

Clase04_2_Procesando_planillas

April 16, 2024

1 Seminario de Lenguajes - Python

1.1 Procesamiento de hojas de cálculo

2 ¿Qué características tienen estos archivos?

Según la Wikipedia ... > Una hoja de cálculo es un tipo de documento que permite manipular datos numéricos y alfanuméricos dispuestos en forma de tablas compuestas por celdas, las cuales se suelen organizar en una matriz de filas y columnas.

El programa más popular para manipular estos archivos es Microsoft Excel

3 DESAFIO 1

Tenemos los datos de afluencia de turistas en la provincia de Córdoba y queremos saber cuántos turistas concurrieron entre los meses de enero y febrero.

Los datos fueron obtenidos del [Portal de datos abiertos de la provincia](#)

Como se puede ver, los datos están en varios formatos, pero no en un único dataset.

Para nuestro ejemplo, armamos un [único dataset](#) con los datos de cada año en una hoja distinta.

4 Procesamos el formato Excel

Una opción: librería [openpyxl](#)

```
[1]: from pathlib import Path
     from openpyxl import load_workbook
```

```
[3]: data_route = Path('ejemplos') / "clase04_2" / "Afluencia_turística_CORDOBA.xlsx"
```

Abrimos el archivo usando la función `load_workbook`

```
[4]: work_book = load_workbook(filename=data_route)
```

5 Obtenemos los nombres de las hojas del archivo

```
[5]: sheet_names = work_book.sheetnames
```

```
[6]: sheet_names
```

```
[6]: ['2023', '2022', '2021', '2020', '2019', '2018']
```

6 Procesamos las hojas de datos

Vamos a crear un diccionario con el nombre de la hoja como clave. En este caso, corresponden a los años desde 2018 a 2023.

Usamos `iter_rows`

```
[7]: summer_tourists = {}
```

```
[8]: for sheet_name in sheet_names:
    sheet = work_book[sheet_name]
    total_tourists = 0
    # Iterar sobre las filas y columnas de la hoja
    for row in sheet.iter_rows(values_only=True):
        if row[0] in ("Ene", "Feb"):
            total_tourists += row[10]

    summer_tourists[sheet_name] = total_tourists
```

7 Armamos el listado pedido

```
[9]: print("TOTAL DE TURISTAS EN ENERO Y FEBRERO")
    print("Año \tTotal de turistas")
    for year, data in summer_tourists.items():
        print(f"{year} \t{data}")
```

TOTAL DE TURISTAS EN ENERO Y FEBRERO

Año	Total de turistas
2023	167788
2022	156111
2021	103723
2020	204083
2019	198448
2018	203122

8 DESAFIO 2

Guardar los datos del listado previo en un archivo en formato json.

```
[10]: import json

      json_file= Path('ejemplos') / "clase04_2" / "turistas.json"

[11]: with open(json_file, "w") as file:
      json.dump(summer_tourists, file, indent=4)
```

9 DESAFIO 3

Queremos generar un archivo con formato csv de siguiente encabezado:

AÑO, VERANO, OTOÑO, INVIERNO, PRIMAVERA

con los totales de turistas en cada estación por año.

Considerar que: - VERANO, corresponde de diciembre a febrero - OTOÑO, corresponde de marzo a mayo - INVIERNO corresponde de junio a agosto - PRIMAVERA, corresponde de septiembre a noviembre.

```
[ ]: # Solución
```