

## FIFA 19 DATASET

# 1. Analisis de los datos

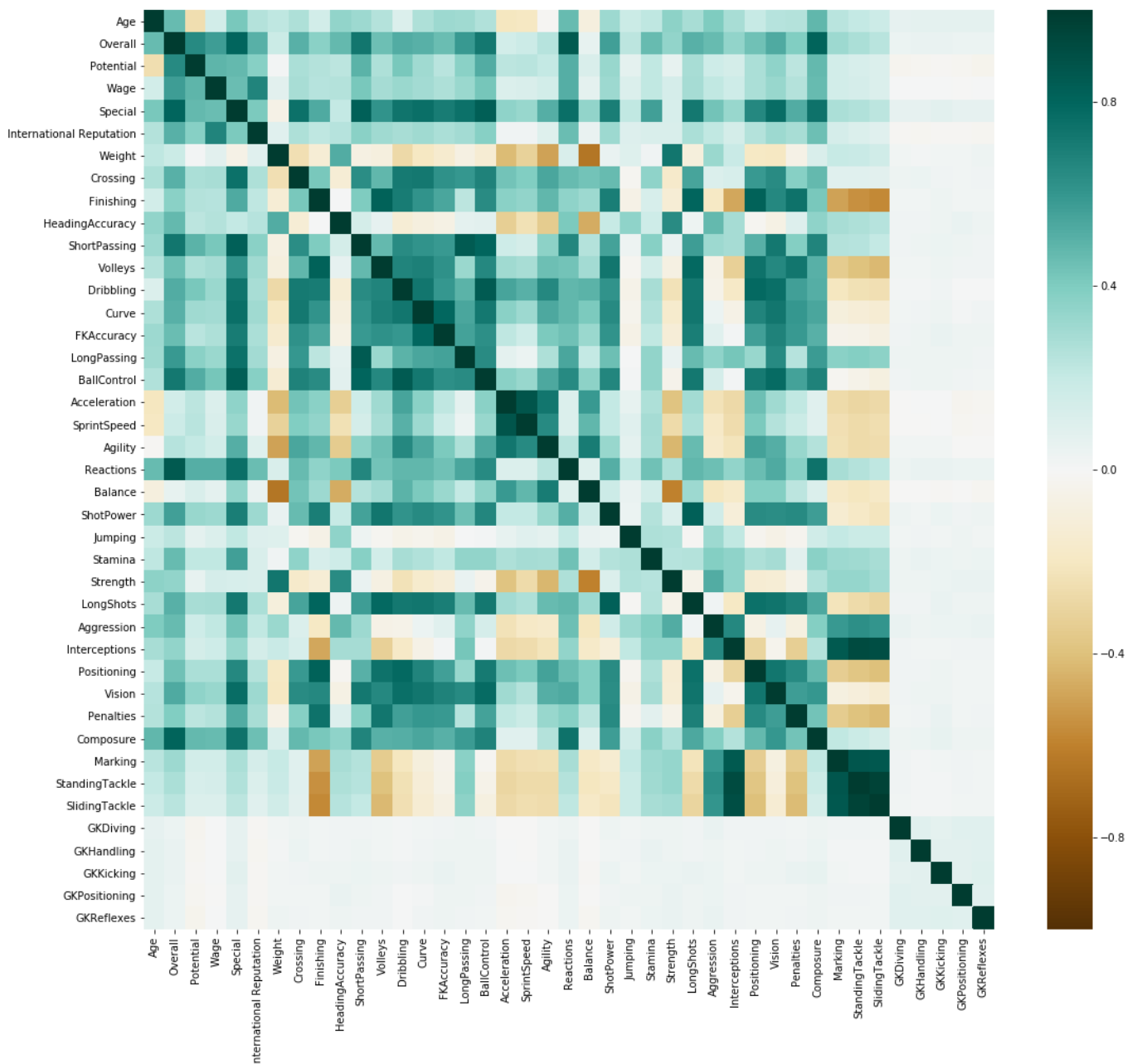
- Tenemos 2 dataset, uno con datos numericos y otro con categoricos.
  - El dataset numerico tiene informacion de las características (skills) de los jugadores.
  - El dataset categorico contiene informacion personal de los jugadores.

**Al analizar estos dataset y el objetivo del ejercicio, nos enteramos de que nos conviene tener unas variables del dataset de categorias con el numerico.**

- Por lo cual empezamos por leer los 2 dataset y crear un 3ero con la combianación de los 2 dataset.
- Una vez que tenemos nuestros dataset, comprobamos si nuestro dataset tiene missing values para tratarlos o no.
  - En este caso los dataset estan limpios de missing values por lo cual no hay tratamiento a hacer por este lado.

# 2. Tratamiento de los datos

- Creamos una matriz de correlation con la correlacion de pearson para poder hacer una selección de las variables continuas y entender mejor nuestros datos



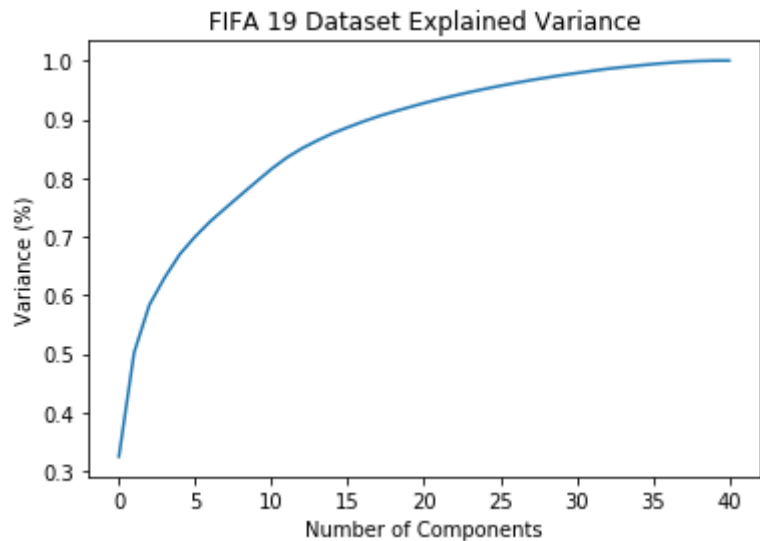
Aquí, empezamos a hacer conclusiones :

- Tenemos algunas correlaciones negativas y otras positivas,

Una de las conclusiones que podemos hacer en este momento es que muchas variables están correladas con otras según la posición del jugador y sus skills, por lo cual podríamos quedar con las variables que tienen el valor absoluto  $\Rightarrow$  a 80%

### 3. PCA

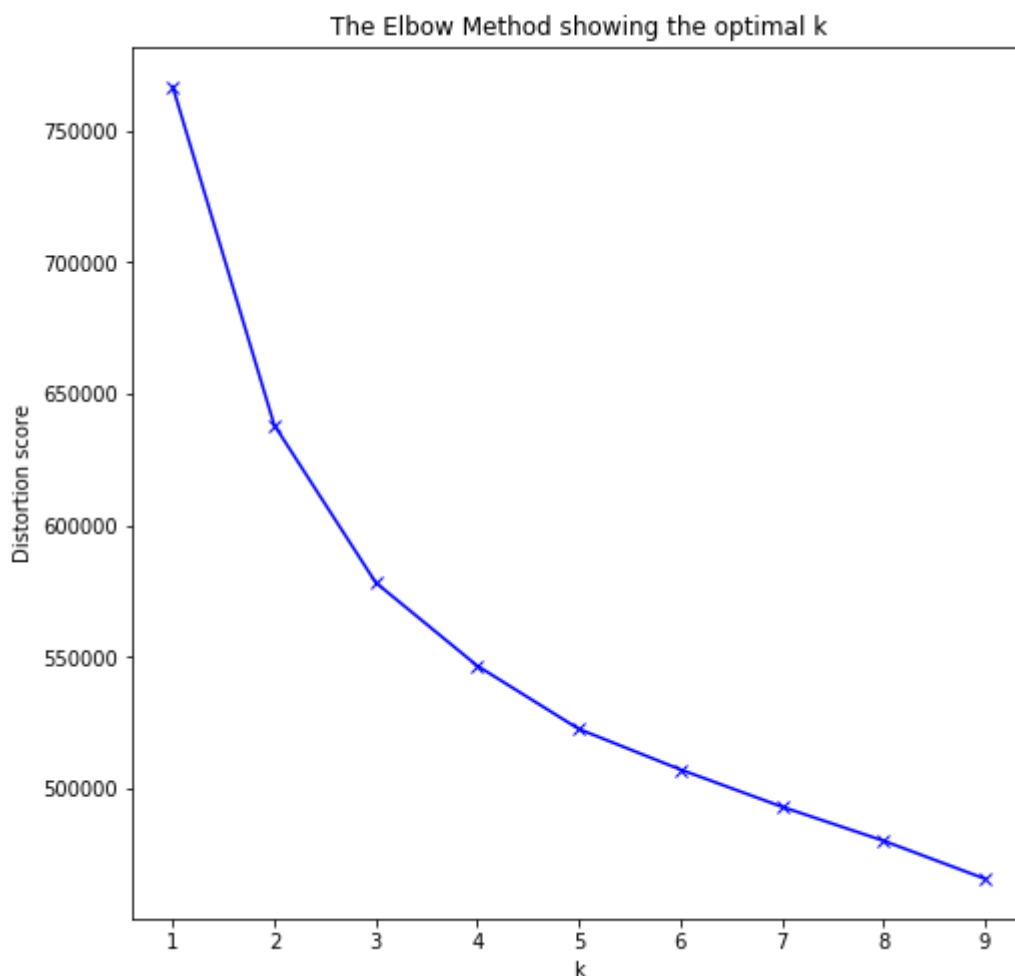
- Empezamos por normalizar nuestras variables continuas para que nuestras variables tengan la misma escala.



- Sacamos la varianza acumulada
- Podemos tener el 95% con 25 dimensiones por lo cual no vale la pena

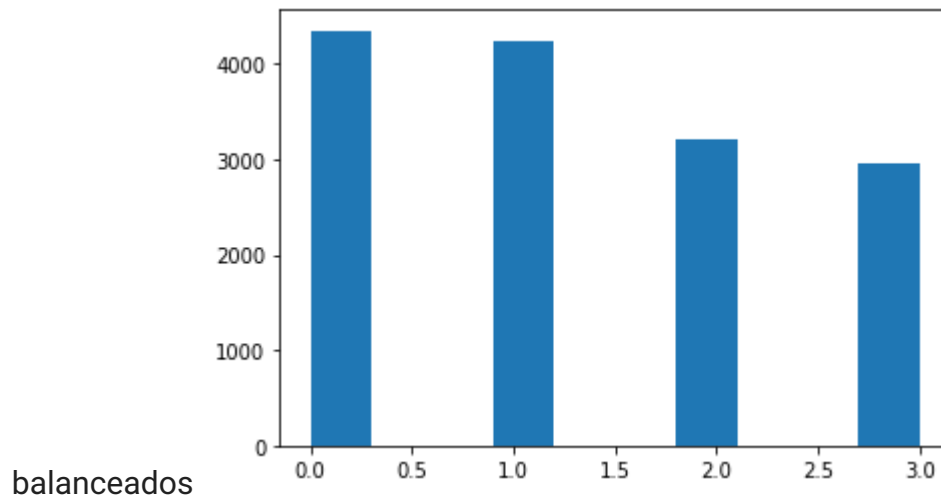
## 4. K-mean

- Empezamos por hacer el metodo del codo (The elbow method) para tener nuestro numero de clusters



- Lanzamos el algoritmo k-mean con numero de clusters = 4.
- Creamos una nueva columna del resultado del cluster (0, 1, 2, 3)

- Comprobamos la densidad de nuestros clusters y vemos que estan bastante bien



## ▼ 5. Primera Segmentación

- Creamos nuestros clusters en nuestro dataset (Portero, Defensa, Mediocampista, Delantero).
- Vamos a aplicar algoritmos de clasificación de arboles para comprobar nuestros clusters

### 5.1 Split y preparación de los datos

- Empezamos por separar nuestros datos entre datos de entrenamiento(80%) y datos de test(20%).

### 5.2 Cross validation

- Definimos un cross validation con 5 muestras para entrenar nuestros algoritmos mas tarde

### 5.3 RandomForestClassifier

- Entrenamos nuestro algoritmo y nos da un resultado de : 93%

### 5.4 XGBClassifier

- Entrenamos nuestro algoritmo y nos da un resultado de : 95%

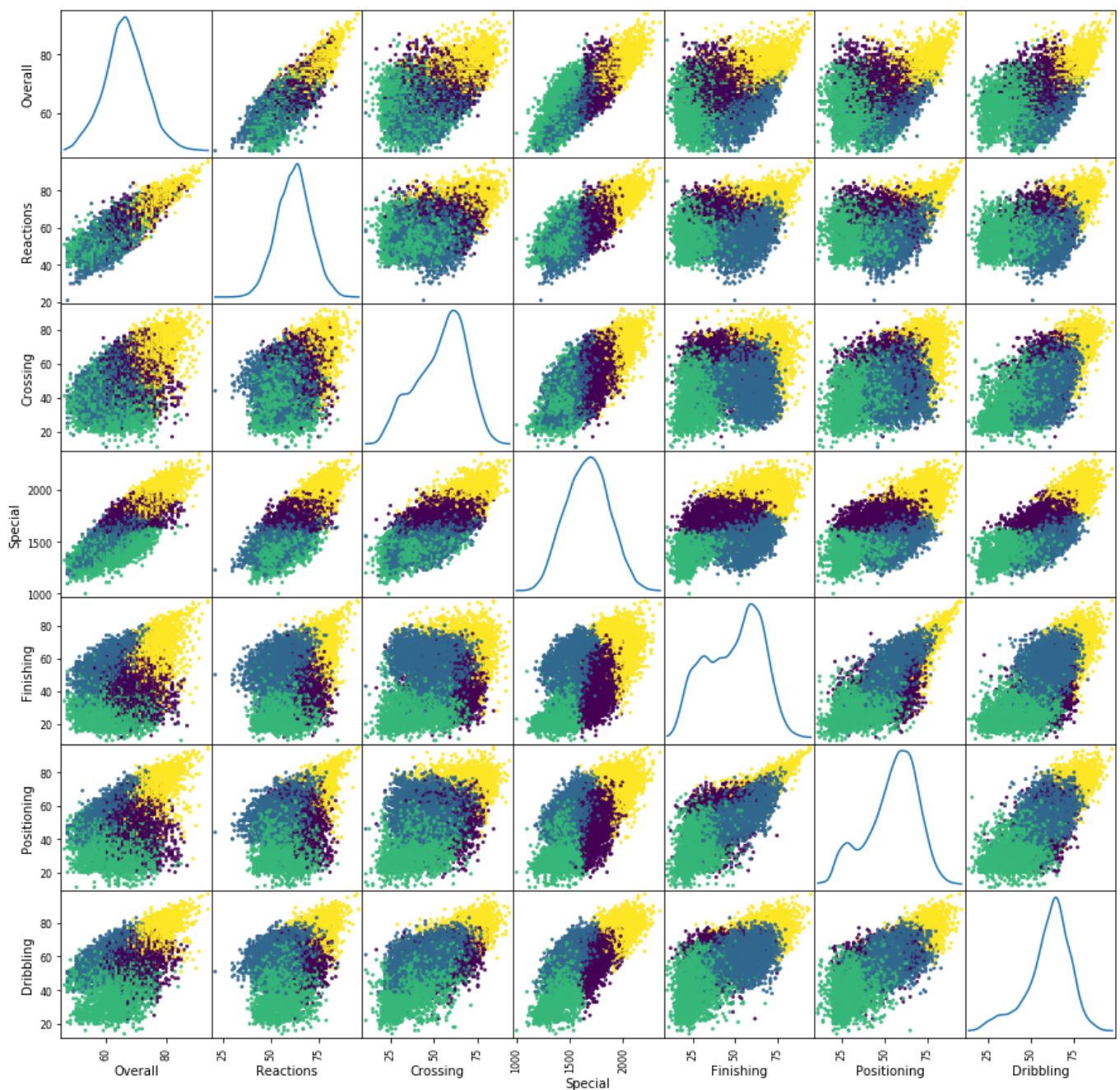
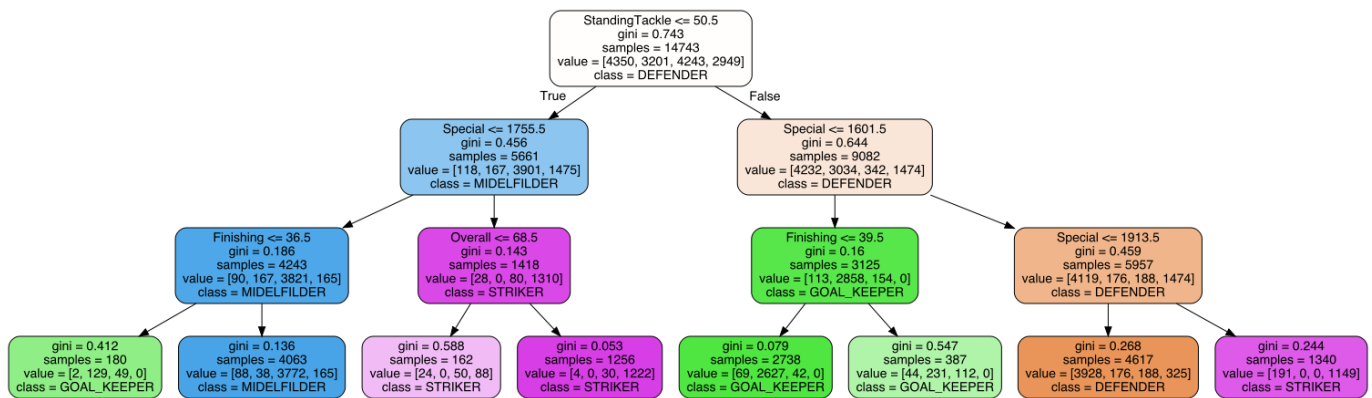
### 5.5 DECISION TREE

- Entrenamos nuestro algoritmo y nos da un resultado de : 89

## 6. Primera conclusión

- El algoritmo esta acertando al 89% por lo cual nuestra clusterización tiene sentido.

- Como el algoritmo consigue clasificar bien, vamos a ver cuales son las variables que mas importancia ha tenido en la clasificación



## 7. Segunda conclusión

- Nuestros clusters estan bien separados y existe una segmentación clara entre los 4 clusters

## 8. Selección de variables

- Basandonos en los resultados del arbol de decisión y la correlación de pearson, hacemos una selección de variables para hacer una reducción de dimensionalidad y volver a probar el k-mean con estas variables y dibujamos los resultados

## ▼ 9. Segunda Segmentación

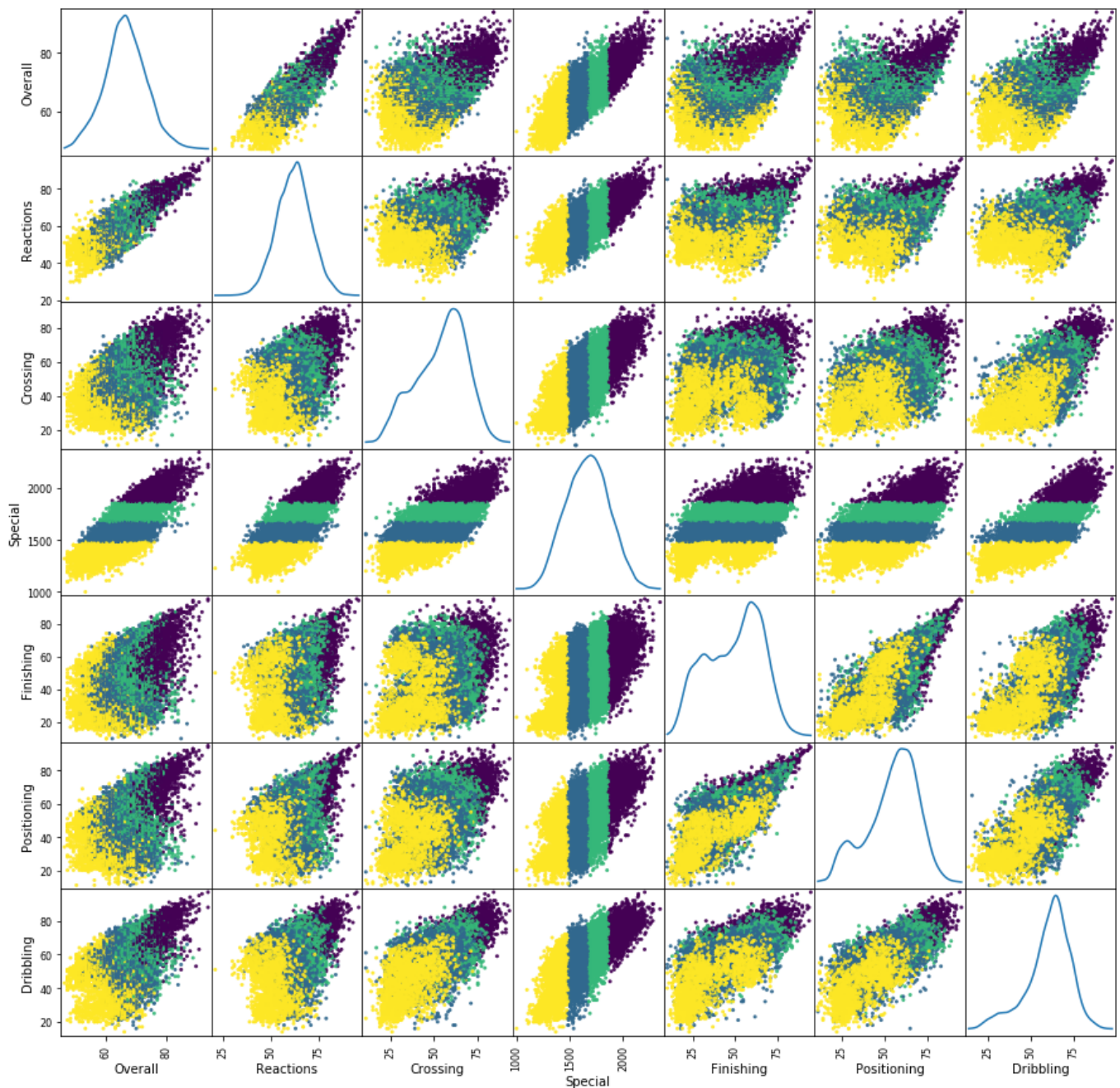
### 9.1 K-mean

- Clusterizamos otra vez a 4 **cogiendo solo las features seleccionadas(a partir de la clasificación y la correlación de pearson)**
- Ponemos la columna de los clusters en nuestro dataset

### 9.2 Decision Tree

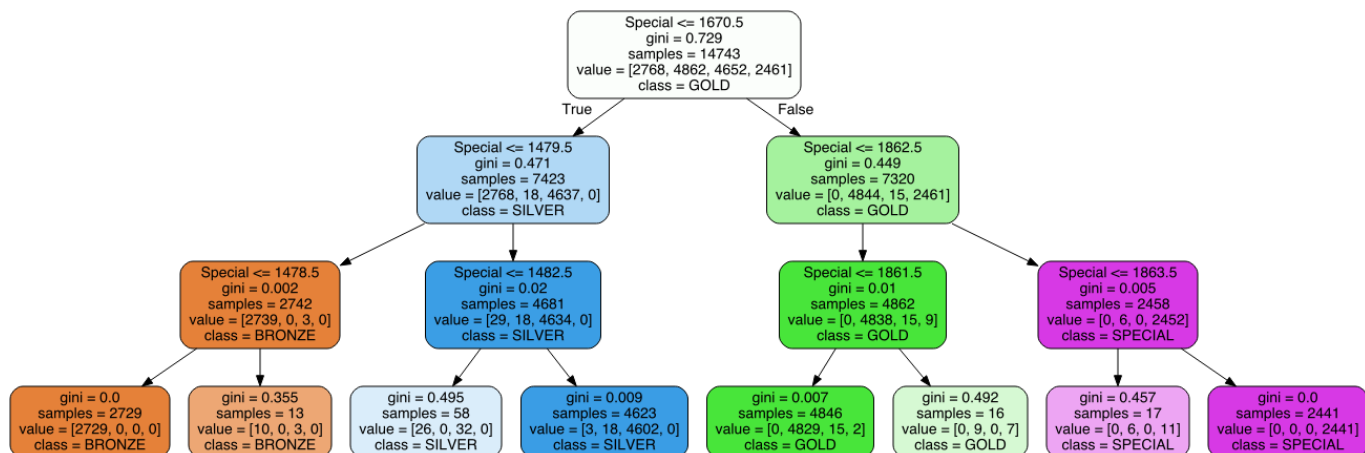
- Dibujamos nuestros clusters :





## Conclusión

- De aquí podemos observar de que nuestros datos se clasifican bien en 4 clusters.
- Si analizamos las variables Special y Crossing podemos observar que crossing va de 25 a 75 y Segun estos resultados, los clusters pueden ser: el tipo de carta que tiene el jugador (Special, Gold, Silver, Bronce)
- Comprobamos el algoritmo de decision tree para confirmar nuestra clasificación y nos da un resultado de : 99%



Double-cliquez (ou appuyez sur Entrée) pour modifier