

# Wizualizacja danych

Dr Anna Muranova  
Semestr letni 2024, UWM w Olsztynie

10. Zajęcie (26.04.2024)

---

Plik z wymaganiami:

<http://wmii.uwm.edu.pl/~muranova/WDI2024/requirements.txt>

Upewnij się, że w projekcie masz podpięta biblioteki `pandas`, `numpy` oraz `matplotlib`.

---

**Ćwiczenie 1** (Praca z plikami `program93.py`). Za pomocą eksplorera plików i notatnika stwórz plik tekstowy `tekst1.txt` w tym samym folderze co projekt, wpisz do pliku dowolny tekst. Następnie za pomocą instrukcji `open` i różnych parametrów sprawdź różne tryby otwarcie plików. Przykładowy kod:

```
f = open('tekst1.txt', 'r+')
s = f.read()
print(s)
print(type(s))
```

Następnie ustaw `break pointa` na trzeciej linijce w powyższym przykładzie i spróbuj zmodyfikować i zapisać zmienioną zawartość pliku tekstowego przez notatnik, WordPad, inny aplikacje.

Napraw sytuację wywołując metodę `close`.

Ew. zmień kod na taki z użyciem `with`.

Spróbuj dodać polskie znaki do pliku tekstowego i dodaj parametr `encoding` do funkcji `open` (poćwicz różne wersje kodowania).

Na koniec spróbuj zapisać coś nowego do pliku. Sprawdź w jakim trybie otwarcia plików jest to możliwe.

**Ćwiczenie 2** (`program101.py`). Pobierz plik z cenami jaj w Polsce w wybranych sieciach. Źródła danych:

<http://www.dlahandlu.pl/koszyk/towar/10-jaj-najtansze,38.html>

<http://wmii.uwm.edu.pl/~muranova/WDI2024/jajka2024.csv>

Zapisz plik w folderze projektu. Załaduj go do środowiska np. poniższa komenda:

```
data = np.genfromtxt('jajka2024.csv', delimiter=";", dtype=('|U16'))
```

<https://numpy.org/doc/stable/reference/generated/numpy.genfromtxt.html>

Dodaj najlepszy twoim zdaniem `encoding`.

Następnie przetwórz dane abyś można było wykonać operacje:

- obliczyć średnią cenę wszystkich jaj.
- wyznaczyć w którym mieście i w jakiej sieci są najtańsze a w jakich najdroższe jajka. Wynik zapisz w postaci dwuwymiarowej tablicy przechowującej pary (Miasto, nazwa sieci).

Wszystkie operacje wykonaj używając funkcji wbudowanych w interpretator lub biblioteki NumPy.

**Ćwiczenie 3.** Przeanalizuj następujący kod:

```
import pandas as pd

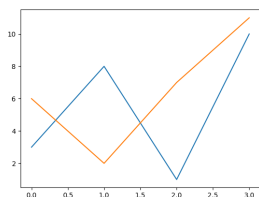
data = pd.read_csv('jajka2024.csv', sep=';', index_col=0,
encoding='UTF-8')

data2 = data.stack()

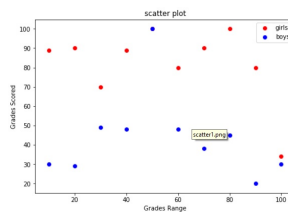
data3 = data2.str.replace(',', '.', astype='float')
srednia = data3.mean()
minCena = data3.min()
maxCena = data3.max()
print(data3[data3 == minCena])
print(data3[data3 == maxCena])
```

<https://www.w3resource.com/pandas/dataframe/dataframe-stack.php>

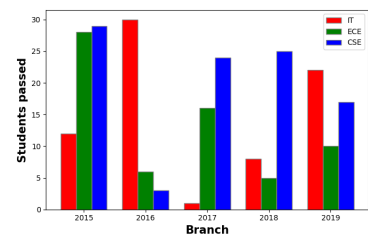
**Ćwiczenie 4.** Wykresy



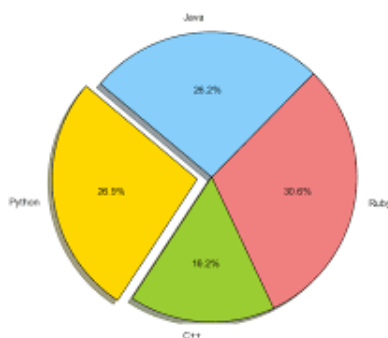
line



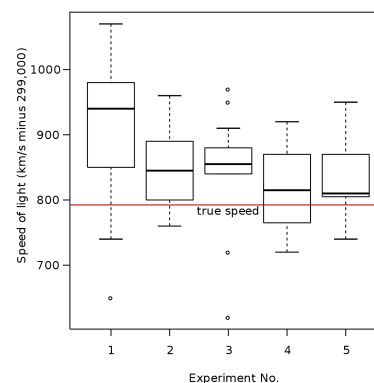
scatter/point



barplot



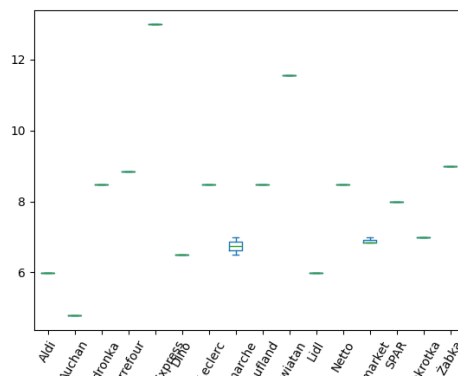
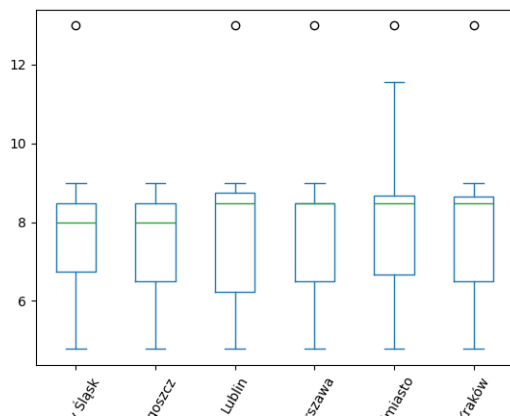
pie



boxplot

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html>

Zrobić wykresy pudełkowy cen na jaja w zależności od sklepu i od miasta (program102.py)



Pamiętaj: `import matplotlib.pyplot as plt` i wykonaj polecenie `plt.show()`

**Ćwiczenie 5** (`program103.py`). Spośród przygotowanych plików `.csv` z poprzednich ćwiczeń wybierz dwa, dla których wykonasz po 10 różnych operacji związanych z grupowaniem i filtrowaniem danych i zrobisz dwa wykresy.

[https://prezydent20200628.pkw.gov.pl/prezydent20200628/pl/dane\\_w\\_arkuszach](https://prezydent20200628.pkw.gov.pl/prezydent20200628/pl/dane_w_arkuszach)

<https://dane.gov.pl/pl/dataset/2582,statystyki-zakazen-i-zgonow-z-powodu-covid-19>

<https://github.com/mwaskom/seaborn-data>