

# Reporte de la actividad 7

Daniela Olmos Velderrain

Grupo 3

24 de marzo de 2019

## 1. Introducción

En esta actividad analizamos los datos en una estación meteorológica de un campo de nogal para observar las correlaciones entre distintos parámetros medidos. Se comparó el uso de las bibliotecas SeaBorn y Matplotlib para la elaboración de gráficos, y también se graficaron las variables que mostraban un alto índice de correlación.

## 2. Desarrollo

### 2.1. Marco teórico

#### 2.1.1. Correlación

Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables (A y B) existe correlación entre ellas si al disminuir los valores de A lo hacen también los de B y viceversa.

#### 2.1.2. Heat Map

Un heatmap o mapa de calor es una representación gráfica que utiliza tonos de color de diferente intensidad para indicar las zonas con mayor incidencia. En esta actividad el uso de mapas de calor será útil para ubicar las variables con mayor correlación del data frame.

### 2.2. Metodología

Para comenzar a trabajar con el archivo, primero fue necesario descargar librerías para el análisis de datos y visualización:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import math
```

Como el archivo a leer contenía caracteres especiales que no podía reconocer la librería de Pandas, utilizamos el lector de python:

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

El documento contaba con varias columnas nombradas "unnamed" que no contenían datos relevantes, así que las eliminamos para quedarnos con un data frame más pequeño:

```
df=df.drop(df.columns[df.columns.str.contains('unnamed:',case = False)],axis = 1)
```

Como el archivo no contaba con una variable de datetime, la creamos a partir de dos columnas que contenían fechas y horas:

```
df["FECHA"] = df["DATE"] + " " + df["TIME"]
```

Convertimos las variables object a float64 para poder operar con los datos de forma numérica:

```
df[df.columns[0:14]] = df[df.columns[0:14]].apply(pd.to_numeric, errors='coerce')
```

Empleamos la función *corr* para encontrar las correlaciones entre las variables:

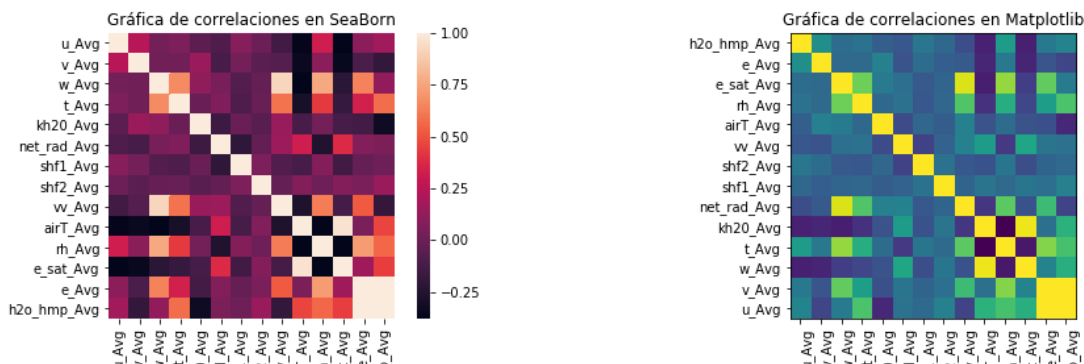
```
df_corr = df.corr(method='pearson', min_periods=1)
```

Con el data frame de las correlaciones creamos dos gráficas del tipo heatmap, una mediante la biblioteca Seaborn y otra mediante Matplotlib.

Por último, observamos las correlaciones cuyo valor absoluto era mayor a 0.5 y las graficamos.

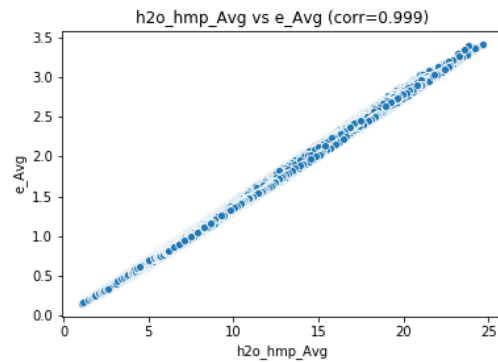
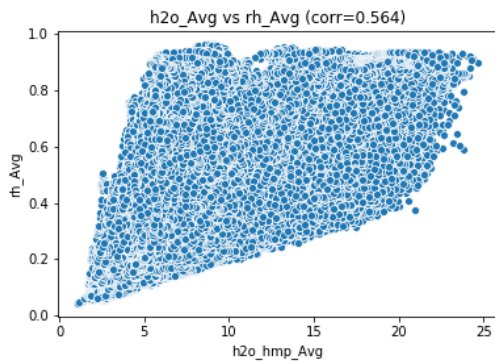
## 2.3. Resultados

A continuación se muestran las gráficas de heatmap elaboradas con Seaborn y Matplotlib:



Aunque las dos son similares en apariencia, el diagrama de Seaborn cuenta con una barra que indica la correlación correspondiente a cada color, lo cual no se pudo lograr en Matplotlib.

Las correlaciones mayores a 0.5 fueron 12 en total. A continuación se mostrarán dos casos, uno en que la correlación es muy cercana a 0.5 y otro en que es muy cercana a 1.



### 3. Conclusiones

Al comparar el uso de ambas bibliotecas al momento de graficar, notamos inmediatamente la gran facilidad que tiene SeaBorn en comparación a Matplotlib, ya que se ocuparon muchas más líneas de código para elaborar el heatmap en Matplotlib, y este no contenía la barra lateral que proporciona SeaBorn.

En cuanto a las gráficas de correlaciones, se puede observar a simple vista que a medida que la correlación se aproxima más a 1 la relación entre las variables es más coherente, mientras que al alejarse de este valor las variables parecen no tener relación entre sí.

### Bibliografía

- Correlación. Recuperado el 24 de marzo de 2019 desde <https://es.wikipedia.org/wiki/Correlaci%C3%B3n>
- Mapa de Calor. Recuperado el 24 de marzo de 2019 desde <https://www.oleoshop.com/blog/mapa-de-calor-que-es>