



Trabajo Práctico N°1

18 de octubre de 2024

Laboratorio de Datos

Grupo 9

Integrante	LU	Correo electrónico
Busso, Tiago	570/23	bussotiago@gmail.com
Dell 'Aringa, Francisco	18/18	fran.dellarlinga@hotmail.com
Tissera, Lucas Hernán	12/17	lucas.tissera979@gmail.com



1. Diagrama Entidad-Relación

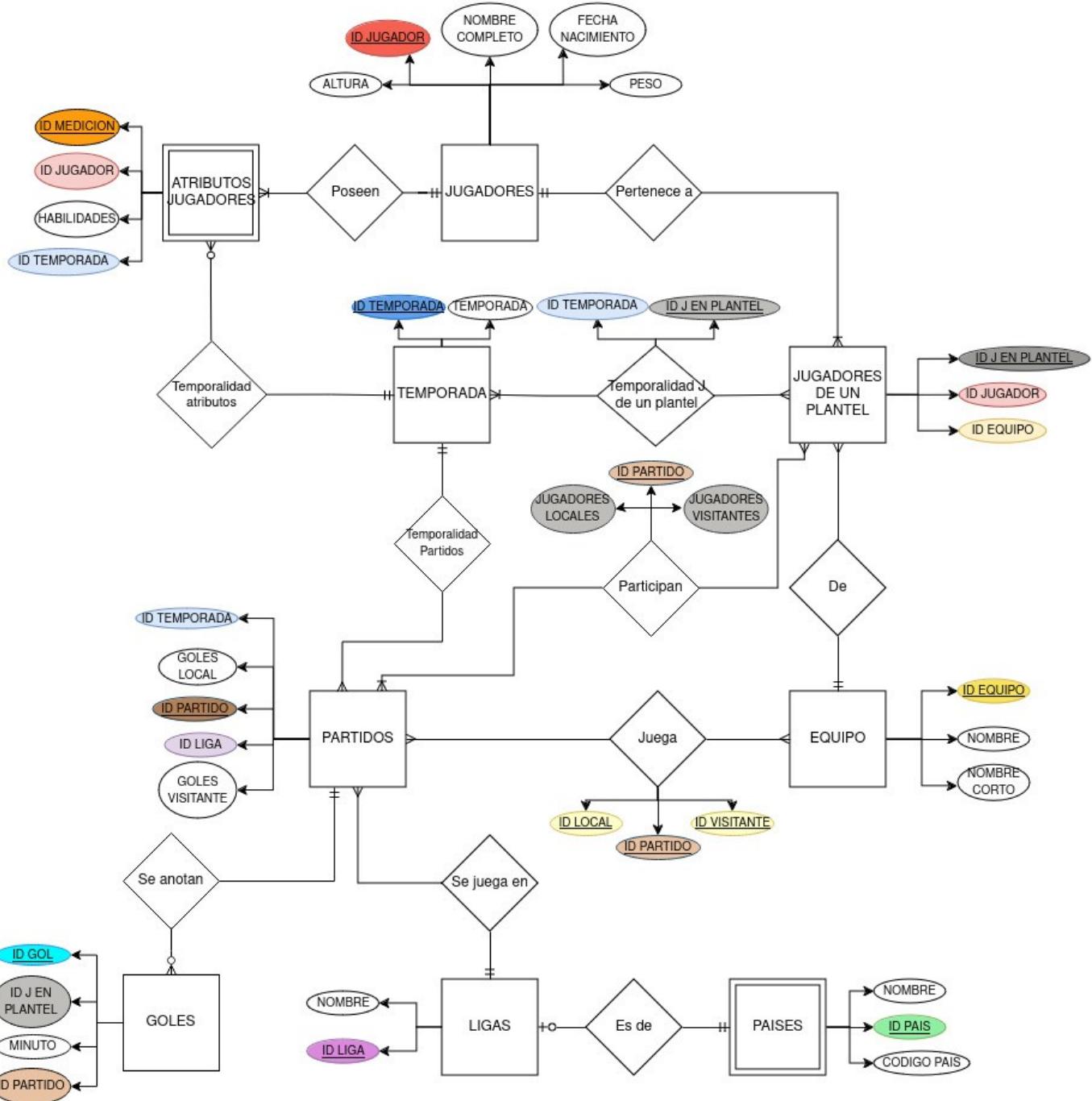


Figura 1: Diagrama entidad-relación del modelo solicitado.

Aclaración: Los atributos pintados de color suave son claves foráneas de los de color fuerte.

Justificación de decisiones tomadas

1.0.1. Justificación de entidades, claves y atributos

Agregaremos las entidades que creemos que son necesarias de explicar.

Entidad Atributos Jugadores: Agregamos un id medición que hace único a cada evaluación de atributos de cualquier jugador. Luego reemplazamos la fecha por el id de temporada ya que a la hora de chequear 3FN llegamos a la siguiente conclusión: como la relación de atributos jugadores con temporada es de (1:N), la misma debe realizarse agregando una clave foránea en atributos. Esta clave foránea es id temporada, pero el problema es que el id temporada dependería funcionalmente del atributo fecha, puesto en los csv originales, el cual no es primo, por lo tanto no se respetaría 2FN. Además, el atributo Habilidades es el conjunto de todos los puntajes realizados en la medición, y está notado así por conveniencia visual.

Entidad Jugadores: Está puesto el atributo Nombre compuesto ya que en el csv no están diferenciados, no figuran como atómicos.

Jugadores de un Plantel: El id J en plantel es el id que tiene un jugador dentro de un plantel. Luego este id está asociado a un jugador mediante la relación Pertenece a la cual nos establece la clave foránea de id jugador en nuestra entidad. También está asociada a un equipo con la misma idea a través de la relación De. Luego al id de J en plantel se le asocia una temporada específica para asignar en que temporadas ese jugador estuvo en el plantel.

Entidad Temporada: Se crea una entidad con todas las temporadas existentes en los registros y le asignamos un id a cada una.

Entidad Partidos: Así como en Atributos Jugadores, la fecha es reemplazada con id temporada para poder estar en 3FN.

Entidad Liga: Cambiamos el id país por id liga, solamente el nombre el id es el mismo. Luego no almacenamos acá el código país como en los csv ya que nos parece irrelevante. Asumimos que el nombre de la liga no es único, ya que sino id liga dependería funcionalmente del nombre, no respetando 3FN.

Entidad País: Id país es igual que el id liga por lo que actúa como clave foránea. Asumimos que código de país no es único para preservar 3FN al igual que en Liga. No vemos que sea conveniente tener una clave compuesta por id país y código país ya que no esclarece las cosas.

Entidad Goles: Esta entidad es la columna de Goles de el csv de partidos. A la hora de descomprimir la columna, vimos que era un gran string con información sobre los goles, lo cual hacia que sea un atributo multivalorado. Esto haría que no este en 1FN, por lo que comenzamos a desglosarlo. Luego nos quedamos solo con la información que creemos que es relevante tanto para nuestra BD como para las respuestas de las preguntas. Ademas chequeamos que este en 3FN la nueva tabla con los atributos seleccionados. Notemos que queda implícita una relación entre goles y jugadores de un plantel. Por un tema de utilidad y legibilidad de nuestro DER decidimos no graficarla, ya que no nos aporta nada para el análisis esa relación y ademas entorpesería visualmente.

1.0.2. Justificación de Relaciones

- **Poseen (N:1):** Consideramos que un jugador tiene que tener sí o sí una medición de atributos como mínima, siendo en el caso de que cuando se inscribe un nuevo jugador se le miden sus atributos. Luego, un jugador con el paso del tiempo puede tener varias reevaluaciones de atributos. Los atributos de un jugador están asignados solamente a 1 jugador de modo estricto.
- **Pertenece a (N:1):** Consideramos que los jugadores pertenecen a los jugadores que conforman un plantel o a muchos, ya que como aclaramos anteriormente esta entidad nos genera una relación entre los jugadores y un equipo en una temporada correspondiente, por lo que puede pertenecer a muchos ya que los id de plantel de sus equipos anteriores se mantienen. Luego jugadores de un plantel corresponde solo a un jugador.
- **De (1:N):** Los jugadores de un plantel pertenecen inequívocamente a un solo equipo. Un equipo tiene sí o sí muchos Jugadores de un plantel ya que sino no tendría la posibilidad de disputar un partido.
- **Juega (N:M):** Un equipo juega sí o sí muchos partidos y un partido es jugado por dos equipos por lo que son muchos equipos.
- **Se juega en (1:N):** Los partidos se juegan en 1 sola liga. Luego, una liga tiene sí o sí muchos partidos.

- **Es de** (1:1): Las ligas están asignadas a un país específico, de hecho más aún el id de liga es igual al id de país al cual está asignada. Consideramos que un país puede tener una liga o no tenerla, dejamos fuera de lado la posibilidad de tener muchas ligas.
- **Participan** (N:M): En un partido participan sí o sí muchos jugadores de un plantel. Los jugadores de un plantel juegan 1 o muchos partidos. Consideramos que como mínimo juegan un partido pues en la consigna dice “En cada temporada, un equipo tiene una serie de jugadores asociados que son quienes juegan los partidos, llamaremos a esta agrupación plantel”, por lo que para ser jugador de un plantel se debe haber jugado como mínimo un partido. A su vez, el único registro de relación de jugadores con equipo en los csv son en los partidos. De jugadores local y jugadores visitante se desprenden 11 jugadores locales y 11 visitantes.
- **Temporalidad partidos** (1:N): Un partido tiene solamente una temporada asignada. Además, una temporada tiene sí o sí muchos partidos.
- **Tiene** (N:M): Una temporada tiene muchos jugadores de un plantel, ya que son todos los jugadores que pertenecen a algún plantel en dicha temporada. Luego, los jugadores de un plantel pertenecen a una temporada o a muchas, ya que si el jugador acaba de llegar a un equipo su id plantel nuevo pertenece solo a una temporada y con el correr del tiempo pertenecerá a más.
- **Temporalidad atributos** (1:N): Una temporada tiene ninguno o muchos atributos de jugador, ya que son todos los atributos de jugadores que se midieron en dicha temporada. Puede suceder que no se hayan tomado atributos a ningún jugador en dicha temporada, que se le haya tomado solo a uno o a muchos. Luego, los atributos de un jugador son medidos en una temporada específica.
- **Se anotan** (1:N): Esta relación se desglosa de la columna de goles de partidos. En un partido se anotan ninguno o muchos goles. En cambio un gol está asociado específicamente a 1 partido.

2. Modelo Relacional

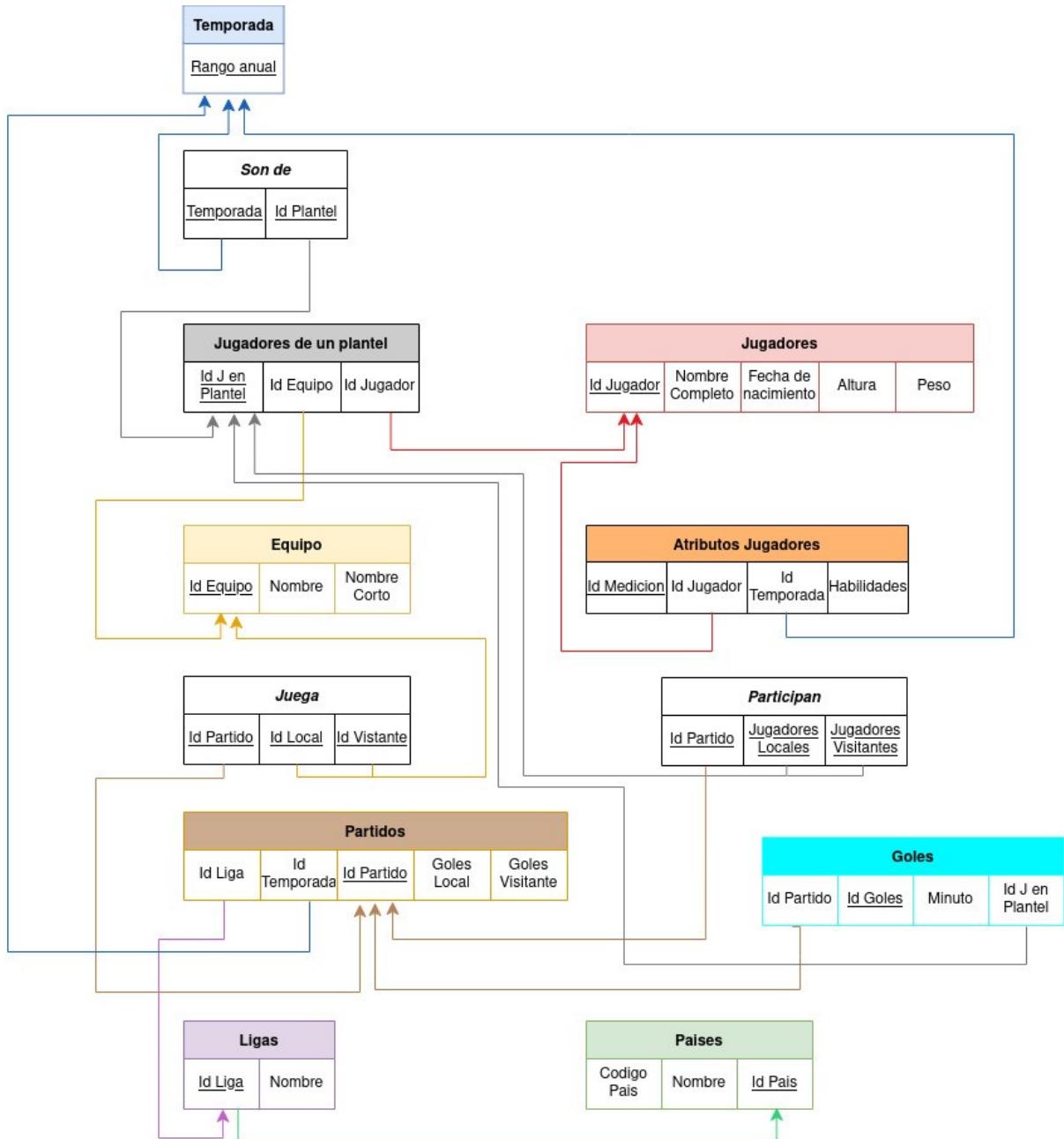


Figura 2: Modelo relacional solicitado.

Aclaración: Las tablas pintadas de color son entidades con el color respectivo de su clave en el DER. Las que no tienen color y tienen letra en itálica son relaciones.

3. Análisis Exploratorio

Nosotros tenemos asignado el país España, por lo que analizaremos la liga española. Más precisamente analizaremos la era de Pep Guardiola en el FC Barcelona, la cual marcó una época dorada para el club. Su dirigencia comienza en la temporada 2008-2009 y se mantuvo al frente del equipo hasta la temporada 2011-2012. En estos años el FC Barcelona ganó 24 títulos logrando conseguir hasta el famoso “Sextete”. En cuanto a las ligas que es lo incluido en nuestras bases de datos ganó 3 de las 4 jugadas, quedando segundo en la 2011/2012. Como si fuera poco bajo este rango de temporadas, Lionel Messi ganó sus primeros 4 balones de oro de forma consecutiva, de 2009 a 2012.

3.1. Consultas SQL

Realizar al menos las siguientes consultas SQL y dar su resultado.

- Sobre equipos del país y la liga asignada:
 - ¿Cuál es el equipo con mayor cantidad de partidos ganados?

Id Equipo	Nombre	Nombre Corto
8633	Real Madrid CF	REA

Nota: Ver que el Real Madrid es el equipo que más partidos ganó a lo largo de las 4 temporadas y aún así solo ganó 1 de ellas.

- ¿Cuál es el equipo con mayor cantidad de partidos perdidos de cada año?

Cantidad de Partidos Perdidos	Nombre	Temporada
23	Real Sporting de Gijón	2008/2009
23	CD Numancia	2008/2009
20	Xerez Club Deportivo	2009/2010
20	CD Tenerife	2009/2010
21	Hércules Club de Fútbol	2010/2011
21	Real Sociedad	2010/2011
21	Real Sporting de Gijón	2011/2012
21	Rayo Vallecano	2011/2012

Nota: Ver que en las 4 temporadas hay 2 equipos con mayor cantidad de partidos perdidos.

- ¿Cuál es el equipo con mayor cantidad de partidos empatabados en el último año?

Nombre	Nombre Corto	Temporada
CA Osasuna	OSA	2011/2012

- ¿Cuál es el equipo con mayor cantidad de goles a favor?

Nombre	Goles a Favor
FC Barcelona	412

- ¿Cuál es el equipo con mayor diferencia de goles?

Nombre	Diferencia de Goles
FC Barcelona	303

- ¿Cuántos jugadores tuvo durante el período de tiempo seleccionado cada equipo en su plantel?

Id Equipo	Nombre	Cantidad de Jugadores
8302	Sevilla FC	90
8581	Levante UD	73
8305	Getafe CF	73
9906	Atlético Madrid	73
10267	Valencia CF	73
8634	FC Barcelona	55
8603	Real Betis Balompié	83
8558	RCD Espanyol	83
9865	UD Almería	59
8370	Rayo Vallecano	69
10281	Real Valladolid	63
8696	Racing Santander	57
9867	CD Tenerife	19
10278	Hércules Club de Fútbol	23
8661	RCD Mallorca	60
9864	Málaga CF	93
9869	Real Sporting de Gijón	55
8633	Real Madrid CF	63
9783	RC Deportivo de La Coruña	84
8371	CA Osasuna	60
10205	Villarreal CF	73
8315	Athletic Club de Bilbao	49
8394	Real Zaragoza	66
8560	Real Sociedad	41
8388	CD Numancia	21
9868	Xerez Club Deportivo	20
7878	Granada CF	69
8479	RC Recreativo	21

- Sobre jugadores del país y la liga asignada:

- ¿Cuál es el jugador con mayor cantidad de goles?

Cantidad de Goles	Nombre
144	Lionel Messi

Nota: Quien iba ser sino, el mejor de todos los tiempos y más en específico el mejor jugador reconocido por la FIFA en estas 4 temporadas.

- ¿Cuáles son los jugadores que más partidos ganó su equipo?

Id Jugador	Nombre	Cantidad de Partidos Ganados
25921	Pepe	117
30962	Sergio Ramos	117
28467	Marcelo	117
23988	Lassana Diarra	117
30657	Iker Casillas	117
25759	Gonzalo Higuaín	117

Nota: Los 6 jugadores pertenecen al Real Madrid, lo cual tiene sentido dado que es la respuesta de la primera consulta.

- ¿Cuál es el jugador que estuvo en más equipos?

Id Jugador	Nombre	Cantidad de Equipos
40501	Adrián González	5
102572	Adrián Colunga	5
154256	Xavier Torres	5
38044	Ikechukwu Uche	5
96623	Sergio Asenjo	5
2805	Ruben Perez	5
74435	Domingo Cisma	5

Nota: Si bien el período de tiempo elegido es de 4 temporadas y se consideró que los planteles no varían a lo largo de cada temporada, se cree que estos jugadores estuvieron en 5 equipos porque la base de datos contempla los mercados de pase de verano y de invierno o bien porque alguno figura en dos equipos por estar a préstamo.

- ¿Cuál es el jugador que menor variación de potencia ha tenido a lo largo de los años? (medida en valor absoluto)

Id Jugador	Nombre	Fecha de Nacimiento	Altura	Peso
30981	Lionel Messi	1987-06-24 00:00:00	170.18	159

Nota: Tiene sentido, fueron los mejores años de Lionel Messi en el Barcelona.

- Joíneate algo (optativa, ver preguntas propias a continuación):

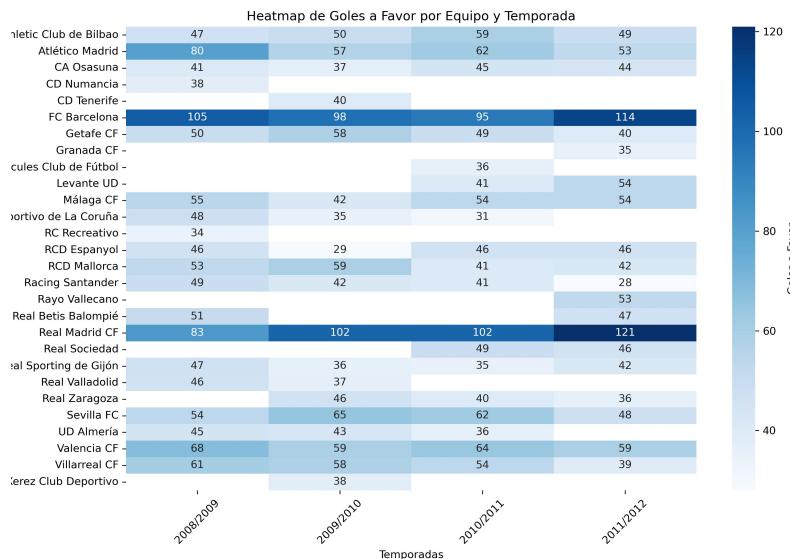
- ¿Hay algún equipo que haya sido a la vez el más goleador y el que tenga mayor valor de alguno de los atributos (considerando la suma de todos los jugadores)?

Ya que todas las tablas creadas están vacías, no hay relación entre ser el mas goleador y la sumatoria de los atributos de los jugadores.

3.2. Visualizaciones

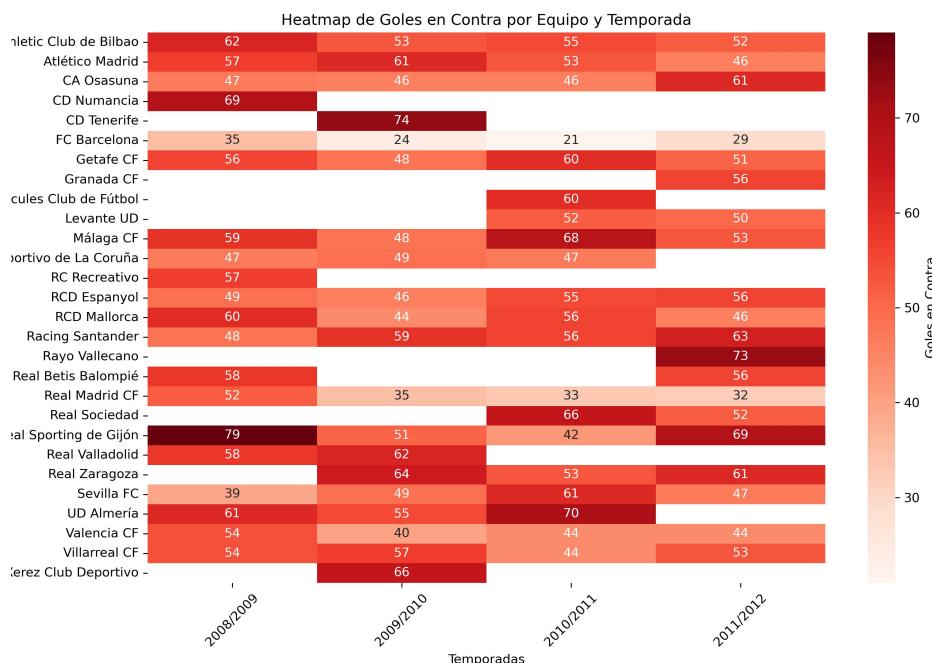
Notar que en los heatmaps las celdas sin valores son debido a que algunos equipos ascendieron o descendieron de categoria. Estos analisis son sobre la 1era.

- Graficar la cantidad de goles a favor y en contra de cada equipo a lo largo de los años que elijan.



(a) Grafico de goles a favor de cada equipo de la liga a lo largo de las temporadas 2008 a 2012.

(b) Nota: Podemos observar como los de mayor cantidad de goles a favor son el Fc barcelona y el Real Madrid, los cuales disputaron siempre 1er y 2do puesto.



(c) Grafico de goles encontra de cada equipo de la liga a lo largo de las temporadas 2008 a 2012

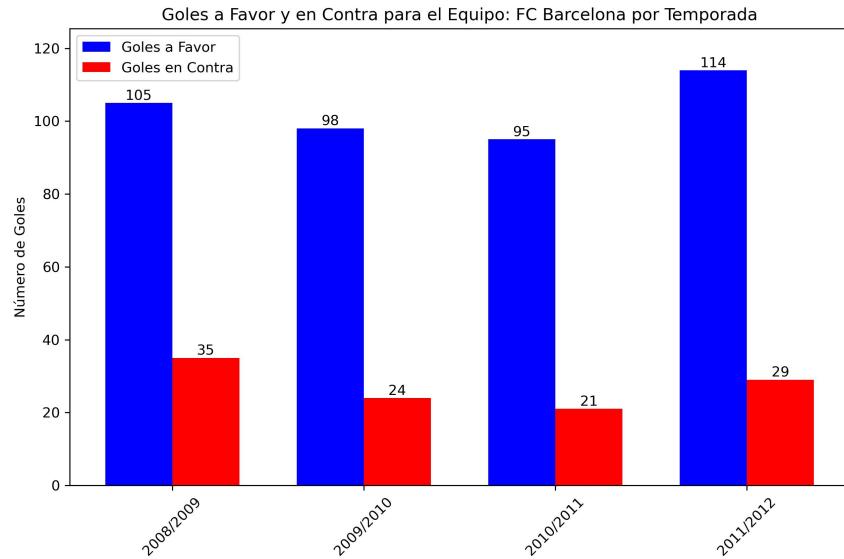
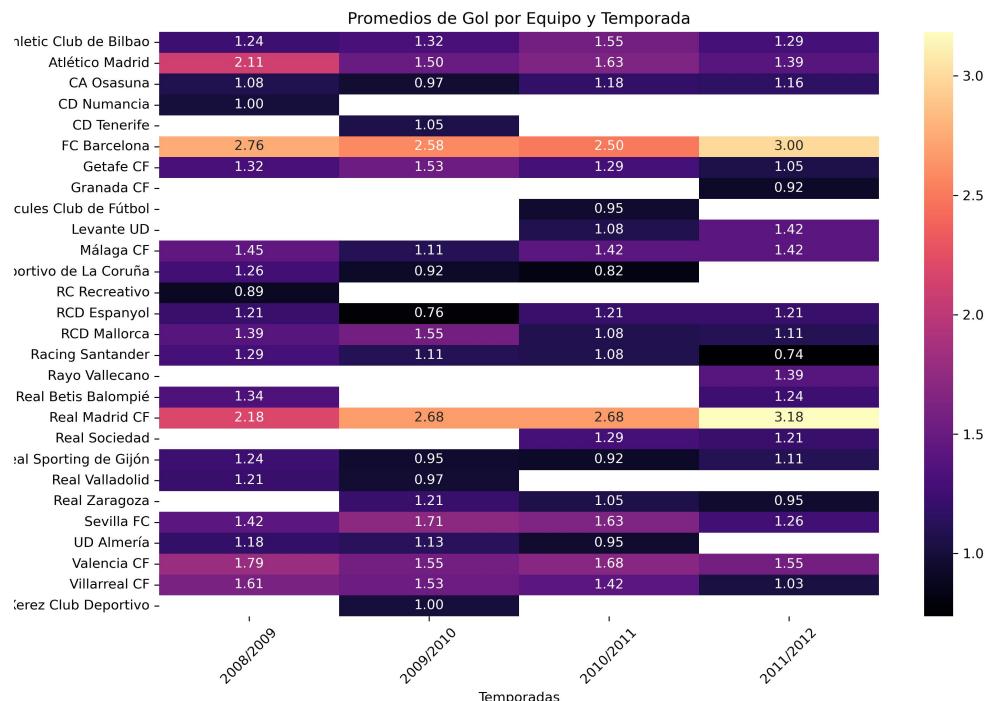


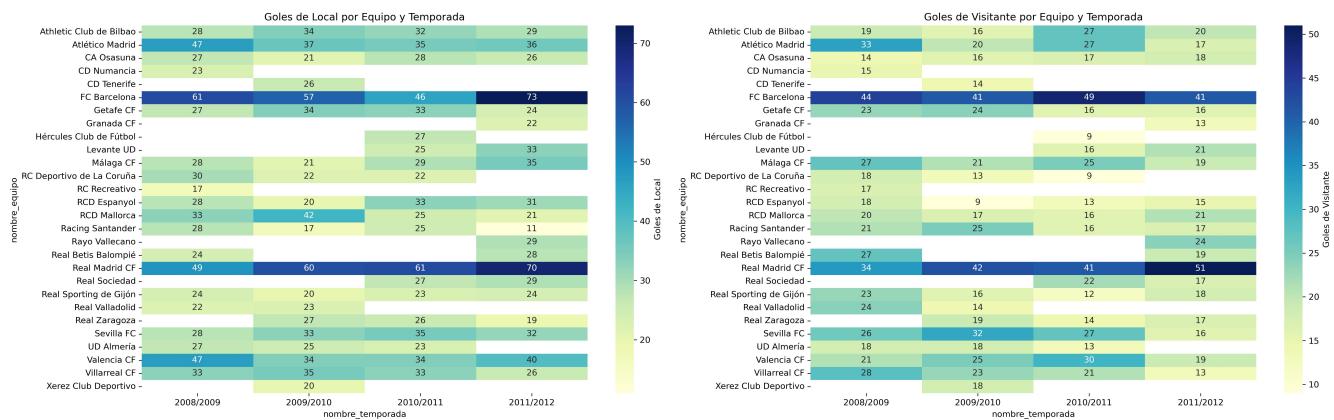
Figura 4: Generamos un grafico para poder analizar a los equipos en individual. En este caso podemos ver los numero del Fc Barcelona.

- Graficar el promedio de gol de los equipos a lo largo de los años que elijan.

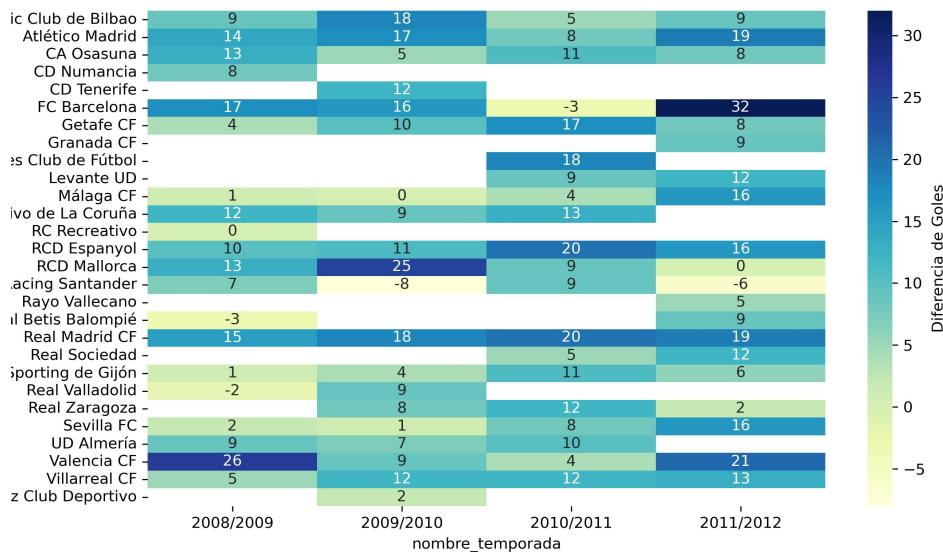


(a) Grafico del promedio de gol por partido de cada equipo en cada una de las temporadas 2008 a 2012

- Graficar la diferencia de goles convertidos jugando de local vs visitante a lo largo del tiempo.



(a) Grafico de goles convertidos como local y visitante de cada equipo de la liga a lo largo de las temporadas 2008 a 2012



(b) Grafico de la diferencia de goles de local vs goles de visitante de cada equipo de la liga a lo largo de las temporadas 2008 a 2012. (Los valores negativos indican que en esa temporada anotó |value| goles de visitante mas que de local)

- Graficar el número de goles convertidos por cada equipo en función de la suma de todos sus atributos.

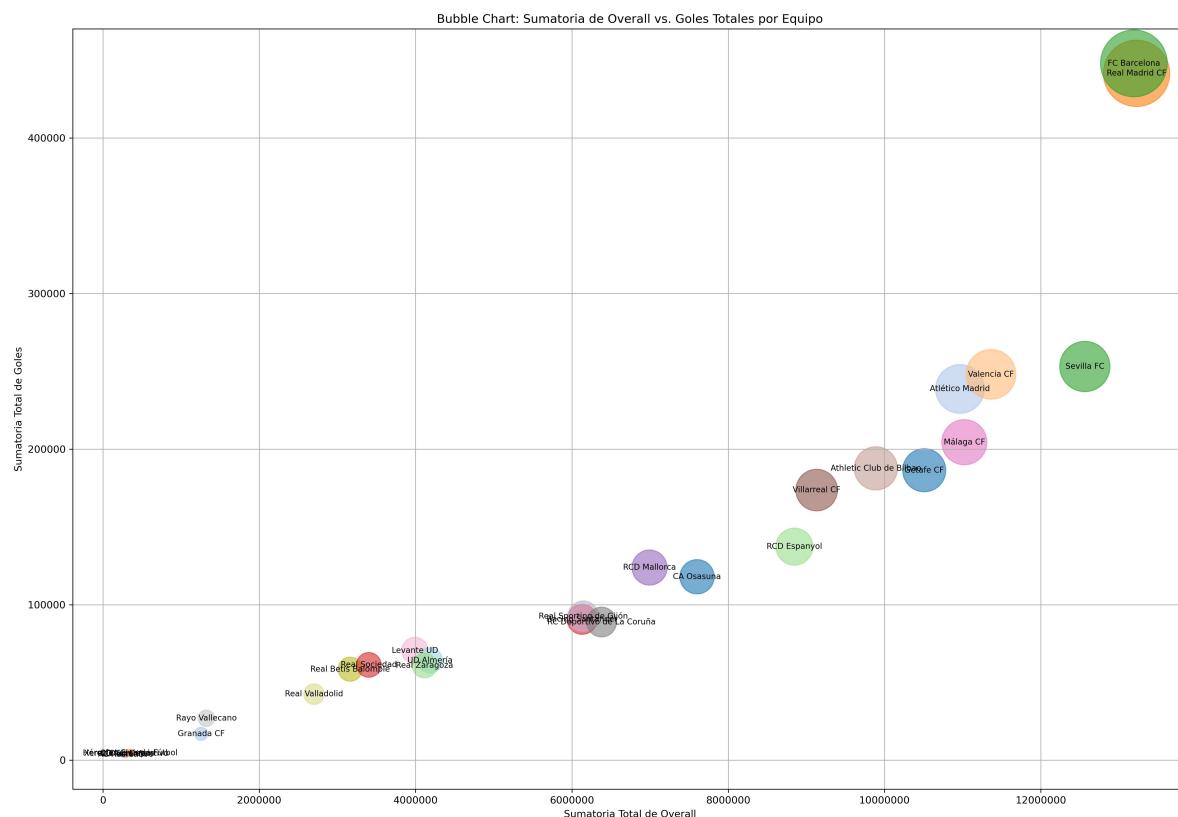


Figura 7: Grafico del numero de goles convertidos en estas 4 temporadas en funcion de la sumatoria de overall de todos los jugadores de cada equipo.