

Práctica 1: Python para la ciencia de datos – Análisis del fichero mundo

El objetivo de esta práctica es utilizar los conocimientos adquiridos en la asignatura para realizar un análisis de distintas características de la población mundial basándonos en el fichero *mundo.csv*. La práctica se realizará en **grupos de 2 personas** y corresponde a **un 50% de la nota de la asignatura**. La fecha límite de entrega será el **23/11/2019** a las **23:59**. Se permite la entrega hasta una semana después (**30/11/2019**), pero las prácticas que se entreguen en fechas **posteriores** al **23/11** tendrán como **máximo** una puntuación de **8 puntos**. La entrega consistirá en **un notebook de Python en el que se incluyan las respuestas a las preguntas planteadas, así como el código desarrollado para responderlas**.

El fichero consta de una tabla con 26 columnas y 110 filas que contienen datos referentes a varias características de la población mundial. Se pide al alumno que, una vez leídos los datos, realice las siguientes operaciones con ellos:

1.- Uno de los primeros pasos para realizar un buen análisis de datos es familiarizarnos con los datos que contiene el fichero a analizar. Para ello, calcularemos los estadísticos descriptivos elementales de las variables del fichero. Una vez cargados los datos en nuestro programa Python (utilizando la librería Pandas), calcula los siguientes valores **para cada una de las variables**:

- Número de muestras (valores distintos de *missing*)
- Media y desviación estándar
- Valor mínimo y valor máximo

2.- Hay datos que nos interesa analizar basándonos en agrupaciones, para darle un sentido a nuestro análisis en base a esa agrupación. Basándonos en las siguientes agrupaciones:

- Por región (columna región)
- Por religión

Calcula los siguientes estadísticos:

- Número de observaciones
- Número de valores ausentes (*missing*)
- Mediana
- Media
- Varianza
- Valores máximo y mínimo

Para las siguientes variables:

- Incremento de la población
- Esperanza de vida
- Tasa de alfabetización (tanto masculina como femenina)
- La mortalidad infantil
- PIB per cápita

¿Qué conclusiones podemos sacar de estos cálculos? Comenta los resultados.

3.- Selecciona los países cuya tasa de mortalidad sea superior a la media. Comenta los resultados obtenidos.

4.- Obtén la tasa media de natalidad y mortalidad por clima. Ordena los países en base a esas tasas medias (de **mayor a menor**). ¿En qué climas se produce mayor natalidad? ¿En qué región se obtiene mayor mortalidad?

5.- Obtén el crecimiento económico de cada país basándote en el PIB y ordénalos de **menor a mayor** crecimiento económico. Realiza el mismo cálculo basándote ahora en el PIB per cápita.

NOTA: La variable PIB **no está incluida en el fichero** por lo que corresponde al alumno realizar su cálculo (basándonos en la población y el PIB per cápita, este cálculo es automático).

6.- Representa la información obtenida en el ejercicio 5 mediante un diagrama de barras, pero sólo con los **20 países con el PIB más alto**.

7.- Representa mediante un *swarmplot* la tasa de natalidad por región y añade como tercera variable categórica el clima. ¿Tiene sentido el gráfico obtenido en relación a los datos? Comenta el resultado.

8.- Representa mediante un *boxplot* la tasa de natalidad por clima. Comenta los resultados.

9.- Selecciona dos o más variables que te llamen la atención y analiza los datos mediante las gráficas o mediciones estadísticas que consideres oportunas y comenta los resultados que obtengas.