

PROYECTO FINAL

SERGIO ADRIAN ORTIZ ORTEGA

2022-05-30

INTRODUCCION

Este conjunto de datos fue descargado por la pagina Kaggle y que incluye 721 Pokémon, incluido su número, nombre, primer y segundo tipo y estadísticas básicas: HP, Ataque, Defensa, Ataque especial, Defensa especial y Velocidad.

Pokémon es una franquicia de medios que originalmente comenzó como un videojuego RPG, pero debido a su popularidad ha logrado expandirse a otros medios de entretenimiento como series de televisión, películas, juegos de cartas, ropa, entre otros, convirtiéndose en una marca que es reconocida en el mercado mundial.

La base de datos fue revisada y modificada antes del analisis, cheque que no hubiera datos faltantes, y elimine datos de manera manual a partir de la fila 151 para tener solo los de la primera generacion de pokemons y tambien elimine las mega-evoluciones.

Con el codigo que nos proporciono la maestra de k-means y un poco de investigacion hare un analisis k-means para ver que numero de clouster son los mas adecuados para la investigacion.

K-MEANS

Cargar las LIBRERIAS

```
library(tidyverse)
library(cluster)
library(factoextra)
library(NbClust)
library(readxl)
library(dplyr)
library(ggplot2)
library(tools)
library(MVN)
library(cluster)
library(carData)
library(car)
library(rJava)
library(xlsx)
library(plotly)
library(fpc)
```

Reconocimiento de la matriz de datos

```
Pokemon <- read_csv("Pokemon.csv")
```

```
## Rows: 151 Columns: 13
```

```
## -- Column specification -----
## Delimiter: ","
## chr (3): Name, Type_1, Type_2
## dbl (9): Pokedex, Total, HP, Attack, Defense, Sp_Atk, Sp_Def, Speed, Generation
## lgl (1): Legendary
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Exploracion de matriz

```
dim(Pokemon)
```

```
## [1] 151 13
```

```
str(Pokemon)
```

```
## spec_tbl_df [151 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Pokedex : num [1:151] 1 2 3 4 5 6 7 8 9 10 ...
## $ Name : chr [1:151] "Bulbasaur" "Ivysaur" "Venusaur" "Charmander" ...
## $ Type_1 : chr [1:151] "Grass" "Grass" "Grass" "Fire" ...
## $ Type_2 : chr [1:151] "Poison" "Poison" "Poison" NA ...
## $ Total : num [1:151] 318 405 525 309 405 534 314 405 530 195 ...
## $ HP : num [1:151] 45 60 80 39 58 78 44 59 79 45 ...
## $ Attack : num [1:151] 49 62 82 52 64 84 48 63 83 30 ...
## $ Defense : num [1:151] 49 63 83 43 58 78 65 80 100 35 ...
## $ Sp_Atk : num [1:151] 65 80 100 60 80 109 50 65 85 20 ...
## $ Sp_Def : num [1:151] 65 80 100 50 65 85 64 80 105 20 ...
## $ Speed : num [1:151] 45 60 80 65 80 100 43 58 78 45 ...
## $ Generation: num [1:151] 1 1 1 1 1 1 1 1 1 1 ...
## $ Legendary : logi [1:151] FALSE FALSE FALSE FALSE FALSE FALSE ...
## - attr(*, "spec")=
## .. cols(
## .. Pokedex = col_double(),
## .. Name = col_character(),
## .. Type_1 = col_character(),
## .. Type_2 = col_character(),
## .. Total = col_double(),
## .. HP = col_double(),
## .. Attack = col_double(),
## .. Defense = col_double(),
## .. Sp_Atk = col_double(),
## .. Sp_Def = col_double(),
## .. Speed = col_double(),
## .. Generation = col_double(),
## .. Legendary = col_logical()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
colnames(Pokemon)
```

```
## [1] "Pokedex" "Name" "Type_1" "Type_2" "Total"
## [6] "HP" "Attack" "Defense" "Sp_Atk" "Sp_Def"
## [11] "Speed" "Generation" "Legendary"
```

```
##cambiar los tipos de las variables
```

```

nombre<-factor(Pokemon$Name)
Tipo<-factor(Pokemon$Type_1,
             levels= c("Grass","Fire","Water","Bug","Normal","Poison","Electric",
                       "Ground","Fairy","Fighting","Flying","Psychic","Rock",
                       "Steel","Ice","Dragon","Dark","Ghost"))
Generacion<-factor(Pokemon$Generation)
legendario<-factor(Pokemon$Legendary)

HP<-as.numeric(Pokemon$HP,strict = TRUE)
Attack<-as.numeric(Pokemon$Attack,strict = TRUE)
Defense<-as.numeric(Pokemon$Defense,strict = TRUE)
Sp_Atk<-as.numeric(Pokemon$Sp_Atk,strict = TRUE)
Sp_Def<-as.numeric(Pokemon$Sp_Def,strict = TRUE)
Speed<-as.numeric(Pokemon$Speed,strict = TRUE)

X<-data.frame(nombre,HP,Attack,Defense,Sp_Atk,Sp_Def,Speed)
BASE=X

```

volver factor variable cualitativa

```
BASE$nombre=as.factor(BASE$nombre)
```

Volvemos “Estado” al marco de los datos

```
DATOS = data.frame(BASE,row.names=BASE$nombre)
```

Eliminacion de la variables

```
DATOS[, 1] <- NULL
```

Separacion de filas y columnas.

```

n<-dim(DATOS)[1]
p<-dim(DATOS)[2]

```

Estandarizacion univariante.

```
X.s<-scale(DATOS)
```

Escalar la base de datos

```
datos.scale = scale (DATOS)
```

Matrix de distancia

```
Mdistancia = get_dist(datos.scale,method = "manhattan")
```

Estimar el numero de cluster

```
#numCluster = NbClust(data=DATOS, method = "median", distance = "euclidean", diss=NULL, min.nc=2, max.
```

-
- Among all indices:
 - 14 proposed 2 as the best number of clusters
 - 1 proposed 3 as the best number of clusters
 - 2 proposed 4 as the best number of clusters
 - 9 proposed 5 as the best number of clusters
 - 1 proposed 6 as the best number of clusters

***** Conclusion *****

- According to the majority rule, the best number of clusters is 2
-

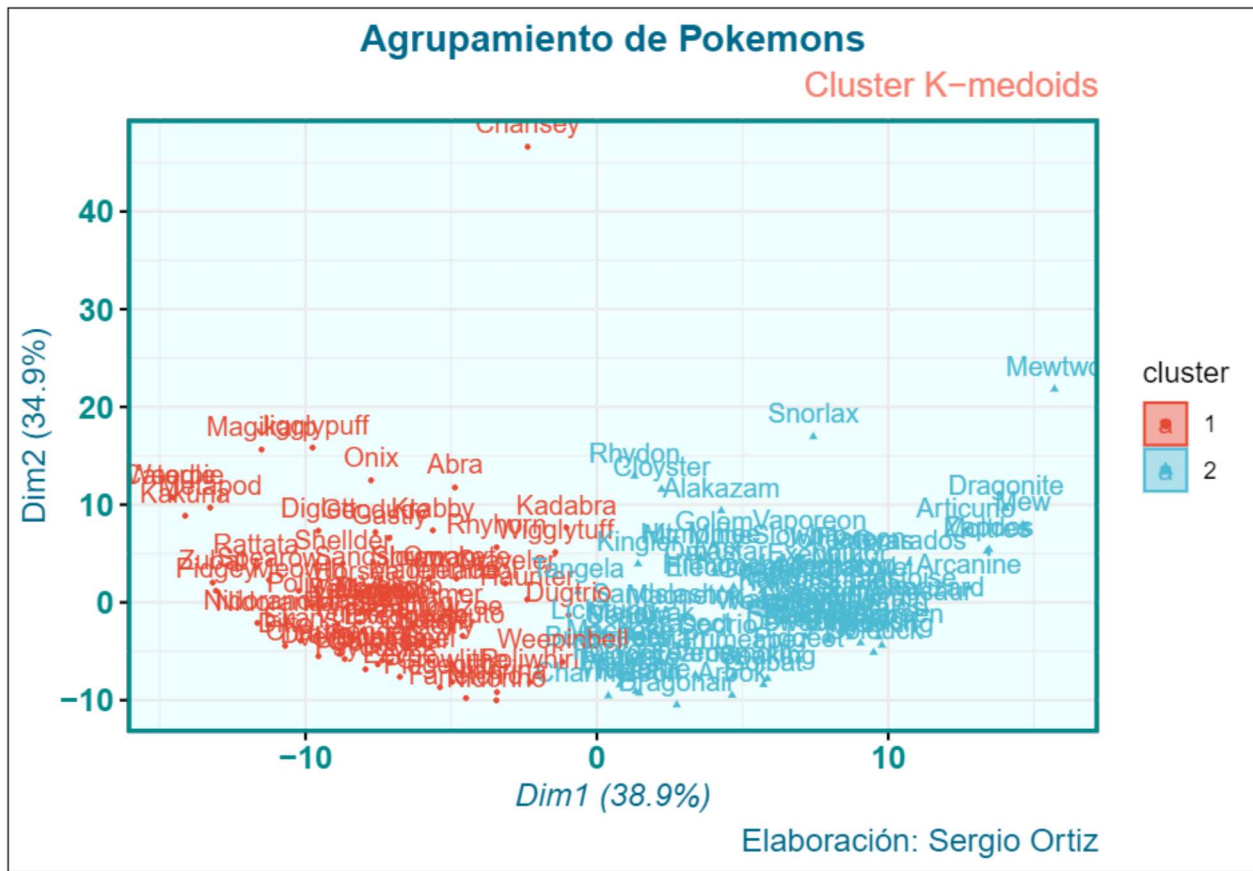
En este resultados me dicen que mi mejor numero de clouster seria 2 (este codigo es muy pesado y cuando intento hacer que el codigo corra en markdown R se cae incluso con rcloud no lo corre asi que pongo el resultado de ese codigo junto con una pequeña interpretasion)

Clusters K-medoids

```
set.seed(10)
#usando una semilla para que siempre den el mismo resultados
method_Cluster = eclust(Mdistancia,FUNcluster = "kmeans", k = 2, nstart = 25, graph = F)
```

Graficar clusters Kmeans

```
Kmedoids = fviz_cluster(method_Cluster, data = Mdistancia, main = "Agrupamiento de Pokemons",
  repel=F,star.plot=F,ellipse = T, ellipse.type="euclid" ,ellipse.level = 0.95,ellipse.alpha=.45,
  palette="npg",ggtheme = theme_minimal(),show.clust.cent=T,pointsize = .8,labels = 11,font.ticks
  font.x = c(12, "italic", "deepskyblue4"),font.y = c(12, "plain", "deepskyblue4"))+
theme(panel.background = element_rect(fill = "azure")) +
theme(plot.background = element_rect(fill = "white"))+
theme(panel.border = element_rect(colour = "darkcyan", fill=NA, size=1.5))+
  theme(plot.title = element_text(size= 14, vjust=.75, color="deepskyblue4", lineheight=1,face="bol
  theme(plot.caption = element_text(size = 12,color = "deepskyblue4",hjust=1))+
  theme(plot.subtitle= element_text(size = 13,color = "salmon",hjust=1,face="plain"))+
labs(subtitle = "Cluster K-medoids", caption = "Elaboración: Sergio Ortiz")
Kmedoids
```



Gracias a la grafica podemos notar como se comportan los datos y se dividen muy bien.

3.- Algoritmo k-medias (2 grupos) cantidad de subconjuntos aleatorios que se escogen para realizar los calculos de algoritmo.

```
Kmeans.3<-kmeans(DATOS, 2, nstart=25)
```

centroides

```
Kmeans.3$centers
```

```
##      HP  Attack Defense  Sp_Atk  Sp_Def  Speed
## 1 76.48276 85.54023 76.50575 81.35632 80.40230 80.13793
## 2 47.53125 54.89062 56.96875 47.81250 46.46875 53.70312
```

cluster de pertenencia

```
Kmeans.3$cluster
```

```
## Bulbasaur Ivysaur Venusaur Charmander Charmeleon Charizard Squirtle
##      2      1      1      2      1      1      2
## Wartortle Blastoise Caterpie Metapod Butterfree Weedle Kakuna
##      1      1      2      2      1      2      2
## Beedrill Pidgey Pidgeotto Pidgeot Rattata Raticate Spearow
##      1      2      2      1      2      1      2
## Fearow Ekans Arbok Pikachu Raichu Sandshrew Sandslash
```


##	1	2	1	2	1	2	1
##	Nidorana	Nidorina	Nidoqueen	Nidorano	Nidorino	Nidoking	Clefairy
##	2	2	1	2	2	1	2
##	Clefable	Vulpix	Ninetales	Jigglypuff	Wigglytuff	Zubat	Golbat
##	1	2	1	2	1	2	1
##	Oddish	Gloom	Vileplume	Paras	Parasect	Venonat	Venomoth
##	2	1	1	2	1	2	1
##	Diglett	Dugtrio	Meowth	Persian	Psyduck	Golduck	Mankey
##	2	1	2	1	2	1	2
##	Primeape	Growlithe	Arcanine	Poliwag	Poliwhirl	Poliwrath	Abra
##	1	2	1	2	2	1	2
##	Kadabra	Alakazam	Machop	Machoke	Machop	Bellsprout	Weepinbell
##	1	1	2	1	1	2	1
##	Victreebel	Tentacool	Tentacruel	Geodude	Graveler	Golem	Ponyta
##	1	2	1	2	2	1	1
##	Rapidash	Slowpoke	Slowbro	Magnemite	Magneton	Farfetch'd	Doduo
##	1	2	1	2	1	2	2
##	Dodrio	Seel	Dewgong	Grimer	Muk	Shellder	Cloyster
##	1	2	1	2	1	2	1
##	Gastly	Haunter	Gengar	Onix	Drowzee	Hypno	Krabby
##	2	1	1	2	2	1	2
##	Kingler	Voltorb	Electrode	Exeggcute	Exeggutor	Cubone	Marowak
##	1	2	1	2	1	2	1
##	Hitmonlee	Hitmonchan	Lickitung	Koffing	Weezing	Rhyhorn	Rhydon
##	1	1	2	2	1	2	1
##	Chansey	Tangela	Kangaskhan	Horsea	Seadra	Goldeen	Seaking
##	1	1	1	2	1	2	1
##	Saryu	Starmie	Mr. Mime	Scyther	Jynx	Electabuzz	Magmar
##	2	1	1	1	1	1	1
##	Pinsir	Tauros	Magikarp	Gyarados	Lapras	Ditto	Eevee
##	1	1	2	1	1	2	2
##	Vaporeon	Jolteon	Flareon	Porygon	Omanyte	Omastar	Kabuto
##	1	1	1	1	2	1	2
##	Kabutops	Aerodactyl	Snorlax	Articuno	Zapdos	Moltres	Dratini
##	1	1	1	1	1	1	2
##	Dragonair	Dragonite	Mewtwo	Mew			
##	1	1	1	1			

SCDG

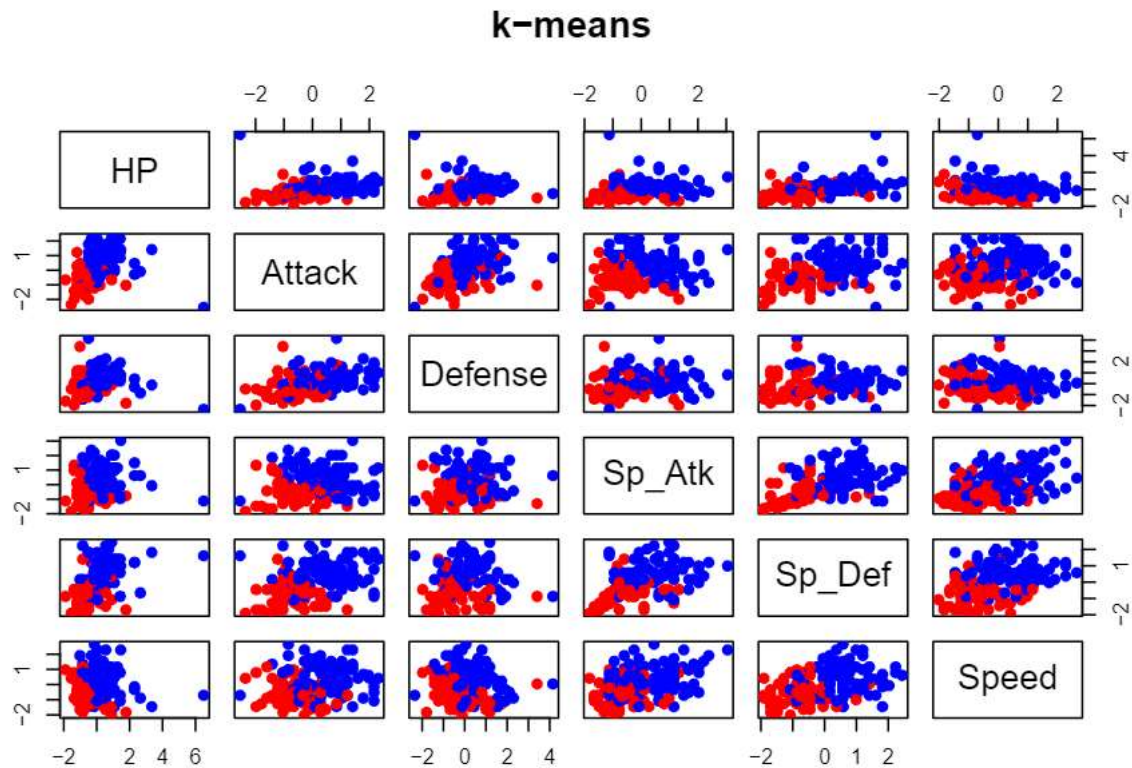
```
SCDG<-sum(Kmeans.3$withinss)
```

Clusters

```
cl.kmeans<-Kmeans.3$cluster
```

Scatter plot con la division de grupos obtenidos (se utiliza la matriz de datos centrados).

```
col.cluster<-c("blue", "red")[cl.kmeans]
pairs(X.s, col=col.cluster, main="k-means", pch=19)
```



Silhouette

Representacion grafica de la eficacia de clasificacion de una observacion dentro de un grupo.

Generacion de los calculos

```
dist.Euc<-dist(X.s, method = "euclidean")
Sil.kmeans<-silhouette(cl.kmeans, dist.Euc)
```

Generacion del grafico

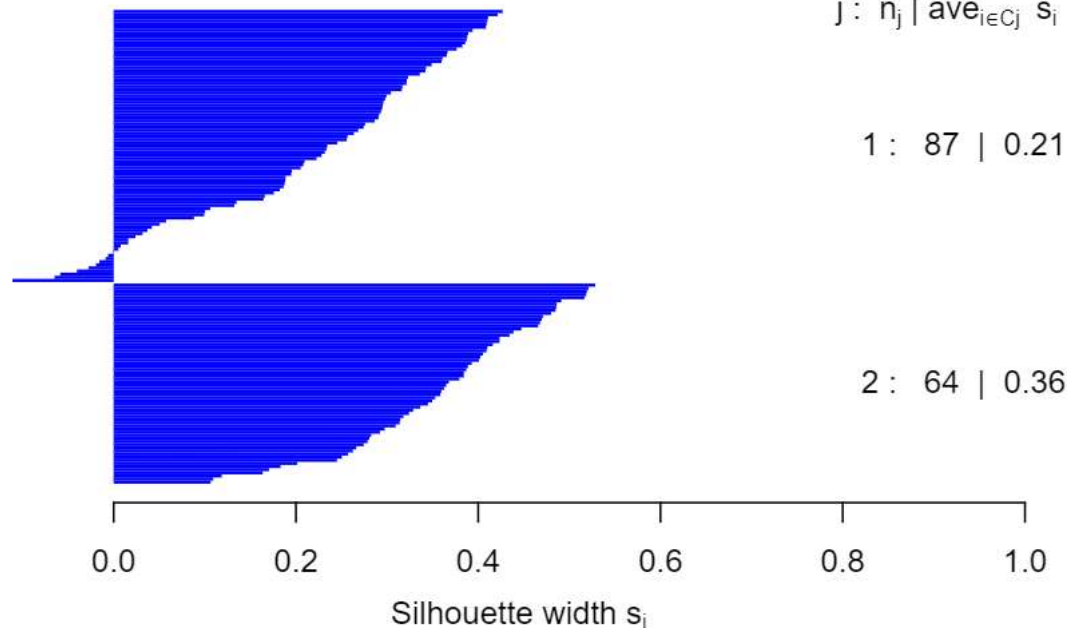
```
plot(Sil.kmeans, main="Silhouette for k-means",
col="blue")
```

Silhouette for k-means

n = 151

2 clusters C_j

$j : n_j \mid \text{ave}_{i \in C_j} s_i$



CONCLUSION

Con los resultados que obtuvimos podemos decir que la cantidad de clusters para una mejor k-medias para esta base de datos de pokedex es de 2 clusters y en Silhouette podemos ver que no es muy grande los porcentajes, pero el segundo cluster es mas alto de los dos.

REFERENCIAS

es, Q. (2016). ¿Qué es Pokémon? Descubre el fenómeno Pokémon en este artículo dedicado. Nintendo of Europe GmbH. <https://www.nintendo.es/Noticias/2016/agosto/-Que-es-Pokemon-Descubre-el-fenomeno-Pokemon-en-este-articulo-dedicado-1128960.html>

de, C. (2003, October 18). franquicia de medios japonesa. Wikipedia.org; Wikimedia Foundation, Inc. <https://es.wikipedia.org/wiki/Pok%C3%A9mon>

Barradas, A. (2016). Pokemon with stats. Kaggle.com. <https://www.kaggle.com/datasets/abcsds/pokemon>

RPubs - Introducción a los Modelos de Agrupamiento en R. (2018, June 23). Rpubs.com. [https://rpubs.com/rdelgado/399475#:~:text=modelos%20de%20agrupamiento.-,Agrupamiento%20por%20K%2DMedios%20\(K%2DMeans%20Clust](https://rpubs.com/rdelgado/399475#:~:text=modelos%20de%20agrupamiento.-,Agrupamiento%20por%20K%2DMedios%20(K%2DMeans%20Clust)

Juan Gabriel Gomila. (2018). 28 - La técnica de k-means en RStudio [YouTube Video]. In YouTube. https://www.youtube.com/watch?v=b-LtNvGXcLo&ab_channel=JuanGabrielGomila