



ISIS-1221

INTRODUCCIÓN A LA PROGRAMACIÓN

Nivel 4 – Laboratorio Pandas y Matplotlib

Objetivos

1. Familiarizarse con las librerías pandas y matplotlib.
2. Ejercitar la capacidad de consultar y entender la documentación de las librerías de Python.
3. Fomentar la habilidad de descomponer un problema en subproblemas y de implementar funciones que los resuelven, lo que se conoce comúnmente como la técnica de “Dividir y Conquistar”.

Contexto: exportaciones de banano

Musa Paradisiaca, es una planta herbácea que fue obtenida tras múltiples cruces entre plantas del género *Musa*, y es el nombre científico para la fruta que es conocida comúnmente como Banano. Colombia es un exponente importante en el mercado global del banano, en el año 2018 se exportaron aproximadamente 1.790.000 toneladas de banano, lo que representa un 9.3% del total mundial de exportaciones para esta fruta; y que además coloca a nuestro país como el 5to exportador a nivel mundial. Para esta tarea, usted deberá trabajar con una base de datos que contiene registros de exportaciones de productos de Banano, respondiendo a diferentes intereses de una empresa de inversiones agrícolas.

En el archivo *exportaciones.csv* se encuentran los registros de exportación de productos de banano desde enero del 2006 hasta agosto del 2018. Cada línea del archivo corresponde a una exportación efectuada. Las columnas de este archivo son:

1. **Anio:** Año en que se realizó la exportación.
2. **Mes:** Mes en que se realizó la exportación.
3. **PaisDestino:** País al cual se realizó la exportación.
4. **DepartamentoOrigen:** Departamento en Colombia donde se produjo el banano que se exportó.
5. **Producto:** Variedad de banano exportada.
6. **Cantidad:** Cantidad en Kilogramos de la exportación.
7. **ValorMilesDolar:** Precio de la exportación en miles de Dólares Americanos.
8. **ValorMilesPesos:** Precio de la exportación en miles de Pesos Colombianos.

Un agricultor emprendedor está interesado en incursionar en el mundo del cultivo del banano, y por esto necesita de su ayuda para determinar algunos detalles sobre su nueva empresa.

Su reto es hacer un programa que permita al usuario realizar las siguientes acciones:

1. Cargar el archivo utilizando la librería *Pandas* y guardar los datos en un *DataFrame*.
2. Graficar en un diagrama de pastel (*pie* en inglés), la distribución de cantidad total de exportaciones en miles de dólares por los tipos de banano. Esto es: para cada tipo de banano, la suma total en miles de dólares correspondiente a las exportaciones de dicho tipo. En seguida, graficar este mismo diagrama de pastel, pero sin tener en cuenta el tipo de banano “Cavendish Valery”. Por último, graficar tres diagramas de pastel, uno para cada uno de los últimos años (2016, 2017 y 2018).

3. Generar un reporte (archivo) con la cantidad de kilos que han sido exportados a lo largo del tiempo (año tras año) por departamento.
4. Graficar en forma de barras, la evolución del mercado del banano a lo largo del tiempo (por año), teniendo en cuenta únicamente los años pares. Esta función debe recibir por parámetro cuál métrica se quiere utilizar para medir la evolución, ya sea la cantidad de kilos, valor en dólares, o valor en pesos.
5. Encontrar y graficar (con barras horizontales) los 5 países a los cuales se exportó una mayor cantidad banano (en kilogramos) y los 5 a los que menos se exportó, indicando la cantidad importada por cada uno de estos países.

Preparación

En los siguientes enlaces se encuentra la documentación y algunos tutoriales muy útiles para familiarizarse con la librería Pandas y con el módulo pyplot de la librería Matplotlib.

Pandas:

Tutorial: http://pandas.pydata.org/pandas-docs/stable/getting_started/tutorials.html

Documentación completa: <http://pandas.pydata.org/pandas-docs/stable/>

Pyplot:

Tutorial: <https://matplotlib.org/tutorials/introductory/pyplot.html>

Ejemplos de tipos de gráficos: http://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html

Documentación completa de matplotlib: <https://matplotlib.org/users/index.html>

Refiérase a estos enlaces cada vez que necesite generar un nuevo tipo de gráfico que no conozca o que desee configurar su visualización.

Instrucciones

Separe su programa en dos archivos: uno para la interfaz del usuario (consola) y otro para las funciones que implemente (librería).

Implementación de la consola

El archivo `consola.py` debe seguir la misma estructura de las consolas que hemos implementado en laboratorios y proyectos anteriores. Esto es: debe existir una función llamada `iniciar_aplicacion()` para que muestre el menú usando la función `mostrar_menu()` y permita al usuario seleccionar una opción. El menú que se despliegue debe permitir al usuario ejecutar todas las acciones descritas en el contexto del problema, así como salir (terminar) del programa. Por cada una de las funciones principales de su programa (las que se enumeraron en el punto anterior), debe existir una función en la consola que la ejecute, pidiendo previamente los datos necesarios al usuario (si aplica) e imprimiendo por pantalla el resultado de la función. Se sugiere nombrar estas funciones como `ejecutar_XX`, donde XX es la respectiva función de su módulo.

Cuando haya implementado la función `iniciar_aplicacion()` corra su programa y verifique que se comporte adecuadamente, permita al usuario seleccionar las opciones que se quieren ejecutar, y termine el programa cuando se le indique.

Implementación de la librería o módulo

Usted puede crear cuántas funciones considere necesarias dentro de su librería o módulo. Mínimo debe haber una función por cada una de las acciones que debe realizar el programa.

Pruebe el correcto funcionamiento de su programa

Puede probar el correcto funcionamiento de su programa cargando la información que se encuentra en el archivo “**exportaciones.csv**” (en sicuaplus). O creando su propio archivo de menor tamaño que le permita corroborar que los resultados arrojados por su programa son correctos.

Problema 1: Cargar el archivo en un DataFrame

Cargar el archivo utilizando la librería *Pandas* y guardar los datos en un *DataFrame*.

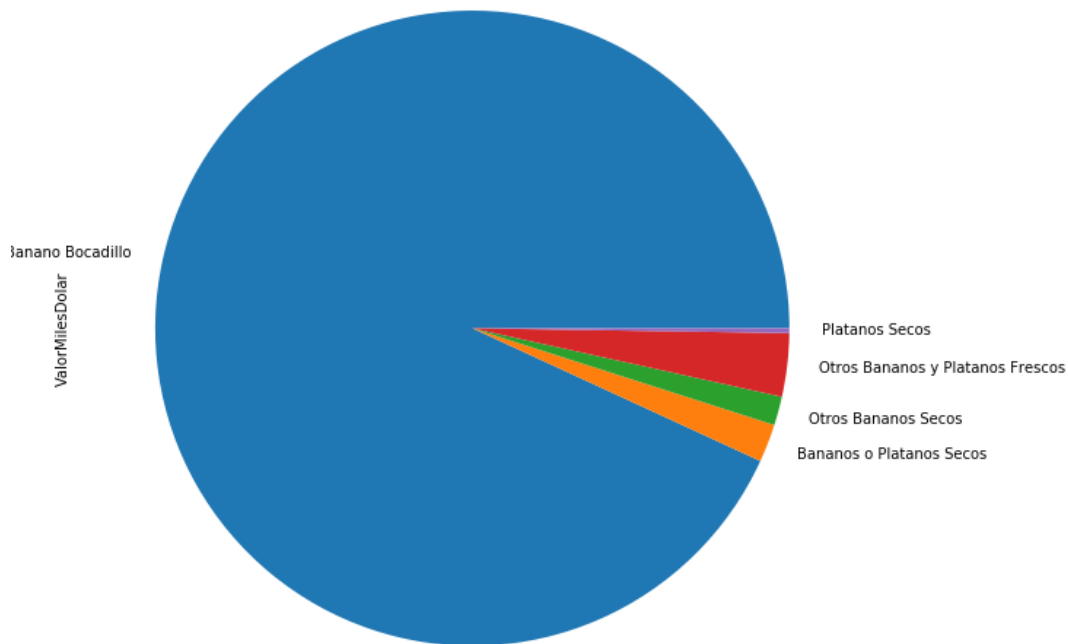
Ayuda: si alguna de sus funciones recibe un parámetro de tipo *DataFrame* o retorna un *DataFrame*, el tipo de este es: *pandas.DataFrame*. Por ejemplo, la función que carga un archivo csv y retorna un *DataFrame* tendría el siguiente encabezado o signatura:

```
def cargar_datos(nombreArchivo: str)->pd.DataFrame:
```

Donde *pd* es el alias que se le dio a la librería *pandas* cuando se importó.

Problema 2: Graficar la distribución de cantidad total de exportaciones en miles de dólares por tipo de banano

El diagrama de pastel (pie en inglés) tiene el siguiente aspecto:



Ayuda:

1. Utilice el método `groupby` sobre el *DataFrame* para agrupar los datos por tipo de banano (columna llamada *Producto*). Suponiendo que el dataframe se llama *bananos*, el agrupamiento se hace de la siguiente manera:

```
bananos.groupby('Producto')
```

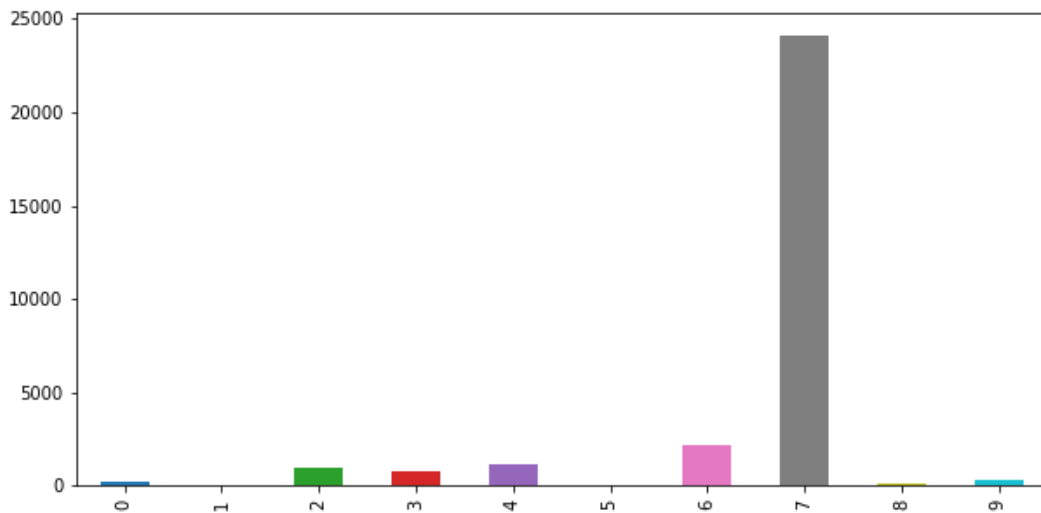
2. Sobre el dataframe que retorna el método `groupby`, efectúe la suma sobre la columna *ValorMilesDolar*, para obtener el total de valor de exportaciones en dólares por tipo de producto:

```
bananos.groupby('Producto').ValorMilesDolar.sum()
```

3. Construya el gráfico de pastel (pie) usando este último dataframe.
4. Repita los pasos anteriores pero sin tener en cuenta el tipo de banana “Cavendish Valery”. Para esto, aplique un filtro al DataFrame inicial que excluya los registros cuya columna Producto es “Cavendish Valery”.
5. Repita los pasos anteriores para graficar tres pasteles, con las exportaciones de los últimos años (2016, 2017 y 2018). Para esto, aplique un filtro al DataFrame inicial que excluya los registros cuya columna Producto es “Cavendish Valery” y además extraiga los registros para cada uno de estos años.
6. Si sus gráficos no se visualizan correctamente, cada vez que vaya a desplegar un nuevo gráfico, cree la figura con la instrucción `plt.figure()` y muestre el gráfico con la instrucción `plt.show()`, donde `plt` es el alias de `matplotlib.pyplot`. Por ejemplo, para generar un nuevo gráfico de barras:

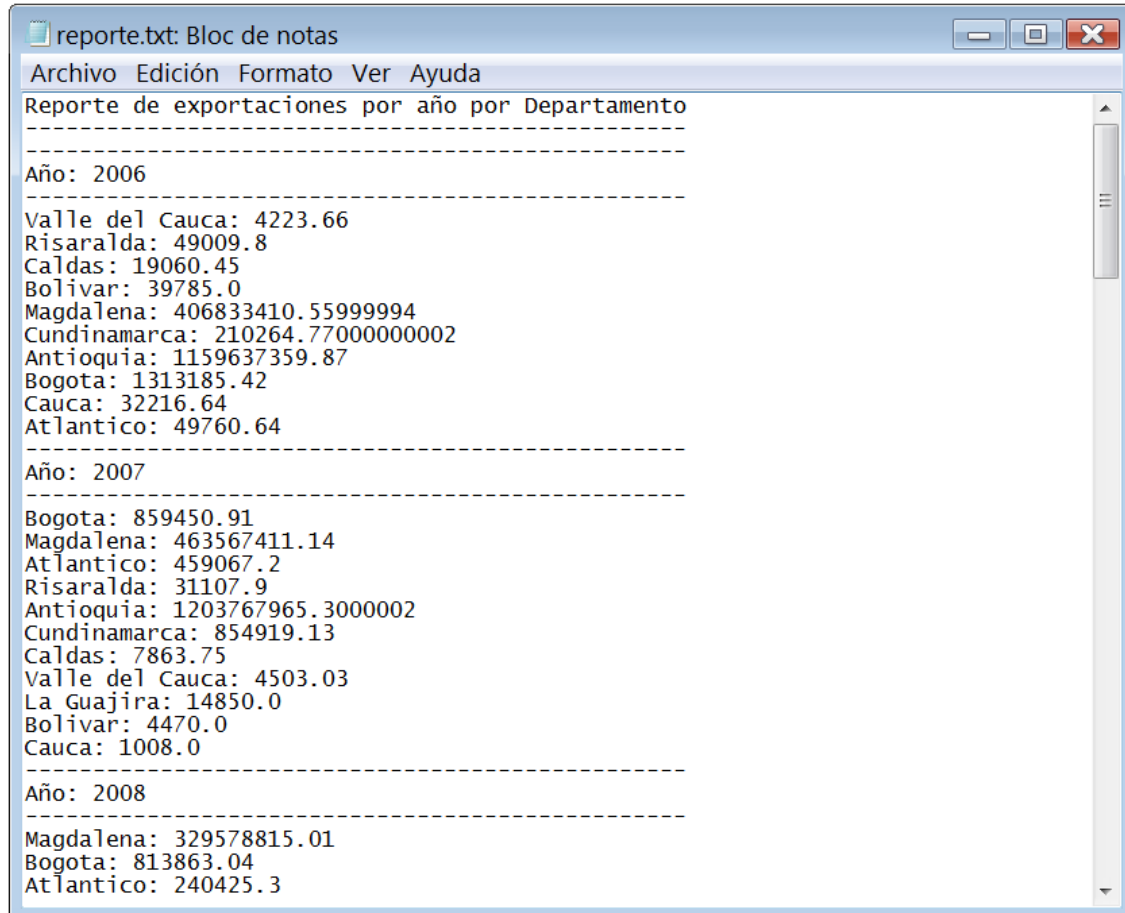
```
plt.figure()  
extracto = bananos.head(10)  
extracto.Cantidad.plot(kind = 'bar', figsize=(10,5))  
plt.show()
```

Cuyo resultado es:



Problema 3: Generar un reporte con la cantidad de kilos que han sido exportados a lo largo del tiempo por departamento

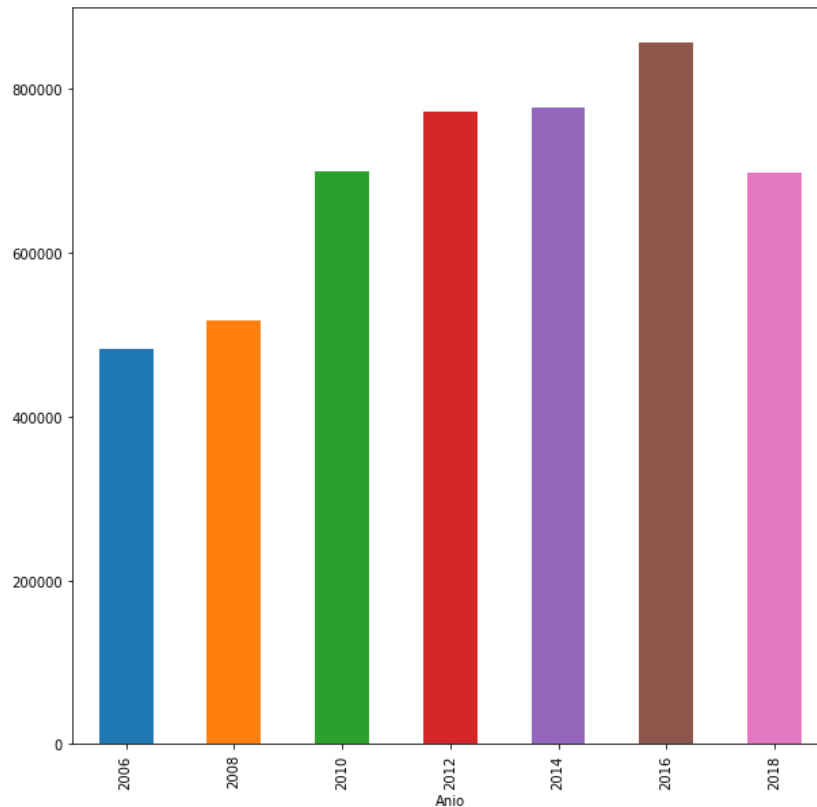
La siguiente figura muestra cómo debe quedar el reporte. En la figura sólo se muestran los primeros años. Su archivo debe contener todos los años. Si no recuerda cómo escribir en un archivo, revise la presentación de la clase N3-C7.



```
reporte.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda
Reporte de exportaciones por año por Departamento
-----
Año: 2006
-----
Valle del Cauca: 4223.66
Risaralda: 49009.8
Caldas: 19060.45
Bolívar: 39785.0
Magdalena: 406833410.55999994
Cundinamarca: 210264.770000000002
Antioquia: 1159637359.87
Bogotá: 1313185.42
Cauca: 32216.64
Atlántico: 49760.64
-----
Año: 2007
-----
Bogotá: 859450.91
Magdalena: 463567411.14
Atlántico: 459067.2
Risaralda: 31107.9
Antioquia: 1203767965.3000002
Cundinamarca: 854919.13
Caldas: 7863.75
Valle del Cauca: 4503.03
La Guajira: 14850.0
Bolívar: 4470.0
Cauca: 1008.0
-----
Año: 2008
-----
Magdalena: 329578815.01
Bogotá: 813863.04
Atlántico: 240425.3
```

Problema 4: Graficar en forma de barras, la evolución del mercado del banano a lo largo del tiempo (para los años pares)

El diagrama de barras (usando como métrica el valor en dólares) tiene el siguiente aspecto:

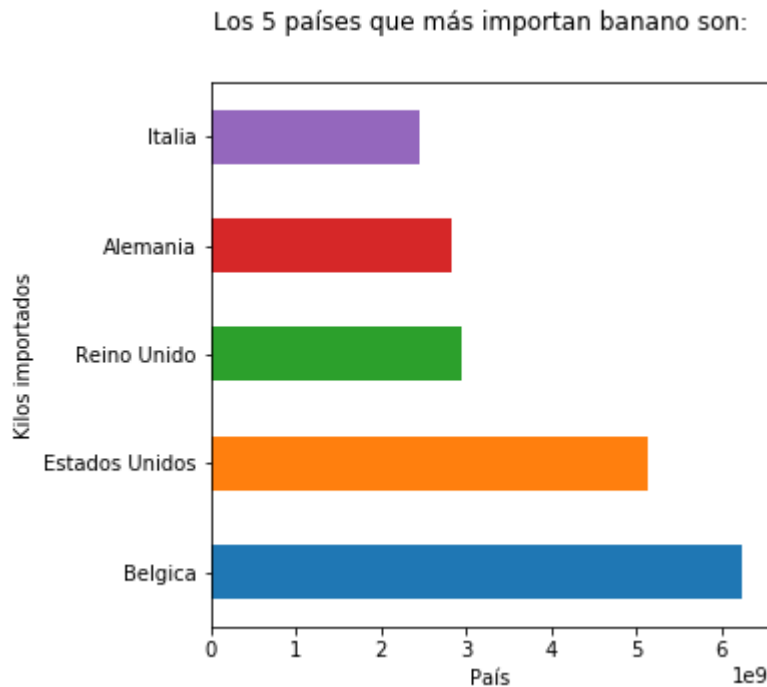


Recuerde que usted debe preguntar al usuario la métrica que desea utilizar para generar la gráfica entre: cantidad, valor en dólares o valor en pesos.

Problema 5: Encontrar y graficar los países que más cantidad importan banano y los que menos importan

Este último requerimiento consiste en informar al usuario cuáles son los 5 países a los cuales se exportó una mayor cantidad banano (en kilogramos) y los 5 a los que menos se exportó.

El resultado con los 5 países que más importaron debe verse así:



Note que al gráfico debe añadirse un título, así como a los ejes.

Ayuda:

- Para configurar el título del gráfico, utilice la función `suptitle` de `pyplot`.
- Para configurar los títulos de los ejes utilice los métodos `set_xlabel` y `set_ylabel`.

Entrega

Entregue el archivo modificado a través de BrightSpace en el laboratorio del nivel 4 designado como “**L2 Librerías Pandas y Matplotlib**”.