Zadanie 1

In [2]:
```python
import pandas as pd
import numpy as np
df = pd.read_csv("data.csv")
print(df.head())
print(df.info())
print(df.describe())
```

```
        observation_id               submitted_time  gender  \
0  wmn_4503683847159808  2020-07-09 23:19:01.982 UTC  Female
1  wmn_4503772699295744  2020-07-09 21:22:15.864 UTC  Female
2  wmn_4504010469146624  2020-07-10 05:09:07.359 UTC  Female
3  wmn_4504035500752896   2020-07-11 16:59:49.85 UTC  Female
4  wmn_4504181395423232  2020-07-11 18:43:35.954 UTC  Female

                 age                         geography  \
0  26 to 35 years old  City center or metropolitan area
1  16 to 25 years old                             Rural
2  16 to 25 years old                             Rural
3  16 to 25 years old              Suburban/Peri-urban
4  26 to 35 years old              Suburban/Peri-urban

                               financial_situation              education  \
0         I cannot afford enough food for my family  College or university
1         I cannot afford enough food for my family  Secondary/high school
2               I can afford food, but nothing else  College or university
3  I can afford food and regular expenses, and bu...  College or university
4  I can afford food and regular expenses, and bu...  College or university

    employment_status   ethnicity      religion  ... wmn_pre_safe_place  \
0          Unemployed     Mestizo   Catholicism  ...                NaN
1             Student     Tagalog        Muslim  ...                NaN
2             Student  Hiligaynon  Christianity  ...                NaN
3          Unemployed        Thai      Buddhism  ...                NaN
4  Employed full-time     African  Christianity  ...                NaN

  wmn_post_safe_place wmn_safe_place_no_access wmn_safe_place_no_access_why  \
0                 NaN                      NaN                          NaN
1                 NaN                      NaN                          NaN
2                 NaN                      NaN                          NaN
3                 NaN                      NaN                          NaN
4                 NaN                      NaN                          NaN

  wmn_pre_help wmn_post_help wmn_post_no_help wmn_no_help_why  \
0          NaN           NaN              NaN             NaN
1          NaN           NaN              NaN             NaN
2          NaN           NaN              NaN             NaN
3          NaN           NaN              NaN             NaN
4          NaN           NaN              NaN             NaN

                     country               user_id
0                    Ecuador  wmn_5900473574883328
1                Philippines  wmn_5702261783658496
2                Philippines  wmn_5652767014387712
3                   Thailand  wmn_6411372690669568
4  United Republic of Tanzania  wmn_6215734184378368

[5 rows x 46 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12354 entries, 0 to 12353
Data columns (total 46 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   observation_id                 12354 non-null  object
 1   submitted_time                 12354 non-null  object
 2   gender                         12354 non-null  object
 3   age                            12354 non-null  object
 4   geography                      12354 non-null  object
 5   financial_situation            12354 non-null  object
 6   education                      12354 non-null  object
 7   employment_status              12354 non-null  object
 8   ethnicity                      12354 non-null  object
 9   religion                       12354 non-null  object
 10  wmn_hh                         12354 non-null  object
 11  wmn_pregnancy_desire           12354 non-null  object
 12  wmn_pregnancy_change           11351 non-null  object
 13  wmn_pregnancy_change_how       3030 non-null   object
 14  wmn_con                        12354 non-null  object
 15  wmn_con_type                   5426 non-null   object
 16  wmn_pre_con_access_difficulty  3460 non-null   object
 17  wmn_pre_missed_dose_pills      186 non-null    object
 18  wmn_pre_con_needed             3185 non-null   object
 19  wmn_pre_con_accessed           1825 non-null   object
 20  wmn_pre_injectable_missed      94 non-null     float64
 21  wmn_pre_iud_missed             75 non-null     float64
 22  wmn_pre_con_missed_why         437 non-null    object
 23  wmn_pre_con_missed_why_other   34 non-null     object
 24  wmn_post_con_access_difficulty 3460 non-null   object
 25  wmn_post_missed_dose_pills     184 non-null    object
 26  wmn_post_con_needed            3133 non-null   object
 27  wmn_post_con_accessed          1825 non-null   object
 28  wmn_post_injectable_missed     110 non-null    float64
 29  wmn_post_iud_missed            60 non-null     float64
 30  wmn_post_con_missed_why        443 non-null    object
 31  wmn_post_con_missed_why_other  23 non-null     object
 32  wmn_alone                      12354 non-null  object
 33  wmn_how_safe                   5282 non-null   object
 34  wmn_safe_change                5282 non-null   object
 35  wmn_safe_place                 5282 non-null   object
 36  wmn_pre_safe_place             2795 non-null   object
 37  wmn_post_safe_place            2795 non-null   object
 38  wmn_safe_place_no_access       2795 non-null   object
 39  wmn_safe_place_no_access_why   849 non-null    object
 40  wmn_pre_help                   5282 non-null   object
 41  wmn_post_help                  5282 non-null   object
 42  wmn_post_no_help               5282 non-null   object
 43  wmn_no_help_why                848 non-null    object
 44  country                        12354 non-null  object
 45  user_id                        12354 non-null  object
dtypes: float64(4), object(42)
memory usage: 4.3+ MB
None
       wmn_pre_injectable_missed  wmn_pre_iud_missed  \
count                  94.000000        7.500000e+01
mean                  214.765957        1.033763e+07
std                  2062.634850        8.952644e+07
min                     0.000000        0.000000e+00
25%                     1.000000        1.000000e+00
50%                     2.000000        2.000000e+00
75%                     2.750000        3.000000e+00
max                 20000.000000        7.753217e+08

       wmn_post_injectable_missed  wmn_post_iud_missed
count                  110.000000         6.000000e+01
mean                   232.845455         1.292203e+07
std                   2418.242269         1.000936e+08
min                      0.000000         0.000000e+00
25%                      1.000000         1.000000e+00
50%                      2.000000         2.000000e+00
75%                      3.000000         3.000000e+00
max                  25365.000000         7.753217e+08
```

Zadanie 2: Obliczanie podstawowych statystyk Oblicz podstawowe statystyki opisowe dla wybranych kolumn, aby zrozumie´c rozk lad danych. Co musisz zrobi´c: • Oblicz ´srednia¸ dla wybranej kolumny, • Oblicz mediane¸ i odchylenie standardowe dla innej kolumny.

Pomijanie wartoście nienumerycznych

In [5]:
```python
df['wmn_hh'] = pd.to_numeric(df['wmn_hh'], errors='coerce')
```

Zadanie 2

In [7]:
```python
#print(df['wmn_hh'])
#mean_wmn_hh = df['wmn_hh'].mean()
#print(mean_wmn_hh)
```

Zadanie 3

In [9]:
```python
missing_values = df.isnull().sum()
print("Brakujace wartosci w kazdej kolumnie:")
print(missing_values)

# Oblicz średnią dla kolumny 'wmn_hh', ignorując brakujące wartości
mean_wmn_hh = df['wmn_hh'].mean()

# Wypełnij brakujące wartości średnią
df['wmn_hh'].fillna(mean_wmn_hh, inplace=True)

print(df['wmn_hh'])

df.dropna(subset=['wmn_pregnancy_change'], inplace=True)
```

```
Brakujace wartosci w kazdej kolumnie:
observation_id                      0
submitted_time                      0
gender                              0
age                                 0
geography                           0
financial_situation                 0
education                           0
employment_status                   0
ethnicity                           0
religion                            0
wmn_hh                             67
wmn_pregnancy_desire                0
wmn_pregnancy_change             1003
wmn_pregnancy_change_how         9324
wmn_con                             0
wmn_con_type                     6928
wmn_pre_con_access_difficulty    8894
wmn_pre_missed_dose_pills       12168
wmn_pre_con_needed               9169
wmn_pre_con_accessed            10529
wmn_pre_injectable_missed       12260
wmn_pre_iud_missed              12279
wmn_pre_con_missed_why          11917
wmn_pre_con_missed_why_other    12320
wmn_post_con_access_difficulty   8894
wmn_post_missed_dose_pills      12170
wmn_post_con_needed              9221
wmn_post_con_accessed           10529
wmn_post_injectable_missed      12244
wmn_post_iud_missed             12294
wmn_post_con_missed_why         11911
wmn_post_con_missed_why_other   12331
wmn_alone                           0
wmn_how_safe                     7072
wmn_safe_change                  7072
wmn_safe_place                   7072
wmn_pre_safe_place               9559
wmn_post_safe_place              9559
wmn_safe_place_no_access         9559
wmn_safe_place_no_access_why    11505
wmn_pre_help                     7072
wmn_post_help                    7072
wmn_post_no_help                 7072
wmn_no_help_why                 11506
country                             0
user_id                             0
dtype: int64
0         3.0
1        13.0
2         5.0
3         7.0
4         3.0
         ...
12349     7.0
12350     3.0
12351     4.0
12352     4.0
12353     3.0
Name: wmn_hh, Length: 12354, dtype: float64
C:\Users\Hubert\AppData\Local\Temp\ipykernel_18780\692307689.py:9: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.


  df['wmn_hh'].fillna(mean_wmn_hh, inplace=True)
```

Zadanie 4

In [11]:
```python
# Oblicz IQR
Q1 = df['wmn_hh'].quantile(0.25)
Q3 = df['wmn_hh'].quantile(0.75)
IQR = Q3 - Q1

# Zidentyfikuj wartoci odstajce
outliers = df[(df['wmn_hh'] < (Q1 - 1.5 * IQR)) | (df['wmn_hh'] > (Q3 + 1.5 * IQR))]
print("Wartoci odstajace:")
print(outliers)
```

```
Wartoci odstajace:
             observation_id               submitted_time  gender  \
1      wmn_4503772699295744  2020-07-09 21:22:15.864 UTC  Female
6      wmn_4504322055602176  2020-07-09 20:43:11.055 UTC  Female
8      wmn_4504469091123200  2020-07-16 16:03:44.066 UTC  Female
10     wmn_4504764873441280  2020-07-23 00:29:41.766 UTC  Female
41     wmn_4511777313521664  2020-07-12 14:06:30.009 UTC  Female
...                     ...                          ...     ...
12224  wmn_6732835697459200  2020-07-11 03:23:37.429 UTC  Female
12253  wmn_6736732706045952  2020-07-10 03:03:14.392 UTC  Female
12279  wmn_6740038290505728  2020-07-10 00:08:04.908 UTC  Female
12302  wmn_6744322450128896   2020-07-14 14:47:04.93 UTC  Female
12308  wmn_6745767575552000  2020-07-14 12:47:51.757 UTC  Female

                      age                         geography  \
1      16 to 25 years old                             Rural
6      16 to 25 years old  City center or metropolitan area
8      36 to 45 years old  City center or metropolitan area
10     16 to 25 years old                             Rural
41     16 to 25 years old              Suburban/Peri-urban
...                   ...                               ...
12224  26 to 35 years old              Suburban/Peri-urban
12253  26 to 35 years old  City center or metropolitan area
12279  26 to 35 years old                             Rural
12302  26 to 35 years old                             Rural
12308  26 to 35 years old  City center or metropolitan area

                                     financial_situation  \
1              I cannot afford enough food for my family
6      I can afford food and regular expenses, and bu...
8      I can afford food and regular expenses, but no...
10                   I can afford food, but nothing else
41             I cannot afford enough food for my family
...                                                  ...
12224  I can afford food and regular expenses, but no...
12253                I can afford food, but nothing else
12279  I can comfortably afford food, clothes, and fu...
12302          I cannot afford enough food for my family
12308          I cannot afford enough food for my family

                   education   employment_status ethnicity        religion  \
1      Secondary/high school             Student   Tagalog          Muslim
6      College or university  Employed full-time   Tagalog    Christianity
8      College or university  Employed part-time     Serer          Muslim
10          Technical school  Employed full-time   Pashtun    Muslim (Shia)
41     College or university             Student   Mestizo     Catholicism
...                      ...                 ...       ...             ...
12224       Technical school  Employed part-time     White     Catholicism
12253          Post graduate       Self-employed    Bisaya    Christianity
12279  College or university  Employed full-time    Bisaya    Christianity
12302  College or university          Unemployed   Mestizo  Evangelicalism
12308  Secondary/high school  Employed full-time      Java          Muslim

       ... wmn_pre_safe_place wmn_post_safe_place wmn_safe_place_no_access  \
1      ...                NaN                 NaN                      NaN
6      ...                NaN                 NaN                      NaN
8      ...                NaN                 NaN                      NaN
10     ...                NaN                 NaN                      NaN
41     ...        Once a week         Once a week                      Yes
...    ...                ...                 ...                      ...
12224  ...                NaN                 NaN                      NaN
12253  ...                NaN                 NaN                      NaN
12279  ...                NaN                 NaN                      NaN
12302  ...                NaN                 NaN                      NaN
12308  ...                NaN                 NaN                      NaN

                            wmn_safe_place_no_access_why wmn_pre_help  \
1                                                    NaN          NaN
6                                                    NaN          NaN
8                                                    NaN          NaN
10                                                   NaN          NaN
41     Place was closed or unavailable for reason oth...           No
...                                                  ...          ...
12224                                                NaN           No
12253                                                NaN          NaN
12279                                                NaN          NaN
12302                                                NaN          NaN
12308                                                NaN          NaN

      wmn_post_help wmn_post_no_help wmn_no_help_why  \
1               NaN              NaN             NaN
6               NaN              NaN             NaN
8               NaN              NaN             NaN
10              NaN              NaN             NaN
41               No               No             NaN
...             ...              ...             ...
12224            No               No             NaN
12253           NaN              NaN             NaN
12279           NaN              NaN             NaN
12302           NaN              NaN             NaN
12308           NaN              NaN             NaN

                                country               user_id
1                           Philippines  wmn_5702261783658496
6                           Philippines  wmn_5846766880817152
8                               Senegal  wmn_6660892420997120
10                          Afghanistan  wmn_5962315779538944
41     Venezuela (Bolivarian Republic of)  wmn_6060988254388224
...                                 ...                   ...
12224                          Colombia  wmn_6146135279534080
12253                       Philippines  wmn_6151534389755904
12279                       Philippines  wmn_4847756230524928
12302                           Ecuador  wmn_5656815158427648
12308                         Indonesia  wmn_5298186865606656

[620 rows x 46 columns]
```

In [13]:
```python
df['random_values'] = np.random.randint(1, 16, size=len(df))

print(df[['wmn_hh', 'random_values']])
```

```
       wmn_hh  random_values
0         3.0              7
1        13.0             14
2         5.0             13
3         7.0             10
4         3.0              3
...       ...            ...
12349     7.0              2
12350     3.0              4
12351     4.0              8
12352     4.0             14
12353     3.0             12

[11351 rows x 2 columns]
```

Zadanie 5

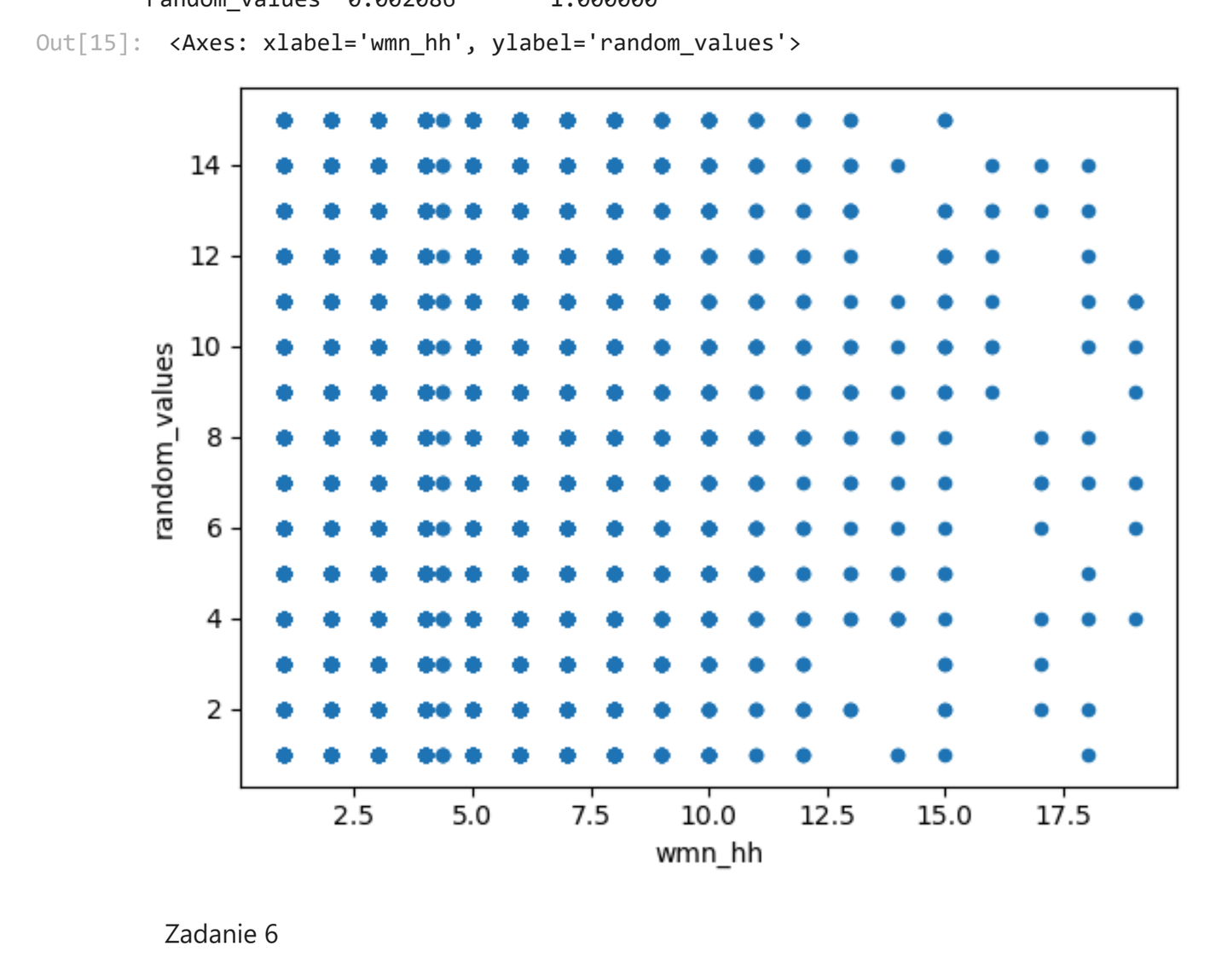
In [15]:
```python
new_df = df[['wmn_hh', 'random_values']].copy()

print(new_df)
# Oblicz macierz korelacji
correlation_matrix = new_df.corr()
print("Macierz korelacji:")
print(correlation_matrix)
# Wykonaj wykres rozrzutu
df.plot.scatter(x='wmn_hh', y='random_values')
```

```
       wmn_hh  random_values
0         3.0              7
1        13.0             14
2         5.0             13
3         7.0             10
4         3.0              3
...       ...            ...
12349     7.0              2
12350     3.0              4
12351     4.0              8
12352     4.0             14
12353     3.0             12

[11351 rows x 2 columns]
Macierz korelacji:
                 wmn_hh  random_values
wmn_hh         1.000000       0.002086
random_values  0.002086       1.000000
```

Out[15]: <Axes: xlabel='wmn_hh', ylabel='random_values'>



Zadanie 6

In [17]:
```python
# Dodaj now kolumn ’ dochd na osob ’
df['wmn_hh_add_5'] = df['wmn_hh'] + 5
# Grupuj dane wedug kolumny ’region’ i oblicz redni dochd
grouped = df.groupby('country')['wmn_hh'].mean()
print(" Srednie wmn_hh w krajach:")
print(grouped)

# Posortuj dane wedug kolumny ’dochd’
df_sorted = df.sort_values(by='wmn_hh', ascending=False)
print("Dane posortowane wedug wmn_hh:")
print(df_sorted.head())
```

```
 Srednie wmn_hh w krajach:
country
Afghanistan                           4.857432
Albania                               4.318397
Algeria                               4.892541
Argentina                             3.774818
Bahrain                               5.333333
                                        ...
Venezuela (Bolivarian Republic of)    4.142104
Viet Nam                              4.365854
Yemen                                 5.516393
Zambia                                6.000000
Zimbabwe                              4.930233
Name: wmn_hh, Length: 76, dtype: float64
Dane posortowane wedug wmn_hh:
            observation_id               submitted_time  gender  \
6086  wmn_5615059754811392  2020-07-16 21:48:51.644 UTC  Female
5435  wmn_5493846818684928  2020-07-10 19:45:01.023 UTC  Female
5016  wmn_5414602256154624  2020-07-09 22:34:15.536 UTC  Female
5622  wmn_5527501175783424  2020-07-15 10:33:36.001 UTC  Female
3858  wmn_5203932164128768  2020-07-11 10:39:59.673 UTC  Female

                     age                         geography  \
6086  16 to 25 years old                             Rural
5435  36 to 45 years old              Suburban/Peri-urban
5016  26 to 35 years old                             Rural
5622  16 to 25 years old  City center or metropolitan area
3858  26 to 35 years old              Suburban/Peri-urban

                                    financial_situation  \
6086                I can afford food, but nothing else
5435  I can afford food and regular expenses, but no...
5016  I can afford food and regular expenses, but no...
5622                I can afford food, but nothing else
3858  I can afford food and regular expenses, and bu...

                  education   employment_status             ethnicity  \
6086  College or university             Student              Sarakole
5435       Technical school  Employed full-time                 White
5016  College or university  Employed full-time                Bisaya
5622  College or university             Student  Prefer not to answer
3858  College or university  Employed part-time                 Tajik

                  religion  ... wmn_safe_place_no_access  \
6086                Muslim  ...                      Yes
5435         Protestantism  ...                      NaN
5016          Christianity  ...                      Yes
5622  Prefer Not To Answer  ...                      NaN
3858        Muslim (Sunni)  ...                       No

                          wmn_safe_place_no_access_why wmn_pre_help  \
6086                                 No transportation           No
5435                                               NaN          NaN
5016  Place was closed or unavailable for reason oth...           No
5622                                               NaN          NaN
3858                                               NaN          Yes

     wmn_post_help wmn_post_no_help wmn_no_help_why      country  \
6086            No               No             NaN         Mali
5435           NaN              NaN             NaN       Brazil
5016           Yes               No             NaN  Philippines
5622           NaN              NaN             NaN  Afghanistan
3858           Yes               No             NaN  Afghanistan

                   user_id random_values wmn_hh_add_5
6086  wmn_4523262962040832             9         24.0
5435  wmn_5995159497474048             6         24.0
5016  wmn_5778690861694976            10         24.0
5622  wmn_5865869400080384            11         24.0
3858  wmn_6547829764128768             7         24.0

[5 rows x 48 columns]
```

In [ ]: