



# **Business Intelligence and Big Data Management**

## **7th Semester**

### **Software and Data Analysis Technologies**

#### **Team 17**

# Business Intelligence Approach



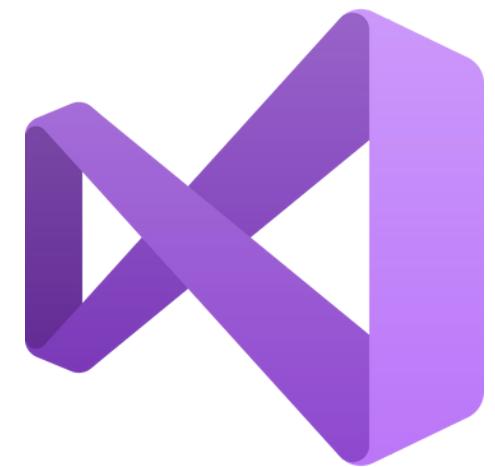
Goal 1: Understand which factors comprise a superhost



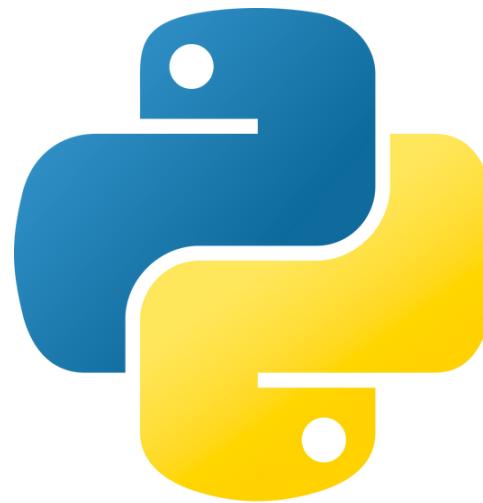
Goal 2: Help hosts and customers make better decisions



## Tools we used



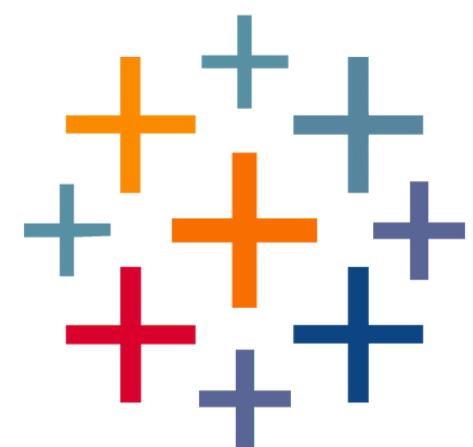
Visual Studio



Python



SSMS 2019



Tableau



Pandas



# The Dataset

296,233 rows

**Listing ID**  
**Name**  
**Host ID**  
**Host Name**  
**Host Response Rate**  
**Host Is Superhost**  
**Host total listings count**  
**Street**  
**City**  
**Neighbourhood cleansed**  
**State**  
**Country**

**Property type**  
**Room type**  
**Accommodates**  
**Bathrooms**  
**Bedrooms**  
**Amenities**  
**Price**  
**Minimum nights**  
**Maximum nights**  
**Availability 365**

**Calendar last scraped**  
**Number of reviews**  
**Last Review Date**  
**Review Scores Rating**  
**Review Scores Accuracy**  
**Review Scores Cleanliness**  
**Review Scores Checkin**  
**Review Scores Communication**  
**Review Scores Location**  
**Review Scores Value**  
**Reviews per month**

■ Kept

■ Dropped

from [kaggle](#)



# The ETL Process

Excel Dataset → Visual Studio → SQL Server

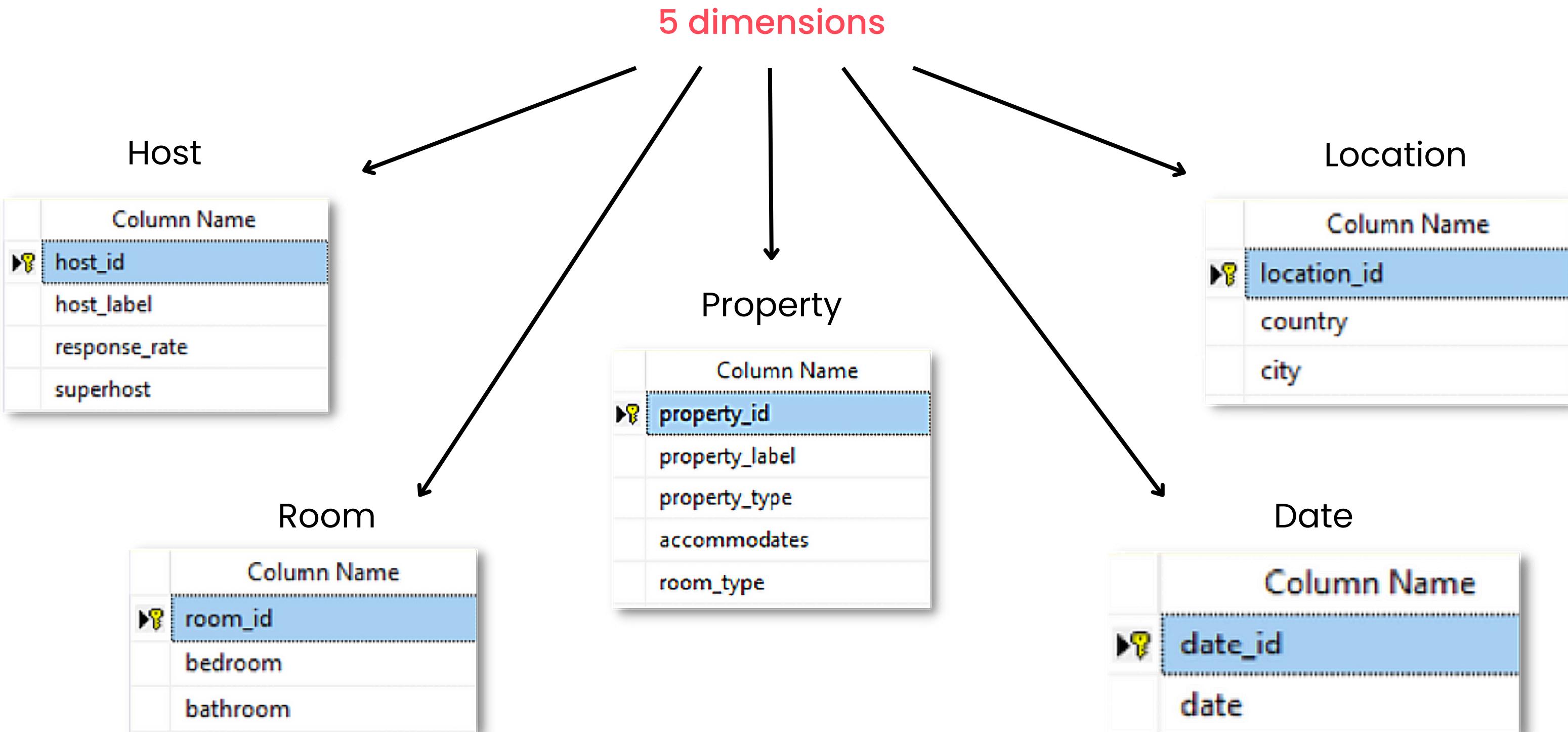
airbnb ←

	Host ID	Host Name	Host Response Rate	Host Is Superhost	Host total listings count
1	6419368	Claudio Antonio	100%	0	1
2	21102685	Ana Belen	100%	0	2
3	34373003	Louise	50%	0	1
4	1927742	Nur	100%	0	3
5	4741190	Henry	100%	1	6
6	71427801	Janning	100%	0	1
7	28206226	Ximena	100%	0	4
8	5624677	Katie	100%	1	2
9	1683600	Carolina	100%	0	2
10	29494768	Darach	100%	0	1
11	83230255	Jpax	100%	0	1
12	39576539	Sam	100%	0	1
13	38313301	Anabel	100%	0	5
14	5433086	Mara	72%	0	90
15	4766870	Michael	100%	0	5
16	2674546	Adriano	100%	0	1
17	42747011	Bernard	100%	0	1

Price	Minimum nights	Maximum nights	Availability 365
80	1	1125	342
50	2	1125	35
59	2	30	230
45	5	60	274
55	8	1125	269
50	1	1125	98
50	2	1125	361
120	1	31	78
...			
90	2	1125	284
200	2	1125	0
43	1	8	2
50	2	1125	28
90	4	1125	326
70	2	1125	76
59	1	1125	365
50	3	1125	324
80	7	1125	243



# The Dimensions(SQL)



# The Fact Table(SQL)

	Column Name	Data Type	Allow Nulls
	Host	int	<input type="checkbox"/>
	Property	int	<input type="checkbox"/>
	Date	int	<input type="checkbox"/>
	Room	int	<input type="checkbox"/>
	Location	int	<input type="checkbox"/>
	Price	float	<input checked="" type="checkbox"/>
	[Number of Reviews]	float	<input checked="" type="checkbox"/>
	[RS Rating]	nvarchar(255)	<input checked="" type="checkbox"/>
	[RS Accuracy]	float	<input checked="" type="checkbox"/>
	[RS Cleanliness]	float	<input checked="" type="checkbox"/>
	[RS Checkin]	float	<input checked="" type="checkbox"/>
	[RS Communication]	float	<input checked="" type="checkbox"/>
	[RS Location]	float	<input checked="" type="checkbox"/>
	[RS Value]	float	<input checked="" type="checkbox"/>

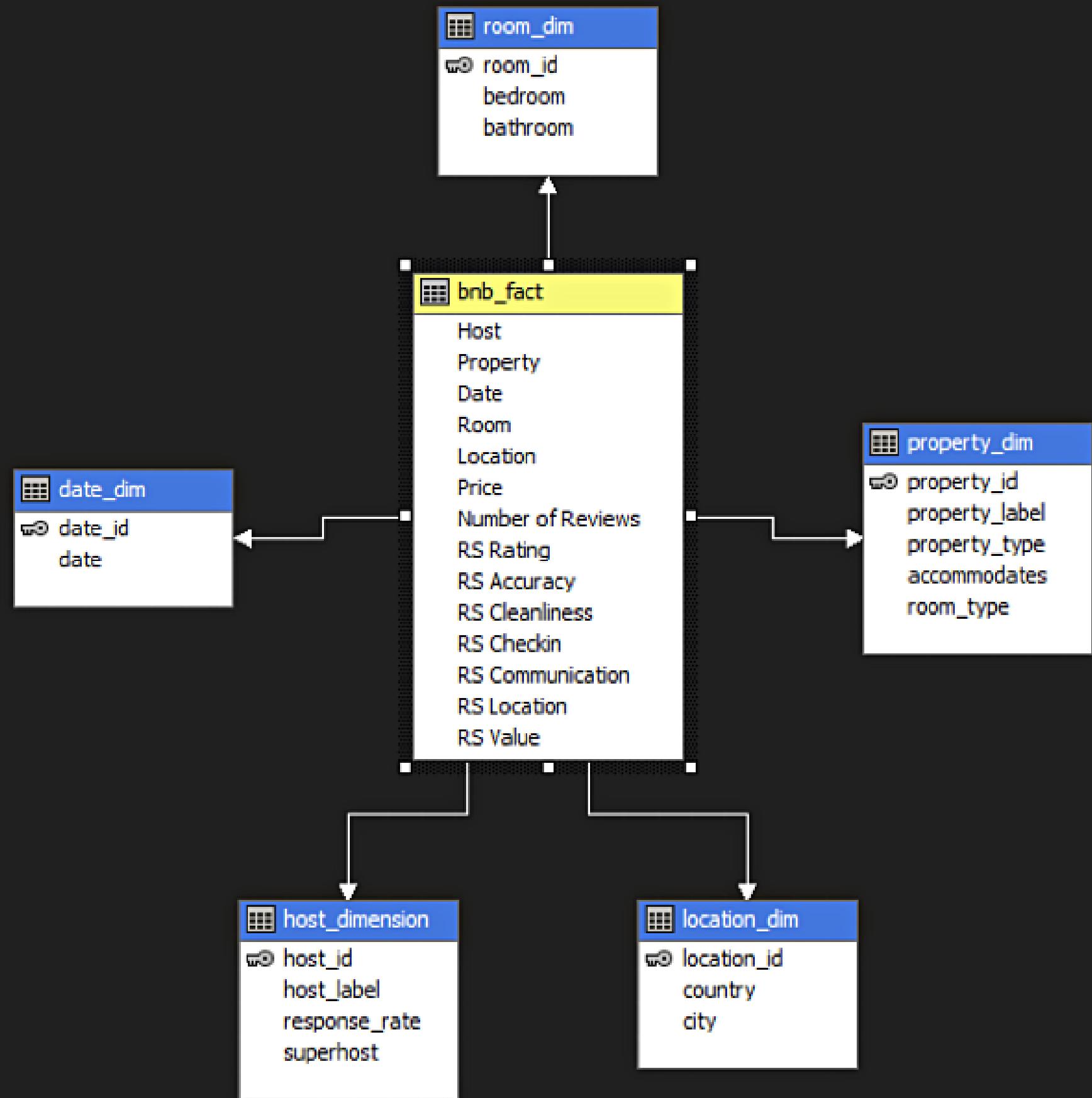
Fact Table Design

	Host	Property	Date	Room	Location	Price	Number of Reviews	RS Ra
1	78886	1	1121	12	380	44	24	97
2	110292	2	784	12	1619	95	19	96
3	196260	3	1321	12	191	110	18	86
4	150004	4	784	25	503	148	116	98
5	190599	5	56	25	1516	179	103	97
6	107753	6	931	12	1090	85	9	90
7	89161	7	1321	12	2069	91	1	100
8	83402	8	94	12	2038	80	4	100
9	138474	9	275	86	429	121	6	100
10	110748	10	1525	12	420	98	24	99
11	35967	11	242	120	2069	215	29	93
12	72002	12	879	12	154	95	1	100
13	64536	13	1527	12	2070	90	14	99
14	123966	14	1454	12	504	70	37	99
15	121860	15	128	25	624	85	22	81
16	13873	16	281	25	40	82	30	99
17	2141	17	242	12	380	125	162	95
18	46687	18	230	12	274	40	44	98
19	81621	19	1345	12	2070	80	1	100
20	168017	20	1460	12	1092	35	8	93

Fact Table Values Sample

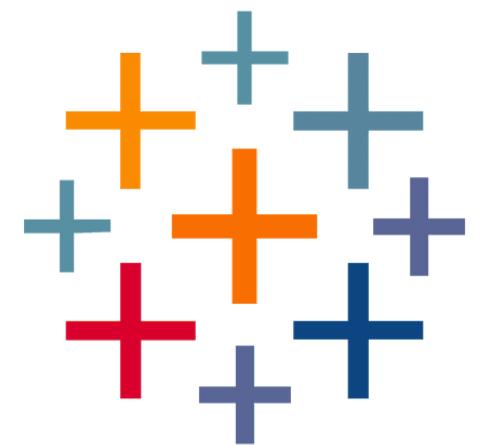


# The Star Schema





Let's Visualise



through Tableau



Global View

Individual Analysis



Amsterdam



Barcelona



London

HOLLYWOOD



Los Angeles



New Orleans



New York



Paris



Portland



San Francisco



Sydney

Select a measure

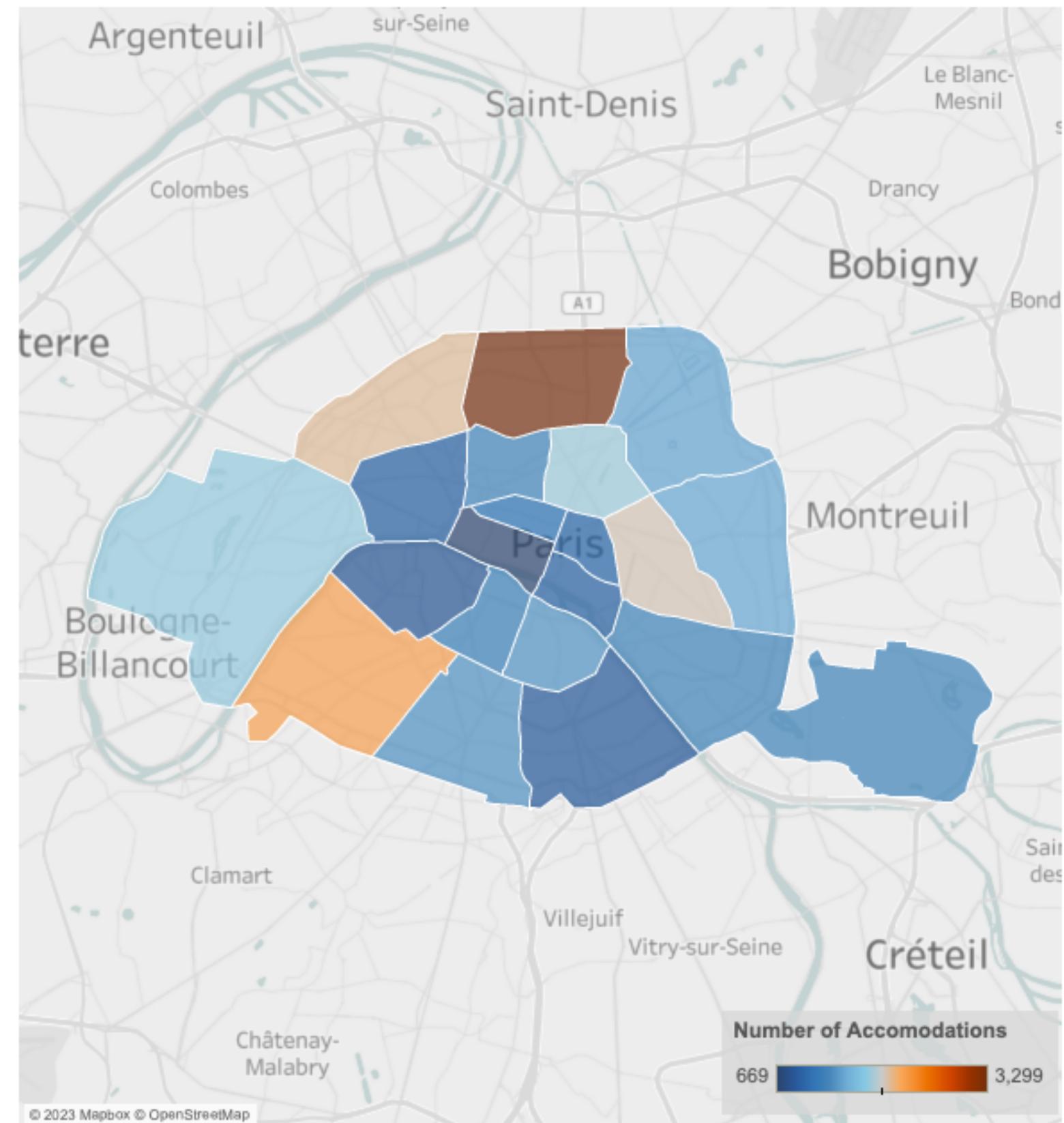
Number of Accomodations

Price

Select a zoom level

300

**24,560 hosts  
29,212 accomodations  
244,465 beds**



Global View

Individual Analysis



Amsterdam



Barcelona



London



Los Angeles



New Orleans



New York



Paris



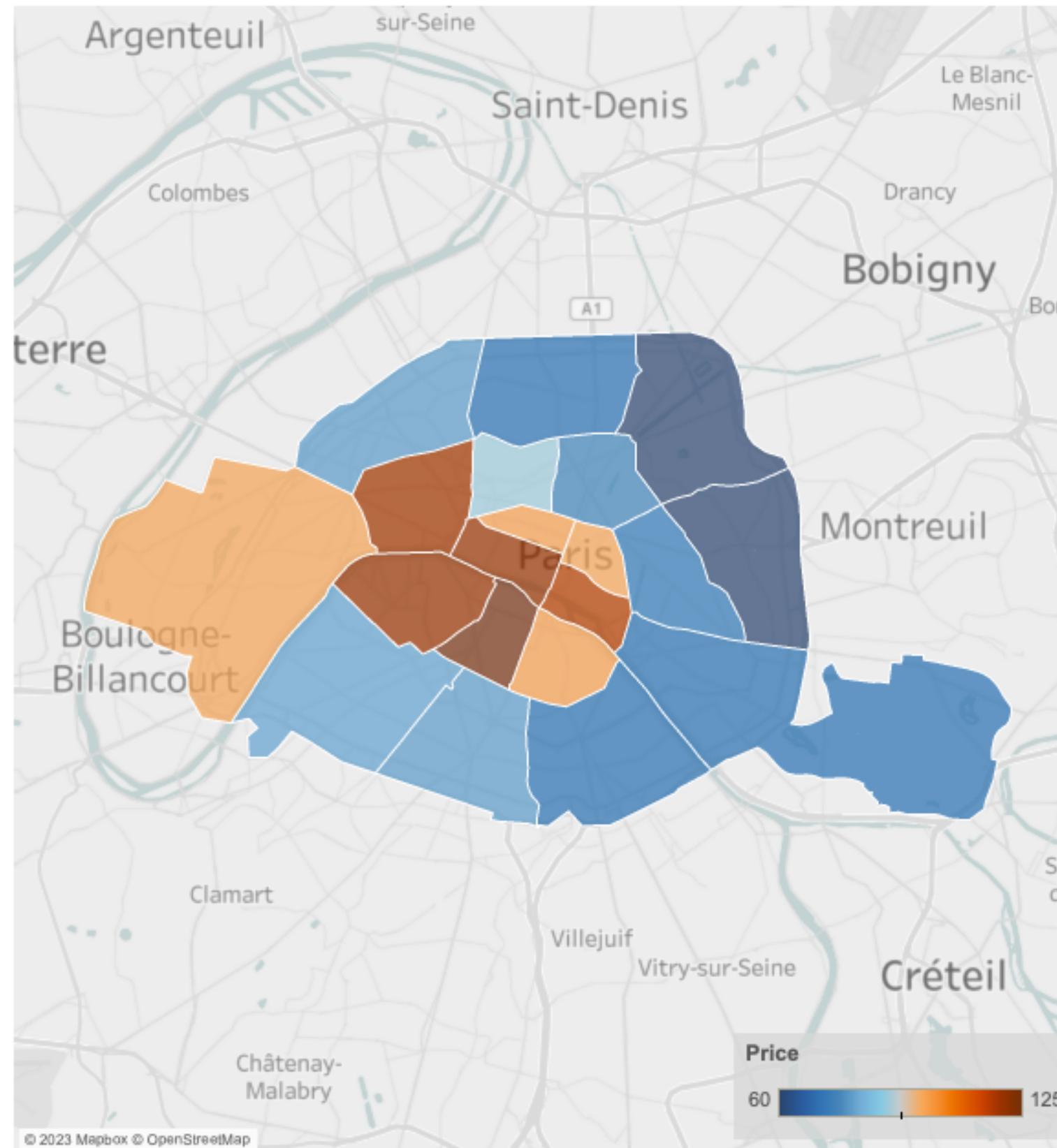
Portland



San Francisco



Sydney



Select a measure

Number of Accomodations

Price

Select a zoom level

300

**24,560 hosts**  
**29,212 accomodations**  
**244,465 beds**



Global View

Individual Analysis



Amsterdam



Barcelona



London

HOLLYWOOD

Los Angeles



New Orleans



New York



Paris



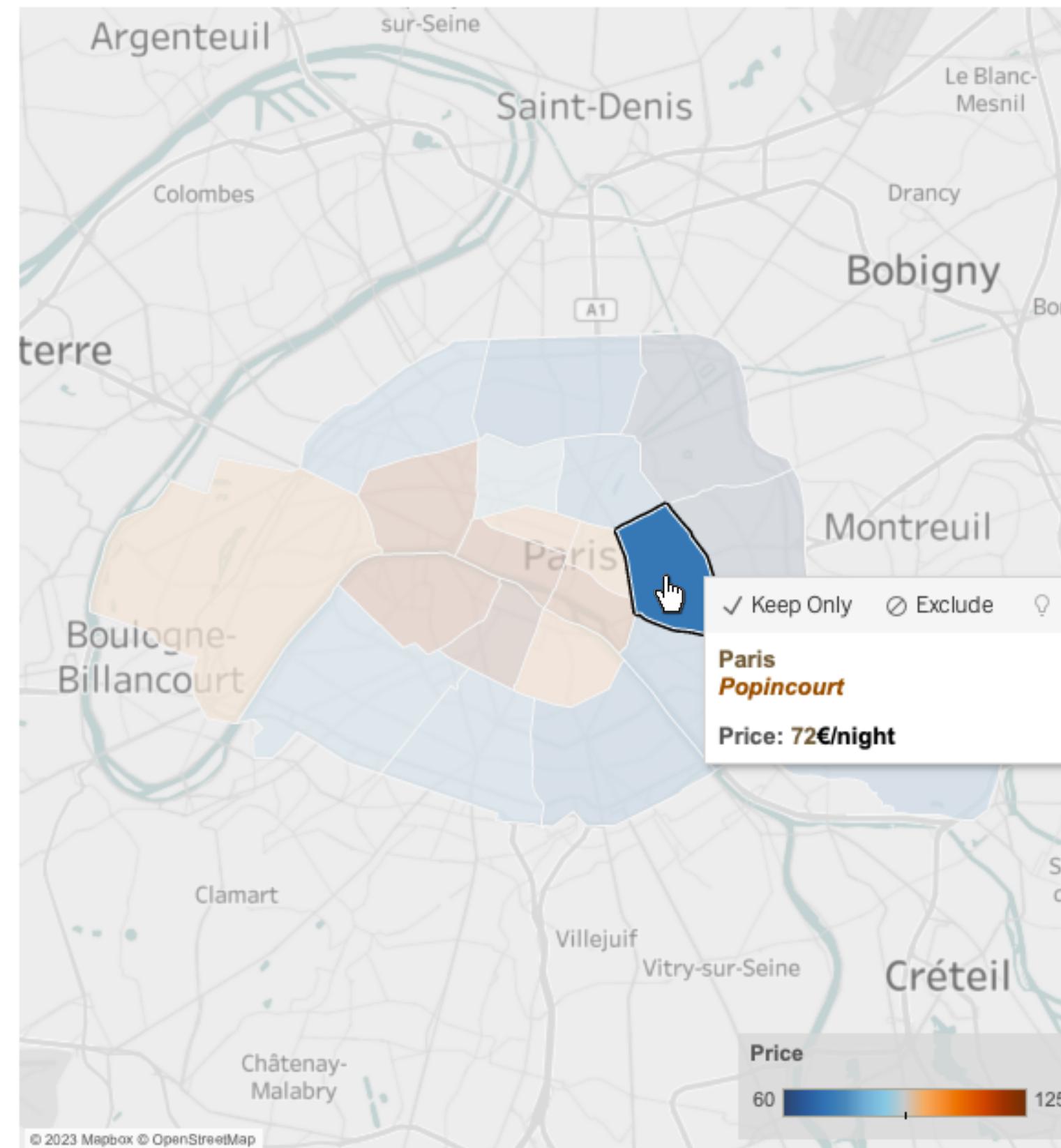
Portland



San Francisco



Sydney



Select a measure

- Number of Accomodations
- Price

Select a zoom level

300

1,916 hosts  
2,024 accomodations  
16,010 beds





Amsterdam



Barcelona



London

HOLLYWOOD



Los Angeles



New Orleans



New York



Paris



San Francisco



Sydney

**Click to select New York!**

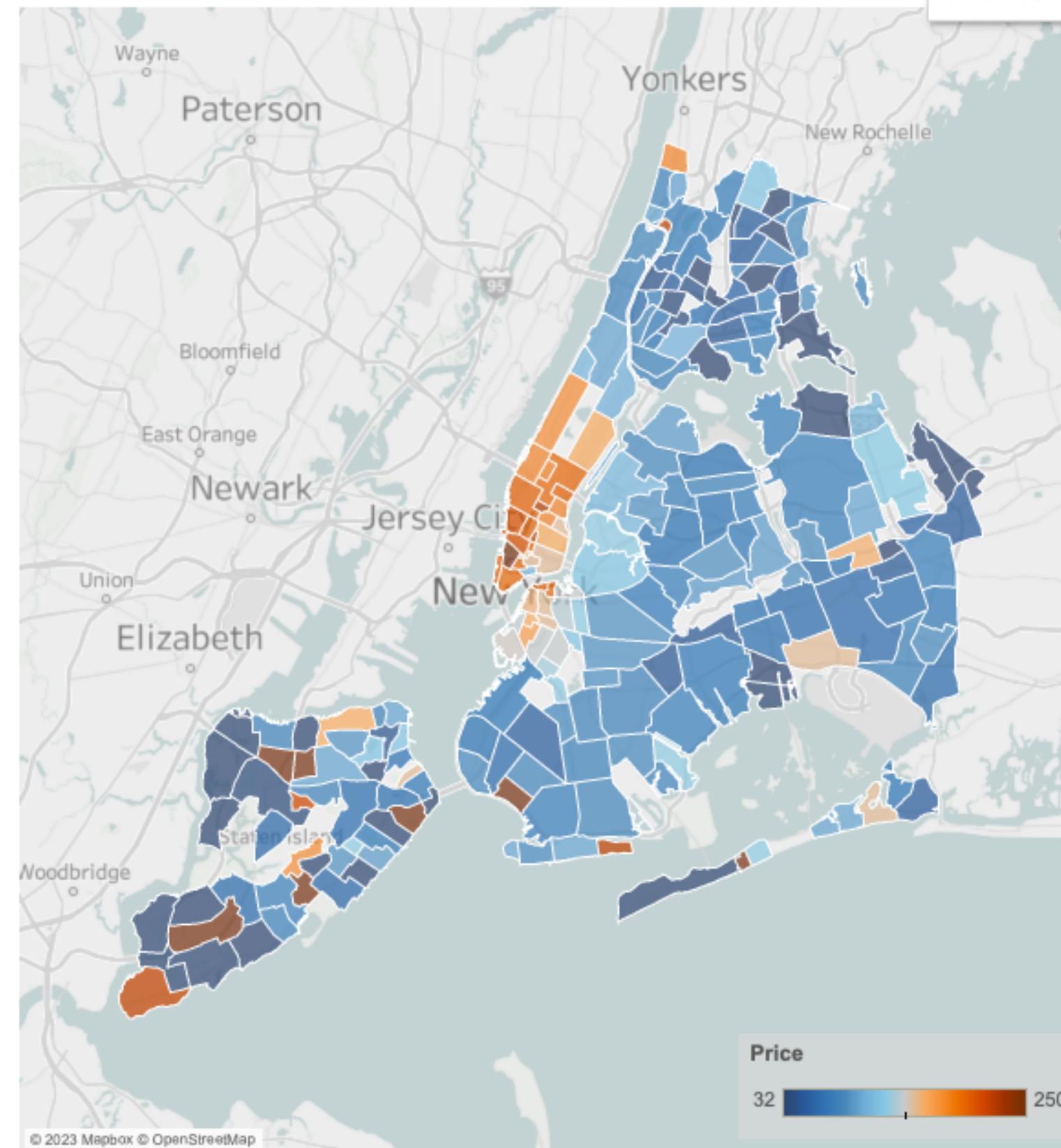
Select a measure

- Number of Accomodations
- Price

Select a zoom level

 ▼

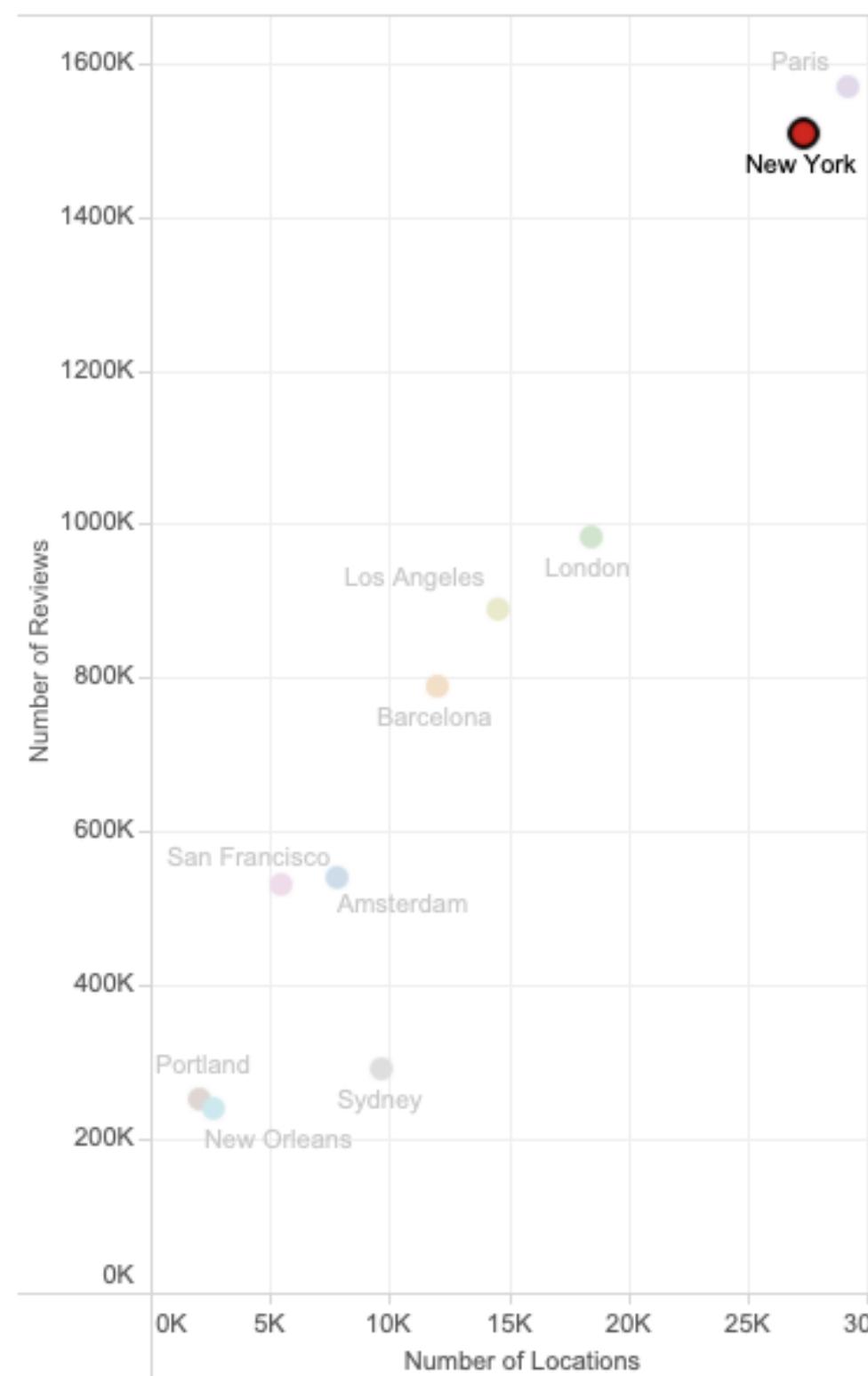
**22,101 hosts  
27,318 accomodations  
207,120 beds**



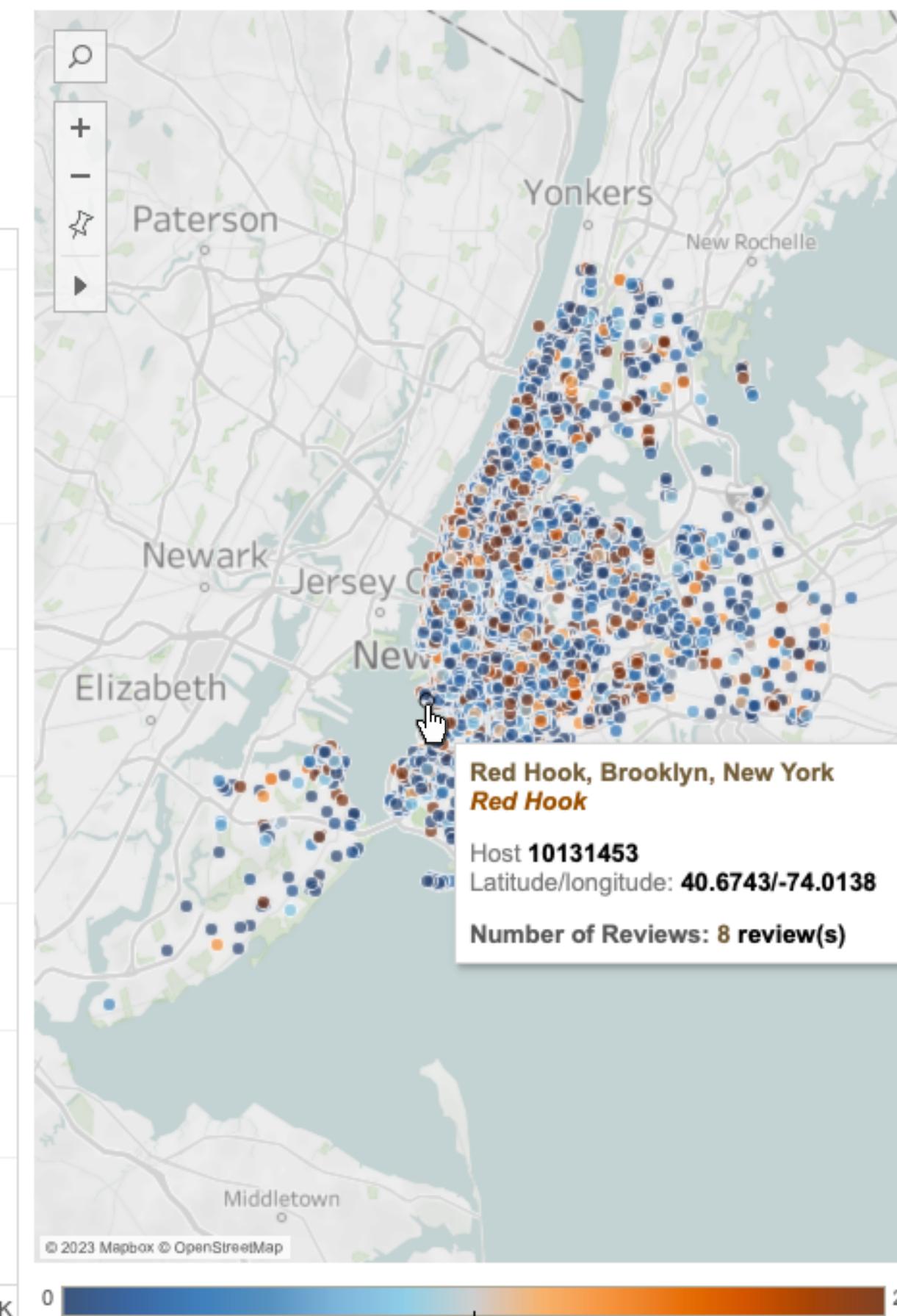
Select a measure for the Map >

- Price
- Number of Reviews

### City by Number of Reviews and Locations



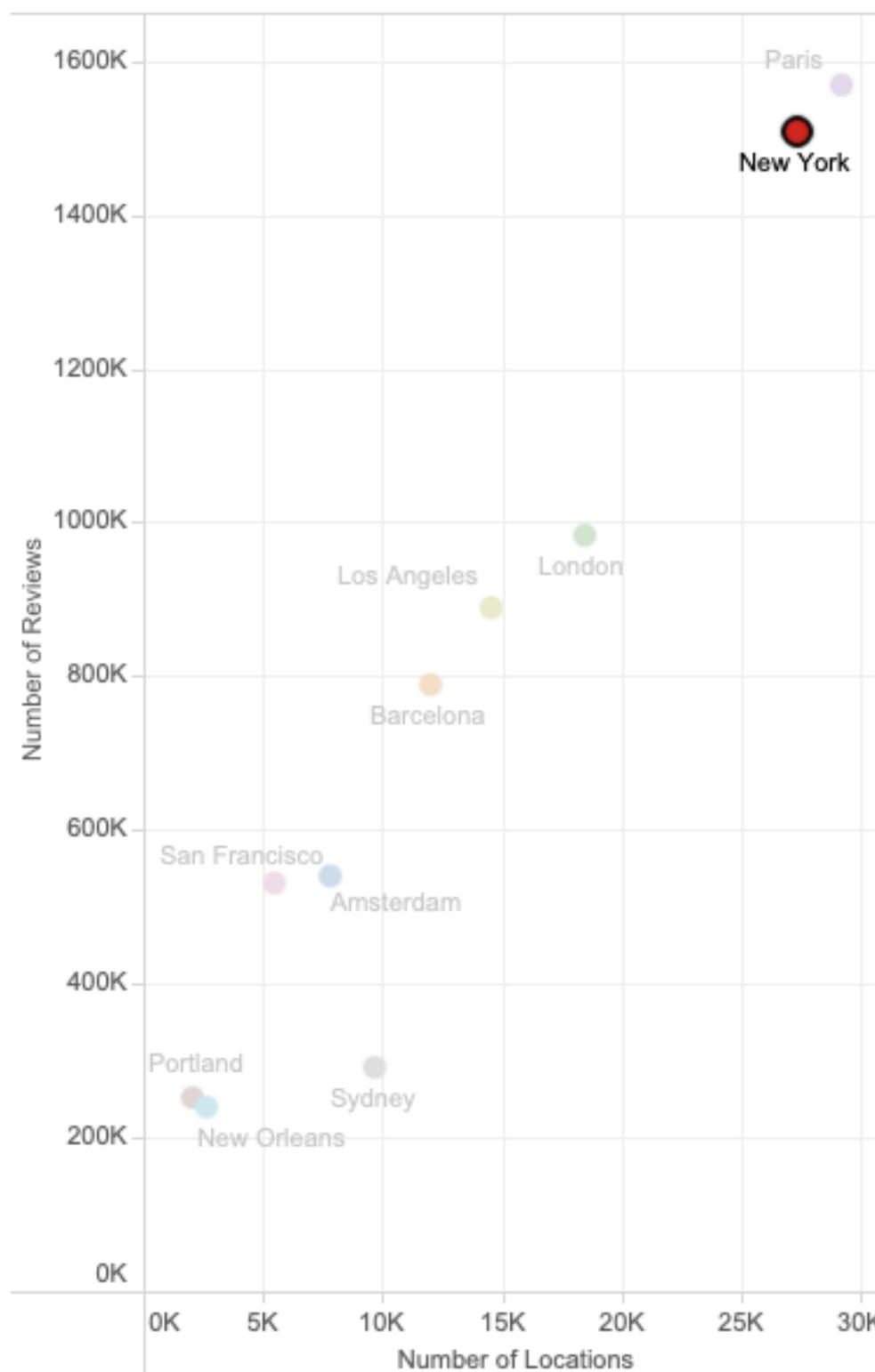
### Location by Number of Reviews



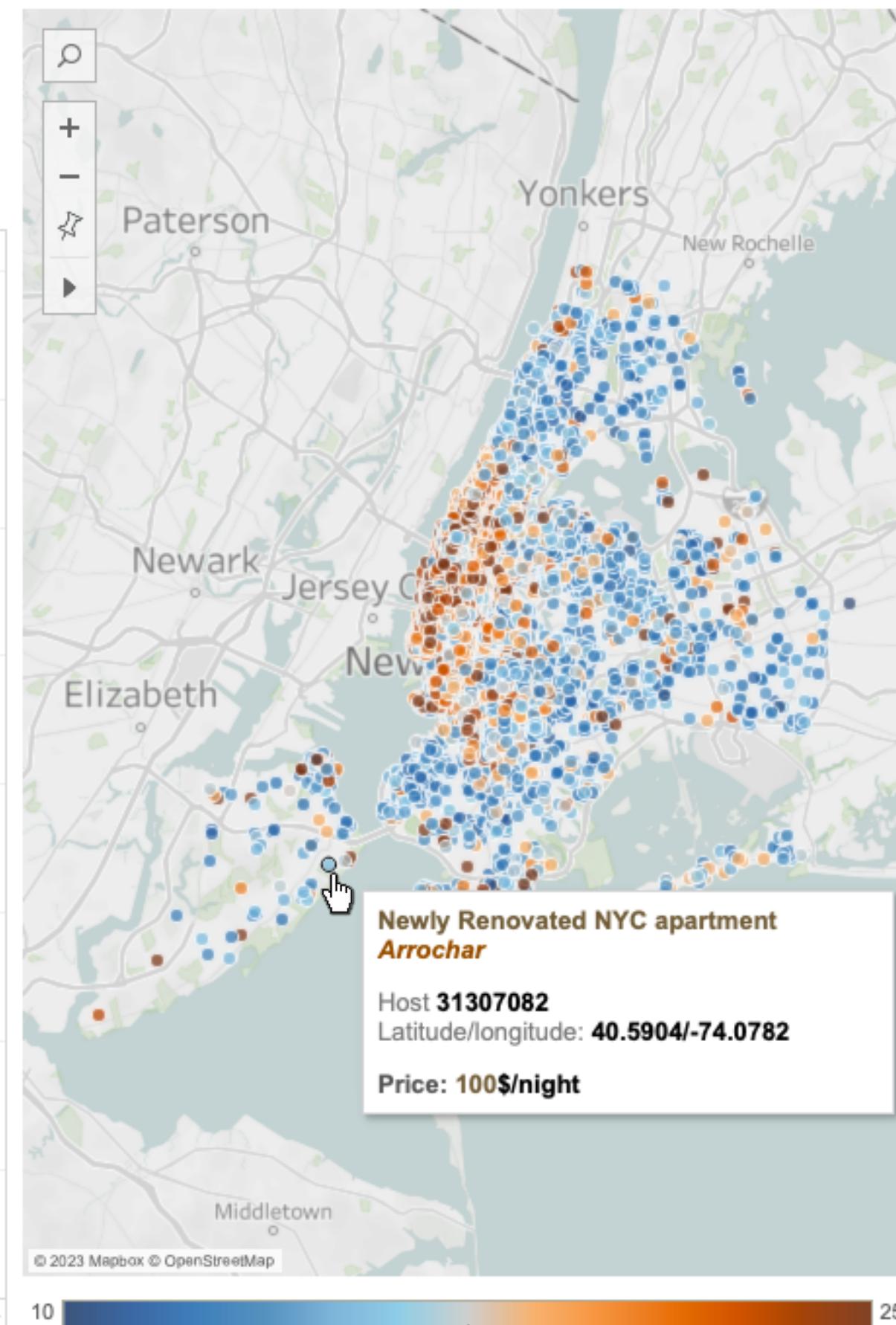
Select a measure for the Map >

- Price
- Number of Reviews

### City by Number of Reviews and Locations



### Location by Price



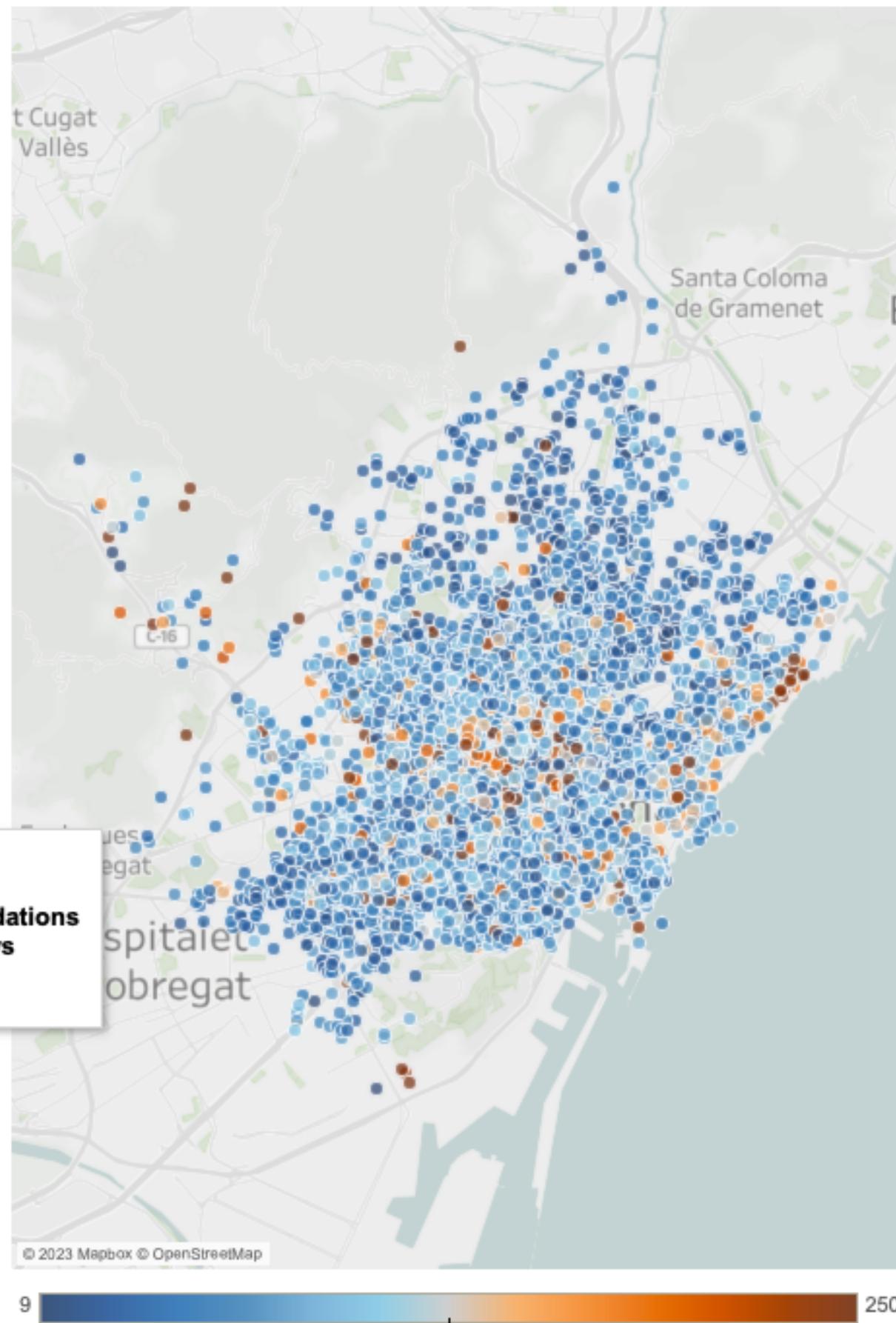
Select a measure for the Map >

- Price
- Number of Reviews

### City by Number of Reviews and Locations



### Location by Price



# Data Mining: Most Frequent Amenity Sets

Which amenitiy sets are the most important  
for review ratings?

Which amenities are the most significant?

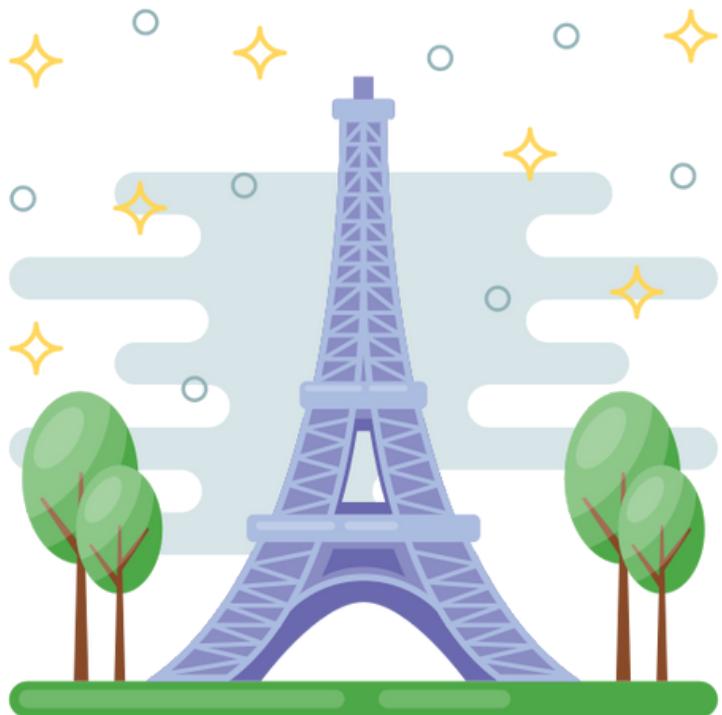
- We used:
- pandas library
  - mlxtend library



# Data Mining: Most Frequent Amenity Sets

Data Samples Used

Paris



Average Listing  
Price  
**97\$**

Global



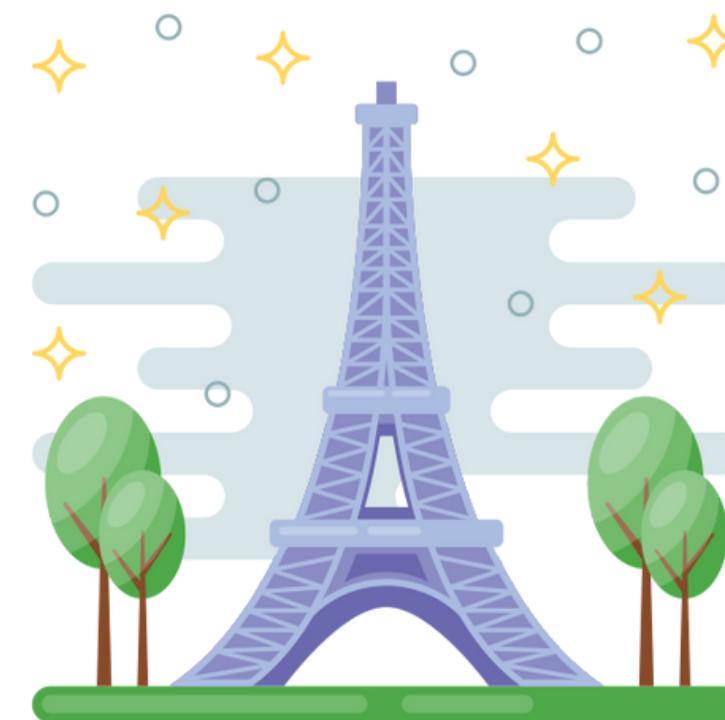
Average Listing  
Price  
**125\$**



# Data Mining: Most Frequent Amenity Sets

Data Samples Used

Paris



Below Average  
Price

Above Average  
Price

All Listings

Global



Below Average  
Price

All Listings

Above Average  
Price

# Data Mining: Most Frequent Amenity Sets

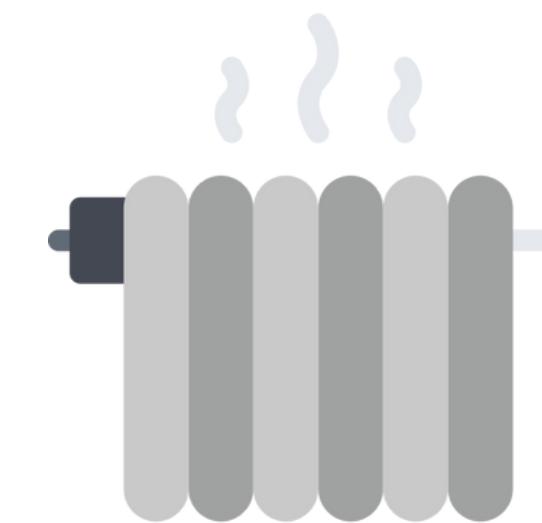
## Apriori Algorithm

Amenity Sets with highest support  
contain all the combinations of:



Wifi

Kitchen



Heating

Essentials



# Data Mining: Most Frequent Amenity Sets

Rating Difference without Amenity Sets

## Paris Below Average Price

<b>Essentials, Wifi</b>	<b>-1.91%</b>
<b>Essentials, Heating, Wifi</b>	<b>-1.75%</b>
<b>Heating, Essentials</b>	<b>-1.72%</b>
<b>Kitchen, Essentials, Wifi</b>	<b>-1.60%</b>

## Paris Above Average Price

<b>Kitchen, Wifi</b>	<b>-1.78%</b>
<b>Kitchen, Essentials, Wifi</b>	<b>-1.64%</b>
<b>Essentials, Wifi</b>	<b>-1.63%</b>
<b>Kitchen, Essentials</b>	<b>-1.54%</b>



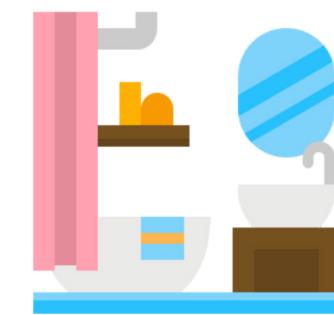
# Data Mining: Price Prediction



# Data Mining: Price Prediction



Linear Regression



Bathrooms

Features:



Location



No. of guests



Property type



Room type



Bedrooms



# Data Mining: Price Prediction

Example:

Location: **Barcelona**

Property type: **Apartment**

No of Guests: **4**

Room type: **Shared**

Bathrooms: **2**

Bedrooms: **2**



Predicted Price:  
**66 EUR/per night**



# Data Mining: Price Prediction

Example:

Location: **Barcelona**

Property type: **Apartment**

No of Guests: **4**

Room type: **Private**

Bathrooms: **2**

Bedrooms: **2**



Predicted Price:  
**136.25 EUR/per night**



# Data Mining: Price Prediction

## 2nd Example:

**Location:** San Francisco

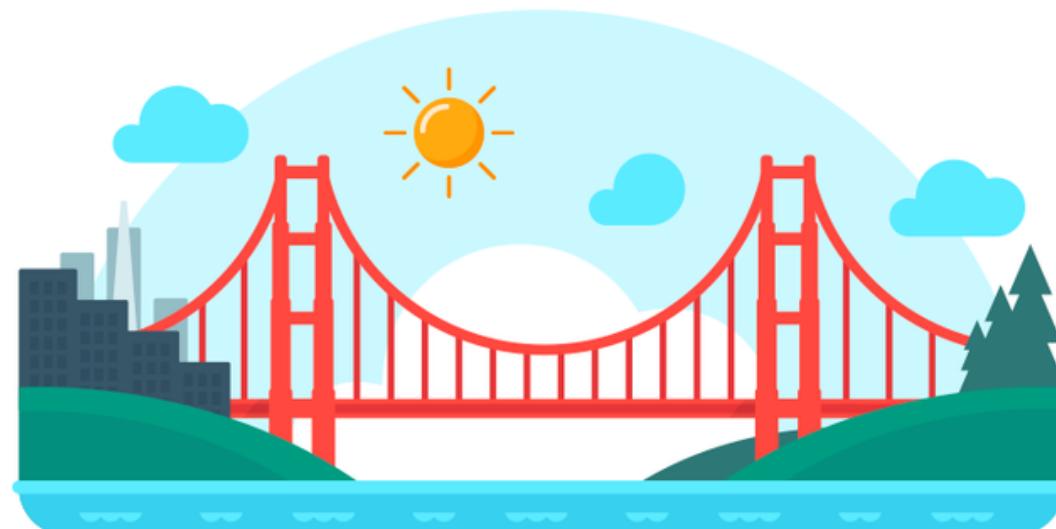
**Property type:** House

**No of Guests:** 8

**Room type:** Entire House

**Bathrooms:** 3

**Bedrooms:** 4



**Predicted Price:**  
**341.81 EUR/per night**



# Thank You!

Bakaros Anastasios

8190111 - t8190111@aueb.gr

Platis Eleftherios

8190144 - t8190144@aueb.gr

**Business Intelligence and Big Data Management**

7th Semester

Software and Data Analysis Technologies

Team 17