Análisis de Datos y Aprendizaje Máquina con Tensorflow 2.0: Pre-procesamiento de Datos para Aprendizaje Máquina

2019/09/30

1 Exploración de datos

Objetivo: Conocer las herramientas de exploración de datos y pre-procesamiento para aprendizaje máquina.

```
In [1]: import pandas as pd
```

```
In [2]: import numpy as np
        import pandas as pd
        from sklearn.datasets import load_iris


        iris = load_iris()
        iris['feature_names']
```

```
Out[2]: ['sepal length (cm)',
         'sepal width (cm)',
         'petal length (cm)',
         'petal width (cm)']
```

```
In [3]: iris['data']
```

```
Out[3]: array([[5.1, 3.5, 1.4, 0.2],
               [4.9, 3. , 1.4, 0.2],
               [4.7, 3.2, 1.3, 0.2],
               [4.6, 3.1, 1.5, 0.2],
               [5. , 3.6, 1.4, 0.2],
               [5.4, 3.9, 1.7, 0.4],
               [4.6, 3.4, 1.4, 0.3],
               [5. , 3.4, 1.5, 0.2],
               [4.4, 2.9, 1.4, 0.2],
               [4.9, 3.1, 1.5, 0.1],
               [5.4, 3.7, 1.5, 0.2],
               [4.8, 3.4, 1.6, 0.2],
               [4.8, 3. , 1.4, 0.1],
[4.3, 3. , 1.1, 0.1],
[5.8, 4. , 1.2, 0.2],
[5.7, 4.4, 1.5, 0.4],
[5.4, 3.9, 1.3, 0.4],
[5.1, 3.5, 1.4, 0.3],
[5.7, 3.8, 1.7, 0.3],
[5.1, 3.8, 1.5, 0.3],
[5.4, 3.4, 1.7, 0.2],
[5.1, 3.7, 1.5, 0.4],
[4.6, 3.6, 1. , 0.2],
[5.1, 3.3, 1.7, 0.5],
[4.8, 3.4, 1.9, 0.2],
[5. , 3. , 1.6, 0.2],
[5. , 3.4, 1.6, 0.4],
[5.2, 3.5, 1.5, 0.2],
[5.2, 3.4, 1.4, 0.2],
[4.7, 3.2, 1.6, 0.2],
[4.8, 3.1, 1.6, 0.2],
[5.4, 3.4, 1.5, 0.4],
[5.2, 4.1, 1.5, 0.1],
[5.5, 4.2, 1.4, 0.2],
[4.9, 3.1, 1.5, 0.2],
[5. , 3.2, 1.2, 0.2],
[5.5, 3.5, 1.3, 0.2],
[4.9, 3.6, 1.4, 0.1],
[4.4, 3. , 1.3, 0.2],
[5.1, 3.4, 1.5, 0.2],
[5. , 3.5, 1.3, 0.3],
[4.5, 2.3, 1.3, 0.3],
[4.4, 3.2, 1.3, 0.2],
[5. , 3.5, 1.6, 0.6],
[5.1, 3.8, 1.9, 0.4],
[4.8, 3. , 1.4, 0.3],
[5.1, 3.8, 1.6, 0.2],
[4.6, 3.2, 1.4, 0.2],
[5.3, 3.7, 1.5, 0.2],
[5. , 3.3, 1.4, 0.2],
[7. , 3.2, 4.7, 1.4],
[6.4, 3.2, 4.5, 1.5],
[6.9, 3.1, 4.9, 1.5],
[5.5, 2.3, 4. , 1.3],
[6.5, 2.8, 4.6, 1.5],
[5.7, 2.8, 4.5, 1.3],
[6.3, 3.3, 4.7, 1.6],
[4.9, 2.4, 3.3, 1. ],
[6.6, 2.9, 4.6, 1.3],
[5.2, 2.7, 3.9, 1.4],
[5. , 2. , 3.5, 1. ],
[5.9, 3. , 4.2, 1.5],
[6. , 2.2, 4. , 1. ],
[6.1, 2.9, 4.7, 1.4],
[5.6, 2.9, 3.6, 1.3],
[6.7, 3.1, 4.4, 1.4],
[5.6, 3. , 4.5, 1.5],
[5.8, 2.7, 4.1, 1. ],
[6.2, 2.2, 4.5, 1.5],
[5.6, 2.5, 3.9, 1.1],
[5.9, 3.2, 4.8, 1.8],
[6.1, 2.8, 4. , 1.3],
[6.3, 2.5, 4.9, 1.5],
[6.1, 2.8, 4.7, 1.2],
[6.4, 2.9, 4.3, 1.3],
[6.6, 3. , 4.4, 1.4],
[6.8, 2.8, 4.8, 1.4],
[6.7, 3. , 5. , 1.7],
[6. , 2.9, 4.5, 1.5],
[5.7, 2.6, 3.5, 1. ],
[5.5, 2.4, 3.8, 1.1],
[5.5, 2.4, 3.7, 1. ],
[5.8, 2.7, 3.9, 1.2],
[6. , 2.7, 5.1, 1.6],
[5.4, 3. , 4.5, 1.5],
[6. , 3.4, 4.5, 1.6],
[6.7, 3.1, 4.7, 1.5],
[6.3, 2.3, 4.4, 1.3],
[5.6, 3. , 4.1, 1.3],
[5.5, 2.5, 4. , 1.3],
[5.5, 2.6, 4.4, 1.2],
[6.1, 3. , 4.6, 1.4],
[5.8, 2.6, 4. , 1.2],
[5. , 2.3, 3.3, 1. ],
[5.6, 2.7, 4.2, 1.3],
[5.7, 3. , 4.2, 1.2],
[5.7, 2.9, 4.2, 1.3],
[6.2, 2.9, 4.3, 1.3],
[5.1, 2.5, 3. , 1.1],
[5.7, 2.8, 4.1, 1.3],
[6.3, 3.3, 6. , 2.5],
[5.8, 2.7, 5.1, 1.9],
[7.1, 3. , 5.9, 2.1],
[6.3, 2.9, 5.6, 1.8],
[6.5, 3. , 5.8, 2.2],
[7.6, 3. , 6.6, 2.1],
[4.9, 2.5, 4.5, 1.7],
[7.3, 2.9, 6.3, 1.8],
[6.7, 2.5, 5.8, 1.8],
[7.2, 3.6, 6.1, 2.5],
[6.5, 3.2, 5.1, 2. ],
[6.4, 2.7, 5.3, 1.9],
[6.8, 3. , 5.5, 2.1],
[5.7, 2.5, 5. , 2. ],
[5.8, 2.8, 5.1, 2.4],
[6.4, 3.2, 5.3, 2.3],
[6.5, 3. , 5.5, 1.8],
[7.7, 3.8, 6.7, 2.2],
[7.7, 2.6, 6.9, 2.3],
[6. , 2.2, 5. , 1.5],
[6.9, 3.2, 5.7, 2.3],
[5.6, 2.8, 4.9, 2. ],
[7.7, 2.8, 6.7, 2. ],
[6.3, 2.7, 4.9, 1.8],
[6.7, 3.3, 5.7, 2.1],
[7.2, 3.2, 6. , 1.8],
[6.2, 2.8, 4.8, 1.8],
[6.1, 3. , 4.9, 1.8],
[6.4, 2.8, 5.6, 2.1],
[7.2, 3. , 5.8, 1.6],
[7.4, 2.8, 6.1, 1.9],
[7.9, 3.8, 6.4, 2. ],
[6.4, 2.8, 5.6, 2.2],
[6.3, 2.8, 5.1, 1.5],
[6.1, 2.6, 5.6, 1.4],
[7.7, 3. , 6.1, 2.3],
[6.3, 3.4, 5.6, 2.4],
[6.4, 3.1, 5.5, 1.8],
[6. , 3. , 4.8, 1.8],
[6.9, 3.1, 5.4, 2.1],
[6.7, 3.1, 5.6, 2.4],
[6.9, 3.1, 5.1, 2.3],
[5.8, 2.7, 5.1, 1.9],
[6.8, 3.2, 5.9, 2.3],
[6.7, 3.3, 5.7, 2.5],
[6.7, 3. , 5.2, 2.3],
[6.3, 2.5, 5. , 1.9],
[6.5, 3. , 5.2, 2. ],
[6.2, 3.4, 5.4, 2.3],
[5.9, 3. , 5.1, 1.8]])
```

- Asignar nombres a columnas

```
In [4]: names = ['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
        'Species']
```

- Cargar datos y etiquetas

```
In [5]: df = pd.DataFrame(data= np.c_[iris['data'], iris['target']],
                          columns= names)
```

- Mostrar dataset

```
In [6]: df

Out[6]:      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
        0              5.1           3.5            1.4           0.2      0.0
        1              4.9           3.0            1.4           0.2      0.0
        2              4.7           3.2            1.3           0.2      0.0
        3              4.6           3.1            1.5           0.2      0.0
        4              5.0           3.6            1.4           0.2      0.0
        ..             ...           ...            ...           ...      ...
        145            6.7           3.0            5.2           2.3      2.0
        146            6.3           2.5            5.0           1.9      2.0
        147            6.5           3.0            5.2           2.0      2.0
        148            6.2           3.4            5.4           2.3      2.0
        149            5.9           3.0            5.1           1.8      2.0

        [150 rows x 5 columns]
```

- Primeros 5 renglones

```
In [7]: df.head(6)

Out[7]:    SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
        0            5.1           3.5            1.4           0.2      0.0
        1            4.9           3.0            1.4           0.2      0.0
        2            4.7           3.2            1.3           0.2      0.0
        3            4.6           3.1            1.5           0.2      0.0
        4            5.0           3.6            1.4           0.2      0.0
        5            5.4           3.9            1.7           0.4      0.0
```

- Ultimos 3

```
In [8]: df.tail(3)

Out[8]:      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
        147            6.5           3.0            5.2           2.0      2.0
        148            6.2           3.4            5.4           2.3      2.0
        149            5.9           3.0            5.1           1.8      2.0
```

- Nombre columnas

```
In [9]: df.columns

Out[9]: Index(['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
               'Species'],
              dtype='object')
```

- Media, std y conteos

```
In [10]: df.describe()

Out[10]:        SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm     Species
         count     150.000000    150.000000     150.000000    150.000000  150.000000
```

```
mean        5.843333     3.057333      3.758000     1.199333    1.000000
std         0.828066     0.435866      1.765298     0.762238    0.819232
min         4.300000     2.000000      1.000000     0.100000    0.000000
25%         5.100000     2.800000      1.600000     0.300000    0.000000
50%         5.800000     3.000000      4.350000     1.300000    1.000000
75%         6.400000     3.300000      5.100000     1.800000    2.000000
max         7.900000     4.400000      6.900000     2.500000    2.000000
```

- Extraer nombres de columnas

```
In [11]: df.columns.values.tolist()
```

```
Out[11]: ['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm', 'Species']
```

```
In [12]: df.isnull().any().any()
```

```
Out[12]: False
```

1.0.1 Crear sub-dataset

```
In [13]: setosa = df[df['Species']==0]
         setosa.shape
```

```
Out[13]: (50, 5)
```

```
In [14]: setosa.head()
```

```
Out[14]:    SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
         0            5.1           3.5            1.4           0.2      0.0
         1            4.9           3.0            1.4           0.2      0.0
         2            4.7           3.2            1.3           0.2      0.0
         3            4.6           3.1            1.5           0.2      0.0
         4            5.0           3.6            1.4           0.2      0.0
```

- Ordenar

```
In [15]: setosa = setosa.sort_values(by=['SepalLengthCm'])
         setosa.head(10)
```

```
Out[15]:     SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
         13            4.3           3.0            1.1           0.1      0.0
         8             4.4           2.9            1.4           0.2      0.0
         38            4.4           3.0            1.3           0.2      0.0
         42            4.4           3.2            1.3           0.2      0.0
         41            4.5           2.3            1.3           0.3      0.0
         22            4.6           3.6            1.0           0.2      0.0
         3             4.6           3.1            1.5           0.2      0.0
         47            4.6           3.2            1.4           0.2      0.0
         6             4.6           3.4            1.4           0.3      0.0
         2             4.7           3.2            1.3           0.2      0.0
```

- inplace para actualizar

```
In [16]: setosa.sort_values(by=['PetalLengthCm'], inplace=True)
         setosa.head(10)
```

```
Out[16]:     SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
         22            4.6           3.6            1.0           0.2      0.0
         13            4.3           3.0            1.1           0.1      0.0
         35            5.0           3.2            1.2           0.2      0.0
         14            5.8           4.0            1.2           0.2      0.0
         41            4.5           2.3            1.3           0.3      0.0
         38            4.4           3.0            1.3           0.2      0.0
         36            5.5           3.5            1.3           0.2      0.0
         16            5.4           3.9            1.3           0.4      0.0
         2             4.7           3.2            1.3           0.2      0.0
         42            4.4           3.2            1.3           0.2      0.0
```

- Condición

```
In [17]: df['class']=np.where(df['SepalLengthCm'] > 4.7 ,  1, 0)

         df.head()
```

```
Out[17]:    SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
         0            5.1           3.5            1.4           0.2      0.0      1
         1            4.9           3.0            1.4           0.2      0.0      1
         2            4.7           3.2            1.3           0.2      0.0      0
         3            4.6           3.1            1.5           0.2      0.0      0
         4            5.0           3.6            1.4           0.2      0.0      1
```

```
In [18]: df.tail()
```

```
Out[18]:      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
         145            6.7           3.0            5.2           2.3      2.0      1
         146            6.3           2.5            5.0           1.9      2.0      1
         147            6.5           3.0            5.2           2.0      2.0      1
         148            6.2           3.4            5.4           2.3      2.0      1
         149            5.9           3.0            5.1           1.8      2.0      1
```

- groupby

```
In [19]: df.groupby('Species')['SepalLengthCm'].apply(lambda x: np.mean(x))
```

```
Out[19]: Species
         0.0    5.006
         1.0    5.936
         2.0    6.588
         Name: SepalLengthCm, dtype: float64
```

- Estadística por columna

```
In [20]: df['SepalLengthCm'].count()
```

```
Out[20]: 150
```

```
In [21]: df['SepalLengthCm'].mean()
```

```
Out[21]: 5.843333333333334
```

```
In [22]: df['SepalLengthCm'].std()
```

```
Out[22]: 0.828066127977863
```

- Máximo y mínimo

```
In [23]: df['SepalLengthCm'].min()
```

```
Out[23]: 4.3
```

```
In [24]: df['SepalLengthCm'].max()
```

```
Out[24]: 7.9
```

- Selección de columna

```
In [25]: df[['SepalLengthCm', 'Species']]
```

```
Out[25]:      SepalLengthCm  Species
         0              5.1      0.0
         1              4.9      0.0
         2              4.7      0.0
         3              4.6      0.0
         4              5.0      0.0
         ..             ...      ...
         145            6.7      2.0
         146            6.3      2.0
         147            6.5      2.0
         148            6.2      2.0
         149            5.9      2.0

         [150 rows x 2 columns]
```

- Suma

```
In [26]: df.sum()
```

```
Out[26]: SepalLengthCm    876.5
         SepalWidthCm     458.6
         PetalLengthCm    563.7
         PetalWidthCm     179.9
         Species          150.0
         class            139.0
         dtype: float64
```

```
In [27]: df.cumsum()
```

```
Out[27]:      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
         0              5.1           3.5            1.4           0.2      0.0    1.0
         1             10.0           6.5            2.8           0.4      0.0    2.0
         2             14.7           9.7            4.1           0.6      0.0    2.0
         3             19.3          12.8            5.6           0.8      0.0    2.0
         4             24.3          16.4            7.0           1.0      0.0    3.0
         ..             ...           ...            ...           ...      ...    ...
         145          851.6         446.7          543.0         171.9    142.0  135.0
         146          857.9         449.2          548.0         173.8    144.0  136.0
         147          864.4         452.2          553.2         175.8    146.0  137.0
         148          870.6         455.6          558.6         178.1    148.0  138.0
         149          876.5         458.6          563.7         179.9    150.0  139.0

         [150 rows x 6 columns]
```

- Tipo de dato

```
In [28]: type(df)
```

```
Out[28]: pandas.core.frame.DataFrame
```

- Forma

```
In [29]: df.shape
```

```
Out[29]: (150, 6)
```

- Índice

```
In [30]: df.index
```

```
Out[30]: RangeIndex(start=0, stop=150, step=1)
```

- Información

```
In [31]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
SepalLengthCm    150 non-null float64
SepalWidthCm     150 non-null float64
PetalLengthCm    150 non-null float64
PetalWidthCm     150 non-null float64
Species          150 non-null float64
class            150 non-null int64
dtypes: float64(5), int64(1)
memory usage: 7.2 KB
```

- Tipo columna

```
In [32]: type(df['SepalLengthCm'])
```

```
Out[32]: pandas.core.series.Series
```

- Columnas específicas

```
In [33]: df[['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm']]
```

```
Out[33]:      SepalLengthCm  SepalWidthCm  PetalLengthCm
         0              5.1           3.5            1.4
         1              4.9           3.0            1.4
         2              4.7           3.2            1.3
         3              4.6           3.1            1.5
         4              5.0           3.6            1.4
         ..             ...           ...            ...
         145            6.7           3.0            5.2
         146            6.3           2.5            5.0
         147            6.5           3.0            5.2
         148            6.2           3.4            5.4
         149            5.9           3.0            5.1

         [150 rows x 3 columns]
```

```
In [34]: df[['Species']].head(15)
```

```
Out[34]:     Species
         0       0.0
         1       0.0
         2       0.0
         3       0.0
         4       0.0
         5       0.0
         6       0.0
         7       0.0
         8       0.0
         9       0.0
         10      0.0
         11      0.0
         12      0.0
         13      0.0
         14      0.0
```

- Renglones específicos

```
In [35]: df[2:6]
```

```
Out[35]:    SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
         2            4.7           3.2            1.3           0.2      0.0      0
         3            4.6           3.1            1.5           0.2      0.0      0
         4            5.0           3.6            1.4           0.2      0.0      1
         5            5.4           3.9            1.7           0.4      0.0      1
```

```
In [36]: df[['SepalLengthCm', 'Species']][2:6]
```

```
Out[36]:    SepalLengthCm  Species
         2            4.7      0.0
         3            4.6      0.0
         4            5.0      0.0
         5            5.4      0.0
```

- iloc

```
In [37]: df.loc[:, ['SepalLengthCm', 'Species']]
```

```
Out[37]:      SepalLengthCm  Species
         0              5.1      0.0
         1              4.9      0.0
         2              4.7      0.0
         3              4.6      0.0
         4              5.0      0.0
         ..             ...      ...
         145            6.7      2.0
         146            6.3      2.0
         147            6.5      2.0
         148            6.2      2.0
         149            5.9      2.0

         [150 rows x 2 columns]
```

- Condición

```
In [38]: df.loc[:, 'Species'] == 0
```

```
Out[38]: 0       True
         1       True
         2       True
         3       True
         4       True
                ...
         145    False
         146    False
         147    False
         148    False
         149    False
         Name: Species, Length: 150, dtype: bool
```

```
In [39]: df.loc[df.loc[:, 'Species'] == 1]
```

```
Out[39]:     SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
         50            7.0           3.2            4.7           1.4      1.0      1
         51            6.4           3.2            4.5           1.5      1.0      1
         52            6.9           3.1            4.9           1.5      1.0      1
         53            5.5           2.3            4.0           1.3      1.0      1
         54            6.5           2.8            4.6           1.5      1.0      1
         55            5.7           2.8            4.5           1.3      1.0      1
```

```
56            6.3           3.3             4.7           1.6      1.0      1
57            4.9           2.4             3.3           1.0      1.0      1
58            6.6           2.9             4.6           1.3      1.0      1
59            5.2           2.7             3.9           1.4      1.0      1
60            5.0           2.0             3.5           1.0      1.0      1
61            5.9           3.0             4.2           1.5      1.0      1
62            6.0           2.2             4.0           1.0      1.0      1
63            6.1           2.9             4.7           1.4      1.0      1
64            5.6           2.9             3.6           1.3      1.0      1
65            6.7           3.1             4.4           1.4      1.0      1
66            5.6           3.0             4.5           1.5      1.0      1
67            5.8           2.7             4.1           1.0      1.0      1
68            6.2           2.2             4.5           1.5      1.0      1
69            5.6           2.5             3.9           1.1      1.0      1
70            5.9           3.2             4.8           1.8      1.0      1
71            6.1           2.8             4.0           1.3      1.0      1
72            6.3           2.5             4.9           1.5      1.0      1
73            6.1           2.8             4.7           1.2      1.0      1
74            6.4           2.9             4.3           1.3      1.0      1
75            6.6           3.0             4.4           1.4      1.0      1
76            6.8           2.8             4.8           1.4      1.0      1
77            6.7           3.0             5.0           1.7      1.0      1
78            6.0           2.9             4.5           1.5      1.0      1
79            5.7           2.6             3.5           1.0      1.0      1
80            5.5           2.4             3.8           1.1      1.0      1
81            5.5           2.4             3.7           1.0      1.0      1
82            5.8           2.7             3.9           1.2      1.0      1
83            6.0           2.7             5.1           1.6      1.0      1
84            5.4           3.0             4.5           1.5      1.0      1
85            6.0           3.4             4.5           1.6      1.0      1
86            6.7           3.1             4.7           1.5      1.0      1
87            6.3           2.3             4.4           1.3      1.0      1
88            5.6           3.0             4.1           1.3      1.0      1
89            5.5           2.5             4.0           1.3      1.0      1
90            5.5           2.6             4.4           1.2      1.0      1
91            6.1           3.0             4.6           1.4      1.0      1
92            5.8           2.6             4.0           1.2      1.0      1
93            5.0           2.3             3.3           1.0      1.0      1
94            5.6           2.7             4.2           1.3      1.0      1
95            5.7           3.0             4.2           1.2      1.0      1
96            5.7           2.9             4.2           1.3      1.0      1
97            6.2           2.9             4.3           1.3      1.0      1
98            5.1           2.5             3.0           1.1      1.0      1
99            5.7           2.8             4.1           1.3      1.0      1
```

- Renglón y columna específicos a nuevo df

```
In [40]: y = df.iloc[:, -1:]
```

```
In [41]: y
```

```
Out[41]:      class
        0        1
        1        1
        2        0
        3        0
        4        1
        ..     ...
        145      1
        146      1
        147      1
        148      1
        149      1

        [150 rows x 1 columns]
```

```
In [42]: type(y)
```

```
Out[42]: pandas.core.frame.DataFrame
```

```
In [43]: df.iloc[[0,2], 0:4]
```

```
Out[43]:    SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
        0             5.1           3.5            1.4           0.2
        2             4.7           3.2            1.3           0.2
```

```
In [44]: df.iloc[0:4, 0:2]
```

```
Out[44]:    SepalLengthCm  SepalWidthCm
        0             5.1           3.5
        1             4.9           3.0
        2             4.7           3.2
        3             4.6           3.1
```

- Condición en columna

```
In [45]: df[df.SepalLengthCm >= 7]
```

```
Out[45]:      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
        50             7.0           3.2            4.7           1.4      1.0      1
        102            7.1           3.0            5.9           2.1      2.0      1
        105            7.6           3.0            6.6           2.1      2.0      1
        107            7.3           2.9            6.3           1.8      2.0      1
        109            7.2           3.6            6.1           2.5      2.0      1
        117            7.7           3.8            6.7           2.2      2.0      1
        118            7.7           2.6            6.9           2.3      2.0      1
        122            7.7           2.8            6.7           2.0      2.0      1
        125            7.2           3.2            6.0           1.8      2.0      1
        129            7.2           3.0            5.8           1.6      2.0      1
        130            7.4           2.8            6.1           1.9      2.0      1
        131            7.9           3.8            6.4           2.0      2.0      1
        135            7.7           3.0            6.1           2.3      2.0      1
```

```
In [46]: df[df.SepalLengthCm == df.SepalLengthCm .max()]

Out[46]:      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species  class
         131            7.9           3.8            6.4           2.0      2.0      1
```

- Explorar otro dataset
- Extraer información de otro dataset