```
#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]
```

```
#wyświetl pierwsze 5 wierszy
print(data_frame.head())
```

```
        observation_id               submitted_time  gender  \
0  gp_4503617949401088   2020-07-07 14:48:29.83 UTC    Male
1  gp_4503631639609344  2020-07-09 13:22:37.107 UTC  Female
2  gp_4503700758593536  2020-07-04 18:53:36.471 UTC    Male
3  gp_4503737805832192  2020-07-12 17:58:20.798 UTC    Male
4  gp_4503819343101952  2020-07-06 00:20:22.983 UTC    Male

                  age                          geography  \
0           Under 16                Suburban/Peri-urban
1  26 to 35 years old  City center or metropolitan area
2  36 to 45 years old  City center or metropolitan area
3  26 to 35 years old                             Rural
4  26 to 35 years old                Suburban/Peri-urban

                                  financial_situation
education  \
0  I can afford food and regular expenses, but no...  Secondary/high
school
1          I cannot afford enough food for my family  College or
university
2  I can comfortably afford food, clothes, and fu...          Primary
school
3  I can afford food and regular expenses, and bu...       Technical
school
4  I can afford food and regular expenses, and bu...       Technical
school

    employment_status           ethnicity      religion  ...  \
0  Employed full-time              Ankole  Christianity  ...
1          Unemployed             Mestizo   Catholicism  ...
```

```
2  Employed full-time  Non-hispanic White   Agnosticism  ...
3            Student             Mestizo  Christianity  ...
4  Employed full-time             Mestizo   Catholicism  ...

  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


                  gp_unemployment_why gp_unemployment_why_other
gp_pre_income  \
0                                 NaN                       NaN
50000.0
1                                 NaN                       NaN
300.0
2                                 NaN                       NaN
3200.0
3  Laid off while business continues                        NaN
12.0
4                                 NaN                       NaN
3000000.0

  gp_post_income                    country    weight
user_id
0       999497.0                     Uganda  2.675598
gp_6372662088826880
1          200.0                    Ecuador  1.997161
gp_5900473574883328
2         3200.0  United States of America  1.898763
gp_4813642242981888
3           12.0                   Colombia  1.691686
gp_4703669741420544
4       400000.0                   Colombia  1.691686
gp_4762741153988608

[5 rows x 48 columns]
```

```python
#sprawdź podstawowe informacje o danych
print(data_frame.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52490 entries, 0 to 52489
Data columns (total 48 columns):
```

```
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   observation_id                    52490 non-null  object
 1   submitted_time                    52490 non-null  object
 2   gender                            52469 non-null  object
 3   age                               52490 non-null  object
 4   geography                         52490 non-null  object
 5   financial_situation               52490 non-null  object
 6   education                         52490 non-null  object
 7   employment_status                 52490 non-null  object
 8   ethnicity                         52490 non-null  object
 9   religion                          52490 non-null  object
 10  gp_hh                             52478 non-null  object
 11  gp_pre_provider_need              52490 non-null  object
 12  gp_pre_provider_condition         21777 non-null  object
 13  gp_pre_provider_condition_other   2982 non-null   object
 14  gp_pre_provider_visit             21777 non-null  object
 15  gp_pre_provider_where             9972 non-null   object
 16  gp_pre_provider_where_other       803 non-null    object
 17  gp_pre_provider_num_visit         19037 non-null  object
 18  gp_pre_provider_why               8363 non-null   object
 19  gp_pre_provider_why_other         398 non-null    object
 20  gp_post_provider_need             52490 non-null  object
 21  gp_post_provider_condition        16598 non-null  object
 22  gp_post_provider_condition_other  1854 non-null   object
 23  gp_post_provider_visit            16598 non-null  object
 24  gp_post_provider_where            9972 non-null   object
 25  gp_post_provider_where_other      327 non-null    object
 26  gp_post_provider_num_visit        13824 non-null  object
 27  gp_post_provider_why              6847 non-null   object
 28  gp_post_provider_why_other        173 non-null    object
 29  gp_medication                     52490 non-null  object
 30  gp_medication_condition           24293 non-null  object
 31  gp_pre_miss_dose                  24293 non-null  object
 32  gp_pre_num_miss_dose              7096 non-null   object
 33  gp_pre_miss_dose_why              7096 non-null   object
 34  gp_pre_miss_dose_why_other        174 non-null    object
 35  gp_post_miss_dose                 24439 non-null  object
 36  gp_post_num_miss_dose             8875 non-null   object
 37  gp_post_miss_dose_why             8875 non-null   object
 38  gp_post_miss_dose_why_other       182 non-null    object
 39  gp_post_labor_force               52490 non-null  object
 40  gp_pre_labor_force                52490 non-null  object
 41  gp_unemployment_why               8691 non-null   object
 42  gp_unemployment_why_other         736 non-null    object
 43  gp_pre_income                     52490 non-null  float64
 44  gp_post_income                    52490 non-null  float64
 45  country                           52490 non-null  object
 46  weight                            52490 non-null  float64
```

```
 47  user_id                          52490 non-null  object
dtypes: float64(3), object(45)
memory usage: 19.2+ MB
None
```

```
# Wywietl podstawowe statystyki opisowe
print(data_frame.describe())
```

```
       gp_pre_income  gp_post_income        weight
count   5.249000e+04    5.249000e+04  52490.000000
mean    1.905125e+56    1.905125e+59      1.795820
std     4.364774e+58    4.364774e+61      0.384507
min     0.000000e+00    0.000000e+00      1.000000
25%     1.500000e+02    7.000000e+01      1.544666
50%     3.425000e+03    2.000000e+03      1.730488
75%     2.500000e+04    2.000000e+04      1.908871
max     1.000000e+61    1.000000e+64      6.411420
```

```
#Oblicz średni dla kolumny
mean_child_count = data_frame["gp_pre_income"].mean()
print(f"Średnia liczba dzieci: {mean_child_count}")
```

```
Średnia liczba dzieci: 1.9051247856734615e+56
```

```
#Oblicz median dla kolumny
median_child_count = data_frame["gp_pre_income"].median()
print(f"Mediana liczby dzieci: {median_child_count}")
```

```
Mediana liczby dzieci: 3425.0
```

```
#Oblicz odchylenie standardowe dla kolumny
std_child_count = data_frame["gp_pre_income"].std()
print(f"Odchylenie standardowe liczby dzieci: {std_child_count}")
```

```
Odchylenie standardowe liczby dzieci: 4.364773517232549e+58
```

```
#Sprawdź brakujące wartości
missing_values = data_frame.isnull().sum()
print("Brakujce wartoci w kadej kolumnie:")
print(missing_values)
```

```
Brakujce wartoci w kadej kolumnie:
observation_id                    0
submitted_time                    0
gender                           21
age                               0
geography                         0
financial_situation               0
education                         0
employment_status                 0
ethnicity                         0
```

```
religion                                0
gp_hh                                  12
gp_pre_provider_need                    0
gp_pre_provider_condition           30713
gp_pre_provider_condition_other     49508
gp_pre_provider_visit               30713
gp_pre_provider_where               42518
gp_pre_provider_where_other         51687
gp_pre_provider_num_visit           33453
gp_pre_provider_why                 44127
gp_pre_provider_why_other           52092
gp_post_provider_need                   0
gp_post_provider_condition          35892
gp_post_provider_condition_other    50636
gp_post_provider_visit              35892
gp_post_provider_where              42518
gp_post_provider_where_other        52163
gp_post_provider_num_visit          38666
gp_post_provider_why                45643
gp_post_provider_why_other          52317
gp_medication                           0
gp_medication_condition             28197
gp_pre_miss_dose                    28197
gp_pre_num_miss_dose                45394
gp_pre_miss_dose_why                45394
gp_pre_miss_dose_why_other          52316
gp_post_miss_dose                   28051
gp_post_num_miss_dose               43615
gp_post_miss_dose_why               43615
gp_post_miss_dose_why_other         52308
gp_post_labor_force                     0
gp_pre_labor_force                      0
gp_unemployment_why                 43799
gp_unemployment_why_other           51754
gp_pre_income                           0
gp_post_income                          0
country                                 0
weight                                  0
user_id                                 0
dtype: int64
```

```
#Uzupełnij brakujce wartoci redni w kolumnie liczba dzieci
data_frame["gp_pre_income"].fillna(data_frame["gp_pre_income"].mean(),
inplace=True)
```

```
C:\Users\Szymon\AppData\Local\Temp\ipykernel_14900\1050778849.py:2:
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.
```

```
data_frame["gp_pre_income"].fillna(data_frame["gp_pre_income"].mean(),
inplace=True)

#Usuń wiersze, gdzie brakuje danych w kolumnie
data_frame.dropna(subset=["gp_pre_income"], inplace=True)

#Oblicz IQR
Q1 = data_frame["gp_pre_income"].quantile(0.25)
Q3 = data_frame["gp_pre_income"].quantile(0.75)
IQR = Q3 - Q1
print(IQR)
```

```
24850.0
```

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

```
Wartoci odstajce :
            observation_id                submitted_time  gender  \
4      gp_4503819343101952  2020-07-06 00:20:22.983 UTC    Male
5      gp_4503868601008128  2020-07-11 17:45:09.631 UTC    Male
18     gp_4504344939724800  2020-07-04 17:45:35.164 UTC    Male
19     gp_4504387688071168  2020-07-12 16:14:12.613 UTC    Male
26     gp_4504536465276928  2020-07-03 18:11:18.738 UTC  Female
...                    ...                          ...     ...
52466  gp_6754360325570560  2020-06-30 20:37:40.345 UTC  Female
52468  gp_6754417841012736  2020-07-03 20:25:54.494 UTC    Male
52474  gp_6754574808645632  2020-07-10 18:04:16.896 UTC    Male
52482  gp_6754988065030144  2020-07-03 03:50:43.956 UTC  Female
52486  gp_6755213279232000  2020-07-02 07:07:44.269 UTC    Male

                       age                          geography  \
4      26 to 35 years old                 Suburban/Peri-urban
5      26 to 35 years old                 Suburban/Peri-urban
18     26 to 35 years old  City center or metropolitan area
19     16 to 25 years old  City center or metropolitan area
26     26 to 35 years old                 Suburban/Peri-urban
...                    ...                                ...
52466  16 to 25 years old  City center or metropolitan area
```

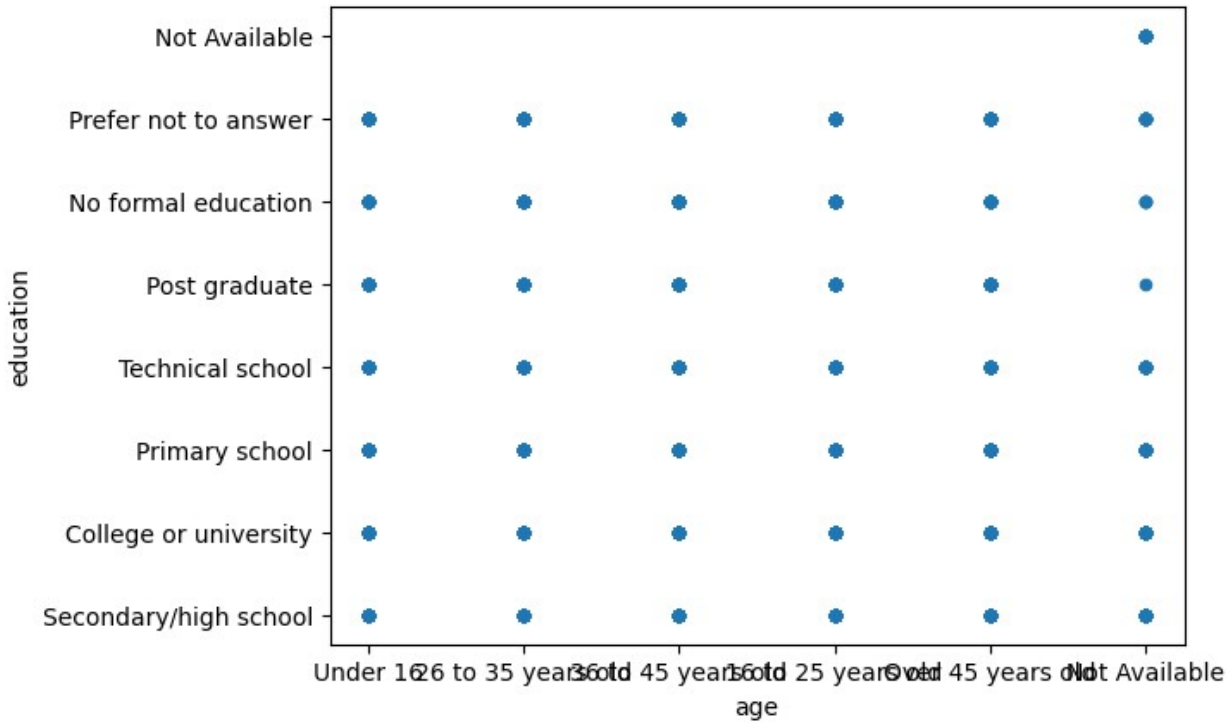
```
52468  26 to 35 years old  City center or metropolitan area
52474  16 to 25 years old                            Rural
52482  16 to 25 years old                            Rural
52486  26 to 35 years old  City center or metropolitan area

                                      financial_situation  \
4      I can afford food and regular expenses, and bu...
5      I can afford food and regular expenses, but no...
18     I can afford food and regular expenses, but no...
19     I can afford food and regular expenses, and bu...
26                  I can afford food, but nothing else
...                                                   ...
52466               I can afford food, but nothing else
52468  I can afford food and regular expenses, but no...
52474         I cannot afford enough food for my family
52482  I can afford food and regular expenses, but no...
52486  I can afford food and regular expenses, but no...

                    education           employment_status
ethnicity  \
4           Technical school          Employed full-time
Mestizo
5      College or university          Employed full-time
Baganda
18     Secondary/high school          Employed part-time  None of the
above
19     College or university          Employed full-time  None of the
above
26             Post graduate          Employed part-time
Soninke
...                      ...                         ...
...
52466  College or university                     Student
Mestizo
52468  College or university  Student and work part-time     Northern
Mande
52474  Secondary/high school                     Student
Songhai
52482  Secondary/high school                  Unemployed
Sunda
52486  College or university          Employed full-time
Arab

             religion  ... gp_post_miss_dose_why_other
gp_post_labor_force  \
4         Catholicism  ...                         NaN
Yes
5               Muslim  ...                         NaN
Yes
18              Muslim  ...                         NaN
```

```
Yes
19               Other  ...                           NaN
No
26              Muslim  ...                           NaN
Yes
...                ...  ...                           ...
...
52466    Christianity  ...                           NaN
Yes
52468          Muslim  ...                           NaN
Yes
52474          Muslim  ...                           NaN
No
52482          Muslim  ...                           NaN
No
52486  Muslim (Sunni)  ...                           NaN
Yes

      gp_pre_labor_force gp_unemployment_why gp_unemployment_why_other
\
4                     Yes                  NaN                         NaN

5                     Yes                  NaN                         NaN

18                     No                  NaN                         NaN

19                    Yes                Other                        XXXX

26                    Yes                  NaN                         NaN

...                   ...                  ...                         ...

52466                 Yes                  NaN                         NaN

52468                  No                  NaN                         NaN

52474                  No                  NaN                         NaN

52482                  No                  NaN                         NaN

52486                 Yes                  NaN                         NaN


      gp_pre_income gp_post_income                          country
\
4         3000000.0       400000.0                         Colombia

5          200000.0       100000.0                           Uganda

18        1500000.0      1500000.0                        Indonesia
```

```
19        5000000.0      5500000.0                                Indonesia

26         100000.0       120000.0                                  Senegal

...             ...            ...                                      ...

52466     1000000.0       200000.0  Venezuela (Bolivarian Republic of)

52468      100000.0       100000.0                             Ivory Coast

52474      150000.0       150000.0                                    Mali

52482     5000000.0      5000000.0                                Indonesia

52486     1200000.0       120000.0                                  Algeria


          weight               user_id
4       1.691686  gp_4762741153988608
5       1.848512  gp_6657644037406720
18      1.502260  gp_5716018716475392
19      1.445650  gp_6160397591904256
26      2.487237  gp_4950520168185856
...          ...                  ...
52466   1.672533  gp_6713107022086144
52468   1.000000  gp_4843290176520192
52474   1.544468  gp_6271730127732736
52482   1.575638  gp_5113627587903488
52486   1.896411  gp_5482738571804672

[9347 rows x 48 columns]
```

```
#Wykonaj wykres rozrzutu
data_frame.plot.scatter(x="age", y="education")
```

```
<Axes: xlabel='age', ylabel='education'>
```




```
#Dodaj nową kolumnę
data_frame["byleco"] = data_frame["u2_child_count"] /
data_frame["u2_child_count"]

#Grupuj dane wedug kolumny ’region’ i oblicz średnią
grouped = data_frame.groupby("gender")["u2_child_count"].mean()
print("Coś:")
print(grouped)
```

```
Coś:
gender
Female                   1.415692
Male                     1.542283
Not Available            1.000000
Prefer not to answer     1.740000
Name: u2_child_count, dtype: float64
```

```
# Posortuj dane wedug kolumny
df_sorted = data_frame.sort_values(by="age", ascending=False)
print("Dane posortowane według wieku:")
print(df_sorted.head())
```

```
Dane posortowane według wieku:
          observation_id               submitted_time  gender
age  \
1290  u2_4925422526791680  2020-07-12 01:51:11.179 UTC  Female  Under
16
2513  u2_5293163465146368  2020-06-30 18:43:51.855 UTC    Male  Under
```

```
16
922   u2_4806821165662208   2020-07-01 17:56:20.74 UTC  Female  Under
16
5389  u2_6185680486268928  2020-07-03 12:48:16.841 UTC    Male  Under
16
5392  u2_6187233382236160  2020-07-02 02:24:18.759 UTC    Male  Under
16

                 geography
financial_situation  \
1290                Rural  I can afford food and regular expenses, but
no...
2513                Rural          I cannot afford enough food for my
family
922                 Rural                 I can afford food, but
nothing else
5389                Rural          I cannot afford enough food for my
family
5392  Suburban/Peri-urban          I cannot afford enough food for my
family

                  education   employment_status
ethnicity  \
1290  Secondary/high school             Student
Thai
2513  Secondary/high school  Employed full-time  Black or African
American
922   Secondary/high school             Student
Pashtun
5389  Secondary/high school             Student
Khmer
5392  Secondary/high school             Student
Khmer

            religion  ... u2_post_provider_need  \
1290        Buddhism  ...                     No
2513    Catholicism  ...                     No
922   Muslim (Sunni)  ...                     No
5389        Buddhism  ...                    Yes
5392        Buddhism  ...                    Yes

                       u2_post_condition  u2_post_provider_visit  \
1290                                 NaN                     NaN
2513                                 NaN                     NaN
922                                  NaN                     NaN
5389  Birth defect or congenital problem                     Yes
5392  Birth defect or congenital problem                     Yes

     u2_post_provider_where  u2_post_provider_num
u2_post_no_provider_why  \
```

```
1290                      NaN                      NaN
NaN
2513                      NaN                      NaN
NaN
922                       NaN                      NaN
NaN
5389        Health facility                        1
NaN
5392        Health facility                        2
NaN

      u2_post_no_provider_why_other         country
user_id byleco
1290                                NaN       Thailand
u2_6480293365284864    1.0
2513                                NaN  United States
u2_4527152061480960    1.0
922                                 NaN    Afghanistan
u2_5818715109326848    1.0
5389                                NaN       Cambodia
u2_5675130892320768    1.0
5392                                NaN       Cambodia
u2_5348126966939648    1.0

[5 rows x 48 columns]
```