

¿Cómo formar al mejor equipo?

Introducción

El fútbol une al mundo y a todas las personas, millones de aficionados ven día a día los partidos del equipo del que son hinchas, o inclusive sin serlos, este deporte logra apasionar a muchas personas. El fútbol engancha, no distingue de sexo ni de edad, y en muchos casos nos da hasta una identidad, pero, sobre todo, el fútbol es sencillo. El mundial, la Eurocopa, la UEFA Champions League y entre muchos más torneos, se presentan como fenómenos sociales y no sólo deportivos con los que se quiere buscar un optimismo desaparecido, una diversión y alegría que impulse a la gente. El fútbol se ha convertido en mucho más que un deporte.

La importancia de nuestro análisis está enfocado a la optimización de un sistema de reclutamiento, el tener una herramienta que apoye a los equipos a identificar que jugador es el que necesitan de acuerdo a sus condiciones sería algo que revolucionara la forma en la que se contratan jugadores, independientemente si son promesas o ya están desarrollados.



Figura 1: En la imagen se muestran algunas de las ligas más importantes a nivel mundial de fútbol soccer. En el presente trabajo se trabajó con datos correspondientes a jugadores pertenecientes a dichas ligas con el fin de evaluar los mejores resultados y usarlos como referencia para la creación de un criterio de selección para nuevos jugadores.

Software y Recursos

Notebook completo:

Base de datos:

Algunas de las librerías utilizadas fueron pandas, numphy, sklearn y matplotlib.



Objetivo

Conocer el perfil del mejor jugador de las diferentes ligas europeas en cuanto a números y estadísticas se refiere de los últimos años para poder realizar la mejor decisión posible al momento de hacer una contratación.

Objetivos secundarios.

- Conocer un promedio de edad de los jugadores que militan en Europa.
- Conocer cuál/cuáles son las mejores ligas del mundo en base a la competitividad de la misma.

Metodología

El primer paso que llevamos acabo fue una limpieza de nuestra base de datos inicial, eliminamos alrededor de 18 columnas de manera inicial, además de crear bases de datos únicos con registros por equipos y por jugadores.

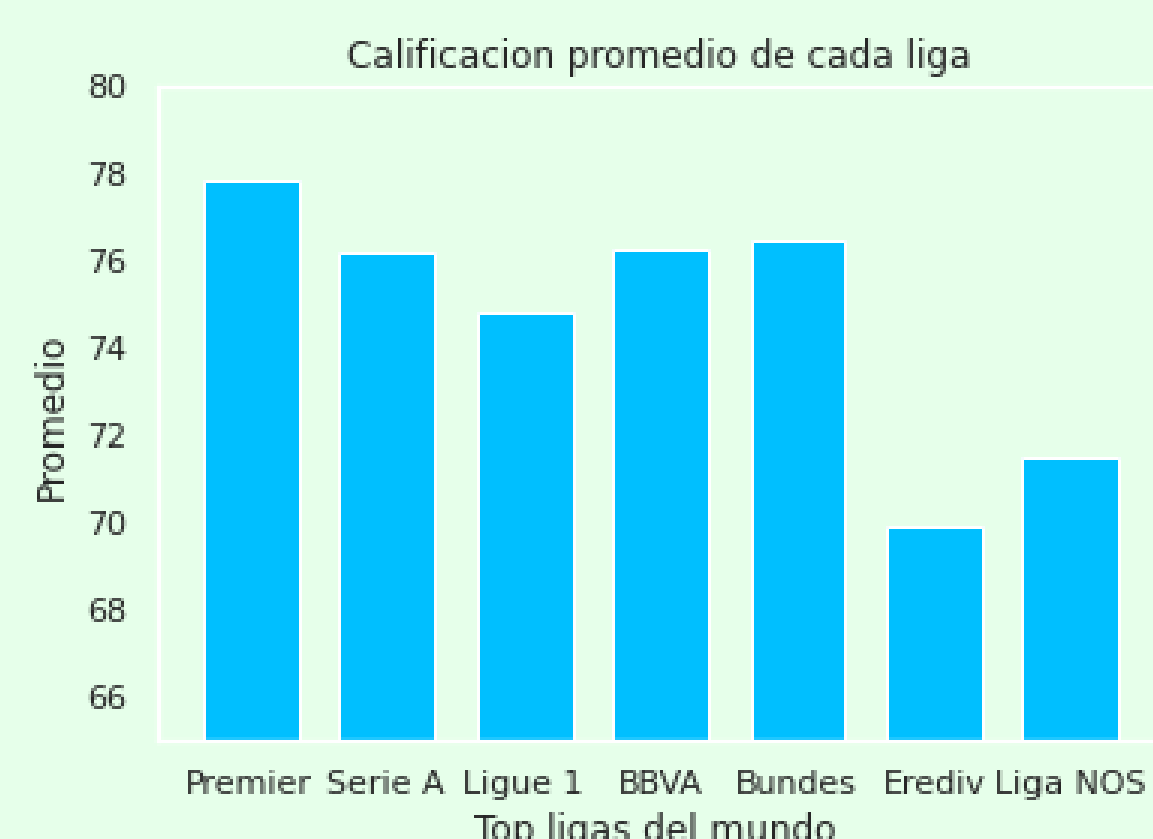


Figura 2: Calificación promedio de la liga en base a los jugadores que juegan en ella.

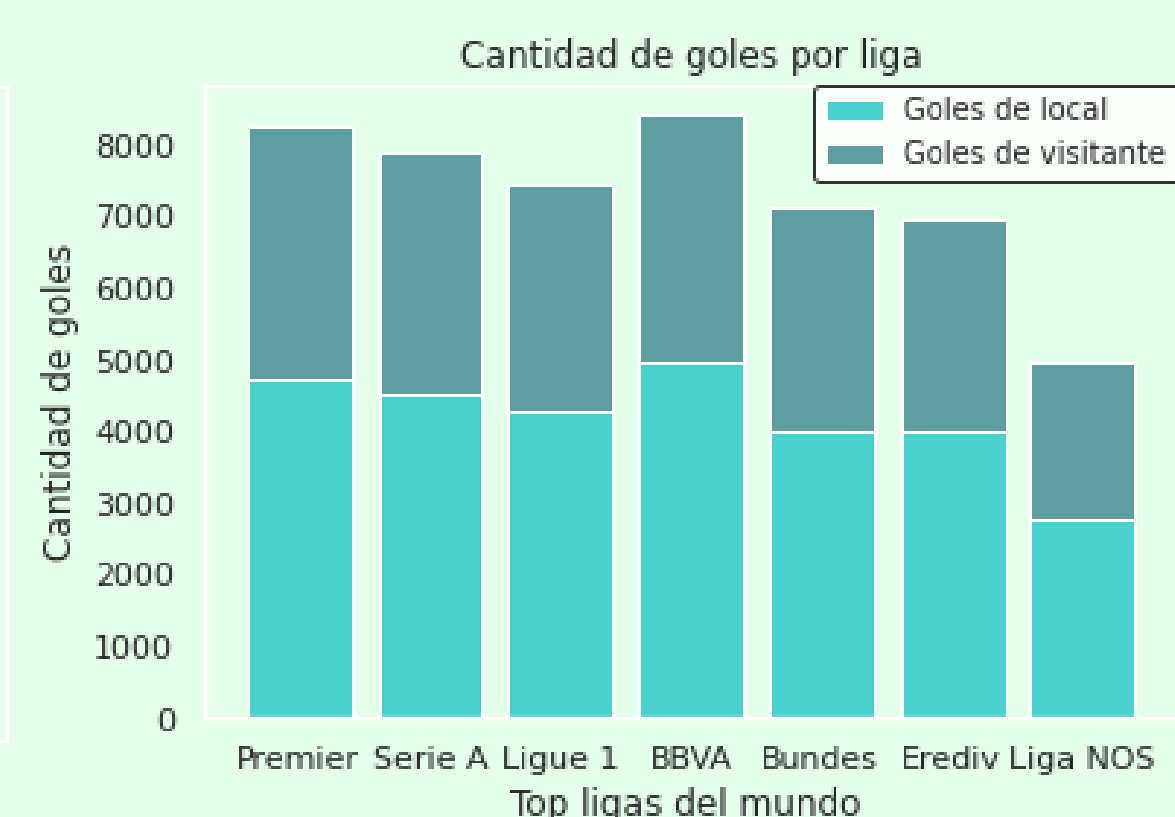


Figura 3: Cantidad de goles de local y visitante por liga.

Las ligas que más destacan son la liga inglesa (Premier League) y la española (La Liga BBVA)

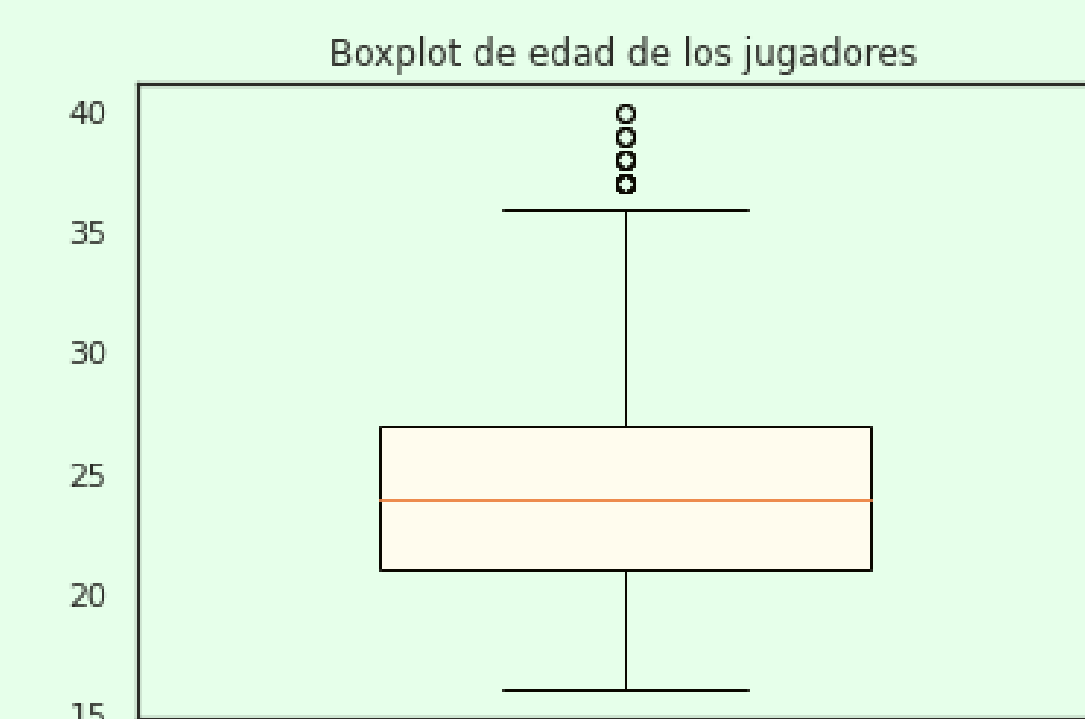


Figura 4: Boxplot de la edad de los jugadores. Notemos que la media de la edad es de 24 años.



Resultados

La técnica de Kmedias agrupara en 4 posiciones a los jugadores, de acuerdo a las características que tenemos de los mismos, ya que si bien buscamos identificar a los mejores jugadores del mundo, para así analizar sus cualidades, en la base de datos que hemos trabajado no se nos agrupan por posiciones, cosa que en la vida real no funciona igual, ya que las "vacantes" surgen para una posición en específico, aunado a esto, probablemente los jugadores con mayor media serán los delanteros, cosa que sesgaría nuestro análisis.

Al aplicar la técnica agruparemos a los jugadores en 4 posiciones:

1-Porteros 2-Defensas 3-Centrocampistas 4-Delanteros

El proceso parte de la base de datos que contiene únicamente las características de los jugadores, se aplicó un filtro previo para eliminar aquellos jugadores con una calificación baja, para después sustituir los valores NA por cero, normalizar la base, y por último, aplicar el algoritmo de K-Medias, con un número de grupos de 4.

Una vez clasificados los jugadores, bastó con obtener los elementos descriptivos de su análisis estadístico para generar estimadores, los cuales servirán como un criterio para la selección de candidatos para formar un equipo de soccer que de resultados favorables. Finalmente, se usó el índice de Silhouette para evaluar la metodología de K-medias, obtuvimos un valor 0.49, lo que nos indica que el agrupamiento fue de una calidad regular, esto debido a que se cuentan con una gran cantidad de características y que en algún punto llegan a ser similares entre diferentes posiciones, lo que hace difícil para el algoritmo agruparlos perfectamente, este es uno de las áreas de mejora para el futuro.

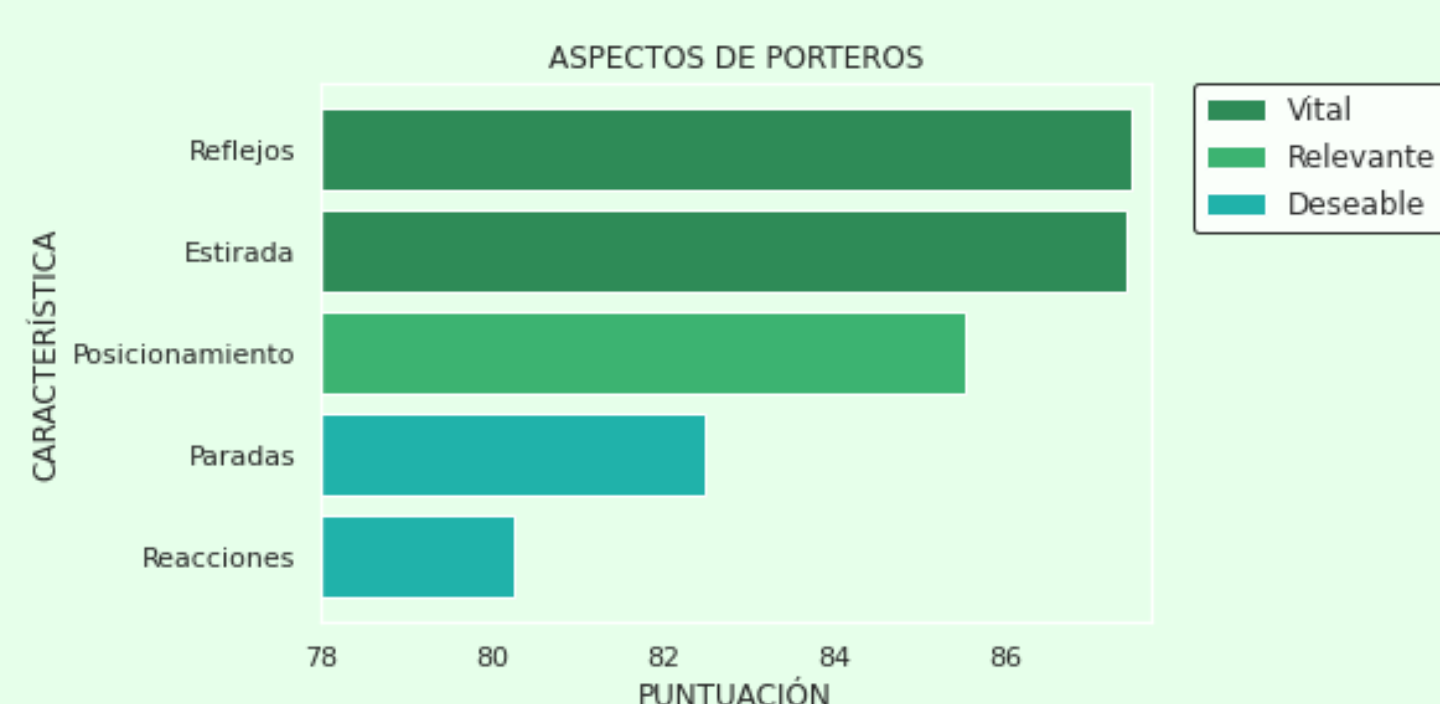


Figura 5.1: Aspectos a evaluar de un portero

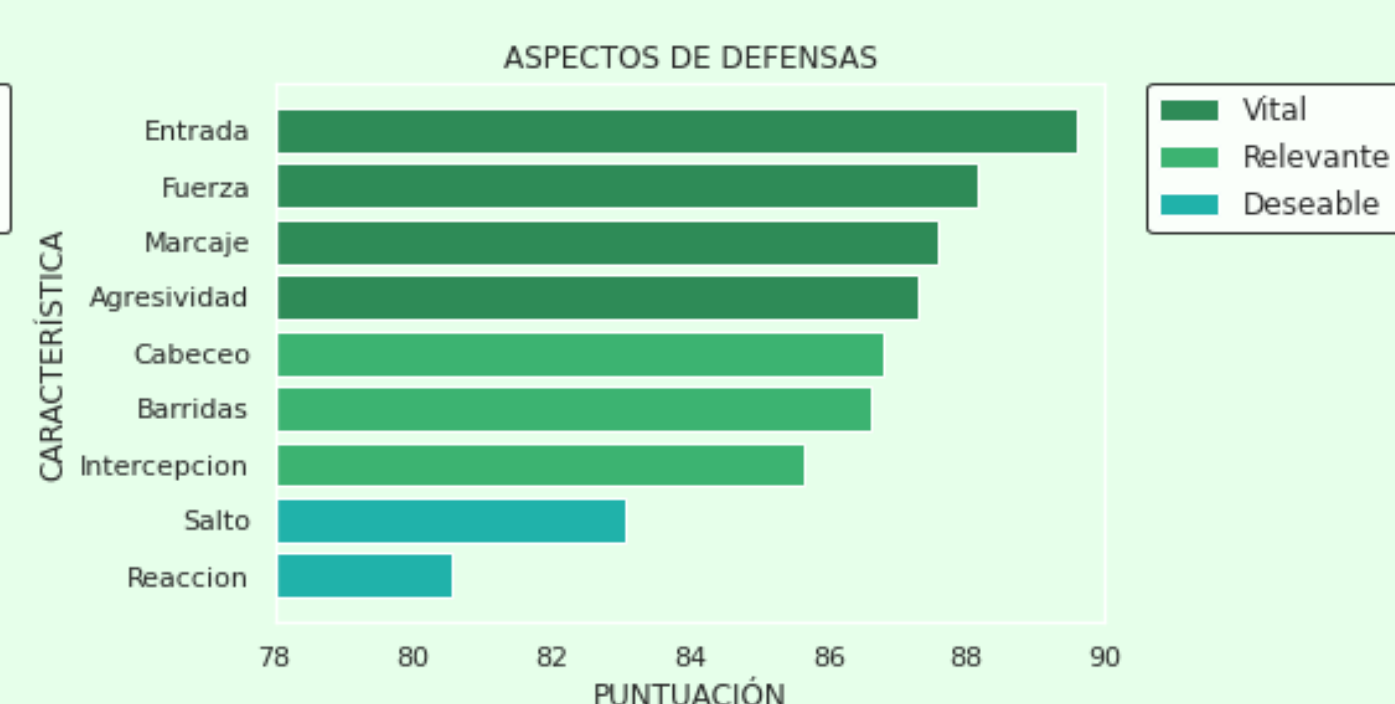


Figura 5.2: Aspectos a evaluar de un defensa

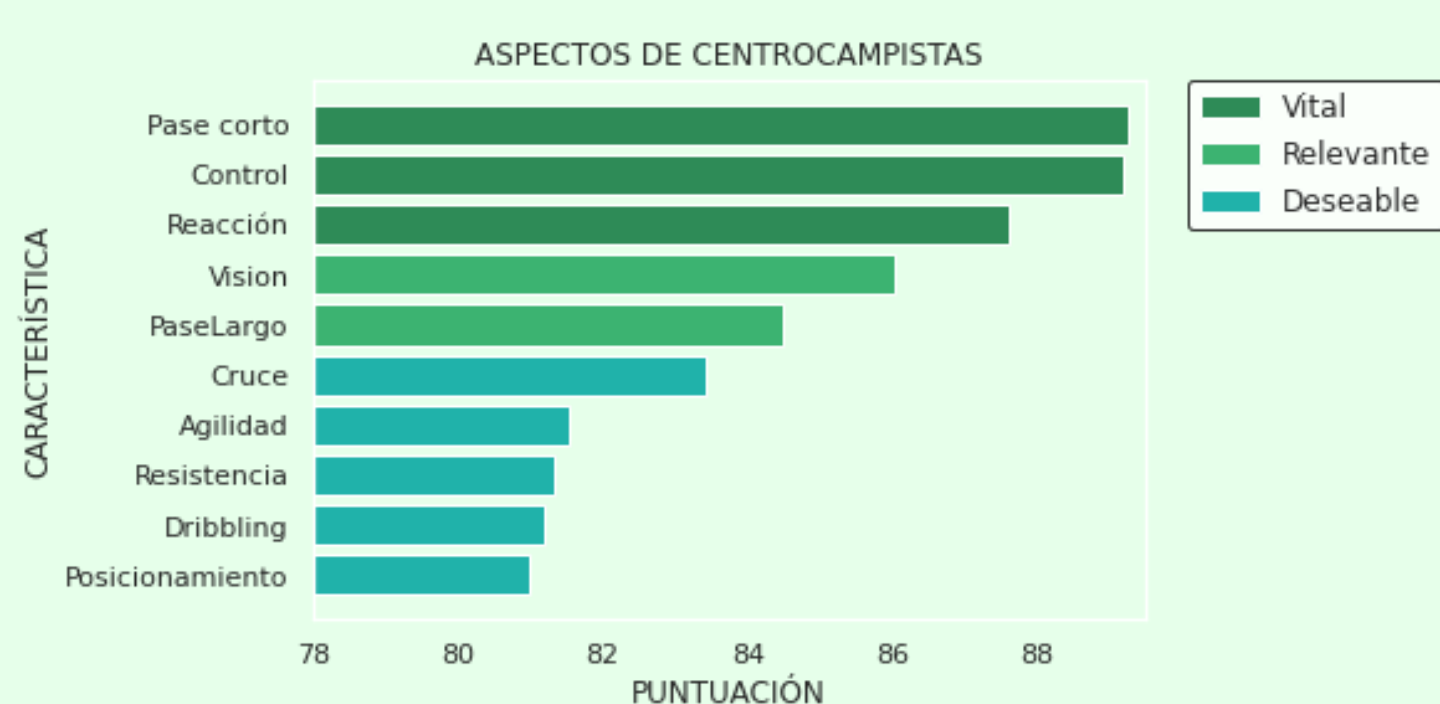


Figura 5.3: Aspectos a evaluar de un centrocampista

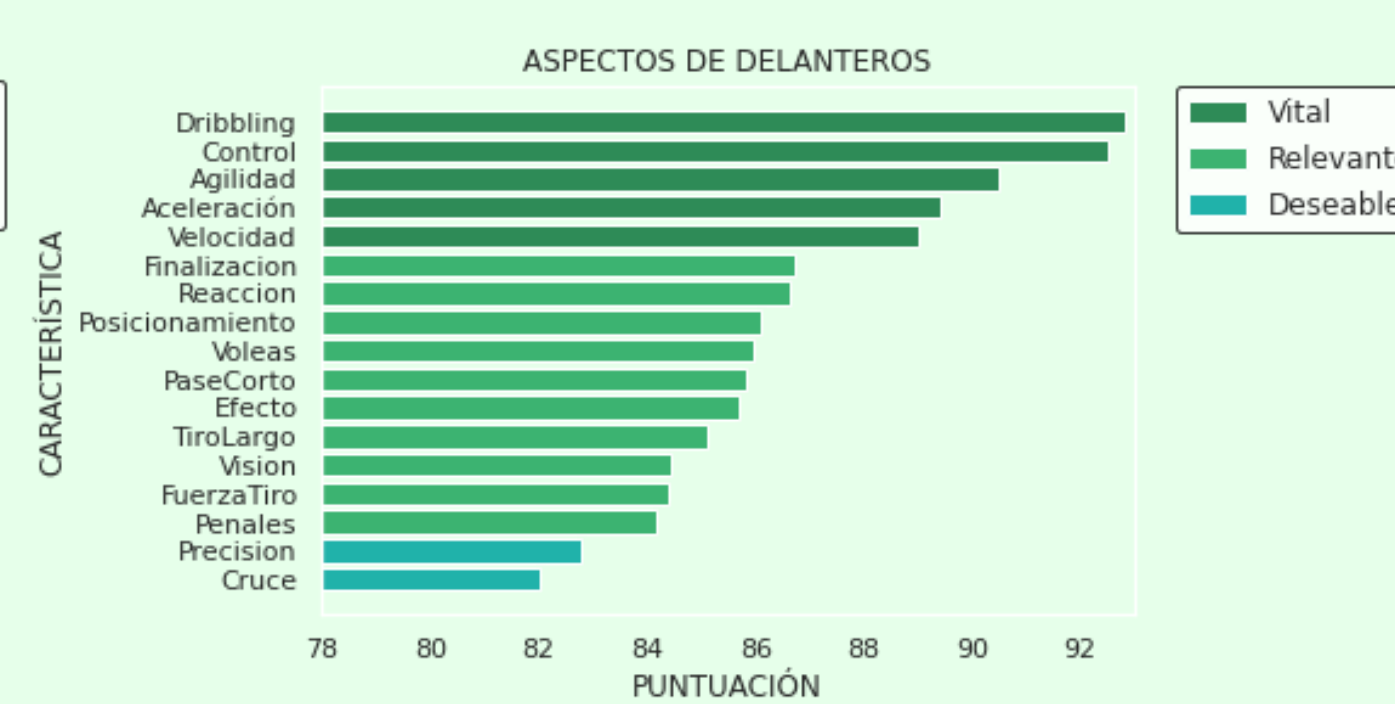


Figura 5.4: Aspectos a evaluar de un delantero

Figura 5: Gráficas de barras con los aspectos a evaluar de acuerdo con la posición del jugador. En ellas vemos los resultados de nuestro análisis.

Conclusión:

Durante la investigación nos dimos cuenta de que encontrar "al mejor jugador" es algo más relativo que numérico, esto debido a la variedad de posiciones que existen en el deporte que estamos analizando, no podemos comparar a un portero con un delantero debido a que su trabajo es el opuesto al otro, gracias a la técnica empleada de K-medias pudimos lograr una mejor aproximación al objetivo principal haciendo las divisiones entre las distintas posiciones y filtrando a los mejores jugadores por posición. Gracias a las figuras 2,3 y 4 pudimos cumplir nuestros objetivos secundarios

Trabajo a futuro:

- Crear nuestra propia base de datos, es decir, con características más específicas, que nos faciliten el manejo de los datos, y con ello poder buscar posiciones más específicas.
- Aumentar el valor de la métrica de evaluación para poder tener una mayor seguridad en nuestros resultados.
- Cuando se tenga lo anterior mencionado, es cuando se podría implementar en los equipos mediante una red de reclutamiento, ya sea para buscar nuevo talento o jugadores ya desarrollados.