



Rzeszów, 2025

ZARZĄDZANIE DANYMI

LABORATORIUM nr 4 (część 1)

Temat: Analiza oraz wizualizacja danych (*pandas* oraz *matplotlib*)

Laboratorium obejmuje implementację skryptów w języku *Python* w zakresie:

- analizy danych z zastosowaniem biblioteki *pandas*, oraz
- wizualizację danych przy pomocy biblioteki *matplotlib*

z poziomu środowiska *Jupyter Notebook/LAB*.

Samodzielne wykonanie zadań z laboratorium będzie wymagane z zastosowaniem środowiska *Jupyter Notebook* oraz/lub *Jupyter LAB* (pliki **.ipynb*).



Zadania

(do wykonania na laboratorium)

1. Import bibliotek (opcjonalny)

Zaimportować biblioteki (opcjonalnie), w tym:

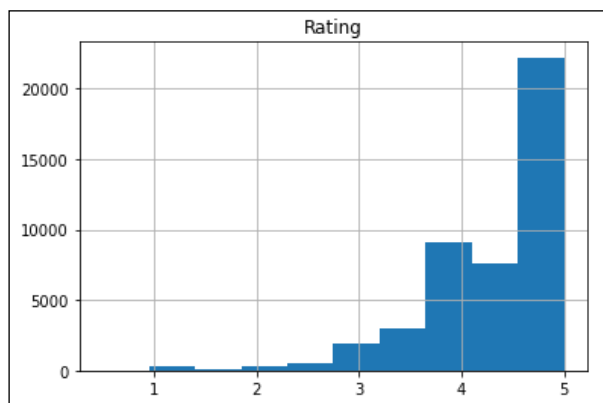
- *justpy* – biblioteka umożliwiająca implementację aplikacji webowych oraz wizualizację danych,
- *pandas* – biblioteka do analizy danych,
- *pytz* – biblioteka do obliczeń wartości daty/godziny (tj. *datetime*) pomiędzy różnymi strefami czasowymi, oraz
- *matplotlib* – biblioteka umożliwiająca wizualizację danych.

2. Wczytywanie oraz podstawowa/prosta analiza zbioru danych (*pandas*)

Wczytać zbiór danych z pliku wejściowego *reviews_courses.csv*, po czym dokonać prostych analiz danych oraz typów danych (np. *DataFrame*, *Series*, etc.) (np. *Rys.2.1-Rys.2.3*).

	Course Name	Timestamp	Rating	Comment
0	The Python Course: AI/ML in Python	2021-04-02 06:25:52+00:00	4.0	NaN
1	The Python Course: AI/ML in Python	2021-04-02 05:12:34+00:00	4.0	NaN
2	The Python Course: AI/ML in Python	2021-04-02 05:11:03+00:00	4.0	NaN
3	The Python Course: AI/ML in Python	2021-04-02 03:33:24+00:00	5.0	NaN
4	The Python Course: AI/ML in Python	2021-04-02 03:31:49+00:00	4.5	NaN

Rys.2.1 Wynik – Część zbioru danych z pliku wejściowego *CSV* zaimportowany do *Jupyter Notebook/LAB*



Rys.2.2 Histogram z rozkładem wartości dla przykładowej kolumny (*Rating(s)*) zbioru danych

```
data['Rating'].mean()  
4.442155555555556
```

Rys.2.3 Średnia z wartości dla przykładowej kolumny (*Rating(s)*) – dla *całego* zbioru danych



3. Filtrowanie danych

	Course Name	Timestamp	Rating	Comment
3	The Python Course: AI/ML in Python	2021-04-02 03:33:24+00:00	5.0	NaN
4	The Python Course: AI/ML in Python	2021-04-02 03:31:49+00:00	4.5	NaN
5	The Python Course: AI/ML in Python	2021-04-02 01:10:06+00:00	4.5	NaN
6	The Python Course: AI/ML in Python	2021-04-02 00:44:54+00:00	4.5	NaN
7	The Python Course: AI/ML in Python	2021-04-01 23:42:02+00:00	5.0	NaN

Rys.3.1 Pojedynczy warunek dla danych (**Rating > 4**) (5 pierwszych wierszy wyniku – początkowy indeks 3 (?)) – jedynie 29 758 wierszy z 45 000

	Course Name	Timestamp	Rating	Comment
31	The Python Course: OpenCV in Python	2021-04-01 01:32:52+00:00	5.0	NaN
34	The Python Course: OpenCV in Python	2021-03-31 22:53:04+00:00	5.0	NaN
43	The Python Course: OpenCV in Python	2021-03-31 19:15:25+00:00	5.0	NaN
45	The Python Course: OpenCV in Python	2021-03-31 17:23:15+00:00	5.0	NaN
101	The Python Course: OpenCV in Python	2021-03-29 21:54:00+00:00	4.5	NaN

Rys.3.2 Połączenie 2-ch warunków (**Rating > 4** oraz nazwa kursu to **The Python Course: OpenCV in Python**) (5 pierwszych wierszy wyniku – początkowy indeks 31 (?)) - jedynie 351 wierszy z 45 000

4. Filtrowanie danych wg daty/czasu (kolumna **Timestamp**)

	Course Name	Timestamp	Rating	Comment
3065	The Python Course: Interactive Visualizations	2020-12-30 23:28:34	3.0	NaN
3066	The Python Course: AI/ML in Python	2020-12-30 22:59:02	4.0	NaN
3067	The Python Course: AI/ML in Python	2020-12-30 22:40:10	4.5	NaN
3068	The Python Course: AI/ML in Python	2020-12-30 21:56:41	4.5	NaN
3069	The Python Course: AI/ML in Python	2020-12-30 21:14:34	4.5	NaN
...
9729	The Python Course: AI/ML in Python	2020-07-01 03:09:44	3.5	NaN
9730	The Python Course: AI/ML in Python	2020-07-01 03:09:12	5.0	NaN
9731	The Python Course: AI/ML in Python	2020-07-01 02:40:58	4.0	NaN
9732	The Python Course: AI/ML in Python	2020-07-01 02:04:02	5.0	nice
9733	The Python Course: AI/ML in Python	2020-07-01 00:01:34	2.0	Hard to follow if u have no experience program...

6669 rows x 4 columns

Rys.4 Wybór jedynie danych z zakresu od 1 lipca 2020 do 31 grudnia 2020 (kolumna **Timestamp**) – jedynie 6 669 wierszy z 45 000



5. Dane ↔ Informacja (tj. analiza danych)

```
print(f'Średnia ocen dla wszystkich kursów wynosi: {average_value}')  
Średnia ocen dla wszystkich kursów wynosi: 4.442155555555556
```

Rys.5.1 Średnia ocen dla *wszystkich* kursów

```
print(f'Średnia ocen dla \n- kursu "The Python Course: From Beginner to Exp  
Średnia ocen dla  
- kursu "The Python Course: From Beginner to Expert",  
- w tym, jedynie od "01.07.2020" do "31.12.2020" wynosi: 4.354066985645933
```

Rys.5.2 Średnia ocen dla „jedynie” kursu *The Python Course: From Beginner to Expert*, w tym w okresie od *1 lipca 2020* do *31 grudnia 2020*

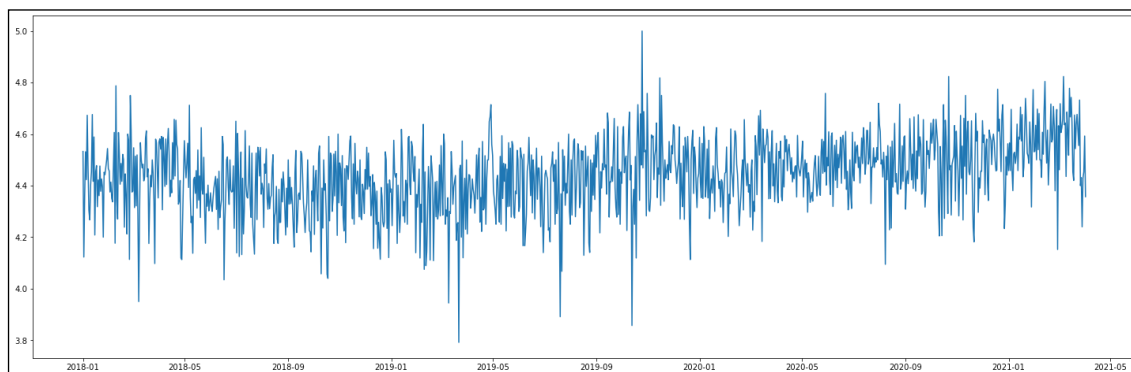
```
print(f'Liczba wierszy bez komentarzy wynosi: {no_of_uncommented_rows}')  
Liczba wierszy bez komentarzy wynosi: 38201
```

Rys.5.3 Liczba wierszy bez komentarzy – *38 201* rekordów z *45 000*

```
print(f'Liczba wierszy ze słowem "accent" w komenarzach wynosi: {no_of_comments_with_accent}')  
Liczba wierszy ze słowem "accent" w komenarzach wynosi: 77
```

Rys.5.4 Liczba wierszy z komentarzami, w tym zawierającymi wybrane słowo (*accent*) – jedynie *77* rekordów z *45 000*

6. Agregacja i wizualizacja danych (*matplotlib*)



Rys.6 Wykres z podziałem poszczególnych (uśrednionych) ocen (kolumna *Rating*) wg poszczególnych *dni*

7. Analiza danych w wybranych okresach czasu (np. wg dni, tygodni, miesiący)

Na kolejnym laboratorium (nr 4 część II)

8. Zadanie dodatkowe

Na kolejnym laboratorium (nr 4 część II)