# ks. Teemu + silhuetti…

# Klusterointi useilla eri muuttujilla. Valitse yksi seuraavista:

1)

appreciation\_v3.csv

Ryhmittele kiinnostuksen kohteiden mukaan

2)

Women Dresses Reviews Dataset.csv

Ryhmittele tuotteet ja asiakkaat yhteisten ominaisuuksien perusteella

[K-means Algorithm With Multiple features](https://www.kaggle.com/code/channaweerakoon/k-means-algorithm-with-multiple-features)

3)

[Wholesale customers - UCI Machine Learning Repository](https://archive.ics.uci.edu/dataset/292/wholesale+customers)

[DBSCAN vs K-Means: visualizing the difference](https://www.kaggle.com/code/ahmedmohameddawoud/dbscan-vs-k-means-visualizing-the-difference)

Ryhmittele tuotteet, kanavat ja alueet

Hae data itse.

4) Valitse oma data.

Huom! Nominaalimuuttujia (vaikka olisivat numeerisia) ei voi käyttää klusteroinnissa, koska ne vääristävät etäisyyksiä

Kategoriset/nominaalimuuttujat voi joko jättää pois tai muokata niitä one-hot-encodingilla tai binary encodingilla eli muuttamalla ne binaarisiksi (0,1).

Muista skaalata aineisto!

Tee 2-3 hengen ryhmissä.