



# UNIVERSIDAD AUTÓNOMA DE CHIAPAS

## FACULTAD DE NEGOCIOS

Lic. Ing. En Desarrollo y Tecnologías de  
Software

**Docente**

Dr. Castillo Estrada Christian Mauricio

**Materia**

Big Data

**Práctica 3**

Aplicar el proceso ETL para generar un Mapa de recomendaciones

**Alumno**

Adamari Martínez Ruiz

**Semestre**

Octavo

**Grupo**

D

17 de septiembre de 2023

# Impresión

```
File Edit Selection View Go Run Terminal Help ← → 🔍 tarea3
```

```
◆ museo.py X ◆ movies.py
```

```
◆ museo.py > ...
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from PIL import Image as PILImage
5 import io
6 import folium
7 from reportlab.lib.pagesizes import letter
8 from reportlab.platypus import SimpleDocTemplate, Image, Paragraph
9 from reportlab.lib.styles import getSampleStyleSheet
10
11 museo = pd.read_csv("new-york-city-art-galleries-1.csv")
12 print("\n\t(LISTADORA - MUSEOS & GALERIAS)\n",museo)
13
14 print("\n1.- Eliminar filas o registros duplicados:")
15 museo = museo.drop_duplicates()
16 print(museo.head(10))
17
18 print("\n2.- Eliminar la columna Address2:")
19 museo = museo.drop(['ADDRESS2'], axis=1) #axis: para indicar que deseo eliminar una columna
20 print(museo.head())
21
22 print("\n3.- Seleccionar los museos o galerías de arte del distrito Brooklyn:")
23 museo = museo.query("CITY == 'Brooklyn'")
24 print(museo['CITY'].head(10))
25
26 print("\n4.- Filtrar a los 3 mejor calificados (museo o galería)")
27 museo = museo.loc[museo['GRADING'] >= 8.2] #loc: nos permite obtener un conjunto de filas >>
28 #se selecciona a la columna y se comprueba el valor dado
29 museo3 = museo.head(3)
30 print(museo3)
31
32 print("\n5.- Generar imagen de Mapa indicando la ubicación de los 3 mejores")
33 mapa_museo = folium.Map(location=[40.7037979,-74.0202391], zoom_start=13)
34
```

```
File Edit Selection View Go Run Terminal Help ← → 🔍 tarea3
```

```
◆ museo.py X ◆ movies.py
```

```
◆ museo.py > ...
34
35 museo3 = museo.head(3)
36 for indice, row in museo3.iterrows():
37     name = row['NAME']
38     address = row['ADDRESS1']
39     lat, lon = map(float, row['the_geom'].replace('POINT (', '').replace(')', '').split())
40     lat, lon = round(lon, 4), round(lat, 4)
41     rating = row['GRADING']
42
43     popup_text = f"Name: {name}\naddress: {address} \nGrade: {rating}"
44     folium.Marker(location=[float(lat), float(lon)], popup=popup_text).add_to(mapa_museo)
45
46 temp_html_file = 'mapa.html'
47 mapa_museo.save(temp_html_file)
48
49 img_data = mapa_museo.to_png()
50 img = PILImage.open(io.BytesIO(img_data))
51 img.save('mapa.png')
52
53 doc = SimpleDocTemplate("ReporteFinal.pdf", pagesize=letter)
54
55 styles = getSampleStyleSheet()
56 contenido = []
57
```

```
File Edit Selection View Go Run Terminal Help ← → 🔍 tarea3
```

```
◆ museo.py X ◆ movies.py
```

```
◆ museo.py > ...
50 img = PILImage.open(io.BytesIO(img_data))
51 img.save('mapa.png')
52
53 doc = SimpleDocTemplate("ReporteFinal.pdf", pagesize=letter)
54
55 styles = getSampleStyleSheet()
56 contenido = []
57
58 titulo = "The best Galleries and Museums\n"
59 contenido.append(Paragraph(titulo, styles["Title"]))
60
61 imagen_mapa = Image("mapa.png", width=400, height=200)
62 contenido.append(imagen_mapa)
63
64 for index, row in museo3.iterrows():
65     name = row['NAME']
66     address = row['ADDRESS1']
67     rating = row['GRADING']
68
69     # Crear un párrafo de texto con los datos
70     texto_datos = f"\nName: {name} \naddress: {address} \nGrade: {rating}"
71     contenido.append(Paragraph(texto_datos, styles["Normal"]))
72
73
74 doc.build(contenido)
75 museo.to_csv('Galerias&Museo.csv', index=False)
```

## 1.- Eliminar filas o registros duplicados

Nos fijamos bien cuales son las duplicadas (duplicates()) y después de comprobarlos se hace la eliminación

```
1.- Eliminar filas o registros duplicados:
0      O'reilly William & Co Ltd POINT (-73.96273074561996 40.773800871637576) (212) 396-1822 ... New York 10021 4.1
1      Organization of Independent Artists - Gallery 402 POINT (-74.0093857886298 40.71646850931738) (212) 219-9213 ... New York 10013 5.2
2      Owen Gallery POINT (-73.96435178145856 40.77408008522409) (212) 879-2415 ... New York 10021 4.3
3      P P O W Gallerie POINT (-74.00389261044238 40.74958540084406) (212) 647-1044 ... New York 10001 7.2
4      P P O W Inc POINT (-74.00176389614576 40.72290742771331) (212) 941-8642 ... New York 10013 5.1
..      ...
155     Sigrid Freunderfer Fine Art POINT (-73.96809592587059 40.76848322657678) (212) 517-9709 ... New York 10021 2.6
156     Silverstein Gallery Inc POINT (-74.00610391839206 40.7465866957298) (212) 929-4300 ... New York 10011 2.6
157     Simply Marvelous Art POINT (-73.79975655329956 40.70412220555415) (718) 291-9862 ... Jamaica 11432 2.5
159     Skoto Gallery POINT (-74.00685030055344 40.746676581928234) (212) 352-8058 ... New York 10011 2.4
160     Sloan McGill collection POINT (-73.95565084132937 40.65961513441343) (718) 469-1930 ... Brooklyn 11225 2.4

[160 rows x 9 columns]
```

## 2.- Eliminar la columna Address 2

Hacemos la eliminación de una columna, como se podrá ver en la anterior columna habían 9 columnas ahora hay 8.

```
2.- Eliminar la columna Address2:
0      O'reilly William & Co Ltd POINT (-73.96273074561996 40.773800871637576) (212) 396-1822 ... New York 10021 4.1
1      Organization of Independent Artists - Gallery 402 POINT (-74.0093857886298 40.71646850931738) (212) 219-9213 ... New York 10013 5.2
2      Owen Gallery POINT (-73.96435178145856 40.77408008522409) (212) 879-2415 ... New York 10021 4.3
3      P P O W Gallerie POINT (-74.00389261044238 40.74958540084406) (212) 647-1044 ... New York 10001 7.2
4      P P O W Inc POINT (-74.00176389614576 40.72290742771331) (212) 941-8642 ... New York 10013 5.1

[5 rows x 8 columns]
```

## 3.- Seleccionar a los museos o galerías de arte del distrito Brooklyn

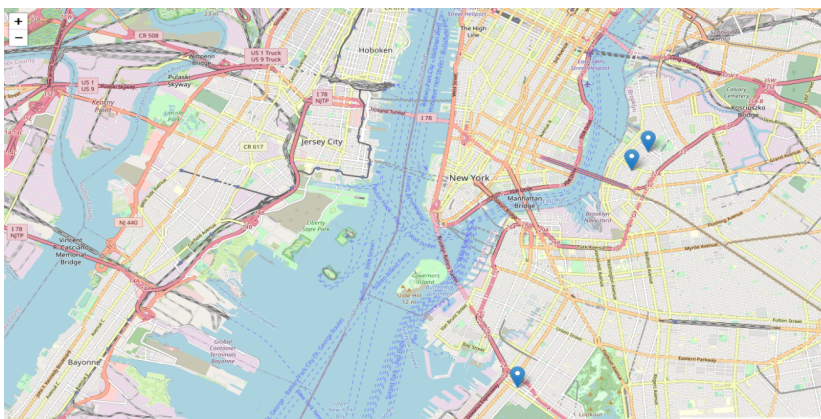
```
3.- Seleccionar los museos o galerías de arte del distrito Brooklyn:
19      Brooklyn
39      Brooklyn
40      Brooklyn
43      Brooklyn
49      Brooklyn
53      Brooklyn
63      Brooklyn
73      Brooklyn
79      Brooklyn
104     Brooklyn
Name: CITY, dtype: object
```

## 4.- Filtrar a los 3 mejor calificados (museo o galería)

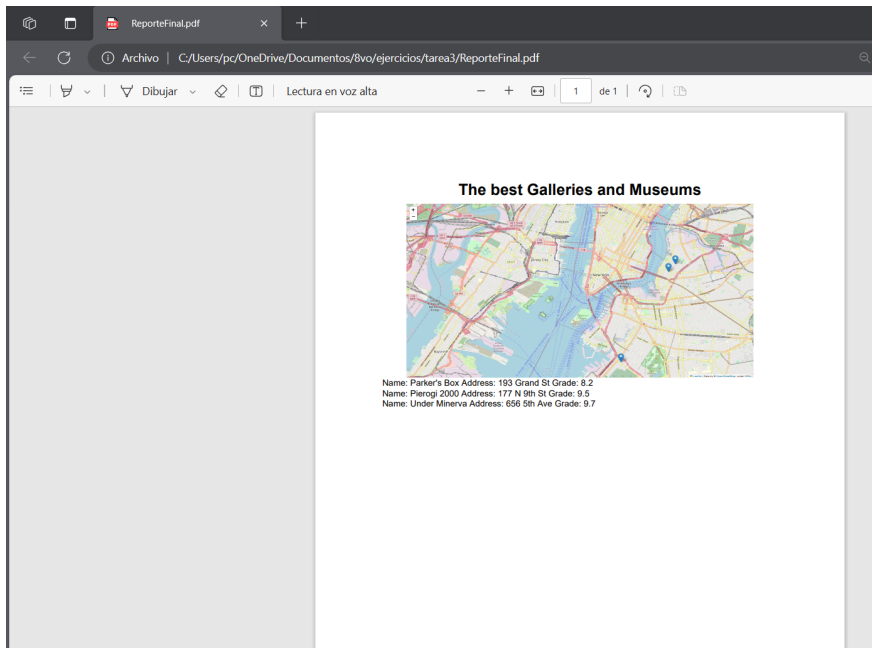
```
4.- Filtrar a los 3 mejor calificados (museo o galería)
19      Panker's Box POINT (-73.9604447082016 40.714300158269266) (718) 388-2882 ... Brooklyn 11211 8.2
40      Pierogi 2000 POINT (-73.95578160004416 40.718750345970065) (718) 599-2144 ... Brooklyn 11211 9.5
242     Under Minerva POINT (-73.99264059463559 40.66242444886783) (718) 788-0170 ... Brooklyn 11215 9.7

[3 rows x 8 columns]
```

## 5. Generar imagen de Mapa indicando la ubicación de los 3 mejores



## 6.- Generar el reporte final en un archivo de tipo PDF



Nombre	Fecha de modificación	Tipo	Tamaño
Galerias&Museo	17/09/2023 23:40	Archivo de valores...	
geckodriver	17/09/2023 23:40	Documento de tex...	
mapa	17/09/2023 23:40	Archivo HTML	
mapa	17/09/2023 23:40	Archivo PNG	
mirror	15/09/2023 00:13	Archivo de origen ...	
museo	17/09/2023 23:40	Archivo de origen ...	
new-york-city-art-galleries-1	12/09/2023 23:34	Archivo de valores...	
Reportefinal	17/09/2023 23:40	Microsoft Edge PD...	

URL GitHub

<https://github.com/adamari-th/BigData>