

# Visualización de datos con la biblioteca Seaborn.

Sealtiel Gallardo Martínez

27 de marzo de 2019

En esta actividad nos concentramos en una biblioteca de Python llamada Seaborn, la cual utilizaremos para visualizar una posible correlación lineal entre pares de variables de un conjunto de datos meteorológicos de obtenidos de un campo de nogal entre los años 2009 y 2010. La tabla de datos con la cual vamos a trabajar contiene doce distintas variables las cuales vamos a comparar entre estas para ver si existe la correlación.

El concepto de correlación consiste en comparar los datos de diferentes columnas y ver alguna relación que sea aproximada a una lineal, para ver de que manera se relacionan esta. El criterio que vamos a tomar es que sea, cómo mínimo, un 60 por ciento de coincidencia de relación.

Empezando con la actividad en Jupyter Lab, tenemos que leer los datos, los cuales haremos con esta sección de código:

---

```
df = pd.DataFrame( pd.read_csv("meteo-nogal-09.csv",  
                             engine="python" ) )  
df.head(10)
```

---

Al pedirle las primeras 10 columnas de la fila, nos podemos da cuenta que hay muchas columnas y filas que no tienen nombre ni datos con los cuales no podemos trabajar, por lo que el siguiente paso es limpiar esta tabla de datos, y lo haremos seleccionando varias columnas a la vez y usar la función drop para deshacernos de ella, de la siguiente manera:

---

```
df.drop( df.columns[18:36], axis=1, inplace=True )  
df.drop( df.columns[2:4], axis=1, inplace=True )
```

---

Teniendo lista la tabla de datos con los cuales vamos a trabajar, tenemos una función de la librería Pandas, la cual nos puede dar el valor numérico de la correlación lineal entre dos conjuntos de datos contenidos en las columnas. Esta función la utilizaremos de la siguiente manera en nuestra actividad.

---

```
df_corr = df.corr(method='pearson', min_periods=1)
```

---

Después de obtener el valor numérico de cada una de las correlaciones, Seaborn nos ayuda con el resto, ayudándonos a visualizar esta correlación en un mapa muy intuitivo que nos facilita ver que relaciones nos interesan. Este mapa lo construiremos con la siguiente línea de código:

---

```
sns_plot =
    sns.heatmap(df_corr,cmap="viridis",robust=True,square=True,
                annot=False)
fig = sns_plot.get_figure()
```

---

Brindándonos el siguiente mapa:

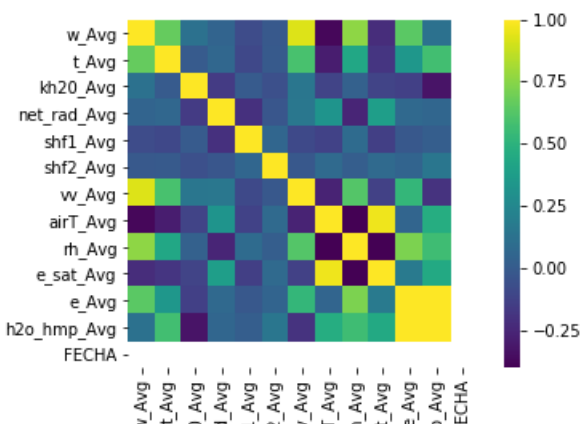


Figura 1: Mapa de correlaciones de la biblioteca Seaborn.

Gracias a ese mapa podemos visualizar que variables cumplen con la característica de tener un 60 por ciento de coincidencia al momento de realizar la correlación, pero con ayuda de una sección de repetición de código for podemos obtener directamente estos pares de variables.

El resultado de estas relaciones fue de ocho gráficas que cumplen con el criterio de coincidencia.

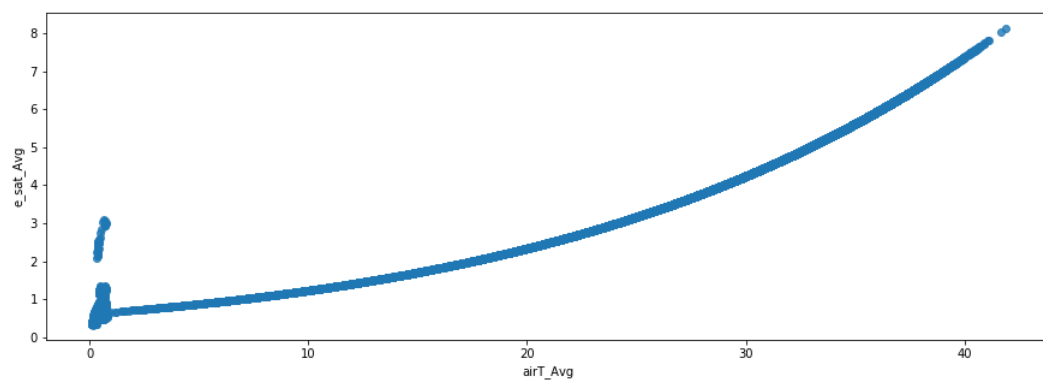


Figura 2: .

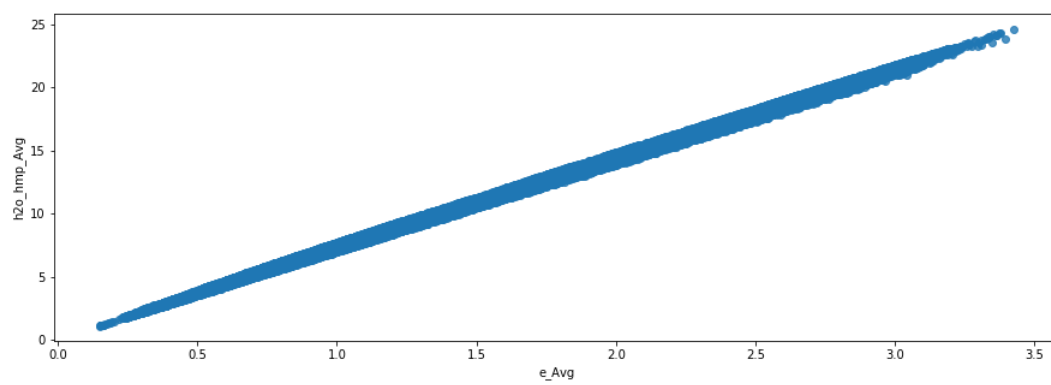


Figura 3: .

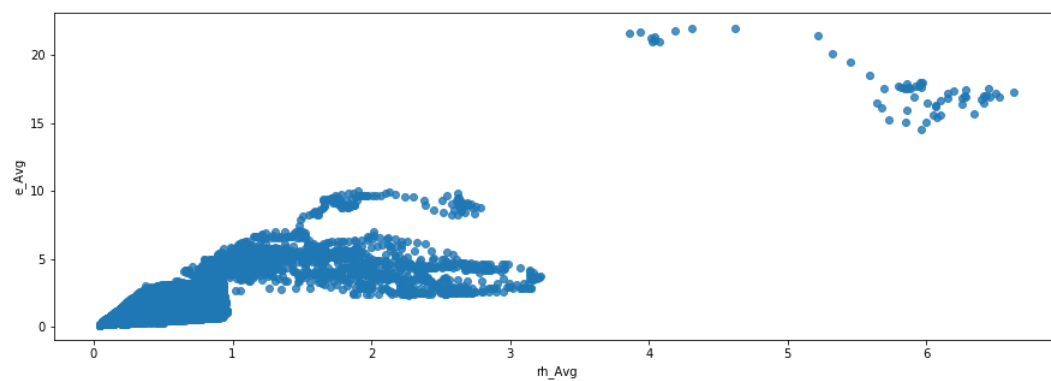


Figura 4: .

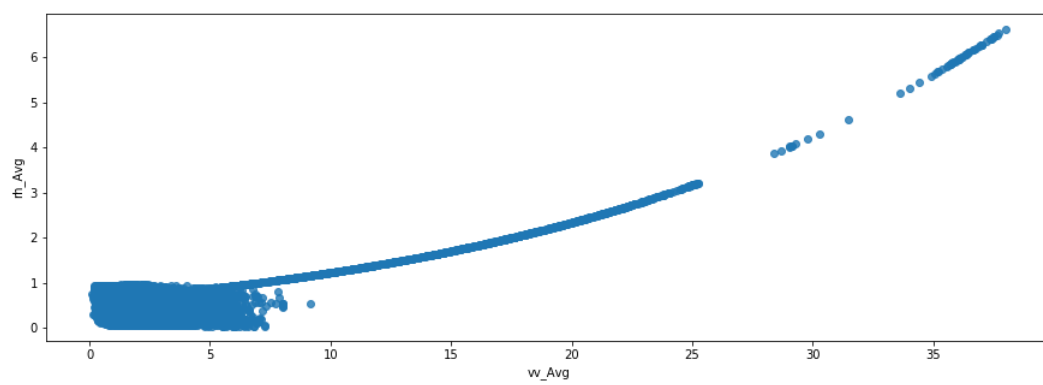


Figura 5: .

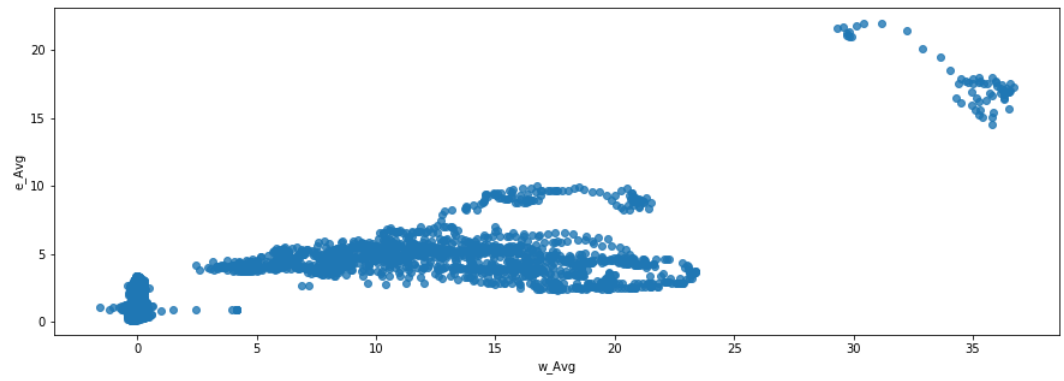


Figura 6: .

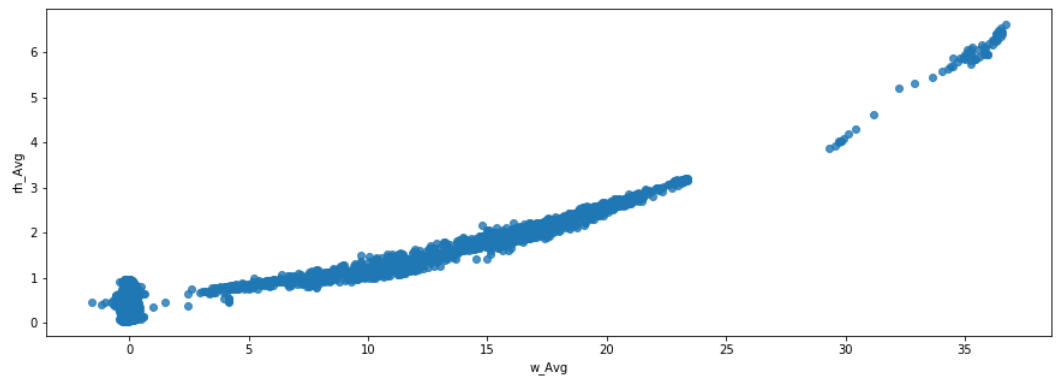


Figura 7: .

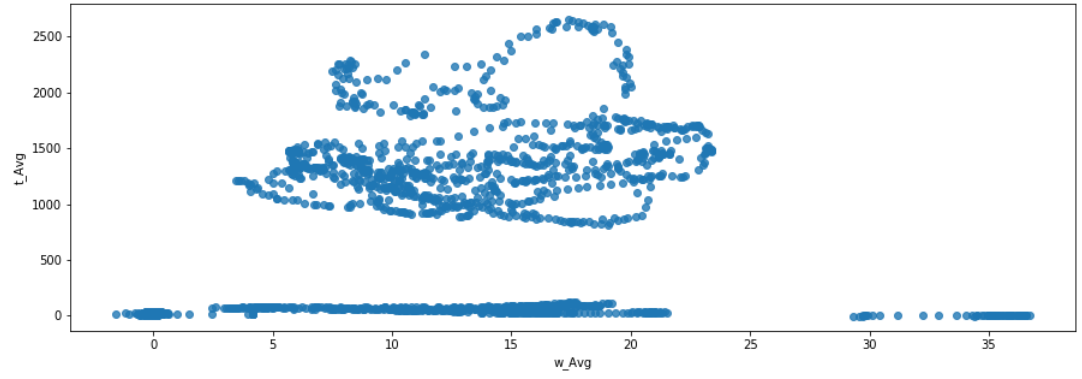


Figura 8: .

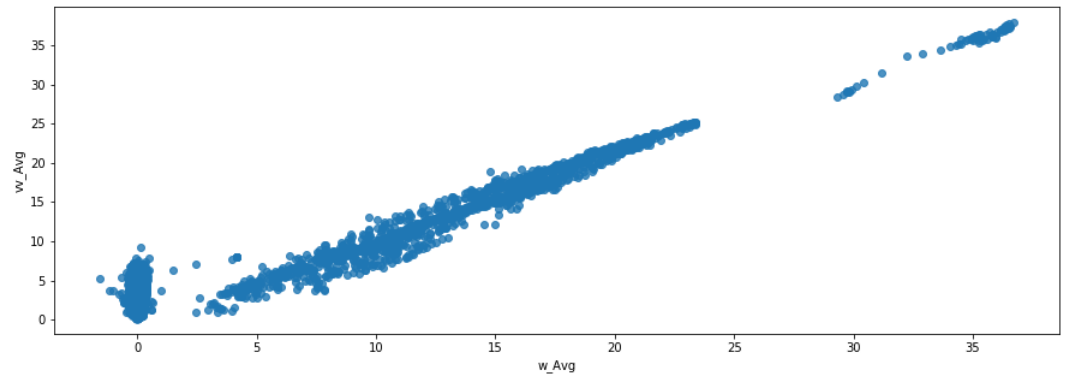


Figura 9: .