

# Práctica de Programación en Python

Fecha de entrega: 14 de noviembre

Como práctica de evaluación del módulo de Programación en Python, se propone la realización de cuatro ejercicios, cada uno de ellos centrado en una parte del material teórico cubierto en clase: Python básico, NumPy y Pandas.

Para cada uno de los ejercicios se entregará un Notebook de Jupyter independiente en el que, además del código que resuelva cada problema, se deberá explicar el funcionamiento, las soluciones/decisiones propuestas (bien con comentarios, bien con celdas *markdown*) y aportar ejemplos que demuestren su correcto funcionamiento.

Puedes usar cualquier recurso a tu alcance como apoyo para la resolución de los ejercicios (Google, StackOverflow, libros, etc.). Eso sí, recuerda hacer mención a tus fuentes en los comentarios sobre la resolución.

Se deberá entregar un fichero zip con tu nombre y apellidos (***miguel\_angel\_corella\_python.zip***), en el que estén todos los archivos necesarios. No es necesario incluir el set de datos de los ejercicios de Pandas.

## Python básico – Ejercicios cortos (2 puntos)

En línea con los ejercicios de Python básico realizados en clase, se deberá dar solución a los siguientes problemas:

1. Escribe un **programa** que calcule todos los enteros entre 4000 y 9600 (ambos incluidos) que sean divisibles entre 5 pero no múltiplos de 7.
2. Escribe una **función** llamada “enteros\_absolutos” que reciba como entrada una lista de números enteros y genere a la salida otra lista con el valor absoluto de cada elemento distinto de 0 de la lista de entrada. Los 0 de la lista original deberán eliminarse en la salida.

**IMPORTANTE:** el ejercicio se debe resolver con comprensiones de listas.

3. Escribe una **función** llamada “comprimir” que reciba como parámetro una lista de elementos y devuelva como salida una lista de tuplas de la forma (elemento, número de repeticiones consecutivas). Por ejemplo:

*llamada: comprimir([1, 1, 1, 2, 2, 3, 4, 4, 1, 1, 1])*  
*output esperado: [(1,3), (2, 2), (3, 1), (4, 2), (1, 3)]*

**IMPORTANTE:** La lista no tiene por qué contener sólo números.

4. Crea una **función** llamada “alumnos\_aprobados” que reciba como parámetro un diccionario cuyas claves son nombres de alumnos y sus valores una lista de calificaciones numéricas. La salida de la función deberá ser una lista que contiene los nombres de los alumnos cuya nota media es igual o superior a 7 y NO tienen ningún suspenso. Por ejemplo:

*llamada: alumnos\_aprobados({'Juan': [8, 7, 6], 'Maria': [5, 5, 5], 'Pedro': [10, 3, 10]})*  
*Salida esperada: ['Juan']*

5. Escribe una **función** “habitantes” que reciba como entrada un nombre de ciudad y un diccionario cuyas claves son nombres de personas y su valor la ciudad en la que viven. La salida de la función deberá ser la lista de nombres de personas ordenados alfabéticamente del diccionario que viven en la ciudad recibida como entrada. Por ejemplo:

*llamada: habitantes('Madrid', {'Juan': 'Madrid', 'Alicia': 'Madrid', 'Pedro': 'Cádiz'})*  
*output esperado: ['Alicia', 'Juan']*

## NumPy – Validar Sudokus (2 puntos)

Escribe una función que reciba como parámetro el nombre de un fichero que contendrá un Sudoku completo (en forma de matriz), y devuelva si está bien solucionado o no.

A lo largo de la función, deberás ir comprobando los distintos requisitos que debe cumplir la solución de un Sudoku, pero, además, también debes hacer comprobaciones más básicas, como que los números de cada celda están en el rango y tipo correcto o que la dimensión de la matriz es la adecuada para un Sudoku.

La función debe devolver un mensaje diciendo que la solución es correcta o, en caso contrario, el listado completo de motivos por el que la solución es incorrecta.

Puedes visitar esta página para repasar las reglas del Sudoku:  
<http://www.sudokumania.com.ar/metodos/reglas-del-sudoku>

## Pandas - Explorando información de seguros (3 puntos)

Para evaluar los conocimientos adquiridos sobre la librería de análisis de datos Pandas, vamos a realizar un pequeño ejercicio de análisis exploratorio sobre un set de datos con información de una compañía ficticia de seguros médicos.

Más concretamente, junto con la práctica se entrega un fichero `seguro_medico.csv` que contiene información sobre contratos de seguros incluyendo región, género, edad, prima pagada... Adicionalmente, se proporcionan dos ficheros de catálogo con los posibles valores de género y condición de fumador.

Con este set de datos, deberás preparar un Notebook donde se resuelvan las siguientes cuestiones.

1. Carga del set de datos en una variable llamada “seguros”.
2. Número de observaciones incluidas en el set de datos.
3. Listado de variables (y sus tipos) incluidas en el set de datos.
4. Estadísticos básicos **sólo de las variables numéricas** del set de datos.
5. Tablas de frecuencia **sólo de las variables categóricas** del set de dato.
6. Identificación de variables con *missing values* y proporción de los mismos.
7. Cruza los datos de seguros con los catálogos de géneros y condición de fumador para tener un set de datos completo y elimina las columnas con los códigos numéricos.

## Pandas - Analizando información de seguros (3 puntos)

Tras la exploración básica del set de datos del ejercicio anterior, debería ser sencillo obtener *insights* o conclusiones sobre los costes de los seguros informados.

El objetivo de este ejercicio será utilizar las diferentes funcionalidades de **agrupación y agregación** que pone a nuestra disposición el módulo **pandas** de Python para responder a un pequeño conjunto de preguntas sobre los seguros disponibles.

**IMPORTANTE:** los ejercicios se deberán utilizar las funcionalidades de **agrupación y agregación de Pandas** y **NO tablas pivote**. Adicionalmente, se espera que los resultados NO presenten valores de códigos numéricos de género y condición de fumador, sino sus descripciones obtenidas de los catálogos tras el punto 7 del ejercicio anterior.

El listado de preguntas que se deberá responder es el siguiente:

1. ¿Qué género es el que tiene el coste medio de seguro más alto?
2. ¿Qué IMC medio tienen los seguros con costes inferiores a 9000?
3. ¿Qué relación hay entre la región de residencia y los costes del seguro?
4. ¿Qué género predomina entre los fumadores?
5. Calcula la edad media de los tomadores de los seguros por región y género.
6. ¿Qué relación existe entre la edad y los costes del seguro?