

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

El conjunto de datos seleccionado y del que hemos recolectado la información consta de estadísticas recogidas de los jugadores de la liga de fútbol en la temporada 2020-2021.

El sitio web del que hemos extraído dicha información la proporciona con el fin de dar a conocer las estadísticas de los diferentes jugadores y mostrar su rendimiento a lo largo de las diferentes temporadas.

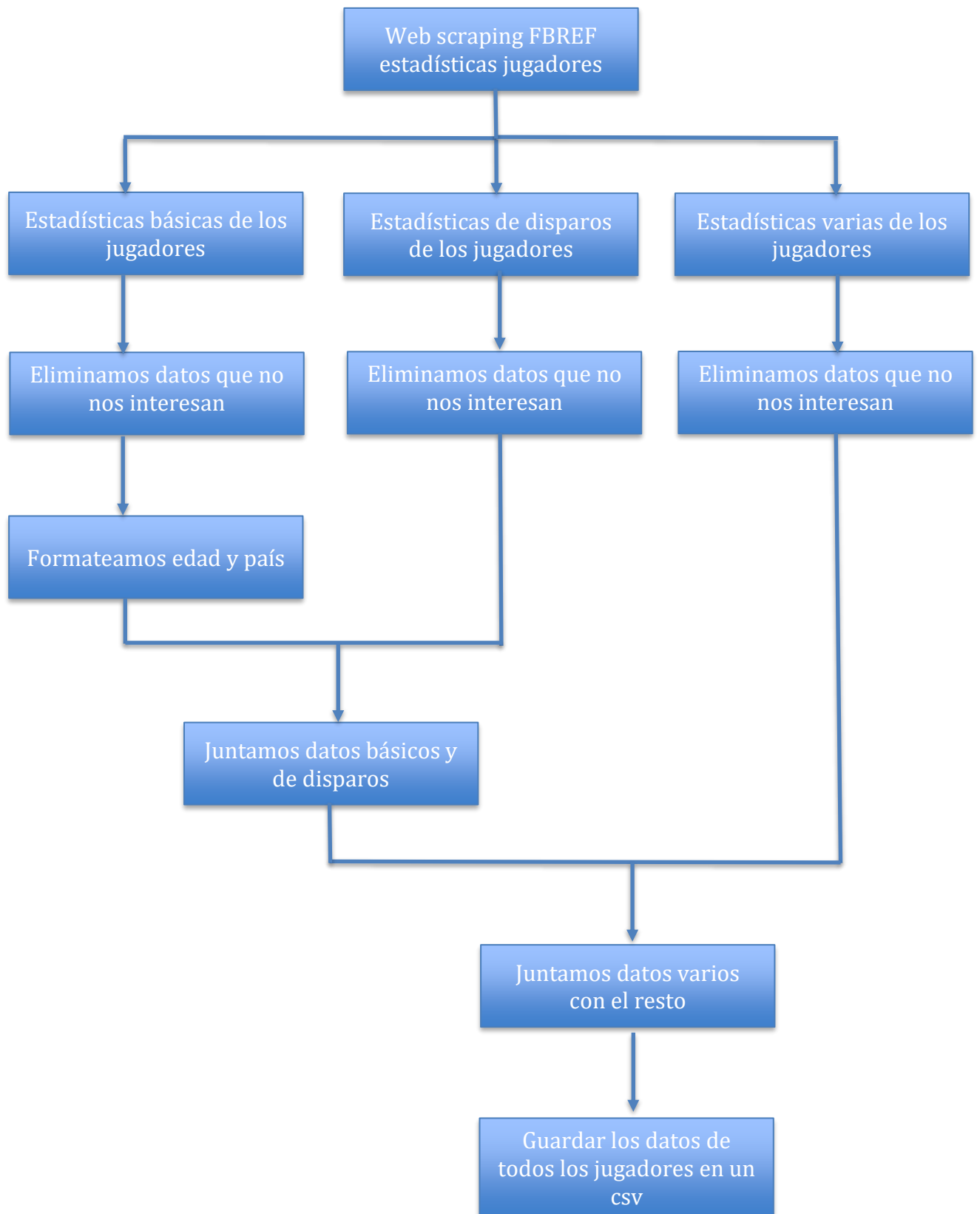
2. Definir un título para el dataset. Elegir un título que sea descriptivo.

El título para el dataset es 'estadisticas_jugadores_lfp_20_21'.

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

Como se ha mencionado, el dataset contiene las estadísticas referentes a los diferentes jugadores de la liga de fútbol. Contiene así información personal de cada uno de estos, así como datos referentes a su juego y los partidos jugados. Posteriormente se explicarán los diferentes campos que presenta.

4. Representación gráfica. Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido



5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

Se han extraído 43 columnas de datos:

Nombre del campo	Significado del campo
RL	Id del jugador y sus estadísticas
Jugador	Nombre del jugador
País	Nacionalidad del jugador
Posc	Posición del jugador, puede ser: <ul style="list-style-type: none">• PO - Portero• DF - Defensores• CC - Centrocampistas• DL – Delanteros• LT - Defensores laterales• LI - Lateral izquierdo• LD - Lateral derecho• DC - Defensa central• MCD - Mediocampistas defensivos• MC - Mediocampistas centrales• CI - Centrocampistas izquierdos• CD - Centrocampistas derechos• INT - Volantes• VI - Volantes izquierdos• VD - Volantes derechos• MCO - Centrocampistas de ataque
Equipo	Equipo donde juega esta temporada el jugador
Edad	Edad actual del jugador
Nacimiento	Año de nacimiento del jugador
PJ	Partidos que ha jugado el jugador
Titular	Partidos que el jugador a iniciado como titular
Min	Minutos que ha jugado esta temporada
90 s	Minutos jugados divididos por 90 min
Gls.	Goles marcados por el jugador
Ass	Asistencias del jugador
G-TP	Goles sin contar los penaltis

TP	Tiros de penalti marcados
TPint	Tiros de penalti
TA	Tarjetas amarillas totales
TR	Tarjetas rojas totales
Gls/90	Goles marcados por 90 minutos
Ast	Asistencias por 90 minutos
G+A	Goles y asistencias por 90 minutos
G-TP	Goles menos penaltis ejecutados por cada 90 minutos
G+A-TP	Goles más asistencias menos tiros de penaltis ejecutados por 90 minutos
Dis	Disparos realizados totales
DaP	Disparos a puerta sin contar penaltis
%% de TT	Porcentaje de tiros a puerta (sin incluir penaltis)
T/90	Tiros a puerta cada 90 minutos
TalArc/90	Tiros a puerta por cada 90 min sin contar penaltis
G/T	Goles por tiro a puerta
G/TalArc	Goles por tiro a puerta sin contar los penaltis
Dist	Distancia promedio en metros, del total de disparos realizados hacia portería
FK	Tiros a puerta hechos por la ejecución de una falta
2ª amarilla	Número de segundas tarjetas amarillas que ha recibido en un mismo partido
Fls	Faltas cometidas
FR	Faltas recibidas
PA	Fueras de juego
Int	Pases interceptados
Penal concedido	Penales realizados

GC	Goles en contra
Recup.	Número de balones sueltos recuperados
Ganados	Duelos aéreos ganados
Perdidos	Duelos aéreos perdidos
% de ganados	Porcentaje de balones aéreos ganados

Este dataset tiene como período de validez desde el 01/09/2020 hasta el día 09/04/2021, ya que no tiene en cuenta la jornada de esta semana ni las restantes hasta la finalización de la temporada.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.

Los datos de este dataset han sido obtenidos mediante la base de datos de la [FBREF](#). Se ha hecho uso de un script realizado con Python y técnicas de Web Scraping para extraer la información alojada en la página.

No se han encontrado análisis anteriores, ni haciendo uso de los datos de esta web, pero se han encontrado análisis haciendo uso de otras webs y clasificando los datos por equipos en vez de por jugadores, por lo que las estadísticas se basan en un ámbito distinto. Se puede comprobar en el siguiente [enlace](#).

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Este conjunto de datos puede llegar a utilizarse para muchas funciones distintas. Al ser datos de estadísticas de jugadores podría llegar a usarse como información para comentaristas deportivos, ya sea para analizar los jugadores de un equipo o como información destacada de los jugadores a lo largo de la temporada.

También puede darse un uso para los ojeadores, los cuales les facilitaría la tarea para ver quien está obteniendo buenos resultados durante la temporada en la posición que quieren reforzar para así tantear posibilidades de fichajes y hacer ofertas.

Otro uso distinto sería el del ocio. Se podría usar como base de datos tanto para videojuegos (juegos de management de equipos y fichajes) como para casas de apuestas, para ver quien tiene más probabilidades de hacer acciones determinadas.

Por último, podría usarse por los propios equipos como modelo para hacer un análisis de los puntos a reforzar y entrenar.

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

- **Released Under CC0: Public Domain License**
- **Released Under CC BY-NC-SA 4.0 License**
- **Released Under CC BY-SA 4.0 License**
- **Database released under Open Database License, individual contents under Database Contents License**
- **Other (specified above)**
- **Unknown License**

La licencia que se ha seleccionado para nuestro dataset es la **CC BY-SA 4.0 License**. Se ha seleccionado esta licencia debido a que las condiciones contractuales que nos aporta la licencia son adecuadas para nuestro dataset. Estas condiciones son:

- *Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado. Se debe de reconocer siempre el trabajo ajeno a medida que se va actualizando el dataset con nuevas aportaciones.*
- *Se permite un uso comercial. Si se permite un uso comercial, permitiría que empresas realizaran trabajos en base a este dataset, haciendo que mejore en calidad de datos y en actualidad de los mismos para poder sacar un mejor provecho del dataset, añadiéndole calidad a los datos.*
- *Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma. Esto hace que el trabajo del autor original continúe distribuyéndose bajo los términos que lo planteamos.*

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El código con el que se ha realizado la extracción de los datos y el dataset generado se puede acceder mediante el siguiente [enlace](#).

10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

Se ha publicado el dataset de las estadísticas de los jugadores de la lfp de la temporada 20/21, con los datos actualizados hasta el día 09/04/2021. El enlace DOI para el dataset es el siguiente:

<https://zenodo.org/record/4674631>