

Visualización de datos con la biblioteca Seaborn

Rolando Salazar

¹ Universidad de Sonora. Hermosillo, Sonora.

3 de mayo de 2019

Resumen de la actividad 7 donde se utiliza la biblioteca Seaborn para realizar gráficos y se compara con Matplotlib.

1. Introducción

La actividad 7 consistió en visualizar datos meteorológicos de un nogal del 2009 mediante las bibliotecas Matplotlib y Seaborn creando un mapa de calor (Heat Map) para comparar ambos resultados y decidir cuál es la mejor para dicha tarea. Luego, se realizan gráficas de dispersión de puntos con Seaborn para las variables (variable1 vs variable 2 cuya correlación sea mayor a 0.6).

2. Desarrollo de la actividad 7

Primero, para poder trabajar con sin problemas, se extraen los datos útiles. Es por ello que se eliminaron las columnas sin nombre y con datos vacíos. Después, se encuentra la correlación entre las variables que contiene el archivo mediante la función `corr()` ejecutando el comando

```
df.corr()
```

y obteniendo la siguiente tabla a partir de ella.

	N	u_Avg	v_Avg	w_Avg	t_Avg	kh20_Avg	net_rad_Avg	shf1_Avg	shf2_Avg
N	1.000000	0.020629	0.031793	0.540191	0.023786	-0.000075	-0.132426	-0.006994	-0.003209
u_Avg	0.020629	1.000000	0.241494	0.015906	0.044331	-0.049445	-0.094394	0.089411	0.003542
v_Avg	0.031793	0.241494	1.000000	0.009572	0.011734	0.155188	-0.117176	0.029321	-0.053846
w_Avg	0.540191	0.015906	0.009572	1.000000	0.601335	0.128426	0.133556	-0.079466	-0.010182
t_Avg	0.023786	0.044331	0.011734	0.601335	1.000000	-0.009929	0.266318	-0.094414	-0.006335
kh20_Avg	-0.000075	-0.049445	0.155188	0.128426	-0.009929	1.000000	-0.157759	-0.005268	-0.061590
net_rad_Avg	-0.132426	-0.094394	-0.117176	0.133556	0.266318	-0.157759	1.000000	-0.206065	-0.027206
shf1_Avg	-0.006994	0.089411	0.029321	-0.079466	-0.094414	-0.005268	-0.206065	1.000000	0.059003
shf2_Avg	-0.003209	0.003542	-0.053846	-0.010182	-0.006335	-0.061590	-0.027206	0.059003	1.000000
vv_Avg	0.463299	-0.163176	-0.088159	0.912885	0.529160	0.152027	0.162470	-0.089608	-0.017229
airT_Avg	-0.268578	-0.363548	-0.344386	-0.338470	-0.231403	-0.122687	0.334375	-0.121637	0.078023
rh_Avg	0.357670	0.331126	0.098515	0.737544	0.367252	0.024498	-0.266374	0.084782	0.017651
e_sat_Avg	-0.162738	-0.370644	-0.347170	-0.192107	-0.139445	-0.123180	0.385945	-0.138901	0.081718
e_Avg	0.387780	0.103343	-0.109168	0.646316	0.306703	-0.126972	0.074294	-0.016448	0.051809
h2o_hmp_Avg	0.768532	0.173845	-0.189840	0.125133	0.576896	-0.326760	0.056376	0.014207	0.151206

Figura 1: Muestra de la tabla de correlaciones.

2.1. Heat Map

El Heat Map es una variante visual de la tabla mostrada anteriormente. Los mapas de calor obtenidos con ambas librerías son los siguientes.

```
plt.imshow(df1)
```

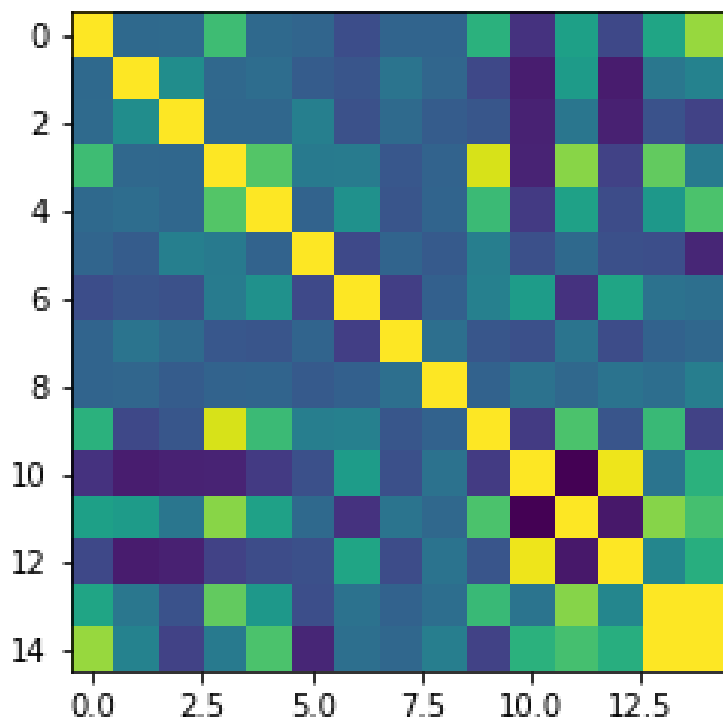


Figura 2: Heat Map con Matplotlib.

```
sns.heatmap(df1)
```

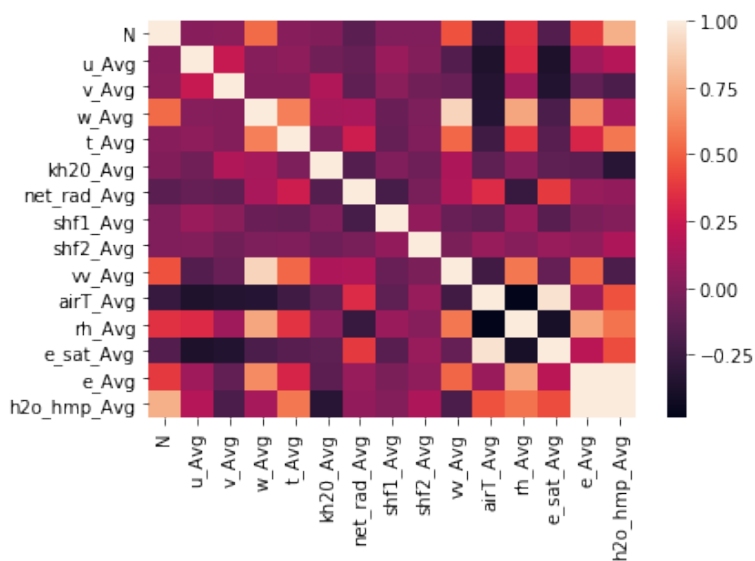


Figura 3: Heat Map con Seaborn.

A mí parecer es mucho mejor el Heat Map obtenido con Seaborn, ya que contiene los nombres de las variables y el nivel de correlación sin la necesidad de incluir atributos extras en el comando principal, o de agregar más comandos. Creo que es una ventaja muy considerable a favor de Seaborn, en este caso.

2.2. Dispersión de datos

Por último, se realizaron gráficas de dispersión de datos para las variables que tenían una correlación mayor de 0.6. Se obtuvieron bastantes gráficas, por lo que aquí mostraremos solo dos para no llenar el reporte de imágenes. Los comandos utilizados para realizar las gráficas son los siguientes:

```
sns.scatterplot(x = df["variable1"], y = df["variable2"])
plt.title("variable1 vs variable2 (corr = number)").
```

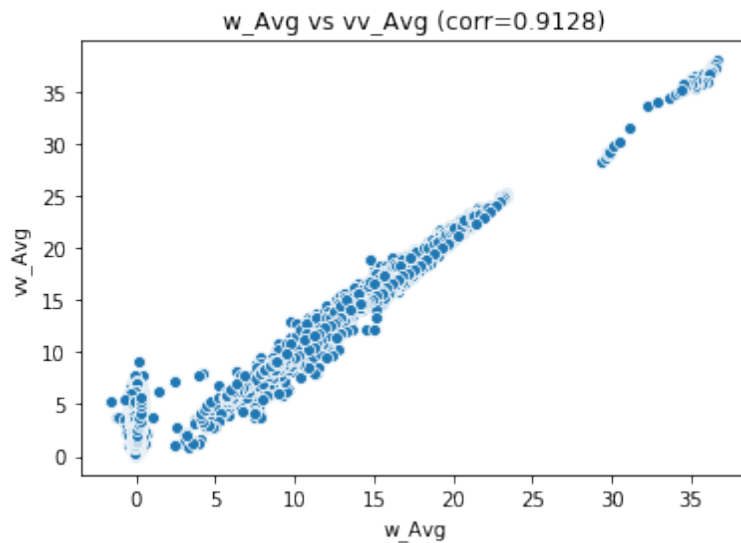


Figura 4: Variable w_Avg vs vv_Avg .

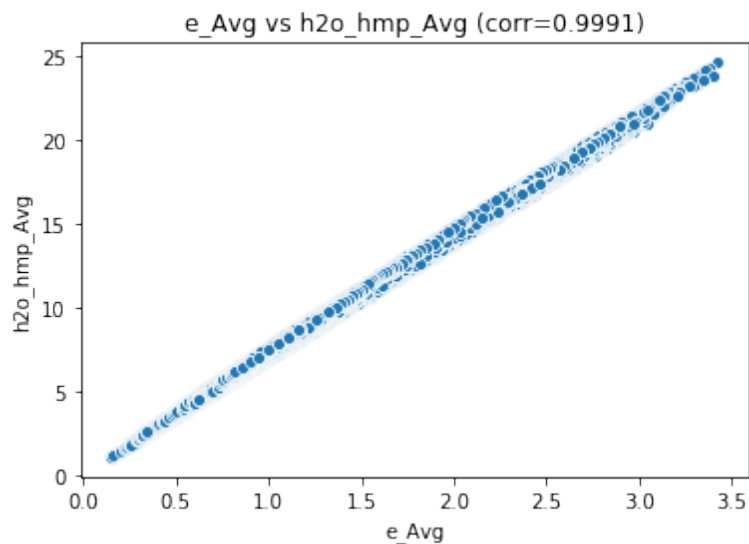


Figura 5: Variable e_Avg vs $h2o_hmlp_Avg$.

3. Conclusiones

Seaborn facilita la visualización de datos en Python y, en específico, la visualización de la correlación entre las variables de un conjunto de datos. La actividad fue solo una introducción a Seaborn. Seguramente tiene más funciones que son de gran ayuda para el manejo de datos visualmente. Además, las gráficas de dispersión de datos también son muy útiles con Seabor, ya que se nota una mejora visual en las gráficas.