eiwd lab3

March 23, 2023

Instalacja wymaganych pakietów

```
[99]: | pip install pyspark==3.0.1 py4j==0.10.9
```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/

Requirement already satisfied: pyspark==3.0.1 in /usr/local/lib/python3.9/dist-packages (3.0.1)

Requirement already satisfied: py4j==0.10.9 in /usr/local/lib/python3.9/dist-packages (0.10.9)

Tworzenie Spark session

Podłączenie dysku google

```
[101]: from google.colab import drive drive.mount('/content/drive')

!ls "/content/drive/My Drive"
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Classroom 'Colab Notebooks' IHME_HEALTH_SPENDING_1995_2018_Y2021M09D22.CSV

Czytanie danych z pliku

```
[102]: # Czytanie z pliku csv
csv_file = "/content/drive/My Drive/IHME_HEALTH_SPENDING_1995_2018_Y2021M09D22.

GCSV"

df = spark.read.csv(csv_file)
```

Strukturyzacja danych za pomocą schematu Spark

```
[109]: data = spark.read.csv(
             "/content/drive/My Drive/IHME_HEALTH_SPENDING_1995_2018_Y2021M09D22.CSV",
             sep=',',
             header=True,
             )
       data.printSchema()
      root
       |-- location_id: string (nullable = true)
       |-- location_name: string (nullable = true)
       |-- iso3: string (nullable = true)
       |-- level: string (nullable = true)
       |-- year: string (nullable = true)
       |-- the_total_mean: string (nullable = true)
       |-- the total lower: string (nullable = true)
       |-- the_total_upper: string (nullable = true)
       |-- the total ppp mean: string (nullable = true)
       |-- the_total_ppp_lower: string (nullable = true)
       |-- the_total_ppp_upper: string (nullable = true)
       |-- ghes_total_mean: string (nullable = true)
       |-- ghes_total_lower: string (nullable = true)
       |-- ghes_total_upper: string (nullable = true)
       |-- ghes_total_ppp_mean: string (nullable = true)
       |-- ghes_total_ppp_lower: string (nullable = true)
       |-- ghes_total_ppp_upper: string (nullable = true)
       |-- ppp_total_mean: string (nullable = true)
       |-- ppp_total_lower: string (nullable = true)
       |-- ppp_total_upper: string (nullable = true)
       |-- ppp_total_ppp_mean: string (nullable = true)
       |-- ppp total ppp lower: string (nullable = true)
       |-- ppp_total_ppp_upper: string (nullable = true)
       |-- oop total mean: string (nullable = true)
       |-- oop_total_lower: string (nullable = true)
       |-- oop_total_upper: string (nullable = true)
       |-- oop_total_ppp_mean: string (nullable = true)
       |-- oop_total_ppp_lower: string (nullable = true)
       |-- oop_total_ppp_upper: string (nullable = true)
       |-- dah_total_mean: string (nullable = true)
       |-- dah_total_ppp_mean: string (nullable = true)
       |-- the_per_cap_mean: string (nullable = true)
       |-- the_per_cap_lower: string (nullable = true)
       |-- the_per_cap_upper: string (nullable = true)
       |-- the_per_cap_ppp_mean: string (nullable = true)
       |-- the_per_cap_ppp_lower: string (nullable = true)
       |-- the per cap ppp upper: string (nullable = true)
       |-- ghes_per_cap_mean: string (nullable = true)
       |-- ghes_per_cap_lower: string (nullable = true)
```

```
|-- ghes_per_cap_upper: string (nullable = true)
|-- ghes_per_cap_ppp_mean: string (nullable = true)
|-- ghes_per_cap_ppp_lower: string (nullable = true)
|-- ghes_per_cap_ppp_upper: string (nullable = true)
|-- ppp_per_cap_mean: string (nullable = true)
|-- ppp_per_cap_lower: string (nullable = true)
|-- ppp_per_cap_upper: string (nullable = true)
|-- ppp_per_cap_ppp_mean: string (nullable = true)
|-- ppp_per_cap_ppp_lower: string (nullable = true)
|-- ppp_per_cap_ppp_upper: string (nullable = true)
|-- oop_per_cap_mean: string (nullable = true)
|-- oop_per_cap_lower: string (nullable = true)
|-- oop_per_cap_upper: string (nullable = true)
|-- oop_per_cap_ppp_mean: string (nullable = true)
|-- oop_per_cap_ppp_lower: string (nullable = true)
|-- oop_per_cap_ppp_upper: string (nullable = true)
|-- dah_per_cap_mean: string (nullable = true)
|-- dah_per_cap_ppp_mean: string (nullable = true)
|-- ghes_per_the_mean: string (nullable = true)
|-- ghes per the lower: string (nullable = true)
|-- ghes_per_the_upper: string (nullable = true)
|-- ppp per the mean: string (nullable = true)
|-- ppp_per_the_lower: string (nullable = true)
|-- ppp_per_the_upper: string (nullable = true)
|-- oop_per_the_mean: string (nullable = true)
|-- oop_per_the_lower: string (nullable = true)
|-- oop_per_the_upper: string (nullable = true)
|-- dah_per_the_mean: string (nullable = true)
|-- dah_per_the_lower: string (nullable = true)
|-- dah_per_the_upper: string (nullable = true)
|-- the_per_gdp_mean: string (nullable = true)
|-- the_per_gdp_lower: string (nullable = true)
|-- the_per_gdp_upper: string (nullable = true)
|-- ghes_per_gdp_mean: string (nullable = true)
|-- ghes per gdp lower: string (nullable = true)
|-- ghes_per_gdp_upper: string (nullable = true)
|-- ppp_per_gdp_mean: string (nullable = true)
|-- ppp_per_gdp_lower: string (nullable = true)
|-- ppp_per_gdp_upper: string (nullable = true)
|-- oop_per_gdp_mean: string (nullable = true)
|-- oop_per_gdp_lower: string (nullable = true)
|-- oop_per_gdp_upper: string (nullable = true)
|-- dah_per_gdp_mean: string (nullable = true)
|-- dah_per_gdp_lower: string (nullable = true)
|-- dah_per_gdp_upper: string (nullable = true)
```

Manualna strukturyzacja danych

```
[104]: from pyspark.sql.types import *
       data_schema = [
           StructField('location_id', IntegerType(), True), #czy dopuszczalna jest_
        →wartość null
           StructField('location_name', StringType(), True),
           StructField('year', IntegerType(), True),
           StructField('the_total_mean', DoubleType(), True),
       ]
       final_struc = StructType(fields = data_schema)
       data = spark.read.csv(
             '/content/drive/My Drive/IHME_HEALTH_SPENDING_1995_2018_Y2021M09D22.CSV',
             sep=',',
             header=True,
             schema=final_struc
       )
       data.printSchema()
      root
       |-- location_id: integer (nullable = true)
       |-- location_name: string (nullable = true)
       |-- year: integer (nullable = true)
       |-- the_total_mean: double (nullable = true)
      Kontrola danych
[105]: data.schema
[105]: StructType(List(StructField(location_id,IntegerType,true),StructField(location_n
       ame, StringType, true), StructField(year, IntegerType, true), StructField(the_total_me
       an,DoubleType,true)))
[106]: data.dtypes
[106]: [('location_id', 'int'),
        ('location_name', 'string'),
        ('year', 'int'),
        ('the_total_mean', 'double')]
[107]: data.head
[107]: <bound method DataFrame.head of DataFrame[location_id: int, location_name:
       string, year: int, the_total_mean: double]>
      Manipulacja kolumnami
```

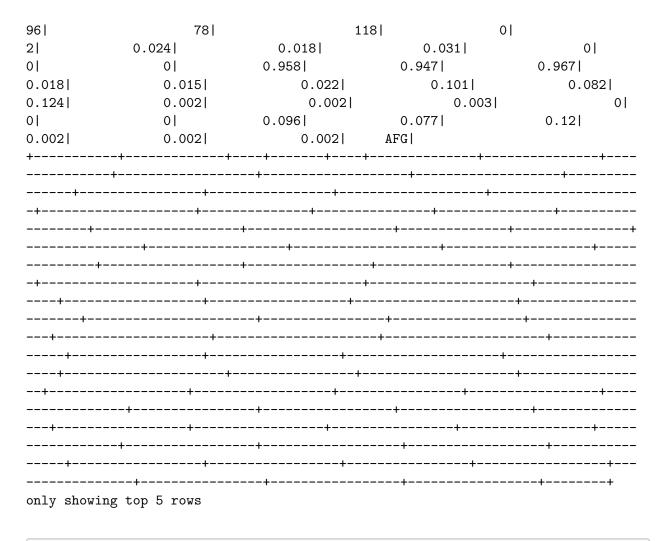
[110]: data = data.withColumn('iso3_new', data.iso3) #kopia data.show(5)

4									
								+	
								+	
	+			+		+-			-+
	+		+-				+-		
-+		+			+			+	
								+	
								-+	
								' +	
								+	
				+		-+		+	
+		+			+		+		-+
	+			+		+		+	
								+	

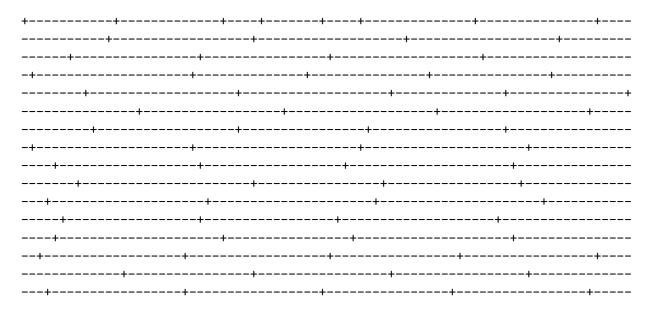
|location_id|location_name|iso3| level|year|the_total_mean|the_total_lower|the_ total_upper|the_total_ppp_mean|the_total_ppp_lower|the_total_ppp_upper|ghes_tota l_mean|ghes_total_lower|ghes_total_upper|ghes_total_ppp_mean|ghes_total_ppp_lowe r|ghes_total_ppp_upper|ppp_total_mean|ppp_total_lower|ppp_total_upper|ppp_total_ ppp_mean|ppp_total_ppp_lower|ppp_total_ppp_upper|oop_total_mean|oop_total_lower| oop_total_upper|oop_total_ppp_mean|oop_total_ppp_lower|oop_total_ppp_upper|dah_t otal_mean|dah_total_ppp_mean|the_per_cap_mean|the_per_cap_lower|the_per_cap_uppe r|the_per_cap_ppp_mean|the_per_cap_ppp_lower|the_per_cap_ppp_upper|ghes_per_cap_ mean|ghes_per_cap_lower|ghes_per_cap_upper|ghes_per_cap_ppp_mean|ghes_per_cap_pp p_lower|ghes_per_cap_ppp_upper|ppp_per_cap_mean|ppp_per_cap_lower|ppp_per_cap_up per|ppp_per_cap_ppp_mean|ppp_per_cap_ppp_lower|ppp_per_cap_ppp_upper|oop_per_cap _mean|oop_per_cap_lower|oop_per_cap_upper|oop_per_cap_ppp_mean|oop_per_cap_ppp_l ower|oop_per_cap_ppp_upper|dah_per_cap_mean|dah_per_cap_ppp_mean|ghes_per_the_me an|ghes_per_the_lower|ghes_per_the_upper|ppp_per_the_mean|ppp_per_the_lower|ppp_ per_the_upper|oop_per_the_mean|oop_per_the_lower|oop_per_the_upper|dah_per_the_m ean|dah_per_the_lower|dah_per_the_upper|the_per_gdp_mean|the_per_gdp_lower|the_p er_gdp_upper|ghes_per_gdp_mean|ghes_per_gdp_lower|ghes_per_gdp_upper|ppp_per_gdp _mean|ppp_per_gdp_lower|ppp_per_gdp_upper|oop_per_gdp_mean|oop_per_gdp_lower|oop _per_gdp_upper|dah_per_gdp_mean|dah_per_gdp_lower|dah_per_gdp_upper|iso3_new|

	-+		-+	-+		+	
-+		·		•			•
•		•	+-			•	
+			+	•		•	
+		•		•			•
'		•				'	
•		·	+			•	
		+	+		+		+
+							
	+-		+		-+		+
+		·	+		•		•
I	 160	•	+AFG Country 1995		-+ 528409		-++ 417121
665425	1001	2193179	173127		·	1865	1111211
14643		11814	17779	·	60775		
49036		73792	20		0		51
81		0	212		509515		398261
647195		2114759	165299	5	268	6205	
4232		17563	30		24		38
125		99		158		1	
1		1	31		0.1	3	0.1
4 0		0	0		0		0
121		0 95	29	154	23	0	37
1		0.028	0.02	1011	0.037	01	01
0		0	0.964		0.953		0.973
0.008		0.006	0.01		0.104		0.08
0.133		0.003	0.002		0.004	.	01
0		0	0.1		0.076		0.129
0.001		0.001	0.001		FG		
1	160	-	AFG Country 1996		516915		412035
651034		2145473	171016	35 I		2136	
13771		11138	16872		57159		FO.1
46227		70026 0	25 217		0 497261		52
102 632006		2063899	162422			3161	391328
5858		24313	28	. 1	23	.51011	36
118		94	201	148	201	1	001
1		1	3	•		3	
4		0	0		0		0
0		0	27		21		35
113		89		144		0	
1		0.027	0.019		0.036		01
0		0	0.961		0.95		0.971
0.011		0.009	0.014		0.099	. 1	0.076
0.128 0		0.003	0.002 0.096		0.003	1	0.124
VΙ		0	0.0301		0.072		U.124

0.001		0.001	0.001	AFG		
0.0011	160		AFG Country 1997			413424
636436	1001	2116248	1715930		1549	4104241
12864		10480	15698	53392	10101	
43499		65155	26	0		51
10133		0	211	485094		387884
610799		2013397	1609925		5139	0070017
11890		49350	27	22	01001	34
112		91	14		1	011
1		1	3	~ 1	2	
3		0	0	0		01
0		01	26	21		32
107		85	13		1	021
3		0.026	0.018	0.034		0
0		0	0.951	0.938		0.962
0.024		0.019	0.029	0.101		0.079
0.129		0.003	0.002	0.003	I	01
0		0	0.096	0.075	•	0.124
0.002		0.002	0.003	AFG		
	160		AFG Country 1998			394629
605041	,	2015335	1637919		1240	
12020		9689	14652	49889		
40216		60812	25	0		49
105		01	202	468699		377467
587578		1945349	1566688		8759	·
3013101		ISTUUTS	10000001	270	01001	
					01091	32
4817 105		19991 85	25	21	1	32
4817		19991	25	21		32
4817 105		19991 85	25 13	21	1	32 0
4817 105 1		19991 85 1	25 13 3	21	1	
4817 105 1 3		19991 85 1 0	25 13 3 0	21 1 0 20	1	0
4817 105 1 3 0		19991 85 1 0 0	25 13 3 0 24 12 0.018	21 1 0 20	1 2	0
4817 105 1 3 0 102		19991 85 1 0 0 82	25 13 3 0 24 12	21 1 0 20 7	1 2	0 31
4817 105 1 3 0 102 1		19991 85 1 0 0 82 0.025	25 13 3 0 24 12 0.018	21 1 0 20 7 0.033	1 2	0 31 0
4817 105 1 3 0 102 1 0		19991 85 1 0 0 82 0.025 0	25 13 3 0 24 0.018 0.965	21 1 0 20 7 0.033 0.955	1 2 0	0 31 0 0.973
4817 105 1 3 0 102 1 0 0.01		19991 85 1 0 0 82 0.025 0 0.008 0.002	25 13 3 3 0 24 12 12 12 12 12 12 12	21 1 0 20 7 0.033 0.955 0.1	1 2 0	0 31 0 0.973 0.08
4817 105 1 3 0 102 1 0 0 .01 0 .126		19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001	25 13 3 3 0 24 12 12 12 12 12 12 12	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG	1 2 0	0 31 0 0.973 0.08 0
4817 105 1 3 0 102 1 0 0 .01 0 .126 0	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan	25 3 0 24 12 0.018 0.965 0.012 0.002 0.096 0.001 AFG Country 1999	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720	1 2 0	0 31 0 0.973 0.08 0
4817 105 1 3 0 102 1 0 0 .01 0 .126 0 0 .001	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001	25 13 3 3 0 24 12 12 12 12 12 12 12	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720	1 2 0	0 31 0 0.973 0.08 0 0.123
4817 105 1 3 0 102 1 0 0 .01 0 .126 0 0 .001 564170 11055	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943	25 13 3 0 24 12 0.018 0.965 0.012 0.002 0.096 0.001 AFG Country 1999 1588504 13521	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234 45884	1 2 0	0 31 0 0.973 0.08 0 0.123 382723
4817 105 1 3 0 102 1 0 0 .01 0 .126 0 0 .001 564170	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119	25 3 0 24 12 0.018 0.965 0.012 0.002 0.096 0.001 AFG Country 1999 1588504	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234	1 2 0	0 31 0 0.973 0.08 0 0.123
4817 105 1 3 0 102 1 0 0 .01 0 .126 0 0 .001 564170 11055 37119 104	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119 0	25 13 3 0 24 12 0.018 0.965 0.012 0.002 0.096 0.001 AFG Country 1999 1588504 13521	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234 45884 0 444298	1 2 0 	0 31 0 0.973 0.08 0 0.123 382723
4817 105 1 3 0 102 1 0 0 102 1 0 0 102 1 0 105 37119 104 545080	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119 0 1844071	25	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234 45884 0 444298 226	1 2 0 	0 31 0 0.973 0.08 0 0.123 382723 46 362573
4817 105 1 3 0 102 1 0 0 102 1 0 0 102 1 0 105 105 105 105 104 105 104 105 104 105 104 105	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119 0 1844071 34626	25	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234 45884 0 444298 226: 20	1 2 0 1603	0 31 0 0.973 0.08 0 0.123 382723
4817 105 1 3 0 102 1 0 0 .01 0 .126 0 0 .001 564170 11055 37119 104 545080 8343 100	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119 0 1844071 34626 83	25	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234 45884 0 444298 226: 20	1 2 0 	0 31 0 0.973 0.08 0 0.123 382723 46 362573
4817 105 1 3 0 102 1 0 0 102 1 0 0 102 1 0 105 37119 104 545080 8343 100 0	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119 0 1844071 34626 83 1	25	21 1	1 2 0 1603	0 31 0,973 0.08 0 0.123 382723 46 362573
4817 105 1 3 0 102 1 0 0 .01 0 .126 0 0 .001 564170 11055 37119 104 545080 8343 100	160	19991 85 1 0 0 82 0.025 0 0.008 0.002 0 0.001 Afghanistan 1924685 8943 56119 0 1844071 34626 83	25	21 1 0 20 7 0.033 0.955 0.1 0.003 0.076 AFG 463720 234 45884 0 444298 226: 20	1 2 0 	0 31 0 0.973 0.08 0 0.123 382723 46 362573







|location_id|location_name|iso3| level|year|the_total_mean|the_total_lower|the_ total_upper|the_total_ppp_mean|the_total_ppp_lower|the_total_ppp_upper|ghes_tota l_mean|ghes_total_lower|ghes_total_upper|ghes_total_ppp_mean|ghes_total_ppp_lowe r|ghes_total_ppp_upper|ppp_total_mean|ppp_total_lower|ppp_total_upper|ppp_total_ ppp_mean|ppp_total_ppp_lower|ppp_total_ppp_upper|oop_total_mean|oop_total_lower| oop_total_upper|oop_total_ppp_mean|oop_total_ppp_lower|oop_total_ppp_upper|dah_t otal_mean|dah_total_ppp_mean|the_per_cap_mean|the_per_cap_lower|the_per_cap_uppe r|the_per_cap_ppp_mean|the_per_cap_ppp_lower|the_per_cap_ppp_upper|ghes_per_cap_ mean|ghes_per_cap_lower|ghes_per_cap_upper|ghes_per_cap_ppp_mean|ghes_per_cap_pp p_lower|ghes_per_cap_ppp_upper|ppp_per_cap_mean|ppp_per_cap_lower|ppp_per_cap_up per|ppp_per_cap_ppp_mean|ppp_per_cap_ppp_lower|ppp_per_cap_ppp_upper|oop_per_cap _mean|oop_per_cap_lower|oop_per_cap_upper|oop_per_cap_ppp_mean|oop_per_cap_ppp_1 ower|oop_per_cap_ppp_upper|dah_per_cap_mean|dah_per_cap_ppp_mean|ghes_per_the_me an|ghes_per_the_lower|ghes_per_the_upper|ppp_per_the_mean|ppp_per_the_lower|ppp_ per_the_upper|oop_per_the_mean|oop_per_the_lower|oop_per_the_upper|dah_per_the_m ean|dah_per_the_lower|dah_per_the_upper|the_per_gdp_mean|the_per_gdp_lower|the_p er_gdp_upper|ghes_per_gdp_mean|ghes_per_gdp_lower|ghes_per_gdp_upper|ppp_per_gdp _mean|ppp_per_gdp_lower|ppp_per_gdp_upper|oop_per_gdp_mean|oop_per_gdp_lower|oop _per_gdp_upper|dah_per_gdp_mean|dah_per_gdp_lower|dah_per_gdp_upper|iso3_new_mar

K2					
+	+-	+	+		
			+		
+-					
-+				+	+
			-+		
		-+			
	-+		-++		
			+		
+			+	+	
+			+	+	
			+-		
+				+	
+		+	+		
+					
	+		+	-++	
	+-		+	+	-+
+			+		
			+		
		T			T
+					
	160	Afghanistan	AFG Country 1995	528409	417121
6654251	•	_	1731272		
000425		21931791	1/312/2	2761865	

201

60775|

0|

51|

17779|

11814

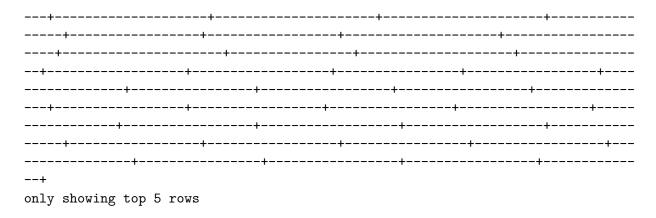
737921

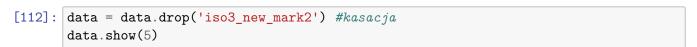
14643|

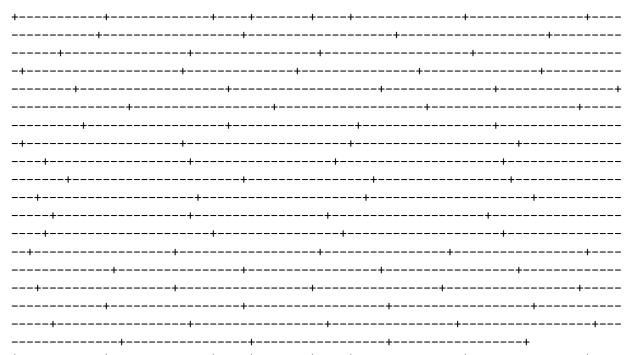
490361

81		01	212		509515		398261
647195		2114759	16529	95	268	6205	
4232		17563	30		24		38
125		99	- 1	158		1	
1		1	3		- 1	3	
4		0	0		0		0
0		0	29	4 - 4	23	0.1	37
121		95	0.001	154	0.0071	0	0.1
1		0.028	0.02		0.037		0
0		0	0.964		0.953		0.973
0.008 0.133		0.006 0.003	0.01 0.002		0.104		0.08 0
0.1331		0.0031	0.002		0.004		0.129
0.001		0.001	0.001		AFG		0.129
0.001	1601		AFG Country 199				412035
651034	1001	2145473	•			2136	4120331
13771		11138	16872		57159	21301	
46227		70026	25		0		52
1022/		01	217		497261		
632006		2063899	16242		262		0310201
5858		24313	28	211	231	01011	36
118		94	201	148	201	1	001
1		1	3			3	
4		0	0		01	01	0
0		0	27		21		35
113		89	•	144		0	
1		0.027	0.019		0.036		0
0		0	0.961		0.95		0.971
0.011		0.009	0.014		0.099		0.076
0.128		0.003	0.002		0.003		0
0		0	0.096		0.072		0.124
0.001		0.001	0.001		AFG		
1	160	Afghanistan	AFG Country 199	7	509874		413424
636436		2116248	17159	30	264	1549	
12864		10480	15698		53392		
43499		65155	26		0		51
108		0	211		485094		387884
610799		2013397	16099	25	253	5139	
11890		49350	27		22		34
112		91		140		1	
1		1	31			2	
3		0	0		01		01
0		0	26		21		32
107		85		134		1	
3		0.026	0.018		0.034		0
0		0	0.951		0.938		0.962
0.024		0.019	0.029		0.101		0.079
0.129		0.003	0.002		0.003	I	01

0		0	0.096		0.075		0.124
0.002		0.002	0.003		AFG		
	160	Afghanistan	AFG Country 19	98	485561		394629
605041		2015335	1637	919	25:	11240	
12020		9689	14652		49889		
40216		60812	25		0		49
105		0	20	2	468699		377467
587578		1945349	1566	688	243	38759	
4817		19991	25		21		32
105		85		131		1	
1		1	3	3		2	
3		0	0		0		0
0		0	24		20		31
102		82		127		0	
1		0.025	0.018		0.033		0
0		0	0.965		0.955		0.973
0.01		0.008	0.012		0.1		0.08
0.126		0.002	0.002	:	0.003	3	0
0		0	0.096		0.076		0.123
0.001		0.001	0.001		AFG		
1	160	Afghanistan	AFG Country 19	99	463720		382723
564170		1924685	1588	504	234	41603	
11055		8943	13521		45884		
37119		56119	25		0		46
104		0	19	2	444298		362573
545080		1844071	1504	872	226	32371	
8343		34626	24		20		29
100		83		122		1	
0		1	2	!		2	
3		0	0		0		0
0		0	23		19		28
96		78		118		0	
2		0.024	0.018		0.031		0
0		0	0.958		0.947		0.967
0.018		0.015	0.022		0.101		0.082
0.124		0.002	0.002	:	0.003	3	0
0		0	0.096		0.077		0.12
0.002		0.002	0.002		AFG		
+	+	+	+	+	+-		
	+-		+		+		+
+-			+-			+	
-+			+		+		+
	+		-+		+	+	+
		+					
	-+		-+	+		+	
-+				+			+
+				+		+-	
+			+	+		+	







|location_id|location_name|iso3| level|year|the_total_mean|the_total_lower|the_total_upper|the_total_ppp_mean|the_total_ppp_lower|the_total_ppp_upper|ghes_total_lower|ghes_total_upper|ghes_total_ppp_mean|ghes_total_ppp_lower|ghes_total_ppp_total_upper|ppp_total_upper|ppp_total_upper|ppp_total_upper|ppp_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|dah_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|dah_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|oop_total_upper|dah_total_upper|oop_mean|the_per_cap_mean|the_per_cap_lower|the_per_cap_lower|the_per_cap_upper|dah_upper|oop_upper|dah_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_cap_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_upper_upper|oop_

an|ghes_per_the_lower|ghes_per_the_upper|ppp_per_the_mean|ppp_per_the_lower|ppp_ per_the_upper|oop_per_the_mean|oop_per_the_lower|oop_per_the_upper|dah_per_the_m ean|dah_per_the_lower|dah_per_the_upper|the_per_gdp_mean|the_per_gdp_lower|the_p er_gdp_upper|ghes_per_gdp_mean|ghes_per_gdp_lower|ghes_per_gdp_upper|ppp_per_gdp _mean|ppp_per_gdp_lower|ppp_per_gdp_upper|oop_per_gdp_mean|oop_per_gdp_lower|oop _per_gdp_upper|dah_per_gdp_mean|dah_per_gdp_lower|dah_per_gdp_upper| 528409 I 160| Afghanistan | AFG | Country | 1995 | 417121| 665425 I 2193179| 1731272| 2761865| 14643| 11814 17779| 60775 490361 737921 0| 51 l 201 81| 01 212| 509515| 398261 647195| 2114759| 1652995 26862051 42321 17563| 30| 24 381 125 991 158 l 1| 1| 1| 3| 3| 41 0| 0| 0| 0| 0| 0| 29| 23| 37| 95| 0| 121 154 1 l 0.028 0.021 0.0371 01 01 0.964| 0.953| 0.973| 0| 0.0081 0.0061 0.01 180.0 0.1041 0.003| 0| 0.133| 0.0021 0.004 01 0.1 0.0761 0.129| 01 0.001| 0.001| 0.001| Afghanistan | AFG | Country | 1996 | 516915| 412035| 651034 l 27021361 2145473| 1710165 13771 l 11138| 16872| 57159 I 46227| 70026| 25 l 01 521 102| 0| 217 497261 391328

632006		2063899	16242	21	262	3161		
5858		24313	28		23		36	3
118		94		148		1		
1		1	3			3		
4		0	0		0		0	
0		0	27		21		35	5
113		89		144		0		
1		0.027	0.019		0.036		0	
0		0	0.961		0.95		0.971	
0.011		0.009	0.014		0.099		0.076	3
0.128		0.003	0.002		0.003			0
0		0	0.096		0.072		0.124	
0.001		0.001	0.001					
	160	${ t Afghanistan} $	AFG Country 199	7	509874		413424	
636436		2116248	17159	30	264	1549		
12864		10480	15698		53392			
43499		65155	26		0		51	
108		0	211		485094		387884	
610799		2013397	16099	25	253	5139		
11890		49350	27		22		3	34
112		91		140		1		
1		1	3			2		
3		0	0		0		0	
0		0	26		21		32	2
107		85		134		1		
3		0.026	0.018		0.034		0	
0		0	0.951		0.938		0.962	
0.024		0.019	0.029		0.101		0.079	9
0.129		0.003	0.002		0.003			0
0		0	0.096		0.075		0.124	
0.002		0.002	0.003					
	160	Afghanistan	AFG Country 199	8	485561		394629	
605041		2015335	16379	19	251	1240		
12020		9689	14652		49889			
40216		60812	25		0		49	
105		0	202		468699		377467	
587578		1945349	15666	88	243	8759		
4817		19991	25		21		32	2
105		85		131		1		
1		1	3			2		
3		0	0		0		0	
0		0	24		20		31	L
102		82		127		0		
1		0.025	0.018		0.033		0	
0		0	0.965		0.955		0.973	
0.01		0.008	0.012		0.1		0.08	
0.126		0.002	0.002		0.003			0
0		0	0.096		0.076		0.123	

```
0.001
       0.001
              0.001
ı
   160|
     Afghanistan | AFG | Country | 1999 |
                     4637201
                           3827231
       1924685 l
564170 l
               1588504 l
                      23416031
11055|
       8943|
             13521
                     45884
371191
        56119 l
               25 l
                      01
                            46 l
                    444298|
104 l
        01
               192 l
                          362573|
545080
       1844071
               1504872
                      2262371
8343 l
       346261
               241
                      201
                             29 l
100 l
         831
                 122 l
                         1 l
                         21
01
       1 l
                21
      0|
             0|
3|
                     0|
                             0|
               23|
                      19|
01
        01
                             281
                        01
961
        781
                 118|
21
     0.024
             0.018
                    0.031
                            01
01
            0.958
                   0.9471
                          0.9671
0.0181
       0.015
              0.0221
                     0.101|
                            0.0821
0.124
       0.0021
              0.002|
                      0.003|
                              01
01
       01
            0.0961
                   0.077
                          0.12
0.0021
       0.002|
              0.0021
           --+----
   _____
 ----+-----+
 -----
 ____+
____+___
 __+____
  ______
 ------
 __+____
 -----+----+
only showing top 5 rows
```

Obróbka brakujących wartości

```
[113]: from pyspark.sql import functions as f

#usuwanie wierszy z brakującymi wartościami
data.na.drop()

#zastąpienie brakującej wartości za pomocą średniej
data.na.fill(data.select(f.mean(data['ghes_per_cap_upper'])).collect()[0][0])
```

```
#zastąpienie brakujących wartości nowymi
#data.na.replace(old_value, new_vallue)
```

```
[113]: DataFrame[location_id: string, location_name: string, iso3: string, level:
      string, year: string, the_total_mean: string, the_total_lower: string,
      the_total_upper: string, the_total_ppp_mean: string, the_total_ppp_lower:
      string, the_total_ppp_upper: string, ghes_total_mean: string, ghes_total_lower:
      string, ghes_total_upper: string, ghes_total_ppp_mean: string,
      ghes_total_ppp_lower: string, ghes_total_ppp_upper: string, ppp_total_mean:
      string, ppp_total_lower: string, ppp_total_upper: string, ppp_total_ppp_mean:
      string, ppp_total_ppp_lower: string, ppp_total_ppp_upper: string,
      oop_total_mean: string, oop_total_lower: string, oop_total_upper: string,
      oop_total_ppp_mean: string, oop_total_ppp_lower: string, oop_total_ppp_upper:
      string, dah_total_mean: string, dah_total_ppp_mean: string, the_per_cap_mean:
      string, the_per_cap_lower: string, the_per_cap_upper: string,
      the_per_cap_ppp_mean: string, the_per_cap_ppp_lower: string,
      the_per_cap_ppp_upper: string, ghes_per_cap_mean: string, ghes_per_cap_lower:
      string, ghes_per_cap_upper: string, ghes_per_cap_ppp_mean: string,
      ghes_per_cap_ppp_lower: string, ghes_per_cap_ppp_upper: string,
      ppp_per_cap_mean: string, ppp_per_cap_lower: string, ppp_per_cap_upper: string,
      ppp_per_cap_ppp_mean: string, ppp_per_cap_ppp_lower: string,
      ppp_per_cap_ppp_upper: string, oop_per_cap_mean: string, oop_per_cap_lower:
      string, oop_per_cap_upper: string, oop_per_cap_ppp_mean: string,
      oop_per_cap_ppp_lower: string, oop_per_cap_ppp_upper: string, dah_per_cap_mean:
      string, dah_per_cap_ppp_mean: string, ghes_per_the_mean: string,
      ghes_per_the_lower: string, ghes_per_the_upper: string, ppp_per_the_mean:
      string, ppp_per_the_lower: string, ppp_per_the_upper: string, oop_per_the_mean:
      string, oop_per_the_lower: string, oop_per_the_upper: string, dah_per_the_mean:
      string, dah_per_the_lower: string, dah_per_the_upper: string, the_per_gdp_mean:
      string, the per_gdp_lower: string, the per_gdp_upper: string, ghes_per_gdp_mean:
      string, ghes_per_gdp_lower: string, ghes_per_gdp_upper: string,
      ppp_per_gdp_mean: string, ppp_per_gdp_lower: string, ppp_per_gdp_upper: string,
      oop_per_gdp_mean: string, oop_per_gdp_lower: string, oop_per_gdp_upper: string,
      dah_per_gdp_mean: string, dah_per_gdp_lower: string, dah_per_gdp_upper: string]
```

Pobieranie danych

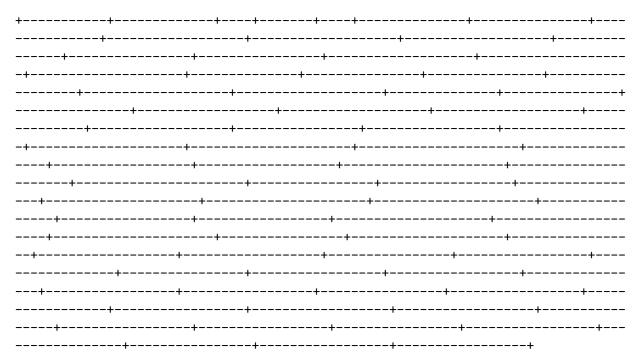
Select

```
[114]: data.select(['location_name', 'year', 'the_total_mean']).show(5)
```

```
+-----+
|location_name|year|the_total_mean|
+-----+
| Afghanistan|1995| 528409|
| Afghanistan|1996| 516915|
| Afghanistan|1997| 509874|
| Afghanistan|1998| 485561|
```

Filter

```
[115]: #filtrowanie na podstawie warunku & / ~(not)
from pyspark.sql.functions import col
data.filter((col('year') >= 2000) & (col('year') <= 2003)).show(5)
```



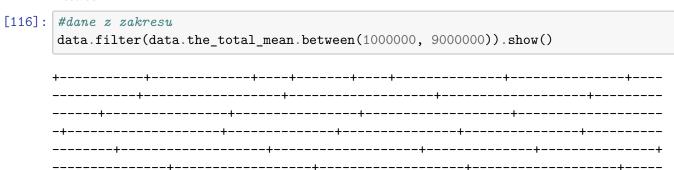
|location_id|location_name|iso3| level|year|the_total_mean|the_total_lower|the_ total_upper|the_total_ppp_mean|the_total_ppp_lower|the_total_ppp_upper|ghes_tota l_mean|ghes_total_lower|ghes_total_upper|ghes_total_ppp_mean|ghes_total_ppp_lowe r|ghes_total_ppp_upper|ppp_total_mean|ppp_total_lower|ppp_total_upper|ppp_total_ ppp_mean|ppp_total_ppp_lower|ppp_total_ppp_upper|oop_total_mean|oop_total_lower| oop_total_upper|oop_total_ppp_mean|oop_total_ppp_lower|oop_total_ppp_upper|dah_t otal_mean|dah_total_ppp_mean|the_per_cap_mean|the_per_cap_lower|the_per_cap_uppe r|the_per_cap_ppp_mean|the_per_cap_ppp_lower|the_per_cap_ppp_upper|ghes_per_cap_ mean|ghes_per_cap_lower|ghes_per_cap_upper|ghes_per_cap_ppp_mean|ghes_per_cap_pp p_lower|ghes_per_cap_ppp_upper|ppp_per_cap_mean|ppp_per_cap_lower|ppp_per_cap_up per|ppp_per_cap_ppp_meam|ppp_per_cap_ppp_lower|ppp_per_cap_ppp_upper|oop_per_cap _mean|oop_per_cap_lower|oop_per_cap_upper|oop_per_cap_ppp_mean|oop_per_cap_ppp_1 ower|oop_per_cap_ppp_upper|dah_per_cap_mean|dah_per_cap_ppp_mean|ghes_per_the_me an|ghes_per_the_lower|ghes_per_the_upper|ppp_per_the_mean|ppp_per_the_lower|ppp_ per_the_upper|oop_per_the_mean|oop_per_the_lower|oop_per_the_upper|dah_per_the_m ean|dah_per_the_lower|dah_per_the_upper|the_per_gdp_mean|the_per_gdp_lower|the_p er_gdp_upper|ghes_per_gdp_mean|ghes_per_gdp_lower|ghes_per_gdp_upper|ppp_per_gdp _mean|ppp_per_gdp_lower|ppp_per_gdp_upper|oop_per_gdp_mean|oop_per_gdp_lower|oop

_per_gdp	_uppe:	r dah_per_gdp_1	mean dah_per_gdp_	lowe	r dah_per_gdp	_uppei	r	
+	+	+	+	+	+-		+	
	+-		+		+		+	
+-			+		+			
-+					+		+	
	+		-+		+	+		+
		+						
	-+		-+	-+		+		
-+				+			+	
+			+-			+		
+			+	+		+-		
+		+		+-			+	
+			+			-+		
•		•	+			•		
			+					
		+	+		+		+	
•		·			·		•	
	+		+		-+		+	
+			+		•		+-	
		-+					-+	
 	160	<u> </u>	AFG Country 2000		446201		372565	
536362		1851971	154634	1		6189		
10382		8407	12686		43091			
34892		52654	25		01		44	
106		01	182		426131		352623	
516002		1768668	146357	1		1683		
9663		40106	23		19		28	
95		79		14		1		
0		1	2		- 1	2	- 1	
3		0	0		0		0	
0		01	221		18	- 1	27	
91		75		10		0		
2		0.023	0.018		0.031		0	
0		0	0.955		0.944		0.963	
0.022		0.018	0.026		0.102		0.084	
0.121		0.002	0.002		0.003			0
0		0	0.097		0.079		0.117	
0.002	4.00	0.002	0.002	1	447004		2040041	
 	1601	-	AFG Country 2001				381894	
524195		· · · ·	158506	Τ		5686		
10191		8281	12450		42298		401	
34369		51674	30		0		43	
126		0	177		421413		354700	
498677		1749090	147219		206	७।।ऽ।	0.0	ı
15686		65106	22		19	4 1	26	I
92		78		07		1		
0		1	2		0.1	2	0.1	
3		0	0		0		01	

0 86		0 73	21	1001	18	4.1	25	5
3		0.023	0.018	102	0.03	1	01	
0		0.023	0.942		0.031		0.951	
0.035		0.03	0.041		0.105		0.088	R I
0.122		0.002	0.002		0.003	3	0.000	0
0.1221		010021	0.099		0.083		0.116	01
0.004		0.004	0.004		0.0001		0.1101	
	160		AFG Country 200	21	534474		464153	l
612575		2218352	19264			12511		•
13506		10994	16525	•	56055	•		
45630		68586	45		0		65	
185		0)	4896961		419200	
567274		2032497	17399			54491		
31228		129614	25		22		6	28
103		90		118		1		
1		1	3			2		
3		0	0		0		0	
0		0	23		19		26	3
95		81		109		1		
6		0.025	0.02		0.033		0	
0		0	0.916		0.902		0.927	
0.059		0.051	0.067		0.082		0.071	L
0.094		0.002	0.002		0.003	3		0
0		0	0.075		0.064		0.087	
0.005		0.005	0.005					
1	160	Afghanistan	AFG Country 200	3	682593		601814	
770072		2833123	24978	349	319	96208		
29302		23856	36109		121617			
99016		149871	24		0		72	
99		0	300		606824		524018	
693301		2518643	21749	955		77567		
46443		192764	30		27		3	34
126		111		142		1		
1		2	5		- 1	4	_	
7		0	0		0		0	
0		0	27	4001	23	0.1	31	LI
112		97	0.0041	128	0.0541	2	0.1	
9		0.043	0.034		0.054		0	
0		0	0.889		0.872		0.903	. .
0.068		0.06	0.077		0.095	- 1	0.082	
0.107		0.004	0.003		0.005) I	0 0071	0
0 0.006		0 0.006	0.084 0.007		0.072		0.097	
1	43		ALB Country 200	\∩ I	549733		513401	ı
ı 587564	431	1532102	14308			37538	313401	ı
200778		177155	228589	/エ/	559567			
493730		637077			15	ı	31	
1001001		031011	221		101		011	

60	41	86	324109	2	297748
350560	903291	829822	97	7010	
24824	69185	172	161		184
480	448	513		63	
56	72	175		155	
200	0	0	0		0
0	0	102	93		110
283	260	306		8	
22	0.365	0.331	0.4		0
0	0	0.59	0.556	(0.623
0.045	0.042	0.048	0.076		0.071
0.081	0.028	0.024	0.032	1	0
0	0	0.045	0.041	(0.048
0.003	0.003	0.003			
•	•	+	•		•
		+	•		•
•	•		•		
·	•		•		•
•		·	•	•	•
	•	+	•		·
•		+ -		•	
•	•	+		'	
•	•	+		•	
•		+	•	•	
· ·	'	+			•
'		+		•	
•	•	+		•	
•	•		•		•
	•	+	•	•	
•	•		•		·
•		+	•		•
,	•	·	·		·
		+	+		F
only showing	top 5 rows				

Between



-+
+
+
+
+
+
+
+
location_id location_name iso3 level year the_total_mean the_total_lower the_total_upper the_total_ppp_mean the_total_ppp_lower the_total_ppp_upper ghes_total_upper ghes_total_upper ghes_total_ppp_upper ghes_total_ppp_lower flower the_total_ppp_upper ppp_total_mean ppp_total_upper ppp_total_ppp_upper oop_total_upper ppp_total_ppp_upper oop_total_upper oop_total_lower oop_total_upper oop_total_ppp_upper dah_total_mean dah_total_ppp_mean the_per_cap_mean the_per_cap_lower the_per_cap_upper dhah_total_mean dah_total_ppp_mean the_per_cap_mean the_per_cap_ppp_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper ghes_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_cap_upper oop_per_the_mean ghes_per_the_mean oop_per_the_mean oop_per_the_mean dah_per_the_mean dah_per_the_upper dah_per_the_mean dah_per_the_upper dah_per_the_mean dah_per_gdp_upper ghes_per_gdp_lower ppp_per_gdp_mean the_per_gdp_lower ppp_per_gdp_upper ppp_per_gdp_upp
+
+
-+
+
+
+
· · · · · · · · · · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·

+		+			+
		•	+	•	+
1	160	•	AFG Country 2007		956086
1141413		4318097	3968264	4737470)
56613		45313	70506	234974	
188072		292635		0	100
170		0	415	784067	700601
880017		3254296	2907866	3652539	
199651		828657	41	37	44
168		155	184		
2		3	9	7	
11		0	0	0	0
0		0	31	27	34
127		113	142	8	
32		0.054	0.044	0.067	0
0		0	0.753	0.73	0.776
0.192		0.175	0.209	0.104	0.091
0.119		0.006	0.004	0.007	0
0		0	0.079	0.067	0.092
0.02		0.018	0.021		
1	160	Afghanistan	AFG Country 2008	1066858	982415
1161963		4428026	4077542	4822762	2
58164		46184	71985	241412	
191690		298775	50	0	103
208		0	429	806163	725220
898419		3346004	3010050	3728915	
202481		840402	40	37	44
166		153	181	2	
2		3	9	7	
11		0	0	0	0
0		0	30	27	34
126		113	140	8	
32		0.055	0.044	0.067	0
0		0	0.755	0.734	0.776
0.19		0.174	0.206	0.103	0.093
0.115		0.006	0.004	0.007	0
0		0	0.078	0.069	0.089
0.02		0.018	0.021		
	160	Afghanistan	AFG Country 2009	1236730	1149054
1337448		5133088	4769186	5551121	1
64523		51524	79114	267803	
213851		328365	54	0	122
226		0	508	872071	784040
970584		3619560	3254182	4028439	
300082		1245499	45	42	49
186		173	202	2	
2		3	10	8	
12		0	0	0	0

0		0	32	28	35
131		118	146	11	
45		0.052	0.042	0.064	0
0		0	0.705	0.679	0.727
0.243		0.224	0.261	0.101	0.094
0.109		0.005	0.004	0.006	0
0		0		0.064	0.079
0.025	4001	0.024	0.025	40004001	44004501
	160	-	AFG Country 2010		
1364153		5238783	4841414	56619	60
65282		52686		270956	
218674		332348		0	135
270		01	562		807082
1006229		3756424	3349819		
291802		1211133	44	41	48
183		169	198		2
2		3	9		8
12		0	0	0	0
0		0	32	28	35
131		117	146	10	l
42		0.052	0.042	0.064	0
0		0	0.717	0.692	0.738
0.232		0.214	0.25	0.093	0.084
0.104		0.005	0.004	0.006	0
0		0	0.067	0.059	0.076
0.022		0.02	0.0021		
		0.021	0.023		
	160		AFG Country 2011	1373092	1273961
 1482720	160			1373092 61540	
 1482720 67603	160	Afghanistan	AFG Country 2011		
	160	Afghanistan 5699060	AFG Country 2011 5287617 82866	61540	
67603	160	Afghanistan 5699060 54439	AFG Country 2011 5287617 82866 71	61540 [°] 280589	77
67603 225952 295	160	Afghanistan 5699060 54439 343937	AFG Country 2011 5287617 82866 71 591	61540 280589 0 943247	77 142 846554
67603 225952	160	Afghanistan 5699060 54439 343937 0	AFG Country 2011 5287617 82866 71	61540 280589 0 943247 43539	77 142 846554
67603 225952 295 1049015 362170	160	Afghanistan 5699060 54439 343937 0 3914978 1503198	AFG Country 2011 5287617 82866 71 591 3513651 46	61540° 280589 0 943247 435390 43	77 142 846554 68 50
67603 225952 295 1049015 362170 192	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178	AFG Country 2011 5287617 82866 71 591 3513651 46 207	61540° 280589 0 943247 43539 43	77 142 846554 68 50
67603 225952 295 1049015 362170 192 2	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3	AFG Country 2011 5287617 82866 71 591 3513651 46 207	61540 280589 0 943247 43539 43	77 142 846554 68 50 2
67603 225952 295 1049015 362170 192 2 12	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9	61540° 280589 0 943247 435390 43 :	77 142 846554 68 50 2 8
67603 225952 295 1049015 362170 192 2 12 0	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32	61540° 280589 0 943247 43539 43 :	77 142 846554 68 50 2 8 0 35
67603 225952 295 1049015 362170 192 2 12 0 132	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32	61540° 280589 0 943247 435390 43 0 28 12	77 142 846554 68 50 2 8 0 35
67603 225952 295 1049015 362170 192 2 12 0 132 51	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04	61540° 280589 0 943247 435396 43 0 28 12	77 142 846554 68 50 2 8 0 35
67603 225952 295 1049015 362170 192 2 12 0 132 51 0	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687	61540° 280589 0 943247 435399 43 50 28 0 0.06 0.662	77 142 846554 68 50 2 8 0 35 0,71
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0 . 264	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284	61540° 280589 0 943247 43539 43 0 28 12 0.06 0.662 0.097	77 142 846554 68 50 2 8 0 35 0,71 0,089
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0 . 264 0 . 105	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244 0.005	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284 0.004	61540° 280589 0 943247 435396 43 0 28 0.06 0.662 0.097 0.006	771
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0.264 0.105 0	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244 0.005 0	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284 0.004 0.066	61540° 280589 0 943247 43539 43 0 28 12 0.06 0.662 0.097	77 142 846554 68 50 2 8 0 35 0,71 0,089
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0 . 264 0 . 105		Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244 0.005 0 0.025	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284 0.004 0.066 0.026	61540° 280589 0 943247 435390 43 0 28 0 0.06 0.662 0.097 0.006 0.059	77 142 846554 68 50 2 8 0 35 0,71 0.089 0,074
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0 . 264 0 . 105 0 0 . 025	160	Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244 0.005 0 0.025 Afghanistan	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284 0.004 0.066 0.026 AFG Country 2012	61540° 280589 0 943247 43539 43 0 28 0,06 0.662 0.097 0.006 0.059	77 142 846554 68 50 2 8 0 35 0,71 0.089 0,074 1299776
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0 . 264 0 . 105 0 0 . 025 1525283		Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244 0.005 0 0.025 Afghanistan 5854216	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284 0.004 0.066 0.026 AFG Country 2012 5394762	61540° 280589 0 943247 435396 43 0 28 0,06 0.06 0.062 0.097 0.006 0.059 1410474 633073	77 142 846554 68 50 2 8 0 35 0,71 0.089 0,074 1299776
67603 225952 295 1049015 362170 192 2 12 0 132 51 0 0 . 264 0 . 105 0 0 . 025		Afghanistan 5699060 54439 343937 0 3914978 1503198 178 3 0 0 118 0.049 0 0.244 0.005 0 0.025 Afghanistan	AFG Country 2011 5287617 82866 71 591 3513651 46 207 9 0 32 147 0.04 0.687 0.284 0.004 0.066 0.026 AFG Country 2012 5394762 91457	61540° 280589 0 943247 43539 43 0 28 0,06 0.662 0.097 0.006 0.059	77 142 846554 68 50 2 8 0 35 0,71 0.089 0,074 1299776

297		0	664		1037194		
1147580		4304908		11	476	30681	
298314		1238162	46	051	42	0.1	49
190		175		05		2	
2		3	10		0.1	81	0.1
12		0	0		0 30		0
0 140		0 126	34	55	301	101	37
40		0.053	0.043	551	0.0651	10	01
0		0.053	0.735	0	0.065 .713		0.756
0.212		0.196	0.735		0.088		0.756
0.212		0.196	0.004		0.006		0.081
0.0981		0.0031	0.004		.057		0.073
0.019			0.003	0.	.037		0.0731
	1601		AFG Country 2013		155//Q//		1435933
1674751		6451977	•		695		
80802		65863			335373)1100	
273368		413863			0		170
320		0	705				
1262524		4756015			524		
327734		1360269	49		45	101401	52
202		187		18		3	521
2		3	11	101		9	
13		0	0		0	91	01
0		01	36		32		40
149		134		64	021	10	101
43		0.052	0.042	011	0.064	101	01
0		01	0.737	0	.715		0.758
0.211		0.196	0.228	0	0.092		0.083
0.101		0.005	0.004		0.006		0.0001
0		0	0.068		0.06		0.076
0.019		0.018	0.02	·			
	160		AFG Country 2014		1664981		1541417
1800126	,	6910557	639770			71482	
87237		71013	107392		362082		
294743		445735			0		173
306		0	719		1283074		1164073
1420629		5325439	4831523	31		96365 I	
294596		1222730	51	•	47		55
211		195		28	·	3	·
2		3	11	•		9	
14		01	0		01		0
01		0	39		36		43
163		148		80	- · ·	9	- ·
37		0.052	0.043	•	0.064	•	0
0		01	0.77	0	.751		0.789
0.177		0.164	0.191		0.096		0.087
0.106		0.005	0.004		0.006		0

0			0.074	0.066	0.083
0.017	160	0.016	0.018	1000001	16020101
1 1960564	1001	7563486	AFG Country 2015 7027729		1693212 37381
92624		75517		384440	0/301
313436		469896		0	173
340		0	718		1306576
1578970		5953073	5422985		53563
295296		1225634	54	501	58
225		209	242		3
2		3	11	•	9
14		0	0	0	0
0		01	43	39	47
177		161	195		9
36		0.051	0.041	0.062	0
0		01	0.787	0.769	0.804
0.162		0.151	0.174	0.106	0.095
0.118		0.005	0.004	0.007	01
0		01	0.083	0.074	0.094
0.017		0.016	0.018		
1	160	Afghanistan	AFG Country 2016	1882415	1747511
2028013		7813023	•		17331
102155		83607	124073	423999	
347012		514970	78	0	178
326		0	737	1610050	1477602
1754298		6682564	6132832	728	31269
170131		706134	54	50	58
225		209	242		3
2		4	12		10
15		0	0	0	0
0		0	46	43	51
193		177	210		5
20		0.054	0.044	0.066	0
0		01	0.855	0.841	0.869
0.091		0.084	0.097	0.106	0.096
0.12		0.006	0.005	0.007	0
0		01	0.091	0.081	0.103
0.01		0.009	0.01		
1	160	-	AFG Country 2017		
2142606		8270915	7709396		92955
107508		87534	130678	446215	1
363311		542384		0	184
342		0	763	1715584	1582352
1857963		7120586	6567601		11533
169562		703771	55	52	60
230		215	248	I	3
2		4	12	0.1	10
15		0	0	0	0

01		01	48	44		52
198		183	215		5	0.1
20		0.054	0.044	0.065		0
0		0	0.861	0.847		0.873
0.085		0.079	0.091	0.109		0.097
0.122		0.006	0.005	0.007		0
0		0		0.083		0.106
0.009	4001	0.009	0.01	00074071		1071070
00070001	160	•	AFG Country 2018			
2207902		8456655	7779658		63967	
106277		86330	128822	441104		4001
358317		534679		0		188
401		0	779	1680810	740701	1514454
1848264		6976254	6285790		71278	
250304		1038896	55	51	0.1	60
228		210	247		3	
2		3	12	0.1	10	0.1
14		0	0	0		0
0		0	45	41	71	50
188		170	207		7	0.1
28		0.052	0.042	0.064		0
0		0	0.825	0.807		0.841
0.123		0.113	0.134	0.109		0.097
0.122		0.006	0.005 0.09	0.007		0.102
0		0	().()91	0.078		0.1021
						0.102
0.013	4001	0.012	0.015			
0.013 	139	0.012 Algeria	0.015 DZA Country 1995	2623765		2299202
0.013 2963485	139	0.012 Algeria 8696294	0.015 DZA Country 1995 7620552	2623765 982	22275	2299202
0.013 2963485 1870428	139	0.012 Algeria 8696294 1590992	0.015 DZA Country 1995 7620552 2190681	2623765 983 6199411	22275	2299202
0.013 2963485 1870428 5273238	139	0.012 Algeria 8696294 1590992 7260867	0.015 DZA Country 1995 7620552 2190681 7 57677	2623765 98: 6199411 39318	22275 	2299202 80587
0.013 2963485 1870428 5273238 191167	139	0.012 Algeria 8696294 1590992 7260867 130318	0.015 DZA Country 1995 7620552 2190681 7 57677 267101	2623765 983 6199411 39318 695600	22275 	2299202
0.013 2963485 1870428 5273238 191167 848769	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523	2623765 983 6199411 39318 695600 2813	22275 	2299202 80587 554702
0.013 2963485 1870428 5273238 191167 848769 59	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92	2623765 983 6199411 39318 695600 2813	22275 	2299202 80587
0.013 2963485 1870428 5273238 191167 848769 59 306	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92 346	2623765 983 6199411 39318 695600 2813	22275 	2299202 80587 554702 104
0.013 2963485 1870428 5273238 191167 848769 59 306 56	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92 346 218	2623765 983 6199411 39318 695600 2813	22275 	2299202 80587 554702 104
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92 346 218 1	2623765 983 6199411 39318 695600 2813 81	22275 	2299202 80587 554702 104 7
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92 346 218 1 24	2623765 983 6199411 39318 695600 2813	22275 	2299202 80587 554702 104
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92 346 218 1 24	2623765 983 6199411 39318 695600 2813 81	22275 	2299202 80587 554702 104 7 30
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713	0.015 DZA Country 1995 7620552 2190681 7 57677 267101 1838523 92 346 218 1 24 99 0.664	2623765 983 6199411 39318 695600 2813 81 3 20	22275 	2299202 80587 554702 104 7 30 0.022
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031	0.015 DZA Country 1995 7620552 2190681 7	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215	22275 3188 66 186	2299202 80587 554702 104 7 30 0.022 0.315
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0	0.015 DZA Country 1995 7620552 2190681 7	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037	22275 	2299202 80587 554702 104 7 30 0.022 0.315 0.032
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0 0.042	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0 0.026	0.015 DZA Country 1995 7620552 2190681 7 267101 1838523 92 346 218 1 24 99 0.664 0.265 0 0.022	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037 0.03	22275 	2299202 80587 554702 104 7 30 0.022 0.315 0.032 0.001
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0 0.042 0.001	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0 0.026 0.001	0.015 DZA Country 1995 7620552 2190681 7	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037	22275 	2299202 80587 554702 104 7 30 0.022 0.315 0.032
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0 0.042		0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0 0.026 0.001 0	0.015 DZA Country 1995 7620552 2190681 7	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037 0.03 0.008	22275 	2299202 80587 554702 104 7 30 0.022 0.315 0.032 0.001 0.012
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0 0.042 0.001 0	139	0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0 0.026 0.001 0 Algeria	0.015 DZA Country 1995 7620552 2190681 7	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037 0.03 0.008 2635043	22275 3188 66 186	2299202 80587 554702 104 7 30 0.022 0.315 0.032 0.001 0.012 2339491
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0 0.042 0.001 0 2970076		0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0 0.026 0.001 0 Algeria 8733676	0.015 DZA Country 1995 7620552 2190681 7 267101 1838523 92 346 218 1 24 99 0.664 0.265 0 0.022 0.01 0 DZA Country 1996 7754089	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037 0.03 0.008 2635043 986	22275 3188 66 186 	2299202 80587 554702 104 7 30 0.022 0.315 0.032 0.001 0.012 2339491
0.013 2963485 1870428 5273238 191167 848769 59 306 56 255 5 81 0 0.015 0 0.042 0.001 0		0.012 Algeria 8696294 1590992 7260867 130318 2305519 197 268 77 2 9 65 0.713 0.031 0 0.026 0.001 0 Algeria	0.015 DZA Country 1995 7620552 2190681 7	2623765 983 6199411 39318 695600 2813 81 3 20 0.763 0.215 0.037 0.03 0.008 2635043	22275 3188 66 186 	2299202 80587 554702 104 7 30 0.022 0.315 0.032 0.001 0.012 2339491

194644		136258		3688	68996		553364
832074		2286845	1834(91	J90 I		57854	1001
138		458	911	2401	81	CE I	102
301 56		267 75	210	340		65	
249		751 2		01	3	187	
2 4 91 5		∠ı 9	1		31 19		7 29
79		63 l	24	95	191	0	291
0		0.716	0.667	951	0.762	O I	0.022
0.015		0.032	0.262		0.7621		0.022
0.0131		0.0321	0.2021		0.035		0.312
0.04		0.025	0.022		0.029	ı	0.001
0.041		0.001	0.0221		0.029	1	0.001
0.0011		0.0011	0.0091		0.0071		0.0111
I	139		DZA Country 199	97 I	2743760		2460967
3073676	1001	9094012	•	6712		187496	
1954422		1702127	2226970	J1 12	647780		
5641588		7381144		21	42644		85114
202386		141341		2106	72769		592157
871828		2411881	19626			89617	0021011
586		1942	93	3001	83	000111	104
308		276	001	345	331	66	1011
58		75	219			191	
250		21	1		3		7
5		10	25		20		30
82		66		98		0	
0		0.712	0.666		0.756		0.022
0.015		0.032	0.265		0.221		0.311
0		0	0		0.036		0.032
0.041		0.026	0.022	l	0.0	3	0.001
0.001		0.001	0.01		0.008		0.012
0		0	0				
	139	Algeria	DZA Country 199	98	2839588		2556079
3162763		9411628	847	1954	10-	482769	
2023466		1766457	2288553		670664	3	
5854805		7585258	62234	1	43199		87510
206271		143179	290	0047	75242	2	623752
897726		2493854	2067	384	29	75455	
1466		4860	94		85		105
313		282		349		67	
59		76	223	3		195	
252		2	1		3		7
5		10	25		21		30
83		691		99		0	
0		0.712	0.669		0.753		0.022
0.015		0.031	0.265		0.224		0.308
0.001		01	0.001		0.036		0.032
0.04		0.025	0.022		0.029	I	0.001

0.001 0		0.001 0	0.009 0		0.008		0.011
1	139	Algeria	DZA Country 199		2930760	7500051	2636841
3244267		9713812	8739	634		752907	
2106985		1838709	2383056		6983463	31	000401
6094281		7898483			44332		88348
212493		146936		825	757650		635624
890141		2511181	210673	34		50314	4001
2014		6674	961	0501	861	20.1	106
318		286	000	352		69	
60		78	229	I	0.1	199	
258		2	1		3		7
5		10	25	071	21	0.1	29
82		69	0.6701	97	0.7501	0	0.0001
0		0.719	0.678		0.756		0.022
0.015		0.03	0.259		0.221		0.3
0.001		0.001	0.001		0.036	Ī	0.032
0.04		0.026	0.022		0.029		0.001
0.001		0.001	0.009		0.008		0.011
0	4001	0	0	0 I	00740001		00000051
1	139	•	DZA Country 200		2974090	047E001	2682085
3293950		9857426	8889	5921		917580	
2169069		1893627	2476199	1	7189236	0	000741
6276300		8207196			44907		90074
217683		148842		544	738032		641214
839929		2446157	21252	621		33888	4001
1312		4349	96	0501	861	701	106
317		286	000	352		70	
61		80	232	I	0.1	202	
264		2	1		3		7
5		10	24	001	21	0.1	27
79		68	0.4001	90	0 7001	0	0.0001
0		0.729	0.689		0.762		0.022
0.015		0.031	0.248		0.217		0.286
0		0	0		0.035		0.031
0.039		0.025	0.022		0.029		0.001
0.001		0.001	0.009		0.008		0.01
0	4001	0	0	4 1	20047001		00704001
05407061	139		DZA Country 200				28794331
3548736		10622032		689		762050	
2391849		2081996			7927625		004701
6900638		9030699			48281		98173
232851		160025	325		74050		669216
816221		2454356	22180	721		05309	4401
2172		7200	102	070 /	91	701	113
337		303	- - ·	373		76	
66		86	251	I	- 1	219	
286		2	2		3		7

5		10	23		21		26
78		70		86		0	
0	0	.746	0.713		0.776		0.022
0.015		0.031	0.232		0.204		0.263
0.001		0.001	0.001		0.036		0.033
0.04		0.027	0.024		0.031		0.001
0.001		0.001	0.008		0.008		0.009
0		0	01				•
i.	139		A Country 2002		3465709		3139612
3802827	1001	11486864	104060			504218	
2619742		2296948	2954669	,011	8682962		•
7613083		9793054	75907		52534	4	106174
						7 I	
251588		174122	3519		768437		715055
822671		2546934	237000)2		26689	4401
1623		5380	108		98		119
359		325		394		82	
72		92	271			238	
306		2	2		3		8
5		11	24		22		26
80		74		85		0	
0	0	.755	0.726		0.781		0.022
0.015		0.031	0.222		0.198		0.25
0		0	0.001		0.037		0.034
0.041		0.028	0.025		0.032	2	0.001
0.001		0.001	0.008		0.008		0.009
01		0	0				
+	+		+	+	+-		
	+		+		+		+
			+				
-+			· ·		•		•
	•	+-			•	•	•
		+					
	-+	+-		-+		+	
•				•			•
			•			•	
+			•	•		•	
+		· 	+			+	·
+		+	+			+-	
+			+				
	+-		+		+		+
+							
	+		-+		-+		+
+			•		·		·
	+	 p 20 rows	•		·		·

```
When
```

```
[117]: data.select('location_name', 'iso3',
          f.when(data.the_total_mean >= '5000000', 1).otherwise(0)
     ).show(5)
     #zwraca 1 gdy prawda, inaczj 0
     +----+
     |location_name|iso3|CASE WHEN (the_total_mean >= 5000000) THEN 1 ELSE 0 END|
       Afghanistan | AFG |
                                                                  11
       Afghanistan | AFG |
                                                                  1 l
     | Afghanistan | AFG |
                                                                  11
       Afghanistan | AFG |
                                                                  01
      Afghanistan | AFG |
                                                                  01
     +-----
     only showing top 5 rows
     Like
[118]: data.select(
     'the_total_mean',
     data.the_total_mean.rlike('^[9,7]').alias('the_total_mean zaczyna sie na 9 lub_

√7')

     ).distinct().show()
     +----+
     |the_total_mean|the_total_mean zaczyna sie na 9 lub 7|
     +----+
             698381
                                            falsel
                                            falsel
          58163376
           2420975|
                                            falsel
           1142944
                                            falsel
          162405032|
                                            false
             44340
                                            false
             68134
                                            false
           23070931
                                            false
             996681
                                             true
             15477|
                                            false
           3587803|
                                            false
                                            false
            534311
            883289|
                                            falsel
                                            false
             28067
           45360241
                                            falsel
            341486|
                                            falsel
          768944101
                                             truel
            5606731
                                            falsel
           9587995|
                                             true
```

```
720561
                                                 true
     +----+
     only showing top 20 rows
     GroupBy
[119]: data.groupBy('ISO3').count().show()
      #grupowanie, po czym wywołanie jakiejś funkcji, tutaj count (może też być mean⊔
       ⇔etc)
     +---+
     |ISO3|count|
     +---+
     | NIU|
             24|
     | HTI|
             241
     | PSE|
             241
     | BRB|
             241
     | LVA|
             241
     | POL|
             241
     | JAM|
             24|
     | ZMB|
             24|
     BRA
             24
     ARM
             24
     | MOZ|
             24
     | CUB|
             24|
     | JOR|
             241
     | FRA|
             24|
     | SOM|
             24|
     BRN
             241
     I CODI
             241
     | FSM|
             241
     | BOL|
             241
     | URY|
             241
     +---+
     only showing top 20 rows
     Agregacja
[120]: from pyspark.sql import functions as f
      data.groupBy("location_name").agg(f.mean("the_total_mean").
       →alias("total_mean_mean")).show() #średnia niezależnie od roku
      #grupowanie i obliczanie wartości dla grup
            location_name
                             total_mean_mean|
     +----+
```

```
South Asia | 6.1906327458333336E7 |
                          329933.83333333333
                 Chadl
|Micronesia (Feder...|
                                21818.375|
             Paraguay |
                                1295237.125
                Yemenl
                         1431750.2083333333
|Southeast Asia, E...| 3.878079027916667E8|
              Senegal |
                                   631454.5
           Cabo Verde
                          67143.708333333333
               Sweden | 3.886377466666664E7 |
              Tokelaul
                         1234.666666666667
             Kiribati|
                         21901.16666666668|
    Republic of Korea
                              5.992538875E7|
               Guyana|
                         144851.70833333334
              Eritrea|
                          89435.1666666667
          Philippines|
                          8087432.9583333331
             Djibouti|
                         58924.541666666641
                Tonga|
                         20671.9583333333332
             Malaysia
                          6636143.958333333|
            Singapore |
                          7900884.4583333331
          High-income | 4.799371008666667E9 |
only showing top 20 rows
```

Wizualizacja danych

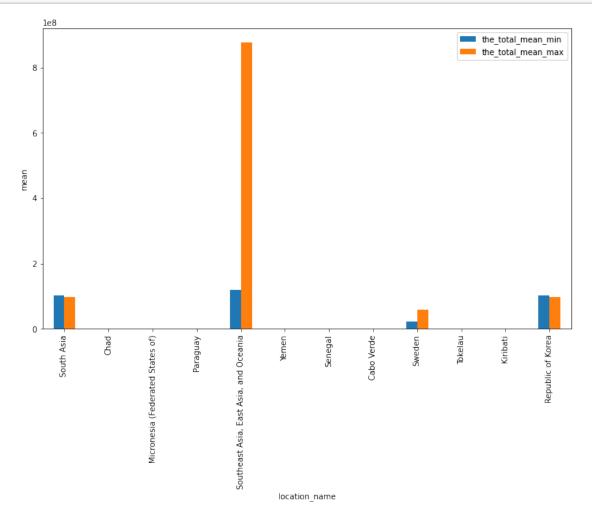
```
[121]:
                                     location_name the_total_mean_min \
       0
                                         South Asia
                                                              103194218
       1
                                               Chad
                                                                 201576
       2
                 Micronesia (Federated States of)
                                                                   10086
       3
                                           Paraguay
                                                                 1008338
       4
          Southeast Asia, East Asia, and Oceania
                                                              118847294
       5
                                              Yemen
                                                                 1015751
       6
                                            Senegal
                                                                  388101
       7
                                         Cabo Verde
                                                                   32597
       8
                                             Sweden
                                                               22629968
       9
                                            Tokelau
                                                                    1060
```

the_total_mean_max

```
0
                    96358386
       1
                      467239
       2
                        9618
       3
                      967137
       4
                   876778676
       5
                      921256
       6
                      972045
       7
                       97929
       8
                    59684355
       9
                         978
[122]:
      df.dtypes
[122]: location_name
                              object
       the_total_mean_min
                              object
       the_total_mean_max
                              object
       dtype: object
[123]: df['location_name']=df['location_name'].astype(str)
       df['the_total_mean_min'] = df['the_total_mean_min'].astype(float)
       df['the_total_mean_max'] = df['the_total_mean_max'].astype(float)
       df = df.head(12)
       df
[123]:
                                      location_name
                                                      the_total_mean_min
       0
                                         South Asia
                                                             103194218.0
       1
                                               Chad
                                                                 201576.0
       2
                 Micronesia (Federated States of)
                                                                  10086.0
       3
                                           Paraguay
                                                                1008338.0
       4
           Southeast Asia, East Asia, and Oceania
                                                             118847294.0
       5
                                              Yemen
                                                                1015751.0
       6
                                            Senegal
                                                                 388101.0
       7
                                         Cabo Verde
                                                                  32597.0
       8
                                             Sweden
                                                               22629968.0
       9
                                            Tokelau
                                                                   1060.0
       10
                                           Kiribati
                                                                  15477.0
                                  Republic of Korea
       11
                                                             102810388.0
           the_total_mean_max
       0
                    96358386.0
       1
                      467239.0
       2
                        9618.0
       3
                      967137.0
       4
                   876778676.0
       5
                      921256.0
       6
                      972045.0
       7
                       97929.0
```

```
8 59684355.0
9 978.0
10 30044.0
11 96960567.0
```

```
ind = list(range(12))
df.iloc[ind ,:].plot(kind='bar', x='location_name', y=df.columns.tolist()[1:],
figsize=(12, 7), ylabel='mean', xlabel='location_name')
plt.show()
```



Zapis do pliku

```
[125]: # error gdy plik już istnieje
data.write.csv('dataset.csv')
data.write.csv('dataset.json', format='json')
```

```
data.write.csv('dataset.parquet', format='parquet')
# wybrane kolumny
data.select(['location_name', 'the_total_mean']).write.csv('dataset.csv')
```

```
AnalysisException
                                                                                                              Traceback (most recent call last)
<ipython-input-125-72e8e4ebc73e> in <module>
                1 # error gdy plik już istnieje
----> 2 data.write.csv('dataset.csv')
               3 data.write.csv('dataset.json', format='json')
               4 data.write.csv('dataset.parquet', format='parquet')
               5 # wybrane kolumny
/usr/local/lib/python3.9/dist-packages/pyspark/sql/readwriter.py in csv(self, u
   →path, mode, compression, sep, quote, escape, header, nullValue, escapeQuotes, quoteAll, dateFormat, timestampFormat, ignoreLeadingWhiteSpace, header, nullValue, escapeQuotes, quoteAll, dateFormat, timestampFormat, ignoreLeadingWhiteSpace, header, nullValue, escapeQuoteEscaping, encoding, emptyValue, procedure of the control of the c
   ⇔lineSep)
       1025
   ⇔charToEscapeQuoteEscaping=charToEscapeQuoteEscaping,
       1026
                                                                                 encoding=encoding, emptyValue=emptyValue, __
   →lineSep=lineSep)
-> 1027
                                         self._jwrite.csv(path)
       1028
       1029
                               @since(1.5)
/usr/local/lib/python3.9/dist-packages/py4j/java_gateway.py in __call__(self,_
   →*args)
       1302
       1303
                                         answer = self.gateway_client.send_command(command)
-> 1304
                                         return value = get return value(
                                                    answer, self.gateway_client, self.target_id, self.name)
       1305
       1306
/usr/local/lib/python3.9/dist-packages/pyspark/sql/utils.py in deco(*a, **kw)
          132
                                                              # Hide where the exception came from that shows a_{\sqcup}
   ⇔non-Pythonic
          133
                                                              # JVM exception message.
--> 134
                                                              raise from(converted)
          135
                                                   else:
          136
                                                              raise
/usr/local/lib/python3.9/dist-packages/pyspark/sql/utils.py in raise from(e)
AnalysisException: path file:/content/dataset.csv already exists.;
```