COMP2200/COMP6200 Lecture 1c – Orange Clustering

Greg Baker

23rd February 2026

MACQUARIE
University

Graphical data science tools

Orange **KNIME** **Weka** **Alteryx** **Tableau**

- **Orange** — **"Easiest for students"**; fastest way to learn workflows and explore data.

- **KNIME** — great for data prep + reproducible pipelines; lots of connectors/plugins.

- **Weka** — classic teaching GUI for ML algorithms (older UI).

- **Alteryx** — common in industry for data prep/analytics; usually paid (enterprise).

- **Tableau** — interactive visualisation and dashboards (communicating results).

Logos are trademarks of their respective owners.

What is Orange?

- Visual data mining suite built on top of Python
- Drag-and-drop widgets let you build workflows quickly
- Great for exploring data without writing code

https://orangedatamining.com/download/

Downloading & installing Orange

- Visit https://orangedatamining.com/download/and download the installer for your operating system
- Install it:
  - Windows: run the installer
  - macOS: drag Orange into *Applications*
  - Linux: follow the download page (AppImage / Conda)
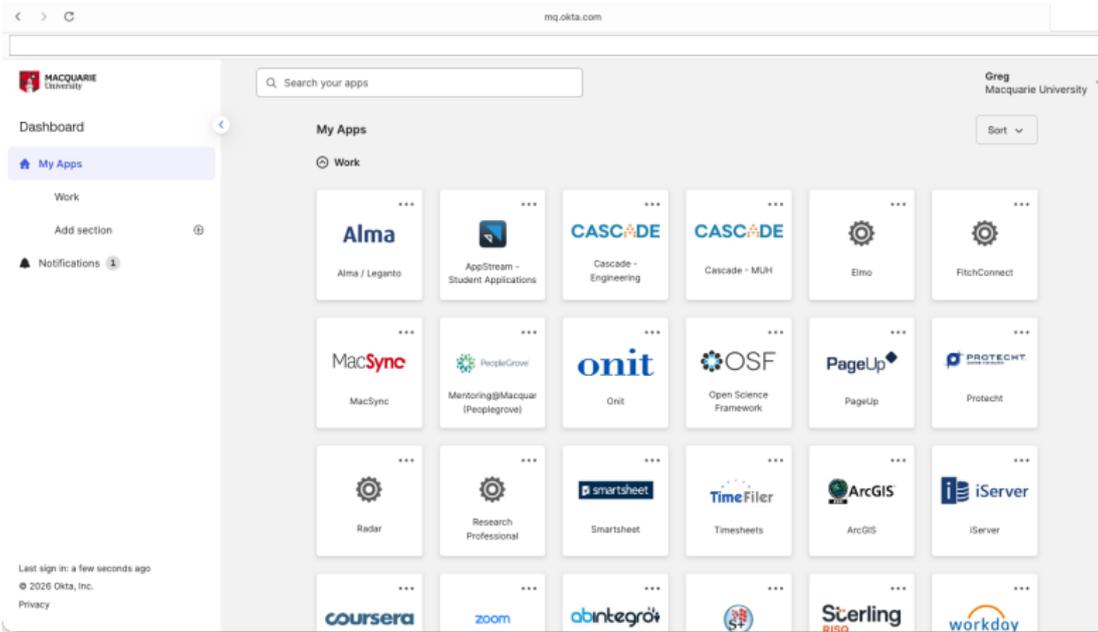- If you can't install locally, use AppStream (next slides)

https://orangedatamining.com/download/

Running Orange on a tablet (AppStream)

- AppStream gives you a remote desktop in your browser.
- Handy if you're on an iPad/tablet, or you can't install Orange on your laptop.
- Go to `https://mq.okta.com/app/UserHome`, sign in, and search for **AppStream**.

`https://mq.okta.com/app/UserHome`

AppStream: 1) Open Okta

https://mq.okta.com/app/UserHome

AppStream: 2) Search for AppStream

https://mq.okta.com/app/UserHome

AppStream: 3) Launch Orange

https://mq.okta.com/app/UserHome

MACQUARIE
University
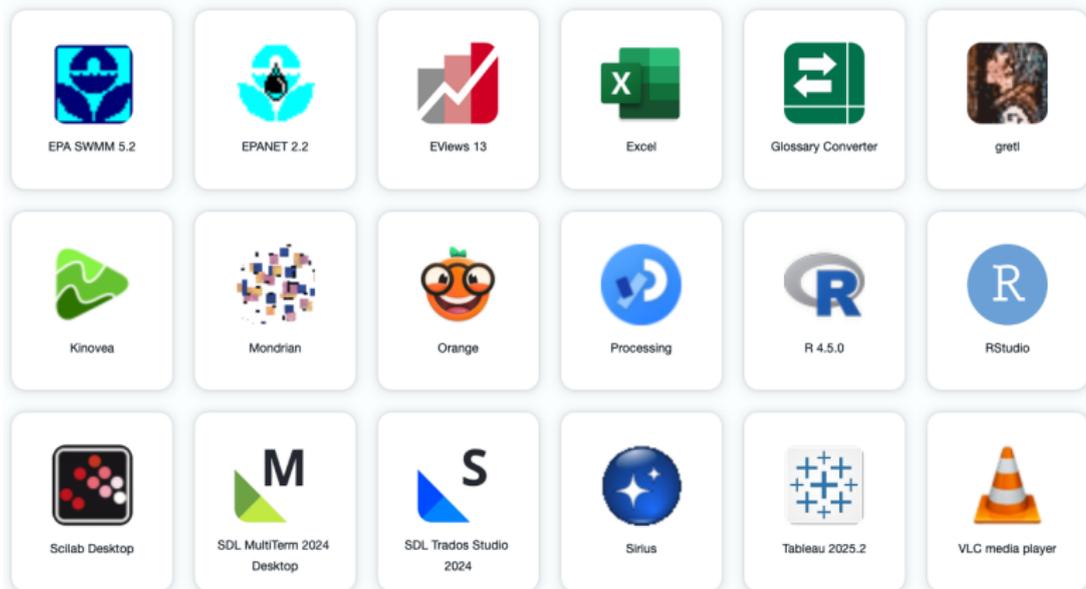
What is Clustering?

Easy Definition

Clustering is a way of grouping similar things together, without being told what the groups should be.

- Imagine sorting socks from the laundry —— you group them by colour or pattern, even if no one told you what the categories are.
- Practical examples:
    - Grouping customers by shopping behaviour (customer segmentation)
    - Detecting communities in social networks
    - Organising news articles into topics

https://orangedatamining.com/download/

Give me a game-ified learning environment

`http://clustering-visualizer.web.app`

Feel free to ignore the lecture and learn on your own from this one web app.

`https://orangedatamining.com/download/`

K-Means Clustering

How it works

- Choose how many clusters you want (say, $k = 3$).

- Drop k points randomly — these are your cluster centres (called **centroids**).

- Assign each point in the dataset to the closest centroid.

- Move the centroids to the **mean** of the points in their cluster.

- Repeat until nothing changes.

https://orangedatamining.com/download/
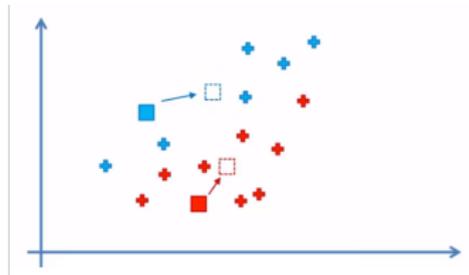
Example: Indian Supermarkets

Pick your scenario:

- You have a business supplying something to Indian supermarkets, and you need to think about distribution depots

- You are looking for a place to rent and you want to see what suburbs to consider so that you not too far away from an Indian supermarket?

Q: Why choose Indian supermarkets?

A: the cluster diagram looked nice
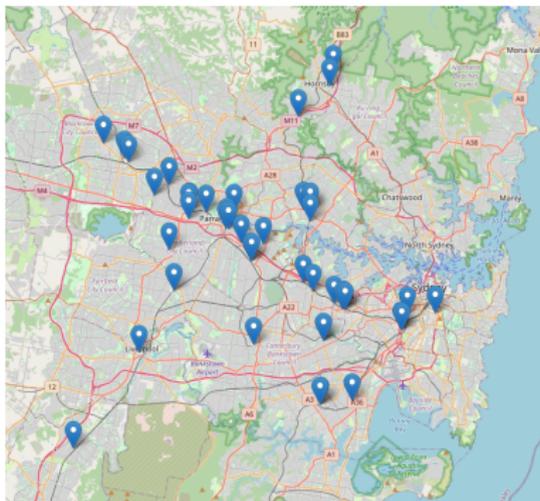
https://orangedatamining.com/download/

Geocoding

Ways to turn a street address into a (latitude,longitude) pair.

- Search in Google maps, then right click
- Government databases
- Pick a programming API from
  https://wiki.openstreetmap.org/wiki/Geocoding
- Google Sheets plug-in

https://orangedatamining.com/download/

Let's visualise it

We'll compress this messy reality into a simple model that's helpful for decisions (like warehouse locations)

https://orangedatamining.com/download/

MACQUARIE
University

Plotting in Orange

① Start Orange and add a **File** widget.

② Load `indian-supermarkets-in-sydn`

③ Add a **Scatter Plot** widget and connect it to the File widget.

④ Choose *latitude* for the x-axis and *longitude* for the y-axis.

⑤ The map of supermarket locations should now appear.

https://orangedatamining.com/download/

MACQUARIE
University
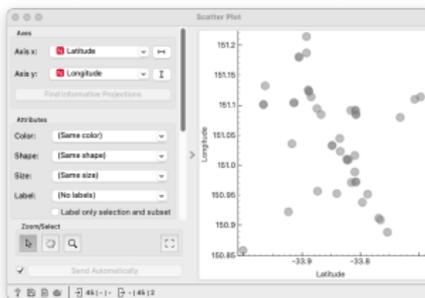
Making it nicer

- Select Columns to ignore the null columns, and make the Latitude and Longitude into Features

- Add a Data Table to see the data

- Add the Geo add-in

- Add a Geo Map
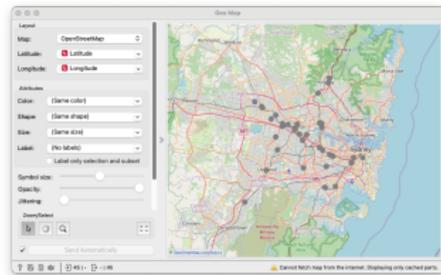
k-Means in Orange

1. Add a **k-Means** widget and connect it to your data. Turn off normalisation

2. Select how many clusters you want to create.

3. Connect the output to a **Scatter Plot** or **Geo Plot** to colour points by cluster.

4. Use the **Silhouette Plot** widget to judge the quality of the clustering.

5. Add another **Scatter Plot** or **Geo Plot** to visualise where the centroids are
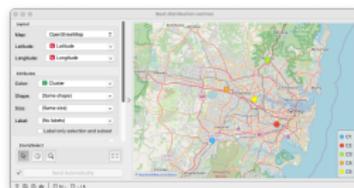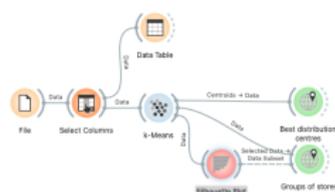
What we haven't talked about

- What's that Silhouette Plot doing?
- How do we calculate a Silhouette score?
- Why did we turn off normalisation?

We'll cover these topics in a few weeks' time

Wrapping up

- Clustering helps you find natural groupings in data, even without explicit labels.
- K-Means clustering is simple and powerful: choose k, assign, average, and iterate.
- Orange makes clustering easy with a visual, drag-and-drop approach.
- Geo visualisations enhance insights when clustering geographic data.

Next steps:

- Try clustering your own datasets.
- Explore different values of k.
- Experiment with other widgets in Orange.