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

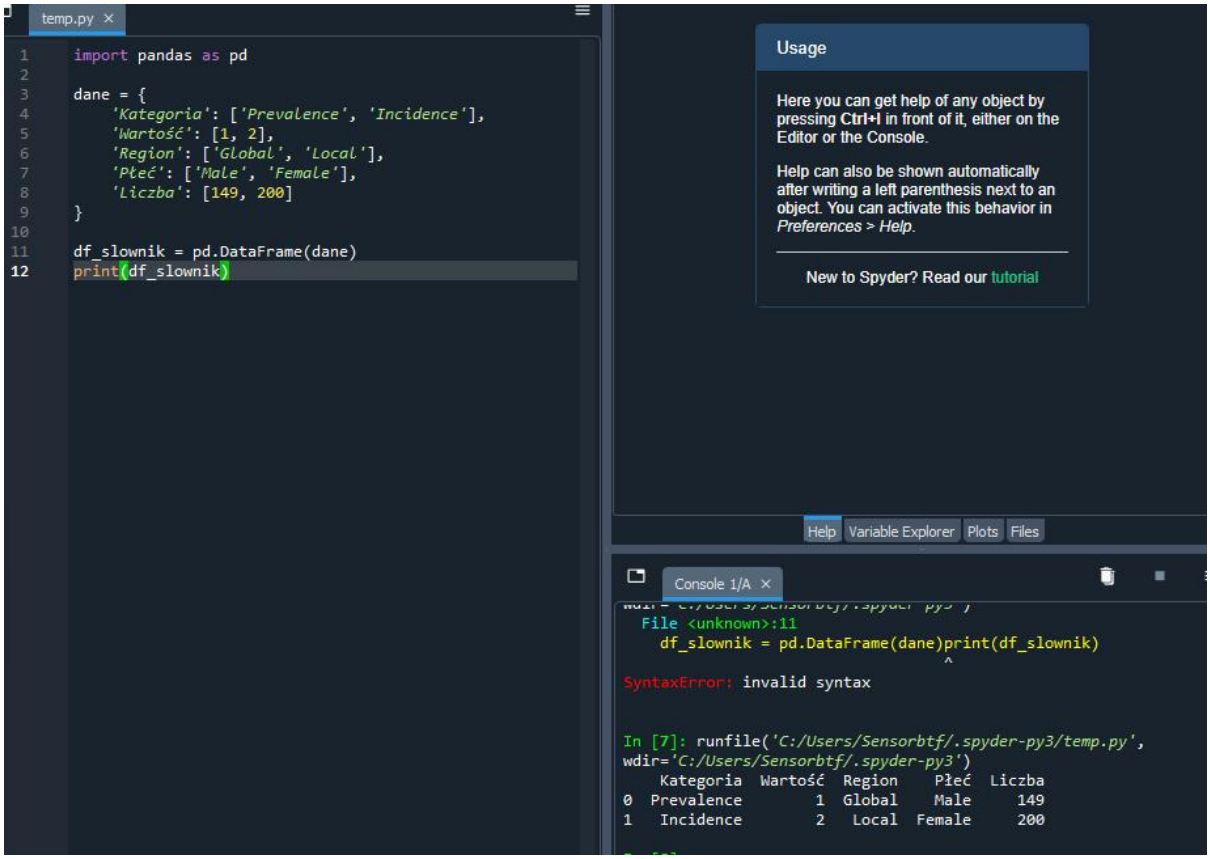
Polecenie: wykonaj polecenia, korzystając ewentualnie z danych w konkretnym wariancie zadania

Opis programu opracowanego (kody źródłowe, rzuty ekranu)

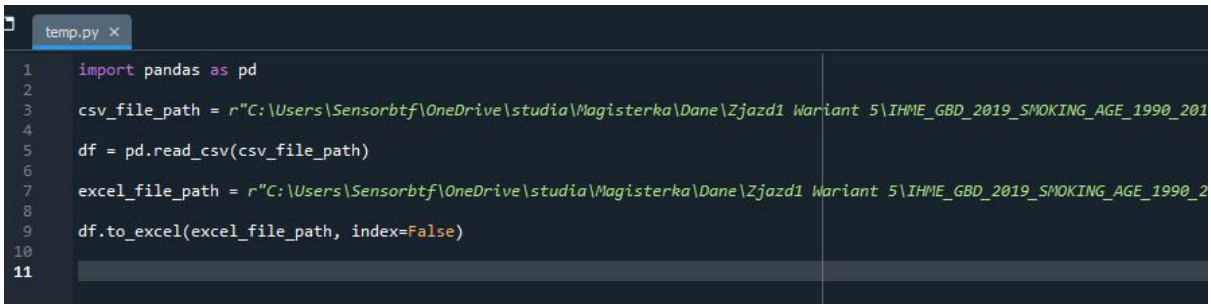
1. Ładowanie biblioteki Pandas

```
import pandas as pd
```

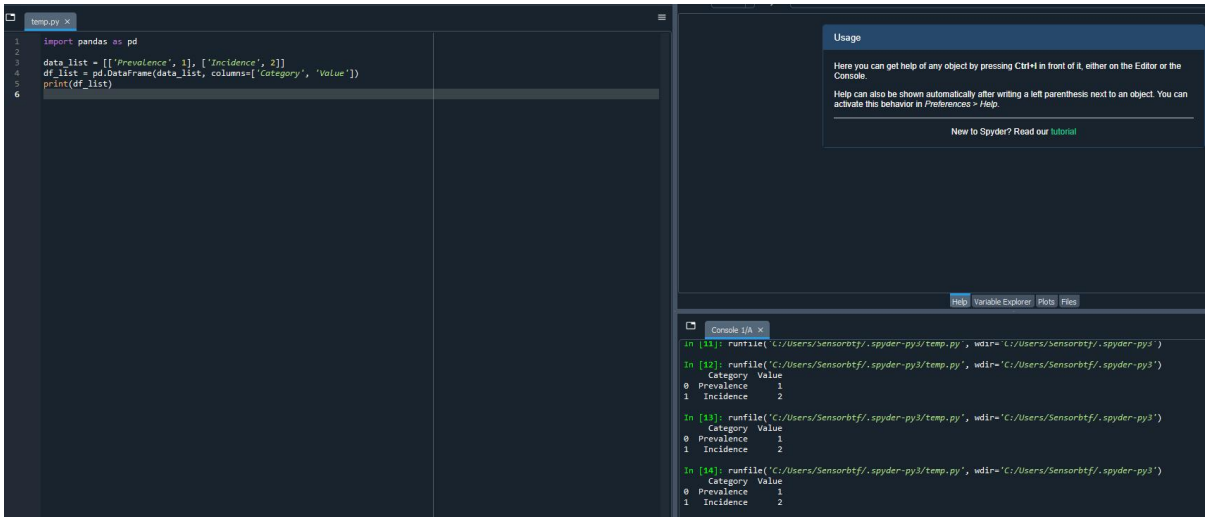
2. Tworzenie ramki danych ze słownika



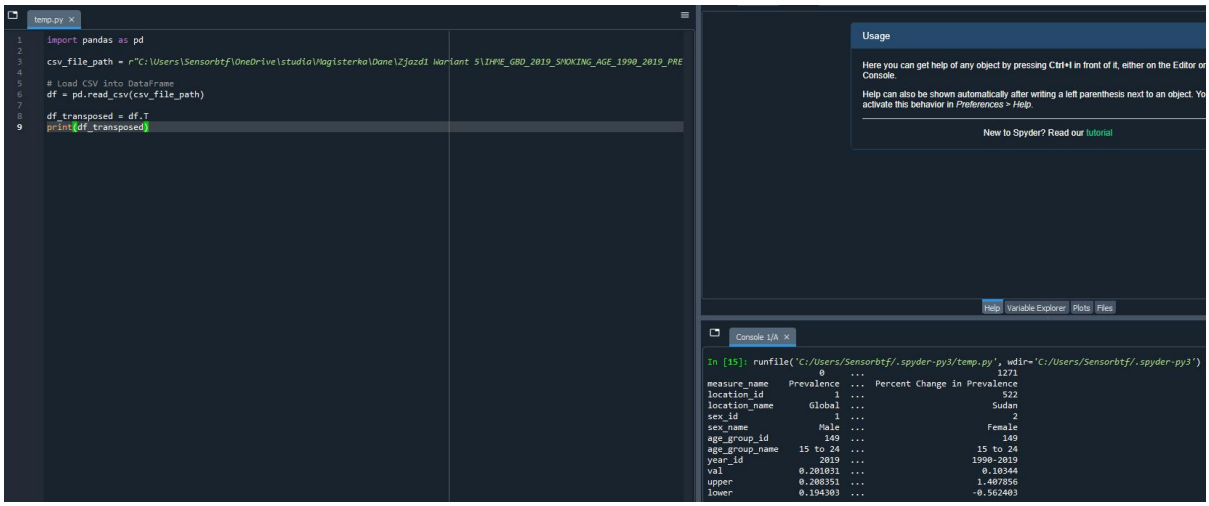
3. Zachowanie ramki danych z pliku CSV do formatu XLSX



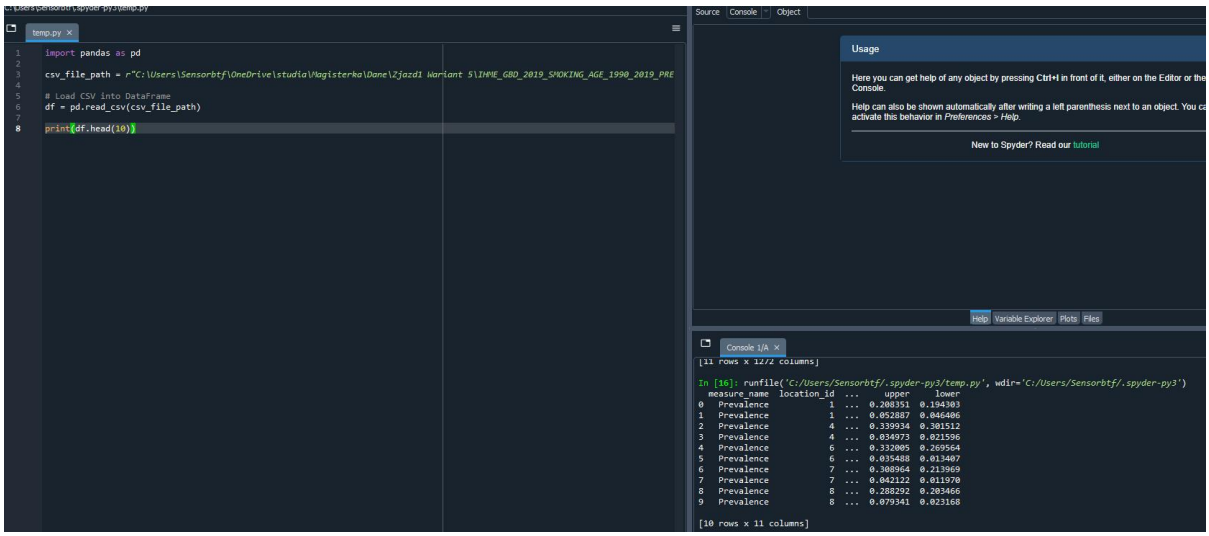
4. Tworzenie ramki danych z listy list



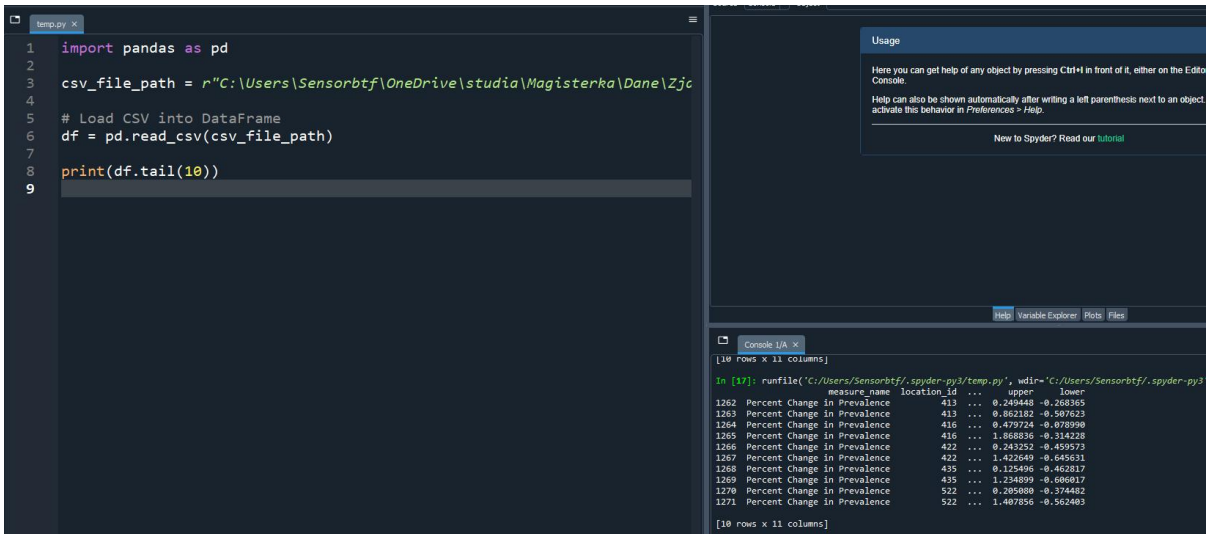
5. Transponowanie ramki danych (zamiana kolumn na wiersze)



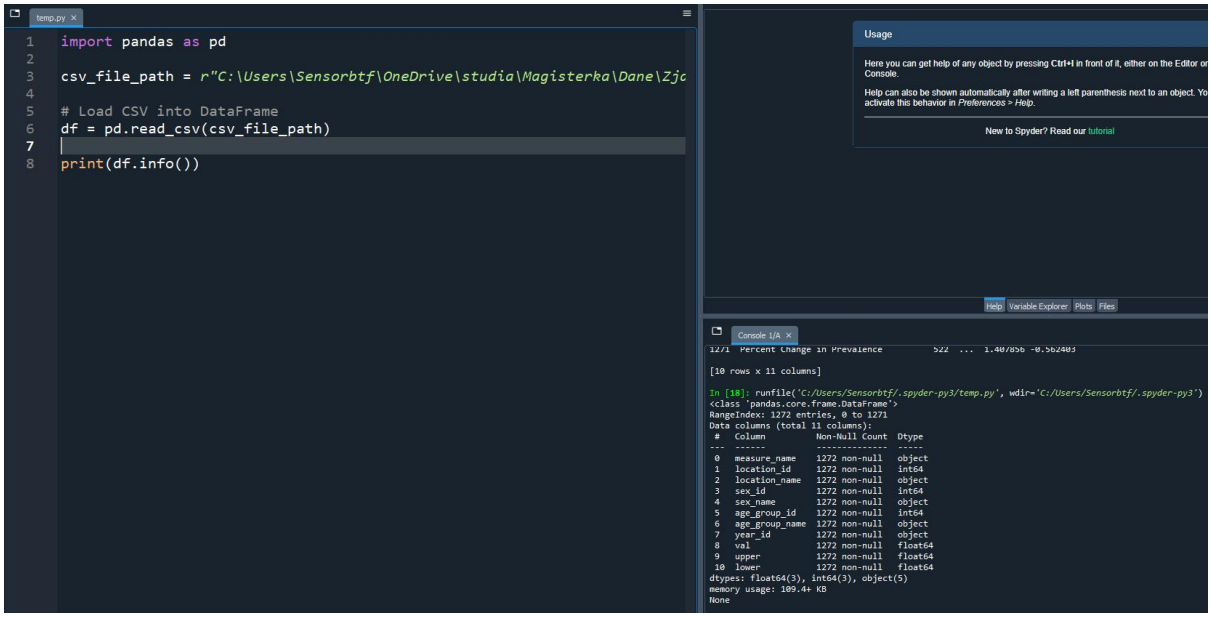
6. Wyświetlenie pierwszych 10 wierszy



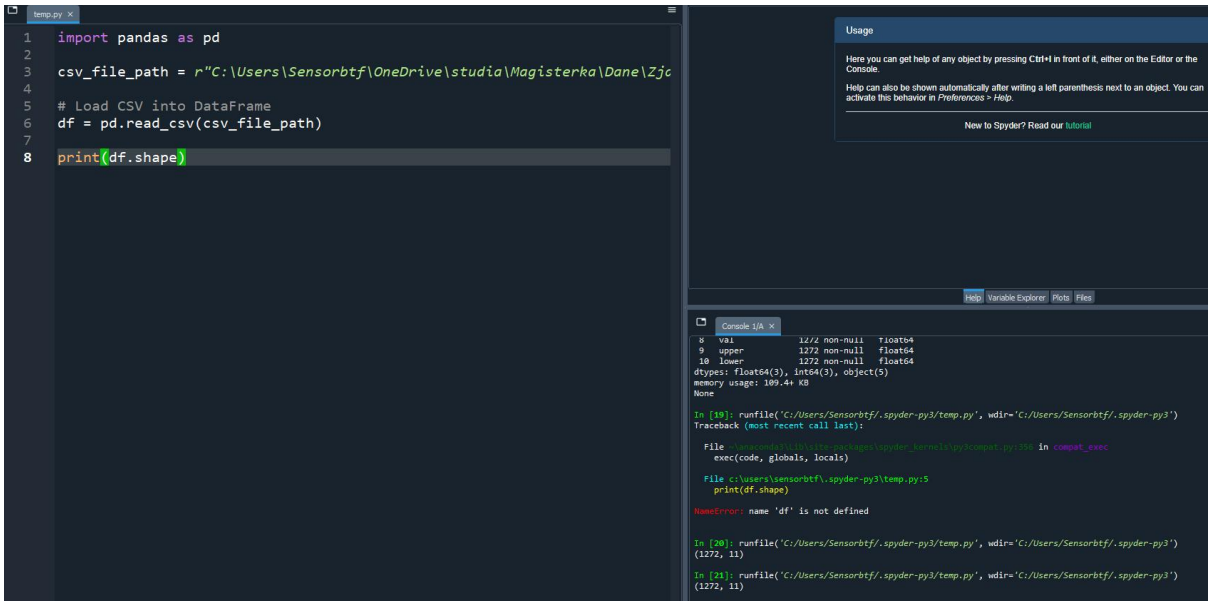
7. Wyświetlenie ostatnich 10 wierszy



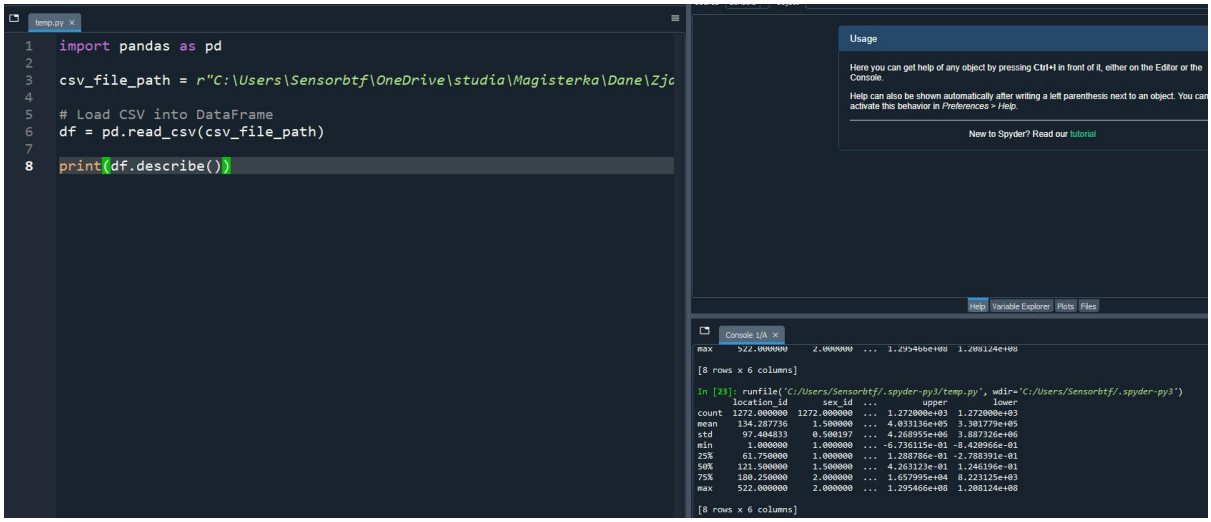
8. Wyświetlenie informacji o ramce danych



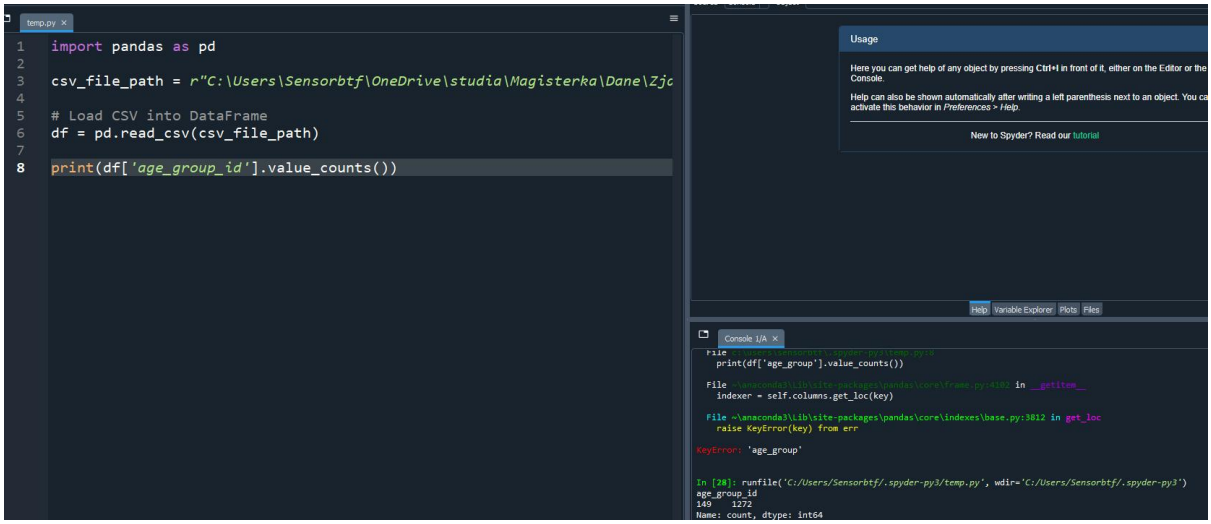
9. Wyświetlenie liczby wierszy i kolumn



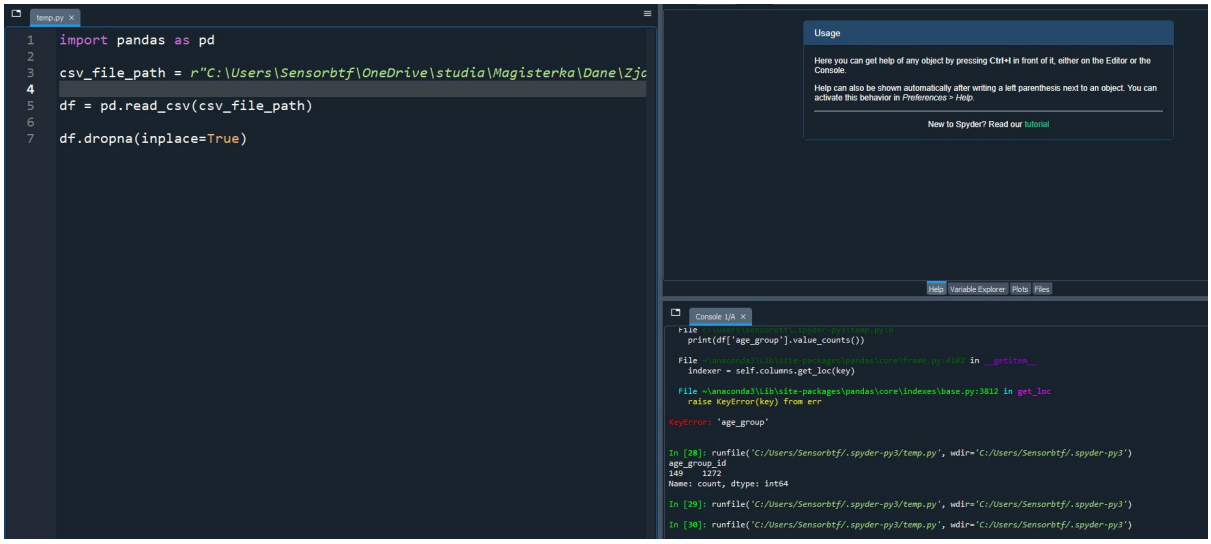
10. Wyświetlenie statystyk liczbowych



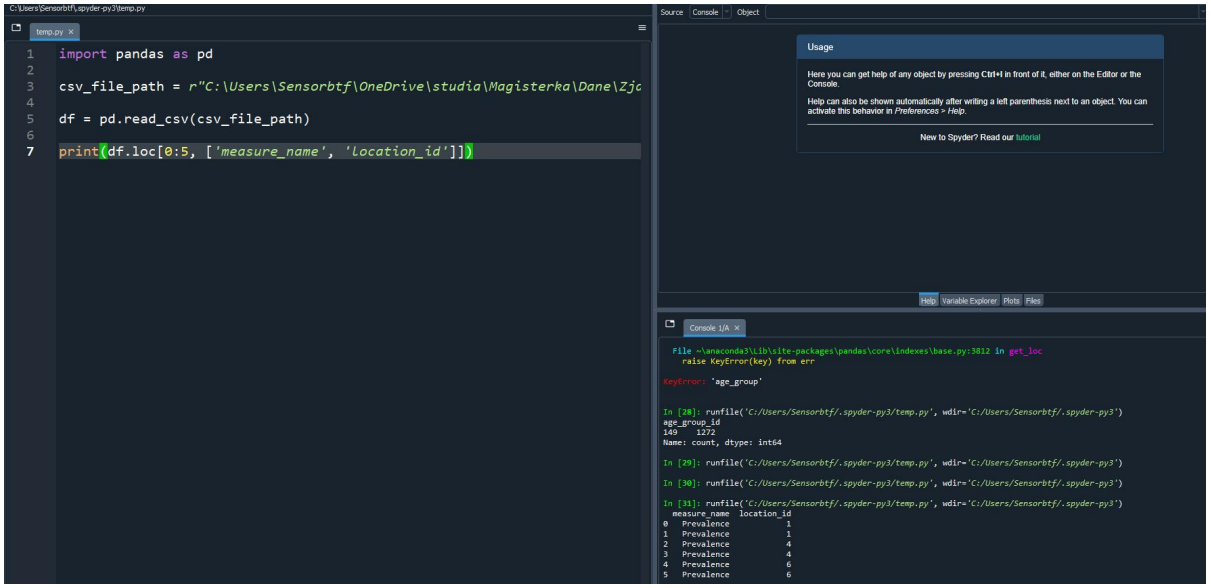
11. Wyświetlenie statystyk dla kolumn kategoryzowanych



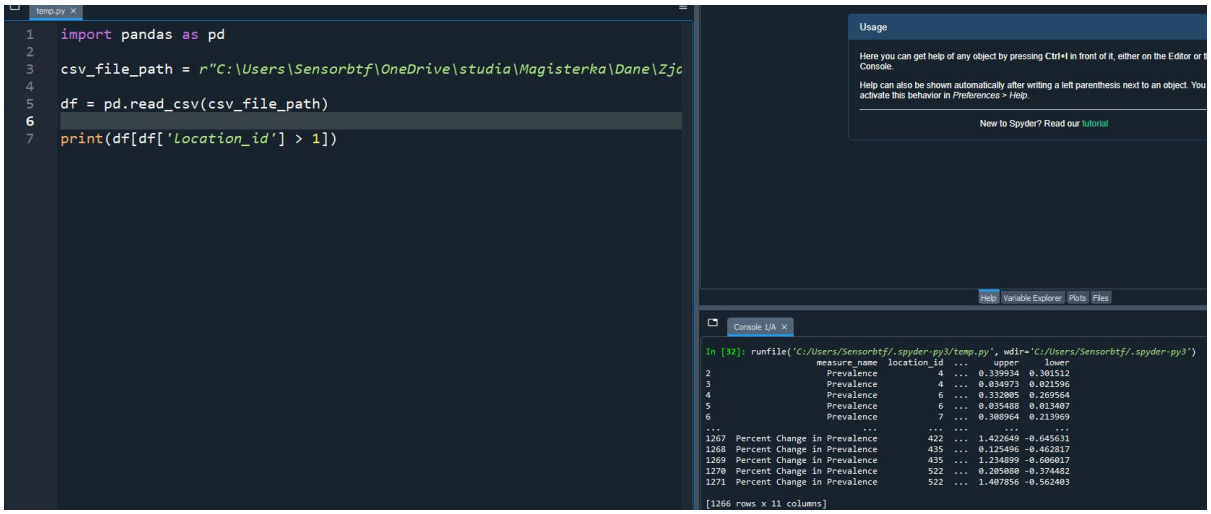
12. Usunięcie brakujących wartości



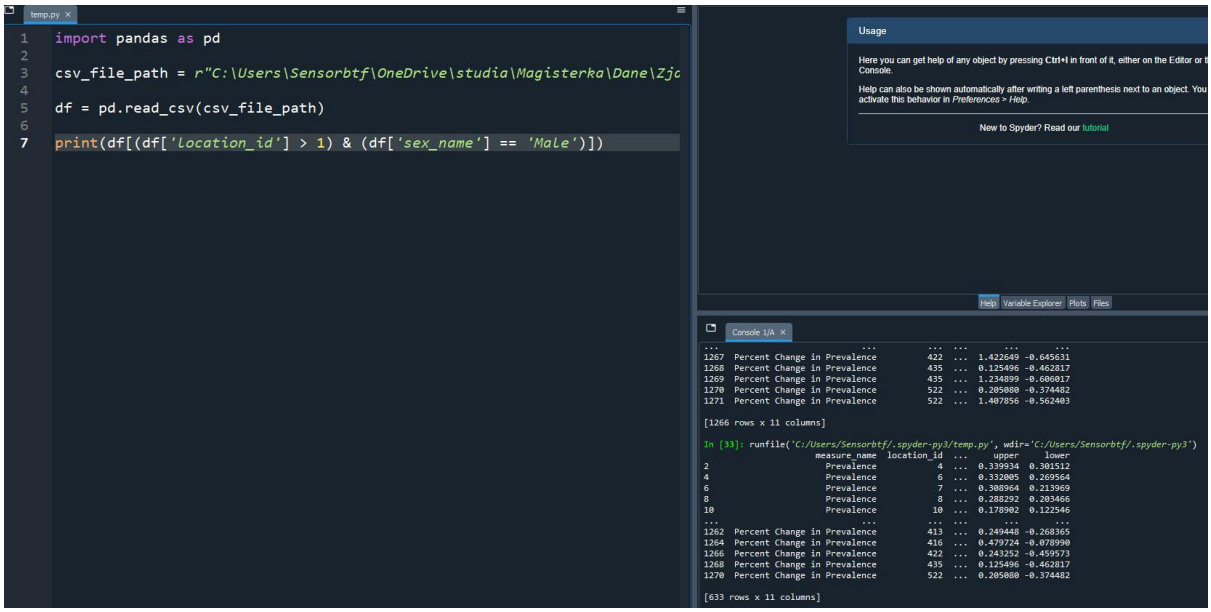
13. Wybór wierszy i kolumn po nazwach oraz indeksach



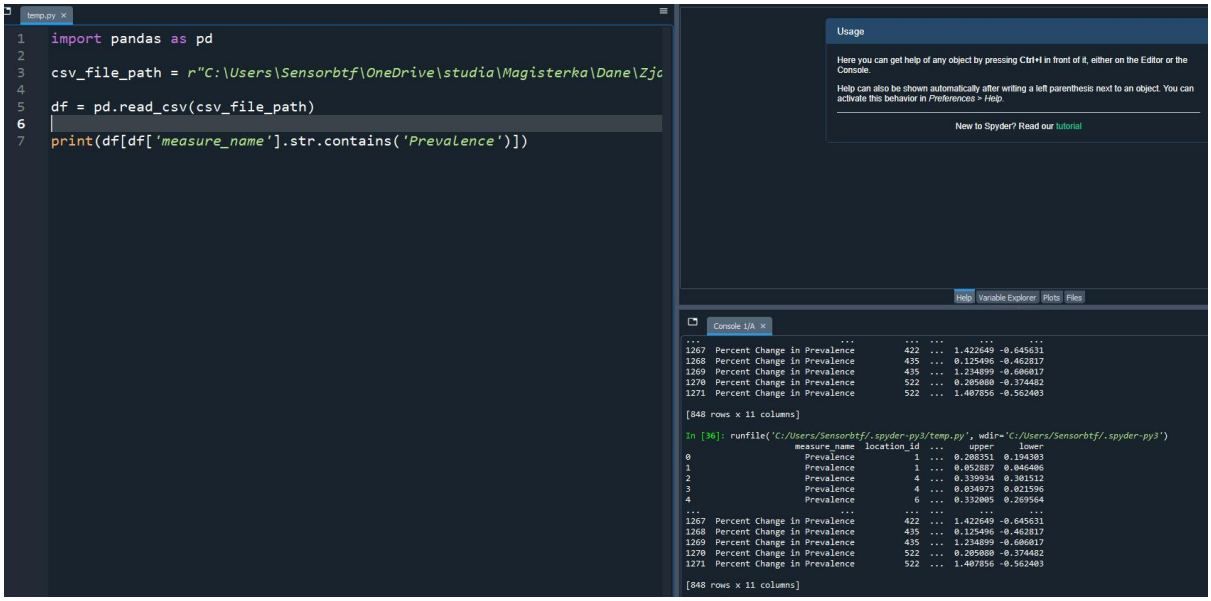
14. Wybór wierszy z warunkiem na wartość kolumny



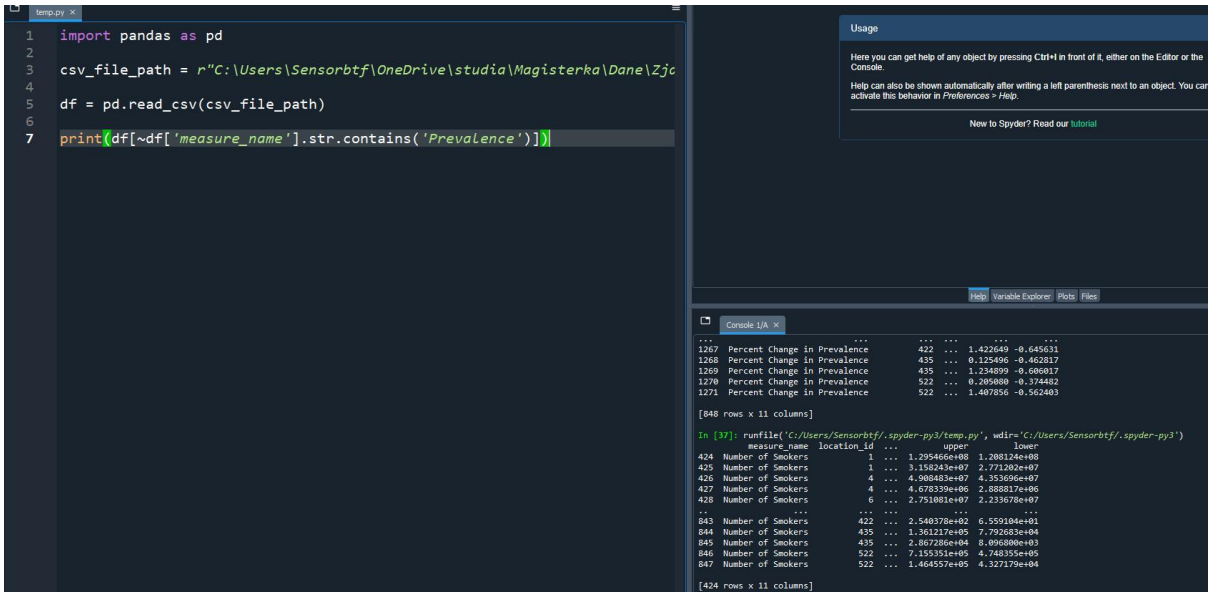
15. Wybór wierszy na podstawie kilku warunków



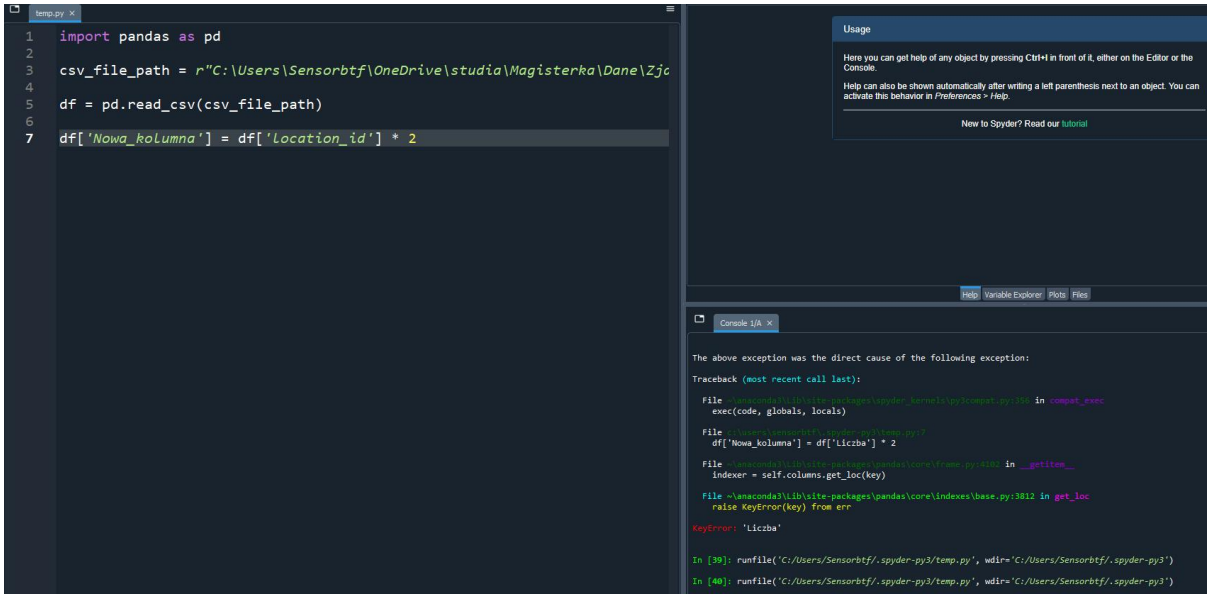
16. Wybór wierszy zawierających określone słowo



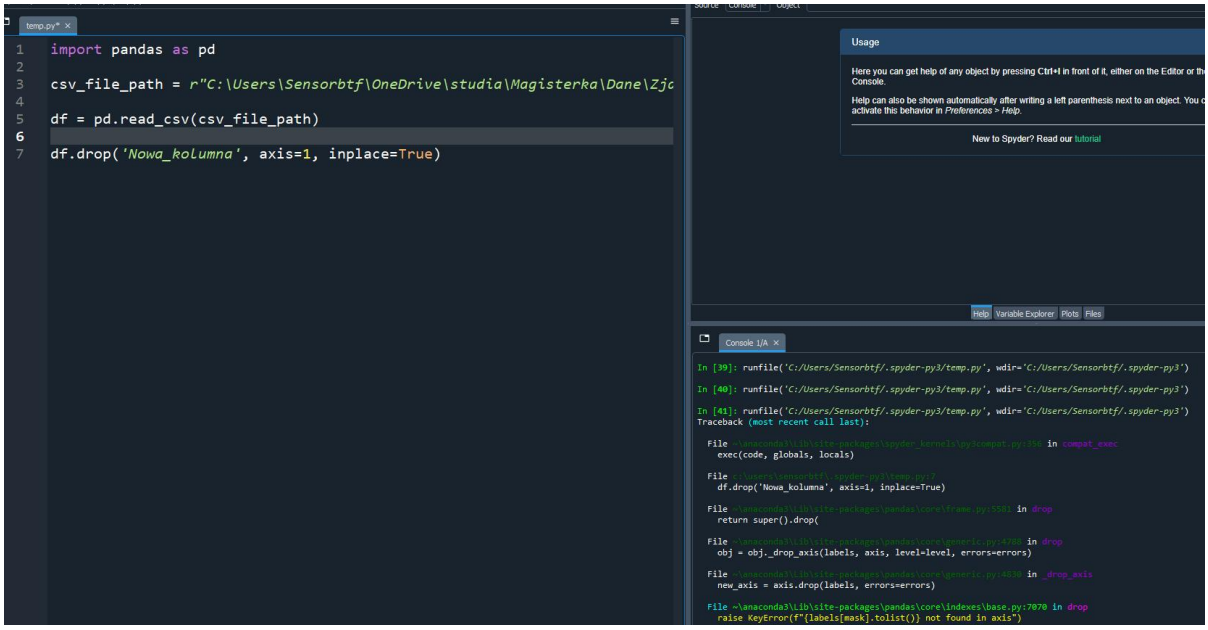
17. Wybór wierszy nie zawierających określonego słowa



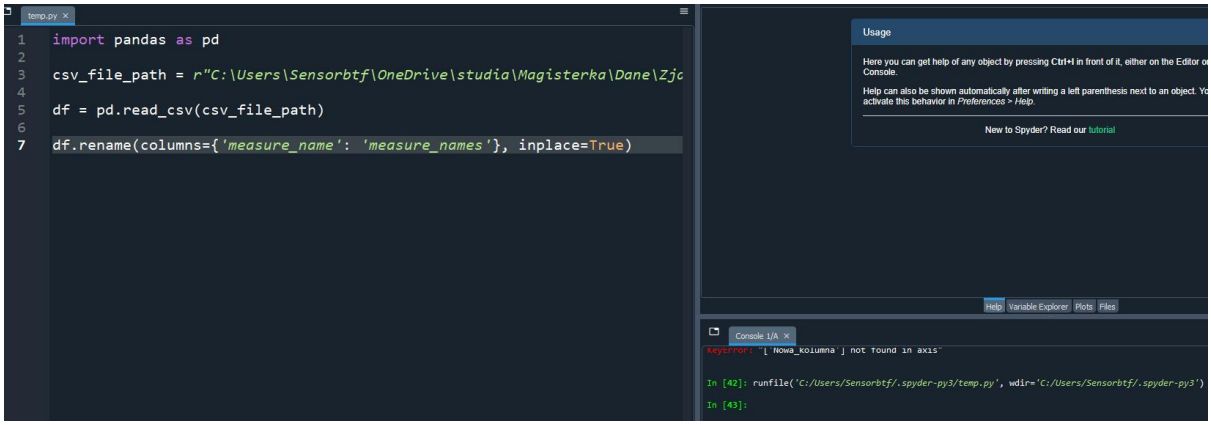
18. Utworzenie nowej kolumny na podstawie istniejących



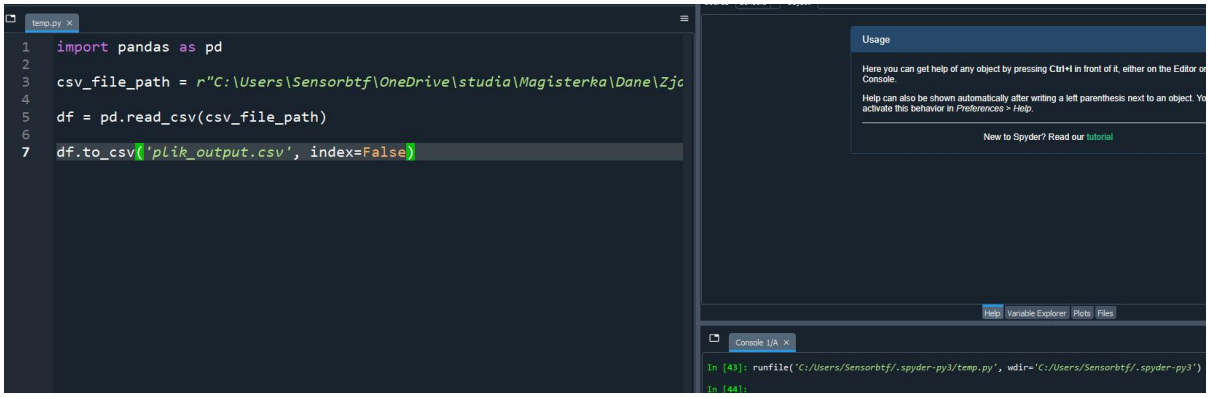
19. Usunięcie kolumny



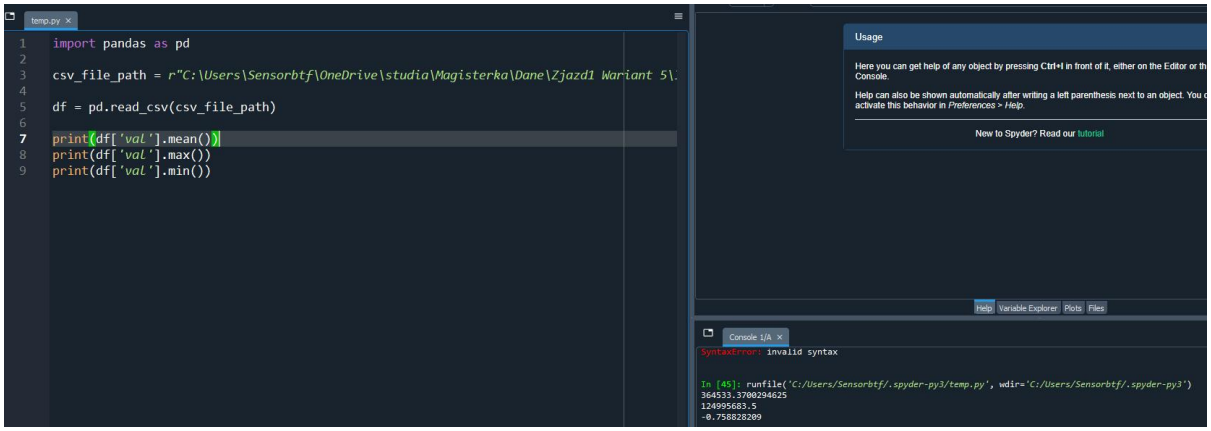
20. Zmiana nazwy kolumny



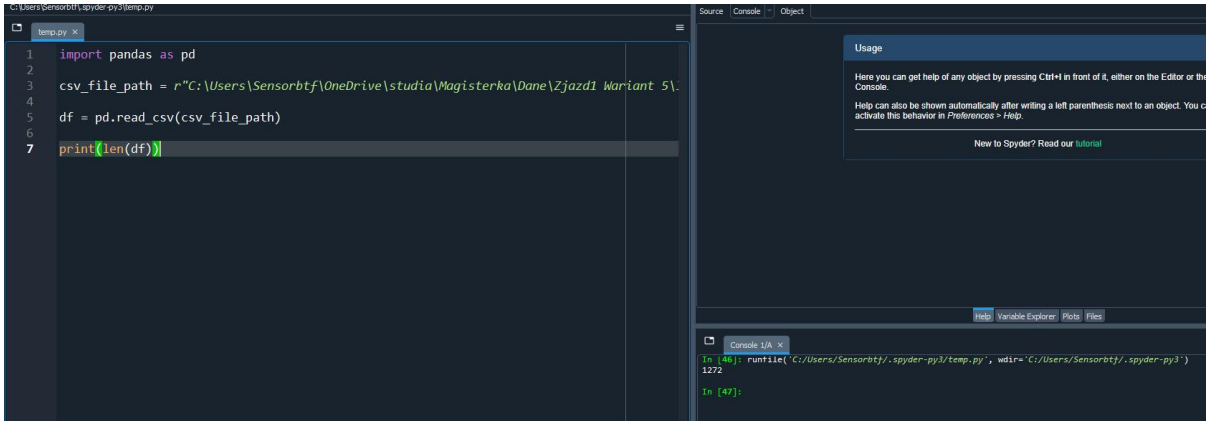
21. Zapis ramki danych jako plik CSV



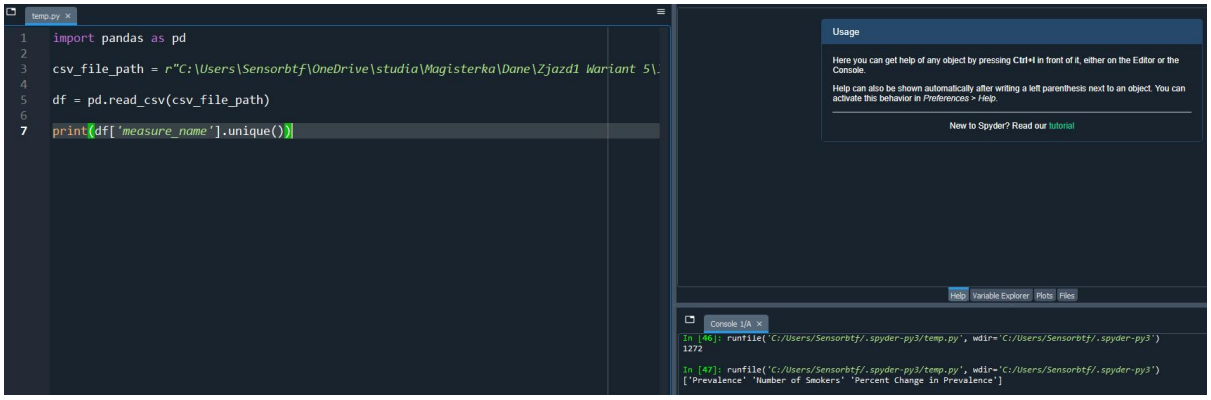
22. Wyświetlenie średniej, maksymalnej i minimalnej wartości z kolumny



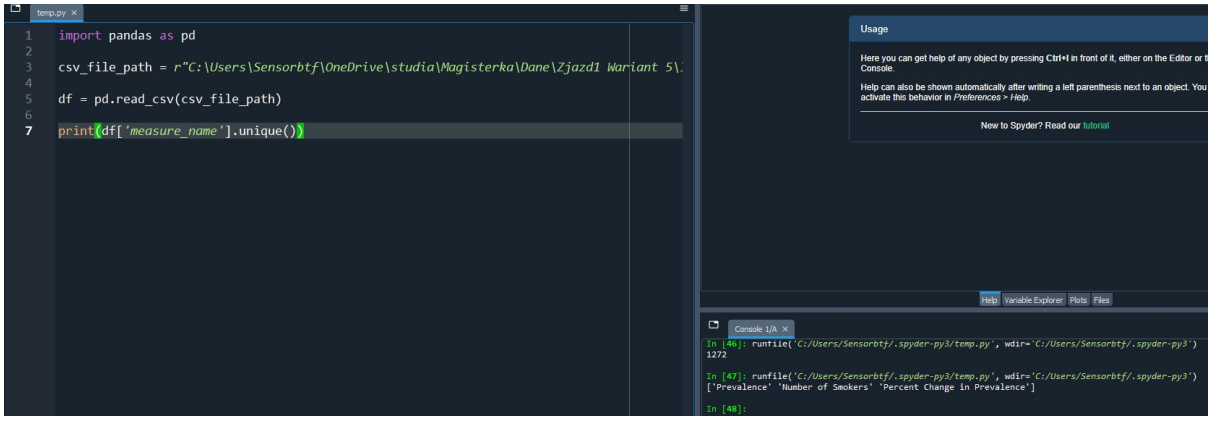
23. Wyświetlenie liczby wierszy



24. Wyświetlenie unikalnych wartości w kolumnie



25. Wyświetlenie liczby rekordów dla wartości w kolumnie



26. Sortowanie wierszy według wartości kolumny

```python
import pandas as pd

csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\

df = pd.read_csv(csv_file_path)

df.sort_values(by='lower', ascending=False, inplace=True)
```

27. Wyświetlenie 10 największych/najmniejszych wartości

```python
import pandas as pd

csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\

df = pd.read_csv(csv_file_path)

print(df.nlargest(10, 'val'))
print(df.nsmallest(10, 'val'))
```

```
In [51]: 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  ...  4.908483e+07  4.353696e+07
425  Number of Smokers            1  ...  3.158243e+07  2.771202e+07
428  Number of Smokers            6  ...  2.751081e+07  2.233678e+07
702  Number of Smokers          158  ...  2.768826e+07  2.157637e+07
710  Number of Smokers          163  ...  2.042377e+07  1.436712e+07
534  Number of Smokers           64  ...  1.652767e+07  1.520471e+07
662  Number of Smokers          137  ...  1.397654e+07  1.293966e+07
535  Number of Smokers           64  ...  1.151689e+07  9.929338e+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 -0.787475
1005  Percent Change in Prevalence           90  ... -0.504751 -0.755938
969   Percent Change in Prevalence           71  ... -0.430627 -0.726910
1230  Percent Change in Prevalence          216  ... -0.445845 -0.688150
991   Percent Change in Prevalence           83  ... -0.361854 -0.725385
1027  Percent Change in Prevalence          103  ... -0.482277 -0.644412
1226  Percent Change in Prevalence          214  ... -0.398855 -0.693001
1064  Percent Change in Prevalence          125  ... -0.439109 -0.660888
1004  Percent Change in Prevalence           90  ... -0.460321 -0.647364
```

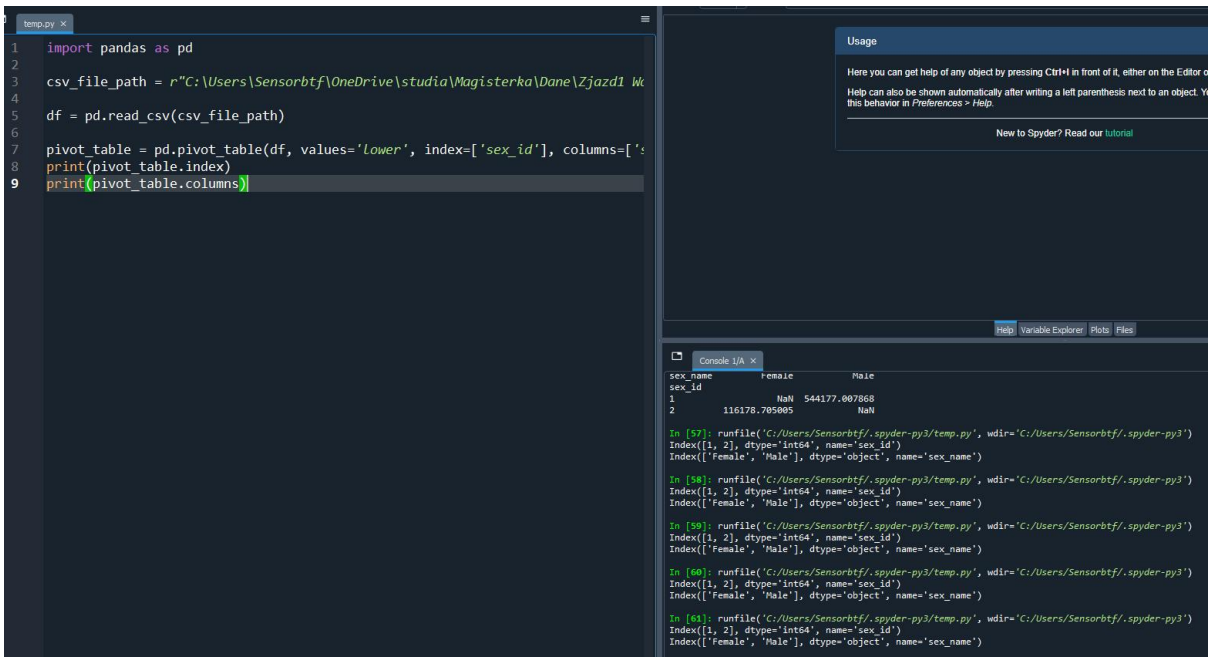
27. Grupowanie wierszy według kolumny

```python
import pandas as pd

csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\Zjazd1 Wariant 5\

df = pd.read_csv(csv_file_path)

df_grupowane = df.groupby('lower').mean()
print(df_grupowane)
```

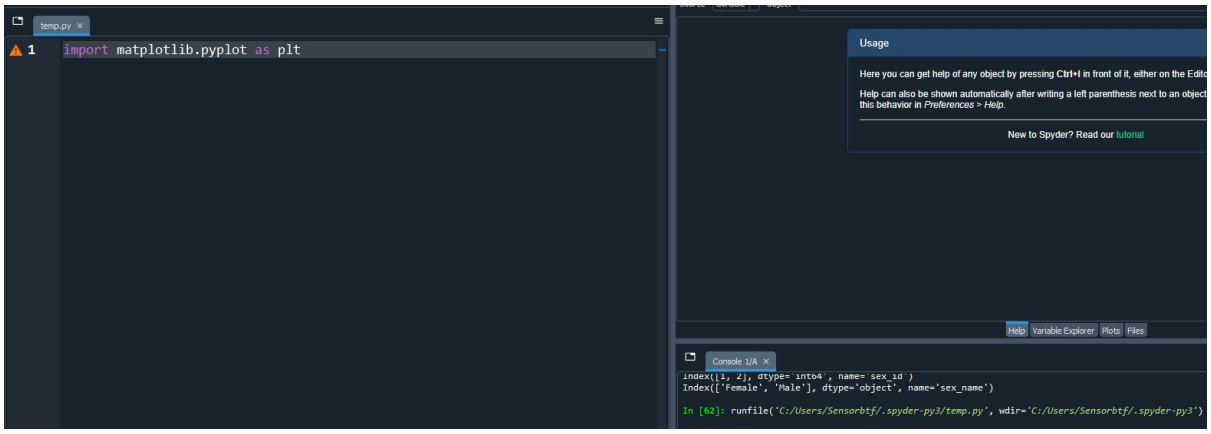
```
Traceback (most recent call last):

  File C:\anaconda3\lib\site-packages\spyder_kernels\py3compat.py:356 in compat_exec
    exec(code, globals, locals)

  File c:\users\sensorbtf\.spyder-py3\temp.py:7
    df_grupowane = df.groupby('lower').mean()

  File C:\anaconda3\lib\site-packages\pandas\core\groupby\groupby.py:2452 in mean
    result = self._cython_agg_general(

  File C:\anaconda3\lib\site-packages\pandas\core\groupby\groupby.py:1998 in _cython_agg_general
    new_mgr = data.grouped_reduce(array_func)

  File C:\anaconda3\lib\site-packages\pandas\core\internals\managers.py:1469 in grouped_reduce
    applied = sb.apply(func)

  File C:\anaconda3\lib\site-packages\pandas\core\internals\blocks.py:393 in apply
    result = func(self.values, **kwargs)

  File C:\anaconda3\lib\site-packages\pandas\core\groupby\groupby.py:1995 in array_func
    result = self._agg_py_fallback(how, values, ndim=data.ndim, alt=alt)

  File C:\anaconda3\lib\site-packages\pandas\core\groupby\groupby.py:1946 in _agg_py_fallback
    raise type(err)(msg) from err
```

29. Tworzenie tabeli przestawnej

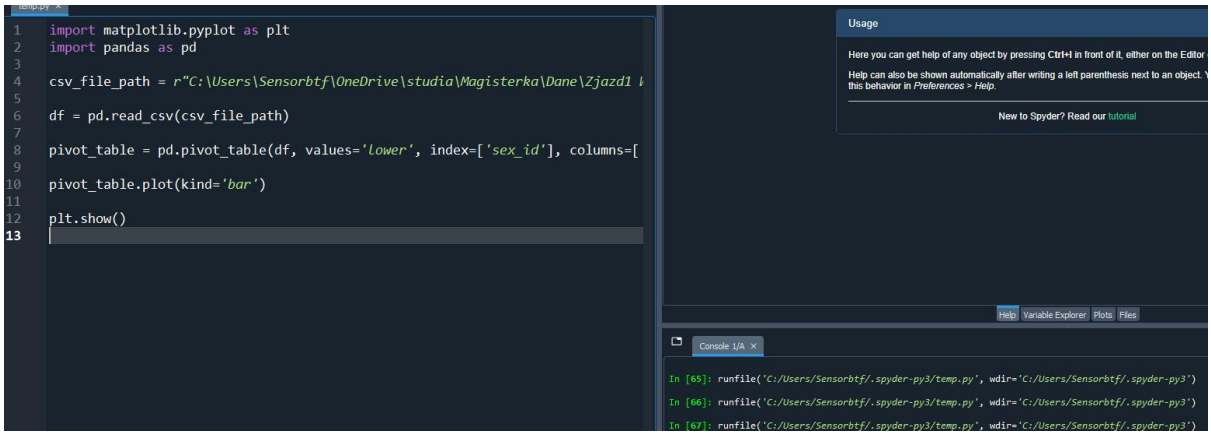
30. Wyświetlenie indeksów i kolumn tabeli przestawnej



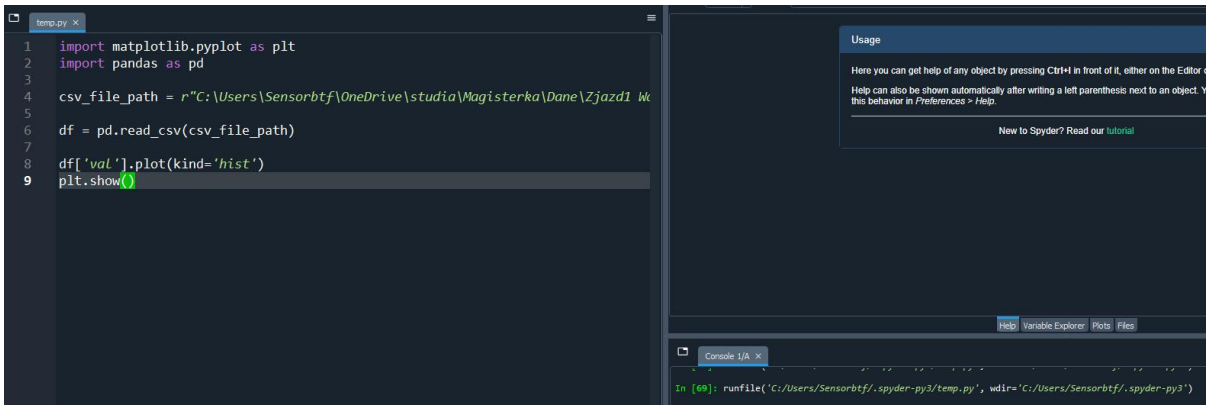
31. Importowanie modułu pyplot z biblioteki matplotlib



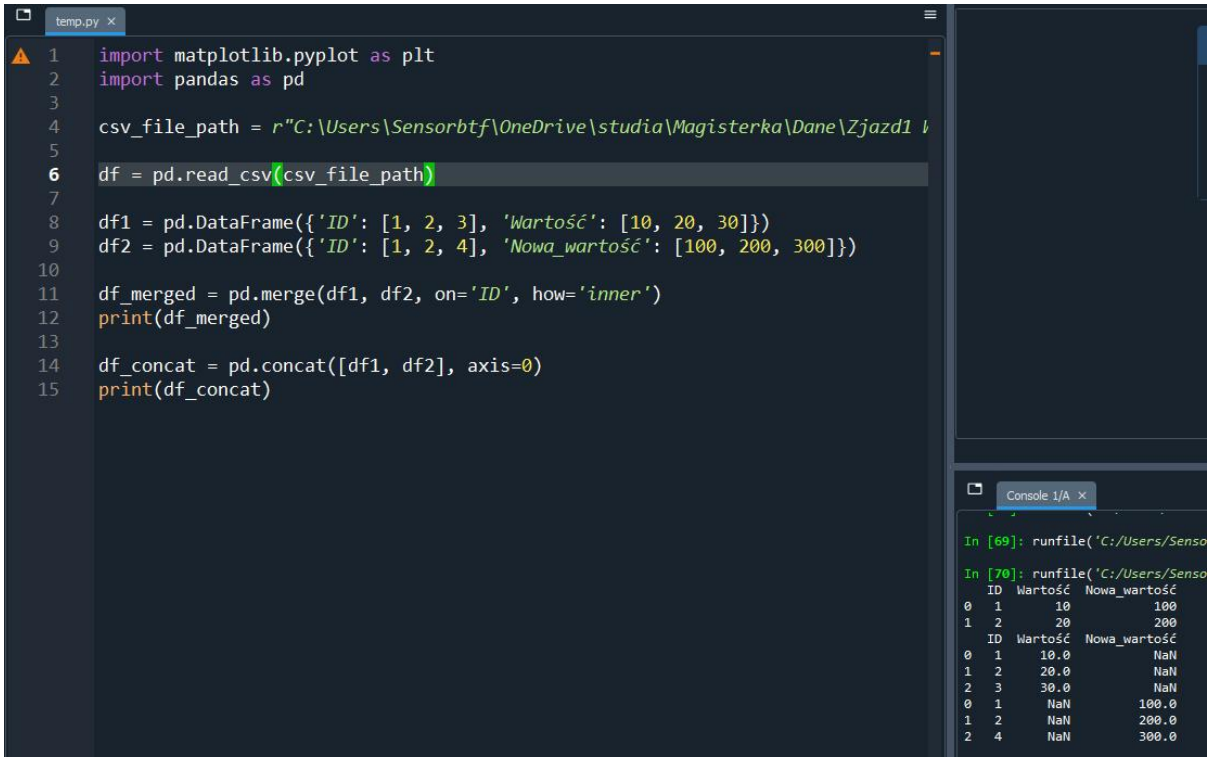
32. Rysowanie wykresu na podstawie tabeli przestawnej



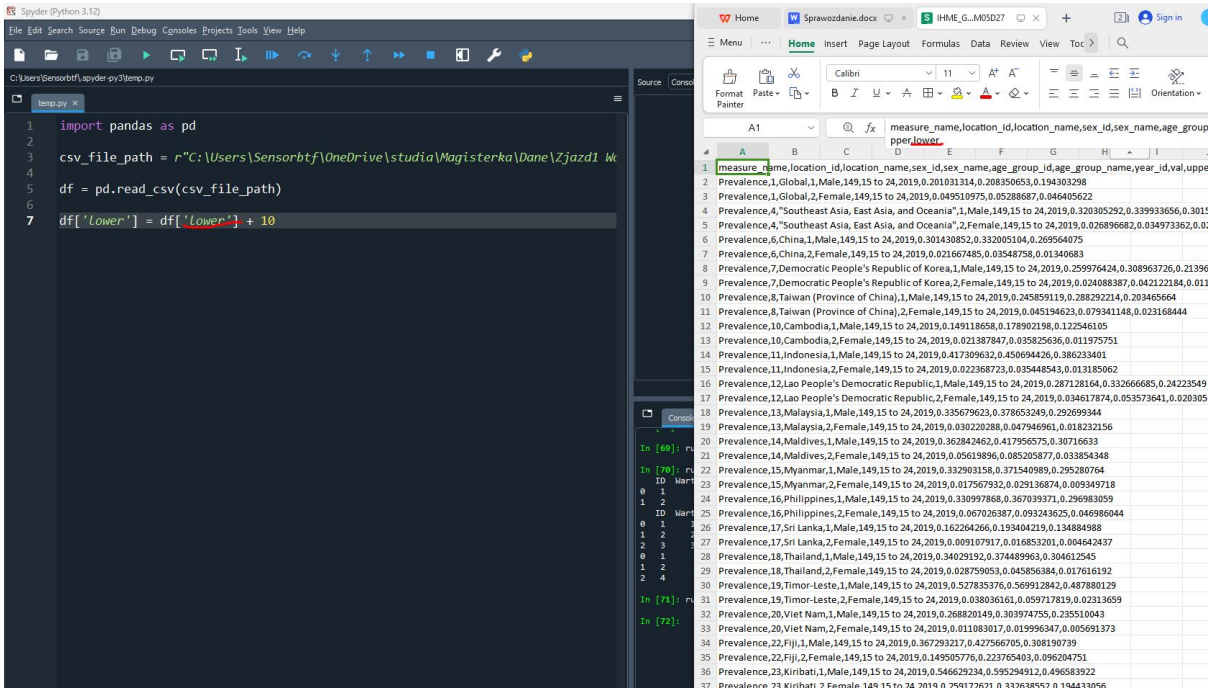
33. Rysowanie histogramu



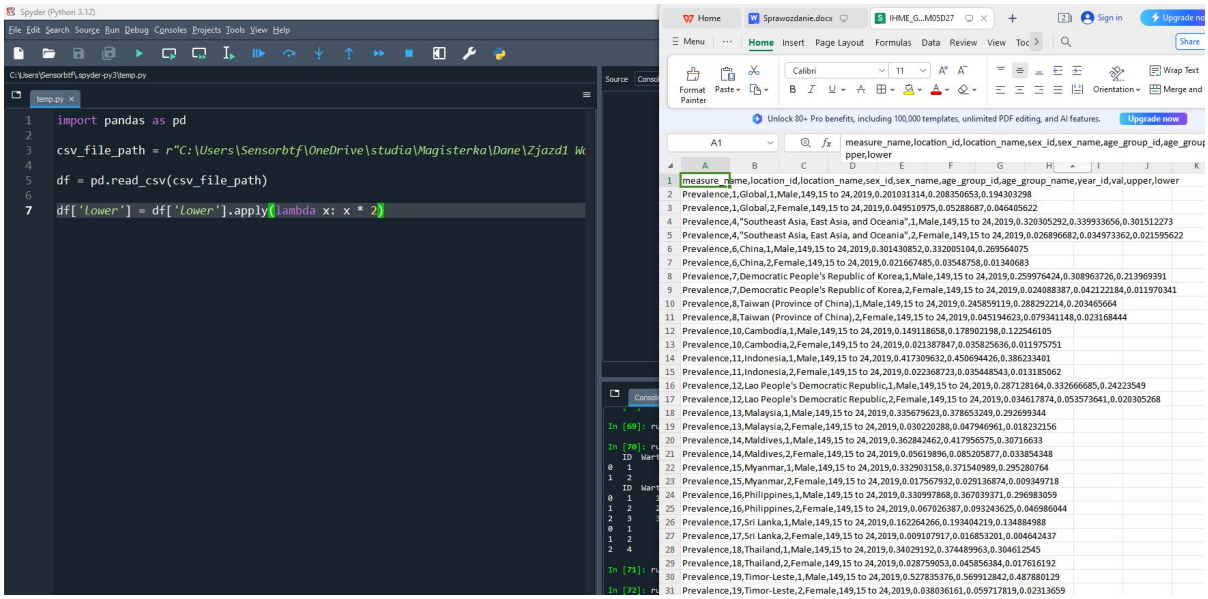
34. Łączenie ramek danych za pomocą merge i concat



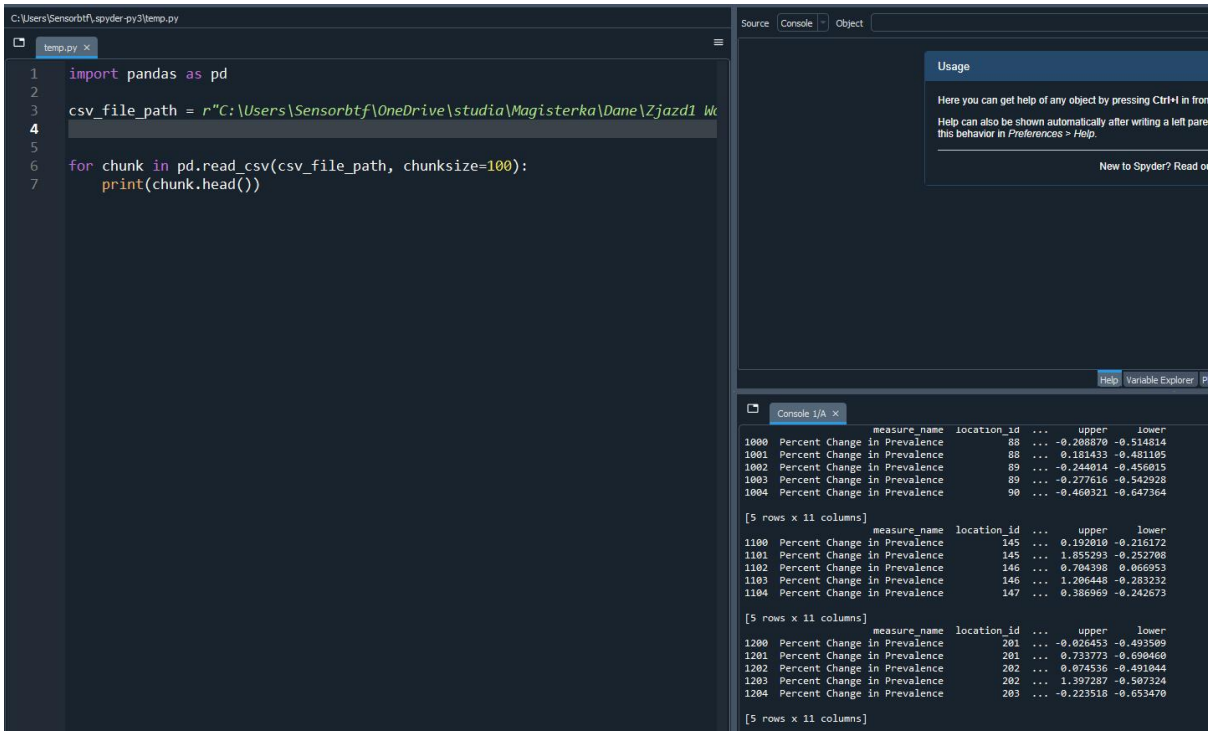
35. Dodawanie nowych kolumn za pomocą operacji matematycznych



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