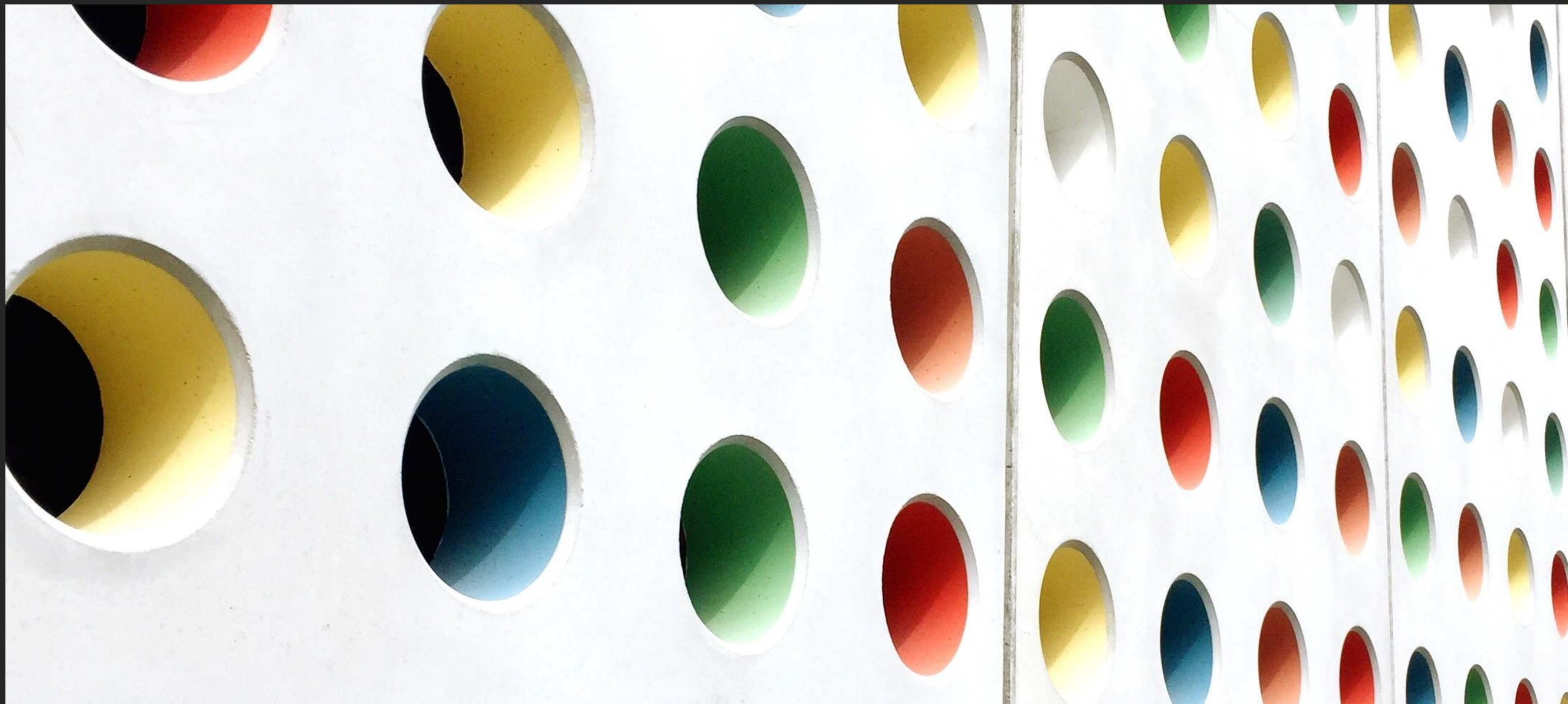
Airbnb - Europe

BY: SEAN & TES



DATA SET SELECTED:

Airbnb-Europe –Dataset(That includes Airbnb properties in(Amsterdam,Athens, Berlin, Barcelona, Rome, Budapest, Paris, Vienna & Lisbon)

(Link: https://www.kaggle.com/datasets/thedevastator/airbnb-prices-in-european-cities)

Project Description & Goals

PROJECT DESCRIPTION

- Analysis on Airbnb -Europe dataset, begin by identifying valuable questions and then to solve the questions.

PROJECT GOAL

- Using EDA & Statistical Modeling, produce an answer or solution to an identified question or problem.
- Communicating Insights using Impactful Dashboards to help stakeholders make decisions.

Project Flow Structures

find dataset & identify problem

EDA & statistical modeling

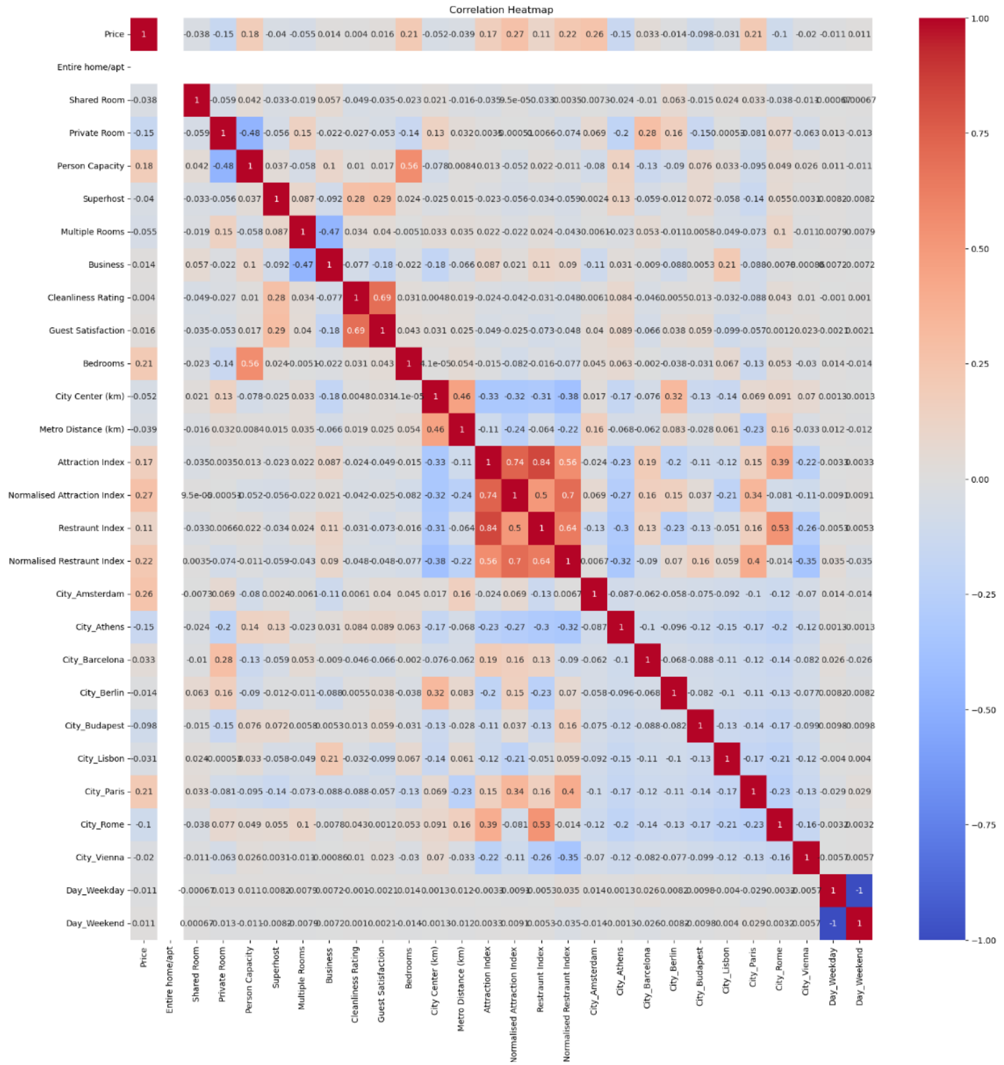
Create visualization & dashboard to communicate insights

PROBLEM IDENTIFIED

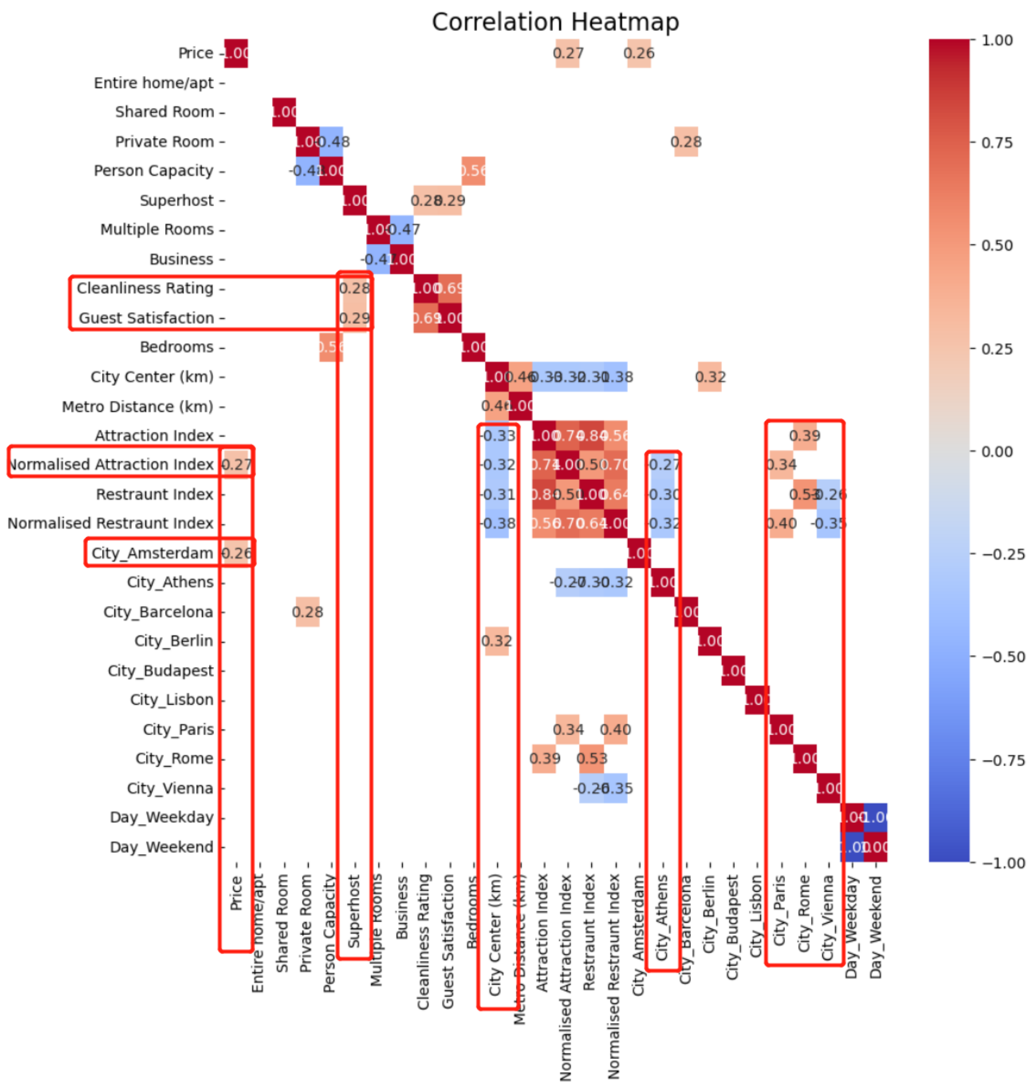
How important are each features to influence the Airbnb property:

1. Price
2. Guest Satisfaction
3. Index

prediction across the 9 European cities?



Using Heatmap



Only the parts of the heatmap with correlation coefficients greater than 0.25 in absolute value were extracted. Here are some pieces of information that can be observed from the heatmap:

1.The city of Amsterdam has the largest positive impact on prices in terms of the attraction index.

2.Superhosts were positively correlated with both cleanliness rating and customer satisfaction in the past.

3.The farther away from the city, the lower all the indices are.

4.The city of Athens has relatively low values for all indices. The restaurant index is also low in the city of Vienna. Both the cities of Paris and Rome show positive correlations with two indices.

Findings @ EDA Level

- Negative correlation between indexes and the distance to the city center and the subway station. We naturally prefer shorter distances.
- There is a significant positive correlation between indexes and price.
- However, it is harder to comprehend the negative correlation between indexes and Guest Satisfaction and Cleanliness Rating, which seems unreasonable for an attraction.

From the results of the regression analysis, the two factors that have the most significant impact on prices are:

1.City

2.Room type

OLS Regression Results

Dep. Variable:	Price	R-squared:	0.260
Model:	OLS	Adj. R-squared:	0.260
Method:	Least Squares	F-statistic:	611.2
Date:	Thu, 13 Jul 2023	Prob (F-statistic):	0.00
Time:	10:22:06	Log-Likelihood:	-2.8786e+05
No. Observations:	41714	AIC:	5.758e+05
Df Residuals:	41689	BIC:	5.760e+05
Df Model:	24		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Entire home/apt	-28.0361	9.741	-2.878	0.004	-47.129	-8.943
Shared Room	-192.1430	13.748	-13.976	0.000	-219.089	-165.197
Private Room	-86.1138	3.228	-26.677	0.000	-92.441	-79.787
Person Capacity	21.7643	1.278	17.033	0.000	19.260	24.269
Superhost	-1.6264	2.813	-0.578	0.563	-7.141	3.888
Multiple Rooms	11.8817	3.034	3.916	0.000	5.935	17.828
Business	40.1648	3.080	13.039	0.000	34.127	46.202
Cleanliness Rating	0.4696	0.149	3.157	0.002	0.178	0.761
Guest Satisfaction	0.7409	0.207	3.575	0.000	0.335	1.147
Bedrooms	7.1052	0.231	30.743	0.000	6.652	7.558
City Center (km)	0.3332	0.941	0.354	0.723	-1.511	2.177
Metro Distance (km)	0.8045	2.076	0.388	0.698	-3.264	4.873
Attraction Index	-0.0486	0.019	-2.616	0.009	-0.085	-0.012
Normalised Attraction Index	5.6468	0.467	12.102	0.000	4.732	6.561
Restraunt Index	-0.0049	0.008	-0.634	0.526	-0.020	0.010
Normalised Restraunt Index	0.8872	0.181	4.895	0.000	0.532	1.242
City_Amsterdam	291.3692	5.379	54.166	0.000	280.826	301.913
City_Athens	-136.8535	4.192	-32.645	0.000	-145.070	-128.637
City_Barcelona	53.6817	5.408	9.927	0.000	43.083	64.281
City_Berlin	-33.0317	6.899	-4.788	0.000	-46.555	-19.509
City_Budapest	-149.1238	4.371	-34.117	0.000	-157.691	-140.557
City_Lisbon	-44.4023	3.925	-11.312	0.000	-52.096	-36.708
City_Paris	80.6592	4.012	20.106	0.000	72.796	88.522
City_Rome	-64.6669	4.792	-13.496	0.000	-74.058	-55.275
City_Vienna	-25.6679	4.588	-5.595	0.000	-34.660	-16.675
Day_Weekday	-17.9140	5.021	-3.568	0.000	-27.756	-8.072
Day_Weekend	-10.1221	5.006	-2.022	0.043	-19.934	-0.310

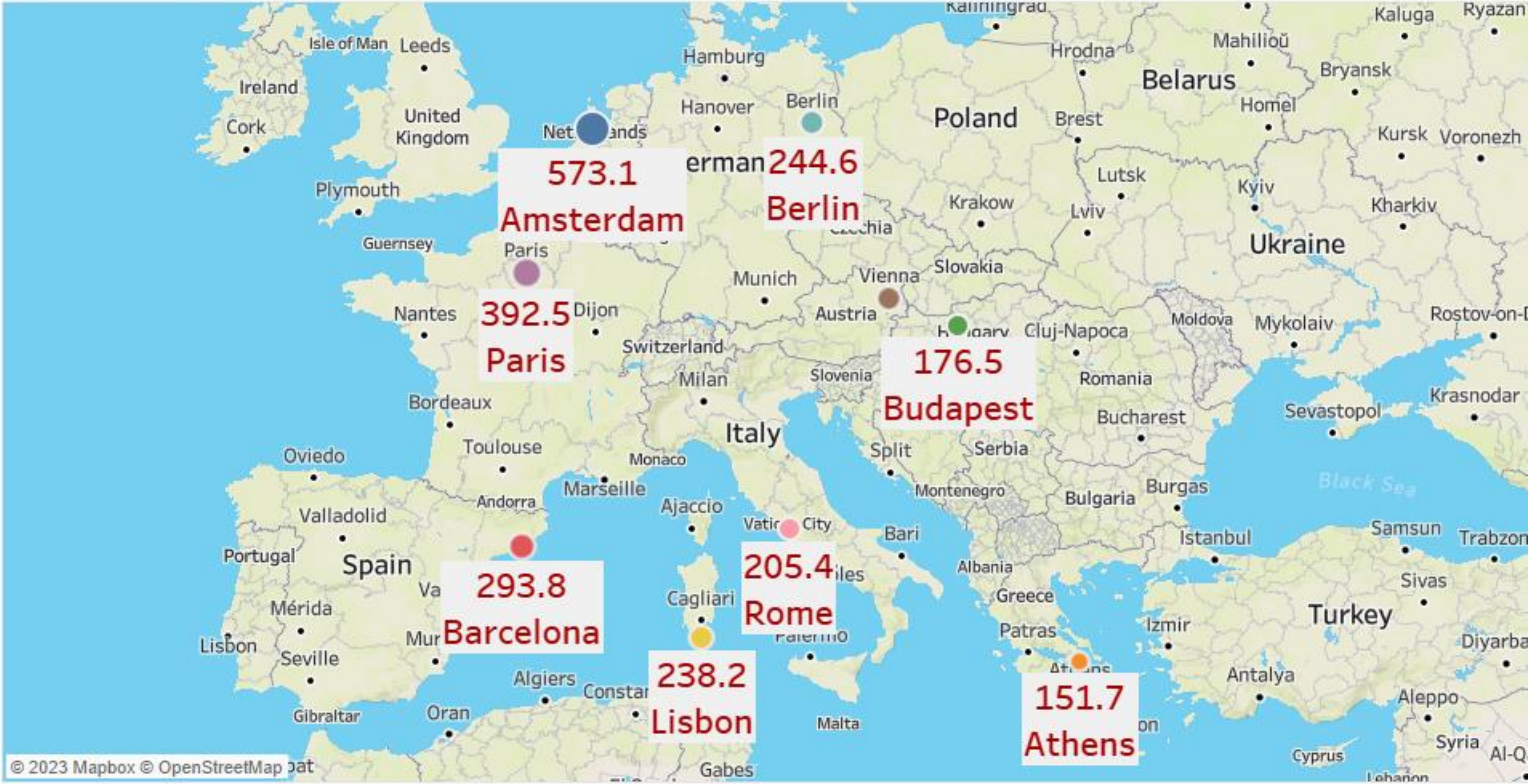
Omnibus:	116159.888	Durbin-Watson:	1.866
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7694417492.001
Skew:	35.828	Prob(JB):	0.00
Kurtosis:	2105.813	Cond. No.	1.27e+16

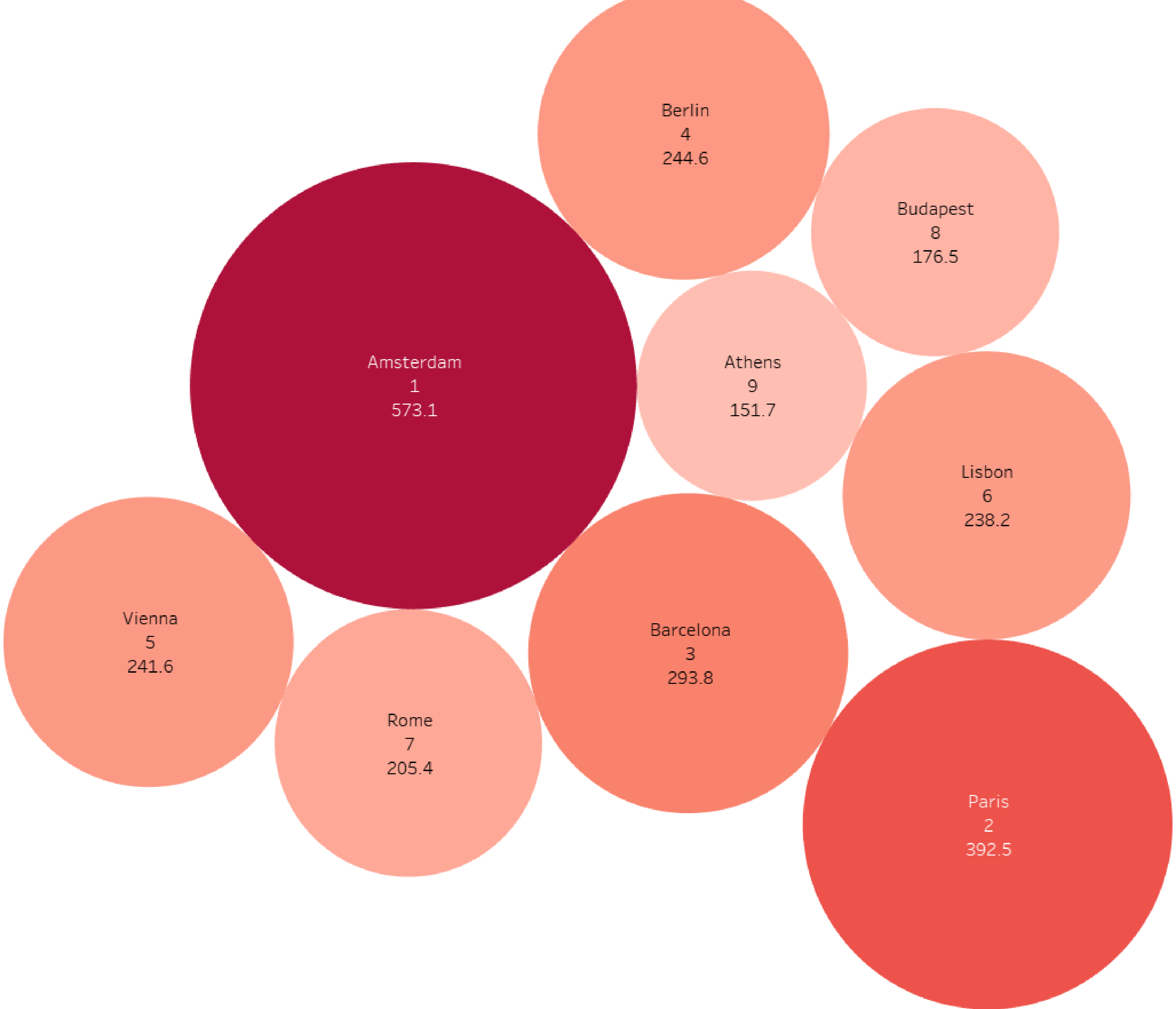
Findings @Modeling level

- R-square indicate approximately 26% of the variation in the Price can be explained by the independent variables included in the model.
- The coefficient -192.143 indicate the negative r/ship and the significant decreases in price when room type is shared room.
- The P-value of most of the variables is less than 0.05, indicating their statistical significance in explaining the variation in price.

Visualization Analysis: Prices

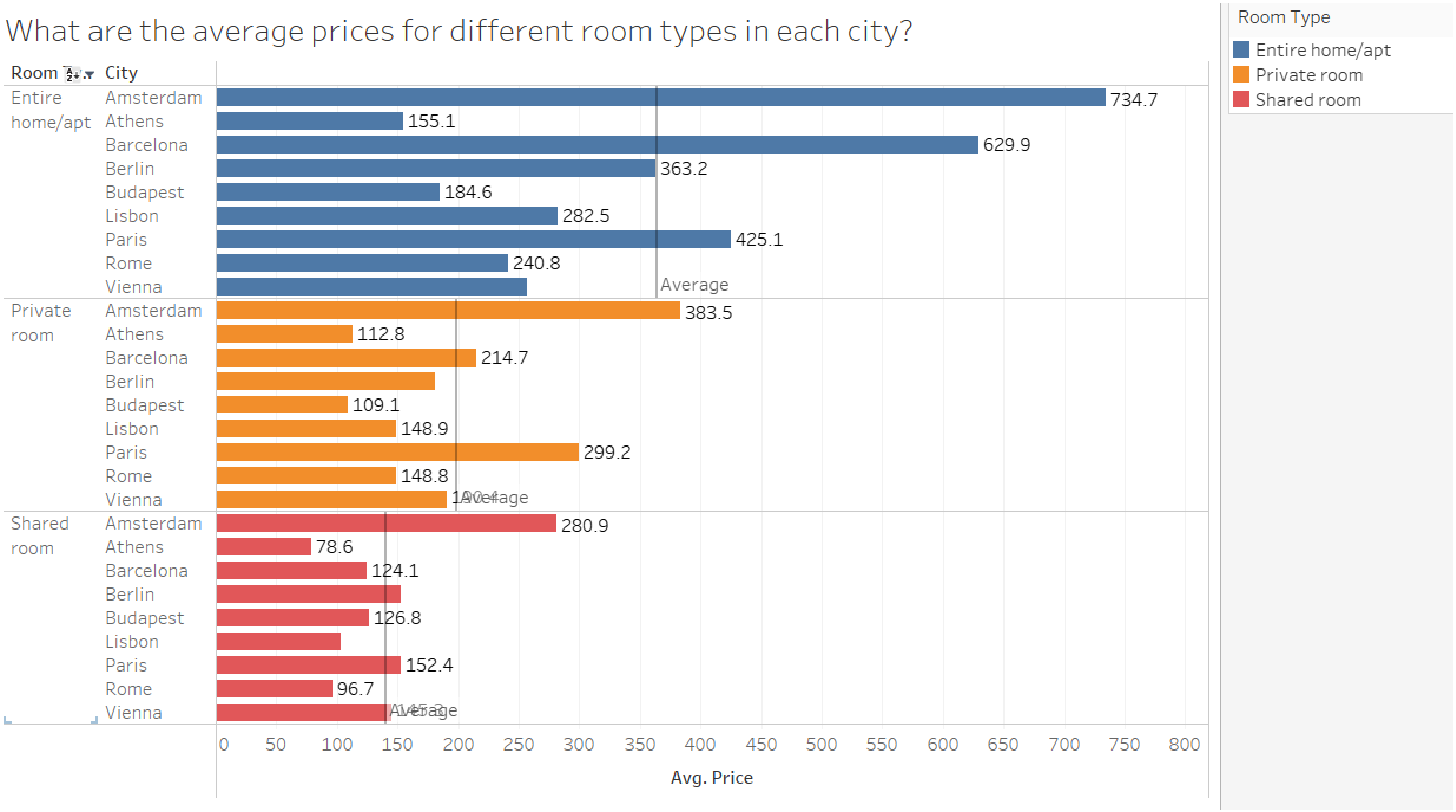
AVG Airbnb Rent by Cities



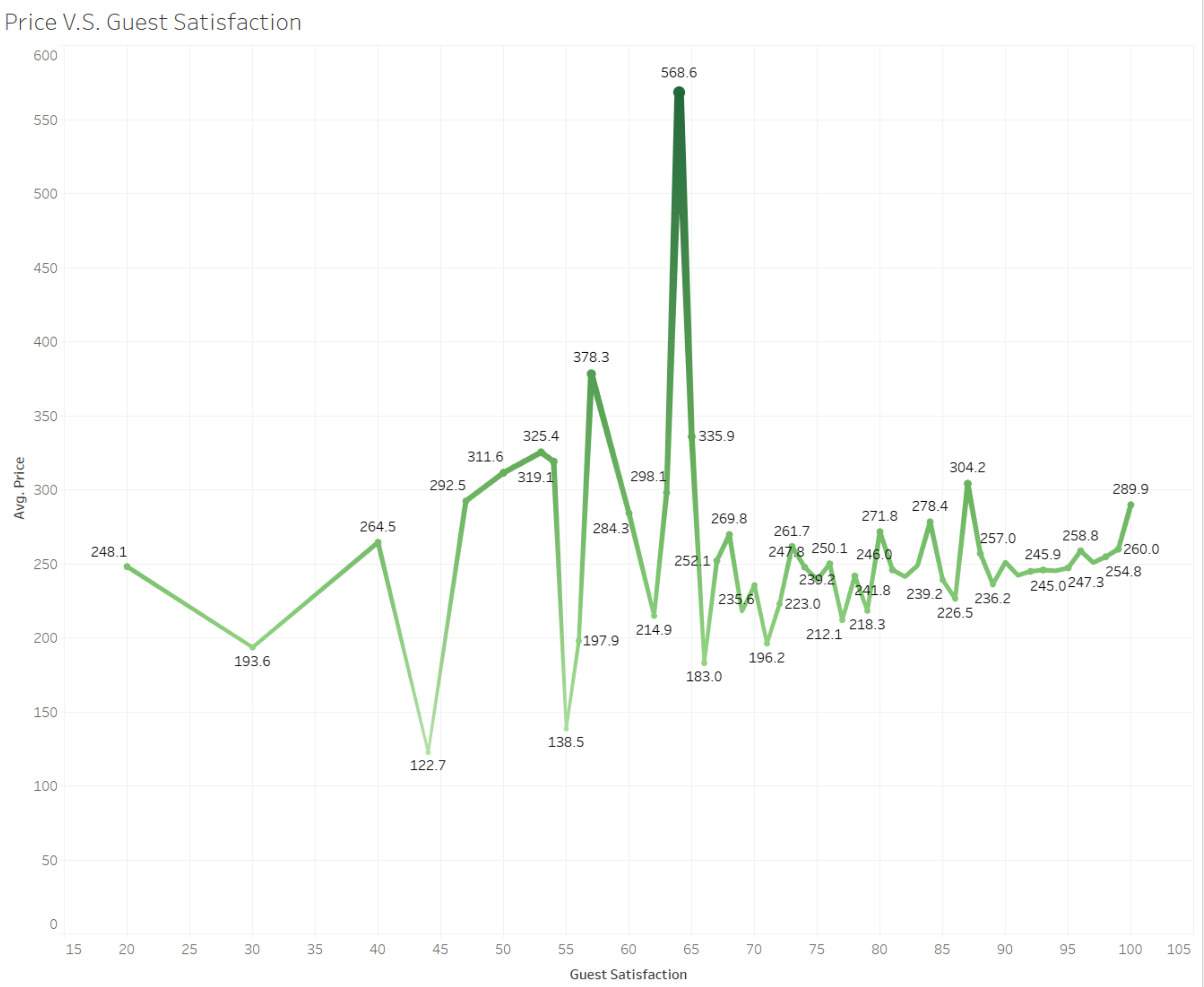


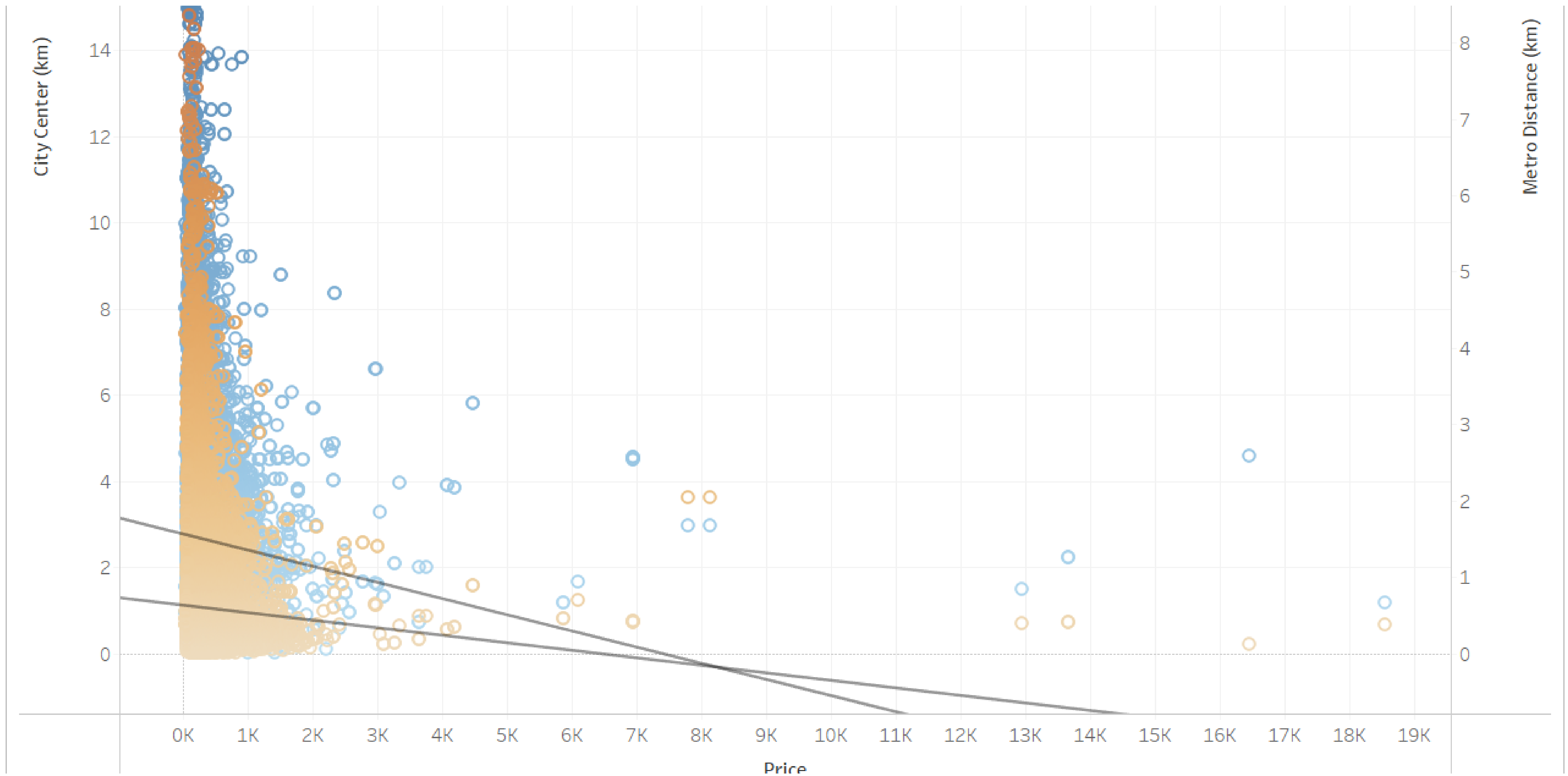
Average Prices and Rankings by City

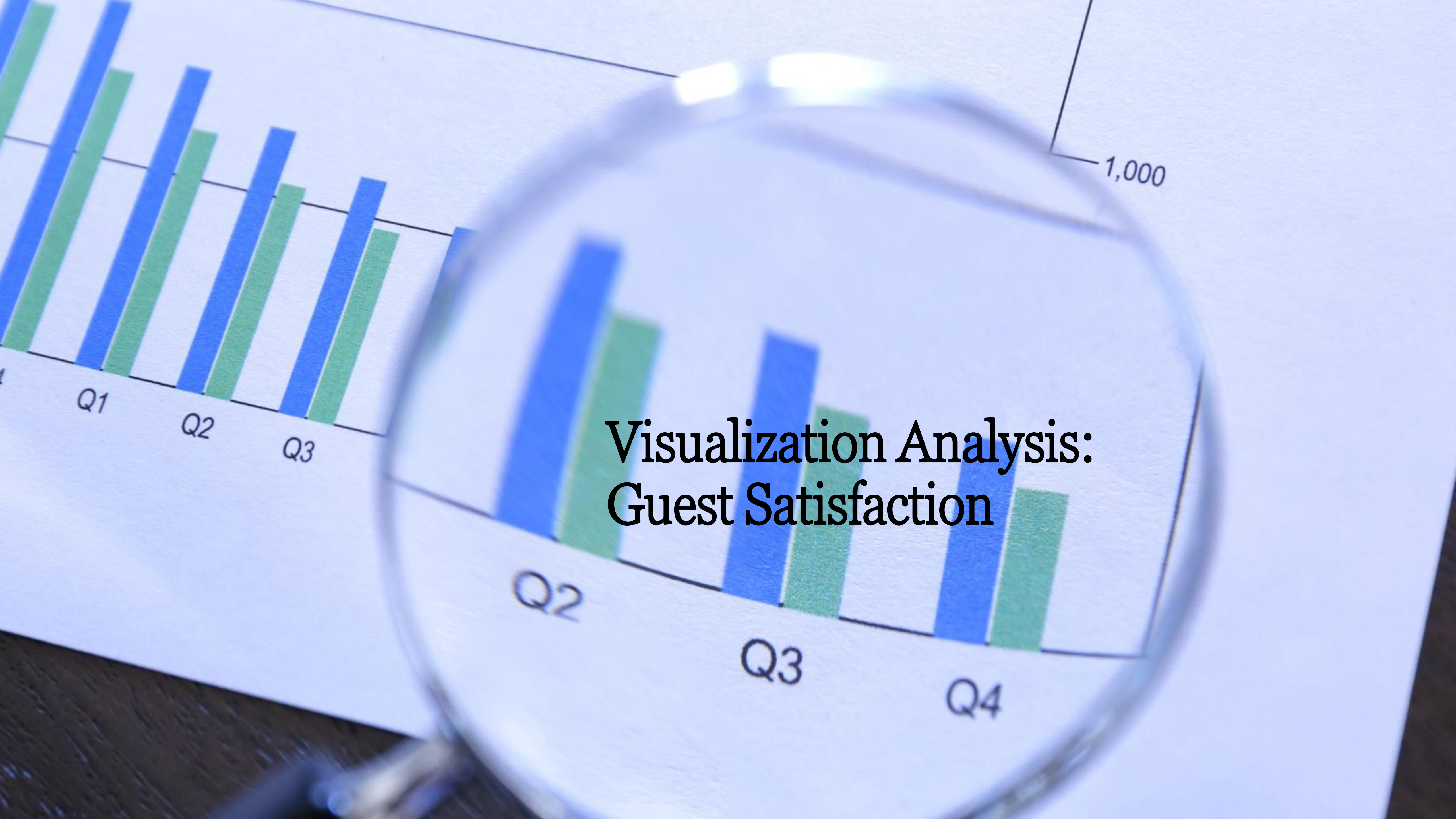
What are the average prices for different room types in each city?



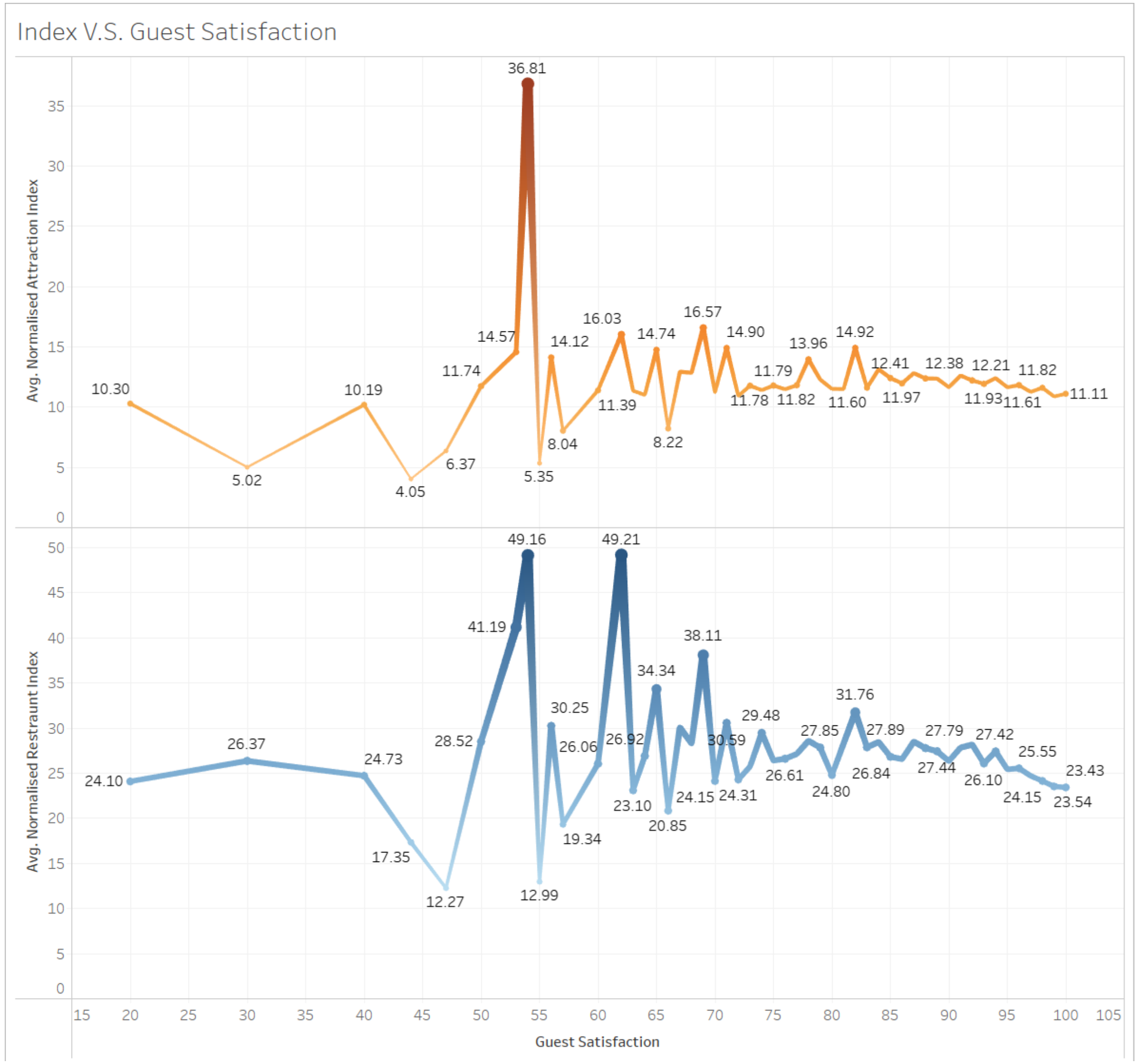
The relationship between customer satisfaction and prices is not clear, but it can be observed that houses with higher average prices tend to have lower overall satisfaction. This could be due to difficulties in meeting customer expectations.

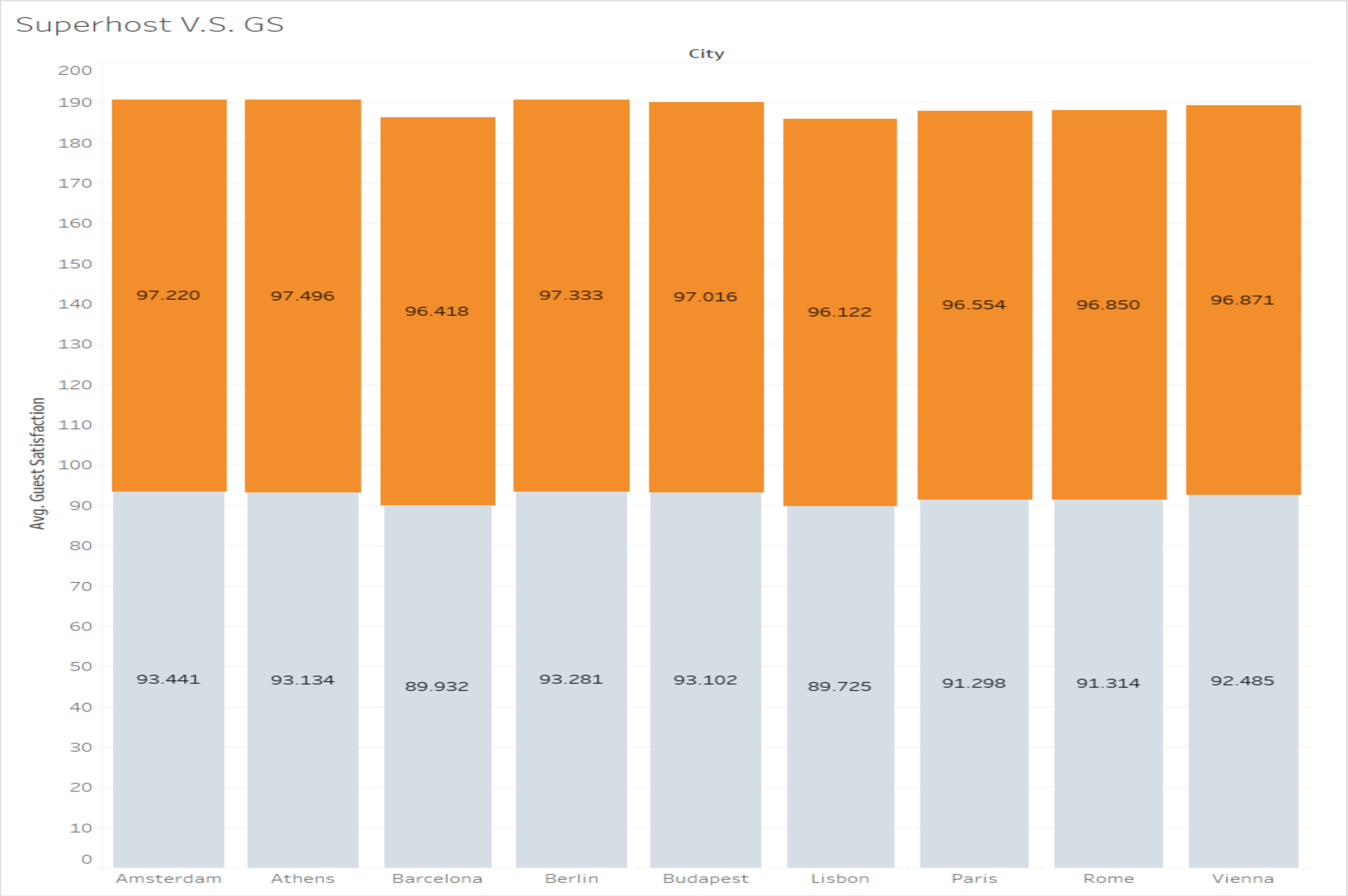



City Center (km)
14
12
10
8
6
4
2
0
Metro Distance (km)
8
7
6
5
4
3
2
1
0
0K
1K
2K
3K
4K
5K
6K
7K
8K
9K
10K
11K
12K
13K
14K
15K
16K
17K
18K
19K
Price



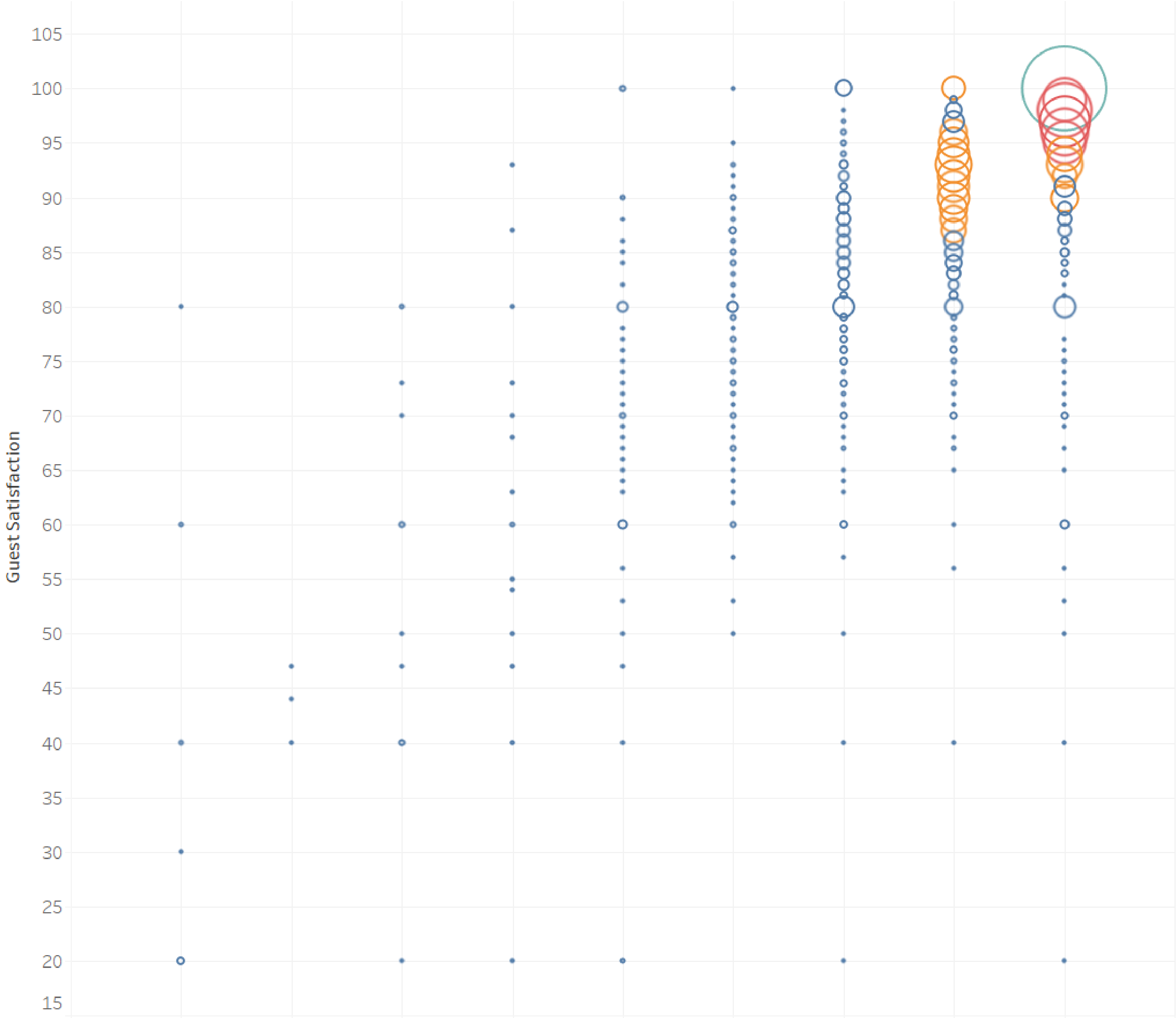
Visualization Analysis: Guest Satisfaction


Index V.S. Guest Satisfaction
Avg. Normalised Attraction Index
Avg. Normalised Restraunt Index
Guest Satisfaction
36.81
10.30
5.02
10.19
4.05
6.37
11.74
14.57
5.35
14.12
8.04
16.03
11.39
14.74
8.22
16.57
14.90
11.78
11.79
11.82
13.96
11.60
14.92
12.41
11.97
12.38
12.21
11.93
11.61
11.82
11.11
49.16
49.21
24.10
26.37
24.73
17.35
12.27
28.52
41.19
12.99
30.25
19.34
26.06
26.92
23.10
34.34
20.85
24.15
38.11
30.59
24.31
29.48
26.61
27.85
24.80
31.76
26.84
27.89
27.44
27.79
26.10
27.42
24.15
25.55
23.54
23.43



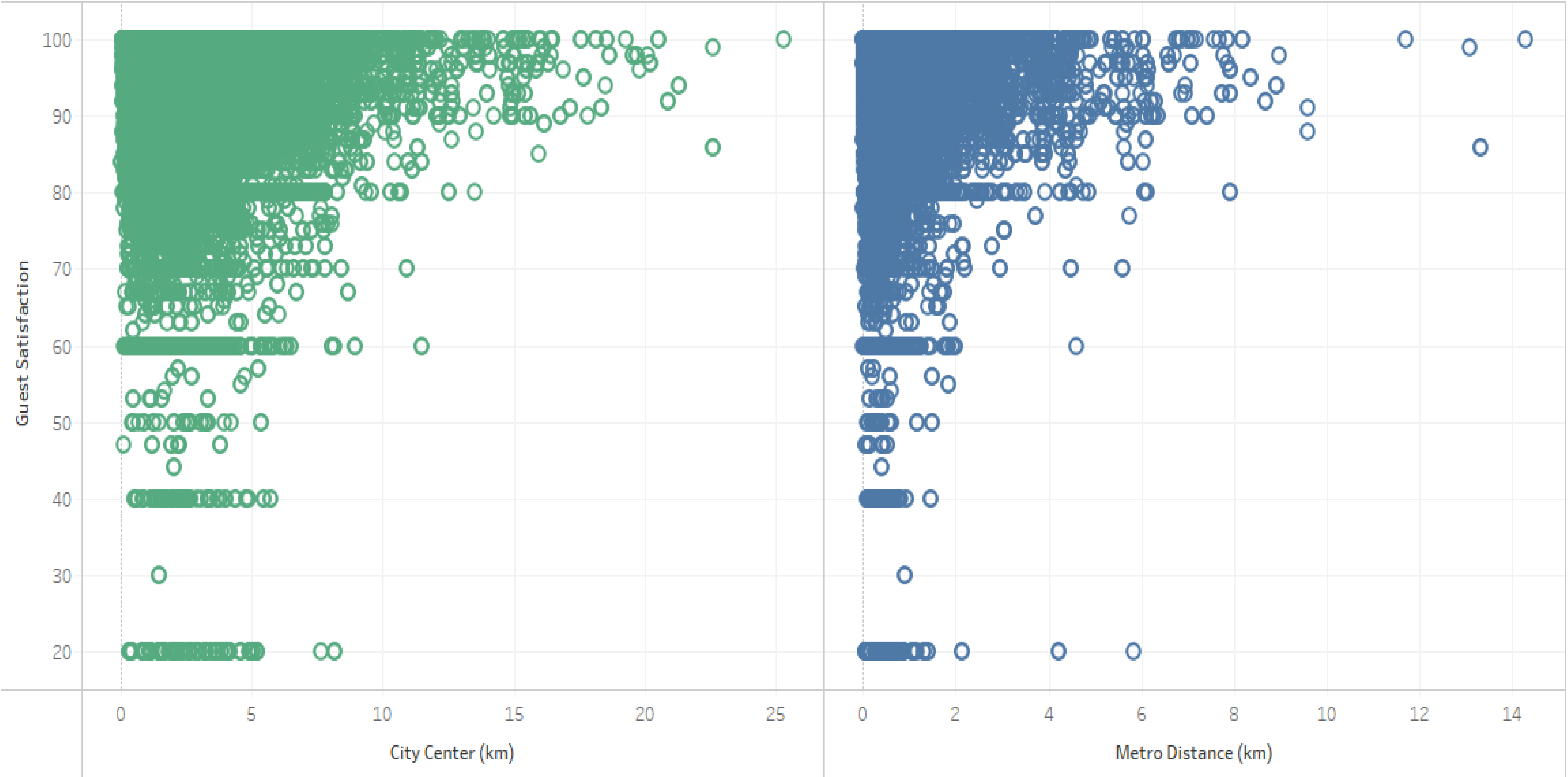
It can be seen that the average guest satisfaction of Superhosts is slightly higher than that of non-Superhosts, but the difference is not significant.

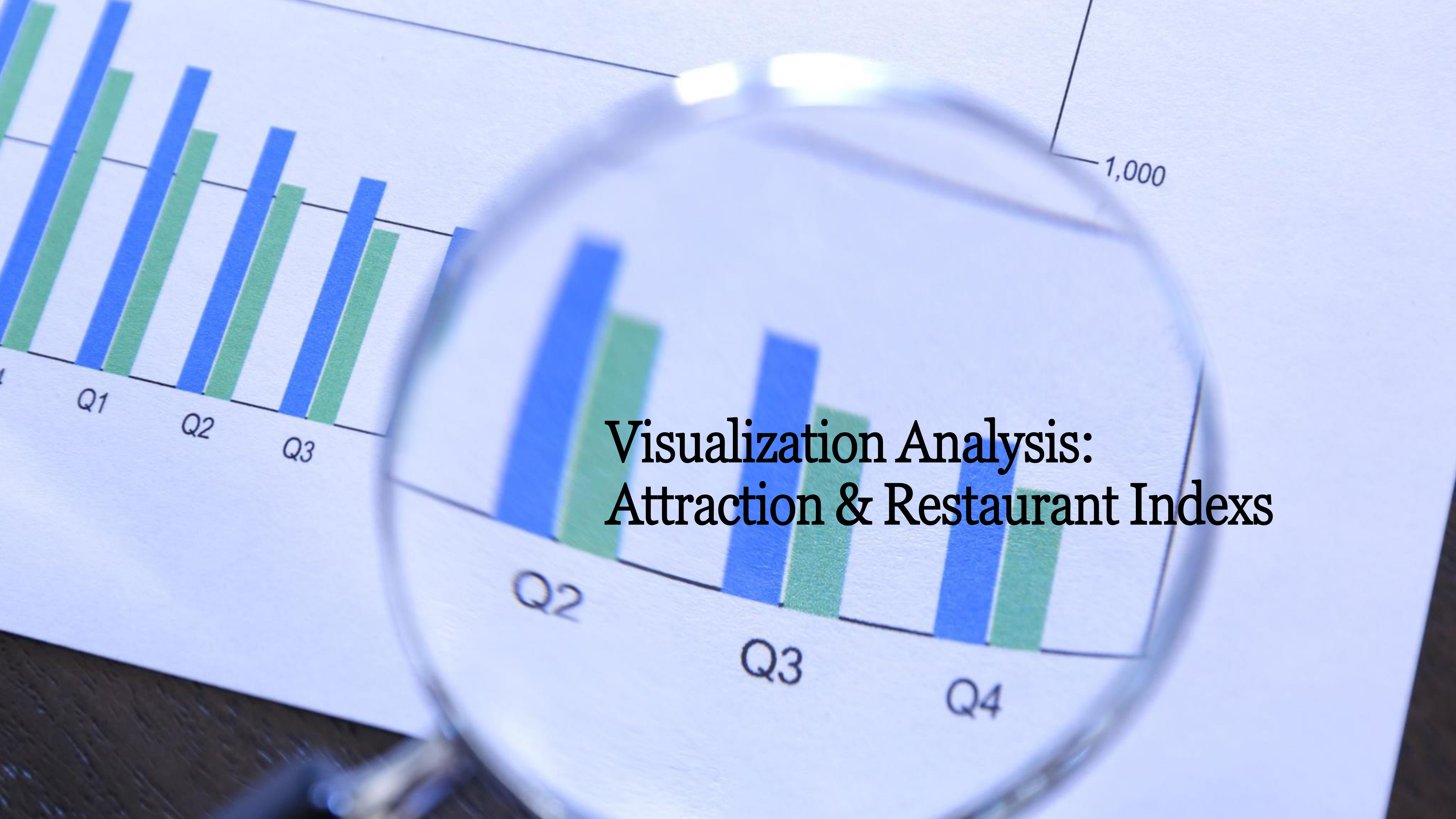
Cleaning V.S. GS



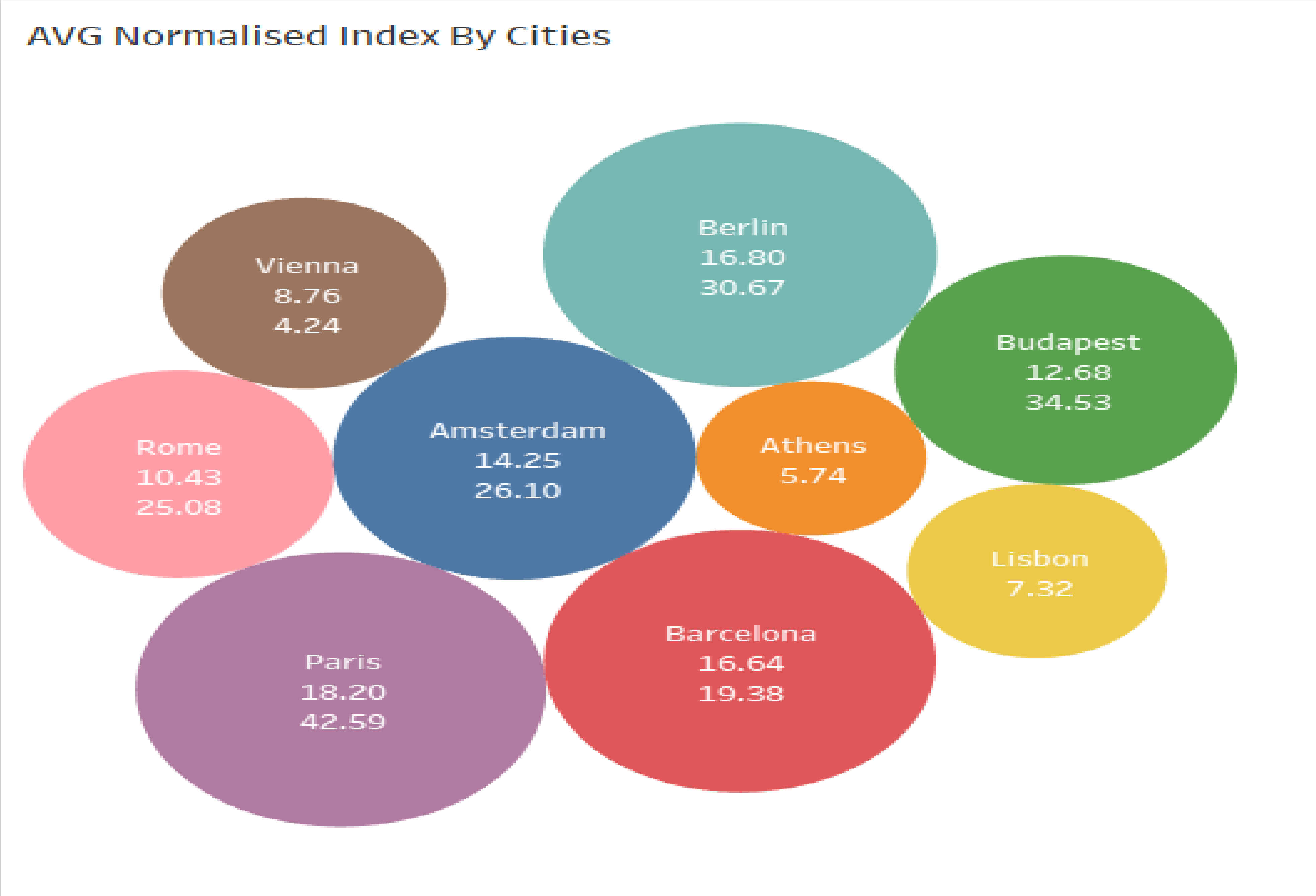
Clearly, higher cleanliness levels lead to higher customer satisfaction.

Distance to City Center/Metro Station VS Guest Satisfaction

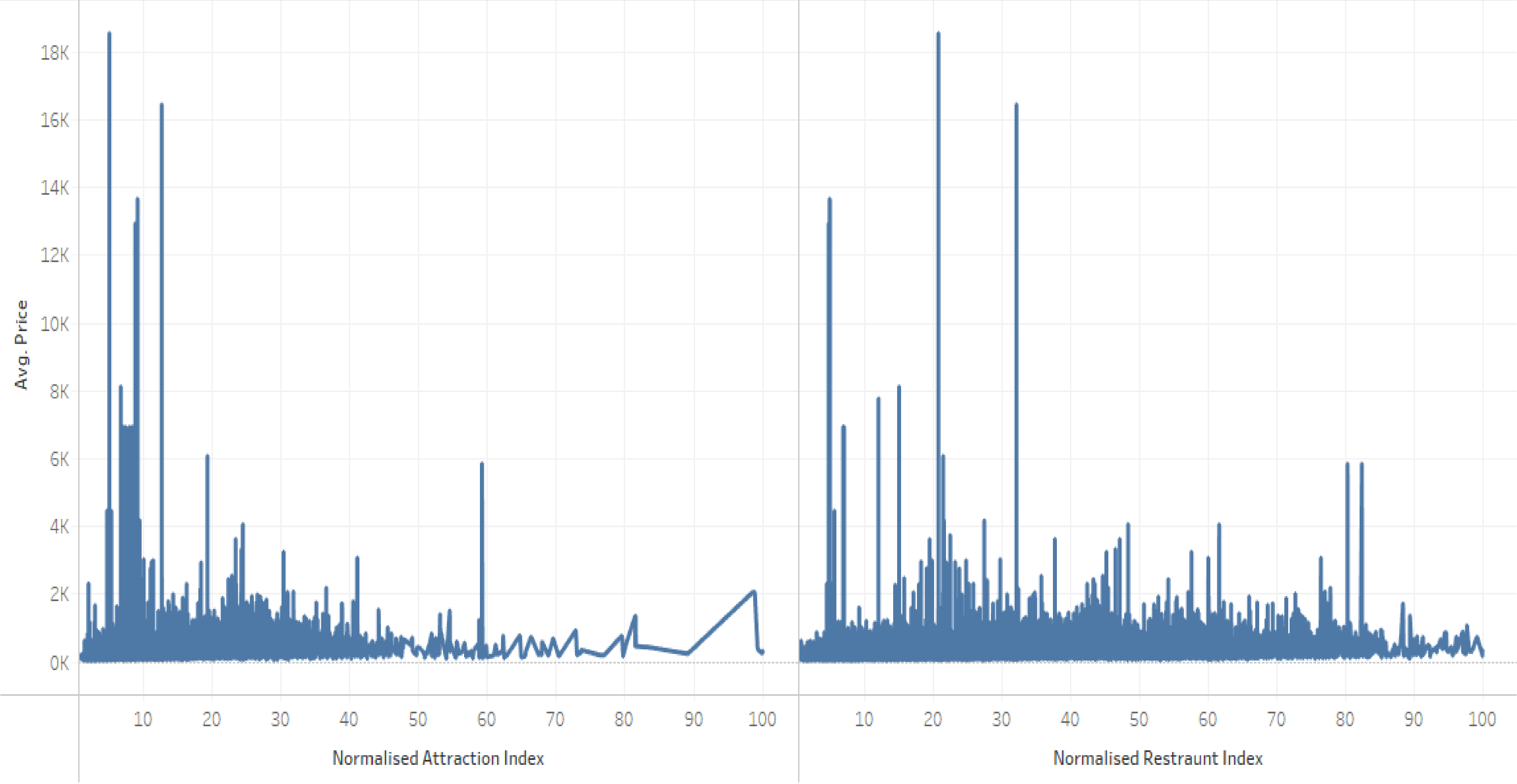


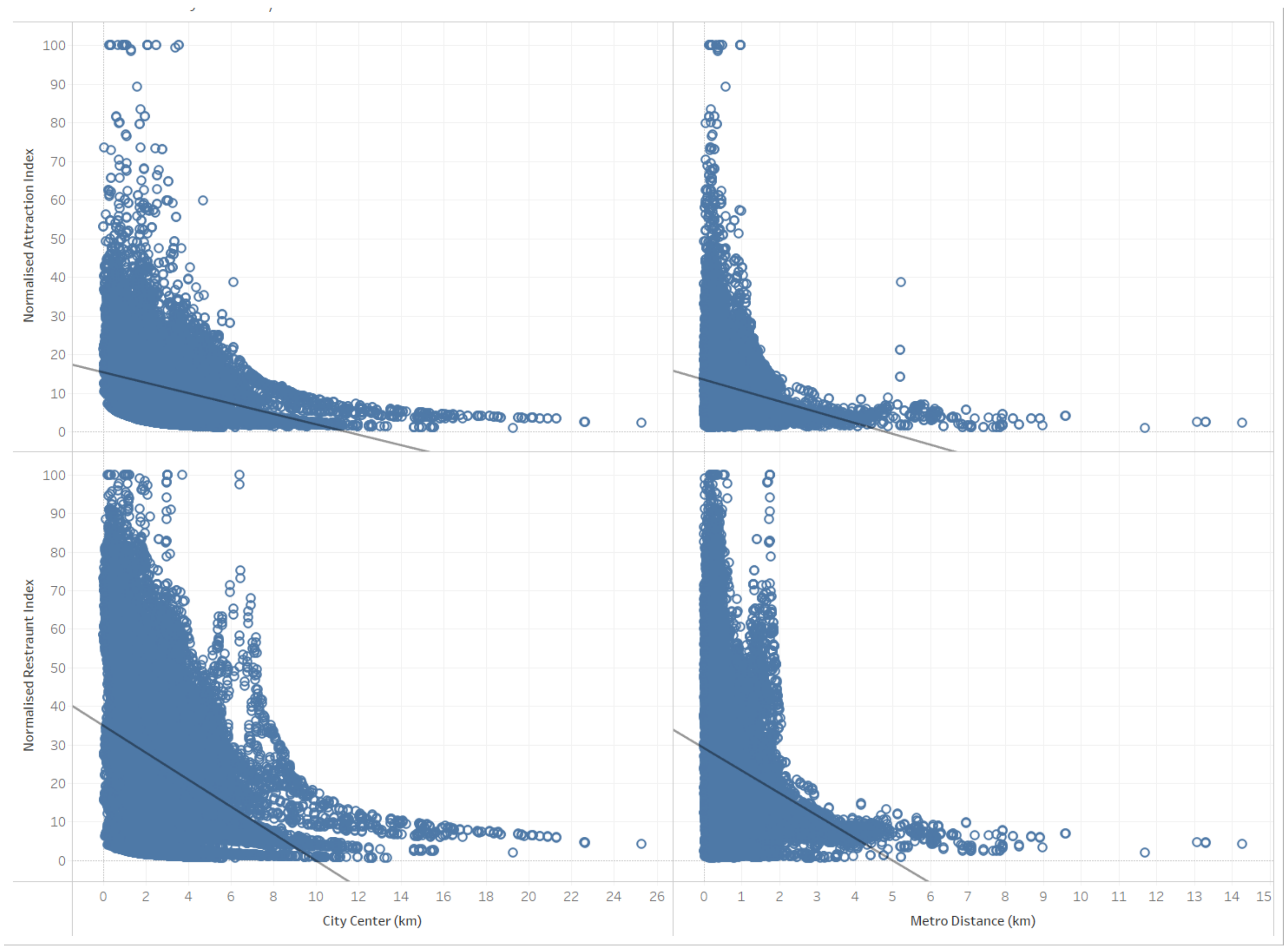


Visualization Analysis: Attraction & Restaurant Indexs

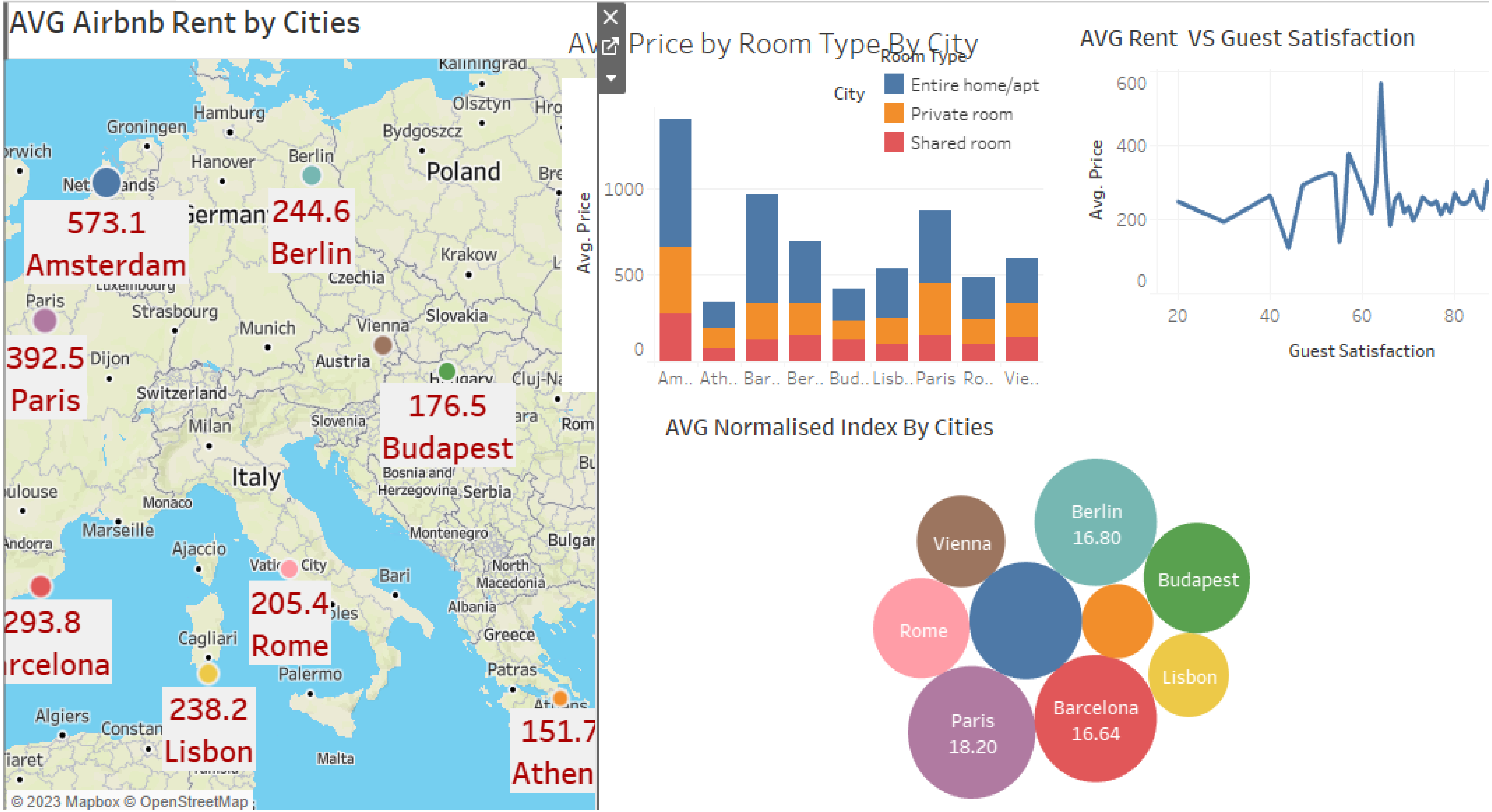

AVG Normalised Index By Cities
Vienna
8.76
4.24
Berlin
16.80
30.67
Budapest
12.68
34.53
Rome
10.43
25.08
Amsterdam
14.25
26.10
Athens
5.74
Lisbon
7.32
Paris
18.20
42.59
Barcelona
16.64
19.38

Price by Normalised Indexes

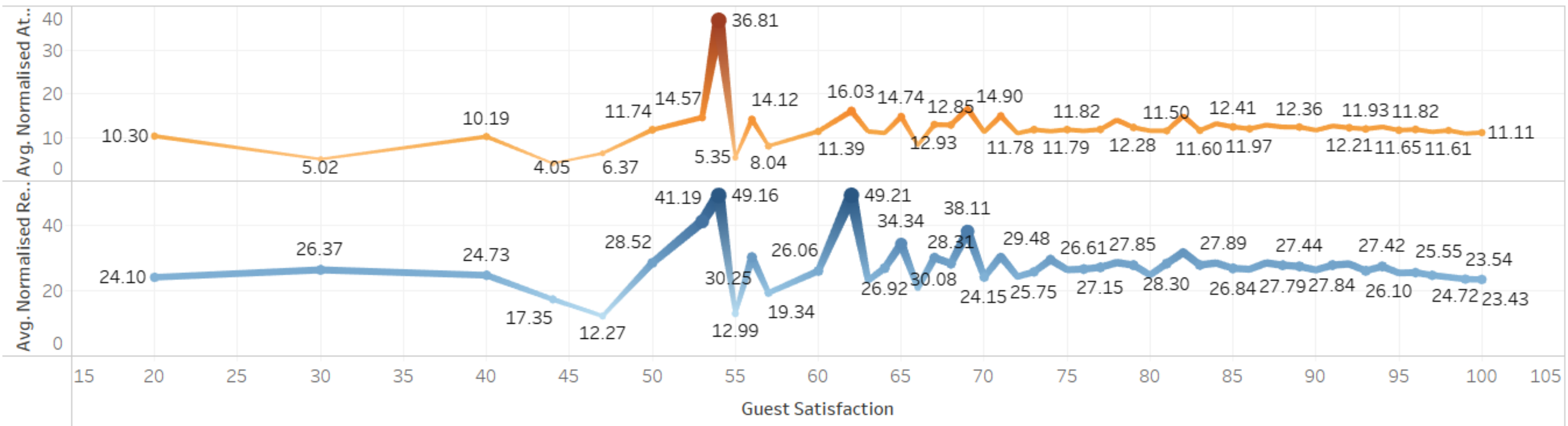




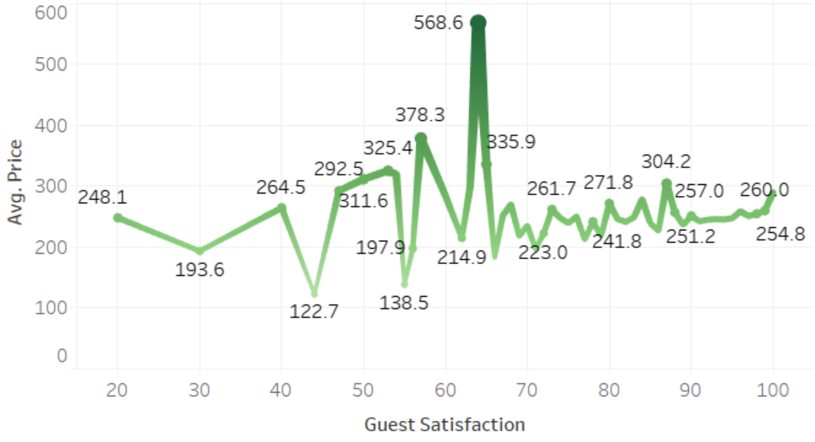
Normalised Attraction Index
Normalised Restraunt Index
City Center (km)
Metro Distance (km)


AVG Airbnb Rent by Cities
573.1
Amsterdam
244.6
Berlin
392.5
Paris
176.5
Budapest
293.8
rcelona
205.4
Rome
238.2
Lisbon
151.7
Athen
© 2023 Mapbox © OpenStreetMap
AVG Price by Room Type By City
Room Type
Entire home/apt
Private room
Shared room
City
Avg. Price
1000
500
0
Am..
Ath..
Bar..
Ber..
Bud..
Lisb..
Paris
Ro..
Vie..
AVG Rent VS Guest Satisfaction
Avg. Price
600
400
200
0
20
40
60
80
Guest Satisfaction
AVG Normalised Index By Cities
Berlin
16.80
Vienna
Budapest
Rome
Lisbon
Paris
18.20
Barcelona
16.64

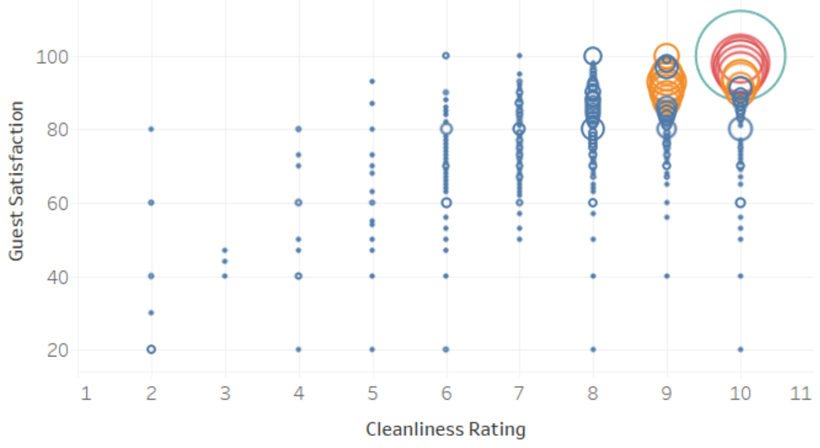
Index V.S. Guest Satisfaction



Price V.S. Guest Satisfaction



Cleaning V.S. GS



Challenges

- Navigating how to work with GitHub as a team member when you must work on the same notebook was not as easy.
- Creating robust and highly interactive dashboard needs time & practice.
- If we had time, we would have tried creating classification(logistic model) specially with those columns with categorical data type.

THANAK YOU