```python
#importowanie biblioteki pandas
import pandas as pd

#tworzenie ramki danych ze słownika
data = pd.read_csv('IHME_PREM_CH_HEALTH_2020_Y2021M10D11.csv',
encoding='utf-8')

data_frame = pd.DataFrame(data)
#wyświetlenie ramki danych
data_frame
```

```
           observation_id                submitted_time  gender  \
0     u2_4503977216704512    2020-07-06 17:46:17.8 UTC    Male
1     u2_4505961390931968  2020-07-07 00:25:56.895 UTC  Female
2     u2_4506421419048960  2020-07-11 07:16:01.196 UTC    Male
3     u2_4506681267716096  2020-07-01 14:32:45.987 UTC    Male
4     u2_4506682207240192  2020-07-01 15:07:48.944 UTC    Male
...                   ...                          ...     ...
7223  u2_6754387638878208   2020-07-01 14:59:06.82 UTC    Male
7224  u2_6754421129347072  2020-07-05 02:06:50.522 UTC    Male
7225  u2_6754616949866496  2020-07-07 06:44:30.974 UTC    Male
7226  u2_6754906729086976  2020-07-02 09:49:21.467 UTC    Male
7227  u2_6754986118873088  2020-07-02 15:45:03.342 UTC    Male

                      age                          geography  \
0     26 to 35 years old  City center or metropolitan area
1     26 to 35 years old              Suburban/Peri-urban
2     26 to 35 years old  City center or metropolitan area
3     26 to 35 years old              Suburban/Peri-urban
4     26 to 35 years old                            Rural
...                   ...                              ...
7223  16 to 25 years old  City center or metropolitan area
7224  16 to 25 years old              Suburban/Peri-urban
7225  26 to 35 years old                            Rural
7226  16 to 25 years old                            Rural
7227  16 to 25 years old                            Rural

                                    financial_situation  \
0     I can afford food and regular expenses, but no...
1             I cannot afford enough food for my family
2     I can afford food and regular expenses, but no...
3     I can afford food and regular expenses, but no...
4                   I can afford food, but nothing else
...                                                 ...
7223          I cannot afford enough food for my family
7224  I can afford food and regular expenses, and bu...
7225                I can afford food, but nothing else
7226          I cannot afford enough food for my family
7227          I cannot afford enough food for my family
```

```
                  education   employment_status      ethnicity  \
0     College or university  Employed full-time  Afro-Jamaican
1          Technical school       Self-employed       Hispanic
2     College or university       Self-employed         Bantou
3     Secondary/high school       Self-employed         Bantou
4     Secondary/high school  Employed part-time        Mestizo
...                     ...                 ...            ...
7223  College or university       Self-employed          Mossi
7224  Secondary/high school  Employed part-time        Baganda
7225       Technical school       Self-employed         Yoruba
7226  Secondary/high school  Employed full-time        Tagalog
7227         Primary school  Employed full-time        Baganda

             religion  ... u2_pre_no_provider_why_other
u2_post_provider_need  \
0     Roman Catholic  ...                          NaN
Yes
1        Catholicism  ...                          NaN
No
2      Protestantism  ...                          NaN
No
3      Protestantism  ...                          NaN
Yes
4              Other  ...                          NaN
Yes
...              ...  ...                          ...
...
7223  Roman Catholic  ...                          NaN
Yes
7224          Muslim  ...                          NaN
No
7225    Christianity  ...                          NaN
No
7226    Christianity  ...                          NaN
No
7227    Christianity  ...                          NaN
Yes

                                  u2_post_condition  \
0              Birth defect or congenital problem
1                                              NaN
2                                              NaN
3                                          Diarrhea
4     Birth defect or congenital problem^Diarrhea
...                                             ...
7223           Birth defect or congenital problem
7224                                           NaN
7225                                           NaN
7226                                           NaN
```

```
7227           Birth defect or congenital problem

                                  u2_post_provider_visit  \
0                                                     No
1                                                    NaN
2                                                    NaN
3                                                     No
4                                                    Yes
...                                                  ...
7223  My child saw a provider during this time, but ...
7224                                                 NaN
7225                                                 NaN
7226                                                 NaN
7227                                                 Yes

          u2_post_provider_where u2_post_provider_num  \
0                            NaN                   NaN
1                            NaN                   NaN
2                            NaN                   NaN
3                            NaN                   NaN
4     At my home^Health facility                     2
...                          ...                   ...
7223                         NaN                     5
7224                         NaN                   NaN
7225                         NaN                   NaN
7226                         NaN                   NaN
7227             Health facility                     1

               u2_post_no_provider_why
u2_post_no_provider_why_other  \
0                Health facility closed                            NaN

1                                   NaN                            NaN

2                                   NaN                            NaN

3     Turned away from health facility                            NaN

4                                   NaN                            NaN

...                                 ...                            ...

7223                Decline to respond                            NaN

7224                                NaN                            NaN

7225                                NaN                            NaN

7226                                NaN                            NaN
```

```
7227                                  NaN                                  NaN


                                country              user_id
0                               Jamaica  u2_4519417576161280
1                         United States  u2_6197364724858880
2     Democratic Republic of the Congo  u2_5202308155834368
3     Democratic Republic of the Congo  u2_6709282991243264
4                             Nicaragua  u2_6190896700194816
...                                 ...                  ...
7223                       Burkina Faso  u2_6532940630851584
7224                             Uganda  u2_5363001946800128
7225                            Nigeria  u2_5779942448562176
7226                        Philippines  u2_6586517350252544
7227                             Uganda  u2_4955652432855040

[7228 rows x 47 columns]
```

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

```
        observation_id               submitted_time  gender  \
0  u2_4503977216704512    2020-07-06 17:46:17.8 UTC    Male
1  u2_4505961390931968  2020-07-07 00:25:56.895 UTC  Female
2  u2_4506421419048960  2020-07-11 07:16:01.196 UTC    Male
3  u2_4506681267716096  2020-07-01 14:32:45.987 UTC    Male
4  u2_4506682207240192  2020-07-01 15:07:48.944 UTC    Male

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

                                  financial_situation
education  \
0  I can afford food and regular expenses, but no...  College or
university
1          I cannot afford enough food for my family       Technical
school
2  I can afford food and regular expenses, but no...  College or
university
3  I can afford food and regular expenses, but no...  Secondary/high
school
4                I can afford food, but nothing else  Secondary/high
school

    employment_status      ethnicity        religion  ...  \
0  Employed full-time  Afro-Jamaican  Roman Catholic  ...
```

```
1       Self-employed       Hispanic     Catholicism  ...
2       Self-employed         Bantou   Protestantism  ...
3       Self-employed         Bantou   Protestantism  ...
4  Employed part-time        Mestizo           Other  ...

  u2_pre_no_provider_why_other  u2_post_provider_need  \
0                          NaN                    Yes
1                          NaN                     No
2                          NaN                     No
3                          NaN                    Yes
4                          NaN                    Yes

                             u2_post_condition u2_post_provider_visit
\
0           Birth defect or congenital problem                     No

1                                          NaN                    NaN

2                                          NaN                    NaN

3                                     Diarrhea                     No

4  Birth defect or congenital problem^Diarrhea                    Yes


       u2_post_provider_where u2_post_provider_num  \
0                         NaN                  NaN
1                         NaN                  NaN
2                         NaN                  NaN
3                         NaN                  NaN
4  At my home^Health facility                    2

            u2_post_no_provider_why u2_post_no_provider_why_other  \
0            Health facility closed                           NaN
1                               NaN                           NaN
2                               NaN                           NaN
3  Turned away from health facility                           NaN
4                               NaN                           NaN

                           country              user_id
0                          Jamaica  u2_4519417576161280
1                    United States  u2_6197364724858880
2  Democratic Republic of the Congo  u2_5202308155834368
3  Democratic Republic of the Congo  u2_6709282991243264
4                         Nicaragua  u2_6190896700194816

[5 rows x 47 columns]
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7228 entries, 0 to 7227
Data columns (total 47 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   observation_id                  7228 non-null   object
 1   submitted_time                  7228 non-null   object
 2   gender                          7228 non-null   object
 3   age                             7228 non-null   object
 4   geography                       7228 non-null   object
 5   financial_situation             7228 non-null   object
 6   education                       7228 non-null   object
 7   employment_status               7228 non-null   object
 8   ethnicity                       7228 non-null   object
 9   religion                        7228 non-null   object
 10  u2_hh                           7227 non-null   object
 11  u2_child_count                  7228 non-null   int64
 12  u2_child1_age_months            7228 non-null   float64
 13  u2_child1_sex                   7228 non-null   object
 14  u2_child2_age_months            2756 non-null   float64
 15  u2_child2_sex                   2756 non-null   object
 16  u2_child3_age_months            922 non-null    float64
 17  u2_child3_sex                   922 non-null    object
 18  u2_vaccine_card                 7228 non-null   object
 19  u2_pre_vaccines                 7228 non-null   object
 20  u2_pre_vaccine_count            5149 non-null   float64
 21  u2_pre_no_vaccine_why           2079 non-null   object
 22  u2_pre_no_vaccine_why_other     222 non-null    object
 23  u2_post_vaccines                7228 non-null   object
 24  u2_post_vaccine_count           4451 non-null   float64
 25  u2_post_no_vaccine_why          2777 non-null   object
 26  u2_post_no_vaccine_why_other    183 non-null    object
 27  u2_dtp_doses                    7228 non-null   object
 28  u2_rotavirus_doses              7228 non-null   object
 29  u2_pneu_doses                   7228 non-null   object
 30  u2_measles_doses                7228 non-null   object
 31  u2_pre_provider_need            7228 non-null   object
 32  u2_pre_condition                3420 non-null   object
 33  u2_pre_provider_visit           3390 non-null   object
 34  u2_pre_provider_where           2070 non-null   object
 35  u2_pre_provider_num             2932 non-null   object
 36  u2_pre_no_provider_why          1320 non-null   object
 37  u2_pre_no_provider_why_other    63 non-null     object
 38  u2_post_provider_need           7174 non-null   object
 39  u2_post_condition               3402 non-null   object
 40  u2_post_provider_visit          3365 non-null   object
 41  u2_post_provider_where          1845 non-null   object
 42  u2_post_provider_num            2909 non-null   object
 43  u2_post_no_provider_why         1520 non-null   object
 44  u2_post_no_provider_why_other   50 non-null     object
```

```
 45  country                       7228 non-null   object
 46  user_id                       7228 non-null   object
dtypes: float64(5), int64(1), object(41)
memory usage: 2.6+ MB
None

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

       u2_child_count  u2_child1_age_months  u2_child2_age_months  \
count     7228.000000           7228.000000           2756.000000
mean         1.508854              7.646526              7.163831
std          0.710711              6.993561              6.906221
min          1.000000              0.000000              0.000000
25%          1.000000              2.000000              2.000000
50%          1.000000              5.000000              5.000000
75%          2.000000             12.000000             10.000000
max          3.000000             25.000000             25.000000

       u2_child3_age_months  u2_pre_vaccine_count
u2_post_vaccine_count  \
count            922.000000          5.149000e+03
4.451000e+03
mean               6.629826          5.839653e+05
3.009780e+06
std                6.477690          2.273222e+07
1.514366e+08
min                0.000000          0.000000e+00
0.000000e+00
25%                2.000000          1.000000e+00
1.000000e+00
50%                5.000000          2.000000e+00
2.000000e+00
75%                9.000000          3.000000e+00
3.000000e+00
max               25.000000          1.124894e+09
9.983722e+09

       byleco
count  7228.0
mean      1.0
std       0.0
min       1.0
25%       1.0
50%       1.0
75%       1.0
max       1.0
```

```python
#Oblicz średni dla kolumny
mean_child_count = data_frame["u2_child_count"].mean()
print(f"Średnia liczba dzieci: {mean_child_count}")
```

```
Średnia liczba dzieci: 1.5088544548976204
```

```python
#Oblicz median dla kolumny
median_child_count = data_frame["u2_child_count"].median()
print(f"Mediana liczby dzieci: {median_child_count}")
```

```
Mediana liczby dzieci: 1.0
```

```python
#Oblicz odchylenie standardowe dla kolumny
std_child_count = data_frame["u2_child_count"].std()
print(f"Odchylenie standardowe liczby dzieci: {std_child_count}")
```

```
Odchylenie standardowe liczby dzieci: 0.7107112373052482
```

```python
#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                          0
age                             0
geography                       0
financial_situation             0
education                       0
employment_status               0
ethnicity                       0
religion                        0
u2_hh                           1
u2_child_count                  0
u2_child1_age_months            0
u2_child1_sex                   0
u2_child2_age_months         4472
u2_child2_sex                4472
u2_child3_age_months         6306
u2_child3_sex                6306
u2_vaccine_card                 0
u2_pre_vaccines                 0
u2_pre_vaccine_count         2079
u2_pre_no_vaccine_why        5149
u2_pre_no_vaccine_why_other  7006
u2_post_vaccines                0
u2_post_vaccine_count        2777
u2_post_no_vaccine_why       4451
u2_post_no_vaccine_why_other 7045
```

```
u2_dtp_doses                        0
u2_rotavirus_doses                  0
u2_pneu_doses                       0
u2_measles_doses                    0
u2_pre_provider_need                0
u2_pre_condition                 3808
u2_pre_provider_visit            3838
u2_pre_provider_where            5158
u2_pre_provider_num              4296
u2_pre_no_provider_why           5908
u2_pre_no_provider_why_other     7165
u2_post_provider_need              54
u2_post_condition                3826
u2_post_provider_visit           3863
u2_post_provider_where           5383
u2_post_provider_num             4319
u2_post_no_provider_why          5708
u2_post_no_provider_why_other    7178
country                             0
user_id                             0
dtype: int64
```

```
#Uzupełnij brakujce wartoci redni w kolumnie liczba dzieci
data_frame["u2_child_count"].fillna(data_frame["u2_child_count"].mean(
), inplace=True)
```

```
C:\Users\Szymon\AppData\Local\Temp\ipykernel_2632\2284081746.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["u2_child_count"].fillna(data_frame["u2_child_count"].mean(
), inplace=True)
```

```
#Usuń wiersze, gdzie brakuje danych w kolumnie
data_frame.dropna(subset=["u2_child_count"], inplace=True)
```

```
#Oblicz IQR
Q1 = data_frame["u2_child_count"].quantile(0.25)
Q3 = data_frame["u2_child_count"].quantile(0.75)
IQR = Q3 - Q1
print(IQR)
```

```
1.0
```

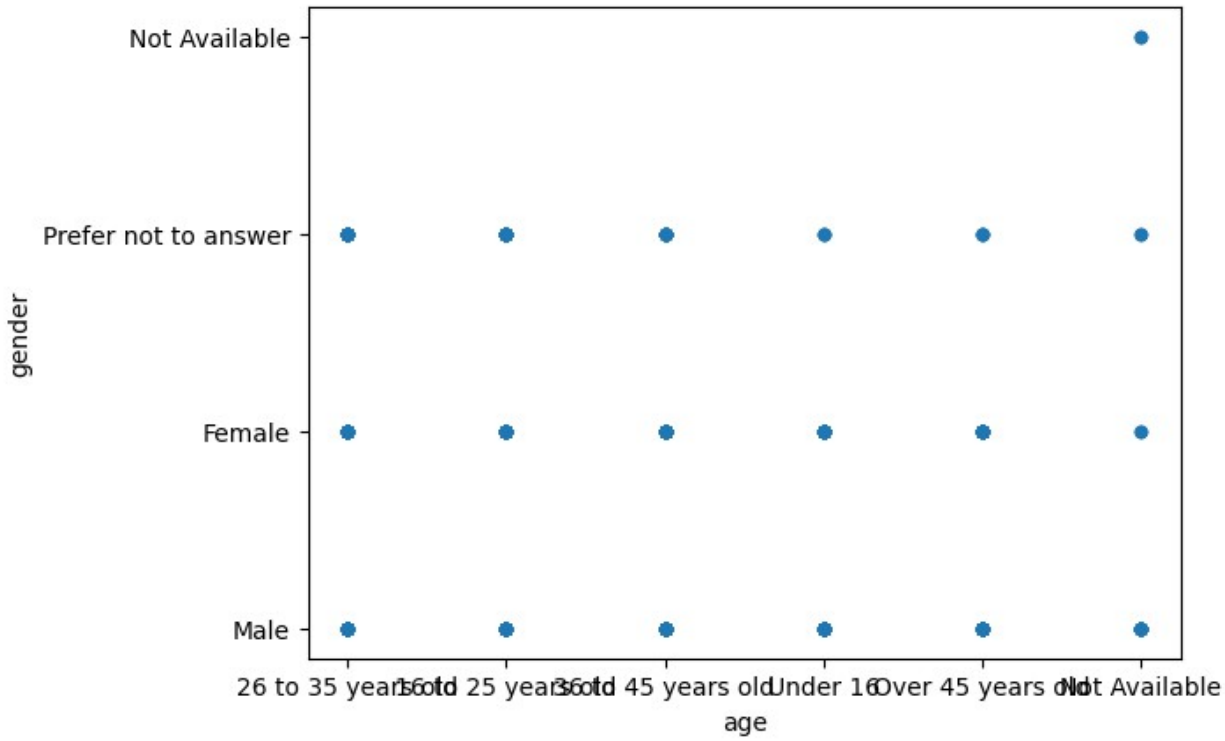
```
#Zidentyfikuj wartoci odstajce
outliers = data_frame[(data_frame["u2_child_count"] < (Q1 - 1.5 *
IQR)) | (data_frame["u2_child_count"] > (Q3 + 1.5 * IQR))]
print("Wartoci odstajce :")
print(outliers)
```

```
Wartoci odstajce :
Empty DataFrame
Columns: [observation_id, submitted_time, gender, age, geography,
financial_situation, education, employment_status, ethnicity,
religion, u2_hh, u2_child_count, u2_child1_age_months, u2_child1_sex,
u2_child2_age_months, u2_child2_sex, u2_child3_age_months,
u2_child3_sex, u2_vaccine_card, u2_pre_vaccines, u2_pre_vaccine_count,
u2_pre_no_vaccine_why, u2_pre_no_vaccine_why_other, u2_post_vaccines,
u2_post_vaccine_count, u2_post_no_vaccine_why,
u2_post_no_vaccine_why_other, u2_dtp_doses, u2_rotavirus_doses,
u2_pneu_doses, u2_measles_doses, u2_pre_provider_need,
u2_pre_condition, u2_pre_provider_visit, u2_pre_provider_where,
u2_pre_provider_num, u2_pre_no_provider_why,
u2_pre_no_provider_why_other, u2_post_provider_need,
u2_post_condition, u2_post_provider_visit, u2_post_provider_where,
u2_post_provider_num, u2_post_no_provider_why,
u2_post_no_provider_why_other, country, user_id]
Index: []

[0 rows x 47 columns]
```

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

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




```
#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]
```