

Co telefony wiedzą o naszej aktywności

Projekt z klasteryzacji

Florek Paweł, Deptuch Mateusz

CEL
BIZNESOWY

CEL BIZNESOWY

Celem projektu jest klasteryzacja danych zebranych z żyroskopu i akcelerometru, znajdujących się w smartfonach trzydziestu osób. Pomiary wykonywane były podczas wykonywania czynności takich jak leżenie, chodzenie czy bieganie.

EDA

EDA

10229

wierszy

561

kolumn

- brak duplikatów,
- brak wartości NA
- tylko wartości numeryczne
- dane ustandardyzowane

sygnały:

- tBodyAcc
- tGravityAccMag
- tBodyGyroJerkMag
- tBodyAccJerk
- tBodyGyro
- tBodyAccJerkMag
- tBodyGyroJerk
- tBodyAccMag
- tGravityAcc
- tBodyGyroMag
- fBodyAcc
- fBodyBodyGyroMag
- fBodyAccJerk
- fBodyBodyAccJerkMag
- fBodyAccMag
- fBodyGyro
- fBodyBodyGyroJerkMag

funkcje:

- max()
- mad()
- min()
- kurtosis()
- bandsEnergy()
- mean()
- meanFreq()
- arCoeff()
- entropy()
- iqr()
- sma()
- std()
- maxInds()
- skewness()
- energy()
- correlation()

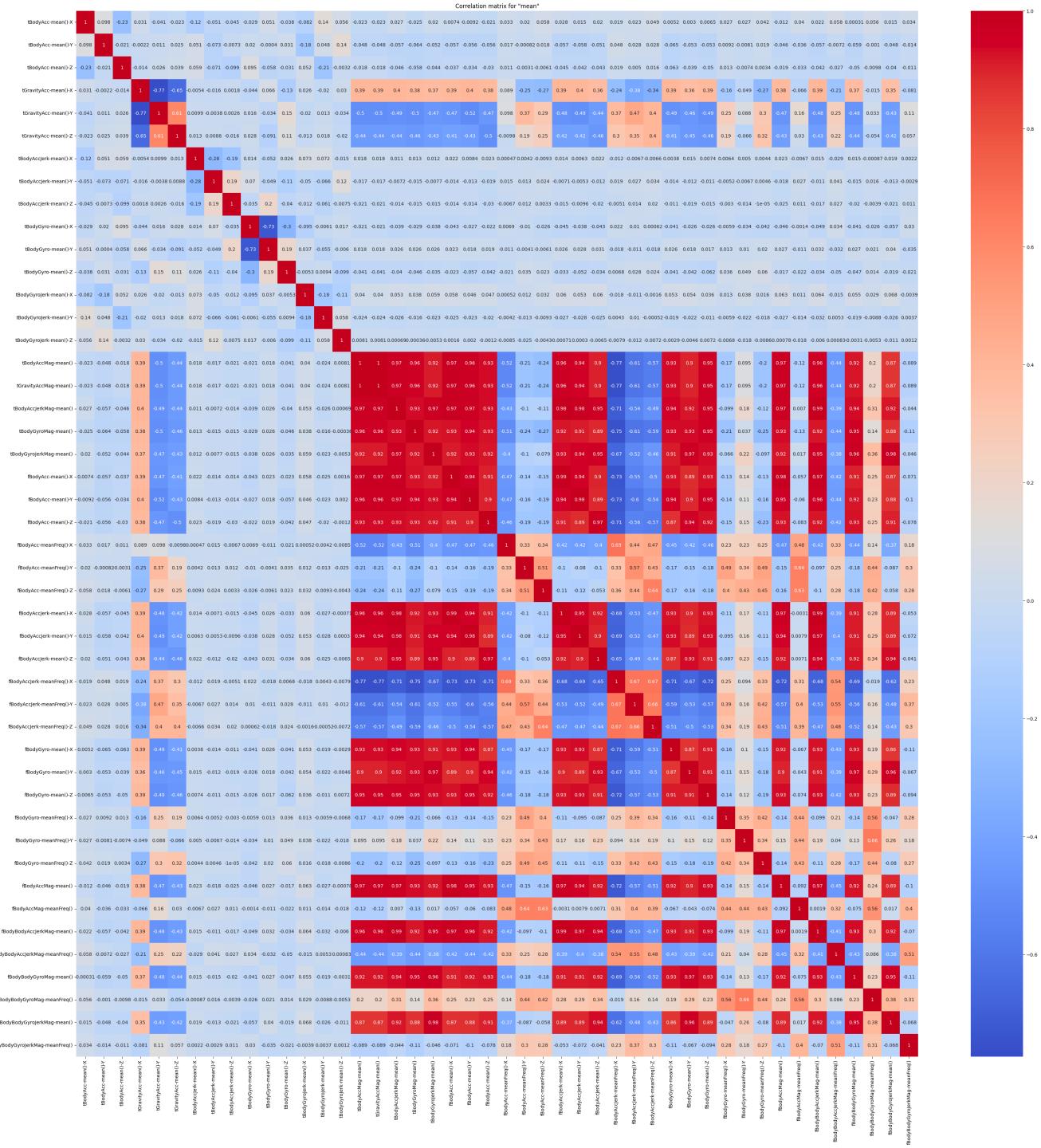
kierunki:

- X
- Y
- Z

rozkłady kolumn zawierających średnie



macierz korelacji dla średnich

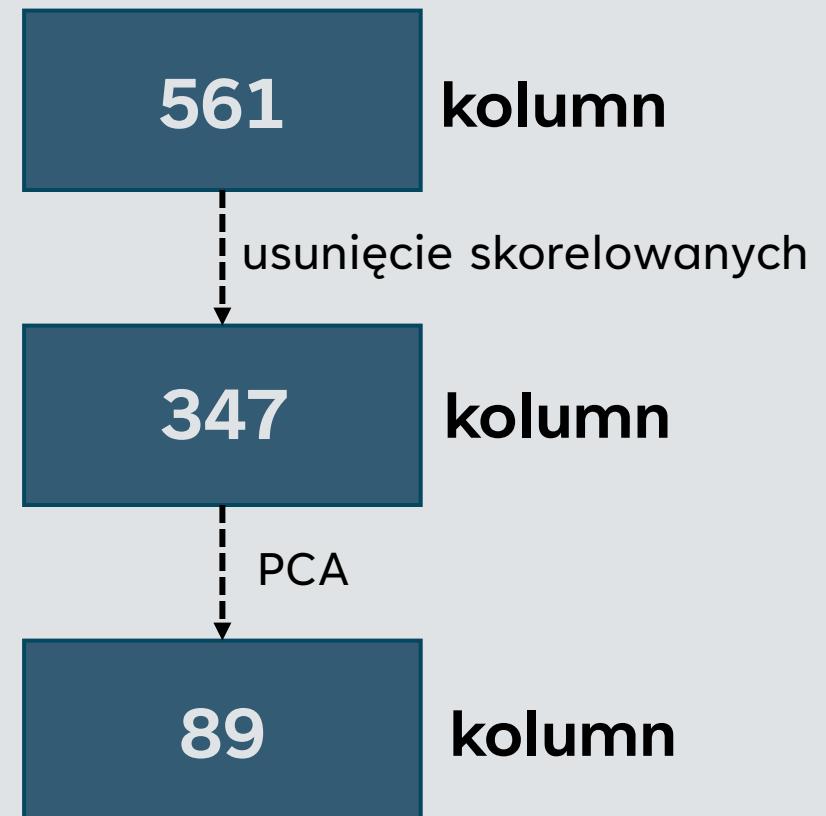


PREPROCESSING

PREPROCESSING

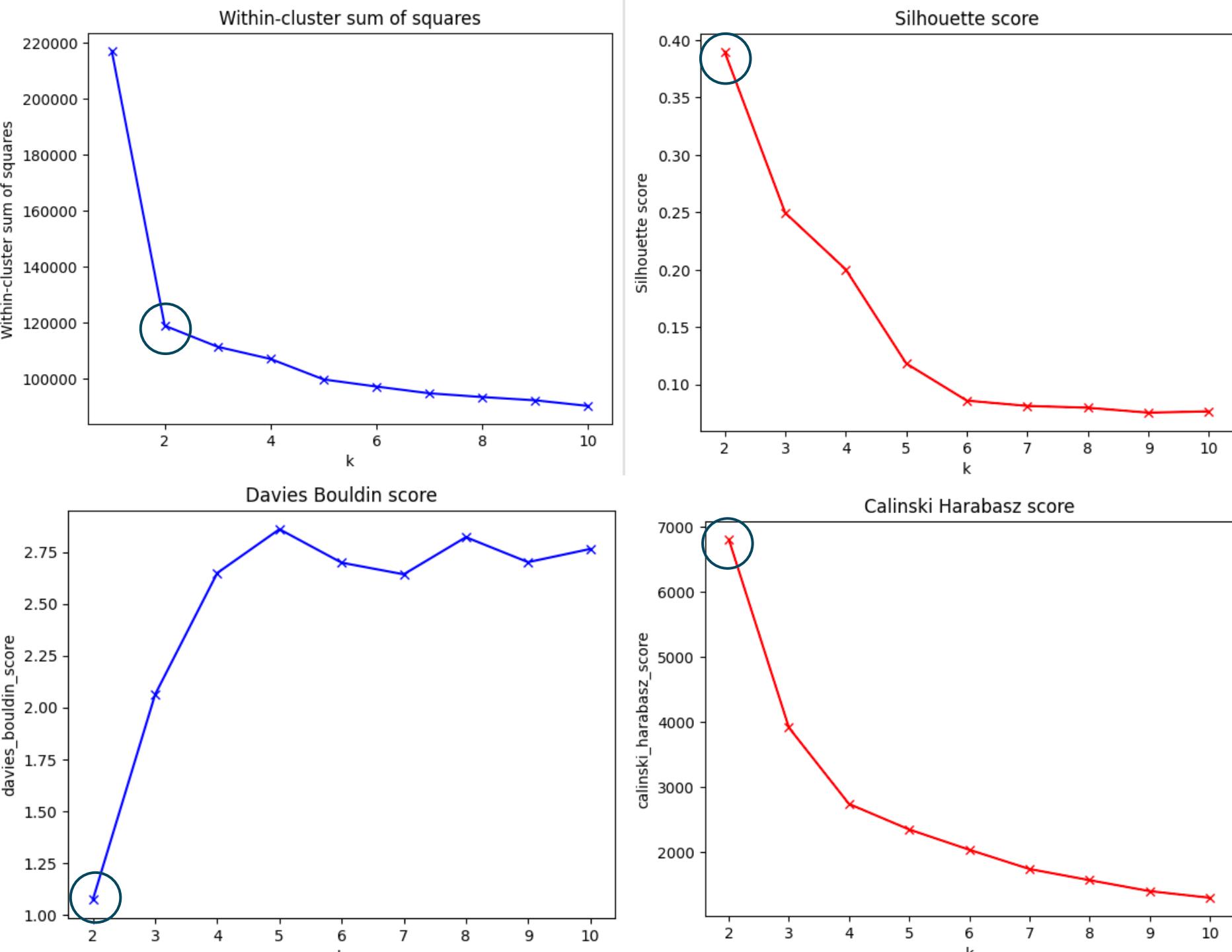
Redukcja liczby cech

Aby zmniejszyć liczbę kolumn, postanowiliśmy usunąć te o najwyższej korelacji (>98%), a następnie zastosować PCA (zachowując 95% wariancji).



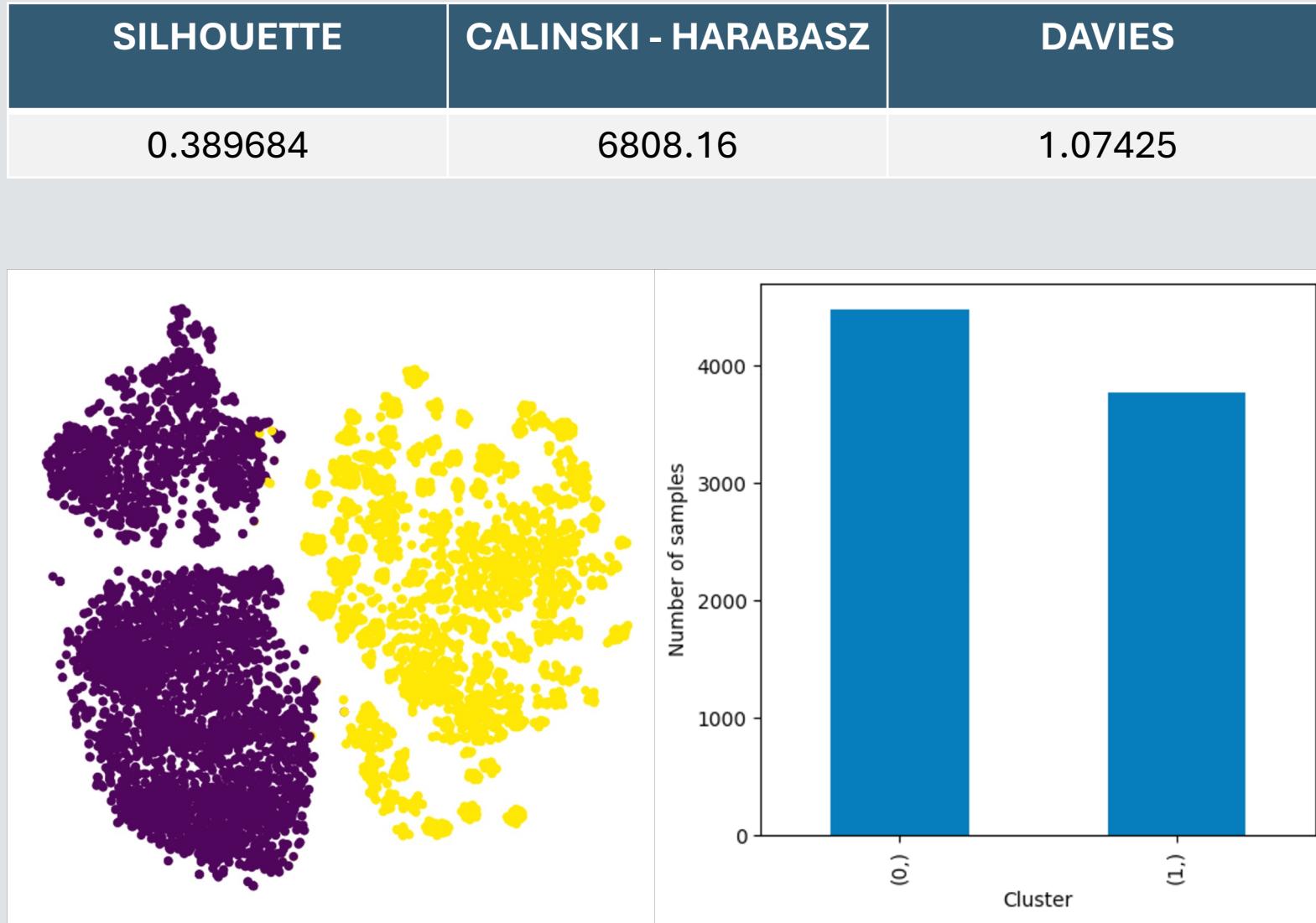
MODELWANIE

K-MEANS



MODELWANIE

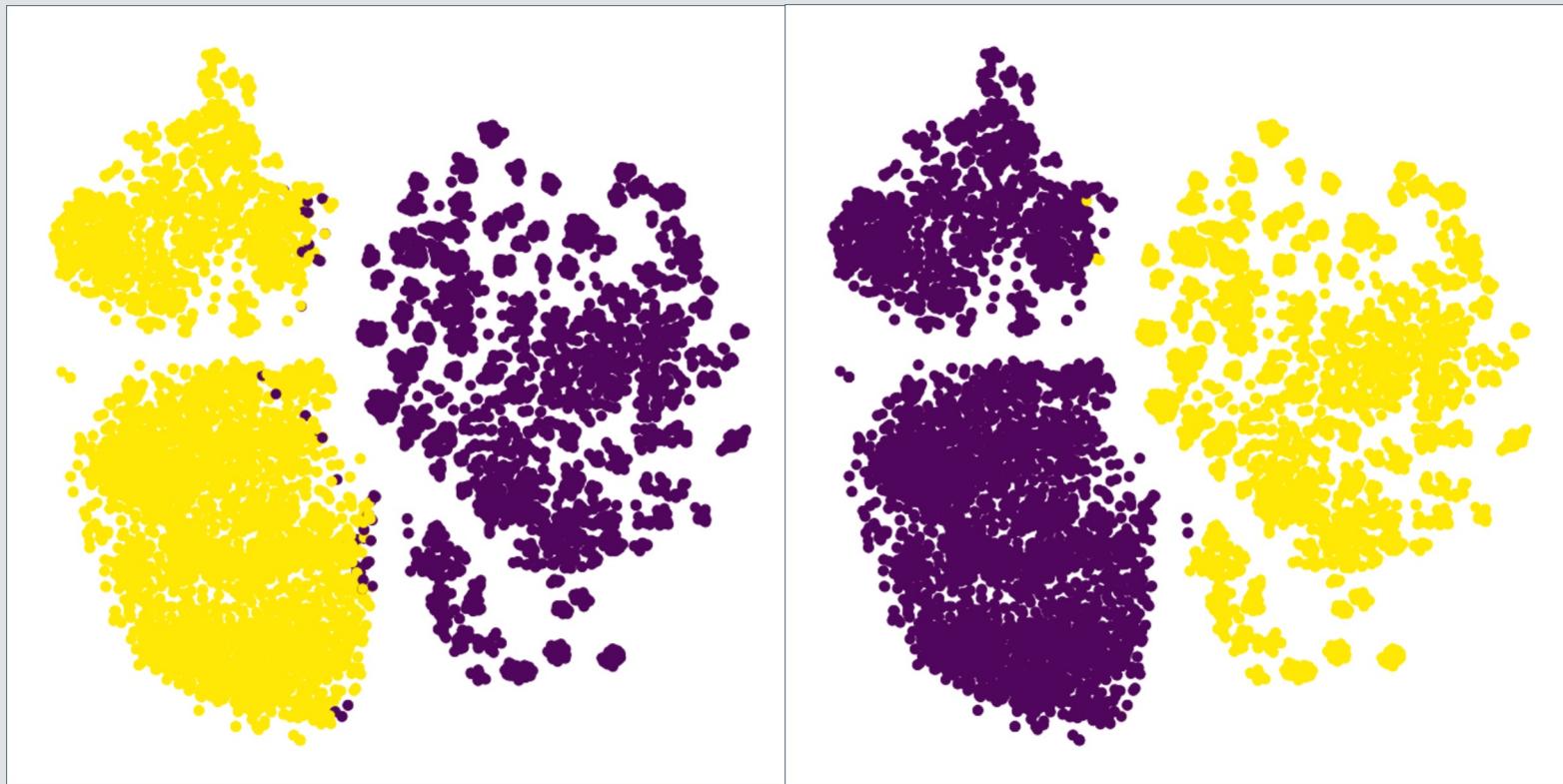
K-MEANS



MODELOWANIE

K-MEDOIDS, GAUSSIAN

	SILHOUETTE	CALINSKI - HARABASZ	DAVIES
K-MEDOIDS	0.388432	6775.65	1.07717
GAUSSIAN	0.389386	6796.79	1.07480

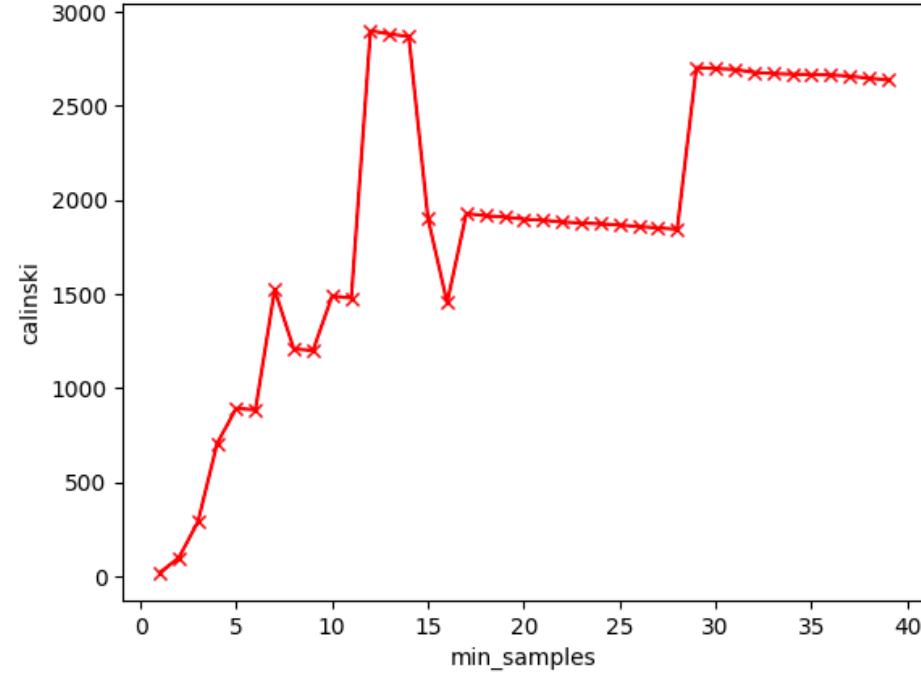
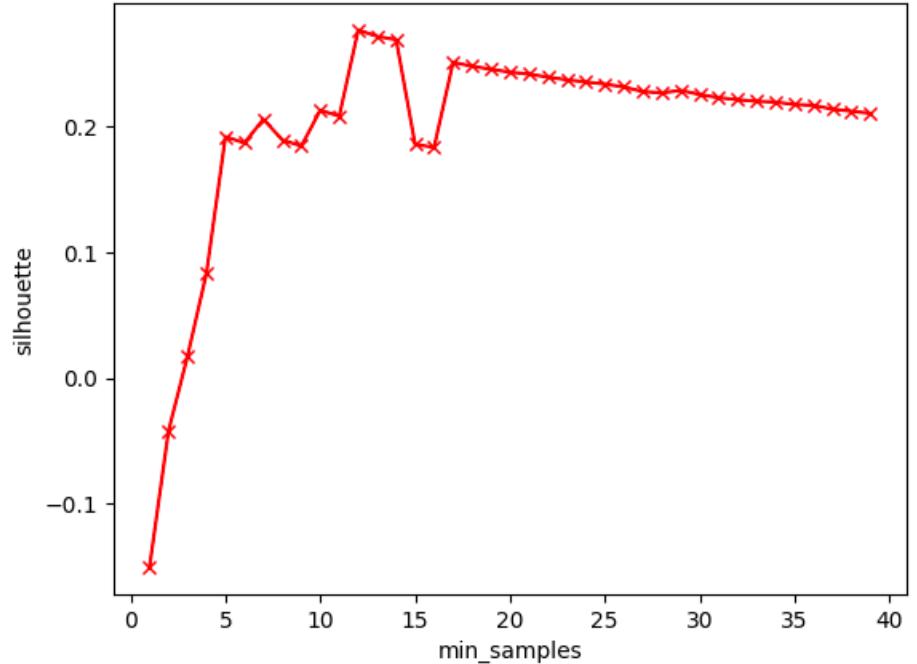
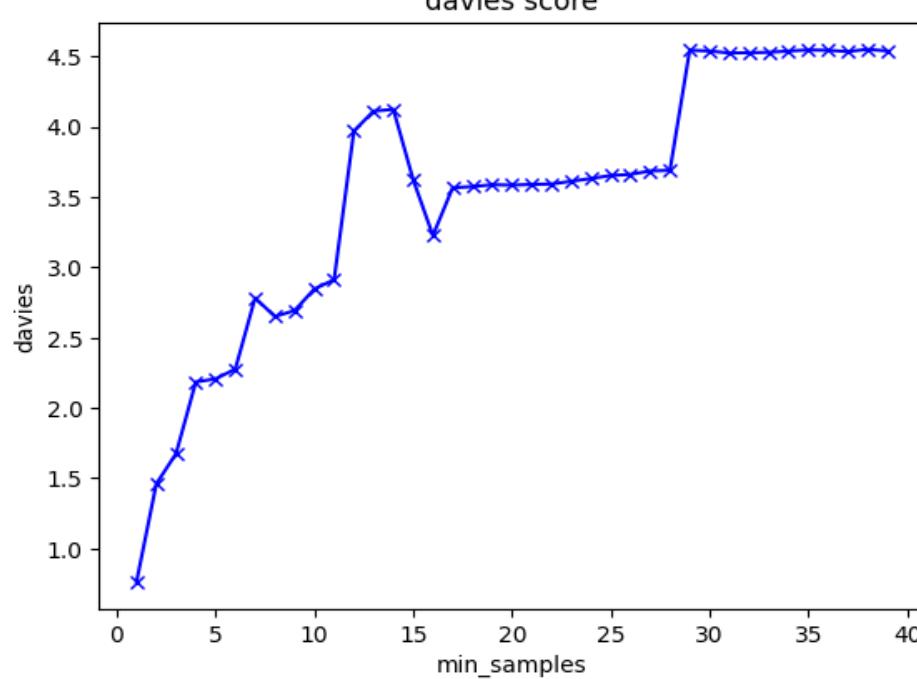


K-MEDOIDS

GAUSSIAN

MODELOWANIE

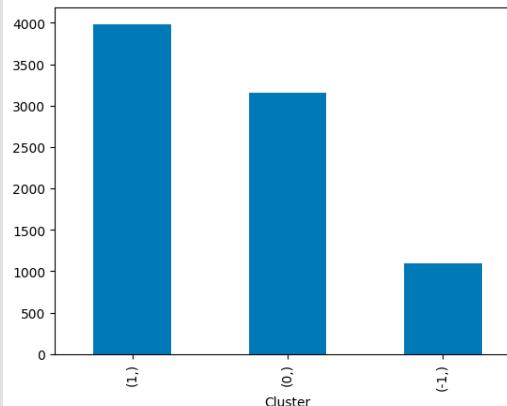
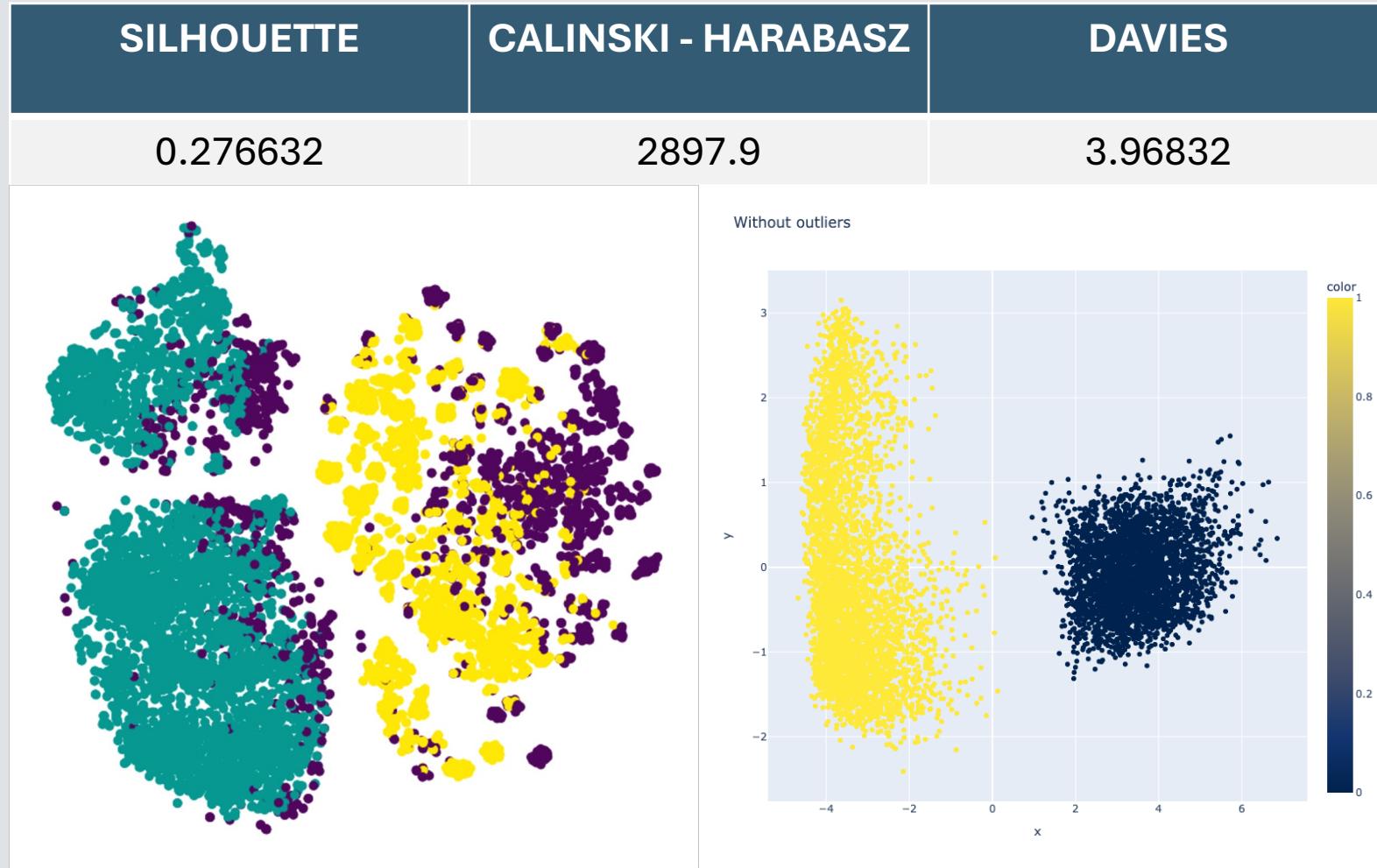
DBSCAN



Wybrałyśmy $\text{min_samples}=12$

MODELOWANIE

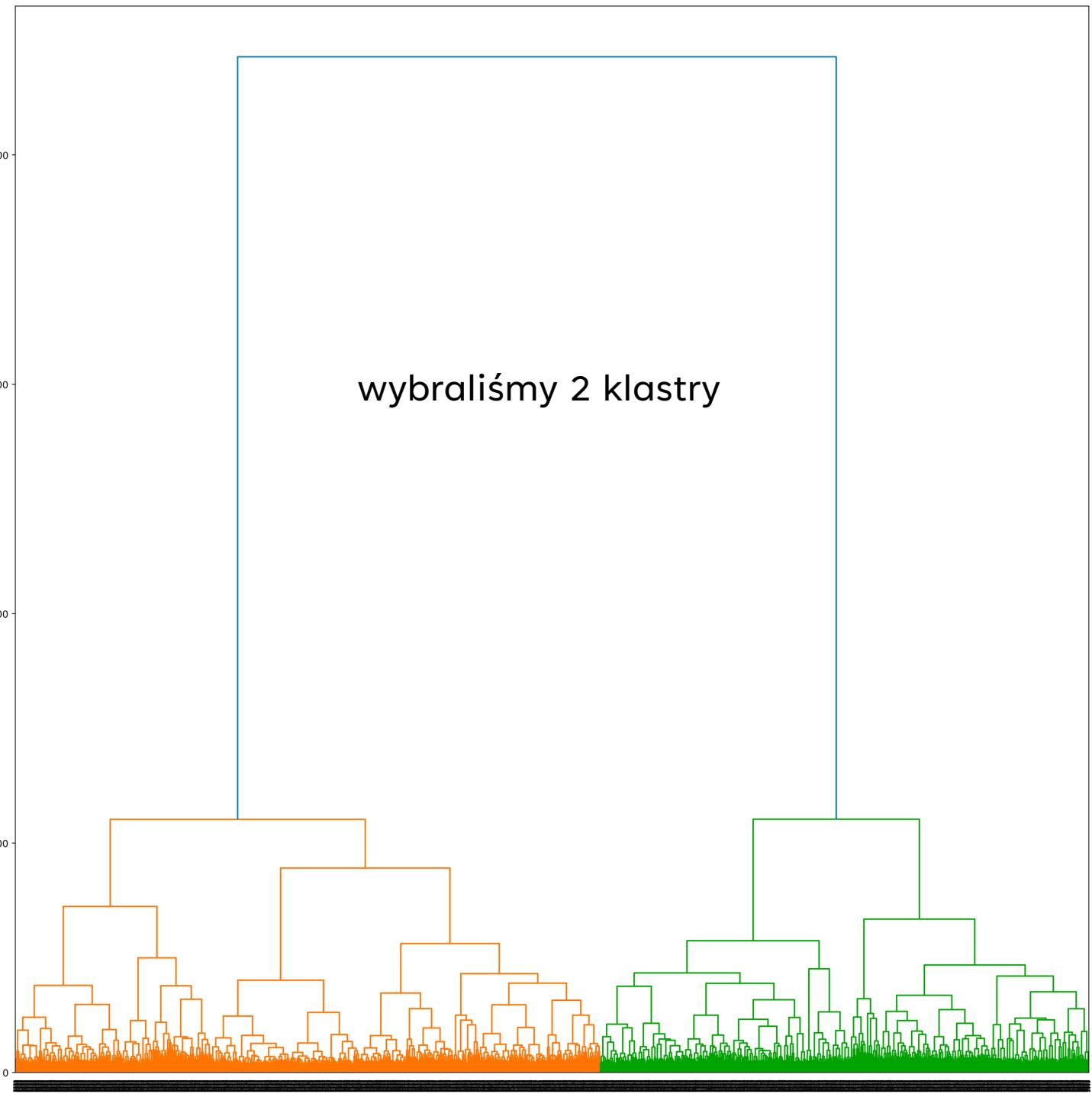
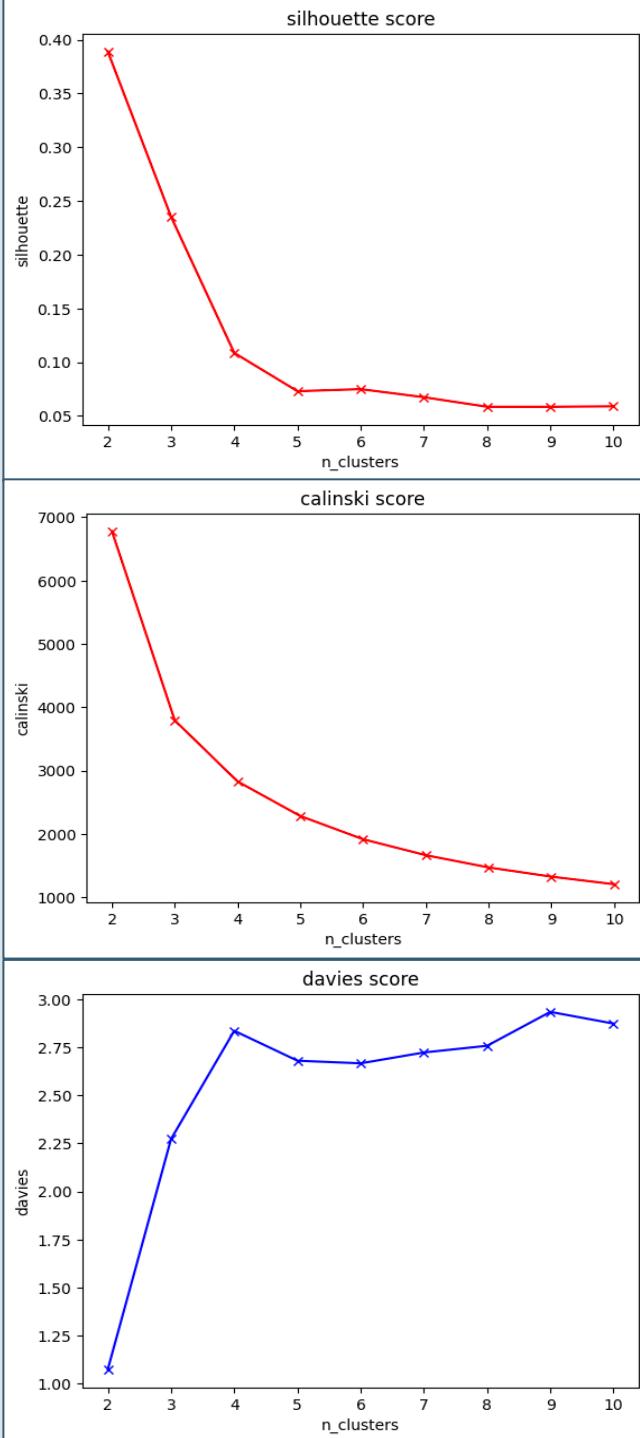
DBSCAN



model odrzuca
ponad 10%
obiektów...

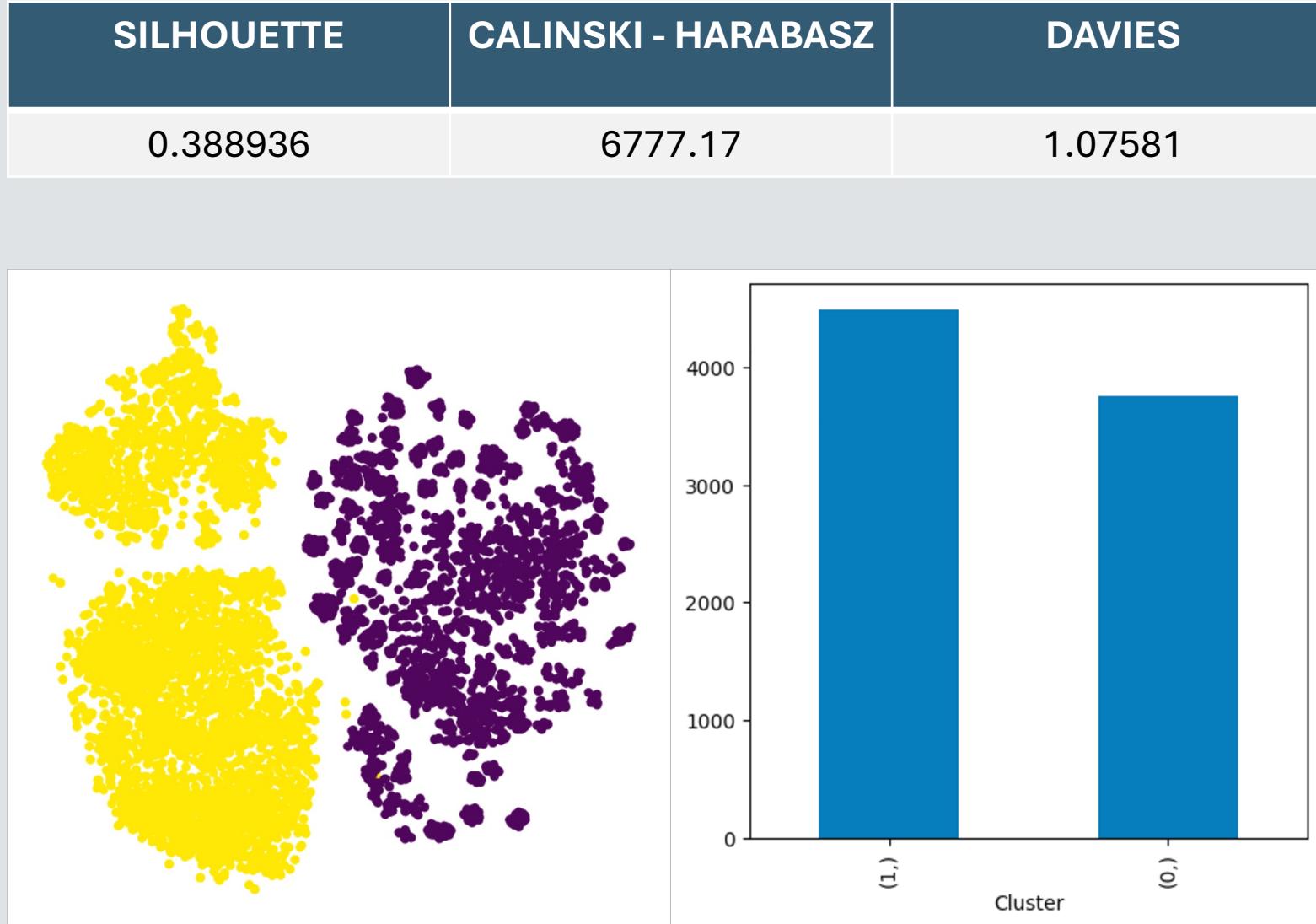
MODELOWANIE

AGGLOMERATIVE



MODELOWANIE

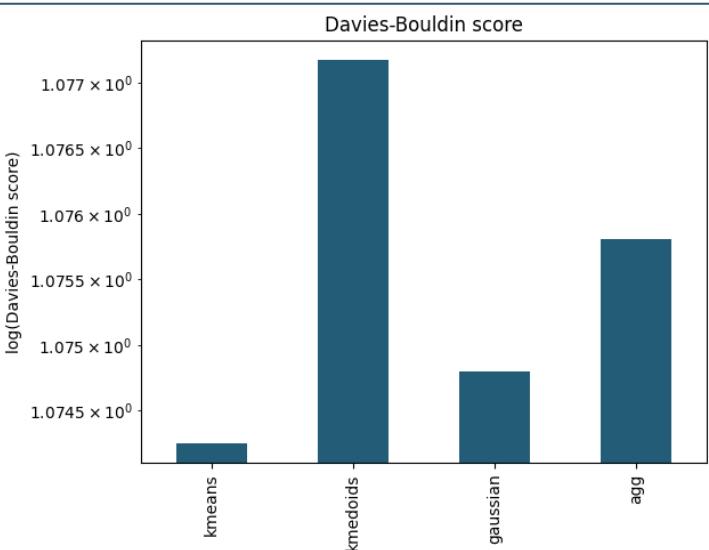
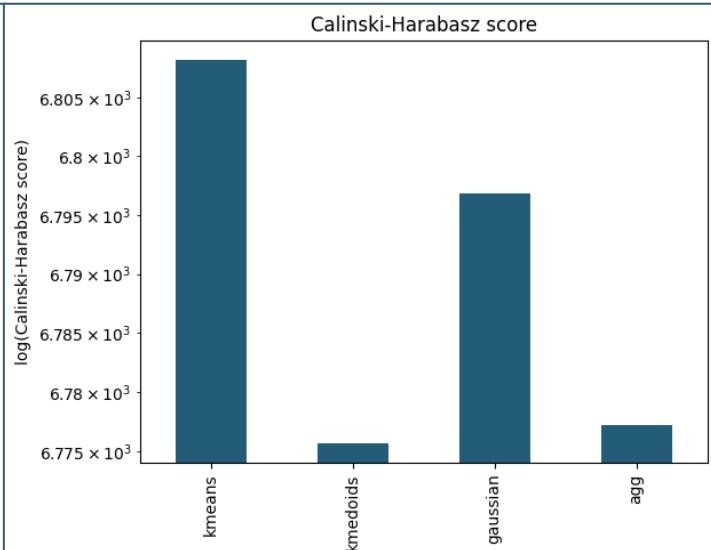
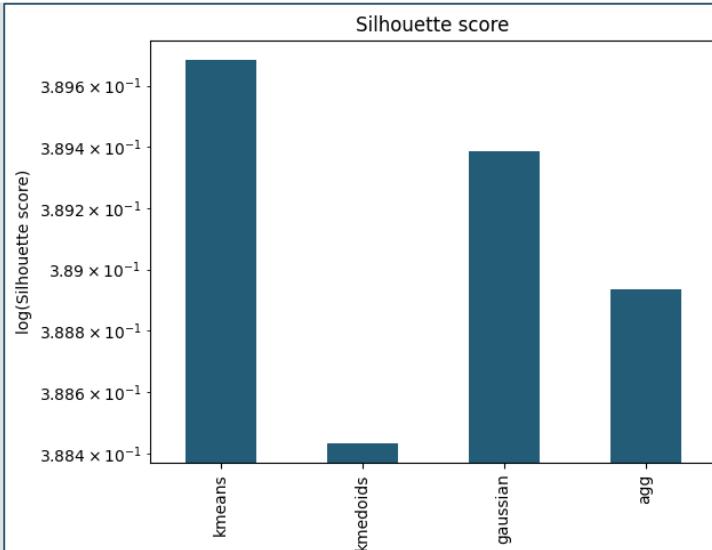
AGGLOMERATIVE



WYNIKI

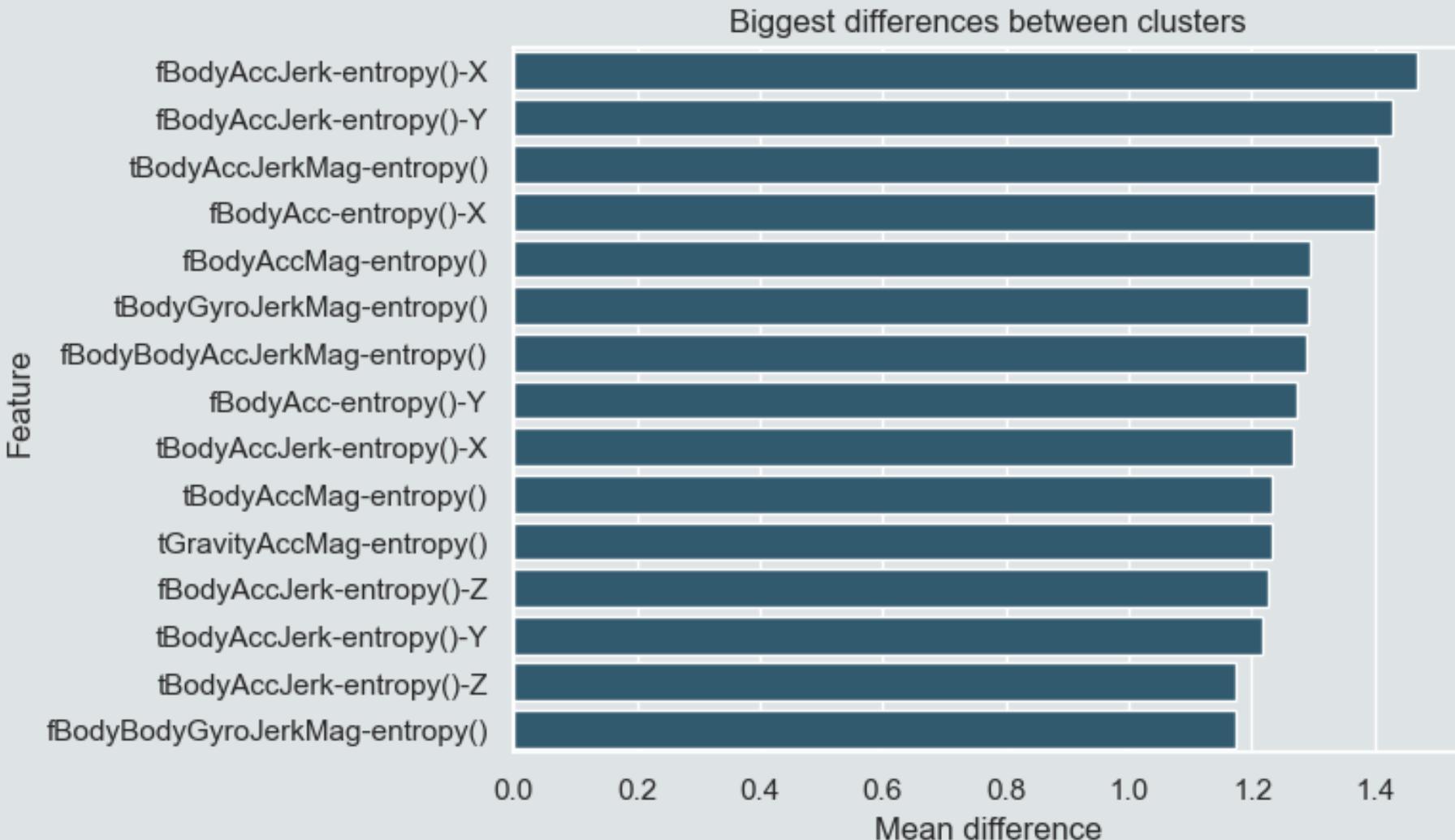
WYNIKI

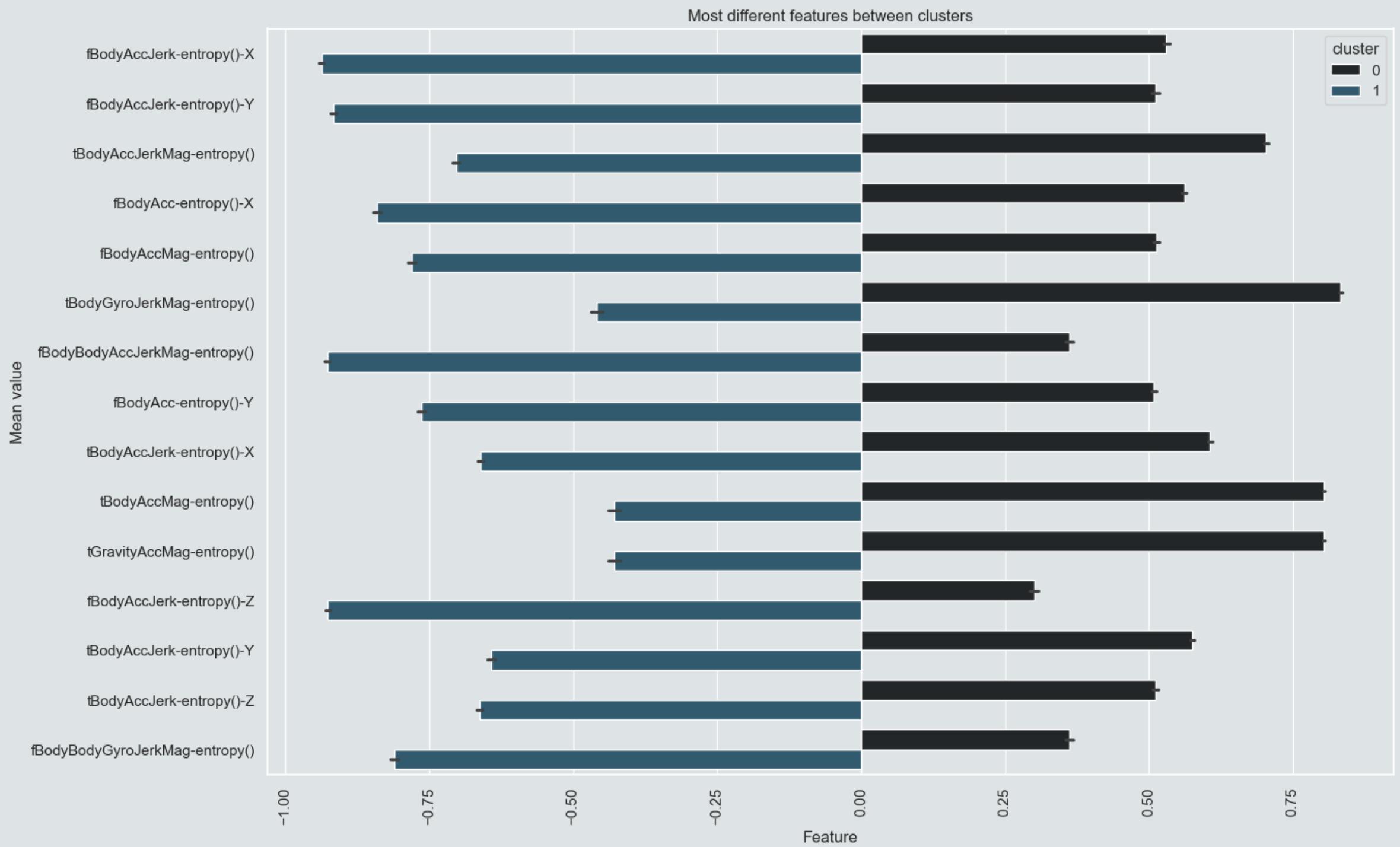
	SILHOUETTE	CALINSKI - HARABASZ	DAVIES - BOULDIN
K-MEANS	0.389684	6808.16	1.07425
K-MEDOIDS	0.388432	6775.65	1.07717
GAUSSIAN	0.389386	6796.79	1.07480
DBSCAN	0.276632	2897.9	3.96832
AGGLOMERATIVE	0.388936	6777.17	1.07581



WYNIKI

RÓŻNICE MIĘDZY KLASTRAMI







mniej aktywni

bardziej aktywni

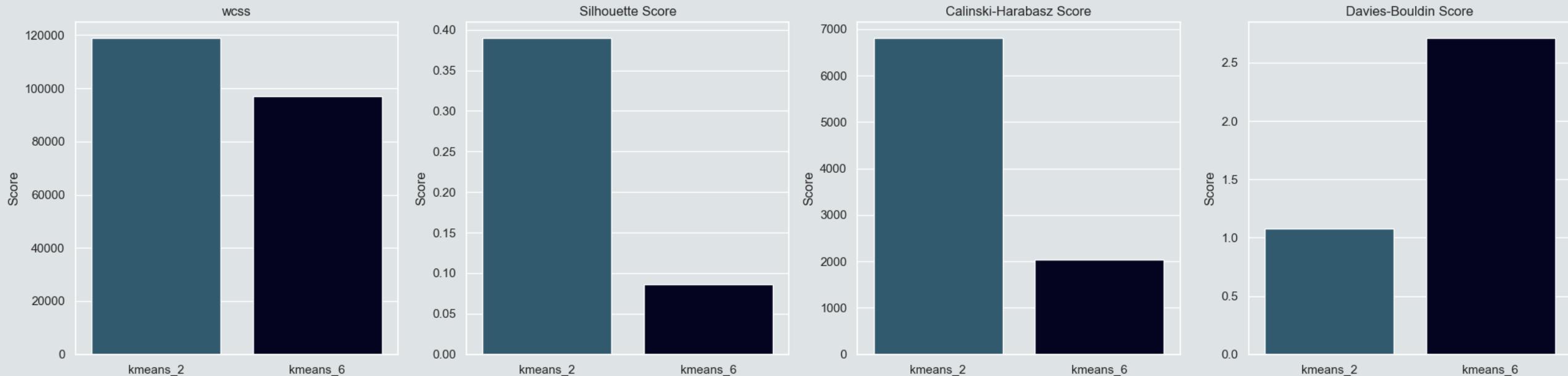


WNIOSZEK

MODEL ROZRÓŻNIA CZYNNOŚCI STATYCZNE OD DYNAMICZNYCH

TEST SZEŚCIU KLASTRÓW

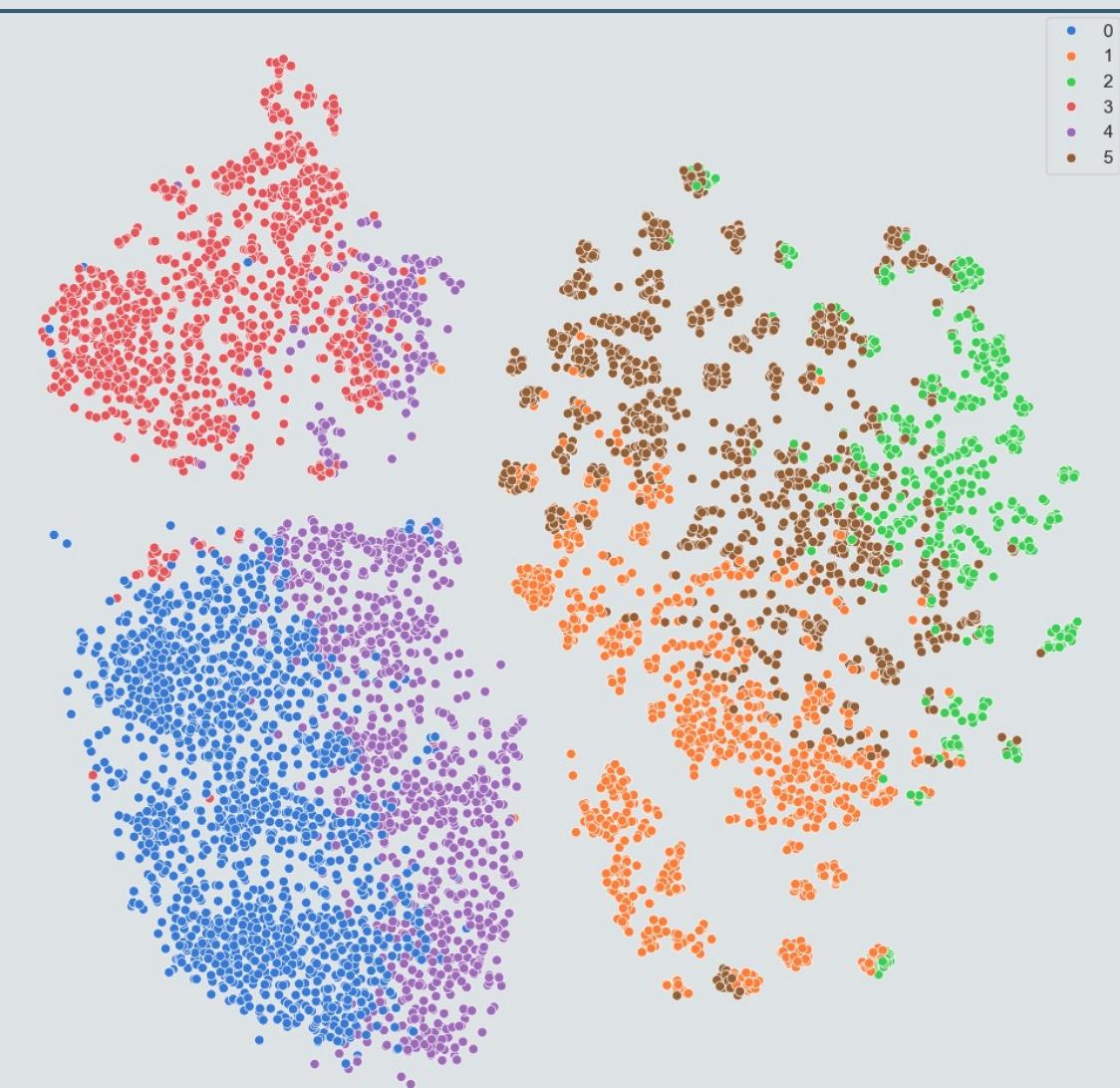
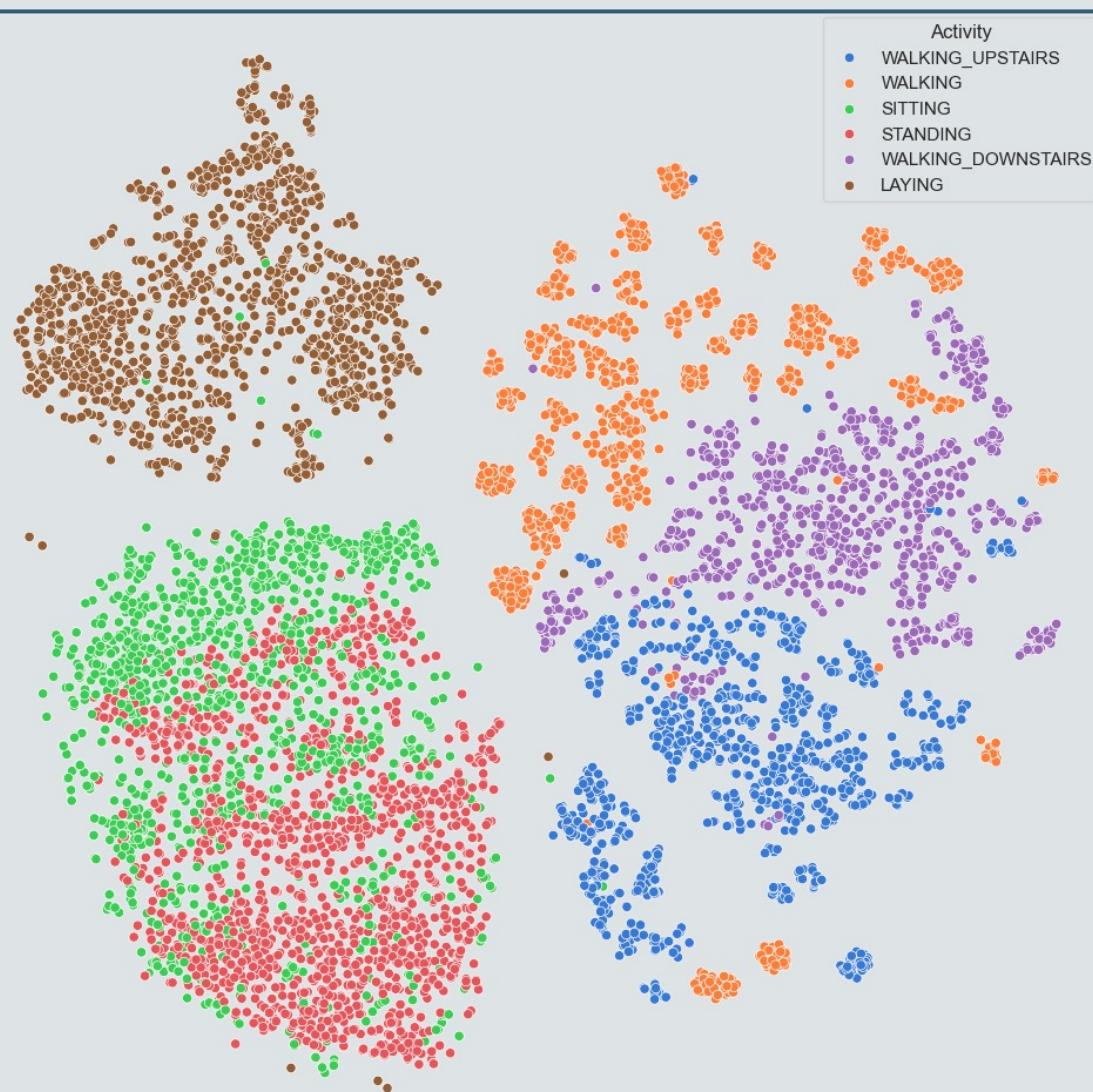
TEST SZEŚCIU KLASTRÓW



TEST SZEŚCIU KLASTRÓW



TEST SZEŚCIU KLASTRÓW



Co telefony wiedzą o naszej aktywności

Projekt z klasteryzacji

Florek Paweł, Deptuch Mateusz