Zadanie: Segmentacja klientów – k-średnich i przypisanie nowego klienta (k-NN)

Kontekst

Jesteś analitykiem w sieci sklepów. Dysponujesz danymi o klientach: wiek, roczne wydatki (w PLN), liczba wizyt w sklepie na miesiąc oraz średnia wartość koszyka. Twoim celem jest podzielenie klientów na **3 segmenty** przy użyciu metody *k-średnich* (k-means), ocenienie jakości segmentacji i przypisanie nowego klienta do jednego z utworzonych segmentów za pomocą metody *k-najbliższych sąsiadów* (k-NN).

Plik z danymi

Dane zostały zapisane w pliku CSV o nazwie:
`clients.csv`

Plik zawiera kolumny (nagłówki):

`id`	identyfikator klienta (liczba całkowita)
`age`	wiek (lata)
`annual_spending`	roczne wydatki (PLN)
`visits_per_month`	liczba wizyt w sklepie na miesiąc
`avg_basket_value`	średnia wartość koszyka (PLN)
`true_segment`	znana etykieta segmentu (0,1,2) – wykorzystywana do oceny

Zadanie (wymagane kroki)

1. Wczytaj dane z pliku `clients.csv` do DataFrame (np. `pandas`). Wybierz cechy: `["age","annual_spending","visits_per_month","avg_basket_value"]`.

2. **Standaryzacja:** dopasuj skalera (np. `StandardScaler` z `sklearn`) do powyższych cech i przeskaluj dane.

3. Przeprowadź klastrowanie metodą **k-średnich** z $k = 3$ (użyj parametru `random_state` dla powtarzalności). Zapisz etykiety klastrów dla każdej obserwacji.

4. Pobierz środki klastrów w przestrzeni znormalizowanej i **odwróć transformację** skalera (użyj `inverse_transform`), aby otrzymać centra w oryginalnych jednostkach (PLN, lata, itp.).

5. Oblicz następujące miary jakości:

- `inertia` (wartość z obiektu KMeans),
- `silhouette_score` (dla przeskalowanych cech),
- na podstawie kolumny `true_segment` obliczenie % zgodnych przypisań.

6. Policz **Within-Cluster Sum of Squares (WSS)** w oryginalnych jednostkach: dla każdego punktu oblicz sumę kwadratów odległości do centrum przypisanego klastra wykorzystując skalę oryginalną, a następnie zsumuj te wartości.

7. **Podpunkt (k-NN):** Przyjmij, że etykiety klastrowe otrzymane z k-means są *celami* trenowania klasyfikatora. Naucz klasyfikator `KNeighborsClassifier` (np. `n_neighbors=5`) na przeskalowanych cechach i przypisz następnie nowego klienta:

   ```
   new_customer = { "age": 33, "annual_spending": 2900,

   "visits_per_month": 5, "avg_basket_value": 360 }
   ```

   W zadaniu podaj:

   - przewidywany klaster dla nowego klienta,
   - centroid przypisanego klastra w oryginalnych jednostkach,
   - odległość nowego klienta do centroidu (zarówno w skali znormalizowanej, jak i w oryginalnej skali).