

Rekrutacja_CodersCrew_Jan_Solarz

July 6, 2021

1 Rekrutacja do działu Data Science CodersCrew

Jan Solarz

```
In [4]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

1.1 Wstępna analiza danych

Wyświetl 5 pierwszych wierszy wczytanego zbioru danych

Jaki jest rozmiar tego zbioru danych? Ile mamy wierszy, a ile kolumn?

```
In [212]: df.shape
```

```
Out[212]: (1704, 6)
```

Jakie są typy poszczególnych kolumn? Czy mamy wartości brakujące?

```
In [213]: df.dtypes
```

```
Out[213]: country      object
year          int64
pop           float64
continent      object
lifeExp       float64
gdpPercap     float64
dtype: object
```

1.2 Statystyki

Ile państw zostało uwzględnionych w tym zbiorze danych?

```
In [214]: df.country.nunique()
```

```
Out[214]: 142
```

W jakich krajach średnia długość życia była najkrótsza w roku 1952?

```
In [215]: shortest=df[df.year==1952]
shortest1=shortest.sort_values(['lifeExp'])
shortest1[0:10]
```

```
Out [215]:
```

	country	year	pop	continent	lifeExp	gdpPercap
0	Afghanistan	1952	8425333.0	Asia	28.801	779.445314
552	Gambia	1952	284320.0	Africa	30.000	485.230659
36	Angola	1952	4232095.0	Africa	30.015	3520.610273
1344	Sierra Leone	1952	2143249.0	Africa	30.331	879.787736
1032	Mozambique	1952	6446316.0	Africa	31.286	468.526038
192	Burkina Faso	1952	4469979.0	Africa	31.975	543.255241
624	Guinea-Bissau	1952	580653.0	Africa	32.500	299.850319
1668	Yemen, Rep.	1952	4963829.0	Asia	32.548	781.717576
1392	Somalia	1952	2526994.0	Africa	32.978	1135.749842
612	Guinea	1952	2664249.0	Africa	33.609	510.196492

W jakich krajach rednia dugo ycia bya najdusza w roku 2007?

```
In [216]: longest=df[df.year==2007]
longest1=longest.sort_values(['lifeExp'], ascending = False)
longest1[0:10]
```

```
Out [216]:
```

	country	year	pop	continent	lifeExp	gdpPercap
803	Japan	2007	127467972.0	Asia	82.603	31656.06806
671	Hong Kong, China	2007	6980412.0	Asia	82.208	39724.97867
695	Iceland	2007	301931.0	Europe	81.757	36180.78919
1487	Switzerland	2007	7554661.0	Europe	81.701	37506.41907
71	Australia	2007	20434176.0	Oceania	81.235	34435.36744
1427	Spain	2007	40448191.0	Europe	80.941	28821.06370
1475	Sweden	2007	9031088.0	Europe	80.884	33859.74835
767	Israel	2007	6426679.0	Asia	80.745	25523.27710
539	France	2007	61083916.0	Europe	80.657	30470.01670
251	Canada	2007	33390141.0	Americas	80.653	36319.23501

W jakich krajach GDP per capita byo najwysze w roku 2007 i najnisze w roku 1952?
Najnisze

```
In [217]: lowest=df[df.year==1952]
lowest1=shortest.sort_values(['gdpPercap'])
lowest1[0:10]
```

```
Out [217]:
```

	country	year	pop	continent	lifeExp	gdpPercap
876	Lesotho	1952	748747.0	Africa	42.138	298.846212
624	Guinea-Bissau	1952	580653.0	Africa	32.500	299.850319
492	Eritrea	1952	1438760.0	Africa	35.928	328.940557
1044	Myanmar	1952	20092996.0	Asia	36.319	331.000000
204	Burundi	1952	2445618.0	Africa	39.031	339.296459
504	Ethiopia	1952	20860941.0	Africa	34.078	362.146280
216	Cambodia	1952	4693836.0	Asia	39.417	368.469286

924	Malawi	1952	2917802.0	Africa	36.256	369.165080
480	Equatorial Guinea	1952	216964.0	Africa	34.482	375.643123
288	China	1952	556263528.0	Asia	44.000	400.448611

Najwysze

```
In [218]: highest=df[df.year==2007]
highest1=highest.sort_values(['gdpPercap'], ascending = False)
highest1[0:10]
```

```
Out [218]:
```

	country	year	pop	continent	lifeExp	gdpPercap
1151	Norway	2007	4627926.0	Europe	80.196	49357.19017
863	Kuwait	2007	2505559.0	Asia	77.588	47306.98978
1367	Singapore	2007	4553009.0	Asia	79.972	47143.17964
1619	United States	2007	301139947.0	Americas	78.242	42951.65309
755	Ireland	2007	4109086.0	Europe	78.885	40675.99635
671	Hong Kong, China	2007	6980412.0	Asia	82.208	39724.97867
1487	Switzerland	2007	7554661.0	Europe	81.701	37506.41907
1091	Netherlands	2007	16570613.0	Europe	79.762	36797.93332
251	Canada	2007	33390141.0	Americas	80.653	36319.23501
695	Iceland	2007	301931.0	Europe	81.757	36180.78919

Jaka jest korelacja midzy GDP, a redni dugoci ycia?

```
In [219]: df['gdpPercap'].corr(df['lifeExp'])
```

```
Out [219]: 0.5837062198659948
```

Gdzie warunki ycia poprawiy si najbardziej? Procentowo według redniej dugoci ycia porównuj rok 1952 i 2007

```
In [220]: table=df.pivot_table(values='lifeExp',index='country',columns='year')
df1 = table[[1952,2007]]
year2=df1[2007].tolist()
year1=df1[1952].tolist()
before=np.array(year1, dtype=np.float)
after=np.array(year2, dtype=np.float)
change_array=(after-before)/before
change = np.array(change_array.tolist())
df1= table[[1952,2007]].copy()
df1['change'] = change
df_sorted=df1.sort_values(['change'],ascending = False)
df_sorted[0:10]
```

```
Out [220]:
```

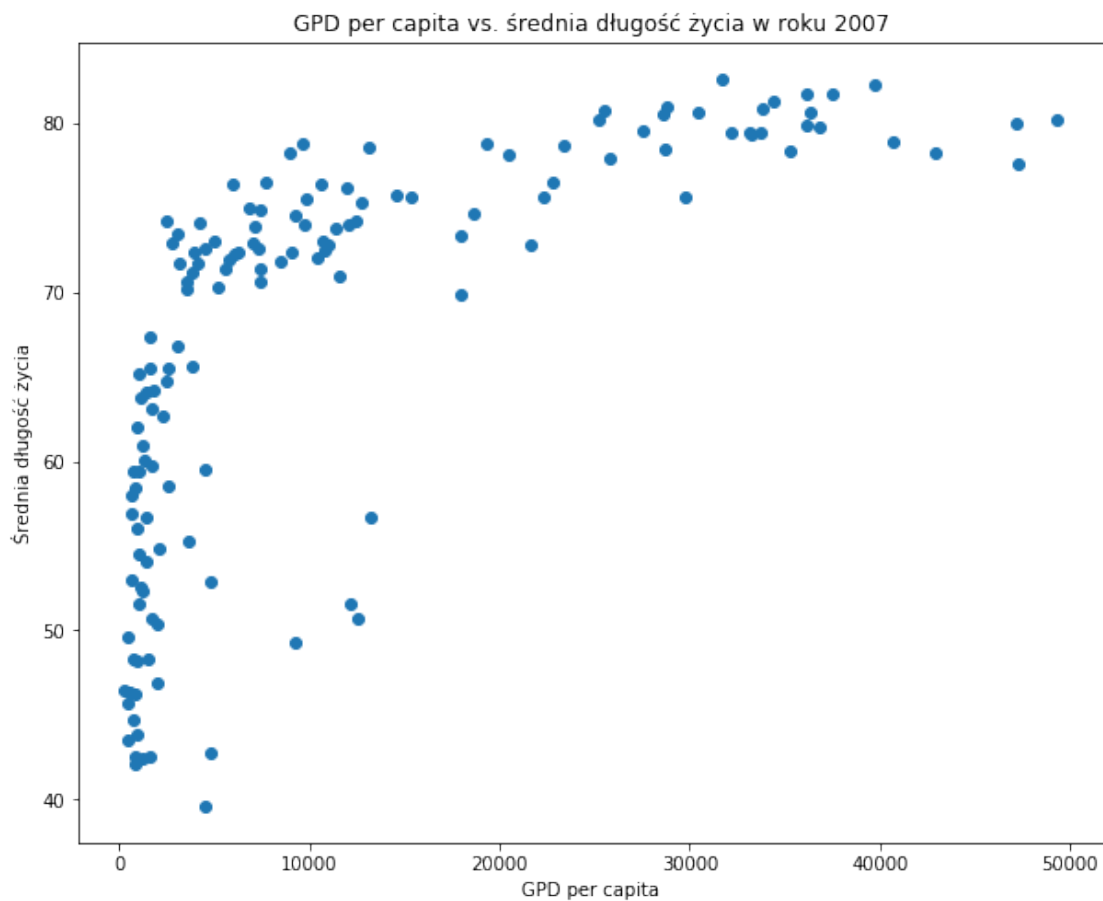
year	1952	2007	change
country			
Oman	37.578	75.640	1.012880
Gambia	30.000	59.448	0.981600
Yemen, Rep.	32.548	62.698	0.926324

Indonesia	37.468	70.650	0.885609
Vietnam	40.412	74.249	0.837301
Saudi Arabia	39.875	72.777	0.825129
Nepal	36.157	63.785	0.764112
India	37.373	64.698	0.731143
Libya	42.723	73.952	0.730965
Nicaragua	42.314	72.899	0.722810

1.3 Wykresy i wizualizacje

Jak rednia dugo ycia zaley od GDP per capita (rok 2007)?

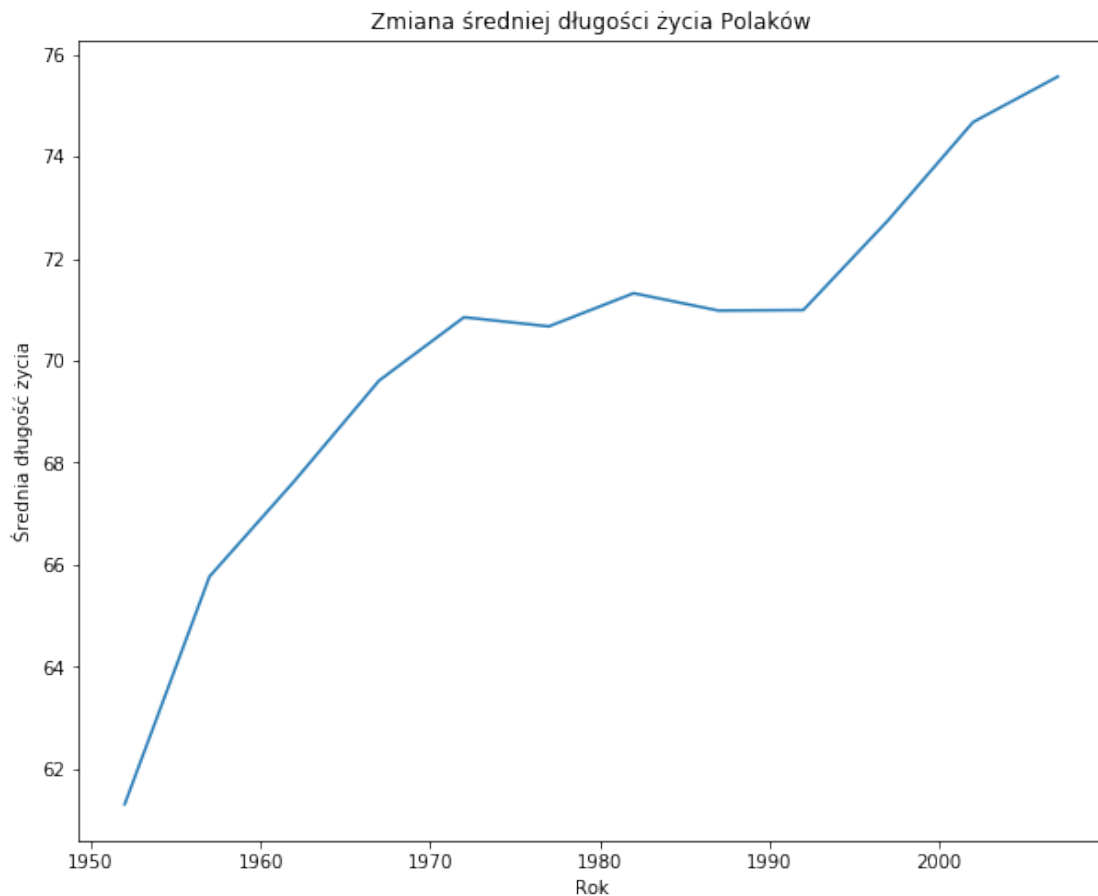
```
In [223]: chart1=df.pivot_table(values='gdpPercap',index='country',columns='year')
plt.figure(figsize=(10,8))
plt.xlabel('GPD per capita')
plt.ylabel('rednia dugo ycia')
plt.title('GPD per capita vs. rednia dugo ycia w roku 2007')
plt.scatter(x,y);
```



1.3.1 Statystyki - Polska

rednia długość życia w Polsce

```
In [224]: chart2=df.pivot_table(values='lifeExp',index='year',columns='country')
Poland=chart2['Poland']
plt.figure(figsize=(10,8))
plt.xlabel('Rok')
plt.ylabel('średnia długość życia')
plt.title('Zmiana średniej długości życia Polaków')
plt.plot(Poland);
```

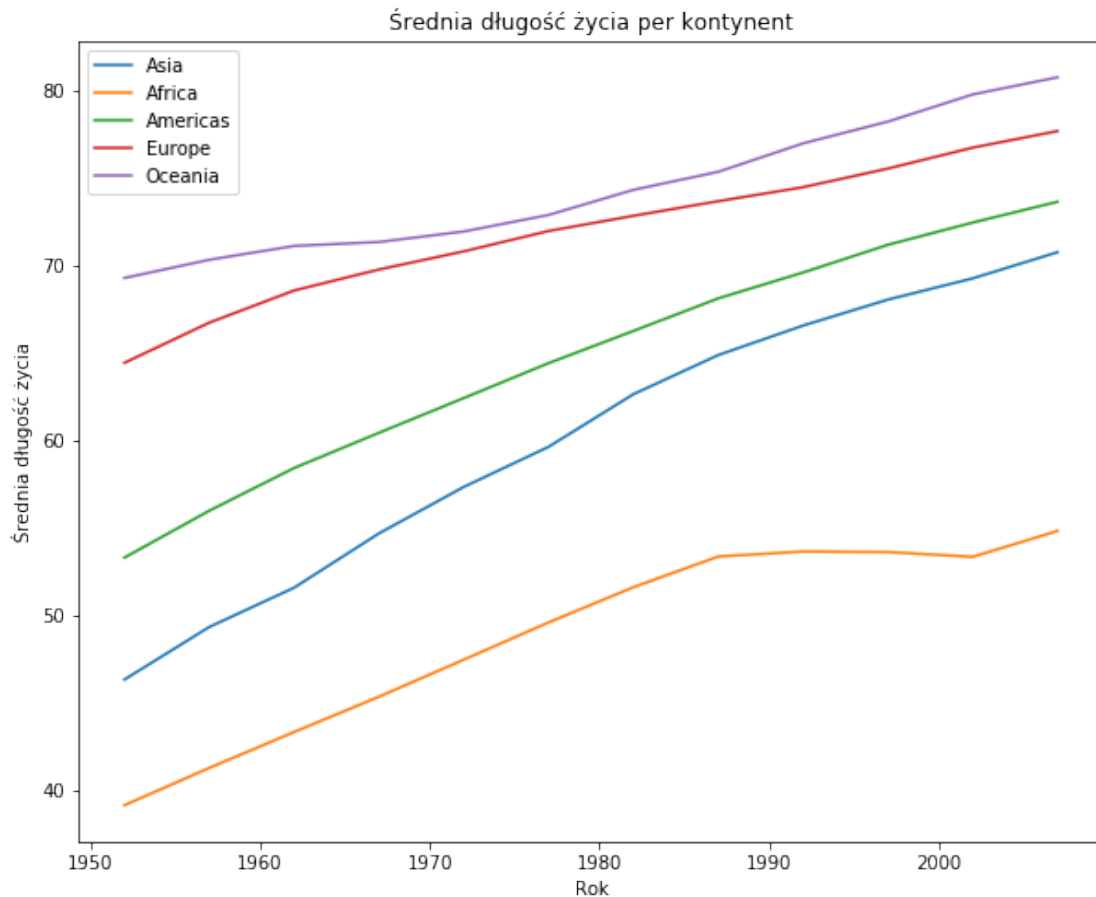


1.3.2 Statystyki - wiat

Zmiana długości życia per kontynent

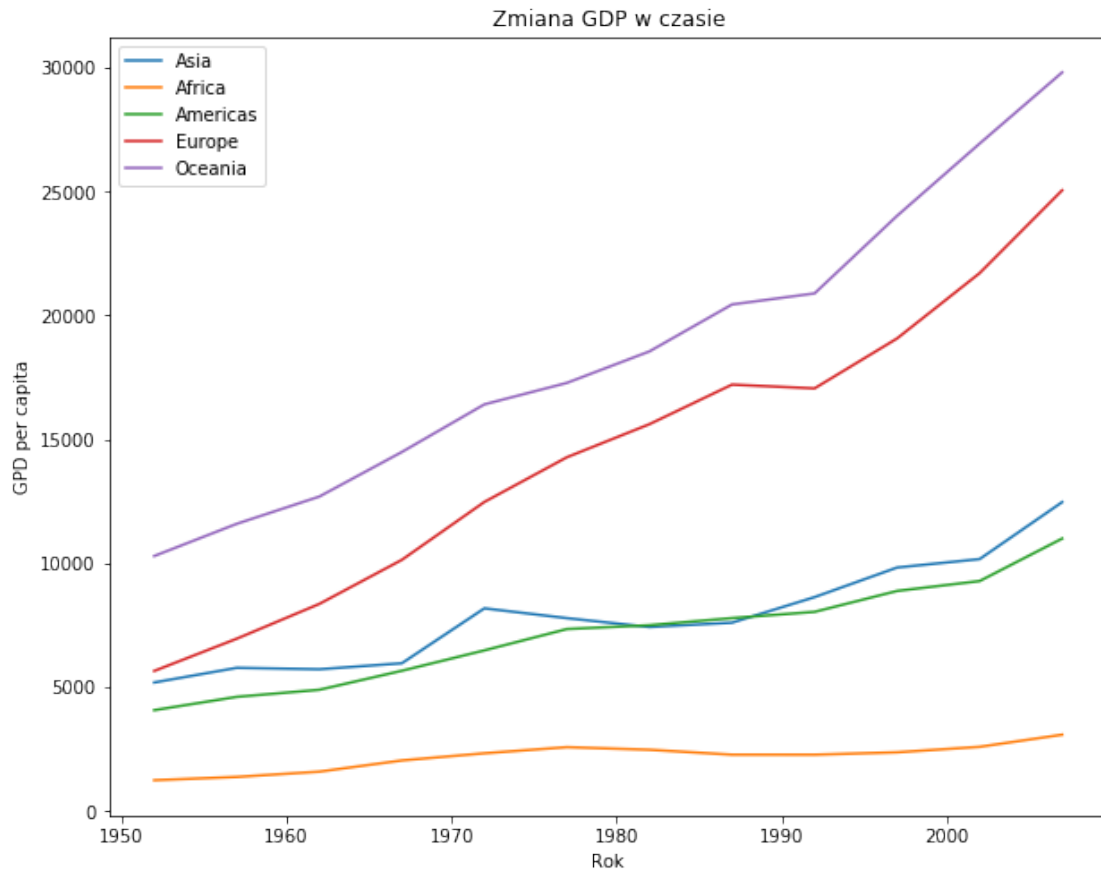
```
In [206]: chart3=df.pivot_table(values='lifeExp',index='year',columns='continent')
continents = chart3[['Asia','Africa', 'Americas', 'Europe', 'Oceania']]
plt.figure(figsize=(10,8))
plt.xlabel('Rok')
```

```
plt.ylabel('rednia dugo ycia')
lines=plt.plot(continents)
plt.title('rednia dugo ycia per kontynent')
plt.legend(['Asia', 'Africa', 'Americas', 'Europe', 'Oceania']);
```



Zmiany w redniej wartoci GDP per capita w zalenoci od kontynentu

```
In [207]: chart4 = df.pivot_table(values='gdpPerCap',index='year',columns='continent')
continents = chart4[['Asia', 'Africa', 'Americas', 'Europe', 'Oceania']]
plt.figure(figsize=(10,8))
plt.xlabel('Rok')
plt.ylabel('GPD per capita')
lines=plt.plot(continents)
plt.title('Zmiana GDP w czasie')
plt.legend(['Asia', 'Africa', 'Americas', 'Europe', 'Oceania']);
```

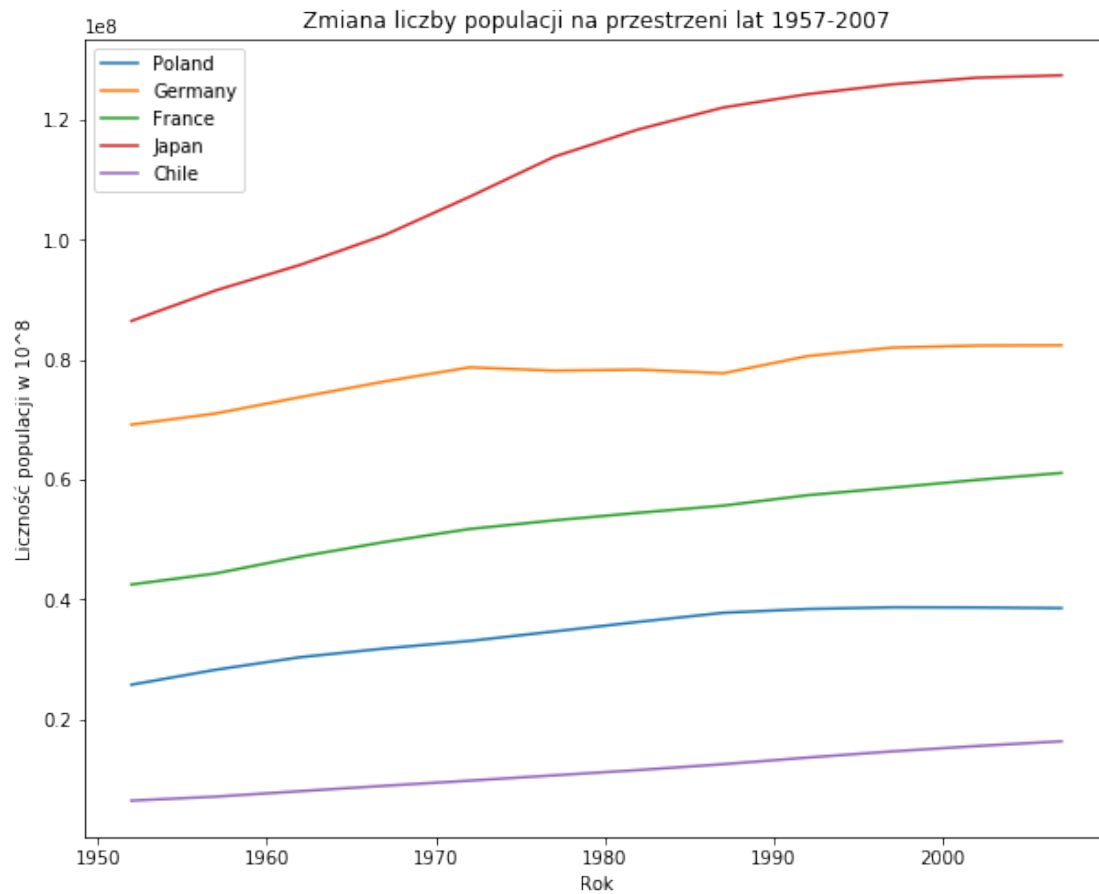


1.4 Zadanie dodatkowe

W tej czci czekamy na Twoje pomysły odnośnie analizy tego zbioru danych. Opisz dokładnie co chcesz sprawdzić i na jakie pytania szukasz odpowiedzi.

Zmiana liczby populacji pięciu krajów w zależności od czasu

```
In [225]: plt.figure(figsize=(10,8))
          chart5=df.pivot_table(values='pop',index='year',columns='country',)
          countries=chart5[['Poland','Germany','France','Japan','Chile']]
          plt.xlabel('Rok')
          plt.ylabel('Liczba populacji w 10^8')
          plt.title('Zmiana liczby populacji na przestrzeni lat 1957-2007')
          lines=plt.plot(countries)
          plt.legend(['Poland','Germany','France','Japan','Chile']);
```



In []: