

# Tipologia y ciclo de vida de los datos

**Alumna:** Mabel Arroyo Adeba

PRA1 Web Scraping (BORRADOR)

**Fecha:** Noviembre 2020

## Indice

1. Contexto
2. Título del dataset
3. Descripción del dataset
4. Representación gráfica
5. Contenido
6. Agradecimientos
7. Inspiración
8. Licencia
9. Código
10. Dataset
11. Firma participantes

## 1. Contexto

La ludopatía es una adicción que sufre un 1% de la población en España. En los últimos años, además, se ha detectado que esta enfermedad mental a aumentado entre los jóvenes de 18 a 25 años.

Basándonos en el cálculo combinatorio, en el sorteo de la Primitiva, la probabilidad de acertar los 6 números y el reintegro es de 139.838.160. Tal vez este dato no sea suficiente para que las personas entiendan que este sorteo no es más que un juego.

En la página de loterías y apuestas del estado, es posible recoger los números de los sorteos que se han celebrado desde el 20/06/1991. Mediante técnicas de Web Scraping se capturarán estos datos, para crear un dataset cuya descripción veremos en los puntos siguientes.

## 2. Título del dataset

El nombre del dataset será: **Historico\_Primitiva.csv**

# Tipologia y ciclo de vida de los datos

Alumna: Mabel Arroyo Adeba

PRA1 Web Scraping (BORRADOR)

Fecha: Noviembre 2020

## 3. Descripción del dataset

El dataset Historico\_Primitiva contendrá tantas líneas como sorteos se hayan celebrado desde la fecha indicada al principio. En cada línea aparecerá la fecha del sorteo y los números de las 6 bolas de la combinación en orden ascendente y la del número complementario. El número del Joker no se recoge.

## 4. Representación gráfica

Pendiente...

## 5. Contenido

Los datos serán recogidos de la página de loterías y apuestas del estado. Se capturarán sorteos celebrados desde el 20/06/1991 hasta el día de hoy.

El fichero de salida será un fichero plano .csv

Los campos definidos serán:

Nombre Campo	Tipo	Descripción	Ejemplo
FEC_SORTEO	Fecha	Fecha del sorteo	20/06/1991
B1	Númérico	Valor Bola 1	
B2	Númérico	Valor Bola 2	
B3	Númérico	Valor Bola 3	
B4	Númérico	Valor Bola 4	
B5	Númérico	Valor Bola 5	
B6	Númérico	Valor Bola 6	
C	Númérico	Valor Complementario	

# Tipologia y ciclo de vida de los datos

Alumna: Mabel Arroyo Adeba

PRA1 Web Scraping (BORRADOR)

Fecha: Noviembre 2020

## 6. Agradecimientos

El propietario de los datos es la Sociedad Estatal Loterías y Apuestas del Estado, S.M.E., S.A. (SELAE), domiciliada en la C/ Poeta Joan Maragall 53, 28020 Madrid. Su NIF es A86171964, inscrita en el Registro Mercantil de Madrid al tomo 28078, folio 202, sección 8ª, hoja M-505970, inscripción 1ª

Se solicitó permiso de acceso el día 28/10/2020. No se ha recibido respuesta.

## 7. Inspiración

El objetivo de la recopilación de los datos es la de hacer una aplicación en la que el usuario pueda introducir una combinación de 6 números y comprobar si en estos últimos años, esa combinación ha sido ganadora. Si los números elegidos han aparecido en algún momento, la probabilidad de que esa combinación vuelva a salir es prácticamente nula por lo que es absurdo seguir apostando por ella. Si, por el contrario, la combinación nunca se ha dado en este tiempo, es una manera de hacer entender que, si en tantos años no hemos acertado, mejor hubiésemos ido guardando el dinero de la apuesta en una hucha, y hoy en día, el dinero acumulado superaría el posible premio.

En conclusión: lo que se pretende es dar a las personas una visión con datos, de la realidad de los juegos de azar. Hacer entender que no hay mucho más allá de la ilusión de que algún día te toque la Primitiva. Por eso, lo que hay que ver es que no es más que un juego para entretenerse. Y tal vez con ello, alguien que lo necesita, vea que jugar de forma compulsiva es absurdo.

## 8. Licencia

Pendiente...

# Tipologia y ciclo de vida de los datos

**Alumna:** Mabel Arroyo Adeba

PRA1 Web Scraping (BORRADOR)

**Fecha:** Noviembre 2020

## 9. Código

La codificación del proyecto se ha realizado con el lenguaje de programación Python utilizando la plataforma Jupiter Notebooks. Para la generación del dataset se ha utilizado técnicas de web scraping. El fichero es Historico\_Primitiva.ipynb y está disponible junto a este documento.

## 10. Obtención del DOI

Pendiente...

# Tipologia y ciclo de vida de los datos

**Alumna:** Mabel Arroyo Adeba

PRA1 Web Scraping (BORRADOR)

**Fecha:** Noviembre 2020

## 11. Firma de participantes:

Contribuciones	Firma
Investigación previa	MAA
Redacción de las respuestas	MAA
Desarrollo código	MAA