



Instituto Tecnológico  
de Buenos Aires

## **Trabajo Práctico 1**

### **82.05 - Análisis Predictivo**

**Abril Schäfer - 62877**

**Ian Dalton - 62345**

**2023 2Q**

# Caso de negocio

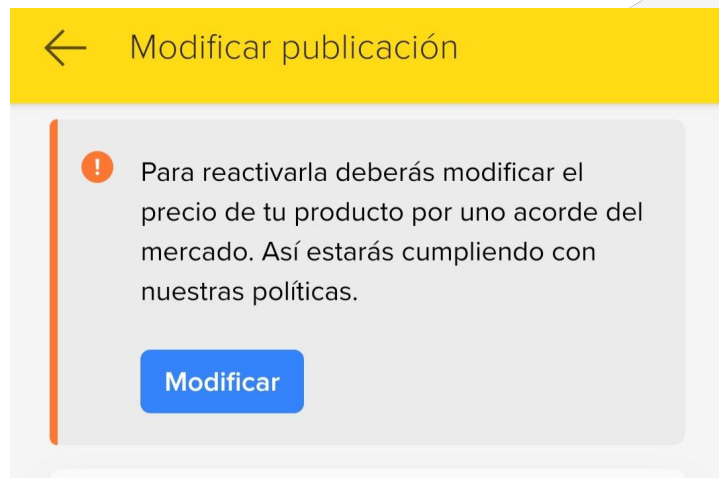


## Introducción y caso de negocio

Predecir el **precio adecuado** para alojamientos a particulares y turísticos.

La tarifa se deduce del cobro del anfitrión.

- **Precio por debajo:** Airbnb pierde ganancias.
- **Precio por encima:** Dificultad para alquilar el alojamiento.



## Modelo de Predicción

**Objetivo:** Predecir el precio que debe tener un alojamiento según sus características.

**Unidad observacional:** Alojamiento disponible en Airbnb en Buenos Aires

**Variable Target:** “Price”

**Modelo:** Regresión

# Dataset

## Información del Dataset

Contiene información de los **alojamientos disponibles** en Airbnb en Buenos Aires.

Cuenta con **29.346** registros y **60** variables.

**Se agregaron 6 nuevas variables:**

**Days\_between\_first\_last\_reviews**

**Host\_has\_about:** Escribe un about de más de 10 caracteres

**is\_shared\_bathroom**

**host\_about\_len**

**Is\_min/max\_nights\_different:** el minimo/maximo que ofrece es diferente a su promedio usual

## Limpieza de la base

**Eliminación de columnas irrelevantes:** listing\_url, scrape\_id, last\_scraped, picture\_url, host\_url, host\_name...

**Corrección de variable Bathroom:** incompleto

**Corrección de variable Licencia:** contenía números de teléfono

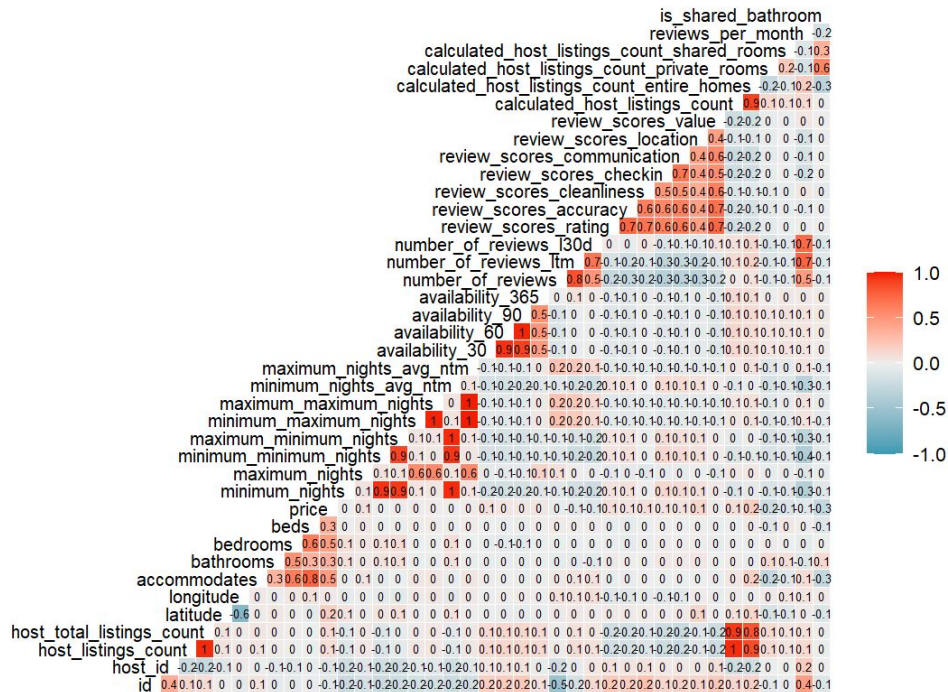
**Corrección de precio:** Se convirtió en número

**Frecuencia de actualización:** mensual

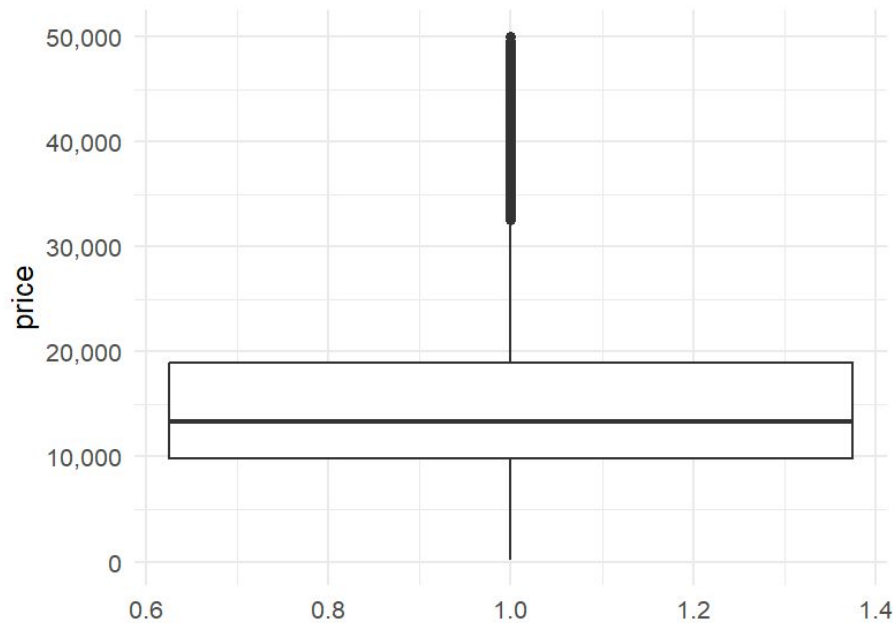
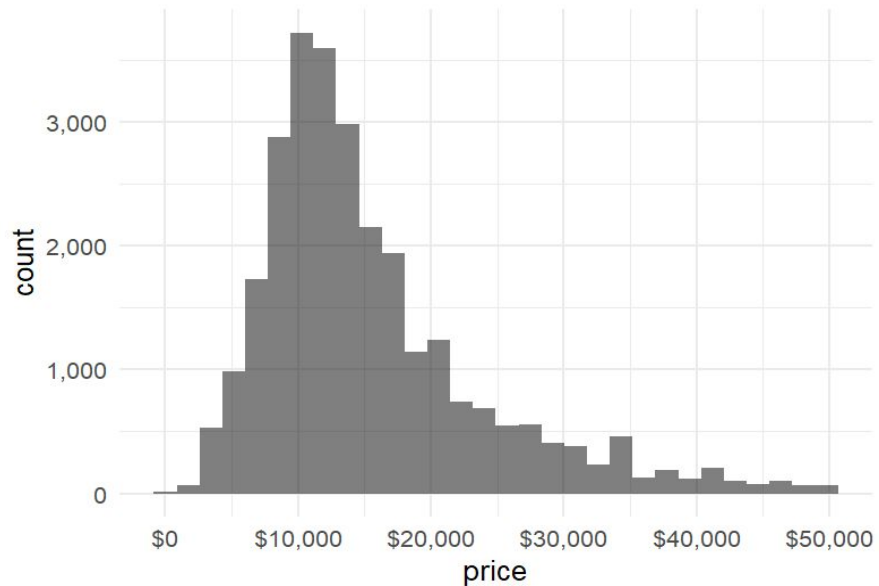
# Análisis exploratorio





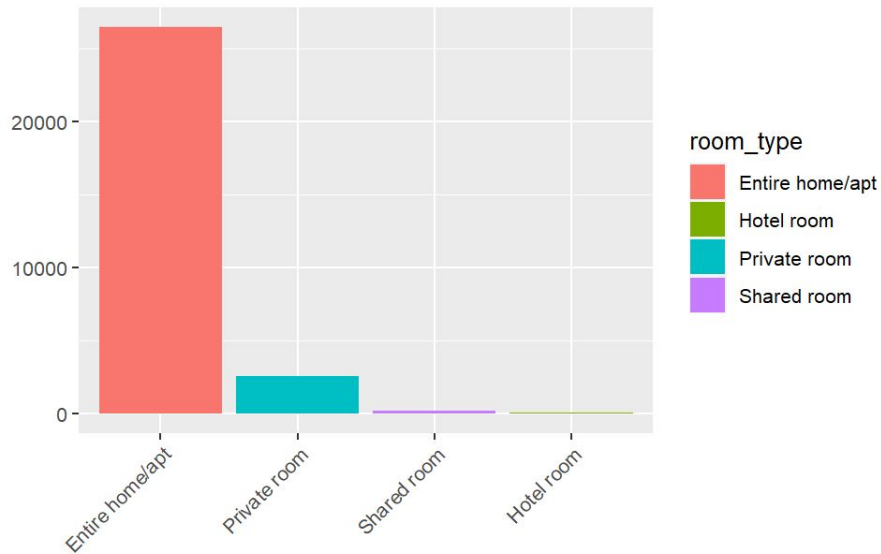
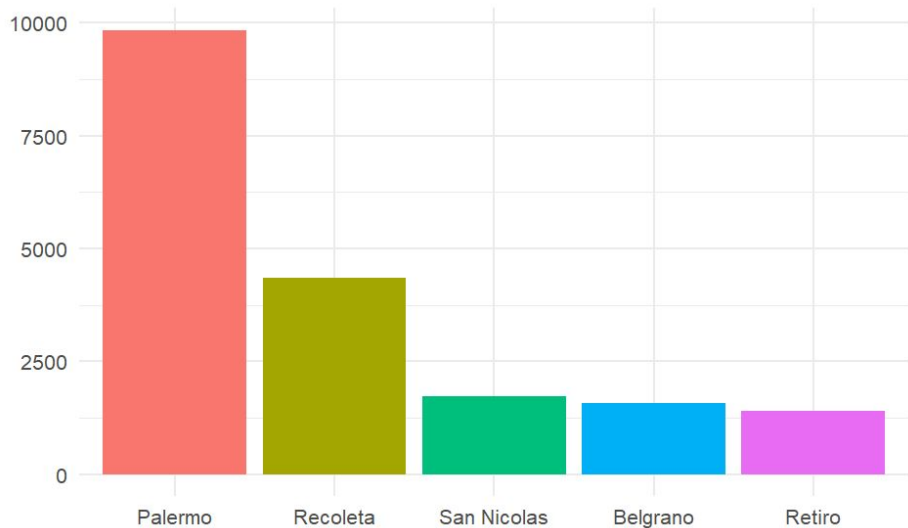


# Target

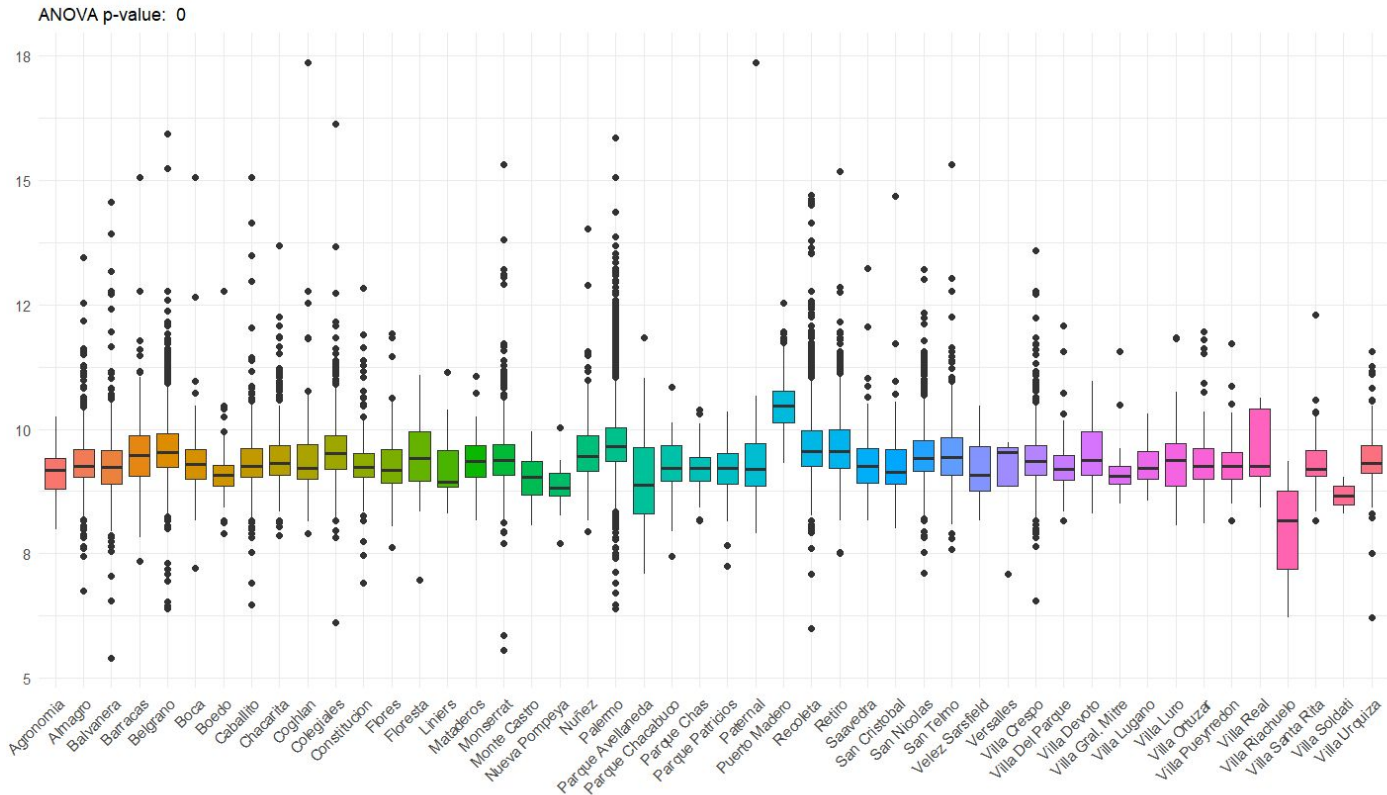


**Precio <= 50.000**

# Barrios y tipo de habitación

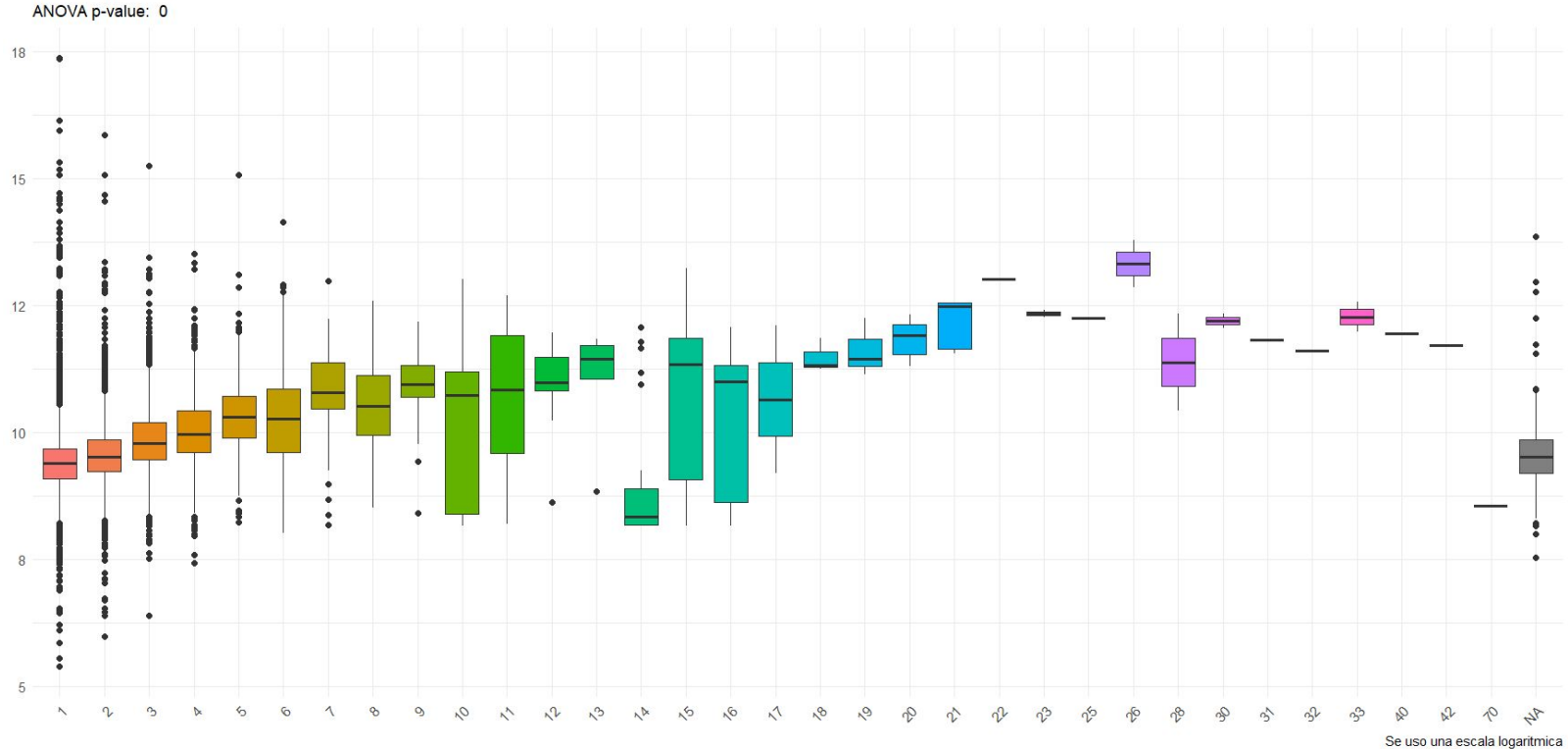


# Barrios

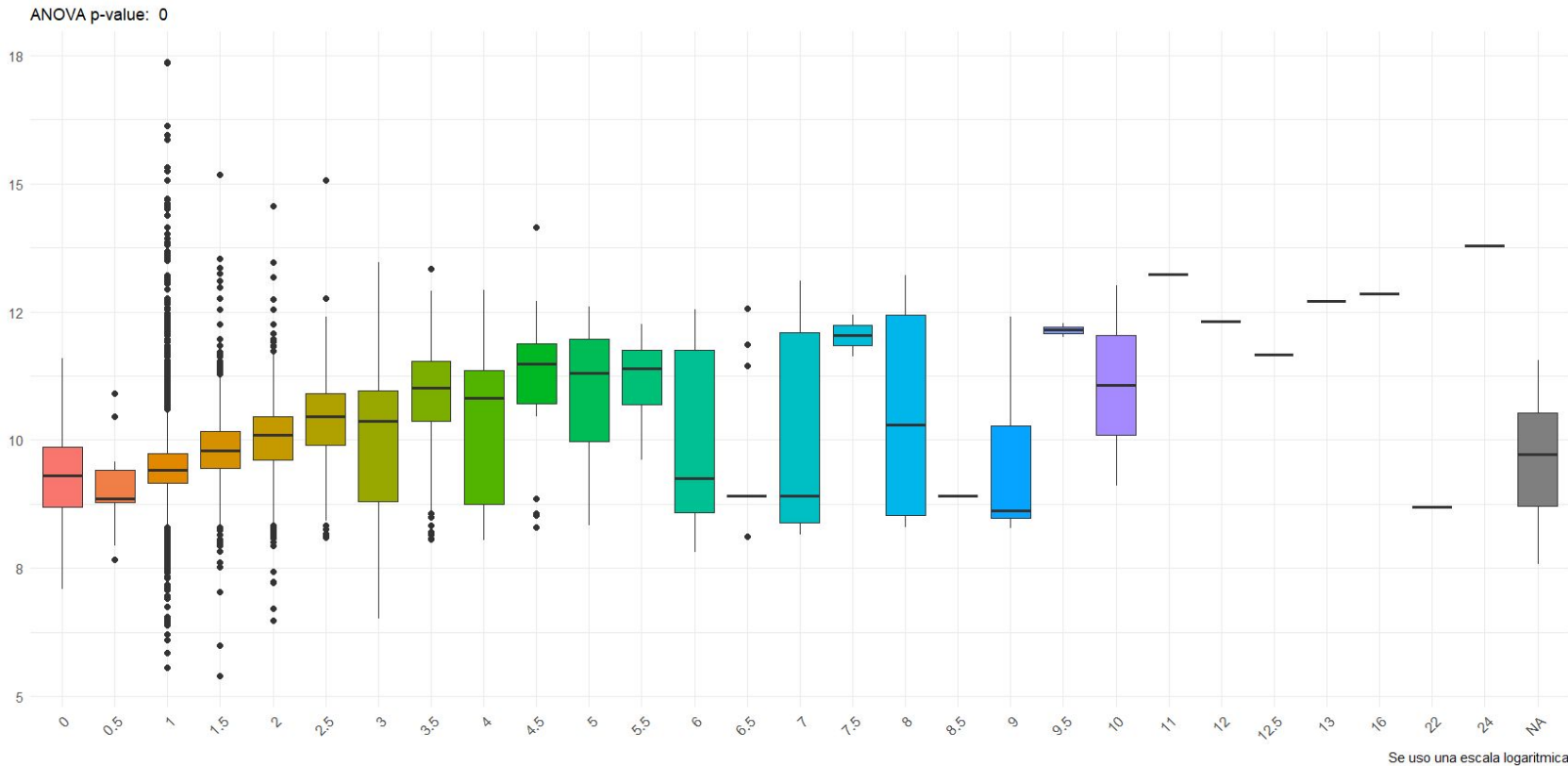


Se uso una escala logaritmica

# Camas vs Precio

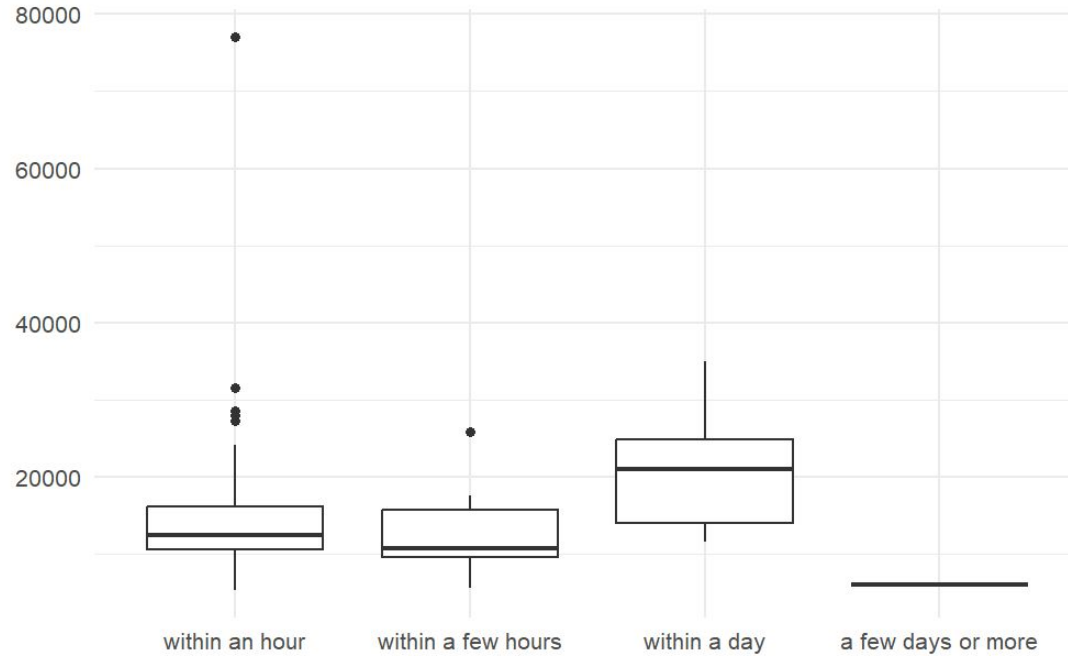


# Baños vs Precio



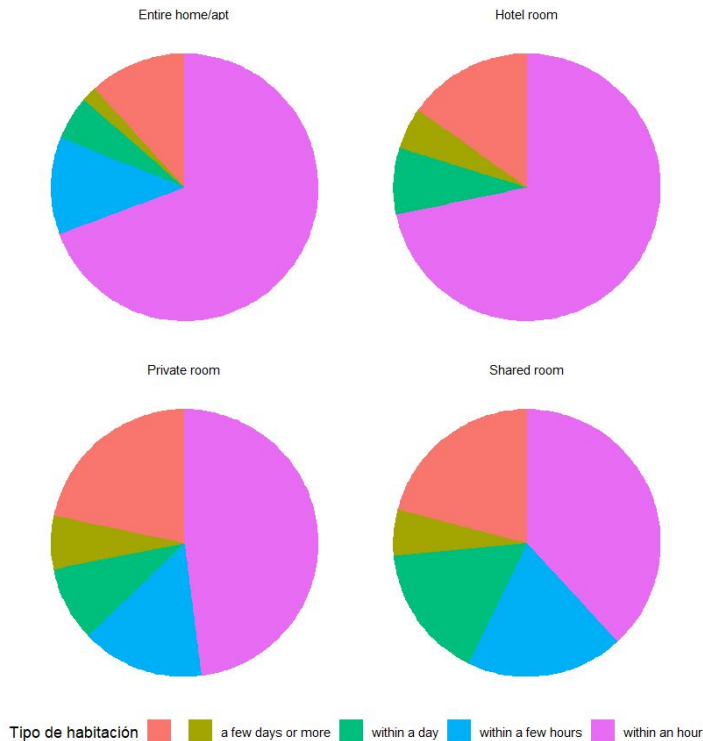
Se uso una escala logarítmica

## Tiempo de respuesta



**ANOVA = 0.115** → No hay evidencia que las medias difieren

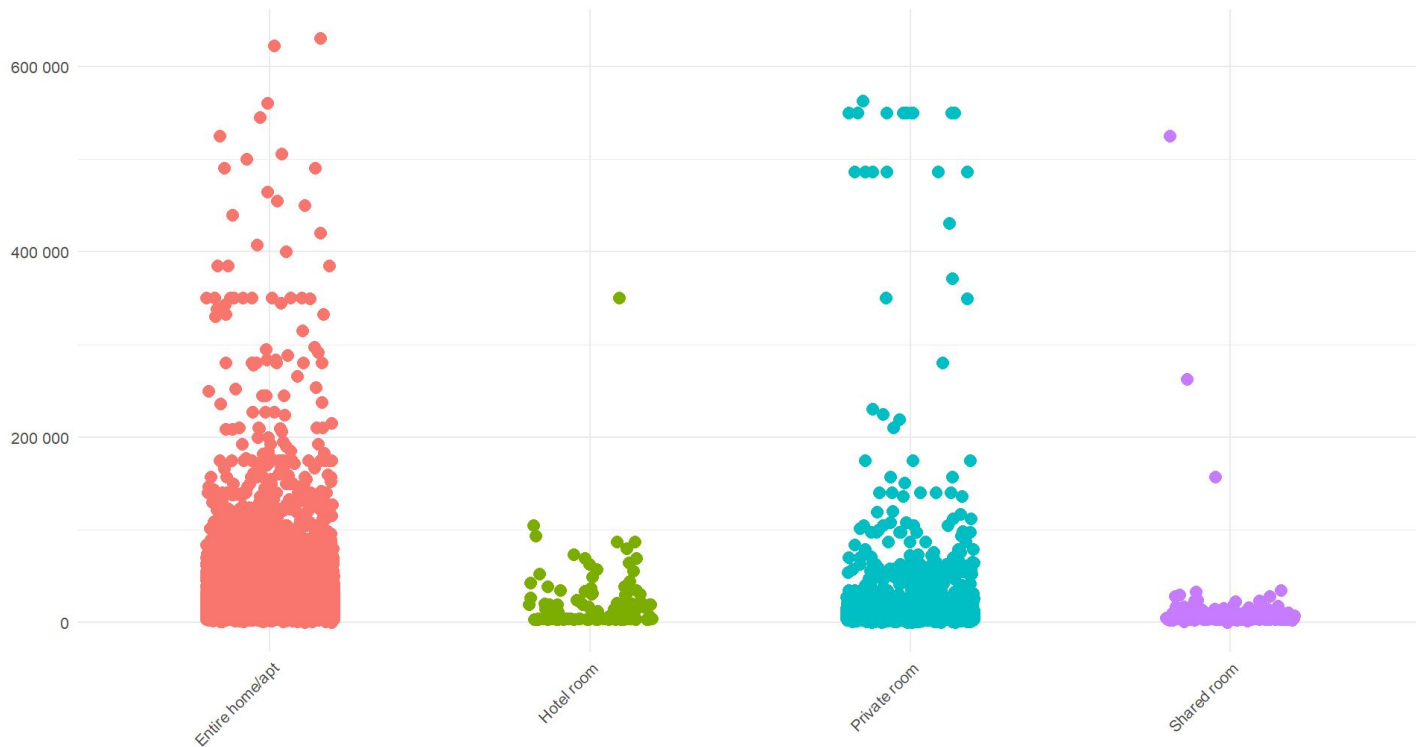
## Tiempo de respuesta



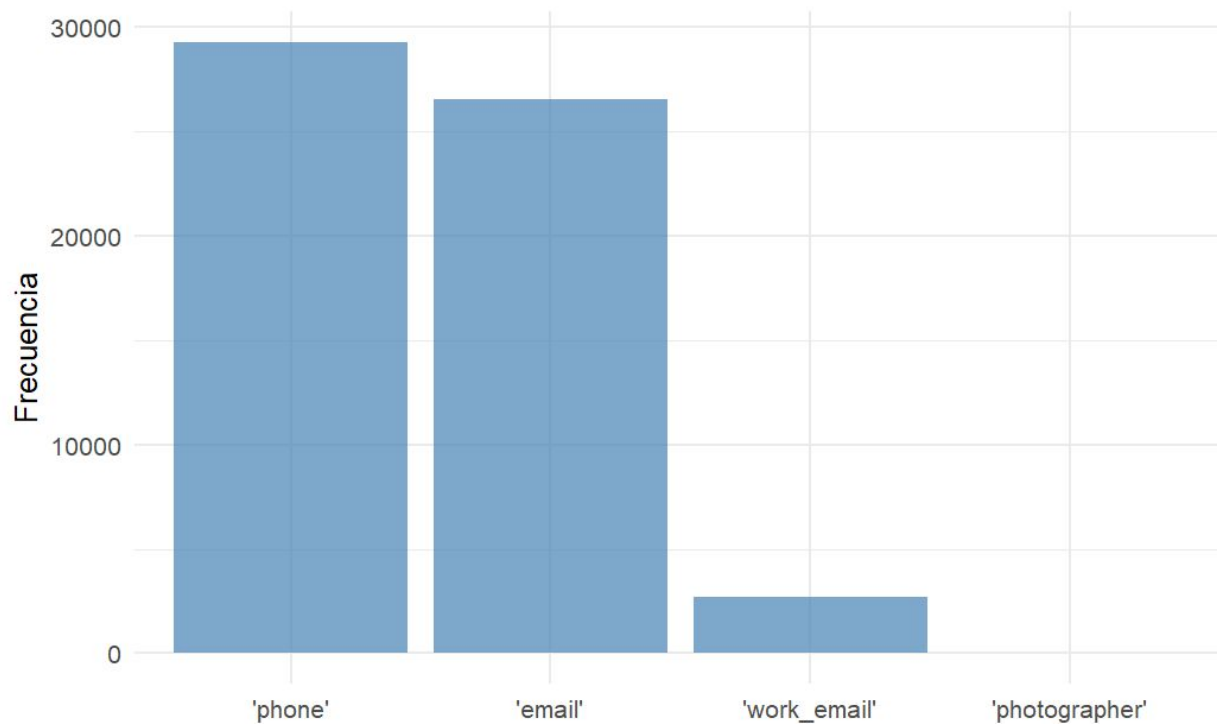
**ANOVA = 0.115** → No hay evidencia que las medias difieren



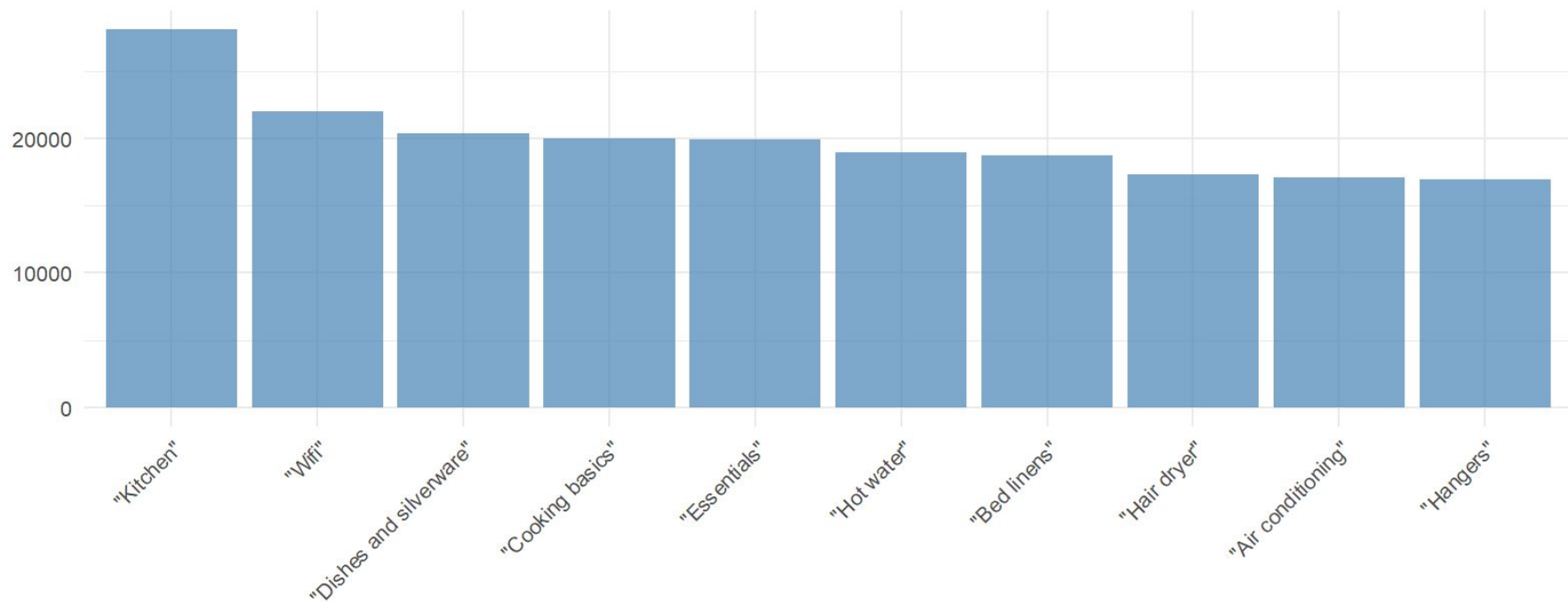
## Precio vs Tipo de habitación



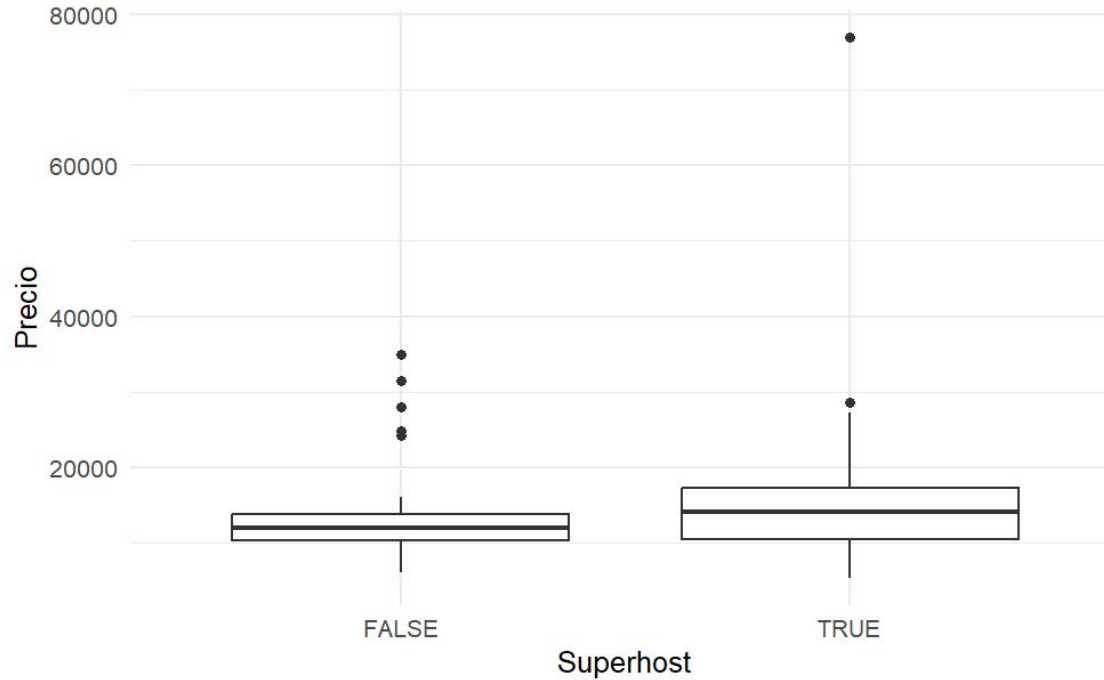
## Frecuencia de forma de verificación



## Frecuencia de los amenities: TOP 10

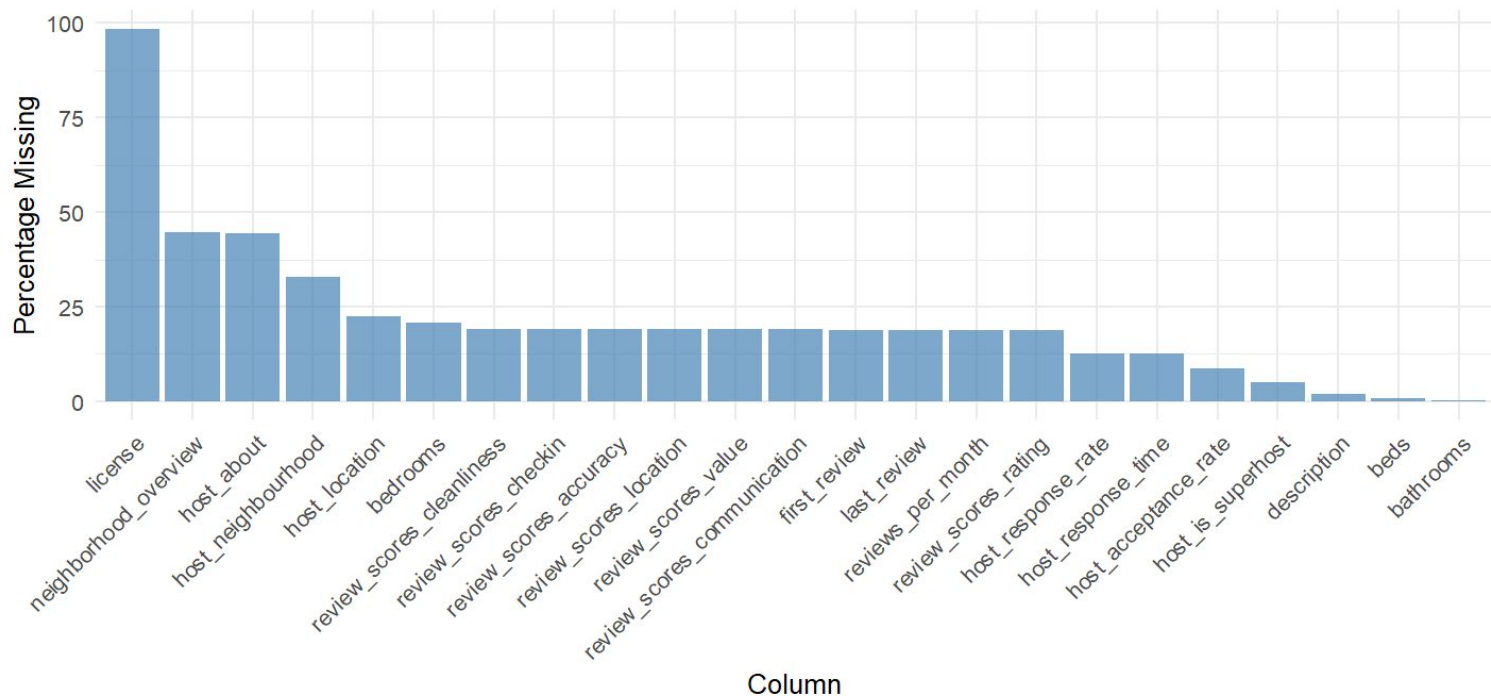


## Superhost



**F - test significativo** → Hay evidencia que las varianzas difieren

# Missings



# Clustering

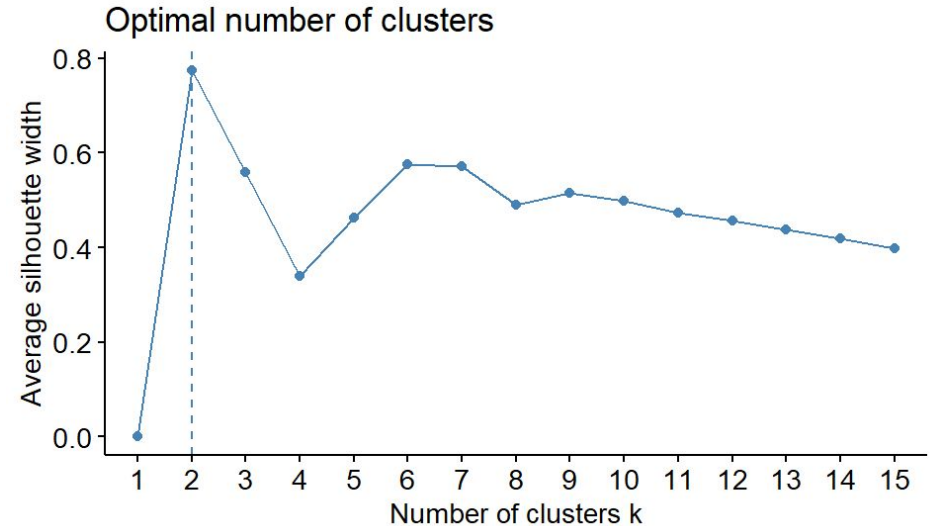
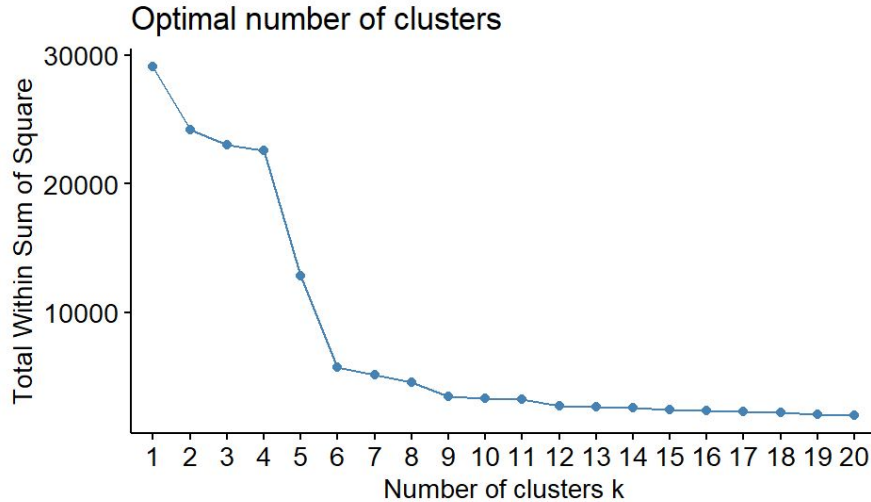
Se realizó con el objetivo de comparar y evaluar diferencias entre los alojamientos para luego determinar si existen **características** similares que puedan **influir en el precio**.

Los análisis se realizarán con respecto a las siguientes **variables numéricas**:

"host\_listings\_count", "accommodates", "bathrooms", "bedrooms", "beds",  
"minimum\_nights", "maximum\_nights", "availability\_30", "availability\_60",  
"availability\_90", "availability\_365", "number\_of\_reviews",  
"calculated\_host\_listings\_count", "reviews\_per\_month", "is\_shared\_bathroom")

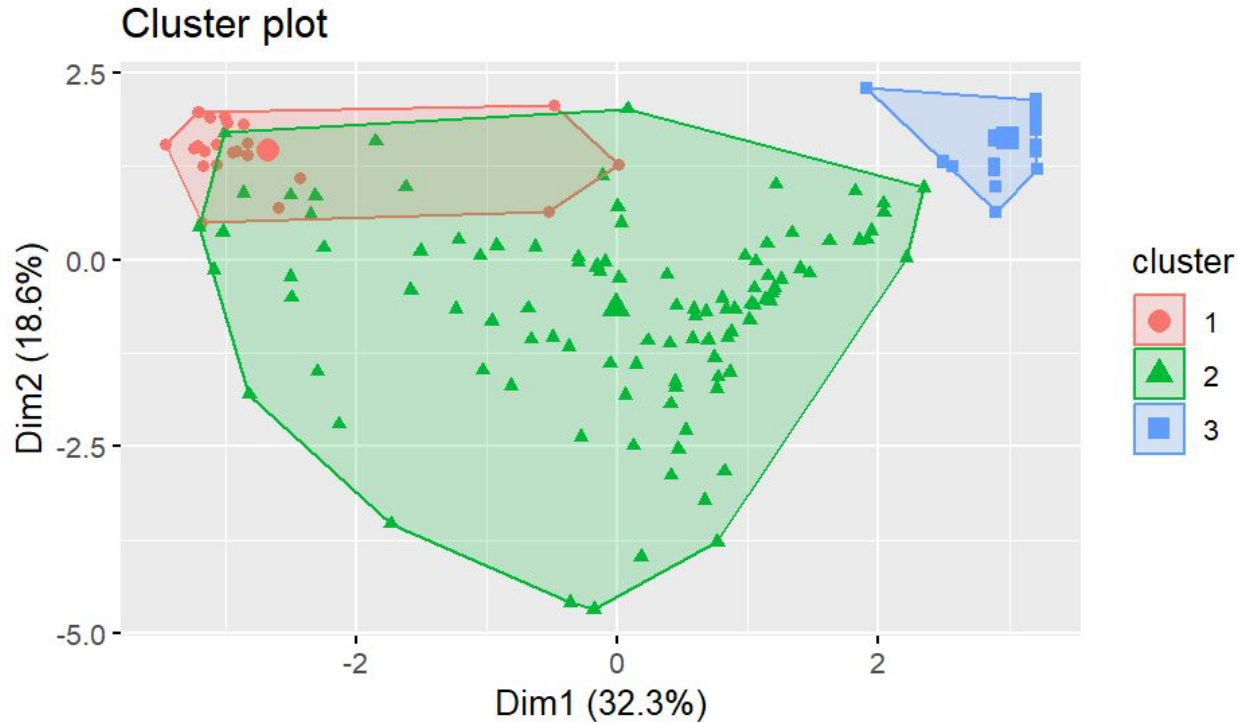
Medimos el valor del **estadístico de Hopkins** para evaluar si es conveniente la clusterización, obteniendo un valor cercano al **92,73%**.

# Clustering - Agrupaciones



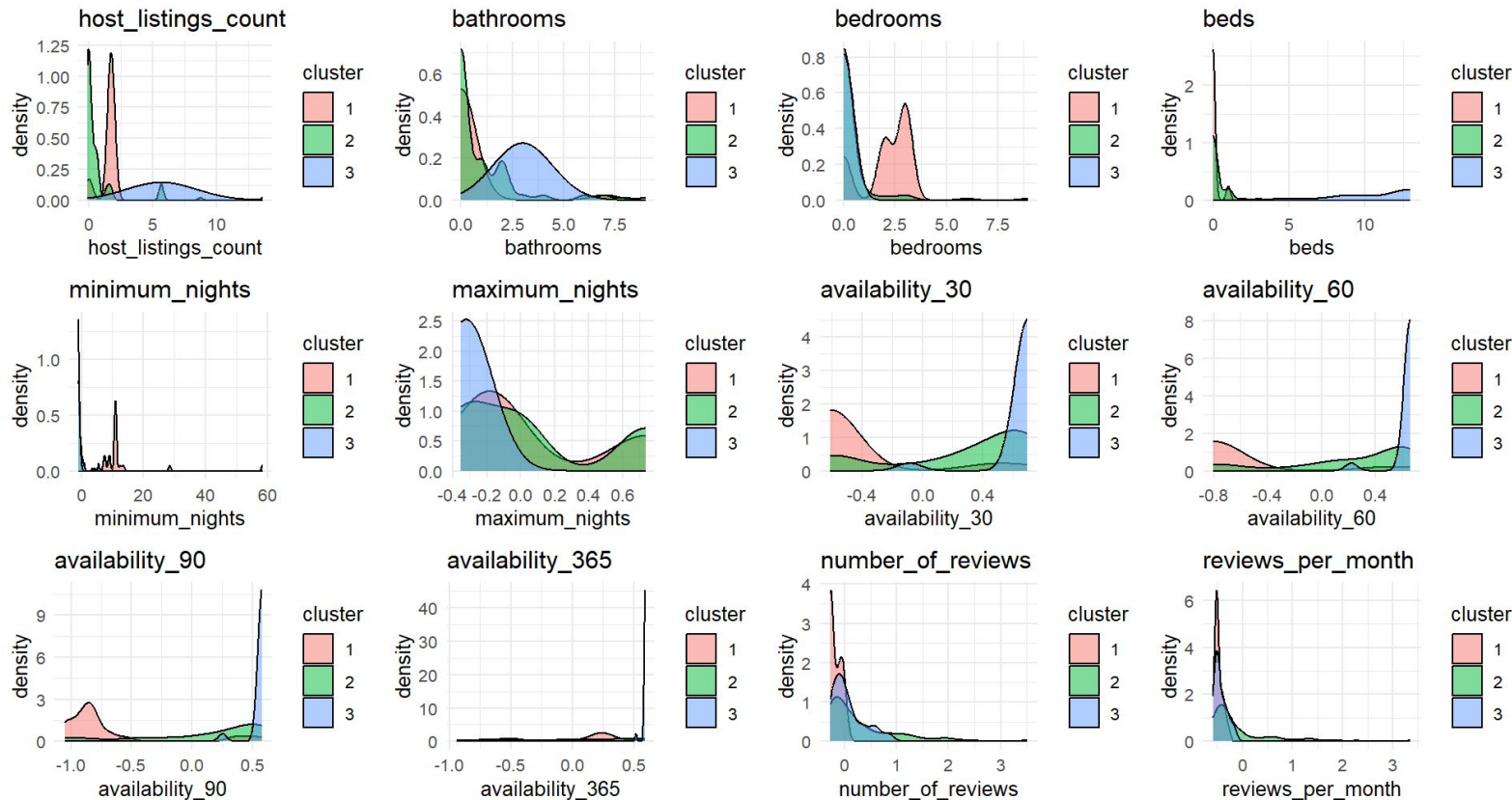
**Número óptimo de clusters: 3**

# Visualización de Clusters





# Análisis de Clusters



# Clusters

## Cluster 1: -> Host experimentado

- 2 o 3 casas medianas con 2-3 cuartos.
- Disponibilidad en el corto y mediano plazo

## Cluster 2: -> Host casual

- Particulares con una sola publicación
- Mínimo bajo cubriendo un área más grande, no tan especializado

## Cluster 3: -> Host profesional

- Muchas propiedades, probablemente administrado profesionalmente
- Con muchas camas y baños pero pocos cuartos
- Pocas noches y mucha disponibilidad



# Conclusiones

- La mayoría de los departamentos están alrededor de los **\$5000** y **\$16000** por día
- La mayoría de los departamentos de AirBnB en CABA estan en **Palermo**
  - Pero los más **caros** están en **Puerto Madero**
- La mayoría de los departamentos de AirBnB son **casas** o **departamentos enteros**
- La cantidad de **camas** y **baños** ofrecidos impactan en el precio
- La mayoría de los hosts verifican su cuenta con **mail** y **teléfono**
- Las hosts que **responden rápido** alquilan la **casa entera** o un **cuarto en un hotel**
- Los **hosts casuales** suelen tener la mayor **variación** en sus precios, mientras que los **experimentados y profesionales** ya conocen a su público y por lo tanto **alquilan más**.



Instituto Tecnológico  
de Buenos Aires

**Gracias!**