

TITANIC SIMPLE

In [1]:

Carga la base de datos titanic train

In [2]:

In [3]:

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2800
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

In [4]:

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2800
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

In [5]:

(891, 4)
(714, 4)

In [6]:

Out[6]:

	Age	Pclass	Survived
Age	1.000000	-0.369226	-0.077221
Pclass	-0.369226	1.000000	-0.359653
Survived	-0.077221	-0.359653	1.000000

Selecciona solo la edad y entrena usando cross-validaton y RandomForestClassifier

In [14]:

In [15]:

Out[15]:

	Age
0	22.0
1	38.0
2	26.0
3	35.0
4	35.0

In [16]:

Añade el sexo transformándolo a números y la clase.

Vuelve a entrenar con las 3 feaures

In [22]:

```
/Library/Python/2.7/site-packages/ipykernel/__main__.py:10: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

Out[22]: 0.81654683344824197

Cambia el valor de n_estimators

¿Qué número de estimadores es el que mejor valor da?

In [21]:

```
Numero estimadores 1
0.778705801241
Numero estimadores 11
0.795548113858
Numero estimadores 21
0.805358022259
Numero estimadores 31
0.815158081355
Numero estimadores 41
0.81372993204
Numero estimadores 51
0.802511572934
Numero estimadores 61
0.812351029252
Numero estimadores 71
0.815158081355
Numero estimadores 81
0.80252142224
Numero estimadores 91
0.812331330641
```

Añade más valores a ver hasta dónde consigues llegar

- Por ejemplo suma los familiares a ver si tener familia ayudaba
- Agrupa los nombres mirando si tienen algún título (mira el ejemplo para ver cómo hacerlo)

<https://www.kaggle.com/mrisdal/exploring-survival-on-the-titanic> (<https://www.kaggle.com/mrisdal/exploring-survival-on-the-titanic>).

In []: