



**SEP**  
SECRETARÍA DE  
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO



**TECNOLÓGICO NACIONAL DE MÉXICO INSTITUTO**

**TECNOLÓGICO DE TIJUANA  
SUBDIRECCIÓN ACADÉMICA**

**DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN**

**SEMESTRE FEBRERO-JUNIO 2022**

**CARRERA**

**Ingeniería en informática e Ingeniería en Sistemas**

**Computacionales**

**MATERIA**

**Minería de datos**

**TÍTULO**

**Práctica evaluatoria, unidad #4**

**Integrantes:**

**Munguía Silva Edgar Geovanny #17212344**

**López Higuera Saúl Alfredo #18210493**

**NOMBRE DEL MAESTRO**

**Jose Christian Romero Hernandez**



**Tijuana Baja California 01 de junio del 2022**

## Introducción.

Esta práctica, es la última del curso de minería de datos, el objetivo principal es usar k means en un marco de datos para categorizar datos sin etiquetas en grupos, antes de comenzar, definamos qué significa el algoritmo: Es un método de agrupamiento, que tiene como objetivo dividir un conjunto de n observaciones en k group en los que cada observación pertenece al grupo cuyo valor medio es más cercano.

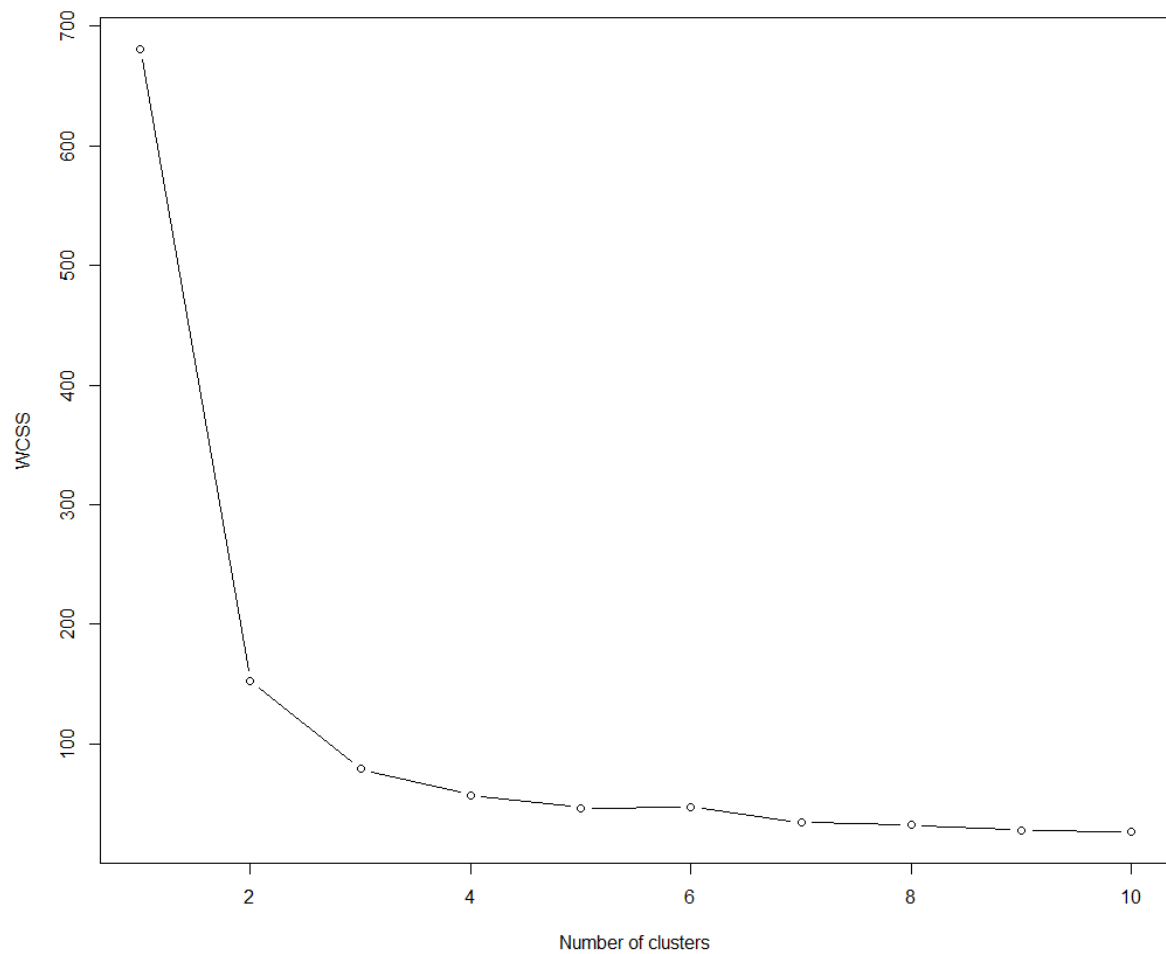
## Código.

```
# Importando el conjunto de datos.
dataset = read.csv(file.choose())
dataset = dataset[1:4]
```

 Iris	11/11/2019 08:52 p. m.	Archivo de valores...	4 KB
 u4-final	26/05/2022 11:06 p. m.	Archivo de origen R	1 KB

```
# Usar el método del codo para encontrar el número óptimo de
grupos.
set.seed(6)
wcss = vector()
for (i in 1:10) wcss[i] = sum(kmeans(dataset, i)$withinss)
plot(1:10,
     wcss,
     type = 'b',
     main = paste('The Elbow Method'),
     xlab = 'Number of clusters',
     ylab = 'WCSS')
```

### The Elbow Method



```
# Ajuste de K-Means al conjunto de datos.  
set.seed(29)  
kmeans = kmeans(x = dataset, centers = 3)  
kmeans
```

20:7 (Top Level) 

Terminal ✕

Jobs ✕

```
> set.seed(29)
```

```
> kmeans
```

cluster means:

1	5.901613	2.748387	4.393548	1.433871
---	----------	----------	----------	----------

3	5.006000	3.418000	1.464000	0.244000
---	----------	----------	----------	----------

```
[1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
```

```
[37] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

[73] 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2

[109] 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2

$$[145] \quad 2 \quad 2 \quad 1 \quad 2 \quad 2 \quad 1$$

within cluster sum of squares by cluster:

[1] 39.82097 23.87947 15.24040

```
(between_SS / total_SS =  88.4 %)
```

Available components:

```
[1] "cluster"      "centers"      "totss"        "withinss"
```

```
[5] "tot.withinss" "betweenss" "size" "iter"
```

```
[9] "ifault"
```

>

```
y kmeans
```

```
> y_kmeans = kmeans$cluster
```

```
> y_kmeans
```

[79] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1

>

```
install.packages('cluster')
```

```
library(cluster)
```

```
clusplot(dataset,
```

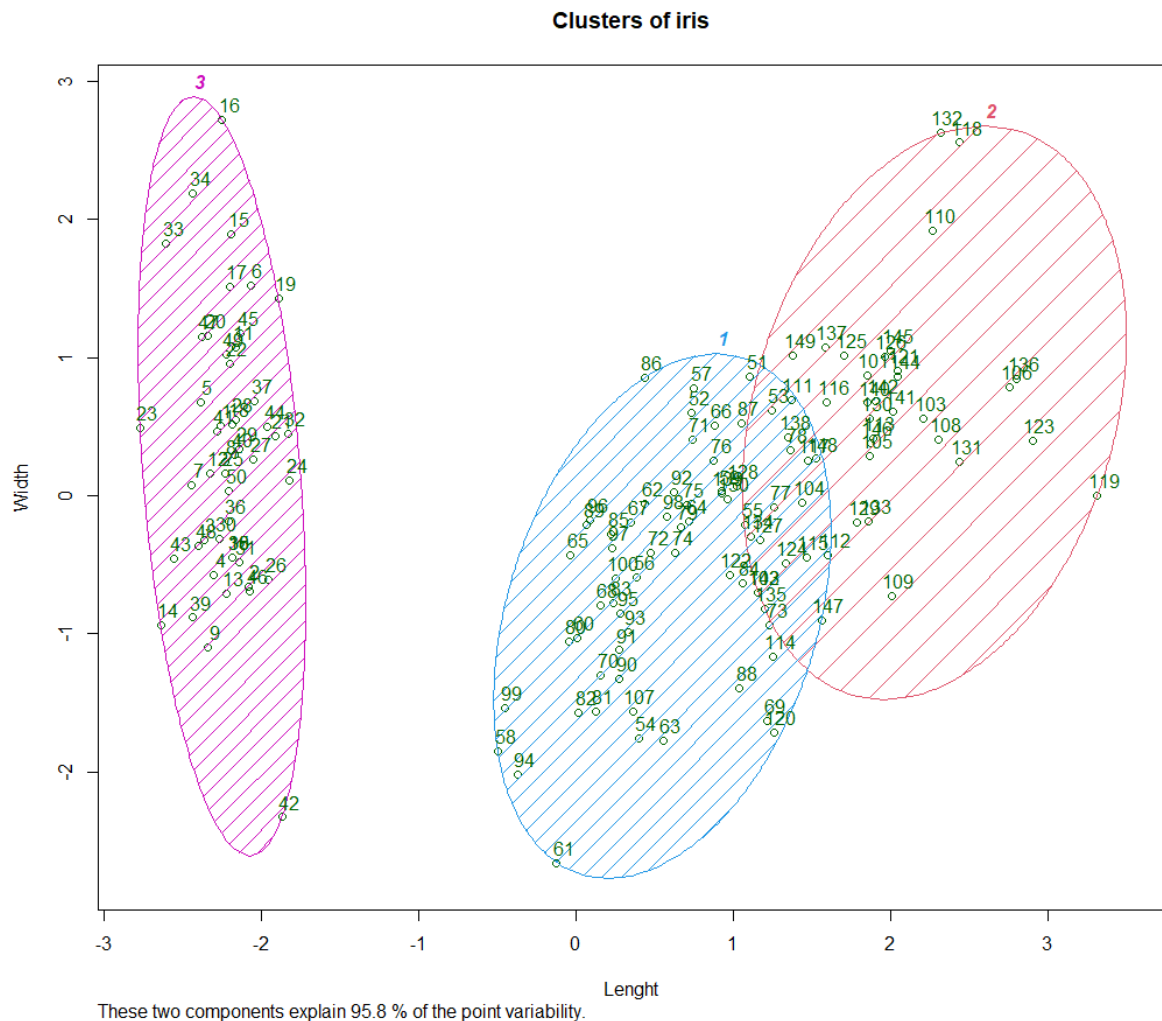
```
y_kmeans,
```

```
lines = 0,
```

```

shade = TRUE,
color = TRUE,
labels = 2,
plotchar = FALSE,
span = TRUE,
main = paste('Clusters of iris'),
xlab = 'Lenght',
ylab = 'Width')

```



## Conclusiones.

### Edgar Munguia:

El uso de este algoritmo nos permite agrupar los datos y encontrar puntos con menos errores, es decir, este modelo querrá precisión sobre todas las cosas, esto es bueno si queremos agrupar los datos con el mínimo error posible. Una vez más, spark es una herramienta poderosa para trabajar con big data, como podemos ver,

nos permite trabajar con mucha información y también nos permite trabajar con menos errores que con otras herramientas.

**Saúl López:**

Este es un método súper útil. Recuerde que la idea básica detrás de los métodos de partición, como la agrupación en clústeres de k-medias, es definir los clústeres de tal manera que se minimice la variación total dentro del clúster. El método Elbow analiza el WSS total en función del número de grupos.

**Link del video (Youtube):** <https://www.youtube.com/watch?v=bvDxlgLKr00>

**Link del repositorio de GitHub:**

[https://github.com/Saul12344/mineria-de-datos/tree/unit\\_4](https://github.com/Saul12344/mineria-de-datos/tree/unit_4)