1. **Contexto.**

La NBA (Nationall Basket Association) es la liga profesional de Baloncesto estadounidense. Está considerada la mejor liga del mundo, tanto a nivel deportivo como de marketing y organización. Desde su sitio web es posible consultar gran cantidad de información sobre los jugadores y equipos que la componen.

En este trabajo vamos a realizar web scraping sobre ella para obtener información de las estadísticas de los jugadores de un equipo en los principales aspectos del juego (puntos, rebotes, asistencias, recuperaciones, tapones y porcentaje de tiros de campo) en una temporada. Con dicha información crearemos un dataset.

Adicionalmente, se realizará web scraping para descargar la “imagen del partido” de un partido concreto. En la página de resumen de cada partido NBA aparece una imagen descriptiva del encuentro (el jugador más destacado, la canasta definitiva, etc).

En esta práctica se realizará un script que cree una carpeta con un nombre identificativo del partido y en su interior la imagen que la NBA ha decidido utilizar como definitoria del mismo. De ejecutar este script diariamente sobre diferentes partidos, podríamos acabar teniendo a final de temporada el conjunto total de la “imagen del partido” de todos los partidos de la temporada.

1. **Título**

El dataset se denominará “Estadísticas de jugadores del equipo *Nombre\_equipo* Temporada *Año\_temporada*”, donde “Nombre\_equipo” es el nombre del equipo, que se introduce por parámetro y “Año\_temprada” la temprada de la que se desean obtener las estadísticas.

1. **Descripción del dataset**

Como hemos comentado con anterioridad, el dataset contendrá las estadísticas de las principales categorías baloncestísticas de una temporada de los jugadores de un determinado equipo.

Contendrá las siguientes columnas:

* **Jugador:** Nombre del jugador
* **Partidos:** Número de partidos disputados
* **Puntos:** Total de puntos anotados
* **FG%:** Porcentaje de acierto en tiros de campo
* **3PT%:** Porcentaje de acierto en tiros de 3 puntos
* **FT%:** Porcentaje de acierto en tiros libres
* **OffReb:** Total de rebotes en ataque capturados
* **DefReb:** Total de rebotes en defensa capturados
* **Rebotes:** Total de rebotes caputados
* **Asistencias:** Total de pases de canasta efectuados
* **Recueperaciones:** Total de recuperaciones logradas
* **Pérdidas:** Total de pérdidas de balón comentidas
* **Faltas:** Total de faltas cometidas

Y una fila por cada jugador que haya militado en dicho equipo durante la temporada seleccionada.

1. **Representación gráfica.**

A continuación ofrecemos una captura de pantalla de la página web de la que hemos obtenido la información:

**Tabla

Descripción generada automáticamente**

Realizando web scraping sobre el código html de la página hemos logrado descargar todos los datos a nuestro dataset. Adicionalmente, para el manejo de contenido audiovisual, hemos descargado la foto del jugador que acompaña a sus estadísticas.

Nuestro dataset tiene el siguiente formato:

1. **Contenido.**

Como hemos comentado en apartados anteriores, nuestro dataset permite almacenar la información estadística sobre el rendimiento de los jugadores NBA de un determinado equipo en una temporada.

Para ello, nuestro script de Python comienza pidiendo el nombre de un equipo y la temporada de la que se quieren obtener los datos.

En la NBA los equipos son conocidos por el nombre de la ciudad a la que representan más un “apodo” por el que son conocidos. Este “apodo” es el que habrá que introducir como parámetro cuando se solicite el nombre del equipo.

Así, por ejemplo, si queremos obtener las estadísticas de Boston Celtics deberemos introducir “Celtics”. Y si queremos obtener las de los jugadores de Los Ángeles Lakers deberemos introducir “Lakers”.

Para la temporada, el formato requerido es “20xx-xx+1”; es decir, si queremos obtener los resultados de la temporada 2020-2021 introduciremos “2020-21”.

Desde el punto de vista técnico, hemos implementado dos ficheros:

* **Scraper.py**.

Es el fichero que realiza el web scraping propiamente dicho en la página "https://www.nba.com/%s/stats?season=%s" . Las variables de esta url son los parámetros que hemos indicado con anterioridad.

Haciendo uso de la librería BeautifulSoup buscamos la tabla de la página donde están representadas las estadísticas de los juegadores e iteramos sobre ella para obtener los datos que luego almacenaremos en nuestro dataset. Para cada fila de la tabla (la correspondiente a un jugador) obtenemos los valores correspondientes a cada apartado estadístico (columnas).

Adicionalmente, descargamos la foto de cada jugador en una carpeta “pictures”.

* main.py. Es el fichero principal que pide los datos al usuario y utiliza el “Scraper.py” para generar el dataset y descargar las imágenes.

1. **Agradecimientos**

La NBA, además de ser deportivamente la mejor liga de baloncesto del mundo, está considerada como una de las mejores marcas deportivas a nivel de publicidad y difusión.

En su página web es posible encontrar una cantidad ingente de información sobre sus jugadores y equipos (tanto actuales como anteriores).

Aprovechando esto, nos ha parecido interesante realizar la práctica de web scraping sobre las estadísticas acumuladas de los jugadores de una plantilla en una temporada. Hemos revisado el archivo “robots.txt” (https://www.nba.com/robots.txt) y aunque se incluyen algunas directivas “disallow”, no afectan a las estadísticas por equipo de los jugadores.

Adjuntamos el contenido del fichero donde se puede comprobar que no hay ningún impedimento para acceder al contenido a partir de /equipo/stats



Buscando trabajos similares realizados anteriormente, hemos encontrado análisis de estadísticas de baloncesto desde la web <https://www.basketball-reference.com/> , que almacena información no solo de la NBA sino de otras ligas de baloncesto.

En nuestro caso, hemos utilizado información pública de la web de la NBA para realizar nuestro dataset.

1. **Inspiración**

En la actualidad, en el mundo del baloncesto se le da una tremenda importancia al análisis del rendimiento de los jugadores.

Desde hace unos años se está utilizando en todos los equipos el concepto de “estadística avanzada”, que analiza a muy bajo nivel de detalle el rendimiento de los jugadores.

Así, los equipos disponen de información referente a desde qué punto de la pista es más efectivo un jugador en sus lanzamientos o si determinado jugador suele fallar más tiros por exceso o defecto de fuerza (lo cual da lugar a que el rebote del balón en el aro pueda salir en una u otra dirección).

Nuestro web scraping no resulta tan ambicioso, pero sí que permite considerar una cantidad de datos que pueden dar lugar a un análisis más exhaustivo del que permite un simple vistazo a las estadísticas generales.

De este modo, la extracción de estos datos nos podría permitir crear columnas “adicionales” en nuestro dataset.

Pongamos que nos interesa ver el ratio de puntos anotados por minuto de juego disputado, ya que no tiene la misma dificultad anotar 10 puntos disputando 40 minutos que disputando 20. Al disponer de la información en un dataset resulta trivial añadir una nueva columna para calcular este cociente