**FUNCIONAMIENTO CODIGO PANDA\_EXAMPLE.PY**

**PRESENTADO POR:**

**JOHAN SEBASTIAN FUENTES ORTEGA**

**COD: 230172002**

**UNIDAD CENTRAL DEL VALLE DEL CAUCA**

**INTELIGENCIA ARTIFICIAL**

**INGENIERÍA DE SISTEMAS**

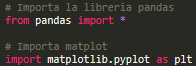
**2021**

**PANDA 1 EXAMPLE**

Como primera medida llamamos la librería Warning para desactivarla, debido que en algún proceso más adelante nos podría generar mensajes de advertencia o nos detiene el programa. Por ende se usa el código:



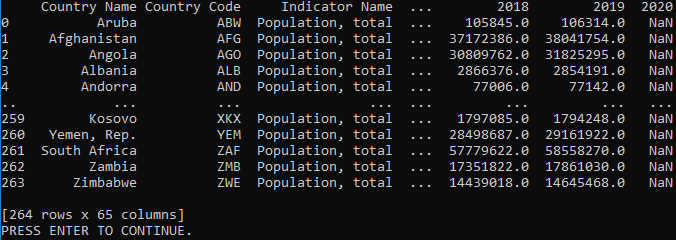
Posteriormente importamos las librerías Pandas y mathplot con el fin de llevar a cabo códigos como datasets y dibujos de gráficos.



Ahora bien el programa comienza creando una variable de tipo objeto para almacenar los datos que descargara de un sitio web, este archivo es un tipo excel.

Luego hacemos un print para mostrar los datos que tenemos, y el programa espera que el usuario digite enter para seguir con el programa.

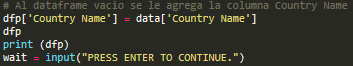




Seguido de eso, creamos una nueva variable de nombre dfp para almacenar un dataframe, en este caso el dataframe al no contener nada entre los paréntesis, se creará como dataframe vacío.

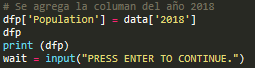


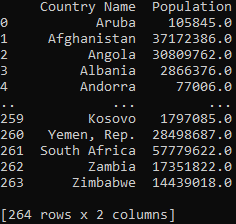
Al data frame vacío le creamos una columna de nombre Country Name el cual contendrá los datos de la columna Country Name que se almacenaron en la variable objeto data (variable que almacenó los datos tomados de la web). Mostramos en pantalla lo que tiene ahora el data frame dfp y esperamos la interacción del usuario.





Ahora creamos una nueva columna de nombre Population en el data frame, que contendrá los datos de la población que habían en el año 2018 que se tomaron de la web. Mostramos la nueva forma del data frame y esperamos la interacción del usuario.

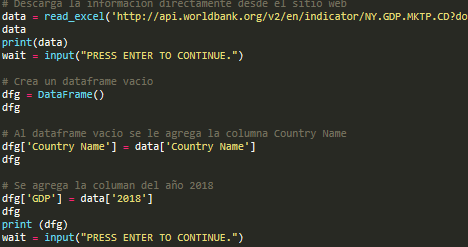


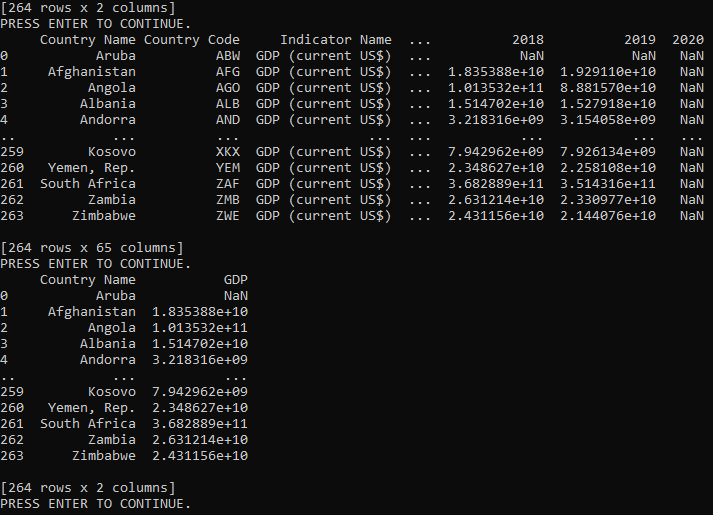


Ahora hacemos una nueva variable que almacenará los datos obtenidos por una dirección web, estos datos son de un archivo tipo excel y mostramos los datos.

Creamos una nueva variable dfg que será un dataframe vacío, creamos las columnas Country Name y GDP, la primera contendrá los datos de Country Name y la segunda los datos de la del producto interno bruto en el 2018 que se obtuvieron en la nueva dirección web que comenté anteriormente.

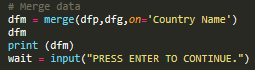
Mostramos el data frame y esperamos la interacción del usuario.

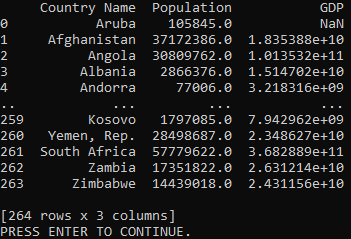




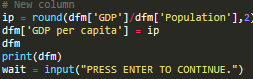
Posteriormente hacemos un Merge entre los dataframe dfp y dfg y su identificador será Country Name. El merge es realizar una combinación entre las dos “tablas”, como un inner join cuyo identificador es el nombre del país.

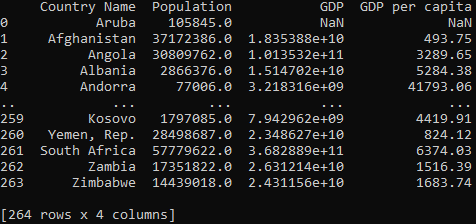
El nuevo data frame obtenido se almacenará en la variable dfm y lo mostramos al usuario.



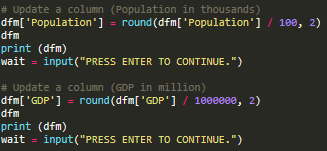


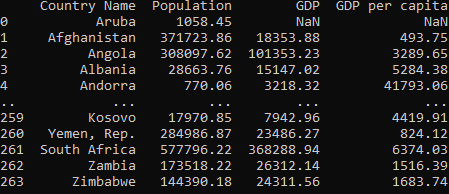
Ahora obtendremos el producto interno bruto per cápita, que se hace dividiendo el GDP sobre la población, al resultado aplicamos la función round, con el fin de redondearla a dos dígitos. Estos datos los almacenamos en una nueva columna del dataframe dfm (data frame que usamos para almacenar el merge entre los data frames anteriores) con el nombre GDP per capita. Mostramos al usuario y esperamos su respuesta.



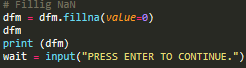


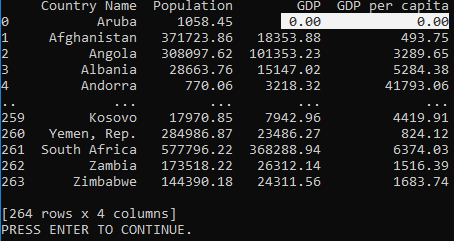
Actualizamos los datos de la columna population para que quede en miles, para ello dividimos entre 100 y lo redondeamos a 2 cifras con la función round; lo mismo hacemos para la columna GDP para actualizarla en millones para ello se dividió en 1 millón y se redondeó en 2 con la misma función round y lo mostramos al usuario mientras esperamos su interacción.



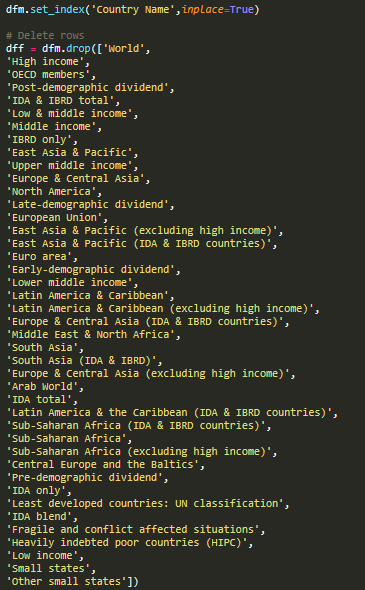


Posteriormente como podemos ver en la imagen anterior algunos campos nos quedan con valor NaN, en pocas palabras, valores nulos, para llenar esos campos nulos y asignarles valor 0 aplicamos la función que brinda panda para los data frames fillna(value=), esta función convierte los campos NaN en el valor que le asignemos, en nuestro caso valor 0.



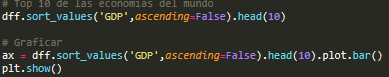


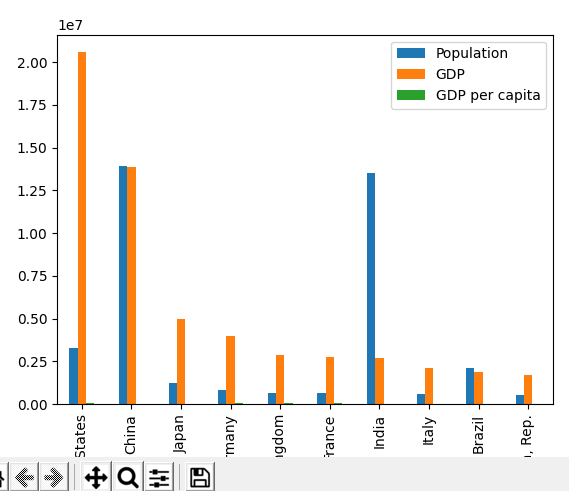
Definimos como índice del dataframe a Country Name, y posteriormente borramos algunas filas que no queremos que estén en nuestro dataframe.



Ahora si queremos obtener solamente los 10 países que encabezan o son líderes en el mercado apartir del producto interno bruto, aplicamos la función sort\_values con los valores de la columna GDP, esa función ordena los datos y luego es decimos que nos tome los 10 mejores.

hacemos una variable ax el cual graficará los valores de los datos en forma de bar con plot.bar(), y los mostramos con plt.show()





Por último aplicamos al data frame el describe() que me retorna un resumen estadístico de los datos que tenemos tales como: cuartiles, media, promedio, cantidad, maximos y minimos.



