

# Taller 1 – Ciencia de datos

## 1 SELECCIÓN DEL DATASET DE TRABAJO:

Se seleccionó la ciudad de **Vancouver en Canadá** para hacer el análisis

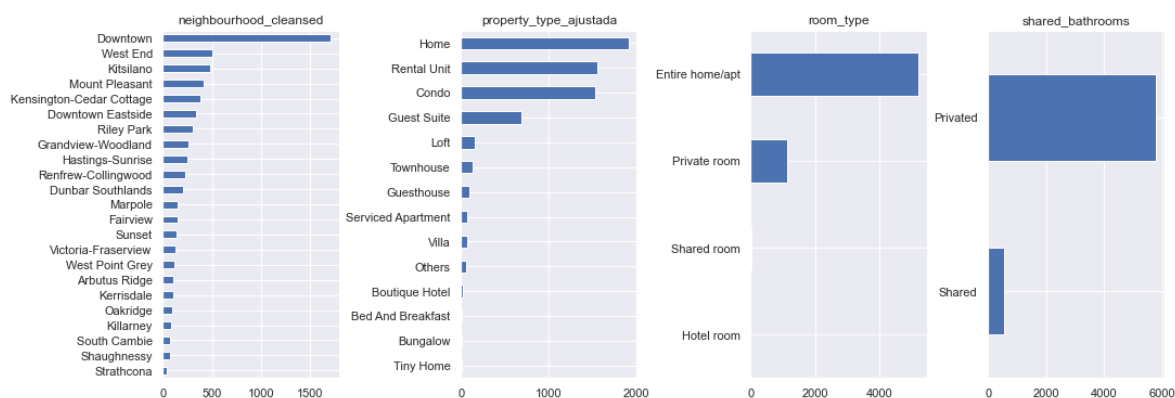
## 2 ENTENDIMIENTO INICIAL DE DATOS

Después de revisar el diccionario de datos y de verificar la cardinalidad de los mismos, se seleccionaron las siguientes variables categóricas:

- neighbourhood\_cleansed
- property\_type
- room\_type

Las demás se descartaron por tener alta cardinalidad, por referirse a características que no son de los inmuebles como al anfitrión o similares. Así mismo, se descartaron aquellas características que requieren un mayor procesamiento o técnicas más avanzadas como es el caso de las imágenes, o los textos que no están estandarizados (ver Notebook).

### 2.1 VARIABLES CATEGÓRICAS



De la gráfica anterior observamos que la gran mayoría de inmuebles están ubicados en el centro de la ciudad, son casas o apartamentos completos y cuentan con baños privados.

### 2.2 VARIABLES CUANTITATIVAS

Con base en el diccionario de datos se excluyeron de este análisis aquellas variables que son ids, se refieren a información del anfitrión, no son relevantes porque no corresponden a características del inmueble o que tienen todos sus valores nulos. Dado que el tiempo para el análisis es limitado

se decide también omitir algunas variables, dado que pueden estar recogidas por otras que se refieren a periodos de tiempo más cercanos, que pueden reflejar mejor la realidad actual del inmueble.

De manera que se decide explorar las siguientes variables:

- latitude
- longitude
- bathrooms
- bedrooms
- beds
- availability\_30
- number\_of\_reviews\_130d
- review\_scores\_location
- review\_scores\_value
- price\_ajustada

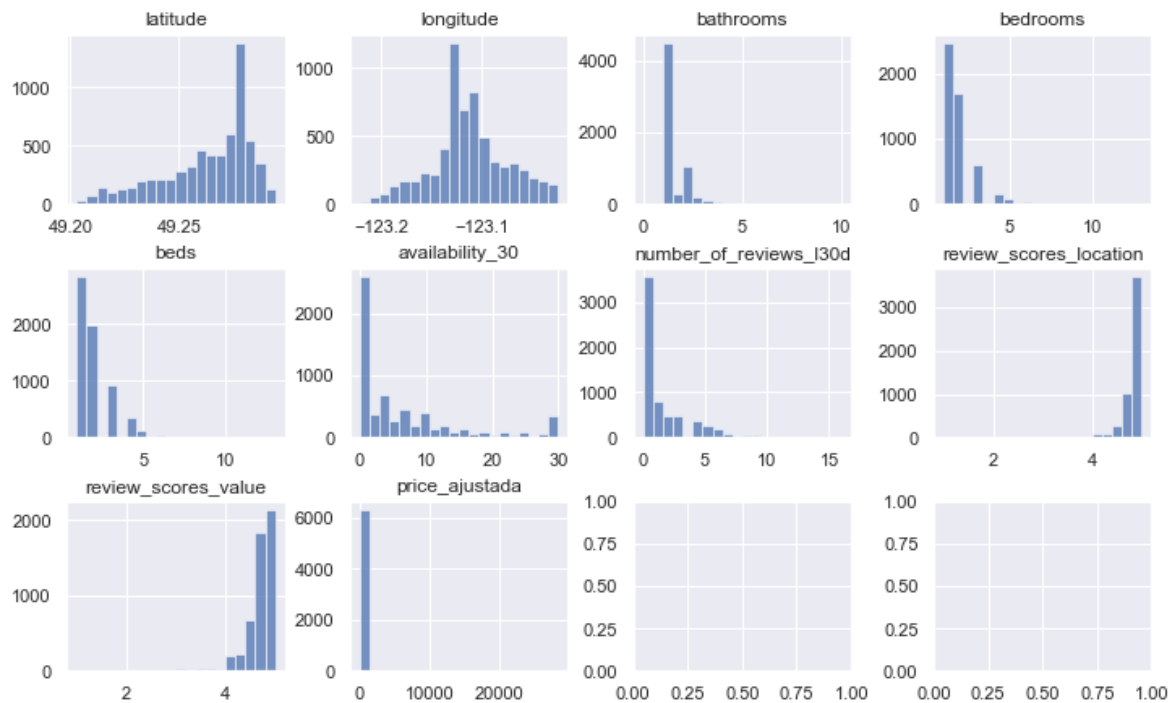
Así mismo, desde el punto de vista del negocio, se considera que las siguientes variables pueden ser las principales:

- **price**: Porque el retorno de una inversión en estos bienes depende en parte del precio al que se pueda arrendar y de la frecuencia de arriendo.
- **availability\_30**: porque es un proxy de qué tanto se arrendará un inmueble (es aproximada porque de acuerdo con el diccionario un inmueble también puede no estar disponible sin estar ocupado).

Estadísticas invariantes de las variables cuantitativas

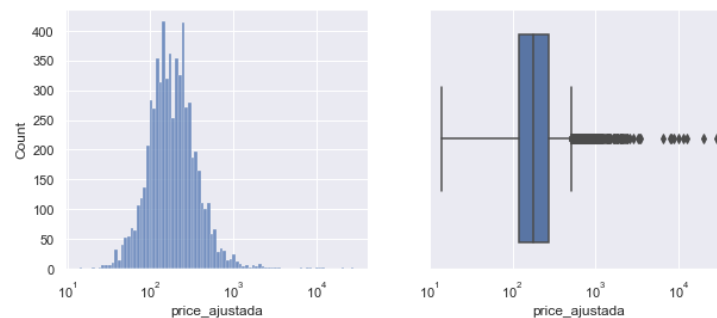
	count	mean	std	min	25%	50%	75%	max
<b>latitude</b>	6355.00	49.26	0.02	49.20	49.25	49.27	49.28	49.29
<b>longitude</b>	6355.00	-123.11	0.04	-123.22	-123.13	-123.12	-123.09	-123.02
<b>bathrooms</b>	6354.00	1.35	0.70	0.00	1.00	1.00	1.50	10.00
<b>bedrooms</b>	5037.00	1.78	1.03	1.00	1.00	2.00	2.00	13.00
<b>beds</b>	6313.00	1.94	1.19	1.00	1.00	2.00	2.00	13.00
<b>availability_30</b>	6355.00	6.50	8.49	0.00	0.00	3.00	9.00	30.00
<b>number_of_reviews_130d</b>	6355.00	1.40	2.17	0.00	0.00	0.00	2.00	16.00
<b>review_scores_location</b>	5233.00	4.83	0.27	1.00	4.78	4.91	5.00	5.00
<b>review_scores_value</b>	5234.00	4.68	0.39	1.00	4.60	4.76	4.89	5.00
<b>price_ajustada</b>	6355.00	250.52	571.81	14.00	120.00	179.00	275.00	28386.00

De la tabla anterior podemos observar que en promedio se trata de inmuebles con 1.9 camas, 1.3 baños y 1.8 habitaciones. Así mismo el precio promedio por noche es de 6.355 dólares la noche y para los siguientes 30 días en promedio tienen disponibilidad de 6.5. La mediana de los precios es de 275 dólares, sin embargo, se observa que algunos de los inmuebles llegan hasta los 28 mil dólares, lo cual corresponde a datos extremos, pero no necesariamente erróneos.



### 2.2.1 Precio

Para visualizar mejor la información de los precios, se optó por representarlos en escala logarítmica dado que, como es normal, existe un gran número de inmuebles con precios similares y algunos poco (exclusivos) con precios muy altos, por lo que la distribución de la variable sin transformar presenta una marcada simetría hacia la izquierda.



### 3 ESTRATEGIA DE ANÁLISIS

---

Dado que se busca dar recomendaciones de inversión, en el análisis se intenta identificar aquellas variables que más afectan el precio de arriendo de los inmuebles y la ocupación. Para ello se realiza un análisis bivariado principalmente gráfico en el cual se busca relacionar las variables cualitativas y cuantitativas, para identificar aquellas que más afectan tanto el precio como la ocupación. Adicionalmente, en el caso de las variables cuantitativas se calculan coeficientes de correlación para hacer esta identificación.

### 4 DESARROLLO DE LA ESTRATEGIA

---

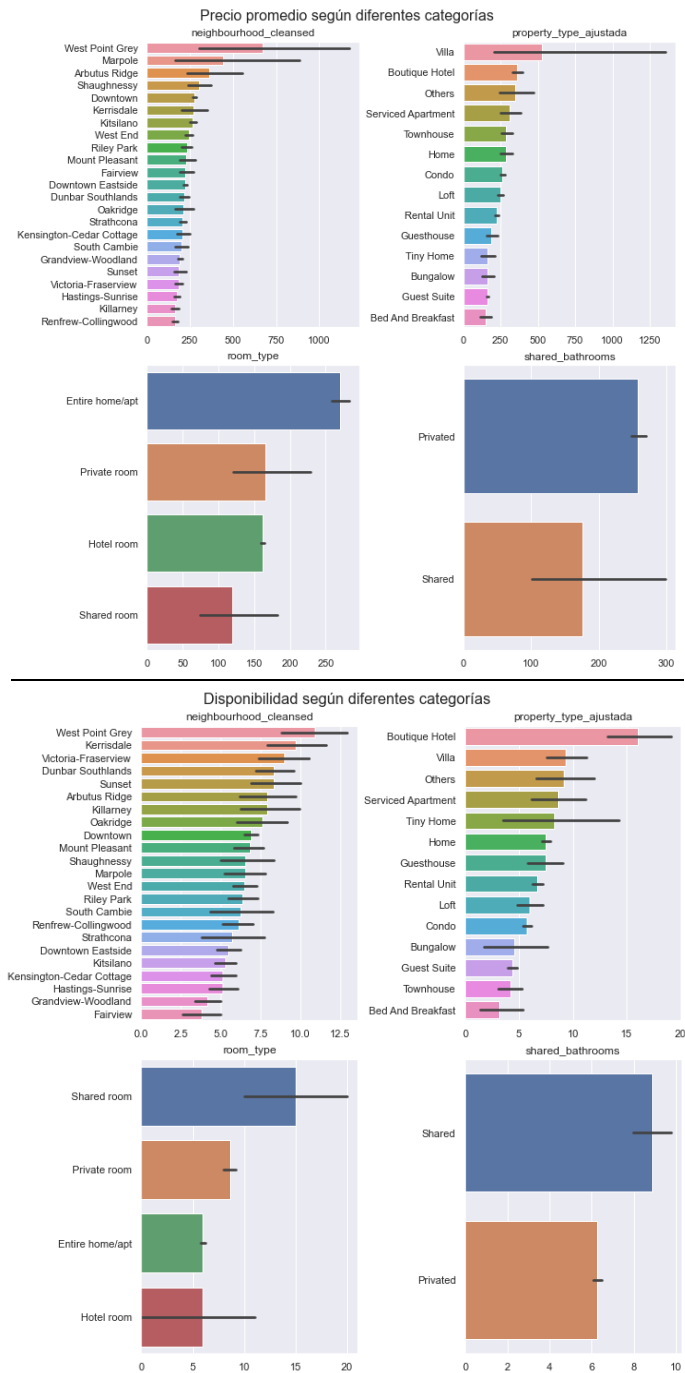
(ver notebook)

### 5 GENERACIÓN DE RESULTADOS

---

#### 5.1 VARIABLES CUALITATIVAS

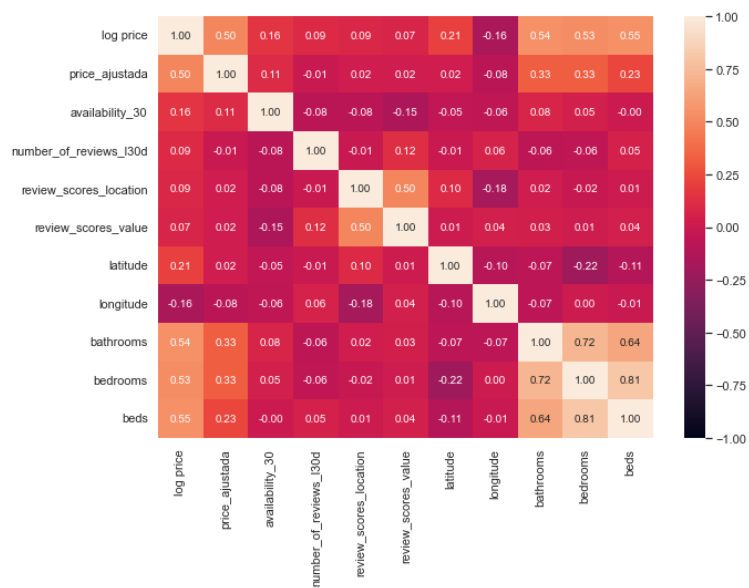
Para analizar estas variables se calculo el valor promedio de la ocupación y del precio para cada categoría junto con un intervalo de confianza calculado con bootstrap:

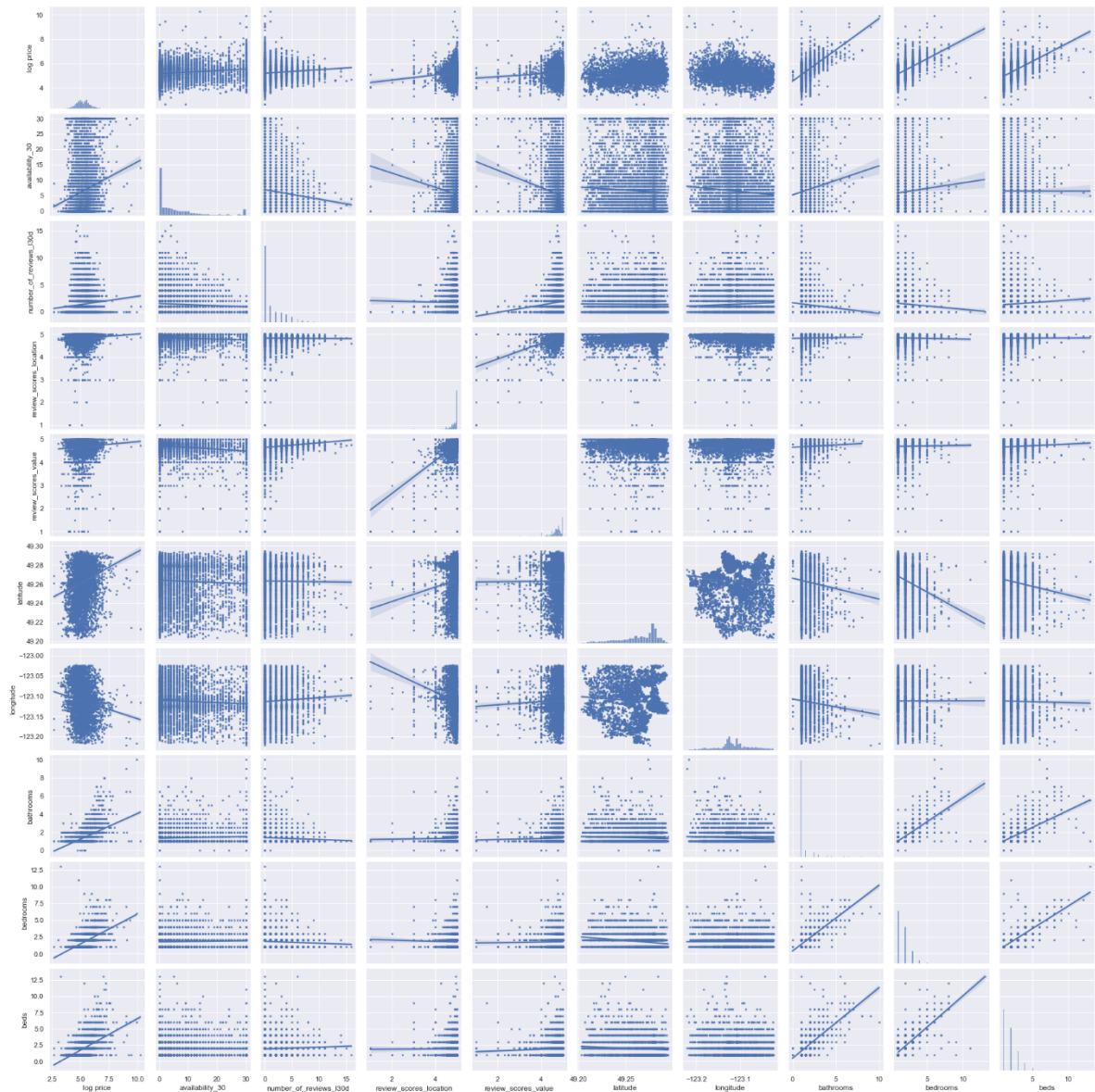


De las gráficas anteriores puede concluirse que los precios más altos se asocian con inmuebles en villas y hoteles boutique, con apartamentos enteros, con baños privados y a los barrios “West Point Grey” y “Marpole”. Igualmente se identifica que la menor disponibilidad (proxy de mayor ocupación) se asocia con baños compartidos, los barrios Fairview y Grandview.

## 5.2 VARIABLES CUANTITATIVAS

A nivel de las variables cuantitativas se observa que la mayor correlación con el precio se encuentra en variables como el número de camas, numero de habitaciones y número de baños. Mientras que la disponibilidad se asocia con el precio, lo cual implica que mayores precios llevan a mayor disponibilidad.





Finalmente, si tenemos en cuenta las variables geográficas es posible observar que los inmuebles con mayor precio se encuentran en el norte de la ciudad. Sin embargo, si nos fijamos en el cuartil más alto de los precios podemos encontrar inmuebles en casi toda la ciudad. Lo cual permitiría identificar inmuebles de alto valor que no necesariamente se encuentren en zonas en que el precio de compra de los inmuebles sea tan alto.



### Recomendaciones:

Es difícil establecer recomendaciones claras de compra de inmuebles si no se cuenta con información del precio de venta de estos, sin embargo, a partir de los datos anteriores es posible identificar algunos patrones que ayuden a tomar estas decisiones:

- 1) Existe una clara relación entre el precio de arriendo y el número de habitaciones, baños y camas. Por lo cual es recomendable invertir en inmuebles con varias habitaciones.
- 2) Es recomendable evaluar la compra de inmuebles en Fairview, este barrio, aunque no representa los precios más altos, tiene unos niveles de ocupación significativamente más altos que todos los demás barrios y maneja precios por encima de la mediana. Esta alta ocupación puede asociarse con una alta demanda o baja oferta de inmuebles en RBnB, lo que representa una oportunidad de negocio.
- 3) Aunque los baños privados se relacionan en promedio con mayores precios, esta diferencia no es estadísticamente significativa (ver intervalos de confianza en la gráfica), sin embargo, si asocia significativamente con una menor ocupación (en promedio 2.5 días).
- 4) El centro de la ciudad concentra la mayor parte de los inmuebles sin embargo allí deben concentrarse los inmuebles de mayor precio de compra. Sin embargo, al verificar la ubicación del 25% de los inmuebles con mayor precio de arriendo, se encuentra que están distribuidos por toda la ciudad, por lo que no necesariamente es una buena idea comprar en el centro si los precios de compra son demasiado altos.

En todo caso, estas recomendaciones deben evaluarse a la luz de los precios de venta de los inmuebles.