

Análisis Dataset

May 27, 2021

1 DATASET

```
[1]: import pandas as pd
import pylab as pl
import numpy as np

# Leer dataset
dataset = pd.read_csv('data.csv')
dataset
```

```
[1]:      col1  col2  col3  col4
0      3.5   1.4   0.2   0.2
1      4.9   3.0   1.4   0.2
2      4.7   3.2   1.3   0.2
3      4.6   3.1   1.5   0.2
4      5.0   3.6   1.4   0.4
..      ...   ...   ...   ...
145     6.7   3.0   5.2   1.9
146     6.3   2.5   5.0   2.0
147     6.5   3.0   5.2   2.3
148     6.2   3.4   5.4   1.8
149     5.9   3.0   5.1   0.0
```

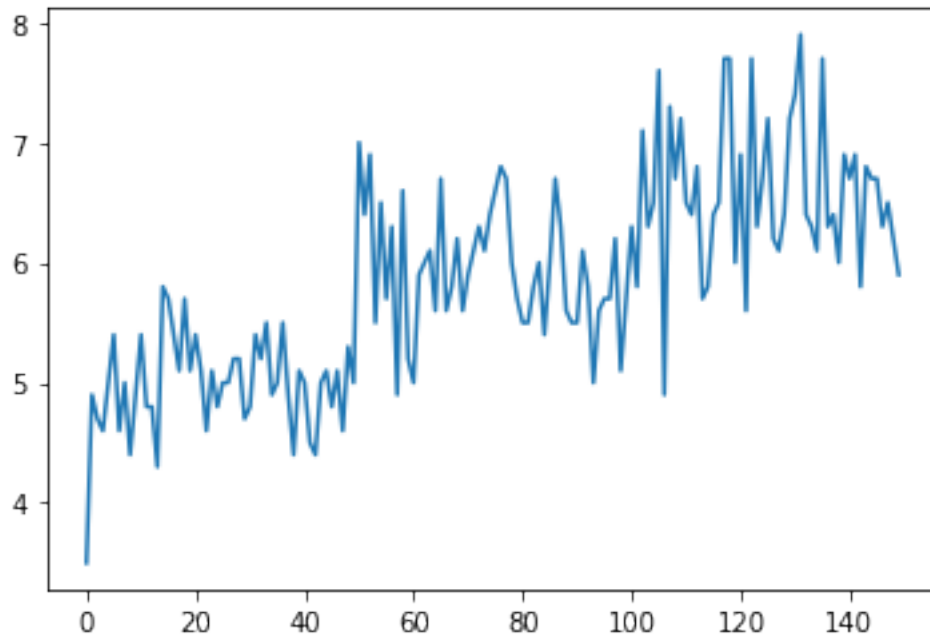
[150 rows x 4 columns]

2 Grafica de la primera columna

Podemos observar que esta primera columna, tiene un incremento muy acelerado al principio y en varios puntos donde los datos no están estables sin no que suben y bajan bruscamente con frecuencia; Pero tiende a subir.

```
[2]: # Columna 1
col1 = dataset.col1.to_numpy()
pl.plot(range(len(col1)), col1)
```

```
[2]: [ <matplotlib.lines.Line2D at 0x215f4e06970> ]
```

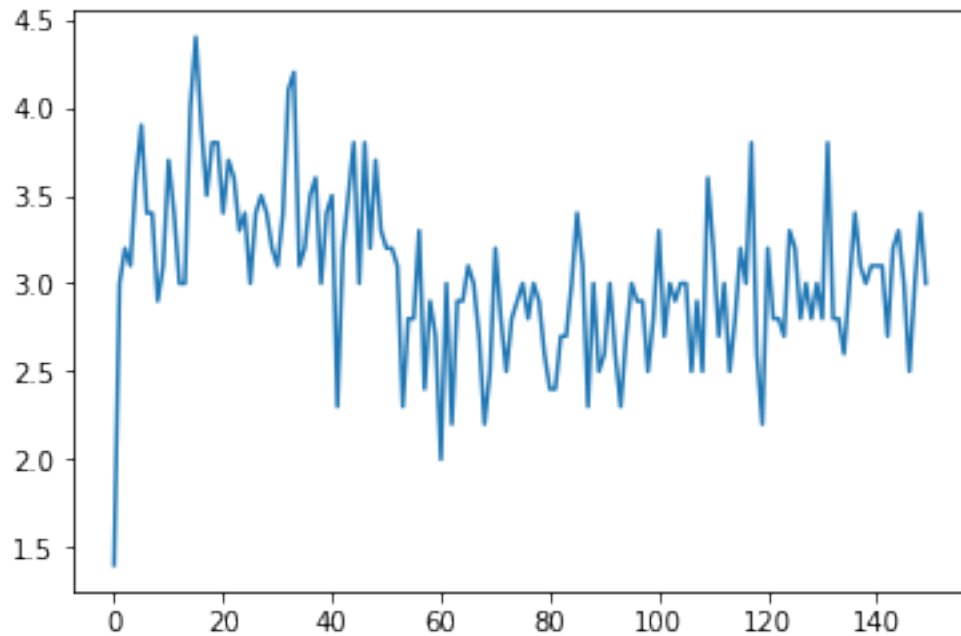


3 Grafica de la segunda columna

Podemos observar que esta segunda columna los datos están igualmente bajando y subiendo bruscamente en el tiempo, pero esta en particular tiene una tendencia a bajar y luego estabilizarse.

```
[3]: # Columna 2
col2 = dataset.col2.to_numpy()
pl.plot(range(len(col2)), col2)
```

```
[3]: [<matplotlib.lines.Line2D at 0x215f4f0a040>]
```

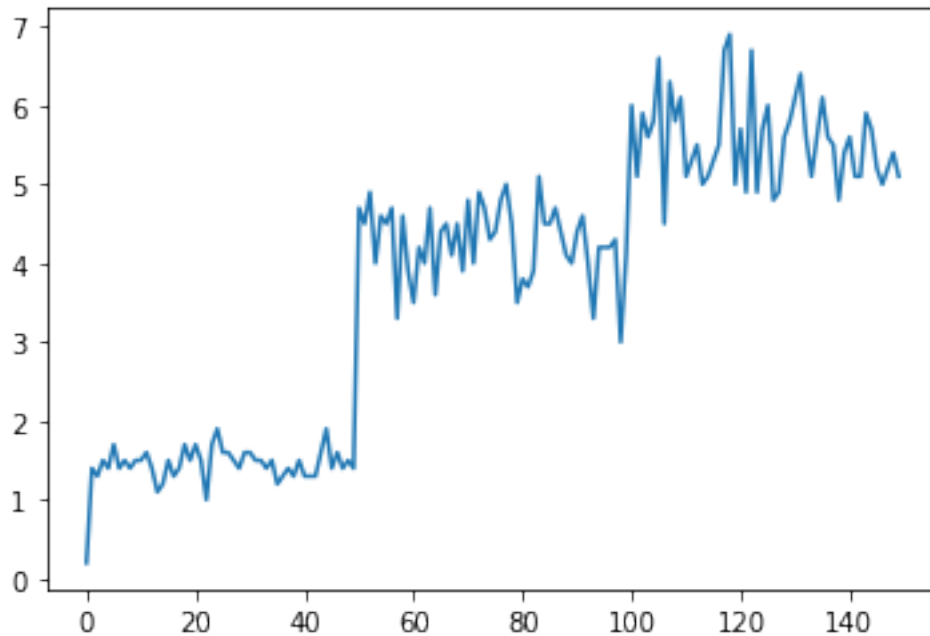


4 Grafica de la tercera columna

Podemos observar que esta tercera columna los datos se mantiene algo estables en 3 segmentos del tiempo, donde existen periodos en lo que tienen un incremento muy acelerado.

```
[4]: # Columna 3  
col3 = dataset.col3.to_numpy()  
pl.plot(range(len(col3)), col3)
```

```
[4]: [<matplotlib.lines.Line2D at 0x215f4f74d00>]
```

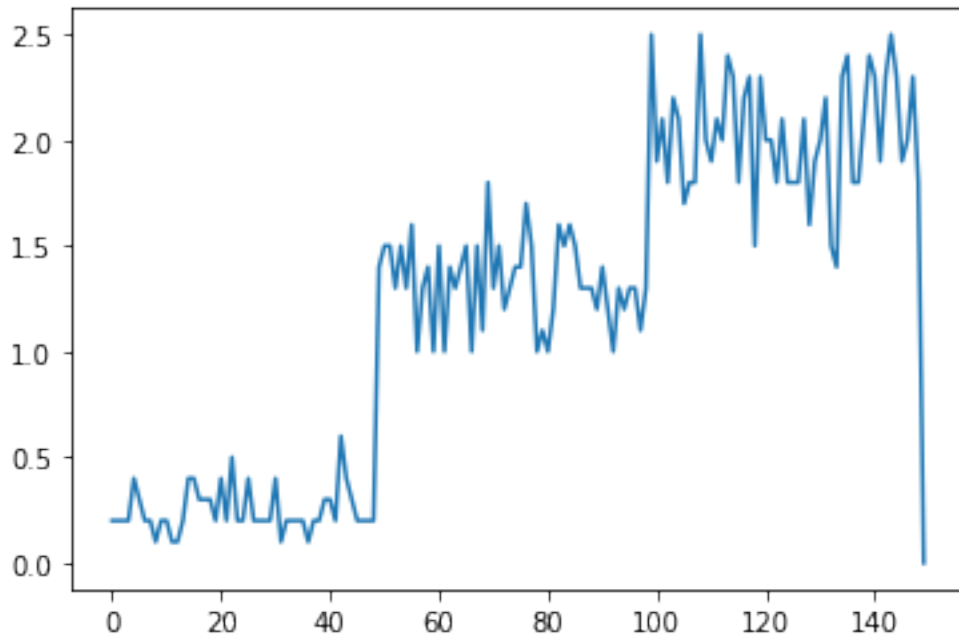


5 Grafica de la cuarta columna

Podemos observar que en la cuarta columna los datos igualmente están algo estables durante 3 periodos de tiempo, y tienen grandes subidas en 2 ocasiones.

```
[5]: # Columna 4  
col4 = dataset.col4.to_numpy()  
pl.plot(range(len(col4)), col4)
```

```
[5]: [<matplotlib.lines.Line2D at 0x215f4fe0b80>]
```

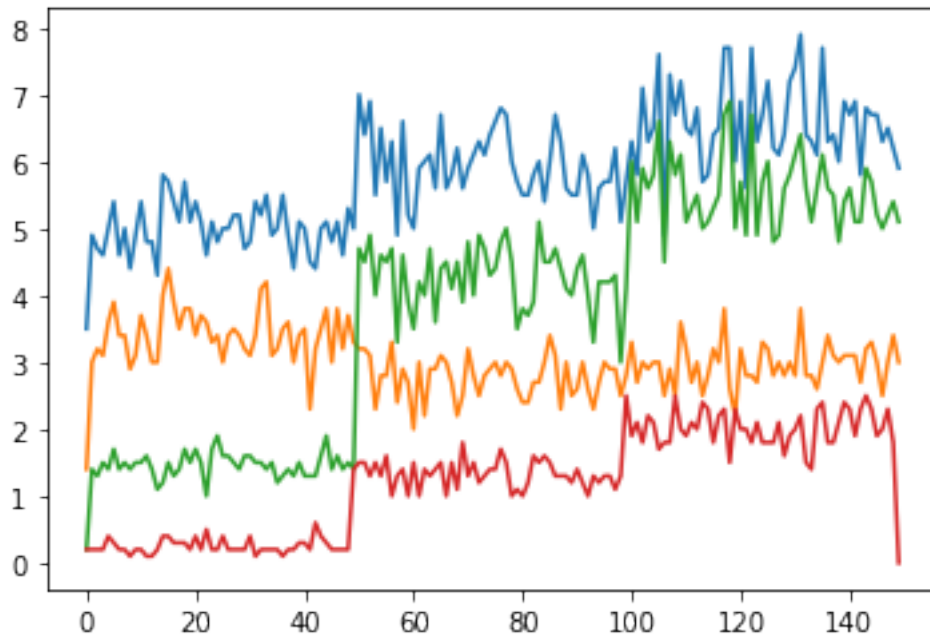


6 Las cuatro graficas

Aquí podemos ver que existe una relación entre la columna 3 y 4 que se parecen mucho ya que ambas tienen un comportamiento similar, pero a escalas diferentes; Y que la columna 1 y 2 a diferencia tienden a ser espejos ya que igualmente tienen subidas y bajadas bruscas pero una tiende a subir mientras la otra a bajar.

```
[6]: # Todas las columnas juntas
pl.plot(range(len(col1)), col1)
pl.plot(range(len(col2)), col2)
pl.plot(range(len(col3)), col3)
pl.plot(range(len(col4)), col4)
```

```
[6]: [<matplotlib.lines.Line2D at 0x215f5051ca0>]
```



Finalmente podemos observar en la última grafica que los datos de la columna 3 tienden a querer alcanzar a los de la columna 1 al final del tiempo, además que los datos de la columna 4 también tienden a alcanzar a los de la columna 3 pero a diferencia de las otras esta es estable y la roja no debe de hacer saltos tan bruscos. Finalmente podría decir que las parejas de la columna 1 y 3 con la de 2 y 4, son semejantes en su comportamiento de que una alcanza a la otra, pero en diferentes escalas; sin obviar que una tiende a subir y otra a bajar.

```
[7]: pl.scatter(range(len(col1)), col1)
      pl.scatter(range(len(col2)), col2)
      pl.scatter(range(len(col3)), col3)
      pl.scatter(range(len(col4)), col4)
```

```
[7]: <matplotlib.collections.PathCollection at 0x215f50c9f70>
```

