

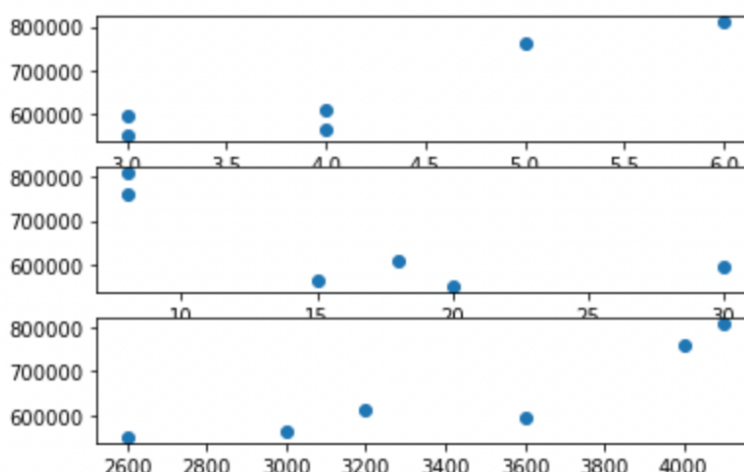
1 MACHINE LEARNING – REGRESIÓN MULTIVARIANTE

El problema planteado es un poco más complejo que el anterior: Tenemos un listado de casas con información básica y queremos predecir el precio de estas según el área, el número de habitaciones y la antigüedad de la casa.

area	bedrooms	age	price
2600	3	20	550000
3000	4	15	565000
3200		18	610000
3600	3	30	595000
4000	5	8	760000
4100	6	8	810000

1º) Debemos reemplazar el valor nulo (podemos hacerlo por la mediana - la media da decimales-)

Podemos analizar los datos y ver el resultado gráfico de cada una de las variables respecto al precio:



Se puede intuir que hay una variación lineal en cada una de las variables, y si las juntamos todas, la ecuación sería:

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + b$$

Y adecuada a nuestro ejemplo:

$$price = m_1 * area + m_2 * bedrooms + m_3 * age + b$$



GENERALITAT
VALENCIANA



UNIÓN EUROPEA
Fondo Social Europeo
El FSE invierte en tu futuro

03008915 C/ Ferrocarril, 22, 03570 La Vila Joiosa Tel 966870140 Fax 966870141 <http://portal.edu.gva.es/iesmarcoszaragoza>

En el ejemplo, price es la variable dependiente y área, bedrooms y age son las variables independientes.

Para la aplicación de algoritmos de machine learning vamos a usar la librería **Scikit-learn (sklearn)**

Dentro de esta librería, usaremos el modelo lineal:

```
from sklearn import linear_model  
reg = linear_model.LinearRegression()  
reg.fit(df.drop('price',axis='columns'),df.price)
```

Con este método (fit) lo que hacemos es entrenar al algoritmo con los valores de x (área, bedrooms y age) e y (precio)

podemos predecir algún valor con:

```
reg.predict([[3000, 3, 40]])
```

El modelo lineal tiene los coeficientes de la línea de regresión que hemos hablado anteriormente. Para conocerlos, podemos acceder a:

```
reg.coef_ y reg.intercept_
```