**Análisis de datos**

**Dataset :** Datos de distintas estaciones meteorológicas de Australia.

1. Los datos no son todos correlativos, hay varios saltos de fecha, por ejemplo varios días o un mes entero dependiendo de la ciudad.
2. Campos nubosidad (Cloud9am y Cloud3pm). Estos campos están en octales, de 0 a 8. Hay algunos datos con valores 9.
   1. El campo “Cloud9am” tiene 2 valores 9
   2. El campo “Cloud3pm” tiene 1 valor 9
3. No se puede inferir el Label RainTomorrow a partir del RainToday del día siguiente porque en todos los casos de fechas correlativas estos datos son iguales, es decir que cuando uno es null el otro también.
4. Datos categóricos :
   1. Cantidad de datos en Location : 49
   2. Cantidad de datos en RainToday : 2
   3. Cantidad de datos en WindGustDir : 16
   4. Cantidad de datos en WindDir9am : 16
   5. Cantidad de datos en WindDir3pm : 16
5. Se propone pasar los siguientes features categóricos a One Hot
   1. Location : 49
   2. RainToday : 2
   3. WindGustDir : 16
   4. WindDir9am : 16
   5. WindDir3pm : 16
6. Los campos Cloud9am y Cloud3pm no se representarán como One Hot ya que estos tienen un valor de peso. Por ejemplo, un valor de cero indica que está totalmente despejado, un valor de 8 indica que está totalmente nublado. Es decir que los valores de estos campos no tienen todos el mismo peso para la predicción.