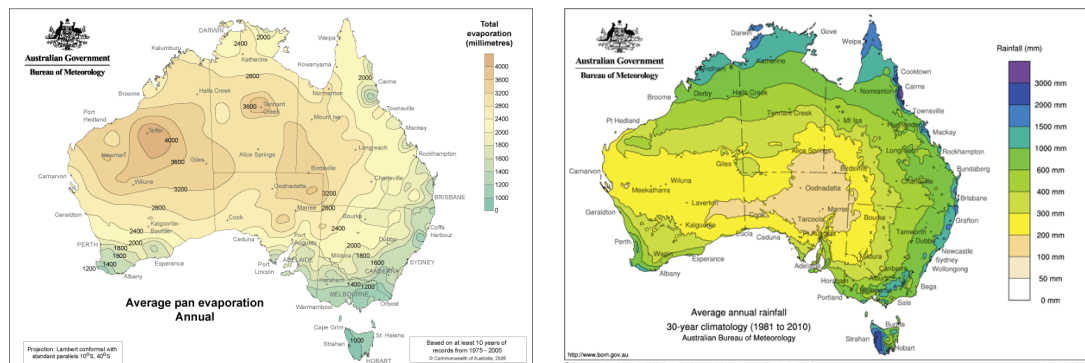


Propuesta de Trabajo Final – Estudio de las áreas climáticas de Australia

El gobierno australiano proporciona datos públicos de distintas variables (Temperatura, lluvia, evaporación, etc.) de todo el territorio. Proporciona esta información a través de la plataforma del 'Bureau of Meteorology' en el siguiente enlace:

<http://www.bom.gov.au/climate/data/>



Los datos que proporciona tienen un formato GIS (Geographic Information System), que serán importados en formato de CSV y luego se procesarán gracias a las librerías de Python 'GeoPandas' y 'Shapely', con la que he tenido experiencia trabajando en el pasado.

Con toda esta información pública recopilare un gran dataset con información sobre variables como:

- Lluvia (mm)
- Días de lluvia
- Duración solar (Horas diarias)
- Exposición solar (Megajoules por metro cuadrado)
- Evaporación
- Evapotranspiración
- Índice UV
- Humedad relativa (A las 9:00 y a las 15:00)
- Distribución de nubes (Nº de nubes al día)
- Temperatura (Max, Min y Media) (°C)
- Días potenciales de helada anuales
- Posición geográfica

Se ofrecen datos de estas variables promediadas para cada mes del año, para cada estación o el promedio del año entero. Para este estudio se tomarán datos mensuales. En total, todos los datos descargados son alrededor de 630 Mb, lo cual es bastante alto para archivos de texto, pero en el proceso de preparación es posible que se tengan que retirar algunos datos no válidos.

Objetivo y Métodos ML

La intención de este estudio es utilizar **PCA** y métodos de **clustering** para delimitar zonas del país con características similares, y poder identificar climas o ecosistemas distintos dentro del país.

Tratamiento de datos

Todos los datasets de las distintas medidas se proporcionan por separado, por lo que este estudio conlleva una gran parte de preparación de los datos, transformando las entradas a un formato aceptable para el estudio, y juntando los datos para asignar los valores correctos de cada medida al punto geográfico que les pertenece, esto es porque las variables como la temperatura obtienen datos mucho más granulares que algunos otros como la humedad relativa. En la imagen inferior se muestran los metadatos de cada fichero, donde se explica la granularidad de los datos geográficos:

ncols	178	ncols	178	ncols	1681
nrows	139	nrows	142	nrows	1361
xllcorner	111.875	xllcorner	111.875	xllcorner	112
yllcorner	-44.625	yllcorner	-44.625	yllcorner	-44
cellsize	0.25	cellsize	0.25	cellsize	0.025
NODATA_value	-9999	NODATA_value	-9999	NODATA_value	-9999

Por esta razón, como el tamaño de los cuadrantes varía, se deben ajustar y preparar antes de realizar el estudio de Machine Learning.