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

```python
[100]: from pyspark.sql import SparkSession
spark = SparkSession.builder\
        .master("local[*]")\
        .appName('PySpark_Tutorial')\
        .getOrCreate()
# gdzie "*" znaczy wszystkie rdzenie procesora.
```

Podłączenie dysku google

```python
[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

```python
[102]: # Czytanie z pliku csv
csv_file = "/content/drive/My Drive/IHME_HEALTH_SPENDING_1995_2018_Y2021M09D22.
 ↪CSV"
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]>
```
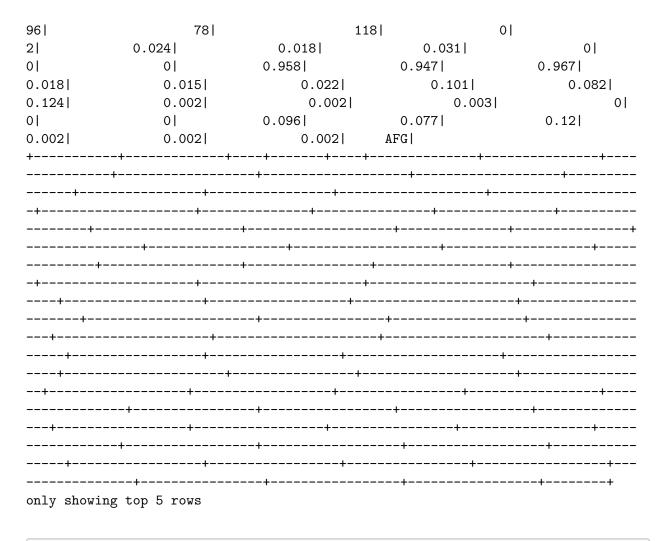
Manipulacja kolumnami

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

```
+----------+-------------+----+-------+----+------------+---------------+----
----------+---------------+---------------+----------------+---------------+---------
------+-------------+--------------+----------------+---------------+---------
-+-----------------+--------------+---------------+----------------+---------
-------+-------------+--------------+-----------+------------+
-------------+---------------+------------------+----------------+-----
--------+-------------------+---------------+----------------+---------------+
-+-----------------+--------------+---------------+---------------+----
---+------------------+---------------+----------------+---------------+-----
-------+-------------------+--------------+---------------+----------------+
---+---------------+-------------+---------------+-----------+
-----+---------------+---------------+----------------+-----------+
---+---------------+---------------+----------------+-----------+
--+----------------+-----------------+------------------+-----------------+---
-----------+-------------------+---------------+----------------+---------
---+---------------+---------------+----------------+----------------+----
----------+-------------------+----------------+----------------+---------
-----+---------------+---------------+-----------------+---------------+---
--------------+----------------+------------------+-----------------+--------+
```
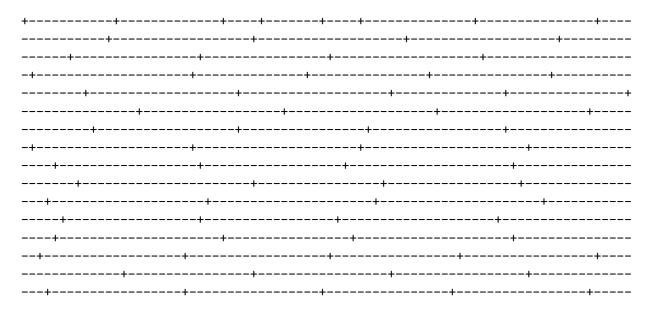
```
|location_id|location_name|iso3|  level|year|the_total_mean|the_total_lower|the_
total_upper|the_total_ppp_mean|the_total_ppp_lower|the_total_ppp_upper|ghes_tota
l_mean|ghes_total_lower|ghes_total_upper|ghes_total_ppp_mean|ghes_total_ppp_lowe
r|ghes_total_ppp_upper|ppp_total_mean|ppp_total_lower|ppp_total_upper|ppp_total_
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|
+----------+-------------+----+-------+----+------------+---------------+----
----------+---------------+---------------+----------------+---------------+---------
------+-------------+--------------+----------------+---------------+---------
-+-----------------+--------------+---------------+----------------+---------
-------+-------------+--------------+-----------+------------+
-------------+---------------+------------------+----------------+-----
--------+-------------------+---------------+----------------+---------------+-----
```

```
---------+-----------------+--------------+---------------+---------------
-+-----------------+--------------+---------------+---------------
----+-----------------+--------------+---------------+---------------
-------+-----------------+--------------+---------------+---------------
---+-----------------+--------------+---------------+----------
-----+-----------------+--------------+---------------+---------------
----+-----------------+--------------+---------------+---------------
--+-----------------+--------------+---------------+----------------+---
------------+-----------------+--------------+---------------+---------------
---+-----------------+--------------+---------------+---------------+-----
------------+-----------------+--------------+---------------+----------
-----+-----------------+--------------+---------------+---------------+---
-------------+-----------------+--------------+---------------+-------+
|        160|  Afghanistan| AFG|Country|1995|         528409|         417121|
665425|         2193179|          1731272|          2761865|
14643|         11814|       17779|          60775|
49036|          73792|          20|           0|           51|
81|           0|          212|      509515|          398261|
647195|         2114759|          1652995|          2686205|
4232|         17563|          30|           24|           38|
125|           99|              158|           1|
1|            1|               3|               3|
4|           0|           0|           0|           0|
0|            0|          29|           23|           37|
121|           95|             154|           0|
1|         0.028|          0.02|          0.037|           0|
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|           0|
0|            0|          0.1|          0.076|         0.129|
0.001|          0.001|          0.001|      AFG|
|        160|  Afghanistan| AFG|Country|1996|         516915|         412035|
651034|         2145473|          1710165|          2702136|
13771|         11138|       16872|          57159|
46227|          70026|          25|           0|           52|
102|           0|          217|      497261|          391328|
632006|         2063899|          1624221|          2623161|
5858|         24313|          28|           23|           36|
118|           94|              148|           1|
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|           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|
|         160|   Afghanistan| AFG|Country|1997|        509874|          413424|
636436|         2116248|          1715930|         2641549|
12864|          10480|            15698|          53392|
43499|            65155|           26|             0|            51|
108|             0|               211|          485094|         387884|
610799|         2013397|          1609925|         2535139|
11890|          49350|            27|             22|           34|
112|             91|              140|            1|
1|              1|               3|              2|
3|              0|               0|              0|            0|
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|          0|
0|              0|               0.096|          0.075|         0.124|
0.002|          0.002|           0.003|     AFG|
|         160|   Afghanistan| AFG|Country|1998|        485561|          394629|
605041|         2015335|          1637919|         2511240|
12020|          9689|            14652|          49889|
40216|           60812|           25|             0|            49|
105|             0|               202|          468699|         377467|
587578|         1945349|          1566688|         2438759|
4817|           19991|            25|             21|           32|
105|             85|              131|            1|
1|              1|               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|          0|
0|              0|               0.096|          0.076|         0.123|
0.001|          0.001|           0.001|     AFG|
|         160|   Afghanistan| AFG|Country|1999|        463720|          382723|
564170|         1924685|          1588504|         2341603|
11055|          8943|            13521|          45884|
37119|           56119|           25|             0|            46|
104|             0|               192|          444298|         362573|
545080|         1844071|          1504872|         2262371|
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|                 0|
     0|               0|            0.096|            0.077|             0.12|
 0.002|           0.002|            0.002|    AFG|
+----------+-------------+----+-------+----+-------------+-------------+----
----------+-------------+----------------+----------------+-------------+---------
------+-------------+-------------+----------------+----------------+----
-+-------------------------+----------------+-------------+-------------+----------
-------+------------------+----------------+----------------+----------------+
-------------+----------------+----------------+----------------+-----
---------+-------------------------+----------------+----------------+----
-+-------------------------+----------------+----------------+----------------+
---+-------------------------+----------------+----------------+-----
-------+-------------------------+----------------+----------------+
---+-------------------------+----------------+----------------+-------------
-----+-------------------------+----------------+----------------+----
----+-------------+-------------+----------------+----------------+
--+-------------------------+----------------+----------------+----
----------+-------------------------+----------------+----------------+-----
---+-------------------------+----------------+----------------+----
----------+-------------------------+----------------+-------------
-----+-------------------------+----------------+----------------+---
-------------+----------------+----------------+----------------+--------+
only showing top 5 rows
```

[111]:
```python
data = data.withColumnRenamed('iso3_new', 'iso3_new_mark2') #zmiana nazwy
data.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_total_mean	ghes_total_lower	ghes_total_upper	ghes_total_ppp_mean	ghes_total_ppp_lower	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_total_mean	dah_total_ppp_mean	the_per_cap_mean	the_per_cap_lower	the_per_cap_upper	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_ppp_lower	ghes_per_cap_ppp_upper	ppp_per_cap_mean	ppp_per_cap_lower	ppp_per_cap_upper	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_lower	oop_per_cap_ppp_upper	dah_per_cap_mean	dah_per_cap_ppp_mean	ghes_per_the_mean	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_mean	dah_per_the_lower	dah_per_the_upper	the_per_gdp_mean	the_per_gdp_lower	the_per_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_mark2
160	Afghanistan	AFG	Country	1995	528409	417121	665425	2193179	1731272	2761865	14643	11814	17779	60775	49036	73792	20	0	51

81| 0| 212| 509515| 398261|
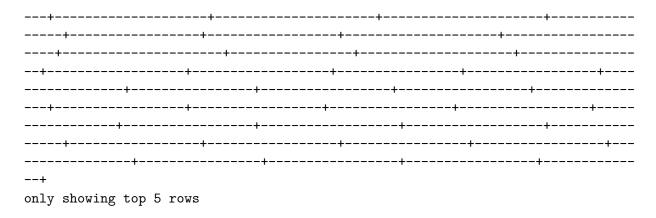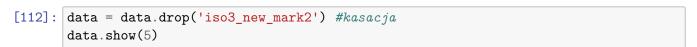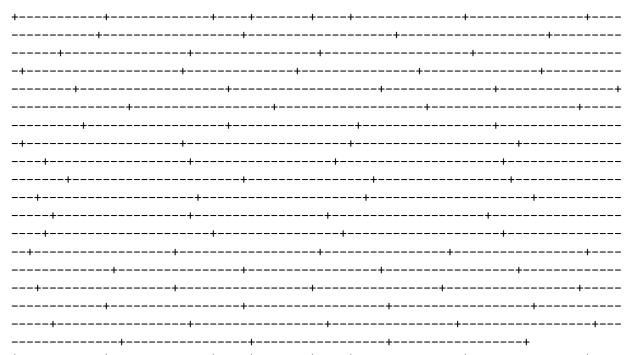647195| 2114759| 1652995| 2686205|
4232| 17563| 30| 24| 38|
125| 99| 158| 1|
1| 1| 3| 3|
4| 0| 0| 0| 0|
0| 0| 29| 23| 37|
121| 95| 154| 0|
1| 0.028| 0.02| 0.037| 0|
0| 0| 0.964| 0.953| 0.973|
0.008| 0.006| 0.01| 0.104| 0.08|
0.133| 0.003| 0.002| 0.004| 0|
0| 0| 0.1| 0.076| 0.129|
0.001| 0.001| 0.001| AFG|
| 160| Afghanistan| AFG|Country|1996| 516915| 412035|
651034| 2145473| 1710165| 2702136|
13771| 11138| 16872| 57159|
46227| 70026| 25| 0| 52|
102| 0| 217| 497261| 391328|
632006| 2063899| 1624221| 2623161|
5858| 24313| 28| 23| 36|
118| 94| 148| 1|
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| 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|
| 160| Afghanistan| AFG|Country|1997| 509874| 413424|
636436| 2116248| 1715930| 2641549|
12864| 10480| 15698| 53392|
43499| 65155| 26| 0| 51|
108| 0| 211| 485094| 387884|
610799| 2013397| 1609925| 2535139|
11890| 49350| 27| 22| 34|
112| 91| 140| 1|
1| 1| 3| 2|
3| 0| 0| 0| 0|
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| 0|

```
0|              0|      0.096|        0.075|          0.124|
0.002|          0.002|      0.003|          AFG|
|      160|  Afghanistan| AFG|Country|1998|      485561|        394629|
605041|        2015335|        1637919|        2511240|
12020|          9689|      14652|        49889|
40216|          60812|      25|          0|          49|
105|            0|        202|      468699|      377467|
587578|        1945349|        1566688|        2438759|
4817|          19991|      25|          21|          32|
105|            85|          131|          1|
1|              1|          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|          0|
0|            0|      0.096|      0.076|        0.123|
0.001|          0.001|      0.001|          AFG|
|      160|  Afghanistan| AFG|Country|1999|      463720|        382723|
564170|        1924685|        1588504|        2341603|
11055|          8943|      13521|        45884|
37119|          56119|      25|          0|          46|
104|            0|        192|      444298|      362573|
545080|        1844071|        1504872|        2262371|
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|          0|
0|            0|      0.096|      0.077|        0.12|
0.002|          0.002|      0.002|          AFG|
+----------+------------+----+-------+----+-------------+--------------+----
----------+---------------+----------------+----------------+----------------+---------
------+---------------+---------------+----------------+---------------+---------
-+---------------+-------------+-------------+---------------+-------------+---------
--------+---------------+-------------+--------------+-------------+-------------+
-------------+---------------+---------------+---------------+---------------+-----
---------+---------------+-------------+---------------+---------------+---------------
-+---------------+---------------+---------------+-----------------+-------------
----+---------------+---------------+--------------+---------------+-------------
-------+---------------+---------------+--------------+---------------+-------------
```

```
---+------------------+------------------+------------------+----------
-----+------------------+------------------+------------------+----------
----+------------------+------------------+------------------+-----------
--+------------------+------------------+------------------+----
-----------+------------------+------------------+------------------+---------
---+------------------+------------------+------------------+-----
-----------+------------------+------------------+------------------+----------
-----+------------------+------------------+------------------+----------+---
-------------+------------------+------------------+------------------+-----------
--+
only showing top 5 rows
```

[112]:
```python
data = data.drop('iso3_new_mark2') #kasacja
data.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_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_mean|dah_per_the_lower|dah_per_the_upper|the_per_gdp_mean|the_per_gdp_lower|the_per_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|

```
+----------+------------+---+-------+---+------------+-------------+--------------+----------------+----------------+-----------------+----------------+----------------+-----------------+----------------+----------------+-----------------+---------------+----------------+----------------+----------------+----------------+----------------+----------------+----------------+----------------+---------------+
```

| 160| Afghanistan| AFG|Country|1995| 528409| 417121|
665425| 2193179| 1731272| 2761865|
14643| 11814| 17779| 60775|
49036| 73792| 20| 0| 51|
81| 0| 212| 509515| 398261|
647195| 2114759| 1652995| 2686205|
4232| 17563| 30| 24| 38|
125| 99| 158| 1|
1| 1| 3| 3|
4| 0| 0| 0| 0|
0| 0| 29| 23| 37|
121| 95| 154| 0|
1| 0.028| 0.02| 0.037| 0|
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| 0|
0| 0| 0.1| 0.076| 0.129|
0.001| 0.001| 0.001|
| 160| Afghanistan| AFG|Country|1996| 516915| 412035|
651034| 2145473| 1710165| 2702136|
13771| 11138| 16872| 57159|
46227| 70026| 25| 0| 52|
102| 0| 217| 497261| 391328|

632006| 2063899| 1624221| 2623161|
5858| 24313| 28| 23| 36|
118| 94| 148| 1|
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| 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|
| 160| Afghanistan| AFG|Country|1997| 509874| 413424|
636436| 2116248| 1715930| 2641549|
12864| 10480| 15698| 53392|
43499| 65155| 26| 0| 51|
108| 0| 211| 485094| 387884|
610799| 2013397| 1609925| 2535139|
11890| 49350| 27| 22| 34|
112| 91| 140| 1|
1| 1| 3| 2|
3| 0| 0| 0| 0|
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| 0|
0| 0| 0.096| 0.075| 0.124|
0.002| 0.002| 0.003|
| 160| Afghanistan| AFG|Country|1998| 485561| 394629|
605041| 2015335| 1637919| 2511240|
12020| 9689| 14652| 49889|
40216| 60812| 25| 0| 49|
105| 0| 202| 468699| 377467|
587578| 1945349| 1566688| 2438759|
4817| 19991| 25| 21| 32|
105| 85| 131| 1|
1| 1| 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| 0|
0| 0| 0.096| 0.076| 0.123|

```
0.001|             0.001|              0.001|
|         160|   Afghanistan|  AFG|Country|1999|             463720|               382723|
564170|          1924685|              1588504|               2341603|
11055|             8943|             13521|              45884|
37119|            56119|               25|               0|                46|
104|                0|               192|          444298|               362573|
545080|          1844071|              1504872|               2262371|
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|                 0|
0|                 0|             0.096|              0.077|               0.12|
0.002|             0.002|             0.002|
+----------+------------+----+-------+----+-------------+--------------+----
----------+-------------+-------------+-------------+-------------+--------
------+-------------+-------------+-------------+-------------+---------
-+-------------+-------------+-------------+-------------+-------------+----------
-------+-------------+-------------+-------------+-------------+-------------+
-------------+-------------+-------------+-------------+-------------+-----
---------+-------------+-------------+-------------+-------------+
-+-------------+-------------+-------------+-------------+-------------+
----+-------------+-------------+-------------+-------------+-------------+------
-------+-------------+-------------+-------------+-------------+-------------+
---+-------------+-------------+-------------+-------------+-------------+
-----+-------------+-------------+-------------+-------------+
----+-------------+-------------+-------------+-------------+
--+-------------+-------------+-------------+-------------+----
----------+-------------+-------------+-------------+-------------+
---+-------------+-------------+-------------+-------------+----
----------+-------------+-------------+-------------+-------------+----------
-----+-------------+-------------+-------------+-------------+---
-------------+-------------+-------------+-------------+
only showing top 5 rows
```

Obróbka brakujących wartości

```python
[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

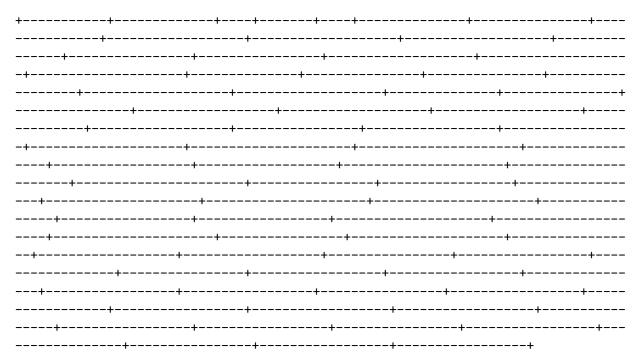Select

[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|
```

```
|   Afghanistan|1999|         463720|
+------------+----+--------------+
only showing top 5 rows
```

Filter

```python
#filtrowanie na podstawie warunku & | ~(not)
from pyspark.sql.functions import col
data.filter((col('year') >= 2000) & (col('year') <= 2003)).show(5)
```
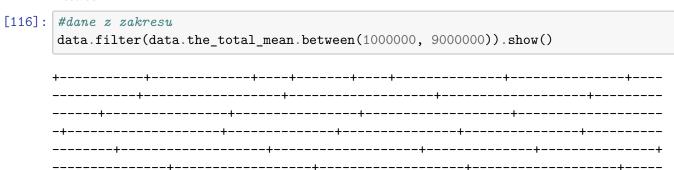
```
+----------+-------------+----+-------+----+------------+--------------+----
----------+---------------+---------------+---------------+---------------+---------
-----+---------------+---------------+---------------+---------------+---------
-+-------------+---------------+---------------+---------------+---------
-------+---------------+---------------+---------------+---------------+
------------+---------------+---------------+---------------+----
--------+---------------+---------------+---------------+---------------
-+---------------+---------------+---------------+---------------+----------
---+---------------+---------------+---------------+---------------+----
------+---------------+---------------+---------------+---------------
---+---------------+---------------+---------------+---------------+----------
-----+---------------+---------------+---------------+---------------+----
---+---------------+---------------+---------------+---------------+----------
--+---------------+---------------+---------------+---------------+----
-----------+---------------+---------------+---------------+---------------
---+---------------+---------------+---------------+---------------+-----
----------+---------------+---------------+---------------+----------
-----+---------------+---------------+---------------+---------------+---
------------+---------------+---------------+---------------+
|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_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|
+----------+-------------+---+------+---+-------------+--------------+----
----------+----------------+-----------------+-----------------+---------
------+-------------+---------------+-----------------+-----------------
-+----------------+-------------+-------------+---------------+----------
--------+-----------------+----------------+----------------+-------------+
--------------+----------------+-----------------+----------------+-----
---------+----------------+----------------+-----------------+----------
-+-----------------+-------------+-------------+---------------+----------
----+----------------+---------------+-----------------+----------------
-------+-----------------+----------------+-----------------+------------
---+----------------+----------------+-----------------+----------------
----+-----------------+----------------+-----------------+---------------
---+----------------+----------------+-----------------+---------------
--+----------------+-------------+-------------+---------------+----
----------+---------------+-------------+---------------+----------
---+----------------+----------------+-----------------+-----
----------+----------------+----------------+---------------+---
--------------+---------------+-----------------+----------------+
|         160|  Afghanistan| AFG|Country|2000|         446201|         372565|
536362|         1851971|         1546341|         2226189|
10382|         8407|         12686|         43091|
34892|         52654|         25|         0|         44|
106|         0|         182|         426131|         352623|
516002|         1768668|         1463571|         2141683|
9663|         40106|         23|         19|         28|
95|         79|         114|         1|
0|         1|         2|         2|
3|         0|         0|         0|         0|
0|         0|         22|         18|         27|
91|         75|         110|         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|         0.002|         0.002|
|         160|  Afghanistan| AFG|Country|2001|         447321|         381894|
524195|         1856620|         1585061|         2175686|
10191|         8281|         12450|         42298|
34369|         51674|         30|         0|         43|
126|         0|         177|         421413|         354700|
498677|         1749090|         1472194|         2069773|
15686|         65106|         22|         19|         26|
92|         78|         107|         1|
0|         1|         2|         2|
3|         0|         0|         0|         0|
```

```
0|                      0|              21|                  18|                25|
86|                     73|             102|                   1|
3|              0.023|          0.018|             0.03|                 0|
0|                      0|          0.942|             0.93|             0.951|
0.035|               0.03|          0.041|            0.105|             0.088|
0.122|              0.002|          0.002|            0.003|                 0|
0|                      0|          0.099|            0.083|             0.116|
0.004|              0.004|          0.004|
|         160|  Afghanistan| AFG|Country|2002|        534474|            464153|
612575|            2218352|        1926481|             2542511|
13506|              10994|          16525|               56055|
45630|              68586|             45|                   0|                65|
185|                    0|            270|              489696|            419200|
567274|            2032497|        1739903|             2354491|
31228|             129614|             25|                  22|                28|
103|                   90|            118|                   1|
1|                      1|              3|                   2|
3|                      0|              0|                   0|                 0|
0|                      0|             23|                  19|                26|
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|
0.094|              0.002|          0.002|            0.003|                 0|
0|                      0|          0.075|            0.064|             0.087|
0.005|              0.005|          0.005|
|         160|  Afghanistan| AFG|Country|2003|        682593|            601814|
770072|            2833123|        2497849|             3196208|
29302|              23856|          36109|              121617|
99016|             149871|             24|                   0|                72|
99|                     0|            300|              606824|            524018|
693301|            2518643|        2174955|             2877567|
46443|             192764|             30|                  27|                34|
126|                  111|            142|                   1|
1|                      2|              5|                   4|
7|                      0|              0|                   0|                 0|
0|                      0|             27|                  23|                31|
112|                    97|            128|                   2|
9|              0.043|          0.034|            0.054|                 0|
0|                      0|          0.889|            0.872|             0.903|
0.068|               0.06|          0.077|            0.095|             0.082|
0.107|              0.004|          0.003|            0.005|                 0|
0|                      0|          0.084|            0.072|             0.097|
0.006|              0.006|          0.007|
|          43|       Albania| ALB|Country|2000|        549733|            513401|
587564|            1532102|        1430847|             1637538|
200778|            177155|         228589|              559567|
493730|             637077|             22|                  15|                31|
```

```
  60|                 41|              86|          324109|          297748|
350560|             903291|          829822|              977010|
 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|               0|
     0|                  0|           0.045|           0.041|           0.048|
 0.003|              0.003|           0.003|
+----------+-------------+---+-------+----+-------------+-------------+----
----------+-------------+----------------+----------------+----------------+---------
------+-------------+-------------+----------------+-------------+----------------
-+-------------+-------------+----------------+-------------+-------------+----------
--------+-------------+----------------+-------------+-------------+-------------+
-------------+----------------+----------------+----------------+-------------+----
---------+-------------+----------------+----------------+-------------+-------------
-+-------------+----------------+----------------+-------------+----------------
----+-------------+-------------+----------------+-------------+
-------+-------------+----------------+-------------+----------------+-------------
---+----------------+-------------+----------------+-------------+----------
----+-------------+----------------+-------------+-------------+
---+-------------+----------------+-------------+-------------+-------------
--+-------------+-------------+----------------+----------------+-------------+----
----------+-------------+----------------+-------------+-------------+----------------
---+-------------+----------------+-------------+----------------+-------------+----
----------+----------------+----------------+----------------+-------------+----------
-----+-------------+----------------+----------------+----------------+---
--------------+----------------+----------------+----------------+

only showing top 5 rows
```

Between

```
[116]:  #dane z zakresu
        data.filter(data.the_total_mean.between(1000000, 9000000)).show()
```

```
+----------+-------------+---+-------+----+-------------+-------------+----
----------+----------------+----------------+----------------+---------
------+-------------+-------------+----------------+-------------+----------------
-+-------------+-------------+----------------+-------------+---------
-------+----------------+----------------+-------------+-------------+-------------+
-------------+----------------+-------------+----------------+-----
---------+----------------+-------------+----------------+-------------
```

```
+------------------+------------------+------------------+------------
---+------------------+------------------+------------------+---------
-------+------------------+------------------+------------------+-----
--+------------------+------------------+------------------+----------
-----+------------------+------------------+------------------+--------
----+------------------+------------------+------------------+---------
--+------------------+------------------+------------------+-----------
------------+------------------+------------------+------------------+-
---+------------------+------------------+------------------+----------
-----+------------------+------------------+------------------+--------
----+------------------+------------------+------------------+----------
-------------+------------------+------------------+---------------+
```

|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_mean|ghes_total_lower|ghes_total_upper|ghes_total_ppp_mean|ghes_total_ppp_lower|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_total_mean|dah_total_ppp_mean|the_per_cap_mean|the_per_cap_lower|the_per_cap_upper|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_ppp_lower|ghes_per_cap_ppp_upper|ppp_per_cap_mean|ppp_per_cap_lower|ppp_per_cap_upper|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_lower|oop_per_cap_ppp_upper|dah_per_cap_mean|dah_per_cap_ppp_mean|ghes_per_the_mean|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_mean|dah_per_the_lower|dah_per_the_upper|the_per_gdp_mean|the_per_gdp_lower|the_per_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|

```
+----------+------------+----+-------+----+-------------+---------------+----
----------+----------------+----------------+----------------+---------
------+---------------+----------------+----------------+----------------+---------
-+----------------+----------------+----------------+----------------+---------
-------+----------------+----------------+----------------+----------------+----------
-------------+----------------+----------------+----------------+-----
---------+----------------+----------------+----------------+----------------
-+----------------+----------------+----------------+----------------+---------
---+----------------+----------------+----------------+----------------+----------
-------+----------------+----------------+----------------+----------------
---+----------------+----------------+----------------+-----------
-----+----------------+----------------+----------------+----------------
----+----------------+----------------+----------------+----------------
--+----------------+----------------+----------------+----------------+----
----------+----------------+----------------+----------------+----------------
---+----------------+----------------+----------------+-----
----------+----------------+----------------+----------------+----------
```

```
-----+----------------+----------------+----------------+----------------+---
-------------+---------------+----------------+----------------+
|         160|   Afghanistan| AFG|Country|2007|         1040372|         956086|
1141413|          4318097|          3968264|          4737470|
56613|          45313|          70506|          234974|
188072|          292635|          41|          0|          100|
170|          0|          415|          784067|          700601|
880017|          3254296|          2907866|          3652539|
199651|          828657|          41|          37|          44|
168|          155|          184|          2|
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|
|         160|   Afghanistan| AFG|Country|2008|         1066858|         982415|
1161963|          4428026|          4077542|          4822762|
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|
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|              0.071|               0.064|               0.079|
0.025|              0.024|              0.025|
|         160|   Afghanistan| AFG|Country|2010|           1262196|            1166456|
1364153|           5238783|            4841414|            5661960|
65282|            52686|             80074|             270956|
218674|             332348|                 65|                   0|                 135|
270|                   0|                562|              905046|              807082|
1006229|           3756424|            3349819|            4176386|
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|
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|              0.023|
|         160|   Afghanistan| AFG|Country|2011|           1373092|            1273961|
1482720|           5699060|            5287617|            6154077|
67603|            54439|             82866|             280589|
225952|             343937|                 71|                   0|                 142|
295|                   0|                591|              943247|              846554|
1049015|           3914978|            3513651|            4353968|
362170|           1503198|                 46|                  43|                  50|
192|                 178|                  207|                   2|
2|                   3|                   9|                   8|
12|                   0|                   0|                   0|                   0|
0|                   0|                 32|                  28|                  35|
132|                 118|                  147|                  12|
51|              0.049|               0.04|                0.06|                   0|
0|                   0|              0.687|               0.662|                0.71|
0.264|              0.244|              0.284|               0.097|               0.089|
0.105|              0.005|              0.004|               0.006|                   0|
0|                   0|              0.066|               0.059|               0.074|
0.025|              0.025|              0.026|
|         160|   Afghanistan| AFG|Country|2012|           1410474|            1299776|
1525283|           5854216|            5394762|            6330734|
74894|            60780|             91457|             310850|
252268|             379594|                 72|                   0|                 160|
```

297| 0| 664| 1037194| 933396|
1147580| 4304908| 3874091| 4763068|
298314| 1238162| 46| 42| 49|
190| 175| 205| 2|
2| 3| 10| 8|
12| 0| 0| 0| 0|
0| 0| 34| 30| 37|
140| 126| 155| 10|
40| 0.053| 0.043| 0.065| 0|
0| 0| 0.735| 0.713| 0.756|
0.212| 0.196| 0.23| 0.088| 0.08|
0.098| 0.005| 0.004| 0.006| 0|
0| 0| 0.065| 0.057| 0.073|
0.019| 0.018| 0.02|
| 160| Afghanistan| AFG|Country|2013| 1554494| 1435933|
1674751| 6451977| 5959885| 6951108|
80802| 65863| 99713| 335373|
273368| 413863| 77| 0| 170|
320| 0| 705| 1145881| 1033726|
1262524| 4756015| 4290513| 5240146|
327734| 1360269| 49| 45| 52|
202| 187| 218| 3|
2| 3| 11| 9|
13| 0| 0| 0| 0|
0| 0| 36| 32| 40|
149| 134| 164| 10|
43| 0.052| 0.042| 0.064| 0|
0| 0| 0.737| 0.715| 0.758|
0.211| 0.196| 0.228| 0.092| 0.083|
0.101| 0.005| 0.004| 0.006| 0|
0| 0| 0.068| 0.06| 0.076|
0.019| 0.018| 0.02|
| 160| Afghanistan| AFG|Country|2014| 1664981| 1541417|
1800126| 6910557| 6397700| 7471482|
87237| 71013| 107392| 362082|
294743| 445735| 74| 0| 173|
306| 0| 719| 1283074| 1164073|
1420629| 5325439| 4831523| 5896365|
294596| 1222730| 51| 47| 55|
211| 195| 228| 3|
2| 3| 11| 9|
14| 0| 0| 0| 0|
0| 0| 39| 36| 43|
163| 148| 180| 9|
37| 0.052| 0.043| 0.064| 0|
0| 0| 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|           0.074|            0.066|            0.083|
0.017|            0.016|            0.018|
|         160|    Afghanistan| AFG|Country|2015|          1822293|           1693212|
1960564|             7563486|               7027729|              8137381|
92624|             75517|          113213|             384440|
313436|             469896|            82|              0|               173|
340|               0|             718|         1434292|            1306576|
1578970|             5953073|              5422985|              6553563|
295296|           1225634|            54|              50|               58|
225|               209|              242|              3|
2|                3|              11|               9|
14|                0|              0|               0|               0|
0|                0|             43|              39|               47|
177|               161|              195|              9|
36|             0.051|          0.041|            0.062|            0|
0|                0|           0.787|           0.769|           0.804|
0.162|            0.151|            0.174|            0.106|            0.095|
0.118|            0.005|            0.004|            0.007|               0|
0|                0|           0.083|           0.074|           0.094|
0.017|            0.016|            0.018|
|         160|    Afghanistan| AFG|Country|2016|          1882415|           1747511|
2028013|             7813023|               7253099|              8417331|
102155|             83607|          124073|             423999|
347012|             514970|            78|              0|               178|
326|               0|             737|         1610050|            1477602|
1754298|             6682564|              6132832|              7281269|
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|                0|           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|                0|           0.091|           0.081|           0.103|
0.01|             0.009|            0.01|
|         160|    Afghanistan| AFG|Country|2017|          1992736|           1857448|
2142606|             8270915|               7709396|              8892955|
107508|             87534|          130678|             446215|
363311|             542384|            82|              0|               184|
342|               0|             763|         1715584|            1582352|
1857963|             7120586|              6567601|              7711533|
169562|             703771|            55|              52|               60|
230|               215|              248|              3|
2|                4|              12|               10|
15|                0|              0|               0|               0|
```

0| 0| 48| 44| 52|
198| 183| 215| 5|
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| 0.094| 0.083| 0.106|
0.009| 0.009| 0.01|
| 160| Afghanistan| AFG|Country|2018| 2037487| 1874376|
2207902| 8456655| 7779658| 9163967|
106277| 86330| 128822| 441104|
358317| 534679| 97| 0| 188|
401| 0| 779| 1680810| 1514454|
1848264| 6976254| 6285790| 7671278|
250304| 1038896| 55| 51| 60|
228| 210| 247| 3|
2| 3| 12| 10|
14| 0| 0| 0| 0|
0| 0| 45| 41| 50|
188| 170| 207| 7|
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| 0.007| 0|
0| 0| 0.09| 0.078| 0.102|
0.013| 0.012| 0.015|
| 139| Algeria| DZA|Country|1995| 2623765| 2299202|
2963485| 8696294| 7620552| 9822275|
1870428| 1590992| 2190681| 6199411|
5273238| 7260867| 57677| 39318| 80587|
191167| 130318| 267101| 695600| 554702|
848769| 2305519| 1838523| 2813188|
59| 197| 92| 81| 104|
306| 268| 346| 66|
56| 77| 218| 186|
255| 2| 1| 3| 7|
5| 9| 24| 20| 30|
81| 65| 99| 0|
0| 0.713| 0.664| 0.763| 0.022|
0.015| 0.031| 0.265| 0.215| 0.315|
0| 0| 0| 0.037| 0.032|
0.042| 0.026| 0.022| 0.03| 0.001|
0.001| 0.001| 0.01| 0.008| 0.012|
0| 0| 0|
| 139| Algeria| DZA|Country|1996| 2635043| 2339491|
2970076| 8733676| 7754089| 9844119|
1886213| 1631909| 2175582| 6251729|
5408856| 7210822| 58726| 41111| 82575|

194644| 136258| 273688| 689965| 553364|
832074| 2286845| 1834090| 2757854|
138| 458| 91| 81| 102|
301| 267| 340| 65|
56| 75| 216| 187|
249| 2| 1| 3| 7|
5| 9| 24| 19| 29|
79| 63| 95| 0|
0| 0.716| 0.667| 0.762| 0.022|
0.015| 0.032| 0.262| 0.215| 0.312|
0| 0| 0| 0.035| 0.031|
0.04| 0.025| 0.022| 0.029| 0.001|
0.001| 0.001| 0.009| 0.007| 0.011|
0| 0| 0|
| 139| Algeria| DZA|Country|1997| 2743760| 2460967|
3073676| 9094012| 8156712| 10187496|
1954422| 1702127| 2226970| 6477803|
5641588| 7381144| 61062| 42644| 85114|
202386| 141341| 282106| 727690| 592157|
871828| 2411881| 1962665| 2889617|
586| 1942| 93| 83| 104|
308| 276| 345| 66|
58| 75| 219| 191|
250| 2| 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| 0.03| 0.001|
0.001| 0.001| 0.01| 0.008| 0.012|
0| 0| 0|
| 139| Algeria| DZA|Country|1998| 2839588| 2556079|
3162763| 9411628| 8471954| 10482769|
2023466| 1766457| 2288553| 6706643|
5854805| 7585258| 62234| 43199| 87510|
206271| 143179| 290047| 752422| 623752|
897726| 2493854| 2067384| 2975455|
1466| 4860| 94| 85| 105|
313| 282| 349| 67|
59| 76| 223| 195|
252| 2| 1| 3| 7|
5| 10| 25| 21| 30|
83| 69| 99| 0|
0| 0.712| 0.669| 0.753| 0.022|
0.015| 0.031| 0.265| 0.224| 0.308|
0.001| 0| 0.001| 0.036| 0.032|
0.04| 0.025| 0.022| 0.029| 0.001|

0.001| 0.001| 0.009| 0.008| 0.011|
0| 0| 0|
| 139| Algeria| DZA|Country|1999| 2930760| 2636841|
3244267| 9713812| 8739634| 10752907|
2106985| 1838709| 2383056| 6983463|
6094281| 7898483| 64111| 44332| 88348|
212493| 146936| 292825| 757650| 635624|
890141| 2511181| 2106734| 2950314|
2014| 6674| 96| 86| 106|
318| 286| 352| 69|
60| 78| 229| 199|
258| 2| 1| 3| 7|
5| 10| 25| 21| 29|
82| 69| 97| 0|
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| 0.032|
0.04| 0.026| 0.022| 0.029| 0.001|
0.001| 0.001| 0.009| 0.008| 0.011|
0| 0| 0|
| 139| Algeria| DZA|Country|2000| 2974090| 2682085|
3293950| 9857426| 8889592| 10917580|
2169069| 1893627| 2476199| 7189236|
6276300| 8207196| 65677| 44907| 90074|
217683| 148842| 298544| 738032| 641214|
839929| 2446157| 2125262| 2783888|
1312| 4349| 96| 86| 106|
317| 286| 352| 70|
61| 80| 232| 202|
264| 2| 1| 3| 7|
5| 10| 24| 21| 27|
79| 68| 90| 0|
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| 0| 0|
| 139| Algeria| DZA|Country|2001| 3204780| 2879433|
3548736| 10622032| 9543689| 11762050|
2391849| 2081996| 2724658| 7927625|
6900638| 9030699| 70254| 48281| 98173|
232851| 160025| 325388| 740505| 669216|
816221| 2454356| 2218072| 2705309|
2172| 7200| 102| 91| 113|
337| 303| 373| 76|
66| 86| 251| 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| 0|
| 139| Algeria| DZA|Country|2002| 3465709| 3139612|
3802827| 11486864| 10406037| 12604218|
2619742| 2296948| 2954669| 8682962|
7613083| 9793054| 75907| 52534| 106174|
251588| 174122| 351905| 768437| 715055|
822671| 2546934| 2370002| 2726689|
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| 0.001|
0.001| 0.001| 0.008| 0.008| 0.009|
0| 0| 0|
+----------+-------------+----+-------+----+-------------+---------------+----
----------+----------------+-------------------+---------------+---------
------+-------------+-------------+---------------+---------------+--------
-+----------------+-------------+---------------+---------------+----------
-------+----------------+-------------+-------------+---------------+
-------------+----------------+-------------+---------------+---------------+-----
--------+----------------+-------------+---------------+---------------
-+----------------+-------------+---------------+---------------+
----+----------------+-------------+---------------+---------------
-------+----------------+-------------+---------------+-------------
---+----------------+-------------+---------------+-------------+----------
-----+----------------+-------------+---------------+---------------
----+----------------+-------------+---------------+---------------
--+----------------+-------------+---------------+---------------+----
-----------+----------------+-------------+---------------+---------------
---+----------------+-------------+---------------+---------------+----
-----------+----------------+-------------+---------------+----------
-----+----------------+-------------+---------------+---------------+---
-------------+----------------+---------------+---------------+

only showing top 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|                                                1|
|  Afghanistan| AFG|                                                1|
|  Afghanistan| AFG|                                                1|
|  Afghanistan| AFG|                                                0|
|  Afghanistan| AFG|                                                0|
+-------------+----+-------------------------------------------------+
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|
+--------------+----------------------------------+
|         69838|                             false|
|      58163376|                             false|
|       2420975|                             false|
|       1142944|                             false|
|     162405032|                             false|
|         44340|                             false|
|         68134|                             false|
|      23070931|                             false|
|         99668|                              true|
|         15477|                             false|
|       3587803|                             false|
|        534311|                             false|
|        883289|                             false|
|         28067|                             false|
|       4536024|                             false|
|        341486|                             false|
|      76894410|                              true|
|        560673|                             false|
|       9587995|                              true|
```

```
|         72056|                                    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|   24|
| PSE|   24|
| BRB|   24|
| LVA|   24|
| POL|   24|
| JAM|   24|
| ZMB|   24|
| BRA|   24|
| ARM|   24|
| MOZ|   24|
| CUB|   24|
| JOR|   24|
| FRA|   24|
| SOM|   24|
| BRN|   24|
| COD|   24|
| FSM|   24|
| BOL|   24|
| URY|   24|
+----+-----+
```
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|
|                Chad|     329933.8333333333|
|Micronesia (Feder…|          21818.375|
|            Paraguay|         1295237.125|
|               Yemen|   1431750.2083333333|
|Southeast Asia, E…| 3.878079027916667E8|
|             Senegal|            631454.5|
|          Cabo Verde|    67143.70833333333|
|              Sweden|3.8863774666666664E7|
|             Tokelau|   1234.6666666666667|
|            Kiribati|   21901.166666666668|
|   Republic of Korea|         5.992538875E7|
|              Guyana|   144851.70833333334|
|             Eritrea|    89435.16666666667|
|         Philippines|    8087432.958333333|
|            Djibouti|   58924.541666666664|
|               Tonga|   20671.958333333332|
|             Malaysia|    6636143.958333333|
|           Singapore|    7900884.458333333|
|         High-income| 4.799371008666667E9|
+------------------+-------------------+
only showing top 20 rows
```

Wizualizacja danych

```python
from pyspark.sql.functions import col, min, max

df = data.select('location_name', 'the_total_mean')\
    .groupBy("location_name")\
    .agg(min("the_total_mean").alias("the_total_mean_min"),
         max("the_total_mean").alias("the_total_mean_max"))\
    .toPandas()
df.head(10)
```

```
[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
11	Republic of Korea	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
```

[124]:
```python
from matplotlib import pyplot as plt

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

```python
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,
 ↪path, mode, compression, sep, quote, escape, header, nullValue, escapeQuotes, ᵣ
 ↪quoteAll, dateFormat, timestampFormat, ignoreLeadingWhiteSpace,
 ↪ignoreTrailingWhiteSpace, charToEscapeQuoteEscaping, encoding, emptyValue,
 ↪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(
   1305                  answer, self.gateway_client, self.target_id, self.name)
   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␣
 ↪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.;
```