

# Pruebe con datos sintéticos los siguientes experimentos\*

luis picon, camila rozales, andres cervantes

Institution(s) of author(s), address(es)  
author@somewhere.host

Received: date / Revised: date / Publised online: data

## 1 resumen

El algoritmo K-means es una técnica de agrupamiento que se encarga de dividir un conjunto de muestras en  $k$  grupos, buscando minimizar la distancia euclidiana ( $d^2$ ) entre los puntos y sus centroides. En esta implementación, vamos a explorar cómo funciona K-means en un espacio tridimensional (3D) utilizando diferentes valores de  $k$ : 2, 3, 5 y 10. Esta variación nos permitirá ver cómo se distribuyen las muestras y cómo se forman los grupos a medida que aumentamos el número de clusters. Es fundamental elegir el valor de  $k$  de manera adecuada para poder interpretar correctamente los resultados y entender mejor la estructura de los datos.

## 2 introducción

El algoritmo K-means es una técnica utilizada en el análisis de datos para agrupar muestras en  $k$  grupos distintos, cada uno representado por un centroide que refleja el promedio de sus puntos.

En esta implementación, se analizará K-means en un espacio tridimensional (3D) con valores de  $k = 2, 3, 5$  y  $10$ , lo que permitirá visualizar cómo se agrupan las muestras según sus características. La selección del valor de  $k$  es fundamental, ya que un valor inapropiado puede llevar a interpretaciones erróneas. x

## 3 objetivos

Analizar las diferentes muestras y clasificarlas en grupos distintos permitiendo una mejor comprensión de sus características y relaciones.

---

\*This research was supported by grant No. xxxx.

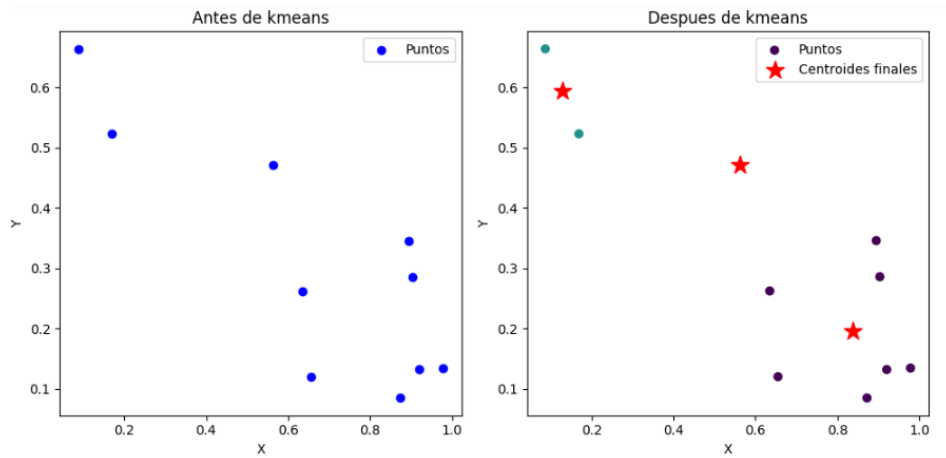


Figure 1: Enter Caption

Analizar el impacto de la distancia euclidiana en la formación de clusters y evaluar la dispersión

Visualizar gráficamente los resultados obtenidos con cada valor de  $k$  colocados por paramestros con las muestras requeridas

### 3.1 imagenes

punto 1 1000 MUESTRAS, 2D,K=3, DISTANCIA EUCLIDIANA, MAX ITERACIONES = 10, 100, 1000, 10000

podemos notar que a medida que a mayor numero de muestras, el color o los puntos de muestran se ven mas intencificados, y con un un grupo de centroides especifico

punto 2: 1000 MUESTRAS, 3D, K=3 DISTANCIA EUCLIDIANA, MAX<sub>I</sub>TERACIONES = 10, 100, 1000, 10000

en las graficas de 3D podemos notar que a medida que a mayor numero de muestras, es imposibles localizar en que punto o cordenada se encuentra los centroides.

PUNTO 3 1000 MUESTRAS, 10D, K=3 DISTANCIA EUCLIDIANA, MAX<sub>I</sub>TERACIONES = 10, 100, 1000, 10000

podemos ver el resultado de los calculos cuando las muestras son, 10,100,1000,10000 pero nos se puede ver graficamente ya que no existen en el momento grficas en 10D

PUNTO 4 1000 MUESTRAS, 100D, K=3 DISTANCIA EUCLIDIANA, MAX<sub>I</sub>TERACIONES = 10, 100, 1000, 10000

podemos ver el resultado de los calculos cuando las muestras son, 10,100,1000,10000 pero nos se puede ver graficamente ya que no existen en el momento grficas en 100D

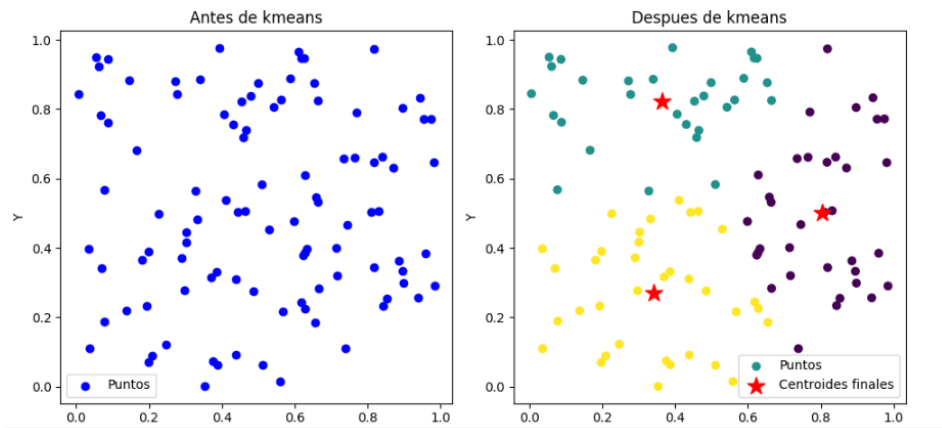


Figure 2: Enter Caption

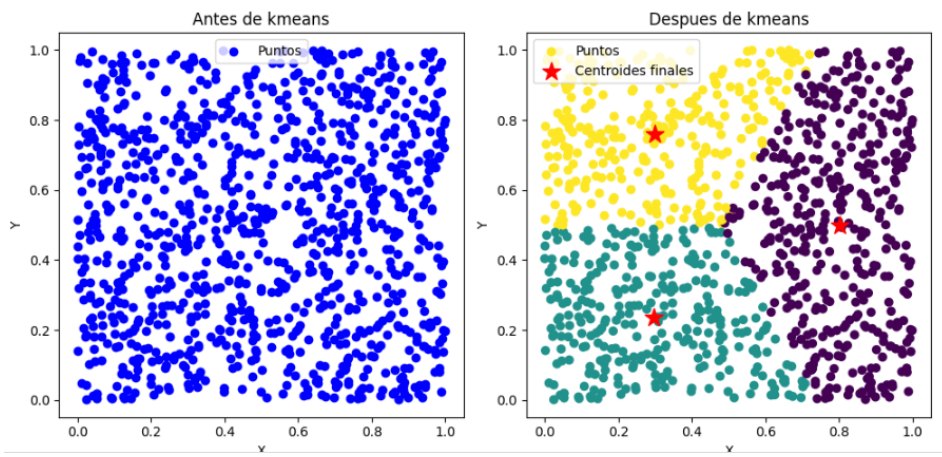


Figure 3: Enter Caption

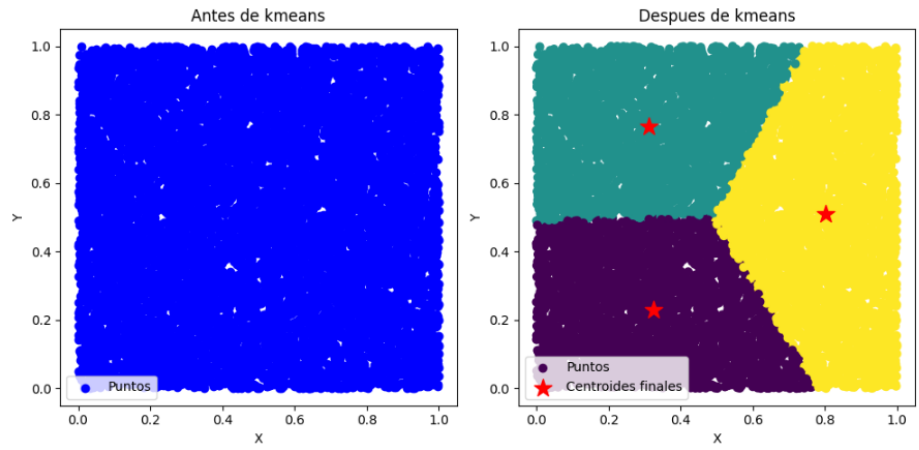


Figure 4: Enter Caption

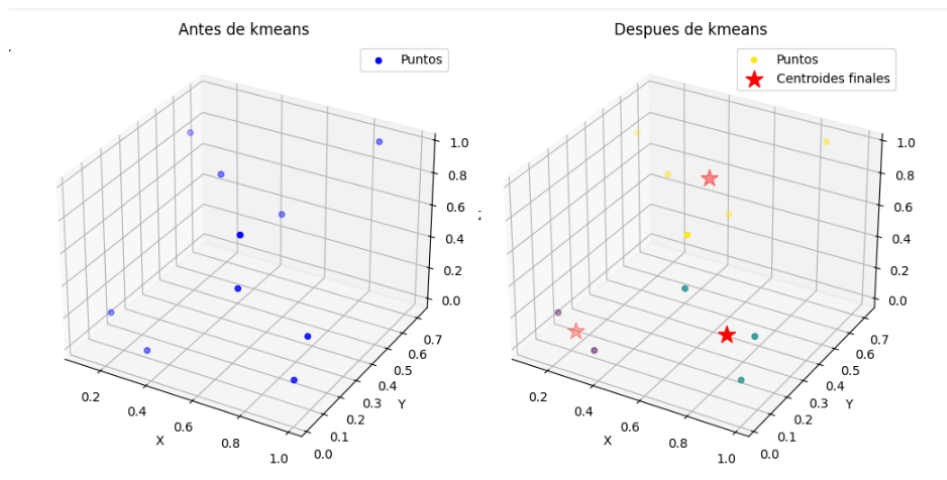


Figure 5: Enter Caption

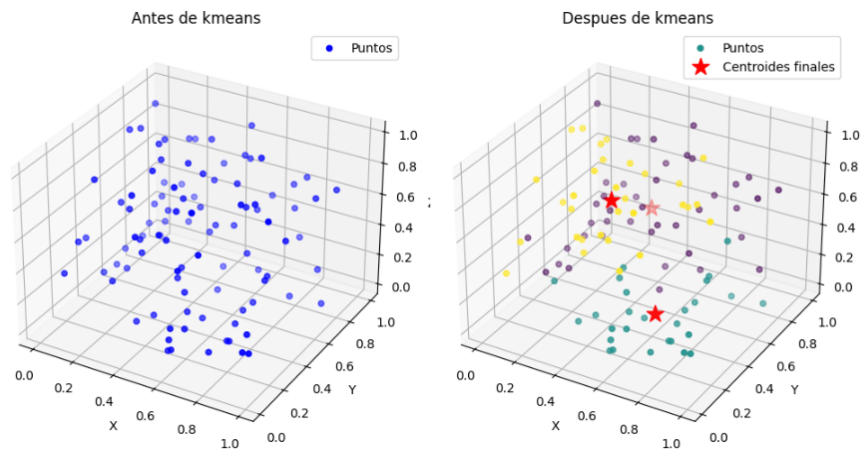


Figure 6: Enter Caption

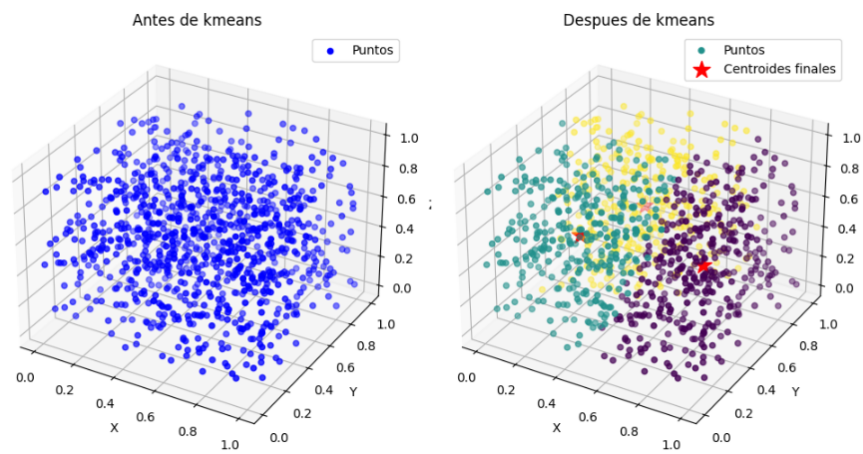


Figure 7: Enter Caption

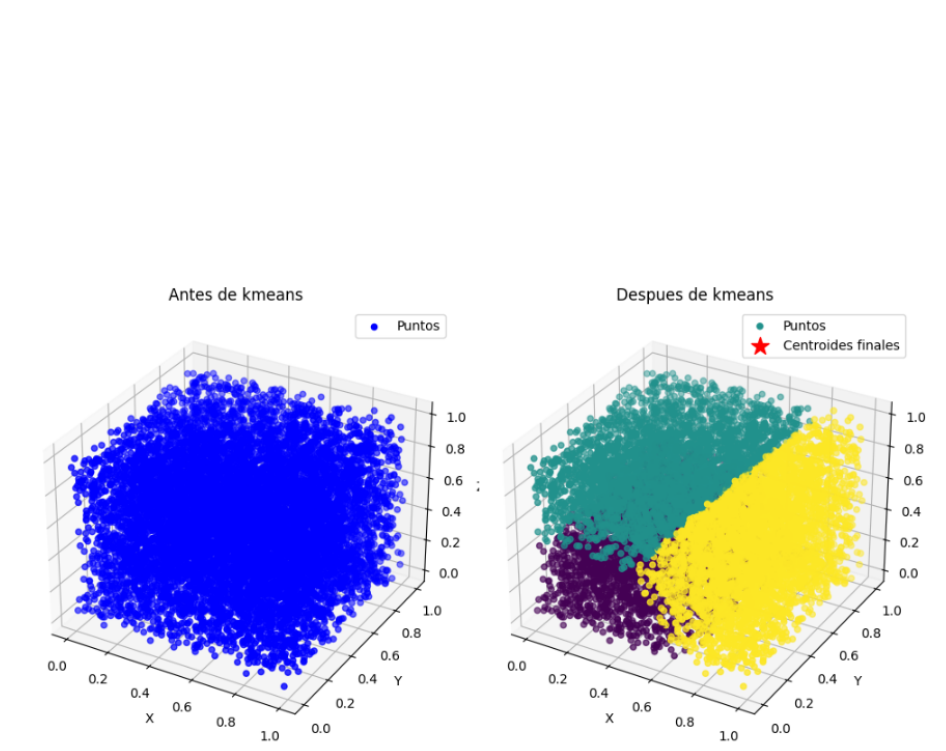


Figure 8: Enter Caption

```

[1 2 2 1 2 2 1 0 0 0]
[[0.66416212 0.7931489 0.63056106 0.43204432 0.52985187 0.28906795
  0.76196204 0.25643537 0.48862519 0.4754988 ]
 [0.38188105 0.35060861 0.47414764 0.6223223 0.23095006 0.40696808
  0.37667827 0.22160388 0.60235289 0.3018301 ]
 [0.30936297 0.47244397 0.52521134 0.77580941 0.42453977 0.44413819
  0.41994818 0.80172075 0.21889352 0.40944003]]

```

Figure 9: Enter Caption

```

[1 0 0 1 0 2 1 2 0 0]
[[0.59323155 0.52750617 0.55650296 0.58464256 0.5390172 0.62928394
0.40063092 0.26562792 0.64149022 0.58917221 0.60513476 0.31256732
0.65795063 0.22536074 0.42880742 0.56287226 0.37988087 0.44995222
0.68088634 0.39903275 0.39911093 0.67032932 0.69235737 0.45051024
0.37568536 0.33356354 0.50283537 0.71100391 0.34659431 0.54355184
0.35286076 0.53483001 0.44452632 0.38315529 0.25617475 0.29619414
0.4226703 0.67791125 0.60902664 0.47169225 0.59577819 0.32031441
0.75805486 0.60055626 0.4810418 0.33246642 0.80617101 0.36838104
0.65082112 0.63035826 0.56207465 0.4638802 0.81361399 0.43834651
0.63535493 0.3177117 0.54865018 0.43575819 0.47774991 0.51030328
0.44299843 0.33348487 0.51324888 0.64253883 0.41094765 0.33098478
0.57875675 0.55301671 0.33418701 0.65396406 0.53199366 0.21369756
0.31018441 0.55397891 0.48562292 0.72545826 0.71499452 0.54775454
0.41696368 0.46796534 0.49529585 0.39007093 0.39819589 0.3297251
0.3322349 0.26747084 0.68979775 0.64908372 0.2882698 0.35494998
0.52577686 0.5795951 0.57082589 0.48999671 0.38714294 0.33459529
0.32488227 0.8189655 0.56780138 0.25330244]
[0.65164941 0.74729991 0.33565794 0.89193706 0.52637785 0.63939411
0.52275059 0.94144355 0.48283662 0.30808916 0.37315561 0.35980545
0.48994355 0.58725941 0.57797741 0.56399008 0.49602917 0.73214463
0.68528888 0.39897587 0.81963028 0.46126284 0.44832394 0.57274084
0.34870041 0.35739522 0.34830173 0.57720092 0.24417385 0.77899501
0.77798504 0.62909388 0.58909223 0.6290394 0.45906215 0.41185326
0.20609061 0.4288807 0.65141259 0.34541842 0.49312132 0.57211576
0.69828204 0.33237591 0.11611503 0.49450944 0.40117752 0.52719444
0.39349432 0.64739157 0.42745285 0.78216492 0.23483307 0.35986099
0.3044020 0.67570751 0.40257257 0.38677055 0.70771762 0.50072777

```

Figure 10: Enter Caption

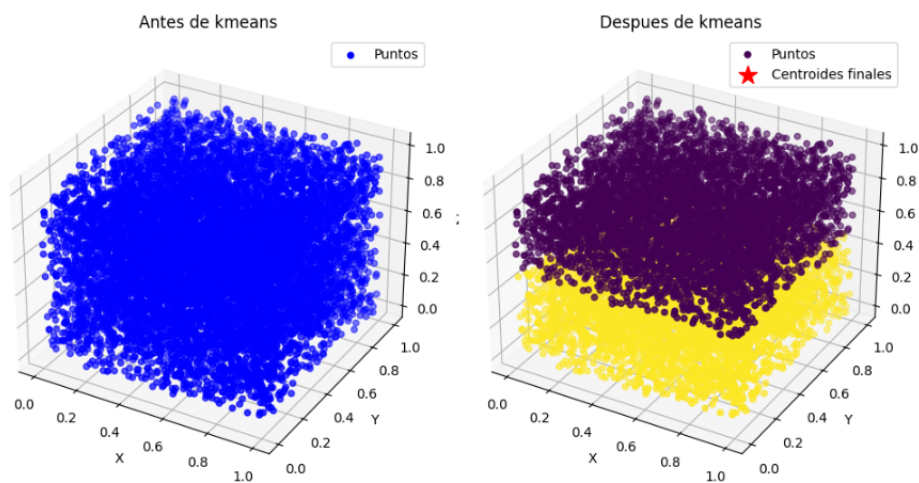


Figure 11: Enter Caption

PUNTO 5 1000 MUESTRAS, 3D, K=2,3,5,10 DISTANCIA EUCLIDIANA,  $MAX\_ITERACIONES = 10, 100, 1000, 10000$

podemos concluir que a medida que se cambia o se crea el numero de k-means para los centroides los datos de muestras tienden a tener una mayor convergencia

**PUNTO #6** 1000 MUESTRAS, 2D, K=5 DISTANCIA EUCLIDIANA, MANHATTAN, MAHALANOBIS,  $MAX\_ITERACIONES = 1000$

podemos notar que en cada centroide la convergencia varia.



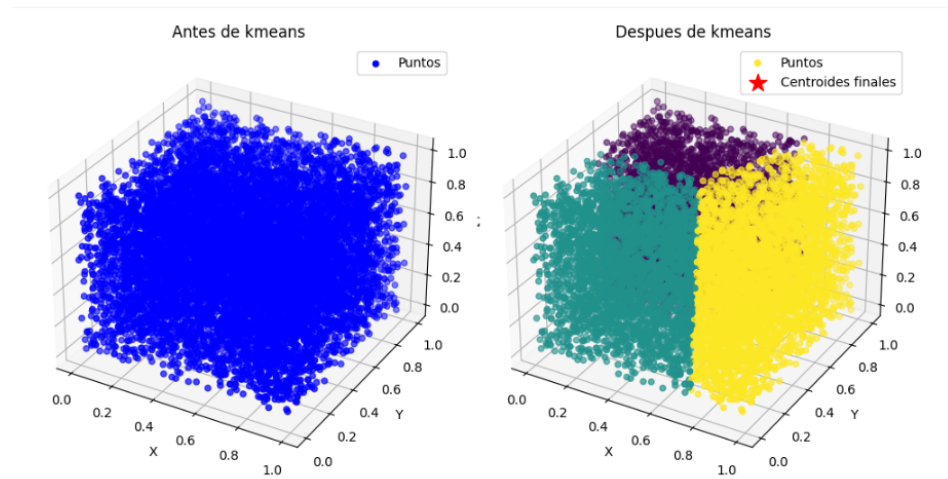


Figure 12: Enter Caption

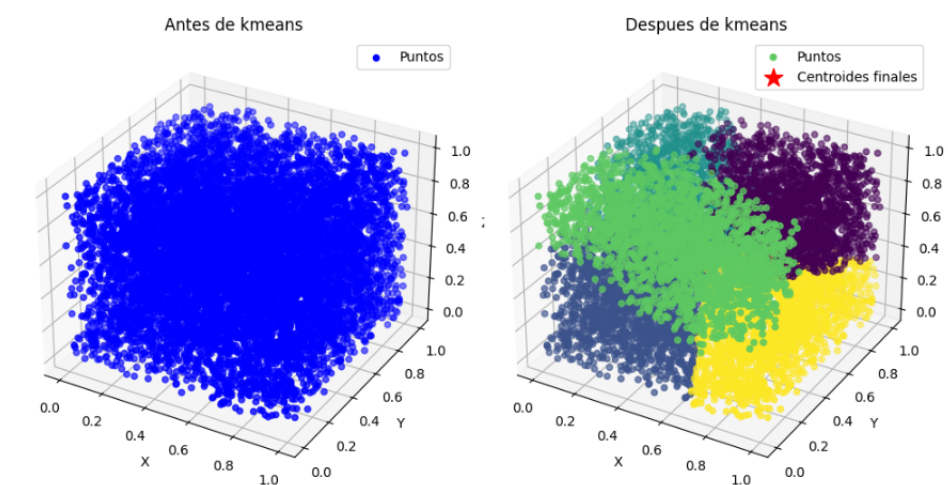


Figure 13: Enter Caption

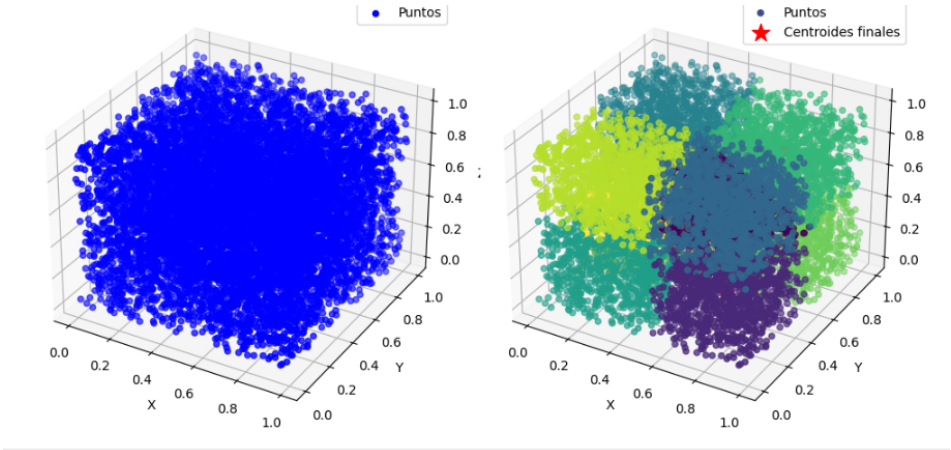


Figure 14: Enter Caption

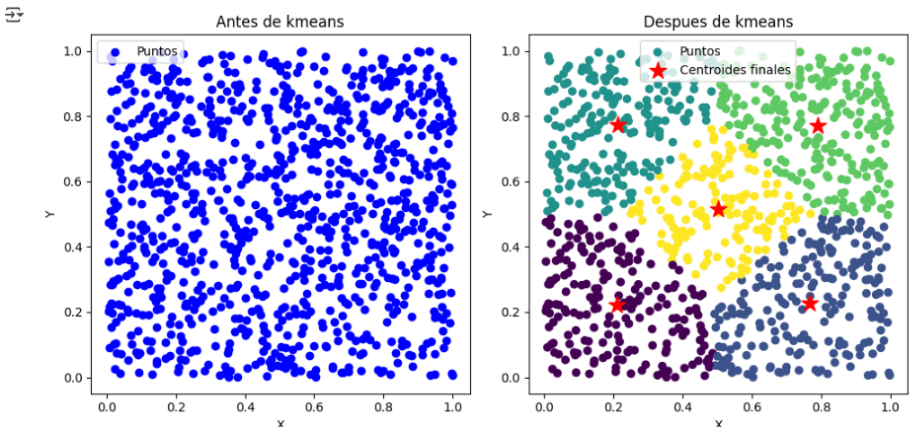


Figure 15: Enter Caption