**Descripción de la Práctica a realizar**

El objetivo de esta actividad será la creación de un dataset a partir de los datos

contenidos en una web. Para su realización, se deben cumplir los siguientes puntos:

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique

por qué el sitio web elegido proporciona dicha información.

 Este conjunto de datos proporciona información sobre los niveles ambientales de diferentes pólenes, medidos en diferentes estaciones de aerobiologia repartidas por España. Estos datos están recogidos en la web de la Sociedad Española de Alergología e Inmunología Clínica SEAIC.

En dicha sociedad, se encuentra el comité de Aerobiología, que es un comité científico, cuyo principal objetivo es potenciar la investigación en el área de la aerobiología (estudio de los pólenes y hongos que se encuentran en el aire) y polinosis (enfermedades alérgicas producidas por los pólenes).

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

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).

Recolección de los niveles ambientales de los principales pólenes, en el último año, en la ciudad de Barcelona

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

dataset visualmente

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

Gracias al uso de los datos proporcionados en la web: <https://www.polenes.com/home>, se ha podido realizar esta base de datos.

En este conjunto de datos se ha recogido la información de la estación de aerobiologia de la ciudad de Barcelona. Los datos están distribuidos en los atributos: tipo de polen (determina el origen de la planta), niveles ambientales (medido en gr/m3) y día. Se han recogido los datos para estas variables a lo largo de la primavera del año 2018 (20 de Marzo al 21 de Junio del 2018).

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario

incluir citas de investigación o análisis anteriores (si los hay).

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué

preguntas se pretenden responder.

<https://www.worldallergy.org/UserFiles/file/WWBOA_Executive-Summary_Spanish.pdf>

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

<https://es.wikipedia.org/wiki/Licencias_Creative_Commons#Licencias>

<https://creativecommons.org/licenses/?lang=es_ES>

<https://opendatacommons.org/licenses/odbl/1-0/index.html>

<https://en.wikipedia.org/wiki/Open_Database_License>

9. Código. Adjuntar el código con el que se ha generado el dataset,

preferiblemente en Python o, alternativamente, en R.

10. Dataset. Presentar el dataset en formato CSV

**Recursos**

Los siguientes recursos son de utilidad por la realización de la PEC:

● Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC.

● Masip, D. El lenguaje Python. Editorial UOC. Tipología y ciclo de vida de los datos Práctica 1 **pág 2**

● Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2.Scraping the Data.

● Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015).

Automated Data Collection with R: A Practical Guide to Web Scraping and Text

Mining. John Wiley & Sons.

● Tutorial de Github https://guides.github.com/activities/hello-world .