Week 6 Quiz

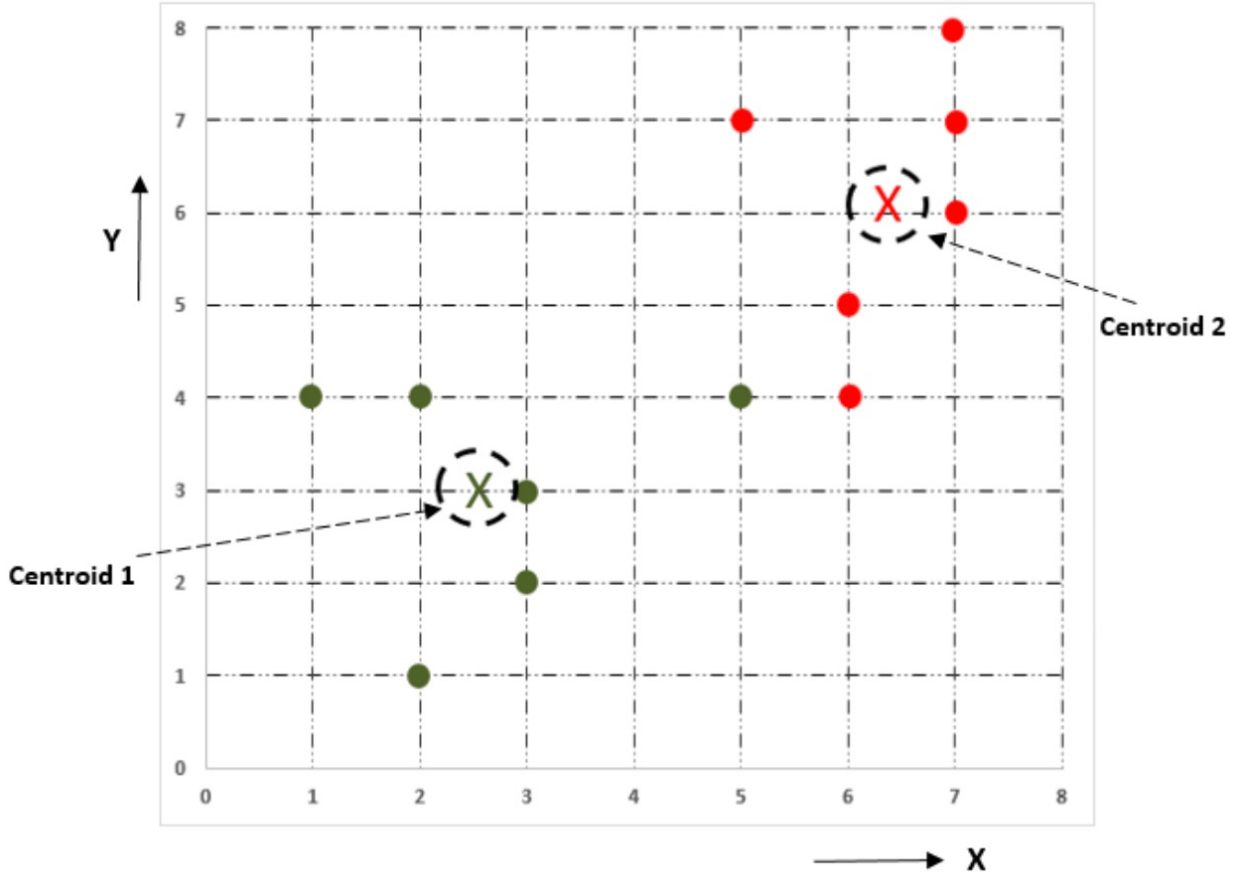
CASA0006

Q1. Which of these is FALSE? - Clustering:

1. reduces the number of observations
2. is an unsupervised learning algorithm
3. groups similar data points together, based on some similarity metric
4. finds the best model to predict a variable
5. requires the standardisation of variables

Q1. Which of these is FALSE? - Clustering:

1. reduces the number of observations
2. is an unsupervised learning algorithm
3. groups similar data points together, based on some similarity metric
4. finds the best model to predict a variable. **FALSE**
5. requires the standardisation of variables



Min-Max Scaling
Limits: $0 \leq X_{scaled} \leq 1$

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Q2. Which of these is TRUE about K-means clustering?

1. the initial placement of the centroids has no impact on the final clustering result

2. it is sensitive to outliers

3. can directly handle categorical variables (without using one-hot)

4. requires the number of clusters to be specified in advance

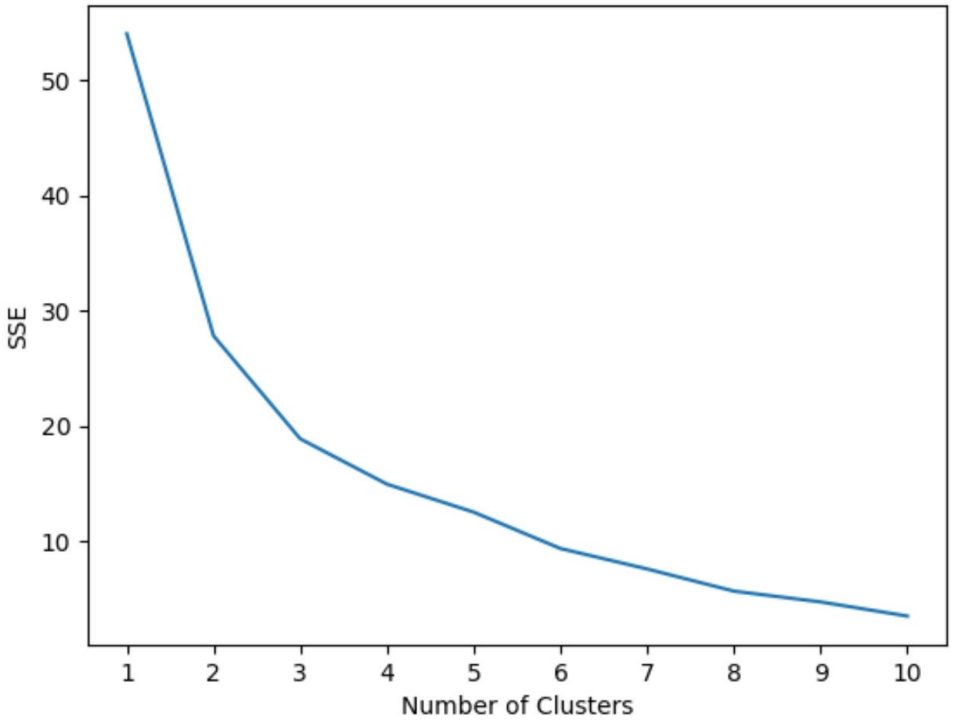
Q2. Which of these is TRUE about K-means clustering?

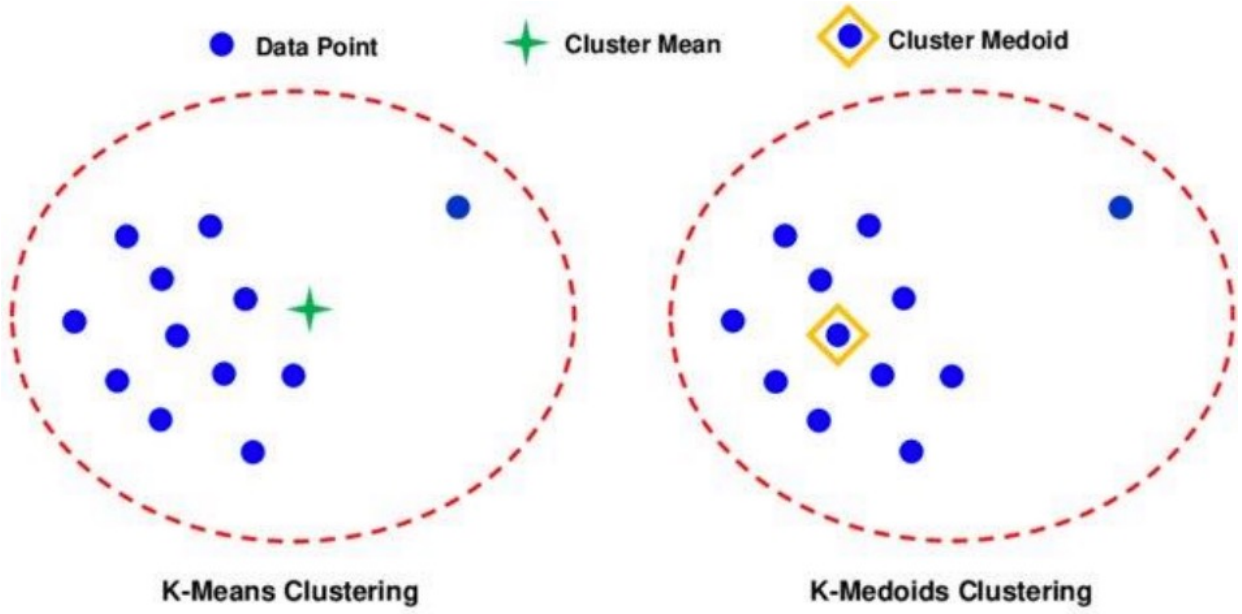
1. the initial placement of the centroids has no impact on the final clustering result

2. it is sensitive to outliers **TRUE**

3. can directly handle categorical variables (without using one-hot)

4. requires the number of clusters to be specified in advance. **TRUE**





Q3. Which of these is FALSE about hierarchical clustering?

1. creates a hierarchy of clusters that can be cut at any level to form different numbers of clusters

2. Agglomerative clustering is a bottom-up approach

3. It requires the number of clusters to be specified in advance.

4. Divisive clustering is a top-down approach

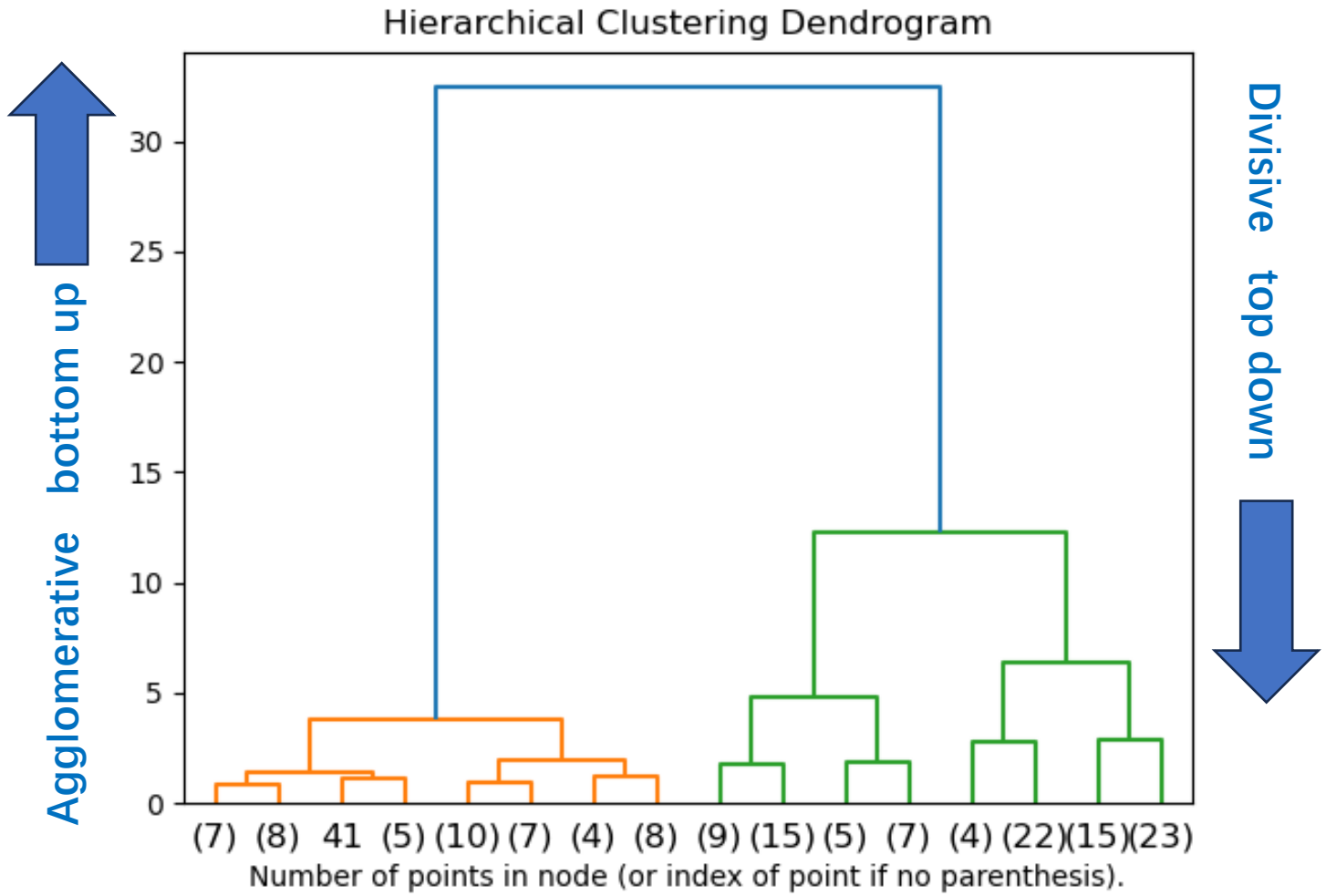
Q3. Which of these is FALSE about hierarchical clustering?

1. It creates a hierarchy of clusters that can be cut at any level to form different numbers of clusters

2. Agglomerative clustering is a bottom-up approach

3. It requires the number of clusters to be specified in advance. **FALSE**

4. Divisive clustering is a top-down approach



Q4. Which of these is FALSE about cluster quality metrics?

1. Silhouette score ranges from -1 to 1.

2. SSE values in the SSE Elbow method can range from 0 to infinity.

3. Homogeneity values lie between 0 and 1.

4. Completeness values lie between -1 and 1.

Q4. Which of these is FALSE about cluster quality metrics?

1. Silhouette score ranges from -1 to 1.

$$\text{Silhouette Score} = \frac{b(i)-a(i)}{\max(a(i),b(i))}$$

Limits: $-1 \leq \text{Silhouette Score} \leq 1$

2. SSE values in the SSE Elbow method can range from 0 to infinity.

$$\text{SSE} = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2$$

Limits: $0 \leq \text{SSE} < \infty$

3. Homogeneity values lie between 0 and 1.

$$h = 1 - \frac{H(Y|C)}{H(Y)}$$

Limits: $0 \leq h \leq 1$

4. Completeness values lie between -1 and 1. **FALSE**

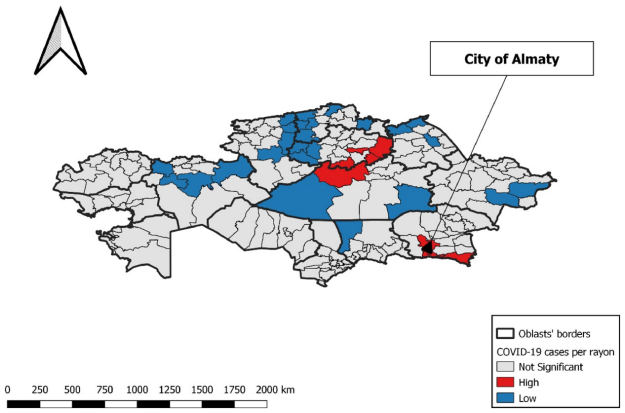
$$c = 1 - \frac{H(C|Y)}{H(C)}$$

Limits: $0 \leq c \leq 1$

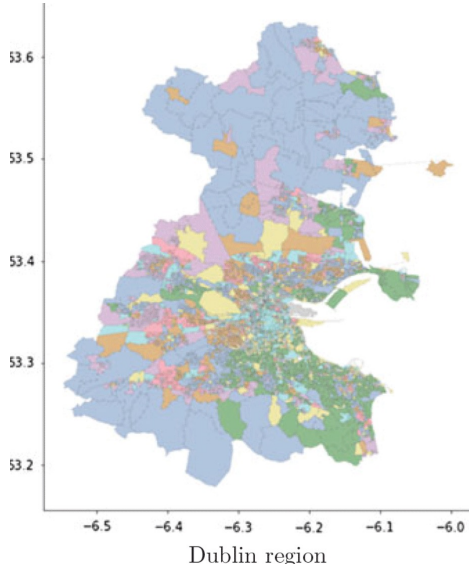
Q5. Spatial clustering can be used to identify:

- hotspots of COVID-19 infections
- groups of customers based on demographic and purchasing data, with the added constraint of geographic proximity
- areas of similar land use patterns, such as residential, commercial, or industrial
- areas of high crime or crime hotspot
- all of the above

Q5. Spatial clustering can be used to identify:

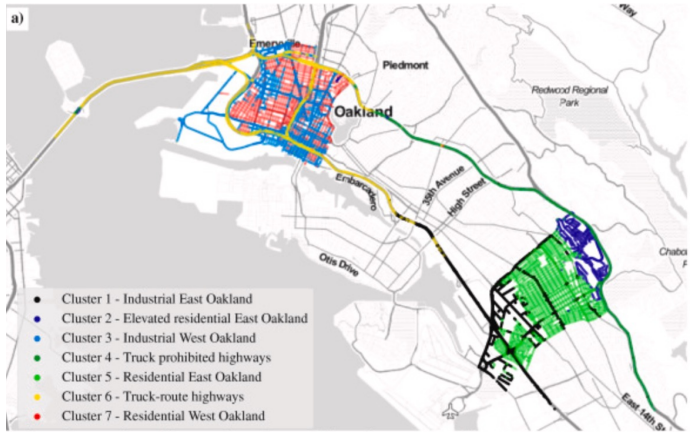


Kuznetsov, A. and Sadovskaya, V., 2021. Spatial variation and hotspot detection of COVID-19 cases in Kazakhstan, 2020. *Spatial and Spatio-Temporal Epidemiology*, *39*

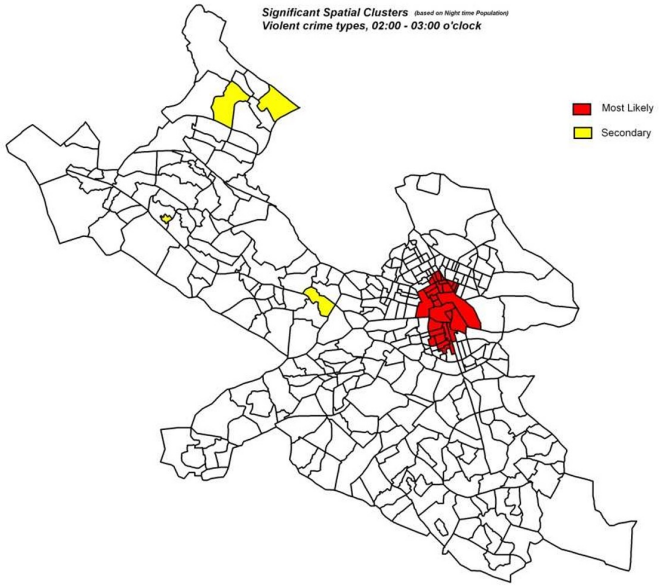


Dublin region

Hu, S., O'Hagan, A., Sweeney, J. and Ghahramani, M., 2021. A spatial machine learning model for analysing customers' lapse behaviour in life insurance. *Annals of Actuarial Science*



Montazeri, A., Lilienthal, A.J. and Albertson, J.D., 2021. A spatial land use clustering framework for investigating the role of land use in mediating the effect of meteorology on urban air quality. *Atmospheric Environment: X*, *12*, p.100126.



Uittenbogaard, A. and Ceccato, V., 2012. Space-time clusters of crime in Stockholm, Sweden. *Rev. Eur. Stud.*, *4*, p.148.

all of the above