**Memoria EDA Alejandro Perdiguero**

Este documento es la memoria de un trabajo de investigación llevado a cabo por Alejandro Perdiguero como alumno de la escuela TheBridge.

El siguiente análisis va a tratar a La Liga española de fútbol desde la temporada 2012-13 a la 2019-20 llevando a cabo un análisis tanto general como por equipo. Partimos de un dataset que nos proporciona las fechas, equipos, resultados etc de todo el campeonato liguero. Contamos con un dataset para cada temporada.

Las fuentes de datos o ‘datasets’ han sido obtenidos de un repositorio de github que contaba con cientos de datasets de las principales ligas de fútbol de todo el mundo. Además contaba con un gran feedback y buena valoración.

Este dataset esta sobretodo centrado en mostrar todos los partidos que se jugaron, sus resultados, así como sus fechas. En este caso yo voy a limpiar el dataset y modificarlo de forma que pueda tratar los datos que yo necesito de la manera más sencilla posible.

Estos datos que yo voy a necesitar para mi análisis son los goles totales por partido separándolos a su vez por ‘GolesLocal’ y ‘GolesVisitante’ lo que me va a permitir llevar a cabo mi análisis sobre la relevancia de los goles fuera de casa y el impacto que pueden tener en la consecución de un campeonato.

Por tanto lo primero que se hace en el archivo ‘main.py’ es importar las librerías de pyhton que vamos a utilizar en nuestro análisis exploratorio.

Una vez estas librerías han sido importadas pasaremos a importar todos los datos que vamos a utilizar. En mi caso son 8 datasets todos en un formato idéntico. Una vez añadidos los datasets el siguiente paso es modificarlos creando una nueva columna partiendo de la columna ‘FT’ que va a ser ‘GolesLocal’ y ‘GolesVisitante’ lo cual lo conseguimos simplemente separando la columna ‘FT’ por su separador ’-‘. Una vez conseguimos estas dos columnas, sumándolas podemos conseguir la columna ‘GolesTotales’ que refleja el número de goles por partido.

Una vez tenemos listos estos datasets vamos a preparar los datasets por cada equipo por cada año.

Para esto agruparemos usando una máscara y la función.loc para extraer del dataset inicial los datos que necesitamos y guardarlos en un nuevo dataset con el nombre de ese equipo. Haremos este proceso 2 veces por equipo, para sus goles como local y como visitante juntandólos finalmente en un dataframe total con las estadísticas locales y visitantes de dicho equipo usando la función .merge.

Esto debemos hacerlo por cada equipo al que queramos analizar en cada uno de los distintos datasets por temporada con los que contamos. He usado siempre 2 equipos Madrid y Barcelona salvando una temporada en la que añado al Atlético de Madrid también.

Una vez que tenemos el dataset para dicho equipo con tanto sus estadísticas locales como visitantes podemos empezar con las visualizaciones.

En primer lugar analizamos con un gráfico de barras el total de goles anotados por temporada a lo largo de los años, dejando una clara marca del impacto que tuvieron en nuestra liga la MSN y BBC los dos tridentes ofensivos de los principales equipos de nuestro campeonato.

Después con los gráficos de violín podemos ver separados en local y visitante la distribución de goles por partido de todos los equipos del campeonato a lo largo de este periodo que estamos analizando.

Es una constante ver como Real Madrid y Barcelona sobre salen por encima del resto todos los años tanto como goleadores locales como visitantes

Lo siguiente que podemos encontrar en el análisis es una comparativa entre estos dos principales equipos (Madrid y Fc Barcelona) desde 2012 hasta 2020 separada dicha comparativa en rendimiento como local y rendimiento como visitante.

A continuación pasamos a comparar las temporadas con mejores registros de los dos principales equipos vs sus temporadas con peores registros para comparar cada uno de estos equipos en su mejor y peor momento.

Finalmente llevamos a cabo una comparación entre la mejor temporada RMCF y la mejor temporada FC Barcelona tanto como local como visitante para comparar ambos equipos.

Esta parte de comparaciones puede ampliarse más y se ha hecho con comparativas también con el Atlético de Madrid campeón de liga por ejemplo.

**Resumen:**

* **Cosas positivas:**
  + El tener una temática libre me ha dado mucha facilidad a la hora de elegir que variables comparar o como entender los datos
  + Como amante del fútbol ha sido muy interesante ir un paso más allá para entender la liga española
  + Crear mis propias comparaciones
* **Cosas negativas:**
  + Primera vez tomando este reto
  + A mayor conocimiento más conclusiones habría podido sacar así como hacer más fáciles y eficaces muchas de las cuestiones de código
  + He sentido cierta limitación algunas veces al intentar representar mezclas de variables pero a la vez aprendido poco a poco a acotar las variables y entender como representarlas de manera eficaz