

Asignatura: **M2.251 TIPOLOGÍA y CICLO DE VIDA DE LOS DATOS**

Alumnos: **TORRES BOLIVAR, Leidy Liliana**

**SOLA VERDÚ, Jose Carlos**

Fecha: **9/11/2020**

## *Dataset: LOTERÍAS BALOTO COLOMBIA*

### Contexto

Se ha seleccionado la web de loterías más utilizada de Colombia, Loterías Baloto, porque se considera que elaborar un *dataset* con números de lotería puede ser una fuente de interés para muchos usuarios, ya que se puede hacer uso de este *dataset* con el objetivo de crear sistemas que sean capaces de predecir qué números tienen más posibilidades de salir premiados.

### Descripción del Dataset

Para el Dataset, se han extraído los siguientes conjuntos de datos. Por un lado, se encuentra la **fecha**, con el objetivo de ver cada cuánto se repiten los números. Después, se encuentra el **resultado**, que obtiene los números premiados, y por último, se encuentra el número del **sorteo**, que vendría a ser el identificador.

### Representación Gráfica

Fecha	Número	Sorteo
31 Octubre 2020	01 - 02 - 23 - 16 - 05 - 11	2029
28 Octubre 2020	19 - 37 - 09 - 20 - 36 - 12	2028
24 Octubre 2020	42 - 20 - 19 - 40 - 28 - 13	2027
21 Octubre 2020	26 - 33 - 03 - 34 - 01 - 10	2026
17 Octubre 2020	27 - 36 - 21 - 02 - 40 - 15	2025
14 Octubre 2020	08 - 11 - 05 - 16 - 32 - 13	2024
10 Octubre 2020	01 - 36 - 39 - 02 - 15 - 04	2023
07 Octubre 2020	40 - 13 - 22 - 41 - 27 - 05	2022
03 Octubre 2020	17 - 28 - 35 - 38 - 09 - 01	2021
30 Septiembre 2020	01 - 16 - 09 - 30 - 13 - 05	2020

## Contenido del Dataset

Para el Dataset, se han extraído los siguientes conjuntos de datos. Por un lado, se encuentra la **fecha**, con el objetivo de ver cada cuánto se repiten los números. Después, se encuentra el **resultado**, que obtiene los números premiados, y por último, se encuentra el número del **sorteo**, que vendría a ser el identificador.

## Agradecimientos

Los datos han sido extraídos de la web <https://www.loteriasdehoy.com/baloto>

## Inspiración

El conjunto de datos de números de lotería puede aplicarse en diversos ámbitos. Por un lado, sería el de crear sistemas predictivos que puedan intentar predecir qué números serán los premiados.

Por otro lado, éste conjunto de datos podría ser de gran utilidad para las administraciones de lotería, ya que podrían “vender” que x números son los más premiados e incitar a los clientes que compren más.

## Licencia

La licencia escogida es la licencia “**Released Under CC BY-NC-SA 4.0 License**”, ya que se considera que no se puede hacer uso comercial de un desarrollo que permita predecir números de lotería, ya que podría haber demasiados intereses de terceros y el gobierno podría tomar medidas.

Las cláusulas detalladas de esta licencia son:

- ❖ El usuario es libre de **compartir**, copiar y redistribuir el material en cualquier medio o formato.
- ❖ Puede **adaptar** y/o transformar el *dataset* a sus necesidades.

Bajo las siguientes condiciones:

- ❖ **Reconocimiento:** El usuario debe reconocer adecuadamente la auditoría, proporcionando un enlace a la licencia e indicar los cambios realizados. Puede realizarlo de forma razonable, pero en ningún caso de forma que parezca que cuenta con el apoyo del licenciador.
- ❖ **Uso no comercial:** No puede utilizar el material con fin comercial ni empresarial.
- ❖ **Compartir Igual:** Los cambios que realice, deberá difundirlos bajo la misma licencia que la original.

## Código fuente y dataset

El código fuente y el dataset generado pueden consultarse a través del siguiente enlace: <https://github.com/lilitorres/Practica1WebScraping>

Se ha hecho uso del lenguaje de programación Python y de las técnicas de Web Scraping, con el objetivo de extraer la información almacenada en la web.

## Dataset

**HAY QUE PUBLICARLO EN ZENODO:** <https://zenodo.org/>

## Recursos consultados

Para la elaboración de la práctica, se han consultado los siguientes enlaces:

1. Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. <https://creativecommons.org/publicdomain/zero/1.0/deed.en>
2. Creative Commons. Attribution – Non Commercial – Share A like 4.0 International (CC BY-NC-SA 4.0) <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>
3. Creative Commons. Attribution – Share A like 4.0 International (CC BY-SA 4.0) <https://creativecommons.org/licenses/by-sa/4.0/deed.en>

## Participación de cada Integrante del Grupo

Contribuciones	Firma
Investigación Previa	LLTB, JCSV
Redacción de las Respuestas	LLTB, JCSV
Desarrollo Código	LLTB, JCSV

