
Lab Book

Bases de Datos Avanzadas

Mario Prez Snchez-Montaez

David Camuas Snchez

Francesco Zingariello

March 2, 2020

Contents

Martes, 11 de Febrero	1
1 Eleccin de la tecnologas	1
2 Eleccin del motor de base de datos	1
Jueves, 13 Febrero 2020	2
1 Obtencin de los datos para el <i>Data Warehouse</i>	2
2 Cambio en diseo del <i>Data Warehouse</i>	2
Viernes, 14 Febrero 2020	3
1 Primer versin de la base de datos	3
2 Primeras consultas en la Base de Datos	3
Lunes, 17 Febrero 2020	4
1 Cambio en la obtencin de datos	4
2 Realizacin de consultas	4
Lunes, 24 Febrero 2020	5
1 Modelo de regresin lineal de jugadores	5
Martes, 25 Febrero 2020	6
1 Documentar el modelo de regresin lineal	6
2 Generar los grficos estadsticos	6
Mircoles, 26 Febrero 2020	7
1 Documentar el modelo de regresin lineal	7
2 Documentar consultas y estadsticas	7
Lunes, 2 Marzo 2020	8
1 Creada presentacin del trabajo	8

Martes, 11 de Febrero

1 Eleccin de la tecnologas

Se ha decidido usar como tecnologa *Python* debido a la multiples librerias que posee relacionadas con el mbito del anlisis de datos.

2 Eleccin 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 Obtencin de los datos para el DataWarehouse

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

2 Cambio en diseo del Data Warehouse

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

Viernes, 14 Febrero 2020

1 Primer versin de la base de datos

Realizamos la construccin de la primera versin 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 ltimos como 'TOT' en la columna de equipo.

Hemos decido mantener esta estructura identificando las estadsticas 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 obtencin de datos

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

2 Realizacin de consultas

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

Lunes, 24 Febrero 2020

1 Modelo de regresin lineal de jugadores

Con los datos obtenidos de nuestra base de datos se buscar predecir cuanto anotar cada jugador. Para ello, se ha creado un modelo de regresin 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 regresin lineal

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

2 Generar los grficos estadsticos

Se han realizado una serie de grficos sobre las consultas realizadas el da: *Lunes, 17 Febrero 2020*.

Mircoles, 26 Febrero 2020

1 Documentar el modelo de regresin lineal

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

2 Documentar consultas y estadsticas

Se ha documentado el apartado que trata sobre las distintas consultas que hemos realizado en nuestra base de datos, para asi poder obtener cierta informacin y mostrarla con la utilizacin de *dataset* y *grficos*.

Lunes, 2 Marzo 2020

1 Creada presentacin del proyecto

Se ha realizado la presentacin del proyecto, el cual se expondr el da: *Martes, 3 Febrero 2020*.