

# STATYSTYKA W ANALIZIE DANYCH

## SPRAWOZDANIE PROJEKT 1

Temat projektu

Problem 1: Inflacja z punktu widzenia studenta

Problem 2: Nakłady na badania i rozwój

Problem 3: Własna analiza i prezentacja

Jakub Dziegielewski

Mikołaj Bańkowski

Prowadzący

Rafał Rytel-Andrianik

6 maja 2024

## Spis treści

<b>1</b>	<b>Problem 1: Inflacja z punktu widzenia studenta</b>	<b>2</b>
1.1	Metoda rozwiązywania problemu . . . . .	2
1.2	Prezentacja wyników . . . . .	2
1.2.1	Wykresy cen produktów . . . . .	2
1.2.2	Pozostałe wskaźniki . . . . .	8
1.3	Interpretacja wyników . . . . .	10
<b>2</b>	<b>Problem 2: Nakłady na badania i rozwój</b>	<b>11</b>
2.1	Metoda rozwiązywania problemu . . . . .	11
2.2	Prezentacja wyników . . . . .	12
2.2.1	Nakłady na B+R . . . . .	13
2.2.2	Nakłady na B+R sektory . . . . .	15
2.2.3	Nakłady na B+R sektor BES . . . . .	15
2.2.4	Nakłady na B+R sektor GOV . . . . .	16
2.2.5	Nakłady na B+R sektor HES . . . . .	16
2.2.6	Nakłady na B+R sektor PNP . . . . .	17
2.2.7	Nakłady na B+R sektory ogółem . . . . .	17
2.3	Interpretacja wyników . . . . .	18
2.3.1	Nakłady B+R względem inflacji . . . . .	18
2.3.2	Nakłady B+R w różnych sektorach . . . . .	18
<b>3</b>	<b>Problem 3: Własna analiza i prezentacja</b>	<b>18</b>
3.1	Metoda rozwiązywania problemu . . . . .	18
3.2	Prezentacja wyników . . . . .	19
3.2.1	Prognoza demograficzna ludności Polski . . . . .	20
3.2.2	Liczba studentów . . . . .	21
3.2.3	Miejsca w domach studenckich, a liczba mieszkańców . . . . .	22
3.2.4	Średni dochód netto na osobę w miesiącu . . . . .	23
3.2.5	Mediana za 1m2 lokali mieszkalnych . . . . .	24
3.3	Interpretacja wyników . . . . .	24
<b>4</b>	<b>Kody źródłowe</b>	<b>25</b>

# 1 Problem 1: Inflacja z punktu widzenia studenta

Proszę wyszukać informacji dotyczących zmian cen towarów i usług. Następnie wybrać tylko te pozycje, które są istotne dla Pana/Pani i wykonać analizę wielkości zmian ich cen w ostatnim czasie (zależnie od dostępności danych). Porównać zakres zmian cen z oficjalną inflacją. Chodzi więc o stworzenie własnego koszyka inflacyjnego i zobrazowanie „swojej studenckiej inflacji”.

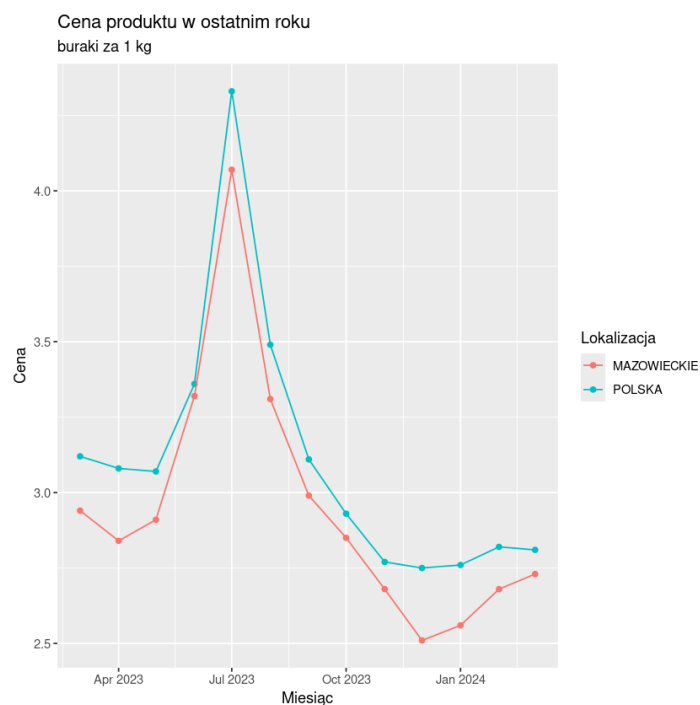
## 1.1 Metoda rozwiązania problemu

Problem numer jeden rozwiązana wykonując analizę zmian cen produktów spożywczych w okresie 03/2023-03/2024, na terenie Polski oraz województwa mazowieckiego. Analizę przeprowadzono z wykorzystaniem języka R, stosując biblioteki *Tidyverse* oraz *ggplot2*. Dane zostały pobrane z oficjalnej strony Głównego Urzędu Statystycznego.

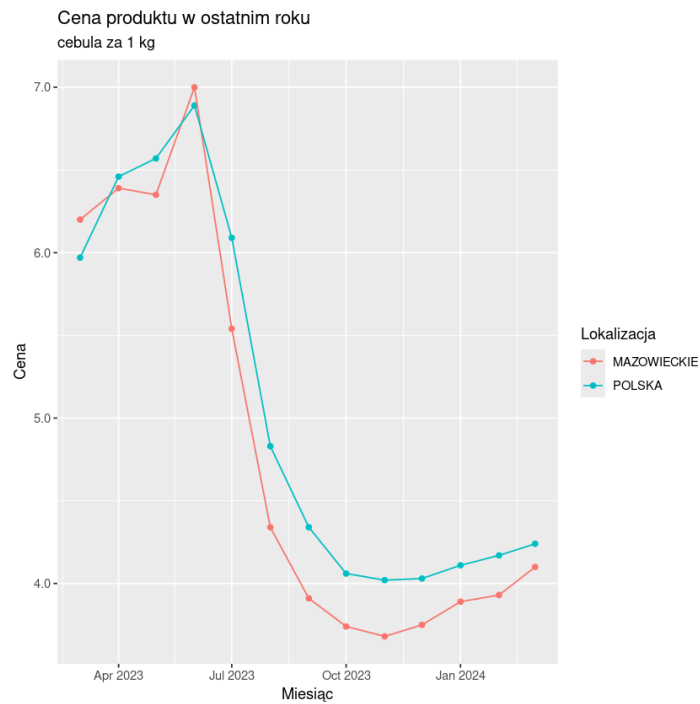
## 1.2 Prezentacja wyników

### 1.2.1 Wykresy cen produktów

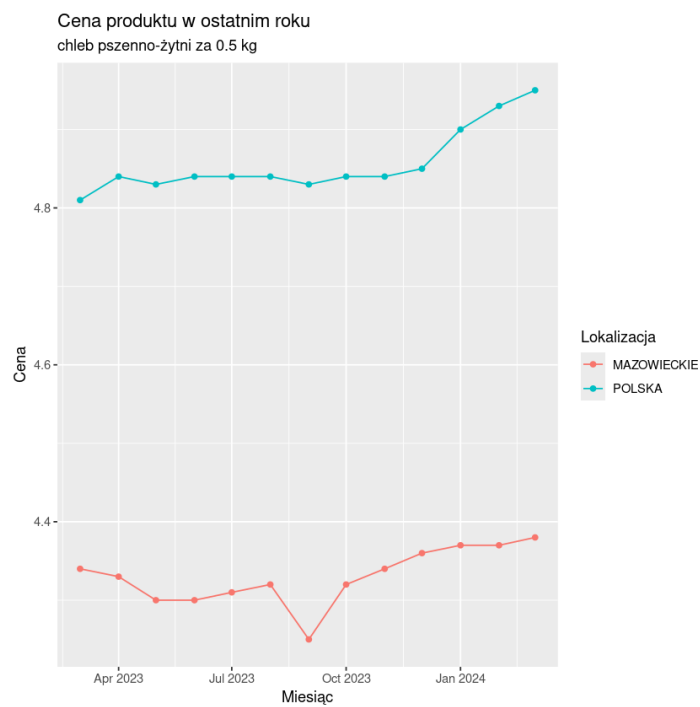
Wykonano wykresy cen dla 12 produktów spożywczych, na terenie Mazowsza oraz całej Polski, w okresie marzec 2023 - marzec 2024.



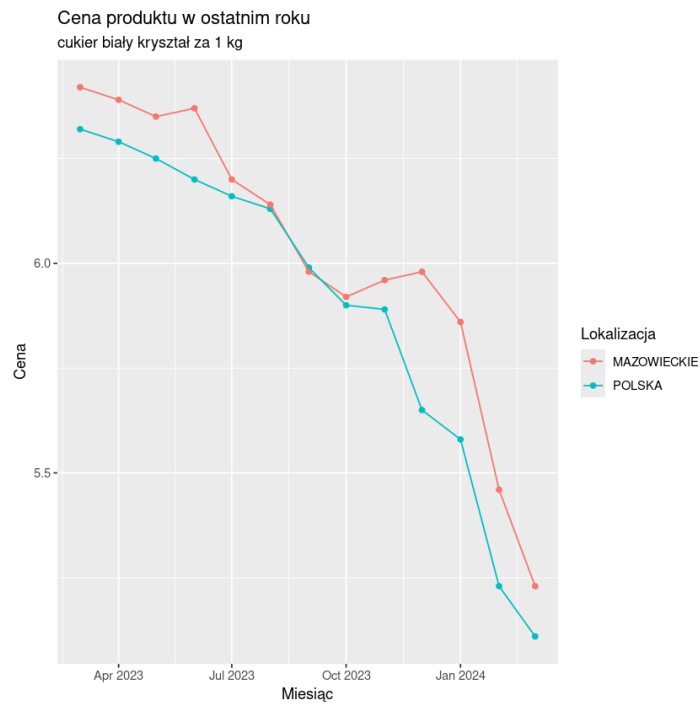
Rysunek 1: Wykres cen dla 1kg buraków



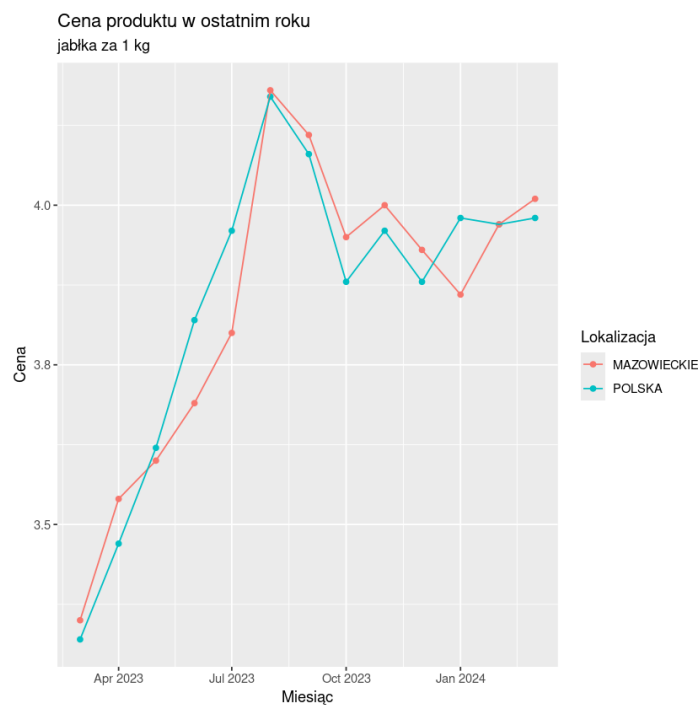
Rysunek 2: Wykres cen dla 1kg cebuli



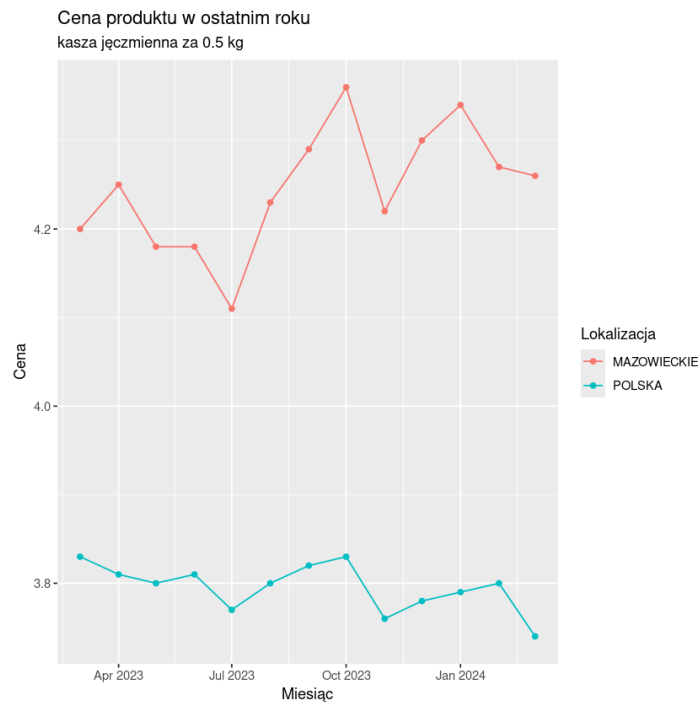
Rysunek 3: Wykres cen dla 1kg buraków



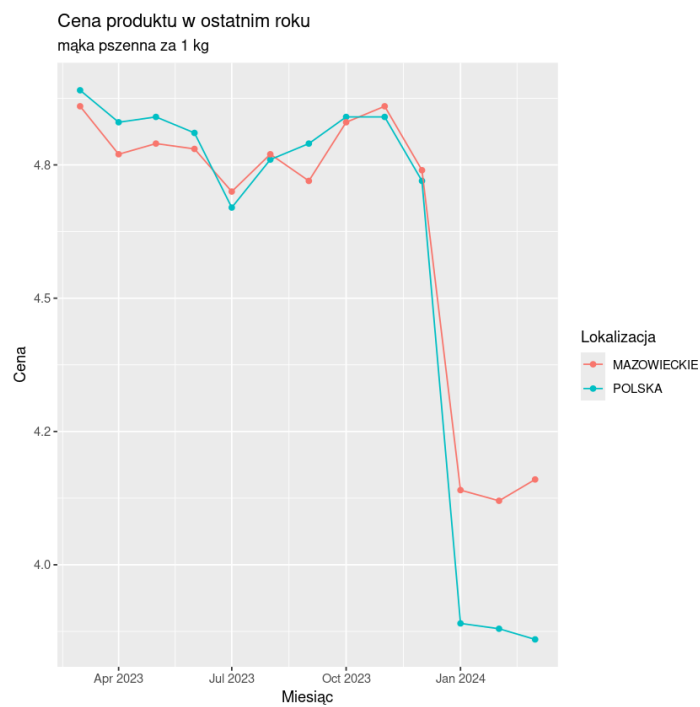
Rysunek 4: Wykres cen dla 1kg buraków



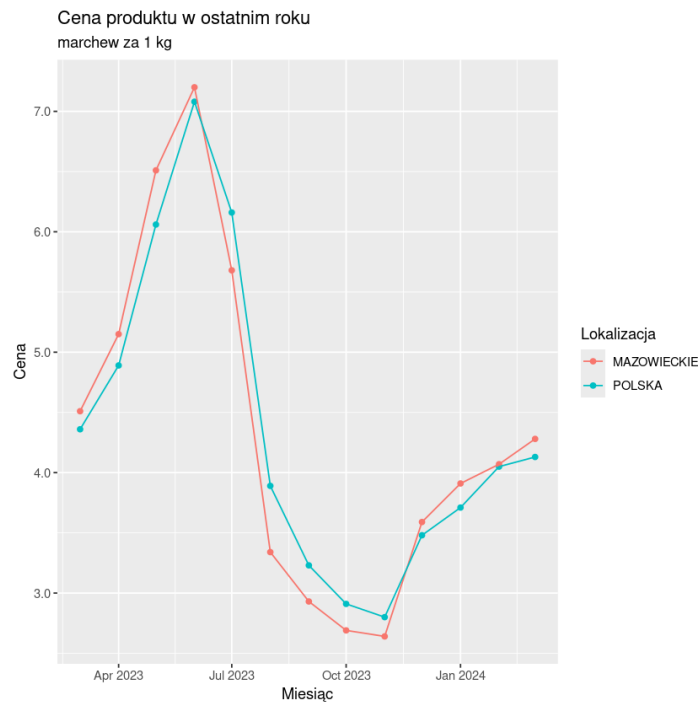
Rysunek 5: Wykres cen dla 1kg buraków



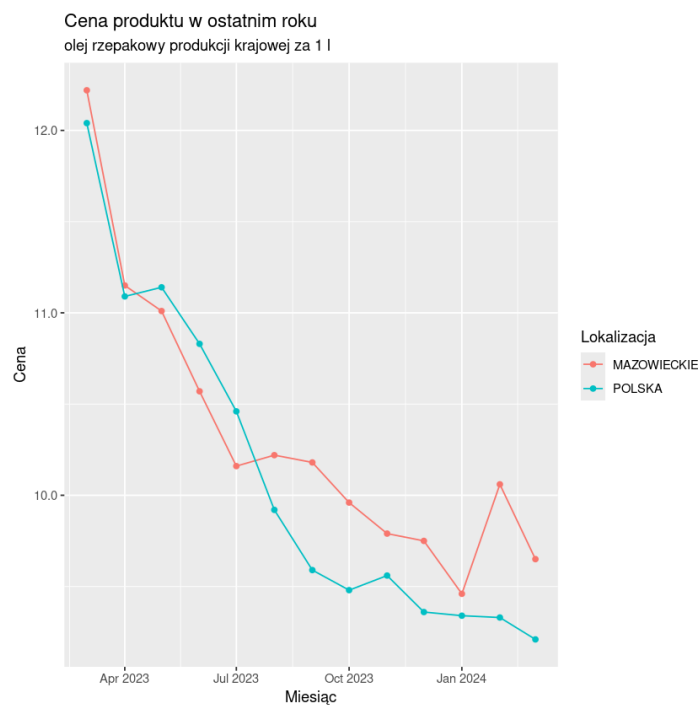
Rysunek 6: Wykres cen dla 1kg buraków



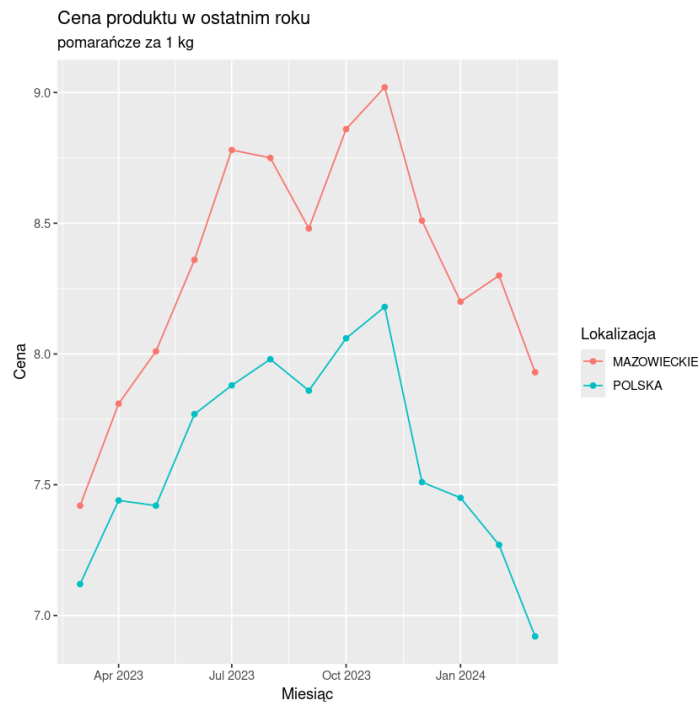
Rysunek 7: Wykres cen dla 1kg buraków



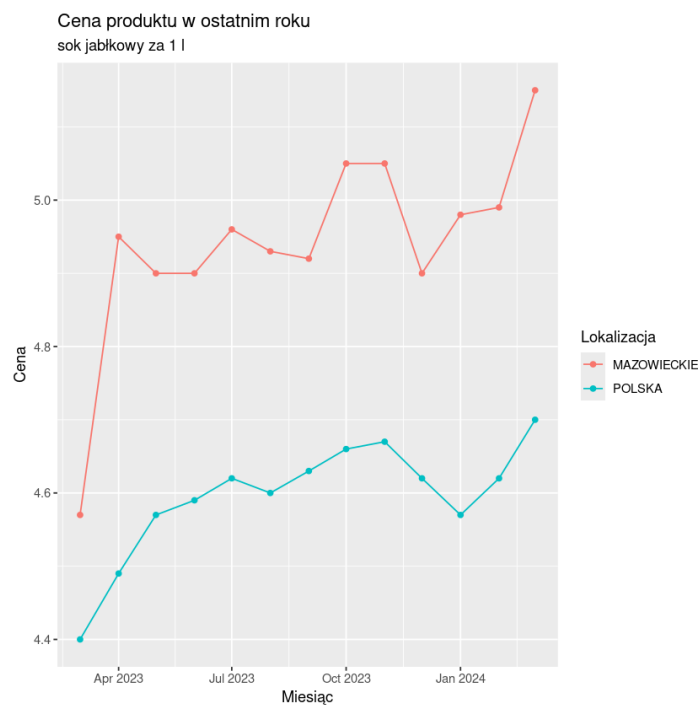
Rysunek 8: Wykres cen dla 1kg buraków



Rysunek 9: Wykres cen dla 1kg buraków

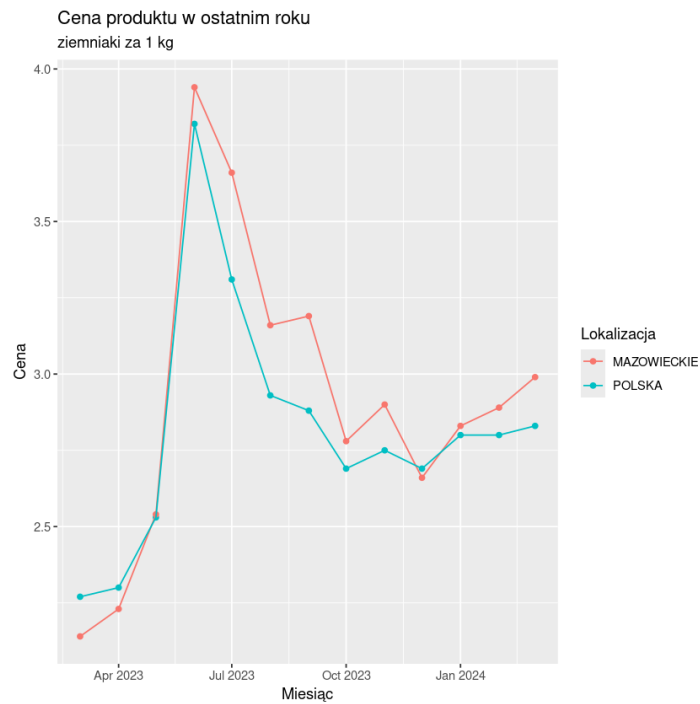


Rysunek 10: Wykres cen dla 1kg buraków



Rysunek 11: Wykres cen dla 1kg buraków

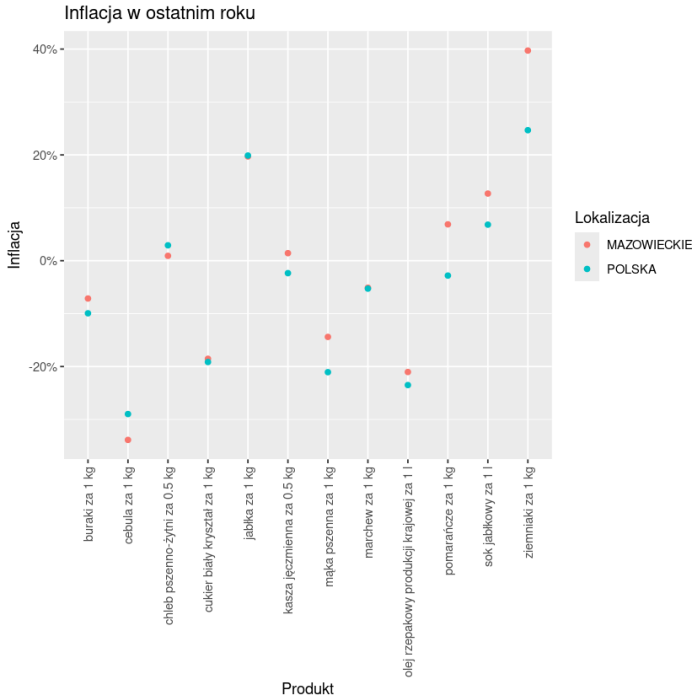




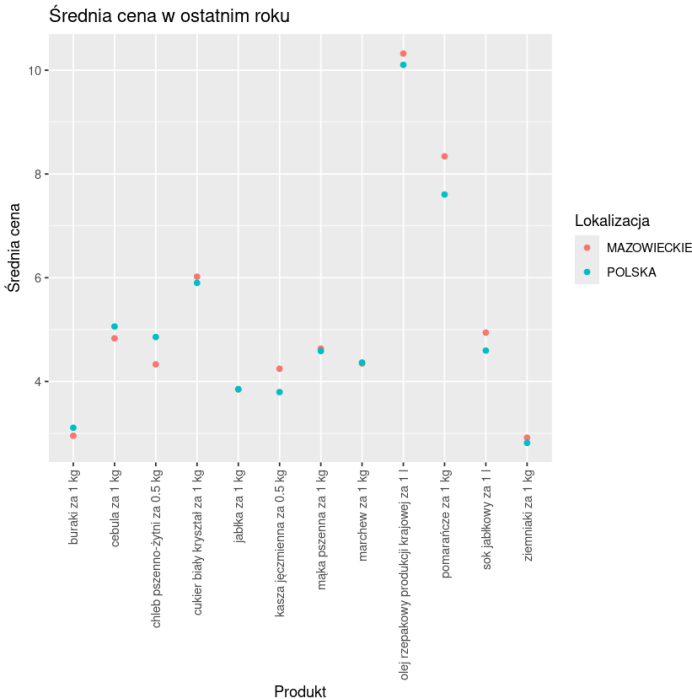
*Rysunek 12: Wykres cen dla 1kg buraków*

### 1.2.2 Pozostałe wskaźniki

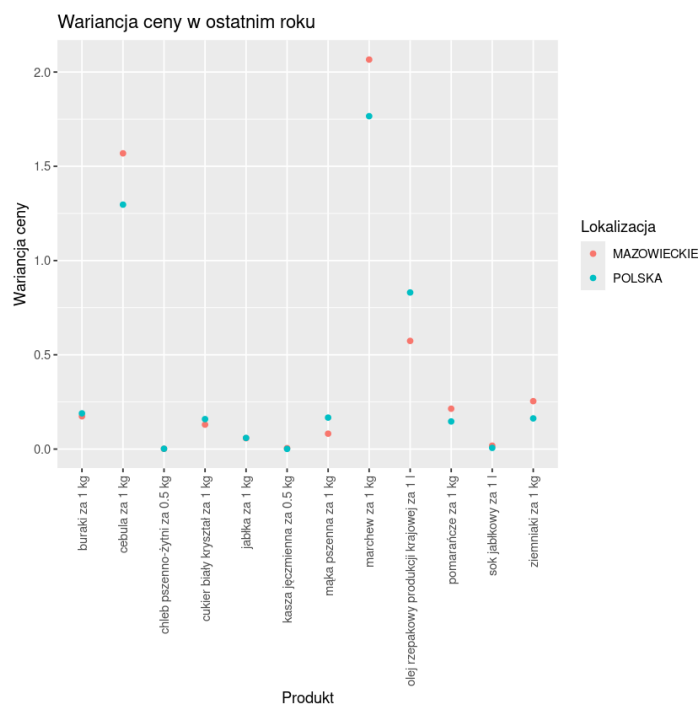
Dla 12 produktów wykonano wykresy wartości inflacji, średniej ceny oraz wariancji w okresie marzec 2023 - marzec 2024. Obliczono także inflację koszyka studenckiego, na który składają się produkty w ilości jednostek, dla których były tworzone wykresy.



Rysunek 13: Wartości inflacji



Rysunek 14: Średnia cena w okresie 03/2023-03/2024



Rysunek 15: Wariancja w okresie 03/2023-03/2024

lokalizajca	obliczona_inflacja
<chr>	<chr>
MAZOWIECKIE	-6.8%
POLSKA	-9.6%

Rysunek 16: Inflacja pełnego koszyka składającego się z jednostek wszystkich wymienionych produktów

### 1.3 Interpretacja wyników

Wg danych GUS-u ceny towarów w marcu 2024 wzrosły o średnio 0.4% w stosunku do ich poziomu z marca 2023. Cena koszyka studenckiego wybranych produktów uległa jednak obniżeniu o 6.8% w województwie mazowieckim, a także o 9.6% na terenie całego kraju, jak pokazuje Rysunek 16. Rozbieżność prawdopodobnie wynika ze zwiększonej dostępności wybranych produktów, co niweluje delikatne zmniejszenie siły nabywczej waluty.

Ceny większości produktów były dość chwiejne, co pasuje do danych GUS-u na temat inflacji. Na wykresach cen można zauważyć, że duża część produktów była najdroższa w okresie wakacyjnym 2023, kiedy inflacja była bardzo wysoka (11.5% w czerwcu 2023).

Odnosnie różnic pomiędzy Polską, a Mazowszem: średnie ceny na Mazowszu są porównywalne do średnich cen w Polsce (z drobnymi wyjątkami jak np. chleb żytni). Mazowsze można podzielić na tereny głównie wiejskie oraz metropolię warszawską. Można zatem zakładać, że średnie ceny na Mazowszu będą niższe od tych w całej Polsce, jeśli dane produkty mogą być

wyprodukowane lokalnie. Natomiast będą wyższe, jeśli produkt musi być importowany. Średnia cena chleba pszenno-żytniego przez cały rok był znacznie niższa w województwie mazowieckim niż średnia cena w całej Polsce. Z kolei sok pomarańczowy był średnio tańszy na terenie całej Polski niż województwa mazowieckiego.

## 2 Problem 2: Nakłady na badania i rozwój

Na podstawie tego samego źródła danych proszę zbadać:

- Jak zmieniały się w Polsce od 1990 roku nakłady na badania i rozwój (B+R)?
- Odnieść te zmiany do inflacji. Czy realne nakłady wzrastały?
- Czy jest widoczny trend w strukturze tych wydatków, tzn. jak zmieniają się te wydatki w sektorach przedsiębiorstw, rządowym, szkolnictwa wyższego (te dane mogą być ograniczone do lat 2020-2022).

### 2.1 Metoda rozwiązania problemu

Do rozwiązania problemu 2 wykorzystano dane statystyczne Głównego Urzędu Statystycznego (GUS), dostępne pod adresem <https://dbw.stat.gov.pl/baza-danych>, w zakładce Nauka i technika > Nakłady na działalność B+R > Nakłady wewnętrzne na działalność badawczą i rozwojową oraz Nauka i technika > Podmioty B+R > Podmioty w działalności badawczej i rozwojowej.

Aby zbadać jak zmieniały się w Polsce nakłady na badania i rozwój (B+R), wykorzystano dane z lat od 1999 roku. Decyzja ta wynikała z faktu, że na stronie GUS dostępne są dane jedynie od roku 1999 w górę. Natomiast w celu zbadania, jak zmieniają się te wydatki w poszczególnych sektorach, użyto danych począwszy od roku 2010. Dłuższy interwał czasowy pozwolił nam lepiej zrozumieć, czy na nakłady na poszczególne sektory miały wpływ zewnętrzne zjawiska, takie jak pandemia COVID-19.

Rozwiązując problem 2 wykonałem następujące kroki:

- Zbieranie danych: Dane dotyczące nakładów na badania i rozwój (B+R) zostały zebrane z odpowiednich źródeł, takich jak Główny Urząd Statystyczny (GUS).
- Analiza danych: Zebrane dane zostały poddane analizie, aby zidentyfikować zmiany w nakładach na B+R oraz poszczególnych sektorach
- Tworzenie wykresów: Na podstawie analizy danych zostały stworzone wykresy, aby zobrazować zmiany w nakładach na B+R oraz w poszczególnych sektorach w kolejnych latach.
- Interpretacja wyników: Wyniki analizy zostały zinterpretowane w celu zrozumienia ewentualnych trendów lub wzorców w strukturze nakładów na B+R oraz w poszczególnych sektorach. Na podstawie tych danych można było określić, czy realne nakłady wzrastały oraz czy widoczny był jakiś trend w strukturze wydatków

Do rozwiązania zadania wykorzystano język programowania Python oraz następujące biblioteki:

- Pandas (pd): Biblioteka do manipulacji i analizy danych, która umożliwia łatwe wczytywanie, przetwarzanie i analizowanie danych z różnych źródeł, w tym plików CSV, arkuszy kalkulacyjnych Excela i baz danych. Pozwala ona na efektywne zarządzanie danymi w postaci tabelarycznej.
- Os: Moduł biblioteki standardowej Pythona, który umożliwia interakcję z systemem operacyjnym. W tym przypadku został wykorzystany do manipulacji ścieżkami do plików, co jest przydatne przy wczytywaniu danych z plików znajdujących się w określonym folderze.
- Matplotlib.pyplot (plt): Biblioteka do tworzenia wykresów i wizualizacji danych w języku Python. Pozwala ona na generowanie różnego rodzaju wykresów, w tym wykresów liniowych, histogramów, wykresów słupkowych itp. Jest często używana w analizie danych do prezentowania wyników i trendów w sposób graficzny.
- Matplotlib.ticker as ticker: Moduł Matplotlib odpowiedzialny za formatowanie etykiet na osiach wykresu. Umożliwia on dostosowanie sposobu, w jaki są prezentowane wartości na osiach wykresu, co może być przydatne w celu czytelniejszego prezentowania danych.

## 2.2 Prezentacja wyników

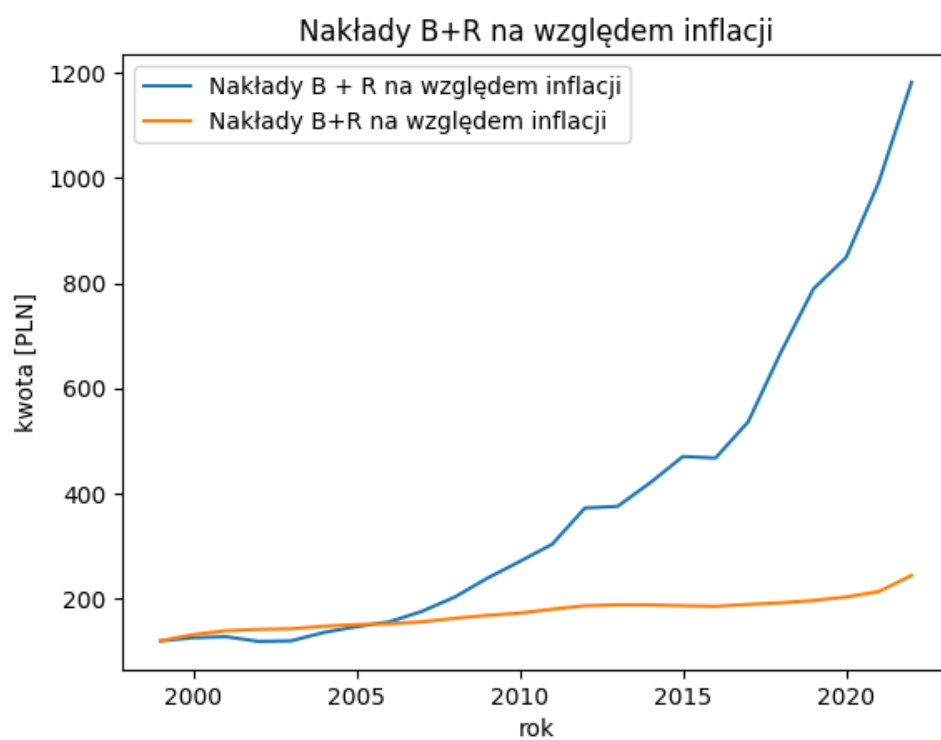
Nakłady B+R względem inflacji, to wykres, na którym przedstawiono zmiany w Polsce od 1999 roku oraz ich odniesienie do inflacji. W celu stworzenia wykresu inflacji, najstarsza kwota z roku 1999 została mnożona przez wskaźnik cen towarów i usług dostarczany przez portal GUS.

Nakłady na B+R w różnych sektorach to wykres pokazujący ich udział w ogólnym nakładzie na B+R. Pozwala to na porównanie, który sektor dominuje pod względem alokacji środków na badania i rozwój.

Nakłady na B+R w różnych sektorach każdy osobno to wykresy, które przedstawiają zmiany w nakładach na B+R dla sektorów BES, GOV, HES i PNP osobno. Pozwalają one na zrozumienie dynamiki wzrostu lub spadku nakładów w każdym z tych sektorów.

Nakłady na B+R ogółem to wykres przedstawiający ogólny trend nakładów na B+R dla wszystkich sektorów razem. Pomaga to zrozumieć, czy ogólny poziom nakładów rośnie, maleje, czy pozostaje stabilny w kontekście wszystkich sektorów.

### 2.2.1 Nakłady na B+R



*Rysunek 17: Nakłady na B+R względem inflacji*

÷	rok	÷	wartosc	÷
0	1999		119	
1	2000		125	
2	2001		127	
3	2002		118	
4	2003		119	
5	2004		135	
6	2005		146	
7	2006		155	
8	2007		175	
9	2008		202	
10	2009		238	
11	2010		270	
12	2011		303	
13	2012		372	
14	2013		375	
15	2014		420	
16	2015		470	
17	2016		467	
18	2017		536	
19	2018		668	
20	2019		789	
21	2020		849	
22	2021		992	
23	2022		1182	

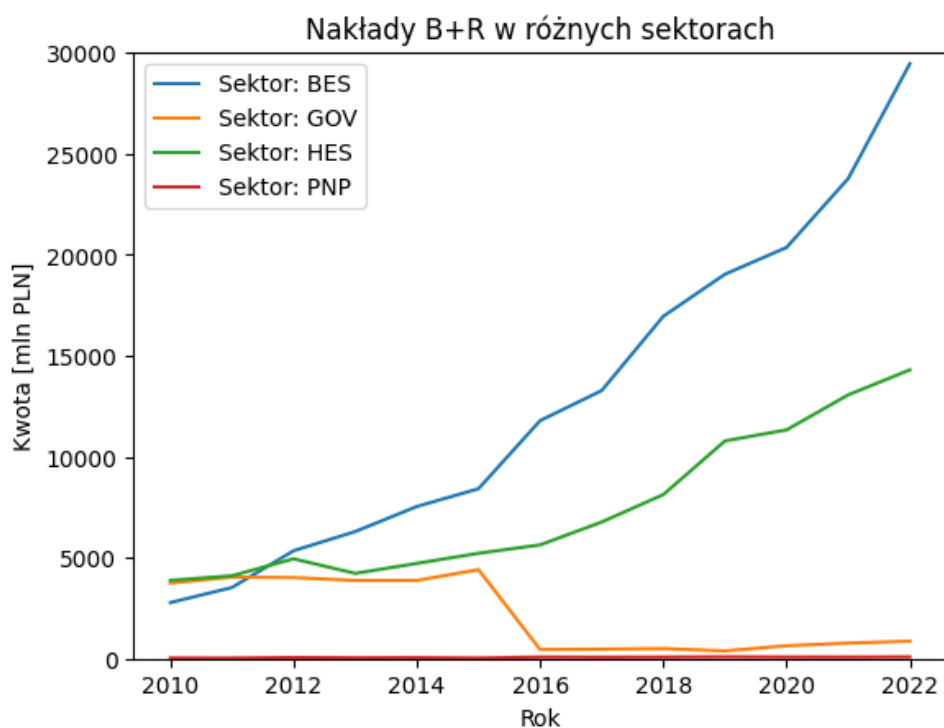
(a) Nakłady B+R

÷	rok	÷	wartosc	÷
0	1999		107.3	
1	2000		110.1	
2	2001		105.5	
3	2002		101.9	
4	2003		100.8	
5	2004		103.5	
6	2005		102.1	
7	2006		101.0	
8	2007		102.5	
9	2008		104.2	
10	2009		103.5	
11	2010		102.6	
12	2011		104.3	
13	2012		103.7	
14	2013		100.9	
15	2014		100.0	
16	2015		99.1	
17	2016		99.4	
18	2017		102.0	
19	2018		101.6	
20	2019		102.3	
21	2020		103.4	
22	2021		105.1	
23	2022		114.4	
24	2023		111.4	

(b) Inflacja

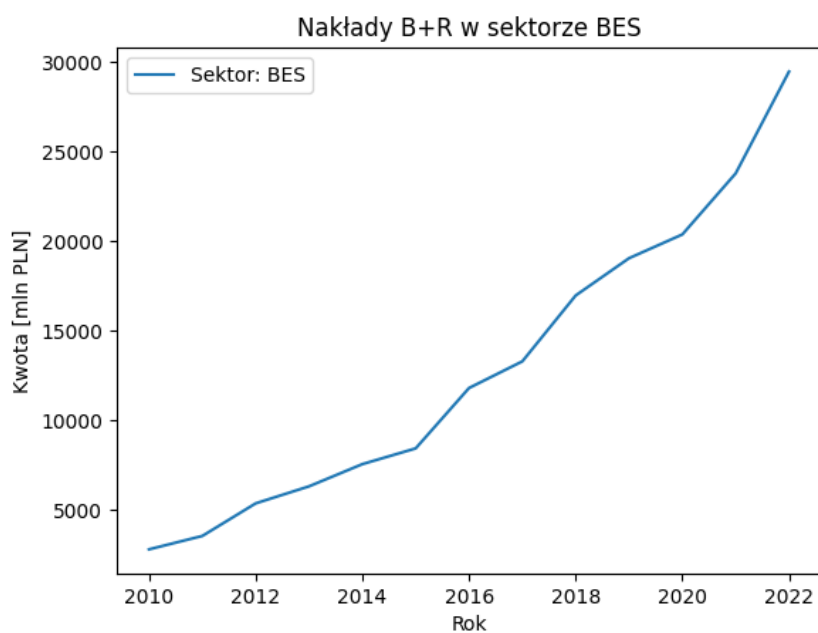
Rysunek 18: Nakłady na B+R względem inflacji

## 2.2.2 Nakłady na B+R sektory



*Rysunek 19: Nakłady na B+R według sektorów każdy osobno*

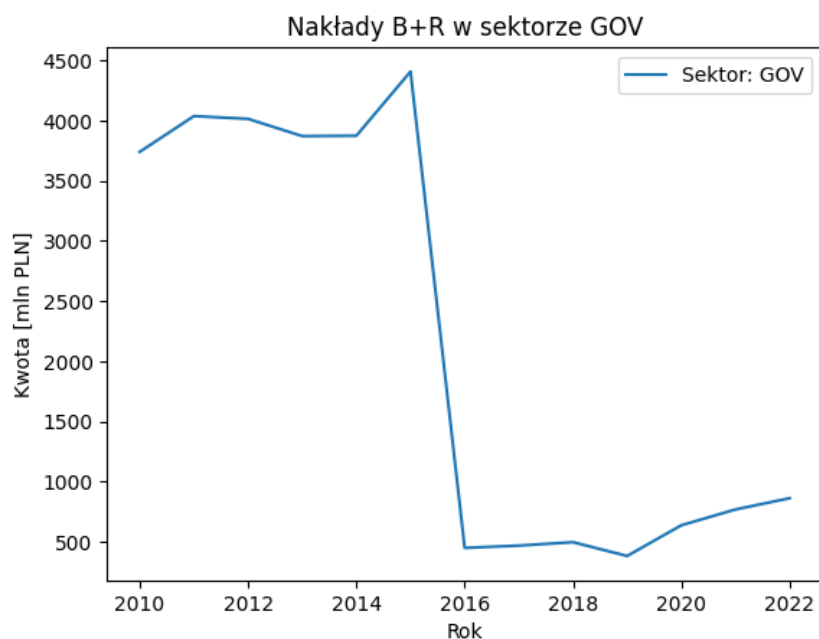
## 2.2.3 Nakłady na B+R sektor BES



*Rysunek 20: Nakłady na B+R sektor BES*

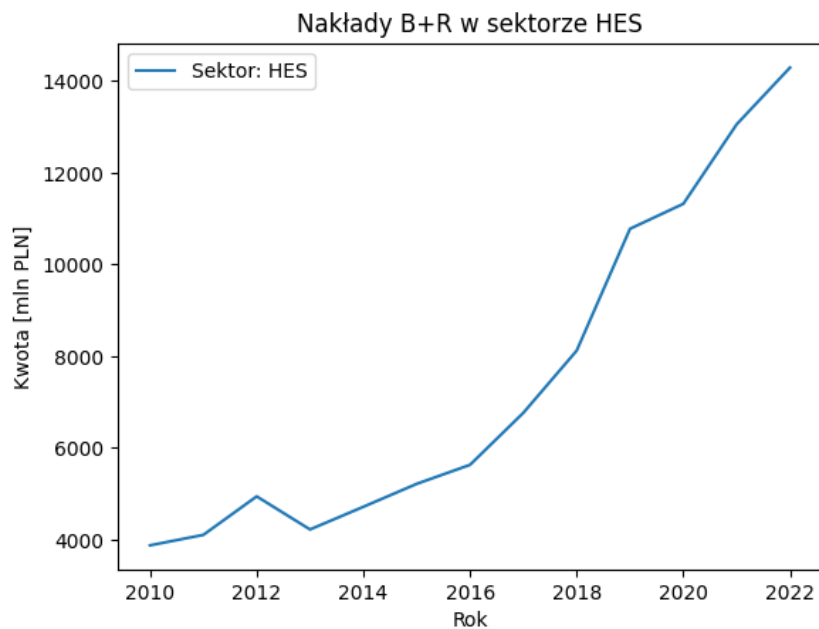


#### 2.2.4 Nakłady na B+R sektor GOV



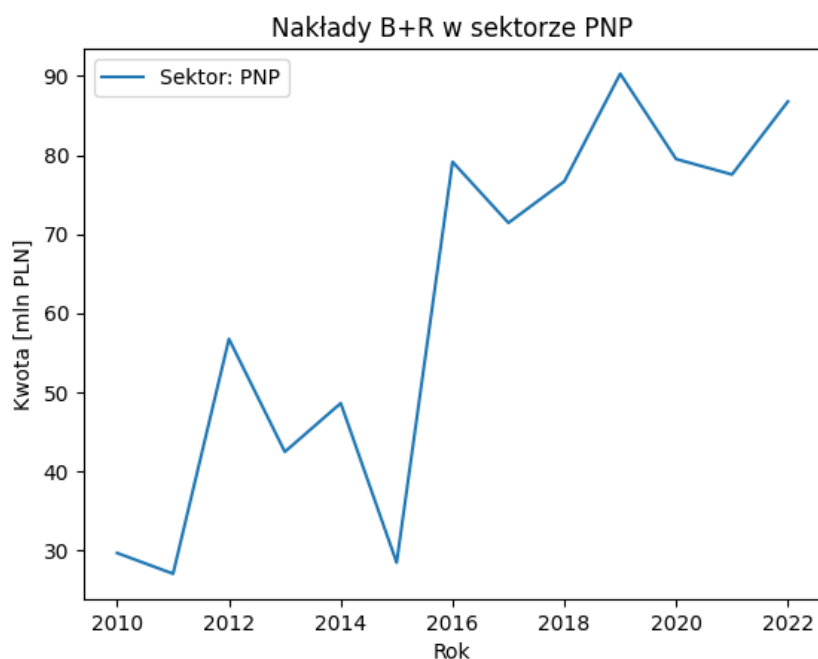
Rysunek 21: Nakłady na B+R sektor GOV

#### 2.2.5 Nakłady na B+R sektor HES



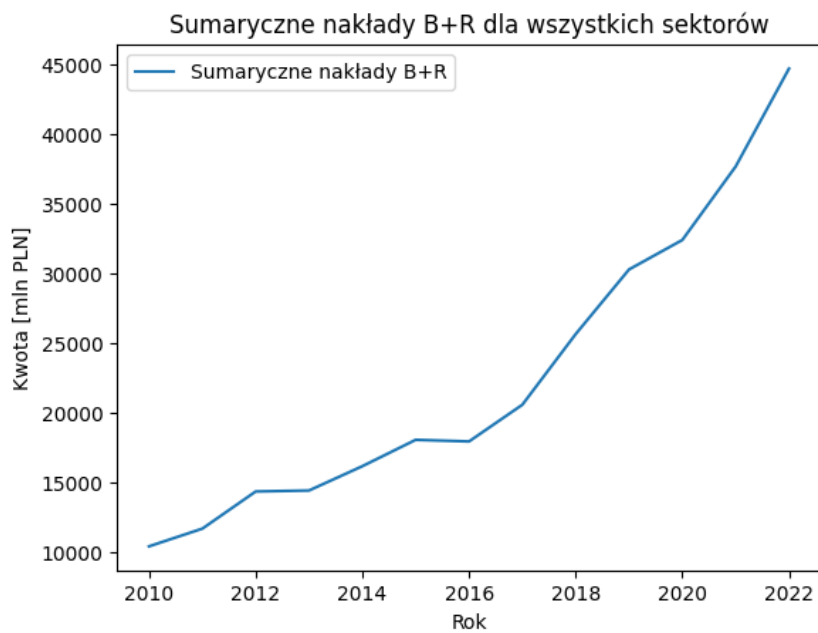
Rysunek 22: Nakłady na B+R sektor HES

### 2.2.6 Nakłady na B+R sektor PNP



Rysunek 23: Nakłady na B+R sektor PNP

### 2.2.7 Nakłady na B+R sektory ogółem



Rysunek 24: Nakłady na B+R sektory ogółem

## 2.3 Interpretacja wyników

Na podstawie wykresów i przeprowadzonej analizy można wyciągnąć następujące wnioski.

### 2.3.1 Nakłady B+R względem inflacji

Od 1999 roku nakłady na badania i rozwój (B+R) w Polsce wykazywały znaczący wzrost, co można zaobserwować na wykresie. Wzrost ten był bardziej dynamiczny niż wzrost wskaźnika inflacji, co wskazuje na rosnące zaangażowanie w rozwój naukowy i technologiczny kraju. W 1999 roku nakłady wynosiły 119 zł na mieszkańca, natomiast w 2022 roku wzrosły do 1182 zł na mieszkańca i mają trend rosnący.

W porównaniu z inflacją, wzrost nakładów na B+R był znacznie wyższy niż wzrost wskaźnika inflacji. Nakłady rozwijały się wykładniczo, co świadczy o coraz większym zaangażowaniu w badania i rozwój.

Analizując te dane, można stwierdzić, że Polska wyraźnie inwestuje w rozwój naukowy i technologiczny kraju, co zwiastuje dalszy postęp i rozwój w przyszłości.

### 2.3.2 Nakłady B+R w różnych sektorach

Nakłady na badania i rozwój (B+R) w Polsce od lat wykazywały tendencję wzrostową, co przyczyniło się do dynamicznego rozwoju kraju na przestrzeni lat. Największe kwoty przeznaczane na badania znajdują się w sektorze przedsiębiorstw, co skutkuje znacznym rozwojem przemysłu. Kolejnym sektorem z istotnymi nakładami jest szkolnictwo wyższe, gdzie stabilny wzrost środków sprzyja edukacji wyspecjalizowanych pracowników, wspierając tym samym rozwój różnych gałęzi przemysłu. Choć kwoty alokowane na sektor rządowy oraz prywatne instytucje niekomercyjne są znacznie niższe, to także odnotowuje się w nich wzrost nakładów w latach 2010-2022, co świadczy o rosnącym zainteresowaniu badaniami naukowymi i rozwojem technologicznym.

## 3 Problem 3: Własna analiza i prezentacja

Na podstawie tego samego źródła danych wykonać dodatkową analizę według własnego pomysłu (stopień skomplikowania powinien być podobny do poprzednich problemów). Wyniki tego zadania będą Państwo prezentować (potrzebne 1-2 slajdy w pdf) na wspólnym spotkaniu prawdopodobnie 8 maja. Na przykład można spróbować przekonać odbiorcę, że w Polsce pod względem analizowanych danych w ostatnich latach działa się dobrze, albo (do wyboru), że działa się źle.

### 3.1 Metoda rozwiązywania problemu

W naszym badaniu skupiliśmy się na analizie różnych aspektów związanych z mieszkalnictwem studenckim. Naszym głównym celem było zrozumienie sytuacji mieszkaniowej w kontekście dostępności miejsc w domach studenckich oraz liczby osób mieszkających w tych domach. Dane, których analizę przeprowadziliśmy, obejmowały okres od 2010 do 2022 roku. Naszym celem było zbadanie, czy na przestrzeni lat na mieszkalnictwo studenckie miały wpływ takie czynniki jak poprawa sytuacji materialnej Polaków, procesy demograficzne, zmiany na rynku nieruchomości oraz pandemia COVID-19.

Wykorzystane zbiory danych:

- Domy studenckie - informacje o dostępnych miejscach w domach studenckich
- Miejsca w domach studenckich - liczba miejsc mieszkalnych dostępnych w domach studenckich
- Korzystający z domów studenckich - informacje dotyczące osób, które aktualnie korzystają z mieszkań w domach studenckich
- Mediana cen za 1m<sup>2</sup> lokali mieszkalnych sprzedanych w ramach transakcji rynkowych - informacja o cenach transakcyjnych, która umożliwia ocenę typowych cen mieszkań w danym okresie czasu.
- Przeciętne miesięczne przychody netto na 1 osobę w gospodarstwach domowych - informacje na temat poziomu dochodów w gospodarstwach domowych, co może być użyteczne do analizy standardu życia

Wybraliśmy te konkretne zbiory danych, ponieważ uważamy są one kluczowe dla zrozumienia sytuacji dotyczącej dostępności miejsc w domach studenckich oraz liczby osób korzystających z tych miejsc. Nasze badanie miało na celu dostarczenie praktycznych informacji na temat mieszkalnictwa studenckiego

Analizę przeprowadzono w języku R, wykorzystując biblioteki Tidyverse i ggplot2, a także dodatkowo jpeg, scales i zoo.

### 3.2 Prezentacja wyników

Piramida demograficzna ludności Polski to wykres, który przedstawia strukturę wiekową populacji Polski, pokazując czy charakter społeczeństwa.

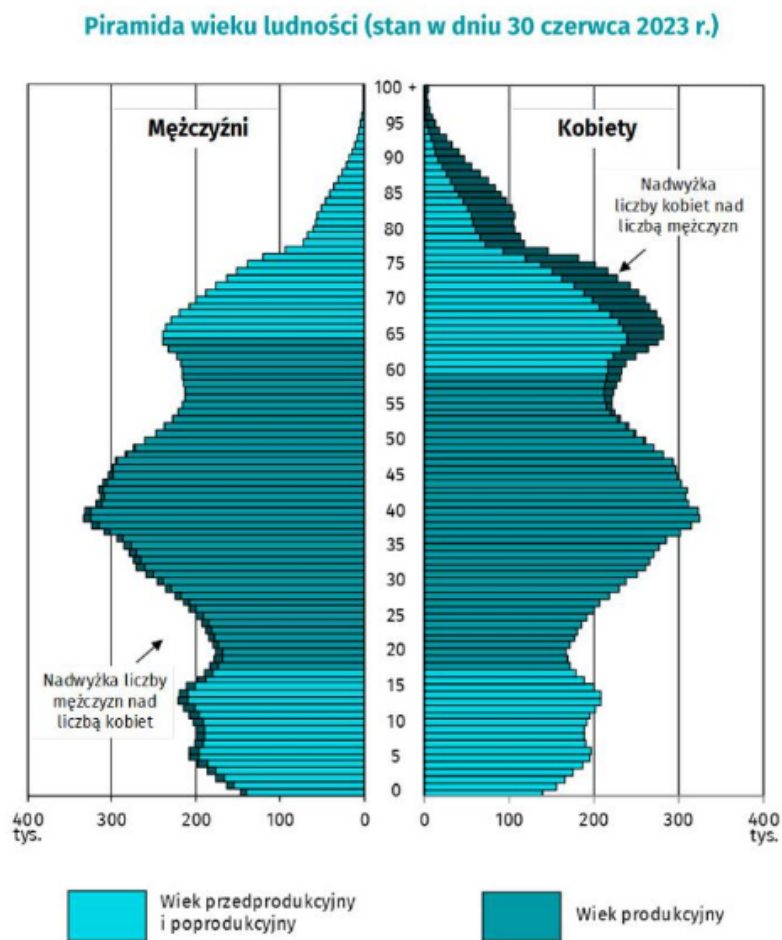
Liczba studentów to wykres, który przedstawia trend zmian liczby studentów na przestrzeni określonego okresu czasu.

Miejsca w domach studenckich, a liczba mieszkańców, to wykres, który porównuje dostępność miejsc w domach studenckich z liczbą osób faktycznie mieszkających w tych miejscach, pozwala zrozumieć jak zmienia się liczba osób korzystających z takiego zakwaterowania w stosunku do dostępnych miejsc

Średni dochód netto na osobę w miesiącu to wykres, który przedstawia średni dochód netto przypadający na jedną osobę w gospodarstwie domowym na miesiąc. prezentuje zmiany poziomu życia i dobrobytu w społeczeństwie w czasie.

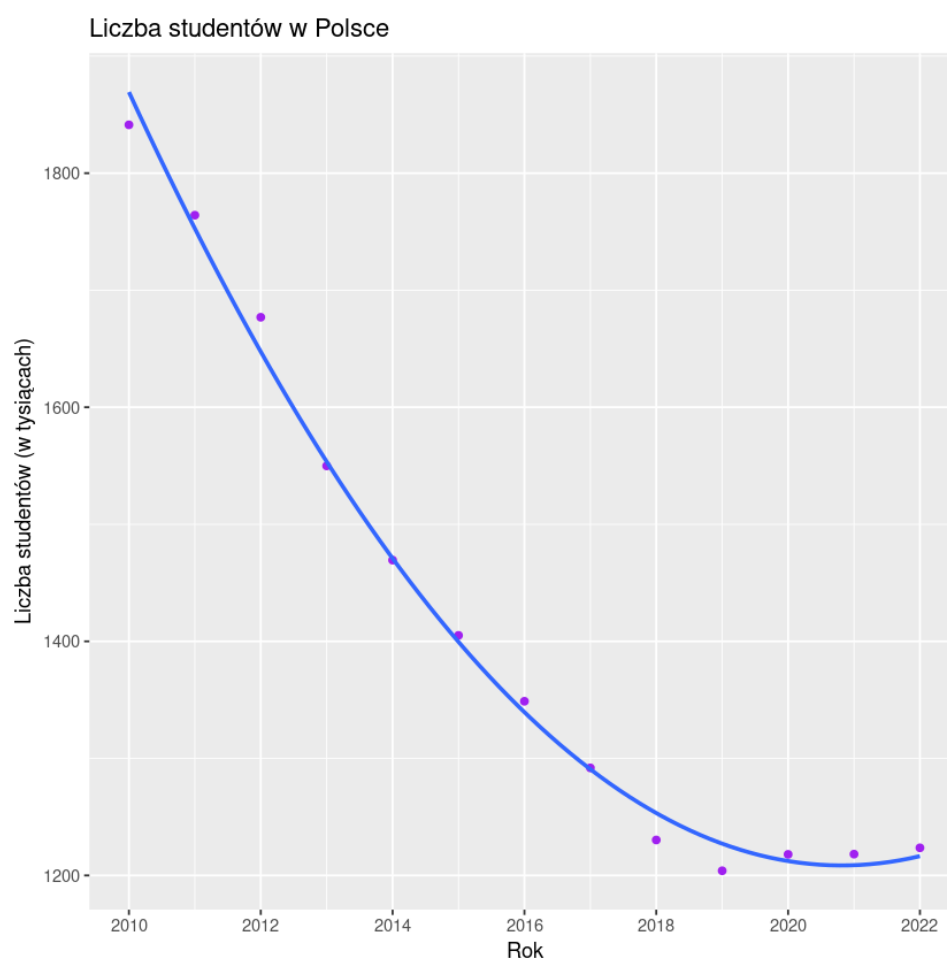
Mediana za 1m<sup>2</sup> lokali mieszkalnych to wykres, który pokazuje to, jakie są typowe ceny nieruchomości na danym rynku oraz jak się zmieniają w zależności od czasu

### 3.2.1 Prognoza demograficzna ludności Polski



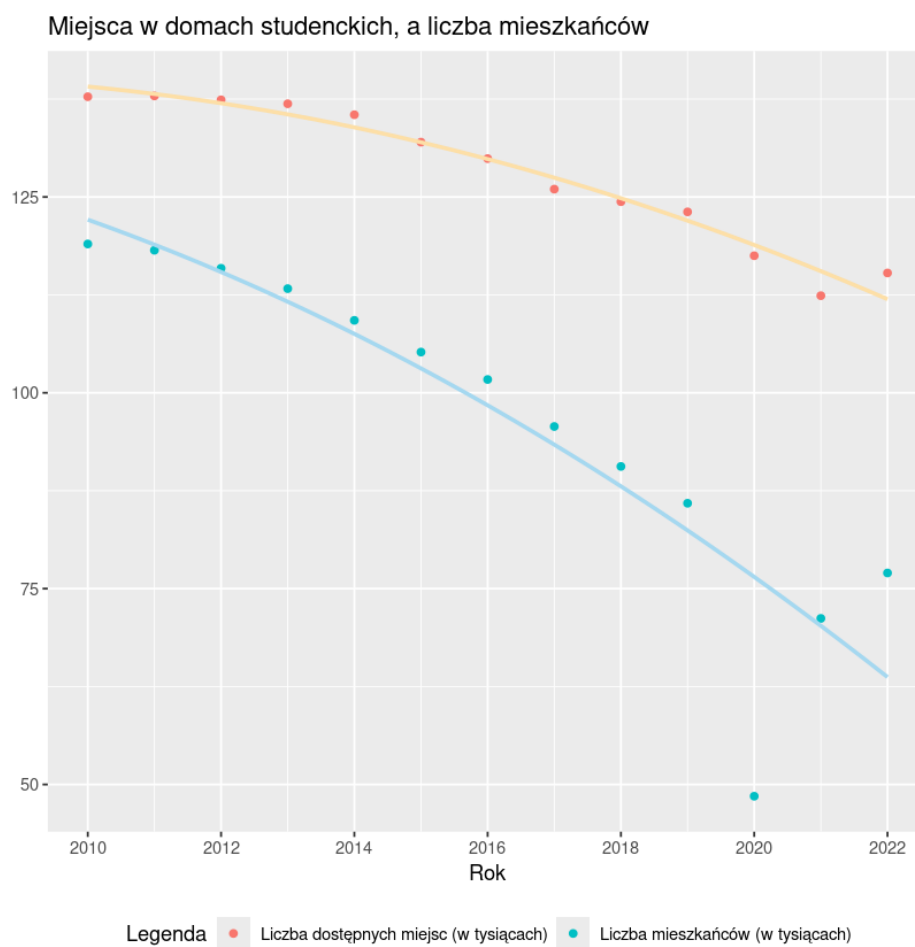
Rysunek 25: Piramida demograficzna ludności Polski

### 3.2.2 Liczba studentów



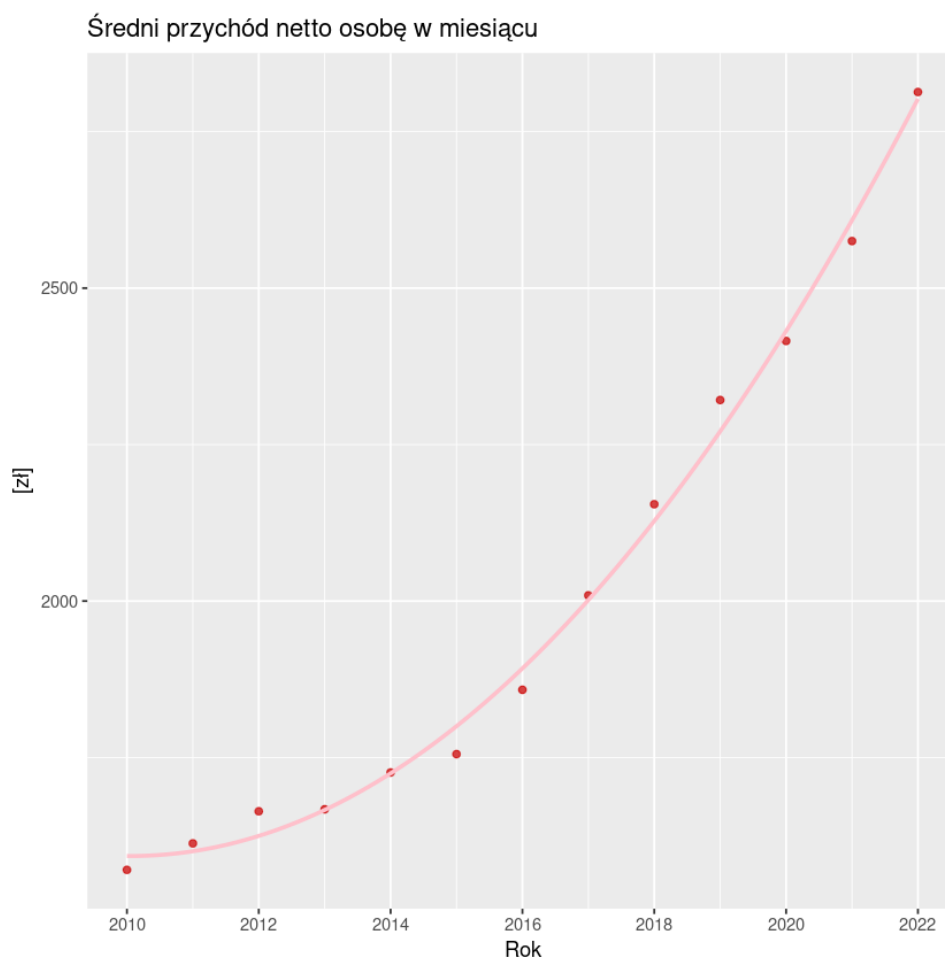
Rysunek 26: Liczba studentów

### 3.2.3 Miejsca w domach studenckich, a liczba mieszkańców



Rysunek 27: Miejsca w domach studenckich, a liczba mieszkańców

### 3.2.4 Średni dochód netto na osobę w miesiącu



Rysunek 28: Średni dochód netto na osobę w miesiącu



### 3.2.5 Mediana za 1m2 lokali mieszkalnych



Rysunek 29: Mediana za 1m2 lokali mieszkalnych

## 3.3 Interpretacja wyników

Liczba dostępnych miejsc w domach w poprzedniej dekadzie ulegała ciągłej redukcji, ze względu na mniejsze zainteresowanie zamieszkaniem w takich placówkach. Grupa wiekowa osób w wieku studenckim stale się zmniejszała, co naturalnie oznacza mniej potencjalnych mieszkańców, a z kolei sytuacja finansowa Polaków stawała się lepsza, przez co mogli się oni decydować na wynajem mieszkań, które gwarantują bardziej komfortowe warunki. W trakcie pandemii koronawirusa, liczba mieszkańców osiągnęła wynik znacznie niższy niż wynikałoby to z prognoz *przedcovidowych*, jednak ten wynik jest pochodną *lockdown'u*, którego nie dało się przewidzieć.

Co ciekawe, 2020 był pierwszym rokiem w badanym okresie, w którym liczba studentów się zwiększyła. Można zatem przypuszczać, że zdalny tryb nauczania jest bardziej przyjazny studentom (tzn. mniej osób rezygnuje ze studiów, a także mniejszy odsetek studentów nie uzyskuje zaliczenia), gdyż wg rysunku 25 ta liczba powinna wciąż jeszcze maleć.

W pierwszej części obecnej dekady liczba mieszkańców akademików rośnie. Razem z końcem pandemii i powrotem do nauczania stacjonarnego, zainteresowanie mieszkaniem w domach

studenckich wraca do normalnego poziomu. W 2022 to zainteresowanie osiągnęło poziom nawet wyższy od przewidywanego przez trend. Można przypuszczać, że to zwiększone zainteresowanie wynika z sytuacji na rynku mieszkaniowym. W chwili obecnej domy studenckie to opcja znacznie tańsza od wynajmu własnego mieszkania czy pokoju. W latach 2017-2022 mediana cen za 1 metr kwadratowy lokali mieszkalnych zwiększyła się o ok. 56%. Drugim potencjalnym czynnikiem zwiększonej liczby mieszkańców akademików jest wojna w Ukrainie i większa liczba ludzi szukających mieszkania na terenie Polski.

## 4 Kody źródłowe

Zadania zostały zrealizowane z wykorzystaniem narzędzia Jupyter Notebook. Pliki wraz z kodem źródłowym znajdują się w repozytorium github'owym `SAD_project_one` w odpowiednich podfolderach.