Placido A. López Ávila

Alex Pardo Ramos

**Tipología y ciclo de vida de los datos**

**Práctica 1**

**1. Contexto**

ESPN Deportes es un canal de televisión por suscripción estadounidense que emite eventos en español, propiedad de Walt Disney Televisión. Por esta razón, en su página web tiene información sobre eventos deportivos de todo tipo, y en este caso nos hemos centrado en el ámbito del baloncesto, en concreto en la liga estadounidense que es la NBA. De la cual dispone de toda la información posible sobre los partidos, equipos, jugadores y demás.

**2. Título**

Información NBA en intervalo de tiempo

**3. Descripción del dataset**

El dataset extraído se basa en la información de los partidos en un tiempo concreto, elegido por el usuario, y a su vez las estadísticas de los equipos y los jugadores que han participado en dichos partidos. Hay todo tipo de estadísticas relacionadas con el juego como puntos, tiros, rebotes, robos, etc. Esta información podría ser muy útil si es bien utilizada.

**4. Representación gráfica**

**ESPN NBA**

**Web Scraping**

**Extracción Partidos**

**Extracción Jugadores**

**Extracción Calendario**

**Dataset**

**5. Contenido**

Una primera extracción la cual está centrada principalmente en los partidos del tiempo escogido la información extraída de estos es la fecha, el equipo visitante, el equipo local, el resultado, máximo anotador del equipo ganador, Puntos del equipo ganador, máximo anotador del equipo perdedor, puntos del equipo perdedor y la url del partido.

Una segunda extracción la cual extrae información más concreta de cada uno de estos partidos, en concreto esta información es el id del partido, y estadísticas tanto para el equipo visitante como para el local que son tiros de campo, porcentaje de acierto en tiros de campo, tiros de tres puntos, porcentaje de acierto en tiros de tres puntos, tiros libres, porcentaje de acierto en tiros libres, rebotes, rebotes ofensivos, rebotes defensivos, asistencias, robos, bloqueos, perdidas, puntos a raíz de perdida, puntos de contraataque, puntos en la pintura, faltas, faltas técnicas, faltas intencionales y ventaja más amplia.

Y, por último, la tercera extracción que se compone de la información por jugador en cada uno de los partidos, en concreto esta información es el id del partido, el equipo y el nombre del jugador, posición, Titular o suplente, minutos, tiros de campo, porcentaje de triples, tiros libres anotados e intentados, rebotes ofensivos, rebotes defensivos, rebotes, asistencias, robos, bloqueos, perdidas, faltas, balance de puntos y puntos.

Tras esto se realiza un merge de las tres extracciones para combinar la información, y generar un dataset con toda esta información.

**6. Agradecimientos**

El propietario de este conjunto de datos como ya hemos comentado antes es la compañía Disney. Hay diversos análisis en repositorios como GitHub que pueden tener algunas semejanzas con este, pero ninguno de ellos trata toda la información de partidos, equipos y jugadores en base a tiempo, por lo tanto, es difícil la comparación con cualquiera de estos. A la hora de realizar el proyecto, hemos tenido en cuenta las pocas restricciones de uso de estos datos que podemos encontrar al final de la web, para no realizar nada fuera del marco permitido.

**7. Inspiración**

Es interesante este conjunto de datos porque nos permitirán el estudio tanto de partidos, equipos y jugadores en un tiempo concreto, siendo el usuario quien elige el tiempo a tener en cuenta lo cual diferencia este estudio a los anteriores a este. Podemos responder casi cualquier pregunta sobre estos tres ámbitos. Esto permitirá tener un acceso mucho más directo e incluso realizar estudios sobre la propia información que se extrae. Es decir, este conjunto de datos nos proporcionara acceso a información temporal muy detallada de la NBA.

**8. Licencia**

Relased Under CC BY-SA 4.0. License. Se tiene la libertad de compartir y adaptar los datos de los cuales se dispone. Las condiciones de esta son que se debe reconocer adecuadamente la autoría y que, aunque se mezcle, transforme o cree a partir del material, hay obligatoriedad de que la licencia sea la misma que la original.

**9. Código**

Enlace a repositorio Git.

<https://github.com/plopezavi/Web-Scraping>

**10. Dataset**

Enlace de los dataset en formato .csv en Zenodo.

FALTA ENLACE

**Firmas**

|  |  |
| --- | --- |
| **Contribuciones** | **Firmas** |
| Investigación previa | P.L.A. , A.P.R. |
| Redacción de las respuestas | P.L.A. , A.P.R. |
| Desarrollo del código | P.L.A. , A.P.R. |