K - Nearest Neighbours

1) Start with dataset with known categories

2) Cluster data

↳ K-Means
→ K-Medoids
} Partitioning Clustering — divide data into non-overlapping clusters

3) Add new data with unknown category to plot

4) Classify unknown cell by looking at nearest neighbour.

↳ if k = 1 : Consider nearest neighbour

if k = 11 : Consider nearest 11 neighbours

↑ Take the most count — eg — 7 red ← choose this
3 orange
1 green

* if equal vote: flip coin to decide category

Determining ideal value for k:

- Need to perform hyperparameters tuning to get ideal value

↳ Test with known labelled data & choose when accuracy is the highest