

C-means

Examples and comparison to k-means

Yazd University

Dr. Derhami

Aref Akbari

2022

Steps

What has been done in this work

- Definitions
- c-means on image
- k-means on image
- c-means example on random generated dataset
- c-means on bank transactions fraud dataset
- k-means on bank transactions fraud dataset

Definitions

Clustering

Definition

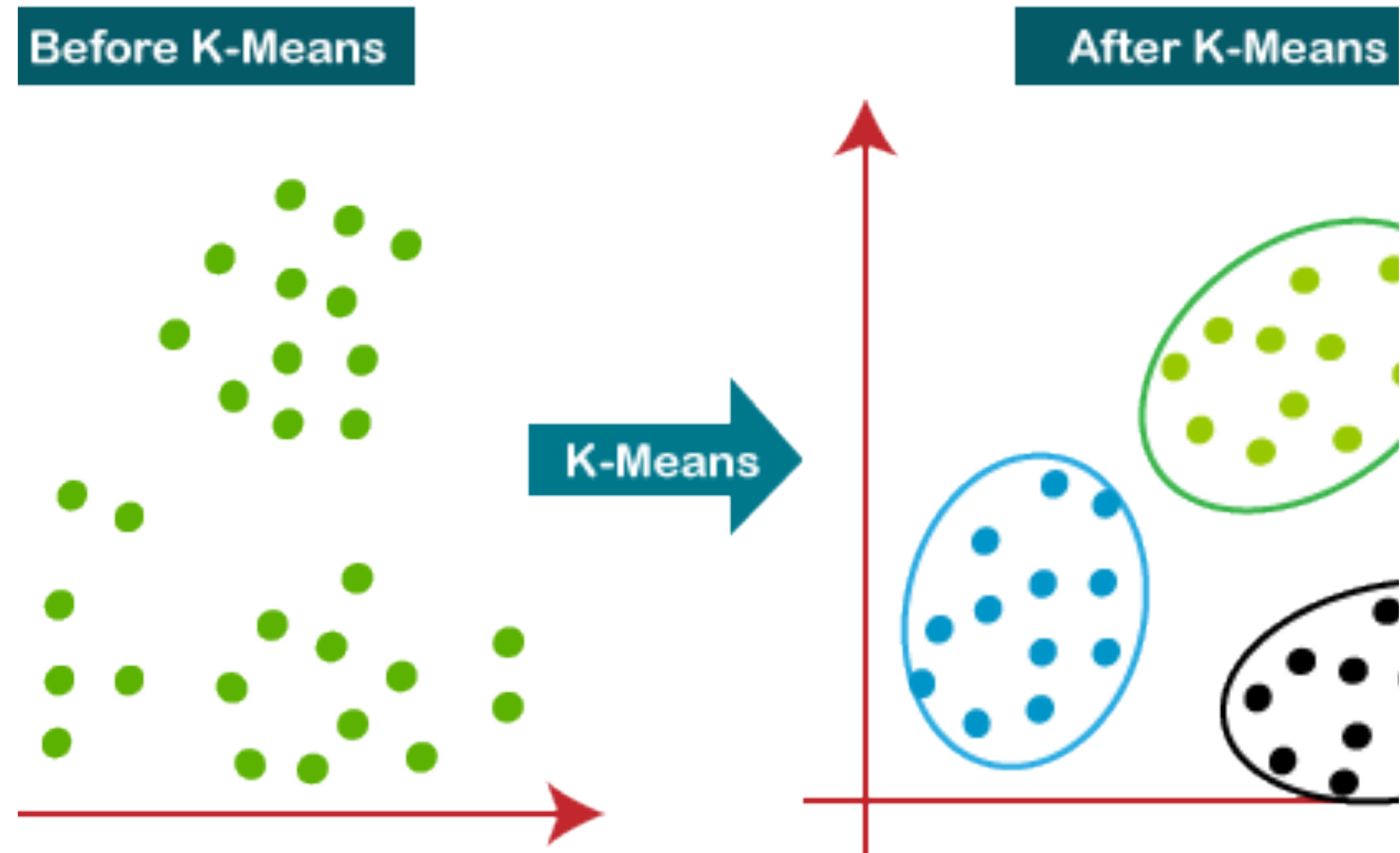
Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.



K-means

Definition

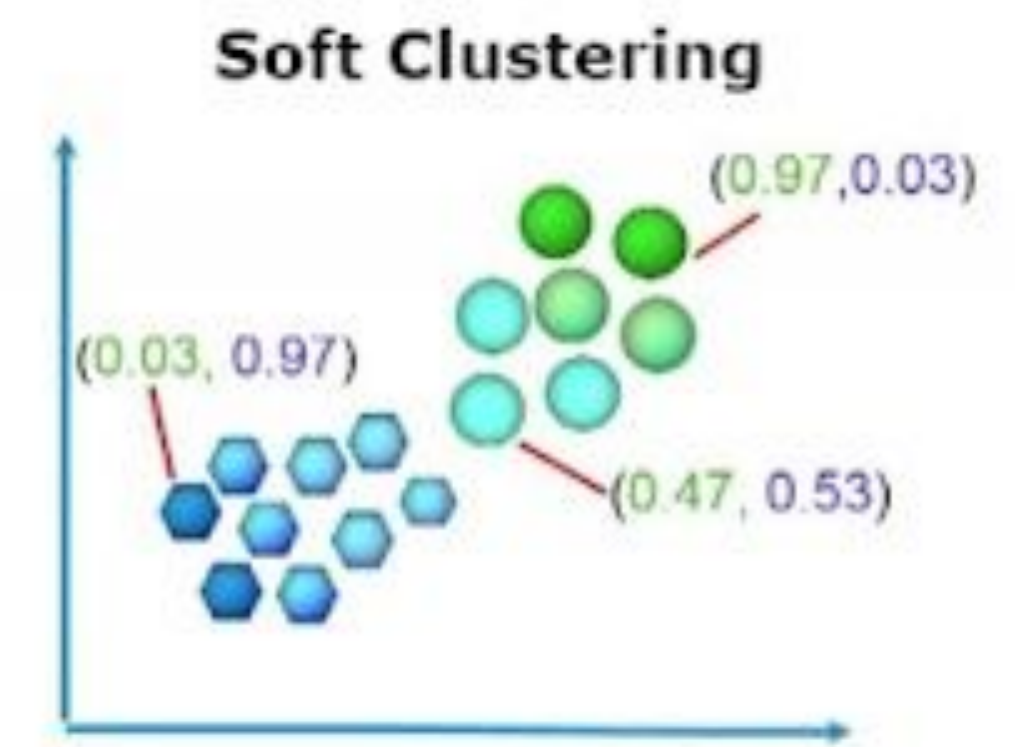
k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.



C-means

Definition

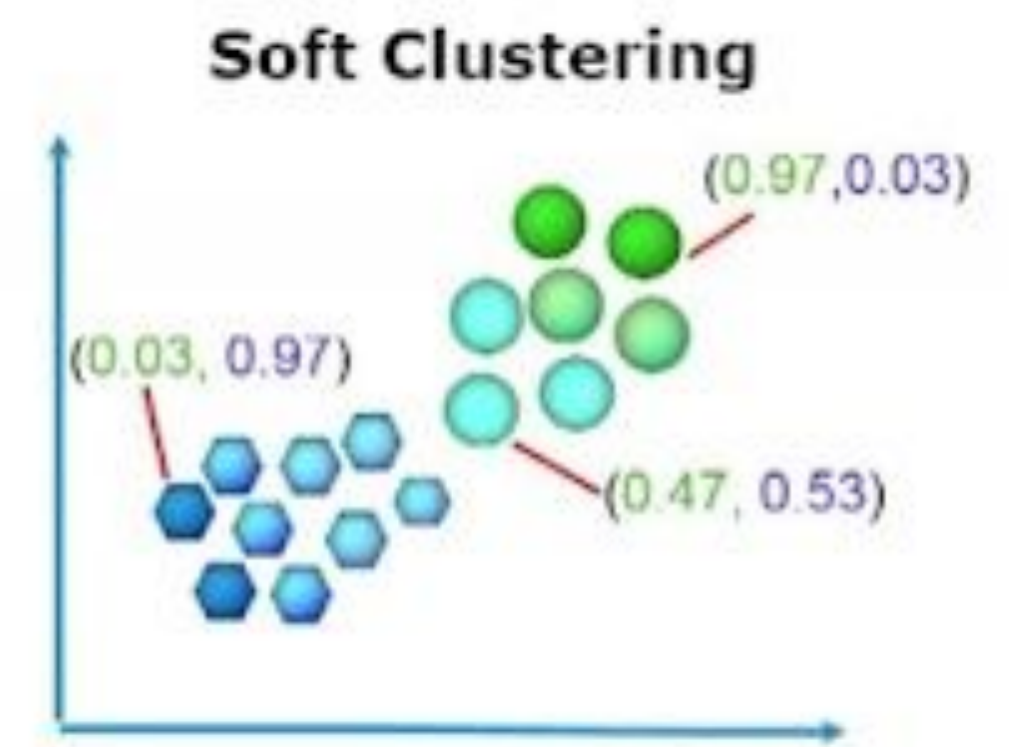
Fuzzy clustering is a form of clustering in which each data point can belong to more than one cluster. Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible.



C-means and k-means

Difference

C-means is fuzzy but k-means is hard (not fuzzy), each point is belonging to a centroid in K-means, but in fuzzy c-means each point can be belonging to two centroids but with different quality. each point either is a part of the first centroids, or the second centroids.



C-means on image

Implementation

Steps

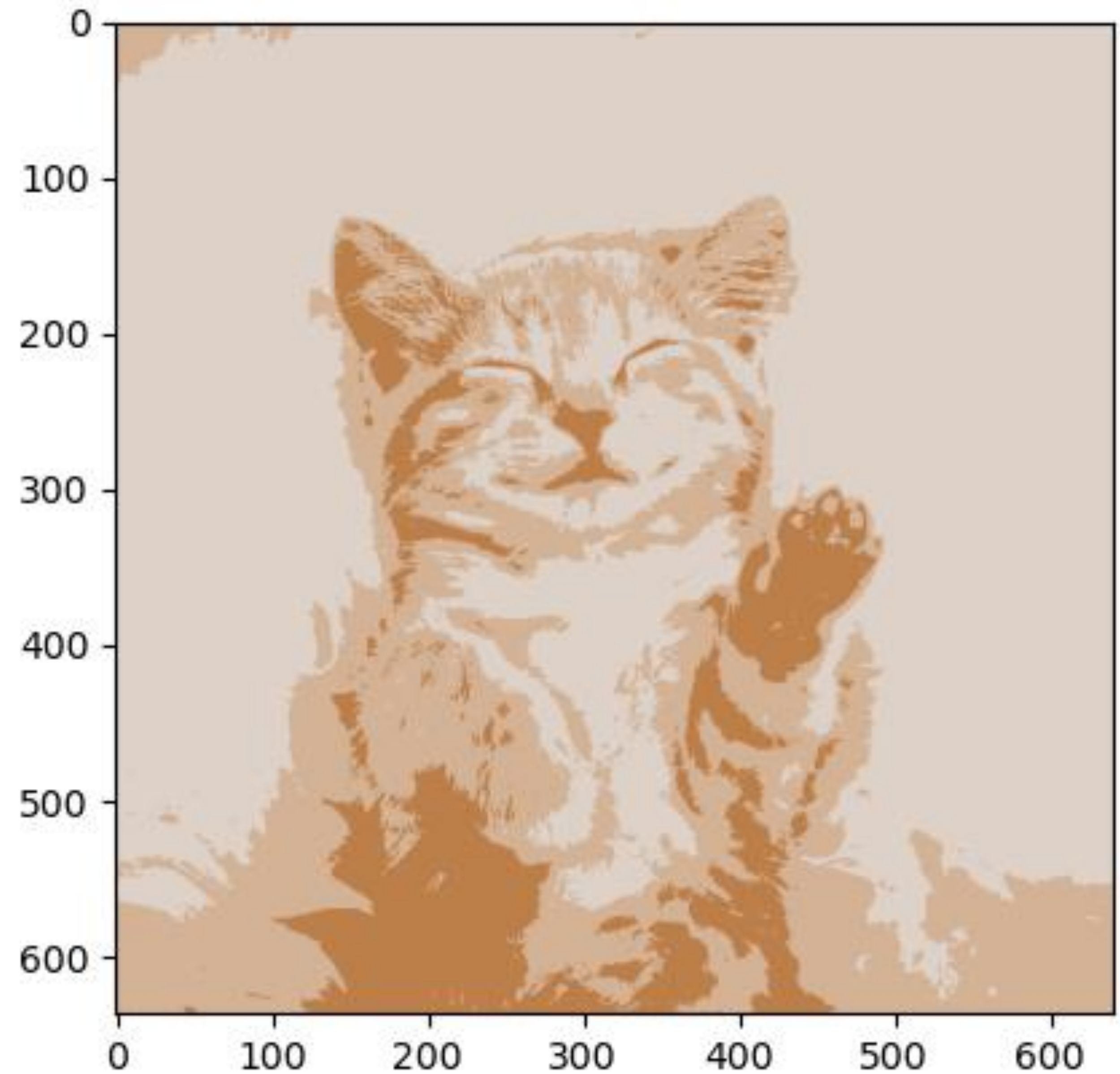
I used a boilerplate code from github that already had implemented c-means image segmentation. The cute cat image was used as input. Same image was also used in k-means implementation.
Url: [Github Repository](#)



Results

Output

Segmented result is provided in output directory.
And also in this slide on right side.



k-means on image

Implementation

Steps

For implementation of k-means image segmentation I used a tutorial from web.

The cute cat image was used as input.

Same image was also used in c-means implementation.

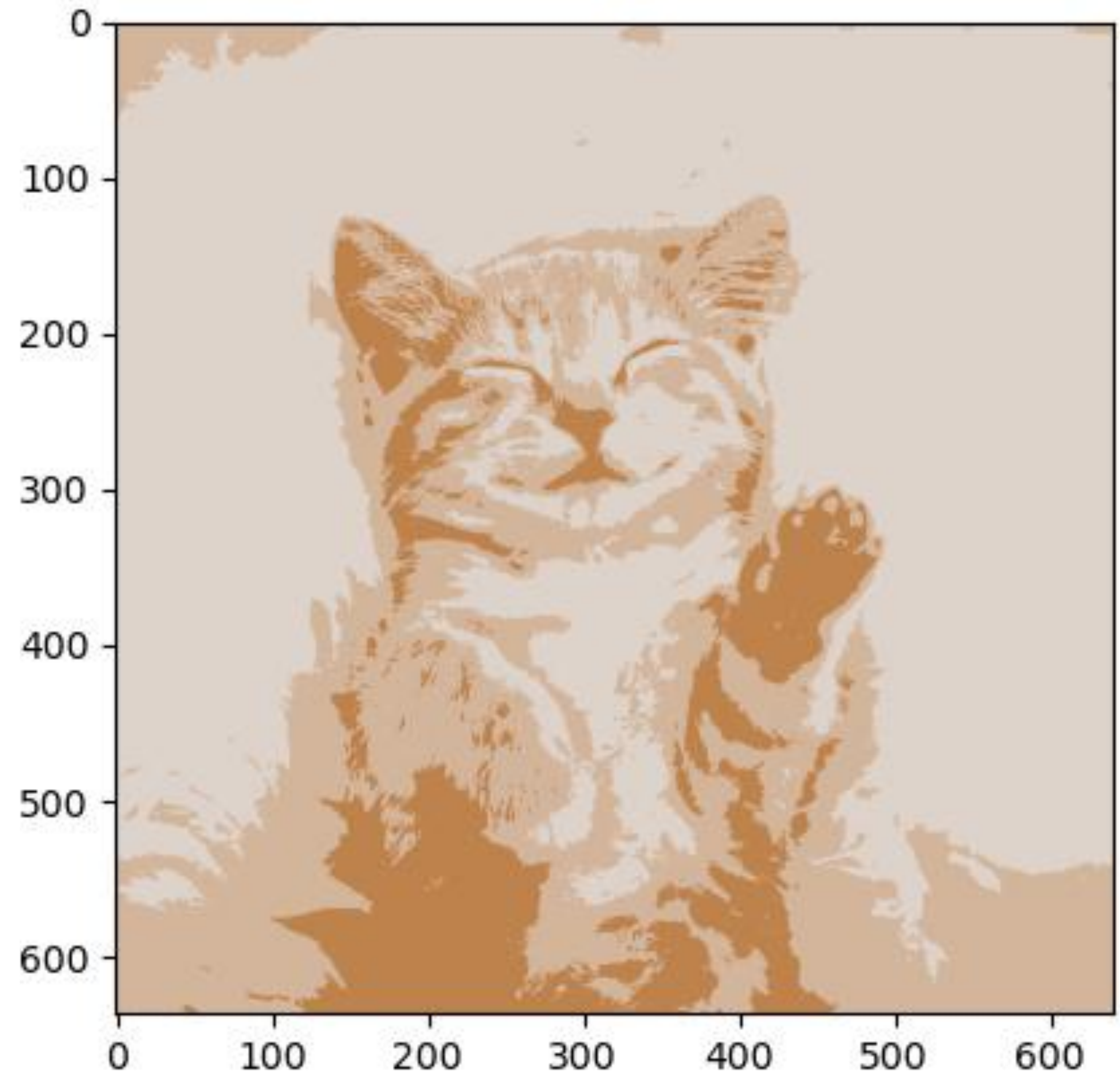
Url: [Website](#)



Results

Output

Segmented result is provided in output directory.
And also in this slide on right side.



C-means on generated dataset

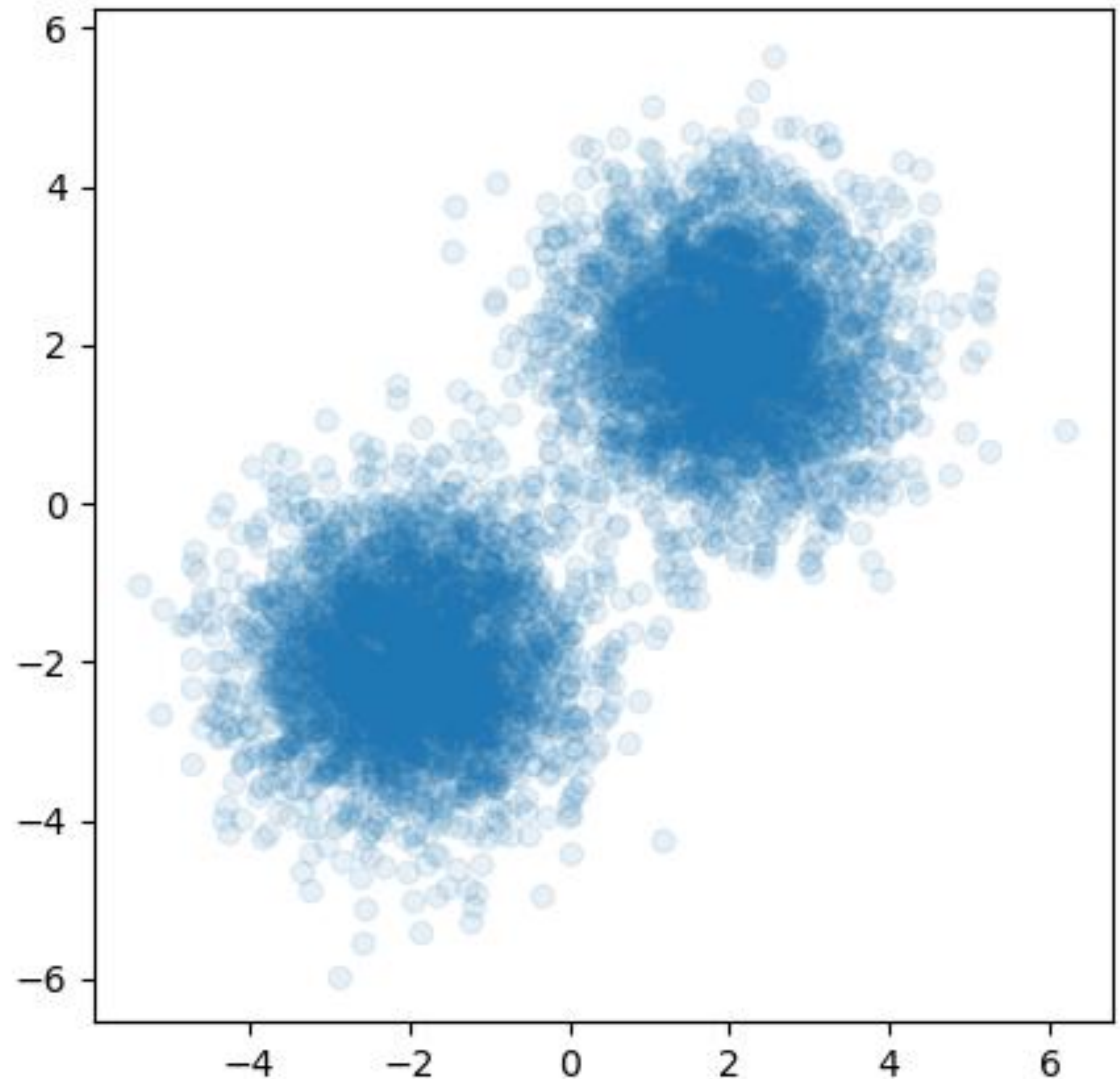
Implementation

Steps

This part was implemented from an example of fuzzy-c-means python module. Fuzzy-c-means is a simple implementation of Fuzzy C-means algorithm.

First step was generating random dataset using numpy library.

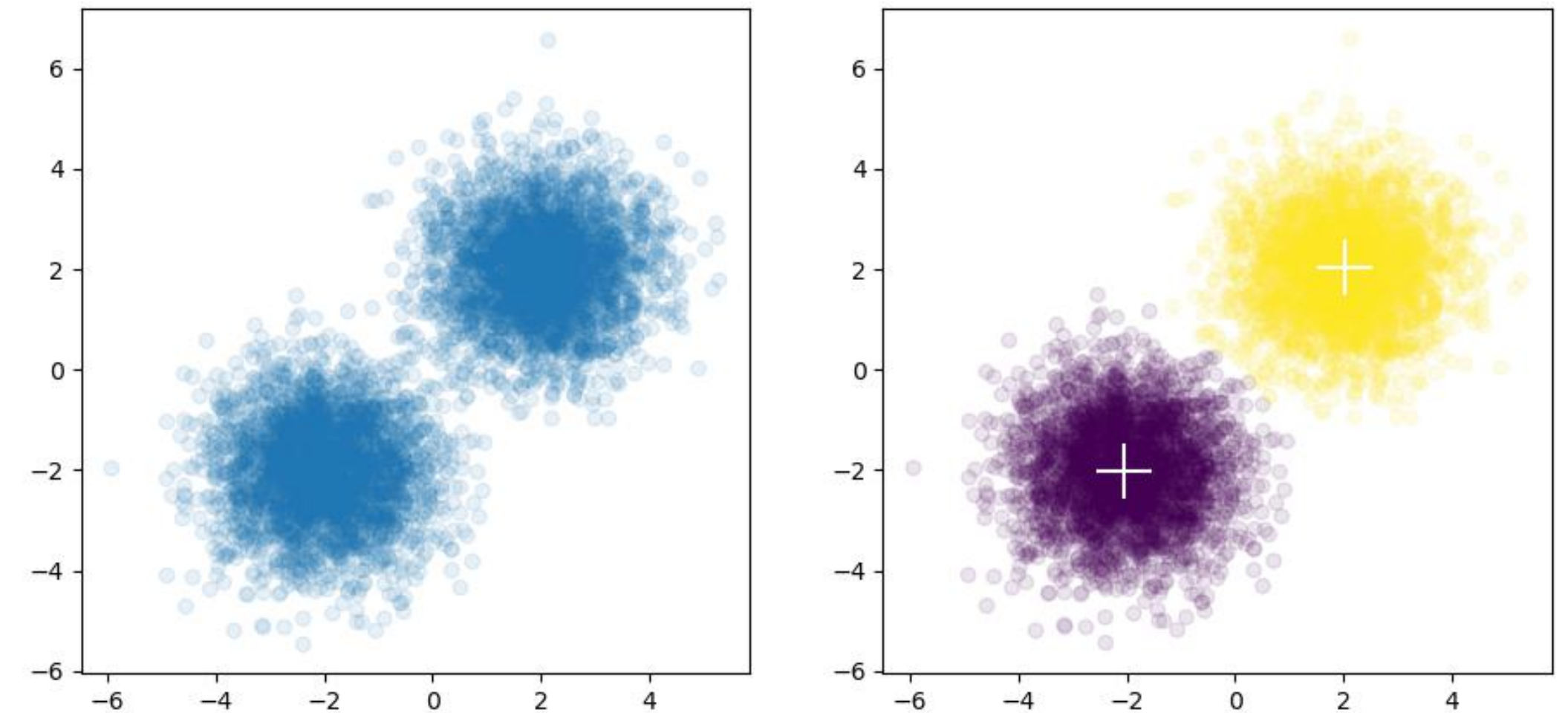
Url: [Website](#)



Results

Output

Clustering result is provided in output directory.
And also in this slide on right side.



C-means on real dataset

Implementation

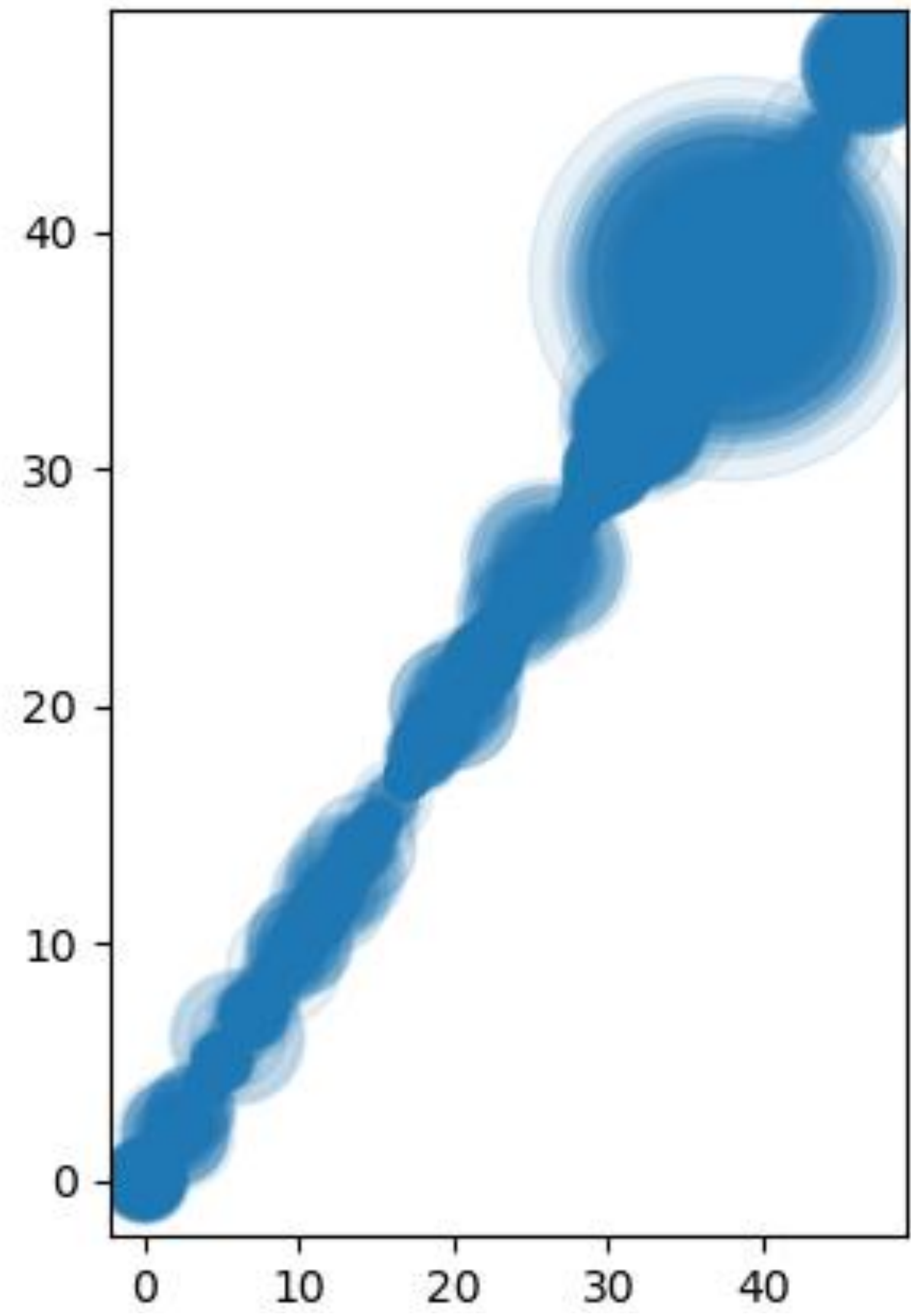
Steps

Dataset was downloaded from kaggle website, titled “Synthetic Financial Datasets For Fraud Detection”.

For a lower resource usage (memory, cpu) only 20 percent of dataset was used.

Also labels on target and source of transactions was converted to numeric value using pandas library.

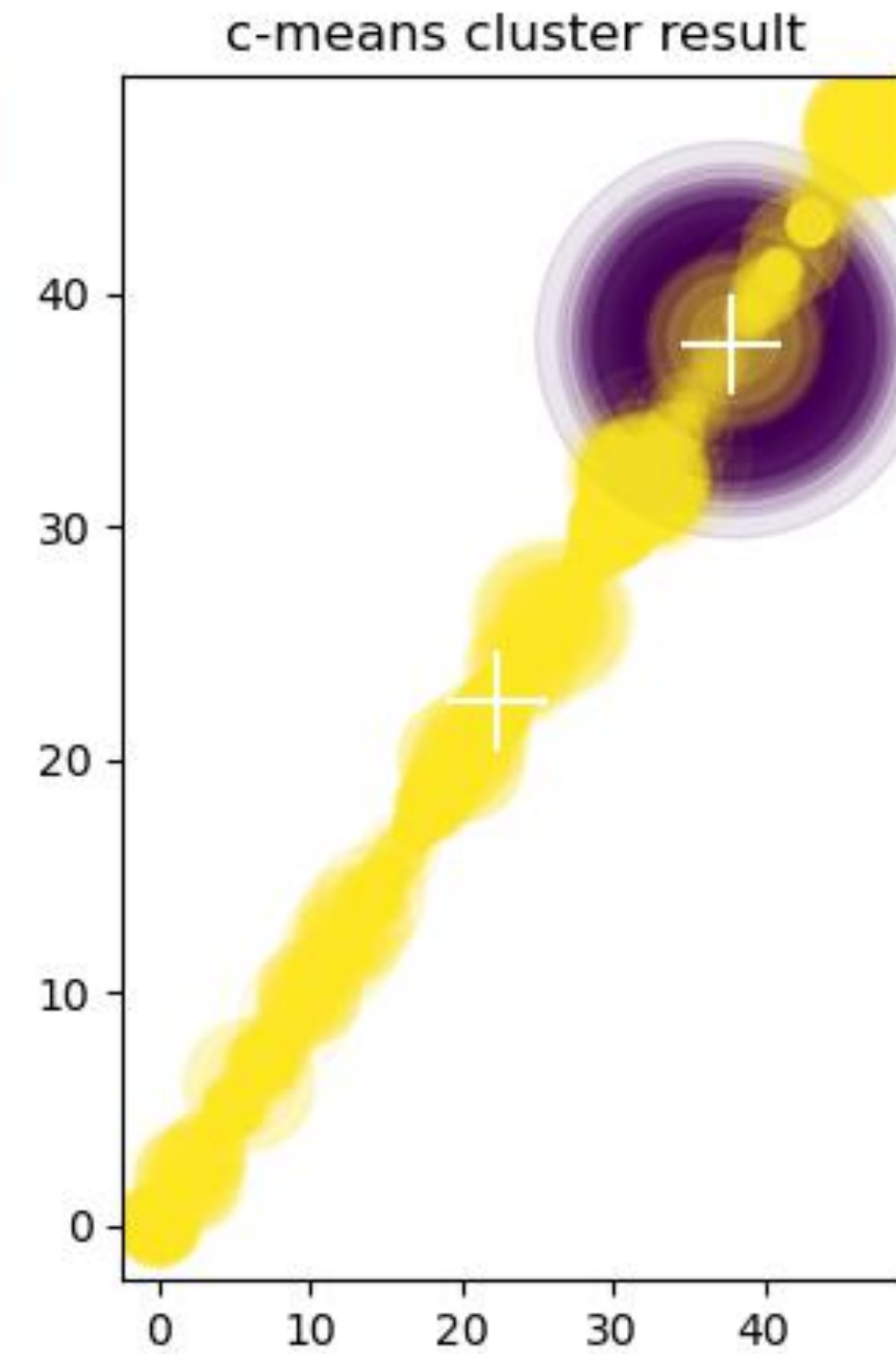
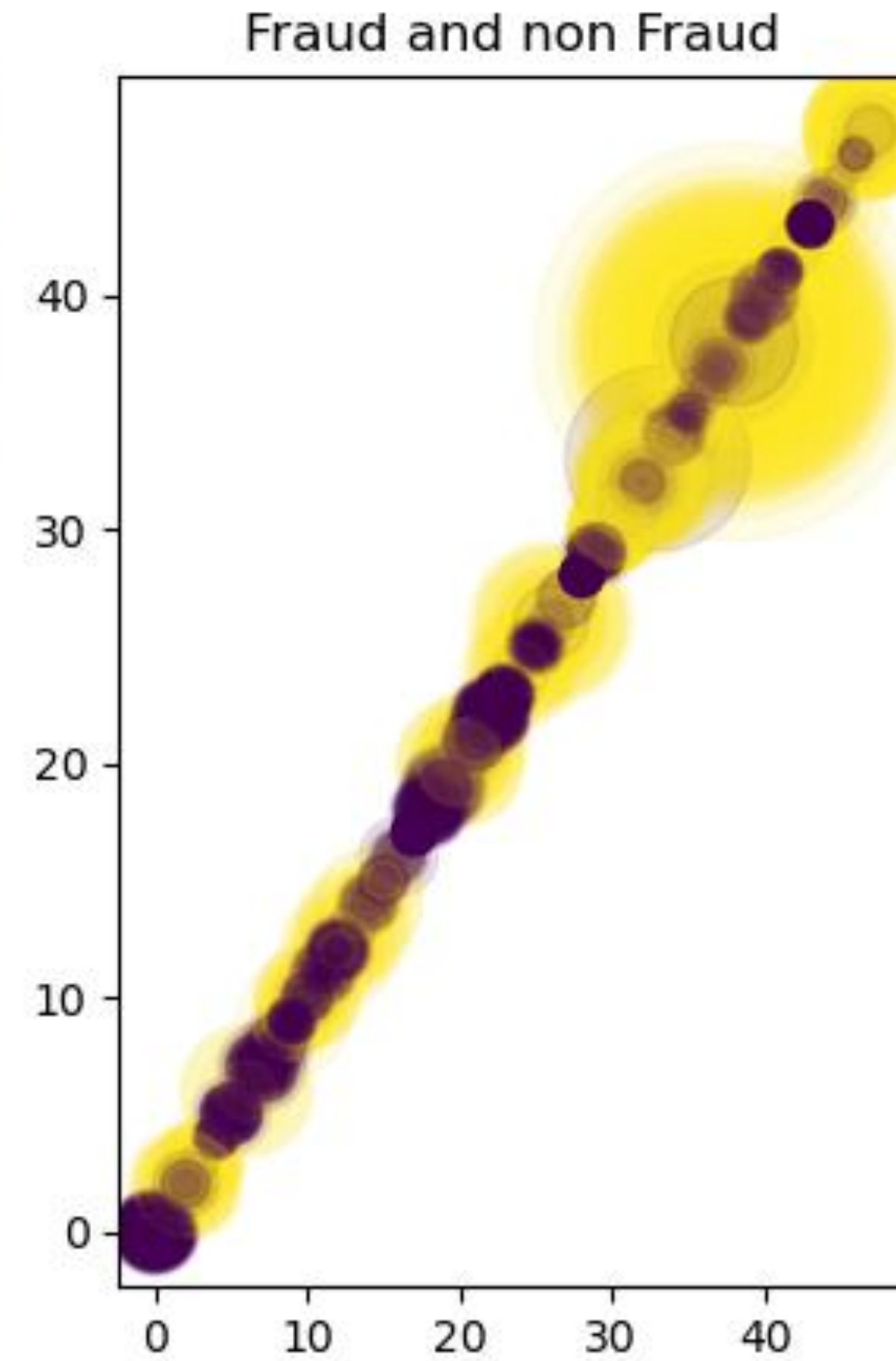
Url: [Website](#)



Results

Output

Clustering result is provided in output directory.
And also in this slide on right side.



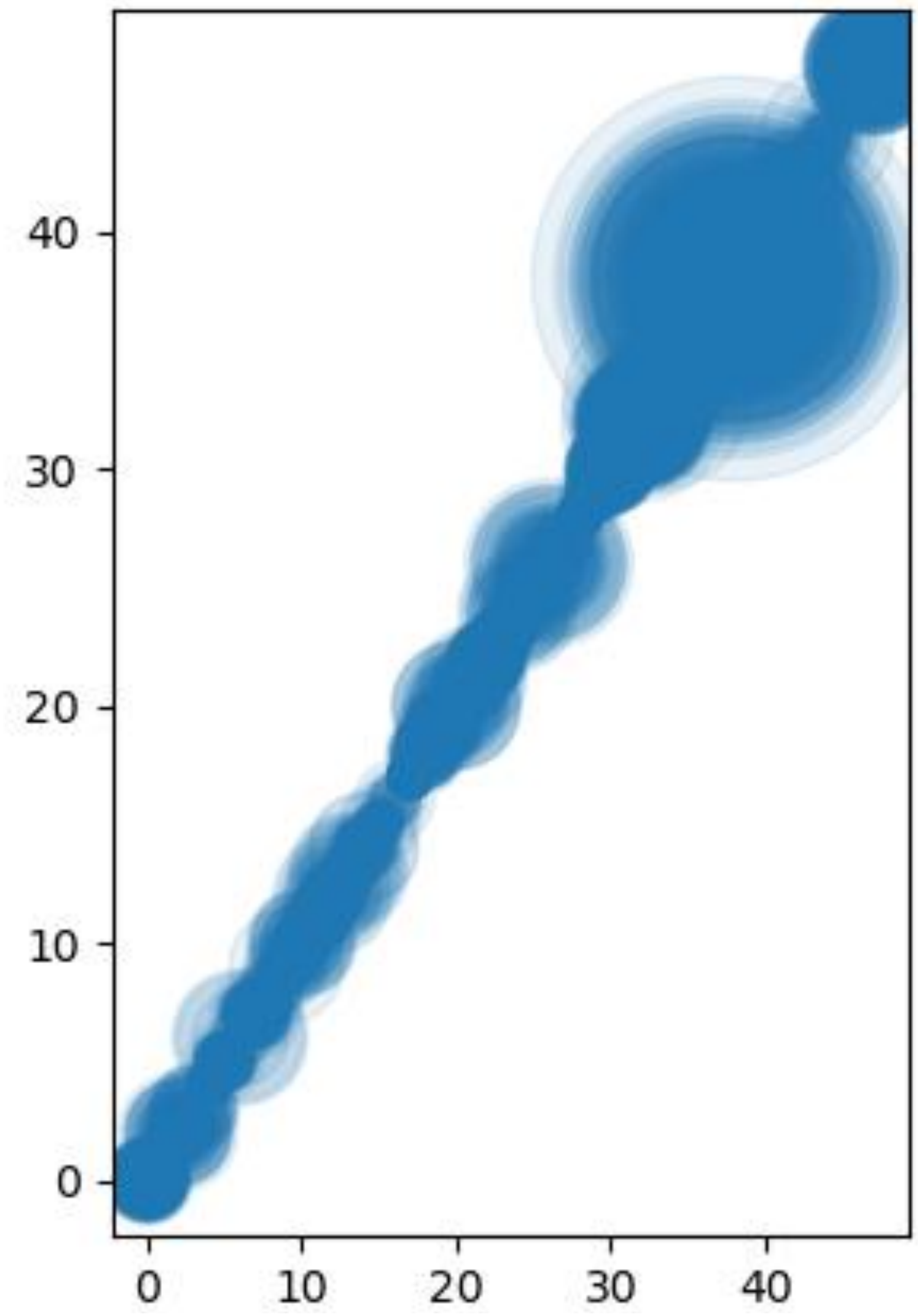
k-means on real dataset

Implementation

Steps

Dataset used is same as last section.

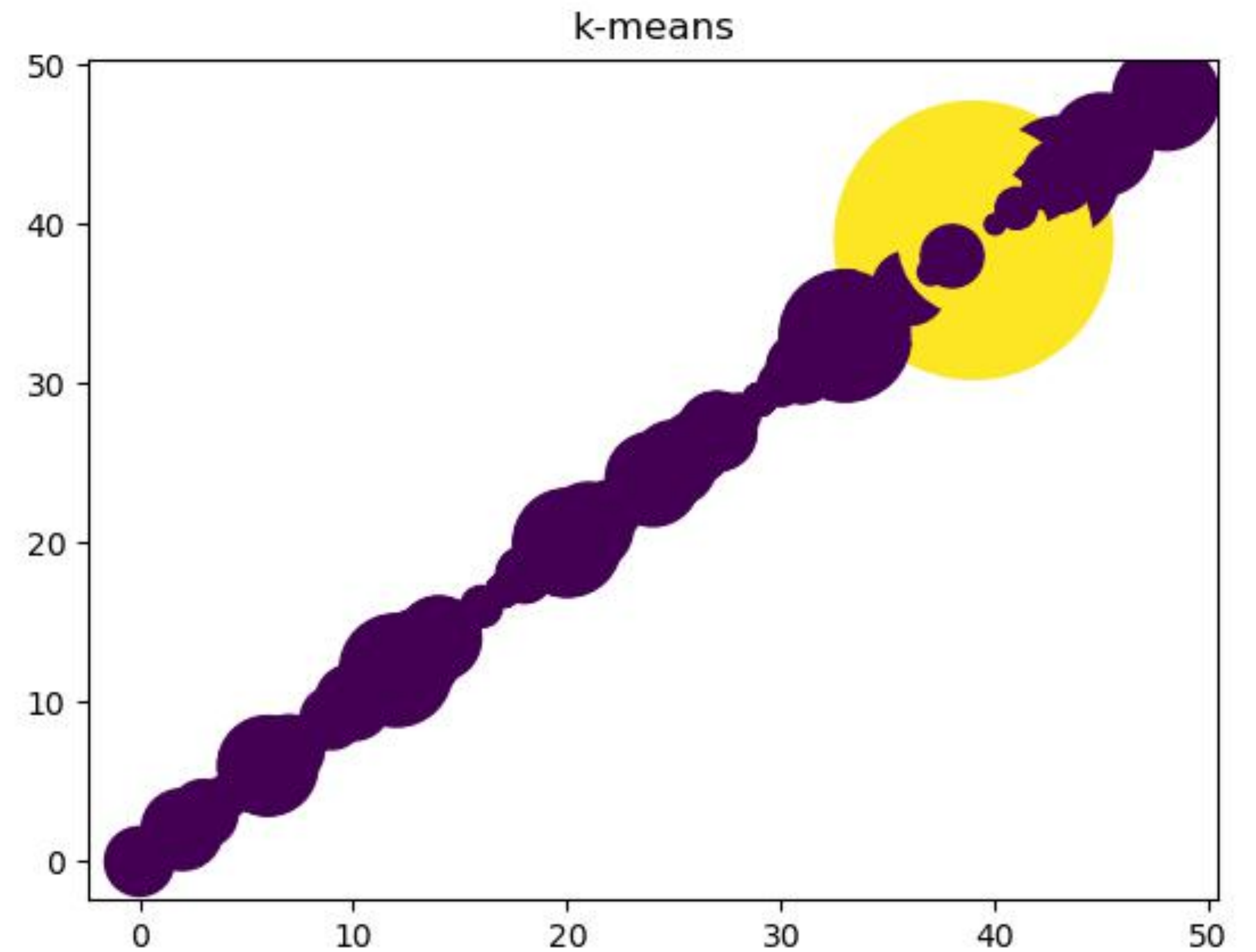
Url: [Website](#)



Results

Output

Clustering result is provided in output directory.
And also in this slide on right side.



**“The Milky Way is
nothing else but a mass
of innumerable stars
planted together in
clusters.”**

Galileo Galilei