

Minería de datos: Clase 22 de agosto

Orden del día

1. Comentarios Tarea 2
2. Más sobre funciones y guardar figuras
3. Primeros ejemplos con bases de datos (csv to dataframe)
4. Explicación Tarea 3

Comentarios Tarea 2

1. A partir de ahora todas las definiciones son "definiciones matemáticas". Por ejemplo pueden usar el libro Curso intermedio de Probabilidad de Luis Rincón (recurso en línea)

Tarea 3: Ejercicio 1

1. Replicar una gráfica con los precios de 5 acciones. Deben utilizar datos diarios del año en curso. Van a poner en el reporte la captura de pantalla de los precios en alguna plataforma (Yahoo finanzas) y su gráfica hecha en Python. En el reporte deben incluir cómo se obtuvieron los datos y los pasaron a Python.
2. Utilizar etiquetas, colores, ejes...

Apple Inc. (AAPL) ☆

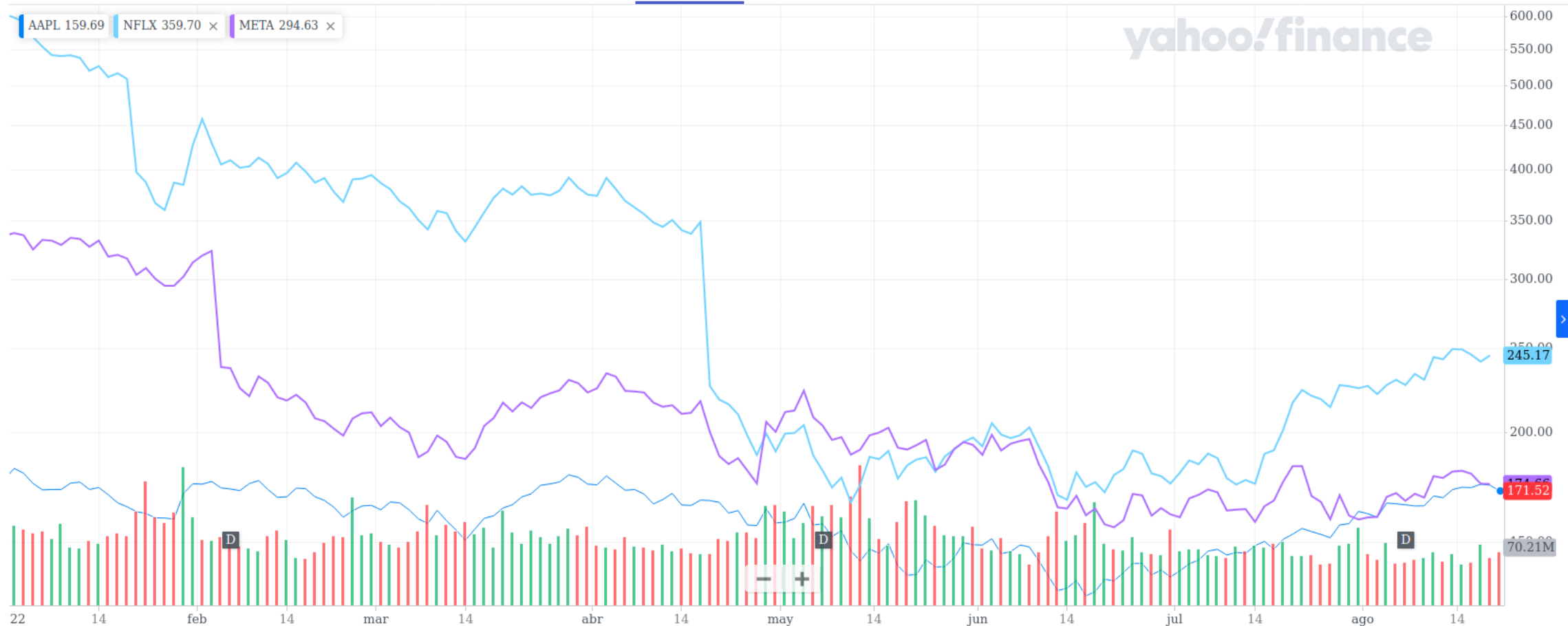
NasdaqGS - NasdaqGS Precio en tiempo real. Divisa en USD.

171.52 -2.63 (-1.51%) **170.75** -0.77 (-0.45%)

Al cierre: 19 de agosto 04:00PM EDT

Fuera de horario: 19 ago., 07:59PM EDT

⊕ Indicadores ⊕ Comparación ⚡ Eventos 📅 Rango de fechas 1 d 5 d 1 M 3 m 6 m **REND. AC. ANUAL** 1A 2A 5A Máx. 📄 Intervalo 1D 📈 De líneas 🖨 Dibujo ⚙ Configuración 📶



Tarea 3: Ejercicio 2

1. Investiga la serie de potencias que caracteriza la función exponencial
2. Crea una función llamada `exp_serie(x,n)` tal que recibe un valor real x y un número natural n y que regrese el valor de los primeros n términos de la serie
3. Grafica la aproximación junto con la función exponencial para los valores de $n=1,2,3,4,5$
4. Cada gráfica debe ser informativa (etiquetas, colores, ejes,...) y debe guardarse en pdf

Tarea 3: Ejercicio 3

1. Investiga la matriz de adyacencia del grafo camino
2. Crea una función llamada `mat_path(n)` tal que recibe un número natural n y regrese la matriz de adyacencia del grafo camino con n vértices.
3. Para $n= 500$ y 1000 , calcula los valores propios de dicha matriz y crea histogramas de dichos valores
4. Genera una lista tal que para cada k desde 1 hasta n tenga el valor de $2 \cos \frac{k\pi}{n+1}$
5. Crea histograma de dicha listas con $n=500$ y 1000
6. Verifica que los histogramas son similares, ¿qué puedes conjeturar al respecto?

Tarea 3

1. Solo se entrega un reporte por equipos (`reporte_tarea3.pdf`) y un script (`tarea3.py`) con todo el código en Python.
2. El reporte debe incluir un resumen de lo que se hace en cada ejercicio
3. Para el ejercicio 1 deben incluir información de cómo descargaron los datos y los pasaron a Python
4. Para el ejercicio 2 deben explicar qué es una aproximación en serie de potencias y usar como ejemplo la función exponencial
5. Para el ejercicio 3 deben explicar qué es un grafo, su matriz de adyacencia, y agregar sus ejemplos explicando el espectro de dicho grafo.

Tarea 3

1. El archivo tarea3.py debe tener todo el código en Python.
2. Al ejecutarlo completo debe guardar 6 gráficas y 4 histogramas en pdf
3. 1 gráfica para ejercicio 1, 5 gráficas para el ejercicio 2, y 4 histogramas para el ejercicio 3
4. El código debe estar comentado
5. Enviar la tarea a más tardar el miércoles 31 de agosto a las 11:59am
6. Enviar ambos archivos en una carpeta comprimida con copia a los integrantes del equipo
7. Equipos de 4 o 5 personas, no voy a recibir tareas individuales

¿Dudas o preguntas?

Spyder (Python 3.9)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\Guillermo\Dropbox\UDLAP\mineria\Codigo\agosto_22

C:\Users\Guillermo\Dropbox\UDLAP\mineria\Codigo\agosto_22\agosto_22.py

agosto_22.py x

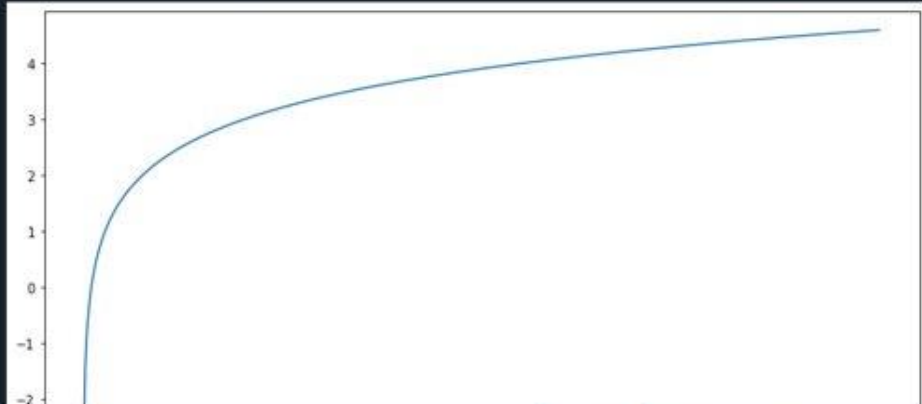
```
1 #!/usr/bin/env python3
2 # -*- coding: utf-8 -*-
3 """
4 Created on Sun Aug 21 14:22:34 2022
5
6 @author: guillermo
7 """
8
9 import numpy as np
10 import pandas as pd
11 import matplotlib.pyplot as plt
12
13 #evaluación de la función por lista
14 valores_x=np.arange(0.1,100,0.1)
15 valores_y=np.log(valores_x)
16 fig = plt.figure(figsize=(12,6))
17 plt.plot(valores_x,valores_y)
18 plt.show()
19 #guardar figura en pdf
20 fig.savefig('logaritmo.pdf', format='pdf', dpi=1200)
21
22
23
24
25
26
27
28
29
30
31
32
33 #cargar base de datos en dataframe
34 df_apoyo = pd.read_csv("mujeres-afiliadas-programa-apoyo-productivo-mujer-poblana-20180228.csv")
35
36 #información por columnas
37 print(df_apoyo[['Municipio', 'Edad']])
38
39 #estadísticas por columnas
40 print("Edad promedio: ", np.mean(df_apoyo["Edad"]))
41 print("Edad máxima: ", np.max(df_apoyo["Edad"]))
42 print("Edad mínima: ", np.min(df_apoyo["Edad"]))
43
44
45 #información por ID
46 print(df_apoyo.loc[0], "\n\n", df_apoyo.loc[1] )
47
48 for i in range(10):
49     print(df_apoyo.loc[i], "\n\n" )
50
51 #Resumen generalizado
52 df_apoyo.describe()
53
54 """
55 Decir cuántas mujeres hay por municipio
56 """
```

Name	Type	Size	Value
df_apoyo	DataFrame	(5010, 5)	Column names: N°, Municipio, Etapa, Edad, Género
df_geo	DataFrame	(2272, 2)	Column names: Latitud, Longitud
df_reconstruccion	DataFrame	(2272, 11)	Column names: No., Nombre de municipio, Nombre de localidad, Clave del ...
fig	figure.Figure	1	Figure object of matplotlib.figure module
i	int	1	9
valores_x	Array of float64	(999,)	[0.1 0.2 0.3 ... 99.7 99.8 99.9]
valores_y	Array of float64	(999,)	[-2.30258509 -1.60943791 -1.2039728 ... 4.60216568 4.60316818 4.6 ...]

Help Variable Explorer Plots Files

Console 1/A x

In [4]: runfile('C:/Users/Guillermo/Dropbox/UDLAP/mineria/Codigo/agosto_22/agosto_22.py', wdir='C:/Users/Guillermo/Dropbox/UDLAP/mineria/Codigo/agosto_22')



Python console History

1 C:\Python\python.exe C:\Python\Scripts\spyder.py --file C:\Users\Guillermo\Dropbox\UDLAP\mineria\Codigo\agosto_22\agosto_22.py --line 23 Col 1 UTF-8 LF RW Mem 51%

```

29
30
31
32
33 #cargar base de datos en dataframe
34 df_apoyo = pd.read_csv("mujeres-afiliadas-programa-apoyo-productivo-mujer-poblana-20180228.csv")
35
36 #información por columnas
37 print(df_apoyo[['Municipio', 'Edad']])
38
39 #estadísticas por columnas
40 print("Edad promedio: ", np.mean(df_apoyo["Edad"]))
41 print("Edad máxima: ", np.max(df_apoyo["Edad"]))
42 print("Edad mínima: ", np.min(df_apoyo["Edad"]))
43
44
45 #información por ID
46 print(df_apoyo.loc[0], "\n\n" , df_apoyo.loc[1] )
47
48 for i in range(10):
49     print(df_apoyo.loc[i], "\n\n" )
50
51 #Resumen generalizado
52 df_apoyo.describe()
53
54 """
55 Decir cuántas mujeres hay por municipio
56 """
57
58 """
59 Decir cuántos municipios están contemplados en la base de datos
60 """
61
62 """
63 Decir la edad promedio por cada etapa
64 """
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83 #Ejemplo datos geográficos
84 df_reconstruccion = pd.read_csv("avance-reconstruccion-escuelas-afectadas-sismo-19s-20191113.csv")

```

Name	Type	Size	Value
df_apoyo	DataFrame	(5010, 5)	Column names: N°, Municipio, Etapa, Edad, Género
df_geo	DataFrame	(2272, 2)	Column names: Latitud, Longitud
df_reconstruccion	DataFrame	(2272, 11)	Column names: No., Nombre de municipio, Nombre de localidad, Clave del ...
fig	figure.Figure	1	Figure object of matplotlib.figure module
i	int	1	9
valores_x	Array of float64	(999,)	[0.1 0.2 0.3 ... 99.7 99.8 99.9]
valores_y	Array of float64	(999,)	[-2.30258509 -1.60943791 -1.2039728 ... 4.60216568 4.60316818 4.6 ...

Help Variable Explorer Plots Files

```

Console I/A x
Municipio Edad
0 ACATLAN DE OSORIO 23
1 ACATLAN DE OSORIO 40
2 ACATLAN DE OSORIO 63
3 ACATLAN DE OSORIO 22
4 ACATLAN DE OSORIO 48
...
5005 ZACATLAN 43
5006 ZACATLAN 22
5007 ZACATLAN 23
5008 ZACATLAN 29
5009 ZACATLAN 57
...
[5010 rows x 2 columns]
Edad promedio: 39.50698602794411
Edad máxima: 97
Edad mínima: 18
N° 1
Municipio ACATLAN DE OSORIO
Etapa 1
Edad 23
Género Femenino
Name: 0, dtype: object

N° 2
Municipio ACATLAN DE OSORIO

```

IPython console History

Spyder (Python 3.9)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\Guillermo\Dropbox\UDLAP\mineria\Codigo\agosto_22

C:\Users\Guillermo\Dropbox\UDLAP\mineria\Codigo\agosto_22\agosto_22.py

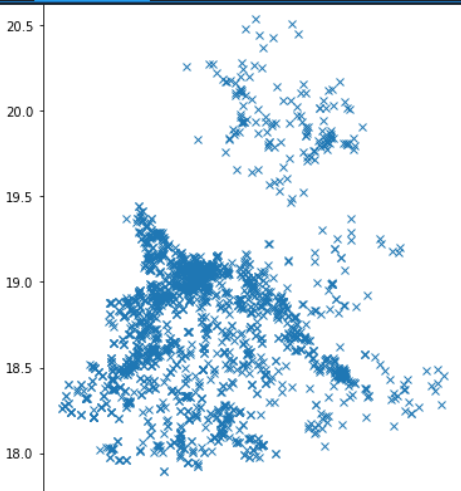
agosto_22.py x

```
51 #Resumen generalizado
52 df_apoyo.describe()
53
54 """
55 Decir cuántas mujeres hay por municipio
56 """
57
58 """
59 Decir cuántos municipios están contemplados en la base de datos
60 """
61
62 """
63 Decir la edad promedio por cada etapa
64 """
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83 #Ejemplo datos geográficos
84 df_reconstruccion = pd.read_csv("avance-reconstruccion-escuelas-afectadas-sismo-19s-20191113.csv")
85 df_geo=df_reconstruccion[["Latitud","Longitud"]]
86
87 fig = plt.figure(figsize=(6,8))
88 plt.plot(df_geo["Longitud"], df_geo["Latitud"], linestyle='None', marker="x")
89 plt.show()
90
91
92 """
93 Ejercicio replicar la figura pero en color rojo los datos que tienen daño mayor en rojo y daño menor en a.
94 """
95
96
97 """
98 Ejercicio replicar la figura pero solo graficar los datos que no tienen el 100% de avance
99 """
100
101
102 fig = plt.figure(figsize=(6,8))
103 plt.plot(df_geo["Longitud"], df_geo["Latitud"], linestyle='None', marker="x", alpha=0.5, color="green")
104 plt.show()
105
```

Name	Type	Size	Value
df_apoyo	DataFrame	(5010, 5)	Column names: N°, Municipio, Etapa, Edad, Género
df_geo	DataFrame	(2272, 2)	Column names: Latitud, Longitud
df_reconstruccion	DataFrame	(2272, 11)	Column names: No., Nombre de municipio, Nombre de localidad, Clave del ...
fig	figure.Figure	1	Figure object of matplotlib.figure module
i	int	1	9
valores_x	Array of float64	(999,)	[0.1 0.2 0.3 ... 99.7 99.8 99.9]
valores_y	Array of float64	(999,)	[-2.30258509 -1.60943791 -1.2039728 ... 4.60216568 4.60316818 4.6 ...]

Help Variable Explorer Plots Files

Console 1/A x



IPython console History

 Data API

Fields

Nº	Muni...	Eta	Edad	Géne...
1	ACAT...	1	23	Feme...
2	ACAT...	1	40	Feme...
3	ACAT...	1	63	Feme...
4	ACAT...	1	22	Feme...
5	ACAT...	1	48	Feme...
6	ACAT...	1	45	Feme...
7	ACAT...	1	48	Feme...
8	ACAT...	1	50	Feme...
9	ACAT...	1	33	Feme...
10	CHILA	1	30	Feme...
11	ACAT...	1	37	Feme...
12	ACAT...	1	39	Feme...
13	ACAT...	1	71	Feme...
14	ACAT...	1	27	Feme...

Home / Datasets / Avance de reconstrucción de escuelas afectadas por el sismo del 19S / 20191113 Avance de reconstrucción de escuelas afectadas por el sismo del 19S

Ver Regresar al dataset Descargar

20191113 Avance de reconstrucción de escuelas afectadas por el sismo del 19S

Nov 13, 2019



avance-reconstruccion-escuelas-afectadas-sismo-19s-20191113.csv

No.	Nom...	Nom...	Clav...	Nom...	Matr...	Sem...	Prog...	Porc...	Latit...	Long...
1	ATLIX...	ATLIX...	21DT...	IGNA...	17	DAÑO...	ESCU...	100	18.91...	-98.5...
2	ATLIX...	TOLO...	21DP...	FERN...	25	DAÑO...	ESCU...	100	18.92...	-98.3...
3	ATLIX...	ATLIX...	21EP...	JOSE ...	36	DAÑO...	ESCU...	100	18.90...	-98.4...
4	TILAPA	SAN F...	21ET...	DEFE...	50	DAÑO...	RECU...	100	18.62...	-98.5...
5	XOCH...	XOCH...	21ET...	SOCR...	63	DAÑO...	RECU...	100	18.65...	-98.3...
6	TOCH...	SANT...	21EB...	BACH...	41	DAÑO...	RECU...	100	18.80...	-98.6...
7	HUAQ...	SANT...	21ET...	AQUI...	19	DAÑO...	RECU...	100	18.83...	-98.4...
8	HUAQ...	CACA...	21DP...	BENIT...	425	DAÑO...	RECU...	100	18.78...	-98.4...
9	ACAT...	ACAT...	21EM...	JOSE ...	17	DAÑO...	RECU...	100	18.61...	-98.1...
10	ACAT...	ACAT...	21EJN...	PREE...	11	DAÑO...	RECU...	100	18.21...	-98.0...
11	ACAT...	ACAT...	21DJ...	JARDI...	102	DAÑO...	RECU...	100	18.22...	-98.0...
12	ACAT...	SAN...	21DE...	JOSE...	64	DAÑO...	RECU...	100	18.21...	-98.1...