Welcome to Week 9 Lecture 2!

Data Science in Python & Machine Learning

Today's Agenda

- ❏ Review unsupervised learning
- ❏ Apply Kmeans clustering to improve a supervised model
- ❏ Clustering Mini Hackathon in breakout rooms
- ❏ Return to share findings (if time permits)

How is Unsupervised Learning Different than Supervised Learning?

1. No target variable
2. No train test split
3. No X, y split

Using KMeans Clusters as a Feature in Supervised Learning

Clusters found with KMeans can be a feature.

KMeans performs feature extraction.

If we are using KMeans for analysis - No Train Test Split

If we are using KMeans for feature extraction for supervised learning we must use a train test split

```
kmeans.fit(X_train)

X_train['cluster'] = kmeans.predict(X_train)
X_test['cluster'] = kmeans.predict(X_test)
```

Combining Supervised and Unsupervised Learning

Code-a-Long

Melbourn Housing Data

Today's Challenge

- You will use the same data set (Melbourne Housing), but you have *different* problem.
- You are tasked with dividing these properties into groups for your marketing team.
- If time permits we will reconvene in the main room for a 2 minute "lightning pitch" with the goal of convincing the class that your clustering results will be useful to the client.
- Note, an example of part b of this challenge to "improve a supervised learning model" was already addressed, so you are only focusing on task a!