



82.05 - Análisis Predictivo

29/03/2023

CASO DE ESTUDIO

PREDICCIÓN DE PRECIO DE ALOJAMIENTO

—

01 Introducción

Composición y preparación de la base, desafíos y objetivos del trabajo.

03 Análisis Geográfico

Utilización de un mapa para analizar interactivamente algunos insights.

02 Análisis Exploratorio

Presentación gráfica y analítica de los datos. Se busca proporcionar un mayor conocimiento de las variables

04 Conclusiones

Conclusiones que se obtienen para el desarrollo del modelo y la implementación del caso de negocios.

La base contiene información sobre alojamientos publicados en Airbnb de **10** capitales europeas. Fueron **20** las bases unidas. En total, **51.707** registros con **17** variables.

- realSum: Precio de listado en la plataforma Airbnb para dos noches y dos personas en EUR. (Numeric). Esta va a ser la **variable a predecir**.
- room_type: Tipo de habitación (e.g. private, shared, etc.). (Categorical)
- person_capacity: Máxima cantidad de personas que se pueden alojar en el lugar. (Numeric)
- host_is_superhost: Si el host es o no un superhost. (Boolean)
- multi: Si el host de la publicación tiene o no de 2 a 4 publicaciones. (Boolean)
- biz: Si el host de la publicación tiene o no más de 4 publicaciones. (Boolean)
- cleanliness_rating: Calificación de limpieza. (Numeric)
- guest_satisfaction_overall: Calificación media de visitantes pasados. (Numeric)
- bedrooms: Número de habitaciones en el alojamiento. (Numeric)
- dist: Distancia hacia el centro de la ciudad en kilómetros. (Numeric)
- metro_dist: Distancia hacia la estación de metro más cercana en kilómetros. (Numeric)
- attr_index: Índice de atracción de la locación del alojamiento. (Numeric)
- rest_index: Índice de restaurantes de la locación del alojamiento. (Numeric)
- lng: Longitud. (Numeric)
- lat: Latitud. (Numeric)
- city: donde está ubicado el alojamiento. (Categorical)
- day: día de semana o fin de semana. (Categorical)

PREPARACIÓN DE LA BASE

- Se unen 20 bases en total. Por cada ciudad (Amsterdam, Athens, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Rome y Vienna) hay una base con datos de publicaciones para los *días de la semana* y otra con datos para los *fines de semana*.
- Las variables categóricas **'day'** y **'city'** son creadas al unir las bases, ya que van a ayudar a diferenciar de qué ciudad es la publicación y si es día de semana o fin de semana.
- Se han ~~eliminado~~ columnas poco relevantes para el estudio.
- Se ha verificado que las variables no contienen nulos.
- Se ha verificado que la base no contiene duplicados.

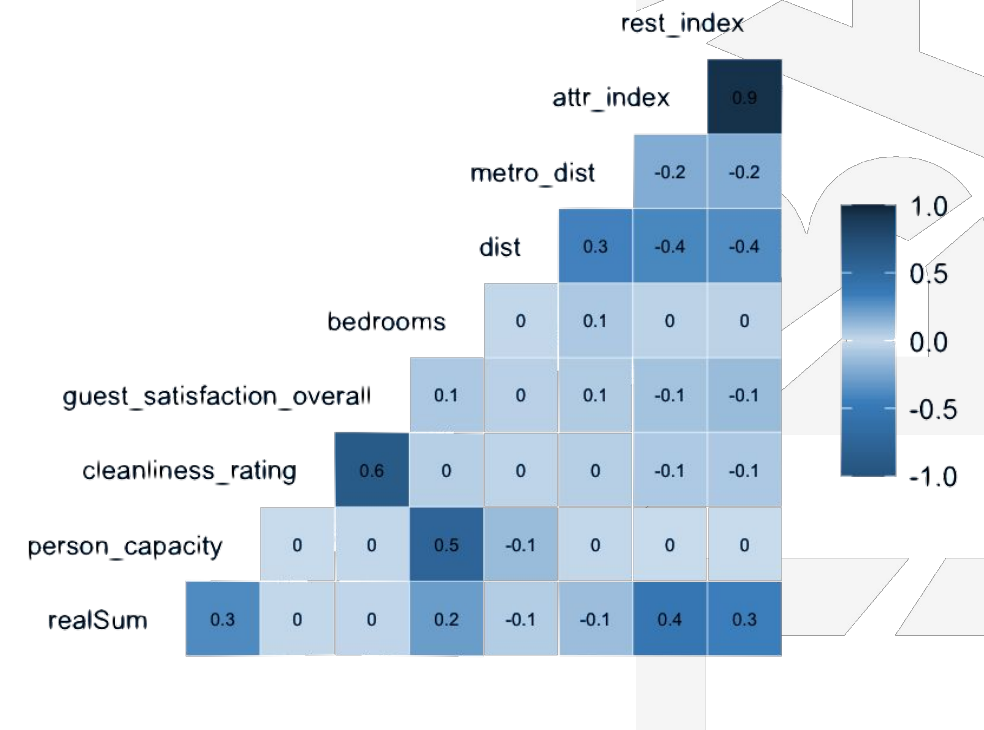
OBJETIVO

El objetivo de este caso de estudio será, a partir de los valores de variables ya mencionadas, llegar a entrenar un modelo para poder predecir cuál debería ser el precio de un alojamiento a publicar en Airbnb. Este modelo servirá luego para asesorar a una *compañía que gestiona alojamientos* en 10 capitales europeas a poner el **precio** correcto en los mismos, para obtener el mejor balance posible para ellos y para quienes los contratan.



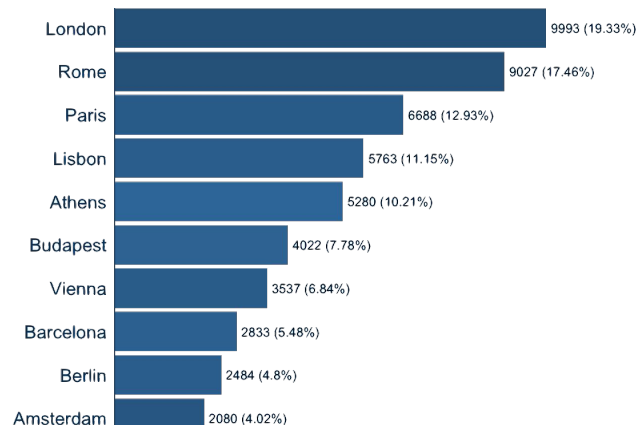
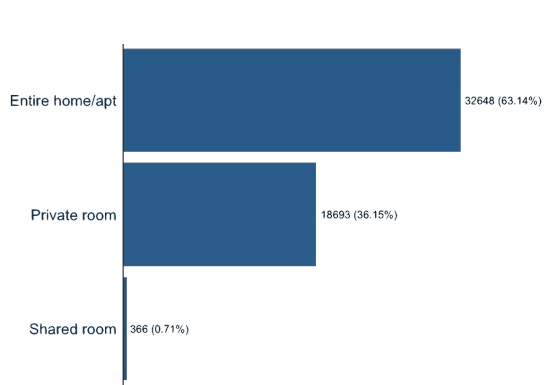
ANÁLISIS EXPLORATORIO: VARIABLES NUMÉRICAS

realSum	person_capacity	cleanliness_rating	guest_satisfaction_overall	bedrooms	dist	metro_dist	attr_index	rest_index
Min. : 34.78	Min. :2.000	Min. : 2.000	Min. : 20.00	Min. : 0.000	Min. : 0.01504	Min. : 0.002301	Min. : 15.15	Min. : 19.58
1st Qu.: 148.75	1st Qu.:2.000	1st Qu.: 9.000	1st Qu.: 90.00	1st Qu.: 1.000	1st Qu.: 1.45314	1st Qu.: 0.248480	1st Qu.: 136.80	1st Qu.: 250.85
Median : 211.34	Median :3.000	Median :10.000	Median : 95.00	Median : 1.000	Median : 2.61354	Median : 0.413269	Median : 234.33	Median : 522.05
Mean : 279.88	Mean :3.162	Mean : 9.391	Mean : 92.63	Mean : 1.159	Mean : 3.19129	Mean : 0.681540	Mean : 294.20	Mean : 626.86
3rd Qu.: 319.69	3rd Qu.:4.000	3rd Qu.:10.000	3rd Qu.: 99.00	3rd Qu.: 1.000	3rd Qu.: 4.26308	3rd Qu.: 0.737840	3rd Qu.: 385.76	3rd Qu.: 832.63
Max. : 18545.45	Max. :6.000	Max. :10.000	Max. :100.00	Max. : 10.000	Max. : 25.28456	Max. : 14.273577	Max. : 4513.56	Max. : 6696.16

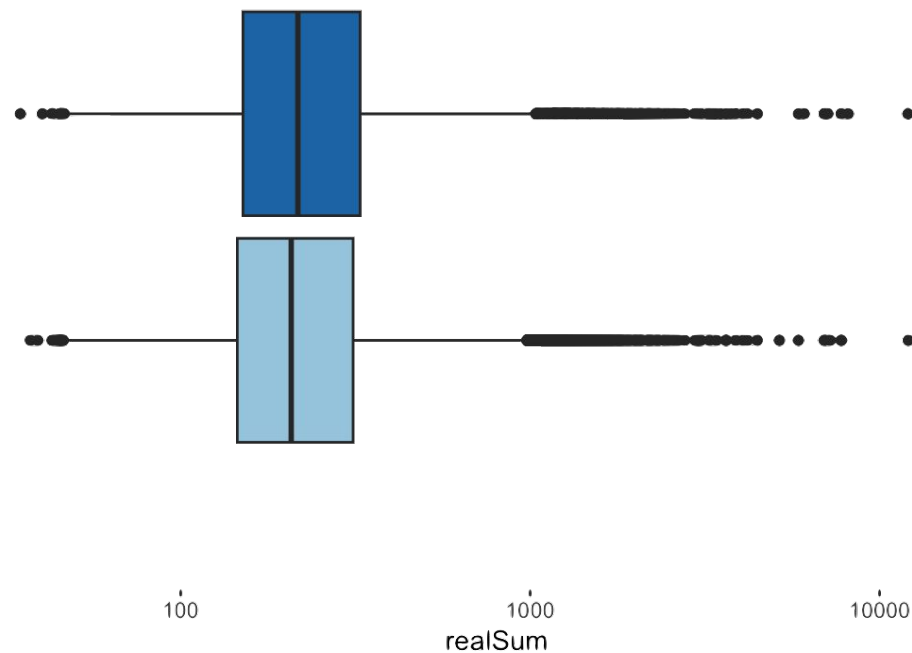
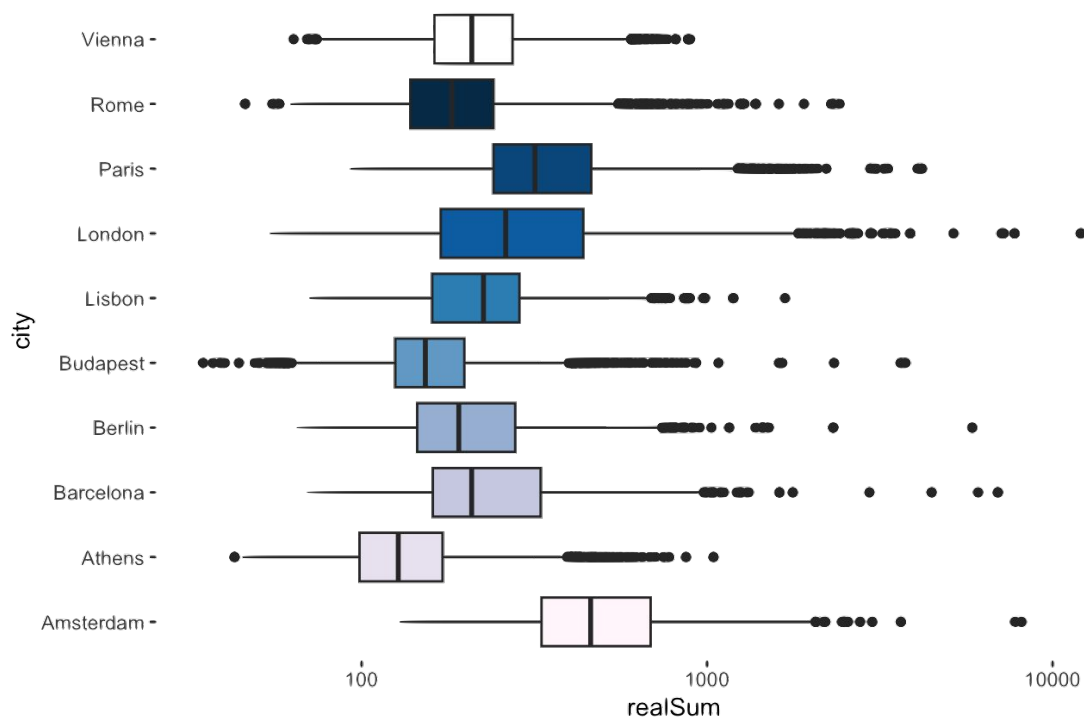


ANÁLISIS EXPLORATORIO: VARIABLES CATEGÓRICAS

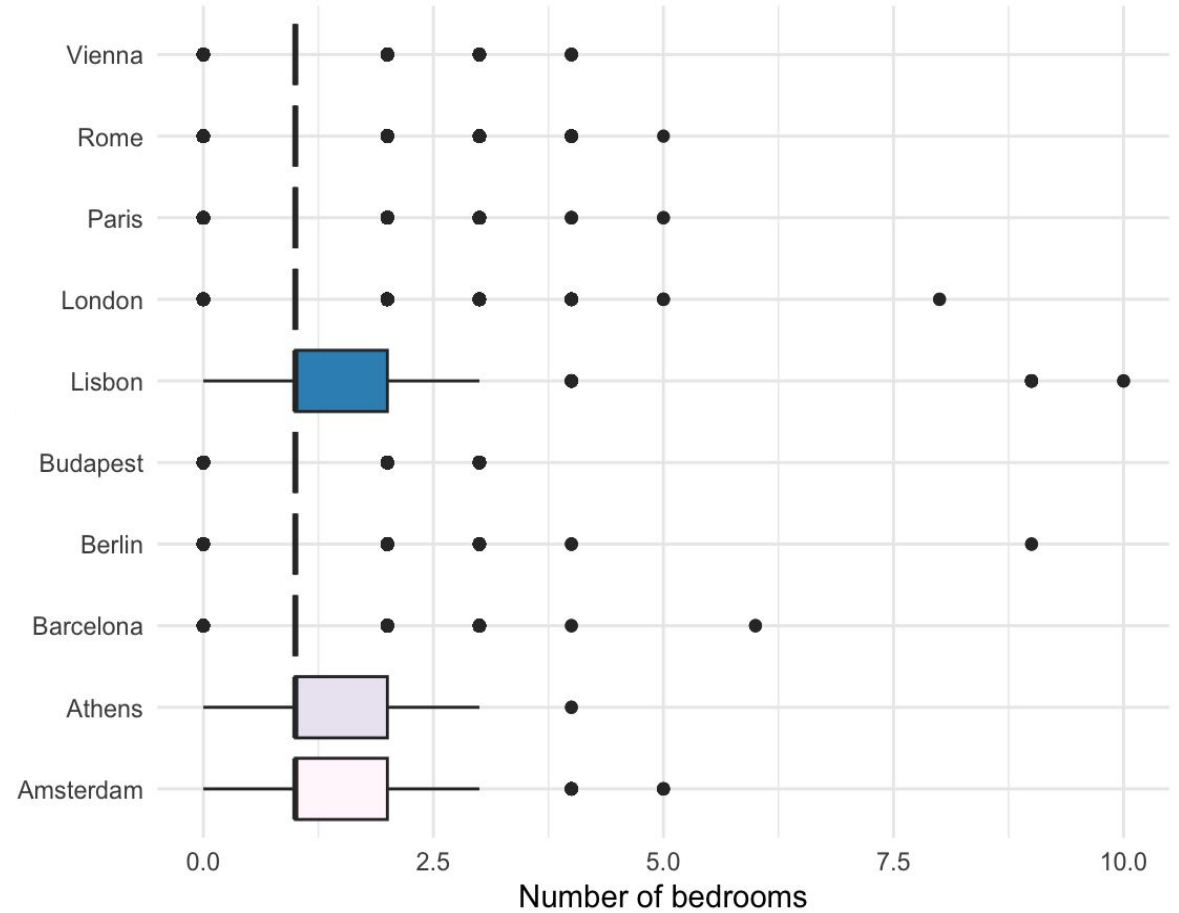
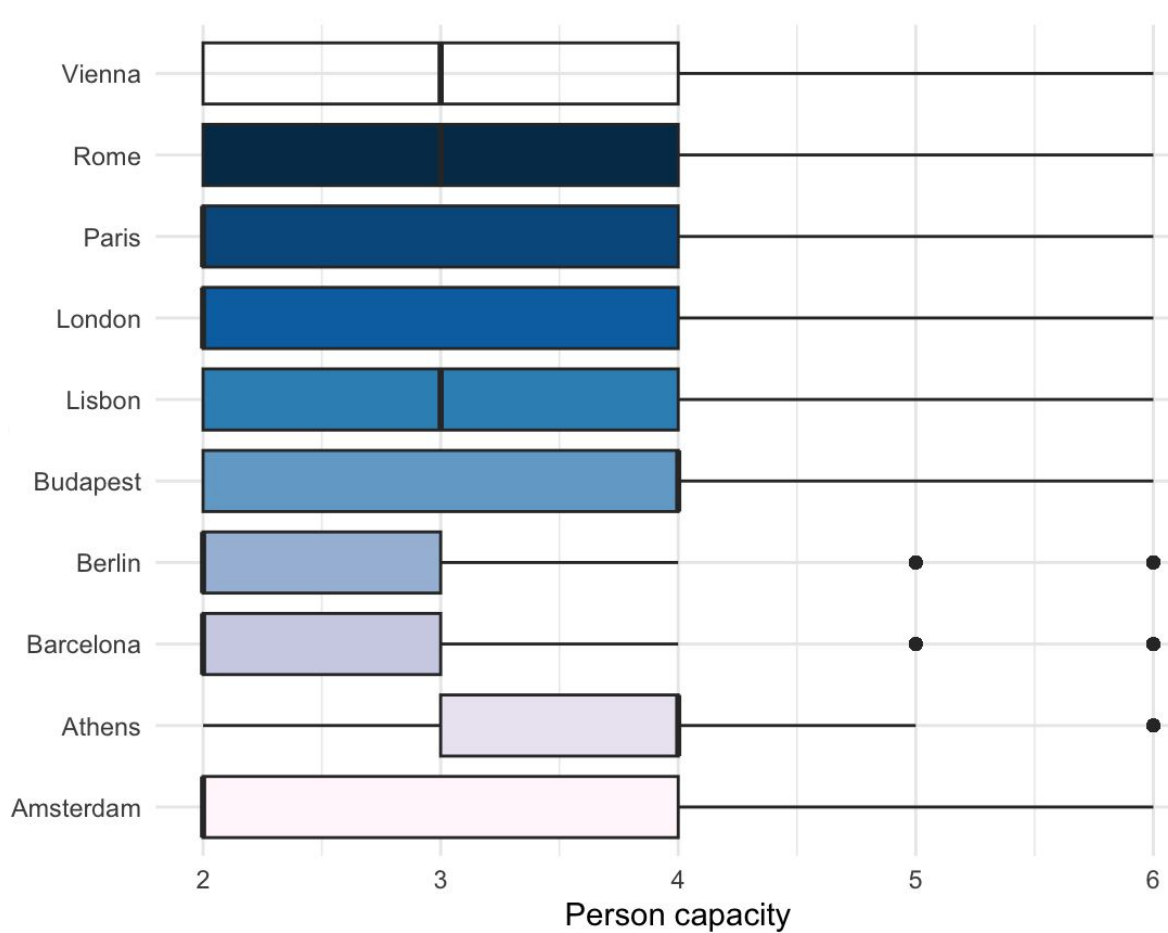
room_type	host_is_superhost	multi	biz	city	day
Length:51707	Mode :logical	Mode :logical	Mode :logical	Length:51707	Length:51707
Class :character	FALSE:38475	FALSE:36642	FALSE:33599	Class :character	Class :character
Mode :character	TRUE :13232	TRUE :15065	TRUE :18108	Mode :character	Mode :character



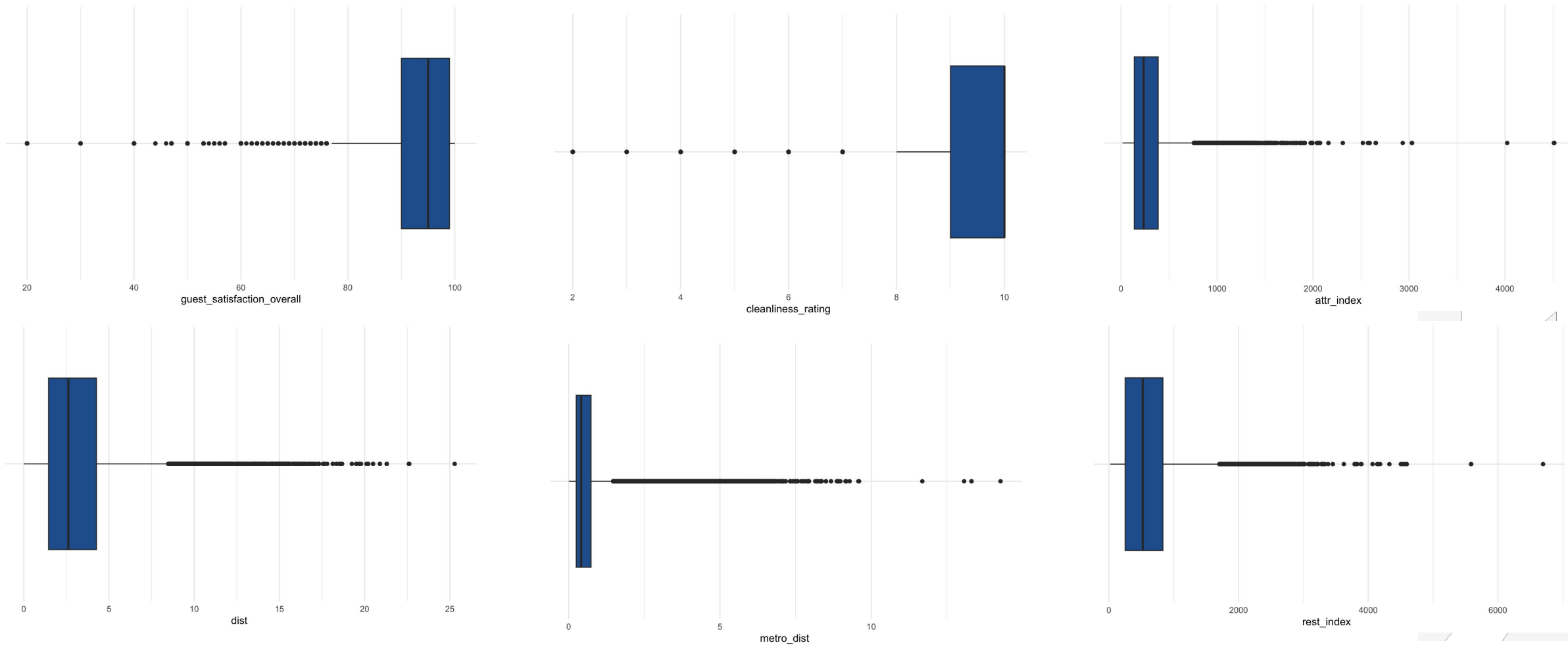
OUTLIERS EN LA VARIABLE A PREDECIR



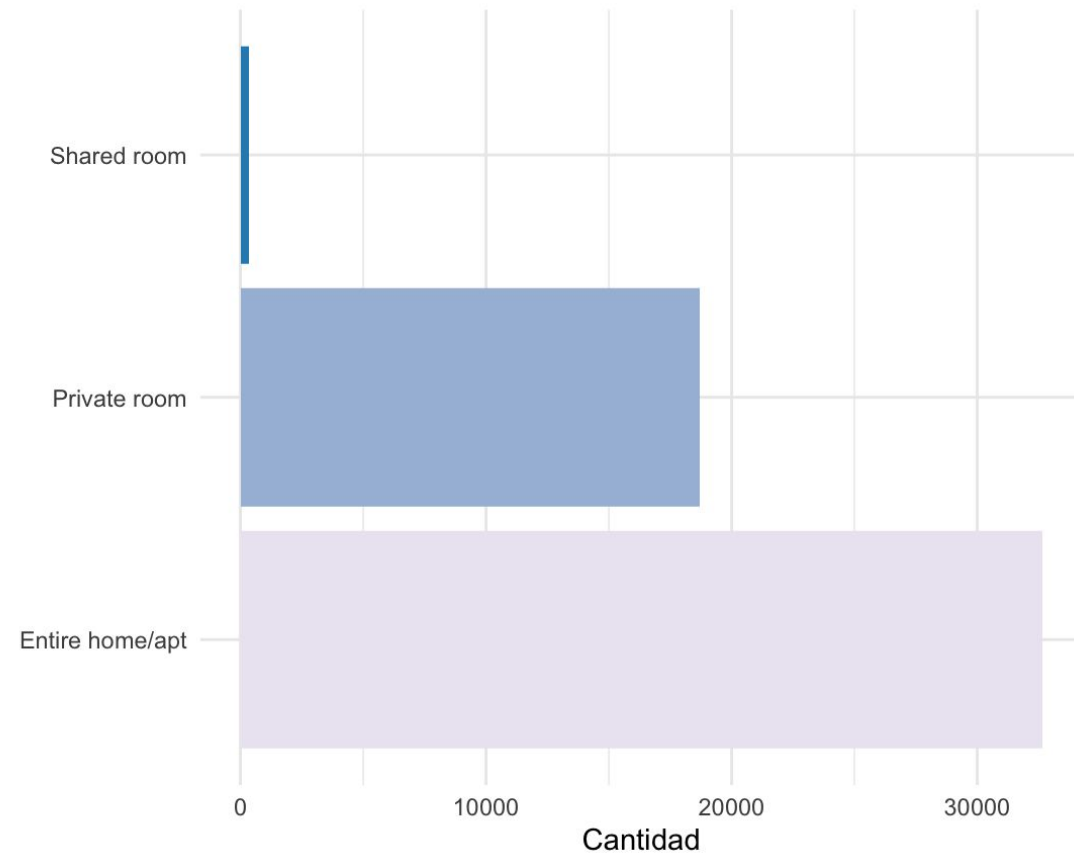
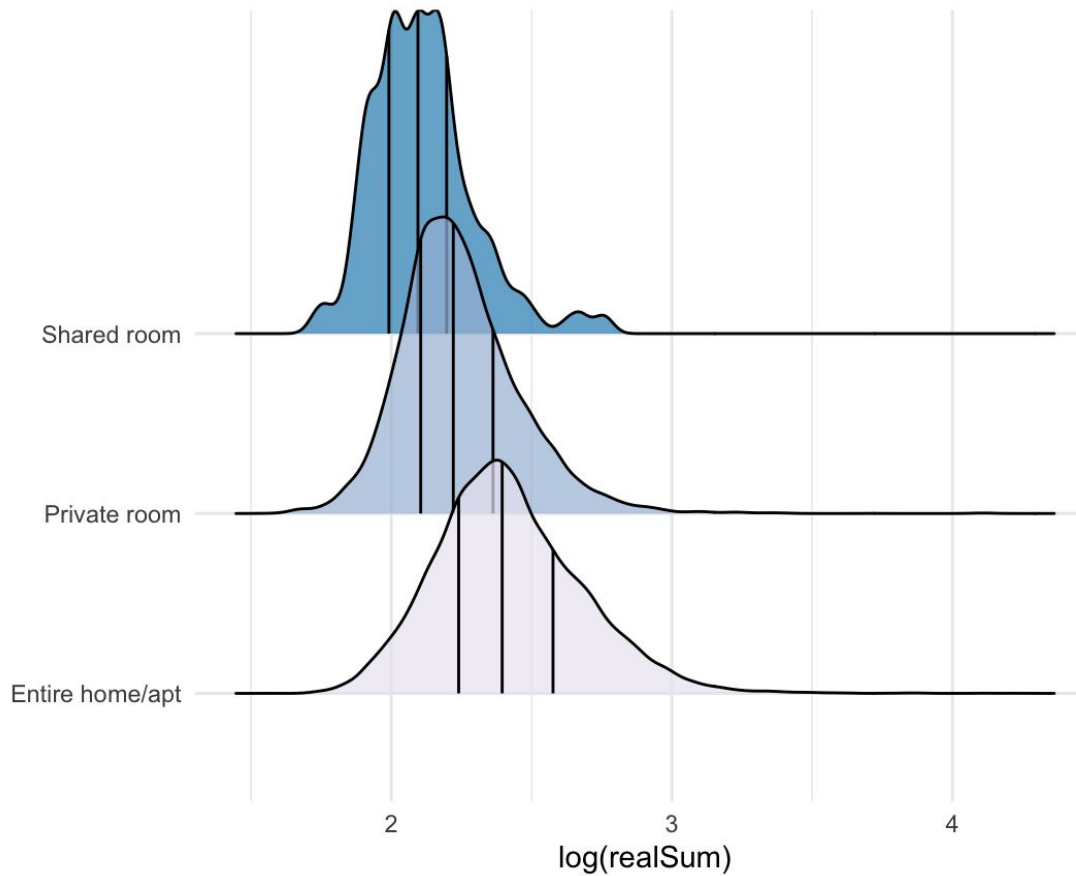
OUTLIERS EN LA CAPACIDAD Y CANTIDAD DE HABITACIONES



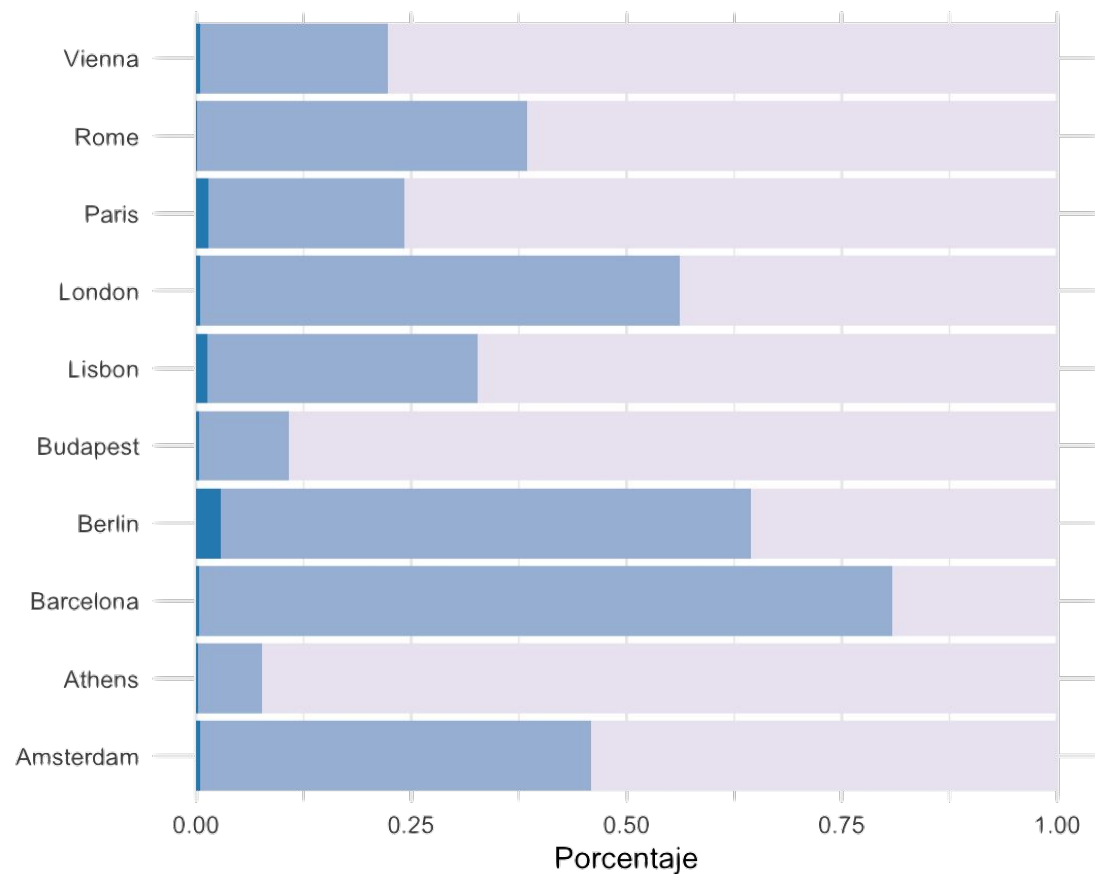
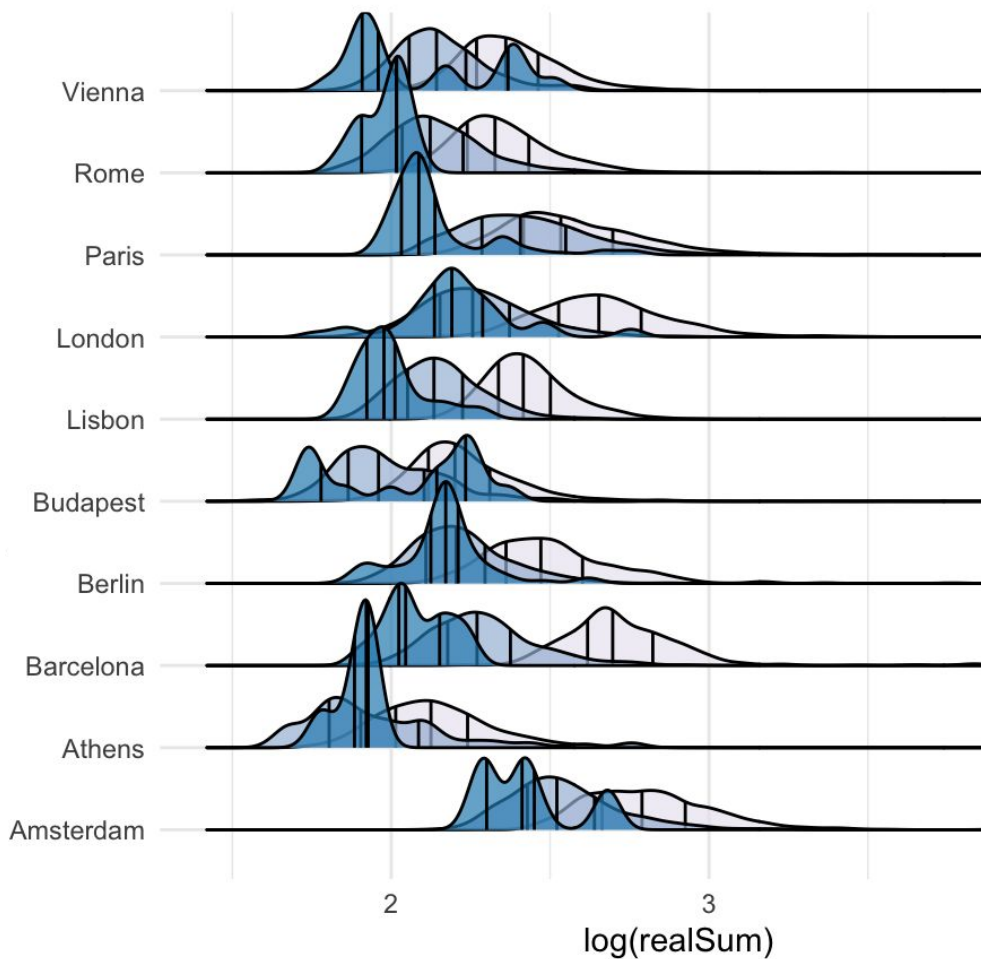
OUTLIERS EN LAS OTRAS VARIABLES NUMÉRICAS



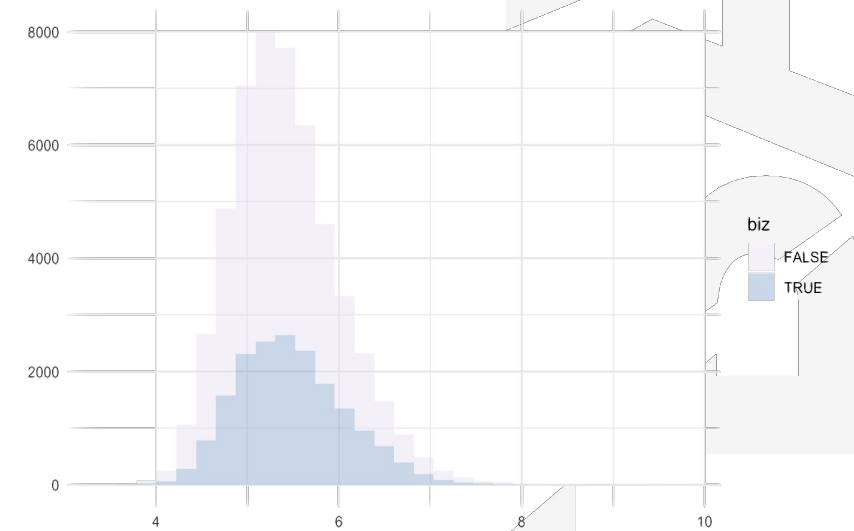
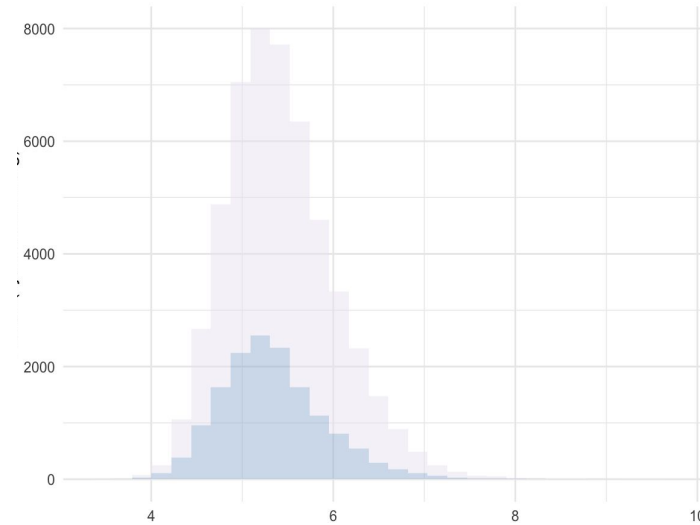
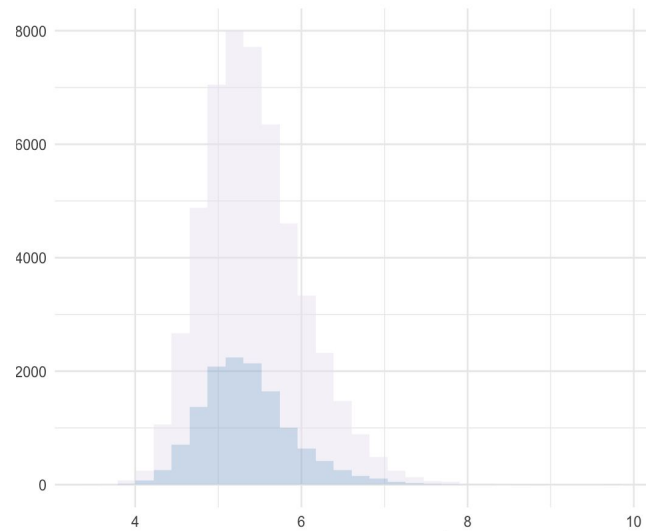
COMPORTAMIENTO DEL PRECIO SEGÚN EL TIPO DE HABITACIÓN



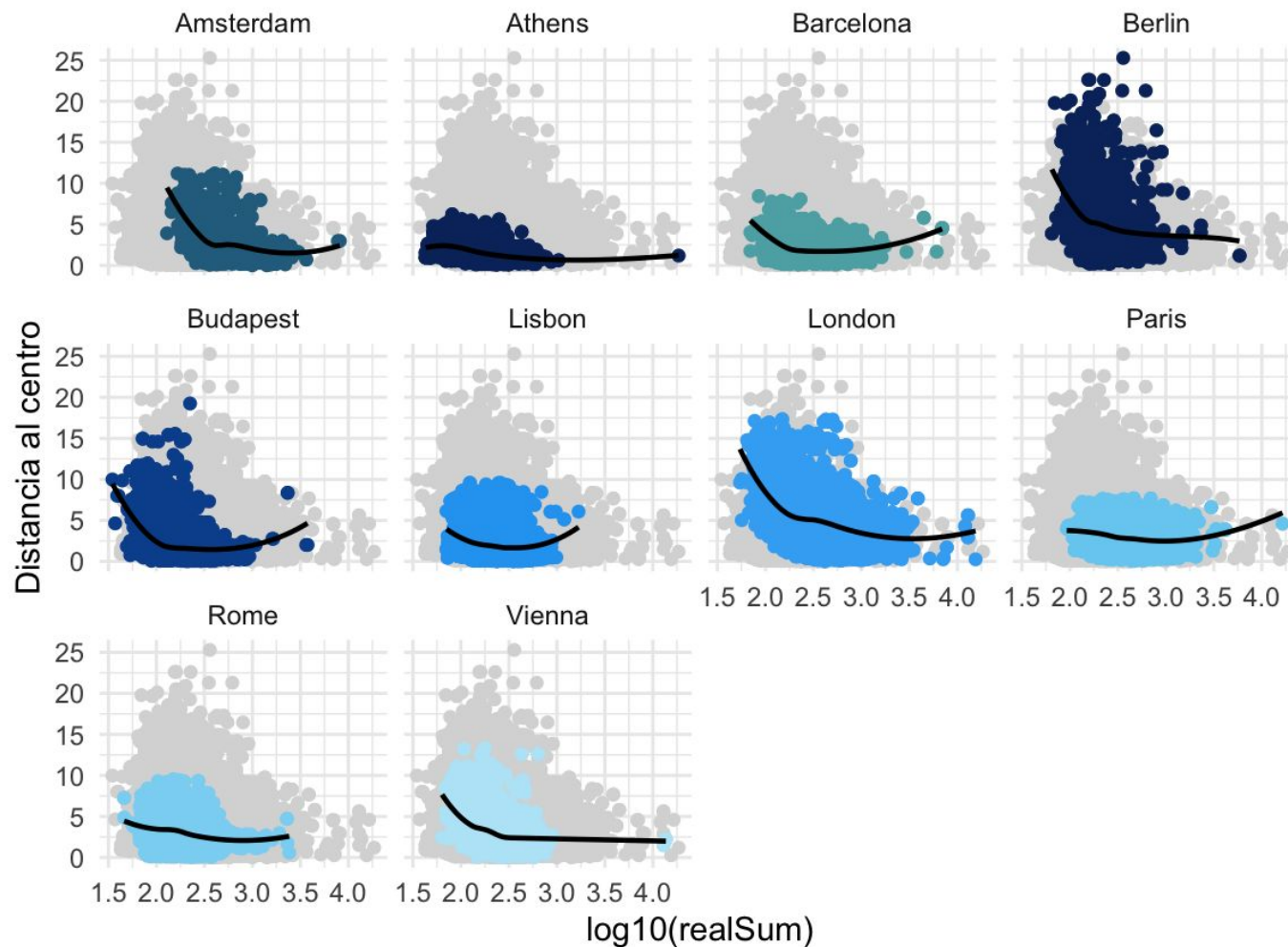
COMPORTAMIENTO DEL PRECIO SEGÚN EL TIPO DE HABITACIÓN Y LA CIUDAD



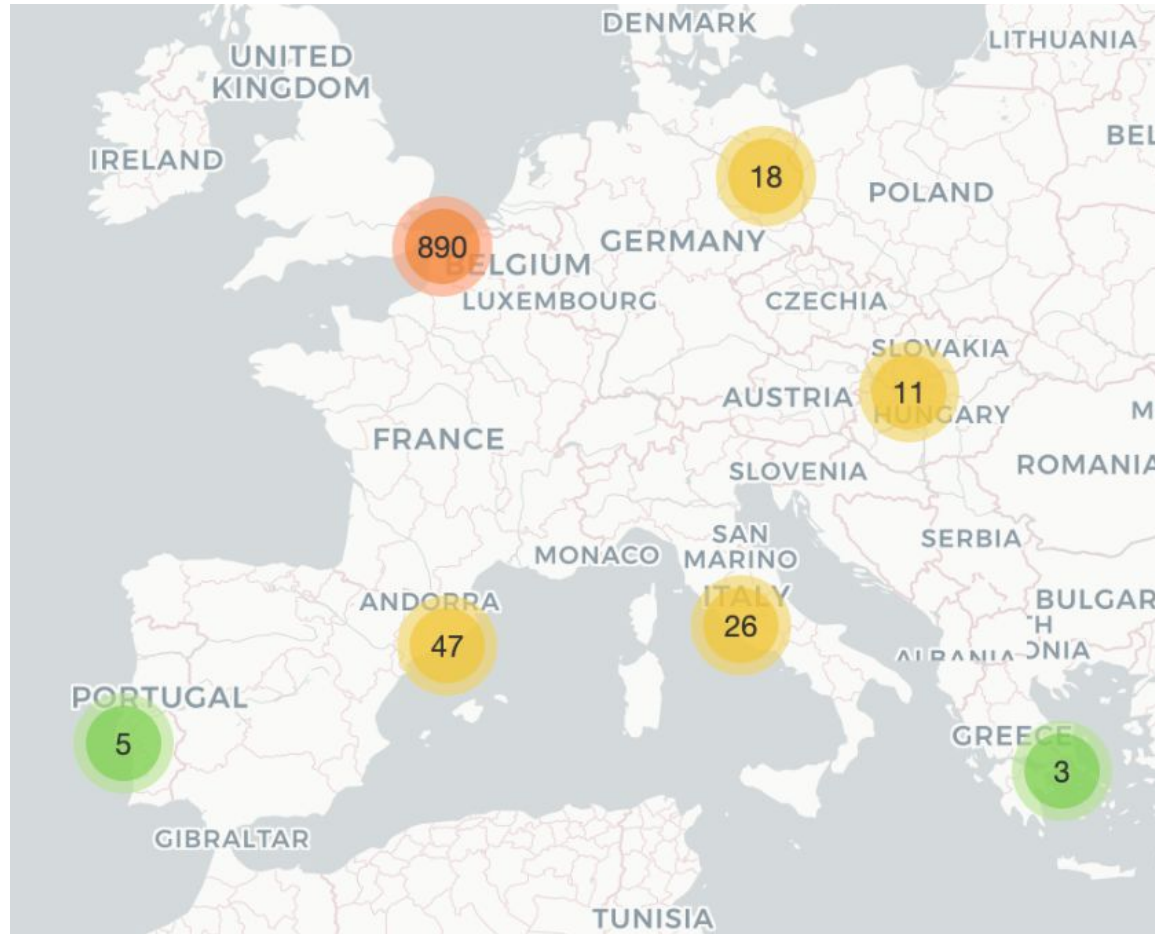
COMPORTAMIENTO DEL PRECIO SEGÚN TIPO DE HOST



COMPORTAMIENTO DEL PRECIO SEGÚN CIUDAD Y DISTANCIA



PRECIOS MÁS ALTOS Y BAJOS CON SUS UBICACIONES



CONCLUSIONES

El dataset va a poder ser utilizado para la realización del modelo. Esto se debe a que tiene variables muy relevantes e interesantes, existe la posibilidad de agregar nuevos registros y se pueden (y es recomendable) agregar nuevas variables que puedan resultar relevantes para el caso de estudio.

Queda para un análisis más exhaustivo poder determinar mediante qué variables se puede predecir el precio, viendo así la feature importance de las variables.

En cuanto a lo observado, las recomendaciones para el negocio son:

- Lanzar promociones, paquetes y más publicidad para alojamientos menos concurridos.
- Promocionar la implementación de alojamientos compartidos.
- Realizar un análisis zonal por ciudad.
- Mejorar la eficiencia del precio publicado teniendo en cuenta las variables.



Instituto Tecnológico
de Buenos Aires



MUCHAS GRACIAS