

THINKFUL Data Science

Capstone Project 3

Unsupervised Learning Analysis on Energy Consumption of Buildings in Chicago

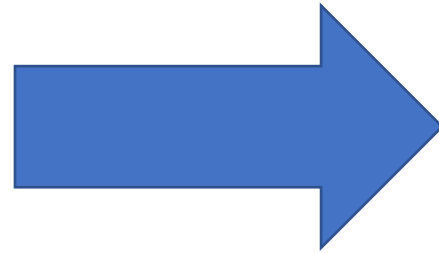
Henry Ssembatya

Energy

It's what powers the world.

Energy Sources

- Fossil Fuels
- Renewables
- Nuclear



- Electric Power
- Natural Gas
- Oil(Gas)

Dataset

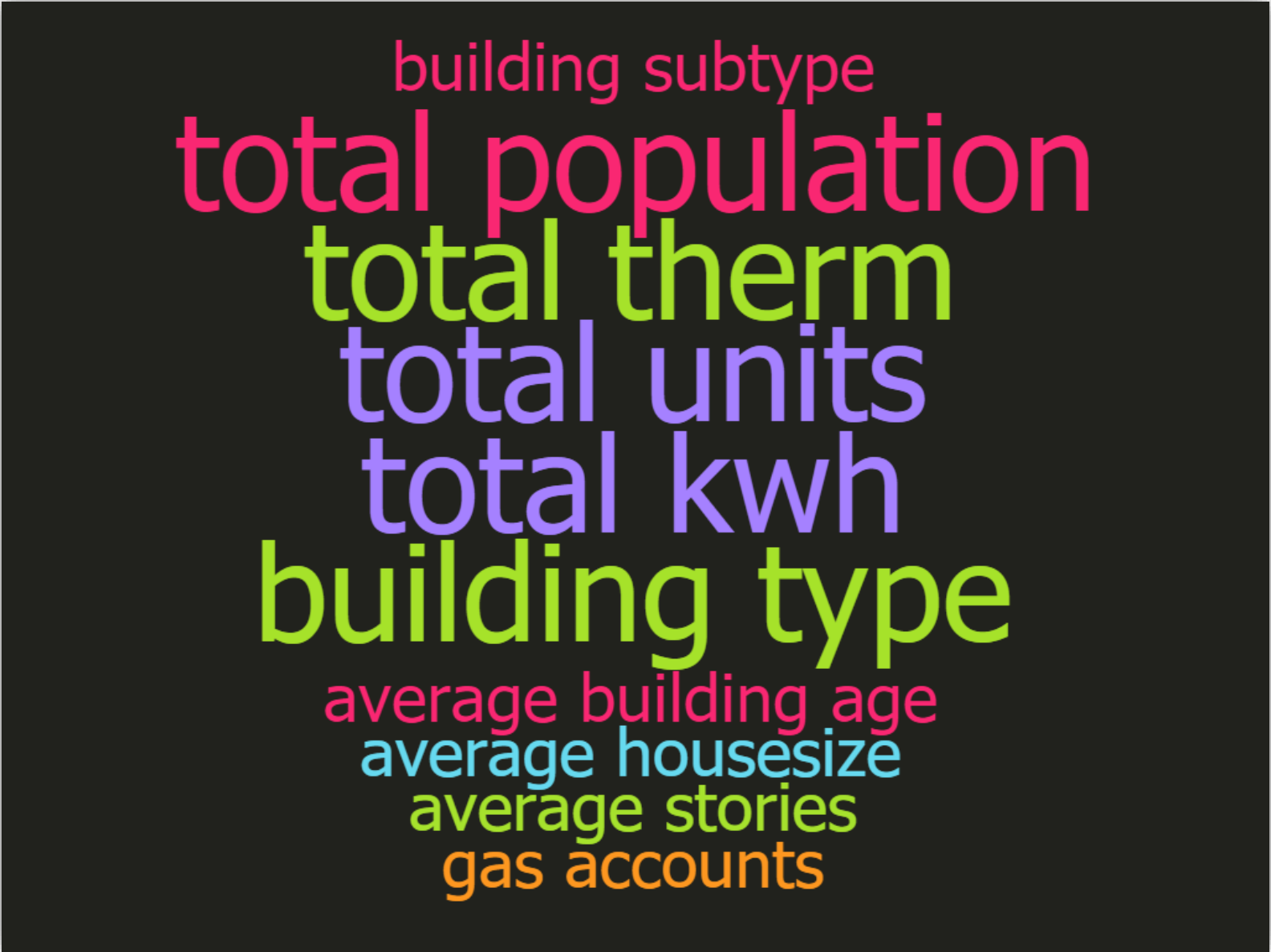
Source: Chicago Data Portal

Period: Jan-Dec 2010

Contents: Energy consumption (power and gas)

Size: 67,051 Observations, 73 Features

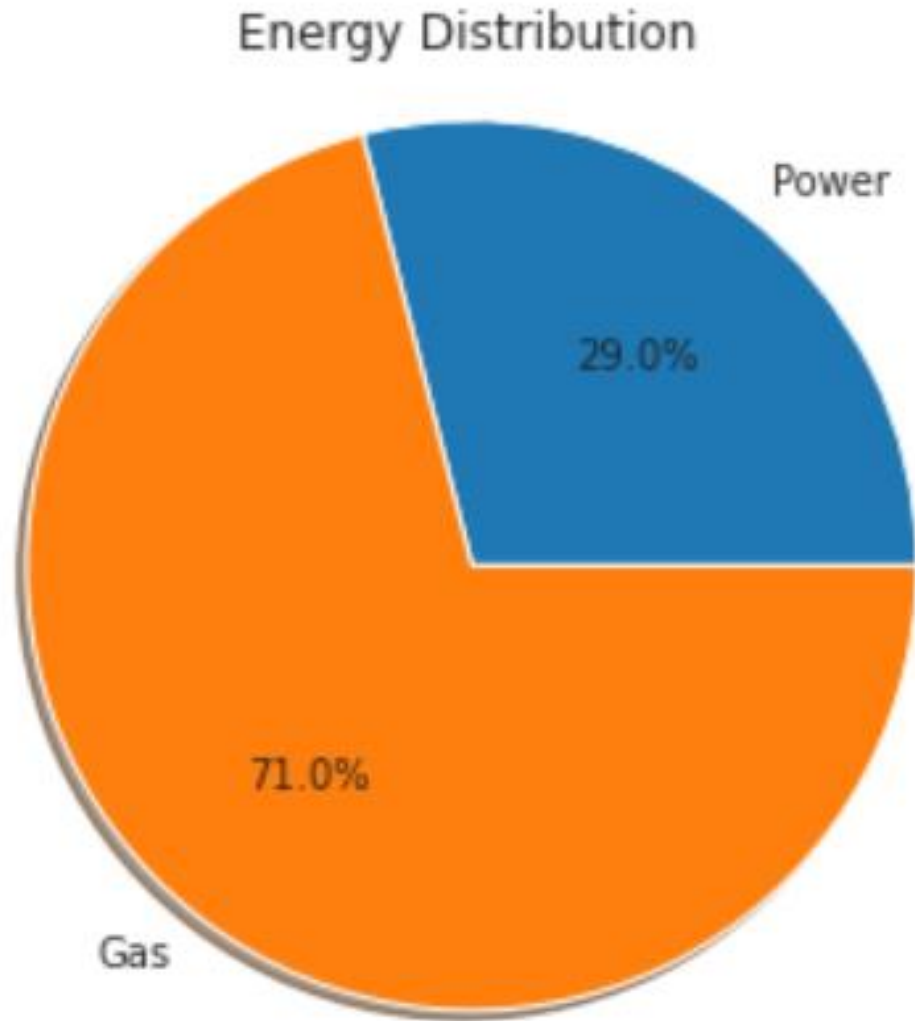
Some Features



A word cloud of building features on a dark background. The words are arranged in a vertical stack, with 'total population' being the largest and most prominent. Other large words include 'total therm', 'total units', 'total kwh', and 'building type'. Smaller words at the bottom include 'average building age', 'average housesize', 'average stories', and 'gas accounts'. The colors of the words vary, including shades of pink, green, purple, and orange.

building subtype
total population
total therm
total units
total kwh
building type
average building age
average housesize
average stories
gas accounts

Analysis



29%

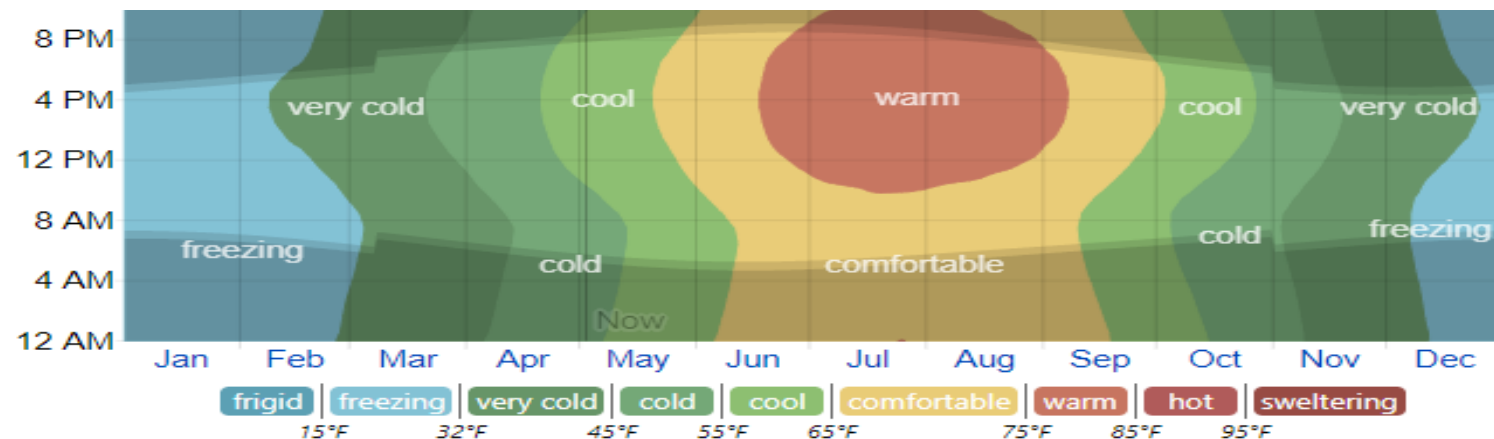
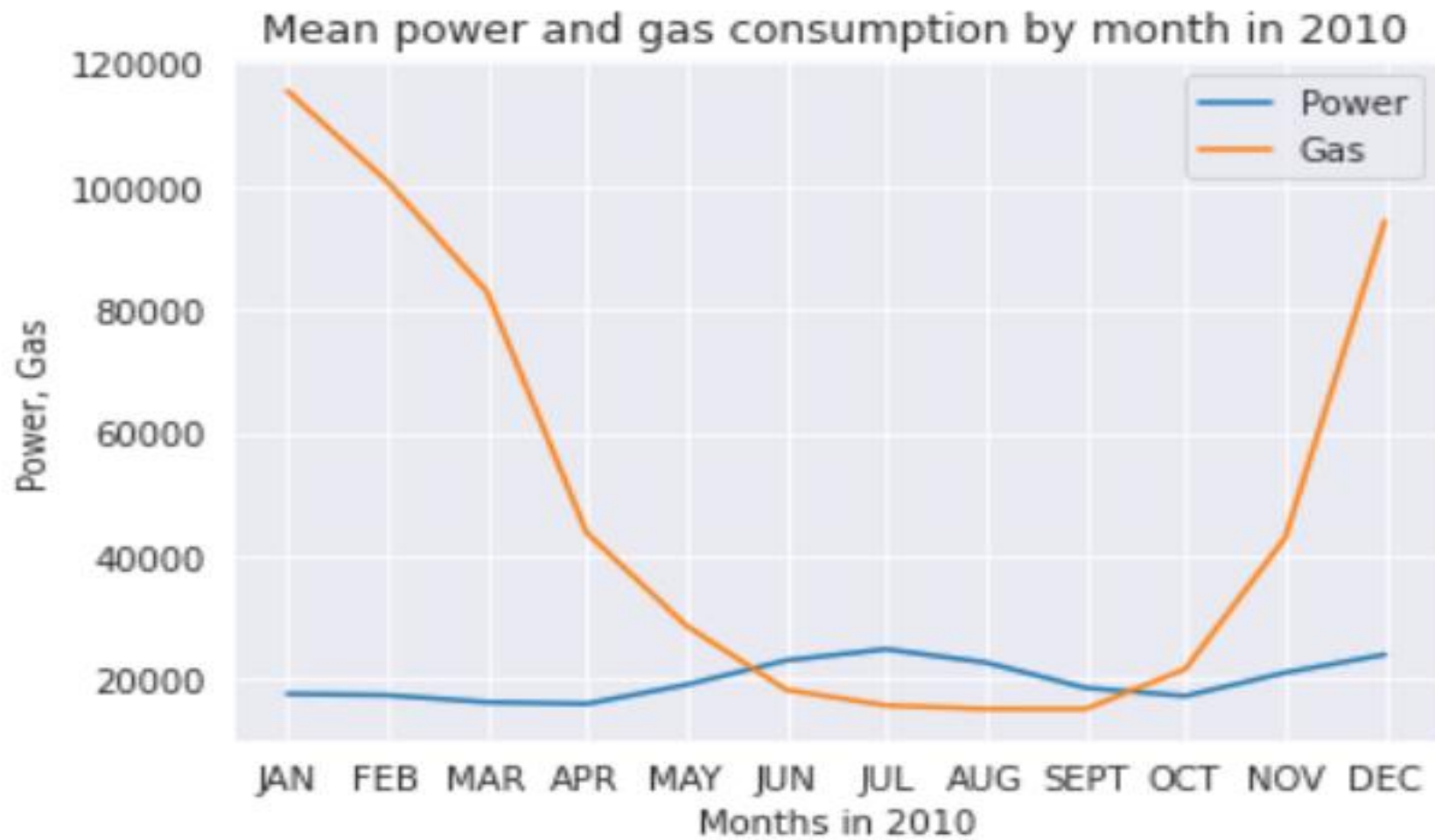


• Power

71%

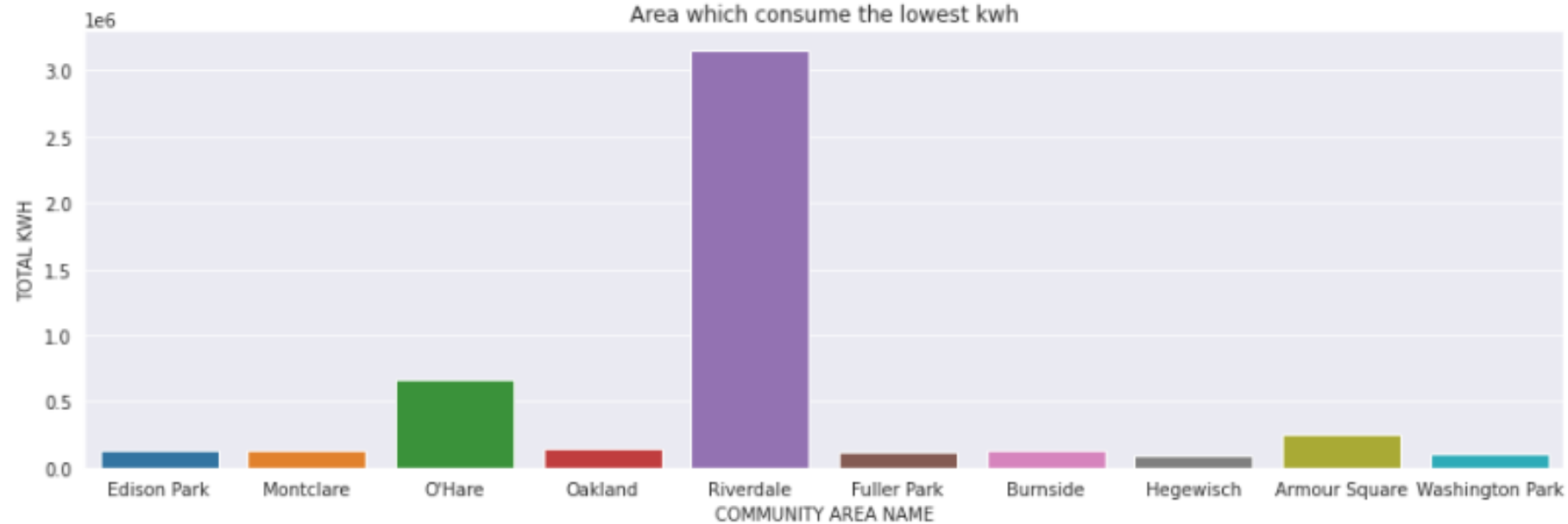
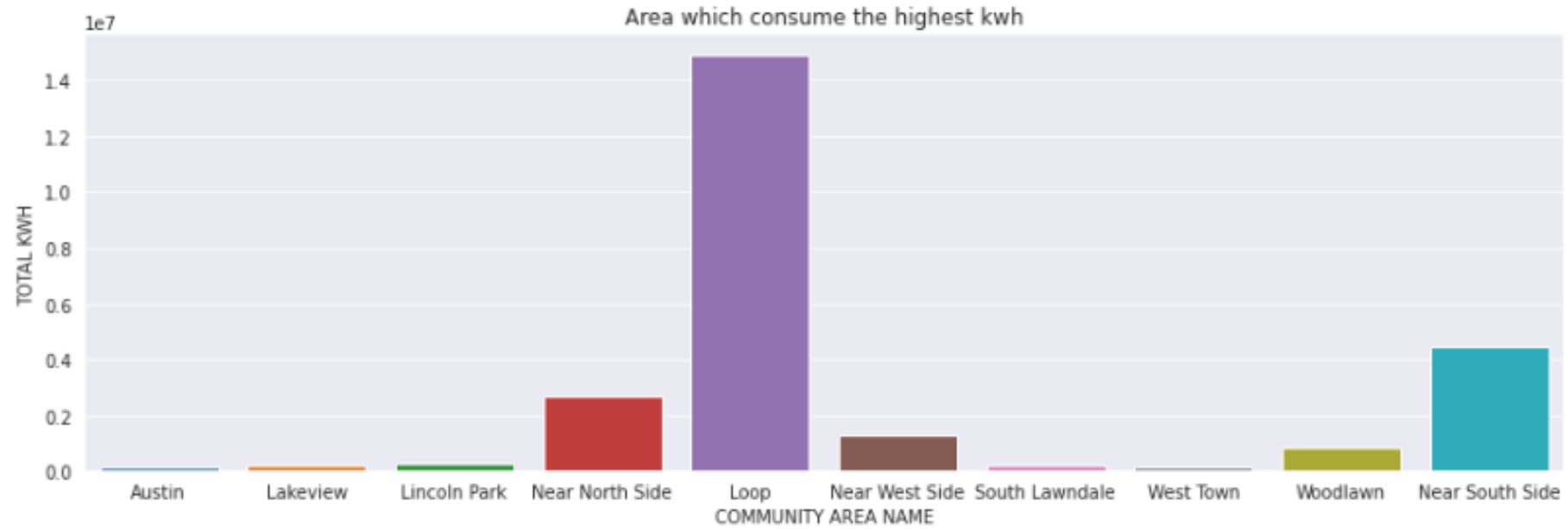


• Gas



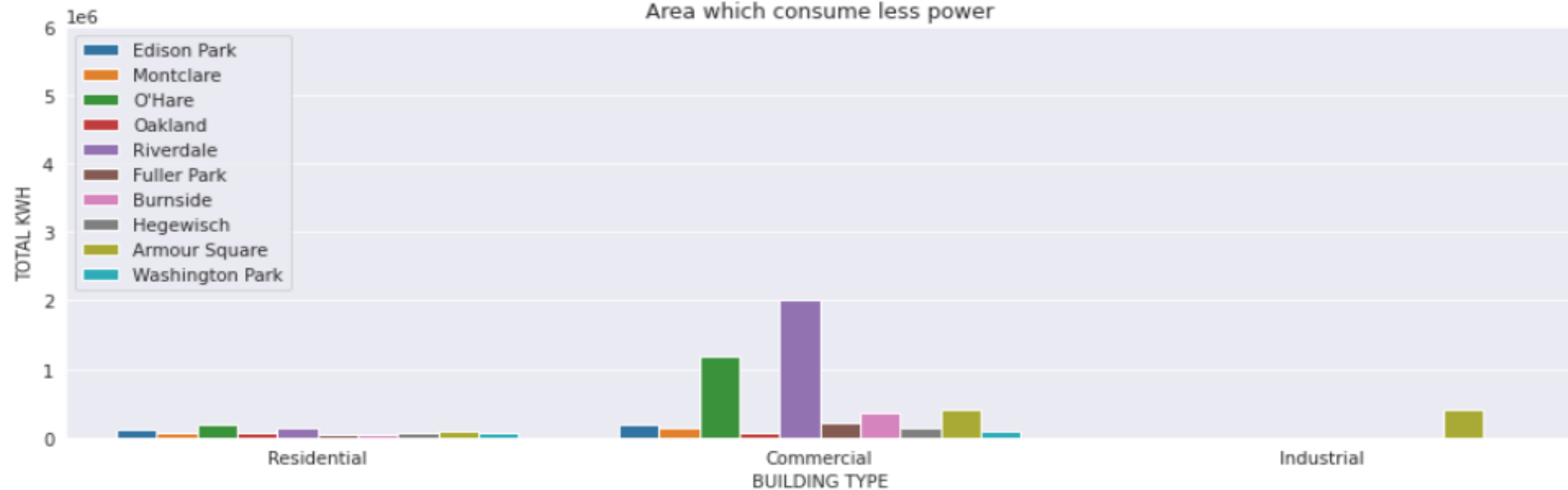
Source:
<https://weatherspark.com>

Energy Consumption

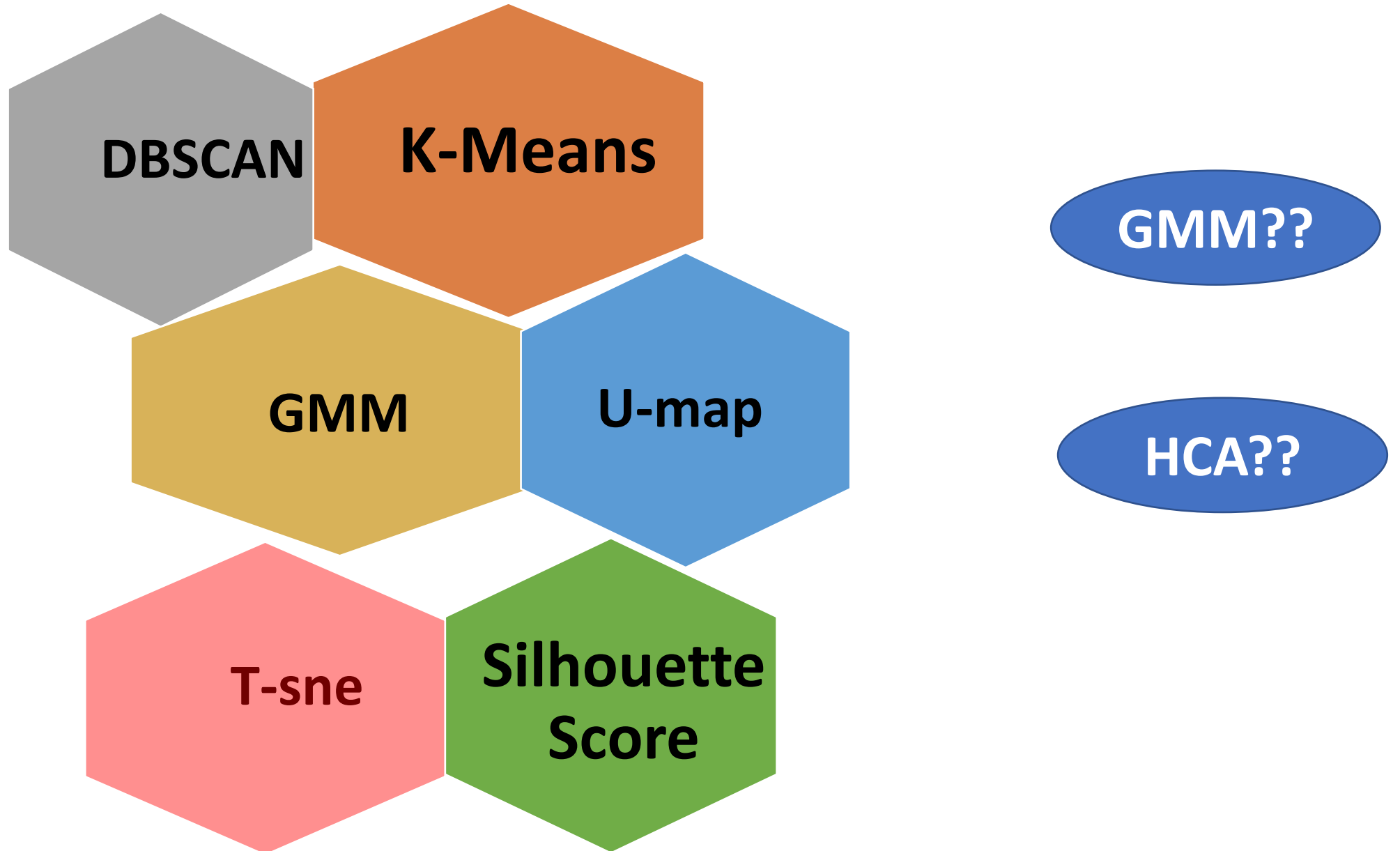


Energy Consumption

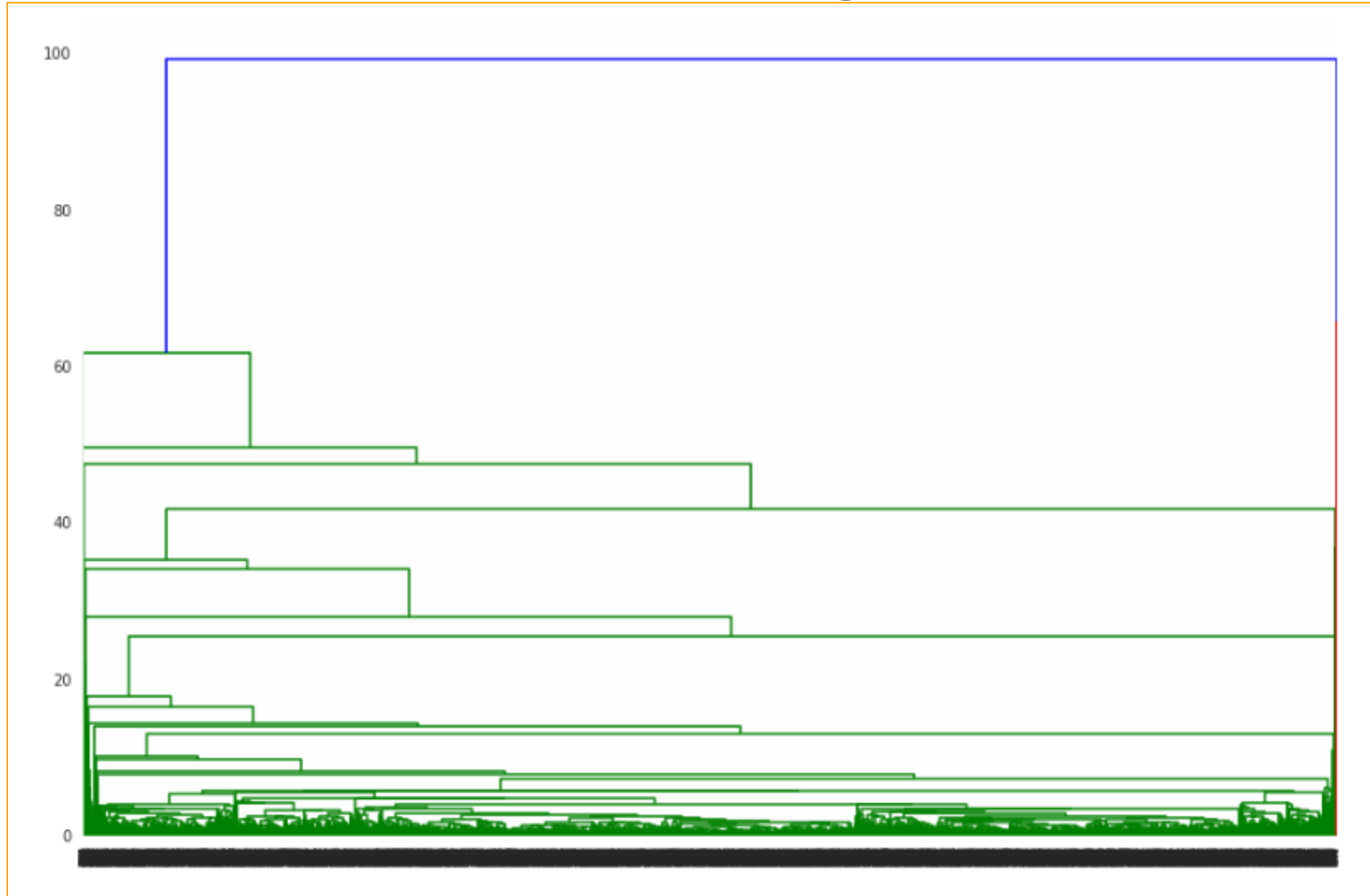
Text(0.5, 1.0, 'Area which consume less power')



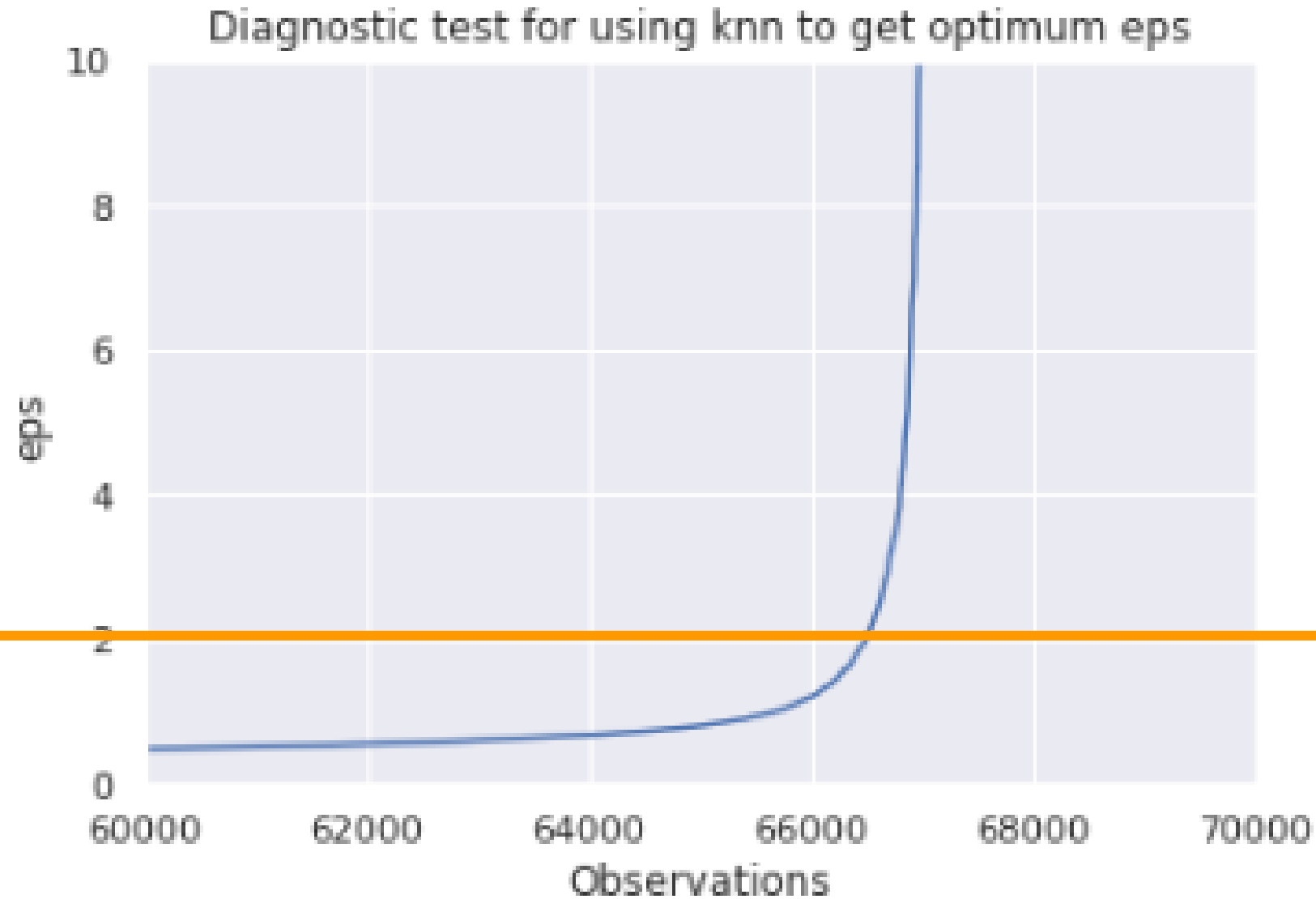
Clustering Analysis



HCA? Why not?

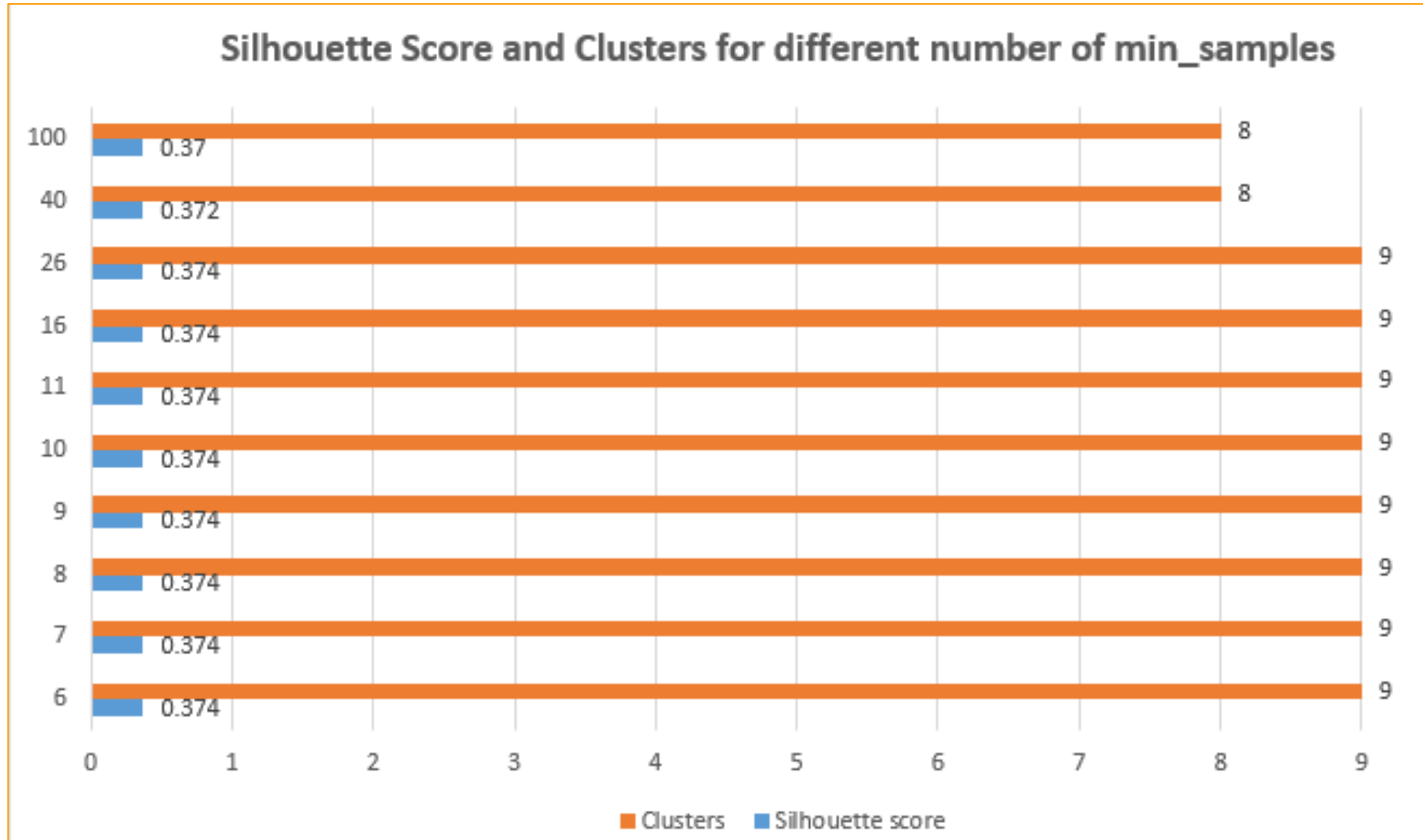


Clusters



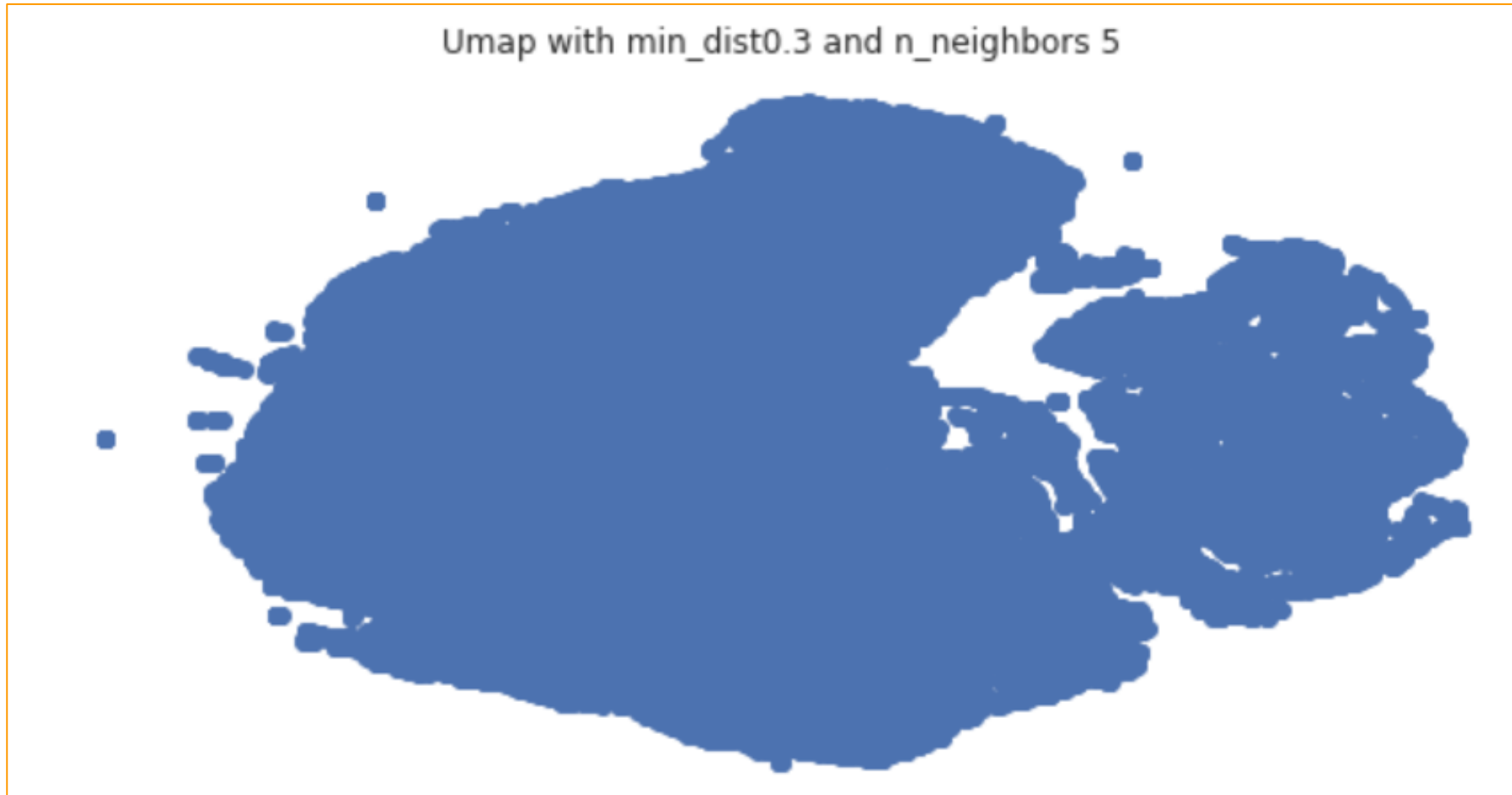
Results

DBSCAN



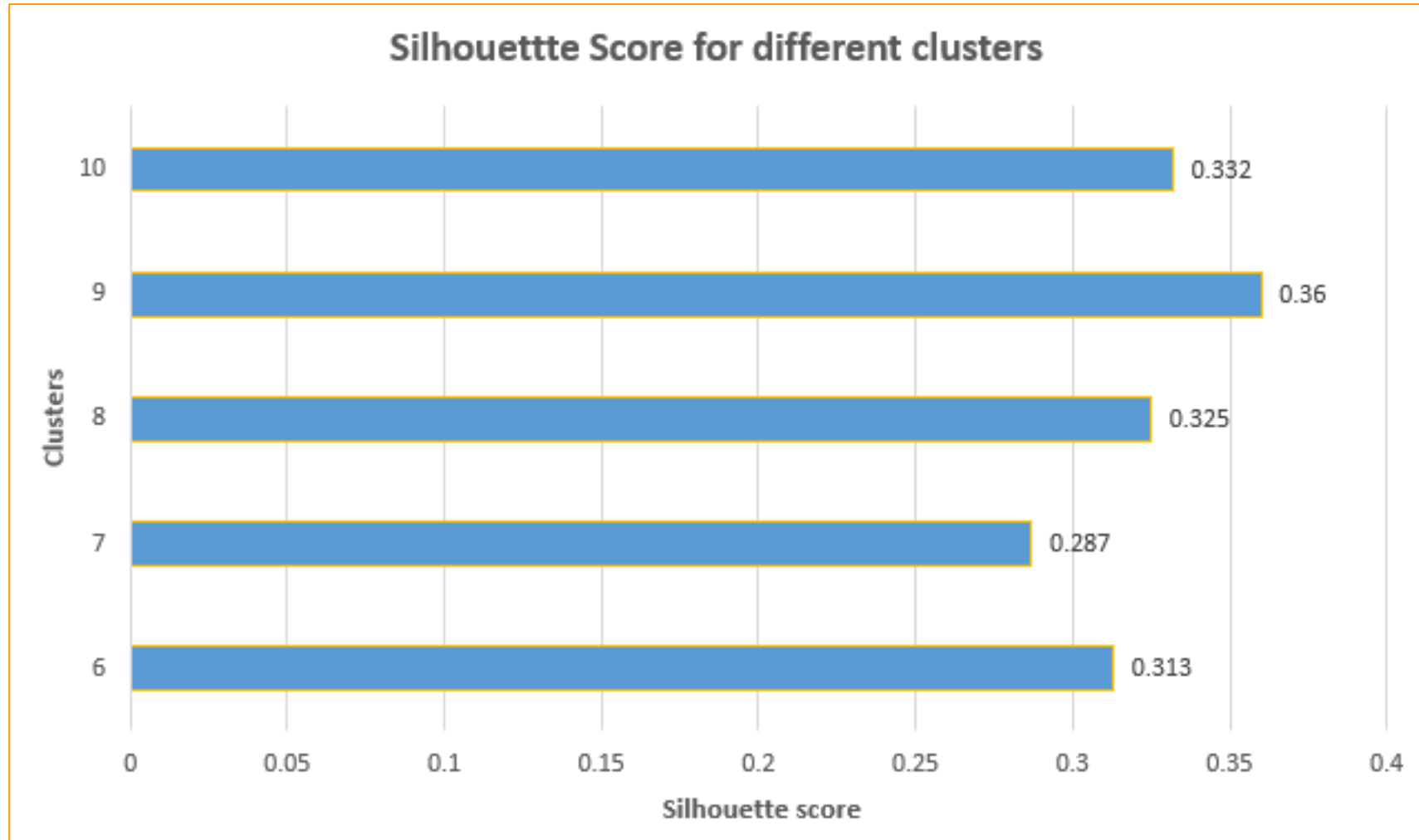
Results

UMAP



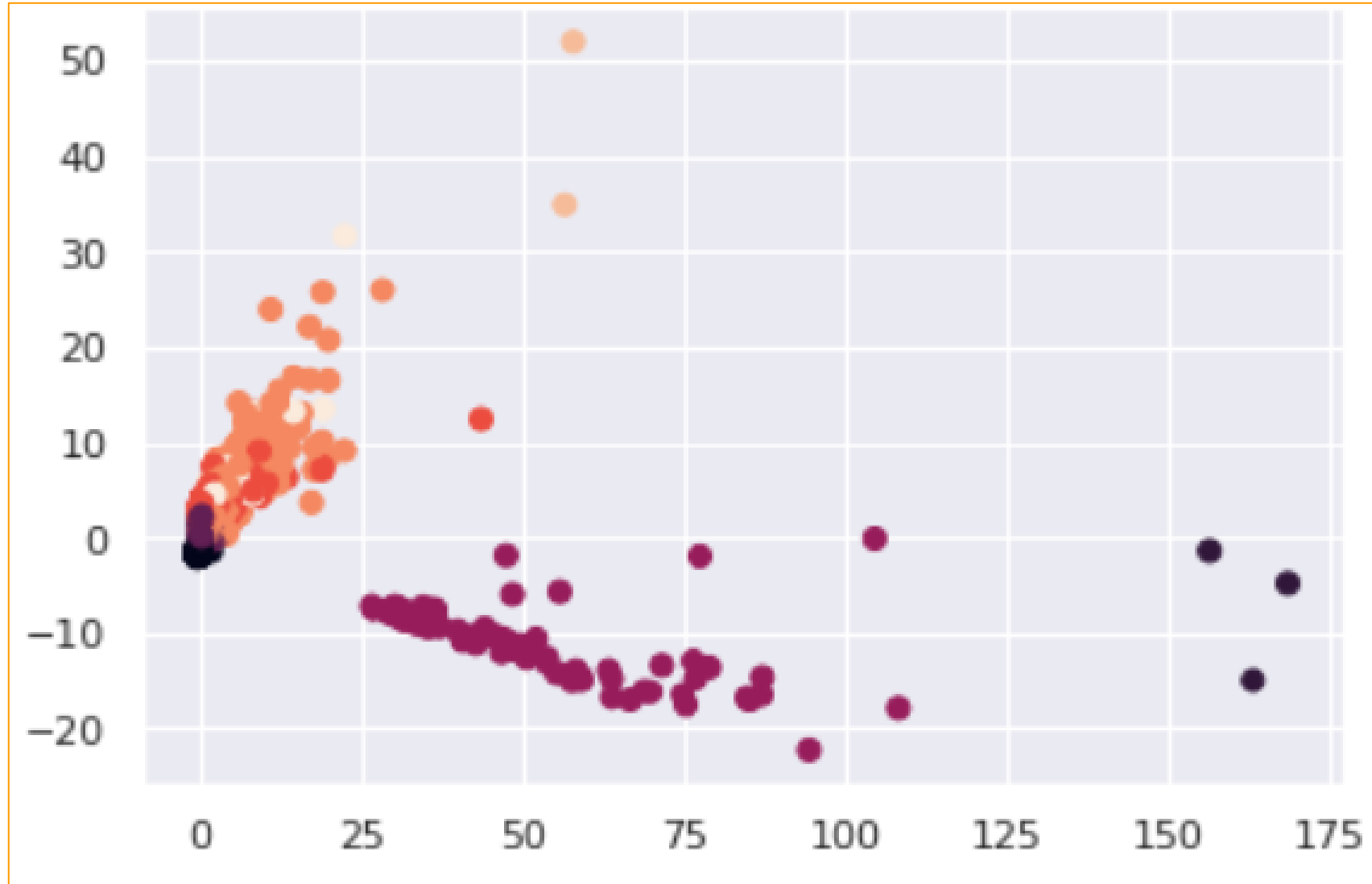
Results

K-means



Results

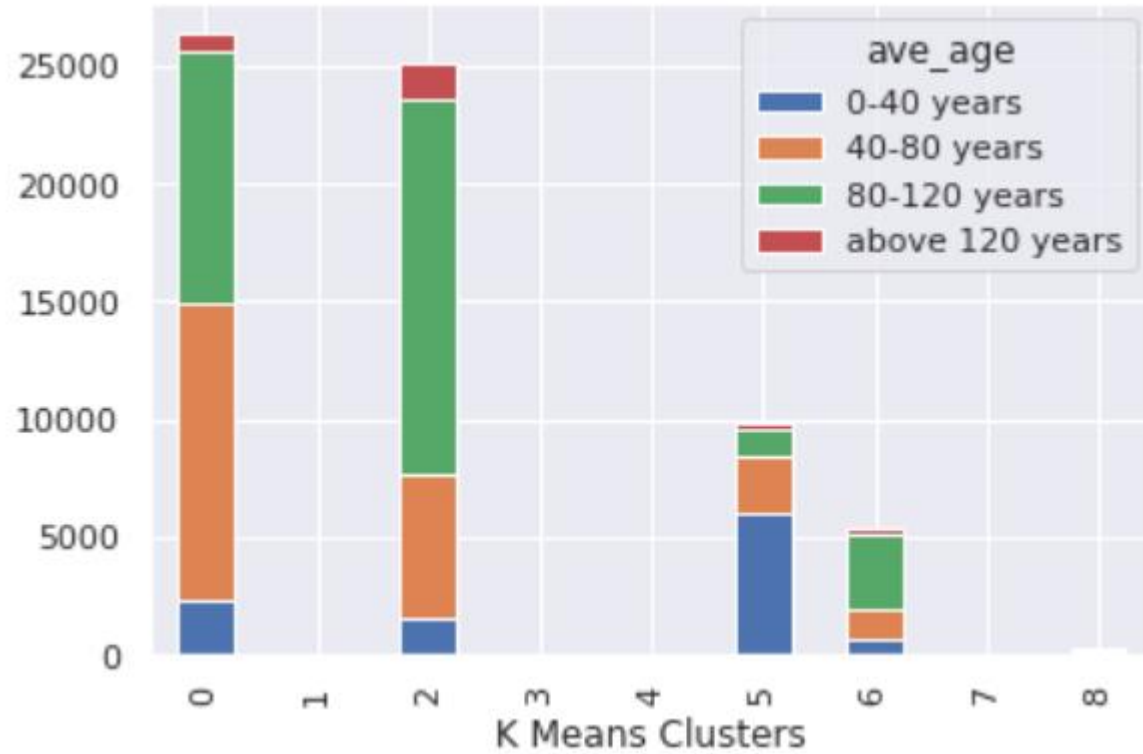
PCA



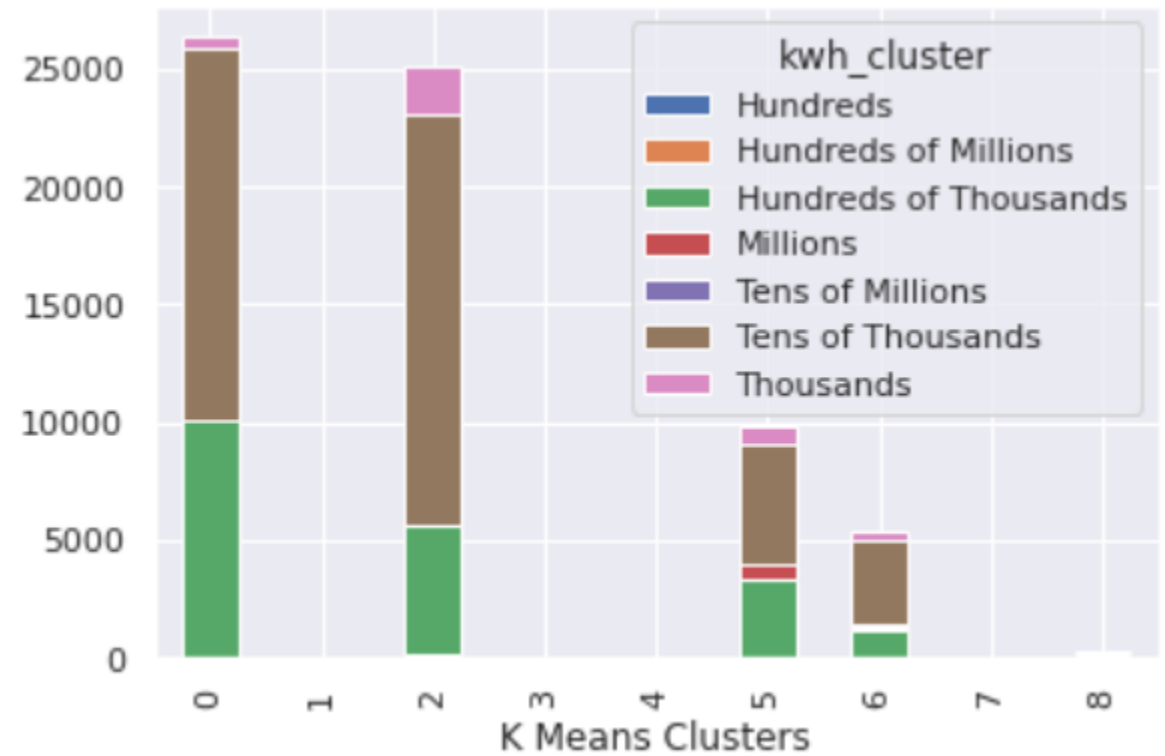
Results

Exploring the Clusters

Average age of buildings



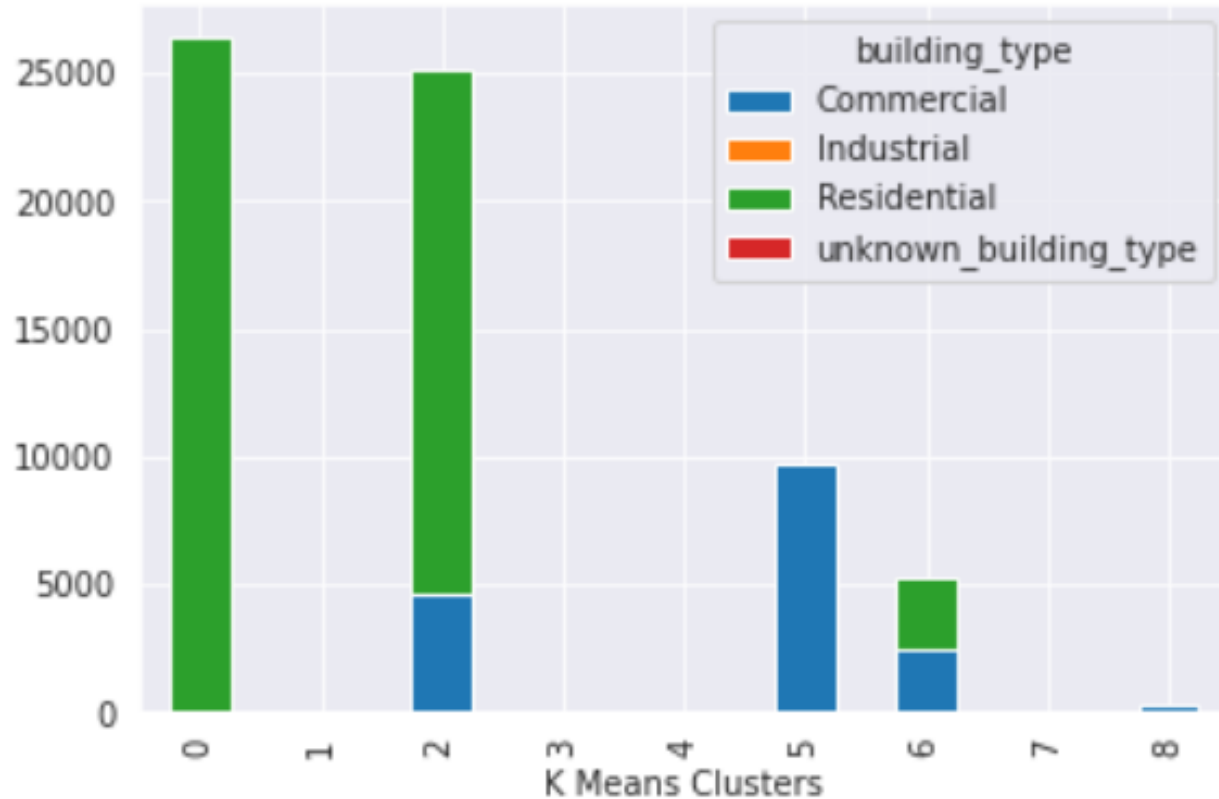
Power consumption



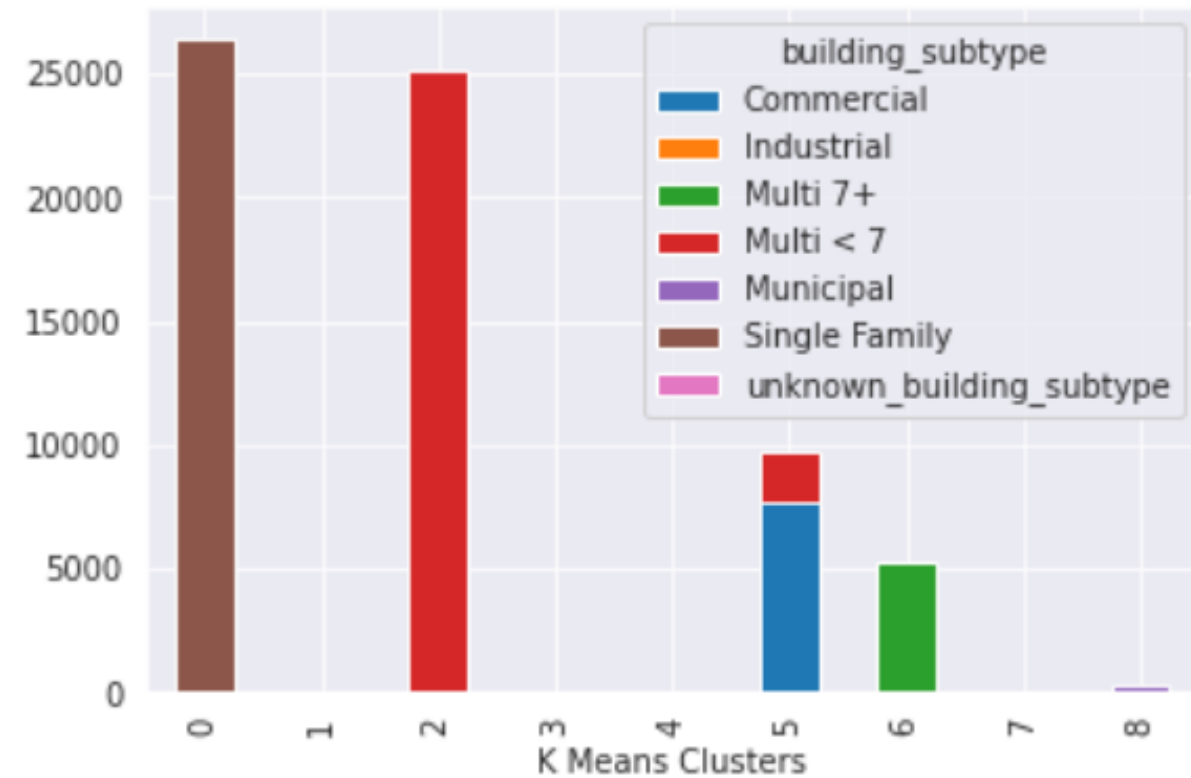
Results

Exploring the Clusters

Building type



Building sub-type



Conclusions

- Only DBSCAN and K-means were the suitable clustering algorithms.
- DBSCAN seems to have the best performance with silhouette score 0.37
- 9 clusters? 4clusters? Basis mainly on household.

Recommendations

- More information on the customers like the age of consumers, breakdown by time, etc.

Future Work

- Further perform clustering on the generated clusters to better understand behavior.
- Supervised learning to predict energy consumption.

Thanks

谢谢

Gracias