

DATASET: ESTADÍSTICAS DEL MEJOR QUINTETO NBA POR AÑO

Germán Yepes Calero y Pablo Mas Cayuelas

Práctica 1

Descripción

La práctica, correspondiente a la asignatura *Tipología y ciclo de vida de los datos*, perteneciente al Máster en Ciencia de Datos de la Universitat Oberta de Catalunya, tiene como objetivo aplicar técnicas de *web scraping*. El código es ejecutado en el lenguaje de programación Python. En ella, reunimos los datos correspondientes a los cinco mejores jugadores de la NBA en base a la posición que ocupan, realizándose una estadística final asociada a la media de cada quinteto desde el año 2000.

Representación gráfica



Figura 1: principales candidatos para ganar el MVP de la temporada 2019-2020.

Miembros del equipo

La actividad ha sido realizada de manera colectiva por **Pablo Mas Cayuelas** y **Germán Yepes Calero**.

Contenido

Para cada temporada, se recopilan los cinco mejores jugadores de la misma en cada una de las posiciones del juego. Además, se indica si ha sido MVP ese jugador en el año en cuestión y se realiza la media de las estadísticas de los cinco jugadores. Las características recogidas son las siguientes:

- **Name:** nombre y apellidos de cada uno de los jugadores de nuestra base de datos.
- **Team:** equipo en el que disputó el jugador esa temporada.
- **Year:** año en el que empezó la temporada.
- **Position:** posición en la que fue el mejor jugador.
- **Games:** partidos disputados durante la temporada regular.
- **Points:** puntos medios por partido del jugador.
- **Percentage:** media del porcentaje en tiros de campo.
- **Rebounds:** media de los rebotes del jugador por partido.
- **Assists:** número de asistencias medias por partido del jugador.
- **Steals:** media del número de robos de balón por partido.
- **Blocks:** media del número de tapones por partidos.
- **MVP:** indica si el jugador ha sido MVP esa temporada o no.

Agradecimientos

La mayoría de los datos han sido extraídos mediante técnicas de *web scraping* de esta [web](#), cambiando en ella el año y las diferentes posiciones disponibles. Adicionalmente, para extraer la información de si en ese año ha recibido el jugador el MVP o no, se ha recurrido a la [wikipedia](#), cambiando también la url para obtener la información de los veinte años para los que hemos recopilado los datos. A pesar de haber obtenido la información de estas páginas, ésta es proporcionada por la propia NBA cada año.

Inspiración

La información del presente juego de datos se puede utilizar sobre todo a nivel periodístico, realizándose así, por medio de las estadísticas medias de cada año, un estudio de cómo ha ido evolucionando el jugador medio a lo largo de los años en la

NBA. De hecho, la mayoría de los equipos tienen analistas dedicados a recopilar este tipo de datos y realizar medias para saber cómo afrontar el próximo partidos en base al tipo de jugador al que se enfrenten.

A nivel personal, además, se juntan con esta base de datos dos de nuestros intereses, que son el análisis de datos y el baloncesto, por lo que decidimos realizar el *web scraping* de este deporte, y vimos estas *webs* adecuadas para ello.

Licencia

La licencia escogida es **Released Under CC BY-SA 4.0 License**. El motivo de la elección es que ésta permite el uso comercial de la obra y de las posibles obras derivadas, pero otras obras que deriven de la misma deben llevarse a cabo con la misma licencia. Así, nos aseguramos de que no se restrinja su utilización. De esta forma, se puede reconocer el trabajo ajeno y en qué medida se ha aportado al trabajo original. Además, es bueno exponer la base de datos a que pueda ser utilizada por una empresa, conociéndose así nuestro trabajo.

Código y dataset

Tanto el código (archivo python) como el dataset (archivo CSV) están accesibles en este mismo portal.

Resumen de la actividad

Contribuciones	Firma
Investigación previa	GYC, PMC
Redacción de las respuestas	GYC, PMC
Desarrollo del código	GYC, PMC