**Plantilla actividad**

**Máster en Big Data & Data Science**

**Módulo 1. Herramientas de Big Data**

## Actividad núm. 4

### Título de la actividad

Indicar un título representativo para la actividad propuesta.

HERRAMIENTAS DE ANALISIS: PROGRAMACIÓN EN PYHTON – MICROACTIVIDAD 4

## 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

Y añadimos los datos de planetas.

# planetas  
planets\_df = entidades\_df['planets'][["orbital\_period","url"]].dropna()  
planets\_df.head()

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 4.1.

Construye una función que diga “buenos días”, “buenas tardes” o “buenas noches” en función de la hora del día.

Truco 1: Para testear la función haz que tenga un parametro de entrada que tenga como valor por defecto: datetime.now() (primero carga from datetime import datetime).

Truco 2: Puedes extraer la hora de un datetime con .hour.

# Solución:

## Ejercicio 4.2.

Elimina del data frame personajes\_df los personajes con valores faltantes en IMC y crea una nueva variable con el trameado de la variable IMC definido en la siguiente tabla:

|  |  |
| --- | --- |
| Categoría de nivel de peso | Intervalo del percentil |
| Bajo peso | < 18.5 |
| Normal | >= 18.5 y < 25 |
| Sobrepeso | >= 25 y <30 |
| Obeso | >= 30 |

Truco: utiliza pd.cut(..., right=False) y modifica las etiquetas con .cat.categories.

# Solución:

## Ejercicio 4.3.

Muestra las frecuencias de la nueva variable definida en 4.2.

# Solución:

## Ejercicio 4.4.

Calcula ahora un trameado de la edad en 5 grupos equiprobables.

Truco: Busca ayuda de la función [pd.qcut()](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.qcut.html)

# Solución:

## Ejercicio 4.5.

Presenta la tabla cruzada de tramos de edad (ej. 4.4) por tramos de IMC (ej. 4.2).

¿Qué tramo de edad tiene un mayor número de personajes con Bajo peso?

# Solución:

## Ejercicio 4.6.

Calcula la media del IMC en cada tramo de edad del ejercicio 4.2.

Truco: En el mismo cálculo del IMC medio, calcula también, la mediana de edad de cada tramo de edad para que sea el eje x del gráfico de líneas.

# Solución:

## Ejercicio 4.7.

Presenta los datos del ej. 4.6. como un gráfico de líneas donde el eje x sea la edad y el eje y el IMC medio.

# Solución:

## Ejercicio 4.8.

Calcula el cociente del IMC sobre la mediana del IMC de su tramo de edad (definidos en el ejercicio 4.2.) mediante la función groupby(…).apply(…).

truco: primero crea la función que devuelva: x / np.nanmedian(x).

Presenta los datos mediante un boxplot de la nueva variable: [pd.boxplot()](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.boxplot.html).

# Solución:

## Ejercicio 4.9.

¿Cuál es planeta con menor IMC medio de sus personajes? ¿Que personaje/s son de ese planeta?

# Solución:

## Ejercicio 4.10.

Convierte a datetime los siguientes strings con la función [datetime.strptime()](https://docs.python.org/3/library/datetime.html#strftime-strptime-behavior) (consulta la ayuda si es necesario) de la libreria datetime:

* “1 Enero, 2020”
* “15-feb.-2017”
* “20190701 22:30” # 1 de julio de 2019

Truco: Modifica la configuración regional:

from datetime import datetime  
import locale  
locale.setlocale(locale.LC\_TIME, 'es\_ES.UTF-8')

# 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”:

* 2.8. Control de flujo
  + 2.8.1. Funciones
  + 2.8.2. Condicionales
  + 2.8.3. Bucles
* 2.9. Gestión de datos
  + 2.9.1. Importar y exportar datos
  + 2.9.2. Cruce entre tablas
  + 2.9.3. Resumenes de agregados
  + 2.9.4. Agregados por subgrupos
  + 2.9.5. Tablas pivote
  + 2.9.6. Datetime: Gestión de fechas y horarios

### Objetivo/s de la actividad

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

* 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.
* 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 % |