PROYECTO CHAMPIONS

Logotipo

Descripción generada automáticamente

*GERMÁN LLORENTE MUÑOZ*

1. OBJETIVOS
2. DESARROLLO

# Búsqueda de datos

Los datos se han sacado de la web <https://fbref.com/en/comps/8/Champions-League-Stats>. Esta página web contiene tablas con datos de los equipos participantes en las últimas temporadas de la competición europea. Pudiendo elegir entre varias temporadas distintas, se ha considerado que una muestra de 10 tablas distintas (desde la temporada 2014 hasta la actual), sería más que suficiente.

## Descripción de los datos

Cada uno de los dataset contiene información sobre equipos de fútbol, basado en las primeras líneas del archivo. A continuación se ofrece una breve descripción de la disposición general de cada una de las columnas:

**Rk**: Ranking del equipo.

**Squad**: Nombre del equipo.

**MP**: Partidos jugados (Matches Played).

**W**: Victorias (Wins).

**D**: Empates (Draws).

**L**: Derrotas (Losses).

**GF**: Goles a favor (Goals For).

**GA**: Goles en contra (Goals Against).

**GD**: Diferencia de goles (Goal Difference).

**Pts**: Puntos acumulados.

**Attendance**: Asistencia promedio a los partidos.

**Top Team Scorer**: Máximo goleador del equipo.

**Goalkeeper**: Portero principal del equipo.

**Notes**: Notas adicionales sobre el rendimiento del equipo o logros específicos.

***Ranking UEFA***

Sin embargo, no solo me he apoyado en fbref para la búsqueda de información, sino que quería hacer uso de una métrica objetiva para medir el valor real de un equipo. ¿Acaso existe tal medición? Por supuesto, el ranking UEFA.

El ranking de la UEFA, conocido oficialmente como los coeficientes de clubes de la UEFA, es un sistema de clasificación utilizado por la Unión de Asociaciones Europeas de Fútbol (UEFA) para clasificar y sembrar equipos en las competiciones de clubes europeos, como la Liga de Campeones de la UEFA y la Liga Europa de la UEFA. Este sistema de ranking se basa en los resultados de los clubes en las competiciones europeas durante un período determinado, generalmente los últimos cinco años.

Para ello, se utiliza una fórmula que tiene en cuenta factores como puntos por resultado, distribución de puntos (no todos los equipos juegan la misma cantidad de partidos), el coeficiente de la liga o bonus especiales por llegar a una determinada fase en una competición UEFA.

# Web scrapping

El web scraping es una técnica utilizada para extraer información de páginas web de manera automática. Esta técnica funciona enviando una solicitud HTTP al servidor de la página, generalmente obteniendo su código HTML. A partir de ahí usamos librerías como BeautifulSoup, que permiten buscar elementos en estructuras HTML. Con la ayuda de selectores específicos como etiquetas HTML, clases o identificadores, se extraen los datos necesarios del contenido analizado. Finalmente, se pasan la información obtenida a una base de datos, csv, Excel, o una estructura de almacenaje de datos similar.

Usando la web <https://kassiesa.net/uefa/data/method5/trank10-2024.html>, se han guardado los resultados de la tabla que se muestra a continuación. Para esta página web, aplicar la técnica que hemos estudiado en criptografía ha sido muy sencillo, dado que los elementos del HTML eran los propios de una tabla. A la hora de scrappear esto ayuda porque al pedir los elementos como “td”, “th” o “tr” nos devuelve las filas de la tabla.

Tabla

Descripción generada automáticamenteCabe destacar que el ser una web relativamente poco desarrollada generalmente mucho la búsqueda de datos de cara al web scrapping.

1. LIMPIEZA DE DATOS

Una vez se han recopilado los datos deseados, se debe realizar la labor de manipulación y limpieza de datos para poder trabajar con ellos como es debido. En nuestro caso, contamos con 11 datasets con datos de distintas temporadas del deporte rey y uno del ranking UEFA.

Las labores que se han realizado son las siguientes:

1. **Eliminación de las columnas inservibles para el análisis.**

Columnas tales como Máximo Goleador del Equipo, o Nombre del Portero

1. **Traducción de los datasets que estaban en español, a inglés.**
2. **Igualar los nombres de los equipos en todos los datasets.**

Esta labor se tuvo que realizar dos veces. Por un lado, debido a la distinción de inglés/español de algunos de los datasets de champions que queríamos utilizar. Además, más tarde también se tuvo que unificar los nombres con la tabla del ranking UEFA, donde había aproximadamente 90 cambios a realizar. Para ello se programó una función que buscaba subexpresiones dentro de cadenas de caracteres, facilitando el trabajo.

1. **Limpieza del dataset de UEFA para que solo se incluyan los equipos más relevantes**
2. **Corrección de los puntos totales de los equipos en los datasets de campeonatos basados en los datos de ranking de la UEFA.**
3. **Transformación de rangos y división de datos por partido para normalizar o ajustar ciertos valores.**
4. **Uso de Matplotlib para la visualización de algunos conjuntos de datos de interés**