



# UT.1 Introducción

IES Abastos

CE Inteligencia Artificial y Big Data

Sistemas de Aprendizaje Automático.

# Bloques de la UT1:



1. Aprendizaje automático.
2. Aprendizaje automático supervisado.
3. Aprendizaje automático no supervisado.
4. Tipos de problemas aprendizaje automático.
5. Conjuntos de entrenamiento.
6. Fundamentos básicos estadísticos.

## 6. Fundamentos básicos estadísticos.

¿ Qué cosas raras os llama la atención del dataset ?

A	B	C	D	E	F	G	H	I
	provincia	comunidad autor	titulo	precio	habitaciones	metros	total inmuebles/comunidad	
0	La Coruña	Galicia					1751	
1	La Coruña	Galicia	Chalet pareado en A Zap	1100	4	243	1751	
2	La Coruña	Galicia	Chalet pareado en Elmña	1100	4	243	1751	
3	La Coruña	Galicia	Ático en avenida Pasaxe	450		Planta 4ª exterior con asce	1751	
4	La Coruña	Galicia	Piso en calle de Fernanc	1000	4	100	1751	
5	La Coruña	Galicia	Piso en calle Manuel Jes	450	2	76	1751	
6	La Coruña	Galicia	Piso en Os Mallos, A Co	800	4	100	1751	
7	La Coruña	Galicia	Piso en ronda de Outeiro	650	3	122	1751	
8	La Coruña	Galicia	Piso en travesía de Meic	400	4	107	1751	
9	La Coruña	Galicia	Piso en calle Concepción	800	2	75	1751	
10	La Coruña	Galicia	Piso en calle da Pía, Ca	450	3	99	1751	
11	La Coruña	Galicia	Piso en ronda de nelle, 1	600	3	87	1751	
12	La Coruña	Galicia	Piso en Somo - Matog	825	3	80	1751	
13	La Coruña	Galicia	Piso en Juan Flórez-San	700	2	80	1751	
14	La Coruña	Galicia	Ático en Editor Francisco	245	1	46	1751	
15	La Coruña	Galicia	Chalet pareado en calle	550	4	90	1751	
16	La Coruña	Galicia	Piso en avenida de Villag	750	3	95	1751	
17	La Coruña	Galicia	Dúplex en calle Garrucha	650	2	90	1751	
18	La Coruña	Galicia	Piso en calle Rosalía de	360	2	65	1751	
19	La Coruña	Galicia	Piso en plaza de la Blas	850	2	110	1751	

## 6. Fundamentos básicos estadísticos.

	A	B	C	D	E	F	G	H	I
1725	1723	La Coruña	Galicia	Dúplex en plaza de Salu	2000	2	Planta 3ª exterior con asce	1751	
1726	1724	La Coruña	Galicia	Piso en calle Arquitecto	800	5		100	1751
1727	1725	La Coruña	Galicia	Piso en Salgueiros de	418	1		50	1751
1728	1726	La Coruña	Galicia	Piso en calle Panasquei	370	2		60	1751
1729	1727	La Coruña	Galicia	Piso en Escorial, 2. Rbe	1200	4		140	1751
1730	1728	La Coruña	Galicia	Piso en calle Antonio VII	615	3		95	1751
1731	1729	La Coruña	Galicia	Piso en ronda de Nella	470	4		115	1751
1732	1730	La Coruña	Galicia	Piso en calle Agri do Mi	485	2		80	1751
1733	1731	La Coruña	Galicia	Piso en carretera Baños	850	3		100	1751
1734	1732	La Coruña	Galicia	Piso en calle Frouxeiras	500	2		64	1751
1735	1733	La Coruña	Galicia	Piso en avenida cristina	350	2		70	1751
1736	1734	La Coruña	Galicia	Piso en Lugar Vixoi, Bar	330	2		60	1751
1737	1735	La Coruña	Galicia	Piso en calle das Oblata	550	2		56	1751
1738	1736	La Coruña	Galicia	Piso en calle Alameda, 4	630	3		72	1751
1739	1737	La Coruña	Galicia	Piso en Xosé Ramón Ba	475	3		105	1751
1740	1738	La Coruña	Galicia	Piso en travesía Praia, 1	1150	1		47	1751
1741	1739	La Coruña	Galicia	Piso en Ciudad Vieja - C	390	1		35	1751
1742	1740	La Coruña	Galicia	Piso en capitán Juan var	700	3		80	1751
1743	1741	La Coruña	Galicia	Piso en calle San Pedro	700	3		100	1751
1744	1742	La Coruña	Galicia	Piso en calle Vázquez d	550	2		78	1751
1745	1743	La Coruña	Galicia	Piso en RÁªa Fanny Gar	650	2		59	1751
1746	1744	La Coruña	Galicia	Dúplex en calle Concepc	680	4		154	1751
1747	1745	La Coruña	Galicia	Piso en calle de San Pe	1100	4		114	1751
1748	1746	La Coruña	Galicia	Piso en Os Matos, A Ce	400	2		80	1751
1749	1747	La Coruña	Galicia	Piso en puerta del Cami	1000	4		200	1751
1750	1748	La Coruña	Galicia	Piso en calle Estanque	590	2		75	1751
1751	1749	La Coruña	Galicia	Piso en Monte Alto - Zali	850	3		110	1751
1752	1750	La Coruña	Galicia	Piso en Doctor Teixeiro,	1200	2		107	1751
1753	1751	La Coruña	Galicia	Piso en O Milladoiro, An	420	2		75	1751
1754	1752	Alava	País Vasco	Piso en calle Sol, 7. Moi	580	1		65	138
1755	1753	Alava	País Vasco	Dúplex en plaza de Salu	2000	2	Planta 3ª exterior con asce	1751	

## 6. Fundamentos básicos estadísticos.



Para solucionar estos problemas que harían que nuestro modelo no fuera correcto, tenemos primero que estar familiarizado con unos conceptos básicos de estadística.

- Media.
- Moda.
- Mediana.

## 6. Fundamentos básicos estadísticos.



La **media aritmética** es el valor que se obtiene al sumar todos los datos que tenemos y dividir el resultado entre el número total de esos datos.

Ejemplo: Calcular la media de: 8,9,10,11,16,17,6

$$\bar{X} = \frac{8 + 9 + 10 + 11 + 16 + 17 + 6}{7} = \frac{77}{7} = 11$$

## 6. Fundamentos básicos estadísticos.



**Media móvil:** Muy parecida a la aritmética salvo que capturamos sólo un subconjunto de los datos para hacer la media. Gracias a las medias móviles se pueden crear líneas de tendencia.

Ejemplo: Calcular la media móvil 5: 8,9,10,11,16,17,6,1,2,3,4,6,8,9

Capturaremos los cinco últimos datos y hallaremos la media.

$$(3+4+6+8+9)/5=30/5= 6$$

## 6. Fundamentos básicos estadísticos.



La **mediana** es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor.

Ejemplo: 8,9,10,11,16,17,6

- 1) Ordenamos: 6,8,9,**10**,11,16,17
- 2) Obtenemos el valor central: 10



## 6. Fundamentos básicos estadísticos.



Se define la **moda** como el número que está representado más veces dentro de esos datos, es decir, aquel número que presenta una mayor frecuencia absoluta dentro de la muestra.

Ejemplo: **8**,9,10,11,**8**,16,17,6,4,**8**,22,2

La moda sería el 8.

## 6. Fundamentos básicos estadísticos.



La **media** se utiliza con variables cuantitativas y tiene utilidad para calcular por ejemplo, la nota promedio de una materia del curso de especialización, mientras que la **mediana** se hace muy útil para evaluar la verdadera tendencia central de un conjunto de datos en donde hay muchos "outliers" o datos atípicos que alterarían mucho la media.

## 6. Fundamentos básicos estadísticos.



La **Moda**, es una medida muy útil para evaluar la tendencia central de variables categóricas, es decir nos permite observar en donde se concentran los datos en las variables que están divididas en niveles.

El cálculo de la moda también puede ser útil en caso de que se utilicen variables "no numéricas" o que se están realizando análisis de documentos de texto.

## 6. Fundamentos básicos estadísticos.

¿ La **media** en qué tipo de aprendizaje es ideal para rellenar huecos del dataset?



## 6. Fundamentos básicos estadísticos.

¿ La **moda** en qué tipo de aprendizaje es ideal para rellenar huecos del dataset?

