
- 2022-04-27: Rosyjski gaz nie płynie już do Polski przez Jamal
- 2022-05-11: Unia Europejska rezygnuje z obowiązku noszenia maseczek w samolotach
- 2022-05-19: Sekretarz generalny NATO: opowiadamy się za szybkim przyjęciem Szwecji i Finlandii
- 2022-06-04: Prezydent Zełenski wzywa amerykańskie miasta do zerwania stosunków z Rosją
- 2022-06-13: Antywojenne protesty w Moskwie

- 2022-08-28: Prezydent Zełenski: powrócimy na Donbas i resztę okupowanych terytoriów
- 2022-09-02: Gazprom wstrzymał dostawy gazu przez NS1
- 2022-09-06: Boris Johnson oficjalnie zrezygnował ze stanowiska premiera
- 2022-09-08: Zmarła królowa Elżbieta II
- 2022-09-14: Rosja zwiększa zakupy broni od Iranu i Korei Płn.
- 2022-09-15: Parlament Europejski zatwierdził pożyczkę dla Ukrainy
- 2022-09-21: Putin ogłosił częściową mobilizację w Rosji
- 2022-09-24: Putin wprowadził kary do 10 lat więzienia za dezercję i poddanie się
- 2022-10-06: Węgry odsuwają głosowanie nad przyjęciem Szwecji i Finlandii do NATO
- 2022-10-14: Rząd przyjął projekt ustawy o maksymalnych cenach prądu
- 2022-11-29: Szef NATO: Putin ponosi porażkę na Ukrainie i odpowiada zwiększoną brutalnością
- 2022-12-01: Unia Europejska przekazuje ponad 1 mln euro na rozminowanie Ukrainy
- 2022-12-21: USA ogłasza, że da Ukrainie najnowszą wersję systemów Patriot
- 2023-01-05: USA i Niemcy przeka Ukrainie wozy bojowe Bradley i Marder
- 2023-01-14: Szef NATO chce zwiększenia produkcji zbrojeniowej, aby nadal wspierać Ukrainę
- 2023-01-16: Białoruś i Rosja rozpoczęły wspólne ćwiczenia lotnictwa wojskowego
- 2023-01-18: USA przekazują Ukrainie amunicję ze swoich magazynów w Izraelu
- 2023-01-23: Przewodniczący Rady Europejskiej: nie ma bezpiecznej Europy bez wolnej i bezpiecznej Ukrainy
- 2023-01-25: Niemcy wysyłają czołgi Leopard Ukrainie
- 2023-01-28: Petr Pavel prezydentem Czech
- 2023-01-30: Boris Johnson: Putin groził mi uderzeniem raketowym
- 2023-02-10: Pocisk rosyjski przeleciał 35 km od granicy Rumunii
- 2023-02-20: Prezydent Zełenski: dziękuję za potęgę ukraińsko-amerykańskiej współpracy
- 2023-02-24: UE zatwierdziła 10. pakiet sankcji wobec Rosji
- 2023-03-09: Francja, Węgry i Bułgaria blokują sankcje UE na rosyjski sektor energii atomowej
- 2023-03-13: Premier Gruzji oskarża prezydenta Ukrainy o ingerowanie w wewnętrzne sprawy państwa
- 2023-03-16: Spadek o 30 procent akcji flagowego szwajcarskiego banku Credit Suisse
- 2023-03-22: Prezydent Ukrainy: zwyciężymy w heroicznej bitwie o Donbas
- 2023-03-28: Rosja nasila działania wojskowe na Dalekim Wschodzie
- 2023-04-04: Finlandia dołącza do NATO
- 2023-04-06: Prezydent Ukrainy: Polacy pokazali, że między ukraińskimi i polskimi sercami nie ma barier
- 2023-04-08: Import zboża z Ukrainy wstrzymany
- 2023-04-21: Stoltenberg: wszyscy członkowie Sojuszu zgadzają się, że Ukraina powinna stać się członkiem NATO
- 2023-04-22: Spadek poparcia dla kanclerza Niemiec Olafa Scholza
- 2023-05-12: USA przekazują Ukrainie czołgi Abrams
- 2023-05-14: Zapowiedź 800 plus od przyszłego roku
- 2023-05-17: Dofinansowanie linii autobusowych. Rząd zwiększa pulę do 1 mld zł

- 2023-05-19: Grupa G7 nakłada kolejne sankcje na Rosję
- 2023-05-20: Ukraina dostaje myśliwce F-16
- 2023-06-03: Prezydent Zełenski: Ukraina gotowa do rozpoczęcia kontrofensywy
- 2023-09-17: Kryzys migracyjny na Lampedusie
- 2023-09-20: Podwyżki stóp procentowych EBC
- 2023-10-11: Sekretarz obrony USA: będziemy wspierać Ukrainę tak długo, jak będzie to konieczne
- 2023-11-02: Putin unieważnił ratyfikację traktatu o zakazie prób nuklearnych
- 2023-11-17: Prezydent Zełenski: Rosja planuje zasiał na Ukrainie chaos
- 2023-11-25: Rekordowy atak rosyjskich dronów na Kijów
- 2023-12-09: Minister finansów podpisał rozporządzenie o zerowym VAT na żywność
- 2024-01-01: Putin zapowiedział zintensyfikowanie ataków na Ukrainę
- 2024-01-03: UE przedstawiła nowe sankcje na Rosję
- 2024-01-08: Zmasowany atak rakietowy w Ukrainie
- 2024-01-11: Trump: to, czy będziemy bronić NATO zależy od tego, jak będą nas traktować
- 2024-01-24: Ogólnopolski protest rolników
- 2024-01-25: Ukraina: w nocy zniszczyliśmy 11 z 14 dronów wystrzelonych przez Rosję
- 2024-02-06: Białoruskie media: Łukaszenka zezwolił wojsku strzelać do cywilów
- 2024-02-09: Morawiecki: przymusowa relokacja nielegalnych migrantów zaszkodzi Europie
- 2024-02-09: Ogólnopolski protest rolników
- 2024-02-11: Trump grozi, że nie będzie chronił przed Rosją krajów NATO zalegających z płatnościami
- 2024-02-12: Rosja przeprowadziła atak dronami na Ukrainę
- 2024-02-16: Śmierć Alekseja Nawalnego
- 2024-02-20: Kanada przekazuje Ukrainie ponad 800 dronów
- 2024-02-20: Ogólnopolski protest rolników
- 2024-02-24: Szef NATO: Ukraina dołączy do Sojuszu Północnoatlantyckiego. To kwestia czasu
- 2024-02-27: Rosja zakazuje eksportu benzyny na przez pół roku

### 3.3. Opis danych

#### **CPI (Consumer Price Index):**

- CPI jest wskaźnikiem cen, który mierzy średnią zmianę cen koszyka towarów i usług konsumpcyjnych zakupionych przez gospodarstwa domowe.
- Skład koszyka towarów i usług w CPI jest ustalany na podstawie wzorcowego zestawu produktów konsumpcyjnych, które typowo nabywa przeciętne gospodarstwo domowe.
- CPI jest szeroko stosowany przez rządy, banki centralne i innych ekonomistów do śledzenia inflacji i podejmowania decyzji politycznych oraz ekonomicznych.

#### **HICP (Harmonized Index of Consumer Prices):**

- HICP jest międzynarodową wersją CPI, stworzoną przez Eurostat, biuro statystyczne Unii Europejskiej.
- Jest to spójny wskaźnik cen, który umożliwia porównywanie inflacji między krajami członkowskimi UE.
- HICP uwzględnia szerszy zakres produktów niż tradycyjne CPI, co pozwala na porównywanie zmian cen w różnych krajach.

### **Wskaźniki makroekonomiczne:**

#### **a) Budownictwo**

- Produkcja budowlano-montażowa (ceny stałe)
  - Uwaga:  
Dane obejmują podmioty gospodarcze o liczbie pracujących powyżej 9 osób.
- Mieszkania oddane do użytkowania (w szt.)

#### **b) Budżet państwa (Dane Ministerstwa Finansów)**

- Dochody budżetu państwa (od początku roku do końca okresu, w mln zł)
- Wydatki budżetu państwa (od początku roku do końca okresu, w mln zł)
- Wynik (saldo) budżetu państwa (od początku roku do końca okresu, w mln zł)

#### **c) Handel wewnętrzny**

- Sprzedaż detaliczna towarów (ceny stałe)
  - Uwaga:  
Dane obejmują podmioty gospodarcze o liczbie pracujących powyżej 9 osób.
- Obroty w handlu detalicznym (Przychody netto ze sprzedaży produktów, towarów i materiałów)

#### **d) Handel zagraniczny**

- Eksport towarów (ceny bieżące, w mln zł)
- Eksport towarów (ceny stałe)
- Import towarów (ceny bieżące, w mln zł)
- Import towarów (ceny stałe)
- Saldo obrotów towarowych handlu zagranicznego ogółem

#### **e) Koniunktura konsumencka i gospodarcza**

- Bieżący wskaźnik ufności konsumenckiej (BWUK)
- Wyprowadzający wskaźnik ufności konsumenckiej (WWUK)
- Wskaźnik ogólnego klimatu koniunktury gospodarczej:
  - przetwórstwo przemysłowe
  - budownictwo

- handel; naprawa pojazdów samochodowych
- transport i gospodarka magazynowa
- zakwaterowanie i gastronomia
- działalność finansowa i ubezpieczeniowa
- obsługa rynku nieruchomości

#### f) Przemysł

- Produkcja sprzedana przemysłu ogółem (ceny stałe)
  - górnictwo i wydobywanie
  - przetwórstwo przemysłowe
  - wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę
  - dostawa wody; gospodarowanie ściekami i odpadami; rekultywacja
- Produkcja sprzedana przemysłu według głównych grupowań przemysłowych (ceny stałe)
  - dobra zaopatrzeniowe
  - dobra inwestycyjne
  - dobra konsumpcyjne trwałe
  - dobra konsumpcyjne nietrwałe
  - dobra związane z energią
- Nowe zamówienia w przemyśle (ceny bieżące)

#### g) Rynek pracy

- Przeciętne zatrudnienie w sektorze przedsiębiorstwa (w tys.)
- Bezrobotni zarejestrowani (stan w końcu okresu)
- Stopa bezrobocia rejestrowanego (stan w końcu okresu)

#### h) Transport

- Przewozy ładunków w transporcie kolejowym (w mln tonokilometrów)
- Przewozy ładunków w transporcie kolejowym (od początku roku do końca okresu, w mln tonokilometrów)

#### i) Usługi biznesowe

- Obroty w usługach biznesowych (ceny bieżące):
  - działalność wydawnicza
  - działalność związana z oprogramowaniem i doradztwem w zakresie informatyki oraz działalność powiązana
  - działalność usługowa w zakresie informacji
  - działalność związana z obsługą rynku nieruchomości
  - działalność prawnicza, rachunkowo-księgowa i doradztwo podatkowe
  - doradztwo związane z zarządzaniem
  - działalność w zakresie architektury i inżynierii; badania i analizy techniczne
  - reklama, badanie rynku i opinii publicznej

- pozostała działalność profesjonalna, naukowa i techniczna
- wynajem i dzierżawa
- działalność związana z zatrudnieniem
- działalność detektywistyczna i ochroniarska
- działalność usługowa związana z utrzymaniem porządku w budynkach i zagospodarowaniem terenów zieleni
- działalność związana z administracyjną obsługą biura i pozostała działalność wspomagająca prowadzenie działalności gospodarczej
- Wskaźnik produkcji usług biznesowych (ceny stałe):
  - działalność wydawnicza
  - działalność związana z oprogramowaniem i doradztwem w zakresie informatyki oraz działalność powiązana
  - działalność usługowa w zakresie informacji
  - działalność związana z obsługą rynku nieruchomości
  - działalność prawnicza, rachunkowo-księgowa i doradztwo podatkowe
  - doradztwo związane z zarządzaniem
  - działalność w zakresie architektury i inżynierii; badania i analizy techniczne
  - reklama, badanie rynku i opinii publicznej
  - pozostała działalność profesjonalna, naukowa i techniczna
  - wynajem i dzierżawa
  - działalność związana z zatrudnieniem
  - działalność detektywistyczna i ochroniarska
  - działalność usługowa związana z utrzymaniem porządku w budynkach i zagospodarowaniem terenów zieleni
  - działalność związana z administracyjną obsługą biura i pozostała działalność wspomagająca prowadzenie działalności gospodarczej

#### j) Wskaźniki cen

- Wskaźniki cen skupu pszenicy (bez ziarna siewnego)
- Wskaźniki cen skupu żyta (bez ziarna siewnego)
- Wskaźniki cen skupu bydła (bez cieląt)
- Wskaźniki cen skupu trzody chlewnej
- Wskaźniki cen skupu mleka
- Relacje cen skupu żywca wieprzowego do cen żyta na targowiskach
- Wskaźniki cen produkcji sprzedanej w przemyśle
- Wskaźniki cen produkcji sprzedanej w przemyśle
  - w górnictwie i wydobywaniu
  - w przetwórstwie przemysłowym
  - w wytwarzaniu i zaopatrywaniu w energię elektryczną, gaz, parę wodną i gorącą wodę
  - w dostawie wody; gospodarowaniu ściekami i odpadami; rekultywacji
- Wskaźniki cen produkcji budowlano-montażowej
- Wskaźniki cen usług transportu i gospodarki magazynowej
- Wskaźniki cen usług telekomunikacji
- Wskaźniki cen towarów i usług konsumpcyjnych
  - żywność i napoje bezalkoholowe

- napoje alkoholowe i wyroby tytoniowe
- odzież i obuwie
- użytkowanie mieszkania lub domu i nośniki energii
- wyposażenie mieszkania i prowadzenie gospodarstwa domowego
- zdrowie
- transport
- łączność
- rekreacja i kultura
- edukacja
- Wskaźniki cen transakcyjnych eksportu
- Wskaźniki cen transakcyjnych importu
- Terms of trade

k) Wynagrodzenia i świadczenia społeczne

- Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstw
- Przeciętne miesięczne realne wynagrodzenie brutto w sektorze przedsiębiorstw
- Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych
- Przeciętna miesięczna realna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych
- Przeciętna miesięczna nominalna emerytura i renta rolników indywidualnych brutto
- Przeciętna miesięczna realna emerytura i renta rolników indywidualnych brutto

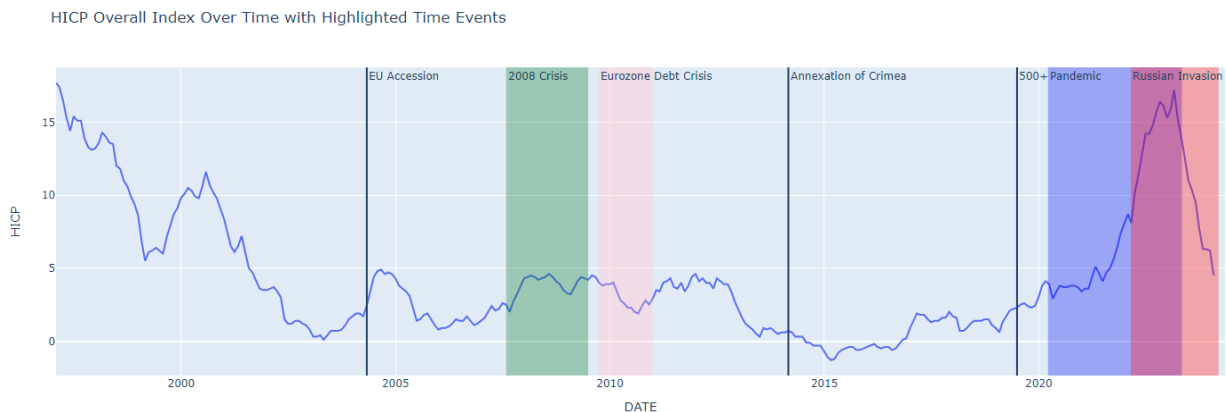
### 3.4. Wstępny przegląd danych

Na [Rysunku 1](#), możemy zaobserwować 325 zebranych obserwacji dla wskaźnika HICP od stycznia 1997 roku.

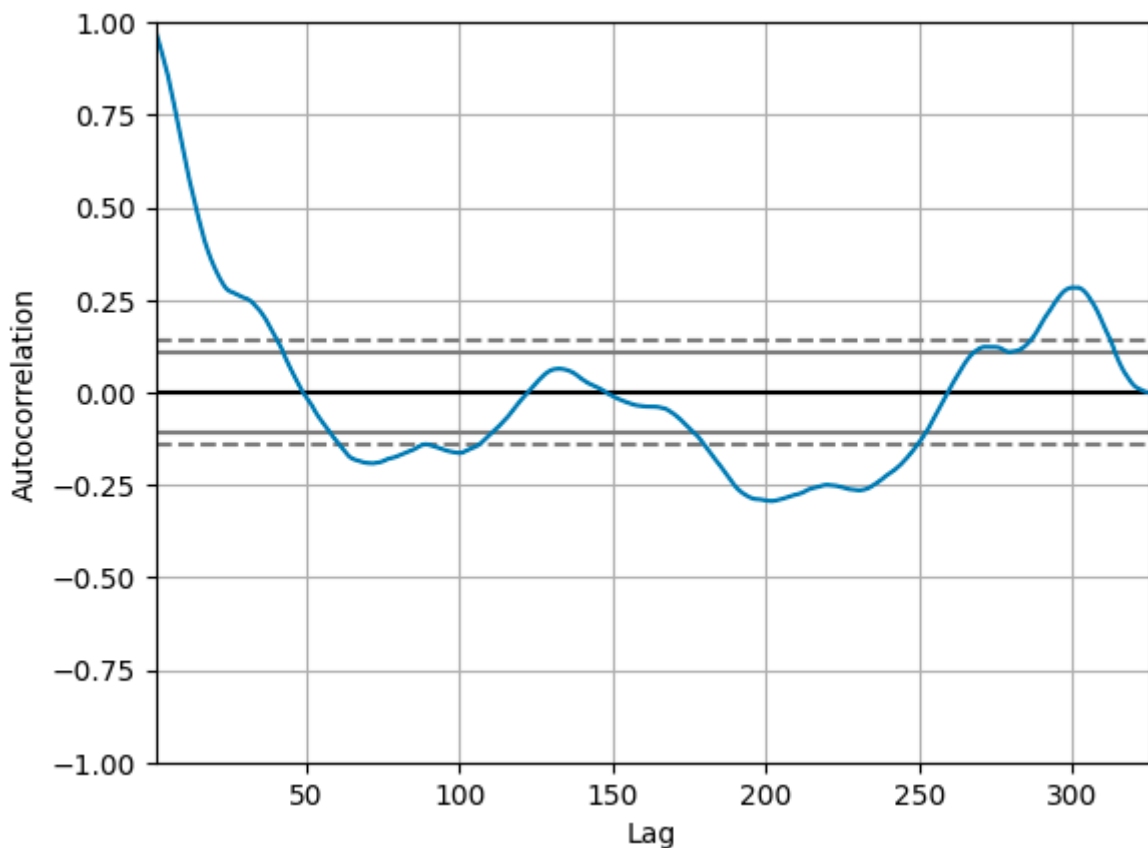
| HICP         |            |
|--------------|------------|
| <b>count</b> | 325.000000 |
| <b>mean</b>  | 4.380923   |
| <b>std</b>   | 4.369960   |
| <b>min</b>   | -1.300000  |
| <b>25%</b>   | 1.400000   |
| <b>50%</b>   | 3.400000   |
| <b>75%</b>   | 5.500000   |
| <b>max</b>   | 17.700000  |



**Rysunek 1:** Metoda *describe* z biblioteki Pandas dla wskaźnika HICP od stycznia 1997 roku



**Rysunek 2:** Wykres wartości HICP z wyszczególnionymi wydarzeniami, które miały największy wpływ na inflację



**Rysunek 3:** Wykres autokorelacji dla wskaźnika HICP

Na [Rysunku 4](#) oraz [Rysunku 5](#), zobrazowana jest prosta analiza statystyczna. Otrzymana ona została przy użyciu funkcji *describe* z biblioteki Pandas. Już na wstępnej eksploracji danych, możemy zaobserwować, że w niektórych kolumnach mogą być dane brakujące. Na przykład Produkcja budowlano-montażowa (ceny stałe) ma 229 wierszy natomiast większość pozostałych kolumn posiada ich 288. Nie są to jednak wszystkie kolumny, a tylko

pierwszych 10 oraz ostatnich 10, jednak już wiemy, że będzie wymagana bardziej dogłębna eksploracja danych.

|       | Dochody budżetu państwa | Wydatki budżetu państwa | Wynik (saldo) budżetu państwa | Bezrobotni zarejestrowani | Bezrobotni zarejestrowani (B) | Produkcja budowlano-montażowa (ceny stałe) | Mieszkania oddane do użytkowania | Dochody budżetu państwa (w mln zł) | Wydatki budżetu państwa (w mln zł) | Wynik (saldo) budżetu państwa (w mln zł) | ... |
|-------|-------------------------|-------------------------|-------------------------------|---------------------------|-------------------------------|--|----------------------------------|------------------------------------|------------------------------------|--|-----|
| count | 288.000000              | 288.000000              | 288.000000                    | 289.000000                | 289.000000                    | 229.000000                                 | 289.000000                       | 288.000000                         | 288.000000                         | 288.000000                               | ... |
| mean  | 154853.935868           | 168259.121562           | 168259.121562                 | 1889.057093               | 99.682353                     | 104.326638                                 | 13043.131488                     | 154853.935868                      | 168259.121562                      | 168259.121562                            | ... |
| std   | 111200.648077           | 113311.114441           | 113311.114441                 | 780.311832                | 2.831163                      | 24.554319                                  | 4815.565980                      | 111200.648077                      | 113311.114441                      | 113311.114441                            | ... |
| min   | 10250.300000            | 12038.600000            | 12038.600000                  | 770.400000                | 94.200000                     | 29.500000                                  | 4991.000000                      | 10250.300000                       | 12038.600000                       | 12038.600000                             | ... |
| 25%   | 66662.900000            | 82545.100000            | 82545.100000                  | 1092.200000               | 97.800000                     | 99.300000                                  | 9545.000000                      | 66662.900000                       | 82545.100000                       | 82545.100000                             | ... |
| 50%   | 129383.900000           | 146473.000000           | 146473.000000                 | 1855.300000               | 99.300000                     | 107.900000                                 | 12156.000000                     | 129383.900000                      | 146473.000000                      | 146473.000000                            | ... |
| 75%   | 214818.125000           | 233661.950000           | 233661.950000                 | 2487.900000               | 101.300000                    | 116.300000                                 | 16183.000000                     | 214818.125000                      | 233661.950000                      | 233661.950000                            | ... |
| max   | 574121.500000           | 659687.100000           | 659687.100000                 | 3344.200000               | 110.900000                    | 151.600000                                 | 43492.000000                     | 574121.500000                      | 659687.100000                      | 659687.100000                            | ... |

**Rysunek 4:** Prosta analiza statystyczna wskaźników makroekonomicznych dla pierwszych 10 kolumn

| ... | Terms of trade (A) | Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (w zł) | Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (B) | Przeciętne miesięczne realne wynagrodzenie brutto w sektorze przedsiębiorstwa (B) | Przeciętne miesięczne nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (w zł) | Przeciętne miesięczne nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (B) | Przeciętne miesięczne realna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (B) | Przeciętne miesięczne nominalna emerytura i renta rolników indywidualnych brutto (w zł) | Przeciętne miesięczne nominalna emerytura i renta rolników indywidualnych brutto (B) | Przeciętne miesięczne realna emerytura i renta rolników indywidualnych brutto (B) |
|-----|--------------------|---|--|---|--|---|--|---|--|---|
| ... | 287.000000         | 229.000000  | 228.000000   | 228.000000  | 289.000000   | 289.000000  | 289.000000   | 289.000000  | 289.000000   | 289.000000  |
| ... | 100.644295         | 4292.313277   | 100.602632   | 100.342394  | 1799.534602  | 100.500346  | 100.201730   | 1048.595398   | 100.476817   | 100.177855  |
| ... | 4.207183           | 1327.439791   | 4.051039   | 4.088249  | 645.237321   | 1.677827  | 1.673413   | 307.939863  | 3.289670   | 3.233612  |
| ... | 90.600000          | 2385.269631   | 88.000000  | 87.100000   | 847.590000   | 98.500000   | 97.100000  | 579.420000  | 79.100000  | 78.500000   |
| ... | 97.750000          | 3332.650000   | 98.900000  | 98.500000   | 1184.210000  | 100.000000  | 99.600000  | 770.990000  | 98.300000  | 98.300000   |
| ... | 100.300000         | 3980.920000   | 100.900000   | 100.500000  | 1746.380000  | 100.100000  | 99.900000  | 1004.940000   | 100.100000   | 99.900000   |
| ... | 103.450000         | 5104.460000   | 102.625000   | 102.525000  | 2172.210000  | 100.300000  | 100.300000   | 1220.300000   | 101.900000   | 101.500000  |
| ... | 112.300000         | 8032.960000   | 110.300000   | 109.919840  | 3463.830000  | 116.600000  | 115.300000   | 1929.170000   | 127.100000   | 126.000000  |

**Rysunek 5:** Prosta analiza statystyczna wskaźników makroekonomicznych dla ostatnich 10 kolumn.

Na [Rysunku 6](#) przedstawiony został fragment macierzy korelacji stworzonej dla wszystkich wskaźników makroekonomicznych. Przy pomocy biblioteki Pandas zostały wyekstraktowane największe unikalne korelacje, które zostały wypisane pod wykresem. Gdybyśmy się przyjrżeli dokładniej wynikom korelacji, to możemy zauważyć, że mają one sens. Na przykład wysokie skorelowanie Dochodów budżetu państwa z Wydatkami budżetu państwa lub Bezrobotni zarejestrowani z Stopa bezrobocia rejestrowanego (%).

|                                     | Dochody<br>budżetu<br>państwa | Wydatki<br>budżetu<br>państwa | Wynik<br>(saldo)<br>budżetu<br>państwa | Bezrobotni<br>zarejestrowani | Bezrobotni<br>zarejestrowani<br>(B) | Produkcja<br>budowlano-<br>montażowa<br>(ceny<br>stałe) | Mieszkania<br>oddane do<br>użytkowania | Dochody<br>budżetu<br>państwa<br>(w mln<br>zł) | Wydatki<br>budżetu<br>państwa<br>(w mln<br>zł) | Wynik<br>(saldo)<br>budżetu<br>państwa<br>(w mln<br>zł) | ... | Terms of<br>trade<br>(A) | Przeciętne<br>miesięczne<br>nominalne<br>wynagrodzenie<br>brutto w<br>sektorze<br>przedsiębiorstwa<br>(w zł) | Przeciętne<br>miesięczne<br>nominalne<br>wynagrodzenie<br>brutto w<br>sektorze<br>przedsiębiorstwa<br>(B) |
|-------------------------------------|-------------------------------|-------------------------------|--|------------------------------|-------------------------------------|---|--|--|--|---|-----|--------------------------|--|---|
| Dochody<br>budżetu<br>państwa       | 1.000000                      | 0.985630                      | 0.985630                               | -0.479322                    | -0.128860                           | 0.333587  | 0.607366                               | 1.000000                                       | 0.985630                                       | 0.985630  | ... | -0.010097                | 0.433192   | 0.436848  |
| Wydatki<br>budżetu<br>państwa       | 0.985630                      | 1.000000                      | 1.000000                               | -0.463166                    | -0.077751                           | 0.365891  | 0.571522                               | 0.985630                                       | 1.000000                                       | 1.000000  | ... | 0.089504                 | 0.447430   | 0.467003  |
| Wynik (saldo)<br>budżetu<br>państwa | 0.985630                      | 1.000000                      | 1.000000                               | -0.463166                    | -0.077751                           | 0.365891  | 0.571522                               | 0.985630                                       | 1.000000                                       | 1.000000  | ... | 0.089504                 | 0.447430   | 0.467003  |
| Bezrobotni<br>zarejestrowani        | -0.479322                     | -0.463166                     | -0.463166                              | 1.000000                     | 0.135761                            | -0.056883   | -0.133005                              | -0.479322                                      | -0.463166                                      | -0.463166   | ... | -0.105558                | -0.686558  | -0.102337   |
| Bezrobotni<br>zarejestrowani<br>(B) | -0.128860                     | -0.077751                     | -0.077751                              | 0.135761                     | 1.000000                            | -0.616320   | -0.057935                              | -0.128860                                      | -0.077751                                      | -0.077751   | ... | 0.215031                 | -0.017330  | -0.362739   |
| ...                                 | ...                           | ...                           | ...                                    | ...                          | ...                                 | ...   | ...                                    | ...  | ...  | ...   | ... | ...                      | ...  | ...   |

**Rysunek 6:** Macierz korelacji dla wskaźników makroekonomicznych

**10 największych unikalnych korelacji** (pomijając korelacje równą 1 oraz drugą stronę korelacji):

1. *Produkcja - przetwórstwo przemysłowe (B) i Produkcja sprzedana przemysłu ogółem (ceny stałe, B): 0.9927*
2. *Dochody budżetu państwa i Wydatki budżetu państwa: 0.9856*
3. *Dochody budżetu państwa i Przewozy ładunków w transporcie kolejowym (od początku roku do końca okresu, w mln tonokilometrów): 0.9840*
4. *Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (B) i Przeciętne miesięczne realne wynagrodzenie brutto w sektorze przedsiębiorstwa (B): 0.9833*
5. *Przeciętna miesięczna nominalna emerytura i renta rolników indywidualnych brutto (B) i Przeciętna miesięczna realna emerytura i renta rolników indywidualnych brutto (B): 0.9801*
6. *Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (w zł) i Przeciętna miesięczna nominalna emerytura i renta rolników indywidualnych brutto (w zł): 0.9777*
7. *Eksport towarów (ceny bieżące w mln zł) i Import towarów (ceny bieżące w mln zł): 0.9762*
8. *Przewozy ładunków w transporcie kolejowym (od początku roku do końca okresu, w mln tonokilometrów) i Wydatki budżetu państwa: 0.9727*
9. *Bezrobotni zarejestrowani i Stopa bezrobocia rejestrowanego (%): 0.9633*
10. *Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (w zł) i Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (w zł): 0.9628*

### 3.5. Weryfikacja jakości danych

Dane w większości są kompletne, niektóre pomiary są wykonywane kwartalnie, a nie miesięcznie.

## 4. Przygotowanie danych

### 4.1. Funkcje oraz kroki w ekstrakcji danych

Ze względu na dużą ilość czynników mających wpływ na inflację, postanowiliśmy ograniczyć dane do pomiarów od stycznia 2019 roku.

Pomiary wskaźnika HICP umieściliśmy w jednej ramce danych wraz ze wskaźnikami makroekonomicznymi.

W pierwszej kolejności stworzyliśmy dwie poniższe funkcje w Pythonie, których celem jest wczytanie danych z plików CSV do ramek danych, zwanych Dataframe w bibliotece Pandas.

```
def excel_range_to_indices(excel_range):
    """
    Convert an Excel-style range into a list of (row_start, row_end, col_start, col_end) in zero-based indexing.

    Parameters:
    - excel_range: str, an Excel-style range, e.g., "A1:B2" or "D13:KF13".

    Returns:
    - A tuple of (row_start, row_end, col_start, col_end).
    """
    import re
    from openpyxl.utils.cell import column_index_from_string

    # Split the range into start and end parts
    start_cell, end_cell = excel_range.split(':')
    # Extract column letters and row numbers from both parts
    col_start, row_start = re.match(r"([A-Z]+)([0-9]+)", start_cell).groups()
    col_end, row_end = re.match(r"([A-Z]+)([0-9]+)", end_cell).groups()

    # Convert column letters to zero-based column indices
    col_start_idx = column_index_from_string(col_start) - 1
    col_end_idx = column_index_from_string(col_end) - 1
    # Convert row numbers to zero-based row indices
    row_start_idx = int(row_start) - 1
    row_end_idx = int(row_end) - 1

    return (row_start_idx, row_end_idx, col_start_idx, col_end_idx)
```

**Rysunek 7:** Funkcja konwertująca tablicę z Excela na listę

```
def add_column_from_excel_to_csv(csv_path, excel_path, sheet_number, excel_range, new_column_name, data_folder_name=''):
    # Convert Excel-style range to indices
    start_row, end_row, start_col, end_col = excel_range_to_indices(excel_range)

    # Load the Excel file
    excel_file_path = f'{data_folder_name}/{excel_path}'
    # Adjust for 0-based indexing in sheet_number
    df_excel_full = pd.read_excel(excel_file_path, sheet_name=sheet_number-1)

    # Extract the specified range
    row_of_interest = df_excel_full.iloc[start_row-1:end_row, start_col-1:end_col]
    # Convert to a DataFrame and reset index if necessary
    df_excel_row = pd.DataFrame(row_of_interest).reset_index(drop=True)

    # Load the CSV file
    df_csv = pd.read_csv(csv_path, parse_dates=['DATE'])
    # Convert 'DATE' column to year and month format
    df_csv['DATE'] = df_csv['DATE'].dt.to_period('M')

    # Add the Excel data as a new column to the CSV DataFrame
    df_csv[new_column_name] = df_excel_row

    return df_csv
```

**Rysunek 8:** Funkcja, dodająca kolumnę z Excela do pliku CSV

```
def add_excel_column_to_df(excel_file_path, sheet_number, excel_range, df, new_column_name, start_month):
    """
    Add data from a specified range in an Excel file as a new column to a pandas DataFrame, aligning with a specified starting month.

    Parameters:
    - excel_file_path: str, path to the Excel file.
    - sheet_number: int, sheet number (1-based indexing).
    - excel_range: str, Excel-style range, e.g., "D13:KF13".
    - df: pandas.DataFrame, DataFrame to add the new column to.
    - new_column_name: str, name of the new column.
    - start_month: str, the starting month for the new data in "YYYY-MM" format.

    Returns:
    - pandas.DataFrame, the DataFrame with the added column.
    """
    # Convert Excel range to indices
    row_start, row_end, col_start, col_end = excel_range_to_indices(excel_range)

    # Read the specified range from the Excel sheet
    df_excel = pd.read_excel(excel_file_path, sheet_name=sheet_number-1, header=None,
                             skiprows=row_start, nrows=row_end-row_start+1,
                             usecols=range(col_start, col_end+1))

    # Since the range might span multiple rows, flatten it to a single list if necessary
    excel_data = df_excel.values.flatten()

    # Find the index of the starting month in the DataFrame
    start_index = df[df['month'] == start_month].index.min()
    # Prepend None values to align the Excel data with the starting month
    excel_data_aligned = [None] * start_index + list(excel_data)

    # Adjust the length of the Excel data to match the DataFrame's length
    length_difference = len(df) - len(excel_data_aligned)
    if length_difference > 0:
        excel_data_aligned += [None] * length_difference

    # Add the Excel data as a new column to the DataFrame
    df[new_column_name] = excel_data_aligned

    return df
```

**Rysunek 9:** Funkcja, dodająca kolumnę z Excela do obiektu DataFrame z biblioteki Pandas

Następnie, korzystając z wcześniej wymienionych funkcji z [Rysunek 7](#), [Rysunek 8](#) i [Rysunek 9](#) umieściliśmy wskaźniki makroekonomiczne razem ze wskaźnikami HICP w jednej ramce danych. Użyliśmy do tego poniższego kodu:

```

csv_file_path = f'{data_folder_name}/ECB Data Portal_20240312120629.csv'

df_csv = pd.read_csv(
    csv_file_path,
    usecols=[0, 2], # Load only columns 0 and 2
    parse_dates=[0], # Parse the first column as dates
    header=0 # Use the first row as header
)

# Convert the 'DATE' column to year and month format without the day
df_csv['DATE'] = df_csv['DATE'].dt.to_period('M')

# Rename columns to 'month' and 'HICP'
df_csv.columns = ['month', 'HICP']

# Load the Excel file
excel_file_path = f'{data_folder_name}/wybrane_miesieczne_wskazniki_makroekonomiczne_cz_i.xlsx'

merged_df = df_csv.copy()

add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='BUDOWNICTWO', excel_ranges=['D12:KF12'], df=merged_df, new_column_name='Mieszkania oddane do użytkowania', start_month='2000-01')
add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='BUDOWNICTWO', excel_ranges=['BL7:KF7'], df=merged_df, new_column_name='Produkcja budowlano-montażowa', start_month='2005-01')

add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='BUDŻET PAŃSTWA', excel_ranges=['D6:KE6'], df=df_csv, new_column_name='Dochody budżetu państwa', start_month='2000-01')
add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='BUDŻET PAŃSTWA', excel_ranges=['D7:KE7'], df=df_csv, new_column_name='Wydatki budżetu państwa', start_month='2000-01')
add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='BUDŻET PAŃSTWA', excel_ranges=['D7:KE7'], df=df_csv, new_column_name='Wynik (saldo) budżetu państwa', start_month='2000-01')

add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='HANDEL WĘWN.', excel_ranges=['M7:HL7'], df=merged_df, new_column_name='Sprzedaż detaliczna towarów (ceny stałe)', start_month='2006-10')
add_excel_column_to_df(excel_file_path=excel_file_path, sheet_name='HANDEL WĘWN.', excel_ranges=['M10:HK10'], df=merged_df, new_column_name='Obroty w handlu detalicznym', start_month='2006-10')

```

**Rysunek 10:** Kod w języku Python do utworzenia ramki danych do pełnej analizy inflacji HICP

```

operations_map = {
    "BUDOWNICTWO": {
        "start_month": "2000-01",
        "operations": [
            {"excel_ranges": ['D7:KF7'], "new_column_name": "Produkcja budowlano-montażowa (ceny stałe)"},
            {"excel_ranges": ['D12:KF12'], "new_column_name": "Mieszkania oddane do użytkowania"},
        ]
    },
    "BUDŻET PAŃSTWA": {
        "start_month": "2000-01",
        "operations": [
            {"excel_ranges": ['D6:KE6'], "new_column_name": "Dochody budżetu państwa (w mln zł)"},
            {"excel_ranges": ['D7:KE7'], "new_column_name": "Wydatki budżetu państwa (w mln zł)"},
            {"excel_ranges": ['D7:KE7'], "new_column_name": "Wynik (saldo) budżetu państwa (w mln zł)"},
        ]
    },
    "HANDEL WĘWN.": {
        "start_month": "2006-01",
        "operations": [
            {"excel_ranges": ['D7:HL7'], "new_column_name": "Sprzedaż detaliczna towarowa (ceny stałe)"},
            {"excel_ranges": ['D10:HL10'], "new_column_name": "Obroty w handlu detalicznym"},
        ]
    },
    "HANDEL ZAGR.": {
        "start_month": "2000-01",
        "operations": [
            {"excel_ranges": ['D6:KE6'], "new_column_name": "Eksport towarów (ceny bieżące w mln zł)"},
            {"excel_ranges": ['D8:KE8'], "new_column_name": "Eksport towarów (ceny stałe B)"},
            {"excel_ranges": ['D9:KE9'], "new_column_name": "Import towarów (ceny bieżące w mln zł)"},
            {"excel_ranges": ['D11:KE11'], "new_column_name": "Import towarów (ceny stałe B)"},
            {"excel_ranges": ['D12:KE12'], "new_column_name": "Saldo obrotów towarowych handlu zagranicznego (w mln zł)"},
        ]
    }
}

```

**Rysunek 11:** Fragment nazw kolumn, które zostały dodane do wcześniejszej ramki danych

Po wykonaniu instrukcji zaprezentowanych na [Rysunek 10](#) otrzymaliśmy następujące dane:

## 4.2. Czyszczenie danych

|    |  |     |          |           |
|----|--|-----|----------|-----------|
| 0  | month  | 325 | non-null | period[M] |
| 1  | HICP   | 325 | non-null | float64   |
| 2  | Dochody budżetu państwa  | 288 | non-null | float64   |
| 3  | Wydatki budżetu państwa  | 288 | non-null | float64   |
| 4  | Wynik (saldo) budżetu państwa  | 288 | non-null | float64   |
| 5  | Bezrobotni zarejestrowani  | 289 | non-null | float64   |
| 6  | Bezrobotni zarejestrowani (B)  | 289 | non-null | float64   |
| 7  | Produkcja budowlano-montażowa (ceny stałe)   | 289 | non-null | object    |
| 8  | Mieszkania oddane do użytkowania   | 289 | non-null | float64   |
| 9  | Dochody budżetu państwa (w mln zł)   | 288 | non-null | float64   |
| 10 | Wydatki budżetu państwa (w mln zł)   | 288 | non-null | float64   |
| 11 | Wynik (saldo) budżetu państwa (w mln zł)   | 288 | non-null | float64   |
| 12 | Sprzedaż detaliczna towarowa (ceny stałe)  | 217 | non-null | object    |
| 13 | Obroty w handlu detalicznym  | 217 | non-null | object    |
| 14 | Eksport towarów (ceny bieżące w mln zł)  | 288 | non-null | float64   |
| 15 | Eksport towarów (ceny stałe B)   | 288 | non-null | object    |
| 16 | Import towarów (ceny bieżące w mln zł)   | 288 | non-null | float64   |
| 17 | Import towarów (ceny stałe B)  | 288 | non-null | object    |
| 18 | Saldo obrotów towarowych handlu zagranicznego (w mln zł)   | 288 | non-null | float64   |
| 19 | Bieżący wskaźnik ufności konsumenckiej (BMUK)  | 289 | non-null | object    |
| 20 | Wyprzedzający wskaźnik ufności konsumenckiej (WMUK)  | 289 | non-null | object    |
| 21 | Koniunktura - przetwórstwo przemysłowe   | 289 | non-null | float64   |
| 22 | Koniunktura - budownictwo  | 289 | non-null | float64   |
| 23 | Koniunktura - handel; naprawa pojazdów samochodowych   | 289 | non-null | float64   |
| 24 | Koniunktura - transport i gospodarka magazynowa  | 289 | non-null | object    |
| 25 | Koniunktura - zakwaterowanie i gastronomia   | 289 | non-null | object    |
| 26 | Koniunktura - działalność finansowa i ubezpieczeniowa  | 289 | non-null | object    |
| 27 | Koniunktura - obsługa rynku nieruchomości  | 289 | non-null | object    |
| 28 | Produkcja sprzedana przemysłu ogółem (ceny stałe, B)   | 229 | non-null | object    |
| 29 | Produkcja - górnictwo i wydobywanie (B)  | 229 | non-null | object    |
| 30 | Produkcja - przetwórstwo przemysłowe (B)   | 229 | non-null | object    |
| 31 | Produkcja - wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę (B) | 229 | non-null | object    |
| 32 | Produkcja - dostawa wody; gospodarowanie ściekami i odpadami; rekultywacja (B)                   | 229 | non-null | object    |

**Rysunek 12:** Zestawienie kilku przykładowych kolumn znajdujących się w omawianej ramce danych oraz liczbie nie pustych wierszy jak i typie przez nie przechowywanym

Tak, jak wcześniej było wspomniane, do analizy wzięte zostały tylko pomiary makroekonomiczne od 2019 roku. Spowodowane to jest 2 argumentami:

- dane sprzed 2019 roku są mocno przestarzałe i nie uwzględniają znaczących wydarzeń gospodarczych takich jak Covid-19 czy wojna na Ukrainie
- dane sprzed 2019 roku są mocno wybrakowane co można było zaobserwować na powyższych Rysunkach

```
data = data[data['month'].dt.year > 2018]
```

**Rysunek 13:** Wyciągnięcie danych tylko od 2019 roku w kodzie w Pythonie

Po wyciągnięciu danych tylko od 2019 roku, liczba wierszy w analizowanej ramce danych zmniejszyła się do około 60, co można zaobserwować na [Rysunek 14](#). Możemy jednak zaobserwować, że pomimo znacznego ograniczenia danych nadal w niektórych kolumnach występują dane brakujące. Po dokładniejszej eksploracji, zauważone zostało, że niektóre dane makroekonomiczne, nie posiadają najnowszych danych za 2024 rok. Zaobserwować to można na [Rysunek 15](#).



| Data columns (total 80 columns): |  |  | Non-Null Count | Dtype     |
|----------------------------------|--|--|----------------|-----------|
| #                                | Column   |  |                |           |
| 0                                | month  |  | 61 non-null    | period[M] |
| 1                                | HICP   |  | 61 non-null    | float64   |
| 2                                | Produkcja budowlano-montażowa (ceny stałe)               |  | 61 non-null    | float64   |
| 3                                | Mieszkania oddane do użytkowania                         |  | 61 non-null    | float64   |
| 4                                | Dochody budżetu państwa (w mln zł)                       |  | 60 non-null    | float64   |
| 5                                | Wydatki budżetu państwa (w mln zł)                       |  | 60 non-null    | float64   |
| 6                                | Wynik (saldo) budżetu państwa (w mln zł)                 |  | 60 non-null    | float64   |
| 7                                | Sprzedaż detaliczna towarowa (ceny stałe)                |  | 61 non-null    | float64   |
| 8                                | Obroty w handlu detalicznym                              |  | 60 non-null    | float64   |
| 9                                | Eksport towarów (ceny bieżące w mln zł)                  |  | 60 non-null    | float64   |
| 10                               | Eksport towarów (ceny stałe B)                           |  | 59 non-null    | float64   |
| 11                               | Import towarów (ceny bieżące w mln zł)                   |  | 60 non-null    | float64   |
| 12                               | Import towarów (ceny stałe B)                            |  | 59 non-null    | float64   |
| 13                               | Saldo obrotów towarowych handlu zagranicznego (w mln zł) |  | 60 non-null    | float64   |
| 14                               | Bieżący wskaźnik ufności konsumenckiej (BWUK)            |  | 61 non-null    | float64   |
| 15                               | Wyrzeczający wskaźnik ufności konsumenckiej (WWUK)       |  | 61 non-null    | float64   |
| 16                               | Koniunktura - przetwórstwo przemysłowe                   |  | 61 non-null    | float64   |
| 17                               | Koniunktura - budownictwo                                |  | 61 non-null    | float64   |
| 18                               | Koniunktura - handel; naprawa pojazdów samochodowych     |  | 61 non-null    | float64   |

**Rysunek 14:** Kilka pierwszych przykładowych kolumn z przefiltrowanej ramki danych wraz z liczbą nie pustych wierszy jak i typie przez nie przechowywanym

| month   | HICP | Produkcja budowlano-montażowa (ceny stałe) | Mieszkania oddane do użytkowania | Dochody budżetu państwa (w mln zł) | Wydatki budżetu państwa (w mln zł) | Wynik (saldo) budżetu państwa (w mln zł) | Sprzedaż detaliczna towarowa (ceny stałe) | Obroty w handlu detalicznym | Eksport towarów (ceny bieżące w mln zł) |
|---------|------|--|----------------------------------|------------------------------------|------------------------------------|--|---|-----------------------------|---|
| 2023-09 | 7.7  | 111.4                                      | 16262.0                          | 417988.3                           | 452682.6                           | 452682.6                                 | 99.7                                      | 99.4                        | 133952.3                                |
| 2023-10 | 6.3  | 102.5                                      | 19549.0                          | 475392.8                           | 511812.3                           | 511812.3                                 | 103.2                                     | 102.3                       | 144134.6                                |
| 2023-11 | 6.3  | 103.1                                      | 18905.0                          | 530330.5                           | 572973.7                           | 572973.7                                 | 97.2                                      | 99.2                        | 137252.7                                |
| 2023-12 | 6.2  | 128.6                                      | 20772.0                          | 574121.5                           | 659687.1                           | 659687.1                                 | 111.0                                     | 98.2                        | 111299.2                                |
| 2024-01 | 4.5  | 36.8                                       | 14762.0                          | NaN                                | NaN                                | NaN                                      | 80.5                                      | NaN                         | NaN                                     |

**Rysunek 15:** Wybrakowane dane dla niektórych kolumn w miesiącu 01.2024

Po usunięciu ostatniego wiersza z ramki danych, udało się otrzymać ostatecznie 59 wierszy oraz 77 kolumn bez danych brakujących w przedziale dat 01.2018 - 12.2023.

Na [Rysunek 16](#) przedstawiona jest ostateczna ramka danych na których będą wykonywane późniejsze operacje jak ekstrakcja cech charakterystycznych, PCA czy predykcja.

| #  | Column   | Non-Null Count | Dtype     |
|----|--|----------------|-----------|
| 0  | month  | 59 non-null    | period[M] |
| 1  | HICP   | 59 non-null    | float64   |
| 2  | Produkcja budowlano-montażowa (ceny stałe)   | 59 non-null    | float64   |
| 3  | Mieszkania oddane do użytkowania   | 59 non-null    | float64   |
| 4  | Dochody budżetu państwa (w mln zł)   | 59 non-null    | float64   |
| 5  | Wydatki budżetu państwa (w mln zł)   | 59 non-null    | float64   |
| 6  | Wynik (saldo) budżetu państwa (w mln zł)   | 59 non-null    | float64   |
| 7  | Sprzedaż detaliczna towarowa (ceny stałe)  | 59 non-null    | float64   |
| 8  | Obroty w handlu detalicznym  | 59 non-null    | float64   |
| 9  | Eksport towarów (ceny bieżące w mln zł)  | 59 non-null    | float64   |
| 10 | Eksport towarów (ceny stałe B)   | 59 non-null    | float64   |
| 11 | Import towarów (ceny bieżące w mln zł)   | 59 non-null    | float64   |
| 12 | Import towarów (ceny stałe B)  | 59 non-null    | float64   |
| 13 | Saldo obrotów towarowych handlu zagranicznego (w mln zł)   | 59 non-null    | float64   |
| 14 | Bieżący wskaźnik ufności konsumenckiej (BWUK)  | 59 non-null    | float64   |
| 15 | Wyprzedzający wskaźnik ufności konsumenckiej (WWUK)  | 59 non-null    | float64   |
| 16 | Koniunktura – przetwórstwo przemysłowe   | 59 non-null    | float64   |
| 17 | Koniunktura – budownictwo  | 59 non-null    | float64   |
| 18 | Koniunktura – handel; naprawa pojazdów samochodowych   | 59 non-null    | float64   |
| 19 | Koniunktura – transport i gospodarka magazynowa  | 59 non-null    | float64   |
| 20 | Koniunktura – zakwaterowanie i gastronomia   | 59 non-null    | float64   |
| 21 | Koniunktura – działalność finansowa i ubezpieczeniowa  | 59 non-null    | float64   |
| 22 | Koniunktura – obsługa rynku nieruchomości  | 59 non-null    | float64   |
| 23 | Produkcja sprzedana przemysłu ogółem (ceny stałe, B)   | 59 non-null    | float64   |
| 24 | Produkcja – górnictwo i wydobywanie (B)  | 59 non-null    | float64   |
| 25 | Produkcja – przetwórstwo przemysłowe (B)   | 59 non-null    | float64   |
| 26 | Produkcja – wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę (B)                 | 59 non-null    | float64   |
| 27 | Produkcja – dostawa wody; gospodarowanie ściekami i odpadami; rekultywacja (B)                                   | 59 non-null    | float64   |
| 28 | Produkcja – dobra zaopatrzeniowe (B)   | 59 non-null    | float64   |
| 29 | Produkcja – dobra inwestycyjne (B)   | 59 non-null    | float64   |
| 30 | Produkcja – dobra konsumpcyjne trwałe (B)  | 59 non-null    | float64   |
| 31 | dProdukcja – obra konsumpcyjne nietrwałe (B)   | 59 non-null    | float64   |
| 32 | Produkcja – dobra związane z energią (B)   | 59 non-null    | float64   |
| 33 | Nowe zamówienia w przemyśle (ceny bieżące, B)  | 59 non-null    | float64   |
| 34 | Przeciętne zatrudnienie w sektorze przedsiębiorstwa (w w tys.)   | 59 non-null    | float64   |
| 35 | Przeciętne zatrudnienie w sektorze przedsiębiorstwa (B)  | 59 non-null    | float64   |
| 36 | Bezrobotni zarejestrowani (w tys.)   | 59 non-null    | float64   |
| 37 | Bezrobotni zarejestrowani (B)  | 59 non-null    | float64   |
| 38 | Stopa bezrobocia rejestrowanego (%)  | 59 non-null    | float64   |
| 39 | Przewozy ładunków w transporcie kolejowym (w mln tonokilometrów)   | 59 non-null    | float64   |
| 40 | Przewozy ładunków w transporcie kolejowym (B)  | 59 non-null    | float64   |
| 41 | Przewozy ładunków w transporcie kolejowym (od początku roku do końca okresu, w mln tonokilometrów)               | 59 non-null    | float64   |
| 42 | Przewozy ładunków w transporcie kolejowym (od początku roku do końca okresu, A)                                  | 59 non-null    | float64   |
| 43 | Wskaźniki cen skupu pszenicy (bez ziarna siewnego) (B)   | 59 non-null    | float64   |
| 44 | Wskaźniki cen skupu żyta (bez ziarna siewnego) (B)   | 59 non-null    | float64   |
| 45 | Wskaźniki cen skupu bydła (bez cieląt) (B)   | 59 non-null    | float64   |
| 46 | Wskaźniki cen skupu trzody chlewnej (B)  | 59 non-null    | float64   |
| 47 | Wskaźniki cen skupu mleka (B)  | 59 non-null    | float64   |
| 48 | Relacje cen skupu żywności do cen żyta na targowiskach   | 59 non-null    | float64   |
| 49 | Wskaźniki cen produkcji sprzedanej w przemyśle (B)   | 59 non-null    | float64   |
| 50 | Wskaźniki cen w górnictwie i wydobywaniu (B)   | 59 non-null    | float64   |
| 51 | Wskaźniki cen w przetwórstwie przemysłowym (B)   | 59 non-null    | float64   |
| 52 | Wskaźniki cen w wytwarzaniu i zaopatrywaniu w energię elektryczną, gaz, parę wodną i gorącą wodę (B)             | 59 non-null    | float64   |
| 53 | Wskaźniki cen w dostawie wody; gospodarowaniu ściekami i odpadami; rekultywacji (B)                              | 59 non-null    | float64   |
| 54 | Wskaźniki cen produkcji budowlano-montażowej (B)   | 59 non-null    | float64   |
| 55 | Wskaźniki cen towarów i usług konsumpcyjnych (B)   | 59 non-null    | float64   |
| 56 | Wskaźniki cen – żywność i napoje bezalkoholowe (B)   | 59 non-null    | float64   |
| 57 | Wskaźniki cen – napoje alkoholowe i wyroby tytoniowe (B)   | 59 non-null    | float64   |
| 58 | Wskaźniki cen – odzież i obuwie (B)  | 59 non-null    | float64   |
| 59 | Wskaźniki cen – użytkowanie mieszkania lub domu i nośniki energii (B)  | 59 non-null    | float64   |
| 60 | Wskaźniki cen – wyposażenie mieszkania i prowadzenie gospodarstwa domowego (B)                                   | 59 non-null    | float64   |
| 61 | Wskaźniki cen – zdrowie (B)  | 59 non-null    | float64   |
| 62 | Wskaźniki cen – transport (B)  | 59 non-null    | float64   |
| 63 | Wskaźniki cen – łączność (B)   | 59 non-null    | float64   |
| 64 | Wskaźniki cen – rekreacja i kultura (B)  | 59 non-null    | float64   |
| 65 | Wskaźniki cen – edukacja (B)   | 59 non-null    | float64   |
| 66 | Wskaźniki cen transakcyjnych eksportu (A)  | 59 non-null    | float64   |
| 67 | Wskaźniki cen transakcyjnych importu (A)   | 59 non-null    | float64   |
| 68 | Terms of trade (A)   | 59 non-null    | float64   |
| 69 | Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (w zł)                          | 59 non-null    | float64   |
| 70 | Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (B)                             | 59 non-null    | float64   |
| 71 | Przeciętne miesięczne realne wynagrodzenie brutto w sektorze przedsiębiorstwa (B)                                | 59 non-null    | float64   |
| 72 | Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (w zł) | 59 non-null    | float64   |
| 73 | Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (B)    | 59 non-null    | float64   |
| 74 | Przeciętna miesięczna realna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (B)       | 59 non-null    | float64   |
| 75 | Przeciętna miesięczna nominalna emerytura i renta rolników indywidualnych brutto (w zł)                          | 59 non-null    | float64   |
| 76 | Przeciętna miesięczna nominalna emerytura i renta rolników indywidualnych brutto (B)                             | 59 non-null    | float64   |
| 77 | Przeciętna miesięczna realna emerytura i renta rolników indywidualnych brutto (B)                                | 59 non-null    | float64   |

**Rysunek 16:** Ostateczne dane na których będą wykonywane dalsze operacje.

Liczba kolumn - 77; liczba wierszy - 59

### 4.3. Feature selection

Aby móc kontynuować dalszą eksplorację danych, niezbędne było wyekstrahowanie najbardziej wartościowych oraz liczących się cech. Z racji dużej liczby kolumn, bo aż 77 niezbędne jest wykonanie PCA lub innej metody analizy głównych składowych. Niektóre cechy są ze sobą bardzo mocno skorelowane, co można było zaobserwować na przykład na [Rysunek 6](#), więc zmniejszenie ich liczby pomoże w zredukowaniu wymiarowości oraz złożoności. W celu zmniejszenia liczby cech oraz wybrania najważniejszych komponentów zostało zastosowane RFE (Recursive Feature Elimination) wraz z regresją liniową. Na początku zostało wybrane 5 cech. Liczba ta była liczbą startową z którą będzie robione więcej eksperymentów, tak aby otrzymać jak najlepsze wyniki w późniejszym etapie.

```
from sklearn.datasets import make_classification
from sklearn.feature_selection import RFE
from sklearn.linear_model import LinearRegression

# Example dataset for regression
df = pd.DataFrame(X)

# Estimator for feature selection in regression
model = LinearRegression()

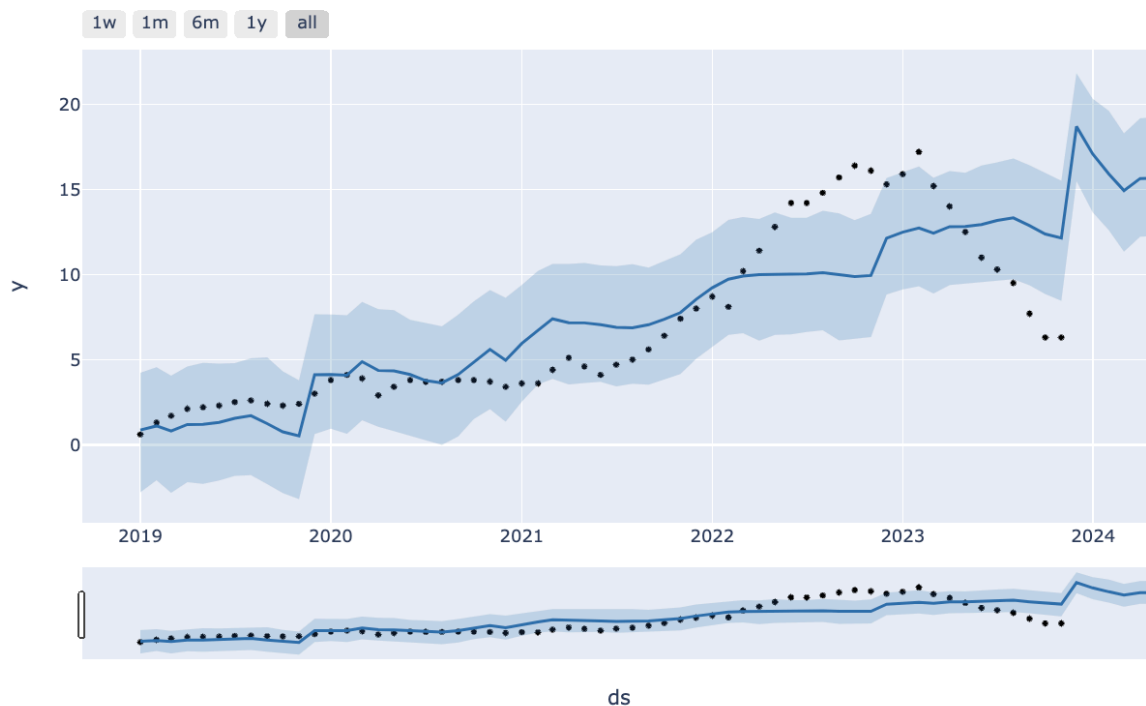
# RFE model
selector = RFE(model, n_features_to_select=5) # Select top 5 features
selector = selector.fit(df, y)

# Get the mask of selected features
selected_features = selector.support_

# Selected features DataFrame
selected_df = df.loc[:, selected_features]
```

**Rysunek 17:** Funkcja w Pythonie do selekcji 5 najlepszych cech przy użyciu modelu RFE wraz z regresją liniową

Testowo również został zastosowany model Prophet z biblioteki prophet do przewidywania szeregów czasowych. Jednak jego wyniki nie są satysfakcjonujące, co można zaobserwować na [Rysunek 18](#).



**Rysunek 18:** Przewidywanie szeregu czasowego dla inflacji HICP przy użyciu modelu Prophet oraz top 5 komponentów według analizy RFE

Po dogłębnej analizie, oraz zastanowieniu się nad problemem, wysunięto wniosek, iż zabrakło wcześniejszej normalizacji danych. W czasie analizy problemu, udało znaleźć się interesującą bibliotekę do przewidywania oraz analizy szeregów czasowych. Bibliotek [Darts](#) oferuje wiele modeli jak i możliwości przetwarzania oraz analizy szeregów czasowych. Znaczna część czasu, została poświęcona na zapoznanie się z tą biblioteką jak również eksperymentowanie. Wydaje się, że jest ona świetnym i bardzo przydatnym narzędziem w analizie inflacji, która jest właśnie szeregiem czasowym.

## 5. Przewidywanie inflacji

### 5.1. Przewidywanie inflacji za pomocą modeli z biblioteki Darts na jednej kolumni **HICP**

Zbiór danych składający się z 59 wierszy oraz 79 kolumn został podzielony i wyciągnięte zostały z niego tylko 2 kolumny:

- month
- HICP

Kolumna month jest nam potrzebna do stworzenia struktury TimeSeries w Darts.

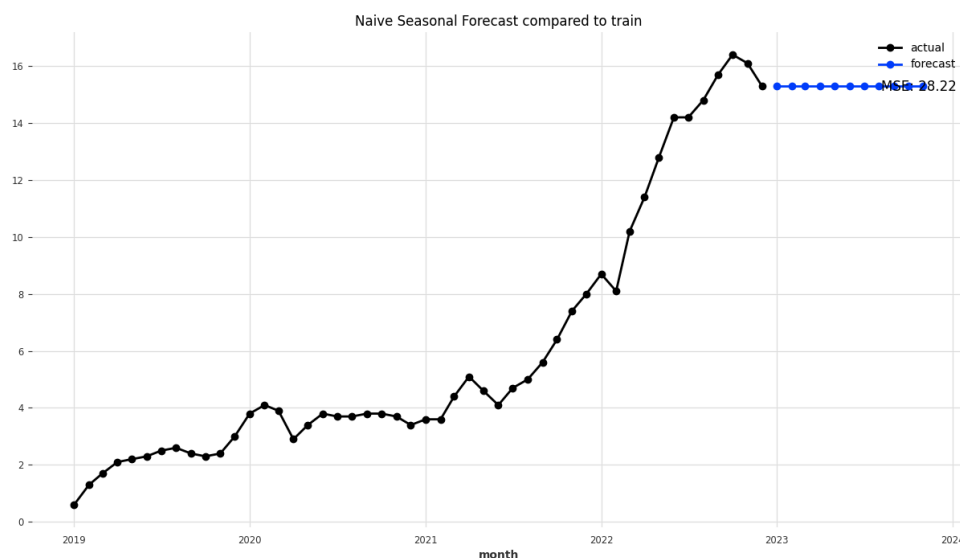
Zbiór został następnie podzielony na treningowy i testowy. Zbiór treningowy zawiera 48 miesięcy z danymi HICP a zbiór walidacyjny ostatnie 11 miesięcy.

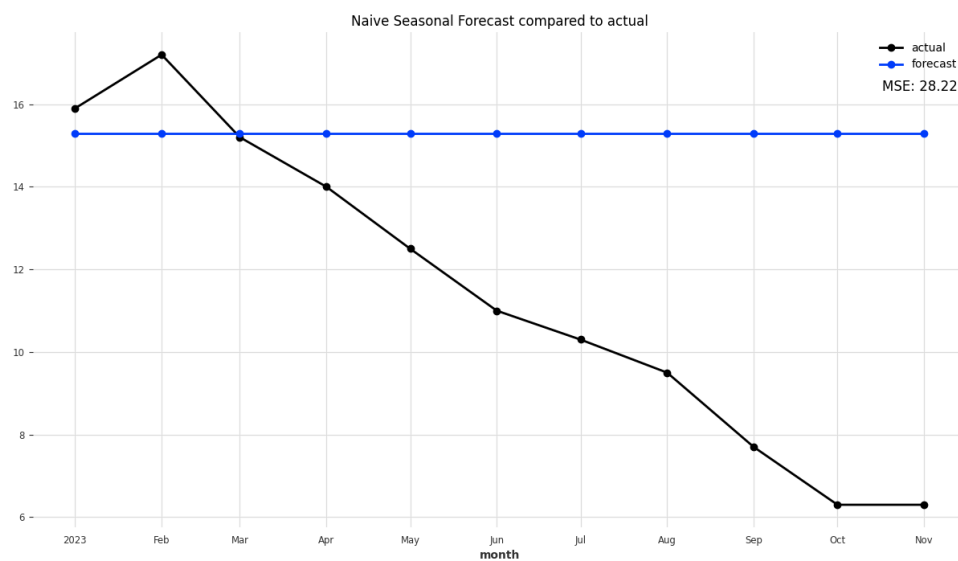
#### 5.1.1. Naive Seasonal

Model Naive Seasonal jest to naiwny model, który powtarza sezonowe dane. Niestety dane o inflacji nie są danymi sezonowymi i model dla 11 miesięcy powtarza dane z ostatniego zanotowanego miesiąca w zbiorze treningowym. Będzie to dla nas baseline, do którego będziemy porównywać inne modele.

Predykcja: [15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3]

MSE: 28.22





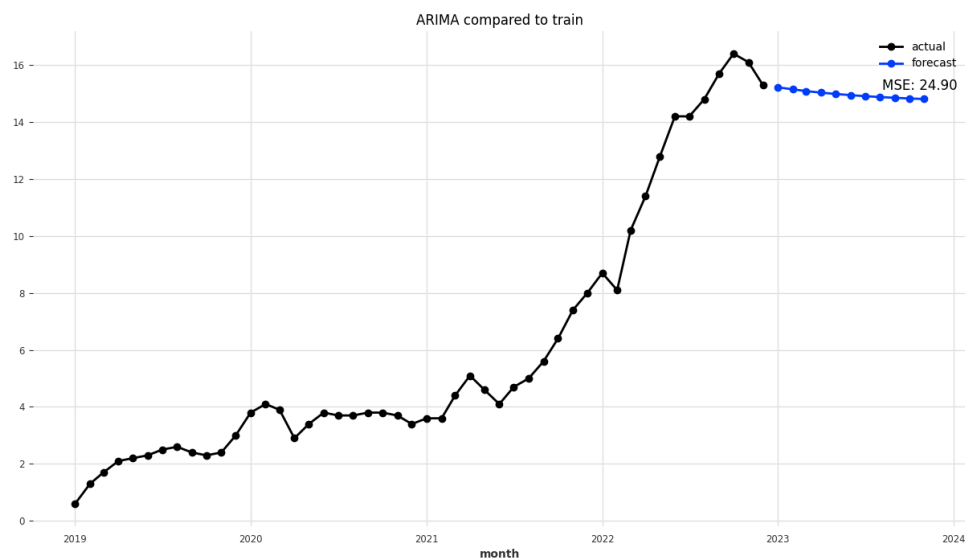
**Rysunek 19:** Przewidywanie inflacji dla Naive Seasonal dla 11 miesięcy

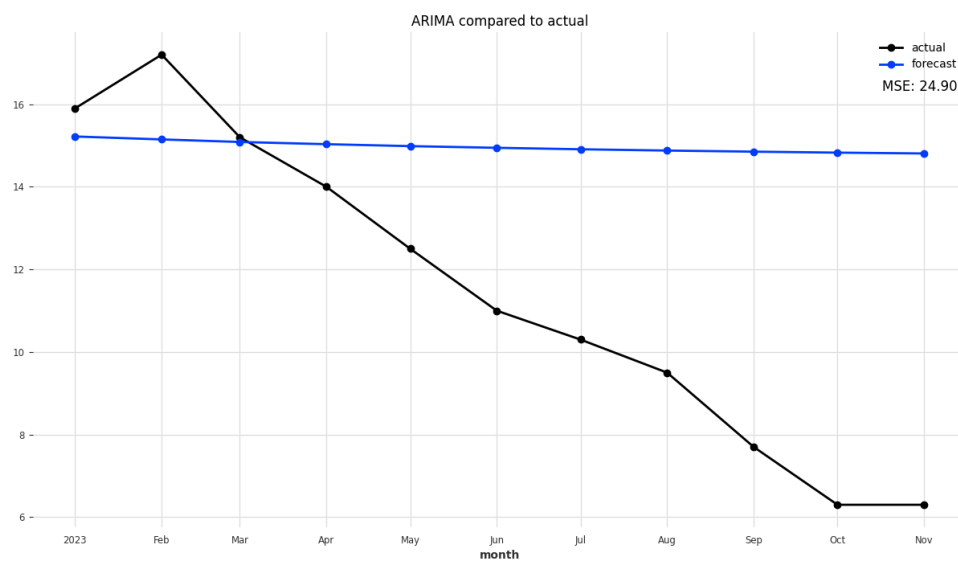
### 5.1.2. ARIMA

Kolejnym testowanym modelem była ARIMA, która zwróciła następujące wyniki:

Predykcja: [15.2, 15.1, 15.1, 15.0, 15.0, 14.9, 14.9, 14.9, 14.9, 14.8, 14.8]

MSE: 24.90



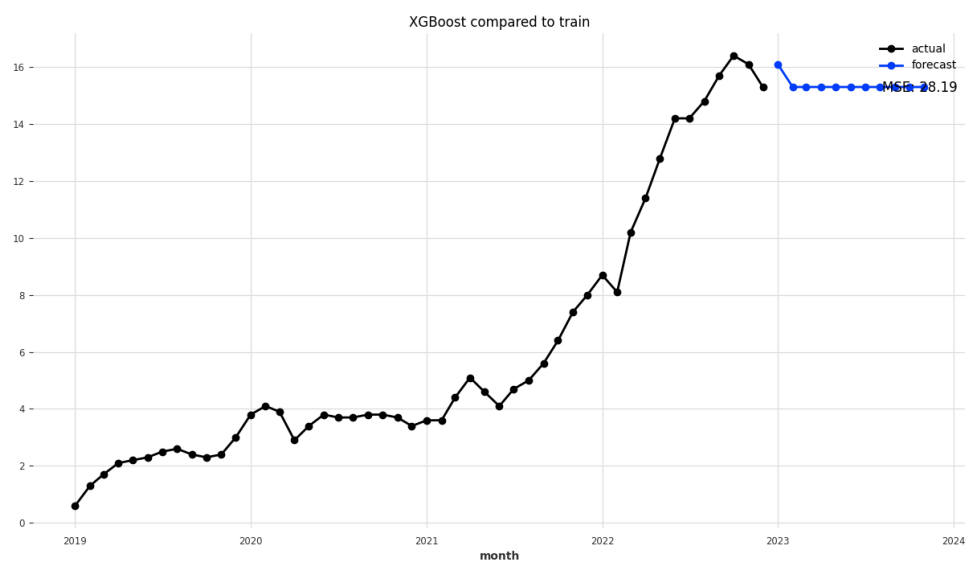


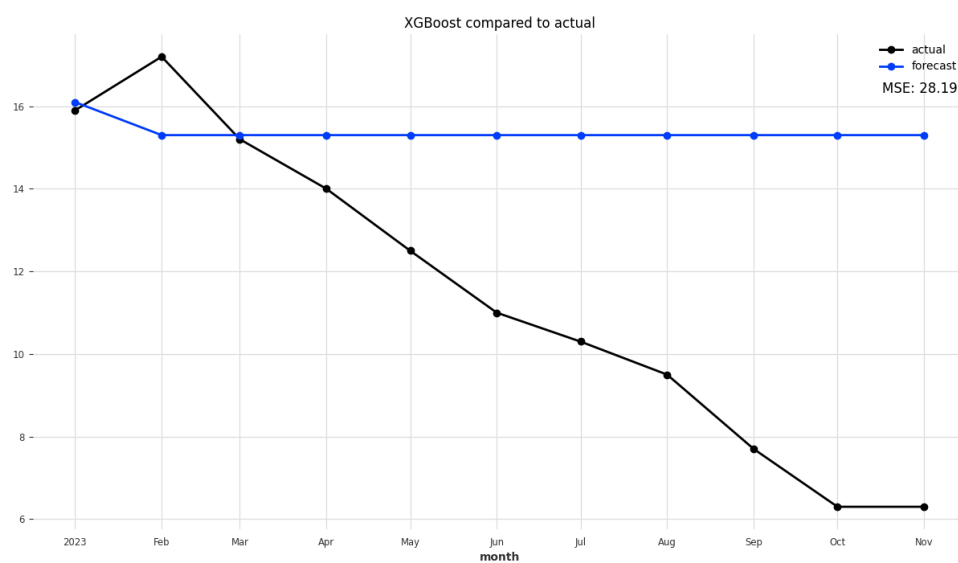
**Rysunek 20:** Przewidywanie inflacji dla ARIMA dla 11 miesięcy

### 5.1.3. XGBoost

Predykcja: [16.1, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3]

MSE: 28.19





**Rysunek 21:** Przewidywanie inflacji dla XGBoost dla 11 miesięcy

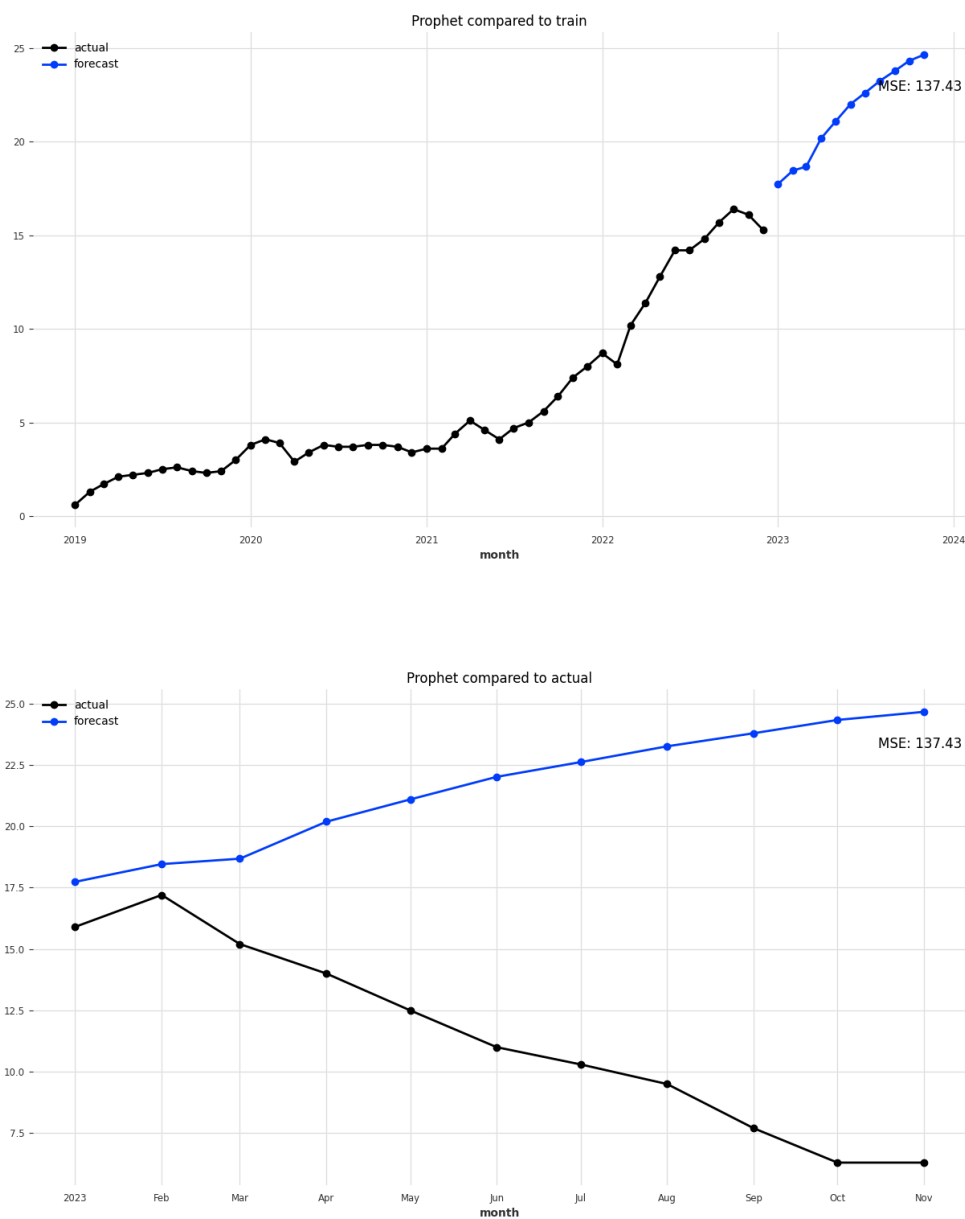


#### 5.1.4. Prophet

Prophet wypadł najgorzej ze wszystkich modeli.

Przybliżenie: [17.7, 18.4, 18.7, 20.2, 21.1, 22.0, 22.6, 23.3, 23.8, 24.3, 24.7]

MSE: 137.43



**Rysunek 22:** Przewidywanie inflacji dla Prophet dla 4 miesięcy

#### 5.1.5. Wnioski

Przewidywanie inflacji w oparciu o same wartości HICP jest bardzo ciężkie. Modele nie uwzględniają innych danych ekonomicznych oraz gospodarczych, które w znacznym stopniu wpływają na dane o inflacji. Poniższa tabela zawiera zestawienie wszystkich 4 modeli, ich predykcje inflacji na 11 kolejnych miesięcy oraz ich Mean Squared Error.

Dla porównania, dane prawdziwe, to:

| Model          | Predykcja  | MSE    |
|----------------|--|--------|
| Naive Seasonal | [15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3] | 28.22  |
| ARIMA          | [15.2, 15.1, 15.1, 15.0, 15.0, 14.9, 14.9, 14.9, 14.9, 14.8] | 24.90  |
| XGBoost        | [16.1, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3, 15.3] | 28.19  |
| Prophet        | [17.7, 18.4, 18.7, 20.2, 21.1, 22.0, 22.6, 23.3, 23.8, 24.3] | 137.43 |

**Tabela 1:** Porównanie modeli dla predykcji inflacji na podstawie danych HICP

Możemy zauważyć, że żaden z modeli nie poradził sobie szczególnie dobrze. Większość z nich na przestrzeni 11 miesięcy często powtarza te same wyniki. Jest to może kwestia błędnych implementacji lub problemu predykcji inflacji w oparciu tylko o jedną tabelę. W dalszej części podjęte będą próby predykcji inflacji o bardziej standardowe modele oraz modele z innych bibliotek. Jednak, z wyników tych możemy zauważyć, że najlepszym modelem z najniższym MSE okazał się ARIMA. Następnie XGBoost i Naive Seasonal jednak nie jest to dobry model, ponieważ powtarza on kolejne wartości z ostatniego zapamiętanego miesiąca ze zbioru treningowe, co na dłuższą metę da oczywiście słabe wyniki. Wniosek jaki na pewno się wysnuwa to brak sezonowości danych inflacyjnych, przez co wiele modeli ma problem z ich poprawną predykcją. Kolejną kwestią jest uwzględnienie tylko jednej kolumny z wartościami HICP bez uwzględnienia innych bardzo istotnych kolumn z danymi ekonomicznymi czy wydarzeniami historycznymi. Kolejnym krokiem jest próba predykcji inflacji ale już z uwzględnieniem znacznie większej liczby kolumn, jednak nie będą to wszystkie kolumny a wybrane w procesie Feature Selection.

## 5.2. Przewidywanie inflacji na kilku kolumnach wybranych w procesie Feature Selection

### 5.2.1. Feature Selection

Przed dokonaniem predykcji na kilku wybranych kolumnach dokonane zostało Feature Selection przy pomocy 3 metody:

- PCA
- RFE
- Feature Importance przy pomocy RandomForest Regression

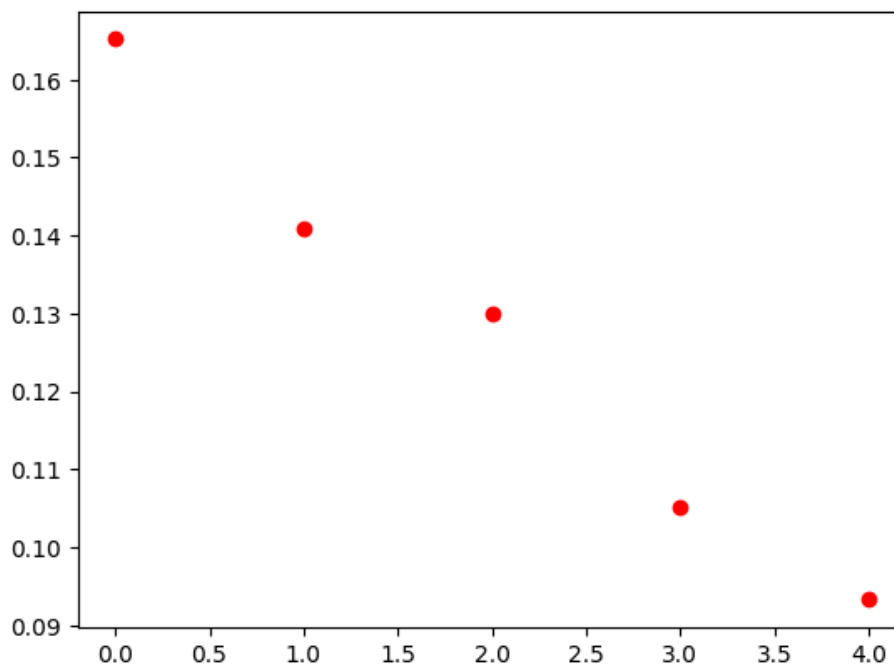
Przed dokonaniem Feature Selection na wszystkich kolumnach dokonana została normalizacja danych przy pomocy Standard Scalera. Przy pomocy każdej z metod chcieliśmy otrzymać 5 najważniejszych cech.

- PCA

PCA zwróciło nam następujące 5 najważniejszych cech:

- Import towarów (ceny bieżące w mln zł)
- Produkcja - dobra konsumpcyjne nietrwałe (B)
- Wydatki budżetu państwa (w mln zł)
- Koniunktura - działalność finansowa i ubezpieczeniowa
- Wskaźniki cen skupu mleka (B)

Tych 5 cech pokrywa około 65% wariancji naszego zbioru danych.



**Rysunek 23:** Explained variance ratio dla 5 cech wybranych przez PCA

- RFE

W tym przypadku, metoda Recursive Feature Elimination zwróciła następujące cechy:

- Wydatki budżetu państwa (w mln zł)
- Bieżący wskaźnik ufności konsumenckiej (BWUK)
- Koniunktura - przetwórstwo przemysłowe
- Produkcja sprzedana przemysłu ogółem (ceny stałe, B)
- Stopa bezrobocia rejestrowanego (%)

- Feature Importance

Cechy wybrane przez metodę Feature Importance z użyciem Random Forest Regression:

- Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (w zł)
- Bieżący wskaźnik ufności konsumenckiej (BWUK)
- Przeciętne zatrudnienie w sektorze przedsiębiorstwa (w w tys.)
- Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (w zł)
- Import towarów (ceny bieżące w mln zł)

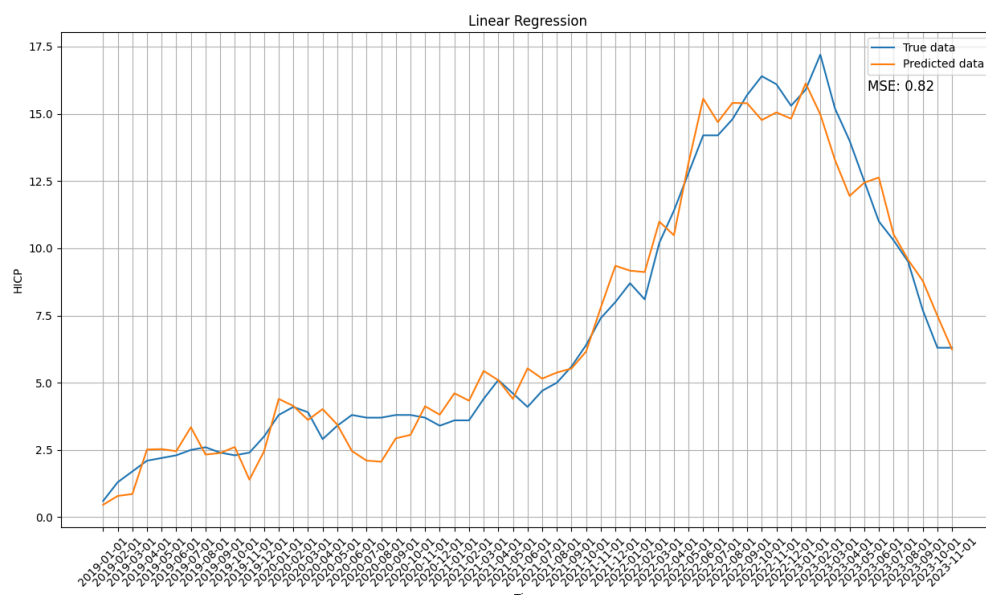
Wyniki z tych 3 metod zostały następnie dodane do jednej listy i przy pomocy struktury danych w Pythonie set, zostało wybranych 12 niepowtarzających się i najbardziej znaczących cech według wyżej wymienionych 3 metod, a są to:

- Produkcja sprzedana przemysłu ogółem (ceny stałe, B)
- Produkcja - dobra konsumpcyjne nietrwałe (B)
- Przeciętne zatrudnienie w sektorze przedsiębiorstwa (w w tys.)
- Wydatki budżetu państwa (w mln zł)
- Bieżący wskaźnik ufności konsumenckiej (BWUK)
- Koniunktura - przetwórstwo przemysłowe
- Import towarów (ceny bieżące w mln zł)
- Stopa bezrobocia rejestrowanego (%)
- Przeciętna miesięczna nominalna emerytura i renta brutto z pozarolniczego systemu ubezpieczeń społecznych (w zł)
- Koniunktura - działalność finansowa i ubezpieczeniowa
- Wskaźniki cen skupu mleka (B)
- Przeciętne miesięczne nominalne wynagrodzenie brutto w sektorze przedsiębiorstwa (w zł)

Na tych kolumnach będą teraz dokonywane predykcje inflacji dla kilku wybranych modeli.

### 5.2.2. Linear Regression

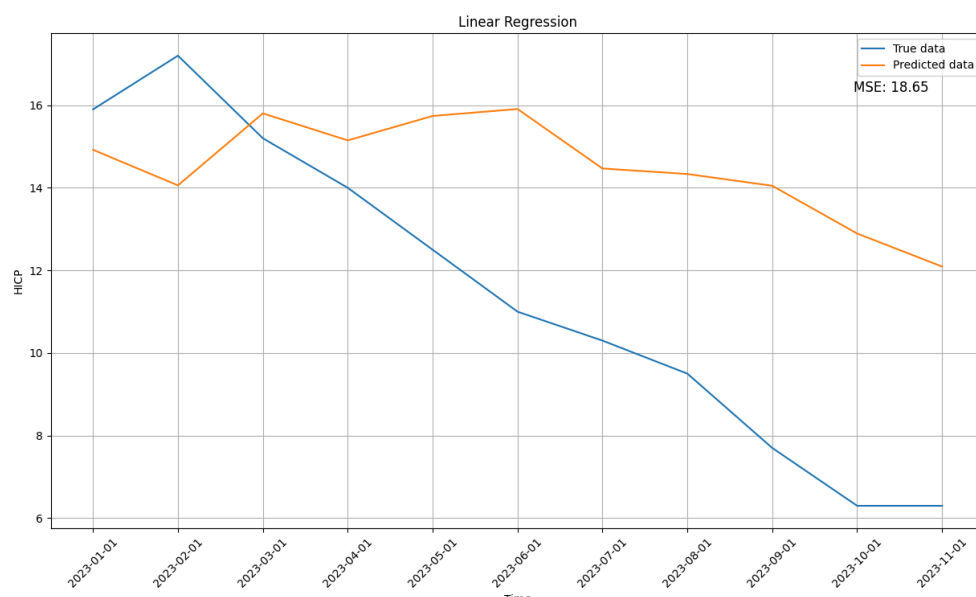
Pierwszym z modeli jest prosta regresja liniowa, która jest naszym baselinem w tym przypadku. W dalszych próbach będziemy chcieli znaleźć modele, które wypadają lepiej.



**Rysunek 24:** Predykcja inflacji dla zbioru treningowego przy użyciu regresji liniowej

Co ciekawe regresja liniowa przy pomocy Feature Selection przynosi całkiem dobre rezultaty porównując to z prawdziwymi wartościami na wykresie oraz wartości MSE.

Tak jak w przypadku modeli z biblioteki Darts, będziemy chcieli przewidzieć inflację dla ostatnich 11 miesięcy.

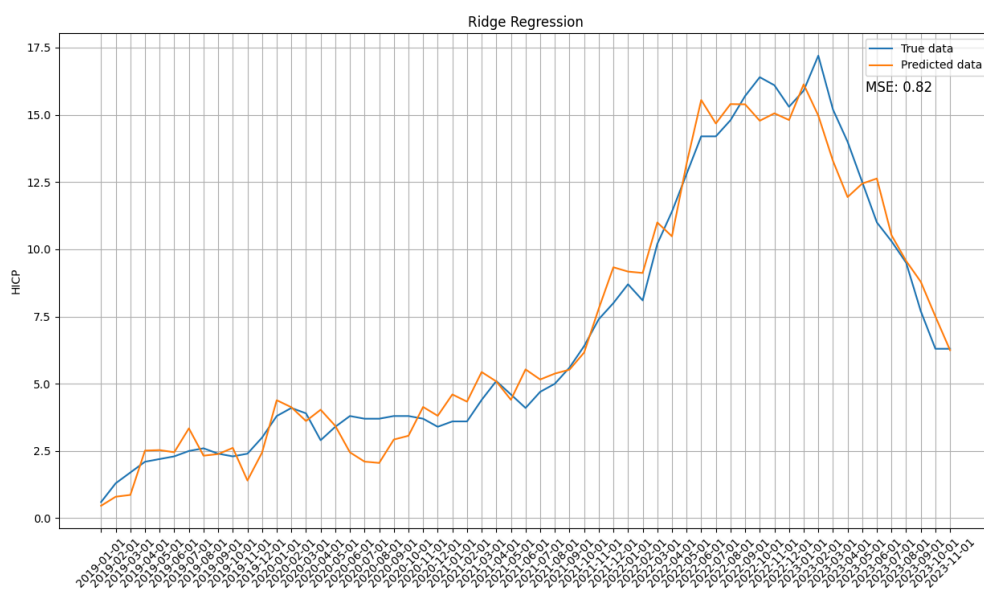


**Rysunek 25:** Predykcja inflacji dla 11 ostatnich miesięcy dla regresji liniowej  
Predykcja: [14.9, 14.1, 15.8, 15.1, 15.7, 15.9, 14.5, 14.3, 14.0, 12.9, 12.1]  
MSE: 18.65

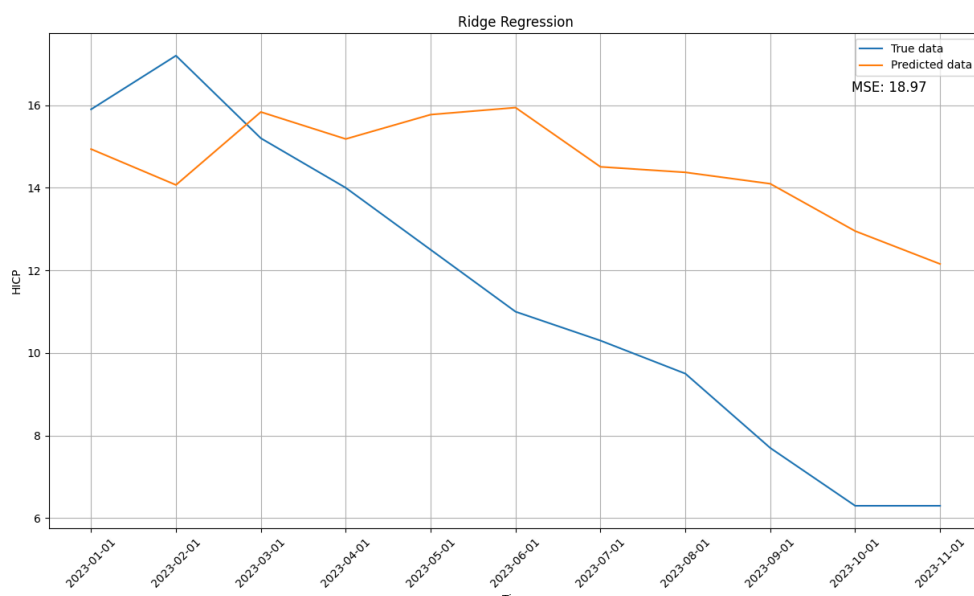
Regresja liniowa osiągnęła całkiem zadowalającą wartość MSE. Już teraz możemy zaobserwować, że MSE ma niższe niż którykolwiek model z wcześniej testowanych modeli z biblioteki Darts na pojedynczej kolumnie HICP. W tym przypadku uwzględniamy 12 różnych kolumn, co daje modelowi większą liczbę danych na których może się lepiej uczyć.

### 5.2.3. Ridge Regression

W przypadku Ridge Regression, wartość MSE oraz predykcje są bardzo podobne do powyższej regresji liniowej.



**Rysunek 26:** Predykcja inflacji dla zbioru treningowego przy użyciu regresji grzbietowej

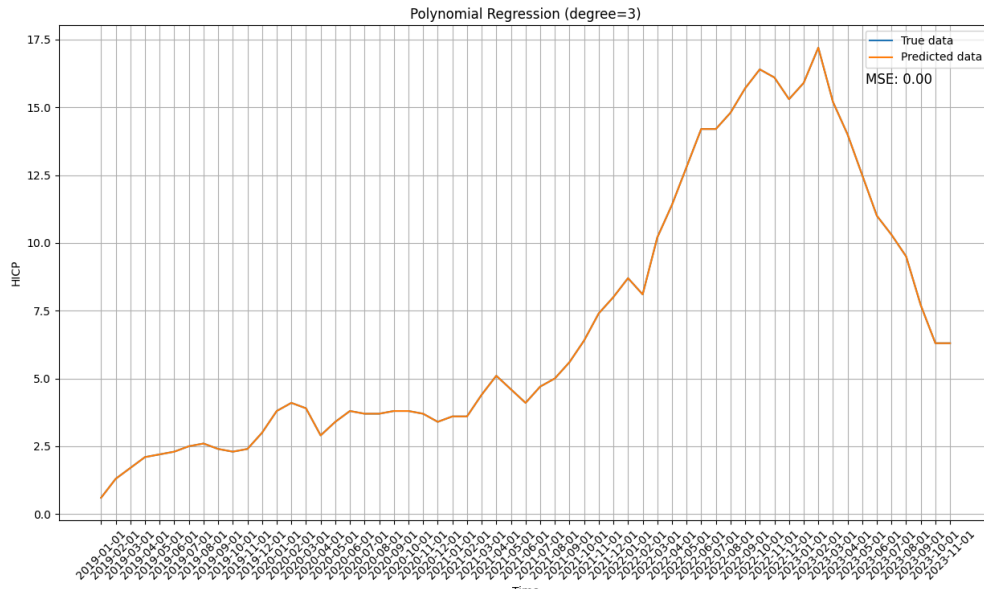


**Rysunek 27:** Predykcja inflacji dla 11 ostatnich miesięcy dla regresji grzbietowej

Można zauważyć, że regresja grzbietowa, czyli regresja liniowa korzystająca z regularyzacji L2 osiąga bardzo podobne wyniki co zwykła regresja liniowa.

#### 5.2.4. Polynomial Regression degree 3

Trzecim testowanym modelem była regresja wielomianowa stopnia 3. Możemy zauważyć, że test modelu na danych treningowych wykazuje się prawie 0 MSE lossem. Oznacza to, że model idealnie dopasował się do danych i prawdopodobnie powstał tutaj overfitting.

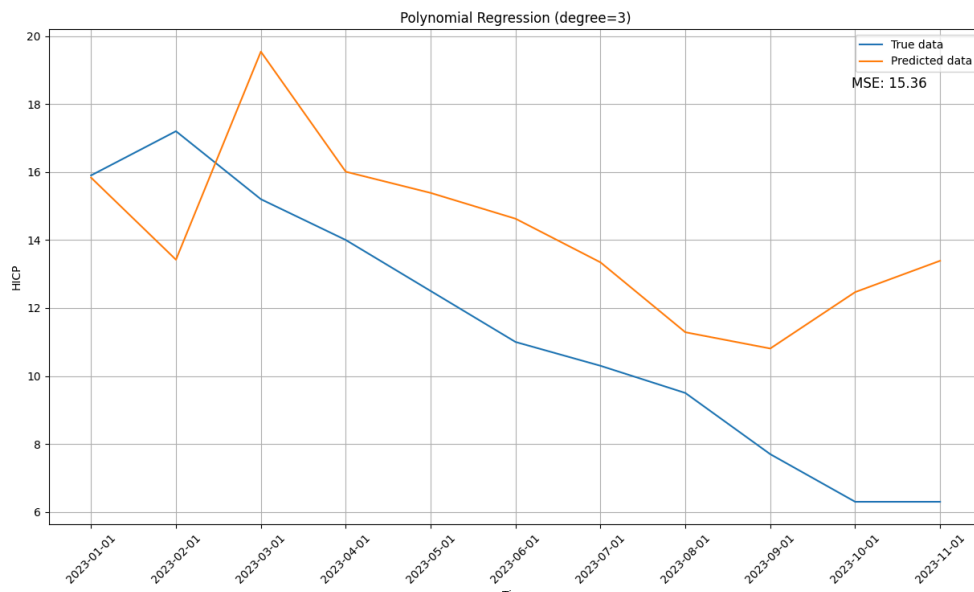


**Rysunek 28:** Predykcja inflacji dla zbioru treningowego przy użyciu regresji wielomianem 3go stopnia

W celu sprawdzenia, jak model radzi sobie na danych, których nie widział, stworzony został nowy model wytrenowany na danych treningowych a następnie dokonana została predykcja na danych których nie widział.

Predykcja: [15.8, 13.4, 19.5, 16.0, 15.3, 14.6, 13.3, 11.3, 10.8, 12.5, 13.4]

MSE: 15.36



**Rysunek 29:** Predykcja inflacji dla 11 ostatnich miesięcy dla regresji wielomianem 3go stopnia

Możemy zauważyć, że ten model ma dotychczasowo najniższe MSE ze wszystkich modeli. Jest on co prawda dalej wysoki i predykcje nie są idealne, ale próby nowych modeli idą w dobrą stronę.

#### 5.2.5. Minutes po spotkaniu 10.05

- Dominik: inna metryka - RMSE
- Karol: różnicowanie i transformacja danych, opis regresji wielorakiej, regresja liniowa w Statistice
- Szymon:
  - napisać prosty model LSTM, który dokonuje predykcji na pojedynczej kolumnie HICP
  - stworzyć nowy model LSTM z bardziej skomplikowaną architekturą, który dokonuje predykcji na pojedynczej kolumnie HICP
  - wykorzystać model z poprawioną architekturą (jeżeli okaże się lepszy od wcześniejszego) do predykcji na danych złożonych z kilku kolumn wybranych przy pomocy feature selection
  - zmienić funkcję losu w przyszłych modelach z MSE na RMSE

#### 5.2.6. LSTM

**Model LSTM (Long Short-Term Memory)** to rodzaj rekurencyjnej sieci neuronowej (RNN), która została zaprojektowana do przetwarzania i prognozowania sekwencji danych, zachowując informacje o długoterminowych zależnościach w danych.

Predykcja z modelem LSTM została podzielona na 2 etapy:

1. Predykcja na pojedynczej kolumnie HICP
2. Predykcja na kolumnach wybranych przy pomocy Feature Selection

W każdej z tych 2 metod, pierwszą wykonaną rzeczą było ustawienie miesięcy jako indeks w DataFrame oraz normalizacja danych przy pomocy Standard Scaler'a z biblioteki sklearn. Model LSTM był samodzielnie zaimplementowany przy pomocy biblioteki tensorflow.

1. Predykcja na pojedynczej kolumnie HICP

Pierwszy wypróbowany model był dość prosty, bo jego architektura wyglądała następująco:

```
model = Sequential([
    Input(shape=(seq_length, 1)),
    LSTM(units=50, activation='relu'),
    Dense(units=1)
])
```

**Rysunek 30:** Architektura pierwszego modelu LSTM

Do kompilacji modelu został użyty optimizer Adam oraz loss RMSE. Następnie model był trenowany przez 200 epok na wcześniej utworzonych sekwencjach danych.



```

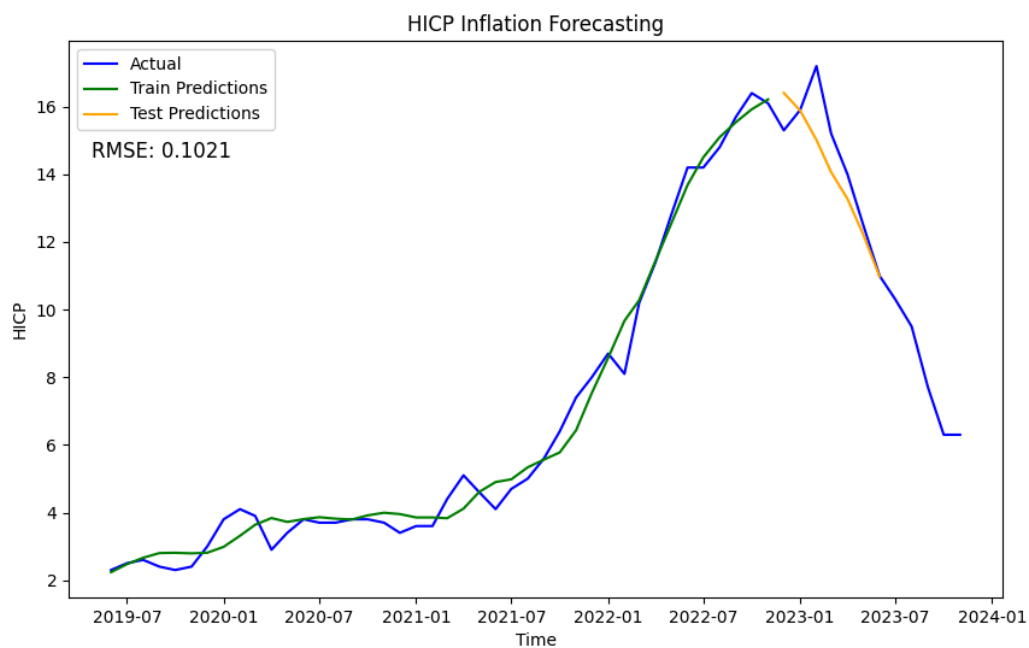
from keras import backend as K

def RMSE(y_true, y_pred):
    return K.sqrt(K.mean(K.square(y_pred - y_true)))

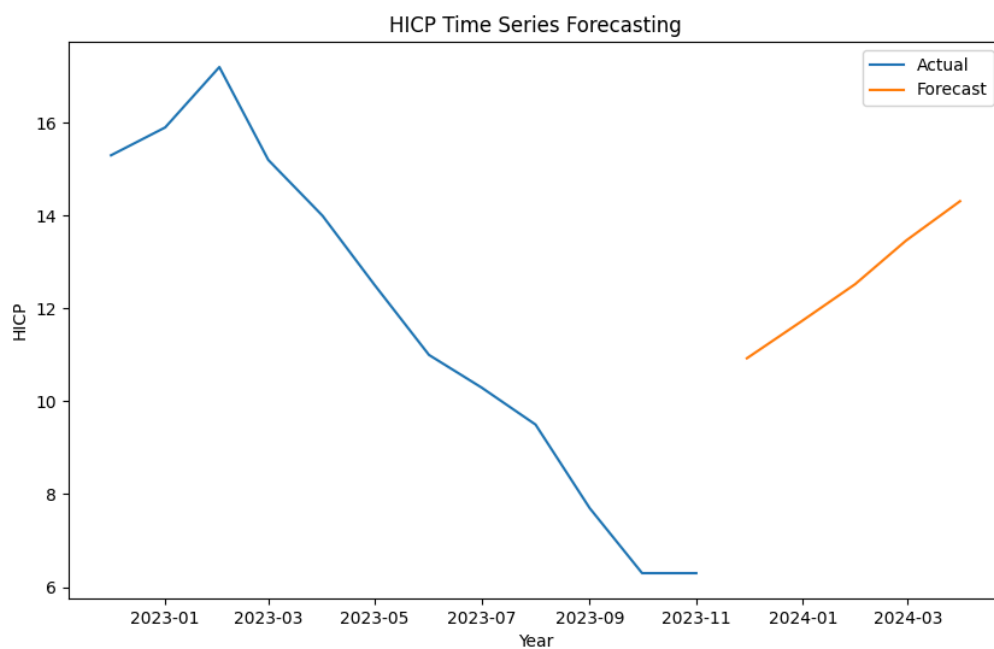
model.compile(optimizer='adam', loss=RMSE)
model.fit(X_train, y_train, epochs=200, batch_size=32, verbose=1)

```

**Rysunek 31:** Kompilacja modelu LSTM oraz trening  
Błąd RMSE po ostatniej epoce wyniósł 0.1021.



**Rysunek 31:** Predykcja modelu LSTM na danych treningowych oraz testowych  
Model LSTM przewidział następujące dane inflacyjne na kolejne 5 miesięcy:  
Predykcja: [10.9, 11.7, 12.5, 13.5, 14.3]



**Rysunek 32:** Predykcja modelu LSTM dla 5 kolejnych miesięcy

Porównując dane modelu LSTM które przewidział na 5 kolejnych miesięcy czyli 12.2023 - 04.2024 i po ogłoszonych wynikach przez NBP wiemy, że model się myli. Przewiduje on wzrost inflacji, kiedy opublikowane już dane pokazały, że inflacja spadła.

### Poprawy:

Aby poprawić predykcje modelu możemy spróbować następujących czynności:

- zwiększyć liczbę warstw modelu LSTM, czyli bardziej go skomplikować
- przeprowadzić trening modelu oraz jego predykcje na większej liczbie kolumn (features)

Pierwszą poprawą jaką spróbujemy zrobić to zwiększenie liczby warstw modelu LSTM.

Nowa architektura modelu wygląda następująco:

| Layer (type)      | Output Shape  | Param # |
|-------------------|---------------|---------|
| lstm (LSTM)       | (None, 2, 50) | 12,800  |
| lstm_1 (LSTM)     | (None, 50)    | 20,200  |
| dropout (Dropout) | (None, 50)    | 0       |
| dense (Dense)     | (None, 1)     | 51      |

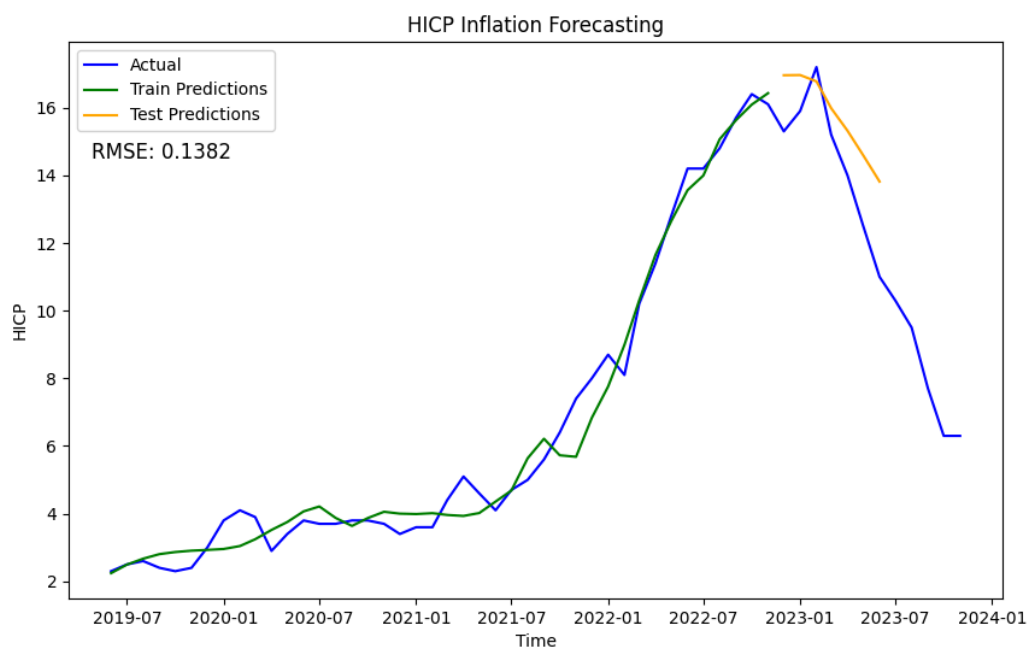
**Total params:** 33,051 (129.11 KB)

**Trainable params:** 33,051 (129.11 KB)

**Non-trainable params:** 0 (0.00 B)

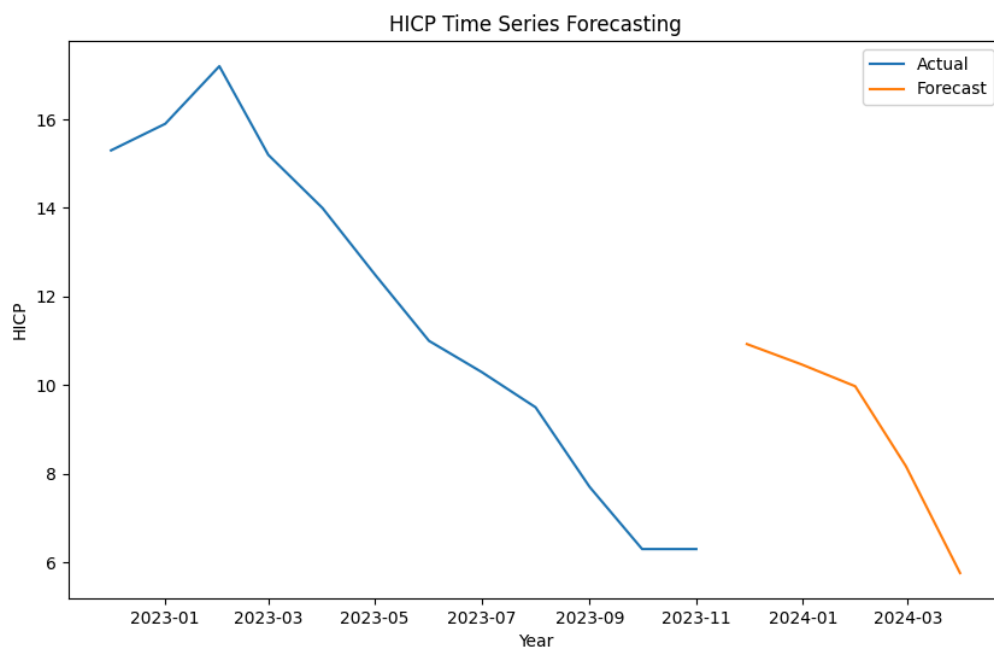
**Rysunek 33:** Nowa architektura modelu LSTM

Kompilacja oraz trening modelu wygląda tak samo jak wcześniej, używamy optimizera Adam oraz funkcji losu RMSE samodzielnie zaimplementowanej. Model jest znowu trenowany na 200 epokach. Po 200 epoche model osiągnął gorszy loss niż wcześniej, bo wynosi on: 0.1382. Nie oznacza to jednak od razu, że model radzi sobie gorzej, mała liczba warstw w poprzednim modelu mogła spowodować overfitting. Zwiększając liczbę warstw w nowym modelu zwiększamy jego złożoność co zapobiega przeuczeniu modelu. Sprawdźmy więc jego wyniki i predykcje dla danych treningowych, testowych oraz dla predykcji na 5 miesięcy do przodu.



**Rysunek 34:** Predykcja rozszerzonego modelu LSTM na danych treningowych oraz testowych

Predykcja: [10.9, 10.5, 9.9, 8.2, 5.8]



**Rysunek 35:** Predykcja rozszerzonego modelu LSTM dla 5 kolejnych miesięcy

Wniosek: Możemy zauważyć, że model LSTM z nową architekturą sprawuje się dużo lepiej w przypadku predykcji inflacji, dla przyszłych danych. Pokazuje on spadek inflacji co rzeczywiście od grudnia 2023 do kwietnia 2024 zgadza się z dotychczasowymi danymi. Model ten daje całkiem satysfakcjonujące wyniki, jednak nie są one idealne. Sprawdźmy jeszcze jak będzie zachowywał się model LSTM z bardziej rozbudowaną architekturą

podczas predykcji danych inflacyjnych ale dla większej liczby kolumn wybranych podczas Feature Selection.

Architektura modelu została lekko zmieniona i wygląda teraz następująco:

| Layer (type)      | Output Shape  | Param # |
|-------------------|---------------|---------|
| lstm (LSTM)       | (None, 1, 50) | 12,800  |
| lstm_1 (LSTM)     | (None, 50)    | 20,200  |
| dropout (Dropout) | (None, 50)    | 0       |
| dense (Dense)     | (None, 1)     | 51      |

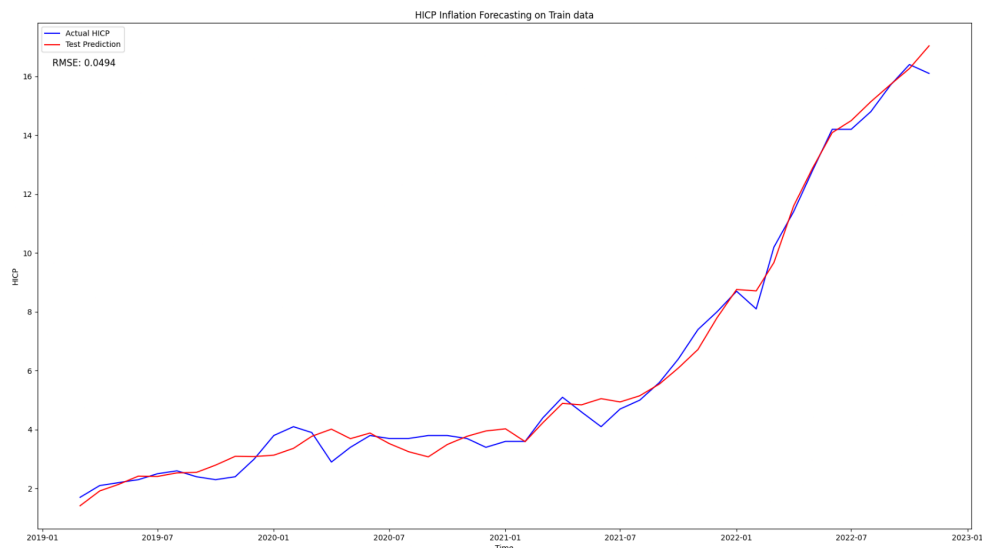
**Total params: 33,051 (129.11 KB)**

**Trainable params: 33,051 (129.11 KB)**

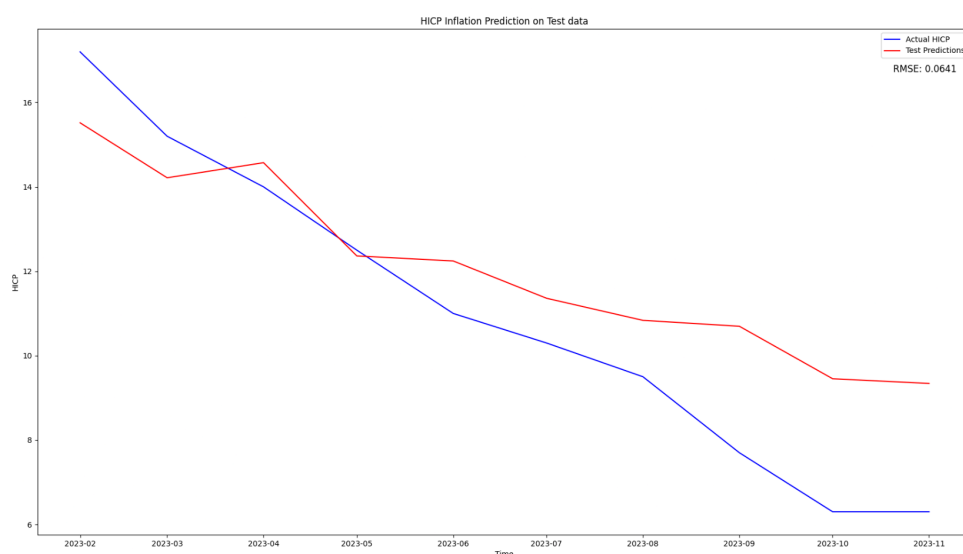
**Non-trainable params: 0 (0.00 B)**

**Rysunek 36:** Nowa architektura modelu LSTM

Zastosowana funkcja losu to znowu RSE, możemy zauważyć, że jej wartość jest niższa od wcześniejszych bo wynosi 0.0494. Może to jednak oznaczać overfitting modelu, dlatego należy sprawdzić jakie model zwraca wartości predykcji dla danych testowych oraz 6 kolejnych miesięcy.

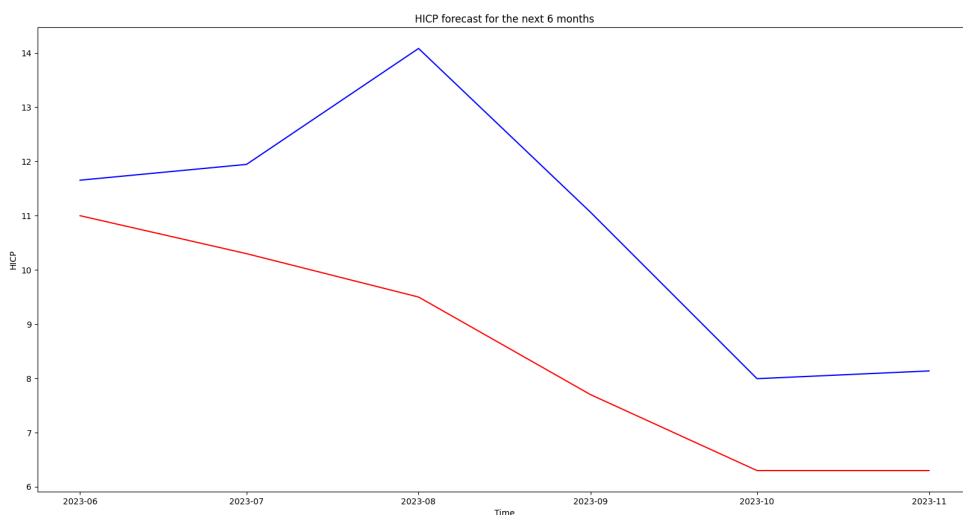


**Rysunek 37:** Predykcja modelu LSTM na danych treningowych dla kolumn wybranych przy pomocy Feature Selection



Predykcja: [15.5, 14.2, 14.6, 12.4, 12.2, 11.4, 10.8, 10.7, 9.5, 9.3]

**Rysunek 37:** Predykcja modelu LSTM dla zbioru testowego dla kolumn wybranych przy pomocy Feature Selection



**Rysunek 38:** Predykcja modelu LSTM dla 6 kolejnych miesięcy

Predykcja: [11.6, 11.9, 14.1, 11.0, 7.9, 8.1]

Wniosek: Model LSTM z nową architekturą zachowuje się lepiej od dwóch wcześniejszych modeli. Jego predykcja dla danych treningowych oraz testowych jest satysfakcjonująca. Również wartości inflacji dla 6 miesięcy w przód począwszy od listopada 2023 dają dobre wyniki, a inflacja w tym czasie spadała, co można było rzadko zaobserwować na danych od 2018 roku gdzie inflacja głównie rosła.

## 6. Prognoza inflacji w Statistica oraz RStudio

### 6.1. Regresja wieloraka

Wynik regresji wielorakiej to zestaw wartości numerycznych, które wskazują, jak dobrze model regresji przewiduje wartości zmiennych zależnych na podstawie zmiennych niezależnych.

Interpretacja wartości współczynników:

**Wielokrotność (Multiple): R** = .999875003 - silny związek między zmienną zależną HICP a zmiennymi niezależnymi w modelu.

**R<sup>2</sup>**: .99975007 - model wyjaśnia prawie całą zmienność zmiennej zależnej. Innymi słowy, model jest bardzo dobrze dopasowany do danych.

**Poprawione R<sup>2</sup> (Adjusted R<sup>2</sup>)**: .99985043 - dodanie dodatkowych zmiennych do modelu poprawiło jego dopasowanie.

**F-statystyka**: 833,3746 - model jest istotny statystycznie. Innymi słowy, istnieje bardzo małe prawdopodobieństwo, że te wyniki są przypadkowe.

**p-wartość (p-value)**: .000000 - model jest istotny statystycznie.

**Błąd standardowy estymacji**: ,186877495 - model jest precyzyjny w swoich prognozach.

```
Wyniki regresji wielorakiej (Krok 48)

Zmn. zależ.HICP          Wielor. R = ,99987503      F = 833,3746
                        R^2= ,99975007      df = 48,10
Liczba przyp. 59        Popraw. R^2= ,99855043      p = ,000000
                        Błąd standardowy estymacji: ,186877495
Wyr. wolny 68,370323878 Błąd std.: 70,82396 t( 10) = ,96536 p = ,3571

12  48  16  22  44  39  19  58  30  69  14  47  40  65  17
18  15  34  36  61  55  74  28  3  25  23  41  62  51  26
4   8  53  31  32  60  35  59  38  37  24  52  57  21  43
49  78  77

(istotne b* są podświetlone na czerwono)
```



Rysunek 38: wyniki regresji wielorakiej

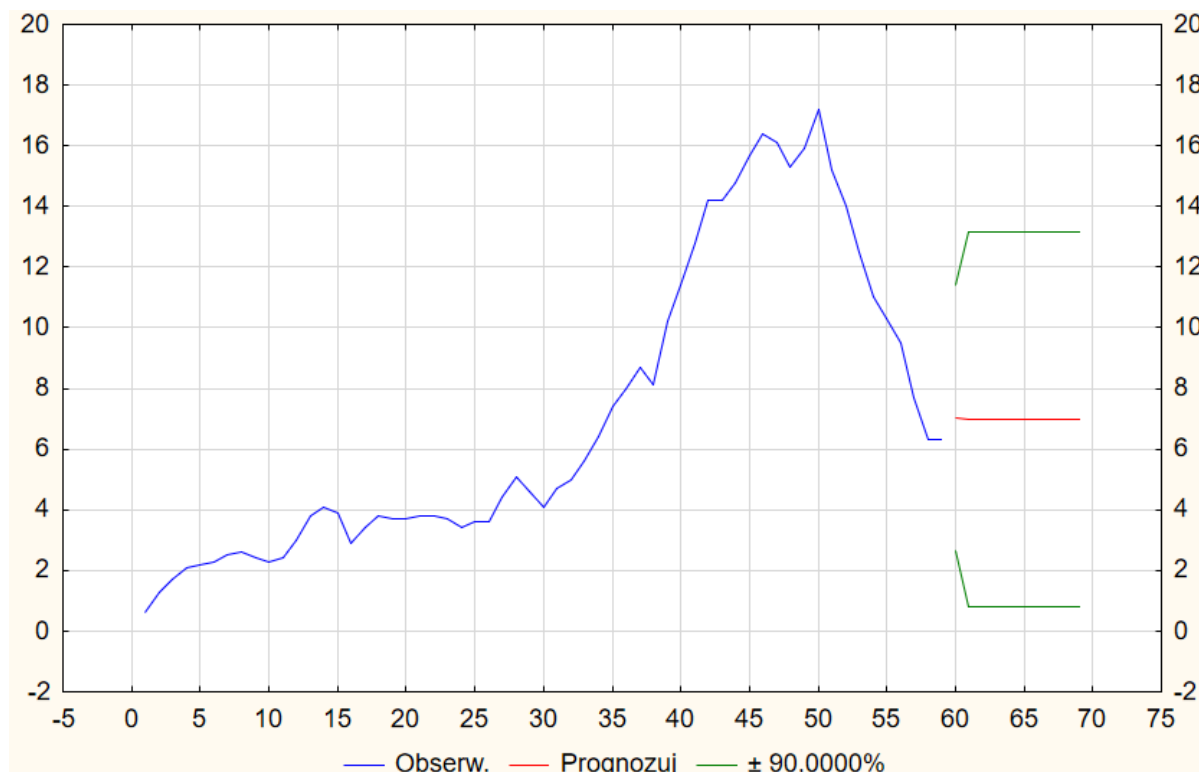
Wniosek: Istotny wpływ na inflację mają następujące wskaźniki makroekonomiczne:

- Wskaźniki cen produkcji sprzedanej w przemyśle
- Przewozy ładunków w transporcie kolejowym
- Wskaźniki cen - napoje alkoholowe i wyroby tytoniowe
- Produkcja - dobra konsumpcyjne nietrwałe
- Przewozy ładunków w transporcie kolejowym
- Koniunktura - przetwórstwo przemysłowe
- Przeciętne zatrudnienie w sektorze przedsiębiorstwa
- Bezrobotni zarejestrowani
- Wskaźniki cen - wyposażenie mieszkania i prowadzenie gospodarstwa domowego
- Wskaźniki cen usług telekomunikacji
- Produkcja - dobra inwestycyjne
- Dochody budżetu państwa
- Produkcja - wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę
- Eksport towarów
- Wskaźniki cen produkcji budowlano-montażowej
- Bezrobotni zarejestrowani
- Produkcja - przetwórstwo przemysłowe
- Wskaźniki cen - żywność i napoje bezalkoholowe
- Koniunktura - obsługa rynku nieruchomości



## 6.2. ARIMA w Statistice

Do predykcji inflacji zostanie wykorzystana ARIMA (Autoregressive Integrated Moving Average)

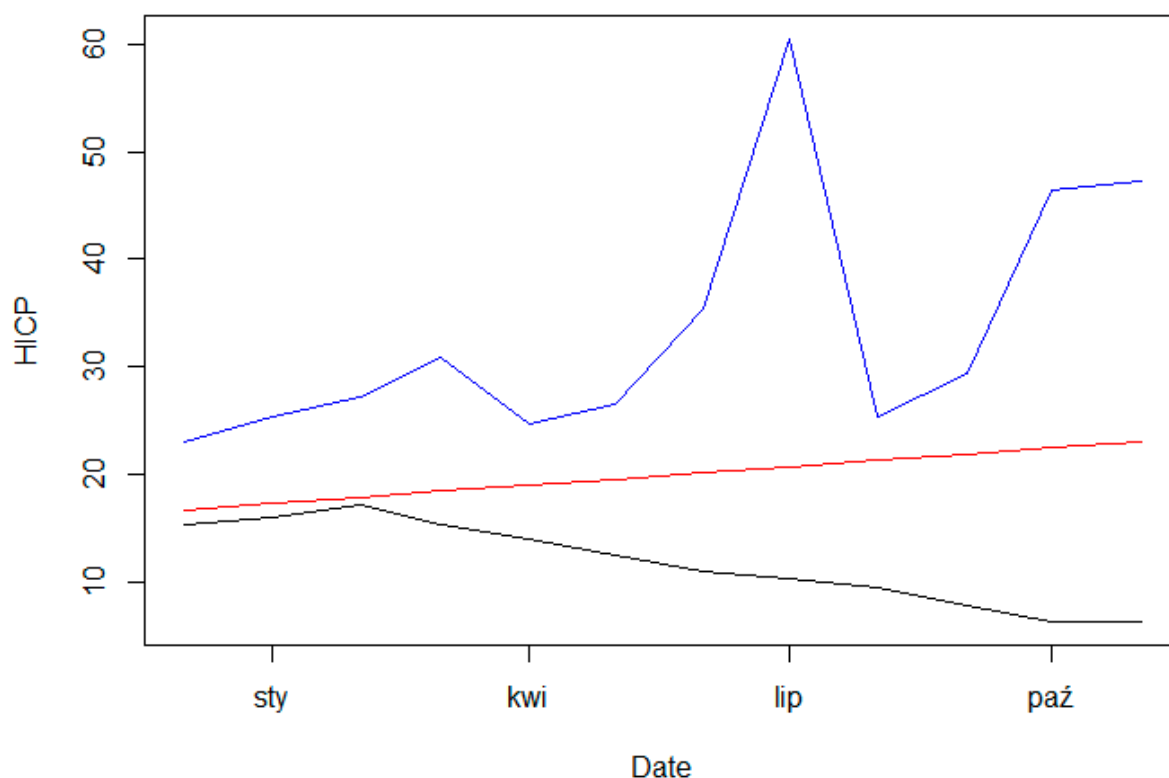


**Rysunek 39:** Predykcja inflacji modelem ARIMA w Statistice

## 6.3. ARIMA i regresja liniowa w RStudio

Czarnym kolorem zaznaczone są dane rzeczywiste, czerwonym Predykcja wartości za pomocą modelu ARIMA, natomiast niebieskim predykcja z użyciem regresji liniowej. Jak widać, wymienione modele nie są zbyt wiarygodnie w aspekcie przewidywania czynnika HICP.

Okazuje się, że szereg czasowy wskaźnika HICP powinien być stacjonarny, aby predykcja miała prawo być dokładną. W następnym kroku zostanie wykonana transformacja danych pod zastosowanie modelu.



**Rysunek 40:** Dane rzeczywiste (kolor niebieski), wartości przewidywane za pomocą ARIMA (kolor czerwony) i regresji liniowej (kolor czarny)

#### 6.4. ARIMA z transformacją danych

ARIMA jest narzędziem służącym do predykcji szeregów stacjonarnych, czyli takich, którego wartości statystyczne takie jak średnia czy mediana są stabilne i nie zmieniają się w czasie.

Aby sprawdzić stacjonarność szeregu czasowego z wartościami wskaźnika HICP, wykorzystuje się test ADF (Augmented Dickey-Fuller). Jeżeli po wykonaniu test p-wartość będzie mniejsza bądź równa 0.05, oznacza to, że szereg jest stacjonarny.

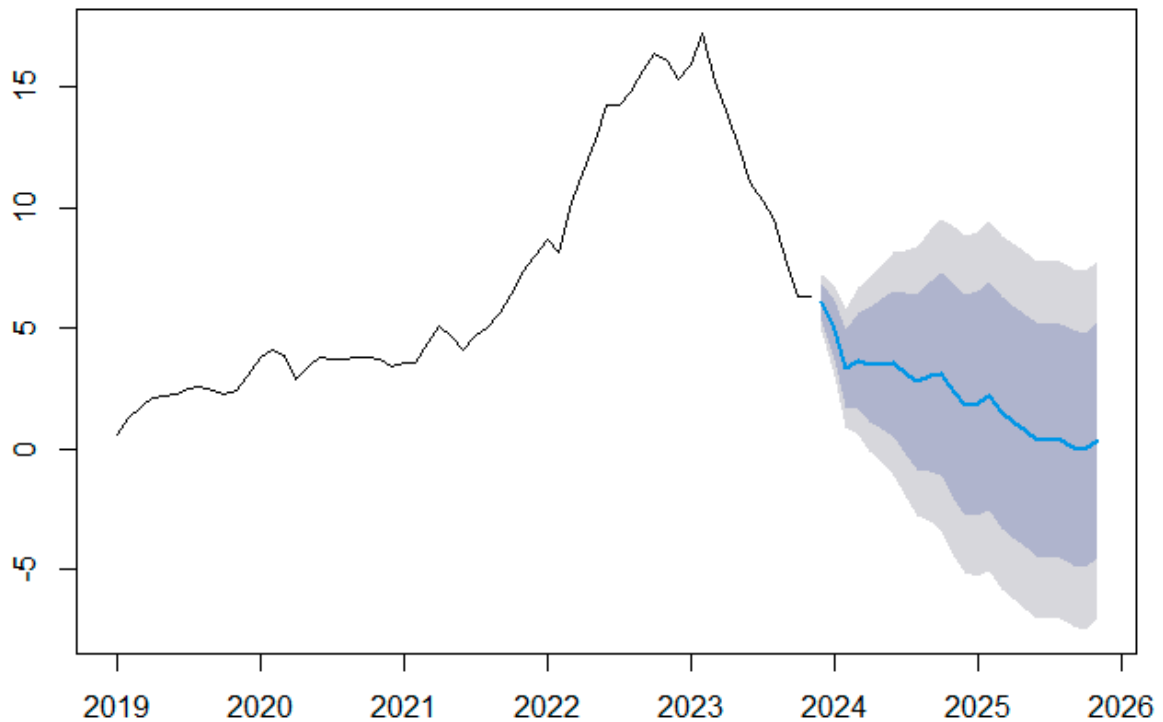
Po wykonaniu test ADF p-wartość wyniosła 0.75, zatem szereg czasowy jest niestacjonarny.

Aby to zmienić, zastosowano różnicowanie danych.

Po zróżnicowaniu p-wartość wyniosła 0.65, więc operację tę należało zastosować podwójnie.

Ostatecznie, po podwójnym różnicowaniu danych p-wartość wyniosła 0.01, więc można było zastosować model ARIMA do predykcji inflacji.

Wyniki predykcji po cofnięciu efektów różnicowania znajduje się na rysunku poniżej:



**Rysunek 41:** Predykcja za pomocą modelu ARIMA po wcześniejszym zróżnicowaniu danych

Wniosek:

Użycie modelu ARIMA bez wcześniejszego zróżnicowania danych doprowadza do predykcji, która nie jest wiarygodna. Jest to model stworzony do przewidywania wartości szeregów czasowych stacjonarnych, czyli takich, dla których p-wartość jest mniejsza bądź równa 0.05. Należy pamiętać, że model ten jest jednoczynnikowy, w związku z czym powyższa predykcja może sprawdzić się tylko w sytuacji, gdy nie będzie miało miejsca wydarzenie mający istotny wpływ na inflację.

## 7. Wnioski

Wniosek, jaki na pewno się wysuwa, to predykcja inflacji jest bardzo trudna nawet przy uwzględnieniu wielu czynników makroekonomicznych oraz wydarzeń historycznych. Na inflację wpływa bardzo wiele niespodziewanych czynników, takich jak wyniki finansowe spółek, wydarzenia w innych krajach oraz publikacje inflacji w innych krajach, które są ze sobą powiązane na przykład łańcuchami dostaw. Nawet eksperci ekonomiczni i specjaliści zajmujący się takimi badaniami na pełen etat, nie są w stanie przewidzieć dokładnie inflacji. Korzystają z wielu wyspecjalizowanych modeli, z których i tak wyniki nie są w 100% pewne, a dodatkowo są analizowane i poprawiane. Na pewno zabrakło nam "domain knowledge", czyli większej wiedzy lub eksperta z dziedziny ekonomii lub inflacji. W takich projektach takie osoby są konieczne, aby móc skonsultować z nimi otrzymane wyniki.

Udało się przetestować kilka modeli z różnych bibliotek czy języków programowania, które dawały lepsze oraz gorsze wyniki. W Pythonie testowano między innymi:

- Regresja Liniowa
- Regresja Grzbietowa
- Regresja Wielomianem 3 go stopnia
- Naive Seasonal
- ARIMA
- XGBoost
- Prophet
- LSTM

W przypadku Pythona, modelem, który radził sobie najlepiej, był LSTM. Model ten zwracał całkiem sensowne dane inflacyjne dla danych treningowych i testowych, a także dla prognoz na kolejne 6 miesięcy. LSTM uwzględniał utworzone sekwencje danych, na których następnie się uczył, co pozwalało mu na skuteczniejsze modelowanie skomplikowanych zależności w danych czasowych.

Również całkiem dobrze wypadła ARIMA, która przewidywała spadek inflacji, jednak zbyt wolny. W przypadku analizy wielu źródeł naukowych, to właśnie te 2 modele otrzymywały najlepsze wyniki dla predykcji inflacji.

Modele regresji liniowej i grzbietowej w standardowej postaci nie były w stanie efektywnie uchwycić sezonowości w danych inflacyjnych. Jest to wynikiem braku zmiennych opóźnionych, które są kluczowe do modelowania wzorców czasowych. Modele te mogą być ulepszone przez wprowadzenie zmiennych opóźnionych, które pozwolą im lepiej zrozumieć zależności czasowe.

Podobnie jak modele regresji liniowej, regresja wielomianowa nie uwzględniała sezonowości ze względu na brak zmiennych opóźnionych. Chociaż model ten może

lepiej dopasować się do nieliniowych wzorców w danych, bez odpowiednich zmiennych opóźnionych nie jest w stanie przewidywać sezonowych zmian.

## 8. Dalsze badania

Warto rozważyć zastosowanie technik łączenia różnych modeli (ensemble methods), takich jak stacking, bagging czy boosting, w celu połączenia wyników uzyskanych przez różne modele. Może to pomóc w uzyskaniu bardziej stabilnych i dokładnych prognoz poprzez wykorzystanie mocnych stron poszczególnych metod. Na przykład model hybrydowy ARIMA-LSTM mógłby wykorzystać mocne strony ARIMA w modelowaniu krótkoterminowych wzorców i LSTM w długoterminowych zależnościach.

Do puli opracowanych modeli można dodać poniższe propozycje:

- Dalsze eksperymenty z zaawansowanymi sieciami neuronowymi, takimi jak Transformer (np. model BERT dostosowany do szeregów czasowych) czy modele GAN, które mogą oferować lepsze możliwości w zakresie modelowania skomplikowanych wzorców w szeregach czasowych.
- Dynamiczne Stochastyczne Modele Równowagi Ogólnej (DSGE) są szczególnie użyteczne w analizie makroekonomicznej, ponieważ umożliwiają modelowanie gospodarki jako całości, uwzględniając wzajemne zależności między różnymi sektorami. Modele te mogą uwzględniać szoki stochastyczne oraz politykę monetarną i fiskalną, co pozwala na bardziej realistyczne i kompleksowe prognozy. Włączenie DSGE do naszego zestawu modeli mogłoby dostarczyć cennych dodatkowych informacji i poprawić jakość prognoz. Z kolei CNN-LSTM mógłby wykazać się lepszą zdolnością do efektywnej ekstrakcji cech z danych sekwencyjnych.
- Implementacja Bayesowskich modeli ARIMA lub innych modeli czasowych może poprawić dokładność prognoz poprzez uwzględnienie niepewności w szacowaniach parametrów modelu. Ponadto, Bayesian Structural Time Series (BSTS) to kolejna metoda, która może być rozważona do analizy trendów i sezonowości w danych inflacyjnych.

Dla rozszerzenia zakresu danych uczących można zastosować Techniki Przetwarzania Języka Naturalnego (NLP) do analizy tekstowych danych makroekonomicznych, takich jak raporty rządowe, artykuły prasowe, i publikacje banków centralnych. Informacje uzyskane z analizy sentymentu, identyfikacji tematów, oraz ekstrakcji kluczowych wskaźników mogą znacząco wzbogacić zestaw danych wejściowych, umożliwiając lepsze zrozumienie czynników wpływających na inflację. Integracja takich danych tekstowych z modelami predykcyjnymi może prowadzić do bardziej informacyjnych i precyzyjnych prognoz.