

PRZETWARZANIE, ANALIZA I WIZUALIZACJA DANYCH W  
SYSTEMIE SAS

# ANALIZA PRZEWIDYWANEJ DŁUGOŚCI ŻYCIA NA ŚWIECIE

NORBERT KABZIŃSKI, GR. GC01\_AD

UNIWERSYTET EKONOMICZNY W KATOWICACH

## Spis treści

1	Wprowadzenie.....	2
1.1	Cel Projektu .....	2
2	Dane źródłowe.....	3
2.1	Pochodzenie oraz charakterystyka.....	3
2.2	Import .....	4
2.3	Przygotowanie danych.....	4
3	Analizy .....	6
3.1	Charakterystyka przewidywanej długości życia w momencie narodzin na świecie w 2021 roku 6	
3.2	Zmiana przewidywanej długości życia na przestrzeni dekady .....	11
3.3	Różnice w przewidywanej długości życia pomiędzy płciami. ....	19
3.4	Wydatki na zdrowie na świecie oraz ich zależność z przewidywaną długością życia .....	25
3.5	Stan powietrza na świecie oraz zależność z przewidywaną długością życia w 2019 roku .....	31
4	Podsumowanie .....	38
4.1	Wnioski.....	38

# 1 Wprowadzenie

## 1.1 Cel Projektu

Przewidywana długość życia w momencie narodzin jest pojęciem istotnym w dziedzinie nauk społecznych i zdrowia publicznego, stanowiącym statystyczne odniesienie do szacowanej liczby lat, jakie jednostka ludzka, społeczeństwo lub populacja może oczekiwać przeżyć. To pojęcie obejmuje szeroki zakres czynników demograficznych, takich jak:

1. Wiek
2. Płeć
3. Stan zdrowia
4. Warunki społeczno-ekonomiczne
5. Dostęp do opieki zdrowotnej.

Niniejszy projekt ma na celu zgłębienie charakterystyki przewidywanej długości życia w momencie narodzin, analizę zmian, jakim ten wskaźnik ulegał w przeszłości, a także identyfikację istniejących różnic między państwami.

Dodatkowo w ramach tego projektu skupiono się również na analizie globalnych wydatków zdrowotnych oraz średniej rocznej ekspozycji na pył zawieszony PM2.5. Ta analiza ma na celu zbadanie ewentualnych związków i korelacji między tymi czynnikami a przewidywaną długością życia.

Analizy wyszczególnione w tym projekcie to:

1. Charakterystyka przewidywanej długości życia w momencie narodzin na świecie w 2021 roku.
2. Zmiana oraz tempo zmian przewidywanej długości życia na przestrzeni lat 2011-2021.
3. Różnice w przewidywanej długości życia pomiędzy płciami.
4. Wydatki na zdrowie na świecie oraz ich zależność z przewidywaną długością życia.
5. Stan powietrza na świecie oraz zależność z przewidywaną długością życia w 2019 roku.

## 2 Dane źródłowe

### 2.1 Pochodzenie oraz charakterystyka

Wybrane dane do analizy zostały pozyskane z zasobów otwartych Banku Światowego, dostępnych na stronie <https://databank.worldbank.org/>. Udostępnione zestawy danych są bezpłatne, a ich wykorzystanie jest zezwolone zarówno w celach komercyjnych, jak i niekomercyjnych.

Databank to narzędzie, dostarczone przez Bank Światowy, które umożliwia dostęp do szerokiego zbioru danych i wskaźników dotyczących każdego kraju na świecie. Przed pobraniem danych istnieje należy wybrać kraje, temat oraz okres. Przygotowany w ten sposób zestaw danych może zostać pobrany w formatach xlsx, csv lub txt. W ramach niniejszego projektu zdecydowano się na format csv.

Przy tworzeniu zestawu danych, zdecydowano się uwzględnić wszystkie dostępne państwa znajdujące się w bazie. Wybrane tematy, które mają zostać zawarte w bazie danych, to:

1. Przewidywaną długość życia od momentu narodzin, uwzględniając dane dla mężczyzn, kobiet oraz ogółem.
2. Wartość GNI per capita.
3. Wydatki zdrowotne per capita.
4. Poziom zanieczyszczenia powietrza pyłem zawieszonym PM2.5.

Informacje obejmują okres od 2011 do 2021 roku. Zastosowanie takiego zakresu czasowego umożliwi zbadanie ewolucji wybranych wskaźników przez ostatnią dekadę. Finalnie, opracowany zestaw danych będzie zawierał kompleksową informację na temat wymienionych aspektów dla wszystkich państw objętych analizą.

Dodatkowo pobrano zestaw danych obejmujący lata 2000-2021 w Niemczech dla tematów:

1. Przewidywaną długość życia od momentu narodzin ogółem.
2. Wydatki zdrowotne per capita.

Aby uzupełnić brakujące informacje o regionach w zestawie danych pobranym z zasobów banku światowego, skorzystano z dwóch dodatkowych zestawów danych. Pierwszy to metadane pobrane banku światowego opisujące do jakiej grupy dochodu należą. Drugi został pobrany ze

strony <https://github.com/>. Ta tabela zawiera kody państw oraz dodatkowe informacje o regionach i kontynentach, na których się znajdują.

## 2.2 Import

Proces importu danych rozpoczęto od utworzenia biblioteki o nazwie „Projekt” za pomocą kodu widocznego na rysunku 2-1. Następnie zamieszczono wszystkie zestawy danych na platformie SASOnDemand korzystając z funkcji o nazwie „Załaduj pliki”.

```
1 LIBNAME PROJEKT '/home/u63619458/Student/Projekt';  
2 run;
```

Rysunek 2-1 Kod tworzący bibliotekę "Projekt"

W kolejnym kroku należało zaimportować dane do biblioteki „Projekt”. W tym celu napisano kod składający się z trzech części widoczny na rysunku 2-2. Każdy jest odpowiedzialny za import jednego zestawu.

```
DATA Projekt.Zestaw;  
INFILE '/home/u63619458/Student/Projekt/Dane/Zestaw.csv' dsd firstobs= 2;  
INPUT Country_Name :$30. Alpha_3 :$3. Topic :$70. Topic_Code $  
      YR2011 YR2012 YR2013 YR2014 YR2015 YR2016 YR2017  
      YR2018 YR2019 YR2020 YR2021;  
DROP Topic_Code;  
RUN;  
PROC IMPORT DATAFILE= '/home/u63619458/Student/Projekt/Dane/KrajeInformacje.csv'  
      OUT=KrajeInformacje  
      DBMS=CSV  
      REPLACE;  
      GETNAMES = YES;  
      DATAROW = 2;  
RUN;  
PROC IMPORT DATAFILE= '/home/u63619458/Student/Projekt/Dane/KrajeInformacjeStatus.csv'  
      OUT=KrajeInformacjeStatus  
      DBMS=CSV  
      REPLACE;  
      GETNAMES = YES;  
      DATAROW = 2;  
RUN;  
PROC IMPORT DATAFILE= '/home/u63619458/Student/Projekt/Dane/Niemcy.csv'  
      OUT=PROJEKT.Niemcy  
      DBMS=CSV  
      REPLACE;  
      GETNAMES = YES;  
      DATAROW = 2;  
RUN;
```

Rysunek 2-2 Kod odpowiedzialny za import danych

## 2.3 Przygotowanie danych

Po udanym imporcie danych można było przejść do następnego kroku jakim jest dostosowanie ich dla potrzeb niniejszego projektu. W tym celu niezbędnym było dostosować dane żeby uniknąć możliwych błędów. Czynności jakie wykonano to dodanie znaku podkreślenia w miejsce spacji lub myślnika w każdym zestawie oraz wyczyszczenie danych poprzez usunięcie kolumn niepotrzebnych. Na potrzeby tej analizy połączono regiony Oceanii w jeden region „Oceania”. Kod odpowiedzialny za ten proces znajduje się na rysunku 2-3.

```

DATA Projekt.KrajeInformacjeStatus;
SET KrajeInformacjeStatus(KEEP = 'Code'n 'Income Group'n '2-alpha code'n);

RENAME 'Income Group'n = Income_Group;
RENAME '2-alpha code'n = Alpha_2;
RENAME 'Code'n = Alpha_3;

RUN;

DATA Projekt.KrajeInformacje;
SET KrajeInformacje;
IF 'sub-region'n IN ('Melanesia', 'Polynesia', 'Micronesia', 'Australia and New Zealand') THEN
  'sub-region'n = 'Oceania';
KEEP 'alpha-3'n 'region'n 'sub-region'n;
RENAME 'alpha-3'n = Alpha_3;
RENAME 'sub-region'n = Region;
RENAME 'region'n = Kontynent;

RUN;

DATA Projekt.Niemcy;
SET Niemcy;

drop 'Country Code'n 'Series Code'n;
RENAME 'Country Name'n = Country_Name;
RENAME 'Series Name'n = Topic;

RUN;

```

*Rysunek 2-3 Kod odpowiedzialny za dostosowanie danych do pracy w środowisku SAS*

Na zakończenie przygotowania danych należało połączyć tabelę w jedną. Ułatwi to prace z danymi na każdym etapie następnych analiz. Kod odpowiedzialny za ten proces znajduje się na rysunku 2-4. Używa procedury sortowania, usuwania oraz instrukcji „merge”. Tworzy tymczasową tabelę WORK.ZESTAWPOLACZONEWORK odpowiadającą połączeniu PROJEKT.ZESTAW z PROJEKT.KRAJEINFORMACJESTATUS. Później tabela WORK.ZESTAWPOLACZONEWORK zostaje połączona z PROJEKT.KRAJEINFORMACJE. Jako, iż w tabeli KRAJEINFORMACJE znajdują się też kody kontynentów dodano proces odpowiedzialny za pozbycie się rekordów, które po połączeniu nie zostały przypisane do żadnego istniejącego kraju.

```

1 proc sort data=PROJEKT.ZESTAW out=tabela1;
2   by Alpha_3;
3 run;
4
5 proc sort data=PROJEKT.KRAJEINFORMACJESTATUS out=tabela2;
6   by Alpha_3;
7 run;
8
9 data ZestawPolaczoneWork;
10  merge tabela1 tabela2;
11  by Alpha_3;
12 run;
13
14 proc delete data=tabela1 tabela2;
15 run;
16
17 proc sort data=ZestawPolaczoneWork out=tabela1;
18   by Alpha_3;
19 run;
20
21 proc sort data=PROJEKT.KRAJEINFORMACJE out=tabela2;
22   by Alpha_3;
23 run;
24
25 data ZestawPolaczoneWork2;
26  merge tabela1 tabela2;
27  by Alpha_3;
28 run;
29
30 data Projekt.ZestawPolaczone;
31  set ZestawPolaczoneWork2;
32  if not missing(Country_Name);
33 run;
34 proc delete data=tabela1 tabela2 ZestawPolaczoneWork ZestawPolaczoneWork2;
35 run;

```

*Rysunek 2-4 Kod odpowiedzialny za połączenie tabel*

Tabela wynikowa składa się z kolumn Country\_Name, Alpha\_3, Topic, Income\_Group, Alpha\_2, Kontynent, Region oraz YR2011-YR2021. Na tak przygotowanej tabeli będą opierać się wszystkie analizy.

### 3 Analizy

#### 3.1 Charakterystyka przewidywanej długości życia w momencie narodzin na świecie w 2021 roku

Celem szczegółowym analizy jest zgłębienie charakterystyki przewidywanej długości życia na skalę światową oraz w kontekście różnych regionów geograficznych w 2021 roku. W ramach tego celu zbadano różne miary, takie jak zróżnicowanie oraz asymetria przewidywanej długości życia w momencie narodzin zarówno na poziomie globalnym, jak i w poszczególnych regionach. Ponadto, analizie poddane zostały ekstrema. Zidentyfikowane zostały państwa, które odpowiadają za skrajne wartości długości życia. Ponadto ustalono, gdzie ludzie żyją najdłużej, a gdzie najkrócej, a także określono zróżnicowanie przewidywanej długości życia w momencie narodzin zależnie od regionu.

Przed rozpoczęciem analizy zbadano czy istnieją rekordy wśród badanych, w których brakuje co najmniej jednej wartości. W tym celu zrealizowano kod obecny na rysunku 3-1. Instrukcja ta sprawdza, czy w danych przedstawiających przewidywaną długość życia lub w kolumnę region znajduje się wartość pusta. Później z pomocą PROC SQL wyświetla te wartości w tablicy. Ta czynność została wykonana żeby uniknąć możliwych błędów, nie pracując na brakujących danych.

```
data TEST;
    set projekt.zestawpolaczone;
    drop Kontynent;
    where Topic = "Life expectancy at birth, total (years)" OR Region = '';
run;

proc sql;
    select *
    from work.test
    where
        Country_Name is null or
        Alpha_3 is null or
        Topic is null or
        YR2011 is null or
        YR2012 is null or
        YR2013 is null or
        YR2014 is null or
        YR2015 is null or
        YR2016 is null or
        YR2017 is null or
        YR2018 is null or
        YR2019 is null or
        YR2020 is null or
        YR2021 is null;
quit;
```

*Rysunek 3-1 Kod odpowiedzialny za sprawdzenie czy komórki są puste*

Spoglądając na tablicę widoczną na rysunku 3-2 można zauważyć, że w danych potrzebnych do tej analizy jedynie 11 rekordów z badanych 217 jest pustych. Brakujące rekordy wynoszą jedynie 3,22% wszystkich danych i przedstawiają głównie wyspy, małe państwa lub terytoria zależne, dlatego można je było wykluczyć z analiz podczas procesu tworzenia tabel bez obawy o znaczący wpływ na wyniki.

Country_Name	Alpha_3	Topic	YR2011	YR2012	YR2013	YR2014	YR2015	YR2016	YR2017	YR2018	YR2019	YR2020	YR2021	Income_Group	Alpha_2	Region
Andorra	AND	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	High income	AD	Southern Europe
American Samoa	ASM	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	High income	AS	Oceania
Channel Islands	CHI	GNI per capita, Atlas method (current US\$)	.	.	.	.	.	.	.	.	.	.	.	High income		
Channel Islands	CHI	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	.	.	.	.	.	.	.	.	.	.	.	High income		
Channel Islands	CHI	Current health expenditure per capita (current US\$)	.	.	.	.	.	.	.	.	.	.	.	High income		
Curacao	CUW	Life expectancy at birth, total (years)	77.47317	77.52439	77.82683	77.97561	78.07561	77.71951	78.01707	.	.	.	.	High income	CW	Latin America and the Caribbean
Cayman Islands	CYM	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	High income	KY	Latin America and the Caribbean
Monaco	MCO	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	High income	MC	Western Europe
Northern Mariana Islands	MNP	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	High income	MP	Oceania
Palau	PLW	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	Upper middle income	PW	Oceania
San Marino	SMR	Life expectancy at birth, total (years)	.	.	.	.	.	.	.	.	.	.	.	High income	SM	Southern Europe

Rysunek 3-2 Tablica z pustymi rekordami

W następnym kroku zrealizowano kod widoczny na rysunku 3-3. Kod ten przygotowuje dwie tabele i dostosowuje je do potrzeb każdego z etapów tej analizy. W pierwszej tabeli widocznej na rysunku 3-4, tworzył zestaw danych składający się z kraju, kodu kraju oraz przewidywanej długości życia. Z tabeli usunięto rekordy, dla których brakuje wartości dla 2021 roku.

```
DATA projekt.Przewidywana2021;
  SET projekt.zestawpolaczone(KEEP= Country_Name Alpha_2 YR2021 Region Topic);
  WHERE Topic = "Life expectancy at birth, total (years)";
  DROP TOPIC;
  RENAME YR2021 = Przewidywana_Dlugosc_Zycia;
  if nmiss(of _numeric_ ) then delete ;
RUN;

proc sort data=projekt.zestawpolaczone out=Sortowane;
  by Region;
run;
proc means data=Sortowane noprint;
  by Region;
  where Topic = 'Life expectancy at birth, total (years)';
  var YR2021;
  output out=Output(drop=_) mean=;
run;
data projekt.PrzewidywanaRegiony2021;
set Output;
if cmiss(of _character_) then delete ;
run;
```

Rysunek 3-3 Kod dostosowujący tabelę na potrzeby analizy



Druga tabela, znajdująca się na rysunku 3-5, zawiera informację o regionie i miarach przewidywanej długości życia. Dodatkowo kod odpowiedzialny za wygenerowanie tego zestawu sprawdza czy wartość nie jest pusta i jeśli jest to nie przekazuje jej do nowej tabeli. Ten wyjątek został dodany ze względu na to, że Kosovo oraz Wyspy Normandzkie nie mają przypisanego regionu. Jako, że są to tylko dwie niewielkie lokalizacje na mapie Europy zdecydowano, że ich usunięcie nie będzie miało ogromnego wpływu na wyniki.

Obs.	Region	YR2021
1	Central Asia	70.7700
2	Eastern Asia	79.3225
3	Eastern Europe	72.6857

Rysunek 3-4 3. Pierwsze rekordy tabeli PROJEKT.PRZEWIDYWANA2021

Obs.	Country_Name	Przewidywana_Dlugosc_Zycia	Alpha_2	Region
1	Aruba	74.6260	AW	Latin America and the Caribbean
2	Afghanistan	61.9820	AF	Southern Asia
3	Angola	61.6430	AO	Sub-Saharan Africa

Rysunek 3-5 3. Pierwsze rekordy tabeli PROJEKT.PRZEWIDYWANAREGIONY2021

Z tak przygotowanymi danymi można przystąpić do pierwszego etapu tej analizy. Zbadano miary tendencji centralnej, zróżnicowania oraz asymetrii dla przewidywanej długości życia. W tym celu wygenerowano tablicę z użyciem PROC MEAN oraz histogram obrazujący rozkład tego wskaźnika na świecie. Kod widoczny na rysunku 3-6.

```

title "Statystyki opisowe zmiennych numerycznych";
proc means data=PROJEKT.PRZEWIDYWANA2021 n nmiss min mean median max std cv q1 q3;
var Przewidywana_Dlugosc_Zycia;
run;

proc sgplot data=PROJEKT.PRZEWIDYWANA2021;
title "Przewidywana długość życia na świecie w momencie narodzin";
title2 'w 2021 roku';
footnote2 justify=left height=12pt "W latach";
histogram Przewidywana_Dlugosc_Zycia / scale=count fillattrs=(color=CX2572e6);
yaxis grid;
run;
```

Rysunek 3-6 Kod odpowiedzialny za wygenerowanie tablicy oraz histogramu

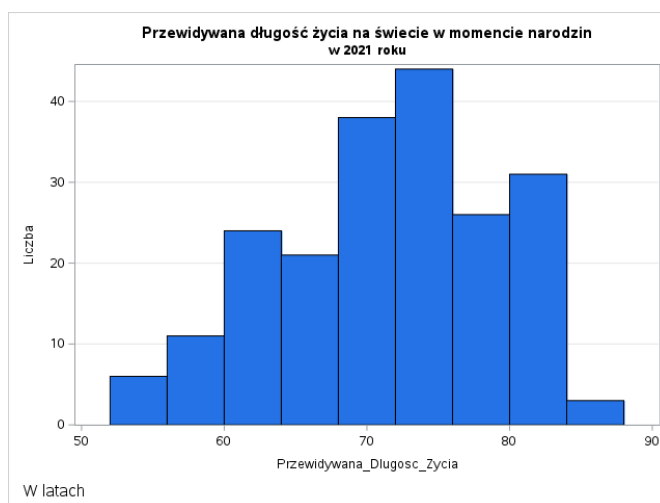
Wygenerowana tablica znajduje się na rysunku 3-7. Procedura ta ujawniła, że analizie zostało poddanych 206 krajów. Średnia długość życia na świecie wynosi około 71,57 lat, a mediana równa się około 72,37 lat. Najkrótsza przewidywana długość życia wynosi około 52,53 lat, a największa około 84,45 lat. Odchylenie standardowe jest równe około 7,59 lat co równa się

10,61 procentowemu współczynnikowi zmienności. Dolny kwartył wynosi około 66,08 lat, a górny w zaokrągleniu 76,94 lat.

Statystyki opisowe zmiennych numerycznych									
Procedura MEANS									
Zmienna analizowana: Przewidywana_Dlugosc_Zycia									
N	N braków	Minimum	Średnia	Mediana	Maksimum	Odch. std.	Wsp. zmienności	Dolny kwartył	Górny kwartył
204	0	52.5250000	71.5364812	72.2168415	85.4926829	7.6448206	10.6866041	65.9465000	76.8387317

Rysunek 3-7 Tablica z informacjami o miarach tendencji centralnej oraz zróżnicowania

Wygenerowany histogram sprawiał wrażenie istnienia asymetrii lewostronnej. W celu dokładnego zbadania tej cechy sprawdzono ile wynosi skośność tego rozkładu. Odkryto, że skośność dla tego rozkładu wynosi -0,34.



Rysunek 3-8 Wygenerowany histogram

Następnie zidentyfikowano, które kraje odpowiadają za wartości ekstremalne przewidywanej długości życia. W tym celu utworzono kod widoczny na rysunku 3-9, sortujący wartości rosnąco i wybierający pierwszą i ostatnią wartość. Stworzono tablicę wyświetlającą te państwa.

```

1 proc sort data=PROJEKT.PRZEWIDYWANA2021 out=work.sortDS0001 equals;
2   by Przewidywana_Dlugosc_Zycia;
3 run;
4
5 data MinMax;
6   set work.sortDS0001 end=last_row;
7
8   if _n_ = 1 then do;
9     output;
10  end;
11
12  if last_row then do;
13    output;
14  end;
15 run;

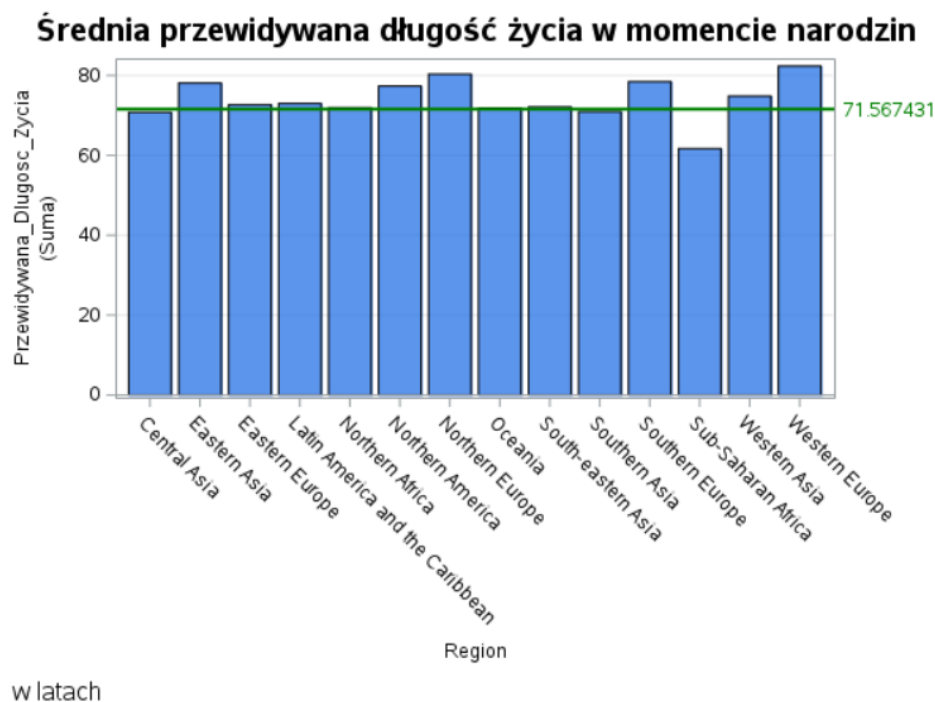
```

Rysunek 3-9 Kod odpowiedzialny za ukazanie krajów o maksymalnej i minimalnej wartości wskaźnika

Obs.	Country_Name	Przewidywana_Dlugosc_Zycia	Alpha_2	Region
1	Chad	52.5250	TD	Sub-Saharan Africa
2	Japan	84.4456	JP	Eastern Asia

Rysunek 3-10 Tablica przedstawiająca państwa o maksymalnej i minimalnej wartości wskaźnika

W ostatnim etapie korzystając z tabeli PRZEWIDYWANAREGIONY2021 wykres słupkowy przedstawiający wszystkie miary w tej tabeli obliczone podczas jej generowania.



Rysunek 3-11 Wykres słupkowy przedstawiający średnią przewidywaną długość życia dla regionów

Podsumowując, analiza ukazała istotne informacje dotyczące przewidywanej długości życia na skalę globalną i w kontekście różnych regionów. Stwierdzono, że co najmniej połowa krajów na świecie posiada przewidywaną długość życia wyższą od około 72,37 lat. Jednocześnie odnotowano, że dla 25% państw na świecie przewidywana długość życia nie przekracza 66,07 lat. Analiza rozkładu tego wskaźnika wykazała, że jest on asymetryczny lewostronny o słabej sile, co sugeruje, że większość krajów ma wartości wskaźnika powyżej średniej, wynoszącej około 71,57 lat. Dodatkowo, przewidywana długość życia na skalę globalną nie wykazuje silnego zróżnicowania, mimo istnienia znacznych różnic między najmniejszymi a największymi wartościami tego wskaźnika. Zauważono, że najniższą przewidywaną długość życia odnotowano w Chadzie, natomiast najwyższą w Japonii.

Podział analizy na regiony wykazał, że trzy z nich (21,4%), mają przewidywaną długość życia poniżej światowej średniej. Są to:

1. Afryka Subsaharyjska.
2. Południowa Azja.
3. Centralna Azja.

Afryka Subsaharyjska wyróżnia się najniższym wynikiem, co sugeruje istniejące wyzwania związane z długością życia w tym regionie. Największą średnią przewidywaną długość życia odnotowano w zachodniej Europie. Z kolei największe zróżnicowanie odnotowano w regionie Oceanii, co wskazuje na istnienie znaczących rozbieżności między wartościami tego wskaźnika w poszczególnych krajach tego obszaru.

### 3.2 Zmiana przewidywanej długości życia na przestrzeni dekady

Celem szczegółowym tej analizy jest zbadanie zmian w długości życia na przestrzeni lat. Analiza ta ma na celu określenie kierunku zmian – czy następuje wzrost, czy ewentualne zmniejszenie przewidywanej długości życia. Dodatkowo, analiza obejmuje identyfikację krajów, które po dekadzie charakteryzują się wyższymi bądź niższymi wartościami tego wskaźnika oraz ilość takich krajów. Badanie obejmuje również analizę trendu zmian na przestrzeni kolejnych lat oraz identyfikację regionów, w których te zmiany występują, a także tempo, w jakim się rozwijają, zarówno na skalę światową, jak i regionalną. W kontekście analizy, istotnym zagadnieniem jest również zbadanie różnic między średnim tempem zmian tego wskaźnika w grupach dochodowych.

Przed rozpoczęciem analizy dokonano wstępnego sprawdzenia rekordów, w których występuje brak co najmniej jednej wartości w kolumnach odpowiadających przewidywanej długości życia.

Country_Name	Alpha_3	Topic	YR2011	YR2012	YR2013	YR2014	YR2015	YR2016	YR2017	YR2018	YR2019	YR2020	YR2021
Andorra	AND	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-
American Samoa	ASM	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-
Curacao	CUW	Life expectancy at birth, total (years)	77.47317073	77.52439024	77.82682927	77.97560976	78.07560976	77.7195122	78.01707317	-	-	-	-
Cayman Islands	CYM	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-
Monaco	MCO	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-
Northern Maria	MNP	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-
Palau	PLW	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-
San Marino	SMR	Life expectancy at birth, total (years)	-	-	-	-	-	-	-	-	-	-	-

Rysunek 3-12 Tablica, przedstawiająca rekordy, w których brakuje danych

Rozpoznano braki danych w 8 rekordach dotyczących lokalizacji, mianowicie w przypadku Andory, Samoa Amerykańskich, Curacao, Kajmanów, Monako, Marianów Północnych, Palau oraz San Marino. Wszystkie wymienione lokalizacje charakteryzują się

statusami geograficznymi, takimi jak wyspy, terytoria zależne lub bardzo małe państwa. W związku z tym, zdecydowano się na usunięcie tych rekordów w następnym kroku, gdyż ich brak nieznacznie wpłynie na ostateczne wyniki analizy, stanowiąc jedynie 3,9% ogółu dostępnych danych.

Następnie przystąpiono do stworzenia tablic niezbędnych do wykonania analiz. W tym celu utworzono kod DataStep widoczny na rysunku 3-13.

```
data projekt.ZestawPrzewidywanaDekada;
set projekt.ZestawPolaczone;
where Topic = 'Life expectancy at birth, total (years)';
if nmiss(of _numeric_) or cmiss(of _character_) then delete ;
run;

proc sort data=projekt.PrzewidywanaDekada out=Sortowanie;
by Region;
run;
proc means data=Sortowanie noprint;
by Region;
where Topic = 'Life expectancy at birth, total (years)';
var YR2021 YR2020 YR2019 YR2018 YR2017 YR2016
    YR2015 YR2014 YR2013 YR2012 YR2011;
output out=Output(drop=_) mean=;
run;
data projekt.PrzewidywanaDekadaRegiony;
set Output;
if cmiss(of _character_) then delete ;
run;

proc sort data=projekt.PrzewidywanaDekada out=Sortowanie2;
by Income_Group;
run;
proc means data=Sortowanie2 noprint;
by Income_Group;
where Topic = 'Life expectancy at birth, total (years)';
var YR2021 YR2020 YR2019 YR2018 YR2017 YR2016
    YR2015 YR2014 YR2013 YR2012 YR2011;
output out=Output2(drop=_) mean=;
run;
data projekt.PrzewidywanaDekadaDochody;
set Output2;
if cmiss(of _character_) then delete ;
run;
```

Rysunek 3-13 Kod odpowiedzialny za utworzenie tablic niezbędnych do przeprowadzenia analizy drugiej

Powstały w ten sposób trzy tabele wynikowe. Pierwsza składała się z kolumn CountryName, Alpha\_3, Topic, YR2011-YR2021, Income\_Group, Alpha 2 i Region. Druga zawierała jedynie Region oraz Srednia\_2021-Srednia\_2011, w których znajdowały się średnie dla regionów w danych latach. Trzecia grupę dochodów i średnią przewidywaną długość życia w niej.

Obs.	Region	YR2021	YR2020	YR2019	YR2018	YR2017	YR2016	YR2015	YR2014	YR2013	YR2012	YR2011
1	Central Asia	70.77	70.0364	71.1986	70.9696	70.7962	70.488443902	70.236843902	69.951887804	69.687487804	69.320487804	68.931287804
2	Eastern Asia	78.088487806	78.295765854	78.114234146	77.762360976	77.437917074	77.229697562	77.015458536	76.80315122	76.391736586	75.968921952	75.495902438
3	Eastern Europe	72.685741464	74.022029269	75.191207317	74.919785366	74.871673171	74.700065853	74.292512194	74.231002439	74.038607318	73.684714634	73.419736586
4	Latin America and the Caribbean	72.968782317	73.943112805	74.752407317	74.797837805	74.502632317	74.558282317	74.428326829	74.339432317	74.157126829	74.052076829	73.850776829

Rysunek 3-14 Pierwsze 4. rekordy tabeli PROJEKT.ZESTAWPRZEWIDYWANADEKADA

Obs.	Region	YR2021	YR2020	YR2019	YR2018	YR2017	YR2016	YR2015	YR2014	YR2013	YR2012	YR2011
1	Central Asia	70.77	70.0364	71.1986	70.9696	70.7962	70.488443902	70.236843902	69.951887804	69.687487804	69.320487804	68.931287804
2	Eastern Asia	78.088487806	78.295765854	78.114234146	77.762360976	77.437917074	77.229697562	77.015458536	76.80315122	76.391736586	75.968921952	75.495902438
3	Eastern Europe	72.685741464	74.022029269	75.191207317	74.919785366	74.871673171	74.700065853	74.292512194	74.231002439	74.038607318	73.684714634	73.419736586
4	Latin America and the Caribbean	72.968782317	73.943112805	74.752407317	74.797837805	74.502632317	74.558282317	74.428326829	74.339432317	74.157126829	74.052076829	73.850776829

Rysunek 3-15 Pierwsze 4. rekordy tabeli PROJEKT.PRZEWIDYWANADEKADAREGIONY

Obs.	Income_Group	YR2021	YR2020	YR2019	YR2018	YR2017	YR2016	YR2015	YR2014	YR2013	YR2012	YR2011
1	High income	78.555195122	78.956888407	79.430304043	79.25238423	79.130382893	79.050491814	78.822942198	78.823419645	78.525871032	78.377931841	78.180634147
2	Low income	61.724307692	62.106076923	62.564769231	62.096423077	61.707461538	61.300807692	60.776576923	60.3095	60.022192308	59.695615385	59.350192308
3	Lower middle income	67.278607498	67.838928184	68.291394761	68.074901536	67.818332882	67.491177055	67.098569106	66.785241192	66.417795393	66.01038121	65.584642728
4	Upper middle income	71.152514071	72.162893996	73.237014071	73.115193246	72.857750938	72.637084428	72.397762195	72.165132739	72.026849437	71.813290807	71.49006379

Rysunek 3-16 Pierwsze 4. rekordy tabeli PROJEKT.PRZEWIDYWANADEKADADOCHODY

Z przygotowanymi danymi można było przejść do pierwszej analizy. Celem jej został utworzony kod DataStep widoczny na rysunku 3-17. Celem tego kodu było dodanie do tabeli nowej kolumny o nazwie "Zmiana". Wartości w tej kolumnie informowały, czy na przestrzeni dekady w danym kraju nastąpił wzrost przewidywanej długości życia. Kolumna ta przyjmie wartości "Urosła" lub "Zmalała" w zależności od obserwowanych zmian. Został również wyświetlony wykres kołowy widoczny na rysunku ilustrujący te zmiany odsetkowo.

```

data projekt.PrzewidywanaDekadaZmiana;
set projekt.przewidywanadekada;

if YR2021<YR2011 then Zmiana = 'Zmalała';
if YR2021>YR2011 then Zmiana = 'Wzrosła';
run;

proc template;
define statgraph SASStudio.Pie;
begingraph;
entrytitle "Zmiana przewidywanej długości życia na świecie" /
textattrs=(size=14);
entryfootnote halign=left "W 2021 roku względem 2011 roku" /
textattrs=(size=12);
layout region;
piechart category=Zmiana / stat=pct;
endlayout;
endgraph;
end;
run;

ods graphics / reset width=6.4in height=4.8in imagemap;

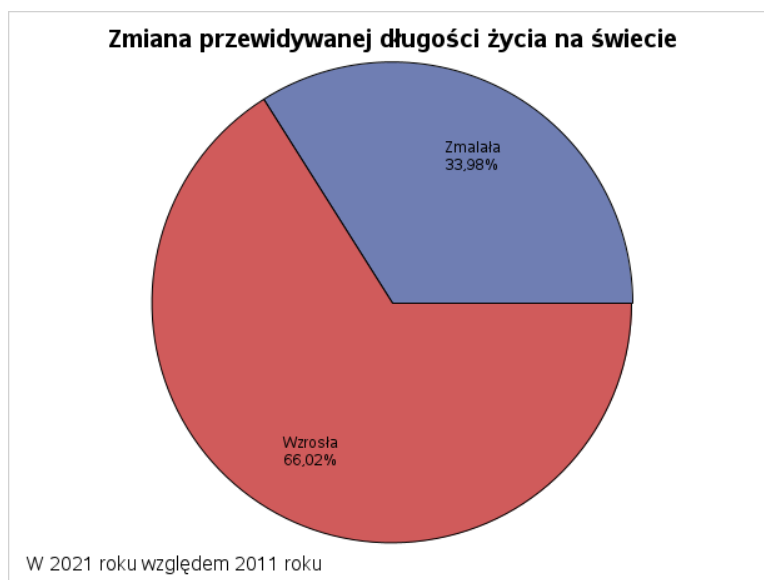
proc sgrender template=SASStudio.Pie data=PROJEKT.PRZEWIDYWANADEKADAZMIANA;
run;

ods graphics / reset;

```

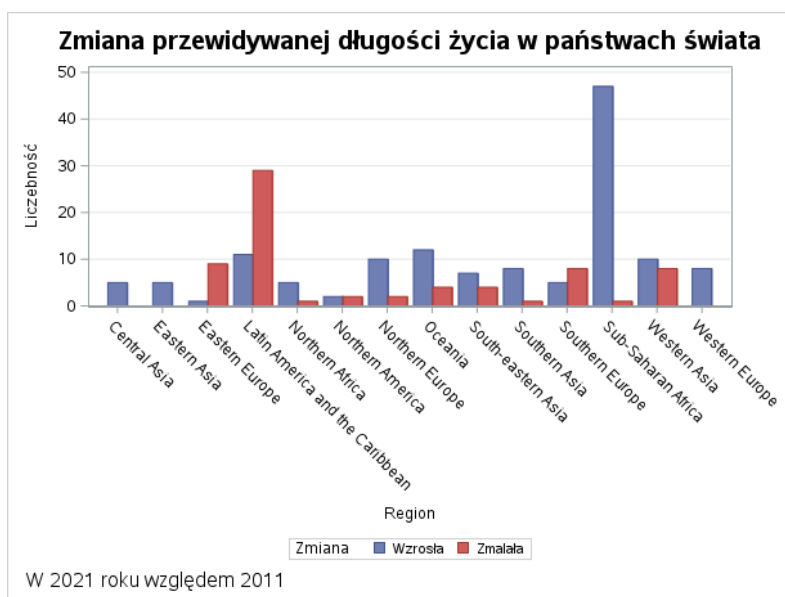
Rysunek 3-17 Kod odpowiedzialny za dodanie kolumny oraz wygenerowanie wykresu

Rozpoznano, że w 66,02% badanych krajów przewidywana długość życia wynosiła więcej w 2021 roku niż w 2011. W przypadku 33,98% zmalała.



*Rysunek 3-18 Wykres kołowy przedstawiający zmianę przewidywanej długości życia na świecie*

W następnym kroku zbadano, w którym regionie wskaźnik ten zmalał oraz wzrósł dla największej ilości państw. W tym wygenerowano wykres słupkowy widoczny na rysunku 3-19 przedstawiający ilość państw, w których przewidywana długość życia zmalała i wzrosła.



*Rysunek 3-19 Wykres słupkowy przedstawiające zmianę ilościowo w regionie*

Największą ilość państw, które doświadczyły spadku zauważono w regionie Ameryka Łacińska i Karaiby. Największy wzrost zanotowano w Afryce Subsaharyjskiej. W następnym kroku sprawdzono dokładnie, jak rozkłada się zmiana w tym regionie. W tym celu utworzono mapę widoczną na rysunku 3-21, za pomocą kodu znajdującego się na rysunku 3-20.



```

1 data projekt.PrzewidywanaDekadaZmianaKaraiby;
2 set projekt.PrzewidywanaDekadaZmiana ( keep = Country_Name Zmiana Region Alpha_2);
3 length Country_Name $40.;
4 where Region = 'Latin America and the Caribbean';
5 run;
6 data Mapka2;
7 set projekt.PrzewidywanaDekadaZmianaKaraiby(rename=(Alpha_2=ISOALPHA2));
8 format Alpha_2 $100.;
9 run;
10
11 proc gmap data=Mapka2 map=mapsghk.world;
12 id ISOALPHA2;
13 choro Zmiana / discrete;
14 title 'Obszary, gdzie przewidywana długość życia urosła lub zmalała';
15 where Zmiana in ('Zmalała', 'Wzrosła');
16 run;

```

Rysunek 3-20 Kod odpowiedzialny za wygenerowanie mapy

W badaniu zaobserwowano, że w większości krajów regionu Ameryka Łacińska i Karaiby prognozowana długość życia uległa zmniejszeniu w porównaniu do roku 2011. Jedynie w wybranych państwach, takich jak Republika Dominikany, Barbados, Antigua i Barbuda, Puerto Rico, Saint Kitts i Nevis, Dominika, Nikaragua, Haiti oraz Surinam, zaobserwowano wzrost tego wskaźnika. Dodatkowo, stwierdzono zjawisko zwiększenia prognozowanej długości życia na terytoriach zależnych od USA i Francji, takich jak Wyspy Dziewicze i Francuski Saint Martin.

Obszary, gdzie przewidywana długość życia urosła lub zmalała



Rysunek 3-21 Mapa przedstawiająca, w których krajach wskaźnik wzrósł, a w których zmalał

Kolejnym krokiem było zbadanie jak na przestrzeni dekady plasowała się przewidywana długość życia na świecie z podziałem na grupy dochodów państw. W tym celu wykorzystano kod widoczny na rysunku 3-22.



```

proc means data=projekt.przewidywanadekada noprint;
  var YR2021 YR2020 YR2019 YR2018 YR2017 YR2016
      YR2015 YR2014 YR2013 YR2012 YR2011;
  output out=swiatsrednia(drop=_) mean=;
run;
data dodanakolumna;
  set swiatsrednia;

  Income_Group = "Swiat";

run;
proc transpose data=dodanakolumna
  out=PROJEKT.PRZEWIDYWANADEKADASREDNIASWIAT
  (rename=(Region = Income_Group _Name_=Rok COL1=Przewidywana));
  by Income_Group;
run;

proc transpose data=projekt.PrzewidywanaDekadaDochody
  out=PROJEKT.PRZEWIDYWANADEKADASREDNIADOCHODY
  (rename=(_Name_=Rok COL1=Przewidywana));
  by Income_Group;
run;

data PROJEKT.PRZEWIDYWANADEKADASREDNIALINIOWY;
set PROJEKT.PRZEWIDYWANADEKADASREDNIADOCHODY
  PROJEKT.PRZEWIDYWANADEKADASREDNIASWIAT;
run;

proc sgplot data=PROJEKT.PRZEWIDYWANADEKADASREDNIALINIOWY;
  title height=14pt
  "Średnia przewidywana długość życia w momencie narodzin na świecie";
  footnote2 justify=left height=12pt "w latach";

  vline Rok / response=Przewidywana group=Income_Group
  lineattrs=(thickness=2);

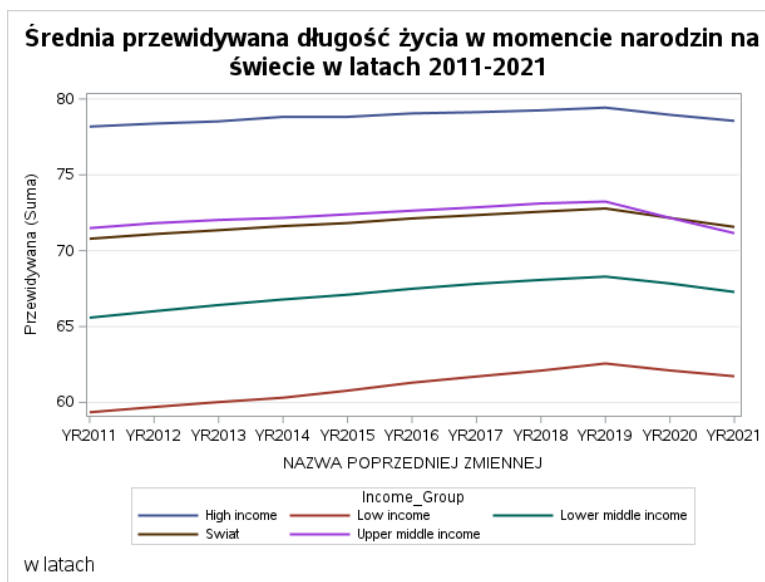
  yaxis grid;
run;

```

*Rysunek 3-22 Kod odpowiedzialny za dostosowanie danych do wygenerowania wykresu oraz utworzenie go*

Kod ten odpowiada za utworzenie tabeli ze średnią wartością wskaźnika dla całego świata. Później dodaje do niej kolumnę o nazwie `Income_Group` i przypisuje jej wartość `Świat`. Następnie zarówno ta tabela jak i tabela średnich podzielonych względu na poziom dochodów państwa ulegają transpozycji. Tabele te są nakładane na siebie tworząc zestaw na podstawie, którego ostatnia procedura generuje wykres liniowy znajdujący się na rysunku 3-23.

Zauważono, że przewidywana długość życia wzrastała w podobnym tempie w każdej grupie dochodowej. Przewidywana długość życia zmalała w każdej z grup w 2020 i w 2021 dalej nie wróciła do poprzedniego poziomu. Grupa dochodowa średnio-wysokich zarobków w wyniku tego spadku znalazła się pod średnią światową.



Rysunek 3-23 Wykres przedstawiający wskaźnik na przestrzeni lat 2011-2021

Chcąc sprawdzić średnie tempo zmian dla regionów oraz czy istnieją różnice pomiędzy grupami dochodowymi utworzono kod widoczny na rysunku 3-24.

```
data x;
  set projekt.PrzewidywanaDekada;
  drop Country_Name Alpha_3 Topic Region Alpha_2;
  Zmiana = (YR2021/YR2011)** (1/8) -1;
  format Zmiana :percent10.2;
  keep Zmiana Income_Group;
run;

proc sgplot data=WORK.X;
  vbar Income_Group / response=Zmiana stat=mean;
  yaxis grid;
run;
```

Rysunek 3-24 Kod dodający kolumnę ze średnim tempem zmian i generujący wykres

Kod ten wytwarza tabelę, której kolumny zawierają informację o grupie dochodowej oraz średnim tempie zmian wyliczonym za pomocą wzoru:

$$\bar{i}_g = \sqrt[n-1]{\frac{y_n}{y_0}}$$

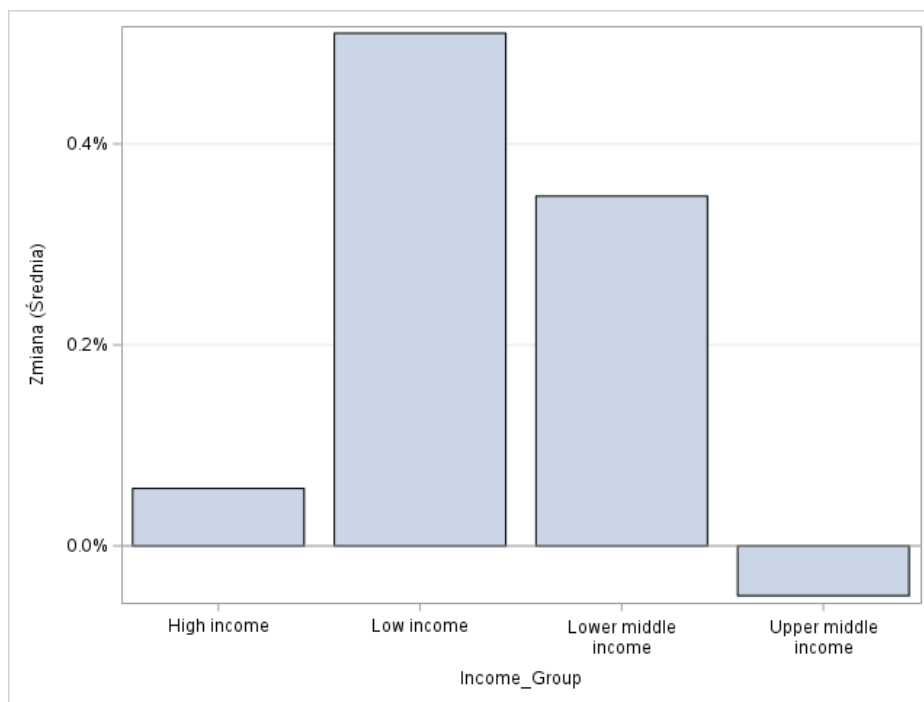
gdzie:

$y_n$  to wartość z ostatniego okresu

$y_0$  z pierwszego okresu

$n$  to liczba okresów

Na końcu generuje wykres widoczny na rysunku 3-25.



*Rysunek 3-25 Wykres przedstawiający średnie tempo zmian w regionach*

Wykres pozwolił zauważyć, że ujemne średnie tempo wzrostu cechuje państwa o dochodach średnio-wysokich, a największe dodatnie państwa o niskich dochodach.

Podsumowując, w okresie od 2011 do 2021 roku zaobserwowano wzrost przewidywanej długości życia w chwili narodzin w większości państw na świecie. W prawie każdym regionie w większości państw ten wskaźnik wzrósł. Niemniej jednak, w Ameryce Łacińskiej i Karaibach oraz w Południowej Europie, większość krajów doświadczyła regresu tego wskaźnika. Szczególnie w kontekście Ameryki Łacińskiej cały ten obszar zanotował spadek wartości.

Ciekawym aspektem obserwowanego zjawiska było spostrzeżenie, że każda grupa dochodowa państw wykazywała podobny wzrost do roku 2020. Wtedy zarówno na poziomie poszczególnych regionów, jak i globalnie, wskaźnik długości życia uległ redukcji. W rezultacie grupa obejmująca państwa o średnio-wysokich dochodach po raz pierwszy znalazła się poniżej światowej średniej tego wskaźnika.

Analiza tempa zmian w długości życia dla różnych regionów ujawniła, że państwa o wysokich dochodach rosły średnio wolniej niż państwa o niskich dochodach. Za to przewidywana długość życia posiada ujemne średnie tempo zmian w krajach o dochodach średnio-wysokich.

### 3.3 Różnice w przewidywanej długości życia pomiędzy płciami.

Celem szczegółowym analizy było zbadanie istniejących różnic w przewidywanej długości życia w chwili narodzin między płciami. Aby osiągnąć ten cel, przeprowadzono porównanie między państwami, mając na uwadze, w których krajach wskaźnik ten dla kobiet przewyższa wartość dla mężczyzn w 2021 roku. Następnie zbadano przeciętną różnicę między płciami w okresie 2011-2012 w celu zbadania czy płcie zmniejszają dystans między sobą. Ostatecznie zbadano związek pomiędzy grupą dochodową państwa, a różnicą między płciami w 2021 roku.

Tak jak w poprzednich analizach przed rozpoczęciem obliczeń stworzono tabele odpowiedzialną za wyświetlenie informacji na temat brakujących rekordów.

Country_Name	Alpha_3	Topic	YR2011	YR2012	YR2013	YR2014	YR2015	YR2016	YR2017	YR2018	YR2019	YR2020	YR2021
Andorra	AND	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
Andorra	AND	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.
American Samoa	ASM	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
American Samoa	ASM	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.
Curacao	CUW	Life expectancy at birth, female (years)	80.7	80.7	80.9	81.1	81.2	81.1	81.5	.	.	.	.
Curacao	CUW	Life expectancy at birth, male (years)	74.4	74.5	74.9	75	75.1	74.5	74.7	.	.	.	.
Cayman Islands	CYM	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
Cayman Islands	CYM	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.
Monaco	MCO	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
Monaco	MCO	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.
Northern Mariana Islands	MNP	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
Northern Mariana Islands	MNP	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.
Palau	PLW	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
Palau	PLW	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.
San Marino	SMR	Life expectancy at birth, female (years)	.	.	.	.	.	.	.	.	.	.	.
San Marino	SMR	Life expectancy at birth, male (years)	.	.	.	.	.	.	.	.	.	.	.

*Rysunek 3-26 Rekordy, w których brakuje danych*

Rozpoznano braki danych w tych samych 8 terytoriach co w analizie 2. W związku z tym, zdecydowano się na usunięcie tych rekordów w następnym kroku, gdyż ich brak nieznacznie wpłynie na ostateczne wyniki analizy, stanowiąc jedynie 3,9% ogółu dostępnych danych.

Następnie wykorzystano kod widoczny na rysunku 3-27 w celu utworzenia tabel niezbędnych do dalszych analiz. Kod ten tworzy trzy tabele wynikowe, w każdej dodając kolumnę „Plec” odpowiadającej płci jakiej dotyczą wartości. Pierwsza z nich zawierała dane przewidywanej długości życia mężczyzn oraz kobiet w okresie 2011-2021. Druga jedynie w latach 2011 oraz 2021. Trzecia posiadała informacje o średniej długości życia w regionach z podziałem na płcie.

```

DATA projekt.PrzewidywanaPlcie2011_2021;
  SET projekt.zestawpolaczone(keep = Country_Name Topic YR2011 YR2021) ;
  WHERE Topic = "Life expectancy at birth, male (years)" OR Topic = "Life expectancy at birth, female (years)";

  IF Topic = "Life expectancy at birth, male (years)" THEN Plec = 'M';
  IF Topic = "Life expectancy at birth, female (years)" THEN Plec = 'K';
  DROP Topic;
  if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
RUN;

DATA projekt.PrzewidywanaPlcie;
  SET projekt.zestawpolaczone;
  WHERE Topic = "Life expectancy at birth, male (years)" OR Topic = "Life expectancy at birth, female (years)";

  IF Topic = "Life expectancy at birth, male (years)" THEN Plec = 'M';
  IF Topic = "Life expectancy at birth, female (years)" THEN Plec = 'K';
  DROP Kontynent alpha_3 Topic;
  if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
RUN;

proc sort data=projekt.PrzewidywanaPlcie out=Sortowanie;
  by Region;
run;
proc means data=Sortowanie noprint;
  by Region;
  var YR2021 YR2020 YR2019 YR2018 YR2017 YR2016
      YR2015 YR2014 YR2013 YR2012 YR2011;
  output out=projekt.PrzewidywanaPlcieRegiony(drop=:) mean=;
run;

```

*Rysunek 3-27 Kod generujący niezbędne tabele*

W następnej kolejności stworzono dwie tabele za pomocą kodu widocznego na rysunku 3-28. Powstały dwie tabele, które podzieliły PRZEWIDYWANAPLCIE2011\_2021 ze względu na płeć, a potem je złączyły. W ten sposób powstała tabela, w której podział na płeć odbywał się kolumnowo, a nie wierszowo.

```

data PrzewidywanaPlecM2021;
  set projekt.przewidywanaplcie2011_2021;
  where Plec = 'M';
  RENAME YR2021 = M2021;
  RENAME YR2011 = M2011;
  drop Plec;
run;

data PrzewidywanaPlecK2021;
  set projekt.przewidywanaplcie2011_2021;
  where Plec = 'K';
  RENAME YR2021 = K2021;
  RENAME YR2011 = K2011;
  drop Plec;
run;
proc sort data=PRZEWIDYWANAPLECK2021 out=x;
  by Country_Name;
run;

proc sort data=PRZEWIDYWANAPLECM2021 out=y;
  by Country_Name;
run;

data projekt.PLCIEKOLUMNY2011_2021;
  merge x y;
  by Country_Name;
  IF M2021 > K2021 THEN Dluzej2021 = 'M';
  ELSE Dluzej2021 = 'K';
  IF M2011 > K2011 THEN Dluzej2011 = 'M';
  ELSE Dluzej2011 = 'K';

run;

proc delete data=x y;
run;

```

*Rysunek 3-28 Kod odpowiedzialny za proces podzielenia i połączenia tabel*

Następnie wykorzystano tę tabelę za pomocą kodu 3-29 do stworzenia tablicy widocznej na rysunku 3-30

```

data test;
set projekt.PLCIEKOLUMNY2011_2021;

run;

proc freq data=projekt.PLCIEKOLUMNY2011_2021;
title 'Odsetki Dluzej2021 - Tabela 1';
tables Dluzej2021 / out=Odsetki (rename=(Dluzej2021=K_M));
run;
proc freq data=projekt.PLCIEKOLUMNY2011_2021;
title 'Odsetki Dluzej2011 - Tabela 1';
tables Dluzej2011 / out=Odsetki2 (rename=(Dluzej2011=K_M));
run;

```

Rysunek 3-29 Kod generujący tablicę

Odsetki Dluzej2021 - Tabela 1				
Procedura FREQ				
Dluzej2021	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
K	204	100.00	204	100.00

Odsetki Dluzej2011 - Tabela 1				
Procedura FREQ				
Dluzej2011	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
K	204	100.00	204	100.00

Rysunek 3-30 Wygenerowana tablica

Zauważono, że zarówno w 2011 roku jak i w 2021 100% badanych państw przewidywana długość życia w momencie narodzin wynosiła więcej dla kobiet niż dla mężczyzn.

W następnym kroku podjęto próbę zbadania różnic pomiędzy średnią przewidywaną długością życia podzielonej na płcie. W tym celu wykonano kod widoczny na rysunku 3-31. Kod ten grupuje wszystkie dane ze względu na płeć i liczy średnią przewidywaną długość życia dla każdej płci na świecie. Później dzieli tabelę ze względu na płcie w celu zmienienia nazwy kolumny Przewidywana na PrzewidywanaM oraz PrzewidywanaK. Łączy tę tabelę pozwalając dodać kolumnę „Różnica” zawierającą informację o różnicy pomiędzy średnią przewidywaną długością życia mężczyzn oraz kobiet na przestrzeni 10 lat. Następnie utworzono wykres liniowy widoczny na rysunku 3-32.

```

data x;
set projekt.przewidywanaplcie;
drop Country_Name Income_Group Alpha_2 Region;
run;
proc sort data=x out=Sortowanie;
by Plec;
run;
proc means data=Sortowanie noprint;
by Plec;
var YR2021 YR2020 YR2019 YR2018 YR2017 YR2016
    YR2015 YR2014 YR2013 YR2012 YR2011;
output out=Output(drop=_) mean=;
run;

proc transpose data=Output
out=projekt.SREDNIAPLCIESWIAT(rename=(COL1=Przewidywana _NAME_=Rok));
by Plec;
run;

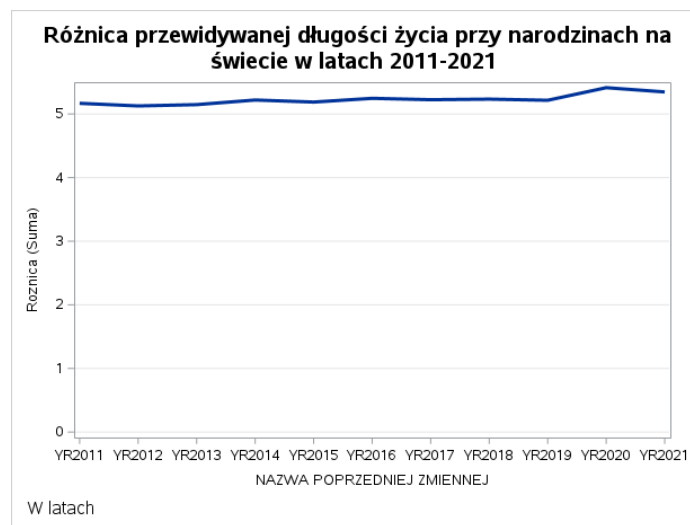
data K;
set projekt.SREDNIAPLCIESWIAT;
where plec = 'K';
drop Przewidywana Plec;
PrzewidywanaK = Przewidywana;
run;
data M;
set projekt.SREDNIAPLCIESWIAT;
where plec = 'M';
PrzewidywanaM = Przewidywana;
drop Przewidywana Plec;
run;
proc sort data=WORK.M out=work._tmpsort1_;
by Rok;
run;

proc sort data=WORK.K out=work._tmpsort2_;
by Rok;
run;

data projekt.SREDNIAPLCIESWIATROZNICE;
merge _tmpsort1_ _tmpsort2_;
by Rok;
Roznica = PrzewidywanaK - PrzewidywanaM;
run;

```

Rysunek 3-31 Kod odpowiedzialny za proces uzyskania kolumny "różnica"



Rysunek 3-32 Wykres przedstawiający różnicę wskaźnika w latach 2011-2021

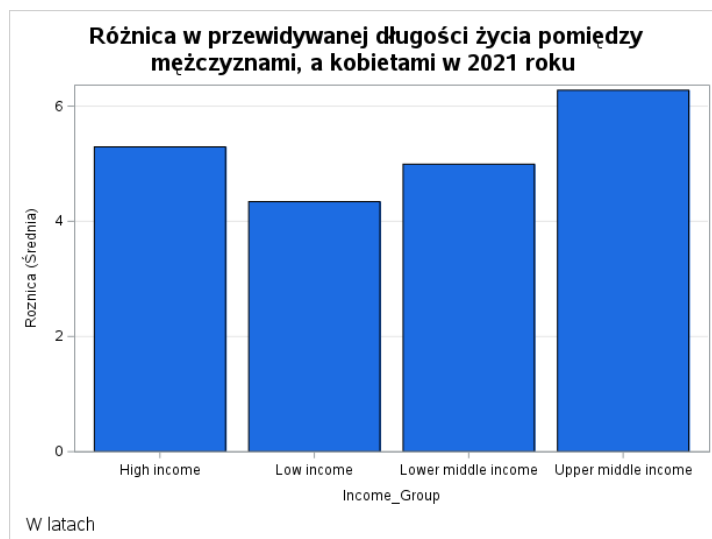
Zauważono, że przeciętna różnica między kobietami i mężczyznami nie zmieniała się znacząco. W kontekście całego świata nie występuje zjawisko nadganiania. W 2020 roku różnica delikatnie wzrosła.

W następnym kroku sprawdzono jak rozkłada się średnia różnica między kobietami, a mężczyznami ze względu na grupę dochodową państwa w 2021 roku. W tym celu utworzono dwie tabele tak jak na rysunku 3-31 z podziałem na płcie w celu ostatecznego połączenia jej w jedną z możliwością obliczenia różnic w kolumnach. Ostateczny wygląd tabeli wynikowej znajduje się na rysunku 3-33.

Obs.	Income_Group	Roznica
1	High income	5.517
2	High income	3.721
3	High income	5.166
4	High income	4.100
5	High income	5.000

Rysunek 3-33 Pierwsze 5. rekordów tabeli wynikowej

Następnie utworzono wykres słupkowy obrazujący tę różnicę.



Rysunek 3-34 Wykres przedstawiający średnią różnicę w grupach dochodowych

Zauważono, że różnice odbiegają od siebie ze względu na grupę dochodową państwa. Najmniejsze różnice występują w krajach niskiego dochodu, a najwyższe w krajach o dochodzie średnio-wysokim. Postanowiono sprawdzić czy istnieje korelacja. W tym celu dodano do tabeli kolumnę odpowiedzialną za GNI per Capita dla roku 2021 z ZESTAWPOŁACZONE i



zastosowano procedurę CORR. Kod odpowiedzialny, za tę czynność znajduje się na rysunku 3-35.

```
data projekt.GNI2021;
set projekt.zestawpolaczone (keep = Country_Name Topic YR2021 );
where topic = 'GNI per capita, Atlas method (current US$)';
if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
drop topic;
rename YR2021 = GNI_per_capita;
run;
proc sort data=PROJEKT.GNI2021 out=work._tmpsort1_;
by Country_Name;
run;

proc sort data=PROJEKT.ROZNICASWIATPLCIE out=work._tmpsort2_;
by Country_Name;
run;

data GNI_Roznica2021;
merge _tmpsort1_ _tmpsort2_;
by Country_Name;
run;

proc delete data=work._tmpsort1_ work._tmpsort2_;
run;
PROC CORR
data = GNI_Roznica2021;
run;
ods graphics off;
VAR Income Group Roznica;
```

Rysunek 3-35 Kod przystosowujący tablicę oraz wywołujący procedurę CORR

W wyniku jej działania powstały tablice oraz wykres widoczne na rysunku 3-36. Wartość współczynnika korelacji Pearsona wynosi -0,10656, ale zależność nie jest statystycznie istotna.

Współczynniki korelacji Pearsona Prawd. >  r  przy H0: rho=0 Liczba obserwacji		
	GNI_per_capita	Roznica
GNI_per_capita	1.00000 194	-0.10410 0.1540 189
Roznica	-0.10410 0.1540 189	1.00000 202

Rysunek 3-36 Wygenerowana macierz prezentująca współczynnik korelacji Pearsona

Podsumowując, w analizowanym w badanych terytoriach obserwowano, że przewidywana długość życia w chwili narodzin była wyższa dla kobiecej populacji we wszystkich badanych krajach. W przypadku płci męskiej, przewidywana długość życia w chwili narodzin utrzymuje

się na poziomie istotnie niższym niż dla płci żeńskiej. Różnice te nie wykazują znaczących zmian w badanym okresie, co sugeruje pewną stabilność tego zjawiska w analizowanym czasie.

Warto zaznaczyć, że różnica w długości życia między płciami nie wydaje się być istotnie związana z poziomem dochodowym badanych państw.

### 3.4 Wydatki na zdrowie na świecie oraz ich zależność z przewidywaną długością życia

Celem szczegółowym analizy jest dokładne zbadanie charakterystyki wydatków na zdrowie na świecie. W pierwszym etapie zidentyfikowano odsetek krajów, w których wydatki wzrosły lub zmalały na przestrzeni lat. Następnie sprawdzono różnice na poziomie regionów. Ostatnim etapem analizy było sprawdzenie ewentualnej zależności pomiędzy wzrostem wydatków na zdrowie, a przewidywaną długością życia w Niemczech w latach 2000-2011.

Przed przystąpieniem do analizy należało przygotować dane. W tym celu sprawdzono za pomocą kodu użytego w poprzednich analizach w jakich rekordach, o temacie przewidywanej długości życia lub wydatków na służbę zdrowia brakuje co najmniej jednej komórki. Okazało się, że w 198 z 424 (46,69%) rekordach brakuje danych. Jednak zdecydowana większość dotyczy się kolumny opisującej 2021 rok. W takim wypadku sprawdzono, czy jeżeli usunięta zostanie ta kolumna, a zakres czasowy ograniczony do 2020 roku, to problem zostanie wyeliminowany.

Po wyeliminowaniu 2021 roku, rekordów, w których brakuje danych pozostało jedynie 36 (8,49%). Były to głównie wyspy Karaibach i Oceanii. Wiedząc, że można je usunąć przystąpiono do przygotowania tabel.

W następnym kroku stworzono dwie tabele. Kod generujący tabele znajduje się na rysunku 3-37.

```
data projekt.Wydatki2011_2020;
set projekt.zestawpolaczone(keep = Country_Name Region YR2011 YR2020 Income_Group Topic);
where Topic = 'Current health expenditure per capita (current US$)';
drop Topic;
if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
run;

data Wydatki2;
set projekt.zestawpolaczone(keep = Country_Name Region YR2011 YR2020 Income_Group Topic);
where Topic = 'Current health expenditure per capita (current US$)';
drop Topic;
if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
run;

proc sort data=Wydatki2 out=Sortowanie;
by Region;
run;
proc means data=Sortowanie noprint;
by Region;
var YR2020 YR2011;
output out=projekt.WydatkiRegiony(drop=_) mean=;
run;
```

Rysunek 3-37 Kod generujący niezbędne tabele

Pierwsza i druga znajdujące się na rysunkach 3-38 i 3-39 zawierają dane z lat 2020 oraz 2011 roku. W przypadku PROJEKT.WYDATKI2011\_2022 jest to nazwa państwa, wydatki zdrowotne, grupa dochodowa oraz region. PROJEKT.WYDATKIREGIONY zawiera uśrednione wydatki zdrowotne w regionach.

Obs.	Country_Name	YR2011	YR2020	Income_Group	Region
1	Afghanistan	52.36	80.29	Low income	Southern Asia
2	Angola	121.91	50.74	Lower middle income	Sub-Saharan Africa
3	Andorra	3956.99	3336.92	High income	Southern Europe
4	United Arab Emirates	1500.80	2191.81	High income	Western Asia
5	Argentina	1202.56	863.71	Upper middle income	Latin America and the Caribbean

Rysunek 3-38 Pierwsze 5. rekordów z tabeli PROJEKT.WYDATKI2011\_2022

Obs.	Region	YR2020	YR2011
1	Central Asia	215.83	160.33
2	Eastern Asia	1953.44	1753.55
3	Eastern Europe	912.73	723.47
4	Latin America and the Caribbean	655.87	548.00
5	Northern Africa	159.65	193.48

Rysunek 3-39 Pierwsze 5. rekordów z tabeli PROJEKT.WYDATKIREGIONY

Z tak przygotowanymi danymi przystąpiono do pierwszego procesu. Było nim zbadanie, w ilu krajach wydatki na zdrowie wzrosły, a w ilu zmalały. W tym celu wykonano kod widoczny na rysunku 3-40.

```
data wydatkiStatus;
set projekt.wydatki2011_2020;
if YR2011>YR2020 then Status = 'Wzrosły';
else Status = 'Zmalały';
run;
proc print data = wydatkiStatus;
```

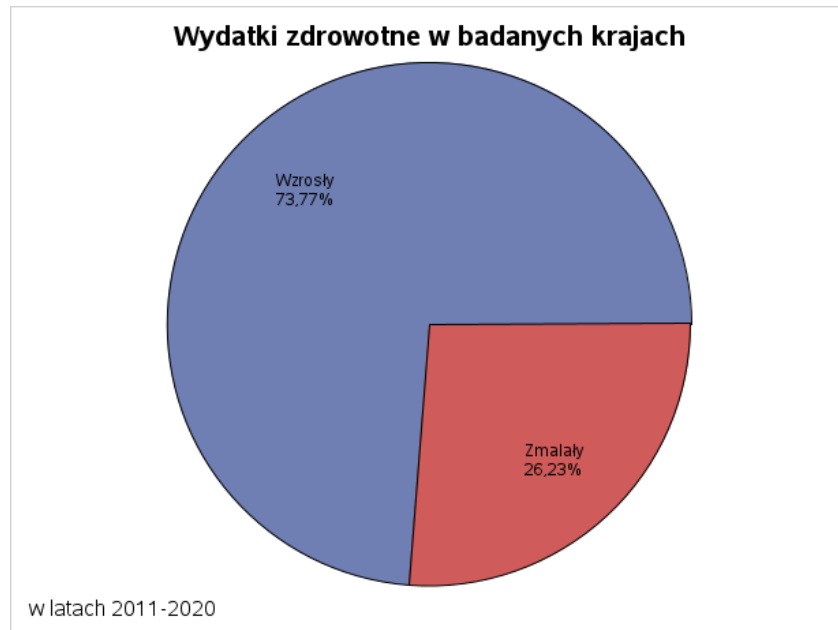
Rysunek 3-40 Kod dodający kolumnę "status" oraz wyświetlający tabelę

W efekcie powstała tabela widoczna na rysunku 3-41.

Obs.	Country_Name	YR2011	YR2020	Income_Group	Region	Status
1	Afghanistan	52.36	80.29	Low income	Southern Asia	Zmalały
2	Angola	121.91	50.74	Lower middle income	Sub-Saharan Africa	Wzrosły
3	Andorra	3956.99	3336.92	High income	Southern Europe	Wzrosły
4	United Arab Emirates	1500.80	2191.81	High income	Western Asia	Zmalały
5	Argentina	1202.56	863.71	Upper middle income	Latin America and the Caribbean	Wzrosły

Rysunek 3-41 Wygenerowana tabela

Następnie w celu sprawdzenia w ilu krajach wzrosła, a w ilu zmalała utworzono wykres słupkowy znajdujący się na rysunku 3-42.



Rysunek 3-42 Wykres słupkowy przedstawiający w ilu krajach wydatki zdrowotne wzrosły, a w ilu zmalały

Zauważono, że w zdecydowanej większości krajów wydatki zdrowotne na przestrzeni lat wzrosły. Postanowiono sprawdzić o ile średnio w regionach wydatki te uległy zmianie. W tym celu, za pomocą kodu widocznego na rysunku 3-43 utworzono dodatkową kolumnę „Zmiana” w tabeli projekt. Kolumna ta wyliczała różnicę pomiędzy średnimi wydatkami w danych latach.

```
data regionywydatki;
set projekt.wydatkiregiony;
Zmiana = YR2020-YR2011;
run;

proc print data = regionywydatki;
```

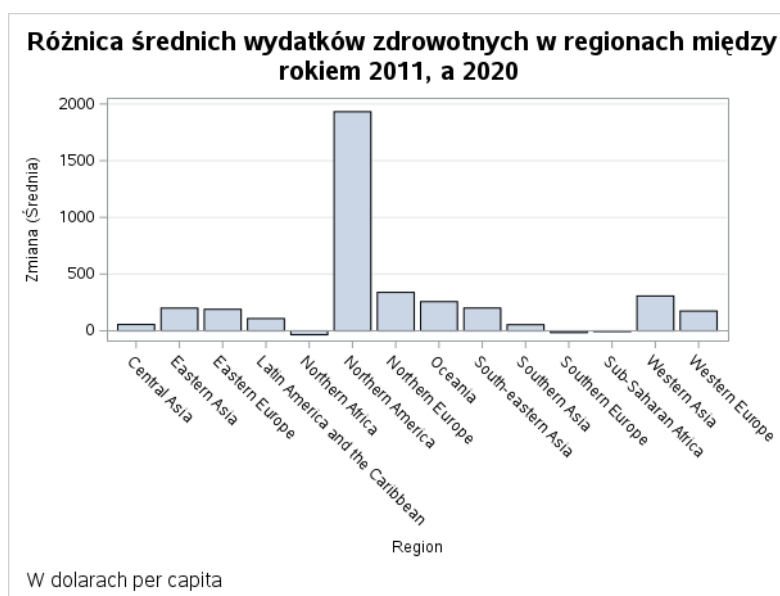
Rysunek 3-43 Kod dodający kolumnę "zmiana" oraz wyświetlający tabelę.

W efekcie powstała tabela widoczna na rysunku 3-44.

Obs.	Region	YR2020	YR2011	Zmiana
1	Central Asia	215.83	160.33	55.50
2	Eastern Asia	1953.44	1753.55	199.89
3	Eastern Europe	912.73	723.47	189.26
4	Latin America and the Caribbean	655.87	548.00	107.87
5	Northern Africa	159.65	193.48	-33.83
6	Northern America	8660.91	6729.15	1931.77
7	Northern Europe	4617.61	4279.09	338.51
8	Oceania	1228.18	971.20	256.98
9	South-eastern Asia	522.93	322.45	200.48
10	Southern Asia	218.64	164.84	53.81
11	Southern Europe	2036.85	2051.69	-14.84
12	Sub-Saharan Africa	115.45	120.89	-5.44
13	Western Asia	1209.54	902.59	306.96
14	Western Europe	5911.50	5737.38	174.11

Rysunek 3-44 Tabela wynikowa

Następnie na podstawie tej tabeli utworzono wykres słupkowy widoczny na rysunku 3-45.



*Rysunek 3-45 Wykres przedstawiający różnicę średnich wydatków zdrowotnych w regionach*

Wydatki zdrowotne najsilniej urosły w Północnej Ameryce. W Południowej Europie i Afryce Subsaharyjskiej oraz Północnej spadły względem 2011 roku.

W ostatnim kroku tej analizy postanowiono sprawdzić czy istnieje zależność pomiędzy wydatkami na zdrowie w Niemczech. W tym celu wykorzystano tabelę PROJEKT.NIEMCY i utworzono jej transponowaną wersję za pomocą kodu na rysunku 3-46.

```

data Przewidywana1;
set PROJEKT.NIEMCY;
where Topic = 'Life expectancy at birth, total (years)';
run;
proc sort data = Przewidywana1 out = Przewidywana2;
by Country_Name Topic;
run;
proc transpose data = Przewidywana2
out = Przewidywana3(rename=( _Name_ = Rok COL1=Przewidywana));
by Country_Name Topic;
run;
data Wydatki1;
set PROJEKT.NIEMCY;
where Topic = 'Current health expenditure per capita (current US$)';
run;
proc sort data = Wydatki1 out = Wydatki2;
by Country_Name Topic;
run;
proc transpose data = Wydatki2
out = Wydatki3(rename=( _Name_ = Rok COL1=Wydatki));
by Country_Name Topic;
run;
proc sort data=WORK.WYDATKI3 out=work._tmpsort1_;
by Rok;
run;

proc sort data=WORK.PRZEWIDYWANA3 out=work._tmpsort2_;
by Rok;
run;

data work.combine;
merge _tmpsort1_ _tmpsort2_;
by Rok;
run;

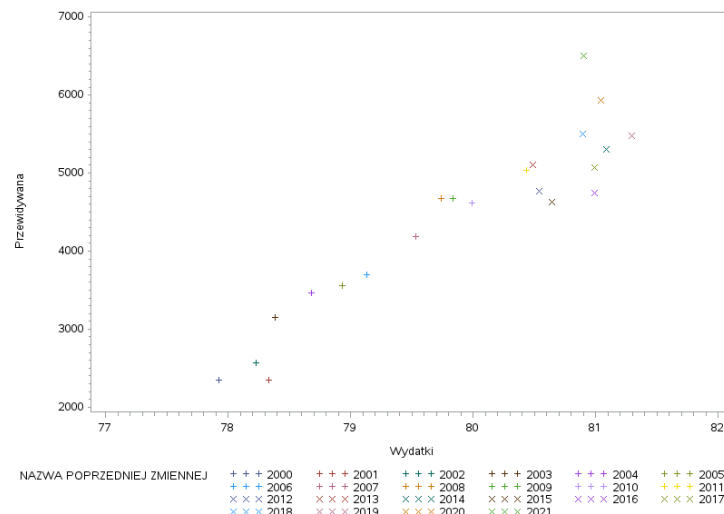
proc delete data=work._tmpsort1_ work._tmpsort2_;
run;
data Projekt.NiemcyZaleznosc;
set work.combine;
Rok = compress(Rok, 'YR');
run;

```

Rysunek 3-46 Kod odpowiadający za transpozycję tabeli PROJEKT.NIEMCY

Następnie stworzono wykres punktowy widoczny na rysunku 3-47.

Wydatki zdrowotne, a przewidywana długość życia w Niemczech w latach 2000-2021



Rysunek 3-47 Wykres zależności pomiędzy przewidywaną długością życia, a wydatkami zdrowotnymi w Niemczech w latach 2000-2021

Zauważono, że na wykresie punkty układają się wzdłuż linii prostej. Następnie, postanowiono sprawdzić siłę zależności. W tym celu wykorzystano PROC CORR za pomocą kodu widocznego na rysunku 3-48.

```
proc corr
data = Projekt.NiemcyZaleznosc;
run;
ods graphics off;
VAR Income_Group Roznica;
```

Rysunek 3-48 Kod wywołujący procedurę CORR

Wyniki tego kodu można zauważyć na rysunku XX. Otrzymaliśmy informację na temat średniej, odchylenia standardowego, sumy, minimum oraz maksimum dla zmiennych. Najważniejsze z punktu widzenia tej analizy jest określić zależność. Współczynnik korelacji wynosi 0,93528, czyli jest to silna korelacja. Można zauważyć również, że jest istotna statystycznie.

Statystyki proste						
Zmienna	N	Średnia	Odch. std.	Suma	Minimum	Maksimum
Wydatki	22	4424	1140	97325	2344	6503
Przewidywana	22	79.90969	1.09696	1758	77.92683	81.29268

Współczynniki korelacji Pearsona, N = 22 Prawd. >  r  przy H0: rho=0		
	Wydatki	Przewidywana
Wydatki	1.00000	0.93528 <.0001
Przewidywana	0.93528 <.0001	1.00000

Rysunek 3-49 Wygenerowana tablica

Podsumowując, odnotowano wzrost wydatków w 73,77% badanych państw. Najbardziej znaczący wzrost średnich wydatków na jednego mieszkańca odnotowano w Ameryce Północnej, gdzie wyniósł imponujące 1931,77\$. W tym kontekście istotne jest zauważenie, że w innych obszarach, takich jak Afryka Subsaharyjska, Północna Afryka oraz Europa Południowa, wydatki per capita uległy obniżeniu.

Przy analizie konkretnego przypadku Niemiec udało się potwierdzić istnienie silnej oraz istotnej statystycznie zależności pomiędzy wydatkami zdrowotnymi a przewidywaną długością życia. Wyniki te sugerują, że wraz ze wzrostem wydatków zdrowotnych w państwie, poprawiają się czynniki wpływające na wskaźnik przewidywanej długości życia w momencie narodzin.

### 3.5 Stan powietrza na świecie oraz zależność z przewidywaną długością życia w 2019 roku

Pył zawieszony PM2.5 to kategoria bardzo drobnych cząstek o średnicy mniejszej niż 2,5 mikrometra, unoszących się w atmosferze. Te mikroskopijne cząstki mają zdolność łatwego przenikania do dróg oddechowych i dostawania się do krwiobiegu. Źródła pyłów PM2.5 obejmują emisje z spalania paliw kopalnych, pojazdów, działalność przemysłową oraz pożary leśne. Wysokie stężenia pyłów PM2.5 stanowią zagrożenie dla zdrowia, zwłaszcza dla układu oddechowego, mogą powodować podrażnienia dróg oddechowych i zwiększać ryzyko chorób tego układu.

Celem szczegółowym analizy jest zbadanie potencjalnej zależności między dochodów narodowymi, a średnią ekspozycją populacji na pyły zawieszone PM2.5 w badanych państwach. W pierwszym etapie zidentyfikowano kraje oraz grupy dochodowe charakteryzujące się najwyższą ekspozycją populacji na PM2.5, w celu zrozumienia rozkładu tych emisji w zależności od poziomu dochodów. Zbadano również, który region cechuje się największym zróżnicowaniem. Kolejnym krokiem było sprawdzenie, w ilu państwach w regionie przekraczana była zalecana przez WHO norma (przed jej zmianą w 2021) przewidująca średnio 10 mikrogramów na metr sześcienny rocznie<sup>1</sup>. W ostatnim etapie analizy przeprowadzono analizę korelacji, aby określić, czy istnieje istotna statystyczna zależność między poziomem dochodów a ekspozycją populacji na PM2.5 na świecie w 2019 roku.

Przed przystąpieniem do analizy należało przygotować dane. W tym celu sprawdzono za pomocą kodu użytego w poprzednich analizach w jakich rekordach brakuje danych dla 2019 roku oraz pozbyto się tych rekordów ze względu na możliwość wpływania na analizę.

Następnie przystąpiono do wygenerowania niezbędnych tablic. W tym celu wykorzystano znajdujący się na rysunku 3-50.

---

<sup>1</sup> <https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines>



```

data Przewidywana2019;
set projekt.zestawpolaczone;
where Topic = 'Life expectancy at birth, total (years)';
drop Alpha_3 Kontynent Region Alpha_2 Income_Group Topic YR2021;
run;

data Zanieczyszczenia2019;
set projekt.zestawpolaczone;
where Topic = 'PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)';
drop Alpha_3 Kontynent Region Alpha_2 Income_Group Topic YR2021;
run;

proc sort data = Zanieczyszczenia2019 out =Zanieczyszczenia2019Sort;
by Country_Name;
run;

proc sort data = Przewidywana2019 out =Przewidywana2019Sort;
by Country_Name;
run;

proc transpose data = Zanieczyszczenia2019Sort out =Zanieczyszczenia2019Transpo
(Rename= (_NAME_ = Rok COL1= Zanieczyszczenia));
by Country_Name;
run;

proc transpose data = Przewidywana2019Sort out =Przewidywana2019Transponowane
(Rename= (_NAME_ = Rok COL1= Przewidywane));
by Country_Name;
run;
data Przewidywana2019Gotowe;
set Przewidywana2019Transponowane;
Rok = compress(Rok, 'YR');
run;
data Zanieczyszczenia2019Gotowe;
set Zanieczyszczenia2019Transpo;
Rok = compress(Rok, 'YR');
run;
proc sort data=Zanieczyszczenia2019Gotowe out=work._tmpsort1_;
by Country_Name Rok;
run;

proc sort data=Przewidywana2019Gotowe out=work._tmpsort2_;
by Country_Name Rok;
run;

data projekt.PrzewidywanaZanie2019;
merge _tmpsort1_ _tmpsort2_;
by Country_Name Rok;
if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
run;

data projekt.Zanieczyszczenia2019;
set projekt.zestawpolaczone(keep = YR2019 Country_Name Region Income_Group Topic);
where Topic = 'PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)';
if nmiss(of _numeric_ ) or cmiss(of _character_) then delete ;
run;

```

Rysunek 3-50 Kod odpowiadający za wygenerowanie tabel

Kod ten odpowiada za utworzenie dwóch tablic. Jedna z nich, znajdująca się na rysunku 3-51, jest połączeniem dwóch zestawów zawierających dane dla przewidywanej długości życia. W wyniku działania kodu ulega transponowaniu. Druga, znajdująca się na rysunku 3-52, posiada informację na temat kraju, tematu, danych dla 2019 roku, grupy dochodowej oraz regionu.

Obs.	Country_Name	Rok	Zanieczyszczenia	Przewidywane
1	Afghanistan	2019	52.4170	63.5650
2	Albania	2019	18.6388	79.2820
3	Algeria	2019	32.8331	76.4740
4	Angola	2019	28.4208	62.4480
5	Antigua and Barbuda	2019	17.5920	78.6910

Rysunek 3-51 Pierwsze 5. rekordów tabeli PROJEKT.PRZEWIDYWANAZANIE2019

Obs.	Country_Name	Topic	YR2019	Income_Group	Region
1	Afghanistan	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	52.4170	Low income	Southern Asia
2	Angola	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	28.4208	Lower middle income	Sub-Saharan Africa
3	Albania	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	18.8388	Upper middle income	Southern Europe
4	Andorra	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	9.0864	High income	Southern Europe
5	United Arab Emirates	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	43.8870	High income	Western Asia

Rysunek 3-52 Pierwsze 5. Rekordów PROJEKT.ZANIECZYSZCZENIA2019

Z tak przygotowanymi danymi przystąpiono do dalszych analiz. W pierwszej kolejności sprawdzono, w których krajach populacja narażona jest na największe stężenia PM 2.5 na metr sześcienny. W tym celu wykonano kod znajdujący się na rysunku 3-53.

```
proc sort data=projekt.zanieczyszczenia2019 out=Sortowane;
    by descending YR2019 ;
run;

proc print data = Sortowane(obs=10);
```

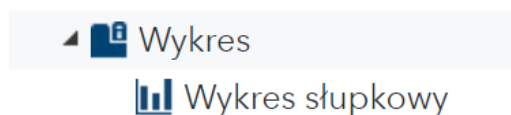
Rysunek 3-53 Kod sortujący tabelę oraz wyświetlający wynik

W efekcie otrzymano tablicę widoczną na rysunku 3-54. Informuje ona o tym, że największą ilość mikrogramów pyłów zawieszonych PM2.5 na metr sześcienny w 2019 znajdowała się w Indiach, Nepalu oraz Nigrze. Następne w kolejności były Katar, Nigeria, Egipt, Maurytania, Kamerun, Bangladesz oraz Pakistan. Zauważywszy, że większość z tych państw należy do jednej grupy dochodowej postanowiono sprawdzić jak rozkłada się ekspozycja na PM 2.5 zależnie od grupy dochodowej.

Obs.	Country_Name	Topic	YR2019	Income_Group	Region
1	India	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	83.1993	Lower middle income	Southern Asia
2	Nepal	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	83.1032	Lower middle income	Southern Asia
3	Niger	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	80.1093	Low income	Sub-Saharan Africa
4	Qatar	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	76.0493	High income	Western Asia
5	Nigeria	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	70.3963	Lower middle income	Sub-Saharan Africa
6	Egypt, Arab Rep.	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	67.8882	Lower middle income	Northern Africa
7	Mauritania	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	66.8118	Lower middle income	Sub-Saharan Africa
8	Cameroon	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	64.4709	Lower middle income	Sub-Saharan Africa
9	Bangladesh	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	63.4255	Lower middle income	Southern Asia
10	Pakistan	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)	62.5742	Lower middle income	Southern Asia

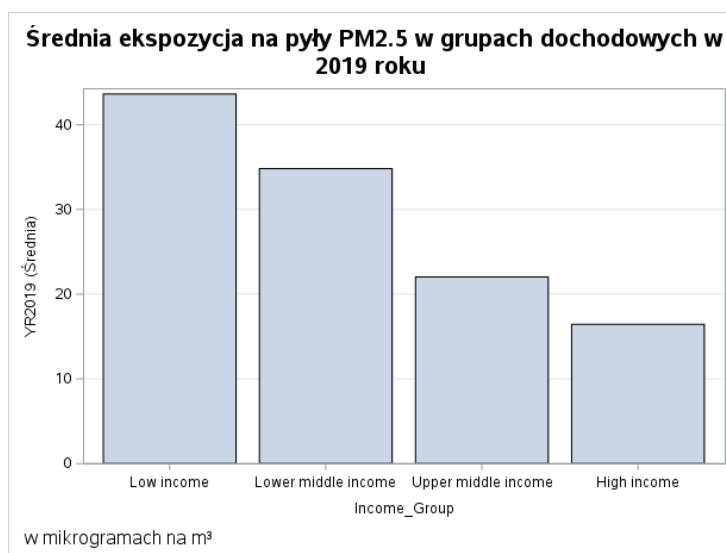
Rysunek 3-54 Tabela wynikowa

W tym celu utworzono wykres słupkowy za pomocą narzędzie wbudowanego w SASOnDemand widocznego na rysunku 3-55.



Rysunek 3-55 Procedura odpowiedzialna za wygenerowanie wykresu

W wyniku tego działania powstał wykres słupkowy widoczny na rysunku 3-56. Zauważono, że największą średnią ekspozycją na PM2.5 charakteryzują się państwa o niskich dochodach, a największa o wysokich.



Rysunek 3-56 Wykres średniej ekspozycji w grupach dochodowych w 2019 roku

W krajach o niskich dochodach średnia ekspozycja na pyły PM2.5 jest największa, a najmniejsza w krajach o wysokich dochodach.

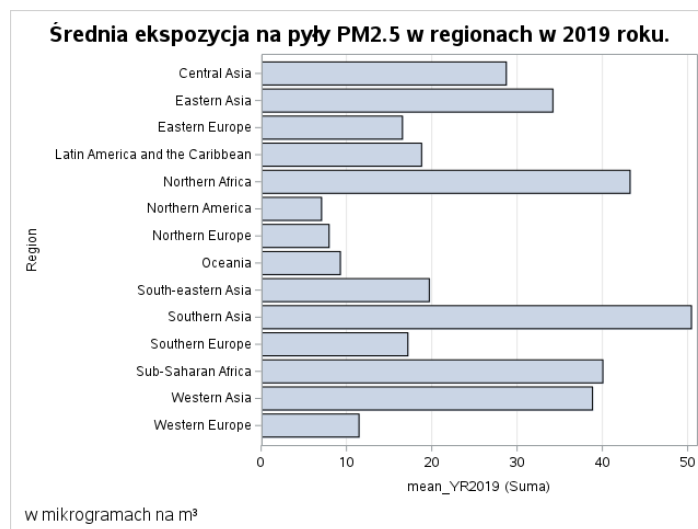
W następnym kroku, korzystając z kodu widocznego na rysunku 3-57, utworzono wykres informujący o średniej ekspozycji na pyły zawieszone PM2.5 na świecie z podziałem na regiony.

```
proc sort data =projekt.zanieczyszczenia2019 out = test;
by Region;
run;
proc means data=test mean std;
var YR2019;
by region;
output out=wyniki mean=mean_YR2019 std=std_YR2019;
run;

proc sgplot data=WYNIKI;
title height=14pt "Średnia ekspozycja na pyły PM2.5 w regionach w 2019 roku.";
footnote2 justify=left height=12pt "w mikrogramach na m³";
hbar Region / response=mean_YR2019;
yaxis discreteorder=data;
xaxis grid;
run;
```

Rysunek 3-57 Kod odpowiedzialny za wygenerowanie wykresu

Korzystając z wykresu widocznego na rysunku 3-58, że największą średnią ekspozycję można zauważyć w południowej Azji, a najmniejszą w Ameryce Północnej.



Rysunek 3-58 Wykres przedstawiający średnią ekspozycję w regionach

Następnie sprawdzono w ilu krajach na świecie przekraczana jest norma WHO. W tym celu wykorzystano kod widoczny na rysunku 3-59.

```
data test;
set projekt.zanieczyszczenia2019;
if YR2019 > 10 then CzyPrzekracza = 'Przekracza';
else YR2019 = 'Nie przekracza';
run;

proc template;
define statgraph SASStudio.Pie;
begingraph;
entrytitle "Odsetek państw, które przekraczają
zalecane normy średniej rocznej ekspozycji na pył PM2.5"
/ textattrs=(size=14);
layout region;
piechart category=CzyPrzekracza / stat=pct;
endlayout;
endgraph;
end;
run;

ods graphics / reset width=6.4in height=4.8in imagemap;

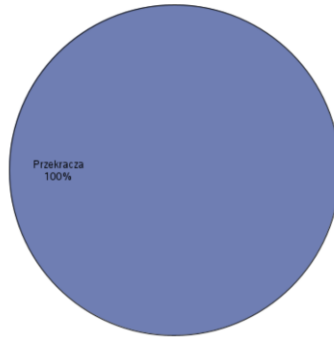
proc sgrender template=SASStudio.Pie data=WORK.TEST;
run;

ods graphics / reset;
```

Rysunek 3-59 Kod generujący wykres kołowy

W wyniku działania tego kodu powstał wykres kołowy widoczny na rysunku 3-60. Okazało się, że w każdym badanym kraju średnia ekspozycja na metr sześcienny była większa od zalecanej.

Odsetek państw, które przekraczają zalecane normy średniej rocznej ekspozycji na pył PM2.5



Rysunek 3-60 Wykres przedstawiający w ilu krajach w 2019 roku przekraczana była zalecana norma WHO

W ostatnim kroku zbadano czy istnieje zależność w badanych krajach pomiędzy przewidywaną długością życia w momencie narodzin, a średnią roczną ekspozycją na PM2.5. W tym celu wykorzystano kod widoczny na rysunku 3-61. Odpowiada on za utworzenie wykresu punktowego oraz wywołanie procedury CORR.

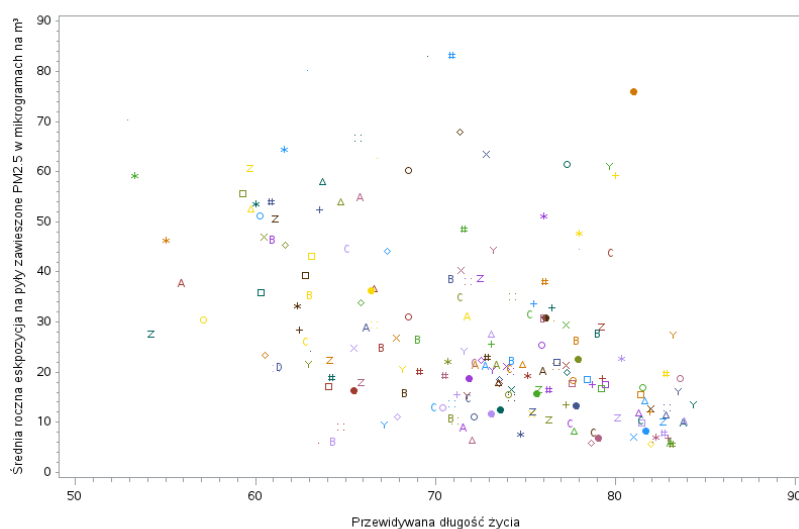
```
proc gplot data = projekt.przewidywanazanie2019;
title 'Wykres zależności pomiędzy średnią roczną
ekspozycją na pyły zawieszone PM 2.5, a przewidywaną długością życia w momencie narodzin';
title2 'w 2019';
plot Zanieczyszczenia * Przewidywane = Country_Name / haxis = axis1 vaxi = axis2;
axis1 label = ("Przewidywana długość życia") offset = (2,2)pcst;
axis2 label = (angle = 90 "Średnia roczna ekspozycja na pyły zawieszone PM2.5 w mikrogramach na m³");
footnote2 justify=left height=12pt "punkt = kraj";
run;

PROC CORR
data = projekt.przewidywanazanie2019;
run;
```

Rysunek 3-61 Kod odpowiedzialny za wygenerowanie wykresu oraz wywołanie procedury CORR

W efekcie powstał wykres widoczny na rysunku 3-62 oraz tablica widoczna na rysunku 3-63. Zauważono, że istnieje statystycznie istotna korelacja ujemna pomiędzy tymi dwoma zmiennymi.

Wykres zależności pomiędzy średnią roczną ekspozycją na pyły zawieszone PM 2.5, a przewidywaną długością życia w momencie narodzin w 2019



punkt = kraj

Rysunek 3-62 Wykres zależności pomiędzy ekspozycją, a przewidywaną długością życia

Współczynniki korelacji Pearsona, N = 193 Prawd. >  r  przy H0: rho=0		
	Zanieczyszczenia	Przewidywane
Zanieczyszczenia	1.00000	-0.46977 <.0001
Przewidywane	-0.46977 <.0001	1.00000

Rysunek 3-63 Macierz wynikowa ze współczynnikiem korelacji Pearsona

Podsumowując, zaobserwowano, że 8 na 10 krajów charakteryzujących się najwyższą średnią roczną ekspozycją na pyły PM2.5 należy do grupy krajów o średnich dochodach. Największa ekspozycja występowała głównie w Indiach i Nepalu. Z utworzonego wykresu słupkowego udało się odczytać, że średnia ekspozycja spada z coraz wyższą grupą dochodową. Podczas analizy regionalnej stwierdzono, że największa ekspozycja na pyły PM2.5 dotyczyła przede wszystkim południowej Azji, podczas gdy najmniejsza występowała w Ameryce Północnej. Wyniki potwierdziły również, że we wszystkich badanych państwach przekraczano zalecaną normę WHO dotyczącą średniej rocznej ekspozycji na pyły PM2.5. Dodatkowo, zaobserwowano ujemną zależność między przewidywaną długością życia a średnią roczną ekspozycją na pyły PM2.5. To oznacza, że kraje o niższej ekspozycji wykazują średniej siły tendencję do posiadania wyższej przewidywanej długości życia.

## 4 Podsumowanie

### 4.1 Wnioski

Przewidywana długość życia na świecie potrafi się znacząco różnić zależnie od regionu oraz grupy dochodowej. Pierwsza analiza udowodniła, że w 2021 roku co najmniej połowa krajów na świecie posiadała przewidywaną długość życia w momencie narodzin wyższą od 72,37 lat. Jednakże, dla 25% państw wskaźnik ten nie przekraczał 66,07 lat. Jest to ogromna różnica, która ilustruje jak warunki życiowe różnią się między krajami oraz regionami wpływając na ten wskaźnik. Mimo to wskaźnik ten nie jest wysoce zróżnicowany na świecie, a współczynnik zmienności wynosi w zaokrągleniu 10,69% co oznacza że średnio o tyle dane odbiegały od średniej. Japonia znajduje się na pierwszym miejscu pod względem przewidywanej długości życia, a Chad na ostatnim. Zauważono, że aż trzy regiony - Afryka Subsaharyjska, Południowa oraz Centralna Azja znajdowały się pod średnią światową, a największa średnia przewidywana długość życia znajduje się w Europie Zachodniej.

Poznawszy ogólną charakterystykę tego wskaźnika na świecie, kolejne analizy skupiały się na bardziej szczegółowych zagadnieniach. Średnia przewidywana długość życia na świecie była wyższa od średniej z 2011 roku, ale w Ameryce Łacińskiej i na Karaibach oraz Południowej Europie wskaźnik ten zmalał. W 2020 roku w każdym możliwym regionie średnia przewidywana długość życia zmalała, czego powodem mogła być pandemia COVID-19. Skutkiem, tych spadków państwa z średnio-wysokiej grupy dochodowej znalazły się poniżej średniej światowej. Zauważono również, że kraje z wyższej grupy dochodowej, miały niższe tempo wzrostu, niż kraje z niskiej grupy dochodowej.

Kobiety w każdym kraju świata mogą oczekiwać dłuższego życia niż mężczyźni. Wskaźnik ten zarówno w 2011 roku jak i w 2021 nie faworyzował mężczyzn w żadnym z badanych krajów. Badając tempo zmian tego wskaźnika, dla każdej płci zauważono pewną stabilność tej sytuacji. Średnia dla mężczyzn nie "nadgania" tej dla kobiet.

W czwartej analizie zauważono, że w większości państw na świecie wydatki zdrowotne wzrosły w 2021 porównując je do 2011 roku. Dalsze badanie pozwoliło zauważyć, że w Niemczech na przestrzeni 21 lat istniała silna dodatnia korelacja pomiędzy wydatkami zdrowotnymi, a przewidywaną długością życia.

W ostatniej analizie przyjrzano się średniej rocznej ekspozycji na pyły zawieszone PM2.5 w 2019 roku. Zauważono, że istnieje możliwa zależność pomiędzy grupą dochodową państwa oraz poziomem średniej ekspozycji. Dodatkowo wartość tej ekspozycji najwyższa była w południowej Azji, a najwyższa w Północnej Ameryce. Każde państwo przekraczało normy zalecane przez WHO dla ekspozycji. Istnieje umiarkowana zależność ujemna pomiędzy średnią roczną ekspozycją na pyły zawieszone PM2.5 w kraju, a przewidywaną długością życia w tym kraju.