

# PREDICCIONES DE PRECIOS EN AIRBNB EN SEATTLE



Ingeniería Industrial  
PONTIFICIA UNIVERSIDAD CATOLICA DE VALPARAISO



Un enfoque basado en un modelo de Regresión Lineal Múltiple

## Objetivo

Desarrollar un modelo predictivo basado en regresión lineal que permita estimar el precio de propiedades en Airbnb utilizando variables clave como ubicación, tipo de propiedad, capacidad de hospedaje y comodidades, con el fin de ayudar a los anfitriones a establecer precios competitivos en la plataforma, ya que esta se ha vuelto muy popular.



## Metodología

### 1. Análisis Exploratorio de Datos

- Limpieza y preparación de la base de datos.
- Identificación de correlaciones entre variables como *accommodates*, *bathrooms* y *bedrooms* con el precio.

### 2. Modelado Predictivo

- Regresión Lineal Simple (RLS): Exploración inicial con variables individuales.
- Regresión Lineal Múltiple (RLM): Inclusión de variables categóricas (tipo de habitación, tipo de propiedad) y geográficas (distancia al centro), entre otras.

### 3. Transformaciones:

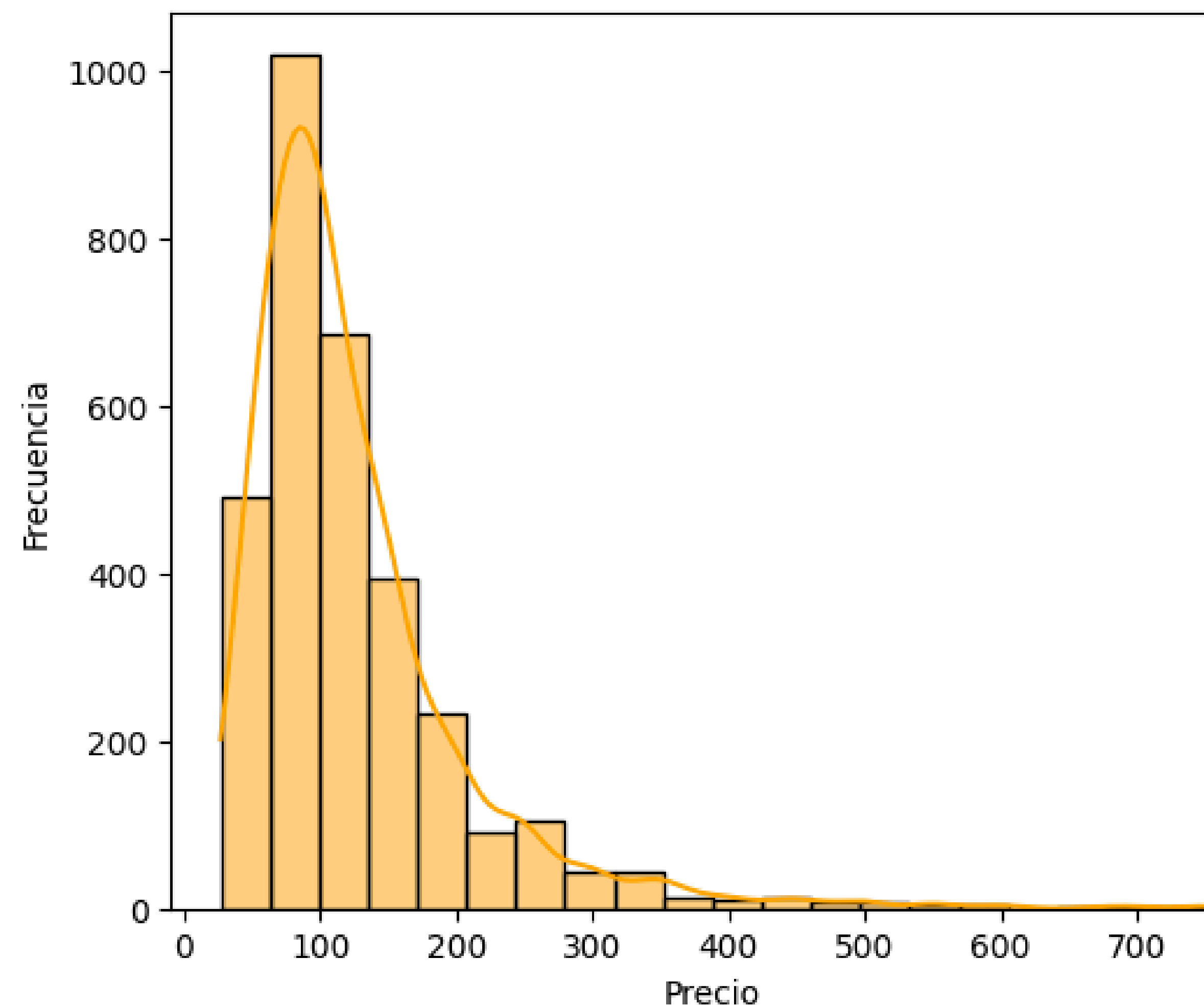
- Del precio utilizando logaritmo para estabilizar la varianza.
- División en rangos de precios (low, mid, high) para analizar patrones específicos.

### 4. Validación del Modelo con evaluación de supuestos de regresión:

- Homocedasticidad, normalidad de residuos, errores independientes e idénticamente distribuidos, valores p-value y estadístico F.



Distribución de Precios en Airbnb



$R^2$

66,8 %

LOW\_PRICE

MSE

5,804

56,1 %

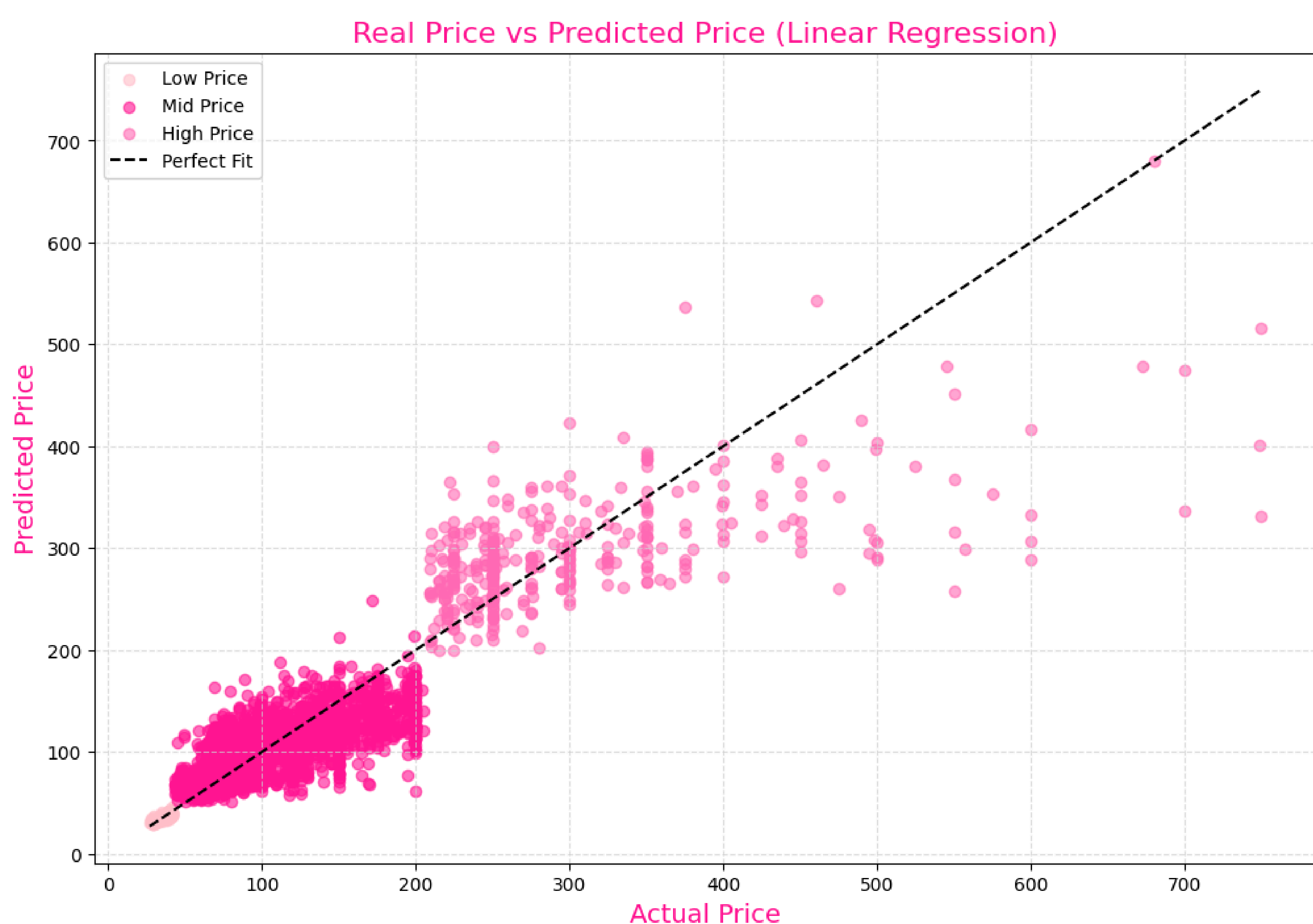
MID\_PRICE

771,591

37,1 %

HIGH\_PRICE

6900,5951



**Autores:**  
-Paulina Acevedo Gonzalez  
-Francisca Ainol Valero  
-Krishna Rojas Tapia



**Contacto:**  
paulina.acevedo.g@mail.pucv.cl

## Resultados

- El modelo de predicción fue más preciso en el rango de precios bajo (low).
- Los precios son más sensibles a la ubicación y los amenities, ya que el  $R^2$  aumentó significativamente al incluir estas variables.
- Para precios medios y altos, el modelo tiene dificultades para predecir el precio (subestima o sobreestima)
- Existe una gran variabilidad en los datos.

### Análisis por Precios

- Bajos: Buen desempeño con residuos uniformes y sin patrones visibles.
- Medios: Ajuste moderado, aunque se detectó leve heterocedasticidad.
- Altos: Dificultades significativas debido a alta variabilidad en los residuos.

### Desafíos Persistentes:

- Heterocedasticidad en precios altos.
- Normalidad parcial en residuos.