

2. Clustering

2.3. Third dataset

Scatter Plot

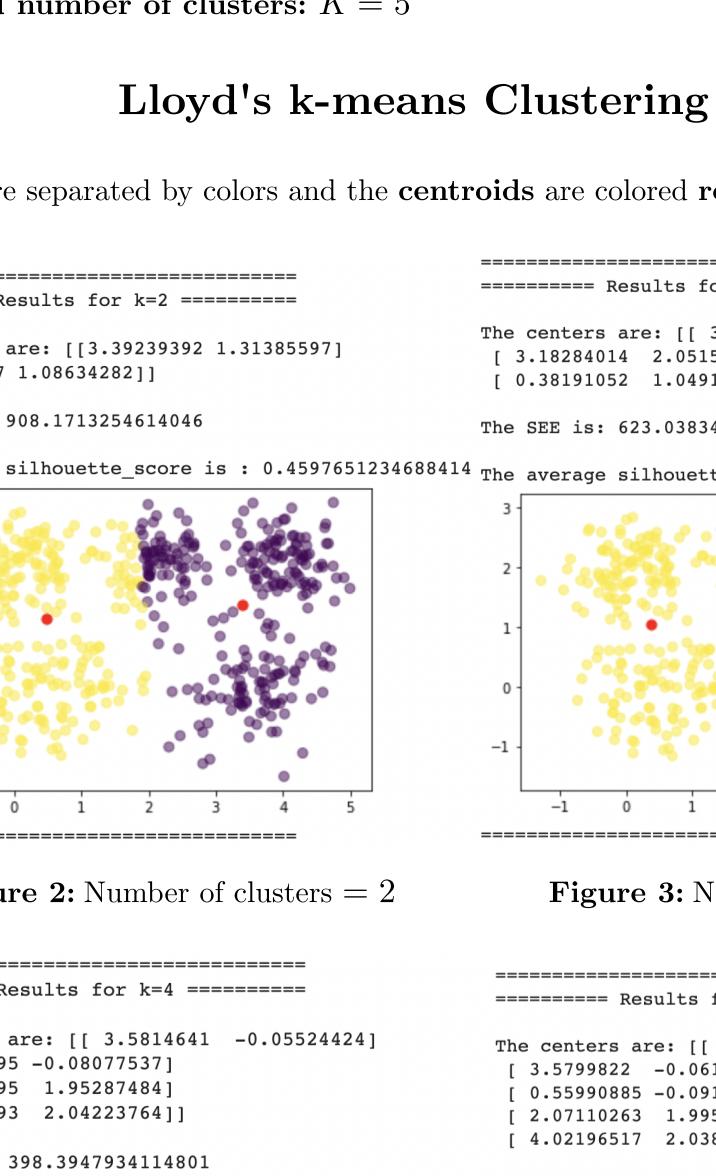


Figure 1: Scatter plot for the third dataset

Type of clusters: Center-based Clusters

- A cluster is a set of objects such that an object in a cluster is closer (more similar) to the “center” of a cluster, than to the center of any other cluster.
- The center of a cluster is often a centroid, the average of all the points in the cluster, or a medoid, the most “representative” point of a cluster.

Predicted number of clusters: $K = 5$

Lloyd's k-means Clustering

Clusters are separated by colors and the centroids are colored red.

===== Results for $k=2$ =====

The centers are: [[3.39239312 1.31385597] [0.48260267 1.08634282]]

The SEE is: 908.1713254614046

The average silhouette_score is : 0.4597651234688414

===== Results for $k=3$ =====

The centers are: [[3.51234268 -0.05959744] [3.19284014 2.0515819] [0.381052 1.04913832]]

The SEE is: 623.0383491955092

The average silhouette_score is : 0.4572516913124048

===== Results for $k=4$ =====

The centers are: [[3.5814641 -0.05524424] [0.37280295 1.95287484] [3.25903593 2.04223764]]

The SEE is: 398.3947934114801

The average silhouette_score is : 0.472599045087266

===== Results for $k=5$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927]]

The SEE is: 222.36598776929088

The average silhouette_score is : 0.5604142087131664

===== Results for $k=6$ =====

The centers are: [[1.3190934 0.10321736] [2.07034875 2.00804262] [4.02587268 2.05882534] [0.06354289 1.96513115] [0.17062224 -0.19507771] [3.62479519 -0.0344875]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=7$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=8$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=9$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=10$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=11$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=12$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=13$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=14$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=15$ =====

The centers are: [[0.07190087 1.95680378] [3.5799822 -0.06140016] [0.55990885 -0.0918281] [2.07110263 1.99524376] [4.02196517 2.0381927] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516] [0.86934362 1.95725167] [0.50963206 0.32898464] [4.03152359 2.02508959] [3.58564321 -0.07948172] [2.08410477 1.957516]]

The SEE is: 193.95255601125345

The average silhouette_score is : 0.5273614457298591

===== Results for $k=16$ =====