

Reporte de Evaluación 1

Valenzuela Terán Jonás

March 9, 2018



1 Introducción

El propósito de la evaluación es verificar que conocemos y sabemos manejar las herramientas que hemos aprendido a utilizar en clase. En este caso, se utilizan para hacer un análisis de datos acerca de el nivel del mar, salinidad y temperatura del agua para encontrar la relación que guardan. Esto es posible haciendo uso de representaciones gráficas que ayudan a interpretar los datos,



Manglar El Sargento -Carlos Lizarraga Celaya

2 Recuperación y limpieza de datos

Los datos fueron tomados en el manglar El Sargento y proporcionados por el profesor, se utilizaron comandos de *emacs* para adaptar el formato, en si, los datos necesitaron solo de cambios menores, de la separación entre variables, con comas y espacios.

3 Análisis de datos

Una vez que los datos se encuentran en el formato deseable, cada archivo de datos es asignado a un dataframe en *python*, con entorno de programación *jupyter notebook*, donde se importan bibliotecas útiles como *numpy* para herramientas numéricas, *pandas* para estadística, *datetime* para formato de tiempo, *matplotlib* y *seaborn* para gráficas.

Se asignó un dataframe que contiene la información de interés para esta práctica, que es la fecha, nivel del mar, salinidad y temperatura, después se representó con boxplots de cada variable, regresiones lineales de Pearson y gráficas de variable contra tiempo individuales y superpuestas.

Para la lectura de datos, se ignoraron las mediciones que no coincidieran en el tiempo de un archivo a otro, y los encabezados, para asignarlos manualmente: (» indica una instrucción o línea)

```
>> df1 = pd.read_csv('sargento_201117.csv', names=['Date-Time', 'AbsPres', 'Temp', 'WaterLevel'],
skipfooter=1, sep=',', engine='python')
>> df2 = pd.read_csv('sargento-salinidad-201117.csv', names=['Date-Time', 'Cond-High-Rng',
'Temp', 'SpecificC', 'Salinity'], skiprows=2, sep=',')
>> df2 = df2.drop(df2.index[:1])
```

Después, se creó un nuevo dataframe con la información de interés:

```
>> df3 = pd.concat([df1['Date-Time'], df1['WaterLevel'], df2['Salinity'], df1['Temp']], axis=1, k
```

Se creó la variable temporal *Ndate* y *month* para poder trabajar por mes en los boxplots:

```
>> df3['Ndate'] = pd.to_datetime(df3['Date-Time'], format='%m/%d/%Y/%H:%M:%S')
>> df3['month'] = df3['Ndate'].dt.month
```

Continuamente se utilizaron las funciones *head()* y *tail()* para comprobar el formato de los datos. Una vez comprobados, se crearon los boxplot de la siguiente manera:

```
>> import seaborn as sns

>> ax = sns.boxplot(x="month", y="WaterLevel", data=df3)
>> plt.show()
```

La regresión lineal de Pearson, en general, fue creada con el código:

```
>> sns.set(style="darkgrid", color_codes=True)

>> g = sns.jointplot("WaterLevel", "Salinity", data=df3, kind="reg",
                    color="r", size=7)
>> plt.show(g)
```

Se crearon gráficas de la variable con respecto al tiempo, creando nuevos dataframe, con:

```
>> df4 = df3[['Date-Time', 'WaterLevel']]
>> plt.figure(); df4.plot(x='Date-Time'); plt.legend(loc='best')
>> plt.title('Variación del Nivel del mar')
>> plt.ylabel('Metros')
>> plt.xlabel('Tiempo')
>> plt.grid(True)
>> plt.show()
```

Después se realizó una superposición de las variables contra el tiempo:

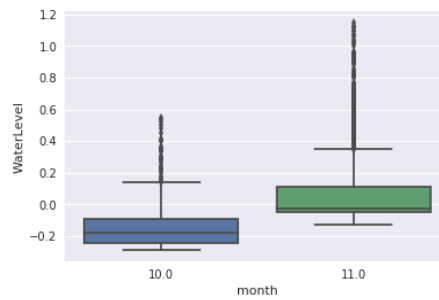
```
>> df7 = df3[['Date-Time', 'WaterLevel', 'Salinity']]
>> plt.figure(); df7.plot(x='Date-Time'); plt.legend(loc='best')
>> plt.title('Variación del Nivel del mar y Salinidad')
>> plt.xlabel('Tiempo')
>> plt.grid(True)
>> plt.show()
```

Y finalmente, se utilizó la función *xlim* para delimitar un rango de tiempo más corto:

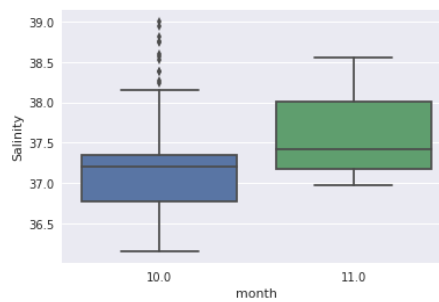
```
>> df9 = df3[['Date-Time', 'WaterLevel', 'Salinity']]
>> df9n= df9.iloc[1:525]
>> plt.figure(); df9n.plot(x='Date-Time'); plt.legend(loc='best')
>> plt.title('Variación del Nivel del mar y Salinidad a 5 días')
>> plt.xlabel('Tiempo')
>> plt.grid(True)
>> plt.show()
```

4 Resultados

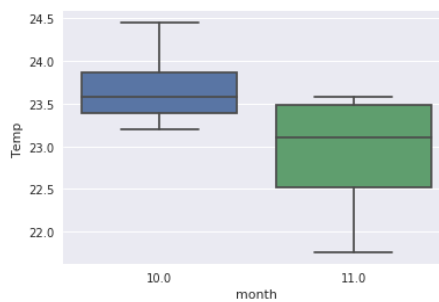
A continuación se muestran las representaciones gráficas obtenidas:



Boxplot de Nivel del Mar por mes

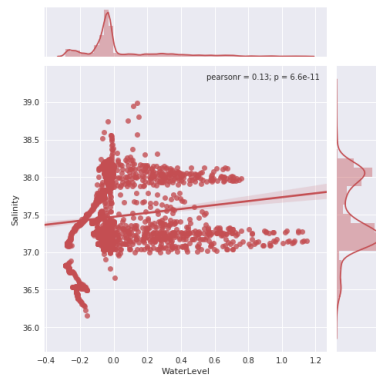


Boxplot de Salinidad por mes

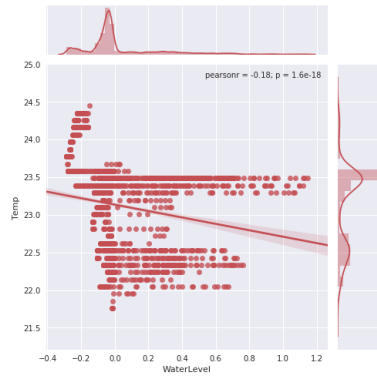


Boxplot de Temperatura por mes

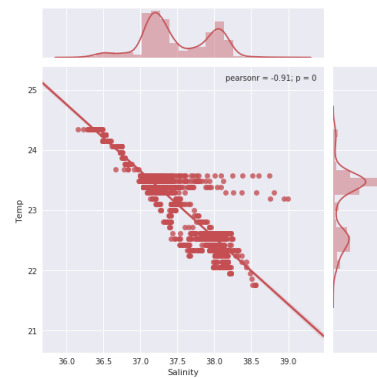
Se observa a través de los boxplots, que al comparar noviembre con octubre en general, el nivel del agua y salinidad se eleva, mientras que las temperaturas disminuyen.



Regresión lineal de Salinidad contra Nivel del mar

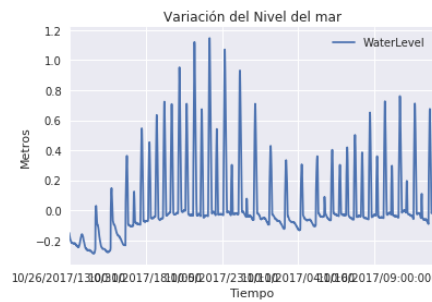


Regresión lineal de la Temperatura contra Nivel del mar

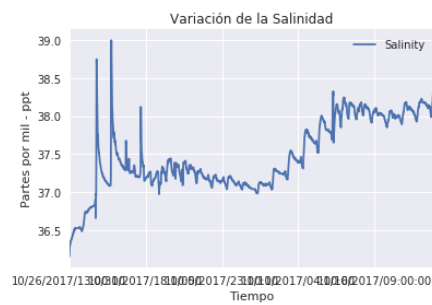


Regresión lineal de Temperatura contra Salinidad

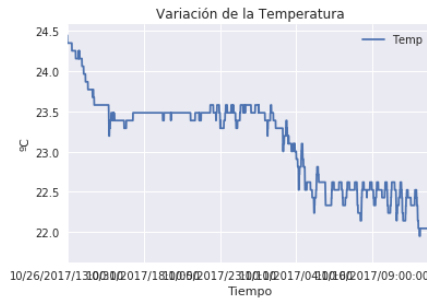
Se puede notar a través de esta representación que el único par de variables que se cuentan con una relación aproximadamente real es la Salinidad y Temperatura.



Gráfica de Nivel del mar contra tiempo

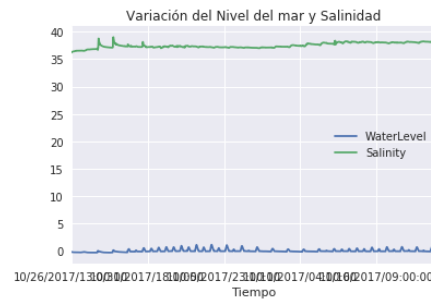


Gráfica de Salinidad contra tiempo

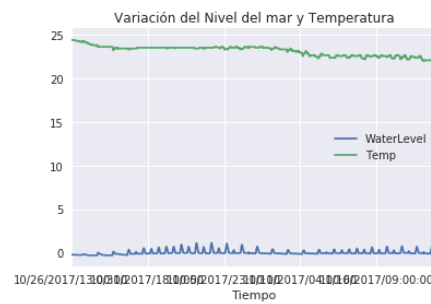


Gráfica de Temperatura contra tiempo

Se muestra mucha variación de cada variable con respecto al tiempo, para poder encontrar una relación, resulta más conveniente hacer una superposición.

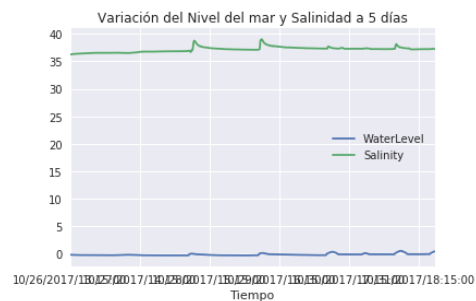


Gráfica del Nivel del mar y Salinidad contra tiempo

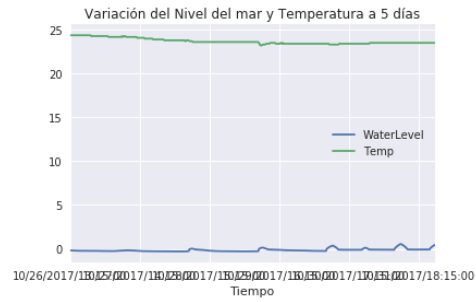


Gráfica de Nivel del mar y Temperatura contra tiempo

Aun no es claro si existe relación o no a través de este tipo de gráfica, debemos analizar un intervalo más pequeño de tiempo.



Gráfica del Nivel del mar y Salinidad contra tiempo en 5 días



Gráfica de Nivel del mar y Temperatura contra tiempo en 5 días

Ahora es notable la relación entre el Nivel del mar y la Salinidad, cuando una variable aumenta, la otra también, pero con magnitud diferente.

5 Final

Ya terminada la actividad, se guardaron todos los archivos en un solo directorio (Evaluación1), y se guardaron en la plataforma de github, siendo útil para futura consulta, y revisar los métodos aprendidos para lectura, acomodo y graficación de datos.