

```
#importowanie biblioteki pandas
```

```
import pandas as pd
```

```
#tworzenie ramki danych ze słownika
```

```
data = pd.read_csv('IHME_PREM_GEN_POP_2020_Y2021M10D11.csv',  
encoding='utf-8')
```

```
data_frame = pd.DataFrame(data)
```

```
#wyświetlenie ramki danych
```

```
data_frame
```

| | observation_id | submitted_time | \ |
|-------|---------------------|-------------------------|-----|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 | UTC |
| 1 | gp_4503631639609344 | 2020-07-09 13:22:37.107 | UTC |
| 2 | gp_4503700758593536 | 2020-07-04 18:53:36.471 | UTC |
| 3 | gp_4503737805832192 | 2020-07-12 17:58:20.798 | UTC |
| 4 | gp_4503819343101952 | 2020-07-06 00:20:22.983 | UTC |
| ... | ... | ... | ... |
| 52485 | gp_6755106844573696 | 2020-07-02 12:37:51.894 | UTC |
| 52486 | gp_6755213279232000 | 2020-07-02 07:07:44.269 | UTC |
| 52487 | gp_6755237508677632 | 2020-07-02 10:26:38.461 | UTC |
| 52488 | gp_6755275458740224 | 2020-07-24 23:16:52.556 | UTC |
| 52489 | gp_6755355351842816 | 2020-07-02 13:33:00.269 | UTC |

| | gender | age | \ |
|-------|-----------------------|--------------------|-----|
| 0 | Male | Under 16 | |
| 1 | Female | 26 to 35 years old | |
| 2 | Male | 36 to 45 years old | |
| 3 | Male | 26 to 35 years old | |
| 4 | Male | 26 to 35 years old | |
| ... | ... | ... | ... |
| 52485 | Female | 36 to 45 years old | |
| 52486 | Male | 26 to 35 years old | |
| 52487 | Prefer not to respond | 16 to 25 years old | |
| 52488 | Prefer not to respond | 26 to 35 years old | |
| 52489 | Female | 26 to 35 years old | |

| | geography | \ |
|-------|----------------------------------|-----|
| 0 | Suburban/Peri-urban | |
| 1 | City center or metropolitan area | |
| 2 | City center or metropolitan area | |
| 3 | Rural | |
| 4 | Suburban/Peri-urban | |
| ... | ... | ... |
| 52485 | Suburban/Peri-urban | |
| 52486 | City center or metropolitan area | |
| 52487 | Suburban/Peri-urban | |
| 52488 | Rural | |

52489 City center or metropolitan area

| | | financial_situation \ |
|-------|---|-----------------------|
| 0 | I can afford food and regular expenses, but no... | |
| 1 | I cannot afford enough food for my family | |
| 2 | I can comfortably afford food, clothes, and fu... | |
| 3 | I can afford food and regular expenses, and bu... | |
| 4 | I can afford food and regular expenses, and bu... | |
| ... | | ... |
| 52485 | I can afford food, but nothing else | |
| 52486 | I can afford food and regular expenses, but no... | |
| 52487 | I can afford food, but nothing else | |
| 52488 | I cannot afford enough food for my family | |
| 52489 | I cannot afford enough food for my family | |

| | education | employment_status \ |
|-------|-----------------------|----------------------------|
| 0 | Secondary/high school | Employed full-time |
| 1 | College or university | Unemployed |
| 2 | Primary school | Employed full-time |
| 3 | Technical school | Student |
| 4 | Technical school | Employed full-time |
| ... | | ... |
| 52485 | Secondary/high school | Employed part-time |
| 52486 | College or university | Employed full-time |
| 52487 | Secondary/high school | Self-employed |
| 52488 | Prefer not to answer | Student and work part-time |
| 52489 | Secondary/high school | Unemployed |

| | ethnicity | religion ... \ |
|-------|----------------------|--------------------------|
| 0 | Ankole | Christianity ... |
| 1 | Mestizo | Catholicism ... |
| 2 | Non-hispanic White | Agnosticism ... |
| 3 | Mestizo | Christianity ... |
| 4 | Mestizo | Catholicism ... |
| ... | | ... |
| 52485 | Not Available | Catholicism ... |
| 52486 | Arab | Muslim (Sunni) ... |
| 52487 | Grebo | Prefer Not To Answer ... |
| 52488 | Prefer not to answer | Other ... |
| 52489 | Tagalog | Christianity ... |

| | gp_post_miss_dose_why_other | gp_post_labor_force |
|----------------------|-----------------------------|---------------------|
| gp_pre_labor_force \ | | |
| 0 | NaN | No |
| No | | |
| 1 | NaN | Yes |
| Yes | | |
| 2 | NaN | Yes |
| Yes | | |
| 3 | NaN | No |

| Yes | | | |
|-------|-----------------------------------|---------------------------|-----------------------------------|
| 4 | NaN | Yes | |
| Yes | | | |
| ... | ... | ... | . |
| .. | | | |
| 52485 | NaN | No | |
| Yes | | | |
| 52486 | NaN | Yes | |
| Yes | | | |
| 52487 | NaN | No | |
| No | | | |
| 52488 | NaN | Yes | |
| No | | | |
| 52489 | NaN | Yes | |
| Yes | | | |
| | | | |
| | gp_unemployment_why | gp_unemployment_why_other | \ |
| 0 | NaN | NaN | |
| 1 | NaN | NaN | |
| 2 | NaN | NaN | |
| 3 | Laid off while business continues | NaN | |
| 4 | NaN | NaN | |
| ... | ... | ... | |
| 52485 | Business / office closed | NaN | |
| 52486 | NaN | NaN | |
| 52487 | NaN | NaN | |
| 52488 | NaN | NaN | |
| 52489 | NaN | NaN | |
| | | | |
| | gp_pre_income | gp_post_income | country weight |
| \ | | | |
| 0 | 50000.0 | 999497.0 | Uganda 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America 1.898763 |
| 3 | 12.0 | 12.0 | Colombia 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia 1.691686 |
| ... | ... | ... | ... |
| 52485 | 900.0 | 900.0 | Italy 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso 1.795822 |

| | | | | |
|-------|---------|---------|-------------|----------|
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |
|-------|---------|---------|-------------|----------|

```
      user_id
0    gp_6372662088826880
1    gp_5900473574883328
2    gp_4813642242981888
3    gp_4703669741420544
4    gp_4762741153988608
...
52485 gp_5444968770699264
52486 gp_5482738571804672
52487 gp_5658409882091520
52488 gp_5003001999917056
52489 gp_5388715507843072
```

```
[52490 rows x 48 columns]
```

```
#zapisanie ramki danych do pliku csv
```

```
data_frame.to_csv('data_frame.csv')
```

```
#tworzenie ramki dancyh z listy list
```

```
data_list = [
    [
        'Warszawa',
        'Łódź',
        'Poznań'
    ],
    [
        123,
        456,
        789
    ]
]
```

```
pd.DataFrame(data_list)
```

```
#transponowanie
```

```
t_data_frame = pd.DataFrame(data_frame).T
t_data_frame
```

```
#wyświetlenie 10 pierwszych wierszy ramki
```

```
data_frame.head(10)
```

```
#wyświetlenie ostatnich 10 wierszy
```

```
data_frame.tail(10)
```

```

#wyświetlić informacje o ramce
data_frame.info()

#wyświetlić liczbę kolumn i wierszy
data_frame.shape

#wyświetlić informację statystyczną o kolumnach liczbowych
data_frame.describe()

#wyświetlić informację statystyczną o kolumnach kategoryzowanych
data_frame.describe(include='all')

#usuwanie brakujących wartości w ramce danych
data_frame.dropna(inplace=True)

#przedstawić wybór wierszy i kolumn używając nazw oraz indeksów na
różne sposoby
#metoda 1
data_frame["observation_id"]

0      gp_4503617949401088
1      gp_4503631639609344
2      gp_4503700758593536
3      gp_4503737805832192
4      gp_4503819343101952
...
52485   gp_6755106844573696
52486   gp_6755213279232000
52487   gp_6755237508677632
52488   gp_6755275458740224
52489   gp_6755355351842816
Name: observation_id, Length: 52490, dtype: object

#metoda 2
data_frame.observation_id

0      gp_4503617949401088
1      gp_4503631639609344
2      gp_4503700758593536
3      gp_4503737805832192
4      gp_4503819343101952
...
52485   gp_6755106844573696
52486   gp_6755213279232000
52487   gp_6755237508677632
52488   gp_6755275458740224

```

```
52489    gp_6755355351842816
Name: observation_id, Length: 52490, dtype: object
```

```
#przedstawianie wyboru wierszy z ramki danych pod warunkiem spełnienia  
kilku warunków jednocześnie
```

```
data_frame[(data_frame.gender == 'Male') & (data_frame.geography ==  
'City center or metropolitan area') & data_frame.employment_status !=  
'Unemployed']
```

| | observation_id | submitted_time \ |
|-------|---------------------|-----------------------------|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 UTC |
| 1 | gp_4503631639609344 | 2020-07-09 13:22:37.107 UTC |
| 2 | gp_4503700758593536 | 2020-07-04 18:53:36.471 UTC |
| 3 | gp_4503737805832192 | 2020-07-12 17:58:20.798 UTC |
| 4 | gp_4503819343101952 | 2020-07-06 00:20:22.983 UTC |
| ... | ... | ... |
| 52485 | gp_6755106844573696 | 2020-07-02 12:37:51.894 UTC |
| 52486 | gp_6755213279232000 | 2020-07-02 07:07:44.269 UTC |
| 52487 | gp_6755237508677632 | 2020-07-02 10:26:38.461 UTC |
| 52488 | gp_6755275458740224 | 2020-07-24 23:16:52.556 UTC |
| 52489 | gp_6755355351842816 | 2020-07-02 13:33:00.269 UTC |

| | gender | age \ |
|-------|-----------------------|--------------------|
| 0 | Male | Under 16 |
| 1 | Female | 26 to 35 years old |
| 2 | Male | 36 to 45 years old |
| 3 | Male | 26 to 35 years old |
| 4 | Male | 26 to 35 years old |
| ... | ... | ... |
| 52485 | Female | 36 to 45 years old |
| 52486 | Male | 26 to 35 years old |
| 52487 | Prefer not to respond | 16 to 25 years old |
| 52488 | Prefer not to respond | 26 to 35 years old |
| 52489 | Female | 26 to 35 years old |

| | geography \ |
|-------|----------------------------------|
| 0 | Suburban/Peri-urban |
| 1 | City center or metropolitan area |
| 2 | City center or metropolitan area |
| 3 | Rural |
| 4 | Suburban/Peri-urban |
| ... | ... |
| 52485 | Suburban/Peri-urban |
| 52486 | City center or metropolitan area |
| 52487 | Suburban/Peri-urban |
| 52488 | Rural |
| 52489 | City center or metropolitan area |

| | financial_situation \ |
|---|---|
| 0 | I can afford food and regular expenses, but no... |

| | |
|-------|---|
| 1 | I cannot afford enough food for my family |
| 2 | I can comfortably afford food, clothes, and fu... |
| 3 | I can afford food and regular expenses, and bu... |
| 4 | I can afford food and regular expenses, and bu... |
| ... | ... |
| 52485 | I can afford food, but nothing else |
| 52486 | I can afford food and regular expenses, but no... |
| 52487 | I can afford food, but nothing else |
| 52488 | I cannot afford enough food for my family |
| 52489 | I cannot afford enough food for my family |

| | education | employment_status \ |
|-------|-----------------------|----------------------------|
| 0 | Secondary/high school | Employed full-time |
| 1 | College or university | Unemployed |
| 2 | Primary school | Employed full-time |
| 3 | Technical school | Student |
| 4 | Technical school | Employed full-time |
| ... | ... | ... |
| 52485 | Secondary/high school | Employed part-time |
| 52486 | College or university | Employed full-time |
| 52487 | Secondary/high school | Self-employed |
| 52488 | Prefer not to answer | Student and work part-time |
| 52489 | Secondary/high school | Unemployed |

| | ethnicity | religion ... \ |
|-------|----------------------|--------------------------|
| 0 | Ankole | Christianity ... |
| 1 | Mestizo | Catholicism ... |
| 2 | Non-hispanic White | Agnosticism ... |
| 3 | Mestizo | Christianity ... |
| 4 | Mestizo | Catholicism ... |
| ... | ... | ... |
| 52485 | Not Available | Catholicism ... |
| 52486 | Arab | Muslim (Sunni) ... |
| 52487 | Grebo | Prefer Not To Answer ... |
| 52488 | Prefer not to answer | Other ... |
| 52489 | Tagalog | Christianity ... |

| | gp_post_miss_dose_why_other | gp_post_labor_force |
|----------------------|-----------------------------|---------------------|
| gp_pre_labor_force \ | | |
| 0 | NaN | No |
| No | | |
| 1 | NaN | Yes |
| Yes | | |
| 2 | NaN | Yes |
| Yes | | |
| 3 | NaN | No |
| Yes | | |
| 4 | NaN | Yes |
| Yes | | |
| ... | ... | ... |

| | | | | |
|-------|-----------------------------------|---------------------------|--------------------------|----------|
| ... | | | | |
| 52485 | | NaN | No | |
| Yes | | | | |
| 52486 | | NaN | Yes | |
| Yes | | | | |
| 52487 | | NaN | No | |
| No | | | | |
| 52488 | | NaN | Yes | |
| No | | | | |
| 52489 | | NaN | Yes | |
| Yes | | | | |
| | | | | |
| | gp_unemployment_why | gp_unemployment_why_other | \ | |
| 0 | | NaN | NaN | |
| 1 | | NaN | NaN | |
| 2 | | NaN | NaN | |
| 3 | Laid off while business continues | | NaN | |
| 4 | | NaN | NaN | |
| ... | | ... | ... | |
| 52485 | Business / office closed | | NaN | |
| 52486 | | NaN | NaN | |
| 52487 | | NaN | NaN | |
| 52488 | | NaN | NaN | |
| 52489 | | NaN | NaN | |
| | | | | |
| | gp_pre_income | gp_post_income | country | weight |
| \ | | | | |
| 0 | 50000.0 | 999497.0 | Uganda | 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador | 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America | 1.898763 |
| 3 | 12.0 | 12.0 | Colombia | 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia | 1.691686 |
| ... | ... | ... | ... | ... |
| 52485 | 900.0 | 900.0 | Italy | 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria | 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia | 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso | 1.795822 |
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |
| | | | | |
| | user_id | | | |


```
0 gp_6372662088826880
1 gp_5900473574883328
2 gp_4813642242981888
3 gp_4703669741420544
4 gp_4762741153988608
```

```
...
52485 gp_5444968770699264
52486 gp_5482738571804672
52487 gp_5658409882091520
52488 gp_5003001999917056
52489 gp_5388715507843072
```

[52490 rows x 48 columns]

#wybór wierszy które zawierają w kolumnie kategoryzowanej określone słowo

```
data_frame[data_frame.country.str.contains('United States of America')]
```

| | observation_id | | submitted_time | gender \ |
|-------|---------------------|------------|------------------|----------|
| 2 | gp_4503700758593536 | 2020-07-04 | 18:53:36.471 UTC | Male |
| 22 | gp_4504459226120192 | 2020-07-02 | 18:06:38.962 UTC | Male |
| 79 | gp_4506455510351872 | 2020-07-08 | 11:28:21.549 UTC | Male |
| 95 | gp_4507104389103616 | 2020-07-02 | 15:03:50.961 UTC | Female |
| 111 | gp_4507856478142464 | 2020-06-30 | 23:29:25.781 UTC | Female |
| ... | ... | ... | ... | ... |
| 52423 | gp_6752600194285568 | 2020-07-08 | 10:37:07.198 UTC | Female |
| 52435 | gp_6752990502584320 | 2020-07-07 | 21:35:26.377 UTC | Female |
| 52437 | gp_6753113110478848 | 2020-07-11 | 22:23:32.729 UTC | Female |
| 52455 | gp_6753989686460416 | 2020-07-09 | 19:31:08.381 UTC | Female |
| 52472 | gp_6754557762994176 | 2020-07-09 | 19:04:15.968 UTC | Female |

| | age | geography \ |
|-------|--------------------|----------------------------------|
| 2 | 36 to 45 years old | City center or metropolitan area |
| 22 | 26 to 35 years old | Rural |
| 79 | 36 to 45 years old | Suburban/Peri-urban |
| 95 | 36 to 45 years old | Rural |
| 111 | 26 to 35 years old | Rural |
| ... | ... | ... |
| 52423 | 26 to 35 years old | Suburban/Peri-urban |
| 52435 | 36 to 45 years old | City center or metropolitan area |
| 52437 | 26 to 35 years old | Suburban/Peri-urban |
| 52455 | Over 45 years old | Suburban/Peri-urban |
| 52472 | Over 45 years old | Rural |

| | financial_situation \ |
|----|---|
| 2 | I can comfortably afford food, clothes, and fu... |
| 22 | I can afford food and regular expenses, but no... |
| 79 | I can afford food, but nothing else |
| 95 | I can afford food, but nothing else |

111 I can afford food and regular expenses, but no...
 ...
 52423 I cannot afford enough food for my family
 52435 I cannot afford enough food for my family
 52437 I can comfortably afford food, clothes, and fu...
 52455 I can comfortably afford food, clothes, and fu...
 52472 I can afford food, but nothing else

| | education | employment_status | |
|-------------|-----------------------|--------------------|------------------|
| ethnicity \ | | | |
| 2 | Primary school | Employed full-time | Non-hispanic |
| White | | | |
| 22 | Post graduate | Self-employed | Black or African |
| American | | | |
| 79 | College or university | Employed full-time | Non-hispanic |
| White | | | |
| 95 | College or university | Employed full-time | Non-hispanic |
| White | | | |
| 111 | College or university | Unemployed | Non-hispanic |
| White | | | |
| ... | ... | ... | |
| ... | | | |
| 52423 | Secondary/high school | Unemployed | Non-hispanic |
| White | | | |
| 52435 | Secondary/high school | Self-employed | Non-hispanic |
| White | | | |
| 52437 | College or university | Employed full-time | Black or African |
| American | | | |
| 52455 | College or university | Retired | Non-hispanic |
| White | | | |
| 52472 | Post graduate | Unemployed | Non-hispanic |
| White | | | |

| | religion | ... gp_post_miss_dose_why_other \ |
|-------|----------------------|-----------------------------------|
| 2 | Agnosticism | NaN |
| 22 | Catholicism | NaN |
| 79 | Protestantism | NaN |
| 95 | Protestantism | NaN |
| 111 | Agnosticism | NaN |
| ... | ... | ... |
| 52423 | Agnosticism | NaN |
| 52435 | Prefer Not To Answer | NaN |
| 52437 | Other | NaN |
| 52455 | Protestantism | NaN |
| 52472 | Other | NaN |

| | gp_post_labor_force | gp_pre_labor_force |
|-----------------------|---------------------|--------------------|
| gp_unemployment_why \ | | |
| 2 | Yes | Yes |
| NaN | | |

| | | |
|----------|-----|----------------------|
| 22 | No | No |
| NaN | | |
| 79 | Yes | Yes |
| NaN | | |
| 95 | No | No |
| NaN | | |
| 111 | No | No |
| NaN | | |
| ... | ... | ... |
| ... | | |
| 52423 | Yes | Yes |
| NaN | | |
| 52435 | Yes | Yes |
| NaN | | |
| 52437 | No | Yes |
| COVID-19 | | To avoid exposure to |
| 52455 | No | No |
| NaN | | |
| 52472 | Yes | Yes |
| NaN | | |

| | gp_unemployment_why_other | gp_pre_income | gp_post_income | \ |
|-------|---------------------------|---------------|----------------|---|
| 2 | NaN | 3200.0 | 3200.0 | |
| 22 | NaN | 6000.0 | 8000.0 | |
| 79 | NaN | 1600.0 | 1600.0 | |
| 95 | NaN | 2400.0 | 1000.0 | |
| 111 | NaN | 1600.0 | 1200.0 | |
| ... | ... | ... | ... | |
| 52423 | NaN | 2000.0 | 3000.0 | |
| 52435 | NaN | 3000.0 | 500.0 | |
| 52437 | NaN | 2500.0 | 1500.0 | |
| 52455 | NaN | 4500.0 | 4500.0 | |
| 52472 | NaN | 2500.0 | 2500.0 | |

| | country | weight | user_id |
|-------|--------------------------|----------|---------------------|
| 2 | United States of America | 1.898763 | gp_4813642242981888 |
| 22 | United States of America | 1.534788 | gp_5773509276205056 |
| 79 | United States of America | 1.600341 | gp_4739309881524224 |
| 95 | United States of America | 1.760472 | gp_4793845916631040 |
| 111 | United States of America | 1.677434 | gp_5304100193566720 |
| ... | ... | ... | ... |
| 52423 | United States of America | 1.690776 | gp_5529499709472768 |
| 52435 | United States of America | 1.775450 | gp_6592924367716352 |
| 52437 | United States of America | 1.677434 | gp_6247699120914432 |
| 52455 | United States of America | 2.091937 | gp_5924949256830976 |
| 52472 | United States of America | 2.091937 | gp_5794485990850560 |

[1943 rows x 48 columns]

#wybór wierszy które nie zawierają w kolumnie kategoryzowanej określonego słowa

```
data_frame[data_frame.country.str.contains('Primary school') == False]
```

| | observation_id | submitted_time \ |
|-------|---------------------|-----------------------------|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 UTC |
| 1 | gp_4503631639609344 | 2020-07-09 13:22:37.107 UTC |
| 2 | gp_4503700758593536 | 2020-07-04 18:53:36.471 UTC |
| 3 | gp_4503737805832192 | 2020-07-12 17:58:20.798 UTC |
| 4 | gp_4503819343101952 | 2020-07-06 00:20:22.983 UTC |
| ... | ... | ... |
| 52485 | gp_6755106844573696 | 2020-07-02 12:37:51.894 UTC |
| 52486 | gp_6755213279232000 | 2020-07-02 07:07:44.269 UTC |
| 52487 | gp_6755237508677632 | 2020-07-02 10:26:38.461 UTC |
| 52488 | gp_6755275458740224 | 2020-07-24 23:16:52.556 UTC |
| 52489 | gp_6755355351842816 | 2020-07-02 13:33:00.269 UTC |

| | gender | age \ |
|-------|-----------------------|--------------------|
| 0 | Male | Under 16 |
| 1 | Female | 26 to 35 years old |
| 2 | Male | 36 to 45 years old |
| 3 | Male | 26 to 35 years old |
| 4 | Male | 26 to 35 years old |
| ... | ... | ... |
| 52485 | Female | 36 to 45 years old |
| 52486 | Male | 26 to 35 years old |
| 52487 | Prefer not to respond | 16 to 25 years old |
| 52488 | Prefer not to respond | 26 to 35 years old |
| 52489 | Female | 26 to 35 years old |

| | geography \ |
|-------|----------------------------------|
| 0 | Suburban/Peri-urban |
| 1 | City center or metropolitan area |
| 2 | City center or metropolitan area |
| 3 | Rural |
| 4 | Suburban/Peri-urban |
| ... | ... |
| 52485 | Suburban/Peri-urban |
| 52486 | City center or metropolitan area |
| 52487 | Suburban/Peri-urban |
| 52488 | Rural |
| 52489 | City center or metropolitan area |

| | financial_situation \ |
|-----|---|
| 0 | I can afford food and regular expenses, but no... |
| 1 | I cannot afford enough food for my family |
| 2 | I can comfortably afford food, clothes, and fu... |
| 3 | I can afford food and regular expenses, and bu... |
| 4 | I can afford food and regular expenses, and bu... |
| ... | ... |

| | |
|-------|---|
| 52485 | I can afford food, but nothing else |
| 52486 | I can afford food and regular expenses, but no... |
| 52487 | I can afford food, but nothing else |
| 52488 | I cannot afford enough food for my family |
| 52489 | I cannot afford enough food for my family |

| | education | employment_status \ |
|-------|-----------------------|----------------------------|
| 0 | Secondary/high school | Employed full-time |
| 1 | College or university | Unemployed |
| 2 | Primary school | Employed full-time |
| 3 | Technical school | Student |
| 4 | Technical school | Employed full-time |
| ... | ... | ... |
| 52485 | Secondary/high school | Employed part-time |
| 52486 | College or university | Employed full-time |
| 52487 | Secondary/high school | Self-employed |
| 52488 | Prefer not to answer | Student and work part-time |
| 52489 | Secondary/high school | Unemployed |

| | ethnicity | religion ... \ |
|-------|----------------------|--------------------------|
| 0 | Ankole | Christianity ... |
| 1 | Mestizo | Catholicism ... |
| 2 | Non-hispanic White | Agnosticism ... |
| 3 | Mestizo | Christianity ... |
| 4 | Mestizo | Catholicism ... |
| ... | ... | ... |
| 52485 | Not Available | Catholicism ... |
| 52486 | Arab | Muslim (Sunni) ... |
| 52487 | Grebo | Prefer Not To Answer ... |
| 52488 | Prefer not to answer | Other ... |
| 52489 | Tagalog | Christianity ... |

| | gp_post_miss_dose_why_other | gp_post_labor_force |
|----------------------|-----------------------------|---------------------|
| gp_pre_labor_force \ | | |
| 0 | NaN | No |
| No | | |
| 1 | NaN | Yes |
| Yes | | |
| 2 | NaN | Yes |
| Yes | | |
| 3 | NaN | No |
| Yes | | |
| 4 | NaN | Yes |
| Yes | | |
| ... | ... | ... |
| .. | | |
| 52485 | NaN | No |
| Yes | | |
| 52486 | NaN | Yes |
| Yes | | |

| | | | | |
|-------|-----------------------------------|-----------------------------|--------------------------|----------|
| 52487 | NaN | No | | |
| No | | | | |
| 52488 | NaN | Yes | | |
| No | | | | |
| 52489 | NaN | Yes | | |
| Yes | | | | |
| | gp_unemployment_why | gp_unemployment_why_other \ | | |
| 0 | NaN | NaN | | |
| 1 | NaN | NaN | | |
| 2 | NaN | NaN | | |
| 3 | Laid off while business continues | NaN | | |
| 4 | NaN | NaN | | |
| ... | ... | ... | | |
| 52485 | Business / office closed | NaN | | |
| 52486 | NaN | NaN | | |
| 52487 | NaN | NaN | | |
| 52488 | NaN | NaN | | |
| 52489 | NaN | NaN | | |
| | gp_pre_income | gp_post_income | country | weight |
| \ | | | | |
| 0 | 50000.0 | 999497.0 | Uganda | 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador | 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America | 1.898763 |
| 3 | 12.0 | 12.0 | Colombia | 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia | 1.691686 |
| ... | ... | ... | ... | ... |
| 52485 | 900.0 | 900.0 | Italy | 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria | 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia | 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso | 1.795822 |
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |
| | user_id | | | |
| 0 | gp_6372662088826880 | | | |
| 1 | gp_5900473574883328 | | | |
| 2 | gp_4813642242981888 | | | |
| 3 | gp_4703669741420544 | | | |
| 4 | gp_4762741153988608 | | | |

```

...
52485 gp_5444968770699264
52486 gp_5482738571804672
52487 gp_5658409882091520
52488 gp_5003001999917056
52489 gp_5388715507843072

```

```
[52490 rows x 48 columns]
```

```

#tworzenie kolumny na podstawie istniejącej - kolumna weight10 =
kolumna weight*10

```

```

data_frame['weight10'] = data_frame['weight']*10
data_frame

```

```

      observation_id      submitted_time \
0      gp_4503617949401088  2020-07-07 14:48:29.83 UTC
1      gp_4503631639609344  2020-07-09 13:22:37.107 UTC
2      gp_4503700758593536  2020-07-04 18:53:36.471 UTC
3      gp_4503737805832192  2020-07-12 17:58:20.798 UTC
4      gp_4503819343101952  2020-07-06 00:20:22.983 UTC
...
52485 gp_6755106844573696  2020-07-02 12:37:51.894 UTC
52486 gp_6755213279232000  2020-07-02 07:07:44.269 UTC
52487 gp_6755237508677632  2020-07-02 10:26:38.461 UTC
52488 gp_6755275458740224  2020-07-24 23:16:52.556 UTC
52489 gp_675535351842816  2020-07-02 13:33:00.269 UTC

```

```

      gender      age \
0      Male      Under 16
1      Female  26 to 35 years old
2      Male  36 to 45 years old
3      Male  26 to 35 years old
4      Male  26 to 35 years old
...
52485      Female  36 to 45 years old
52486      Male  26 to 35 years old
52487 Prefer not to respond  16 to 25 years old
52488 Prefer not to respond  26 to 35 years old
52489      Female  26 to 35 years old

```

```

      geography \
0      Suburban/Peri-urban
1      City center or metropolitan area
2      City center or metropolitan area
3      Rural
4      Suburban/Peri-urban
...
52485      Suburban/Peri-urban
52486      City center or metropolitan area
52487      Suburban/Peri-urban

```

52488 Rural
 52489 City center or metropolitan area

financial_situation \

0 I can afford food and regular expenses, but no...
 1 I cannot afford enough food for my family
 2 I can comfortably afford food, clothes, and fu...
 3 I can afford food and regular expenses, and bu...
 4 I can afford food and regular expenses, and bu...
 ...
 52485 I can afford food, but nothing else
 52486 I can afford food and regular expenses, but no...
 52487 I can afford food, but nothing else
 52488 I cannot afford enough food for my family
 52489 I cannot afford enough food for my family

education employment_status \

0 Secondary/high school Employed full-time
 1 College or university Unemployed
 2 Primary school Employed full-time
 3 Technical school Student
 4 Technical school Employed full-time
 ...
 52485 Secondary/high school Employed part-time
 52486 College or university Employed full-time
 52487 Secondary/high school Self-employed
 52488 Prefer not to answer Student and work part-time
 52489 Secondary/high school Unemployed

ethnicity religion ...

gp_post_labor_force \

0 Ankole Christianity ...
 No
 1 Mestizo Catholicism ...
 Yes
 2 Non-hispanic White Agnosticism ...
 Yes
 3 Mestizo Christianity ...
 No
 4 Mestizo Catholicism ...
 Yes
 ...
 ...
 52485 Not Available Catholicism ...
 No
 52486 Arab Muslim (Sunni) ...
 Yes
 52487 Grebo Prefer Not To Answer ...
 No
 52488 Prefer not to answer Other ...

Yes
52489 Tagalog Christianity ...
Yes

| | gp_pre_labor_force | gp_unemployment_why | \ |
|-------|--------------------|-----------------------------------|---|
| 0 | No | NaN | |
| 1 | Yes | NaN | |
| 2 | Yes | NaN | |
| 3 | Yes | Laid off while business continues | |
| 4 | Yes | NaN | |
| ... | ... | ... | |
| 52485 | Yes | Business / office closed | |
| 52486 | Yes | NaN | |
| 52487 | No | NaN | |
| 52488 | No | NaN | |
| 52489 | Yes | NaN | |

| | gp_unemployment_why_other | gp_pre_income | gp_post_income | \ |
|-------|---------------------------|---------------|----------------|---|
| 0 | NaN | 50000.0 | 999497.0 | |
| 1 | NaN | 300.0 | 200.0 | |
| 2 | NaN | 3200.0 | 3200.0 | |
| 3 | NaN | 12.0 | 12.0 | |
| 4 | NaN | 3000000.0 | 400000.0 | |
| ... | ... | ... | ... | |
| 52485 | NaN | 900.0 | 900.0 | |
| 52486 | NaN | 1200000.0 | 120000.0 | |
| 52487 | NaN | 0.0 | 0.0 | |
| 52488 | NaN | 15.0 | 58.0 | |
| 52489 | NaN | 14000.0 | 16000.0 | |

| | country | weight | user_id |
|-----------|--------------------------|----------|---------------------|
| weight10 | | | |
| 0 | Uganda | 2.675598 | gp_6372662088826880 |
| 26.755980 | | | |
| 1 | Ecuador | 1.997161 | gp_5900473574883328 |
| 19.971608 | | | |
| 2 | United States of America | 1.898763 | gp_4813642242981888 |
| 18.987634 | | | |
| 3 | Colombia | 1.691686 | gp_4703669741420544 |
| 16.916865 | | | |
| 4 | Colombia | 1.691686 | gp_4762741153988608 |
| 16.916865 | | | |
| ... | ... | ... | ... |
| ... | | | |
| 52485 | Italy | 2.260748 | gp_5444968770699264 |
| 22.607481 | | | |
| 52486 | Algeria | 1.896411 | gp_5482738571804672 |
| 18.964113 | | | |
| 52487 | Liberia | 1.795822 | gp_5658409882091520 |
| 17.958216 | | | |

```

52488          Burkina Faso  1.795822  gp_5003001999917056
17.958216
52489          Philippines  1.608296  gp_5388715507843072
16.082961

```

```
[52490 rows x 49 columns]
```

```
#usuwanie kolumny
```

```
data_frame.drop("weight10", axis=1, inplace=True)
data_frame
```

```

      observation_id      submitted_time \
0      gp_4503617949401088  2020-07-07 14:48:29.83 UTC
1      gp_4503631639609344  2020-07-09 13:22:37.107 UTC
2      gp_4503700758593536  2020-07-04 18:53:36.471 UTC
3      gp_4503737805832192  2020-07-12 17:58:20.798 UTC
4      gp_4503819343101952  2020-07-06 00:20:22.983 UTC
...
52485  gp_6755106844573696  2020-07-02 12:37:51.894 UTC
52486  gp_6755213279232000  2020-07-02 07:07:44.269 UTC
52487  gp_6755237508677632  2020-07-02 10:26:38.461 UTC
52488  gp_6755275458740224  2020-07-24 23:16:52.556 UTC
52489  gp_6755355351842816  2020-07-02 13:33:00.269 UTC

```

```

      gender      age \
0      Male      Under 16
1      Female  26 to 35 years old
2      Male  36 to 45 years old
3      Male  26 to 35 years old
4      Male  26 to 35 years old
...
52485  Female  36 to 45 years old
52486  Male  26 to 35 years old
52487  Prefer not to respond  16 to 25 years old
52488  Prefer not to respond  26 to 35 years old
52489  Female  26 to 35 years old

```

```

      geography \
0      Suburban/Peri-urban
1      City center or metropolitan area
2      City center or metropolitan area
3      Rural
4      Suburban/Peri-urban
...
52485  Suburban/Peri-urban
52486  City center or metropolitan area
52487  Suburban/Peri-urban
52488  Rural
52489  City center or metropolitan area

```

| | | financial_situation \ |
|-------|---|-----------------------|
| 0 | I can afford food and regular expenses, but no... | |
| 1 | I cannot afford enough food for my family | |
| 2 | I can comfortably afford food, clothes, and fu... | |
| 3 | I can afford food and regular expenses, and bu... | |
| 4 | I can afford food and regular expenses, and bu... | |
| ... | | ... |
| 52485 | I can afford food, but nothing else | |
| 52486 | I can afford food and regular expenses, but no... | |
| 52487 | I can afford food, but nothing else | |
| 52488 | I cannot afford enough food for my family | |
| 52489 | I cannot afford enough food for my family | |

| | education | employment_status \ |
|-------|-----------------------|----------------------------|
| 0 | Secondary/high school | Employed full-time |
| 1 | College or university | Unemployed |
| 2 | Primary school | Employed full-time |
| 3 | Technical school | Student |
| 4 | Technical school | Employed full-time |
| ... | | ... |
| 52485 | Secondary/high school | Employed part-time |
| 52486 | College or university | Employed full-time |
| 52487 | Secondary/high school | Self-employed |
| 52488 | Prefer not to answer | Student and work part-time |
| 52489 | Secondary/high school | Unemployed |

| | ethnicity | religion ... \ |
|-------|----------------------|--------------------------|
| 0 | Ankole | Christianity ... |
| 1 | Mestizo | Catholicism ... |
| 2 | Non-hispanic White | Agnosticism ... |
| 3 | Mestizo | Christianity ... |
| 4 | Mestizo | Catholicism ... |
| ... | | ... |
| 52485 | Not Available | Catholicism ... |
| 52486 | Arab | Muslim (Sunni) ... |
| 52487 | Grebo | Prefer Not To Answer ... |
| 52488 | Prefer not to answer | Other ... |
| 52489 | Tagalog | Christianity ... |

| | gp_post_miss_dose_why_other | gp_post_labor_force |
|----------------------|-----------------------------|---------------------|
| gp_pre_labor_force \ | | |
| 0 | NaN | No |
| No | | |
| 1 | NaN | Yes |
| Yes | | |
| 2 | NaN | Yes |
| Yes | | |
| 3 | NaN | No |
| Yes | | |
| 4 | NaN | Yes |

| | | |
|-------|-----|-----|
| Yes | | |
| ... | ... | ... |
| .. | | |
| 52485 | NaN | No |
| Yes | | |
| 52486 | NaN | Yes |
| Yes | | |
| 52487 | NaN | No |
| No | | |
| 52488 | NaN | Yes |
| No | | |
| 52489 | NaN | Yes |
| Yes | | |

| | gp_unemployment_why | gp_unemployment_why_other | \ |
|-------|-----------------------------------|---------------------------|---|
| 0 | NaN | NaN | |
| 1 | NaN | NaN | |
| 2 | NaN | NaN | |
| 3 | Laid off while business continues | NaN | |
| 4 | NaN | NaN | |
| ... | ... | ... | |
| 52485 | Business / office closed | NaN | |
| 52486 | NaN | NaN | |
| 52487 | NaN | NaN | |
| 52488 | NaN | NaN | |
| 52489 | NaN | NaN | |

| | gp_pre_income | gp_post_income | country | weight |
|-------|---------------|----------------|--------------------------|----------|
| \ | | | | |
| 0 | 50000.0 | 999497.0 | Uganda | 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador | 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America | 1.898763 |
| 3 | 12.0 | 12.0 | Colombia | 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia | 1.691686 |
| ... | ... | ... | ... | ... |
| 52485 | 900.0 | 900.0 | Italy | 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria | 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia | 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso | 1.795822 |
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |

| | user_id |
|-------|---------------------|
| 0 | gp_6372662088826880 |
| 1 | gp_5900473574883328 |
| 2 | gp_4813642242981888 |
| 3 | gp_4703669741420544 |
| 4 | gp_4762741153988608 |
| ... | ... |
| 52485 | gp_5444968770699264 |
| 52486 | gp_5482738571804672 |
| 52487 | gp_5658409882091520 |
| 52488 | gp_5003001999917056 |
| 52489 | gp_5388715507843072 |

[52490 rows x 48 columns]

#zmiana nazwy kolumny - kolumna education -> school

```
data_frame.rename(columns={"education": "school"}, inplace = True)
data_frame
```

| | observation_id | submitted_time | \ |
|-------|---------------------|-------------------------|-----|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 | UTC |
| 1 | gp_4503631639609344 | 2020-07-09 13:22:37.107 | UTC |
| 2 | gp_4503700758593536 | 2020-07-04 18:53:36.471 | UTC |
| 3 | gp_4503737805832192 | 2020-07-12 17:58:20.798 | UTC |
| 4 | gp_4503819343101952 | 2020-07-06 00:20:22.983 | UTC |
| ... | ... | ... | ... |
| 52485 | gp_6755106844573696 | 2020-07-02 12:37:51.894 | UTC |
| 52486 | gp_6755213279232000 | 2020-07-02 07:07:44.269 | UTC |
| 52487 | gp_6755237508677632 | 2020-07-02 10:26:38.461 | UTC |
| 52488 | gp_6755275458740224 | 2020-07-24 23:16:52.556 | UTC |
| 52489 | gp_6755355351842816 | 2020-07-02 13:33:00.269 | UTC |

| | gender | age | \ |
|-------|-----------------------|--------------------|-----|
| 0 | Male | Under 16 | |
| 1 | Female | 26 to 35 years old | |
| 2 | Male | 36 to 45 years old | |
| 3 | Male | 26 to 35 years old | |
| 4 | Male | 26 to 35 years old | |
| ... | ... | ... | ... |
| 52485 | Female | 36 to 45 years old | |
| 52486 | Male | 26 to 35 years old | |
| 52487 | Prefer not to respond | 16 to 25 years old | |
| 52488 | Prefer not to respond | 26 to 35 years old | |
| 52489 | Female | 26 to 35 years old | |

| | geography | \ |
|---|----------------------------------|---|
| 0 | Suburban/Peri-urban | |
| 1 | City center or metropolitan area | |

| | |
|-------|----------------------------------|
| 2 | City center or metropolitan area |
| 3 | Rural |
| 4 | Suburban/Peri-urban |
| ... | ... |
| 52485 | Suburban/Peri-urban |
| 52486 | City center or metropolitan area |
| 52487 | Suburban/Peri-urban |
| 52488 | Rural |
| 52489 | City center or metropolitan area |

| | financial_situation | \ |
|-------|---|---|
| 0 | I can afford food and regular expenses, but no... | |
| 1 | I cannot afford enough food for my family | |
| 2 | I can comfortably afford food, clothes, and fu... | |
| 3 | I can afford food and regular expenses, and bu... | |
| 4 | I can afford food and regular expenses, and bu... | |
| ... | ... | |
| 52485 | I can afford food, but nothing else | |
| 52486 | I can afford food and regular expenses, but no... | |
| 52487 | I can afford food, but nothing else | |
| 52488 | I cannot afford enough food for my family | |
| 52489 | I cannot afford enough food for my family | |

| | school | employment_status | \ |
|-------|-----------------------|----------------------------|---|
| 0 | Secondary/high school | Employed full-time | |
| 1 | College or university | Unemployed | |
| 2 | Primary school | Employed full-time | |
| 3 | Technical school | Student | |
| 4 | Technical school | Employed full-time | |
| ... | ... | ... | |
| 52485 | Secondary/high school | Employed part-time | |
| 52486 | College or university | Employed full-time | |
| 52487 | Secondary/high school | Self-employed | |
| 52488 | Prefer not to answer | Student and work part-time | |
| 52489 | Secondary/high school | Unemployed | |

| | ethnicity | religion | ... | \ |
|-------|----------------------|----------------------|-----|---|
| 0 | Ankole | Christianity | ... | |
| 1 | Mestizo | Catholicism | ... | |
| 2 | Non-hispanic White | Agnosticism | ... | |
| 3 | Mestizo | Christianity | ... | |
| 4 | Mestizo | Catholicism | ... | |
| ... | ... | ... | ... | |
| 52485 | Not Available | Catholicism | ... | |
| 52486 | Arab | Muslim (Sunni) | ... | |
| 52487 | Grebo | Prefer Not To Answer | ... | |
| 52488 | Prefer not to answer | Other | ... | |
| 52489 | Tagalog | Christianity | ... | |

gp_post_miss_dose_why_other gp_post_labor_force

| gp_pre_labor_force \ | | |
|----------------------|-----|-----|
| 0 | NaN | No |
| No | | |
| 1 | NaN | Yes |
| Yes | | |
| 2 | NaN | Yes |
| Yes | | |
| 3 | NaN | No |
| Yes | | |
| 4 | NaN | Yes |
| Yes | | |
| ... | ... | ... |
| .. | | |
| 52485 | NaN | No |
| Yes | | |
| 52486 | NaN | Yes |
| Yes | | |
| 52487 | NaN | No |
| No | | |
| 52488 | NaN | Yes |
| No | | |
| 52489 | NaN | Yes |
| Yes | | |

| | gp_unemployment_why | gp_unemployment_why_other \ |
|-------|-----------------------------------|-----------------------------|
| 0 | NaN | NaN |
| 1 | NaN | NaN |
| 2 | NaN | NaN |
| 3 | Laid off while business continues | NaN |
| 4 | NaN | NaN |
| ... | ... | ... |
| 52485 | Business / office closed | NaN |
| 52486 | NaN | NaN |
| 52487 | NaN | NaN |
| 52488 | NaN | NaN |
| 52489 | NaN | NaN |

| | gp_pre_income | gp_post_income | country | weight |
|-----|---------------|----------------|--------------------------|----------|
| \ | | | | |
| 0 | 50000.0 | 999497.0 | Uganda | 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador | 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America | 1.898763 |
| 3 | 12.0 | 12.0 | Colombia | 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia | 1.691686 |
| ... | ... | ... | ... | ... |

| | | | | |
|-------|-----------|----------|--------------|----------|
| 52485 | 900.0 | 900.0 | Italy | 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria | 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia | 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso | 1.795822 |
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |

| | user_id |
|-------|---------------------|
| 0 | gp_6372662088826880 |
| 1 | gp_5900473574883328 |
| 2 | gp_4813642242981888 |
| 3 | gp_4703669741420544 |
| 4 | gp_4762741153988608 |
| ... | ... |
| 52485 | gp_5444968770699264 |
| 52486 | gp_5482738571804672 |
| 52487 | gp_5658409882091520 |
| 52488 | gp_5003001999917056 |
| 52489 | gp_5388715507843072 |

[52490 rows x 48 columns]

#zapisanie ramki danych do pliku csv

`data_frame.to_csv('data_frame.csv')`

#wyświetlenie średniej

`data_frame.weight.mean()`

`np.float64(1.7958203667990285)`

#wyświetlenie maksymalnej wartości kolumny

`data_frame.weight.max()`

`np.float64(6.411419833)`

#wyświetlenie minimalnej wartości kolumny

`data_frame.weight.min()`

`np.float64(1.000000002)`

#wyświetlenie liczby wierszy

#wyświetlenie maksymalnej wartości kolumny

`data_frame.weight.count()`

`np.int64(52490)`


```
#wyświetlenie wartości unikalnych
```

```
data_frame.age.unique()
```

```
array(['Under 16', '26 to 35 years old', '36 to 45 years old',  
      '16 to 25 years old', 'Over 45 years old', 'Not Available'],  
      dtype=object)
```

```
#wyświetlenie liczby rekordów odpowiadających wartości
```

```
data_frame.country.value_counts()
```

```
country  
Philippines          5864  
Afghanistan          3836  
Indonesia            2989  
Venezuela (Bolivarian Republic of)  2474  
Mali                 2440  
...  
South Africa         77  
Saudi Arabia         64  
Niger                55  
Bahrain              52  
Oman                 49
```

```
Name: count, Length: 76, dtype: int64
```

```
#sortowanie wierszy ramki danych według wartości określonej kolumny
```

```
data_frame.sort_values(['age'], ascending = False)
```

| | observation_id | submitted_time | gender \ |
|-------|---------------------|-----------------------------|----------|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 UTC | Male |
| 39805 | gp_6212113459838976 | 2020-07-01 16:26:44.48 UTC | Male |
| 39774 | gp_6210954120658944 | 2020-07-18 20:27:53.831 UTC | Female |
| 7745 | gp_4835939500425216 | 2020-07-03 07:36:54.128 UTC | Male |
| 39769 | gp_6210720447594496 | 2020-07-07 19:59:10.754 UTC | Male |
| ... | ... | ... | ... |
| 31870 | gp_5876032570916864 | 2020-07-18 01:37:13.994 UTC | Male |
| 31869 | gp_5875986900189184 | 2020-07-01 17:16:36.293 UTC | Male |
| 31868 | gp_5875928183078912 | 2020-07-02 01:42:03.266 UTC | Female |
| 12528 | gp_5039420454731776 | 2020-07-11 00:38:04.472 UTC | Male |
| 26245 | gp_5637891361603584 | 2020-06-30 23:40:51.831 UTC | Female |

| | age | geography \ |
|-------|--------------------|----------------------------------|
| 0 | Under 16 | Suburban/Peri-urban |
| 39805 | Under 16 | Rural |
| 39774 | Under 16 | Rural |
| 7745 | Under 16 | Rural |
| 39769 | Under 16 | Rural |
| ... | ... | ... |
| 31870 | 16 to 25 years old | City center or metropolitan area |
| 31869 | 16 to 25 years old | Rural |
| 31868 | 16 to 25 years old | City center or metropolitan area |

| | | |
|-------|--------------------|----------------------------------|
| 12528 | 16 to 25 years old | City center or metropolitan area |
| 26245 | 16 to 25 years old | Rural |

| | financial_situation \ |
|-------|---|
| 0 | I can afford food and regular expenses, but no... |
| 39805 | I can afford food, but nothing else |
| 39774 | I can afford food and regular expenses, and bu... |
| 7745 | I can afford food and regular expenses, but no... |
| 39769 | I cannot afford enough food for my family |
| ... | ... |
| 31870 | I cannot afford enough food for my family |
| 31869 | I can afford food and regular expenses, and bu... |
| 31868 | I can afford food and regular expenses, but no... |
| 12528 | I cannot afford enough food for my family |
| 26245 | I can afford food and regular expenses, but no... |

| | school | employment_status |
|-------------|-----------------------|----------------------------|
| ethnicity \ | | |
| 0 | Secondary/high school | Employed full-time |
| Ankole | | |
| 39805 | Secondary/high school | Student |
| Sarakole | | |
| 39774 | Technical school | Employed part-time |
| Igbo | | |
| 7745 | Secondary/high school | Self-employed |
| Bisaya | | |
| 39769 | Primary school | Self-employed |
| Malinke | | |
| ... | ... | ... |
| ... | | |
| 31870 | College or university | Unemployed |
| to say | | Prefer not |
| 31869 | Secondary/high school | Employed full-time |
| Northern | | |
| Mande | | |
| 31868 | Secondary/high school | Student and work part-time |
| Tagalog | | |
| 12528 | Technical school | Employed full-time |
| Arab | | |
| 26245 | Technical school | Student |
| Other | | |

| | religion | ... gp_post_miss_dose_why_other \ |
|-------|----------------------|-----------------------------------|
| 0 | Christianity | NaN |
| 39805 | Muslim | NaN |
| 39774 | Christianity | NaN |
| 7745 | Other | NaN |
| 39769 | Muslim | NaN |
| ... | ... | ... |
| 31870 | Prefer Not To Answer | NaN |
| 31869 | Muslim | NaN |

| | | | | |
|-------------|---------------------------|--------------------|--------------------------|-------|
| 31868 | Christianity | ... | | NaN |
| 12528 | Muslim (Shia) | ... | | NaN |
| 26245 | Christianity | ... | | NaN |
| | | | | |
| | gp_post_labor_force | gp_pre_labor_force | gp_unemployment_why | |
| \ | | | | |
| 0 | No | No | | NaN |
| 39805 | No | No | | NaN |
| 39774 | No | No | | NaN |
| 7745 | Yes | Yes | | NaN |
| 39769 | No | No | | NaN |
| ... | ... | ... | | ... |
| 31870 | No | Yes | Business / office closed | |
| 31869 | Yes | Yes | | NaN |
| 31868 | No | Yes | Business / office closed | |
| 12528 | No | No | | NaN |
| 26245 | No | No | | NaN |
| | | | | |
| | gp_unemployment_why_other | gp_pre_income | gp_post_income | |
| country \ | | | | |
| 0 | NaN | 50000.0 | 999497.0 | |
| Uganda | | | | |
| 39805 | NaN | 82475574.0 | 8247557.0 | |
| Mali | | | | |
| 39774 | NaN | 1.0 | 1.0 | |
| Nigeria | | | | |
| 7745 | NaN | 5000.0 | 5000.0 | |
| Philippines | | | | |
| 39769 | NaN | 94066792.0 | 94066792.0 | |
| Mali | | | | |
| ... | ... | ... | ... | |
| ... | | | | |
| 31870 | NaN | 800.0 | 808000.0 | |
| Colombia | | | | |
| 31869 | NaN | 170000.0 | 50000.0 | Ivory |
| Coast | | | | |
| 31868 | NaN | 40000.0 | 35000.0 | |
| Philippines | | | | |
| 12528 | NaN | 0.0 | 0.0 | |

| | | | |
|----------|-----|-------|-------|
| Iraq | | | |
| 26245 | NaN | 800.0 | 800.0 |
| Colombia | | | |

| | weight | user_id |
|-------|----------|---------------------|
| 0 | 2.675598 | gp_6372662088826880 |
| 39805 | 2.165410 | gp_4816278315597824 |
| 39774 | 2.207765 | gp_5680568787795968 |
| 7745 | 1.929975 | gp_6559930632372224 |
| 39769 | 2.711022 | gp_4556741239111680 |
| ... | ... | ... |
| 31870 | 1.625813 | gp_4860486959235072 |
| 31869 | 1.000000 | gp_4655799541956608 |
| 31868 | 1.550364 | gp_4886213289050112 |
| 12528 | 1.655397 | gp_5669009610768384 |
| 26245 | 1.792738 | gp_5200759995236352 |

[52490 rows x 48 columns]

#wyświetlanie wierszy dla 10 największych wartości określonej kolumny
data_frame.nlargest(10, 'weight')

| | observation_id | submitted_time | gender | \ |
|-------|---------------------|-----------------------------|--------|---|
| 3774 | gp_4665227603083264 | 2020-07-24 03:57:54.877 UTC | Female | |
| 4482 | gp_4696567945887744 | 2020-07-15 10:33:37.864 UTC | Male | |
| 7265 | gp_4816317741006848 | 2020-07-14 19:20:11.775 UTC | Female | |
| 7940 | gp_4845288906031104 | 2020-07-05 22:12:21.834 UTC | Female | |
| 9809 | gp_4924182455648256 | 2020-07-05 09:29:28.805 UTC | Female | |
| 10722 | gp_4962030073413632 | 2020-07-02 16:09:01.896 UTC | Male | |
| 17281 | gp_5248369841930240 | 2020-07-02 22:17:57.454 UTC | Male | |
| 18131 | gp_5283973979963392 | 2020-07-02 07:32:48.826 UTC | Female | |
| 21289 | gp_5419180758401024 | 2020-07-05 13:26:40.929 UTC | Male | |
| 26953 | gp_5669895447773184 | 2020-07-15 23:50:54.245 UTC | Female | |

| | age | geography | \ |
|-------|--------------------|----------------------------------|---|
| 3774 | Under 16 | City center or metropolitan area | |
| 4482 | 26 to 35 years old | City center or metropolitan area | |
| 7265 | Under 16 | Suburban/Peri-urban | |
| 7940 | Under 16 | City center or metropolitan area | |
| 9809 | 16 to 25 years old | City center or metropolitan area | |
| 10722 | 36 to 45 years old | Suburban/Peri-urban | |
| 17281 | Under 16 | City center or metropolitan area | |
| 18131 | 26 to 35 years old | City center or metropolitan area | |
| 21289 | 36 to 45 years old | Rural | |
| 26953 | Under 16 | Rural | |

| | financial_situation |
|--------|---|
| school | \ |
| 3774 | I can afford food and regular expenses, but no... No formal education |

| | | |
|-------|---|---------------------|
| 4482 | I can afford food and regular expenses, and bu... | No formal education |
| 7265 | I cannot afford enough food for my family | No formal education |
| 7940 | I can afford food, but nothing else | No formal education |
| 9809 | I cannot afford enough food for my family | No formal education |
| 10722 | I can afford food and regular expenses, and bu... | No formal education |
| 17281 | I can afford food, but nothing else | No formal education |
| 18131 | I cannot afford enough food for my family | No formal education |
| 21289 | I can afford food, but nothing else | Primary school |
| 26953 | I cannot afford enough food for my family | Primary school |

| | employment_status | ethnicity | religion |
|-------|--------------------|----------------------|----------------------|
| ... | \ | | |
| 3774 | Unemployed | Ukrainian | Prefer Not To Answer |
| ... | | | |
| 4482 | Self-employed | Russian | Agnosticism |
| ... | | | |
| 7265 | Retired | Mestizo | Evangelicalism |
| ... | | | |
| 7940 | Student | White | Catholicism |
| ... | | | |
| 9809 | Unemployed | Somali | Muslim |
| ... | | | |
| 10722 | Retired | Afro-Asian | Other |
| ... | | | |
| 17281 | Self-employed | Prefer not to answer | Vodoun |
| ... | | | |
| 18131 | Unemployed | Arab | Muslim (Shia) |
| ... | | | |
| 21289 | Employed full-time | Fulani | Muslim |
| ... | | | |
| 26953 | Employed part-time | Arab (Other) | Muslim (Sunni) |
| ... | | | |

| | gp_post_miss_dose_why_other | gp_post_labor_force |
|----------------------|-----------------------------|---------------------|
| gp_pre_labor_force \ | | |
| 3774 | NaN | Yes |
| No | | |
| 4482 | NaN | Yes |
| No | | |
| 7265 | NaN | Yes |

| | | |
|-------|-----|-----|
| Yes | | |
| 7940 | NaN | No |
| No | | |
| 9809 | NaN | Yes |
| No | | |
| 10722 | NaN | Yes |
| No | | |
| 17281 | NaN | Yes |
| No | | |
| 18131 | NaN | Yes |
| Yes | | |
| 21289 | NaN | Yes |
| Yes | | |
| 26953 | NaN | Yes |
| Yes | | |

| | gp_unemployment_why | gp_unemployment_why_other | gp_pre_income | \ |
|-------|---------------------|---------------------------|---------------|---|
| 3774 | NaN | NaN | 1.5 | |
| 4482 | NaN | NaN | 0.0 | |
| 7265 | NaN | NaN | 800.0 | |
| 7940 | NaN | NaN | 0.0 | |
| 9809 | NaN | NaN | 10.0 | |
| 10722 | NaN | NaN | 222.0 | |
| 17281 | NaN | NaN | 50.0 | |
| 18131 | NaN | NaN | 80.0 | |
| 21289 | NaN | NaN | 70000.0 | |
| 26953 | NaN | NaN | 300.0 | |

| | gp_post_income | country | weight | user_id |
|-------|----------------|--------------|---------|---------------------|
| 3774 | 0.0 | Ukraine | 6.41142 | gp_5850331822161920 |
| 4482 | 150000.0 | Kazakhstan | 6.41142 | gp_5140308383236096 |
| 7265 | 530.0 | Panama | 6.41142 | gp_6286689484144640 |
| 7940 | 0.0 | Colombia | 6.41142 | gp_5154587281719296 |
| 9809 | 10.0 | Somalia | 6.41142 | gp_5852393440018432 |
| 10722 | 222.0 | Saudi Arabia | 6.41142 | gp_4797705280290816 |
| 17281 | 50.0 | Benin | 6.41142 | gp_5616480036585472 |
| 18131 | 80.0 | Lebanon | 6.41142 | gp_6566692340105216 |
| 21289 | 70000.0 | Niger | 6.41142 | gp_4658733716340736 |
| 26953 | 0.0 | Bahrain | 6.41142 | gp_4801118388617216 |

[10 rows x 48 columns]

#wyświetlanie wierszy dla 10 najmniejszych wartości określonej kolumny
data_frame.nsmallest(10, 'weight')

| | observation_id | submitted_time | gender | \ |
|-----|---------------------|-----------------------------|--------|---|
| 144 | gp_4509416591458304 | 2020-07-25 11:16:23.805 UTC | Female | |
| 520 | gp_4526482610454528 | 2020-07-17 12:58:56.165 UTC | Male | |
| 852 | gp_4540126480039936 | 2020-07-07 18:23:05.845 UTC | Male | |
| 960 | gp_4544756689666048 | 2020-07-02 12:36:58.115 UTC | Male | |

| | | | |
|------|---------------------|-----------------------------|--------|
| 964 | gp_4544937145401344 | 2020-07-12 14:25:59.905 UTC | Male |
| 1363 | gp_4561830661062656 | 2020-07-02 16:05:59.404 UTC | Male |
| 1398 | gp_4563247127855104 | 2020-07-14 10:33:08.86 UTC | Female |
| 1567 | gp_4569919158222848 | 2020-07-05 11:00:07.838 UTC | Female |
| 1684 | gp_4574351463809024 | 2020-07-05 16:04:36.891 UTC | Male |
| 1872 | gp_4583458099953664 | 2020-07-03 14:11:13.9 UTC | Male |

| | age | geography \ |
|------|--------------------|----------------------------------|
| 144 | 16 to 25 years old | City center or metropolitan area |
| 520 | 16 to 25 years old | Suburban/Peri-urban |
| 852 | 16 to 25 years old | Suburban/Peri-urban |
| 960 | 16 to 25 years old | City center or metropolitan area |
| 964 | 16 to 25 years old | City center or metropolitan area |
| 1363 | 26 to 35 years old | Rural |
| 1398 | 16 to 25 years old | City center or metropolitan area |
| 1567 | 16 to 25 years old | City center or metropolitan area |
| 1684 | 16 to 25 years old | Rural |
| 1872 | 16 to 25 years old | City center or metropolitan area |

| | financial_situation \ |
|------|---|
| 144 | I can afford food and regular expenses, and bu... |
| 520 | I can afford food and regular expenses, and bu... |
| 852 | I cannot afford enough food for my family |
| 960 | I can afford food and regular expenses, and bu... |
| 964 | I can afford food and regular expenses, and bu... |
| 1363 | I can afford food, but nothing else |
| 1398 | I can afford food and regular expenses, but no... |
| 1567 | I cannot afford enough food for my family |
| 1684 | I can afford food and regular expenses, but no... |
| 1872 | I can afford food and regular expenses, and bu... |

| | school | employment_status |
|--------------|-----------------------|----------------------------|
| ethnicity \ | | |
| 144 | College or university | Student |
| Chinese | | |
| 520 | College or university | Student |
| answer | | Prefer not to |
| 852 | Secondary/high school | Student |
| Other | | |
| 960 | College or university | Student and work part-time |
| Voltaire/Gur | | |
| 964 | Secondary/high school | Student |
| Kru | | |
| 1363 | Secondary/high school | Unemployed |
| Akan | | |
| 1398 | College or university | Employed part-time |
| Chinese | | |
| 1567 | Secondary/high school | Student |
| Chinese | | |
| 1684 | College or university | Student |

Other
 1872 College or university Student
 Northern Mande

| | religion | ... | gp_post_miss_dose_why_other | \ |
|------|----------------------|-----|-----------------------------|---|
| 144 | Agnosticism | ... | NaN | |
| 520 | Catholicism | ... | NaN | |
| 852 | Catholicism | ... | NaN | |
| 960 | Catholicism | ... | NaN | |
| 964 | Catholicism | ... | NaN | |
| 1363 | Catholicism | ... | NaN | |
| 1398 | Other | ... | NaN | |
| 1567 | Buddhism | ... | NaN | |
| 1684 | Prefer Not To Answer | ... | NaN | |
| 1872 | Muslim | ... | NaN | |

| | gp_post_labor_force | gp_pre_labor_force | gp_unemployment_why | \ |
|------|---------------------|--------------------|---------------------|---|
| 144 | No | No | NaN | |
| 520 | Yes | Yes | NaN | |
| 852 | No | No | NaN | |
| 960 | No | No | NaN | |
| 964 | Yes | Yes | NaN | |
| 1363 | No | No | NaN | |
| 1398 | Yes | Yes | NaN | |
| 1567 | No | No | NaN | |
| 1684 | Yes | No | NaN | |
| 1872 | No | No | NaN | |

| | gp_unemployment_why_other | gp_pre_income | gp_post_income | |
|-----------|---------------------------|---------------|----------------|-------|
| country \ | | | | |
| 144 | NaN | 0.0 | 0.0 | Hong |
| Kong | | | | |
| 520 | NaN | 0.0 | 0.0 | Ivory |
| Coast | | | | |
| 852 | NaN | 1.0 | 1000.0 | Ivory |
| Coast | | | | |
| 960 | NaN | 0.0 | 0.0 | Ivory |
| Coast | | | | |
| 964 | NaN | 21000.0 | 27000.0 | Ivory |
| Coast | | | | |
| 1363 | NaN | 150000.0 | 150000.0 | Ivory |
| Coast | | | | |
| 1398 | NaN | 6000.0 | 6000.0 | Hong |
| Kong | | | | |
| 1567 | NaN | 0.0 | 0.0 | Hong |
| Kong | | | | |
| 1684 | NaN | 80000.0 | 80000.0 | Ivory |
| Coast | | | | |
| 1872 | NaN | 0.0 | 0.0 | Ivory |
| Coast | | | | |

| | weight | user_id |
|------|--------|---------------------|
| 144 | 1.0 | gp_5058356632354816 |
| 520 | 1.0 | gp_6142385918836736 |
| 852 | 1.0 | gp_4852694114369536 |
| 960 | 1.0 | gp_6358894674182144 |
| 964 | 1.0 | gp_6238940295331840 |
| 1363 | 1.0 | gp_4934027729698816 |
| 1398 | 1.0 | gp_5776393856352256 |
| 1567 | 1.0 | gp_5899587211755520 |
| 1684 | 1.0 | gp_5676395832803328 |
| 1872 | 1.0 | gp_6429813654421504 |

[10 rows x 48 columns]

#wyświetlenie wierszy da 10 największych wartości określonej kolumny pod warunkiem określonych wartości innej kolumny

```
data_frame[data_frame['ethnicity'] == 'Chinese'].nlargest(10, 'weight')
```

| | observation_id | submitted_time | gender \ |
|-------|---------------------|-----------------------------|----------|
| 52010 | gp_6735853616627712 | 2020-07-26 00:14:59.342 UTC | Male |
| 26194 | gp_5635674017628160 | 2020-07-01 22:36:13.34 UTC | Female |
| 33994 | gp_5964867221848064 | 2020-07-02 01:46:35.224 UTC | Female |
| 37808 | gp_6126128483008512 | 2020-07-02 04:10:15.291 UTC | Female |
| 19992 | gp_5365734990675968 | 2020-07-04 10:10:16.268 UTC | Female |
| 5229 | gp_4729703346995200 | 2020-07-01 23:39:03.512 UTC | Male |
| 27229 | gp_5681794151350272 | 2020-07-25 14:54:52.485 UTC | Male |
| 28872 | gp_5750585669451776 | 2020-07-06 12:55:10.953 UTC | Male |
| 47189 | gp_6528578931130368 | 2020-07-25 10:51:26.623 UTC | Male |
| 52101 | gp_6739600402022400 | 2020-07-05 10:20:31.806 UTC | Male |

| | age | geography \ |
|-------|-------------------|----------------------------------|
| 52010 | Under 16 | City center or metropolitan area |
| 26194 | Over 45 years old | Suburban/Peri-urban |
| 33994 | Over 45 years old | Suburban/Peri-urban |
| 37808 | Over 45 years old | Suburban/Peri-urban |
| 19992 | Over 45 years old | Suburban/Peri-urban |
| 5229 | Under 16 | Suburban/Peri-urban |
| 27229 | Under 16 | City center or metropolitan area |
| 28872 | Under 16 | City center or metropolitan area |
| 47189 | Under 16 | City center or metropolitan area |
| 52101 | Under 16 | Suburban/Peri-urban |

| | financial_situation \ |
|-------|---|
| 52010 | I can comfortably afford food, clothes, and fu... |
| 26194 | I can afford food and regular expenses, and bu... |
| 33994 | I can afford food and regular expenses, and bu... |
| 37808 | I can comfortably afford food, clothes, and fu... |
| 19992 | I can afford food, but nothing else |

5229 I can afford food and regular expenses, and bu...
 27229 I cannot afford enough food for my family
 28872 I cannot afford enough food for my family
 47189 I can afford food and regular expenses, and bu...
 52101 I can afford food and regular expenses, but no...

| | school | employment_status | ethnicity \ |
|-------|-----------------------|----------------------------|-------------|
| 52010 | Primary school | Student | Chinese |
| 26194 | Secondary/high school | Employed full-time | Chinese |
| 33994 | Secondary/high school | Retired | Chinese |
| 37808 | Secondary/high school | Retired | Chinese |
| 19992 | College or university | Unemployed | Chinese |
| 5229 | Secondary/high school | Student | Chinese |
| 27229 | Secondary/high school | Student | Chinese |
| 28872 | Secondary/high school | Student | Chinese |
| 47189 | Secondary/high school | Student and work part-time | Chinese |
| 52101 | Secondary/high school | Student | Chinese |

| | religion | ... gp_post_miss_dose_why_other |
|-----------------------|--------------|---------------------------------|
| gp_post_labor_force \ | | |
| 52010 | Buddhism | NaN |
| Yes | | |
| 26194 | Buddhism | NaN |
| Yes | | |
| 33994 | Christianity | NaN |
| No | | |
| 37808 | Christianity | NaN |
| Yes | | |
| 19992 | Buddhism | NaN |
| No | | |
| 5229 | Christianity | NaN |
| No | | |
| 27229 | Buddhism | NaN |
| No | | |
| 28872 | Christianity | NaN |
| Yes | | |
| 47189 | Buddhism | NaN |
| No | | |
| 52101 | Buddhism | NaN |
| No | | |

| | gp_pre_labor_force |
|-----------------------|--------------------|
| gp_unemployment_why \ | |
| 52010 | No |
| NaN | |
| 26194 | Yes |
| NaN | |
| 33994 | No |
| NaN | |
| 37808 | Yes |

| | |
|-------|-----|
| NaN | |
| 19992 | No |
| NaN | |
| 5229 | No |
| NaN | |
| 27229 | No |
| NaN | |
| 28872 | Yes |
| NaN | |
| 47189 | No |
| NaN | |
| 52101 | Yes |

Not able to go to work due to government
manda...

| | gp_unemployment_why_other | gp_pre_income | gp_post_income | country |
|-------|---------------------------|---------------|----------------|----------|
| \ | | | | |
| 52010 | NaN | 0.0 | 0.0 | Malaysia |
| 26194 | NaN | 1000.0 | 1200.0 | Malaysia |
| 33994 | NaN | 0.0 | 0.0 | Malaysia |
| 37808 | NaN | 500.0 | 500.0 | Malaysia |
| 19992 | NaN | 0.0 | 0.0 | Malaysia |
| 5229 | NaN | 4000.0 | 4000.0 | Malaysia |
| 27229 | NaN | 0.0 | 0.0 | Malaysia |
| 28872 | NaN | 5000.0 | 4500.0 | Malaysia |
| 47189 | NaN | 0.0 | 0.0 | Malaysia |
| 52101 | NaN | 3500.0 | 3000.0 | Malaysia |

| | weight | user_id |
|-------|----------|---------------------|
| 52010 | 3.206059 | gp_4567382006956032 |
| 26194 | 2.584217 | gp_5337201199546368 |
| 33994 | 2.584217 | gp_5707177534947328 |
| 37808 | 2.584217 | gp_6211090023186432 |
| 19992 | 2.553618 | gp_6165029812961280 |
| 5229 | 2.502590 | gp_5431670490005504 |
| 27229 | 2.502590 | gp_5746967461494784 |
| 28872 | 2.502590 | gp_5013102881406976 |
| 47189 | 2.502590 | gp_6552871429996544 |
| 52101 | 2.502590 | gp_6123853283065856 |

[10 rows x 48 columns]

#grupowanie wierszy według wartości kolumny kategoryzowanej, potem uśrednienie wartości dla pewnych kolumn, liczba wartości i mediana dla pozostałych kolumn w grupach

```
data_frame.groupby(['age', 'country']).agg({'weight': 'mean'})
```

| | | weight |
|---------------------------|------------------------------------|----------|
| age 16 to 25 years old | country Afghanistan | 1.519903 |
| | Albania | 2.512004 |
| | Algeria | 1.926268 |
| | Argentina | 1.766030 |
| | Bahrain | 2.646164 |
| ... | | ... |
| Under 16 | Uruguay | 2.977481 |
| | Venezuela (Bolivarian Republic of) | 2.325178 |
| | Viet Nam | 2.784022 |
| | Yemen | 2.552718 |
| | Zambia | 3.676628 |

```
[396 rows x 1 columns]
```

#grupowanie wierszy według wartości kolumny kategoryzowanej, potem uśrednienie wartości dla pewnych kolumn, liczba wartości i mediana dla pozostałych kolumn w grupach

```
group = data_frame.groupby(['age', 'country']).agg({'weight': ['mean', 'median']})
```

#wyświetlenie nazwy kolumn indeksu złożonego

```
group.index.names
```

```
FrozenList(['age', 'country'])
```

#sortowanie kolumn indeksu złożonego

```
group.sort_index()
```

| | | weight |
|--------------------------------|-------------|----------|
| | | mean |
| median age | country | |
| | | |
| 16 to 25 years old 1.467960 | Afghanistan | 1.519903 |
| | Albania | 2.512004 |
| | Algeria | 1.926268 |
| | Argentina | 1.766030 |
| | Bahrain | 2.646164 |
| 2.362396 | | |

```

...
...
Under 16          Uruguay          2.977481
2.872425
          Venezuela (Bolivarian Republic of) 2.325178
2.158791
          Viet Nam          2.784022
2.604028
          Yemen          2.552718
2.525756
          Zambia          3.676628
3.990096

```

```
[396 rows x 2 columns]
```

```
#tworzenie tabeli przestawnej
```

```

df_pivot = data_frame.pivot_table(values = 'age',
index='financial_situation', columns='gender', aggfunc='sum',
margins=False, dropna=True, fill_value=None)
df_pivot

```

```

gender
Female \
financial_situation

```

```

I can afford food and regular expenses, and buy... 26 to 35 years
old26 to 35 years old16 to 25 y...
I can afford food and regular expenses, but not... 26 to 35 years
old36 to 45 years old26 to 35 y...
I can afford food, but nothing else          26 to 35 years
old26 to 35 years old16 to 25 y...
I can comfortably afford food, clothes, and fur... 26 to 35 years
old16 to 25 years old26 to 35 y...
I cannot afford enough food for my family          26 to 35 years
old16 to 25 years old16 to 25 y...

```

```

gender
Male \
financial_situation

```

```

I can afford food and regular expenses, and buy... 26 to 35 years
old26 to 35 years old16 to 25 y...
I can afford food and regular expenses, but not... Under 1626 to 35
years old26 to 35 years old26...
I can afford food, but nothing else          16 to 25 years
old16 to 25 years old26 to 35 y...
I can comfortably afford food, clothes, and fur... 36 to 45 years
old0ver 45 years old0ver 45 yea...
I cannot afford enough food for my family          Under 1616 to 25

```

years old16 to 25 years old16...

gender

Prefer not to respond

financial_situation

I can afford food and regular expenses, and buy... 16 to 25 years

old26 to 35 years old16 to 25 y...

I can afford food and regular expenses, but not... 36 to 45 years

old26 to 35 years old16 to 25 y...

I can afford food, but nothing else 16 to 25 years

old36 to 45 years old36 to 45 y...

I can comfortably afford food, clothes, and fur... Over 45 years

oldOver 45 years old16 to 25 yea...

I cannot afford enough food for my family 16 to 25 years

old26 to 35 years old26 to 35 y...

#wyświetlenie indeksów i kolumn tabelu przestawnej

df_pivot.index

Index(['I can afford food and regular expenses, and buy new clothes
once a year, but nothing else',

'I can afford food and regular expenses, but nothing else',

'I can afford food, but nothing else',

'I can comfortably afford food, clothes, and furniture, and I
have savings',

'I cannot afford enough food for my family'],

dtype='object', name='financial_situation')

#wyświetlenie indeksów i kolumn tabelu przestawnej

df_pivot.columns

Index(['Female', 'Male', 'Prefer not to respond'], dtype='object',
name='gender')

#indeks złożony tabeli przestawnej

```
df_pivot = data_frame.pivot_table(values = 'age',  
index=['financial_situation', 'country'], columns='gender',  
aggfunc='sum', margins=False, dropna=True, fill_value=None)  
df_pivot
```

gender

Female \

financial_situation country

I can afford food and regular expenses, and buy... Afghanistan

16 to 25 years old26 to 35 years old16 to 25 y...

Albania

| | | | |
|--|--------------------|--------------------|---------------|
| 26 to 35 years old | 36 to 45 years old | 26 to 35 y... | Algeria |
| 26 to 35 years old | 26 to 35 years old | 16 to 25 y... | Argentina |
| 16 to 25 years old | 26 to 35 years old | 36 to 45 y... | Bahrain |
| 26 to 35 years old | 26 to 35 years old | 16 to 25 y... | |
| ... | | | |
| ... | | | |
| I cannot afford enough food for my family | | | Venezuela |
| (Bolivarian Republic of) | 36 to 45 years old | 16 to 25 years old | 16 to 25 y... |
| | | | |
| | | | Viet Nam |
| 16 to 25 years old | 26 to 35 years old | 16 to 25 y... | Yemen |
| 36 to 45 years old | 26 to 35 years old | 26 to 35 y... | Zambia |
| 16 to 25 years old | 16 to 25 years old | 16 to 25 y... | Zimbabwe |
| 16 to 25 years old | 16 to 25 years old | 36 to 45 y... | |
| gender | | | |
| Male \ | | | |
| financial_situation | | | country |
| I can afford food and regular expenses, and buy... | | | Afghanistan |
| 16 to 25 years old | 26 to 35 years old | 16 to 25 y... | Albania |
| 36 to 45 years old | 16 to 25 years old | 16 to 25 y... | Algeria |
| 26 to 35 years old | 36 to 45 years old | 16 to 25 y... | Argentina |
| 16 to 25 years old | 16 to 25 years old | 16 to 25 y... | Bahrain |
| 36 to 45 years old | 16 to 25 years old | 36 to 45 y... | |
| ... | | | |
| ... | | | |
| I cannot afford enough food for my family | | | Venezuela |
| (Bolivarian Republic of) | 16 to 25 years old | 16 to 25 years old | 16 to 25 y... |
| | | | |
| | | | Viet Nam |
| 36 to 45 years old | 26 to 35 years old | 16 to 25 y... | Yemen |
| 26 to 35 years old | 16 to 25 years old | 16 to 25 y... | Zambia |
| 16 to 25 years old | 16 to 25 years old | 16 to 25 y... | Zimbabwe |
| 36 to 45 years old | 36 to 45 years old | 16 to 25 y... | |
| gender | | | |

| financial_situation | country |
|--|-------------|
| Prefer not to respond | |
| I can afford food and regular expenses, and buy... | Afghanistan |
| 36 to 45 years old26 to 35 years old26 to 35 y... | Albania |
| NaN | Algeria |
| NaN | Argentina |
| NaN | Bahrain |
| NaN | |
| ... | |
| ... | |
| I cannot afford enough food for my family | Venezuela |
| (Bolivarian Republic of) 16 to 25 years old36 to 45 years oldUnder | |
| 1616... | Viet Nam |
| NaN | Yemen |
| 16 to 25 years old16 to 25 years old16 to 25 y... | Zambia |
| Under 16 | Zimbabwe |
| NaN | |

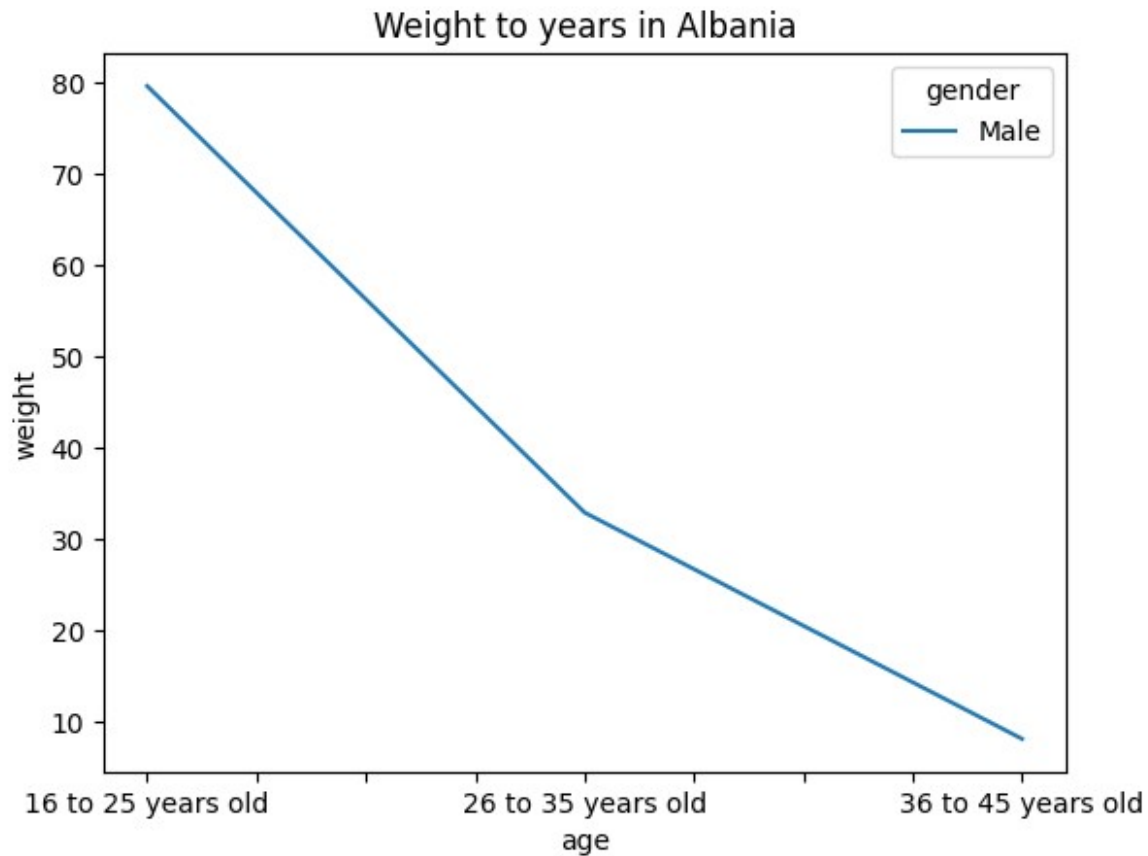
[380 rows x 3 columns]

```
#import matplotlib + wskazanie że należy wykres rysować w zeszycie a
nie w zakładce
import matplotlib.pyplot as plt
%matplotlib inline

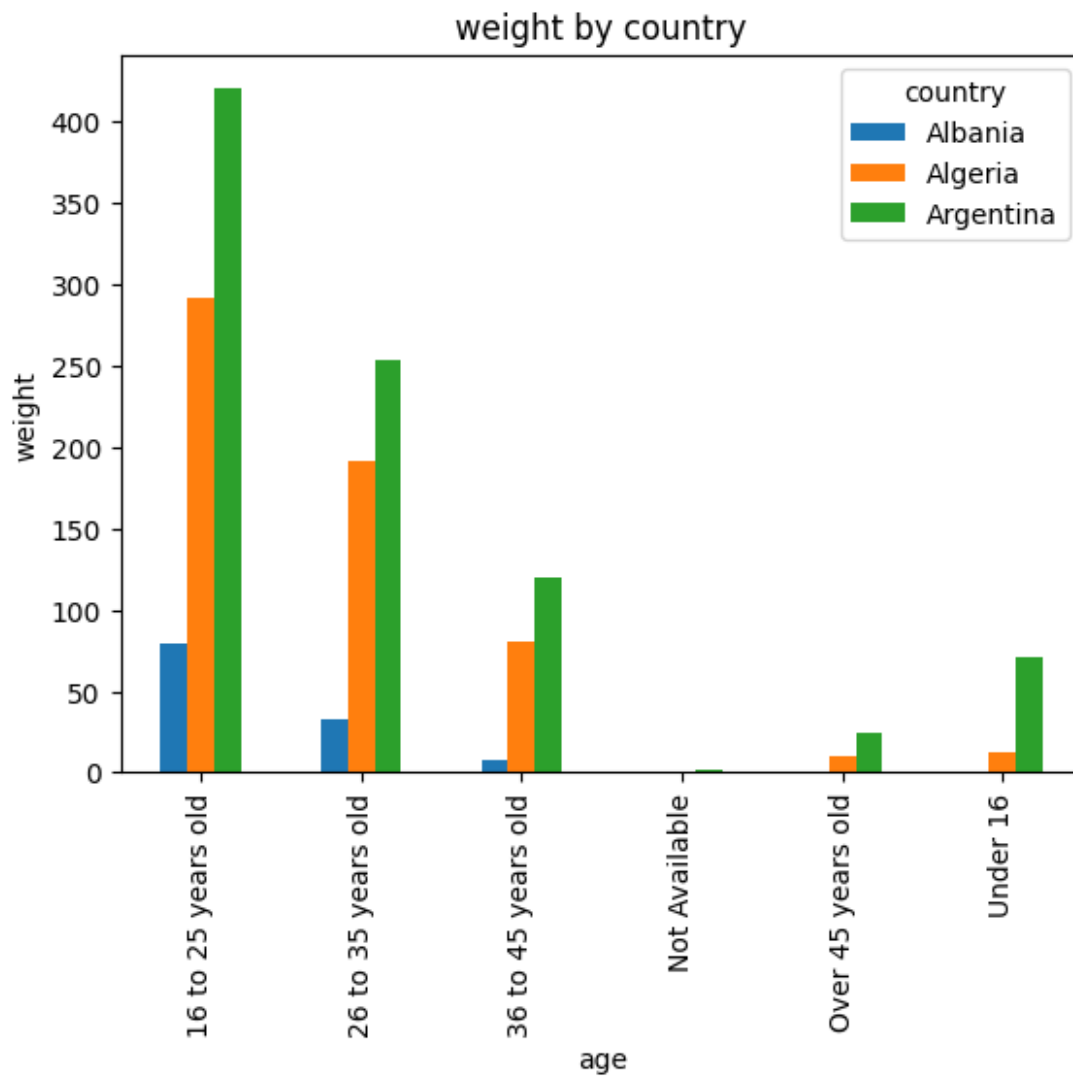
#wykres na podstawie tabeli przestawnej
data_frame[(data_frame['country'] == 'Albania') &
(data_frame['gender'] == 'Male')].pivot_table(values='weight',
index='age', columns='gender', aggfunc='sum',
fill_value=None,
margins=False, dropna=True).plot(kind = 'line')

plt.ylabel('weight') # etykieta osi y
plt.title('Weight to years in Albania') # tytuł wykresu

Text(0.5, 1.0, 'Weight to years in Albania')
```

```
#histogram z tabeli przestawnej
df_bar = data_frame[(data_frame['country'].isin(['Albania', 'Algeria',
'Argentina'])) & (data_frame['gender'] ==
'Male')].pivot_table(values='weight', index='age', columns='country',
aggfunc='sum',
                                fill_value=None,
                                margins=False, dropna=True)
df_bar.plot(kind = 'bar')
plt.ylabel('weight')
plt.title('weight by country')
Text(0.5, 1.0, 'weight by country')
```



```
#łączenie za pomocą merge
df1 = pd.read_csv('data1.csv')

df2 = pd.read_csv('data2.csv')

merged_df = pd.merge(df1, df2, on='ID')
merged_df
```

| | ID | Imię | Miasto |
|---|----|-------|----------|
| 0 | 1 | Jan | Warszawa |
| 1 | 2 | Anna | Kraków |
| 2 | 3 | Piotr | Gdańsk |

```
concat_df = pd.merge(df1, df2)
concat_df
```

| | ID | Imię | Miasto |
|---|----|-------|----------|
| 0 | 1 | Jan | Warszawa |
| 1 | 2 | Anna | Kraków |
| 2 | 3 | Piotr | Gdańsk |

#dodawanie kolumn za pomocą operacji matematycznych

```
data_frame['weight+gp_pre_income'] = data_frame['weight']
+data_frame['gp_pre_income']
data_frame
```

| | observation_id | submitted_time | \ |
|-------|---------------------|-------------------------|-----|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 | UTC |
| 1 | gp_4503631639609344 | 2020-07-09 13:22:37.107 | UTC |
| 2 | gp_4503700758593536 | 2020-07-04 18:53:36.471 | UTC |
| 3 | gp_4503737805832192 | 2020-07-12 17:58:20.798 | UTC |
| 4 | gp_4503819343101952 | 2020-07-06 00:20:22.983 | UTC |
| ... | ... | ... | ... |
| 52485 | gp_6755106844573696 | 2020-07-02 12:37:51.894 | UTC |
| 52486 | gp_6755213279232000 | 2020-07-02 07:07:44.269 | UTC |
| 52487 | gp_6755237508677632 | 2020-07-02 10:26:38.461 | UTC |
| 52488 | gp_6755275458740224 | 2020-07-24 23:16:52.556 | UTC |
| 52489 | gp_6755355351842816 | 2020-07-02 13:33:00.269 | UTC |

| | gender | age | \ |
|-------|-----------------------|--------------------|-----|
| 0 | Male | Under 16 | |
| 1 | Female | 26 to 35 years old | |
| 2 | Male | 36 to 45 years old | |
| 3 | Male | 26 to 35 years old | |
| 4 | Male | 26 to 35 years old | |
| ... | ... | ... | ... |
| 52485 | Female | 36 to 45 years old | |
| 52486 | Male | 26 to 35 years old | |
| 52487 | Prefer not to respond | 16 to 25 years old | |
| 52488 | Prefer not to respond | 26 to 35 years old | |
| 52489 | Female | 26 to 35 years old | |

| | geography | \ |
|-------|----------------------------------|-----|
| 0 | Suburban/Peri-urban | |
| 1 | City center or metropolitan area | |
| 2 | City center or metropolitan area | |
| 3 | Rural | |
| 4 | Suburban/Peri-urban | |
| ... | ... | ... |
| 52485 | Suburban/Peri-urban | |
| 52486 | City center or metropolitan area | |
| 52487 | Suburban/Peri-urban | |
| 52488 | Rural | |
| 52489 | City center or metropolitan area | |

| | | financial_situation | \ |
|-------|---|---------------------|---|
| 0 | I can afford food and regular expenses, but no... | | |
| 1 | I cannot afford enough food for my family | | |
| 2 | I can comfortably afford food, clothes, and fu... | | |
| 3 | I can afford food and regular expenses, and bu... | | |
| 4 | I can afford food and regular expenses, and bu... | | |
| ... | | | |
| 52485 | I can afford food, but nothing else | | |
| 52486 | I can afford food and regular expenses, but no... | | |
| 52487 | I can afford food, but nothing else | | |
| 52488 | I cannot afford enough food for my family | | |
| 52489 | I cannot afford enough food for my family | | |

| | education | employment_status | \ |
|-------|-----------------------|----------------------------|---|
| 0 | Secondary/high school | Employed full-time | |
| 1 | College or university | Unemployed | |
| 2 | Primary school | Employed full-time | |
| 3 | Technical school | Student | |
| 4 | Technical school | Employed full-time | |
| ... | | | |
| 52485 | Secondary/high school | Employed part-time | |
| 52486 | College or university | Employed full-time | |
| 52487 | Secondary/high school | Self-employed | |
| 52488 | Prefer not to answer | Student and work part-time | |
| 52489 | Secondary/high school | Unemployed | |

| | ethnicity | religion | ... |
|--------------------|----------------------|----------------------|-----|
| gp_pre_labor_force | \ | | |
| 0 | Ankole | Christianity | ... |
| No | | | |
| 1 | Mestizo | Catholicism | ... |
| Yes | | | |
| 2 | Non-hispanic White | Agnosticism | ... |
| Yes | | | |
| 3 | Mestizo | Christianity | ... |
| Yes | | | |
| 4 | Mestizo | Catholicism | ... |
| Yes | | | |
| ... | ... | ... | ... |
| ... | | | |
| 52485 | Not Available | Catholicism | ... |
| Yes | | | |
| 52486 | Arab | Muslim (Sunni) | ... |
| Yes | | | |
| 52487 | Grebo | Prefer Not To Answer | ... |
| No | | | |
| 52488 | Prefer not to answer | Other | ... |
| No | | | |
| 52489 | Tagalog | Christianity | ... |

Yes

| | gp_unemployment_why | gp_unemployment_why_other | \ |
|-------|-----------------------------------|---------------------------|---|
| 0 | NaN | NaN | |
| 1 | NaN | NaN | |
| 2 | NaN | NaN | |
| 3 | Laid off while business continues | NaN | |
| 4 | NaN | NaN | |
| ... | ... | ... | |
| 52485 | Business / office closed | NaN | |
| 52486 | NaN | NaN | |
| 52487 | NaN | NaN | |
| 52488 | NaN | NaN | |
| 52489 | NaN | NaN | |

| | gp_pre_income | gp_post_income | country | weight |
|-------|---------------|----------------|--------------------------|----------|
| \ | | | | |
| 0 | 50000.0 | 999497.0 | Uganda | 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador | 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America | 1.898763 |
| 3 | 12.0 | 12.0 | Colombia | 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia | 1.691686 |
| ... | ... | ... | ... | ... |
| 52485 | 900.0 | 900.0 | Italy | 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria | 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia | 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso | 1.795822 |
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |

| | user_id | weight-gp_pre_income | weight+gp_pre_income |
|-------|---------------------|----------------------|----------------------|
| 0 | gp_6372662088826880 | -4.999732e+04 | 5.000268e+04 |
| 1 | gp_5900473574883328 | -2.980028e+02 | 3.019972e+02 |
| 2 | gp_4813642242981888 | -3.198101e+03 | 3.201899e+03 |
| 3 | gp_4703669741420544 | -1.030831e+01 | 1.369169e+01 |
| 4 | gp_4762741153988608 | -2.999998e+06 | 3.000002e+06 |
| ... | ... | ... | ... |
| 52485 | gp_5444968770699264 | -8.977393e+02 | 9.022607e+02 |
| 52486 | gp_5482738571804672 | -1.199998e+06 | 1.200002e+06 |
| 52487 | gp_5658409882091520 | 1.795822e+00 | 1.795822e+00 |
| 52488 | gp_5003001999917056 | -1.320418e+01 | 1.679582e+01 |

```
52489 gp_5388715507843072 -1.399839e+04 1.400161e+04
```

```
[52490 rows x 50 columns]
```

```
#dodawanie kolumn za pomocą lambdy
```

```
data_frame['weight-gp_pre_income'] = data_frame.apply(lambda row:  
row['weight'] - row['gp_pre_income'], axis=1)
```

```
data_frame
```

| | observation_id | submitted_time \ |
|-------|---------------------|-----------------------------|
| 0 | gp_4503617949401088 | 2020-07-07 14:48:29.83 UTC |
| 1 | gp_4503631639609344 | 2020-07-09 13:22:37.107 UTC |
| 2 | gp_4503700758593536 | 2020-07-04 18:53:36.471 UTC |
| 3 | gp_4503737805832192 | 2020-07-12 17:58:20.798 UTC |
| 4 | gp_4503819343101952 | 2020-07-06 00:20:22.983 UTC |
| ... | ... | ... |
| 52485 | gp_6755106844573696 | 2020-07-02 12:37:51.894 UTC |
| 52486 | gp_6755213279232000 | 2020-07-02 07:07:44.269 UTC |
| 52487 | gp_6755237508677632 | 2020-07-02 10:26:38.461 UTC |
| 52488 | gp_6755275458740224 | 2020-07-24 23:16:52.556 UTC |
| 52489 | gp_6755355351842816 | 2020-07-02 13:33:00.269 UTC |

| | gender | age \ |
|-------|-----------------------|--------------------|
| 0 | Male | Under 16 |
| 1 | Female | 26 to 35 years old |
| 2 | Male | 36 to 45 years old |
| 3 | Male | 26 to 35 years old |
| 4 | Male | 26 to 35 years old |
| ... | ... | ... |
| 52485 | Female | 36 to 45 years old |
| 52486 | Male | 26 to 35 years old |
| 52487 | Prefer not to respond | 16 to 25 years old |
| 52488 | Prefer not to respond | 26 to 35 years old |
| 52489 | Female | 26 to 35 years old |

| | geography \ |
|-------|----------------------------------|
| 0 | Suburban/Peri-urban |
| 1 | City center or metropolitan area |
| 2 | City center or metropolitan area |
| 3 | Rural |
| 4 | Suburban/Peri-urban |
| ... | ... |
| 52485 | Suburban/Peri-urban |
| 52486 | City center or metropolitan area |
| 52487 | Suburban/Peri-urban |
| 52488 | Rural |
| 52489 | City center or metropolitan area |

| | financial_situation \ |
|---|---|
| 0 | I can afford food and regular expenses, but no... |

| | |
|-------|---|
| 1 | I cannot afford enough food for my family |
| 2 | I can comfortably afford food, clothes, and fu... |
| 3 | I can afford food and regular expenses, and bu... |
| 4 | I can afford food and regular expenses, and bu... |
| ... | ... |
| 52485 | I can afford food, but nothing else |
| 52486 | I can afford food and regular expenses, but no... |
| 52487 | I can afford food, but nothing else |
| 52488 | I cannot afford enough food for my family |
| 52489 | I cannot afford enough food for my family |

| | education | employment_status \ |
|-------|-----------------------|----------------------------|
| 0 | Secondary/high school | Employed full-time |
| 1 | College or university | Unemployed |
| 2 | Primary school | Employed full-time |
| 3 | Technical school | Student |
| 4 | Technical school | Employed full-time |
| ... | ... | ... |
| 52485 | Secondary/high school | Employed part-time |
| 52486 | College or university | Employed full-time |
| 52487 | Secondary/high school | Self-employed |
| 52488 | Prefer not to answer | Student and work part-time |
| 52489 | Secondary/high school | Unemployed |

| | ethnicity | religion ... |
|----------------------|----------------------|--------------------------|
| gp_pre_labor_force \ | | |
| 0 | Ankole | Christianity ... |
| No | | |
| 1 | Mestizo | Catholicism ... |
| Yes | | |
| 2 | Non-hispanic White | Agnosticism ... |
| Yes | | |
| 3 | Mestizo | Christianity ... |
| Yes | | |
| 4 | Mestizo | Catholicism ... |
| Yes | | |
| ... | ... | ... |
| ... | | |
| 52485 | Not Available | Catholicism ... |
| Yes | | |
| 52486 | Arab | Muslim (Sunni) ... |
| Yes | | |
| 52487 | Grebo | Prefer Not To Answer ... |
| No | | |
| 52488 | Prefer not to answer | Other ... |
| No | | |
| 52489 | Tagalog | Christianity ... |
| Yes | | |

| gp_unemployment_why | gp_unemployment_why_other \ |
|---------------------|-----------------------------|
|---------------------|-----------------------------|

| | | | |
|-------|-----------------------------------|-----|-----|
| 0 | | NaN | NaN |
| 1 | | NaN | NaN |
| 2 | | NaN | NaN |
| 3 | Laid off while business continues | | NaN |
| 4 | | NaN | NaN |
| ... | | ... | ... |
| 52485 | Business / office closed | | NaN |
| 52486 | | NaN | NaN |
| 52487 | | NaN | NaN |
| 52488 | | NaN | NaN |
| 52489 | | NaN | NaN |

| | gp_pre_income | gp_post_income | country | weight |
|-------|---------------|----------------|--------------------------|----------|
| \ | | | | |
| 0 | 50000.0 | 999497.0 | Uganda | 2.675598 |
| 1 | 300.0 | 200.0 | Ecuador | 1.997161 |
| 2 | 3200.0 | 3200.0 | United States of America | 1.898763 |
| 3 | 12.0 | 12.0 | Colombia | 1.691686 |
| 4 | 3000000.0 | 400000.0 | Colombia | 1.691686 |
| ... | ... | ... | ... | ... |
| 52485 | 900.0 | 900.0 | Italy | 2.260748 |
| 52486 | 1200000.0 | 120000.0 | Algeria | 1.896411 |
| 52487 | 0.0 | 0.0 | Liberia | 1.795822 |
| 52488 | 15.0 | 58.0 | Burkina Faso | 1.795822 |
| 52489 | 14000.0 | 16000.0 | Philippines | 1.608296 |

| | user_id | weight-gp_pre_income | weight+gp_pre_income |
|-------|---------------------|----------------------|----------------------|
| 0 | gp_6372662088826880 | -4.999732e+04 | 5.000268e+04 |
| 1 | gp_5900473574883328 | -2.980028e+02 | 3.019972e+02 |
| 2 | gp_4813642242981888 | -3.198101e+03 | 3.201899e+03 |
| 3 | gp_4703669741420544 | -1.030831e+01 | 1.369169e+01 |
| 4 | gp_4762741153988608 | -2.999998e+06 | 3.000002e+06 |
| ... | | ... | ... |
| 52485 | gp_5444968770699264 | -8.977393e+02 | 9.022607e+02 |
| 52486 | gp_5482738571804672 | -1.199998e+06 | 1.200002e+06 |
| 52487 | gp_5658409882091520 | 1.795822e+00 | 1.795822e+00 |
| 52488 | gp_5003001999917056 | -1.320418e+01 | 1.679582e+01 |
| 52489 | gp_5388715507843072 | -1.399839e+04 | 1.400161e+04 |

[52490 rows x 50 columns]

[illegible]

```
(1000, 48)
(1000, 48)
(1000, 48)
(1000, 48)
(1000, 48)
(1000, 48)
(1000, 48)
(1000, 48)
(490, 48)
```