

# MongoDB Queries

Análisis de los precios de vivienda de Madrid - Justin Mitchel Calleja

Máster en Big Data, Data Science, y Business Analytics

6 de marzo de 2025

## Introducción

En este trabajo, se nos asignó la tarea de seleccionar un conjunto de datos y utilizar MongoDB para llevar a cabo un análisis de datos. Para este estudio, he elegido un dataset sobre los precios de la vivienda en Madrid, recopilado el 14 de agosto de 2022. Este conjunto de datos contiene información detallada sobre inmuebles en distintos distritos y barrios de la ciudad, incluyendo variables como el precio, el tamaño en metros cuadrados, el número de habitaciones, la presencia de ascensor o garaje y el tipo de vivienda.

El objetivo de este análisis es identificar patrones de precios en función de las características de las viviendas y su ubicación. Para ello, realicé varias consultas en MongoDB que permiten responder preguntas clave, como la distribución de precios según el número de habitaciones, la relación entre el precio por metro cuadrado y los distritos, el impacto de factores como la presencia de garaje o ascensor en los precios y la comparación entre distintos barrios y distritos.

En este documento, presento 10 consultas que utilicé para extraer información relevante, comparar datos y obtener conclusiones significativas sobre el mercado inmobiliario madrileño.

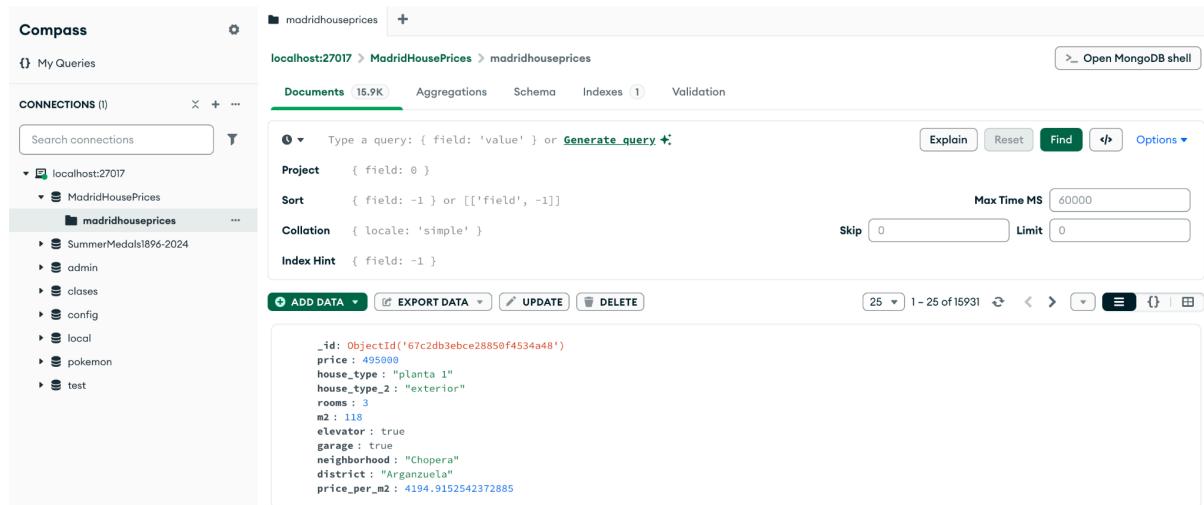
Los campos que componen cada registro en la base de datos son los siguientes:

- **price**: Precio de la vivienda.
- **house\_type**: Tipo de vivienda (ej. planta baja, primer piso).
- **house\_type\_2**: Características adicionales (ej. exterior, interior).
- **rooms**: Número de habitaciones.
- **m2**: Tamaño en metros cuadrados.
- **elevator**: Indica si el edificio tiene ascensor (true/false).
- **garage**: Indica si la vivienda incluye garaje (true/false).
- **neighborhood**: Nombre del barrio.
- **district**: Nombre del distrito.

Con esta información, es posible analizar la variación de precios en Madrid y los factores que pueden influir en ellos.

# 1. Importación de datos:

La base de datos se importó a la aplicación MongoDB Compass para su análisis. Los datos fueron obtenidos del sitio web Kaggle y contienen exactamente 15.931 registros del mercado inmobiliario de Madrid.



The screenshot shows the MongoDB Compass interface. On the left, the 'CONNECTIONS' sidebar lists connections to 'localhost:27017' and the 'MadridHousePrices' database, specifically the 'madridhouseprices' collection which contains 15.9K documents. The main workspace displays a query builder with various filters and sorting options. Below the builder, a results panel shows a single document's details, including its ID, price, type, size, and location information.

```
_id: ObjectId('67c2db3ebce28850f4534a48')
price: 495000
house_type: "planta 1"
house_type_2: "exterior"
rooms: 3
m2: 118
elevator: true
garage: true
neighborhood: "Chopera"
district: "Arganzuela"
price_per_m2: 4194.915254237285
```

Figura 1. Captura de pantalla de la base de datos cargada en MongoDB Compass.

## 2. Queries:

Para conocer la información contenida en la base de datos, utilicé los siguientes códigos para examinar y visualizar los resultados obtenidos. Mediante el uso de `.find({})` en la colección `madridhouseprices`, identifiqué los diferentes campos del documento y la estructura de los datos. Quiero reiterar que cuando se describe un precio o un piso como barato o más barato, se refiere a que es más económico en comparación con un piso que se describe como caro.

Tras un análisis preliminar, se pueden desarrollar diversas consultas específicas para extraer la información requerida. Estas queries son las siguientes:

## 2.1 Limpieza y Preparación de Datos

### 2.1.1 Exploración inicial del dataset

Se realiza una consulta simple para obtener una vista general de la base de datos y verificar la estructura de los documentos almacenados.

The screenshot shows a MongoDB query interface with a code editor at the top containing the following command:

```
March4MadridHousePrices.js x
localhost ▾ MadridHousePrices ▾
1 db.madridhouseprices.find({})
2   .projection({})
3   .sort({})
4   .limit(0)
5
```

Below the code is a table titled "madridhouseprices" with 15,931 documents. The table has columns: \_id, price, house\_type, house\_type\_2, rooms, m2, elevator, garage, neighborhood, district, and price. The data shows various house types (planta 1, planta 2, planta 3, planta 4) with their respective details like exterior/interior, room counts, and sizes. Neighborhoods like Arganzuela, Palos de Moguer, Legazpi, and Delicias are listed, along with districts like Imperial and Acacias.

_id	price	house_type	house_type_2	rooms	m2	elevator	garage	neighborhood	district	price	
1	67c2db3ebcce28850f4534a4b	495,000 (0.49M)	planta 1	exterior	3	118	true	true	Chopera	Arganzuela	4,1€
2	67c2db3ebcce28850f4534a49	485,000 (0.48M)	planta 2	exterior	2	82	true	true	Palos de Moguer	Arganzuela	5,9€
3	67c2db3ebcce28850f4534a4a	315,000 (0.32M)	planta 2	exterior	2	72	false	false	Legazpi	Arganzuela	4,3€
4	67c2db3ebcce28850f4534a4b	585,000 (0.58M)	planta 4	exterior	2	174	true	true	Palos de Moguer	Arganzuela	3,3€
5	67c2db3ebcce28850f4534a4c	255,000 (0.26M)	bajo	exterior	3	75	false	false	Acacias	Arganzuela	3,4€
6	67c2db3ebcce28850f4534a4d	299,000 (0.30M)	planta 1	exterior	1	69	true	false	Chopera	Arganzuela	4,3€
7	67c2db3ebcce28850f4534a4e	265,000 (0.27M)	planta 3	exterior	2	54	true	false	Delicias	Arganzuela	4,9€
8	67c2db3ebcce28850f4534a4f	290,000 (0.29M)	planta 3	interior	4	69	true	false	Chopera	Arganzuela	4,2€
9	67c2db3ebcce28850f4534a50	660,220 (0.66M)	planta 2	exterior	3	129	true	true	Imperial	Arganzuela	5,1€
10	67c2db3ebcce28850f4534a51	525,000 (0.53M)	planta 4	exterior	4	111	true	false	Chopera	Arganzuela	4,7€

Figura 2. Conjunto de datos madridhouseprices mostrado en forma de tabla.

## 2.2 Encontrar los Barrios Más Baratos y Más Caros

### 2.2.1. Limpieza y preparación de datos

Los códigos eliminan los valores atípicos o incompletos mejorando así la calidad del conjunto de datos y del análisis. Además, se crea una nueva columna price\_per\_m<sup>2</sup> para facilitar las comparaciones entre propiedades.

The screenshot shows a MongoDB query interface with a code editor at the top containing the following command:

```
March4MadridHousePrices.js x
localhost ▾ MadridHousePrices ▾
5
6 db.madridhouseprices.deleteMany({
7   $or: [
8     { price: { $lt: 50000 } },
9     { price: { $gt: 1000000 } },
10    { m2: { $lt: 10 } },
11    { m2: { $exists: false } }
12  ]
13 });
14
15
16 db.madridhouseprices.updateMany({}, [
17   { $set: { price_per_m2: { $divide: ["$price", "$m2"] } } }
18 ]);
19
20
```

Below the code is a JSON result showing the execution time (0.272 s) and the updated document count:

```
Find x Result x
0.272 s
1 + {
2   "acknowledged" : true,
3   "matchedCount" : 15931,
4   "modifiedCount" : 0
5 }
```

Figura 3. Resultado de los códigos en formato JSON.

El resultado sugiere que, aunque 15.931 documentos coincidían con los criterios de actualización, ninguno se modificó, probablemente porque todos los documentos ya tenían el valor correcto de precio\_por\_m2 o porque la operación \$divide no produjo ningún cambio.

## 2.3. Encontrar los Pisos Más Baratos y Más Caros

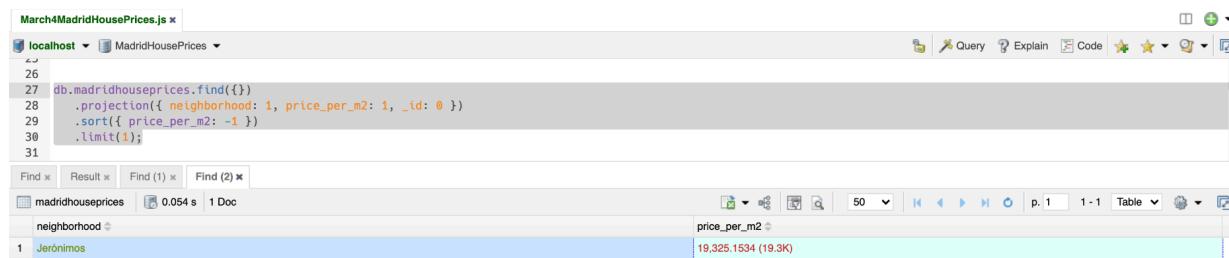
### 2.3.1. Identificación de los barrios más baratos y más caros

Se calculan los precios medios por metro cuadrado en cada barrio para determinar cuáles tienen los costos más altos y más bajos.



```
March4MadridHousePrices.js x
localhost ▾ MadridHousePrices ▾
20
21 db.madridhouseprices.find({})
22   .projection({ neighborhood: 1, price_per_m2: 1, _id: 0 })
23   .sort({ price_per_m2: 1 })
24   .limit(1);
25
26
Find x Result x Find (1) x
madridhouseprices 0.120 s 1 Doc
neighborhood
1 Pilar
price_per_m2
543.4783
```

Figura 4. El barrio con el precio más barato por m<sup>2</sup> es Pilar.



```
March4MadridHousePrices.js x
localhost ▾ MadridHousePrices ▾
26
27 db.madridhouseprices.find({})
28   .projection({ neighborhood: 1, price_per_m2: 1, _id: 0 })
29   .sort({ price_per_m2: -1 })
30   .limit(1);
31
Find x Result x Find (1) x Find (2) x
madridhouseprices 0.054 s 1 Doc
neighborhood
1 Jerónimos
price_per_m2
19,325.1534 (19.3K)
```

Figura 5. El barrio con el precio más caro por m<sup>2</sup> es Jerónimos.

Existe una gran diferencia en el precio entre Pilar y Jerónimos. El precio de Jerónimos es elevado debido a la presencia de importantes edificios gubernamentales nacionales en el distrito, como la Real Academia Española, el Ayuntamiento de Madrid, el Hotel Ritz, el Palacio de la Bolsa de Madrid, el Cuartel General de la Armada, entre otros. Por otro lado, el barrio de Pilar tiene un precio más asequible, ya que históricamente se desarrolló como un barrio para la clase trabajadora.

### 2.3.2. Identificación de los pisos más baratos y más caros

Se buscan los inmuebles con el precio total más bajo y más alto en toda la base de datos.



The screenshot shows the MongoDB Compass interface. The code in the query editor is:

```
db.madridhouseprices.find({})  
.projection({ neighborhood: 1, price: 1, m2: 1, rooms: 1, _id: 0 })  
.sort({ price: 1 })  
.limit(1);
```

The results pane shows one document:

price	rooms	m2	neighborhood
53.800 (53.8K)	3	62	San Cristóbal

Figura 6. El piso más barato del conjunto de datos es un piso en San Cristóbal con 3 habitaciones y una superficie de 62 m<sup>2</sup> por 53.800 euros.



The screenshot shows the MongoDB Compass interface. The code in the query editor is:

```
db.madridhouseprices.find({})  
.projection({ neighborhood: 1, price: 1, m2: 1, rooms: 1, _id: 0 })  
.sort({ price: -1 })  
.limit(1);
```

The results pane shows one document:

price	rooms	m2	neighborhood
10.000.000 (10.0M)	14	849	El Viso

Figura 7. El piso más caro del conjunto de datos es un piso en El Viso con 14 habitaciones y una superficie de 849 m<sup>2</sup> por 10.000.000 euros.

San Cristóbal tiene pisos más baratos debido a su carácter de barrio obrero, con viviendas más pequeñas y accesibles, y su ubicación al sur de la ciudad, fuera de la M-30, lo que lo aleja del centro. Por otro lado, El Viso es un barrio exclusivo de alto poder adquisitivo, con propiedades de lujo y una ubicación privilegiada cerca del Paseo de la Castellana. Este barrio atrae a personas con mayor capacidad económica, lo que eleva los precios. Además, alberga numerosas embajadas y oficinas.

## 2.4. Aggregation Queries

### 2.4.1. Análisis de la distribución de precios

Se analiza cómo varía el precio total en función del número de habitaciones de cada vivienda según el barrio y su distrito.

*En los siguientes resultados, se utiliza Arganzuela como punto de referencia, ya que es el primero en orden alfabético en el conjunto de datos.*

The screenshot shows the MongoDB Compass interface with a query named "March4MadridHousePrices.js". The query uses the "aggregate" method on the "madridhouseprices" collection. It groups documents by district and neighborhood, calculates the average price per flat (avgPricePerFlat) using the \$avg operator, and sorts the results by district and neighborhood. The results table displays 15 rows of data, each representing a different neighborhood within the Arganzuela district. The columns are labeled: \_id, district, neighborhood, rooms, and avgPricePerFlat. The prices range from 229,111.1111 (0.23M) to 180,000.0 (0.18M), with a general upward trend as the number of rooms increases.

_id	district	neighborhood	rooms	avgPricePerFlat
1	Arganzuela	Acacias	1	229,111.1111 (0.23M)
2	Arganzuela	Acacias	2	260,603.5714 (0.26M)
3	Arganzuela	Acacias	3	447,043.1818 (0.45M)
4	Arganzuela	Acacias	4	631,111.1111 (0.63M)
5	Arganzuela	Chopera	1	217,624.9375 (0.22M)
6	Arganzuela	Chopera	2	290,622.2222 (0.29M)
7	Arganzuela	Chopera	3	326,954.0541 (0.33M)
8	Arganzuela	Chopera	4	453,545.4545 (0.45M)
9	Arganzuela	Delicias	1	226,021.0526 (0.23M)
10	Arganzuela	Delicias	2	266,846.875 (0.27M)
11	Arganzuela	Delicias	3	391,930.0 (0.39M)
12	Arganzuela	Delicias	4	475,727.2727 (0.48M)
13	Arganzuela	Delicias	5	467,500.0 (0.47M)
14	Arganzuela	Delicias	7	675,000.0 (0.68M)
15	Arganzuela	Dúplex en Imperial	1	180,000.0 (0.18M)

Figura 8. Ejemplo de precios medios por piso entre los barrios del distrito de Arganzuela.

En todos los barrios de Arganzuela, los pisos de una habitación son los más baratos, y el precio aumenta a medida que crece el número de habitaciones.

### 2.4.2. Precio por metro cuadrado por distrito

Se examina cómo se distribuyen los precios por metro cuadrado en cada distrito de Madrid.

The screenshot shows the MongoDB Compass interface with a query named "March4MadridHousePrices.js". The query uses the "aggregate" method on the "madridhouseprices" collection. It groups documents by district, calculates the average price per square meter (avgPricePerM2) using the \$avg operator, and sorts the results by avgPricePerM2 in descending order. The results table displays 7 rows of data, each representing a different district. The districts are listed in ascending order of price: barrio de salamanca, chamberi, chamartin, retiro, centro, moncloa, and Arganzuela. The prices range from 7,336.354 (7.3K) to 4,426.1773 (4.4K).

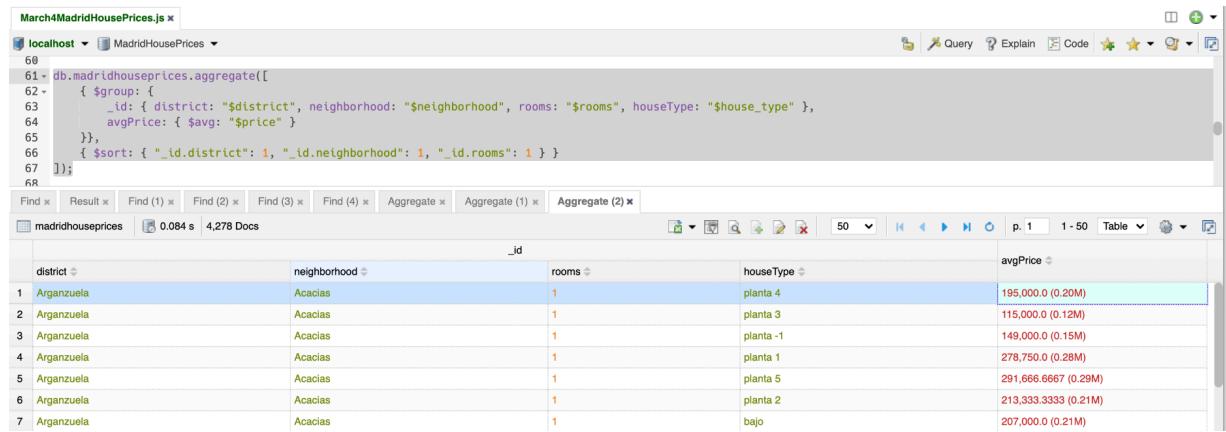
_id	avgPricePerM2	
1	barrio de salamanca	7,336.354 (7.3K)
2	chamberi	6,192.5535 (6.2K)
3	chamartin	5,881.3274 (5.9K)
4	retiro	5,765.965 (5.8K)
5	centro	5,525.6441 (5.5K)
6	moncloa	4,763.6699 (4.8K)
7	Arganzuela	4,426.1773 (4.4K)

Figura 9. Precio medio por m<sup>2</sup> en cada distrito en orden decreciente.

Barrio de Salamanca tiene el precio medio por m<sup>2</sup> más alto, mientras que Arganzuela ocupa el séptimo lugar en la lista. Cabe destacar que los tres distritos con los precios más elevados—Barrio de Salamanca, Chamberí y Chamartín—se caracterizan por su abundancia de edificios históricos, oficinas, comercios y viviendas bien establecidas. En cambio, Arganzuela es un distrito más residencial, lo que podría explicar su precio más moderado en comparación con los tres primeros.

#### 2.4.3. Análisis de correlación (habitaciones, barrio, distrito, tipo de vivienda)

Se investiga la relación entre el número de habitaciones, el barrio, el distrito y el tipo de vivienda para identificar patrones.



*Figura 10. Precio medio de un piso según el número de habitaciones y la planta en la que se encuentra.*

Se observa que el precio de los pisos de una habitación varía según la planta. En algunos casos, las plantas bajas son más caras que las superiores, lo que resulta curioso y plantea la pregunta de por qué se producen estas diferencias de precio.

#### 2.4.4. Diferencia de precio de cada distrito en comparación con el distrito más caro

Se comparan los precios promedio de cada distrito con el del distrito más caro para medir la brecha de costos.

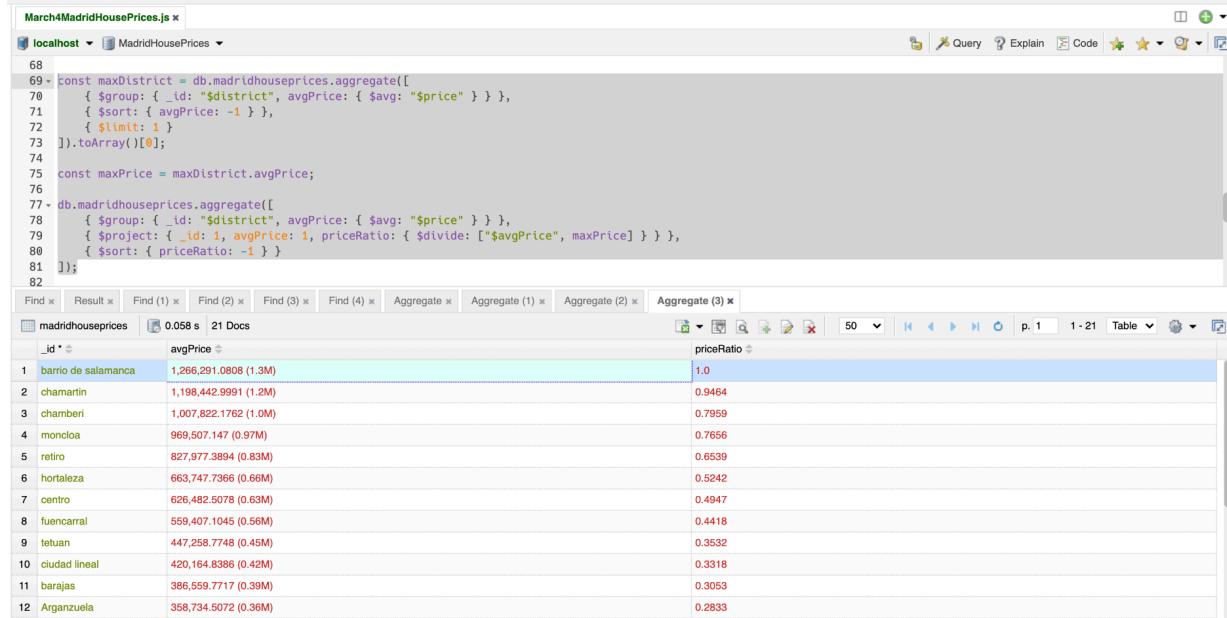


Figura 11. Los 5 distritos con los precios promedio más caros de Madrid.

Con un precio medio de 1,27 millones de euros, el Barrio de Salamanca encabeza nuevamente la lista, mientras que Arganzuela se encuentra en el puesto 12.

#### 2.4.5. Comparación de precios por tipo de vivienda por distrito.

Se analizan las diferencias de precios entre los distintos tipos de viviendas dentro de cada distrito.

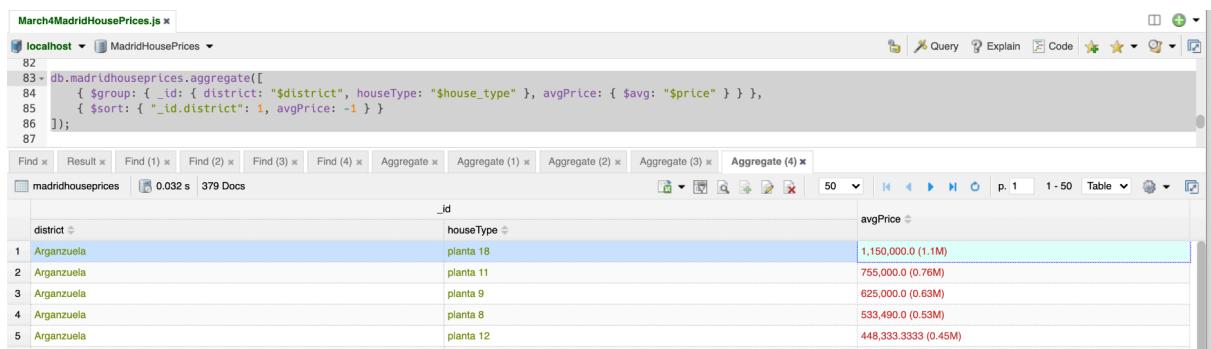


Figura 12. La planta 18 tiene el precio promedio mas caro en el distrito de Arganzuela.

En todo Arganzuela, el piso en la planta 18 es el más caro. Se puede asumir que estos pisos son más nuevos, con un diseño moderno y posiblemente ubicados en una urbanización.

## 2.4.6. Número de habitaciones más caro por distrito

Se identifica cuál es el número de habitaciones con el precio más alto en cada distrito.

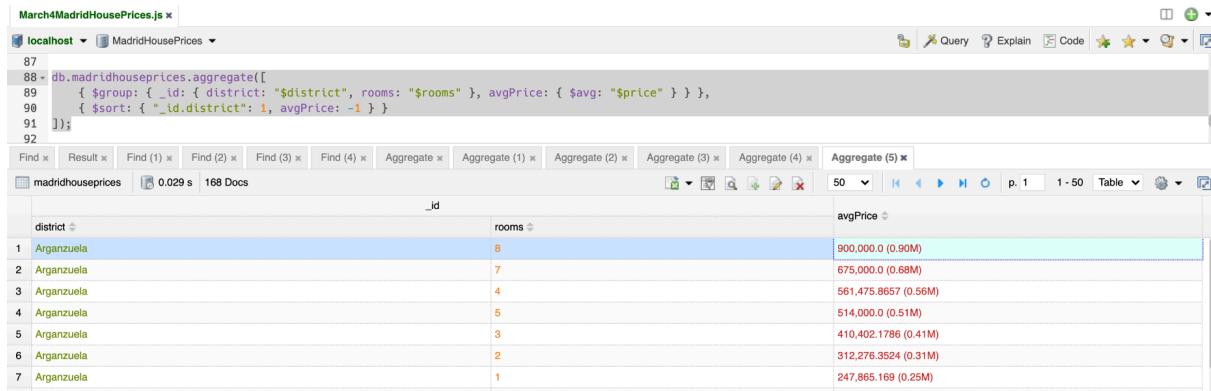


Figura 13. En Arganzuela, un piso de 8 habitaciones es en el primer lugar y cuesta 900.000 euros.

Un piso de 8 habitaciones tiene un precio medio de 900.000 euros; a medida que el número de habitaciones disminuye, el precio también lo hace.

## 2.4.7. Distrito con el precio más barato en comparación con la mayor área en metros cuadrados

Se busca el distrito con el precio más bajo considerando también el tamaño de las propiedades en metros cuadrados.

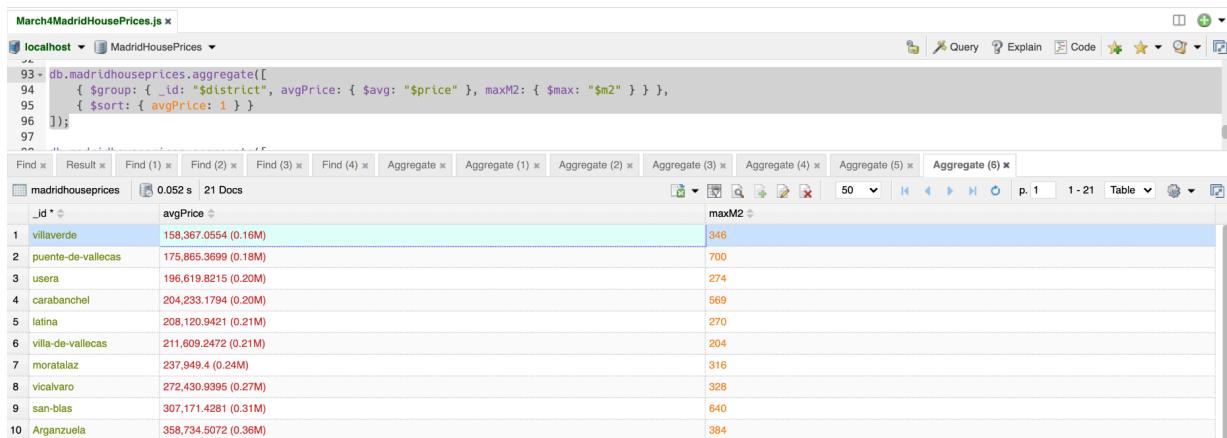


Figura 14. Los 5 distritos más baratos comparados con el piso más grande que se puede comprar en cada uno de ellos.

Se puede observar que los 5 distritos con los precios medios más baratos, donde se pueden comprar pisos grandes, son barrios de clase trabajadora o ciudades dormitorio, y todos se encuentran fuera de la M-30.

#### 2.4.8. Piso con garaje vs. sin garaje por distrito

Se compara los precios de viviendas con y sin garaje dentro de cada distrito para evaluar el impacto del estacionamiento en los costos.

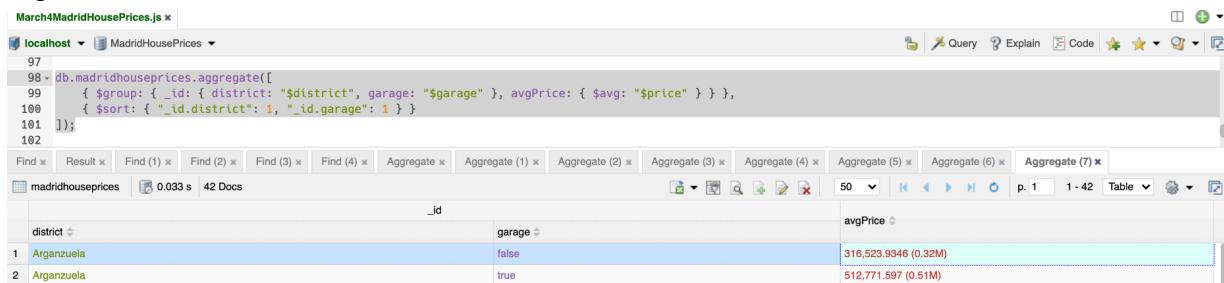


Figura 15. Diferencia de los precios medios de los pisos con y sin aparcamiento.

En estos resultados, podemos ver que un piso con garaje es más caro, lo cual tiene sentido ya que el garaje representa un terreno adicional en el título de propiedad.

#### 2.4.9. Barrio más económico por distrito (Precio por m<sup>2</sup>)

Se identifica el barrio más asequible dentro de cada distrito según el precio por metro cuadrado.

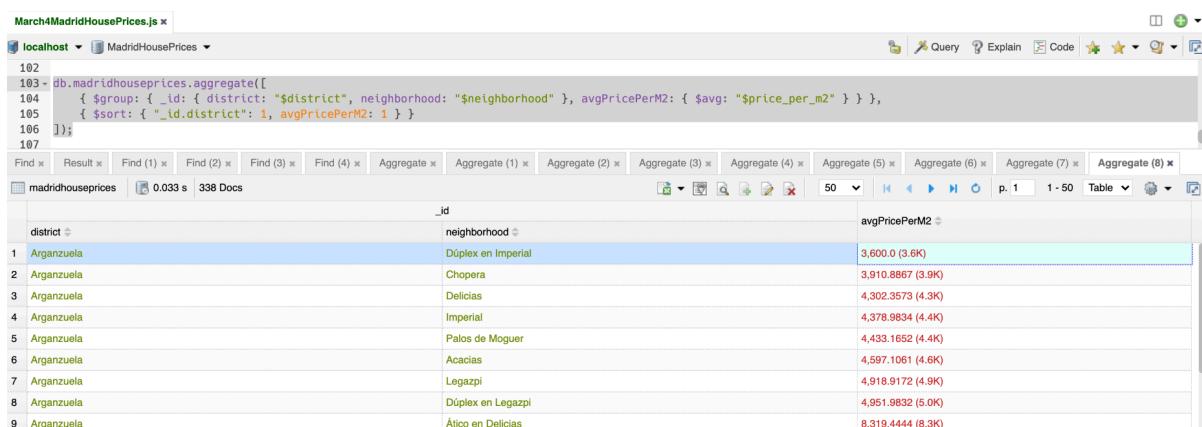


Figura 16. Precios medios por m<sup>2</sup> de cada barrio del distrito de Arganzuela.

Se observa que un dúplex en Imperial tiene el precio medio por m<sup>2</sup> más económico del distrito de Arganzuela, mientras que un ático en Delicias resulta más caro. Esto es interesante, ya que un ático suele ser más pequeño que un dúplex. Sin embargo, se puede inferir que el barrio de Imperial no cuenta con estación de metro ni de trenes de cercanías, lo que podría hacer que el precio de los pisos sea más bajo en comparación con Delicias, que está más cerca del centro y tiene una mejor conectividad de transporte público.

#### 2.4.10. Barrio con el precio por m<sup>2</sup> mas caro en relación con el número de habitaciones

Se analiza cuál es el barrio donde el precio por metro cuadrado es más caro en función del número de habitaciones.

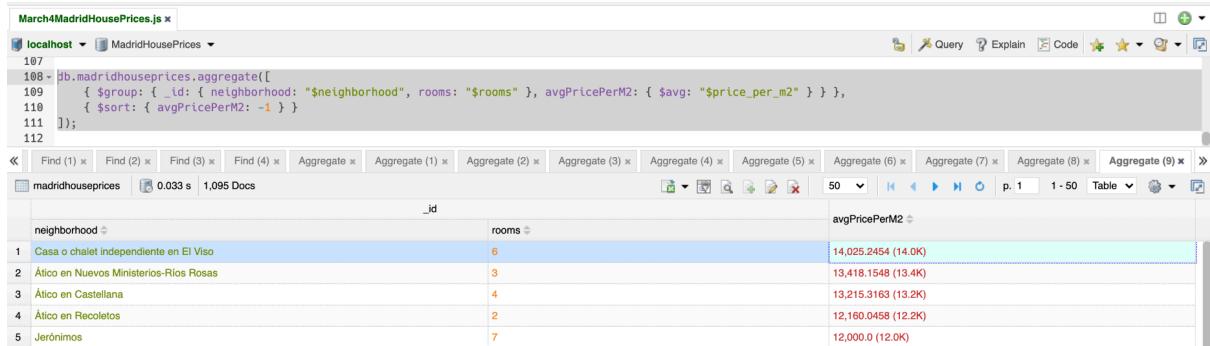


Figura 17. Los 5 barrios de Madrid con la peor relación entre el precio por m<sup>2</sup> y el número de habitaciones.

Se puede observar en la lista que los 5 primeros distritos son los conocidos como caros. Arganzuela ni siquiera se encuentra en el top 20 de la lista.

### 3. Conclusion

El mercado inmobiliario de Madrid varía mucho según el barrio, el tipo de vivienda y el número de habitaciones. Zonas como Salamanca, El Viso y Recoletos tienen los precios más altos por metro cuadrado, mientras que barrios como El Plantío y Conde Orgaz-Piovera ofrecen opciones más asequibles, incluso para casas grandes. En general, vivir en el centro es más caro, pero en las afueras se puede conseguir más espacio por menos dinero.

Los pisos de lujo en el centro mantienen su valor, con precios que superan los 10.000 €/m<sup>2</sup> debido a su ubicación y exclusividad. Los chalets tienen un rango de precios más amplio: en zonas exclusivas pueden costar más de 4M €, mientras que en áreas menos céntricas hay opciones por menos de 1M €. Esto confirma que la ubicación es el factor clave en el precio de la vivienda.

El número de habitaciones influye en el precio total, pero no siempre en el precio por metro cuadrado. Algunas casas grandes en barrios periféricos son más asequibles que pisos pequeños en zonas exclusivas como Recoletos o Castellana. Esto demuestra que la demanda y la exclusividad pesan más en el precio que el tamaño.

Para compradores e inversores, es clave considerar no solo el precio por metro cuadrado, sino también la ubicación, los servicios y las tendencias del mercado. Características como el garaje y el ascensor también influyen en el valor. En resumen, Madrid tiene un mercado diverso, donde las zonas exclusivas mantienen precios altos, mientras que en la periferia se pueden encontrar viviendas más grandes a mejor precio.