3. Conclusion

Error Comparison Table

Algorithm Dataset	K-means	Fuzzy c-means	DBSCAN

First Dataset	K = 3
	SC = 0.493
	SSE = 80.71
	PC = *NA*
	S/F= **Succeeds**
		K = 2
		SC = 0.536
		SSE = *NA*
		PC = 0.408
		S/F= **Succeeds**
			K = 1
			SC = 0.516
			SSE = *NA*
			PC = *NA*
			S/F= **Fails**

Second Dataset	K = 3
	SC = 0.724
	SSE = 6487
	PC = *NA*
	S/F= **Succeeds**
		K = 3
		SC = 0.544
		SSE = *NA*
		PC = 0.28
		S/F= **Succeeds**
			K = 9
			SC = 0.617
			SSE = *NA*
			PC = *NA*
			S/F= **Succeeds**

Third Dataset	K = 5
	SC = 0.560
	SSE = 222
	PC = *NA*
	S/F= **Succeeds**
		K = 4
		SC = 0.365
		SSE = *NA*
		PC = 0.16
		S/F= **Fails**
			K = 5
			SC = 0.501
			SSE = *NA*
			PC = *NA*
			S/F= **Succeeds**

K = number of clusters
SC = Silhouette Coefficient
SSE = Error Sum of Squares
PC = Partition Coefficient
S/F = **Succeeds** or **fails** to cluster

Validation Comparison Table

Algorithm Dataset	K-means	Fuzzy c-means	DBSCAN

First Dataset	SC = **Good**
	SSE = **Good**
	PC = *NA*
		SC = **Good**
		SSE = *NA*
		PC = **Good**
			SC = **Bad**
			SSE = *NA*
			PC = *NA*

Second Dataset	SC = **Good**
	SSE = **Good**
	PC = *NA*
		SC = **Good**
		SSE = *NA*
		PC = **So-so**
			SC = **Good**
			SSE = *NA*
			PC = *NA*

Third Dataset	SC = **Good**
	SSE = **Good**
	PC = *NA*
		SC = **So-so**
		SSE = *NA*
		PC = **Bad**
			SC = **Good**
			SSE = *NA*
			PC = *NA*

K = number of clusters
SC = Silhouette Coefficient
SSE = Error Sum of Squares
PC = Partition Coefficient

Good: Using this validation we can beautifully cluster our data.
So-so: Using this validation we can somehow cluster our data.
Bad: Using this validation we can not cluster our data.

Winners of the Clustering Competition

First dataset:

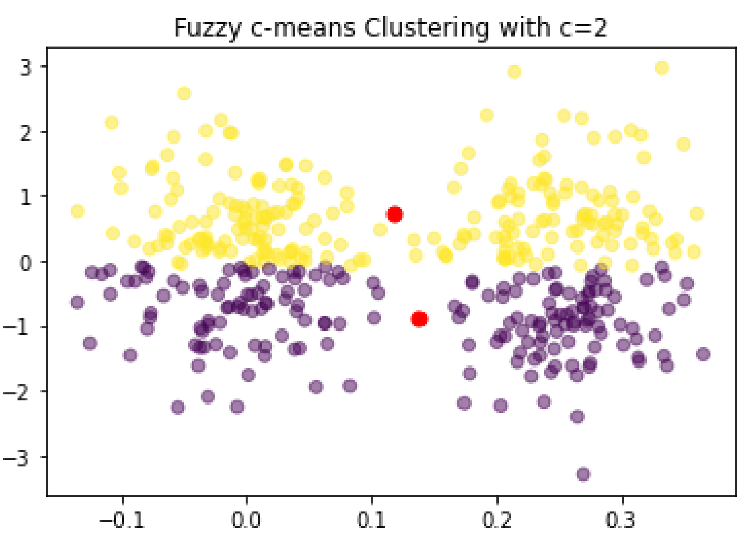

Figure 1: Fuzzy c-means with Silhouette Coefficient of **0.536**.

Although depending on the nature of our clustering problem, it might be reasonable to cluster the data into **3 clusters** using the **k-means** algorithm. The Silhouette Coefficient for these two are very close.

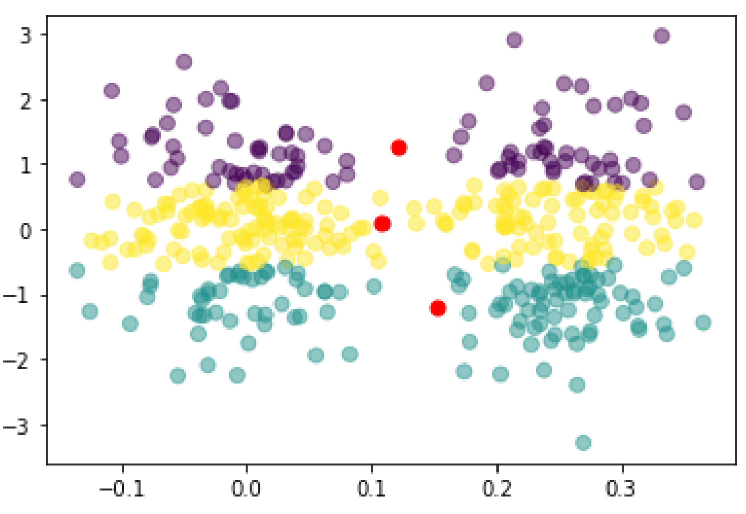

Figure 2: k-means with Silhouette Coefficient of **0.493**.

Second dataset:

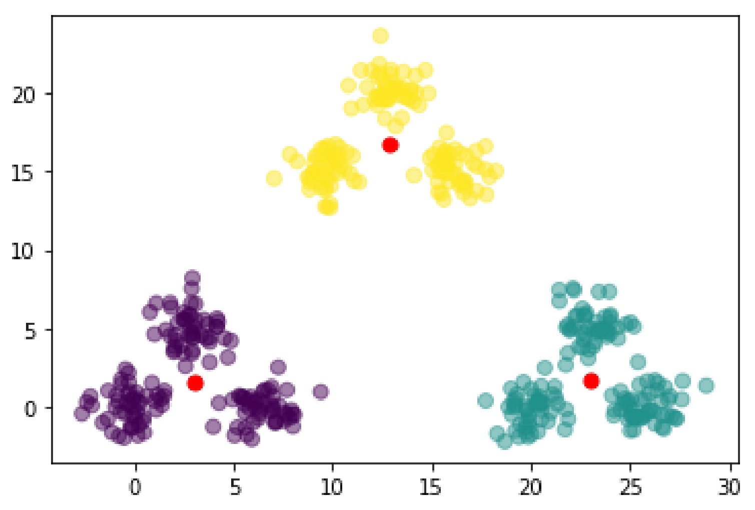

Figure 3: k-means with Silhouette Coefficient of **0.724**.

Although depending on the nature of our clustering problem, it might be reasonable to cluster the data into **9 clusters** using the **DBSCAN** algorithm. The Silhouette Coefficient for these two are very close.

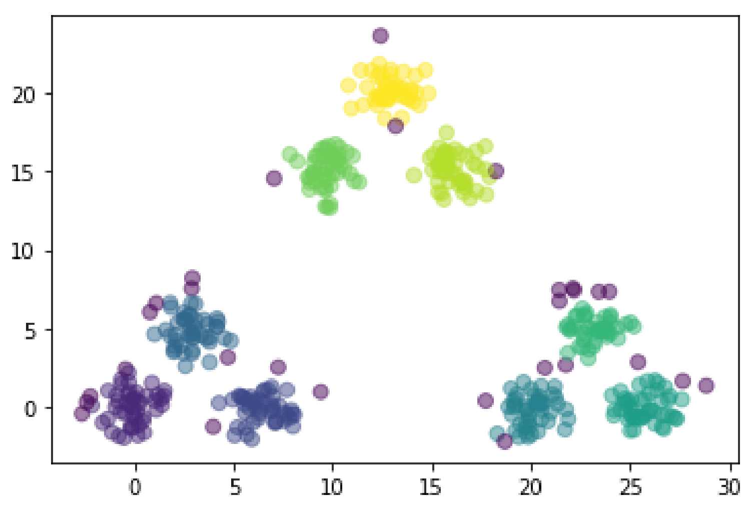

Figure 4: DBSCAN with Silhouette Coefficient of **0.617**.

Third dataset:

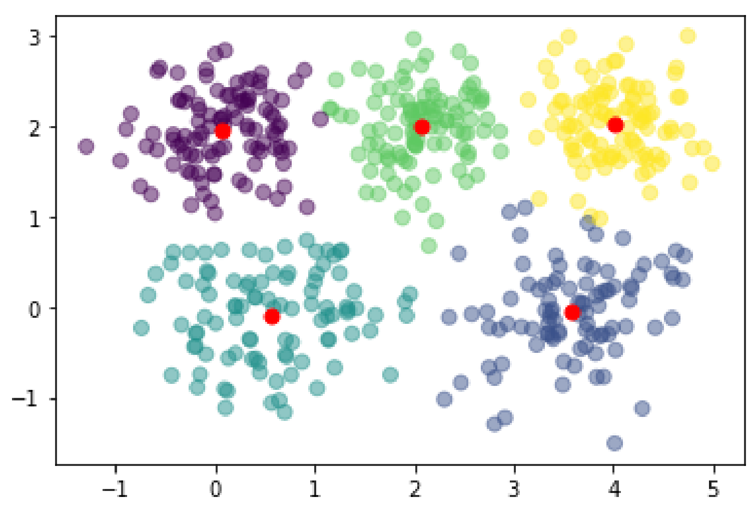

Figure 5: k-means with Silhouette Coefficient of **0.560**.

Ties with:

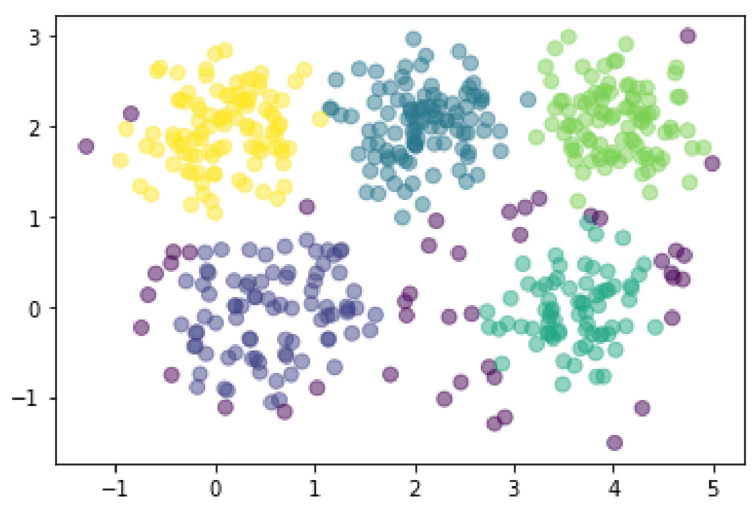

Figure 6: DBSCAN with Silhouette Coefficient of **0.501**.

Final thoughts:

As there are no one-size-fits-all, there are also no one-algorithm-clusters-all.
Having a sense of how each clustering algorithm works, how each method validates the clusters and how our data is scattered, can guide us into the right direction of clustering.
Knowing the nature of the clustering problem can also be very helpful.