

Informe DAVD

El problema habitacional en Madrid

Examen Final

Rim TYOUSS

El objetivo de nuestro análisis es ver que variables afectan el problema habitacional en Madrid, en particular la subida de precios por metro cuadrado y por qué y en función de que factores tiene lugar.

Análisis descriptivo

El dataset se compone de 18 variables. Primero comprobamos si hay duplicados, o variables faltantes (NAs).

Para realizar este análisis, nuestra variable “target” sería el precio del metro cuadrado: m2_price. El objetivo es ver que variables afectan más a la subida de precios en la comunidad de Madrid.

Empezamos por ver estadísticas generales:

Existen 18 distritos en Madrid. El precio medio por distrito es el siguiente:

neighbourhood_group	
Arganzuela	13.823507
Barajas	11.194508
Carabanchel	10.881736
Centro	16.231677
Chamartín	15.048115
Chamberí	16.279343
Ciudad Lineal	12.453770
Hortaleza	11.935000
Latina	10.803206
Moratalaz	10.579722
Puente de Vallecas	11.004237
Retiro	14.442734
Salamanca	16.915515
Tetuán	14.246279
Usera	10.917080
Vicálvaro	10.347778
Villa de Vallecas	11.041978
Villaverde	10.023661

Para poder hacer un análisis comparativo, tenemos que comprobar que todos los distritos tengan más o menos el mismo número de observaciones.

neighbourhood_group	
Centro	155
Retiro	139
Hortaleza	138
Chamberí	137
Salamanca	136
Arganzuela	134
Latina	131
Tetuán	129
Ciudad Lineal	122
Barajas	122
Chamartín	122
Carabanchel	121
Puente de Vallecas	118
Usera	113
Villaverde	112
Moratalaz	108
Villa de Vallecas	91
Vicálvaro	90
Name: count, dtype: int64	

Aquí vemos que deberíamos tener más a la hora de analizar Vicálvaro y Villa de Vallecas.

Empezamos por ver el precio del metro cuadrado por distrito, y como podemos ver en la figura siguiente, de 2010 a 2024 tenemos una tendencia de crecimiento de precios, independientemente del distrito. Podemos ver también a partir de esta grafica que los distritos con precios más altos son el distrito de Salamanca, el Centro y Chamberí. Los distritos con precios más bajos son Arganzuela, Vicálvaro, Villa de Vallecas, Ciudad Lineal y Moratalaz.

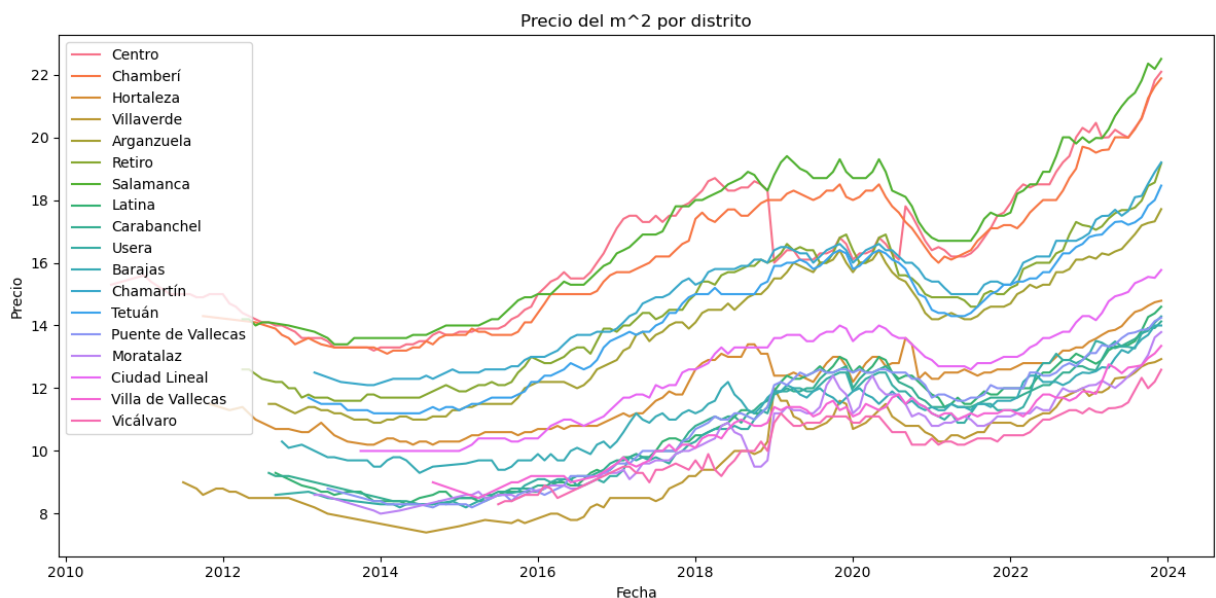


Figura 1: Precio del m² por distrito

Otra variable que tenemos que tener en cuenta es la inflación, ya si los precios suben acorde a la inflación, no podemos concluir que los precios por metro cuadrado suben realmente.

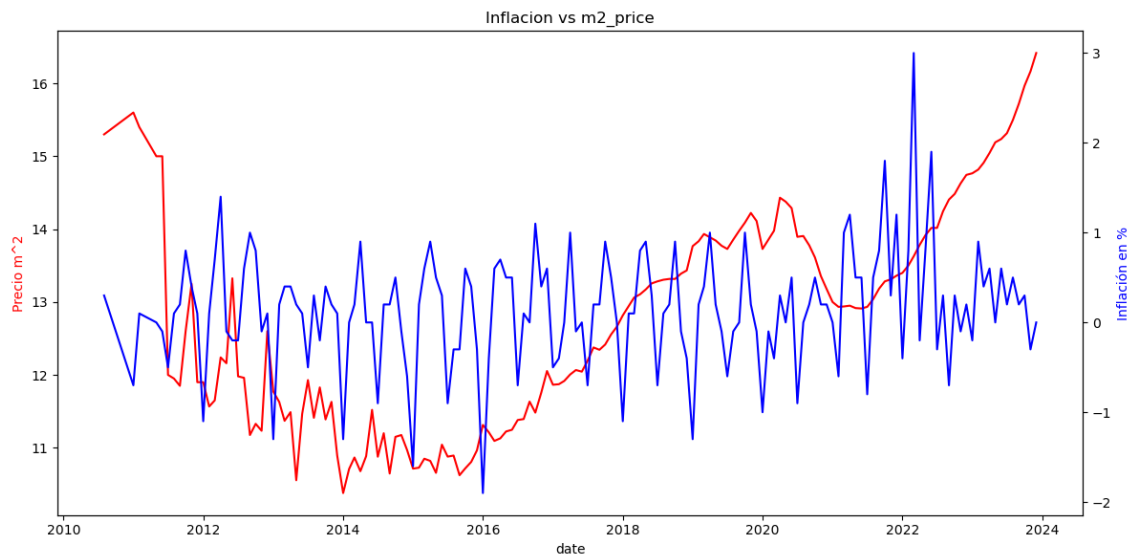


Figura 2: Inflación y precio por metro cuadrado

Sin embargo, podemos ver en la figura que el precio por metro cuadrado no sigue tanto la línea de inflación: de 2012 a 2016, después de la crisis, los precios fueron mas bajos que la inflación. Después hasta día de hoy, han tenido un aumento significativo, más que el porcentaje de inflación.

Con un mapa de calor, podemos ver más que variables tienen más relación con la subida de precios en Madrid y centrarnos más en ellas:

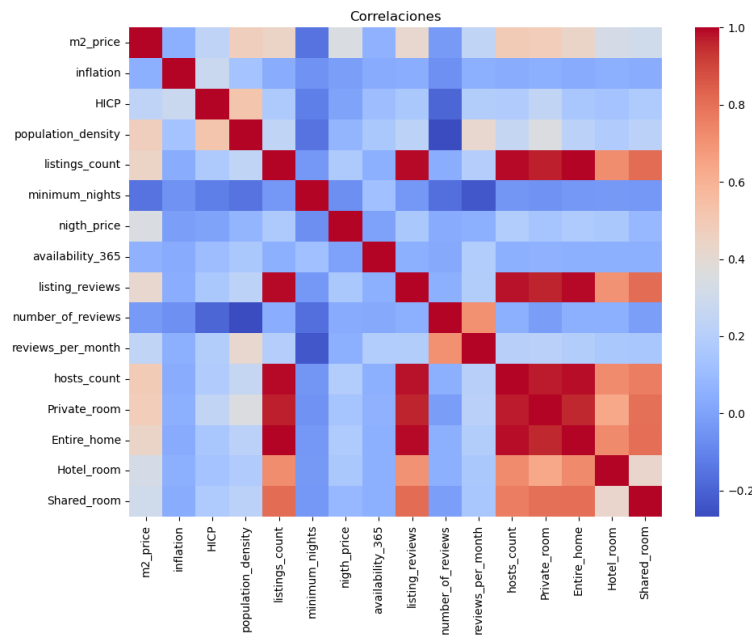


Figura 3: Mapa de calor

Podemos ver que las variables con una correlación fuerte (> 0.5) con el precio por metro cuadrado son:

- La densidad de población (population_density)
- El número total de propiedades listadas en Airbnb (listings_count)
- El número de anfitriones activos en un vecindario (host_counts)

Comparamos entonces el número total de propiedades listadas en el tiempo con el precio por metro cuadrado:

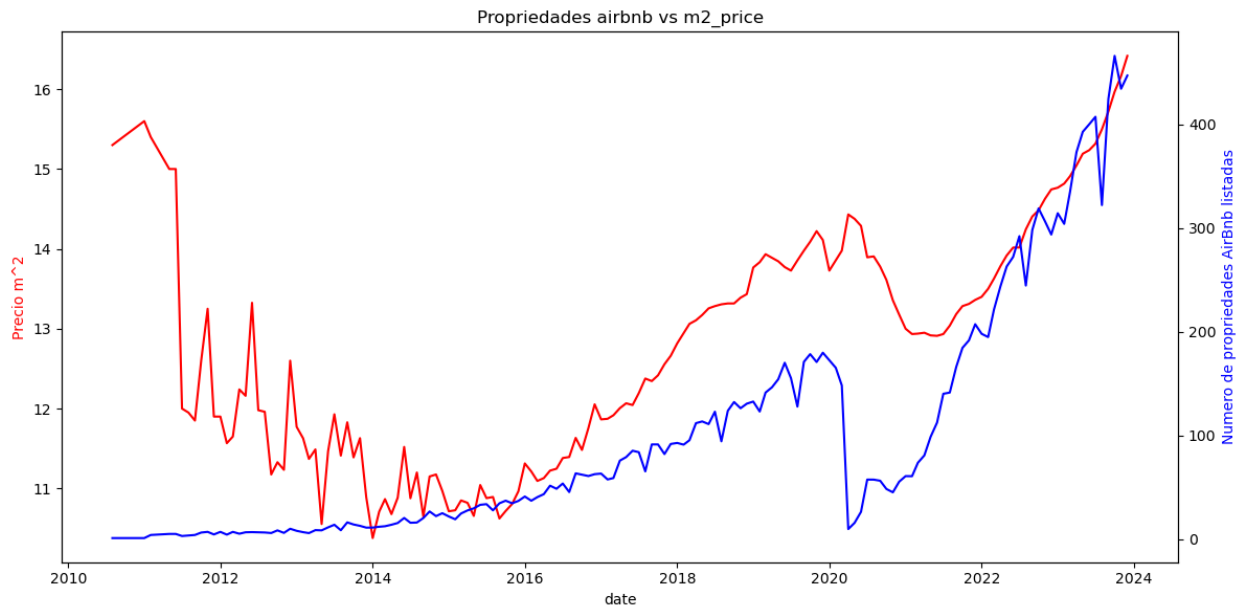


Figura 4: Propiedades listadas en Airbnb y precio por metro cuadrado

Efectivamente, podemos ver que, a partir de 2014, las dos líneas siguen casi la misma tendencia: cuando sube el número de propiedades listadas en Airbnb, sube el precio por metro cuadrado.

Finalmente, a mayor densidad de población, mayor será la demanda de vivienda, servicios y comercios, y se puede traducir en precios más altos. Lo comprobamos con la siguiente gráfica:

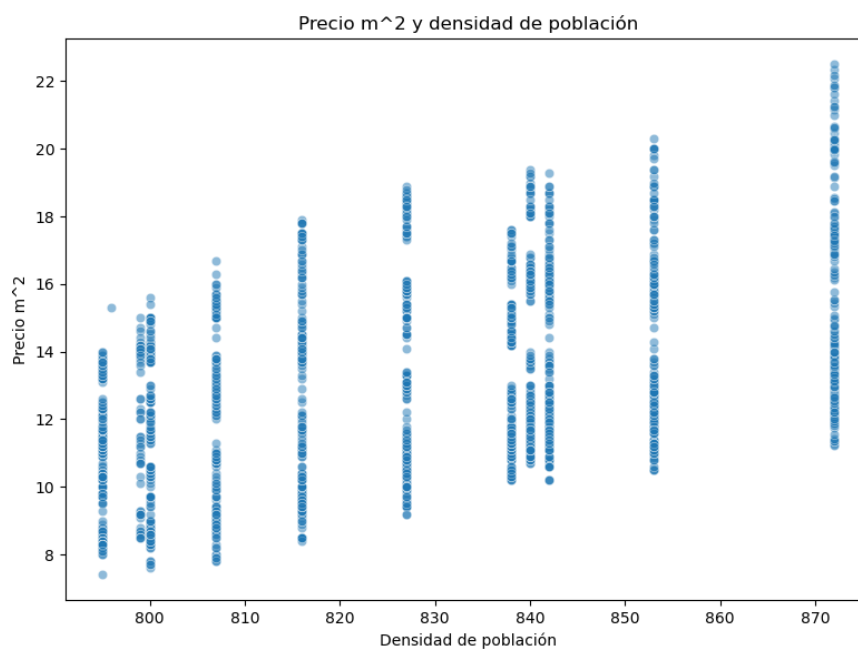


Figura 5: Densidad de población y precio por metro cuadrado

Finalmente, a mayor número de anfitriones, podemos ver una clara relación con el precio del metro cuadrado, que aumenta. Sin embargo, parece que esta relación afecta más a los distritos del “centro”: Centro, Salamanca, Chamberí, Tetuán, Retiro y no tanto en los que están más alejados.

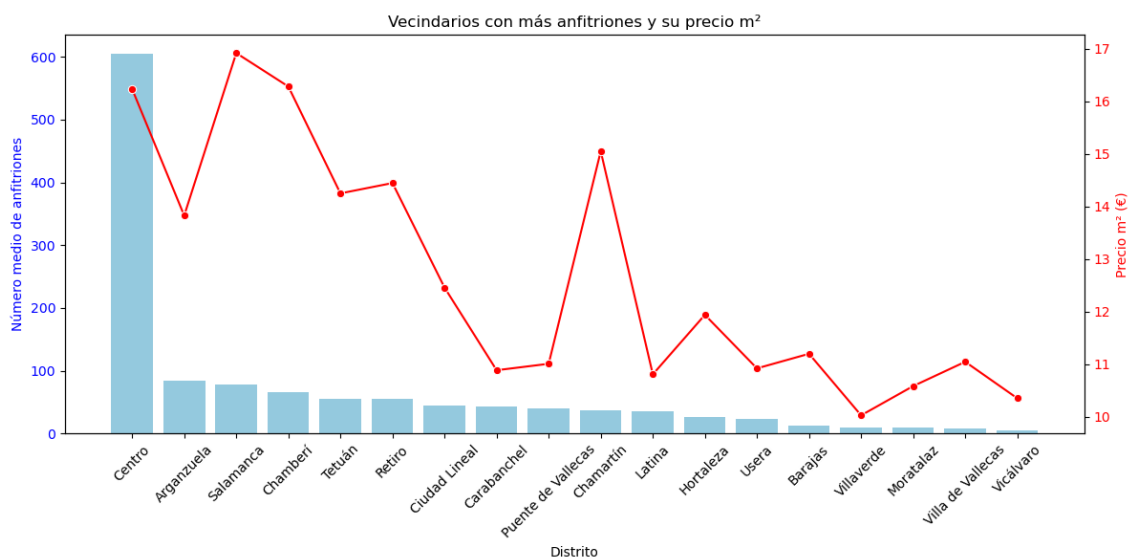


Figura 6: Anfitriones y precio por m²

Para terminar, comparo el precio por noche de las propiedades Airbnb con el precio del metro cuadrado para ver si tienen alguna relación.

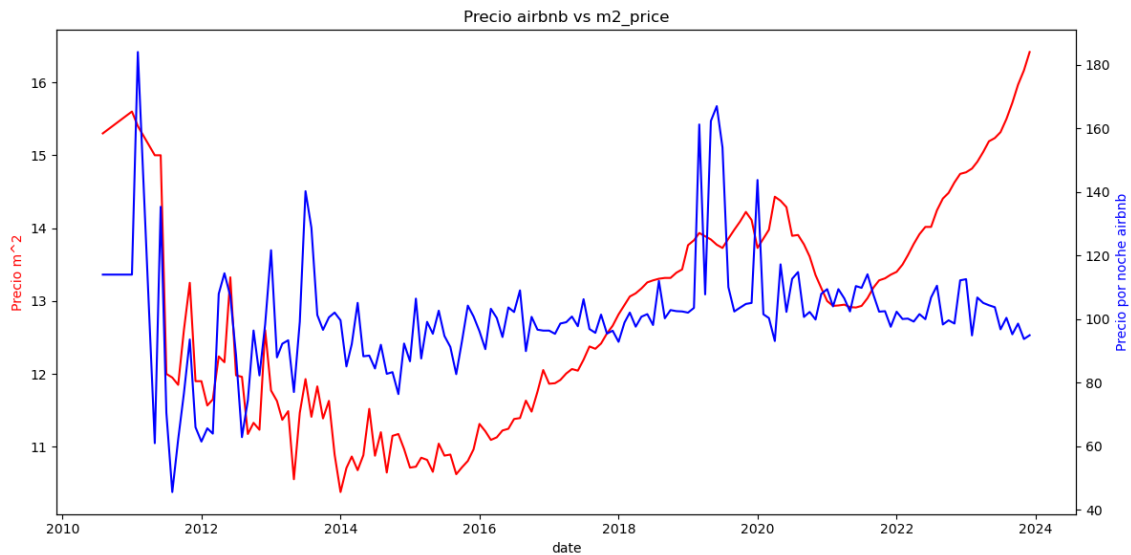


Figura 7: Precio propiedades Airbnb y precio m^2

No parece que el precio de los Airbnb tenga tanta relación con el precio por metro cuadrado. Podemos ver algunas subidas puntuales conjuntas, pero no se comportan de la misma manera. Vemos que el precio por metro cuadrado a partir de 2022 se dispara, comparado con el precio por noche de los Airbnb. Sin embargo, creo que si tendrían la misma tendencia en ciertos distritos en particular: el centro, salamanca, Chamberí, y en el Retiro.

Efectivamente, podemos ver que el precio por metro cuadrado sigue el precio por noche de los Airbnb de manera mucho más clara cuando el distrito es un distrito más “céntrico”: el Centro, Chamberí, Salamanca.

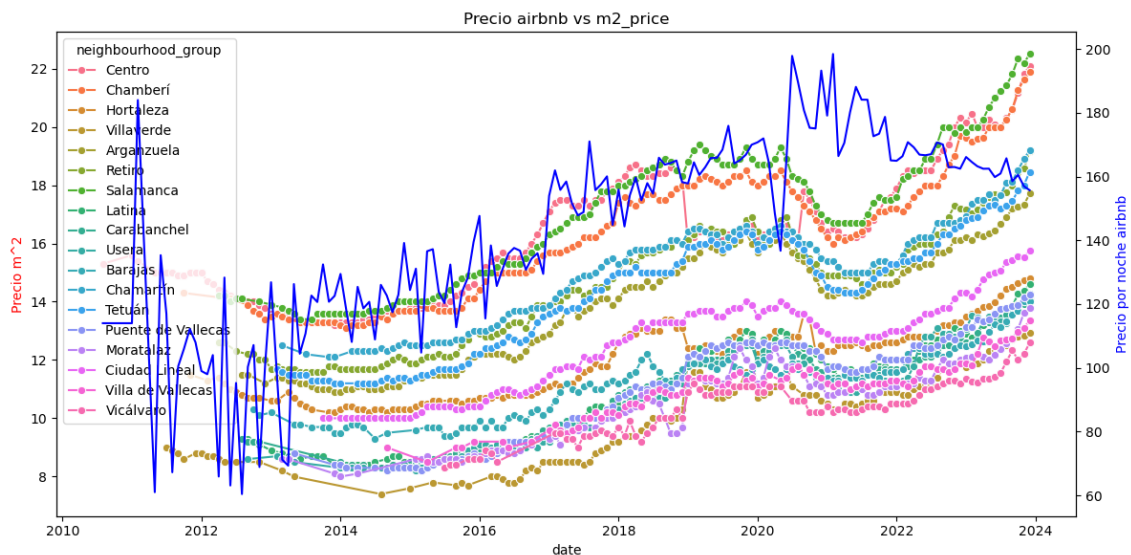


Figura 8: Precio propiedades Airbnb y precio m^2 por distrito

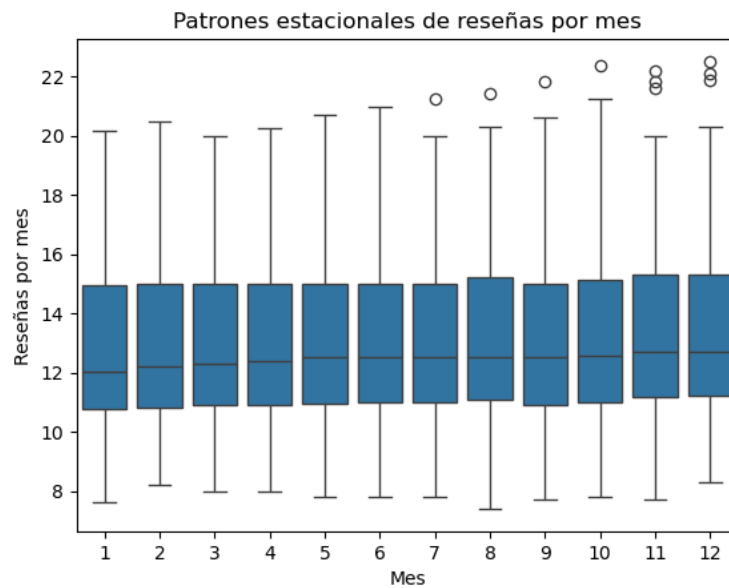


Figura 9: Estacionalidad de los precios

Los precios tienen una mediana bastante parecida, con bastantes outliers en épocas festivas en finales de año. a

Hasta ahora, mi hipótesis es la siguiente: las variables que más afectan a los precios de las viviendas son la densidad de población del distrito, el número de propiedades listadas, la inflación, y la densidad de población.

2. Regresión lineal

Probamos un modelo de regresión lineal con todas las variables que tenemos en el conjunto de datos:

OLS Regression Results						
Dep. Variable:	m2_price	R-squared:	0.921			
Model:	OLS	Adj. R-squared:	0.911			
Method:	Least Squares	F-statistic:	90.68			
Date:	Thu, 19 Dec 2024	Prob (F-statistic):	2.84e-53			
Time:	15:55:04	Log-Likelihood:	-69.855			
No. Observations:	124	AIC:	169.7			
Df Residuals:	109	BIC:	212.0			
Df Model:	14					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-32.3165	5.623	-5.747	0.000	-43.461	-21.172
inflation	-0.0116	0.065	-0.177	0.859	-0.141	0.118
HICP	0.0137	0.027	0.505	0.614	-0.040	0.067
population_density	0.0520	0.007	7.076	0.000	0.037	0.067
listings_count	-0.0033	0.004	-0.875	0.383	-0.011	0.004
minimum_nights	-0.0047	0.009	-0.507	0.613	-0.023	0.014
nigth_price	0.0164	0.003	6.424	0.000	0.011	0.021
availability_365	0.0091	0.001	6.749	0.000	0.006	0.012
listing_reviews	8.371e-05	0.001	0.086	0.932	-0.002	0.002
number_of_reviews	-0.0013	9.68e-05	-13.822	0.000	-0.002	-0.001
reviews_per_month	1.4680	0.264	5.556	0.000	0.944	1.992

Sin embargo, como podemos ver en las gráficas anteriores, tenemos mucho ruido en las gráficas antes del 2014. Podemos relacionar ese ruido por las crisis en España, pero también porque Airbnb todavía no tenía tanta fama ni estaba tan usado (Figura 4). Hacemos una regresión lineal con fechas posteriores a 2014:

OLS Regression Results						
=====						
Dep. Variable:	m2_price	R-squared:	0.976			
Model:	OLS	Adj. R-squared:	0.972			
Method:	Least Squares	F-statistic:	237.6			
Date:	Thu, 19 Dec 2024	Prob (F-statistic):	1.47e-59			
Time:	16:33:08	Log-Likelihood:	4.2746			
No. Observations:	96	AIC:	21.45			
Df Residuals:	81	BIC:	59.92			
Df Model:	14					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-17.3558	4.135	-4.197	0.000	-25.583	-9.128
inflation	-0.0251	0.041	-0.610	0.544	-0.107	0.057
HICP	-0.0007	0.017	-0.039	0.969	-0.035	0.033
population_density	0.0335	0.006	6.041	0.000	0.022	0.045
listings_count	-0.0057	0.002	-2.801	0.006	-0.010	-0.002
minimum_nights	0.0021	0.006	0.333	0.740	-0.010	0.014
nigth_price	0.0004	0.003	0.108	0.915	-0.007	0.007
availability_365	0.0036	0.002	1.562	0.122	-0.001	0.008
listing_reviews	0.0012	0.001	2.135	0.036	8.18e-05	0.002
number_of_reviews	-0.0010	0.000	-5.794	0.000	-0.001	-0.001
reviews_per_month	2.2203	0.273	8.143	0.000	1.678	2.763
hosts_count	0.0049	0.001	7.803	0.000	0.004	0.006
Private_room	0.0027	0.002	1.179	0.242	-0.002	0.007
Entire_home	0.0026	0.002	1.217	0.227	-0.002	0.007
Hotel_room	-0.0081	0.005	-1.475	0.144	-0.019	0.003
Shared_room	-0.0029	0.006	-0.494	0.623	-0.015	0.009

Efectivamente, nuestro modelo explica mucho más el precio por metro cuadrado que antes, con R^2 ajustado de 97.2%. Quitamos las variables que no tienen mucha significancia:

- La inflación,

- El HICP,
- El mínimo de noches,
- La disponibilidad
- Si la habitación es privada, o hotel o compartida.

Finalmente nos quedamos con este modelo:

OLS Regression Results

Dep. Variable:	m2_price	R-squared:	0.974
Model:	OLS	Adj. R-squared:	0.972
Method:	Least Squares	F-statistic:	467.7
Date:	Thu, 19 Dec 2024	Prob (F-statistic):	9.71e-67
Time:	16:51:38	Log-Likelihood:	-0.35468
No. Observations:	96	AIC:	16.71
Df Residuals:	88	BIC:	37.22
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-15.3269	3.759	-4.077	0.000	-22.797	-7.857
population_density	0.0321	0.005	6.369	0.000	0.022	0.042
listings_count	-0.0032	0.000	-10.733	0.000	-0.004	-0.003
nigth_price	-0.0016	0.003	-0.494	0.623	-0.008	0.005
listing_reviews	0.0011	0.000	2.431	0.017	0.000	0.002
number_of_reviews	-0.0011	0.000	-7.198	0.000	-0.001	-0.001
reviews_per_month	2.2733	0.238	9.538	0.000	1.800	2.747
hosts_count	0.0051	0.000	11.100	0.000	0.004	0.006

Omnibus: 10.805 Durbin-Watson: 2.042

Prob(Omnibus): 0.005 Jarque-Bera (JB): 19.716

...

Notes:

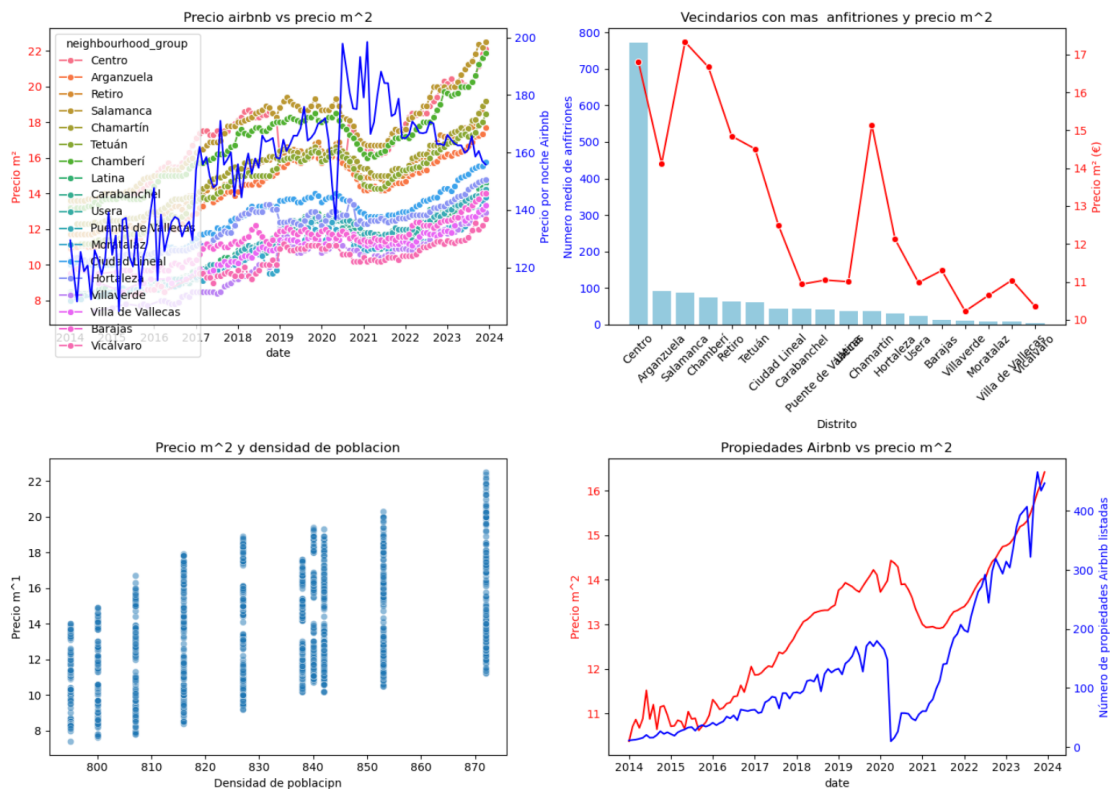
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 6.34e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

Explica lo mismo, pero es más sencillo. Además, al quitar cualquier variable bajaría el R^2 .

En conclusión, el precio de las viviendas en Madrid esta siendo afectado por variables como la densidad de la población, pero también por varios factores acerca de la existencia de alquileres vacacionales. Las principales son : el número de propiedades Airbnbs en la zona, el precio de la propiedad por noche, y .el número de anfitriones. Estas afirmaciones están confirmadas por las graficas que hemos sacado a lo largo de este análisis pero también por el modelo de regresión lineal.



En este dashboard final, resumimos las observaciones más importantes sobre el problema de la vivienda en la comunidad de Madrid:

- El precio del metro cuadrado en Madrid sigue completamente el número de propiedades Airbnb. Se ve muy claramente que cuantas más propiedades vacacionales, más sube el precio por metro cuadrado.
- Además, parece que el precio de los Airbnb, pero sobre todo en los barrios más céntricos, afectan completamente al precio de la vivienda en Madrid, llegando a precios exorbitantes y cada vez más altos.
- A mayor densidad de población, mayor será la demanda de vivienda, servicios y comercios, que se traduce en precios más altos.
- Los vecindarios que tienen más concentración de anfitriones y por ende más propiedades, tienden a tener un mayor precio.

En conclusión, hemos observado que los precios de las viviendas han aumentado en general, pero hay diferencias significativas entre los diferentes distritos de Madrid. Los distritos más afectados, como Salamanca, Chamberí, y el Centro, tienen un estrés habitacional significativamente mayor, con precios que superan mucho a los de distritos periféricos como Vicálvaro o Villa de Vallecas.

Para acabar, se puede decir que las variables vinculadas al alquiler vacacional (listings_count, hosts_count, nighth_price) parecen influir en el precio de la vivienda y ser, **definitivamente, un parte del problema habitacional.**

La comunidad de Madrid podría focalizar medidas de política de vivienda en los barrios con mayor densidad y precios más altos, ya que allí el estrés habitacional es mayor con precios mucho más altos. La monitorización de la oferta de alquiler vacacional, ya que hemos visto que están relacionadas con el incremento del precio por m² debería ser monitoreada por la comunidad de Madrid. Existe la necesidad de acciones políticas y regulatorias específicas para equilibrar el mercado de vivienda en Madrid, por ejemplo, implementado límites al número de propiedades listadas en plataformas como Airbnb en los distritos más céntricos y afectados por el incremento de precios, también incentivar el alquiler a largo plazo frente al alquiler turístico en zonas de alta demanda. Descentralizar la demanda que vemos que se centra mucho en 5 distritos: desarrollo de infraestructura y servicios en distritos periféricos para redistribuir la demanda de vivienda y reducir la presión sobre los distritos céntricos. También sería interesante comparar los precios de las viviendas en Madrid con los de otras ciudades de España y Europa para ver cómo se comparan.