

Actividad 3: Pandas

Borbón Fragoso Julio Cèsar

11 de febrero del 2019

Comandos en Pandas

Pandas es una librería para facilitar el manejo de datos en el lenguaje Python, tiene varias funciones a continuación se describen algunas de estas:

pd.read_csv : Con esta función podemos leer un archivo en formato CSV (comma-separated values)
pd.head(): Está función sirve para leer los primeros archivos de una tabla, en el parentesis se puede ingresar el número de archivos que deseas leer.
df.tail(): Parecido al head pero muestra los últimos archivos
df.dtypes: Nos muestra como se han leído de manera automática nuestras variables.
pd.to_datetime: Cambia una columna escogida a la variable de DATETIME para poder tratarla como tiempo
pd.options: Podemos configurar la configuración de pandas, como por ejemplo cambiar el número predefinido de columnas que se muestran, etc.
pd.to_numeric: Cambia una columna para que puede ser detectada como variable numérica.
df.mean(): Nos muestra la media de las columnas
df.std/median/max/min : Nos muestra la varianza, la mediana, el máximo y el mínimo respectivamente.
df.describe(): Muestra un resumen estadístico con información variada
df.count: Nos muestra la cantidad de datos en una columna
df.drop: Deshecha una columna de los datos.
pd.dataframe: Da estructura de datos a nuestro archivo
df.resample: Método de conveniencia para transformación de frecuencia, solo se puede usar en variables relacionadas con la fecha.

Preguntas

Se incluyen las siguientes preguntas con un fragmento de código de como se resolvieron.

¿Cómo le podrás determinar cuáles son los meses más lluviosos?

```
#La precipitación esta sumamente relacionada con la lluvia así que asumiremos
#que toda la cantidad de precipitación es debido únicamente a las lluvias
#Declarando una nueva variable df10, usamos el comando set index y resample para sacar
#la media de cada mes
df10 = df9.set_index('FECHA').resample('M')['PRECIP'].mean()
#Se muestra df10 que es una tabla de datos y con esto podemos ver cuales meses han tenido mayor
#precipitación
df10
```

—
—
¿Cuáles son los meses más fríos y cuáles son los más cálidos?

```
#se declara una tabla df11 que tenga la media de las temperaturas
df11 = df9.set_index('FECHA').resample('M')['TMAX', "TMIN"].mean()
#Se lee la tabla y se checan los valores
df11
```

¿Cuáles han sido años muy húmedos/secos?

```
#De la misma manera que la primer pregunta, pero cambiamos
#la frecuencia a anual.
df12 = df9.set_index('FECHA').resample('Y')['PRECIP'].mean()
#Se imprime df12 para revisar sus datos
df12
```

¿Cuáles años han tenido inviernos fríos/calidos?

```
#DE la misma manera que la pregunta anterior
df13 = df9.set_index('FECHA').resample('Y')['TMAX', "TMIN"].mean()
df13
```

¿Cómo ha venido siendo la temperatura mensual promedio en los últimos 20 años?

```
#Se crea una nueva columna de la suma de los promedios
#de Tmin y Tmax y se divide entre 2
df12['prom'] = (df12['TMAX'] + df12['TMIN'])/2
#Se muestran únicamente los últimos 20 datos
df12.tail(20)
```

¿Qué ha pasado con la precipitación en los últimos 20 años de datos?

```
#Se modifica df10 para que tenga la media de precipitación anual
df10 = df9.set_index('FECHA').resample('Y')['PRECIP'].mean()
#Se muestran los últimos 20 datos
df10.tail(20)
```

Con esto se da concluida la actividad 3 realizada para conocer mejor la libreria para Python, Pandas.