

Clustering Discovery Workshop

In-Class Activity Sheet - Week 1

BSc Machine Learning for Innovation

Team: _____

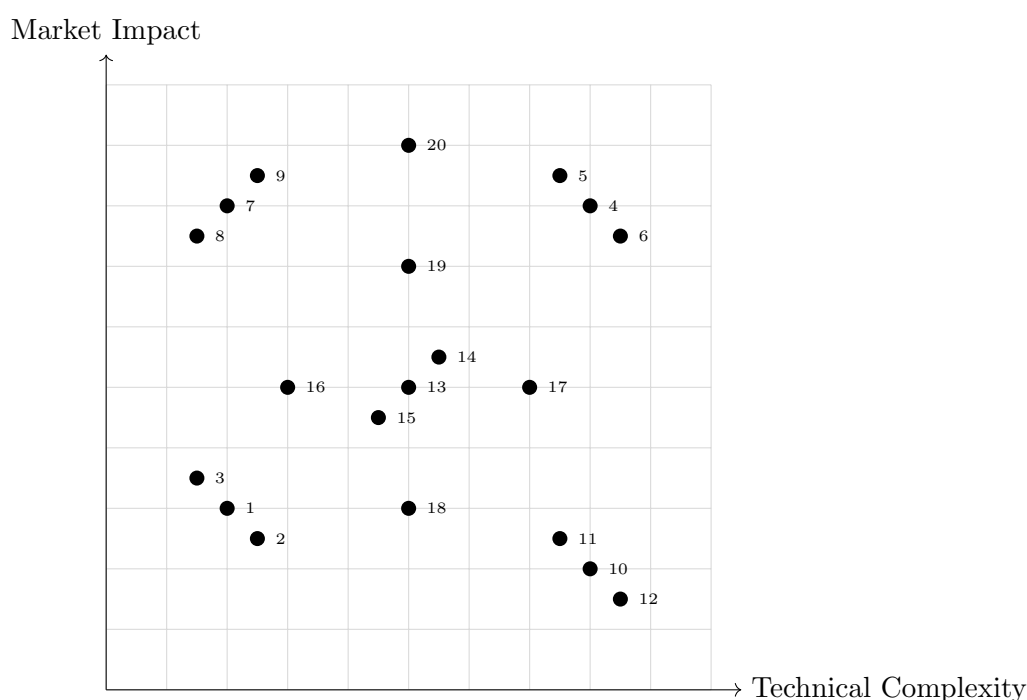
Workshop Objective

Today you'll discover clustering algorithms by building them yourself! Work in teams of 2-3 and document your process as you go.

Activity 1: Manual Clustering Challenge (15 minutes)

Your Data: Innovation Ideas from a Startup

Below are 20 innovation ideas plotted by **Technical Complexity** (x-axis) and **Market Impact** (y-axis):



Task 1: Create Your Clusters

1. Draw circles around groups you think belong together in the plot above.
2. How many clusters did you find? _____
3. Name each cluster based on its characteristics:

- Cluster 1: _____
- Cluster 2: _____
- Cluster 3: _____

- Cluster 4: _____

- Cluster 5: _____

4. **Which ideas don't fit any cluster?** (Outliers): _____

Activity 2: Build K-means Algorithm (20 minutes)

Step-by-Step K-means Discovery

You'll now recreate the K-means algorithm using the same data!

Round 1: Initial Centers

1. Choose $K=3$ (we want 3 groups)
2. Pick 3 random points as starting centers. Circle them with a different color.
3. Centers chosen: Point _____, Point _____, Point _____

Round 2: Assign Points

1. For each point, measure distance to all 3 centers
2. Assign each point to its nearest center
3. Color-code your assignments:

Point	Distance to C1	Distance to C2	Distance to C3	Assigned to
1				
2				
3				
...

Round 3: Update Centers

1. Calculate the average position of all points in each cluster
2. New Center 1: (_____, _____)
3. New Center 2: (_____, _____)
4. New Center 3: (_____, _____)

Round 4: Check Convergence

- Did any points change clusters? YES / NO
- If YES, repeat Round 2-3
- If NO, you're done!

Discovery Question

How is your K-means result different from your manual clustering? Why?

Activity 3: Distance Metrics Exploration (15 minutes)

Different Ways to Measure “Close”

Consider these two innovation profiles:

- **Idea A:** Tech=8, Impact=3, Cost=5
- **Idea B:** Tech=5, Impact=7, Cost=6

Calculate Different Distances:

1. Euclidean Distance (straight line):

$$d = \sqrt{(8-5)^2 + (3-7)^2 + (5-6)^2} = \sqrt{\quad + \quad + \quad} = \quad$$

2. Manhattan Distance (city blocks):

$$d = |8-5| + |3-7| + |5-6| = \quad + \quad + \quad = \quad$$

3. Maximum Distance (biggest difference):

$$d = \max(|8-5|, |3-7|, |5-6|) = \max(\quad, \quad, \quad) = \quad$$

Innovation Insight

Which distance metric would you use for:

- Finding similar products for customers? _____
- Grouping projects by resource needs? _____
- Identifying competing innovations? _____

Activity 4: Cluster Quality Assessment (10 minutes)

How Good Are Your Clusters?

Rate your clusters on these criteria (1-5 scale):

Criterion	Score (1-5)	Why?
Cohesion Points in same cluster are similar		
Separation Different clusters are distinct		
Coverage Most points belong to a cluster		
Balance Clusters have similar sizes		
Interpretability Clusters make business sense		

Overall Quality Score: _____ / 25

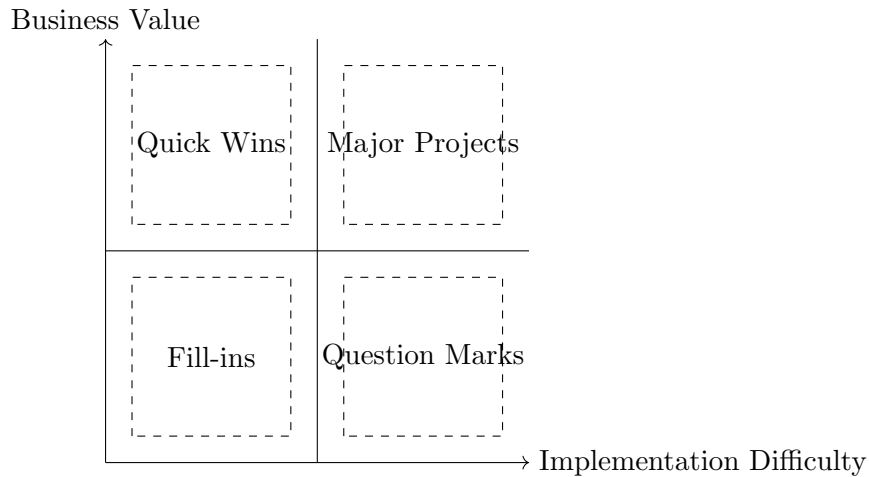
Activity 5: Innovation Application (15 minutes)

From Clusters to Strategy

Based on your clustering results, create an innovation strategy:

1. Priority Matrix

Place your clusters in the appropriate quadrant:



2. Resource Allocation

How would you allocate 100 points of resources?

- Cluster 1: _____ points
- Cluster 2: _____ points
- Cluster 3: _____ points
- Outliers/Other: _____ points

3. Innovation Roadmap

Order your clusters for implementation:

1. First: _____
2. Second: _____
3. Third: _____

Reflection & Synthesis (5 minutes)

Key Discoveries

1. What surprised you about clustering?
2. How would clustering help your organization?
3. What challenges did you encounter?

Team Members:

- _____
- _____
- _____