

DIPLOMADO DE PROGRAMACIÓN CON PYTHON V3

TEMA 4 VISUALIZAR DATOS CON PANDAS

UNIBE

EDUCACIÓN
CONTINUA



Eliezer Figueroa
MCT, MOS, MCSA, MCSE, ITIL, SCF
3 Septiembre, 2019

COMPONENTES DEL TEMA

- ✓ VISUALIZAR DATOS CON PANDAS
- ✓ LIMPIEZA DE DATOS
- ✓ LEER CSV
- ✓ DATAFRAMES
- ✓ LEER EXCEL CON PANDAS
- ✓ ESCRIBIR EXCEL CON PANDAS



- Siendo Excel tan omnipresente, los profesionales de datos deben estar familiarizados con él.
- Trabajar con datos en Python o R ofrece ventajas serias sobre la interfaz de usuario de Excel
- Encontrar una manera de trabajar con Excel usando código es fundamental
- Ya hay una gran herramienta para usar Excel con Python llamada pandas.

A	H	I	J	K	L	M	N	O
President	National total votes	Total electoral votes	Rating points	Political Party	Occupation	College	% electoral	% popular
George Washington	NA()	69	842	None	Planter	None	100.0	NA()
John Adams	NA()	139	598	Federalist	Lawyer	Harvard	95.0	NA()
Thomas Jefferson	NA()	137	711	Democratic-Republican	Planter, Lawyer	William and Mary	53.3	NA()
James Madison	NA()	176	567	Democratic-Republican	Lawyer	Princeton	69.3	NA()
James Monroe	NA()	221	602	Democratic-Republican	Lawyer	William and Mary	82.8	NA()
John Quincy Adams	NA()	261	564	Democratic-Republican	Lawyer	Harvard	32.2	NA()
Andrew Jackson	1,148,018	261	632	Democrat	Lawyer	None	68.2	56.0
Martin Van Buren	1,503,534	294	429	Democrat	Lawyer	None	57.8	50.8
William Henry Harrison	2,411,808	294	329	Whig	Soldier	Hampden-Sydney	79.6	52.9
John Tyler	NA()	NA()	369	Whig	Lawyer	William and Mary	NA()	NA()
James K. Polk	2,703,659	275	632	Democrat	Lawyer	U. of North Carolina	61.8	49.5

- Este conjunto de datos contiene una lista de presidentes de EE. UU., partidos asociados, profesión y más.

```
from pandas import DataFrame, read_csv
import matplotlib.pyplot as plt
import pandas as pd

file = r'Presidents.xls'
df = pd.read_excel(file)

# plot data
colors = ['yellowgreen', 'gold', 'lightskyblue', 'lightcoral', 'red', 'green', 'blue', 'orange', 'white', 'brown']
df['Occupation'].value_counts().plot(kind='pie', title='Occupation by President', colors=colors)
plt.show()
```

- Podemos trazar datos de este gran archivo de Excel con unas pocas líneas de código. Seleccionamos la columna "Ocupación" para esta demostración.

```
from pandas import DataFrame, read_csv
import matplotlib.pyplot as plt
import pandas as pd

file = r'data/Presidents.xls'
df = pd.read_excel(file)

# plot data
#df = df.replace('NA()',0)
df = df[df['% popular'] != 'NA()']

print( df['% popular'] )
df['% popular'].plot(kind='hist', bins=8, title='Popularity by President', facecolor='blue', alpha=0.5, normed=1)
plt.show()
```

- También podemos trazar la popularidad del presidente. Antes de hacerlo, necesitamos limpiar algunos datos.
- Algunas de las celdas no contienen valores numéricos, podemos eliminarlas o reemplazarlas.

```
from pandas import DataFrame, read_csv
import matplotlib.pyplot as plt
import pandas as pd

file = r'highscore.csv'
df = pd.read_csv(file)
print(df)
```

- Las primeras líneas importan el módulo Pandas. El método **read_csv** carga los datos en una trama de datos Pandas que hemos denominado **df**.

```
from pandas import DataFrame, read_csv
import matplotlib.pyplot as plt
import pandas as pd

file = r'highscore.csv'
df = pd.read_csv(file)
print('Max', df['Highscore'].max())
print('Min', df['Highscore'].min())
```

- Una trama de datos se puede manipular
- Utilizando métodos, el mínimo y el máximo se pueden extraer fácilmente.


```
import pandas as pd
from pandas import ExcelWriter
from pandas import ExcelFile

df = pd.read_excel('File.xlsx', sheetname='Sheet1')

print("Column headings:")
print(df.columns)
```

- Importamos el módulo pandas, incluyendo **ExcelFile**.
- El método **read_excel()** lee los datos en un marco de datos Pandas, donde el primer parámetro es el nombre de archivo y el segundo parámetro es la hoja.

```
import pandas as pd
from pandas import ExcelWriter
from pandas import ExcelFile
import numpy as np

df = pd.DataFrame({'a': [1,3,5,7,4,5,6,4,7,8,9], 'b': [3,5,6,2,4,6,7,8,7,8,9]})

writer = ExcelWriter('Pandas-Example2.xlsx')
df.to_excel(writer, 'Sheet1', index=False)
writer.save()
```

- Comenzamos importando el módulo pandas. Desde el módulo importamos **ExcelWriter** y **ExcelFile**.
- El siguiente paso es crear un marco de datos. En el marco de datos ponemos una lista, con el nombre de la lista como primer argumento.

- Pythonspot. (2019). Visualize data with Pandas. Retrieved 16 September, 2019, from <https://pythonspot.com/visualize-data-with-pandas/>
- Pythonspot. (2019). Pandas read CSV. Retrieved 16 September, 2019, from <https://pythonspot.com/pandas-read-csv/>

MUCHAS GRACIAS

Eliezer Figueroa
MCT,MOS,MCSA,MCSE,ITIL,SCF

UNIBE

EDUCACIÓN
CONTINUA

