Patryk Ostrowski

Moduł 4, zadanie 1 – Irysy

1. Zaczytanie pliku i ‘przegląd’ zawartości:

A screen shot of a computer code

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

1. Wylistowanie kolumn zastanych i zastąpienie ich własnymi:

A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

1. Zliczenie wartości dla każdej klasy i zliczenie unikatów w każdej z kolumn:

A screen shot of a computer

AI-generated content may be incorrect.

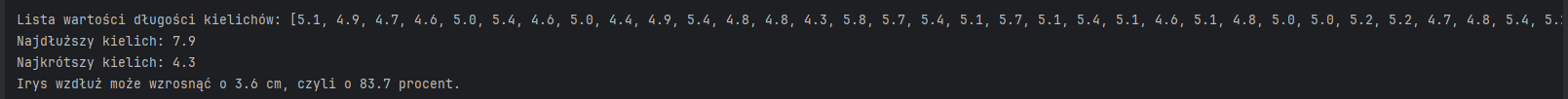
A screenshot of a computer

AI-generated content may be incorrect.

1. Praca na danych dot. kolumny długości kielichów:

A computer screen with text on it

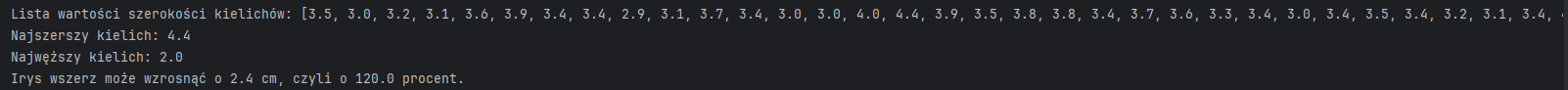
AI-generated content may be incorrect.



1. Praca na danych dot. kolumny szerokości kielichów:

A computer screen with text

AI-generated content may be incorrect.



1. Praca na danych dot. kolumny długości płatków – na Data Frame a nie na liście tym razem:

A blurry image of a city

AI-generated content may be incorrect.

A black background with white text

AI-generated content may be incorrect.

1. Praca na danych dot. kolumny szerokości płatków – na Data Frame a nie na liście tym razem:

A black background with white text

AI-generated content may be incorrect.

A black background with white text

AI-generated content may be incorrect.

1. Analiza grup w ramach agregacji:

A computer screen with text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

1. Napisałem funkcję, która jest następnie wywoływana w pętli i pokazuje wyliczenia – w tym przypadku średniej dla każdego z wymiarów irysa – dł. i szer. kielicha i odpowiednio płatka.

A black screen with many lines

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

1. Wykresy

A screen shot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

Podział na pięć binsów przrejżyście pokazuje udział próbek każdej klasy – akurat tutaj jest ich po 50 co sugeruje, że wcześniejsze i dalsze analizy są miarodajne, gdyż nie mamy klasy, której ilość próbek jest nadzwyczajnie duża lub nadzwyczajnie niewielka.

Wykres kołowy:

A black background with orange and green text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Przy zastanej konfiguracji Data Frame program pokazał co umie, czyli po prostu poszalał, ale to sugeruje poprawne wczytanie rzędów i serii danych. Dalszym krokiem byłoby wyłuskanie od biznesu co byłoby dla nich wartościowe i głębsze poznanie modułu wykresów przez developera 😊

Wykres słupkowy:

A black background with white text

AI-generated content may be incorrect.

A line graph with orange and blue lines

AI-generated content may be incorrect.

Wykres – pomimo swojej pozornej nieczytelności – jest bardzo czytelny w pokazaniu tendencji wzrostowych dla poszczególnych odmian/typów irysów.

Co można z tego wyczytać? Na przykład przekrojowy zakres rozmiarów kwiata, co przy prowadzeniu biznesu może wpłynąć na strategię kształtowania cen, gdyż mniejsze egzemplarze wystarczy wpakować do mniejszego opakowania, mniej też miejsca mogą one zająć na ekspozycji, a co za tym idzie: 1) będziemy mogli/musieli zamówić różnego rodzaju/wymiaru opakowania, 2) będziemy różnicować cenę wliczoną za transport w zależności od tej wartości. Otóż większych kwiatów zmieści się mnie na ciężarówce transportowej, a więc trzeba rozważyć konwersję konkretnego wolumenu zamówienia na zysk; do rozważenia jest także większa – a co za tym idzie droższa – skrzynka w automacie paczkowym, gdyż gdzieś na którejś wartości osi Y możemy przekraczać progową wartość dla rozmiaru przesyłki M i zamawiać tylko skrytki L lub w ogóle być zmuszeni do omijania automatów i bardzo duże egzemplarze wysyłać wyłącznie kurierem z dostawą osobistą.

Wykresy punktowe:

A screenshot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

Powyższe wykresy ukazują rozstrzał wartości dla każdego z analizowanych wymiarów każdej z klas irysów. Możemy dokładniej przyjrzeć się wartościom skrajnym i pogłębić analizę w oparciu o poprzedni wykres słupkowy rozważając opakowania ‘nie mniejsze niż’ i ‘nie większe niż’.

Wykres pudełkowy:

A computer screen shot of text

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

Ten wykres pokazuje wartości min. i max, oznakowuje kwartyle 1-wszy (percentyl 25-ty), 2-gi (percentyl 50-ty) i 3-ci (percentyl 75-ty), a także wartości odstające dla każdego z analizaowanych wymiarów irysa z poddanego obserwacji zbioru próbek.

Analizę przeprowadzono w IDE PyCharm firmy JetBrains.

Repozytorium dostępne jest pod adresem: https://github.com/racibornio/Python-lessons/blob/master/akademia/zadania/mod\_4/mod\_4\_zad\_1.py