

Trabajo CBD

Estadísticas de jugadores de la NBA con Pentaho



Realizado por:

Flores Rodríguez, Román

Danko, Siamion

Índice

Índice	2
Introducción	3
Marco teórico	4
Pentaho	4
Oracle Database	4
Manual de instalación	5
Base de datos inicial	6
Implementación de filtros en Pentaho	8
Ranking de jugadores	8
Obtención de plantillas ganadoras	13
Conclusión	16
Referencias	17
Bibliografía	17
Apéndice	18

Introducción

La NBA es una de las ligas de baloncesto más populares en el mundo. Cada temporada, los equipos compiten por lograr los mejores resultados y para ello requieren de un análisis exhaustivo de las estadísticas tanto de sus propios jugadores como de los adversarios para encontrar las estrategias adecuadas. Estas estadísticas son obtenidas a partir de datos recolectados por numerosas fuentes y tienen naturalezas muy distintas.

El objetivo final de nuestro proyecto es procesar la información obtenida de una base de datos histórica de la NBA en datos estadísticos sobre los resultados y rendimiento de los jugadores a lo largo de los años. Estos datos permitirán a los equipos analizar y ajustar sus estrategias acorde a los resultados obtenidos con anterioridad de manera similar a cómo se hace en la liga actual.

En este trabajo partimos de una base de datos obtenida de internet que contiene información sobre los equipos, partidos jugados, jugadores y estadísticas de estos desde 1998 hasta la actualidad. Esta información la procesaremos aplicando una sucesión de filtros y transformaciones con Pentaho que nos permitirán obtener datos estadísticos relevantes para su análisis por parte de los equipos para mejorar sus estrategias y rendimientos.

Al final de nuestro proyecto, hemos obtenido los ranking de los jugadores que más destacaron en cada temporada por sus tiros, rebotes, asistencias y tapones, así como las plantillas de los equipos que ganaron la liga en cada una de las temporadas. Estos datos permitirán a los equipos saber qué estrategias han funcionado con anterioridad, qué jugadores trabajan bien con otros o cuáles jugadores tener en cuenta a la hora de fichar. También se pueden utilizar todos estos datos para entrenar una IA para la predicción de resultados de partidos.

Marco teórico

A continuación, explicaremos brevemente las tecnologías usadas durante el desarrollo de nuestro proyecto.

Pentaho

Pentaho es un conjunto de herramientas de business intelligence que permite a las empresas procesar, gestionar y analizar datos relevantes para el funcionamiento de sus negocios y estrategias.

En concreto, nosotros hemos usado Pentaho Data Integration que permite la extracción, transformación y carga de datos con cientos de filtros que podemos aplicar para estos fines (1).

Además, Pentaho nos permite trabajar con numerosas bases de datos distintas, entre ellas la base de datos de Oracle que es la que hemos elegido para este trabajo.

Oracle Database

Oracle Database es un sistema de gestión de base de datos objeto-relacional que permite acceder y modificar gran volumen de datos de forma rápida y eficiente (2).

El motivo principal por el que hemos decidido utilizar esta base de datos es debido a que el script que contiene los datos de la NBA que hemos obtenido es de esta base de datos. Además, es un sistema con el que ya hemos trabajado tanto en esta como en otras asignaturas.

Manual de instalación

Para poder ejecutar y visualizar el proyecto en nuestro equipo, debemos seguir los siguientes pasos.

1. Descargamos el proyecto del repositorio de github disponible en el apéndice.
2. Descargamos e instalamos Oracle Database 21c Express Edition.
3. Descargamos e instalamos la versión 9.3 de Pentaho.
4. En SQL Developer, creamos una nueva conexión llamada “cbd” con el usuario “system”. Lo hacemos este usuario porque es el que hemos utilizado durante el desarrollo del proyecto, pero se puede crear uno nuevo.
5. Con la conexión anterior, importamos y ejecutamos el script “nba.sql” para poblar la base de datos inicial.
6. Crear las siguientes tablas con conexión system; para la tabla de ganadores, una tabla vacía llamada CATALOGO_TEMPORADAS, para la tabla de top 10 jugadores por puntos por temporada, una tabla vacía llamada TOP_PUNTOS, para la tabla de top 10 jugadores por asistencias por temporada, una tabla vacía llamada TOP_ASISTENCIAS, para la tabla de top 10 jugadores por rebotes por temporada, una tabla vacía llamada TOP_REBOTES, para la tabla de top 10 jugadores por tapones por temporada, una tabla vacía llamada TOP_TAPONES.

Atte: Según la versión de Spoon y SQLDeveloper usadas(recomendamos 9.3 de Spoon y 23.1 de SQLDeveloper) podría haber problema en el atributo esperado de salida, para ello en las salidas de tablas del script, aparece la opción de ejecutar un script sql, el cual nos dará el formato de los atributos necesarios para cada tabla.

7. Abrimos spoon e importamos el archivo “trabajo.ktr”.
8. Hacemos click en “Conexiones a base de datos” y acto seguido hacemos doble click sobre la conexión “cbd” que aparece.
9. En la conexión, cambiamos el usuario y la contraseña por las que hemos creado en el paso 4, tal y como se ve en la *figura X*.
10. Una vez guardados los cambios, podemos ejecutar el proyecto con la tecla “F9”.

Base de datos inicial

Antes de explicar el desarrollo de nuestro proyecto, vamos a explicar brevemente el contenido de la base de datos histórica que usaremos como punto de partida. La página web de la que hemos obtenido dicha base de datos viene especificada en las referencias del trabajo (3).

La base de datos viene dada por un solo script SQL que genera y puebla las siguientes tablas:

- Equipos: contiene el nombre, la ciudad, la conferencia y la división de los distintos equipos que participan en la liga. Tiene como clave primaria el nombre del equipo. Véase la figura 1.

	NOMBRE	CIUDAD	CONFERENCIA	DIVISION
1	Celtics	Boston	East	Atlantic
2	Raptors	Toronto	East	Atlantic
3	76ers	Philadelphia	East	Atlantic
4	Nets	New Jersey	East	Atlantic
5	Knicks	New York	East	Atlantic
6	Pistons	Detroit	East	Central
7	Cavaliers	Cleveland	East	Central
8	Pacers	Indiana	East	Central
9	Bulls	Chicago	East	Central
10	Bucks	Milwaukee	East	Central

Figura 1. Primeras 10 entradas de la tabla equipos.

- Jugadores: contiene un id, el nombre, la procedencia, la altura, el peso, la posición y el nombre del equipo al que pertenece el jugador. El id es asignado de forma automática para su uso como primary key. Véase la figura 2.

❖ CODIGO	❖ NOMBRE	❖ PROCEDENCIA	❖ ALTURA	❖ PESO	❖ POSICION	❖ NOMBRE_EQUIPO
1	1 Corey Brever	Florida	6-9	185	F-G	Timberwolves
2	2 Greg Buckner	Clemson	6-4	210	G-F	Timberwolves
3	3 Michael Doleac	Utah	6-11	262	C	Timberwolves
4	4 Randy Foye	Villanova	6-4	213	G	Timberwolves
5	5 Ryan Gomes	Providence	6-7	250	F	Timberwolves
6	6 Marko Jaric	Serbia	6-7	224	G	Timberwolves
7	7 Al Jefferson	Prentiss Hs	6-10	265	C-F	Timberwolves
8	8 Mark Madsen	Stanford	6-9	250	C-F	Timberwolves
9	9 Rashard McCants	North Carolina	6-4	21	G	Timberwolves
10	10 Chris Richard	Florida	6-9	270	F	Timberwolves

Figura 2. Primeras 10 entradas de la tabla jugadores.

- Estadísticas: contiene un id, tapones, puntos, rebotes y asistencias por partido, equipo, de cada jugador ordenado por temporada. Al igual que en la tabla de jugadores, el id se asigna automáticamente y se usa como primary key. Véase la figura 3.

❖ TEMPORADA	❖ JUGADOR	❖ PUNTOS_POR_PARTIDO	❖ ASISTENCIAS_POR_PARTIDO	❖ TAPONES_POR_PARTIDO	❖ REBOTES_POR_PARTIDO
1 03/04	283	5,9	2,2	1,54	1,6
2 04/05	283	4,6	1,8	0,98	1,6
3 05/06	283	12	4,7	2,4	2,9
4 06/07	283	4,9	1,3	0,89	0,8
5 05/06	85	5,5	4,5	0,1	2,2
6 06/07	85	8,7	5	0,1	1,7
7 07/08	283	9,5	3	1,33	2,1
8 07/08	85	11,6	8,2	0,1	2,9
9 03/04	580	7,9	2,4	0,1	1,8
10 05/06	580	13,1	2,8	0,1	2,6

Figura 3. Primeras 10 entradas de la tabla estadísticas.

- Partidos: contiene un id, el equipo local, el visitante y los puntos que anotaron cada uno y la temporada en la que se disputó el partido. Véase la figura 4.

	⚡ CODIGO	⚡ EQUIPO_LOCAL	⚡ EQUIPO_VISITANTE	⚡ PUNTOS_LOCAL	⚡ PUNTOS_VISITANTE	⚡ TEMPORADA
1	1	Raptors	Lakers	74	117	98/99
2	2	Raptors	Grizzlies	70	118	98/99
3	3	Raptors	Clippers	105	76	98/99
4	4	Raptors	Knicks	62	124	98/99
5	5	Raptors	Timberwolves	66	68	98/99
6	6	Raptors	Celtics	52	69	98/99
7	7	Raptors	76ers	74	103	98/99
8	8	Raptors	Nets	121	129	98/99
9	9	Raptors	Pistons	100	85	98/99
10	10	Raptors	Cavaliers	129	158	98/99

Figura 4. Primeras 10 entradas de la tabla partidos.

Implementación de filtros en Pentaho

A continuación procedemos a explicar los pasos que hemos seguido para procesar los datos en información que consideramos más importante para el análisis. Nuestro trabajo se puede dividir en dos grandes partes: obtención del ranking de jugadores para cada estadística por temporada y recopilación de las plantillas de los equipos que más partidos ganaron en cada temporada.

Ranking de jugadores

El primer paso es, obviamente, la lectura de datos. Véase la figura 5.

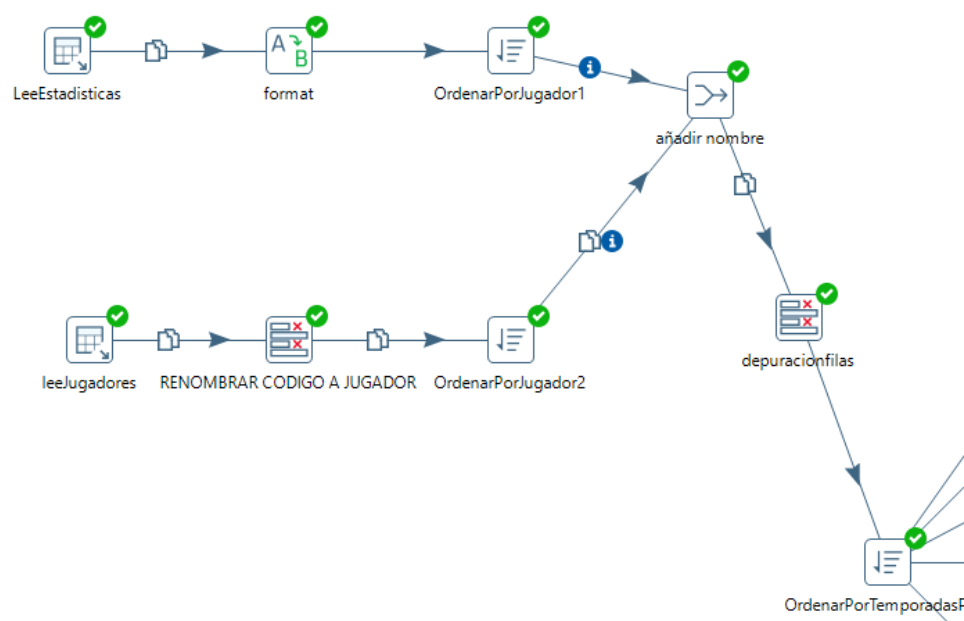


Figura 5. Lectura de las tablas estadísticas y jugadores.

Como dijimos anteriormente, la tabla estadísticas contiene los valores para cada uno de los jugadores por temporada. Por lo tanto, durante la lectura, podemos añadir nuevas columnas que contengan la posición en el ranking de jugadores para cada estadística. Esto lo conseguimos comparando el valor de cada estadística para un jugador

con los demás jugadores en la misma temporada y asignándole un valor dependiendo de si obtuvo mejores o peores resultados. Véase la figura 6.

```
SELECT
TEMPORADA,
JUGADOR,
PUNTOS_POR_PARTIDO,
ASISTENCIAS_POR_PARTIDO,
TAPONES_POR_PARTIDO,
REBOTES_POR_PARTIDO,
(SELECT COUNT(*)
FROM SYSTEM.ESTADISTICAS e2
WHERE e2.TEMPORADA = e1.TEMPORADA
AND e2.PUNTOS_POR_PARTIDO > e1.PUNTOS_POR_PARTIDO) + 1 AS PUNTOS_RANK,
(SELECT COUNT(*)
FROM SYSTEM.ESTADISTICAS e2
WHERE e2.TEMPORADA = e1.TEMPORADA
AND e2.ASISTENCIAS_POR_PARTIDO > e1.ASISTENCIAS_POR_PARTIDO) + 1 AS ASISTENCIAS_RANK,
(SELECT COUNT(*)
FROM SYSTEM.ESTADISTICAS e2
WHERE e2.TEMPORADA = e1.TEMPORADA
AND e2.TAPONES_POR_PARTIDO > e1.TAPONES_POR_PARTIDO) + 1 AS TAPONES_RANK,
(SELECT COUNT(*)
FROM SYSTEM.ESTADISTICAS e2
WHERE e2.TEMPORADA = e1.TEMPORADA
AND e2.REBOTES_POR_PARTIDO > e1.REBOTES_POR_PARTIDO) + 1 AS REBOTES_RANK
FROM SYSTEM.ESTADISTICAS e1
ORDER BY TEMPORADA, PUNTOS_RANK, ASISTENCIAS_RANK, TAPONES_RANK, REBOTES_RANK
```

Figura 6. Lectura de datos de “LeeEstadisticas”.

En estos datos, nos encontramos con algunos errores de consistencia en el formato de la temporada. En concreto, hay 2 entradas que en vez de seguir el formato “XX/YY” para denominar los años de la temporada, tienen asignado el valor “5.6” para denominar la temporada “05/06” y “02.03” para denominar la temporada “02/03”. Debido a esto, debemos realizar una transformación tipo “mapeo de valores” y normalizar los datos.

Por otro lado, nos encontramos con que no contienen el nombre de cada jugador, sino un id que hace referencia a la tabla de jugadores. Por ello, realizamos en paralelo una carga de los datos de la tabla jugadores para poder hacer “join” de ambas, aunque antes debemos renombrar el nombre del campo “código” en la tabla de jugadores a “jugador” para que coincida con el nombre que tiene el mismo campo en la de estadísticas y ordenar ambas tablas por dicho valor. El orden de estas operaciones se puede ver en la figura 5.

La parte final de la lectura, es un depurado de las filas obtenidas del “join” con el fin de eliminar algunas columnas con datos repetidos. Se puede ver una previsualización de los datos obtenidos en la figura 7.

#	TEMPORADA	JUGADOR	PUNTOS_POR_PARTIDO	ASISTENCIAS_POR_PARTIDO	TAPONES_POR_PARTIDO	REBOTES_POR_PARTIDO	PUNTOS_RANK	ASISTENCIAS_RANK	TAPONES_RANK	REBOTES_RANK	NOMBRE	NOMBRE_EQUIPO
1	05/06	1	5.6	1.4	0.3	3.6	1.0	1.0	1.0	1.0	Corey Brewer	Timberwolves
2	00/01	2	6.2	1.3	0.2	4.2	93.0	80.0	77.0	62.0	Greg Buckner	Timberwolves
3	01/02	2	5.8	1.1	0.4	3.9	120.0	108.0	81.0	88.0	Greg Buckner	Timberwolves
4	02/03	2	6.0	2.9	0.2	2.9	127.0	53.0	116.0	125.0	Greg Buckner	Timberwolves
5	03/04	2	3.1	0.8	0.1	1.9	202.0	168.0	173.0	190.0	Greg Buckner	Timberwolves
6	04/05	2	6.2	1.9	0.1	3.0	186.0	109.0	199.0	162.0	Greg Buckner	Timberwolves
7	05/06	2	6.7	1.7	0.3	2.9	193.0	137.0	133.0	189.0	Greg Buckner	Timberwolves
8	06/07	2	4.0	0.9	0.1	4.0	275.0	224.0	240.0	139.0	Greg Buckner	Timberwolves
9	07/08	2	4.1	1.3	0.1	2.2	293.0	205.0	285.0	274.0	Greg Buckner	Timberwolves
10	99/00	2	5.7	1.1	0.4	3.6	81.0	74.0	44.0	63.0	Greg Buckner	Timberwolves
11	00/01	3	6.4	0.8	0.5	3.5	90.0	98.0	57.0	78.0	Michael Doleac	Timberwolves
12	01/02	3	4.6	0.6	0.3	4.0	131.0	134.0	98.0	84.0	Michael Doleac	Timberwolves
13	02/03	3	4.4	0.6	0.2	2.9	149.0	155.0	116.0	125.0	Michael Doleac	Timberwolves
14	03/04	3	5.0	0.7	0.6	4.1	168.0	177.0	62.0	108.0	Michael Doleac	Timberwolves
15	04/05	3	4.0	0.6	0.3	3.2	224.0	222.0	123.0	153.0	Michael Doleac	Timberwolves
16	05/06	3	3.2	0.3	0.2	2.7	267.0	275.0	173.0	201.0	Michael Doleac	Timberwolves
17	06/07	3	3.6	0.4	0.3	2.8	288.0	283.0	137.0	209.0	Michael Doleac	Timberwolves
18	07/08	3	2.4	0.3	0.4	2.0	350.0	348.0	133.0	297.0	Michael Doleac	Timberwolves
19	98/99	3	6.2	0.4	0.3	3.0	55.0	69.0	42.0	50.0	Michael Doleac	Timberwolves

Figura 7. Previsualización de datos del join depurado.

Una vez tenemos los jugadores con sus estadísticas y su posición en el ranking para cada una de ellas por temporada, pasamos a obtener el top 10 de los jugadores por temporada en cada modalidad, tal y como se ve en la figura 8.

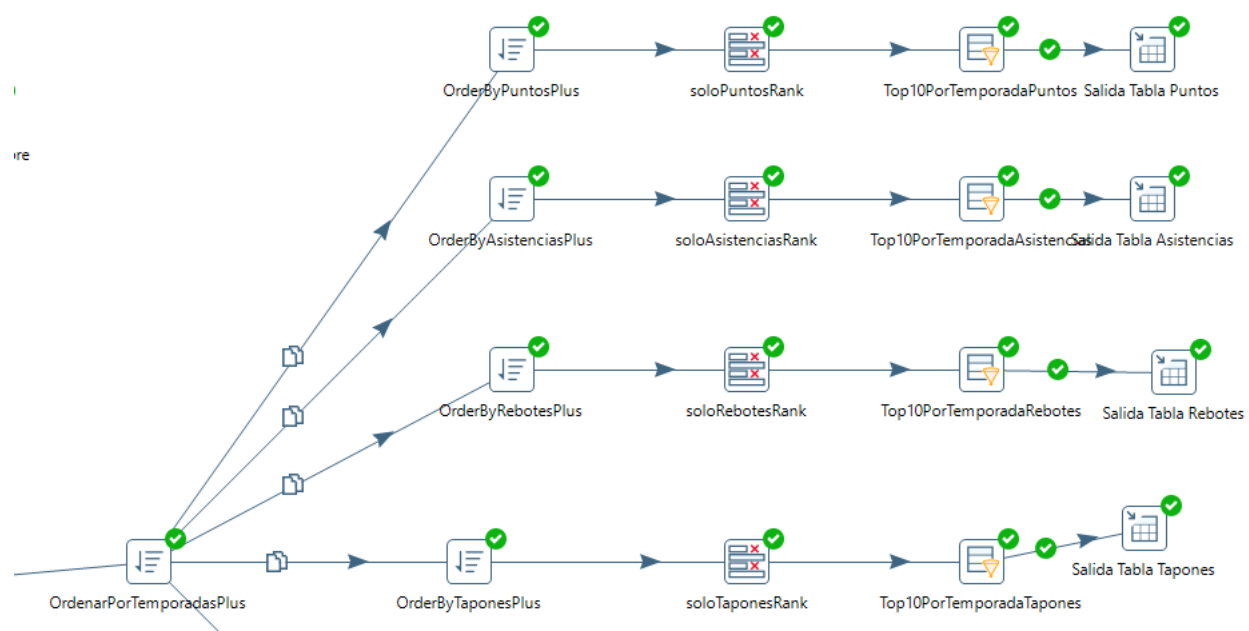


Figura 8. Ordenación y formato por atributo y temporada.

Lo primero que hacemos es ordenar todas las entradas por temporada para tener una especie de orden cronológico de los datos y mejorar así su legibilidad. Acto seguido, dividimos el flujo de las transformaciones en 4 ramas y ordenamos en cada una por una estadística distinta de forma descendente: puntos anotados, asistencias, tapones y rebotes. Finalmente, obtenemos las 10 primeras entradas de cada ramificación para obtener nuestro top 10 por temporada, que podemos ahora guardar en una nueva tabla en nuestra base de datos.

Podemos ver una previsualización de los datos obtenidos para el ranking de puntos en la *figura 9*.

#	TEMPORADA	JUGADOR	PUNTOS_POR_PARTIDO	ASISTENCIAS_POR_PARTIDO	TAPONES_POR_PARTIDO	REBOTES_POR_PARTIDO	NOMBRE	NOMBRE EQUIPO	PUNTOS_RANK
1	00/01	451	29.8	5.1	0.7	3.9	Jerry Stackhouse	Mavericks	1.0
2	00/01	587	28.7	3.7	2.8	12.0	Shaquille O'Neal	Suns	2.0
3	00/01	62	28.5	5.0	0.6	5.9	Kobe Bryant	Lakers	3.0
4	00/01	464	27.6	3.9	1.1	5.5	Vince Carter	Nets	4.0
5	00/01	349	26.8	4.6	1.5	7.5	Tracy McGrady	Rockets	5.0
6	00/01	248	25.3	3.1	0.8	6.4	Paul Pierce	Celtics	6.0
7	00/01	405	24.9	2.0	0.3	8.7	Antawn Jamison	Wizards	7.0
8	00/01	14	23.4	5.5	0.6	8.9	Antoine Walker	Timberwolves	8.0
9	00/01	109	22.2	8.4	0.2	3.2	Stephon Marbury	Knicks	9.0
10	00/01	364	22.2	3.0	2.3	12.2	Tim Duncan	Spurs	9.0
11	01/02	587	27.2	3.0	2.0	10.7	Shaquille O'Neal	Suns	1.0
12	01/02	248	26.1	3.2	1.0	6.9	Paul Pierce	Celtics	2.0
13	01/02	349	25.6	5.3	1.0	7.9	Tracy McGrady	Rockets	3.0
14	01/02	364	25.5	3.7	2.5	12.7	Tim Duncan	Spurs	4.0
15	01/02	62	25.2	5.5	0.4	5.5	Kobe Bryant	Lakers	5.0
16	01/02	464	24.7	4.0	0.7	5.2	Vince Carter	Nets	6.0
17	01/02	450	23.4	2.4	1.0	9.9	Dirk Nowitzki	Mavericks	7.0
18	01/02	365	22.6	5.3	0.4	6.3	Michael Finley	Spurs	8.0
19	01/02	14	22.1	5.0	0.5	8.8	Antoine Walker	Timberwolves	9.0
20	01/02	240	21.8	3.9	0.3	4.5	Ray Allen	Celtics	10.0
21	02/03	349	32.1	5.5	0.8	6.5	Tracy McGrady	Rockets	1.0
22	02/03	62	30.0	5.9	0.8	6.9	Kobe Bryant	Lakers	2.0
23	02/03	587	27.5	3.1	2.4	11.1	Shaquille O'Neal	Suns	3.0
24	02/03	248	25.9	4.4	0.8	7.3	Paul Pierce	Celtics	4.0
25	02/03	450	25.1	3.0	1.0	9.9	Dirk Nowitzki	Mavericks	5.0
26	02/03	240	24.5	5.9	0.1	5.6	Ray Allen	Celtics	6.0
27	02/03	364	23.3	3.9	2.9	12.9	Tim Duncan	Spurs	7.0
28	02/03	245	23.0	6.0	1.6	13.4	Kevin Garnett	Celtics	8.0
29	02/03	605	22.9	7.5	0.4	4.3	Baron Davis	Warriors	9.0
30	02/03	109	22.3	8.1	0.2	3.2	Stephon Marbury	Knicks	10.0
31	02/03	324	9.2	1.0	1.4	6.9	Tyson Chandler	Hornets	1.0
32	03/04	349	28.0	5.5	0.6	6.0	Tracy McGrady	Rockets	1.0
33	03/04	245	24.2	5.0	2.2	13.9	Kevin Garnett	Celtics	2.0
34	03/04	330	24.2	2.1	0.2	6.3	Peja Stojakovic	Hornets	2.0
35	03/04	62	24.0	5.1	0.4	5.5	Kobe Bryant	Lakers	4.0

Figura 9. Previsualización de datos obtenidos para TOP 10 JUGADORES POR PUNTOS BY TEMPORADA.

Obtención de plantillas ganadoras

A continuación, vamos a obtener las plantillas de jugadores de los equipos que ganaron más partidos durante cada temporada. Veamos la lectura de datos, que sigue la estructura que se puede observar en la *figura 10*.

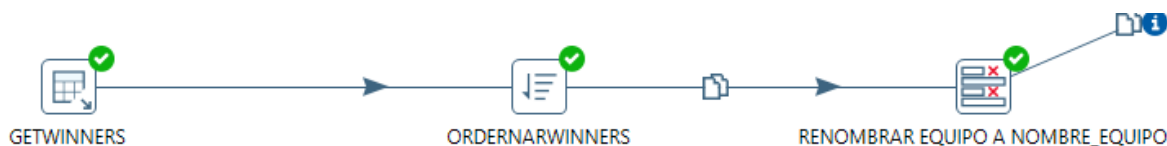


Figura 10. Lectura de datos de los equipos ganadores.

Como ya hemos explicado, la tabla partidos contiene los resultados de todos los partidos de cada temporada almacenados como los puntos anotados por cada uno de los equipos. Debido a esto, durante la lectura debemos crear un nuevo atributo que almacena el número de partidos que ha ganado cada equipo durante una temporada (que supere en puntos al equipo rival en el resultado) y quedarnos solamente con el equipo que más victorias tenga. En la *figura 11* podemos observar el proceso que hemos seguido para conseguirlo.

```

SELECT TEMPORADA, EQUIPO, VICTORIAS
FROM (
  SELECT TEMPORADA, GANADOR AS EQUIPO, COUNT(*) AS VICTORIAS,
    RANK() OVER (PARTITION BY TEMPORADA ORDER BY COUNT(*) DESC) AS RANK
  FROM (
    SELECT
      CODIGO,
      EQUIPO_LOCAL,
      EQUIPO_VISITANTE,
      PUNTOS_LOCAL,
      PUNTOS_VISITANTE,
      TEMPORADA,
      CASE
        WHEN PUNTOS_LOCAL > PUNTOS_VISITANTE THEN EQUIPO_LOCAL
        WHEN PUNTOS_LOCAL < PUNTOS_VISITANTE THEN EQUIPO_VISITANTE
        ELSE 'Empate'
      END AS GANADOR
    FROM SYSTEM.PARTIDOS
  ) T
  GROUP BY TEMPORADA, GANADOR
) T
WHERE RANK = 1

```

Figura 11. Obtención de los equipos con más victorias en cada temporada.

Además podemos ver una previsualización de la secuencia anterior en la figura 12.

#	TEMPORADA	EQUIPO	VICTORIAS
1	00/01	Pacers	68
2	01/02	Supersonics	69
3	02/03	Nuggets	71
4	03/04	Rockets	72
5	04/05	Wizards	67
6	05/06	Clippers	37
7	06/07	Pacers	68
8	07/08	Mavericks	37
9	07/08	Knicks	37
10	98/99	Pistons	68
11	99/00	Bulls	68

Figura 12. Datos obtenidos de la lectura.

A continuación, ordenamos las entradas obtenidas por temporada y cambiamos el nombre de la columna "EQUIPO" a "NOMBRE_EQUIPO" para poder realizar el siguiente paso, ilustrado en la figura 13.



Figura 13. Obtención de ganadores.

Como podemos observar, lo siguiente es hacer un “join” de los equipos ganadores con los datos de los jugadores obtenidos con anterioridad. Este paso nos crea algunas entradas erróneas en las tablas debido a que algunos jugadores cambian de equipo en las distintas temporadas, por lo que debemos filtrar el contenido y quedarnos sólo con las entradas en las que coincidan el nombre del equipo y la temporada en ambas tablas, jugadores y equipos vencedores. Véase en la figura 14.

#	TEMPORADA	JUGADOR	PUNTOS_POR_PARTIDO	ASISTENCIAS_POR_PARTIDO	TAPONES_POR_PARTIDO	REBOTES_POR_PARTIDO	PUNTOS_RANK	ASISTENCIAS_RANK	TAPONES_RANK	REBOTES_RANK	NOMBRE	NOMBRE_EQUIPO
1	00/01	524	3.5	0.5	0.4	5.5	110.0	112.0	60.0	42.0	Jeff Foster	Pacers
2	00/01	530	12.9	1.2	2.8	9.8	44.0	83.0	3.0	9.0	Jermaine O'Neal	Pacers
3	01/02	566	7.2	1.8	0.2	3.9	103.0	79.0	113.0	88.0	Adrian Griffin	Supersonics
4	01/02	567	14.8	1.7	1.2	7.6	39.0	84.0	26.0	23.0	Donyell Marshall	Supersonics
5	01/02	572	3.6	2.0	0.1	1.3	143.0	70.0	133.0	154.0	Earl Watson	Supersonics
6	02/03	482	3.3	2.7	0.1	1.5	164.0	63.0	150.0	164.0	Chuckie Atkins	Nuggets
7	02/03	483	7.6	1.6	1.4	7.2	107.0	96.0	19.0	26.0	Marcus Camby	Nuggets
8	02/03	484	4.1	4.1	0.1	1.7	153.0	29.0	150.0	162.0	Anthony Carter	Nuggets
9	02/03	487	3.9	0.2	1.1	2.8	156.0	179.0	25.0	132.0	Steven Hunter	Nuggets
10	02/03	490	16.7	2.4	0.9	8.3	39.0	72.0	38.0	16.0	Kenyon Martin	Nuggets
11	02/03	491	6.7	1.0	0.5	4.6	119.0	129.0	68.0	77.0	Eduardo Najera	Nuggets
12	02/03	492	10.5	1.9	0.8	6.1	78.0	84.0	40.0	50.0	Nene	Nuggets
13	03/04	340	10.2	4.5	0.2	2.8	101.0	34.0	131.0	158.0	Rafer ALston	Rockets
14	03/04	341	8.5	1.3	0.7	3.8	121.0	136.0	54.0	114.0	Shane Battier	Rockets
15	03/04	343	16.6	6.2	0.4	5.5	44.0	10.0	84.0	63.0	Steve Francis	Rockets
16	03/04	347	13.8	2.1	0.2	3.5	65.0	91.0	131.0	127.0	Bobby Jackson	Rockets
17	03/04	349	28.0	5.5	0.6	6.0	1.0	18.0	62.0	57.0	Tracy McGrady	Rockets
18	03/04	350	5.6	0.4	1.9	6.7	165.0	205.0	12.0	36.0	Dikembre Mutombo	Rockets
19	03/04	353	3.2	0.3	0.5	3.5	200.0	214.0	73.0	127.0	Loren Woods	Rockets
20	03/04	354	17.5	1.5	1.9	9.0	33.0	120.0	12.0	18.0	Yao Ming	Rockets
21	04/05	400	25.5	5.1	0.3	4.7	6.0	27.0	123.0	101.0	Gilbert Arenas	Wizards
22	04/05	402	15.5	1.9	0.3	5.8	56.0	109.0	123.0	72.0	Caron Butler	Wizards
23	04/05	403	11.2	4.1	0.0	2.3	102.0	40.0	244.0	201.0	Antonio Daniels	Wizards
24	04/05	404	9.4	0.8	1.7	6.8	137.0	197.0	17.0	48.0	Brendan Haywood	Wizards
25	04/05	405	19.6	2.3	0.2	7.6	28.0	85.0	156.0	35.0	Antawn Jamison	Wizards
26	04/05	409	7.5	1.4	0.2	4.2	160.0	150.0	156.0	115.0	Darius Songaila	Wizards
27	04/05	410	7.8	1.3	0.2	1.9	156.0	159.0	156.0	219.0	DeShawn Stevenson	Wizards
28	04/05	411	7.1	0.4	1.1	5.2	168.0	235.0	39.0	87.0	Etan Thomas	Wizards
29	05/06	21	24.7	2.6	2.5	10.0	10.0	91.0	6.0	9.0	Elton brand	Clippers
30	05/06	23	3.3	2.1	0.1	0.8	263.0	114.0	226.0	301.0	Dan Dickau	Clippers
31	05/06	25	11.9	1.0	1.4	9.6	104.0	189.0	24.0	12.0	Chris Kaman	Clippers
32	05/06	26	12.6	8.8	0.1	3.2	89.0	4.0	226.0	175.0	Brevin Knight	Clippers
33	05/06	27	4.5	5.8	0.5	3.0	208.0	38.0	90.0	186.0	Shawn Livingston	Clippers
34	05/06	28	17.8	2.1	0.1	5.3	38.0	114.0	236.0	75.0	Cory Maggette	Clippers
35	05/06	29	14.8	3.0	0.5	4.3	64.0	70.0	90.0	125.0	Cuttino Mobley	Clippers
36	05/06	30	11.5	3.7	0.2	3.3	110.0	56.0	173.0	169.0	Smush Parker	Clippers
37	05/06	31	3.0	0.2	0.1	2.2	269.0	289.0	226.0	224.0	Josh Powell	Clippers
38	05/06	32	4.7	1.2	0.2	2.5	237.0	172.0	173.0	211.0	Quinton Ross	Clippers
39	05/06	33	4.3	0.7	0.3	1.3	247.0	217.0	133.0	281.0	Tim Thomas	Clippers
40	06/07	520	7.1	1.3	0.2	1.8	196.0	177.0	187.0	281.0	Marquis Daniels	Pacers

Figura 14. Previsualización de ganadores.

Así pues, ya hemos obtenido las plantillas de los equipos que ganaron cada temporada, ahora procedemos a eliminar las columnas "NOMBRE_EQUIPO_1" y "TEMPORADA_1" para evitar tener datos repetidos y podemos guardarlas en una nueva tabla en nuestra base de datos, para así poder usar estos datos para futuros análisis.

Conclusión

Para finalizar, vamos a hacer un repaso de los resultados que hemos obtenido, comentar las dificultades que presentaba el proceso de conseguirlas y los posibles usos que se le pueden dar a estos.

En primer lugar, hemos obtenido un ranking de los jugadores que más puntos han anotado, los que más asistencias hicieron, los que más rebotes cogieron y los que más tapones realizaron por temporada. Para poder obtener estos datos, tuvimos que limpiar algunas entradas de nuestra base de datos debido a los errores que contenían, asignar posiciones en el ranking de cada temporada a cada uno de los jugadores y unir esos datos a los obtenidos de la tabla jugadores para poder mostrar el nombre. Estos datos son cruciales para ver los jugadores con mejor rendimiento y también su evolución a lo largo de los años, permitiendo a los equipos saber que jugadores fichar o adaptar su estrategia según las habilidades del jugador.

Por otro lado, hemos obtenido las plantillas de los equipos que dominaron en cada una de las temporadas. Durante la obtención de estos datos tuvimos que calcular el número de victorias de cada equipo en cada temporada con una consulta bastante compleja, unirlo con los datos de los jugadores que obtuvimos con anterioridad y limpiar la información de datos erróneos y repetidos. El poder saber qué plantillas y alineaciones, junto con las estadísticas de los jugadores, han funcionado con anterioridad le permite a los equipos analizar y razonar por qué triunfaron y poder crear nuevas estrategias en torno

al rendimiento que consiguen los distintos jugadores cuando trabajan en conjunto con otros.

Por último, estos datos también pueden ser utilizados para el entrenamiento de IAs que puedan predecir futuros resultados de partido, una herramienta útil tanto para los equipos de la liga como para medios de comunicación, espectadores o apuestas.

Sería importante mencionar que se ha realizado el trabajo buscando la auditoría de calidad en él. Se ha realizado una normalización de datos óptimas, y las pruebas de carga, rendimiento eficacia y estrés han sido buenas, se ha conseguido desde el punto de vista estadístico buenos números en estos campos.

Referencias

- (1) hitachivantara.com. 2023. Pentaho Data Integration - Hitachi Vantara Lumada and Pentaho Documentation. Recuperado el 5 de mayo de 2023 de :
<https://help.hitachivantara.com/Documentation/Pentaho/8.3/Products/Pentaho_Data_Integration>
- (2) oracle.com. 2023. Introduction to Oracle Database. Recuperado el 5 de mayo de 2023 de:
<<https://docs.oracle.com/en/database/oracle/oracle-database/19/cncpt/introduction-to-oracle-database.html#GUID-A42A6EF0-20F8-4F4B-AFF7-09C100AE581E>>
- (3) discoduroderoer.es. 2022. Ejercicios propuestos y resueltos consultas SQL (NBA) (discoduroderoer.es). Recuperado el 5 de mayo de 2023 de:
<<https://www.discoduroderoer.es/ejercicios-propuestos-y-resueltos-consultas-sql-nba/>>

Bibliografía

1. IAs consultadas: ChatGPT, BING AI & Github Copilot
2. Ayuda para setup:
<https://www.forosdelweb.com/f100/no-puedo-conectar-sql-developer-604017/>

<https://gobansaor.wordpress.com/2006/12/18/sqlite-jdbc-and-kettle-pentaho-data-integration-etl/>

https://asktom.oracle.com/pls/apex/f?p=100:11:::P11_QUESTION_ID:9542646700346383126

3. Búsquedas puntuales para resolución de dudas y problemas:

https://mundodb.es/funciones-analiticas-de-ordenacion-en-oracle-row_number-rank-y-dense_rank

https://www.youtube.com/watch?v=eTs7sVYlts&list=PLPgjON4ZM0JAjcvkWW63eQeoU3Dre2E-k&ab_channel=LEARNING-BI

Apéndice

1. Github: https://github.com/romflorod/CBD_PENTAHO_DATAWAREHOUSE_NBA.git