

## Práctica de Open Data y Visualización Dinámica

La práctica que se ha desarrollado trata de recopilar estadísticas de la liga profesional de baloncesto NBA y elaborar un dashboard para mostrar esta información.

### Open Data:

Para recoger los datos de las estadísticas utilizamos dos métodos: solicitamos la información a una API para la mayoría de las estadísticas y, por otro lado, hacemos web scraping para recolectar los sueldos de los jugadores.

- Conexión a API:  
El sitio web <https://www.balldontlie.io/> recopila información de todas las temporadas de la NBA y la dispone al acceso de los usuarios a través de una API con diversos endpoints. La API no requiere Authentication y tiene un límite de peticiones de 60 por minuto.
- Web Scraping:  
En la web <https://hoopshype.com/salaries/> encontramos los salarios de todos los jugadores de la NBA por temporada. En la página de cada temporada se muestra una tabla con el sueldo al lado del nombre del jugador que serán los valores que recopilaremos.
- Código:  
El fichero de código que debemos ejecutar es `get_data/get_data.py`. Lo primero que hará la ejecución del código será hacer peticiones al endpoint que nos devolverá todas las estadísticas de cada jugador por cada partido. El endpoint tiene una paginación y un número máximo de elementos a devolver de 100. Por lo tanto, hace un total de más de 300 peticiones para obtener toda la información de una temporada (unos 15 minutos).  
Por otro lado, no debería haber ningún problema con el límite de peticiones ya que cada petición tarda más de 1 segundo. Sin embargo, este caso se ha tratado para que si nos hemos pasado del límite haga un sleep de varios segundos y a partir de ese momento, por cada petición haga una espera de 1 segundo.  
Con todos los datos recogidos creamos un dataframe que utilizaremos más tarde.  
A continuación, invocamos a una función que hace web scraping a la página que hemos mencionado antes con Selenium. Como resultado, generamos otro dataframe con los nombres y sueldos.

Hacemos un merge de ambos dataframes a través del nombre del jugador (al que antes hemos eliminado caracteres no alfabéticos para conseguir la máxima igualdad entre los nombres de los dos dataframes). De los más de 500 jugadores de la NBA sólo se quedan sin sueldo asignado unos 30 debido a este problema de mergear a través del nombre. Este último dataframe lo guardamos como csv en la carpeta de data y en la de visualitation.

## Visualización Dinámica

El dashboard se encuentra en subido en la dirección:

<https://javierherraezalbarran.shinyapps.io/statsNBA/>

El dashboard se encuentra dividido en tres pestañas:

- Estadísticas promedio semanales  
En esta pestaña mostramos la evolución temporal de las estadísticas medias por partido agrupadas por semanas. Además, podemos optar por mostrar la media de todos los equipo o elegir uno de ellos.  
Vemos que lo más normal es que las estadísticas se mantengan bastante estables a lo largo de la temporada. Podemos ver rachas malas o buenas de partidos para ciertos equipos. Un ejemplo de esto puede ser las primeras semanas de enero de los Cavaliers donde no pasaban de 100 puntos de media.
- Salarios por Equipos  
Mostramos en un mapa la localización de los equipos junto información acerca del salario de los jugadores. Podemos filtrar tanto para que se incluyan en las estadísticas distintas posiciones y para elegir que conferencia/división queremos mostrar. Podemos ver que equipos como Oklahoma City Thunder tiene salarios por lo general más bajos que otros como Golden State Warriors.
- Comparación de Estadísticas  
En esta pestaña mostramos un gráfico de barras en el que podemos seleccionar que queremos comparar entre Equipos, Posiciones o jugadores y a través de que estadística los vamos a comparar. De esta manera podremos hacer distintas comparaciones como por ejemplo ver que equipo anota más puntos o ver si los jugadores de una posición cobran más de media que en otras.