In [97]:

```python
# ladowanie biblioteki Pandas
import pandas as pd

# tworzenie ramki danych ze słownika
countries = {
    "Aridia": {
        "iso": "ARX",
        "capital": "Solara",
        "continent": "Oceania",
        "population": 4_200_000,
        "area_km2": 125_400,
        "gdp_usd_billion": 28.6,
        "official_languages": ["Aridian", "English"],
        "neighbors": ["Belvar", "Cascadia"],
        "independence_year": 1978,
        "currency": "Aridian Crown"
    },
    "Belvar": {
        "iso": "BVL",
        "capital": "Novo",
        "continent": "Europe",
        "population": 12_750_000,
        "area_km2": 89_700,
        "gdp_usd_billion": 215.1,
        "official_languages": ["Belvarian"],
        "neighbors": ["Aridia", "Denshire", "Kandor"],
        "independence_year": 1850,
        "currency": "Belvar Lira"
    },
    "Cascadia": {
        "iso": "CSC",
        "capital": "Rivermouth",
        "continent": "North America",
        "population": 6_980_000,
        "area_km2": 410_200,
        "gdp_usd_billion": 145.3,
        "official_languages": ["English", "Cascadian"],
        "neighbors": ["Aridia"],
        "independence_year": 1949,
        "currency": "Cascade Dollar"
    },
    "Denshire": {
        "iso": "DNS",
        "capital": "Highgate",
        "continent": "Asia",
        "population": 83_400_000,
        "area_km2": 650_000,
        "gdp_usd_billion": 1_120.7,
        "official_languages": ["Denshirian", "Mandarin"],
        "neighbors": ["Belvar", "Kandor"],
        "independence_year": 1921,
        "currency": "Denshire Rupee"
    },
    "Kandor": {
        "iso": "KDR",
        "capital": "Halem",
        "continent": "Africa",
        "population": 27_600_000,
```

```
        "area_km2": 310_500,
        "gdp_usd_billion": 89.4,
        "official_languages": ["Kandori", "French"],
        "neighbors": ["Belvar", "Denshire"],
        "independence_year": 1963,
        "currency": "Kandor Franc"
    },
    "Zephyria": {
        "iso": "ZPH",
        "capital": "Skylight",
        "continent": "South America",
        "population": 3_120_000,
        "area_km2": 210_900,
        "gdp_usd_billion": 34.9,
        "official_languages": ["Zephyrian", "Spanish"],
        "neighbors": [],
        "independence_year": 1987,
        "currency": "Zeph"
    }
}

df = pd.DataFrame(countries)

# zachowanie ramki danych pobranych z pliku w formacie csv (xlsx)
df.to_csv("cities.csv")

# tworzenie ramki danych z listy list
samples = [
    ["id", "name", "age", "city", "active"],
    [1, "Alice", 25, "Solara", True],
    [2, "Bob", 30, "Novo", True],
    [3, "Charlie", 35, "Rivermouth", False],
    [4, "Diana", 28, "Highgate", True],
    [5, "Ethan", 42, "Halem", False],
]

samples_df = pd.DataFrame(samples[1:], columns=samples[0])
samples_df

# transponowanie
pd.DataFrame(samples).T
```

Out[97]:

	0	1	2	3	4	5
0	id	1	2	3	4	5
1	name	Alice	Bob	Charlie	Diana	Ethan
2	age	25	30	35	28	42
3	city	Solara	Novo	Rivermouth	Highgate	Halem
4	active	True	True	False	True	False

In [98]:

```
#wyswietlic pierwsze 10 wierszy ramki danych
df = pd.read_csv('data.csv')
df.head(10)
```

Out[98]:

	observation_id	submitted_time	gender	age	geography	financial_situa
0	wmn_4503683847159808	2020-07-09 23:19:01.982 UTC	Female	26 to 35 years old	City center or metropolitan area	I cannot a enough foo my fa
1	wmn_4503772699295744	2020-07-09 21:22:15.864 UTC	Female	16 to 25 years old	Rural	I cannot a enough foo my fa
2	wmn_4504010469146624	2020-07-10 05:09:07.359 UTC	Female	16 to 25 years old	Rural	I can afford but nothing
3	wmn_4504035500752896	2020-07-11 16:59:49.85 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can afford and re expenses, and
4	wmn_4504181395423232	2020-07-11 18:43:35.954 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can afford and re expenses, and
5	wmn_4504301990051840	2020-07-10 11:27:16.581 UTC	Female	26 to 35 years old	City center or metropolitan area	I can comfor afford clothes, and
6	wmn_4504322055602176	2020-07-09 20:43:11.055 UTC	Female	16 to 25 years old	City center or metropolitan area	I can afford and re expenses, and
7	wmn_4504369904222208	2020-07-18 12:52:31.482 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I cannot a enough foo my fa
8	wmn_4504469091123200	2020-07-16 16:03:44.066 UTC	Female	36 to 45 years old	City center or metropolitan area	I can afford and re expenses, but
9	wmn_4504687899574272	2020-07-17 07:16:32.082 UTC	Female	36 to 45 years old	City center or metropolitan area	I can afford and re expenses, but

10 rows × 46 columns

In [99]:
```
# wyswietlic ostatnie 10 wierszy ramki danych
df.tail(10)
```

Out[99]:

	observation_id	submitted_time	gender	age	geography	financial_
12344	wmn_6752309616050176	2020-07-11 14:37:07.551 UTC	Female	26 to 35 years old	City center or metropolitan area	I can cc af clothe
12345	wmn_6752631872815104	2020-07-10 02:25:50.01 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expense
12346	wmn_6752968893530112	2020-07-14 05:19:46.429 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expenses
12347	wmn_6753819934588928	2020-07-25 17:34:27.837 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
12348	wmn_6753897143336960	2020-07-10 11:49:14.64 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
12349	wmn_6754210441068544	2020-07-16 15:46:12.095 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
12350	wmn_6754415891709952	2020-07-10 09:57:24.863 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expense
12351	wmn_6754483574145024	2020-07-19 17:50:01.295 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expense
12352	wmn_6755256899993600	2020-07-11 16:09:09.78 UTC	Female	36 to 45 years old	Rural	I can cc af clothe
12353	wmn_6755376524689408	2020-07-17 03:19:00.388 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expenses

10 rows × 46 columns

In [100…

```
#wyswietlic informacje, o ramce danych
df.describe
```

Out[100…

```
<bound method NDFrame.describe of              observation_id              sub
mitted_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
...                     ...                          ...     ...
12349  wmn_6754210441068544  2020-07-16 15:46:12.095 UTC  Female
12350  wmn_6754415891709952  2020-07-10 09:57:24.863 UTC  Female
12351  wmn_6754483574145024  2020-07-19 17:50:01.295 UTC  Female
12352  wmn_6755256899993600   2020-07-11 16:09:09.78 UTC  Female
12353  wmn_6755376524689408  2020-07-17 03:19:00.388 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
...                    ...                              ...
12349  26 to 35 years old             Suburban/Peri-urban
12350  16 to 25 years old  City center or metropolitan area
12351  26 to 35 years old  City center or metropolitan area
12352  36 to 45 years old                             Rural
12353  26 to 35 years old  City center or metropolitan area

                                    financial_situation  \
0              I cannot afford enough food for my family
1              I cannot afford enough food for my family
2                    I can afford food, but nothing else
3      I can afford food and regular expenses, and bu...
4      I can afford food and regular expenses, and bu...
...                                                  ...
12349  I can afford food and regular expenses, but no...
12350  I can afford food and regular expenses, but no...
12351  I can afford food and regular expenses, but no...
12352  I can comfortably afford food, clothes, and fu...
12353  I can afford food and regular expenses, and bu...

                   education   employment_status           ethnicity  \
0      College or university          Unemployed            Mestizo
1      Secondary/high school             Student            Tagalog
2      College or university             Student          Hiligaynon
3      College or university          Unemployed                Thai
4      College or university  Employed full-time             African
...                      ...                 ...                 ...
12349  College or university          Unemployed             African
12350  Secondary/high school       Self-employed                 Kru
12351       Technical school          Unemployed  Non-hispanic White
12352  College or university       Self-employed                Thai
12353  Secondary/high school  Employed full-time             Mestizo

                   religion  ... wmn_pre_safe_place wmn_post_safe_place  \
0               Catholicism  ...                NaN                 NaN
1                    Muslim  ...                NaN                 NaN
2              Christianity  ...                NaN                 NaN
3                  Buddhism  ...                NaN                 NaN
4              Christianity  ...                NaN                 NaN
...                     ...  ...                ...                 ...
```

```
12349               Muslim  ...               NaN               NaN
12350          Catholicism  ...               NaN               NaN
12351  Prefer Not To Answer  ...         Every day            Rarely
12352             Buddhism  ...               NaN               NaN
12353          Catholicism  ...               NaN               NaN

      wmn_safe_place_no_access wmn_safe_place_no_access_why wmn_pre_help  \
0                          NaN                          NaN          NaN
1                          NaN                          NaN          NaN
2                          NaN                          NaN          NaN
3                          NaN                          NaN          NaN
4                          NaN                          NaN          NaN
...                        ...                          ...          ...
12349                      NaN                          NaN          NaN
12350                      NaN                          NaN          NaN
12351                       No                          NaN          Yes
12352                      NaN                          NaN          NaN
12353                      NaN                          NaN          NaN

      wmn_post_help wmn_post_no_help    wmn_no_help_why  \
0               NaN              NaN                NaN
1               NaN              NaN                NaN
2               NaN              NaN                NaN
3               NaN              NaN                NaN
4               NaN              NaN                NaN
...             ...              ...                ...
12349           NaN              NaN                NaN
12350           NaN              NaN                NaN
12351           Yes              Yes  Decline to Answer
12352           NaN              NaN                NaN
12353           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
...                             ...                   ...
12349  United Republic of Tanzania  wmn_6151550260215808
12350                 Ivory Coast  wmn_5222308327456768
12351               United States  wmn_4706368994148352
12352                    Thailand  wmn_6730228637892608
12353                     Ecuador  wmn_6585614141489152

[12354 rows x 46 columns]>
```

In [101…

```
# wyswietlic, ile wierszy i kolumn znajduje sie¸ w ramce danych
df.shape
```

Out[101…

```
(12354, 46)
```

In [102…

```
#  wyswietlic informacje¸ statystyczna¸ o kolumnach liczbowych (wartosci
# niepowtarzalne, srednia, odchylenie standardowe, minimum, kwartyle, maksimum )
df.info
```

Out[102…

```
<bound method DataFrame.info of              observation_id              submi
tted_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
...                     ...                          ...     ...
12349  wmn_6754210441068544  2020-07-16 15:46:12.095 UTC  Female
12350  wmn_6754415891709952  2020-07-10 09:57:24.863 UTC  Female
12351  wmn_6754483574145024  2020-07-19 17:50:01.295 UTC  Female
12352  wmn_6755256899993600   2020-07-11 16:09:09.78 UTC  Female
12353  wmn_6755376524689408  2020-07-17 03:19:00.388 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
...                    ...                               ...
12349  26 to 35 years old             Suburban/Peri-urban
12350  16 to 25 years old  City center or metropolitan area
12351  26 to 35 years old  City center or metropolitan area
12352  36 to 45 years old                             Rural
12353  26 to 35 years old  City center or metropolitan area

                                    financial_situation  \
0              I cannot afford enough food for my family
1              I cannot afford enough food for my family
2                    I can afford food, but nothing else
3      I can afford food and regular expenses, and bu...
4      I can afford food and regular expenses, and bu...
...                                                  ...
12349  I can afford food and regular expenses, but no...
12350  I can afford food and regular expenses, but no...
12351  I can afford food and regular expenses, but no...
12352  I can comfortably afford food, clothes, and fu...
12353  I can afford food and regular expenses, and bu...

                   education   employment_status           ethnicity  \
0      College or university          Unemployed            Mestizo
1      Secondary/high school             Student            Tagalog
2      College or university             Student          Hiligaynon
3      College or university          Unemployed                Thai
4      College or university  Employed full-time             African
...                      ...                 ...                 ...
12349  College or university          Unemployed             African
12350  Secondary/high school       Self-employed                 Kru
12351       Technical school          Unemployed  Non-hispanic White
12352  College or university       Self-employed                Thai
12353  Secondary/high school  Employed full-time             Mestizo

                   religion  ... wmn_pre_safe_place wmn_post_safe_place  \
0               Catholicism  ...                NaN                 NaN
1                    Muslim  ...                NaN                 NaN
2              Christianity  ...                NaN                 NaN
3                  Buddhism  ...                NaN                 NaN
4              Christianity  ...                NaN                 NaN
...                     ...  ...                ...                 ...
```

```
12349               Muslim  ...                NaN                NaN
12350          Catholicism  ...                NaN                NaN
12351  Prefer Not To Answer  ...          Every day             Rarely
12352             Buddhism  ...                NaN                NaN
12353          Catholicism  ...                NaN                NaN

      wmn_safe_place_no_access wmn_safe_place_no_access_why wmn_pre_help  \
0                          NaN                          NaN          NaN
1                          NaN                          NaN          NaN
2                          NaN                          NaN          NaN
3                          NaN                          NaN          NaN
4                          NaN                          NaN          NaN
...                        ...                          ...          ...
12349                      NaN                          NaN          NaN
12350                      NaN                          NaN          NaN
12351                       No                          NaN          Yes
12352                      NaN                          NaN          NaN
12353                      NaN                          NaN          NaN

      wmn_post_help wmn_post_no_help    wmn_no_help_why  \
0               NaN              NaN                NaN
1               NaN              NaN                NaN
2               NaN              NaN                NaN
3               NaN              NaN                NaN
4               NaN              NaN                NaN
...             ...              ...                ...
12349           NaN              NaN                NaN
12350           NaN              NaN                NaN
12351           Yes              Yes  Decline to Answer
12352           NaN              NaN                NaN
12353           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
...                            ...                   ...
12349 United Republic of Tanzania  wmn_6151550260215808
12350                Ivory Coast  wmn_5222308327456768
12351              United States  wmn_4706368994148352
12352                   Thailand  wmn_6730228637892608
12353                    Ecuador  wmn_6585614141489152

[12354 rows x 46 columns]>
```

In [103…

```python
# wyswietlic informacje statystyczna o kolumnach kategoryzowanych (ile
# unikalnych wartosci, top - jaka jest najpopularniejsza wartosc, freq -
# jak czesto najpopularniejsza)

df.describe(include="all")
```

Out[103…

	observation_id	submitted_time	gender	age	geography	financial_
count	12354	12354	12354	12354	12354	
unique	12354	12354	3	6	4	
top	wmn_4503683847159808	2020-07-09 23:19:01.982 UTC	Female	16 to 25 years old	City center or metropolitan area	I can af ar expenses
freq	1	1	12331	5672	5274	
mean	NaN	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	NaN	

11 rows × 46 columns

In [104…

```
# usunac brakujace wartosci w ramce danych
df.fillna(value=0)
```

Out[104…

	observation_id	submitted_time	gender	age	geography	financial_
0	wmn_4503683847159808	2020-07-09 23:19:01.982 UTC	Female	26 to 35 years old	City center or metropolitan area	I can enoug
1	wmn_4503772699295744	2020-07-09 21:22:15.864 UTC	Female	16 to 25 years old	Rural	I can enoug
2	wmn_4504010469146624	2020-07-10 05:09:07.359 UTC	Female	16 to 25 years old	Rural	I can af but nc
3	wmn_4504035500752896	2020-07-11 16:59:49.85 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can a a expense
4	wmn_4504181395423232	2020-07-11 18:43:35.954 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
...	...	...	...	...	...	
12349	wmn_6754210441068544	2020-07-16 15:46:12.095 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
12350	wmn_6754415891709952	2020-07-10 09:57:24.863 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expense
12351	wmn_6754483574145024	2020-07-19 17:50:01.295 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expense
12352	wmn_6755256899993600	2020-07-11 16:09:09.78 UTC	Female	36 to 45 years old	Rural	I can cc af clothe
12353	wmn_6755376524689408	2020-07-17 03:19:00.388 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expense

12354 rows × 46 columns

In [105…

```
# Przedstawienie wyboru wierszy i kolumny używając nazw oraz indeksów na różne s
df["wmn_alone"]
```

Out[105…

```
0        Yes
1         No
2         No
3         No
4        Yes
        ...
12349     No
12350     No
12351     No
12352    Yes
12353    Yes
Name: wmn_alone, Length: 12354, dtype: object
```

In [106…

```
df.gender
```

Out[106…

```
0        Female
1        Female
2        Female
3        Female
4        Female
         ...
12349    Female
12350    Female
12351    Female
12352    Female
12353    Female
Name: gender, Length: 12354, dtype: object
```

In [107…

```
df[["gender","country","user_id"]]
```

Out[107…

	gender	country	user_id
0	Female	Ecuador	wmn_5900473574883328
1	Female	Philippines	wmn_5702261783658496
2	Female	Philippines	wmn_5652767014387712
3	Female	Thailand	wmn_6411372690669568
4	Female	United Republic of Tanzania	wmn_6215734184378368
...	...	...	...
12349	Female	United Republic of Tanzania	wmn_6151550260215808
12350	Female	Ivory Coast	wmn_5222308327456768
12351	Female	United States	wmn_4706368994148352
12352	Female	Thailand	wmn_6730228637892608
12353	Female	Ecuador	wmn_6585614141489152

12354 rows × 3 columns

In [108…

```
df.loc[:, "gender":"country"]
```

Out[108…

	gender	age	geography	financial_situation	education	employment_st
0	Female	26 to 35 years old	City center or metropolitan area	I cannot afford enough food for my family	College or university	Unempl
1	Female	16 to 25 years old	Rural	I cannot afford enough food for my family	Secondary/high school	Stu
2	Female	16 to 25 years old	Rural	I can afford food, but nothing else	College or university	Stu
3	Female	16 to 25 years old	Suburban/Peri-urban	I can afford food and regular expenses, and bu...	College or university	Unempl
4	Female	26 to 35 years old	Suburban/Peri-urban	I can afford food and regular expenses, and bu...	College or university	Employed full-
...	...	...	...	...	...	
12349	Female	26 to 35 years old	Suburban/Peri-urban	I can afford food and regular expenses, but no...	College or university	Unempl
12350	Female	16 to 25 years old	City center or metropolitan area	I can afford food and regular expenses, but no...	Secondary/high school	Self-empl
12351	Female	26 to 35 years old	City center or metropolitan area	I can afford food and regular expenses, but no...	Technical school	Unempl
12352	Female	36 to 45 years old	Rural	I can comfortably afford food, clothes, and fu...	College or university	Self-empl
12353	Female	26 to 35 years old	City center or metropolitan area	I can afford food and regular expenses, and bu...	Secondary/high school	Employed full-

12354 rows × 43 columns

In [109…

```
df.iloc[10:15, 0:3]
```

Out[109…

	observation_id	submitted_time	gender
10	wmn_4504764873441280	2020-07-23 00:29:41.766 UTC	Female
11	wmn_4505151219171328	2020-07-14 04:43:18.445 UTC	Female
12	wmn_4505272015126528	2020-07-16 17:21:24.686 UTC	Female
13	wmn_4505544242233344	2020-07-10 01:19:58.4 UTC	Female
14	wmn_4505874686279680	2020-07-18 06:43:04.442 UTC	Female

In [110…

```
df.loc[100:110, "gender":"country"]
```

Out[110...

	gender	age	geography	financial_situation	education	employment_stat
100	Female	16 to 25 years old	City center or metropolitan area	I can comfortably afford food, clothes, and fu...	Technical school	Student and wo part-tin
101	Female	16 to 25 years old	Suburban/Peri-urban	I can afford food, but nothing else	College or university	Employed part-tin
102	Female	36 to 45 years old	Suburban/Peri-urban	I can afford food and regular expenses, but no...	College or university	Unemploye
103	Female	36 to 45 years old	City center or metropolitan area	I can afford food and regular expenses, and bu...	Post graduate	Retire
104	Female	16 to 25 years old	City center or metropolitan area	I can afford food, but nothing else	Secondary/high school	Stude
105	Female	16 to 25 years old	City center or metropolitan area	I can comfortably afford food, clothes, and fu...	College or university	Student and wo part-tin
106	Female	16 to 25 years old	City center or metropolitan area	I can comfortably afford food, clothes, and fu...	College or university	Self-employe
107	Female	16 to 25 years old	Rural	I can comfortably afford food, clothes, and fu...	College or university	Stude
108	Female	26 to 35 years old	Suburban/Peri-urban	I can comfortably afford food, clothes, and fu...	College or university	Employed full-tin
109	Female	Over 45 years old	City center or metropolitan area	I can afford food, but nothing else	Secondary/high school	Self-employe
110	Female	26 to 35 years old	City center or metropolitan area	I can afford food, but nothing else	College or university	Employed full-tin

11 rows × 43 columns

In [111…

```
# Przedstawienie wyboru wierszy z ramki danych pod warunkiem odnośnie określonej
df.wmn_hh = pd.to_numeric(df['wmn_hh'], errors='coerce')
over_value = df[(df["wmn_hh"] >= 10)]
over_value
```

Out[111…

	observation_id	submitted_time	gender	age	geography	financia
1	wmn_4503772699295744	2020-07-09 21:22:15.864 UTC	Female	16 to 25 years old	Rural	I ca enou
6	wmn_4504322055602176	2020-07-09 20:43:11.055 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a expense
8	wmn_4504469091123200	2020-07-16 16:03:44.066 UTC	Female	36 to 45 years old	City center or metropolitan area	I can a expens
10	wmn_4504764873441280	2020-07-23 00:29:41.766 UTC	Female	16 to 25 years old	Rural	I can a but n
52	wmn_4513376917258240	2020-07-10 01:31:47.575 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a expens
…	…	…	…	…	…	
12040	wmn_6699298844835840	2020-07-10 21:13:02.67 UTC	Female	16 to 25 years old	City center or metropolitan area	I ca enou
12075	wmn_6705767504936960	2020-07-10 01:19:59.82 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can a but n
12153	wmn_6721338137116672	2020-07-12 18:34:01.47 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a but n
12190	wmn_6726386669846528	2020-07-10 00:57:12.129 UTC	Female	Under 16	Rural	I can c a cloth
12253	wmn_6736732706045952	2020-07-10 03:03:14.392 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a but n

455 rows × 46 columns

In [112…

```
# Przedstawienie wyboru wierszy z ramki danych pod warunkiem spełnienia kilku wa
sample_data = df[(df["gender"] == "Female")&(df["age"] == "16 to 25 years old")
sample_data
```

Out[112…

	observation_id	submitted_time	gender	age	geography	financial_situ
1	wmn_4503772699295744	2020-07-09 21:22:15.864 UTC	Female	16 to 25 years old	Rural	I cannot enough fo my
2	wmn_4504010469146624	2020-07-10 05:09:07.359 UTC	Female	16 to 25 years old	Rural	I can afford but nothir
10	wmn_4504764873441280	2020-07-23 00:29:41.766 UTC	Female	16 to 25 years old	Rural	I can afford but nothir
13	wmn_4505544242233344	2020-07-10 01:19:58.4 UTC	Female	16 to 25 years old	Rural	I cannot enough fo my
14	wmn_4505874686279680	2020-07-18 06:43:04.442 UTC	Female	16 to 25 years old	Rural	I can afford and r expenses, bu
...	...	...	...	...	...	
12235	wmn_6733738311680000	2020-07-15 08:21:35.974 UTC	Female	16 to 25 years old	Rural	I can afford and r expenses, ar
12276	wmn_6739594566696960	2020-07-11 02:09:48.315 UTC	Female	16 to 25 years old	Rural	I can afford and r expenses, bu
12288	wmn_6742010620018688	2020-07-24 08:01:47.993 UTC	Female	16 to 25 years old	Rural	I can afford and r expenses, bu
12309	wmn_6745823410126848	2020-07-14 12:52:46.253 UTC	Female	16 to 25 years old	Rural	I can afford and r expenses, bu
12328	wmn_6749915977089024	2020-07-10 09:24:53.523 UTC	Female	16 to 25 years old	Rural	I can afford and r expenses, ar

1519 rows × 46 columns

In [113…

```
#Wybranie wierszy, które zawierają w kolumnie kategoryzowanej określone słowo
df[df["gender"] == "Female"]
```

Out[113…

	observation_id	submitted_time	gender	age	geography	financial_
0	wmn_4503683847159808	2020-07-09 23:19:01.982 UTC	Female	26 to 35 years old	City center or metropolitan area	I can enoug
1	wmn_4503772699295744	2020-07-09 21:22:15.864 UTC	Female	16 to 25 years old	Rural	I can enoug
2	wmn_4504010469146624	2020-07-10 05:09:07.359 UTC	Female	16 to 25 years old	Rural	I can af but nc
3	wmn_4504035500752896	2020-07-11 16:59:49.85 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can a a expense
4	wmn_4504181395423232	2020-07-11 18:43:35.954 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
...	...	...	...	...	...	
12349	wmn_6754210441068544	2020-07-16 15:46:12.095 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can a a expense
12350	wmn_6754415891709952	2020-07-10 09:57:24.863 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expense
12351	wmn_6754483574145024	2020-07-19 17:50:01.295 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expense
12352	wmn_6755256899993600	2020-07-11 16:09:09.78 UTC	Female	36 to 45 years old	Rural	I can cc af clothe
12353	wmn_6755376524689408	2020-07-17 03:19:00.388 UTC	Female	26 to 35 years old	City center or metropolitan area	I can a a expense

12331 rows × 46 columns

In [114…

```
# Wybranie wierszy, które nie zawierają w kolumnie kategoryzowanej określone sło
df[df["age"] != "26 to 35 years old"]
```

Out[114…

	observation_id	submitted_time	gender	age	geography	financial_
1	wmn_4503772699295744	2020-07-09 21:22:15.864 UTC	Female	16 to 25 years old	Rural	I can enoug
2	wmn_4504010469146624	2020-07-10 05:09:07.359 UTC	Female	16 to 25 years old	Rural	I can af but nc
3	wmn_4504035500752896	2020-07-11 16:59:49.85 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can a a expense
6	wmn_4504322055602176	2020-07-09 20:43:11.055 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expense
7	wmn_4504369904222208	2020-07-18 12:52:31.482 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can enoug
...	...	...	...	...	...	
12342	wmn_6751879750221824	2020-07-12 20:14:41.671 UTC	Female	36 to 45 years old	City center or metropolitan area	I can cc af clothe
12343	wmn_6752208617209856	2020-07-09 20:27:36.992 UTC	Female	36 to 45 years old	Suburban/Peri-urban	I can enoug
12346	wmn_6752968893530112	2020-07-14 05:19:46.429 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expense
12350	wmn_6754415891709952	2020-07-10 09:57:24.863 UTC	Female	16 to 25 years old	City center or metropolitan area	I can a a expense
12352	wmn_6755256899993600	2020-07-11 16:09:09.78 UTC	Female	36 to 45 years old	Rural	I can cc af clothe

7998 rows × 46 columns

In [115…
```python
# Utworzenie kolumny na podstawie istniejących
df["new_creation"] = df["gender"] + df["age"]
df["new_creation"]
```

Out[115…
```
0        Female26 to 35 years old
1        Female16 to 25 years old
2        Female16 to 25 years old
3        Female16 to 25 years old
4        Female26 to 35 years old
                   ...
12349    Female26 to 35 years old
12350    Female16 to 25 years old
12351    Female26 to 35 years old
12352    Female36 to 45 years old
12353    Female26 to 35 years old
Name: new_creation, Length: 12354, dtype: object
```

In [116…
```python
# Usunięcie kolumn
df.drop("new_creation", axis=1, inplace = True)
```

In [117…
```python
# Zmiania nazwy kolumny
df.rename(columns = {"gender": "sex"}, inplace = True)
```

In [118…
```python
# Zachowanie ramki danych jako CSV
df.to_csv('output.csv')
```

In [119…
```python
# Wyświetlanie Unikatowych wartosci w kolumnie
df['geography'].unique()
```

Out[119…
```
array(['City center or metropolitan area', 'Rural', 'Suburban/Peri-urban',
       'Not Available'], dtype=object)
```

In [120…
```python
# Liczba rekordów odpowiadającej wartosci
df['geography'].value_counts()
```

Out[120…
```
geography
City center or metropolitan area    5274
Suburban/Peri-urban                 4061
Rural                               3011
Not Available                          8
Name: count, dtype: int64
```

In [121…
```python
# Sortowanie wg kolumny
df.sort_values(['geography'], ascending = False)
```

Out[121…

	observation_id	submitted_time	sex	age	geography	financial_
11251	wmn_6558814225825792	2020-07-11 22:33:23.007 UTC	Female	36 to 45 years old	Suburban/Peri-urban	I can af ar expenses
2839	wmn_5016187936440320	2020-07-10 21:44:08.956 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can enough
2805	wmn_5010192094986240	2020-07-10 02:59:29.663 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can af ar expenses
11173	wmn_6545336752668672	2020-07-10 01:24:52.7 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can af ar expenses
6869	wmn_5759099301265408	2020-07-10 05:47:16.176 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can af but no
...	...	...	...	...	...	
5785	wmn_5560947663175680	2020-07-24 11:12:07.109 UTC	Female	26 to 35 years old	City center or metropolitan area	I can af ar expenses
5784	wmn_5560881862934528	2020-07-18 00:06:12.302 UTC	Female	16 to 25 years old	City center or metropolitan area	I can af but no
5783	wmn_5560816901554176	2020-07-15 13:24:04.574 UTC	Female	26 to 35 years old	City center or metropolitan area	I can af ar expenses
5775	wmn_5559023484272640	2020-07-15 02:29:06.661 UTC	Female	26 to 35 years old	City center or metropolitan area	I can af ar expenses
12353	wmn_6755376524689408	2020-07-17 03:19:00.388 UTC	Female	26 to 35 years old	City center or metropolitan area	I can af ar expenses

12354 rows × 46 columns

In [122…

```
# Wyświetlenie wierszy dla 10 największych (najmniejszych) wartości określonej k
df.nlargest(10,'wmn_hh')
```

Out[122…

	observation_id	submitted_time	sex	age	geography	financial_s
143	wmn_4530439953055744	2020-07-10 08:36:13.544 UTC	Female	16 to 25 years old	Suburban/Peri-urban	I can affc but noth
3858	wmn_5203932164128768	2020-07-11 10:39:59.673 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can affc and expenses,
5016	wmn_5414602256154624	2020-07-09 22:34:15.536 UTC	Female	26 to 35 years old	Rural	I can affc and expenses,
5234	wmn_5456310247358464	2020-07-18 01:22:15.4 UTC	Female	16 to 25 years old	City center or metropolitan area	I can affc but noth
5371	wmn_5482101425307648	2020-07-11 13:33:20.521 UTC	Female	16 to 25 years old	Rural	I can affc but noth
5435	wmn_5493846818684928	2020-07-10 19:45:01.023 UTC	Female	36 to 45 years old	Suburban/Peri-urban	I can affc and expenses,
5622	wmn_5527501175783424	2020-07-15 10:33:36.001 UTC	Female	16 to 25 years old	City center or metropolitan area	I can affc but noth
5823	wmn_5566857434365952	2020-07-14 13:45:27.272 UTC	Female	26 to 35 years old	Suburban/Peri-urban	I can affc and expenses,
6086	wmn_5615059754811392	2020-07-16 21:48:51.644 UTC	Female	16 to 25 years old	Rural	I can affc but noth
8449	wmn_6050421933342720	2020-07-10 03:32:46.227 UTC	Female	16 to 25 years old	City center or metropolitan area	I cann enough m

10 rows × 46 columns

In [123…

```
# Wyświetlenie wierszy dla 10 największych wartości określonej kolumny pod warun
df[df['geography'] == 'Rural'].nlargest(10,'wmn_hh')
```

Out[123…

	observation_id	submitted_time	sex	age	geography	financial_situ
5016	wmn_5414602256154624	2020-07-09 22:34:15.536 UTC	Female	26 to 35 years old	Rural	I can afford and re expenses, bu
5371	wmn_5482101425307648	2020-07-11 13:33:20.521 UTC	Female	16 to 25 years old	Rural	I can afford but nothin
6086	wmn_5615059754811392	2020-07-16 21:48:51.644 UTC	Female	16 to 25 years old	Rural	I can afford but nothin
2341	wmn_4931271416610816	2020-07-27 00:41:18.251 UTC	Female	Under 16	Rural	I can afford but nothin
2741	wmn_4998252085903360	2020-07-10 04:49:25.759 UTC	Female	16 to 25 years old	Rural	I can afford but nothin
4603	wmn_5335257969852416	2020-07-12 19:27:19.415 UTC	Female	16 to 25 years old	Rural	I can afford but nothin
6730	wmn_5731160941658112	2020-07-11 13:24:35.86 UTC	Female	26 to 35 years old	Rural	I cannot enough foo my
3398	wmn_5116410931183616	2020-07-09 22:41:38.341 UTC	Female	16 to 25 years old	Rural	I cannot enough foo my
6429	wmn_5678895682813952	2020-07-13 17:18:44.042 UTC	Female	16 to 25 years old	Rural	I can afford but nothin
8546	wmn_6066783175049216	2020-07-10 01:13:27.741 UTC	Female	26 to 35 years old	Rural	I cannot enough foo my

10 rows × 46 columns

In [124…

```
# Grupowanie wierszy według wartości kolumny kategoryzowanej, potem uśrednienie
outcome = df.groupby(['ethnicity','age']).mean(numeric_only=True)
outcome
```

Out[124...

ethnicity	age	wmn_hh	wmn_pre_injectable_missed	wmn_pre_iud_missed	wmn_post
Aboriginal Taiwanese	**16 to 25 years old**	6.000000	NaN	NaN	
	26 to 35 years old	2.400000	NaN	NaN	
Acholi	**16 to 25 years old**	5.000000	1.0	NaN	
	26 to 35 years old	5.000000	NaN	NaN	
Adja and related	**16 to 25 years old**	2.000000	NaN	NaN	
...	**...**	...	...	...	
kurdish	**26 to 35 years old**	4.000000	NaN	NaN	
turkish	**16 to 25 years old**	3.666667	NaN	NaN	
	26 to 35 years old	3.400000	NaN	NaN	
	36 to 45 years old	3.833333	NaN	NaN	
	Over 45 years old	5.000000	NaN	NaN	

498 rows × 5 columns

In [125...

```
# Grupowanie wierszy według wartości kolumny kategoryzowanej, potem uśrednienie
agregations = {
```

```
    'wmn_hh' : 'mean',
    'wmn_pre_iud_missed' : 'median'
}
outcome = df.groupby('age').agg(agregations)
outcome
```

Out[125...

age	wmn_hh	wmn_pre_iud_missed
16 to 25 years old	4.609591	2.0
26 to 35 years old	4.217562	2.0
36 to 45 years old	4.051518	1.0
Not Available	4.307692	NaN
Over 45 years old	3.974026	4.0
Under 16	4.352941	NaN

In [126...

```
# Stworzenie tabeli przestawnej (pivot table) na podstawie ramki danych

tabela_przestawna = pd.pivot_table(
    data = df,
    index = 'geography',
    columns='age',
    values = 'ethnicity',
    aggfunc= 'sum'

)
tabela_przestawna
```

Out[126...

geography	age	16 to 25 years old	
City center or metropolitan area	TagalogArabPashtunMulatoOtherWhiteTagalogWhite...		MestizoBisayaBlack o
Not Available		NaN	
Rural	TagalogHiligaynonPashtunIlocanoBedouinMestizoO...		YorubaBisayaTagalogBis
Suburban/Peri-urban	ThaiShonaJavaOtherKikuyuJavaHausaMestizoMestiz...		AfricanWhiteMestizoKa

In [127...

```
# Wyświetlenie indeksów i kolumn tabeli przestawnej
tabela_przestawna.index
```

Out[127…
```
Index(['City center or metropolitan area', 'Not Available', 'Rural',
       'Suburban/Peri-urban'],
      dtype='object', name='geography')
```

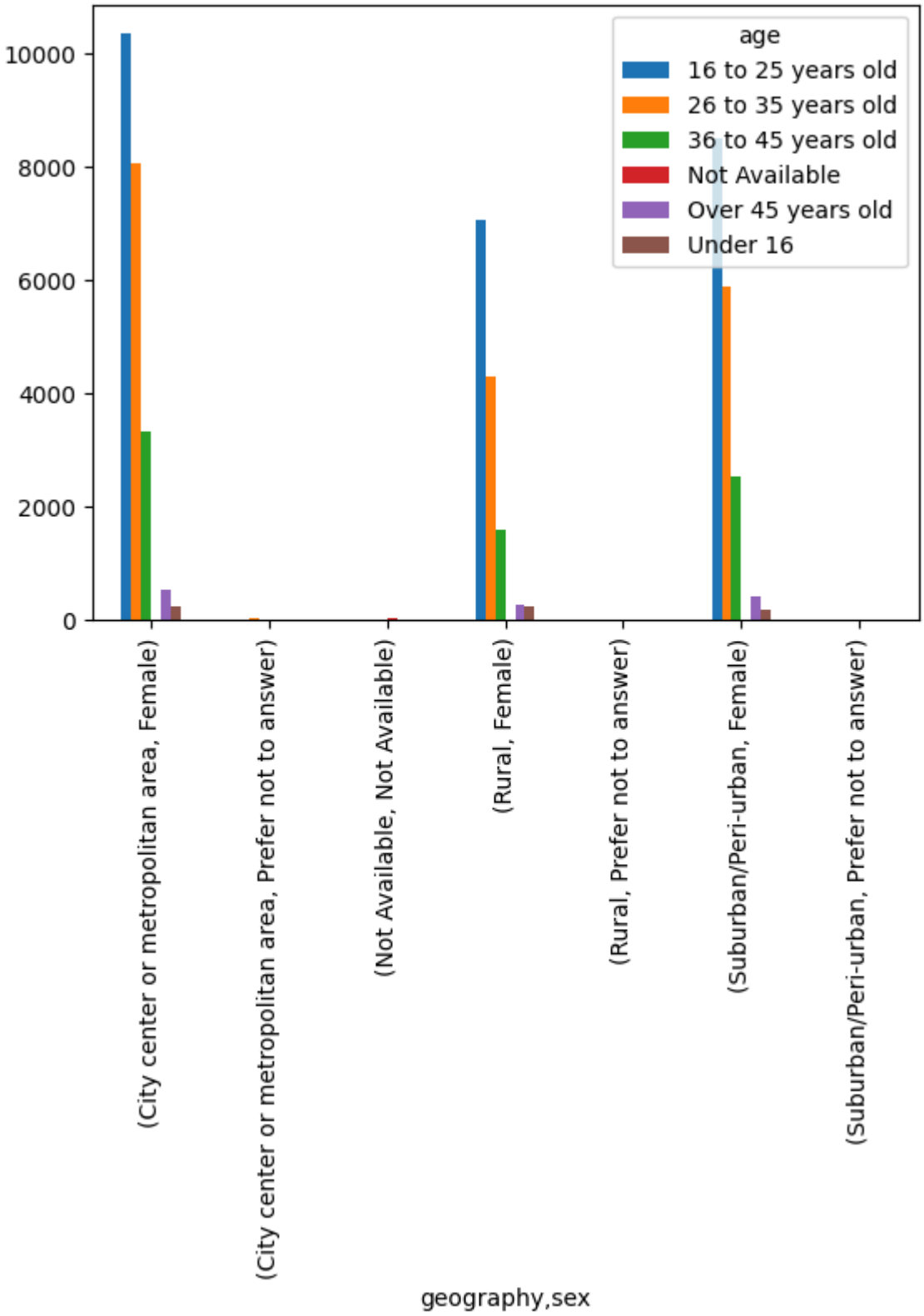
In [128…
```
# Utworzenie indeksu złożonego tabeli przestawnej i wyświetlenie go
tabela_przestawna_2 = pd.pivot_table(
    data = df,
    index = ['geography','sex'],
    columns='age',
    values = 'wmn_hh',
    aggfunc= 'sum'

)
tabela_przestawna_2
```

Out[128…

	age	16 to 25 years old	26 to 35 years old	36 to 45 years old	Not Available	Over 45 years old	Under 16
geography	**sex**						
City center or metropolitan area	**Female**	10360.0	8081.0	3331.0	15.0	529.0	240.0
	Prefer not to answer	NaN	29.0	NaN	NaN	4.0	NaN
Not Available	**Not Available**	NaN	NaN	NaN	35.0	NaN	NaN
Rural	**Female**	7079.0	4290.0	1586.0	3.0	262.0	237.0
	Prefer not to answer	5.0	NaN	2.0	NaN	NaN	NaN
Suburban/Peri-urban	**Female**	8508.0	5892.0	2546.0	3.0	429.0	189.0
	Prefer not to answer	NaN	8.0	6.0	NaN	NaN	NaN

In [129…
```
# Zaimportowanie modułu pyplot z biblioteki matplotlib

import matplotlib.pyplot as plt

%matplotlib inline

# Wyświetlenie wykresu na podstawie tabeli przestawnej

tabela_przestawna_2.plot(kind='bar')
```

Out[129…
```
<Axes: xlabel='geography,sex'>
```

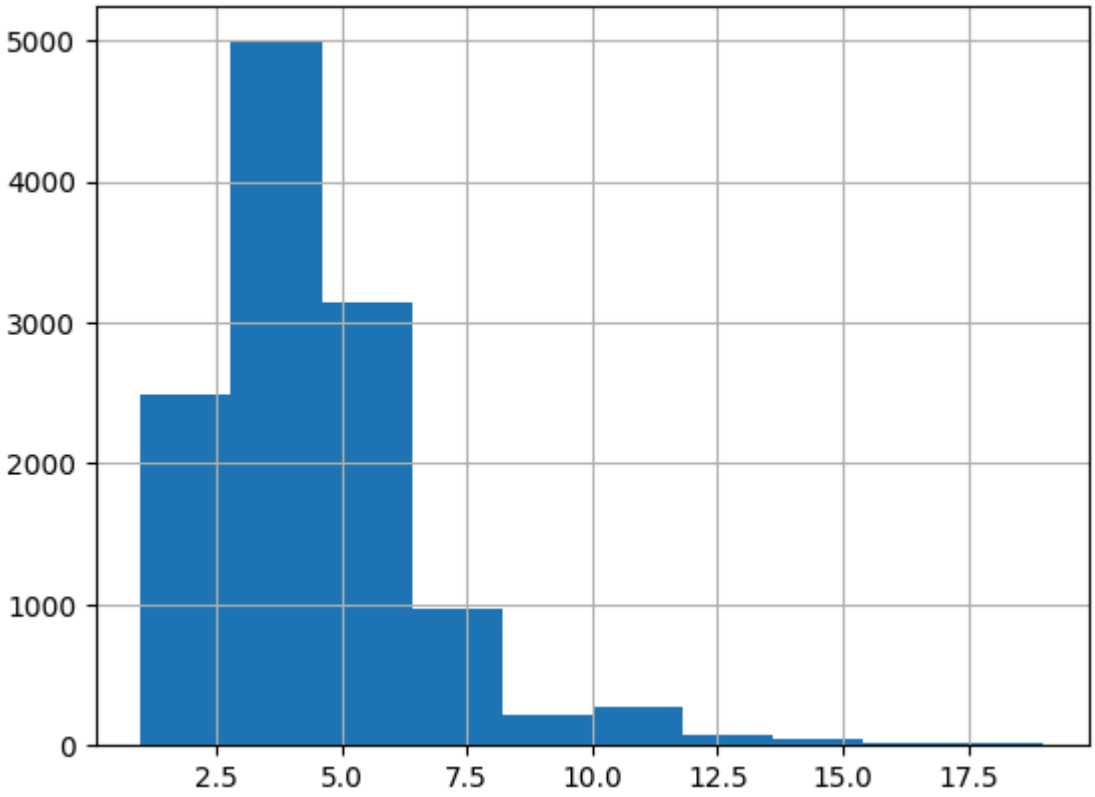


In [130…

```
# Narysowanie histogramu na podstawie wartości kolumny
df.wmn_hh.hist()
```

Out[130…

```
<Axes: >
```



In [131…

```
# Przedstawienie sposobów łączenia ramek danych za pomocą metod merge i concat

df_pracownicy = pd.DataFrame({'ID_Pracownika': [1, 2, 3], 'Imię': ['Anna', 'Piot
df_dzialy = pd.DataFrame({'ID_Pracownika': [2, 3, 4], 'Dział': ['IT', 'HR', 'Mar

# Złączenie wewnętrzne (inner join) na podstawie 'ID_Pracownika'
wynik_merge = pd.merge(df_pracownicy, df_dzialy, on='ID_Pracownika', how='inner'
```

In [132…

```
# Pokazanie dodawania nowych kolumn za pomocą operacji matematycznych
df = pd.DataFrame({
    'Cena': [100, 150, 200],
    'Ilość': [5, 3, 2]
})

# Dodanie nowej kolumny 'Przychod'
df['Przychod'] = df['Cena'] * df['Ilość']

df
```

Out[132…

	Cena	Ilość	Przychod
0	100	5	500
1	150	3	450
2	200	2	400

In [133…

```
# Przedstawienie na przykładzie dodawania nowych kolumn z pomocą funkcji lambda

df = pd.DataFrame({
    'Przychód': [1500, 2200, 900, 3100],
    'Koszt': [1000, 1500, 1200, 2500]
})
df['Wskaźnik_Zysku'] = df.apply(
```

```
    lambda x: (x['Przychód'] - x['Koszt']) / x['Koszt'],
    axis=1  # KLUCZOWE: Uruchamia funkcję wzdłuż wierszy
)
df
```

Out[133…

	Przychód	Koszt	Wskaźnik_Zysku
0	1500	1000	0.500000
1	2200	1500	0.466667
2	900	1200	-0.250000
3	3100	2500	0.240000

In [ ]:
```
# Przedstawienie możliwości pracy z dużymi plikami przy użyciu argumentu

big_file = 'path/to/big/file'

test_file = pd.read_csv(big_file, chunksize=100_000)
```