

ANALÍTICA DE NEGOCIO

Análisis del precio de la vivienda en España

Lucía Cordero y Jorge Garcelán

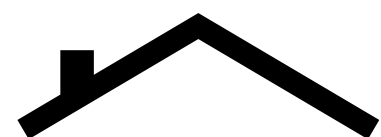


pesimista

01

pesimista

RESPONSABILIDAD AUTORES



ANALÍTICA DE NEGOCIO

**Análisis del
precio de la
vivienda en
España**



Responsabilidad



01

Brainstorming

02

Conjunto de datos

03

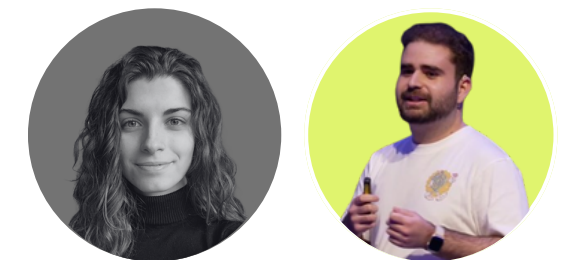
Visualización de datos

04

Método paramétrico

05

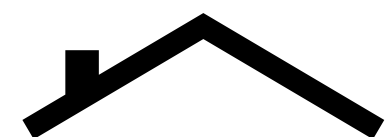
Método no paramétrico



02

pesimista

CONJUNTO DE DATOS



ANALÍTICA DE NEGOCIO

**Análisis del
precio de la
vivienda en
España**



Conjunto de datos

02

idealista

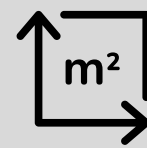
Anuncios publicados
en Idealista para toda
España



2018

Fuente: [Idealista](#)

Ejemplos de variables



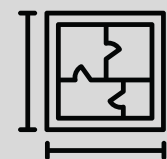
Metros
cuadrados



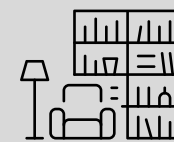
Contiene
Jardín



Garaje
incluido



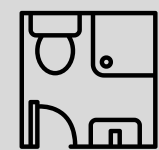
Número de
habitaciones



Está
amueblado



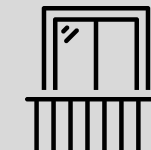
Tipo de
inmueble



Número de
baños



Contiene
piscina



Contiene
terrazza

Variable objetivo

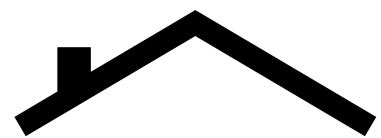


Precio del
inmueble

03

pesimista

VISUALIZACIÓN DE DATOS

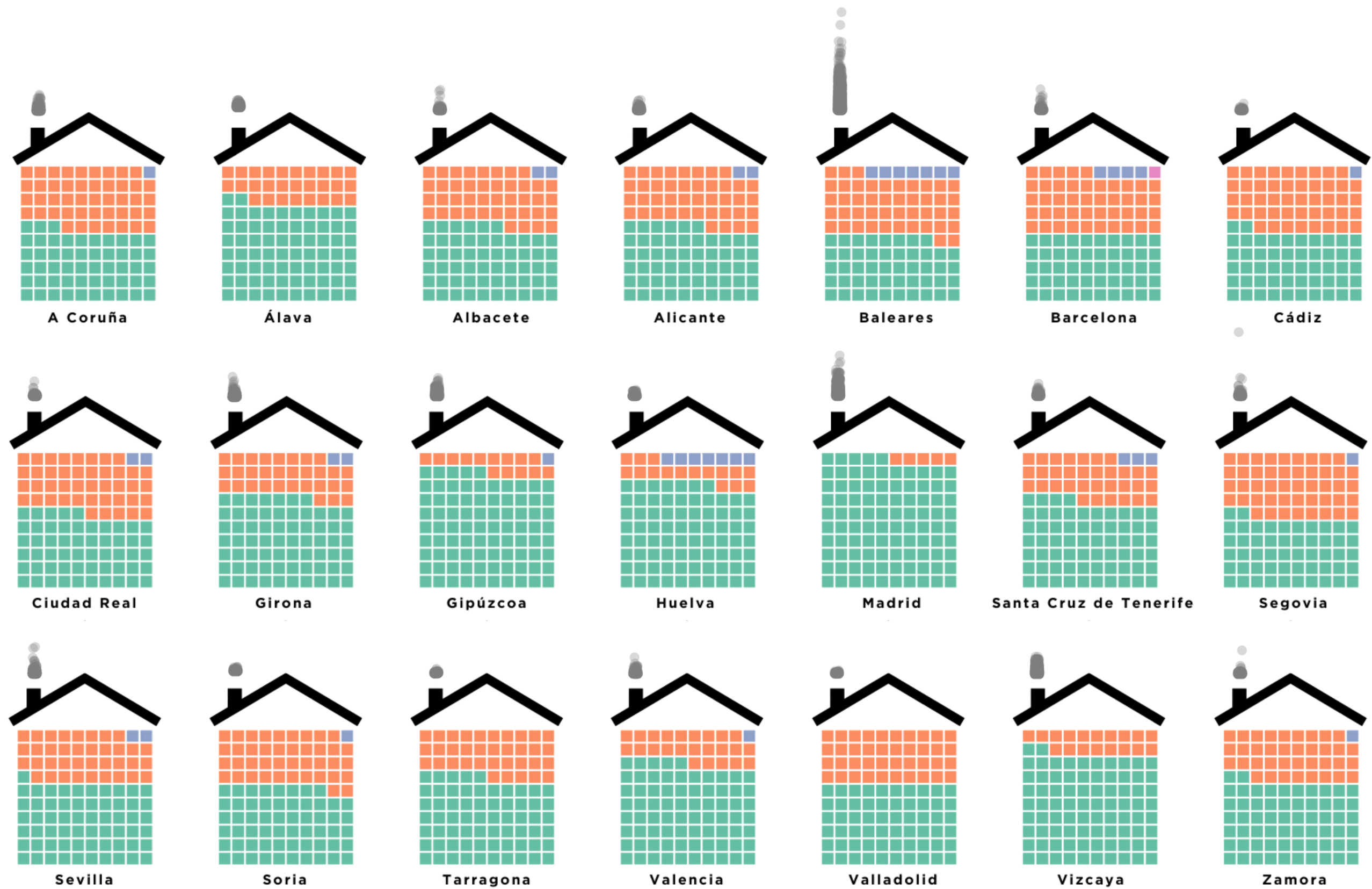


ANALÍTICA DE NEGOCIO

**Análisis del
precio de la
vivienda en
España**

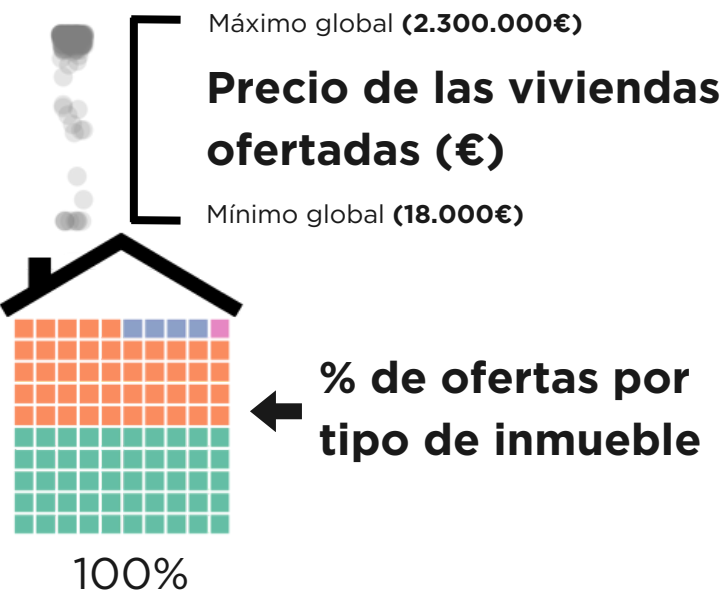
Visualización de datos

Distribución del tipo de inmueble ofertado por provincia en 2018



Tipo de inmueble

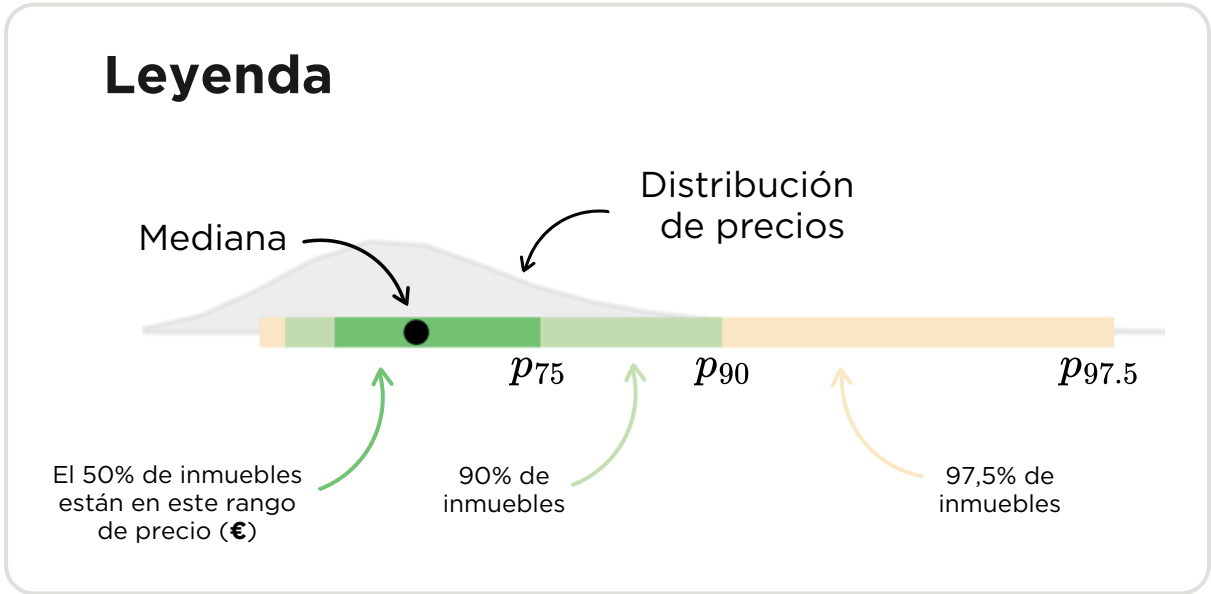
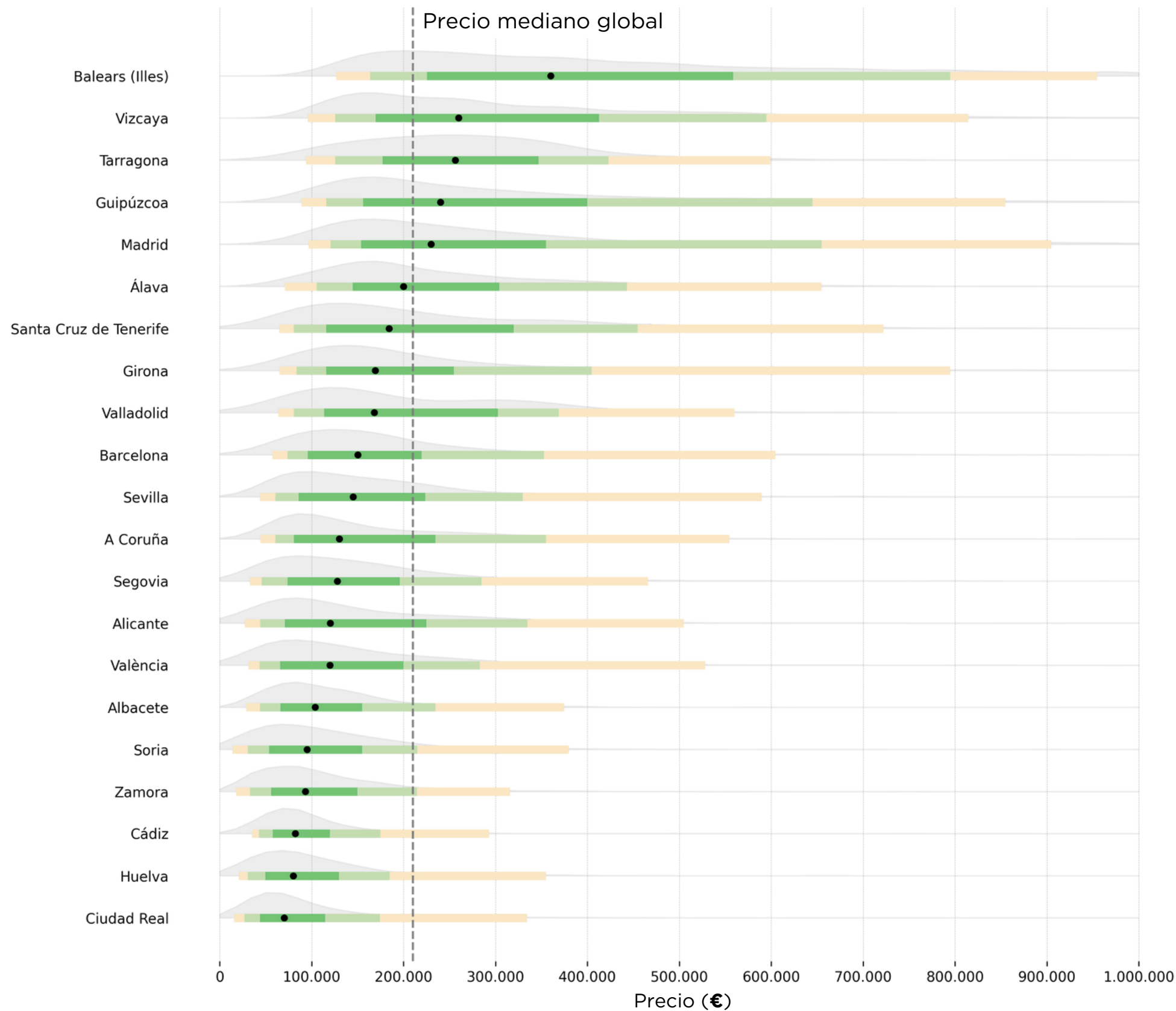
- Piso / Ático
- Casa / Chalet
- Casa rural / rústica
- Vivienda singular



Fuente: [Idealista](#)

Visualización de datos

Distribución del viviendas por precio y provincia en 2018

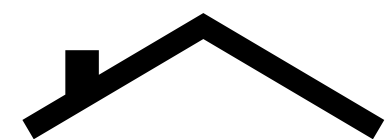


Fuente: [Idealista](#)

04

pesimista

PREDICCIÓN



ANALÍTICA DE NEGOCIO

**Análisis del
precio de la
vivienda en
España**



Pipeline de los modelos



01

Preprocesamiento

- Eliminamos outliers
- Generamos variables dummy
- Recategorizar las variables (p.e. $\text{floor} \in \{1, 2, 3, 4, 5+\}$)
- Tratamiento de observaciones nulas (NAs)

02

Matriz de correlación

- `air conditioner` está positivamente correlacionado con `Islas Baleares`
- `num_bath` está positivamente correlacionado con `num_rooms`

03

Selección de variables

- Evaluamos la relación individual de cada variable con la variable objetivo (precio)
- Aplicamos `forward selection`

04

Selección de modelos

- Método paramétrico: `Linear Regression`
- Método no paramétrico: `Random Forest Regressor`

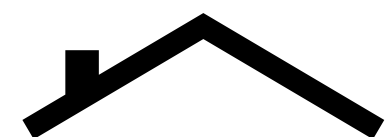
Evaluación de resultados

	Escenario	R^2	RMSE	Nº Variables
Método paramétrico	Modelo con todas las variables	0.5913	116586.81	39
	Provincias + tipo de inmueble + variables estructurales (núm. habitaciones, núm. baños, terraza, etc.)	0.5702	116786.42	33
	Top 5 variables más contribuidoras + provincias	0.5492	119556.91	25
Método no paramétrico	Modelo con todas las variables	0.7149	95080.40	39
	Todas las variables menos top 5 menos contribuidoras	0.6940	98495.38	35
	Provincias + tipo de inmueble + variables estructurales (núm. habitaciones, núm. baños, terraza, etc.)	0.6819	100422.96	33

05

pesimista

CONCLUSIONES



ANALÍTICA DE NEGOCIO

**Análisis del
precio de la
vivienda en
España**



Conclusiones

05

- 1 En todo el territorio la distribución de precios está sesgada a la derecha, con una cola más larga en los precios altos.
- 2 Islas Baleares es el territorio con precios de la vivienda más variables.
- 3 En el dataset, los datos de Barcelona están sesgados con respecto a la situación real de precios en la zona.
- 4 Provincias como Álava, Soria y Zamora (rurales) tienen una mayor proporción de casas/chalets respecto a pisos.