



Rzeszów, 2025

ZARZĄDZANIE DANYMI

LABORATORIUM nr 4 (część 2)

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

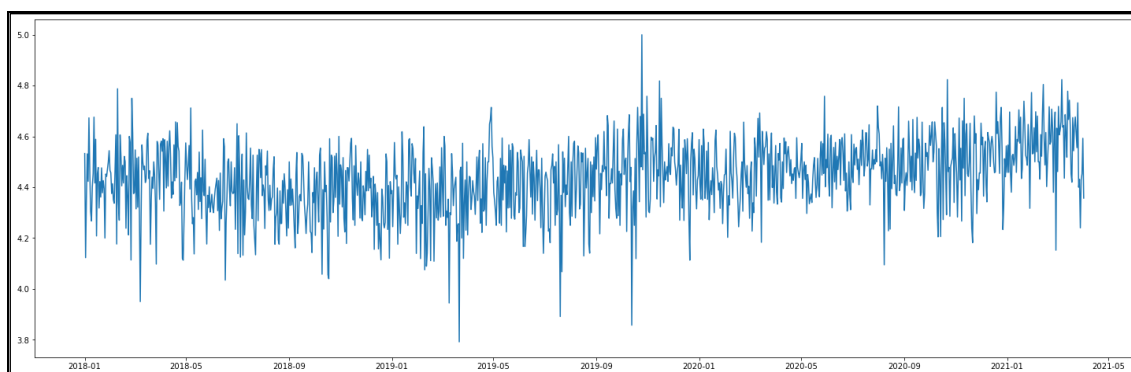
(do wykonania na laboratorium)

1. Analiza danych w wybranych okresach czasu (np. wg dni, tygodni, miesięcy)

1.1 Analiza wg *dni*

Zadania do wykonania:

- zaimportować wymagane biblioteki,
- wczytać zbiór danych wejściowych (plik *reviews_courses.csv*),
- pogrupować dane (wg kolumny *Timestamp* – dla *dni*),
- w oparciu o w/w dane utworzyć wykres z rozkładem ocen wg poszczególnych *dni*. Po wykonaniu wszystkich zadań wykres powinien wyglądać jak poniżej (**Rys.1.1**).



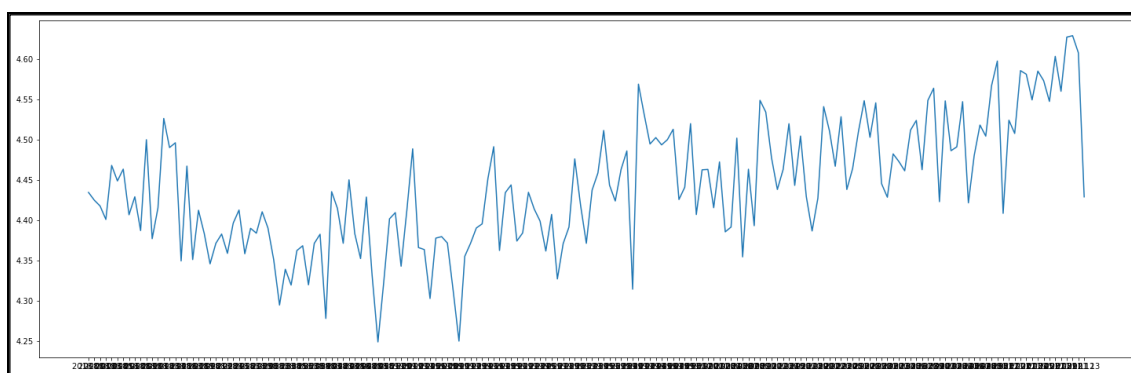
Rys.1.1 Wykres z podziałem poszczególnych (uśrednionych) ocen (kolumna *Rating*) wg poszczególnych *dni*



1.2 Analiza wg tygodni

Zadania do wykonania:

- zaimportować wymagane biblioteki,
- wczytać zbiór danych wejściowych (plik *reviews_courses.csv*),
- pogrupować dane (wg kolumny *Timestamp* – dla *tygodni*),
- w oparciu o w/w dane utworzyć wykres z rozkładem ocen wg poszczególnych *tygodni*. Po wykonaniu wszystkich zadań wykres powinien wyglądać jak poniżej (**Rys.1.2**).



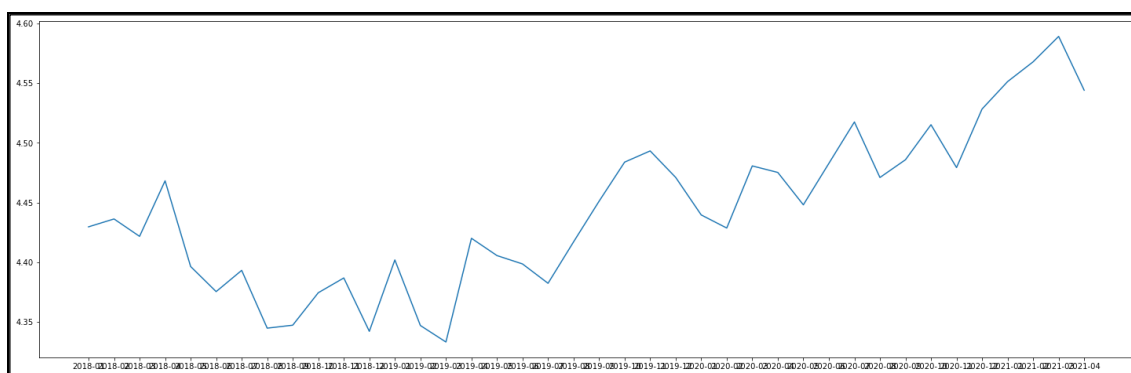
Rys.1.2 Wykres z podziałem poszczególnych (uśrednionych) ocen (kolumna *Rating*) wg poszczególnych *tygodni*



1.3 Analiza wg *miesięcy*

Zadania do wykonania:

- zaimportować wymagane biblioteki,
- wczytać zbiór danych wejściowych (plik *reviews_courses.csv*),
- pogrupować dane (wg kolumny *Timestamp* – dla *miesięcy*),
- w oparciu o w/w dane utworzyć wykres z rozkładem średnich ocen wg poszczególnych *miesięcy*. Po wykonaniu wszystkich zadań wykres powinien wyglądać jak poniżej (*Rys.1.3*).



Rys.1.3 Wykres z podziałem poszczególnych (uśrednionych) ocen (kolumna *Rating*) wg poszczególnych *miesięcy*



1.4 Analiza wg *miesiący* (dla każdego z kursów oddzielnie)

Zadania do wykonania:

- zaimportować wymagane biblioteki,
- wczytać zbiór danych wejściowych (plik *reviews_courses.csv*),
- pogrupować dane (dla poszczególnych *kursów* według *miesiący*) (*Rys.1.4* oraz/lub *Rys.1.5*),
- w oparciu o w/w dane utworzyć wykres z rozkładem uśrednionych ocen wg poszczególnych *miesiący*, w tym oddzielnie dla każdego *kursu*. Po wykonaniu wszystkich zadań wykres powinien wyglądać jak poniżej (*Rys.1.6*), oraz
- utworzyć wykres (wg danych jak w poprzednim punkcie), ale sumujący liczbę ocen kursów w poszczególnych miesiącach (*Rys.1.7*).

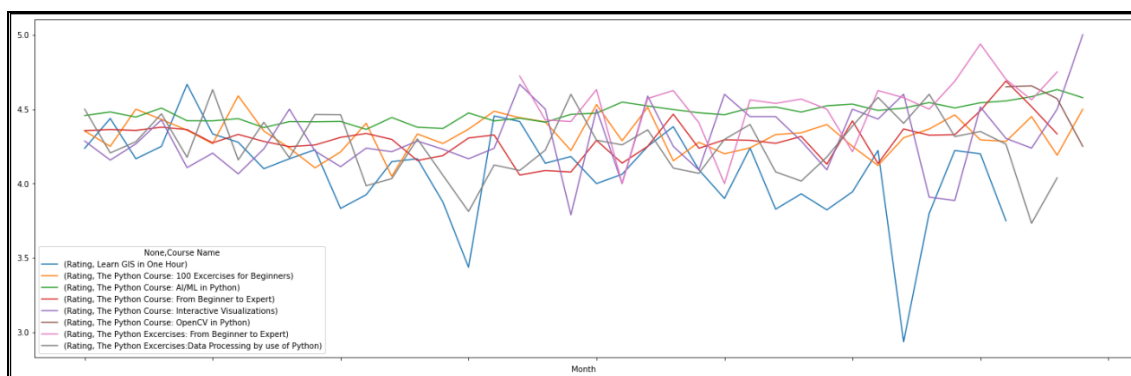
Month	Course Name	Rating
2018-01	Learn GIS in One Hour	4.236842
	The Python Course: 100 Exercises for Beginners	4.353448
	The Python Course: AI/ML in Python	4.457368
	The Python Course: From Beginner to Expert	4.355422
	The Python Course: Interactive Visualizations	4.285714
	The Python Exercises: Data Processing by use of Python	4.500000
2018-02	Learn GIS in One Hour	4.437500
	The Python Course: 100 Exercises for Beginners	4.250000
	The Python Course: AI/ML in Python	4.481070
	The Python Course: From Beginner to Expert	4.363095
	The Python Course: Interactive Visualizations	4.157895
	The Python Exercises: Data Processing by use of Python	4.205882
2018-03	Learn GIS in One Hour	4.166667
	The Python Course: 100 Exercises for Beginners	4.500000
	The Python Course: AI/ML in Python	4.447037

Rys.1.4 Dane dla poszczególnych (uśrednionych) ocen (kolumna *Rating*) wg poszczególnych *miesiący* oraz *kursów*

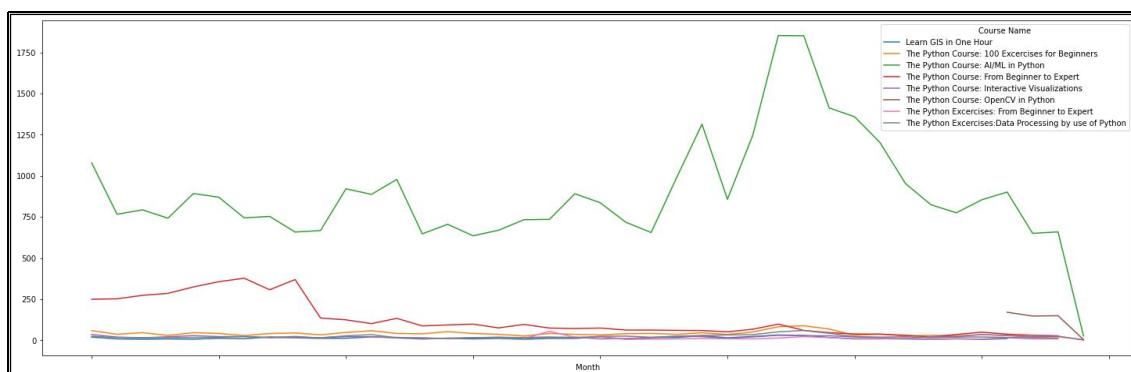


Course Name	Learn GIS in One Hour	The Python Course: 100 Exercises for Beginners	The Python Course: AI/ML in Python	The Python Course: From Beginner to Expert	The Python Course: Interactive Visualizations	The Python Course: OpenCV in Python	The Python Exercises: From Beginner to Expert	The Python Exercises: Data Processing by use of Python	Rating
Month									
2018-01	4.236842	4.353448	4.457368	4.355422	4.285714	NaN	NaN	4.500000	
2018-02	4.437500	4.250000	4.481070	4.363095	4.157895	NaN	NaN	4.205882	
2018-03	4.166667	4.500000	4.447037	4.357143	4.266667	NaN	NaN	4.281250	
2018-04	4.250000	4.431034	4.507412	4.378947	4.428571	NaN	NaN	4.468750	
2018-05	4.666667	4.358696	4.422085	4.362654	4.107143	NaN	NaN	4.176471	

Rys.1.5 Dane dla poszczególnych (uśrednionych) ocen (kolumna **Rating**) wg poszczególnych **miesięcy** oraz **kursów** w formie tabeli przestawnej



Rys.1.6 Wykres z podziałem poszczególnych (uśrednionych) ocen (kolumna **Rating**) wg poszczególnych **miesięcy** (oddzielnie dla każdego z **kursów**)

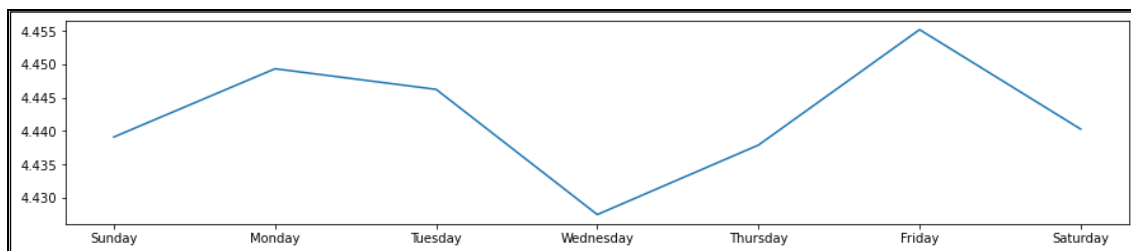


Rys.1.7 Wykres z sumą ocen dla poszczególnych **kursów** wg **miesięcy**



2. Zadanie dodatkowe – przykładowa analiza wg wybranych cech

Zadanie – na podstawie uśrednionych wartości *ocen* oraz dat – tj., *dni tygodnia*, kiedy były wystawiane – odpowiedzieć na pytanie „*w którym dniu użytkownicy byli najbardziej zadowoleni z kursów?*”. Wynik powinien wyglądać jak poniżej (*Rys.2.1*).



Rys.2.1. Wynik analizy zadowolenia użytkowników (wg *dni* oraz wystawianych (uśrednionych) *ocen*)