**Plantilla actividad**

**Máster en Big Data & Data Science**

**Módulo 1. Herramientas de Big Data**

## Actividad núm. 3

### Título de la actividad

Indicar un título representativo para la actividad propuesta.

HERRAMIENTAS DE ANALISIS: PROGRAMACIÓN EN PYHTON – MICROACTIVIDAD 3

## Tipo de actividad

Marcar si la actividad será individual o grupal.

|  |  |  |  |
| --- | --- | --- | --- |
| **Individual** | **X** | **Grupal** |  |

## Instrucciones para el alumnado

### Descripción de la tarea

Describir de forma clara el enunciado de la actividad.

*Redactad en segunda persona del singular (trato de tú) para actividades individuales y para grupos segunda persona del plural (trato de vosotros).*

**Important:** En aquest manual explicarem les macros programades que creiem són les més utilitzades.

Para realizar el ejercicio cargamos los datos de *Especies* en STARWARS SWAPI y las librerías principales.

import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns; sns.set() # para el estilo de graficos  
  
entidades = ['planets','starships','vehicles','people','species']  
entidades\_df = {x: pd.read\_pickle('www/' + x + '\_df.pkl') for x in entidades}  
  
# Datos principales  
people\_df = entidades\_df['people'][["height","mass","birth\_year","gender","homeworld"]].dropna()  
  
people\_df

Utiliza la plantilla del ejercicio en [**Colab**](https://colab.research.google.com/github/griu/mbdds_fc20/blob/master/Python/modulo1_tema4_Py_Ejercicio_1.ipynb) .

## Ejercicio 3.1.

Construye un gráfico de dispersión de los personajes donde se visualice: la altura (height), el peso (mass), la edad en años BBY (birth\_year) y el género (gender). Para ello utiliza la función sns.scatterplot() de la librería seaborn. Aprovecha los parámetros: x, y, size, hue y style (consulta la ayuda de la función [.scatterplot()](https://seaborn.pydata.org/generated/seaborn.scatterplot.html).

# Solución:

## Ejercicio 3.2.

Sobre el gráfico del ejercicio 3.1:

* Pon título al gráfico y a los ejes x e y.
* Modifica los límites del eje y para que aparezcan sólo personajes de menos de 150 Kg de peso.
* Sitúa en el gráfico los nombres de “Darth Vader” y “Anakin Skywalker”. ¿Cómo es posible tengan un peso y altura tan distintos si eran la misma persona?

# Solución:

## Ejercicio 3.3.

Utiliza las *list comprehension* para calcular el cuadrado de los valores positivos de la siguiente lista:

Muestra el resultado por pantalla.

val = [5, 6, -1, 2, -3, -7, 9, 1]

# Solución:

## Ejercicio 3.4.

Construye un diccionario donde se identifique, mediante claves y valores, las siguientes características del personaje Yoda: “nombre”, “altura”, “peso”, “edad” y “genero”. Utiliza los datos de people\_df.

Muestra el diccionario por pantalla.

# Solución:

## Ejercicio 3.5.

Calcula, a partir de los vectores numpy de altura y peso, definidos a continuación, el [índice de masa corporal (IMC)](https://es.wikipedia.org/wiki/%C3%8Dndice_de_masa_corporal) de los personajes de star wars contenidos en people\_df:

donde altura está en metros y el peso en kg.

Muestra los datos por pantalla.

altura = people\_df.height.values  
peso = people\_df.mass.values

# Solución:

## Ejercicio 3.6.

A partir del IMC que has calculado en el ejercicio 3.5. Construye un panel con dos histogramas:

* Un histograma con toda la muestra
* Un histograma seleccionando los valores con un IMC inferior a 100.

# Solución:

## Ejercicio 3.7.

A partir del vector 1 y 2 que se definen a continuación contesta las siguientes preguntas:

* Calcula el shape, ndim, size del vector1 y vector2
* Explica cuál es la diferencia entre vector1 y vector2 a partir de los que hayas observado

vector1 = np.hstack([altura,peso])  
vector2 = np.vstack([altura,peso])

# Solución:

## Ejercicio 3.8.

Crea una copia de people\_df llamada personajes\_df y renombra las columnas con su traducción al castellano.

Truco: Puedes renombrar el índice con:

personajes\_df.index = personajes\_df.index.rename("nombre")

Muestra los 5 primeros registros del nuevo data frame con .head().

# Solución:

## Ejercicio 3.9.

Haz el mismo cálculo de IMC que has hecho en 3.5. pero directamente sobre el objeto personajes\_df.

Ordena el data frame de mayor a menor IMC y muestra los personajes con IMC por encima de 30.

¿Sabías que IMC por encima de 30 se considera obeso?

# Solución:

## Ejercicio 3.10.

Inserta un valor faltante en los valores de IMC que sean superiores a 100 y dibuja el histograma con .plot.hist().

# Solución:

### Pasos para la realización de la actividad

Detallar los pasos que se seguirá en la elaboración de la actividad que servirán de pauta para el tutor/docente durante su impartición y para el alumno como guía para su realización.

Esta información será parte del enunciado de la actividad y por lo tanto visible para el alumnado.

|  |  |  |
| --- | --- | --- |
| 1. | Presentación de la actividad. | Día XX/XX/2020. |
| 2. | Elaboración y entrega del informe/resultado. | Envío a través del espacio habilitado en el aula virtual (antes del XX/XX/2020). |
| 3. | Evaluación de la actividad por parte del tutor(a). | La nota se publicará en el aula virtual a partir del día XX/XX/2020. |

**Descripción del entregable**

Se deberá detallar:

* El tipo de documento que se espera (informe, artículo, vídeo, etc.), si debe tener una estructura determinada (se detallarán los apartados, los bloques, etc.), así como otros aspectos o consideraciones a tener en cuenta.
* Formato: si el documento a entregar será un word, pdf, gif, avi, etc.
* Nombre de archivo: Se indicará cómo deberá nombrarse el archivo. Como criterio general se detallará: “El archivo a entregar deberá nombrarse siguiendo la siguiente estructura (sin acentos): apellido1\_apellido 2\_moduloX\_actX”.
* Extensión: longitud mínima y máxima (en páginas o caracteres).
* Tipo de documento: Notebook de Jupyter/Colab. El alumno debe entregar un documento .ipynb con el código Python solicitado. El notebook debe estar guardado con los resultados ya generados, aunque debe poderse re-ejecutar por parte del profesor.
* Formato: ipynb
* Nombre de archivo: El archivo a entregar deberá nombrarse siguiendo la siguiente estructura (sin acentos): apellido1\_apellido 2\_moduloX\_temaX\_actX
* Extensión: Libre
* La entrega se deberá realizar en la correspondiente tarea del aula del curso.

**Materiales y/o recursos requeridos para la realización de la actividad (si procede)**

* Listado de los contenidos del temario que hace falta leer, consultar o trabajar para la elaboración de la tarea.
* Si la actividad requiere de algún material específico (consultar un informe, una web, un software, también se debe indicar aquí).
* Si aplica, adjuntar los documentos que se entregarán a los alumnos para realizar la actividad. Por ejemplo, herramientas, texto para trabajar, documentos modelo, tablas, organigramas, gráficos numéricos, esquemas…

Se recomienda leer los siguientes apartados del tema “Programación en R”:

* 2.3. Elementos básicos de Python
* 2.4. Gráficos con Matplotlib
* 2.5. Colecciones de objetos
* 2.6. Numpy
* 2.7. Pandas

### Objetivo/s de la actividad

¿Qué se pretende que el alumno consiga al finalizar esta actividad?

* Desarrollar gráficos con todos los elementos que lo definen: titulo, título de los ejes, si aplica, leyenda, uso de elementos gráficos como la tipología, colores, tamaños y textos.
* Identificar la estructura de datos más adecuada a cada problema.
* Utilizar las funciones adecuadas de Python para resolver el problema de la forma más simple posible.
* Presentar los resultados de forma limpia y organizada en el contenido y la forma.
* Redactar de forma clara y comprensible las dudas planteadas.

### Criterios que tendrá en cuenta el tutor para evaluar la actividad

* Los criterios de evaluación serán públicos en el aula y visibles para el alumno.
* Se deben explicitar todos los criterios a tener en cuenta.
* Deben ser acordes con los objetivos de la actividad, y han de tener en cuenta no sólo el punto de vista de un profesor que evalúa sino también el punto de vista de quien recibiría la actividad en el lugar de trabajo (un responsable al dar feedback a un colaborador, o un cliente, o un compañero, o un proveedor) siguiendo la lógica y la cultura del sector. Responden por tanto a la pregunta: ¿Qué será lo que mirará y tendrá en cuenta el tutor y el profesional para evaluar la actividad?
* Se deberá cumplimentar el porcentaje de evaluación, es decir qué peso tendrá la actividad respecto del módulo.
* Indicar si la actividad es recuperable marcando la selección SÍ | NO
* En el caso de marcar SÍ indicar cuál va a ser la manera de recuperar la actividad propuesta por el tutor.
* Programación de gráficos con todos los elementos solicitados, que sean visualmente limpios (con todos los elementos legibles), que utilicen toda la variedad de elementos gráficos disponibles.
* Precisión en los resultados obtenidos y explicaciones que se piden.
* Uso de funciones de Python correctas y de la forma más simple posible.
* Presentación de los resultados de forma limpia y organizada en el contenido y en la forma.
* Redacción de clara y comprensible las dudas planteadas.

|  |
| --- |
| Porcentaje de evaluación de la actividad respecto al módulo: xx % |