
Lab Book

Bases de Datos Avanzadas

Mario Pérez Sánchez-Montañez

David Camuñas Sánchez

Francesco Zingariello

February 26, 2020

Contents

Martes, 11 de Febrero	1
1 Elección de la tecnologías	1
2 Elección del motor de base de datos	1
Jueves, 13 Febrero 2020	2
1 Obtención de los datos para el <i>DataWarehouse</i>	2
2 Cambio en diseño del <i>Data Warehouse</i>	2
Viernes, 14 Febrero 2020	3
1 Primer versión de la base de datos	3
2 Primeras consultas en la Base de Datos	3
Lunes, 17 Febrero 2020	4
1 Cambio en la obtención de datos	4
2 Realización de consultas	4
Lunes, 24 Febrero 2020	5
1 Modelo de regresión lineal de jugadores	5
Martes, 25 Febrero 2020	6
1 Documentar el modelo de regresión lineal	6

Martes, 11 de Febrero

1 Elección de la tecnologías

Se ha decidido usar como tecnología *Python* debido a la multiples librerías que posee relacionadas con el ámbito del análisis de datos.

2 Elección del motor de base de datos

Se ha elegido como motor de nuestra base de datos *SQLite*, ya que soporta funciones SQL y podemos almacenar la base de datos completa en un solo archivo.

Jueves, 13 Febrero 2020

1 Obtención de los datos para el DataWarehouse

Desarrollamos tres *scripts* en *Python* utilizando técnicas de *web scrapping* para obtener datos de jugadores y equipos de *Basketball Reference*.

2 Cambio en diseño del Data Warehouse

Decidimos que vamos a tener una tabla con todos los jugadores, con los 30 equipos y de *rookies* (jugadores nóveles). Todo ello estará relacionado en una tabla central, siguiendo el modelo de estrella.

Viernes, 14 Febrero 2020

1 Primer versión de la base de datos

Realizamos la construcción de la primera versión de la base de datos. Uno de los problemas que hemos encontrado es con los jugadores que han jugado en varios equipos en la misma temporada, ya que en la fuente que hemos usado se muestran datos en cada equipo donde han jugado y lo totales, identificando estos últimos como 'TOT' en la columna de equipo.

Hemos decido mantener esta estructura identificando las estadísticas totales con -1, para que en la base de datos no se utilice un id de un equipo existente.

2 Primeras consultas en la Base de Datos

Hemos realizado las primeras las consultas en la base de datos de nuestro Data Warehouse. Cuyo resultado ha salido satisfactoriamente.

Lunes, 17 Febrero 2020

1 Cambio en la obtención de datos

Se ha realizado unos cambios en cuanto a la forma de obtener los datos. Estos cambios han afectado al número de *scripts*. Ahora los datos se obtienen mediante un solo *script* (en *Python*), cuyo objetivo es obtener los datos mediante la técnica conocida como *Web Scrapping*. Además para la creación de las tablas que compondrán nuestra base de datos, se utiliza otro nuevo *script* (en *SQL*).

2 Realización de consultas

Se han realizado una serie de consultas, cuyo objetivo es obtener la información necesaria de nuestra base de datos. Y así obtener información para poder llevar a cabo un estudio estadístico. Este estudio se realizara mediante el uso de consultas (dicho anteriormente) y el uso de *dataframes*.

Lunes, 24 Febrero 2020

1 Modelo de regresión lineal de jugadores

Con los datos obtenidos de nuestra base de datos se buscará predecir cuanto anotará cada jugador. Para ello, se ha creado un modelo de regresión lineal basado en los intentos de tiro de la temporada almacenada en la base de datos

Martes, 25 Febrero 2020

1 Documentar el modelo de regresión lineal

Se ha documentado el modelo creado anteriormente, para entender como podemos usarlo para predecir cuantos puntos conseguirá un jugador según sus intentos de tiro.

Miércoles, 26 Febrero 2020

1 Documentar el modelo de regresión lineal

Se ha terminado de documentar el modelo de regresión lineal, introduciendo tablas y gráficas para interpretarlo mejor.