

Sesión 7: Repaso

¿Qué hemos visto hasta el momento?

Tema 1: Estructuras de datos*

1. Integer: edad=18
2. Float: altura=1.76
3. String: nombre="Maria"
4. Lista: clases=["Python", "Intro_CPOL", "Taller_Metodología"]
5. Diccionario: estudiante = {'nombre' : 'Maria', 'edad' : 18, 'altura': 1.76, 'cursos': ["Python", "Intro_CPOL", "Taller_Metodología"]}
6. tuplas: ("juan","carolina","camila")

*Más información en los scripts de la primera y segunda sesión

Tema 1: Estructuras de datos

Objeto	métodos	indexación	mutabilidad
Listas:	.append() , .index() , .sort()	list[0]	si
String	.upper(), .count(), .split()	string[0]	no
Diccionarios	.keys() .values()	dic["key"]	si
Tuplas	.count(), .index()	tpla[0]	no

*Más información en los scripts de la primera y segunda sesión

Tema 2: Funciones usando Loops y Condicionales

Estructura de una función:

```
def NOMBRE(LISTA DE ARGUMENTOS):  
    ENUNCIADOS  
    print(VALOR)
```

Elementos para tener en cuenta:

- Cuidado con las sangrías
- if, elif, else
- no se necesita terminar con print.

Ejemplo de una función usando
condicionales

```
def mayor_o_menor(x, y):  
    if x < y:  
        print(x, "es menor que", y)  
    elif x > y:  
        print(x, "es mayor que", y)  
    else:  
        print(x, "y", y, "son iguales")
```

*Más información en los scripts de la segunda y la sesión

Tema 2: Funciones usando Loops y Condicionales

Ejemplo de una función usando loops

Estructura de una función:

```
def NOMBRE(LISTA DE ARGUMENTOS):  
    ENUNCIADOS  
    return VALOR
```

Elementos para tener en cuenta:

- Cuidado con las sangrías
- for __ in ____
- varias formas de hacerlo

```
names=["María", "Margarita"]
```

```
def familia_forero(lista):  
    for i in range(0, len (lista)):  
        names[i]= names[i] + " Forero"  
    return(lista)
```

```
familia_forero(names)
```

```
['María Forero', 'Margarita Forero']
```

*Más información en los scripts de la segunda y cuarta sesión

Tema 3: Manejo de Bases de Datos con Pandas

Pandas es un paquete de python para manipular bases de datos

Sesión 3:

- Los elementos básicos de los Dataframes
- Aprendimos a Filtrar por columna y por fila
- Aprendimos a crear nuevas columnas
- Sacar estadísticas descriptivas
- Cargar bases de datos (excel)

Sesión 4:

- Manipulación de datos (.groupby, .reset_index, .sum, .npwhere, .stack ,etc)

	names	ideology	age	votes
0	Duque	D	43	7.6
1	Petro	I	59	4.8
2	Fajardo	C	63	4.6
3	DeLaCalle	None	73	0.4
4	Vargas	D	57	1.7

Tema 3: Manejo de Bases de Datos con Pandas

Ejemplos de comandos aprendidos: Crear nuevas variables

```
elecciones_2018["votes_m"] = elecciones_2018.votes * 1000000  
elecciones_2018
```

	names	ideology	age	votes	votes_m
0	Duque	D	43	7.6	7600000.0
1	Petro	I	59	4.8	4800000.0
2	Fajardo	C	63	4.6	4600000.0
3	DeLaCalle	None	73	0.4	400000.0
4	Vargas	D	57	1.7	1700000.0

Tema 3: Manejo de Bases de Datos con Pandas

Ejemplos de comandos aprendidos: filtrar por filas y columnas

```
elecciones_2018[['names']]
```

	names
0	Duque
1	Petro
2	Fajardo
3	DeLaCalle
4	Vargas

```
elecciones_2018.loc[elecciones_2018['names']=='Vargas']
```

	names	ideology	age	votes
4	Vargas	D	57	1.7

Tema 4: Visualización de Datos

matplotlib es un paquete de python para visualizar bases de datos

Sesión 4: matplotlib

- Pie charts de datos cualitativos (umbral)
- Gráficos de barra (Votos por Partido)
- Pie Charts de datos cuantitativos (distribución de Curules)

Sesión 5: altair

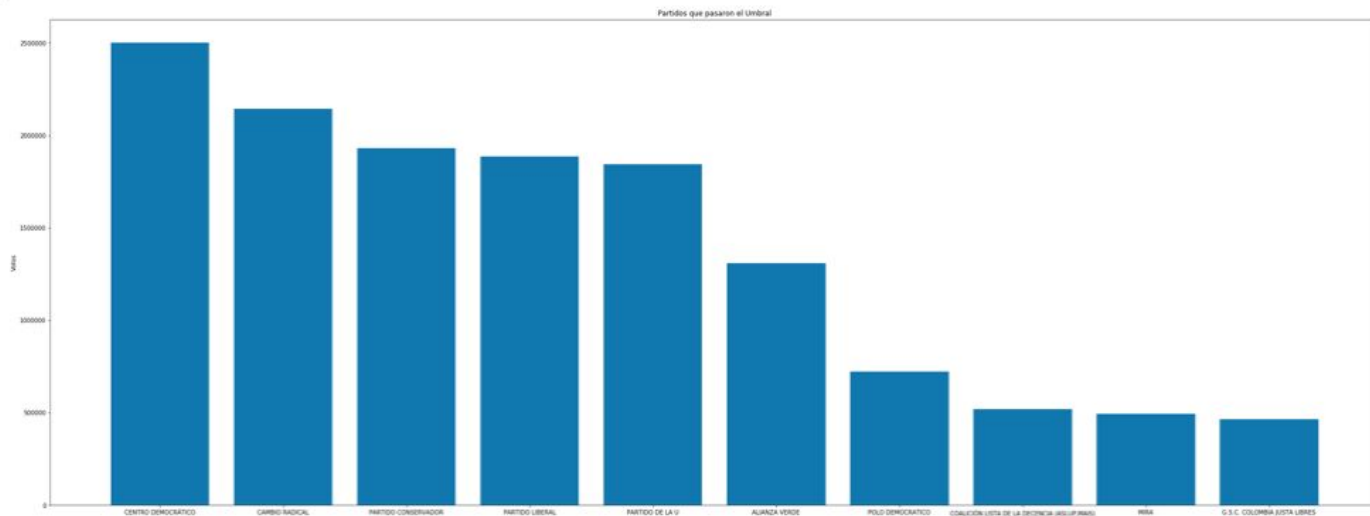
- Series de tiempo

Más información: [Matplotlib.org](https://matplotlib.org), altair-vlз.github.io

Tema 4: Visualización de Datos

Ejemplos de comandos aprendidos: Barras

```
plt.figure(figsize=(40,15))  
plt.bar(pasaron_umbral.codpartido, pasaron_umbral.votos, align='center', alpha=1)  
plt.ylabel('Votos')  
plt.title('Partidos que pasaron el Umbral')  
plt.show()
```

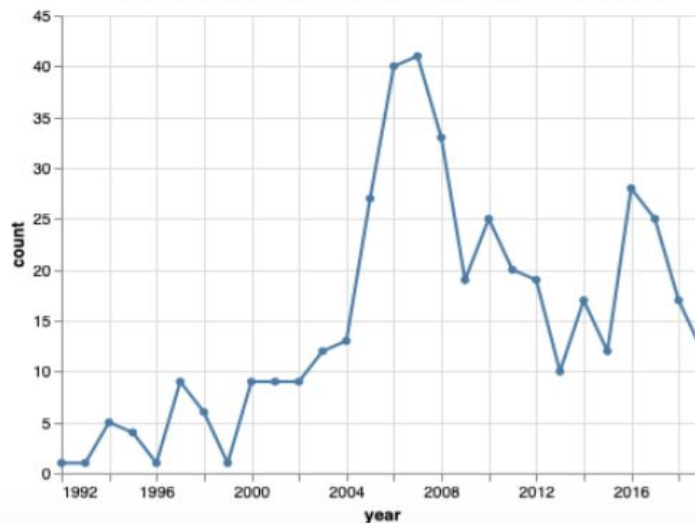


Tema 4: Visualización de Datos

Ejemplos de comandos aprendidos: Serie de tiempo con altair

```
In [25]: alt.Chart(counts_df).mark_line(point=True).encode(  
    x='year',  
    y='count',  
).properties(  
    title='Fallos de la Corte Constitucional de Colombia sobre aborto'  
)
```

Out[25]: **Fallos de la Corte Constitucional de Colombia sobre aborto**



Tema 5: WebScrapping

BeautifulSoup es un paquete de python para descargar datos de Internet

Sesión 5:

- Manipular carpetas de nuestro computador
- Leer y conocer código fuente de las páginas web
- Descargar archivos desde una página web
- Almacenar automáticamente esos documentos

Tema 6: Text Mining

Spacy es un paquete de python para hacer análisis de lenguaje

WordCloud es un paquete de python para hacer nubes de palabras

Sesión 6:

- Extraer el texto de archivos tipo html
- Hacer nubes de palabras por frecuencias
- Utilizar modelos de NLP para determinar el tipo de palabra (Verbo, Sustantivo, etc)

Tema 6: Text Mining

Ejemplos de comandos aprendidos: WordCloud con Spacy en la sesión 6

```
In [29]: wordcloud = WordCloud(background_color='white', stopwords=stopwords_es).generate(texto_limpio)

# plot the WordCloud image
plt.figure()
plt.imshow(wordcloud)
plt.axis("off")
plt.savefig('./gráfica1.png')
plt.show()
```

