PRAC 1. Web Scraping

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

   El trabajo realizado corresponde al equipo de *scouting* de un equipo de la primera división de la liga profesional de fútbol de España, que desea tener unas estadísticas generales de todos los jugadores de la segunda división de la liga profesional de fútbol de la última temporada (2017 – 2018) para planificar su temporada 2018 – 2019.

   Para obtener estos datos, se ha recurrido a las estadísticas históricas de *LaLiga*, en su web: https://www.laliga.es/estadisticas-historicas/plantillas/segunda/2017-18/, pues es el organismo que regula el fútbol profesional en España y, por ello, es la información más fidedigna al recoger estos datos oficiales de las actas arbitrales.

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

   El nombre del dataset donde que recopila los datos es **LaLiga123_17-18_stats**.

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).

   El dataset generado por el código implementado se guarda en un *.csv* que contiene las estadísticas por jugador más representativas referentes al juego durante una temporada completa, 2017 - 2018. Esta información es almacenada en 11 columnas, donde cada fila hace referencia a un jugador.

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.

   Representación gráfica de las 10 primeras filas del dataset **LaLiga123_17-18_stats**.

```
> head(finaldata, 10)
        Jugador P.Jug. P.Compl. P.Tit. P.Sust.  Min Tarj. Expul. Goles Penalti   Equipo
1     ALVARO G.     29       18     24       5 2157     6      1     6       0 alcorcon
2   ALVARO PEÑA     40       14     34       6 2963     6      0     6       1 alcorcon
3      ASDRUBAL     11        1      5       6  429     5      0     1       0 alcorcon
4       BELLVIS     28       25     25       3 2304     6      0     1       0 alcorcon
5      BORJA D.     22        3     10      12  925     3      0     0       0 alcorcon
6  BORJA LAZARO      7        1      2       5  261     2      0     0       0 alcorcon
7    BRUNO GAMA     15        8     12       3 1115     1      0     1       0 alcorcon
8    BURGOS. E.     31       26     28       3 2565     8      0     1       0 alcorcon
9      CASTO E.     39       37     39       0 3427     6      1     0       0 alcorcon
10     CESAR S.     17       15     17       0 1449     2      0     0       0 alcorcon
```

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

   Los atributos reflejados en el dataset corresponden a datos de las siguientes columnas, recogidos de la temporada completa 2017 - 2018:

   - Jugador (*Jugador*).
   - Partidos jugados (*P.Jug.*).
   - Partidos completados (*P.Compl.*).
   - Partidos titular (*P.Tit.*).
   - Partidos sustituido (*P.Sust.*).

- Minutos (*Min*).
- Tarjetas (*Tarj.*).
- Expulsiones (*Expul.*).
- Goles marcados (*Goles*).
- Penaltis transformados (*Penalti*).
- Equipo (Equipo).

Estos datos han sido generados a partir de la información que ha reflejado cada árbitro a la conclusión de cada partido en el acta arbitral. Posteriormente, *LaLiga* los ha hecho públicos en su web, filtrando por liga, temporada y equipo. De esta manera, el código accede directamente a la temporada a estudio, 2017 - 2018, y es capaz de generar el filtrado de todos los equipos que han participado en ella, para acceder a los datos mostrados de cada equipo y almacenarlos en una única tabla dataset, pues lo que interesa es el análisis por jugador.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

El propietario de los datos es la **Liga de Fútbol Profesional** con sede en *Calle Torrelaguna 60, Madrid*, teléfono *912 055 000*, email *prensa@laliga.es* y *webmaster@laliga.es*. Su actual presidente es Javier Tebas (*@tebasjavier*).
Agradecimientos a los árbitros por cumplimentar las actas arbitrales y a *LaLiga* por hacer públicos los datos y permitir acceder a ellos.

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

El propósito de la recopilación de estos datos es conocer qué jugadores has destacado durante la temporada 2017 - 2018, en base a las estadísticas recogidas que facilitan el *scouting* para el equipo de fútbol implicado.

- Máximos goleadores.
- Ratio goles por partido.
- Ratio goles por minutos jugados.
- Jugadores más amonestados.
- Jugadores con más minutos jugados.
- Partidos totales completados.

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

Evaluado las restricciones de la web *LaLiga* cuando se pretende rastrearla, y con la intención de preservar su declaración, con el fin de reducir así las posibilidades de ser bloqueados, se ha decido escoger para la publicación de este conjunto de datos una licencia **CC BY-SA 4.0 License**.
Los motivos de esta elección, como se ha indicado, están relacionados con la idoneidad de las cláusulas que esta presenta en relación con el trabajo realizado.

- *Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado*. De esta manera, se reconoce el trabajo ajeno y en qué medida se han realizado aportaciones en relación con el trabajo original.
- *Se permite un uso comercial*. Esto haría que incrementen las probabilidades de que una empresa utilice los datos generados y realicen trabajos de calidad que reporten cierto reconocimiento al autor original.

- *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 él mismo planteó.

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

   Adjunto código en R *LaLiga123_17-18_stats.Rmd*.

10. Dataset. Presentar el dataset en formato CSV.

    Adjunto el dataset *LaLiga123_17-18_stats* en formato *.csv*. Los campos están separados por ",".

Contribuciones al trabajo.

Contribucciones	Firma
Investigación previa	LAB* y MABA*
Redacción de las respuestas	LAB* y MABA*
Desarrollo del código	LAB* y MABA*

LAB*, Luis Alberto Bayo.
MABA*, Miguel Ángel Bermejo Águeda