

SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 1 Data 28.09.2024 Temat: Wprowadzenie do narzędzi i środowiska pracy w analizie danych Wariant: 5	Mateusz Łysoń Informatyka II stopień, niestacjonarne, 1 semestr, gr. b
--	---

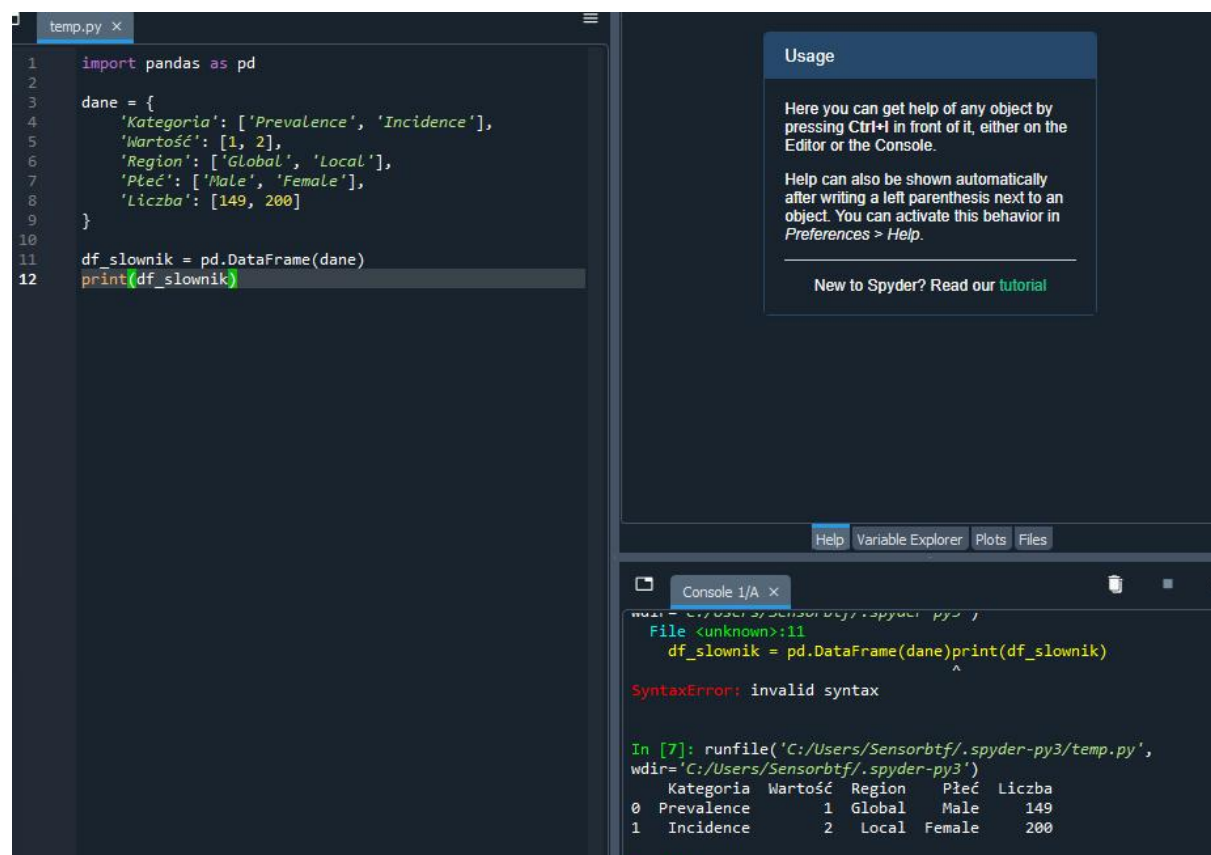
Polecenie: wykonaj polecenia, korzystając ewentualnie z danych w konkretnym wariacie zadania

Opis programu opracowanego (kody źródłowe, rzuty ekranu)

1. Ładowanie biblioteki Pandas

```
import pandas as pd
```

2. Tworzenie ramki danych ze słownika



The screenshot shows the Spyder IDE interface. The editor on the left contains the following Python code:

```
1 import pandas as pd
2
3 dane = {
4     'Kategoria': ['Prevalence', 'Incidence'],
5     'Wartość': [1, 2],
6     'Region': ['Global', 'Local'],
7     'Płeć': ['Male', 'Female'],
8     'Liczba': [149, 200]
9 }
10
11 df_slownik = pd.DataFrame(dane)
12 print(df_slownik)
```

The console on the right shows the following output:

```
File <unknown>:11
df_slownik = pd.DataFrame(dane)print(df_slownik)
^
SyntaxError: invalid syntax
```

Below the error message, the console shows the execution of the script:

```
In [7]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py',
wdir='C:/Users/Sensorbtf/.spyder-py3')
Kategoria  Wartość  Region  Płeć  Liczba
0  Prevalence      1  Global  Male    149
1  Incidence      2   Local  Female  200
```

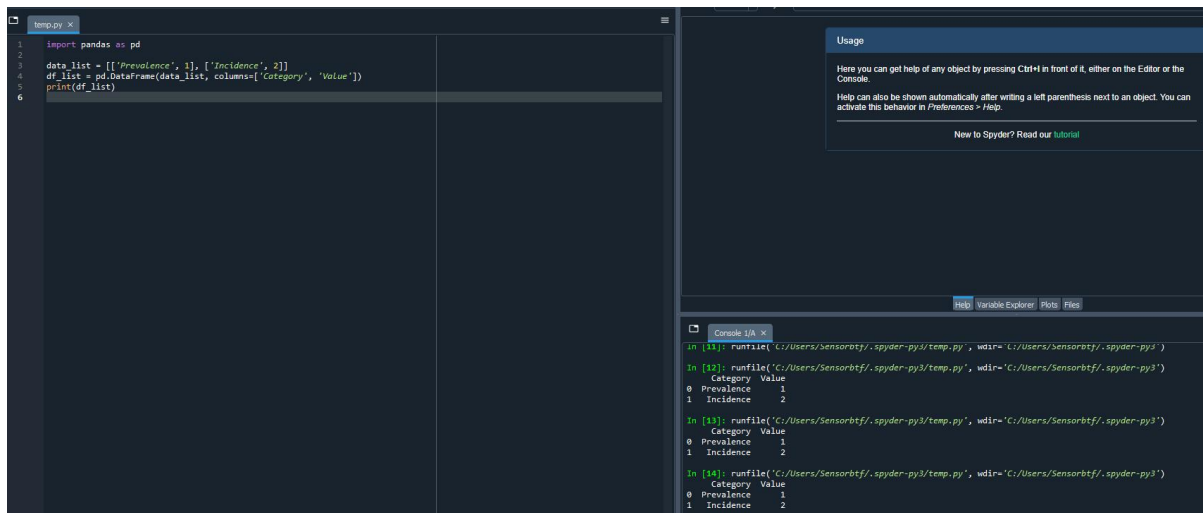
3. Zachowanie ramki danych z pliku CSV do formatu XLSX



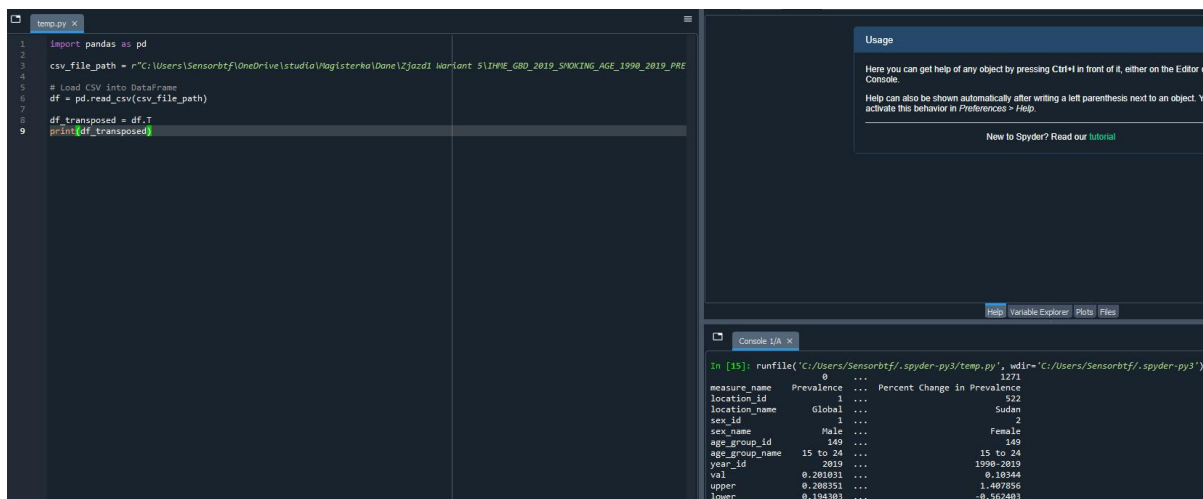
The screenshot shows the Spyder IDE editor with the following Python code:

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\IHME_GBD_2019_SMOKING_AGE_1990_201
4
5 df = pd.read_csv(csv_file_path)
6
7 excel_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\IHME_GBD_2019_SMOKING_AGE_1990_2
8
9 df.to_excel(excel_file_path, index=False)
10
11
```

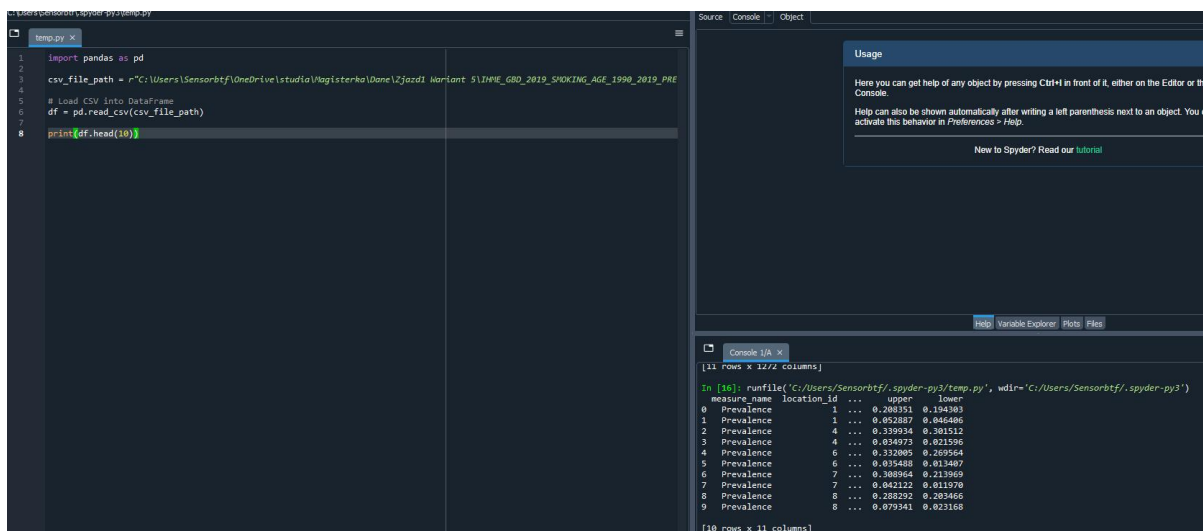
4. Tworzenie ramki danych z listy list



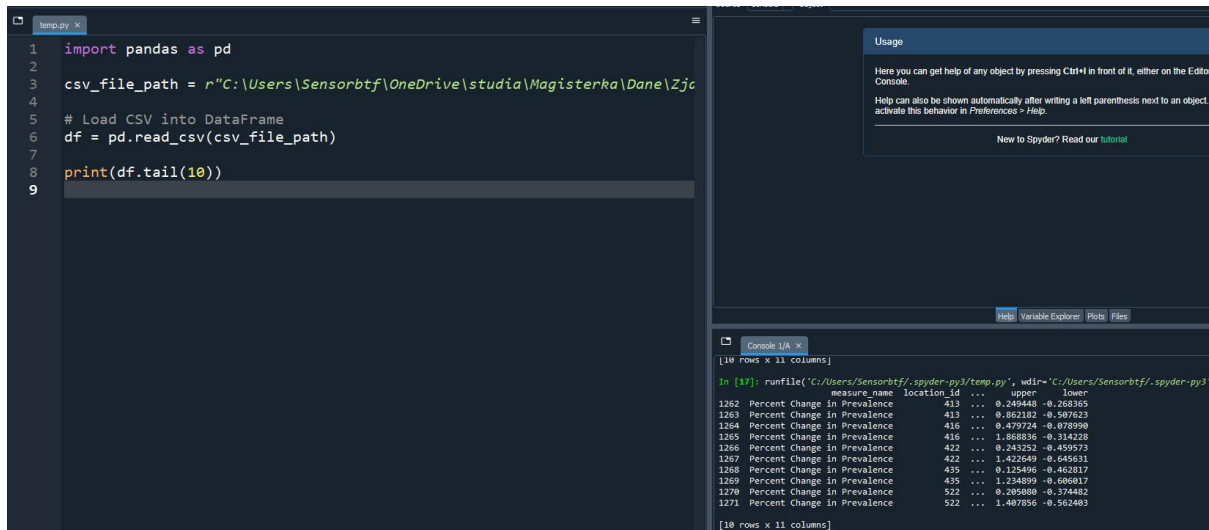
5. Transponowanie ramki danych (zamiana kolumn na wiersze)



6. Wyświetlenie pierwszych 10 wierszy



7. Wyświetlenie ostatnich 10 wierszy



The screenshot shows the Spyder IDE interface. The editor on the left contains a Python script that imports pandas as pd, defines a file path, loads a CSV file into a DataFrame, and prints the last 10 rows. The console on the right shows the output of the script, which is a table of 10 rows and 11 columns. The table has columns: measure_name, location_id, upper, and lower. The rows show data for 'Percent Change in Prevalence' at various locations.

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 # Load CSV into DataFrame
6 df = pd.read_csv(csv_file_path)
7
8 print(df.tail(10))
9
```

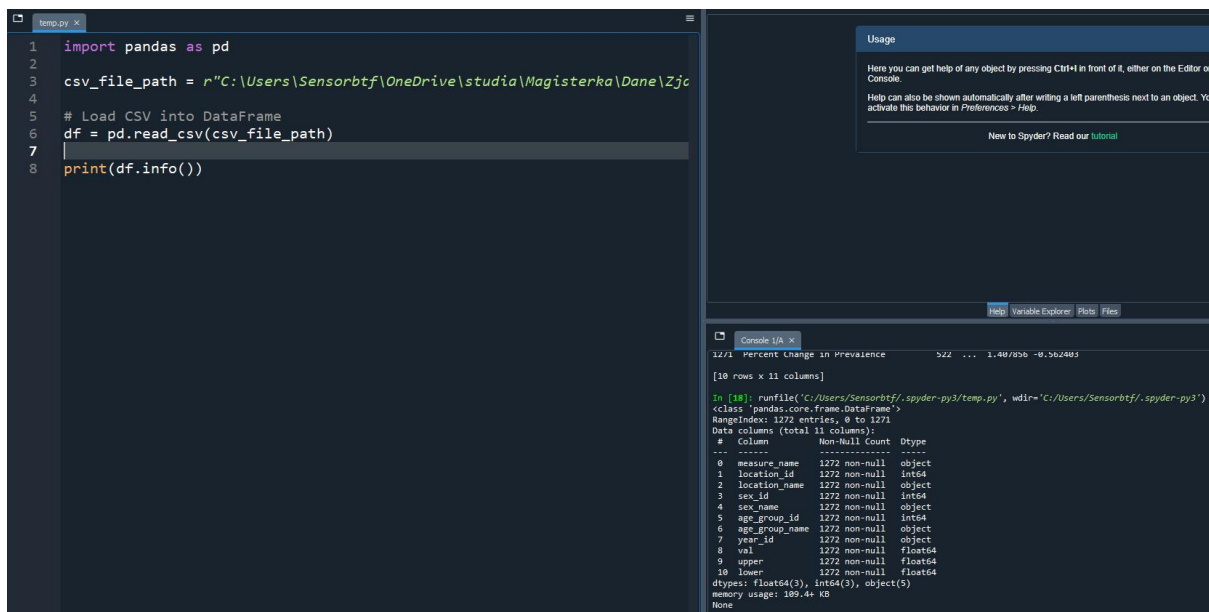
Console I/A X

[10 rows x 11 columns]

```
In [17]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
measure_name location_id ... upper lower
1262 Percent Change in Prevalence 413 ... 0.249448 -0.268365
1263 Percent Change in Prevalence 413 ... 0.862182 -0.587623
1264 Percent Change in Prevalence 416 ... 0.479724 -0.878998
1265 Percent Change in Prevalence 416 ... 1.868836 -0.314228
1266 Percent Change in Prevalence 422 ... 0.243252 -0.459573
1267 Percent Change in Prevalence 422 ... 1.422649 -0.645631
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1269 Percent Change in Prevalence 435 ... 1.234899 -0.686817
1270 Percent Change in Prevalence 522 ... 0.285888 -0.374482
1271 Percent Change in Prevalence 522 ... 1.487856 -0.562483
```

[10 rows x 11 columns]

8. Wyświetlenie informacji o ramce danych



The screenshot shows the Spyder IDE interface. The editor on the left contains a Python script that imports pandas as pd, defines a file path, loads a CSV file into a DataFrame, and prints the information about the DataFrame. The console on the right shows the output of the script, which is a detailed summary of the DataFrame, including the number of rows and columns, the data types, and the memory usage.

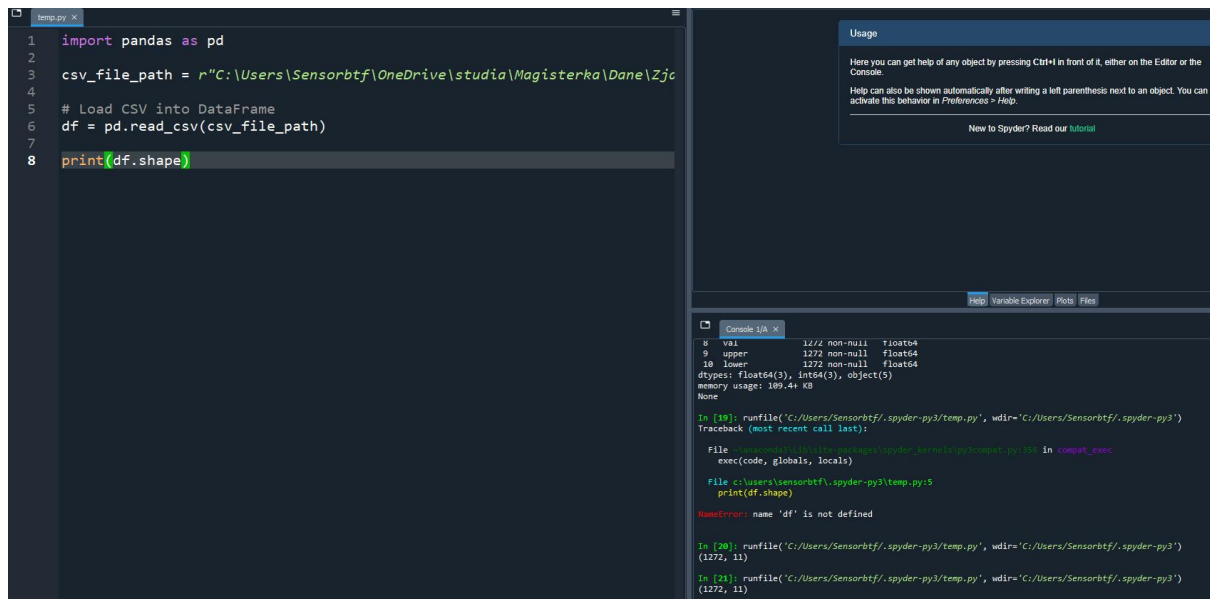
```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 # Load CSV into DataFrame
6 df = pd.read_csv(csv_file_path)
7
8 print(df.info())
```

Console I/A X

[10 rows x 11 columns]

```
In [18]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1272 entries, 0 to 1271
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   measure_name        1272 non-null   object
1   location_id         1272 non-null   int64
2   location_name       1272 non-null   object
3   sex_id              1272 non-null   int64
4   sex_name            1272 non-null   object
5   age_group_id        1272 non-null   int64
6   age_group_name      1272 non-null   object
7   year_id             1272 non-null   object
8   val                 1272 non-null   float64
9   upper               1272 non-null   float64
10  lower               1272 non-null   float64
dtypes: float64(3), int64(3), object(5)
memory usage: 169.4+ KB
None
```

9. Wyświetlenie liczby wierszy i kolumn



The screenshot shows the Spyder IDE interface. The editor on the left contains a Python script that imports pandas, defines a file path, loads a CSV file into a DataFrame, and prints its shape. The console on the right shows the output of the script, which is a tuple (1272, 11). The variable explorer on the right shows the DataFrame object.

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 # Load CSV into DataFrame
6 df = pd.read_csv(csv_file_path)
7
8 print(df.shape)
```

Console (1/1) x

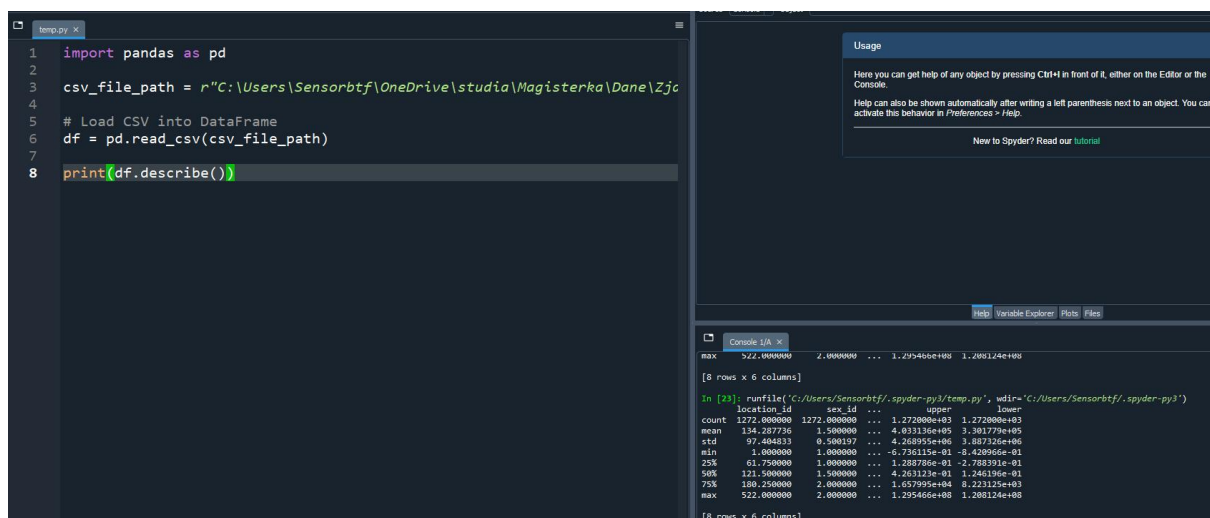
```
0 val 1272 non-null float64
9 upper 1272 non-null float64
10 lower 1272 non-null float64
dtypes: float64(3), int64(3), object(5)
memory usage: 109.4+ KB
None

In [19]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Traceback (most recent call last):
  File "C:/Users/Sensorbtf/.spyder-py3/temp.py", line 5, in <module>
    print(df.shape)
NameError: name 'df' is not defined

In [20]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
(1272, 11)

In [21]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
(1272, 11)
```

10. Wyświetlenie statystyk liczbowych



The screenshot shows the Spyder IDE interface. The editor on the left contains a Python script that imports pandas, defines a file path, loads a CSV file into a DataFrame, and prints its statistical summary. The console on the right shows the output of the script, which is a summary of the DataFrame's statistics. The variable explorer on the right shows the DataFrame object.

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 # Load CSV into DataFrame
6 df = pd.read_csv(csv_file_path)
7
8 print(df.describe())
```

Console (1/1) x

```
max 522.000000 1.000000 ... 1.295466e+08 1.288124e+08
[8 rows x 6 columns]

In [23]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
count 1272.000000 1272.000000 ... 1.272000e+03 1.272000e+03
mean 134.287736 1.500000 ... 4.033136e+05 3.301779e+05
std 97.404813 0.500197 ... 4.268955e+06 3.887326e+06
min 1.000000 1.000000 ... -5.736115e-01 -8.429566e-01
25% 61.750000 1.000000 ... 1.288786e-01 -2.788391e-01
50% 121.500000 1.500000 ... 4.263123e-01 1.246196e-01
75% 180.250000 2.000000 ... 1.657895e+04 8.221125e+03
max 522.000000 2.000000 ... 1.295466e+08 1.288124e+08
[8 rows x 6 columns]
```

11. Wyświetlenie statystyk dla kolumn kategoryzowanych

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 # Load CSV into DataFrame
6 df = pd.read_csv(csv_file_path)
7
8 print(df['age_group_id'].value_counts())
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console Jupyter x

```
File ~\anaconda3\lib\site-packages\spyder-py3\temp.py:10
print(df['age_group'].value_counts())
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:4100 in __getitem__
Indexer = self.columns.get_loc(key)
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:3812 in get_loc
raise KeyError(key) from err
KeyError: 'age_group'
```

In [28]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

```
age_group_id
149    1272
Name: count, dtype: int64
```

12. Usunięcie brakujących wartości

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 df.dropna(inplace=True)
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console Jupyter x

```
File ~\anaconda3\lib\site-packages\spyder-py3\temp.py:10
print(df['age_group'].value_counts())
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:4100 in __getitem__
Indexer = self.columns.get_loc(key)
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:3812 in get_loc
raise KeyError(key) from err
KeyError: 'age_group'
```

In [28]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

```
age_group_id
149    1272
Name: count, dtype: int64
```

In [29]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

In [30]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

13. Wybór wierszy i kolumn po nazwach oraz indeksach

```
C:\Users\Sensorbtf\.spyder-py3\temp.py
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df.loc[0:5, ['measure_name', 'Location_id']])
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console Jupyter x

```
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:3812 in get_loc
raise KeyError(key) from err
KeyError: 'age_group'
```

In [28]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

```
age_group_id
149    1272
Name: count, dtype: int64
```

In [29]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

In [30]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

In [31]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

```
measure_name location_id
0  Prevalence          1
1  Prevalence          1
2  Prevalence          4
3  Prevalence          4
4  Prevalence          6
5  Prevalence          6
```

14. Wybór wierszy z warunkiem na wartość kolumny

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df[df['Location_id'] > 1])
```

Usage

Here you can get help of any object by pressing Ctrl+H in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in [Preferences > Help](#).

[New to Spyder? Read our tutorial](#)

Help Variable Explorer Plots Files

```
Console I/A x
In [32]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
measure_name location_id ... upper lower
2 Prevalence 4 ... 0.339934 0.301512
3 Prevalence 4 ... 0.034973 0.021596
4 Prevalence 6 ... 0.332005 0.269564
5 Prevalence 6 ... 0.635408 0.033407
6 Prevalence 7 ... 0.308964 0.213969
...
1267 Percent Change in Prevalence 422 ... 1.422649 -0.646531
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1269 Percent Change in Prevalence 435 ... 1.234899 -0.606817
1270 Percent Change in Prevalence 522 ... 0.205808 -0.374482
1271 Percent Change in Prevalence 522 ... 1.407856 -0.562483

[1266 rows x 11 columns]
```

15. Wybór wierszy na podstawie kilku warunków

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df[(df['Location_id'] > 1) & (df['sex_name'] == 'Male')])
```

Usage

Here you can get help of any object by pressing Ctrl+H in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in [Preferences > Help](#).

[New to Spyder? Read our tutorial](#)

Help Variable Explorer Plots Files

```
Console I/A x
...
1267 Percent Change in Prevalence 422 ... 1.422649 -0.646531
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1269 Percent Change in Prevalence 435 ... 1.234899 -0.606817
1270 Percent Change in Prevalence 522 ... 0.205808 -0.374482
1271 Percent Change in Prevalence 522 ... 1.407856 -0.562483

[1266 rows x 11 columns]

In [33]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
measure_name location_id ... upper lower
2 Prevalence 4 ... 0.339934 0.301512
4 Prevalence 6 ... 0.332005 0.269564
6 Prevalence 7 ... 0.308964 0.213969
8 Prevalence 8 ... 0.288292 0.203466
10 Prevalence 10 ... 0.178902 0.122546
...
1262 Percent Change in Prevalence 413 ... 0.249448 -0.268365
1264 Percent Change in Prevalence 416 ... 0.479724 -0.078999
1266 Percent Change in Prevalence 422 ... 0.243252 -0.459573
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1270 Percent Change in Prevalence 522 ... 0.205808 -0.374482

[633 rows x 11 columns]
```

16. Wybór wierszy zawierających określone słowo


```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df[df['measure_name'].str.contains('Prevalence')])
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console I/A x

```
... ..
1267 Percent Change in Prevalence 422 ... 1.422649 -0.645631
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1269 Percent Change in Prevalence 435 ... 1.234899 -0.606017
1270 Percent Change in Prevalence 522 ... 0.205800 -0.374482
1271 Percent Change in Prevalence 522 ... 1.407856 -0.562403

[848 rows x 11 columns]

In [36]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
measure_name location_id ... upper lower
0 Prevalence 1 ... 0.268351 0.194303
1 Prevalence 1 ... 0.403207 0.364686
2 Prevalence 4 ... 0.339934 0.301512
3 Prevalence 4 ... 0.634973 0.621596
4 Prevalence 6 ... 0.332005 0.209564
... ..
1267 Percent Change in Prevalence 422 ... 1.422649 -0.645631
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1269 Percent Change in Prevalence 435 ... 1.234899 -0.606017
1270 Percent Change in Prevalence 522 ... 0.205800 -0.374482
1271 Percent Change in Prevalence 522 ... 1.407856 -0.562403

[848 rows x 11 columns]
```

17. Wybór wierszy nie zawierających określonego słowa

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df[~df['measure_name'].str.contains('Prevalence')])
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console I/A x

```
... ..
1267 Percent Change in Prevalence 422 ... 1.422649 -0.645631
1268 Percent Change in Prevalence 435 ... 0.125496 -0.462817
1269 Percent Change in Prevalence 435 ... 1.234899 -0.606017
1270 Percent Change in Prevalence 522 ... 0.205800 -0.374482
1271 Percent Change in Prevalence 522 ... 1.407856 -0.562403

[848 rows x 11 columns]

In [32]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
measure_name location_id ... upper lower
424 Number of Smokers 1 ... 1.295466e+08 1.208124e+08
425 Number of Smokers 1 ... 3.158243e+07 2.771202e+07
426 Number of Smokers 4 ... 4.988433e+07 4.353696e+07
427 Number of Smokers 4 ... 4.678339e+06 2.030817e+06
428 Number of Smokers 6 ... 2.751081e+07 2.233678e+07
... ..
843 Number of Smokers 422 ... 2.540378e+02 6.559104e+01
844 Number of Smokers 435 ... 1.361217e+05 7.792683e+04
845 Number of Smokers 435 ... 2.867286e+04 8.096800e+03
846 Number of Smokers 522 ... 7.153351e+05 4.748355e+05
847 Number of Smokers 522 ... 1.464557e+05 4.327179e+04

[424 rows x 11 columns]
```

18. Utworzenie nowej kolumny na podstawie istniejących


```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 df['Nowa_kolumna'] = df['Location_id'] * 2
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console 1/A x

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

```
File ~\anaconda3\lib\site-packages\spyder_kernels\ipythonconsole.py:126 in _run_cell_execute
    exec(code, globals, locals)
File ~\Users\Sensorbtf\spyder-py3\temp.py:7
    df['Nowa_kolumna'] = df['Liczba'] * 2
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:4100 in _getitem_i
    indexer = self.columns.get_loc(key)
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:3812 in get_loc
    raise KeyError(key) from err
KeyError: 'Liczba'
```

In [39]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
In [40]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

19. Usunięcie kolumny

```
temp.py* x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 df.drop('Nowa_kolumna', axis=1, inplace=True)
```

Usage

Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console 1/A x

```
In [39]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
In [40]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
In [41]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Traceback (most recent call last):
File ~\anaconda3\lib\site-packages\spyder_kernels\ipythonconsole.py:126 in _run_cell_execute
    exec(code, globals, locals)
File ~\Users\Sensorbtf\spyder-py3\temp.py:7
    df.drop('Nowa_kolumna', axis=1, inplace=True)
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:5571 in drop
    return super().drop(
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:4902 in drop
    obj = obj._drop_axis(labels, axis, level=level, errors=errors)
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:4530 in _drop_axis
    new_axis = axis.drop(labels, errors=errors)
File ~\anaconda3\lib\site-packages\pandas\core\indexes\base.py:7070 in drop
    raise KeyError("{}[labels[mask].tolist()] not found in axis")
KeyError: "['Liczba'] not found in axis"
```

20. Zmiana nazwy kolumny

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 df.rename(columns={'measure_name': 'measure_names'}, inplace=True)
```

Console I/A x

```
KeyError: '[ Nowa_kolumna ] not found in axis'
```

In [42]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

In [43]:

21. Zapis ramki danych jako plik CSV

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjc
4
5 df = pd.read_csv(csv_file_path)
6
7 df.to_csv('plik_output.csv', index=False)
```

Console I/A x

```
In [43]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
```

In [44]:

22. Wyświetlenie średniej, maksymalnej i minimalnej wartości z kolumny

```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df['val'].mean())
8 print(df['val'].max())
9 print(df['val'].min())
```

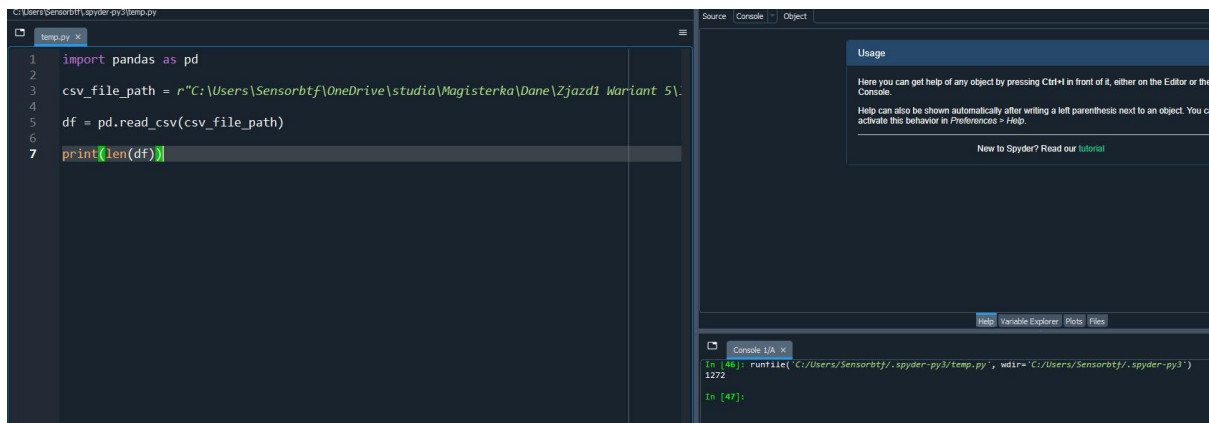
Console I/A x

```
SyntaxError: invalid syntax
```

In [45]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')

```
364533.3700294625
124995683.5
-8.758232309
```

23. Wyświetlenie liczby wierszy

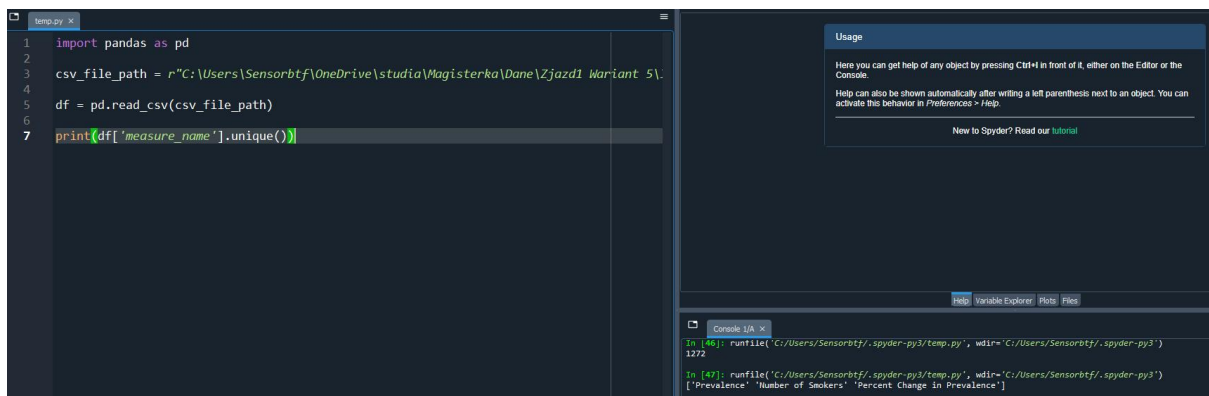


```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1_Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 print(len(df))
```

Console I/A x

```
In [46]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
1272
In [47]:
```

24. Wyświetlenie unikalnych wartości w kolumnie

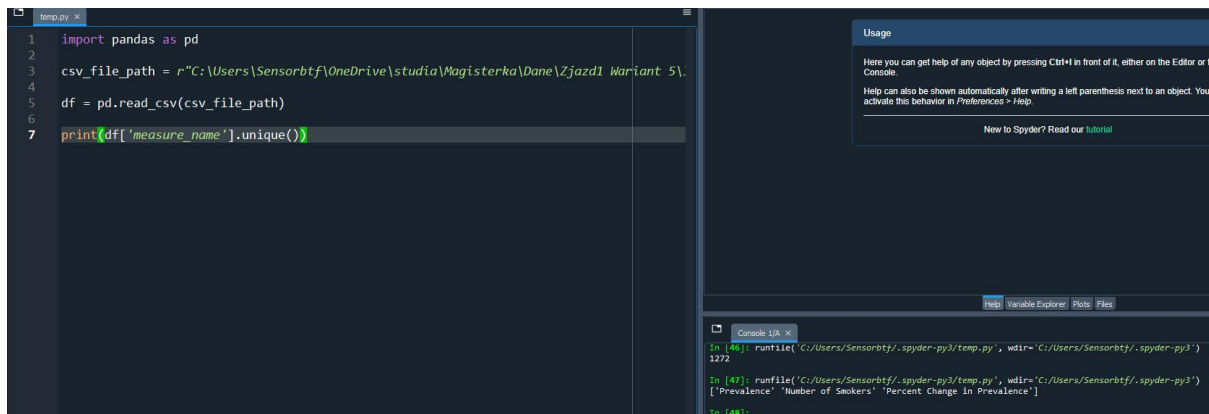


```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1_Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df['measure_name'].unique())
```

Console I/A x

```
In [46]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
1272
In [47]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
['Prevalence' 'Number of Smokers' 'Percent Change in Prevalence']
```

25. Wyświetlenie liczby rekordów dla wartości w kolumnie



```
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1_Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df['measure_name'].unique())
```

Console I/A x

```
In [46]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
1272
In [47]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
['Prevalence' 'Number of Smokers' 'Percent Change in Prevalence']
In [48]:
```

26. Sortowanie wierszy według wartości kolumny

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 df.sort_values(by='lower', ascending=False, inplace=True)
```

27. Wyświetlenie 10 największych/najmniejszych wartości

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 print(df.nlargest(10, 'val'))
8 print(df.nsmallest(10, 'val'))
```

Usage

Here you can get help of any object by pressing Ctrl+H in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in [Preferences > Help](#).

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

```
In [31]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
measure_name location_id ... upper lower
424 Number of Smokers 1 ... 1.295466e+08 1.208124e+08
426 Number of Smokers 4 ... 4.908463e+07 4.353096e+07
425 Number of Smokers 1 ... 3.158243e+07 2.771202e+07
428 Number of Smokers 6 ... 2.751081e+07 2.233678e+07
782 Number of Smokers 158 ... 2.768826e+07 2.157657e+07
719 Number of Smokers 163 ... 2.048377e+07 1.436712e+07
534 Number of Smokers 64 ... 1.652767e+07 1.520471e+07
662 Number of Smokers 137 ... 1.397654e+07 1.253966e+07
535 Number of Smokers 64 ... 1.151609e+07 9.299330e+06
716 Number of Smokers 166 ... 1.081519e+07 9.863906e+06

[10 rows x 11 columns]
measure_name location_id ... upper lower
1083 Percent Change in Prevalence 135 ... -0.654538 -0.842097
1082 Percent Change in Prevalence 135 ... -0.673611 -0.787495
1095 Percent Change in Prevalence 90 ... -0.504751 -0.755938
969 Percent Change in Prevalence 71 ... -0.430627 -0.726910
1230 Percent Change in Prevalence 210 ... -0.445845 -0.688156
921 Percent Change in Prevalence 83 ... -0.361954 -0.725385
1027 Percent Change in Prevalence 103 ... -0.482277 -0.644412
1226 Percent Change in Prevalence 214 ... -0.398855 -0.693081
1064 Percent Change in Prevalence 125 ... -0.433190 -0.668080
1004 Percent Change in Prevalence 90 ... -0.468321 -0.647364
```

27. Grupowanie wierszy według kolumny

```
temp.py x
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\
4
5 df = pd.read_csv(csv_file_path)
6
7 df_grupowane = df.groupby('lower').mean()
8 print(df_grupowane)
```

Usage

Here you can get help of any object by pressing Ctrl+H in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in [Preferences > Help](#).

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

```
Traceback (most recent call last):
  File ~\anaconda3\lib\site-packages\spyder\internal\pycompat.py:39 in compat_exec
    exec(code, globals, locals)
  File ~\anaconda3\lib\site-packages\spyder\internal\pycompat.py:39 in compat_exec
    exec(code, globals, locals)
  File ~\anaconda3\lib\site-packages\pandas\core\groupby.py:1462 in mean
    result = self._cython_agg_general(
  File ~\anaconda3\lib\site-packages\pandas\core\groupby.py:1500 in _cython_agg_general
    new_mgr = data.grouped_reduce(array_func)
  File ~\anaconda3\lib\site-packages\pandas\core\groupby.py:1460 in grouped_reduce
    applied = sb.apply(func)
  File ~\anaconda3\lib\site-packages\pandas\core\internals\blocks.py:161 in apply
    result = func(self.values, **kwargs)
  File ~\anaconda3\lib\site-packages\pandas\core\internals\blocks.py:161 in apply
    result = self._agg_py_fallback(how, values, ndim=data.ndim, alt=alt)
  File ~\anaconda3\lib\site-packages\pandas\core\groupby.py:1946 in _agg_py_fallback
    raise TypeError(msg) from err
```

29. Tworzenie tabeli przestawnej

```
temp.py X
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\IHME_GBD_2019_SMOKING_AGE_
4
5 df = pd.read_csv(csv_file_path)
6
7 pivot_table = pd.pivot_table(df, values='Lower', index=['sex_id'], columns=['sex_name'], aggfunc='mean')
8 print(pivot_table)
```

Usage

Here you can get help of any object by pressing Ctrl+H either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console I/A X

```
In [56]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='
py3')
sex_name      Female      Male
sex_id
1             NaN  544177.007868
2      116178.705005      NaN

In [57]:
```

30. Wyświetlenie indeksów i kolumn tabeli przestawnej

```
temp.py X
1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 W
4
5 df = pd.read_csv(csv_file_path)
6
7 pivot_table = pd.pivot_table(df, values='Lower', index=['sex_id'], columns=['
8 print(pivot_table.index)
9 print(pivot_table.columns)
```

Usage

Here you can get help of any object by pressing Ctrl+H in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

Help Variable Explorer Plots Files

Console I/A X

```
sex_name      Female      Male
sex_id
1             NaN  544177.007868
2      116178.705005      NaN

In [57]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Index([1, 2], dtype='int64', name='sex_id')
Index(['Female', 'Male'], dtype='object', name='sex_name')

In [58]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Index([1, 2], dtype='int64', name='sex_id')
Index(['Female', 'Male'], dtype='object', name='sex_name')

In [59]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Index([1, 2], dtype='int64', name='sex_id')
Index(['Female', 'Male'], dtype='object', name='sex_name')

In [60]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Index([1, 2], dtype='int64', name='sex_id')
Index(['Female', 'Male'], dtype='object', name='sex_name')

In [61]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
Index([1, 2], dtype='int64', name='sex_id')
Index(['Female', 'Male'], dtype='object', name='sex_name')
```

31. Importowanie modułu pyplot z biblioteki matplotlib

```
temp.py X
1 import matplotlib.pyplot as plt
```

Usage

Here you can get help of any object by pressing Ctrl+H in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our [tutorial](#)

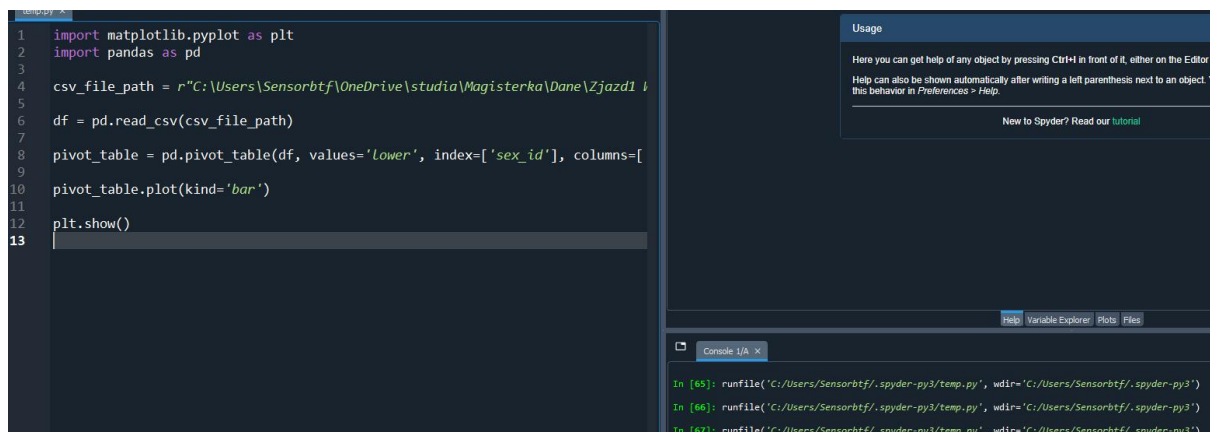
Help Variable Explorer Plots Files

Console I/A X

```
Index([1, 2], dtype='int64', name='sex_id')
Index(['Female', 'Male'], dtype='object', name='sex_name')

In [62]: runfile('C:/Users/Sensorbtf/.spyder-py3/temp.py', wdir='C:/Users/Sensorbtf/.spyder-py3')
```

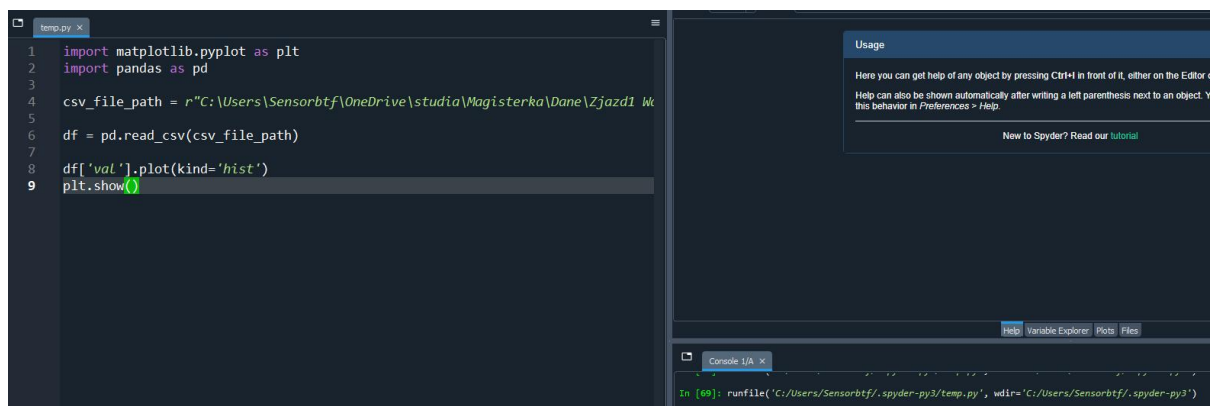
32. Rysowanie wykresu na podstawie tabeli przestawnej



The screenshot shows the Spyder IDE interface. The editor window displays a Python script that imports matplotlib.pyplot as plt and pandas as pd. It defines a CSV file path, reads the data into a DataFrame, creates a pivot table with 'sex_id' as the index and 'Lower' as the values, and then plots the pivot table as a bar chart using plt.plot(kind='bar'). The console window at the bottom shows the execution of the script, with the output of the plot function being displayed.

```
1 import matplotlib.pyplot as plt
2 import pandas as pd
3
4 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 k
5
6 df = pd.read_csv(csv_file_path)
7
8 pivot_table = pd.pivot_table(df, values='Lower', index=['sex_id'], columns=[
9
10 pivot_table.plot(kind='bar')
11
12 plt.show()
13
```

33. Rysowanie histogramu



The screenshot shows the Spyder IDE interface. The editor window displays a Python script that imports matplotlib.pyplot as plt and pandas as pd. It defines a CSV file path, reads the data into a DataFrame, and then plots a histogram of the 'val' column using df['val'].plot(kind='hist'). The console window at the bottom shows the execution of the script, with the output of the plot function being displayed.

```
1 import matplotlib.pyplot as plt
2 import pandas as pd
3
4 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 k
5
6 df = pd.read_csv(csv_file_path)
7
8 df['val'].plot(kind='hist')
9 plt.show()
```

34. Łączenie ramek danych za pomocą merge i concat


```

1 import matplotlib.pyplot as plt
2 import pandas as pd
3
4 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1\Wartości.csv"
5
6 df = pd.read_csv(csv_file_path)
7
8 df1 = pd.DataFrame({'ID': [1, 2, 3], 'Wartość': [10, 20, 30]})
9 df2 = pd.DataFrame({'ID': [1, 2, 4], 'Nowa_wartość': [100, 200, 300]})
10
11 df_merged = pd.merge(df1, df2, on='ID', how='inner')
12 print(df_merged)
13
14 df_concat = pd.concat([df1, df2], axis=0)
15 print(df_concat)

```

Console 1/A X

```

In [69]: runfile('C:/Users/Sensorbtf/OneDrive/...')
In [70]: runfile('C:/Users/Sensorbtf/OneDrive/...')
ID Wartość Nowa_wartość
0 1 10 100
1 2 20 200
ID Wartość Nowa_wartość
0 1 10.0 NaN
1 2 20.0 NaN
2 3 30.0 NaN
0 1 NaN 100.0
1 2 NaN 200.0
2 4 NaN 300.0

```

35. Dodawanie nowych kolumn za pomocą operacji matematycznych

```

1 import pandas as pd
2
3 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1\Wartości.csv"
4 df = pd.read_csv(csv_file_path)
5
6 df['Lower'] = df['Lower'] + 10

```

Console

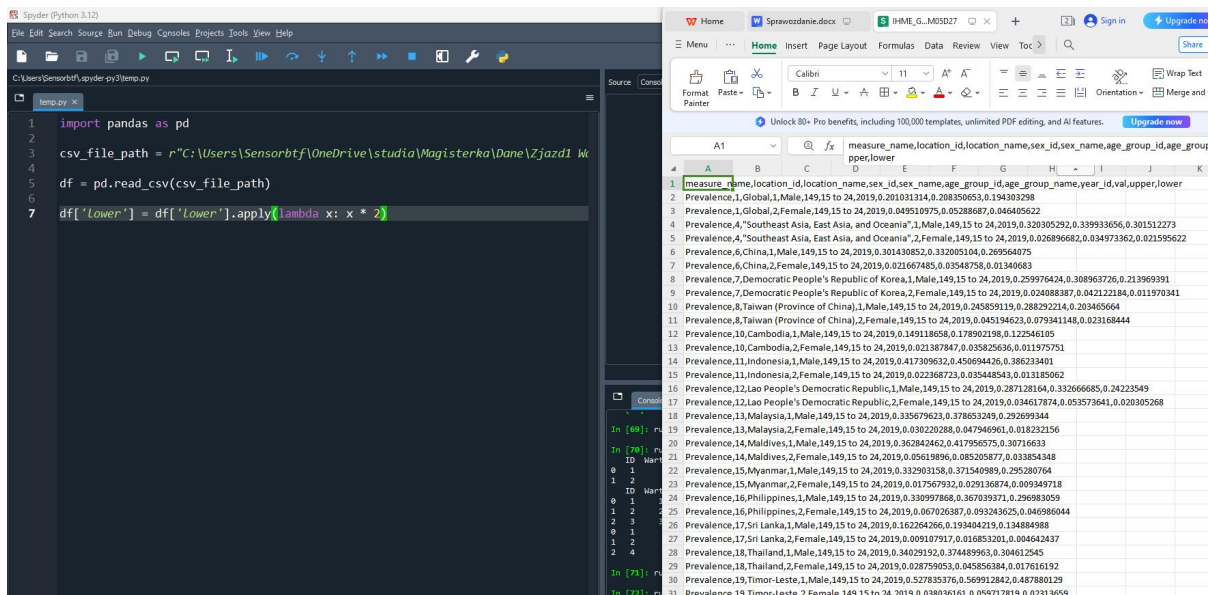
```

In [69]: runfile('C:/Users/Sensorbtf/OneDrive/...')
In [70]: runfile('C:/Users/Sensorbtf/OneDrive/...')
In [71]: runfile('C:/Users/Sensorbtf/OneDrive/...')
In [72]: runfile('C:/Users/Sensorbtf/OneDrive/...')

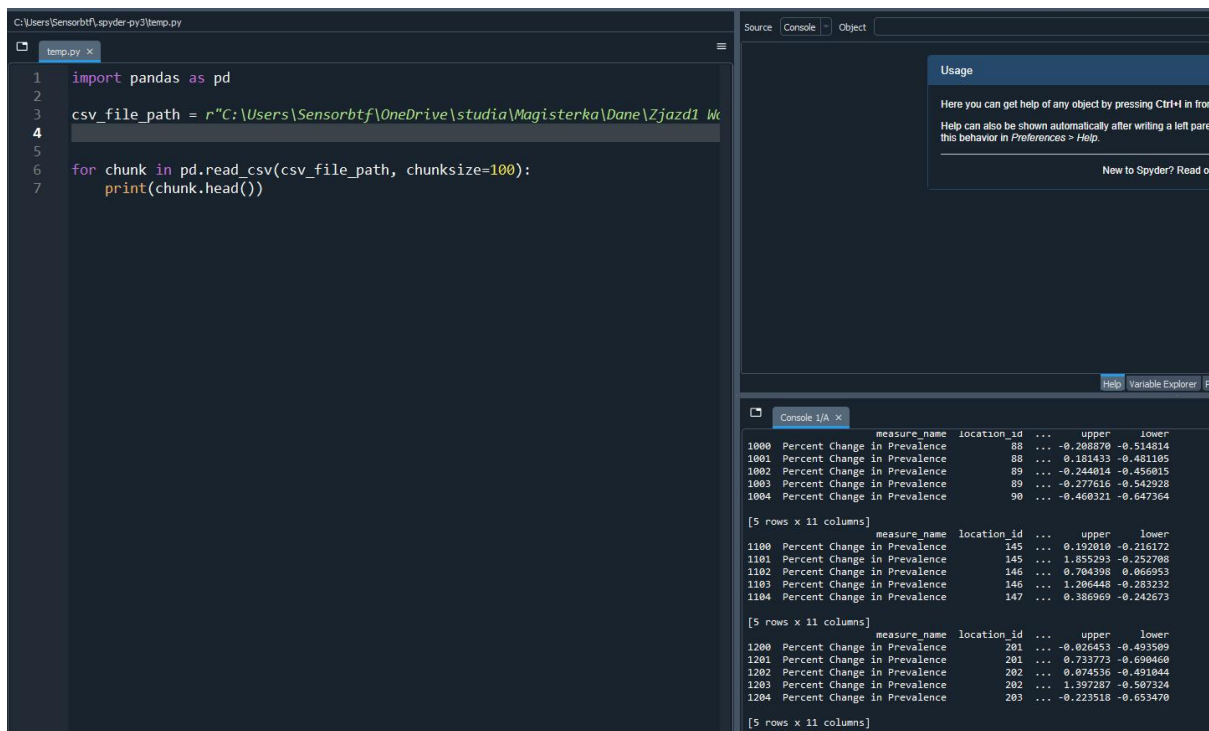
```

Worksheet showing a table of COVID-19 prevalence data with columns: measure_name, location_id, location_name, sex_id, sex_name, age_group_id, age_group_name, year_id, val_upper, val_lower.

36. Dodawanie kolumny za pomocą funkcji lambda



37. Praca z dużymi plikami przy użyciu chunksize



Wnioski

Wykorzystując podstawowe komendy w języku Python, możemy wyciągać, modyfikować i wyświetlać dane, a dodatkowo tworzyć i modyfikować tabele.

Github: <https://github.com/sensorbtf/Nauka-O-Danych/tree/main/Wprowadzenie%20do%20narz%C4%99dzi%20i%20%C5%9Brodowiska%20pracy%20w%20analizie%20danych>