

Przetwarzanie, analiza i wizualizacja danych w systemie SAS

Analiza danych dotyczących wpływu ekonomicznego i zdrowotnego lockdownu oraz pandemii wirusa Covid-19

Konrad Komor

Rok 2, gc01, st

Informatyka i Ekonometria

Wprowadzenie

Niniejsze opracowanie ma na celu analizę skutków pandemii wywołanej wirusem Covid-19 oraz lockdownu wprowadzonego ze względu na walkę z tym wirusem. W analizie będę odnosił się do danych dotyczących wielu państw jednak w przypadku niektórych pytań badawczych posiadam dane dotyczące tylko jednego państwa czyli Stanów Zjednoczonych Ameryki. Punktem docelowym mojej pracy jest zbadanie efektów i wpływów Pandemii na społeczeństwo.

Cele szczegółowe

1. Analiza zależności wprowadzenia lockdownu pomiędzy wzrostem bezrobocia w danym kraju. W którym kraju bezrobocie wzrosło najbardziej a w którym najmniej.
2. Analiza wpływu lockdownu na ilość osób które poszukują pomocy związanej z chorobami psychicznymi takimi jak depresja, ataki paniki oraz bezsenność. Zbadanie korelacji pomiędzy liczbą zakażonych a ilością osób które szukają takiej pomocy.
3. Analiza zmian cen w sieci sklepów Walmart pomiędzy rokiem 2020 a 2019. Czy Covid wpłynął na zmianę cen produktów codziennego użytku?
4. Analiza wpływu diety na szansę zachorowania na Covid, a także sprawdzenie jaki wpływ na tę szansę ma otyłość, sprawdzenie także jaki wpływ ma otyłość na śmiertelność choroby.
5. Analiza zależności pomiędzy wprowadzeniem lockdownu a liczbą wypadków samochodowych na drogach w USA. Czy fakt, że przez lockdown mniej osób podróżuje na co dzień do pracy oraz szkoły zmniejszył ilość wypadków samochodowych?

W celu zwiększenia czytelności pracy każda nowa analiza będzie się zaczynać od nowej strony.

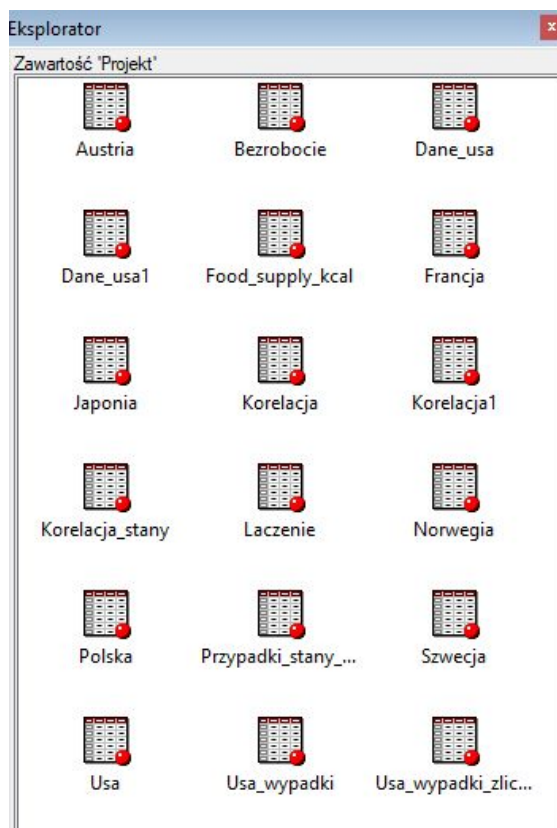
Dane źródłowe

Dane którymi się posłużę pochodzą ze strony kaggle.com. Są one dostępne po pobraniu dla każdej osoby bez konieczności logowania. Ze względu na to, iż w przypadku niektórych badań dane podzielone są na osobne tabele dla każdego innego kraju ilość moich tabel wejściowych jest dość spora. Zanim rozpocznę analizę w systemie SAS utworzę bibliotekę o nazwie Projekt, do której zaimportuję swoje dane. W tym celu użyję kodu języka 4gl aczkolwiek można to wykonać także ręcznie.

```
libname Projekt 'C:\Users\konra\Desktop\Projekt';
```

kod 1 - Tworzenie biblioteki

Dzięki kodowi widocznemu powyżej w bibliotece roboczej nazwanej przez nas Projekt możemy zapisywać swoje pliki, które zawierają wszystkie potrzebne nam informacje które wcześniej pobraliśmy musimy pamiętać żeby nazwy naszych danych nie przekraczały 32 znaków.



Część danych wejściowych potrzebna do analiz

Analiza zależności wprowadzenia lockdownu pomiędzy wzrostem bezrobocia w danym kraju.

1. Wyjaśnienie celu szczegółowego

Celem analizy jest określenie czy wprowadzenie lockdownu wpłynęło na bezrobocie w różnych krajach oraz sprawdzenie który kraj ucierpiało najmniej a który najbardziej.

2. Opis procesu uzyskania danych wynikowych.

Pierwszym etapem analizy zależności pomiędzy wprowadzeniem lockdownu a wzrostem bezrobocia jest zaimportowanie danych dane importuję z pliku o rozszerzeniu csv. W tym celu wykonam poniższy kod.

```
proc import datafile = 'C:\Users\konra\2020 unemployment.csv'
    out = Projekt.Bezrobocie
    dbms= csv replace;
run;
```

kod 2 - Import danych do biblioteki Projekt

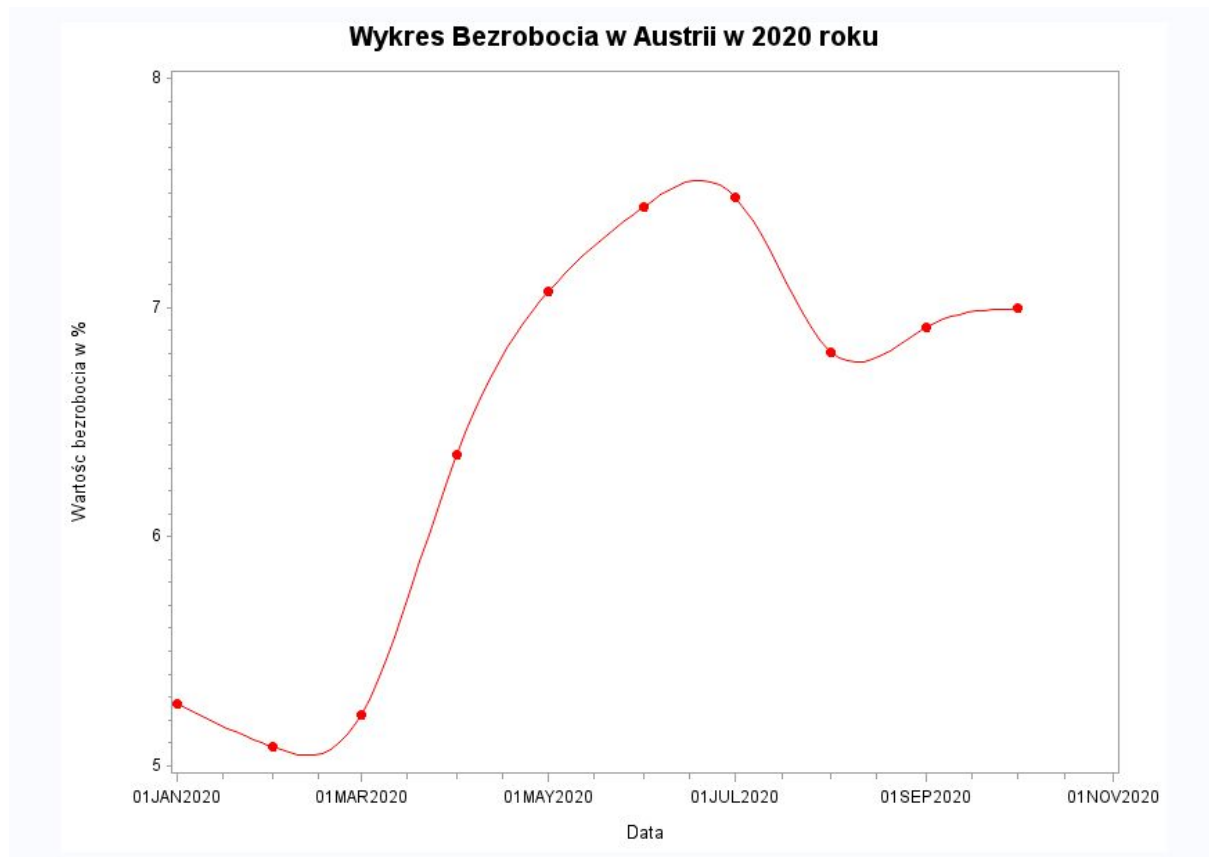
W wyniku wykonania tego kodu otrzymujemy tabelę którą prezentuje się następująco

	LOCATION	INDICATOR	SUBJECT	MEASURE	FREQUENCY	TIME	Value
6	AUS	HUR	TOT	PC_LF	M	01JUN2020	7.440282
7	AUS	HUR	TOT	PC_LF	M	01JUL2020	7.479592
8	AUS	HUR	TOT	PC_LF	M	01AUG2020	6.807751
9	AUS	HUR	TOT	PC_LF	M	01SEP2020	6.913334
10	AUS	HUR	TOT	PC_LF	M	01OCT2020	6.996009
11	AUT	HUR	TOT	PC_LF	M	01JAN2020	4.4
12	AUT	HUR	TOT	PC_LF	M	01FEB2020	4.5
13	AUT	HUR	TOT	PC_LF	M	01MAR2020	4.6
14	AUT	HUR	TOT	PC_LF	M	01APR2020	4.7
15	AUT	HUR	TOT	PC_LF	M	01MAY2020	5.7
16	AUT	HUR	TOT	PC_LF	M	01JUN2020	5.9
17	AUT	HUR	TOT	PC_LF	M	01JUL2020	5.7
18	AUT	HUR	TOT	PC_LF	M	01AUG2020	5.5
19	AUT	HUR	TOT	PC_LF	M	01SEP2020	5.4
20	AUT	HUR	TOT	PC_LF	M	01OCT2020	5.4
21	BEL	HUR	TOT	PC_LF	M	01JAN2020	5.1
22	BEL	HUR	TOT	PC_LF	M	01FEB2020	5
23	BEL	HUR	TOT	PC_LF	M	01MAR2020	5
24	BEL	HUR	TOT	PC_LF	M	01APR2020	5.1
25	BEL	HUR	TOT	PC_LF	M	01MAY2020	5

Tabela 1 - dane do analizy 1

Jak widzimy Kraj którego będzie dotyczyła nasza analiza jest przedstawiony 3 pierwszymi literami angielskiej nazwy kraju. Kolumna Subject opisuje czy mamy do czynienia z bezrobociem mężczyzn, kobiet czy bezrobociem w sumie. Kolumna Time pokazuje miesiąc którego dotyczy bezrobocie a Value wartość tego bezrobocia w %.

Pierwsza w naszej tabeli jest Austria oznaczona skrótem AUS zobaczmy jak prezentuje się wykres bezrobocia w Austrii w 2020 roku.



Wykres 1 - Wykres bezrobocia w Austrii

Wykres ten tworzymy poniższym kodem.

```

≡ data Projekt.Austria;
    set Projekt.Bezrobocie;
    if LOCATION='AUS' and SUBJECT='TOT';

    run;

≡ proc print data= Projekt.Japonia;
    run;

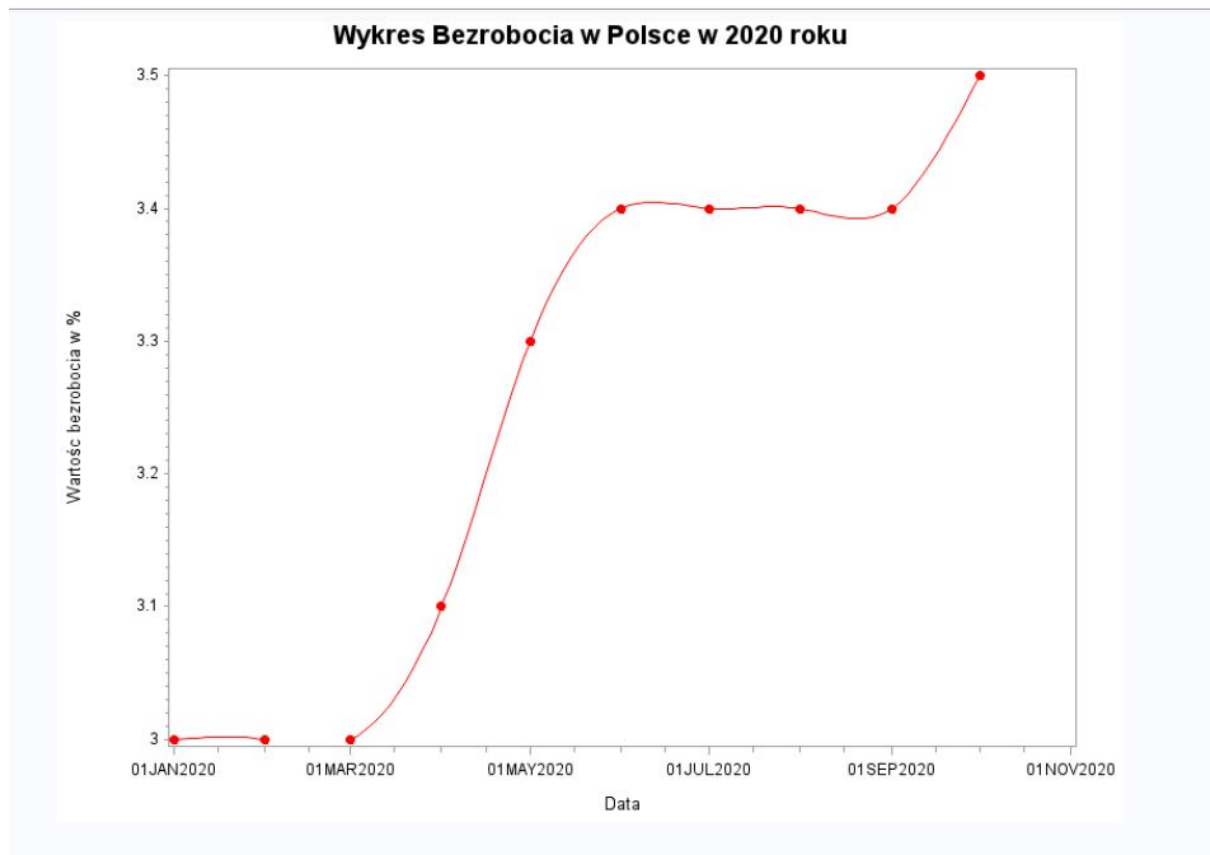
≡ proc gplot data= Projekt.Austria;
    title 'Wykres Bezrobocia w Austrii w 2020 roku';
    symbol1 color= red value=dot interpol =spline;
    plot Value*TIME / haxis=axis1 vaxis=axis2;
    axis1 label=("Data");
    axis2 label=(angle=90 "Wartość bezrobocia w %");
    run;
quit;

```

Kod 3 Tworzenie Osobnej tabeli oraz wykonanie wykresu bezrobocia

Jak widzimy na wykresie na przełomie marca i kwietnia odesetek osób bezrobotnych bardzo mocno się zwiększył 16 marca w kraju tym zaczął się lockdown możemy więc zakładać, że te dwa zdarzenia są ze sobą powiązane

Analogicznie do Austrii możemy wykonać wykresy dla Wielu krajów świata dzięki temu możemy zobaczyć jak sytuacja wyglądała na przykład w Polsce po wykonaniu odpowiedniego kodu otrzymujemy taki wykres:



Wykres 2 - Bezrobocie w Polsce

Jak widzimy także w naszym kraju po przekroczeniu marca bezrobocie zaczyna rosnąć aczkolwiek spadek ten nie jest aż mocne jak w Austrii ponieważ rośnie maksymalnie o 0,5%.

Jeśli chcemy zobaczyć jak prezentują się wszystkie kraje w naszej bazie możemy wykonać wykres punktowy w tym celu zaimplementujemy poniższy kod

```

data Projekt.Świat;
    set Projekt.Bezrobocie;
    if SUBJECT='TOT';

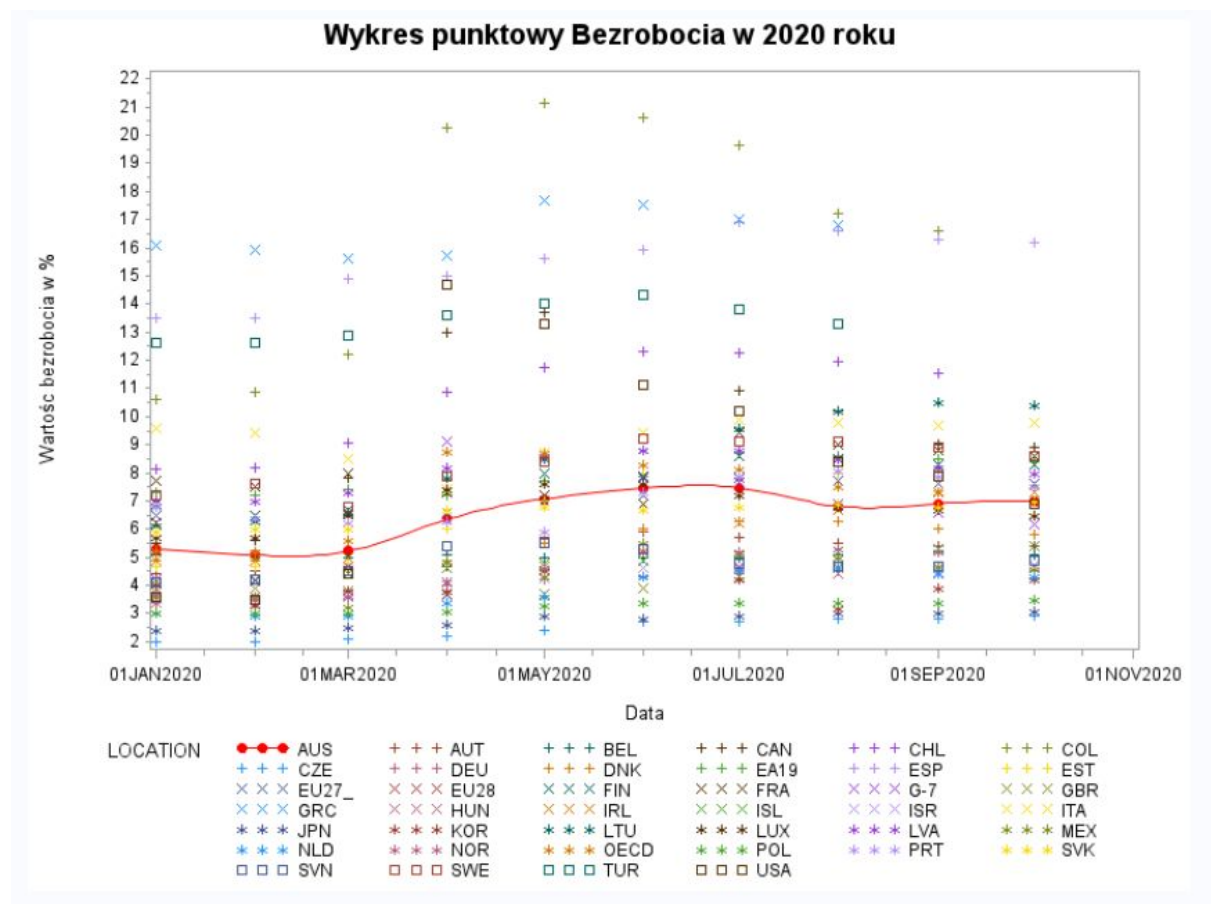
run;

proc gplot data= Projekt.Świat;
    title 'Wykres punktowy Bezrobocia w 2020 roku';
    plot Value*TIME=Location / haxis=axis1 vaxis=axis2;
    axis1 label= ("Data");
    axis2 label= (angle=90 "Wartość bezrobocia w %");
run;
quit;

```

kod 4 - Wykres punktowy bezrobocia

Po wykonaniu powyższego kodu otrzymujemy taki wykres:



wykres 3 - Wykres punktowy bezrobocia na świecie

3. Wnioski

Z wykresu wynika, że najmniejsze bezrobocie było w Czechach (2%) a największe w Kolumbii (20,62%). Na wykresie widać też, że w większości państw na przełomie marca i kwietnia poziom bezrobocia wzrósł znacząco co zbiegło się z początkiem lockdownów i globalnej pandemii. Potwierdza to tezę, że lockdown miał znaczący wpływ na bezrobocie na świecie.

Analiza wpływu lockdownu na ilość osób które poszukują pomocy związanej z chorobami psychicznymi takimi jak depresja, ataki paniki oraz bezsenność.

1. Wyjaśnienie celu szczegółowego

Głównym celem analizy jest sprawdzenie czy występuje zależność pomiędzy liczbą osób chorych na covid oraz wprowadzonym lockdownem a ilością osób które w danym kraju poszukują informacji i pomocy związanej z walką z daną chorobą psychiczną.

2. Opis procesu uzyskania danych wynikowych

Żeby dokonać analizy przede wszystkim potrzebujemy informacji o tym ile osób w danym momencie czasu szukało pomocy związanej z chorobami psychicznymi. W tym celu znalazłem bazę danych na serwisie kaggle.com w której wpisane reprezentatywne wartości o tym ile osób szukało takich informacji w wyszukiwarce google (Wartości w przedziale od 1-100 im większa wartość w danym tygodniu tym więcej osób szukało informacji o pomocy). Na początek zaimportujemy dane dotyczące USA i zobaczymy jak się one prezentują.

	Week	depression	anxiety	obsessive compulsive disorder	ocd	insomnia	panic attack	mental health	counseling	psychiatrist
1	16JUN2019	70	89	37	69	77	73	61	88	73
2	23JUN2019	70	91	51	73	83	71	62	89	66
3	30JUN2019	63	87	41	70	74	75	53	72	76
4	07JUL2019	74	92	60	74	84	77	63	88	89
5	14JUL2019	70	92	70	77	81	77	60	88	61
6	21JUL2019	75	93	42	72	82	81	61	90	75
7	28JUL2019	74	92	53	68	80	78	58	89	83
8	04AUG2019	69	93	47	69	83	79	63	90	49
9	11AUG2019	70	95	41	68	87	76	64	95	64
10	18AUG2019	72	94	56	74	91	77	62	96	36
11	25AUG2019	71	91	57	73	86	71	61	96	23
12	01SEP2019	71	89	54	71	87	73	62	90	60
13	08SEP2019	82	93	49	72	86	75	72	99	51

Tabela 2 - Częstotliwość wyszukiwania pomocy związanej z chorobami psychicznymi

Jak widzimy dane ukazują nam konkretny tydzień i ilość konkretnych wyszukiwań w tym tygodniu. Nie mamy jednak żadnej informacji o tym ile na terenie USA było w danym tygodniu przypadków w tym celu pobrałem inną tabelę z dziennymi przypadkami w USA i w innym programie za pomocą wbudowanej w język Python funkcji przebudowałem przypadki dzienne na tygodniowe a następnie bazę danych zaimportowałem do systemu SAS. Następnie w systemie SAS łączymy obydwie table za pomocą poniższego kodu.

```
data Projekt.Laczenie;  
set Projekt.Zdrowie_USA;  
set Projekt.Przypadki_USA;  
run;
```

kod 5 - Łączenie 2 tabel

Dzięki temu połączeniu otrzymujemy table w której znajdują się zarówno informacje o o ilości wyszukiwań informacji o pomocy jak i ilości zachorowań na Covid-19.

	Week	cases_Sum	depression	anxiety	obsessive compulsive disorder	ocd	insomnia	panic attack	mental health	counseling	psychiatrist
1	16JUN2019	0	70	89	37	69	77	73	61	88	73
2	23JUN2019	0	70	91	51	73	83	71	62	89	66
3	30JUN2019	0	63	87	41	70	74	75	53	72	76
4	07JUL2019	0	74	92	60	74	84	77	63	88	89
5	14JUL2019	0	70	92	70	77	81	77	60	88	61
6	21JUL2019	0	75	93	42	72	82	81	61	90	75
7	28JUL2019	0	74	92	53	68	80	78	58	89	83
8	04AUG2019	0	69	93	47	69	83	79	63	90	49
9	11AUG2019	0	70	95	41	68	87	76	64	95	64
10	18AUG2019	0	72	94	56	74	91	77	62	96	36
11	25AUG2019	0	71	91	57	73	86	74	61	96	33

Tabela 3 - Połączona tabela przypadków i wyszukiwań informacji o zdrowiu psychicznym w USA.

Dzięki tej tabeli możemy teraz wykonać wykres zależności pomiędzy liczbą osób szukających pomocy związanej z depresją a datą w USA mając na uwadze, że lockdown rozpoczął się tam w połowie Marca. Tak prezentuje się kod który wpisujemy żeby otrzymać taki wykres.

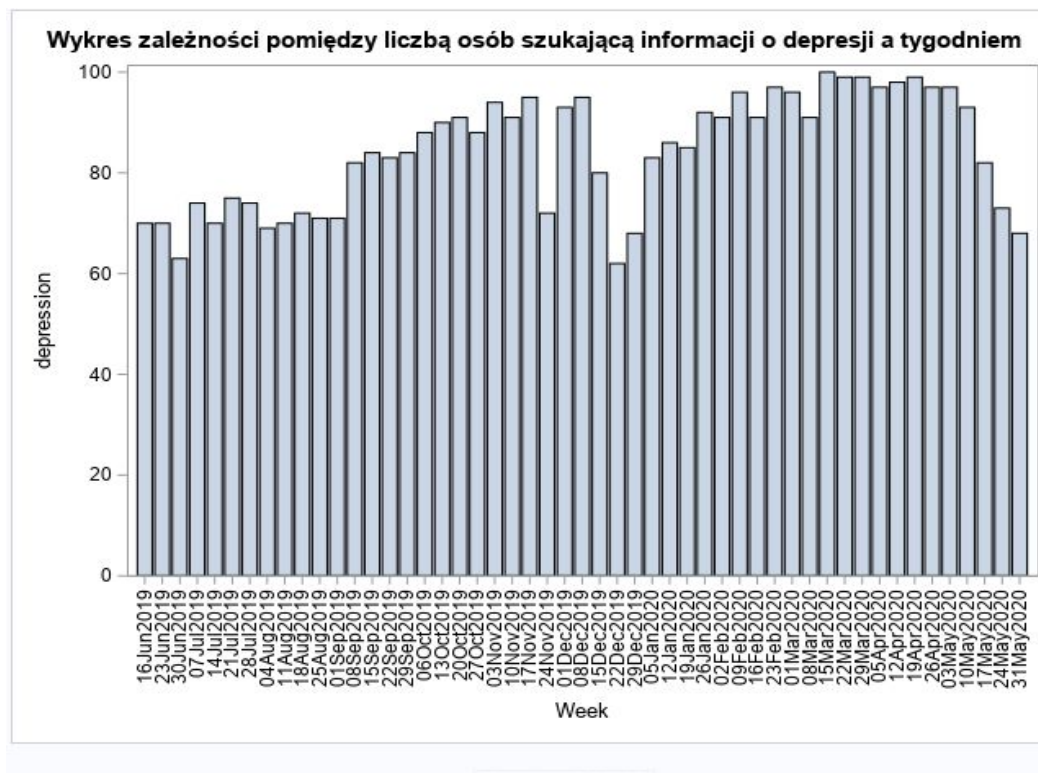
```

proc sgplot data= Projekt.Laczenie;
  title 'Wykres zależności pomiędzy liczbą osób szukającą informacji o depresji a tygodniem';
  symbol color= red value=dot interpol =spline;
  vbar Week/responbse=Depression;
  xaxis valuesrotate=vertical;
run;
quit;

```

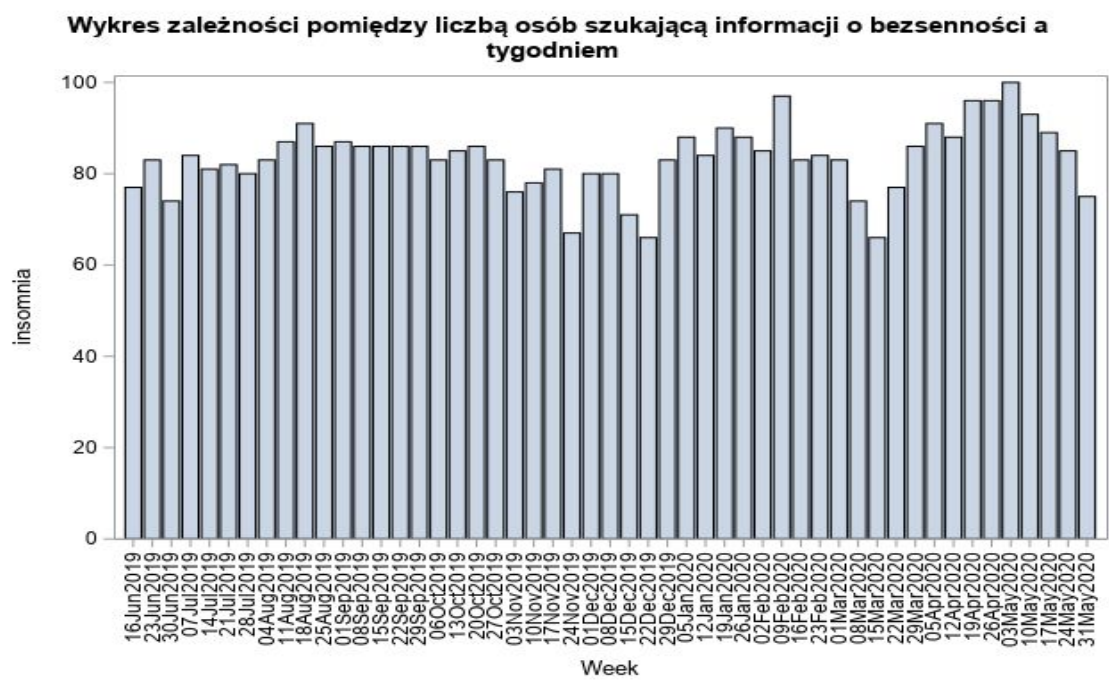
kod 6 - Stworzenie wykresu słupkowego pokazującego zależność pomiędzy tygodniem a liczbą osób poszukujących informacji o depresji

Wynikiem wpisania tego kodu będzie wykres słupkowy prezentujący się w tak:

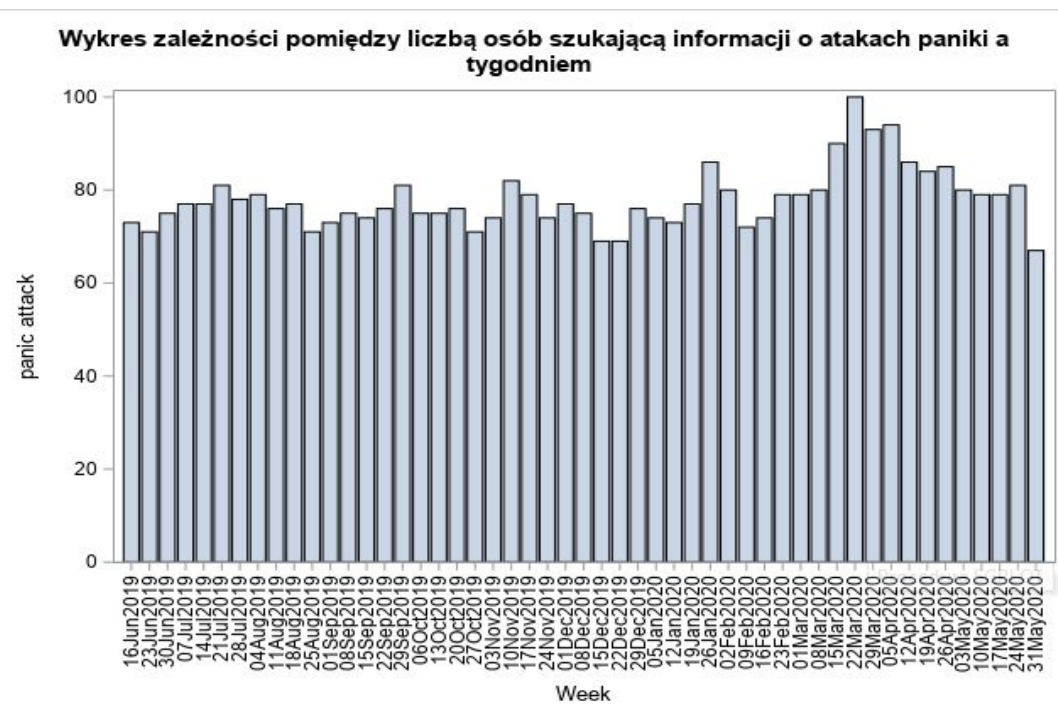


Wykres 4 - wykres przedstawiający zależność pomiędzy liczbą osób szukających informacji o pomocy związanej z depresją a tygodniem

Z wykresu widzimy że w momencie rozpoczęcia lockdownu ilość osób szukających informacji o pomocy związanej z depresją nie zmieniła się zbyt mocno. Możemy sprawdzić czy tak samo stało się z atakami paniki i bezsennością wykonując podobne wykresy.

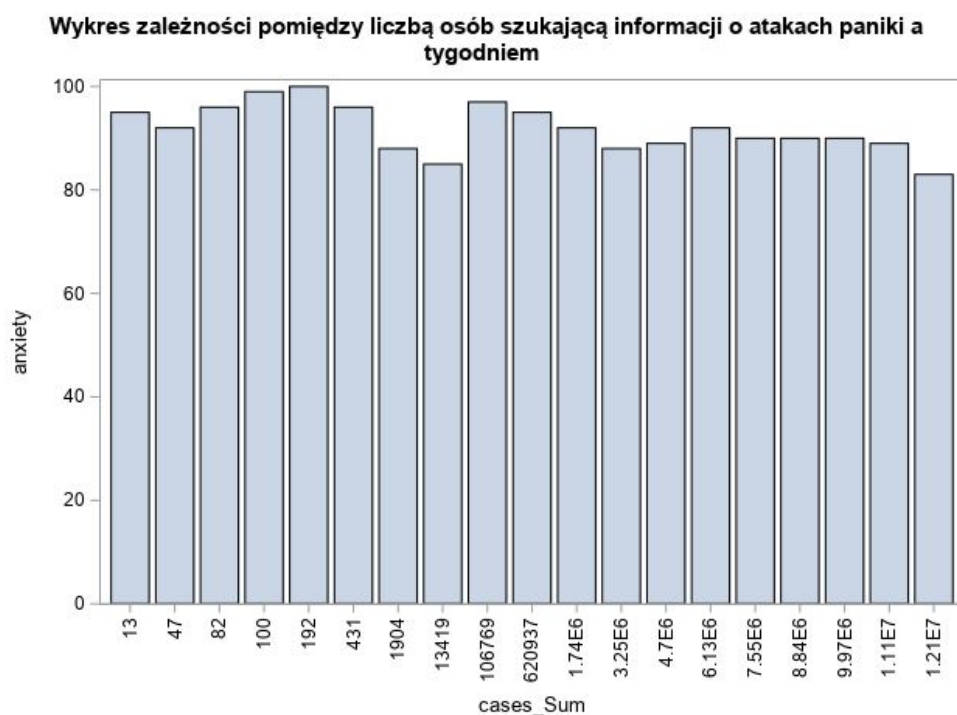


Wykres 5 - wykres przedstawiający zależność pomiędzy liczbą osób szukających informacji o pomocy związanej z bezsennością a tygodniem.

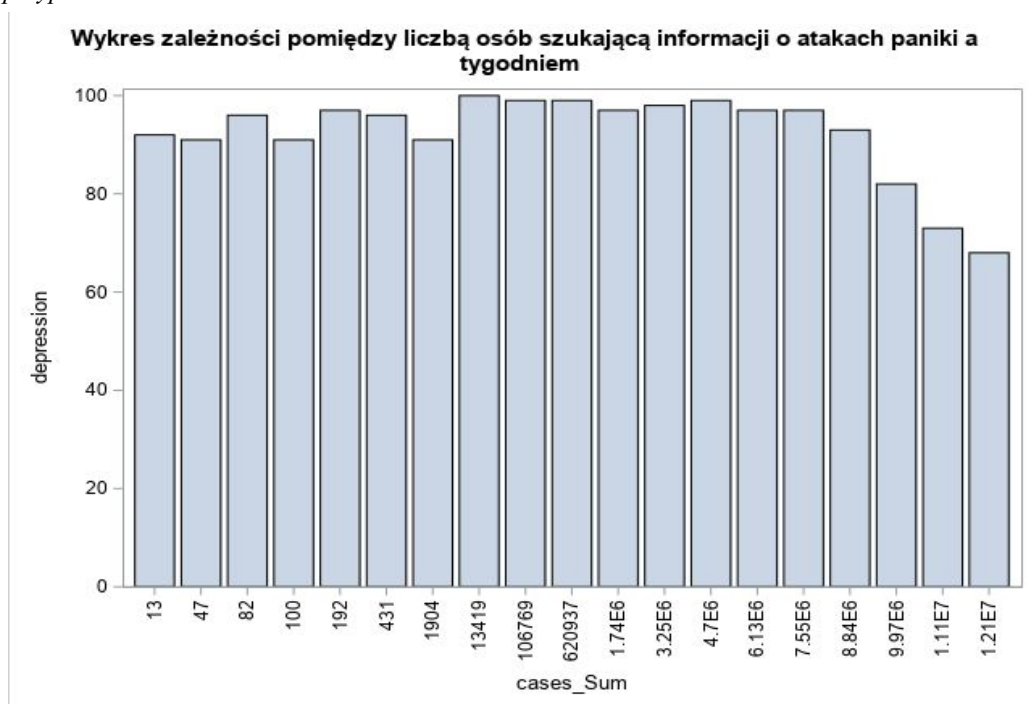


Wykres 6 - wykres przedstawiający zależność pomiędzy liczbą osób szukających informacji o pomocy związanej z atakami paniki a tygodniem.

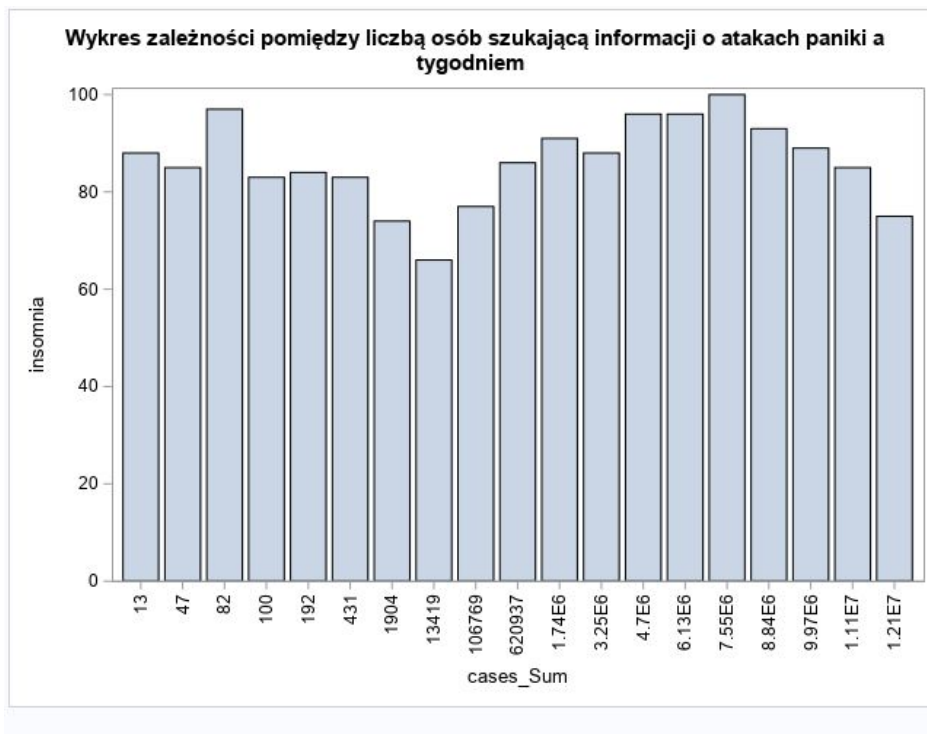
Jak widzimy na wykresach ilość osób szukających pomocy związanej z atakami panik lekko zwiększyła się w momencie wprowadzenia lockdownu podobnie zresztą stało się z ilością osób poszukujących informacji o bezsenności. Możemy teraz sprawdzić jak wygląda wykres ilości osób poszukujących informacji o danych chorobach psychicznych a ilością chorych na Covid-19.



wykres 7 - wykres zależności pomiędzy liczbą osób poszukujących informacji o atakach paniki a ilością przypadków SARS Cov 2.



wykres 8 - wykres zależności pomiędzy liczbą osób poszukujących informacji o depresji a ilością przypadków SARS Cov 2.

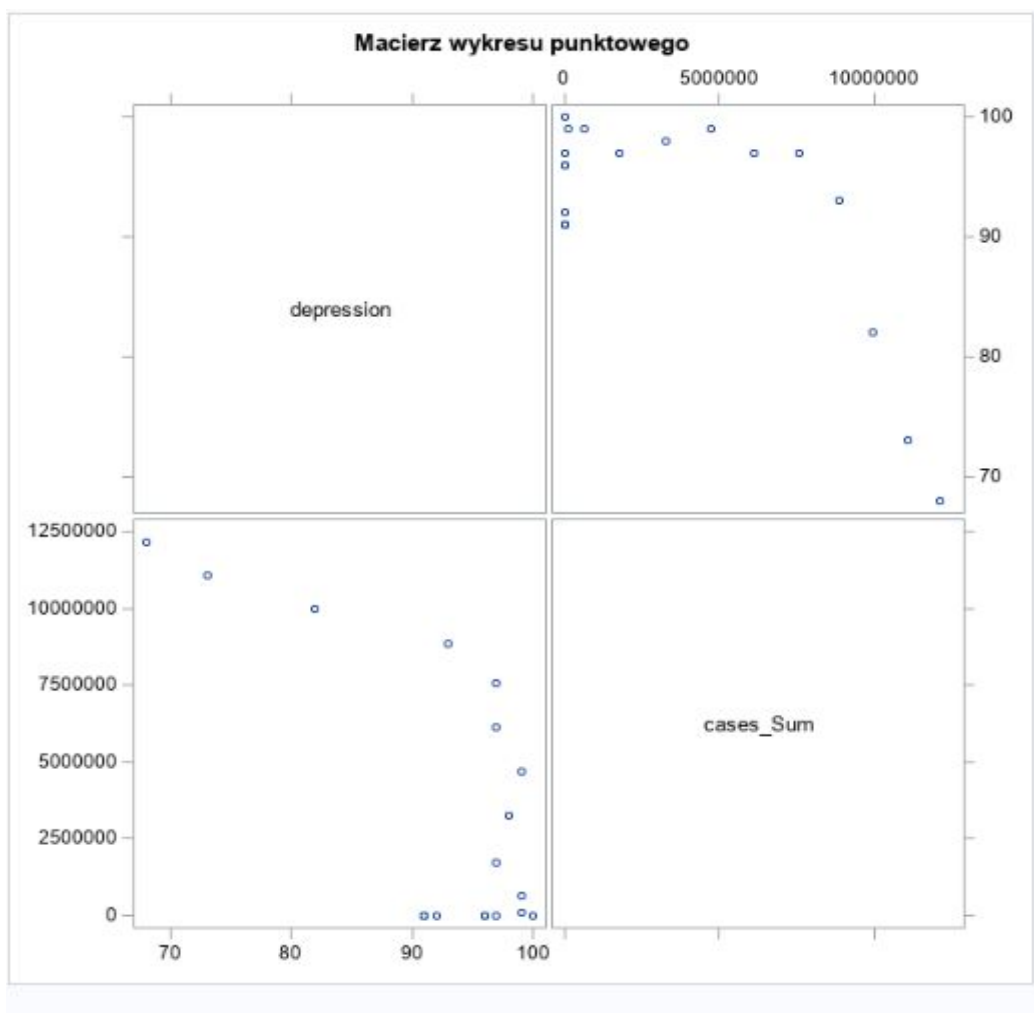


wykres 9 - wykres zależności pomiędzy liczbą osób poszukujących informacji o bezsenności a ilością przypadków SARS Cov 2.

Możemy także sprawdzić czy istnieje korelacja pomiędzy ilością osób chorych a ilością osób które szukają informacji o tych chorobach, widzimy że wykresy nie posiadają rozkładu normalnego dlatego sprawdzimy korelację rang spearmana w tym celu użyjemy poniższego kodu.

```
ods graphics on;
proc corr data=Projekt.Korelacja_stany plots= matrix spearman;
var depression Cases_sum;
run;
ods graphics off;|
```

kod 6 - korelacja rang spearmana pomiędzy depresją a ilością przypadków



Wykres 10 macierz wykresu punktowego korelacji między ilością zakażeń a depresją

Procedura CORR

2 Zmienne: depression cases_Sum

Statystyki proste

Zmienna	N	Średnia	Odch. std.	Mediana	Minimum	Maksimum	Etykieta
depression	19	92.42105	8.87140	96.00000	68.00000	100.00000	depression
cases_Sum	19	3482112	4398040	620937	13.00000	12132213	cases_Sum

Współczynniki korelacji Pearsona, N = 19 Prawd. > |r| przy H0: rho=0

	depression	cases_Sum
depression	1.00000	-0.69182
depression		0.0010
cases_Sum	-0.69182	1.00000
cases_Sum		0.0010

Współczynniki korelacji Spearmana, N = 19 Prawd. > |r| przy H0: rho=0

	depression	cases_Sum
depression	1.00000	-0.12384
depression		0.6135
cases_Sum	-0.12384	1.00000
cases_Sum		0.6135

Tabela 4 Wygenerowane wyniki procedur CORR

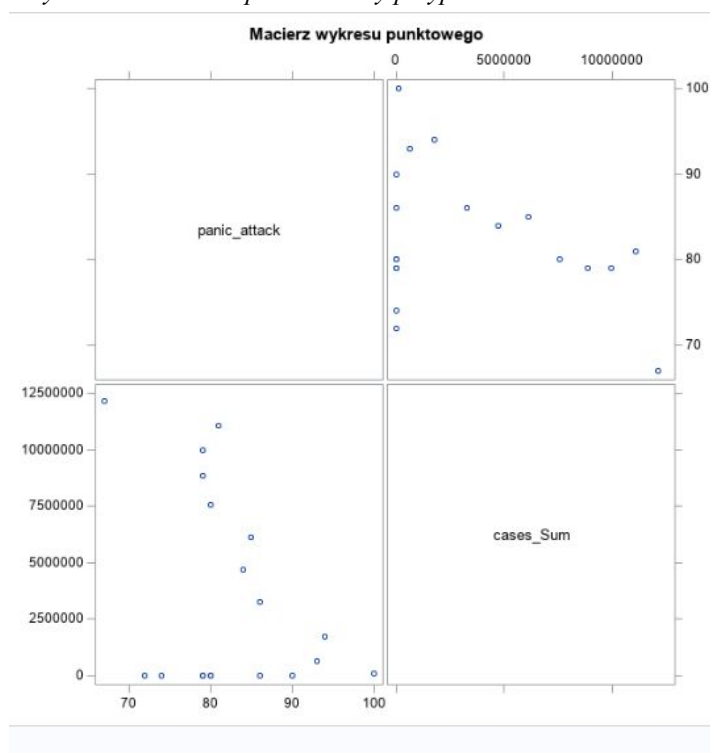
Dzięki powyższemu zabiegowi możemy odczytać średnią, odchylenie standardowe, minimum i maksimum oraz zależność pomiędzy zmiennymi. Możemy odczytać, że współczynnik korelacji Spearmana wynosi $-0,12$ a więc nie ma bezpośredniego związku pomiędzy ilością osób chorych a ilością osób szukających pomocy związanej z depresją analogicznie możemy sprawdzić czy występuje korelacja pomiędzy ilością zakażeń a bezsennością i atakami paniki

Procedura CORR							
2 Zmienne: panic_attack cases_Sum							
Statystyki proste							
Zmienna	N	Średnia	Odch. std.	Mediana	Minimum	Maksimum	Etykieta
panic_attack	19	82.52632	7.92583	80.00000	67.00000	100.00000	panic_attack
cases_Sum	19	3482112	4398040	620937	13.00000	12132213	cases_Sum

Współczynniki korelacji Pearsona, N = 19 Prawd. > r przy H0: rho=0			
	panic_attack	cases_Sum	
panic_attack	1.00000	-0.36477	
panic_attack		0.1247	
cases_Sum	-0.36477	1.00000	
cases_Sum		0.1247	

Współczynniki korelacji Spearmana, N = 19 Prawd. > r przy H0: rho=0			
	panic_attack	cases_Sum	
panic_attack	1.00000	-0.02384	
panic_attack		0.9228	
cases_Sum	-0.02384	1.00000	
cases_Sum		0.9228	

Tabela 5 Wyniki procedury Corr dla ataków paniki i sumy przypadków



Wykres 11 macierz wykresu punktowego korelacji między ilością zakażeń a atakami paniki

Procedura CORR

2 Zmienne: insomnia cases_Sum

Statystyki proste							
Zmienna	N	Średnia	Odch. std.	Mediana	Minimum	Maksimum	Etykieta
insomnia	19	86.10526	8.74893	86.00000	66.00000	100.00000	insomnia
cases_Sum	19	3482112	4398040	620937	13.00000	12132213	cases_Sum

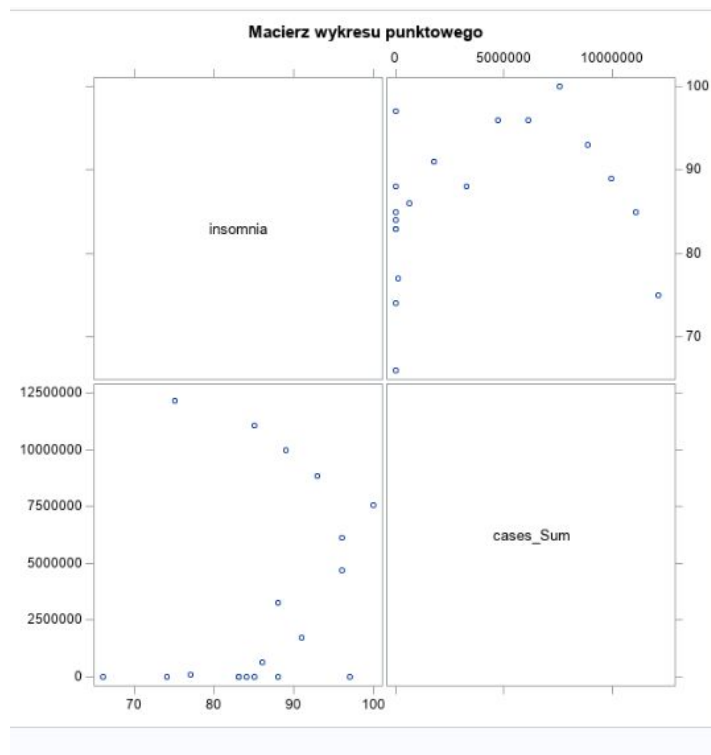
Współczynniki korelacji Pearsona, N = 19
Prawd. > |r| przy H0: rho=0

	insomnia	cases_Sum
insomnia	1.00000	0.24299
insomnia		0.3162
cases_Sum	0.24299	1.00000
cases_Sum		0.3162

Współczynniki korelacji Spearmana, N = 19
Prawd. > |r| przy H0: rho=0

	insomnia	cases_Sum
insomnia	1.00000	0.21793
insomnia		0.3701
cases_Sum	0.21793	1.00000
cases_Sum		0.3701

Tabela 6- Wyniki procedury Corr dla bezsenności i i sumy przypadków



Wykres 12 macierz wykresu punktowego korelacji między ilością zakażeń a bezsennością

3. Wnioski uzyskane z przeprowadzonej analizy

Na wykresach widać, że w momencie wprowadzenia lockdownu nastąpił lekki skok jeśli chodzi o liczbę osób które szukały pomocy związanej z chorobami psychicznymi jednak był to skok niewielki. Analiza wykazała, że jedyna korelacja jakiej można się dopatrzeć jest korelacją pomiędzy ilością zachorowań a bezsennością a i w tym wypadku korelacja ta jest stosunkowo słaba.

Analiza wpływu diety na szansę zachorowania na Covid, a także sprawdzenie jaki wpływ na tę szansę ma otyłość, sprawdzenie także jaki wpływ ma otyłość na śmiertelność choroby.

1. Wyjaśnienie celu szczegółowego

Celem analizy jest zbadanie zależności pomiędzy sposobem odżywiania się w danej populacji a ilością zachorowań na Covid w tej populacji. Chcę także sprawdzić jaki wpływ ma otyłość na śmiertelność tej choroby.

2. Opis procesu uzyskania danych wynikowych

Zacniemy tak jak zawsze od zaimportowania i przejrzenia naszych danych. Importujemy je z pliku CSV a po zaimportowaniu do systemu SAS dane prezentują się następująco.

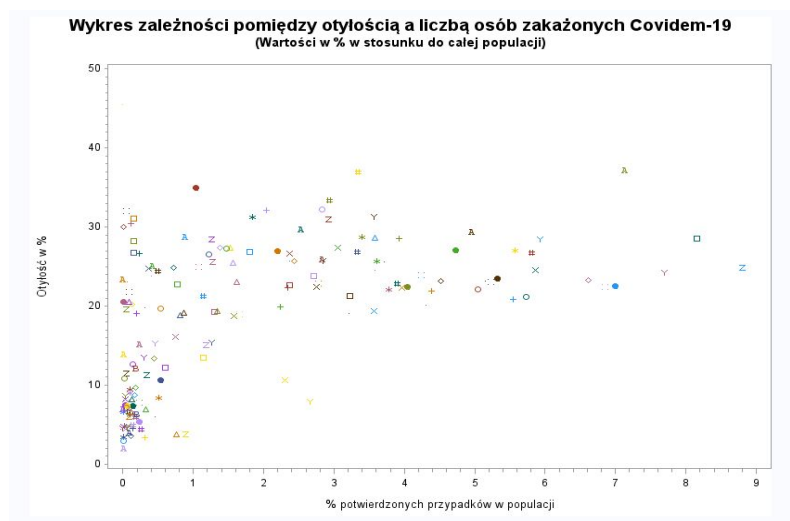
	Country	Alcoholic_Beverages	Animal_Products	Animal_Fats	Aquatic_Products	Other	Cereals	Excluding_Beer	Eggs	Fish_Seafood	Fruits	Excluding_Wine	Meat	Milk	Excluding_Butter	Miscellaneous	Offals	Olives	Pulses
1	Kiribati	0.1145	5.9555	0.2291	0	0.0336	16.9993	0.1309	2.6014	1.8815	2.6832	0.2781	0.1145	0.0164	10.2912	0			
2	Samoa	0.521	13.4454	2			9.6134	0.1513	1.7479		3.8992	7.9832	1.4286	0.2857	0.1176	9.9024	0		
3	United States of America	2.1373	13.6201	1.3009	0		10.7792	0.7832	0.4779	1.7788	6.1065	4.925	0	0.0133	0.9027	0			
4	Kuwait	0	9.1727	0.5806	0		20.4209	1.016	0.3048	2.5109	4.1655	2.9753	0.3193	1.1306	0.4644	0			
5	Saudi Arabia	0	7.59	0.7825	0		23.4585	0.4069	0.3286	3.2707	3.349	2.6448	0.2973	0.0782	0.2347	0			
6	Jordan	0.0553	6.0796	0.5527	0		22.0892	0.2395	0.2211	1.308	2.8924	2.1002	0.1105	0.0737	0.9211	0			
7	Turkey	0.3107	7.5989	0.9322	0		19.2938	0.4237	0.1271	2.2881	2.1328	3.9286	0.0282	0.0706	0.6921	0			

Tabela 7 - Dane dotyczące uśrednionego sposobu odżywiania się w danym kraju i przypadkami zachorowań na covid część 1

Spices	Starchy_Roots	Stimulants	Sugar_Crops	Sugar_Sweeteners	Treenuts	Vegetal_Products	Vegetable_Oils	Vegetables	Obesity	Undemourished	Confirmed	Deaths	Recovered	Active	Population
0.0164	3.6649	0.0164	0	7.9188	0.2945	44.0609	2.1924	0.5399	45.6 2.7						125000
0.1345	4.7395	0.2857	0	4.1849	0.0336	36.5378	3.4454	0.3193	45.5 2.7		0.001	0	0.001	0	200000
0.1062	1.2478	0.292	0	8.0048	0.4779	36.3733	9.2792	0.9292	37.3 <2.5		7.1330801084	0.1188639436	3.2046258793	7.0142161648	329878000
0.3774	1.2337	0.6096	0	6.1393	0.4209	40.8418	5.762	1.7271	37 2.8		3.3460669367	0.0201875933	0.0201875933	0.1212534641	4691000
0.4069	0.5321	0.5008	0	4.6479	0.266	42.41	7.2457	0.7668	35 7.1		1.040532519	0.0180160384	1.0170428926	0.0054735881	35041000
0.0921	0.8106	0.3316	0	6.9823	0.35	43.9204	8.6035	1.1422	33.4 12.2		2.9295907857	0.0385897556	2.769032681	0.1219683491	10679000
0.2966	1.2853	0.0989	0	4.3644	1.0734	42.4011	8.7571	2.1751	32.2 <2.5		2.8340100553	0.028259909	2.6822635155	0.1234866308	83737000
0.2204	0.7103	0.4164	0	8.6701	0.1714	35.2437	3.1594	1.5185	32.1 NA		2.0409669211	0.0445292621	1.6557251908	0.3407124682	393000

Tabela 8 - Dane dotyczące uśrednionego sposobu odżywiania się w danym kraju i przypadkami zachorowań na covid część 2

Jak widzimy dane są dosyć obszerne i pokazują procentowe źródła spożycia kalorii w danym kraju oraz populację danego kraju a także dane dotyczące wirusa w tym kraju. Na samym początku wykonajmy wykres zależności pomiędzy procentową ilością osób otyłych w populacji a ilością osób chorych na Covid



Wykres 13 - Zależność pomiędzy ilością osób otyłych a ilością osób chorych

Jak widzimy po wykresie w pewnym momencie następuje wzrost. Być może jedną z jego przyczyn jest fakt iż osoby otyłe są bardziej podatne na zarażenie się chorobą. Sprawdźmy czy istnieje korelacja pomiędzy otyłością a ilością zakażeń. W tym celu posłużymy się procedurą CORR a sama korelacja ze względu na fakt iż nasz rozkład nie jest rozkładem normalnym będzie korelacją spearmana.

Procedura CORR

2 Zmienne:	Obesity Confirmed
------------	-------------------

Statystyki proste						
Zmienna	N	Średnia	Odch. std.	Mediana	Minimum	Maksimum
Obesity	167	18.70778	9.63356	21.20000	2.10000	45.60000
Confirmed	164	1.76915	2.07381	0.84073	0.0003115	8.79904

Współczynniki korelacji Pearsona Prawd. > r przy H0: rho=0 Liczba obserwacji		
	Obesity	Confirmed
Obesity	1.00000 167	0.51184 <.0001 163
Confirmed	0.51184 <.0001 163	1.00000 164

Współczynniki korelacji Spearmana Prawd. > r przy H0: rho=0 Liczba obserwacji		
	Obesity	Confirmed
Obesity	1.00000 167	0.56759 <.0001 163
Confirmed	0.56759 <.0001 163	1.00000 164

Tabela 9 - Korelacja rang Spearmana pomiędzy otyłością a ilością zakażeń.

Jak widzimy korelacja pomiędzy otyłością a ilością potwierdzonych przypadków jest stosunkowo silną korelacją więc możemy wnioskować, że rzeczywiście te rzeczy są ze sobą powiązane. Możemy też sprawdzić czy otyłość ma wpływ na śmiertelność choroby

Procedura CORR

2 Zmienne: Obesity Deaths

Statystyki proste						
Zmienna	N	Średnia	Odch. std.	Mediana	Minimum	Maksimum
Obesity	167	18.70778	9.63356	21.20000	2.10000	45.60000
Deaths	164	0.03384	0.04247	0.01023	0	0.17674

Współczynniki korelacji Spearmana Prawd. > r przy H0: rho=0 Liczba obserwacji		
	Obesity	Deaths
Obesity	1.00000 167	0.53540 <.0001 163
Deaths	0.53540 <.0001 163	1.00000 164

Tabela 10 - Korelacja pomiędzy śmiertelnością a otyłością.

Jak widzimy korelacja pomiędzy śmiertelnością a otyłością jest trochę słabsza niż korelacją pomiędzy ilością zakażeń a otyłością nie zmienia to jednak faktu, iż cały czas jest to widoczna korelacja. Możemy teraz zadać pytanie który czynnik diety najbardziej wpływa na szansę zachorowania na Covid-19 w tym celu sprawdzimy korelację pomiędzy sposobem w jaki przyjmujemy kalorie do naszego organizmu a ilością zachorowań na Covid-19.

Procedura CORR

9 Zmienne: Alcoholic_Beverages Animal_Products Animal_fats Cereals___Excluding_Beer Fish___Seafood Fruits___Excluding_Wine Meat Vegetables Confirmed

Statystyki proste						
Zmienna	N	Średnia	Odch. std.	Mediana	Minimum	Maksimum
Alcoholic_Beverages	170	1.32520	1.06236	1.24460	0	5.15740
Animal_Products	170	9.29455	4.75417	9.03410	1.62370	22.29110
Animal_fats	170	1.26742	1.28372	0.87750	0	7.80070
Cereals___Excluding_Beer	170	20.36536	6.46794	19.61995	8.95650	37.52650
Fish___Seafood	170	0.63150	0.57901	0.47830	0	4.41830
Fruits___Excluding_Wine	170	2.01197	1.41615	1.69485	0.14710	8.85400
Meat	170	3.89637	2.21669	3.68745	0.29800	10.56740
Vegetables	170	1.08633	0.64823	1.00310	0.09570	3.35240
Confirmed	164	1.76915	2.07381	0.84073	0.0003115	8.79904

Tabela 11 Statystyki dotyczące sposobu odżywiania i przypadków Covida

Współczynniki korelacji Spearmana Prawd. > r przy H0: rho=0 Liczba obserwacji									
	Alcoholic_Beverages	Animal_Products	Animal_fats	Cereals__Excluding_Beer	Fish__Seafood	Fruits__Excluding_Wine	Meat	Vegetables	Confirmed
Alcoholic_Beverages	1.00000 170	0.59805 <.0001 170	0.52358 <.0001 170	-0.61044 <.0001 170	0.21583 0.0047 170	0.02134 0.7824 170	0.55582 <.0001 170	0.09123 0.2368 170	0.43324 <.0001 164
Animal_Products	0.59805 <.0001 170	1.00000 170	0.85129 <.0001 170	-0.68881 <.0001 170	0.21999 0.0039 170	0.10285 0.1820 170	0.87263 <.0001 170	0.36463 <.0001 170	0.58894 <.0001 164
Animal_fats	0.52358 <.0001 170	0.85129 <.0001 170	1.00000 170	-0.56721 <.0001 170	0.10354 0.1790 170	0.01083 0.8885 170	0.66660 <.0001 170	0.27755 0.0002 170	0.58216 <.0001 164
Cereals__Excluding_Beer	-0.61044 <.0001 170	-0.68881 <.0001 170	-0.56721 <.0001 170	1.00000 170	-0.37719 <.0001 170	-0.35532 <.0001 170	-0.64548 <.0001 170	-0.12479 0.1049 170	-0.40718 <.0001 164
Fish__Seafood	0.21583 0.0047 170	0.21999 0.0039 170	0.10354 0.1790 170	-0.37719 <.0001 170	1.00000 170	0.15449 0.0443 170	0.28938 0.0001 170	0.01008 0.8962 170	-0.06532 0.4060 164
Fruits__Excluding_Wine	0.02134 0.7824 170	0.10285 0.1820 170	0.01083 0.8885 170	-0.35532 <.0001 170	0.15449 0.0443 170	1.00000 170	0.16384 0.0328 170	0.19712 0.0100 170	0.13605 0.0824 164
Meat	0.55582 <.0001 170	0.87263 <.0001 170	0.66660 <.0001 170	-0.64548 <.0001 170	0.28938 0.0001 170	0.16384 0.0328 170	1.00000 170	0.23786 0.0018 170	0.40212 <.0001 164
Vegetables	0.09123 0.2368 170	0.36463 <.0001 170	0.27755 0.0002 170	-0.12479 0.1049 170	0.01008 0.8962 170	0.19712 0.0100 170	0.23786 0.0018 170	1.00000 170	0.36016 <.0001 164
Confirmed	0.43324 <.0001 164	0.58894 <.0001 164	0.58216 <.0001 164	-0.40718 <.0001 164	-0.06532 0.4060 164	0.13605 0.0824 164	0.40212 <.0001 164	0.36016 <.0001 164	1.00000 164

Tabela 12 Korelacja pomiędzy sposobem odżywiania się a ilością zachorowań na Covid

3. Wnioski

Jak widzimy istnieje korelacja pomiędzy sposobem odżywiania się a szansą zachorowania na Covid. Szczególnie silna jest korelacja pomiędzy jedzeniem tłuszczów zwierzęcych a ilością zachorowań. Może to wynikać z faktu, iż tłuszcze zwierzęce w nadmiarze mogą być niezdrowe i powodować otyłość która także zwiększa ryzyko zachorowania oraz śmiertelność choroby. Nie należy jednak wykluczać faktu iż produkty zwierzęce w tym tłuszcze spożywa się częściej w krajach bardziej rozwiniętych z większym zagęszczeniem ludności co także może wpływać na szanse zakażenia. Widać także korelację odwrotną w przypadku osób które spożywają w swojej diecie wiele produktów zbożowych. To także może wiązać się z faktem, iż produkty zbożowe spożywa się w większych ilościach na terenach mniej rozwiniętych

Analiza zależności pomiędzy wprowadzeniem lockdownu a liczbą wypadków samochodowych na drogach w USA.

1. Wyjaśnienie celu szczegółowego.

Celem analizy jest zbadanie czy istnieje związek pomiędzy wprowadzeniem lockdownu a zmniejszeniem się ilości wypadków samochodowych na drogach w stanach

2. Opis procesu uzyskania danych wynikowych

Na początku zaimportujemy potrzebne nam dane. W tym wypadku będą to dane dotyczące wypadków samochodowych w latach 2016 - 2020 z tym że nas będą interesowały wyłącznie wypadki z okresu wprowadzenia lockdownu a także rok przed lockdownem. W bazie danych którą znalazłem znajduje się wiele informacji które nie będą nam potrzebne ponieważ każdy wypadek jest osobno opisywany dlatego możemy pozbyć się kolumn które nas nie interesują. Interesuje nas tylko data kiedy wypadek się wydarzył oraz kolumna w której wpisujemy 1 żeby potem móc zsumować miesięczną sumę wypadków. Posiadając tę informację za pomocą funkcji `reshape()` w języku python 3 znalazłem miesięczną sumę wypadków w USA w latach 2019-2020 tabela prezentuje się następująco

	Start_Time	Quantity
1	2019-01-31	79276
2	2019-02-28	74559
3	2019-03-31	69401
4	2019-04-30	73001
5	2019-05-31	73864
6	2019-06-30	65501
7	2019-07-31	68234
8	2019-08-31	75246
9	2019-09-30	87659
10	2019-10-31	100017

Tabela 13 - Miesięczna ilość wypadków samochodowych w USA

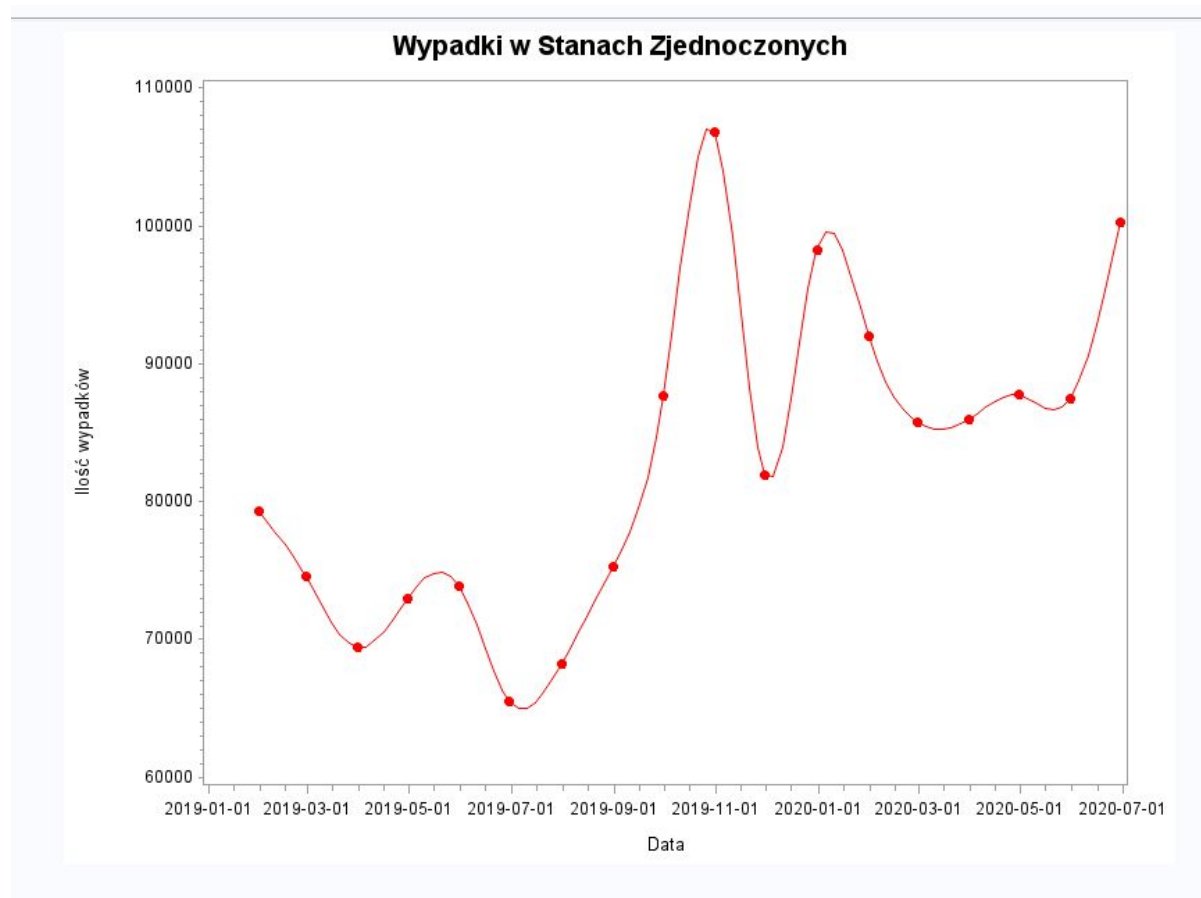
Dzięki posiadaniu takich informacji możemy teraz wykonać wykres ilości wypadków która wydarzyła się danego miesiąca. Dzięki temu będziemy mogli określić czy w momencie wprowadzenia w USA lockdownu ilość ta znacząco się zmieniła zrobimy to za pomocą prostego kodu.

```

proc gplot data= Projekt.Usa_wypadki_zliczone_miesieczne;
  symbol1 color= red value=dot interpol =spline;
  title 'Wypadki w Stanach Zjednoczonych';
  plot Quantity*Start_Time / haxis=axis1 vaxis=axis2;
  axis1 label=("Data");
  axis2 label=(angle=90 "Ilość wypadków");
run;
quit;

```

Kod 7 - tworzenie wykresu dotyczącego wypadków samochodowych w USA



Wykres 14 - Wykres przedstawiający ilość wypadków samochodowych w USA w danym miesiącu

3. Wnioski

Lockdown w USA w stanach zjednoczonych rozpoczął się w połowie Marca 2020 roku. Jak widzimy na wykresie nie doszło do gwałtownego spadku wypadków w tym czasie. Więc możemy wnioskować, że lockdown który zakłada ograniczenie przemieszczania się osób nie miał wpływu na ilość wypadków samochodowych na drogach w USA co ciekawe widać za to wyraźny wzrost wypadków wraz z rozpoczęciem się jesieni. Być może istnieje korelacja pomiędzy warunkami atmosferycznymi a ilością wypadków.

Analiza zmian cen w sieci sklepów Walmart pomiędzy rokiem 2020 a 2019. Czy Covid wpłynął na zmianę cen produktów codziennego użytku?

1. Wyjaśnienie celu szczegółowego.

Celem analizy jest sprawdzenie jak zmieniły się ceny towarów w sieci sklepów Walmart uwzględniając ceny przed pandemią jak i po jej rozpoczęciu.

2. Opis procesu uzyskania danych wynikowych.

Dane do tej analizy pobrałem z platformy kaggle. Znalazłem dwie bazy danych dotyczące produktów sprzedawanych w sieci Walmart jedna dotyczyła produktów sprzedawanych w 2019 roku a druga dotyczyła produktów sprzedawanych w roku 2020. Po implementacji baz danych nasze dane prezentowały się następująco

Crawl_Timestamp	Product_Name	Sale_Price	Brand
2019-12-18	La Costena Chipotle Peppers, 7 OZ (Pack of 12)	31.93	La Coste
2019-12-18	Equate Triamcinolone Acetonide Nasal Allergy Spray, 55 mcg per spray, 0.37 fl oz	10.48	Equate
2019-12-18	AduroSmart ERIA Soft White Smart A19 Light Bulb CRI 90+, 60W Equivalent, Hub Required	10.99	AduroSmart ERIA
2019-12-18	24" Classic Adjustable Balloon Fender Set Chrome. for bicycles, bikes, for 24" lowrider bike, beach cruiser.	38.59	lowrider

Tabela 14 - Baza danych dotycząca produktów w roku 2019

Crawl_Timestamp	Product_Name	Sale_Price	Brand
2020-01-24	Allegiance Economy Dual-scale Digital Thermometer	11.11	Cardinal Health
2020-01-24	Kenneth Cole Reaction Eau De Parfum Spray For Women 3.40 Oz	23.99	Kenneth Cole
2020-01-24	Kid Tough Fitness Inflatable Free-Standing Punching Bag + Machine Washable Fabric Cover South Carolina Gamecocks Kids Workout Buddy by Bonk Fit	30.76	BONK FIT

Tabela 15 - Baza danych dotycząca produktów w 2020 roku

Za pomocą prostego polecenia możemy obie tabele połączyć w jedną tabelę a następnie wyeksportować z systemu SAS żeby skorzystać z możliwości znalezienia tylko rekordów które posiadały taką samą nazwę produktu ale inną cenę i stworzenia z nich tabeli przestawnej dzięki której wyświetli się nam nazwa produktu jego cena w roku 2020 i w roku 2019 a także policzymy na jej podstawie różnice cen.

crawl	product	price	
		2019	2020
	(2 Pack) Heinz Hot Dog Relish, 12.7 oz	4.98	4.83
	(Case of 10) F32T8/SPX65 32 Watt Daylight T8 Linear Fluorescent Tube, 4 Foot, 32W FO32 865 Fluorescent Light Bulbs	39.99	43.99
	(Price/Case)Bobs Red Mill Natural Foods Inc 1357 S116 Bobs Red Mill Lemon Blueberry Granola	45.31	48.84
	(Price/Case)Del Monte 2004703 Diced Pears In 100% Juice Delmonte 48/4.4oz Plastic Cups	62.79	62.80
	1 Pair / 2 Packs Powerful Suction Cup Bath & Shower Safety Handle Bathtub Grab Bar Grip Support Anti-Slip Waterproof Bathroom Toilet Hand Rail Bathroom Aids & Safety	17.93	23.67

	under armour mens highlight rm football cleats	143.75	43.99
	under armour mens leadoff rm baseball cleats	399.46	137.77
	under armour mens showdown vented golf pants	212.46	0.00
	[[Follure]] Remover Foot File Perfect Health Foot Care Pedicure Tools Remove Hard Dead Skin	13.06	0.00

tabela 16 - Pomocnicza tabela przestawna wykonania w języku python

Następnie ponownie implementujemy tabelę tym razem z wyliczoną już różnicą do systemu SAS.

	product	crawl1	_crawl2	price1	_price2	_diff
1	Baby Brezza Transitional Soft Spout Trainer Sippy Cup	10MAR20:00:36:02	18DEC19:06:50:16	8.09	8.99	-0.9
2	Under Armour Mens Magnetico Premiere Firm Ground Soccer Shoe	24JAN20:20:31:18	19DEC19:09:02:24	206	71.77	134.23
3	Roces Mens RSK 2 Ice Skate Superior Italian Design 450572 00001	24JAN20:19:22:43	19DEC19:02:13:33	73.26	99.95	-26.69
4	Bacati - Tribal Noah Mint/Navy Cotton Percale Fabric covered Storage,Collapsible Hamper, 18 H x 13 W x 13 L inches	10MAR20:01:00:02	18DEC19:08:30:32	12.29	22.68	-10.39

Tabela 17 - Tabela na podstawie której dokonamy dalszej analizy zmian cen produktów

Jak widzimy dla tego samego produktu mamy teraz dwie daty sprzedaży price1-Cena w 2019 roku i price2 Cena w 2020 roku. Możemy posortować nasze dane i zobaczyć jakie produkty podrożały najmocniej robimy to za pomocą prostego kodu.

```
proc sort data= Projekt.Walmart;
by _diff
;
run;
```

Kod 8 -Sortowanie różnicy cen od najmniejszej

Po wykonaniu tego kodu nasza tabela prezentuje się następująco

1	Framed Art For Your Wall Perca Fluvialilis Fishing Perch Fishing Boat Fish 10x13 Frame	24JAN20:18:34:54	19DEC19:07:49:14	12.99	5000	-4987.01
2	5-Row Powder Coated Bleachers (Navy)	24JAN20:20:36:54	19DEC19:07:25:19	4036.68	4708.24	-671.56
3	Tip N Roll Bleachers-Type:4 Rows	24JAN20:20:36:48	19DEC19:07:24:57	1592.63	2161.99	-569.36
4	Tip N Roll Bleachers-Type:2 Rows	10MAR20:11:03:35	19DEC19:07:24:47	1890.95	2324.99	-434.04
5	Driftsun Gromp Wakesurf Board 3ftin long Custom Grom Kids Wakesurfer	10MAR20:10:06:24	19DEC19:04:10:03	0	319.99	-319.99
6	RollaSport II Steel-Acrylic Portable Basketball System, Royal Blue	10MAR20:10:22:05	19DEC19:05:10:05	1844.9	2098.9	-254
7	Merchandise	10MAR20:08:20:01	18DEC19:07:44:10	20.31	222.43	-202.12

Tabela 18 - Produkty posegregowane według tego który zdrożał najbardziej.

Jak widzimy pierwszy produkt zdrożał 4987\$ co jest bardzo nie prawdopodobne i prawdopodobnie jest błędem w naszej bazie za to kolejny produkt którym jest zestaw ławek

w kolorze niebieskim zdrożał już o kwotę realną czyli o 671\$ co jest dużą kwotą pieniędzy. Sprawdźmy jeszcze o ile średnio zmieniła się cena produktów pomiędzy tymi dwoma latami zrobimy to za pomocą procedury means

```
proc MEANS data=Projekt.Walmart;
run;
```

Kod 9 - Znalezienie średniej

Procedura MEANS					
Zmienna	N	Średnia	Odch. std.	Minimum	Maksimum
crawl1	970	1897460865	1970800.75	1895480342	1899461201
_crawl2	970	1892327866	31374.84	1892268180	1892365642
price1	970	70.3667216	274.0523540	0	4036.68
_price2	970	72.0821959	301.2513731	0	5000.00
_diff	970	-1.7154742	184.9817327	-4987.01	2106.04

Tabela 19 - Rezultat wywołania procedur means

Dzięki wywołaniu tej procedury możemy zobaczyć że średnia cena produktu w 2019 roku to 70,36\$ a w 2020 roku to już 72,08\$ a średnia różnica pomiędzy cenami to wzrost ceny o 1,71\$

3. Wnioski

Po rozpoczęciu się pandemii w 2020 roku ceny produktów wzrosły w stosunku do roku poprzedniego. Może to mieć związek problemami jakie sprawia Pandemia dla gospodarki.

Podsumowanie analiz

Podsumowując wszystkie uzyskane informacje z przeprowadzonych przeze mnie analiz możemy stwierdzić, że Covid-19 miał wpływ na niektóre dziedziny naszego życia. Z przeprowadzonych analiz wynika, że Covid mocno wpłynął na wzrost bezrobocia w większości krajów świata. W Kolumbii bezrobocie osiągnęło aż 20 %. Pomimo tendencji wzrostowych we wszystkich krajach kraje takie jak Czechy i Japonia doskonale poradziły sobie z problemem bezrobocia i w tych krajach wzrosło one nieznacznie.

Jeśli chodzi o wpływ lockdownu na zdrowie psychiczne ludzi okazuje się, że nie widać bezpośredniego wzrostu pomiędzy ilością zakażonych a ilością osób poszukujących pomocy psychiatrycznej. W przypadku depresji współczynnik korelacji Spearmana wyszedł nawet ujemny (-0,12). Tylko w przypadku bezsenności widać wyraźny wzrost na początku wprowadzenia lockdownu który i bardzo szybko się normuje. Może to świadczyć o tym, że jako ludzie szybko przyzwyczajamy się do nowych standardów.

Dieta ma wpływ na szansę zarażenia się Covidem-19. Przede wszystkim większą szansę zakażenia się mają osoby otyłe oraz osoby które większość kalorii w swojej diecie przyjmują z produktów zwierzęcych. Najmniejszą za to szansę mają osoby które przyjmują w diecie jak najwięcej posiłko z produktów zbożowych. Może to jednak mieć też związek z tym, że osoby mieszkające w miastach w których gęstość zaludnienia jest większa częściej jedzą produkty pochodzenia zwierzęcego.

Lockdown miał ograniczyć się nasze poruszanie co przełożyło się na zmniejszenie się ruchu drogowego w momencie wprowadzenia go. Czy jednak te zmniejszenie się miało rzeczywisty wpływ na spadek liczby wypadków drogowych. Według przeprowadzonej przeze ze mnie analizy ilość wypadków drogowych wcale nie spadła w momencie wprowadzenia lockdownu.

Ceny w sklepie Walmart pomiędzy rokiem przed pandemią a rokiem podczas którego pandemia już trwała zmieniły się o wyraźną kwotę średnio produkty kosztowały 70,36\$ w roku 2019 a w roku 2020 cena ta wzrosła do 72,08\$ czyli średnio o 1,71\$.