

Introducción a pandas

En Computación y Ciencia de datos, pandas es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. Es un software libre distribuido bajo la licencia BSD versión tres cláusulas. El de la biblioteca deriva del término "datos de panel", término de econometría que designa datos que combinan una dimensión temporal con otra dimensión transversal.

1. Características

1. Tipo de datos DataFrame para manipulación de datos con indexación integrada.
2. Herramientas para leer y escribir datos entre estructuras de dato en-memoria y formatos de archivo variados.
3. Alineación de dato y manejo integrado de datos faltantes.
4. Reestructuración y segmentación de conjuntos de datos.
5. Segmentación vertical basada en etiquetas, indexación elegante, y segmentación horizontal de grandes conjuntos de datos.
6. Inserción y eliminación de columnas en estructuras de datos.
7. Mezcla y unión de datos.
8. Indexación jerárquica de ejes para trabajar con datos de altas dimensiones en estructuras de datos de menor dimensión.

Desventajas

1. Lenguaje de interpretación. Como python es un lenguaje de interpretación, esto quiere decir que es as lento comparado con otros como Java o C++.

2. Información climatológicas de la ciudad De Agua Prieta

Esta actividad consiste en ver la informacion del clima de una ciudad en particular. La ciudad que escogi es la ciudad de Agua Prieta. Estos datos van desde el año 1961 hasta el año 2009 y van desde el dia a dia.

Esta información climatologica consta de los siguientes parametros.

- Fecha
- Precipitación
- Evaporación
- Temperatura maxima
- Temperatura minima

Para hacer uso de esta actividad tendremos que hacer uso de los siguientes comandos.

- Comando para leer los archivos

```
In [227]: df0=pd.read_csv('aguaprieta.txt',skiprows=0,sep='\s+', na_values=SENTINELS)
```

El comando que difiere a la actividad pasada sera NA-VALUES que sirve para que insertarle una variable que especificara que valores se tienen que omitir.

- Cambiar para que la fecha sea tipo DateTime Este comando convierte el tipo de los

```
In [232]: df['FECHAS'] = pd.to_datetime(df0['FECHA'], dayfirst=True)
```

datos de la fecha de objeto a date. Esto servira para poder manipular los datos mas facilmente.

- Agrupar datos, aplicarles una funcion y graficarlas. Con el comando groupby podemos

```
In [226]: fig, ax = plt.subplots(figsize=(15,7))  
df0.groupby(df0['FECHAS'].dt.month)['PRECIPn'].agg(['mean']).plot(ax=ax)
```

agrupar los datos de la fecha con respecto ahora sea un dia, mes o año. Ademas se le puede aplicar una funcion, ya sea sumar todos los valores, aplicarles una regresión o promediarlos. Por ultimo podemos graficarlos.

Apéndice

Para que la actividad sea completada tendremos que responder las siguientes preguntas.

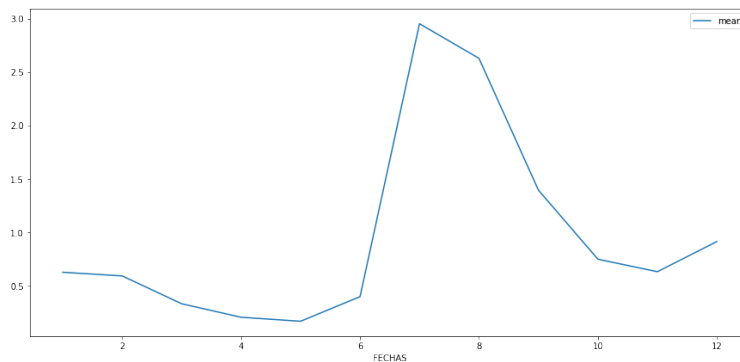
1.- ¿Cómo le podrás determinar cuáles son los meses más lluviosos?

Para determinar los meses mas lluviosos usamos el siguiente comando.

```
df0.groupby(df0['FECHAS'].dt.month)['PRECIP'].agg(['mean'])
```

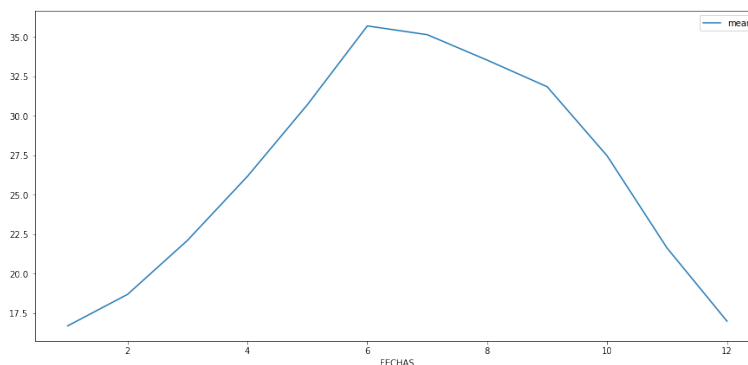
Con el `dt.month` nos dara los meses y su precipitacion promediada.

Podemos ver que los meses con mas precipitación son Junio, Julio y Agosto. Para verlo mejor mostraremos la siguiente grafica.



¿Cuáles son los meses más fríos y cuáles son los más cálidos? Usaremos el siguiente comando `df0.groupby(df0['FECHAS'].dt.month)['TMAX'].agg(['mean']).plot(ax=ax)`

Podemos ver que los meses con mas temperatura son Junio, Julio y Agosto; Los mas frios son Enero, febrero y diciembre, Para verlo mejor mostraremos la siguiente grafica.



¿Cuáles han sido años muy húmedos? ¿ y Cuáles han sido años muy secos?

Para responder esta pregunta usaremos el siguiente comando.

```
df0.groupby(df0['FECHAS'].dt.year)['EVAP'].agg(['mean']).plot(ax=ax)
```

Donde el `dt.year` es para poner los años y `evap` para saber los niveles de evaporación. La siguiente grafica muestra que hay algunos espacios huecos ya que no hay muestras en algunos años, pero los años mas secos son en la decada de los 70 y los mas humedos son el decada de los 80.

Las demas preguntas se contestan analogamente usando los mismos codigos.

