

## 1-. Subir el archivo csv en la plataforma de BigML.

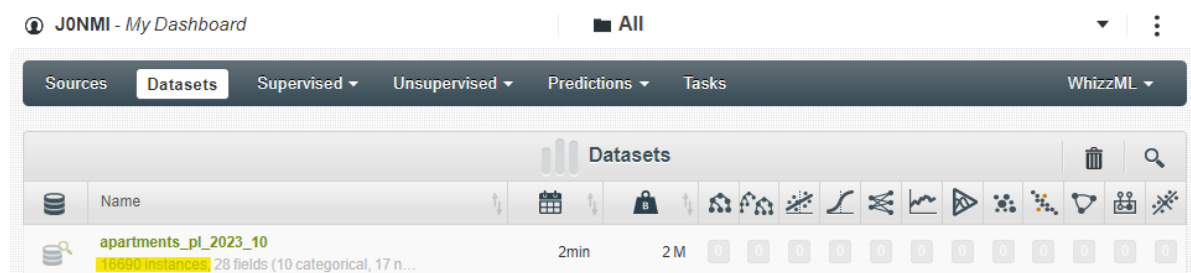
Type	Name	Details	Age	Size
csv	apartments_pt_2023_10.csv	In progress: 10%		
csv	credit_customers.csv	closed, table, 21 fields (14 categorical, 7 numeric)	5d 5h	149.4 KB
image	grape-strawberry classification	open, image, 70 sources, 237 fields (1 categorical, 234 numeric, 1 path, 1 image)	6d 6h	13.3 MB
image	Firetruck Anomalies	open, image, 118 sources, 514 fields (512 numeric, 1 path, 1 image)	6d 6h	8.5 MB

## 2-. Analizar y justificar el tipo de problema que se enfrentan.

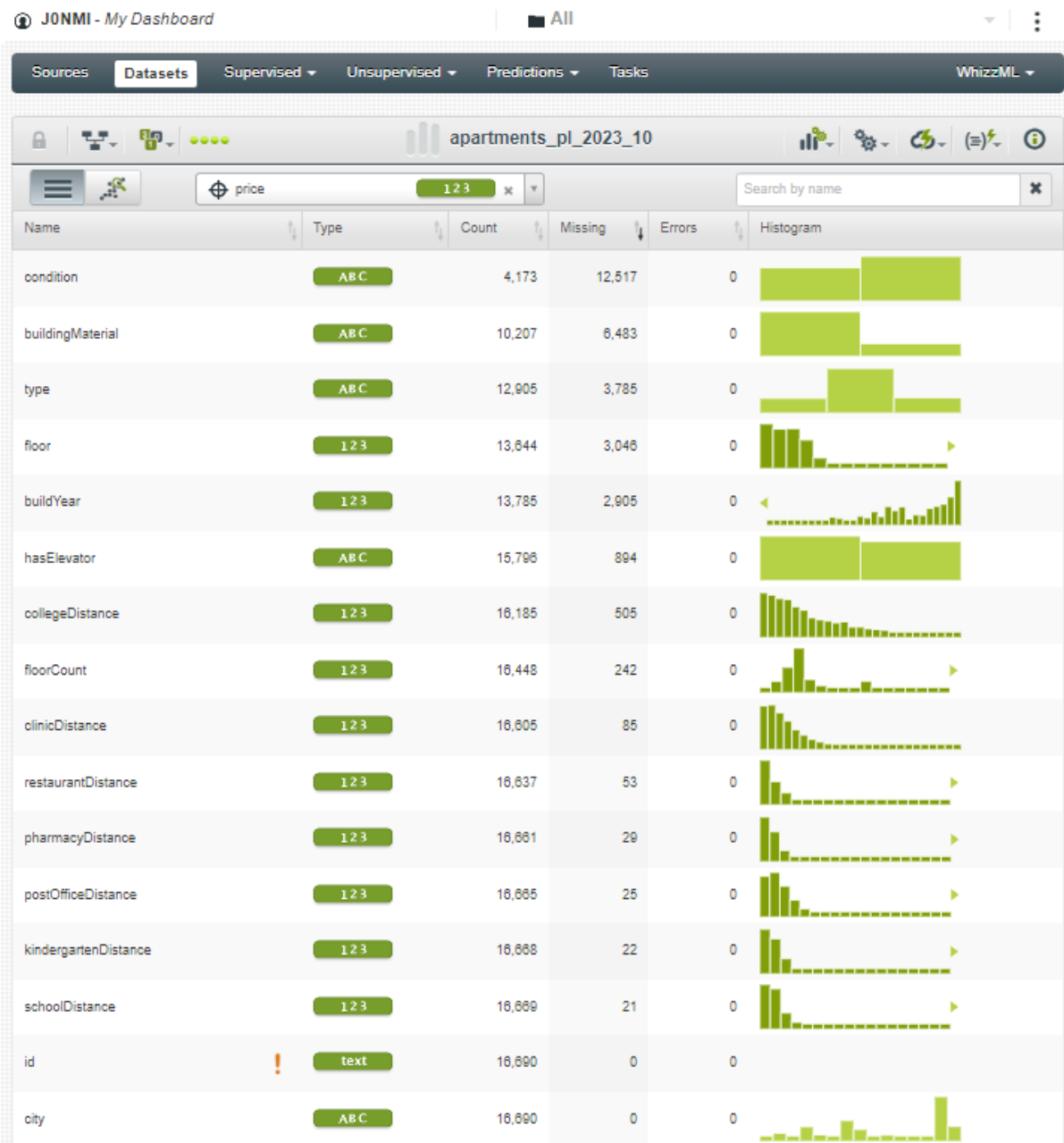
Estamos ante un problema interesante: tenemos datos sobre varias características de casas y queremos predecir sus precios. Para hacer esto, vamos a utilizar herramientas de inteligencia artificial y análisis de datos para encontrar patrones en estos datos. La idea es entender cómo las diferentes características de una casa influyen en su precio. Esto no solo es emocionante desde el punto de vista técnico, sino que también tiene aplicaciones prácticas en el mundo real, como ayudar en decisiones de compra y venta de propiedades.

### 3-. ¿Cuántos datos contiene el conjunto de datos? ¿En qué características faltan valores?

Como vemos en la captura, hay 16690 datos y 28 campos en el dataset.



Como vemos, hay 14 de 28 campos con “Missing Values”.



**4. Limpiar las características que contienen missing values y características que el BigML por defecto no considera útiles. (Por ejemplo, si una característica tiene missing values, tenéis que eliminarla.)**

Desde las hojas de cálculo de google, nos posicionamos en el campo que deseamos eliminar, clic derecho y eliminamos la columna. Realizaremos este paso con todos los campos que contienen "Missing values".

R	S	T	U	V	W	X	
nacyDistan	ownership	building				hasSecurity	has
73	cooperative	brick				no	no
	condominium	brick				no	no
205	cooperative	concret				yes	yes
177	condominium	brick				no	no
43	condominium	brick				yes	yes
239	cooperative	brick				no	no
471	condominium	brick				no	no
165	condominium	brick				no	yes
	cooperative	concret				no	yes
	condominium	brick				no	yes
149	condominium	brick				no	yes
103	condominium					no	yes
222	condominium	brick				no	yes
92	cooperative	concret				no	yes
267	cooperative					no	no
	condominium					no	no
258	condominium	brick				no	no

**5. Dividir los datos en 80% de entrenamiento y 20% de prueba.**

Hacemos clic en el icono del gráfico > "TRAINING | TEST SPLIT".

JONMI - My Dashboard

All

Sources Datasets Supervised Unsupervised Predictions Tasks WhizzML

apartaments 2023 10 CLEAN

price

Name Type

city ABC

squareMeters 123

rooms 123 16,690 0 0

SAMPLE AND FILTER DATASET

TRAINING | TEST SPLIT

SAMPLE

FILTER

REMOVE DUPLICATES

TRANSFORM DATASET

ADD FIELDS

AGGREGATE INSTANCES

JOIN DATASETS

MERGE DATASETS

ORDER INSTANCES

El siguiente paso será configurar el split para el dataset. En este caso, dejamos los porcentajes por defecto ya que coinciden con las indicaciones. Una vez configurados los porcentajes hacemos clic en el botón de abajo a la derecha “Create Training | Test”.

## 6-. Eliminar la variable dependiente de los datos de prueba.

Vamos a el Dataset de Test y hacemos clic en descargar csv eliminando la variable price.

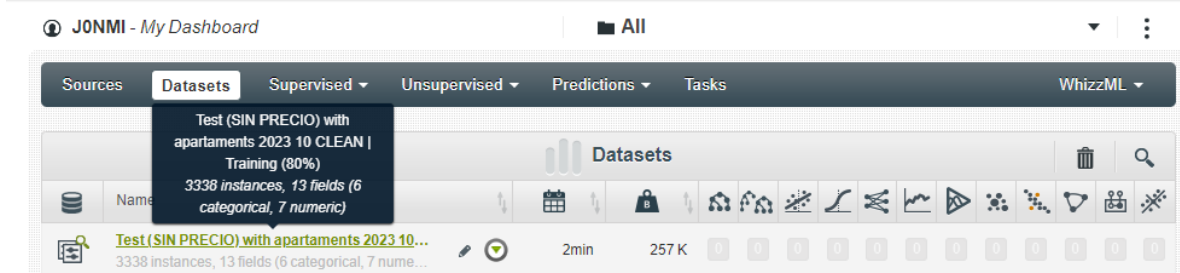
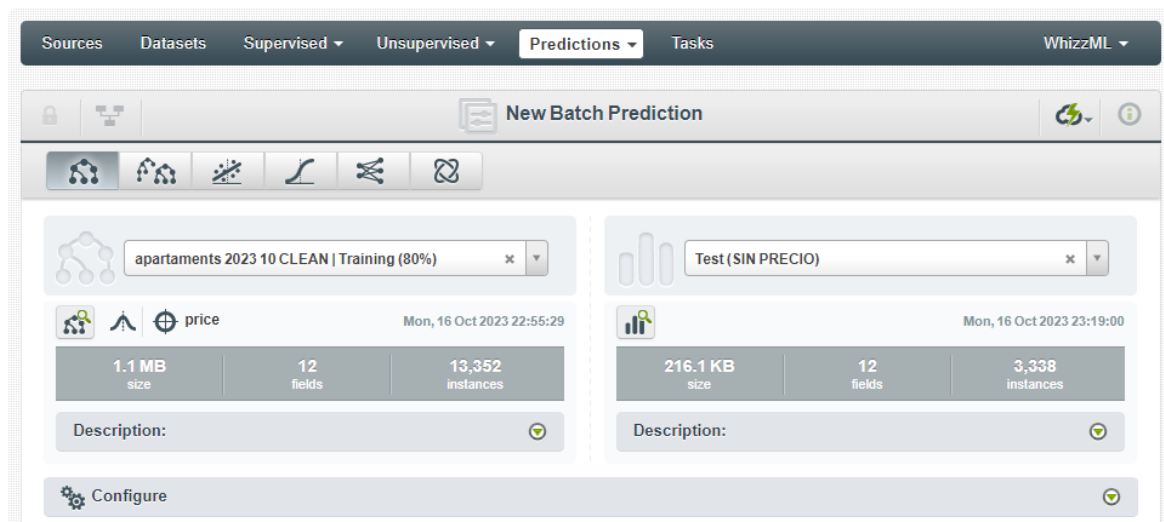
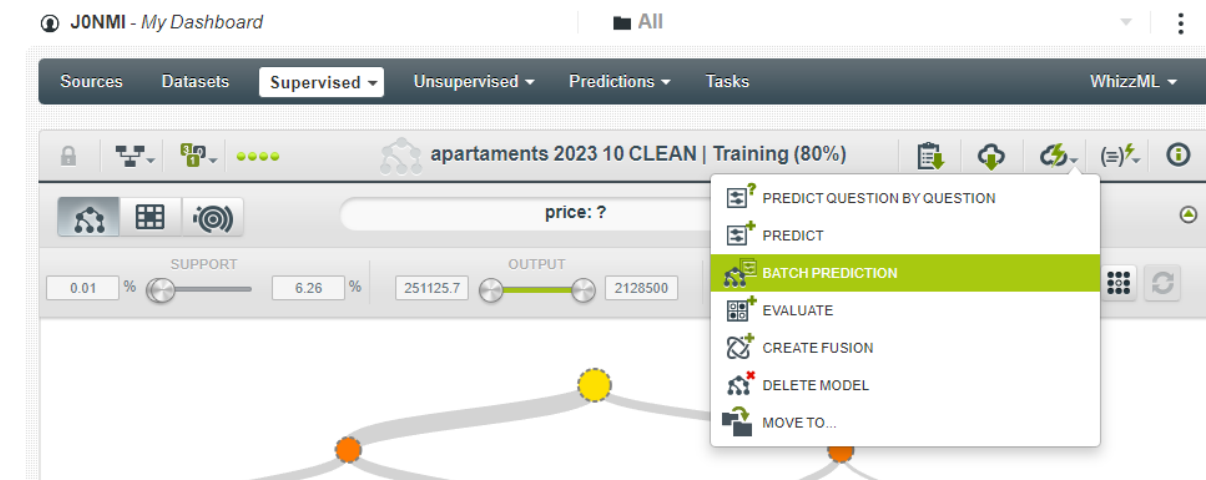
Subimos el csv que se nos ha descargado y listo.

JONMI - My Dashboard | All

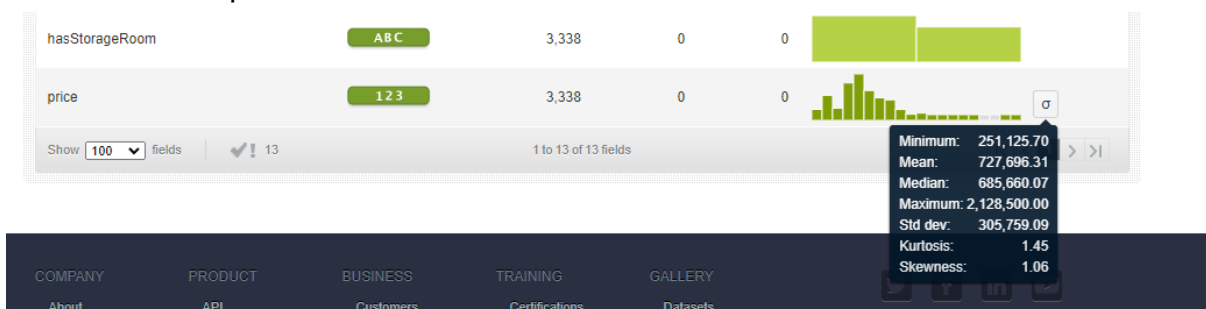
Sources		Datasets	Supervised	Unsupervised	Predictions	Tasks	WhizzML
Name							
Test (SIN PRECIO)	In progress: 50%	0min	216 K	0	0	0	0
TRAINING	13352 instances, 13 fields (6 categorical, 7 num...)	32min	1 M	1	0	0	0

## 7-. Predecir los valores de la variable dependiente de los datos de prueba.

Creamos un modelo con los datos de Training y creamos un "Batch Prediction".

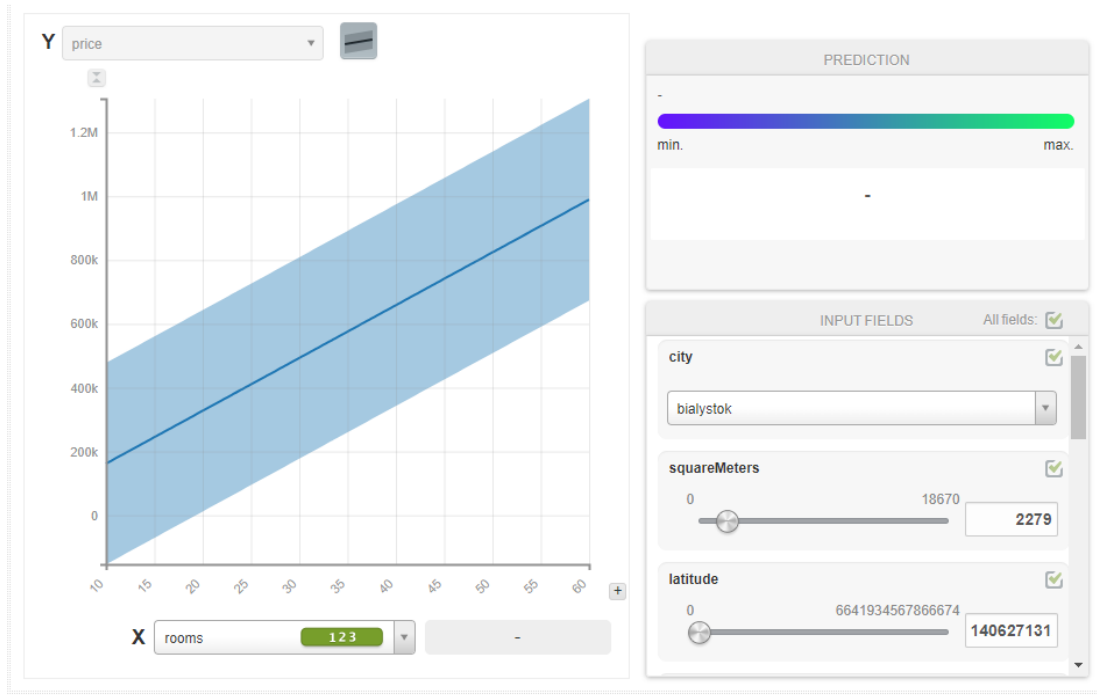


Ya tendríamos la predicción.

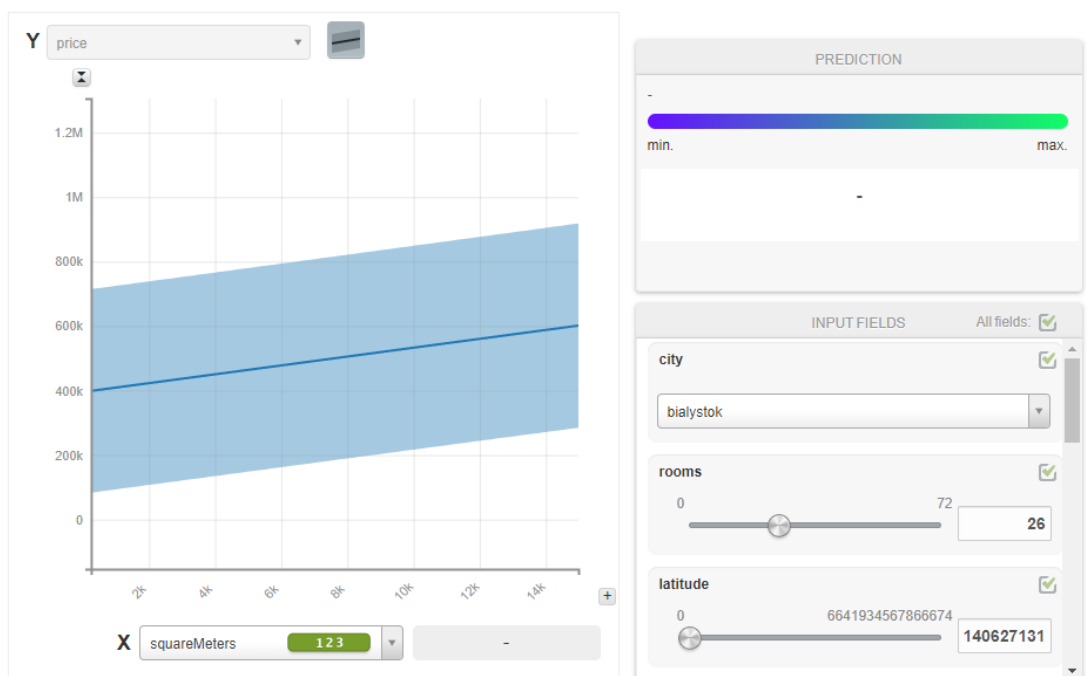


**8-. Descargar la gráfica de los datos de prueba en formato png y describir un par de ellos.**

Como vemos la gráfica del precio respecto al número de habitaciones es lineal ascendente, algo que es completamente lógico. Cuantas más habitaciones, mayor será el precio.



Lo mismo sucede si hacemos una gráfica, cambiando el **EJE X: Metros cuadrados**



Por último, por sacar otra conclusión, si vemos el precio respecto a la distancia a la que está la vivienda del centro, cuanto más lejos, más barata.

