



Klasteryzacja zakupów w internecie



Piotr Sieńko, Konrad Welkier, Jacek Wiśniewski



Zbiór danych

Online Shoppers Purchasing Intention

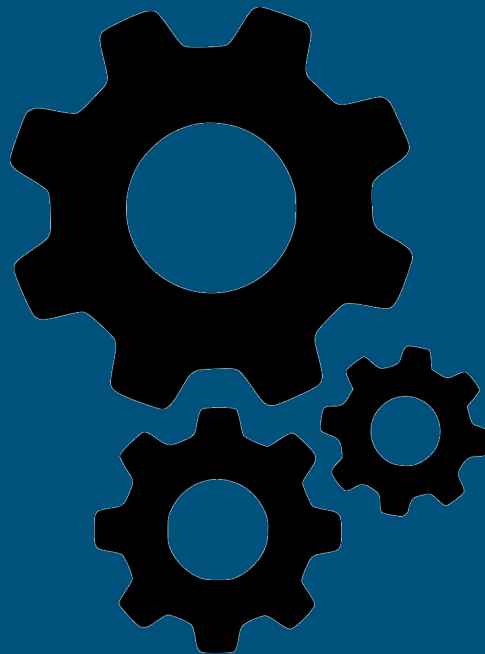
- 18 kolumn
- 12330 wierszy
- Bez braków danych



Metodyka

Algorytmy

- K-średnich
- GMM
- Aglomeracyjny



Metryki

- Indeks Silhouette
- Indeks Daviesa-Bouldina
- Indeks Randa

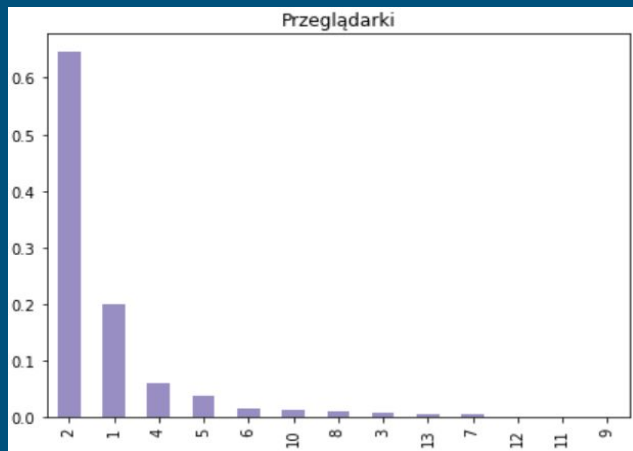
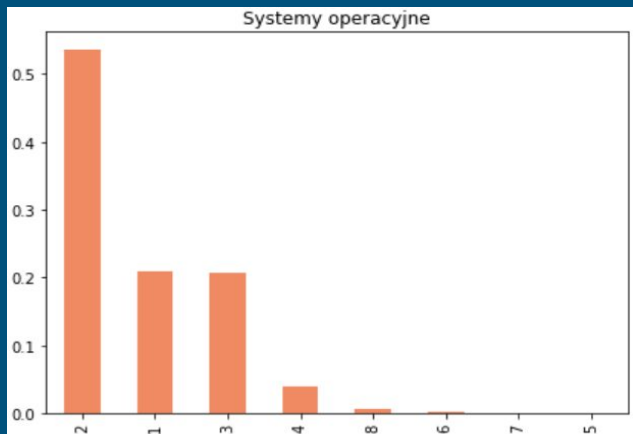
Pytanie badawcze

Czy podział zbioru przy pomocy klasteryzacji odzwierciedla rzeczywiste grupy klientów?



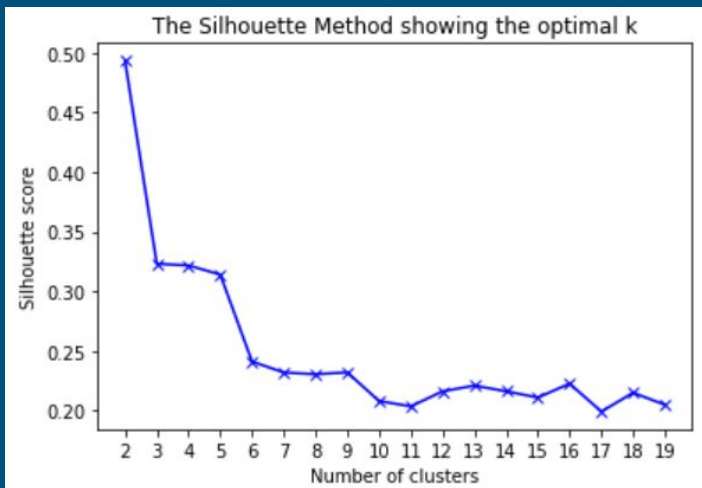
Przygotowanie danych

- Usunięcie kolumn
- Zgrupowanie zmiennych
- Standaryzacja

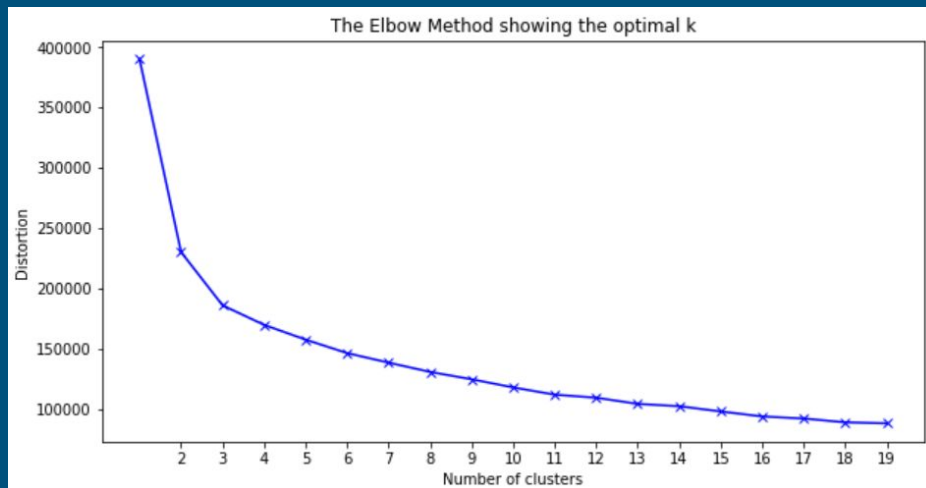


Dobór liczby klastrow

Metoda Silhouette



Metoda Łokcia



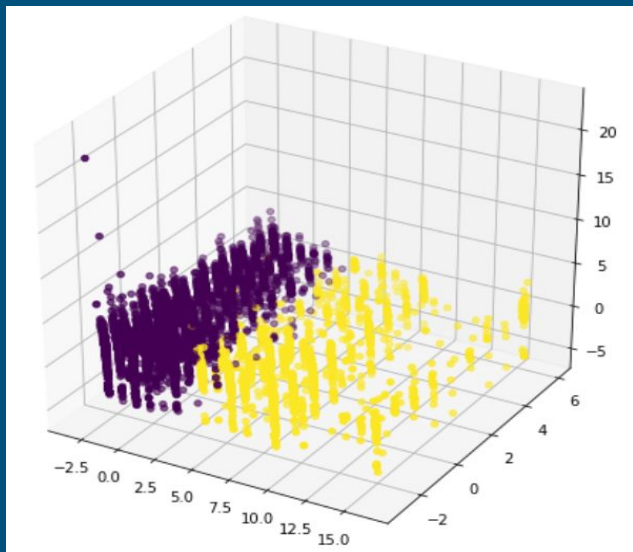
Nowe pytanie badawcze

Czy zbiór jest naturalnie podzielony na klientów kupujących i odwiedzających?

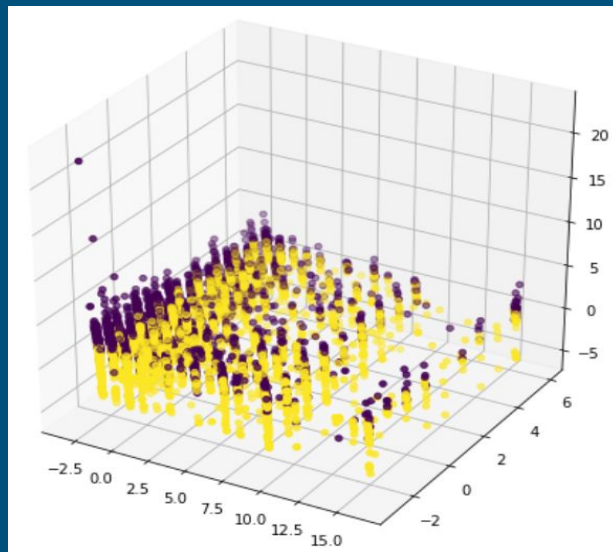


Pierwsze modele

Model k-średnich

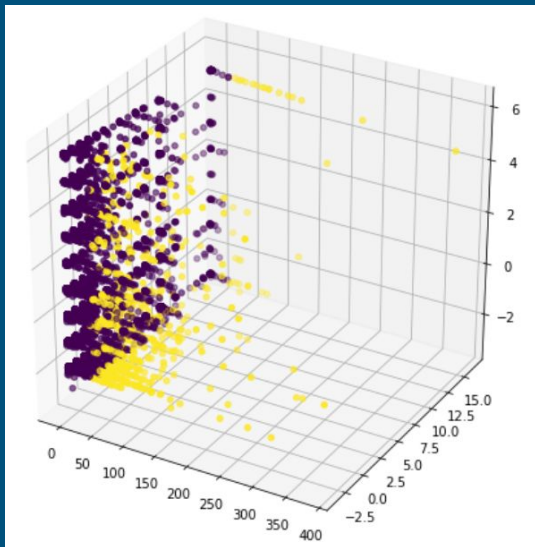


Model GMM

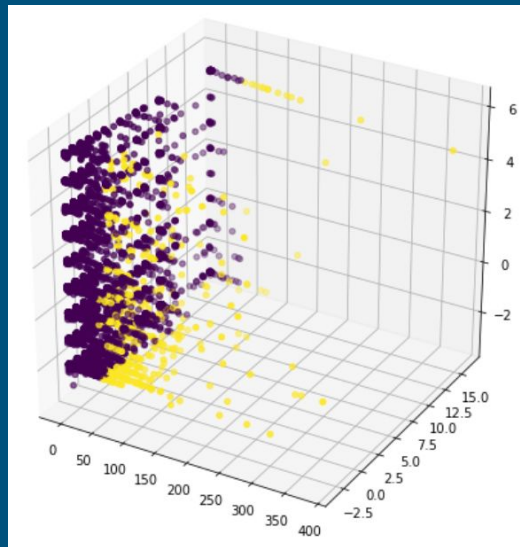


Modele ze zwiększonym page value

Model k-średnich

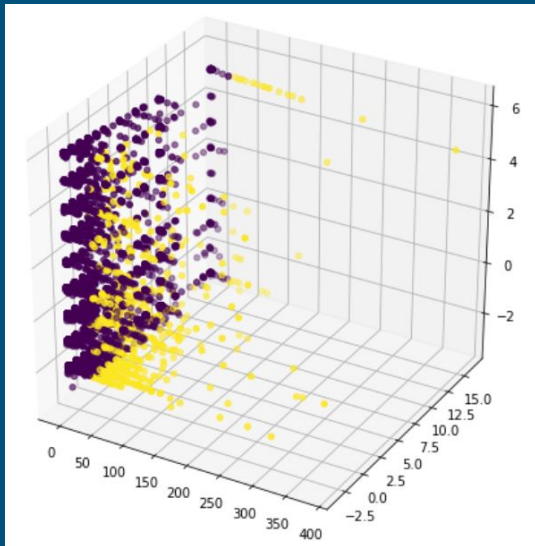


Model aglomeracyjny

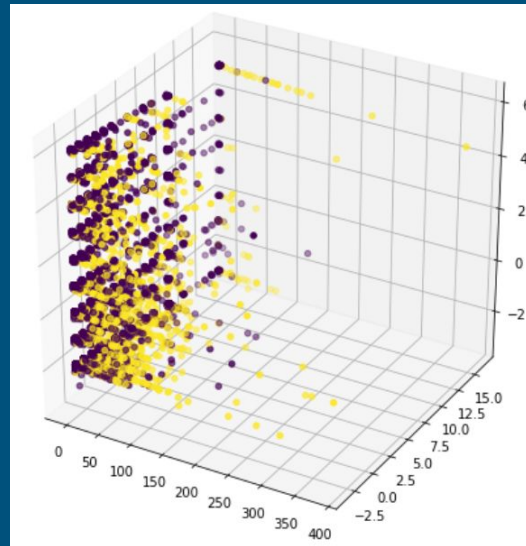


Weryfikacja hipotezy

Model k-średnich



Rzeczywistość



Wnioski

- W zbiorze znajdują się 2 klastry
- Ich podział nie jest oczywisty
- Prawdopodobnie nie odzwierciedlają rzeczywistych grup klientów



Pytania?



Dziękujemy
