

# PREDICTING AIRBNB PRICES WITH MACHINE LEARNING



David Arroyo

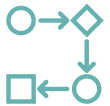
# FLUJO DE TRABAJO

## 1. EXPLORATORY DATA ANALYSIS (EDA)



1. Manejo de Nulos
2. Exploración de Variables
3. Feature Engineering (nuevas variables y reducción de características)
4. Manejo de Outliers

## 3. BUSQUEDA Y ENTRENAMIENTO DE MODELOS



1. Exploración y evaluación de diversos modelos
2. Acotación a los modelos con mejor **RMSE**
3. Búsqueda de hiperparámetros (GridSearchCV)
4. Proceso iterativo

## 2. PREPROCESAMIENTO DE DATOS



1. Búsqueda de correlaciones
2. One Hot Encoding
3. Normalización de características numéricas

## 4. PREDICCIONES

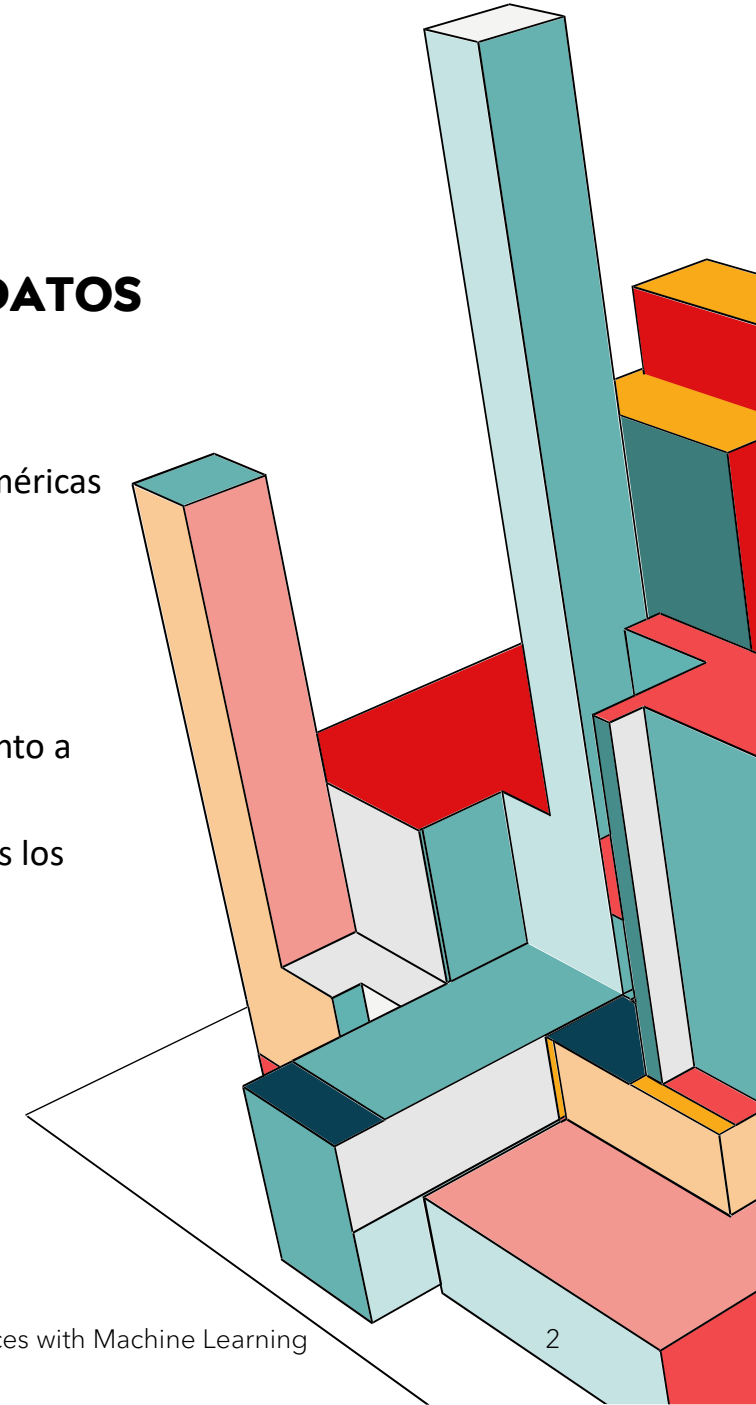


1. Transformación de datos Test (conjunto a predecir)
2. Entrenamiento del modelo con todos los datos
3. Evaluación del modelo y predicción

## 5. AUTOML



H2o.ai



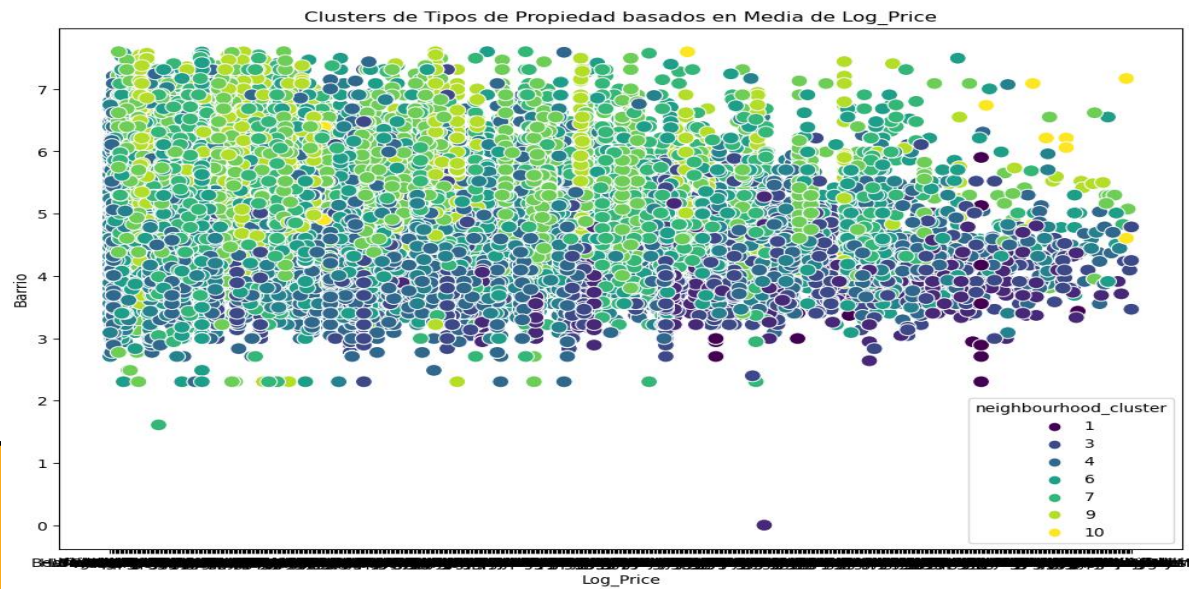
# EDA

## VALORES NULOS

1. Relleno de nulos de variables numéricas con las medias según el número de accommodates
2. Aplicación de **NearestNeighbors** con las columnas Latitude y Longitude para encontrar el barrio más cercano.

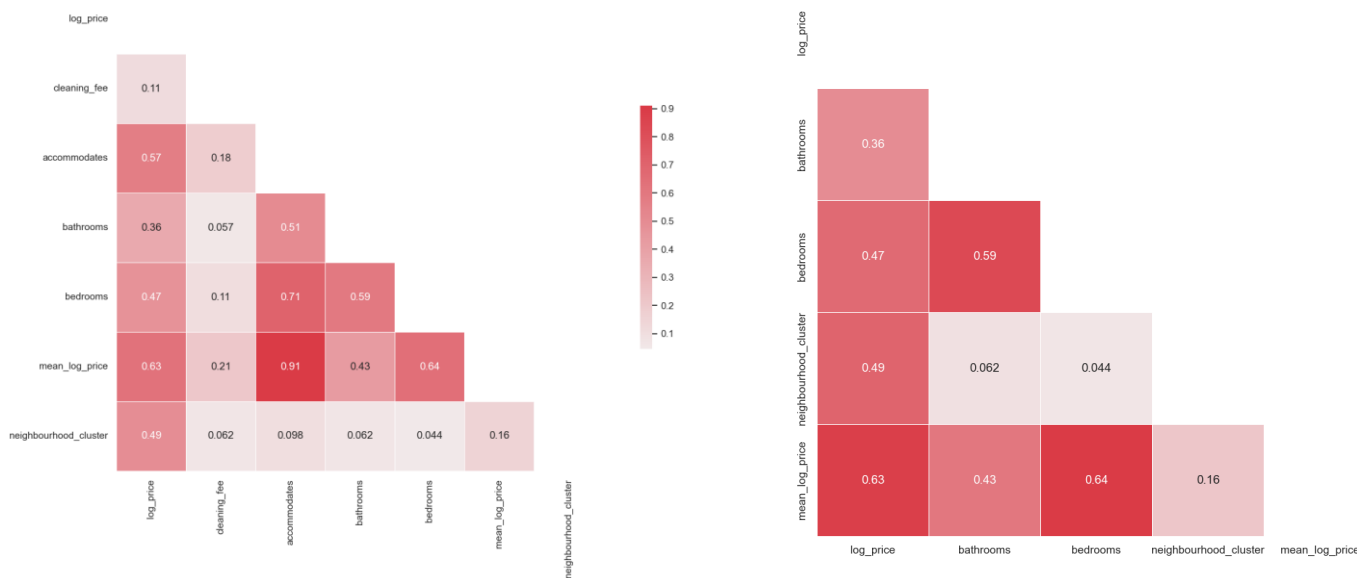
## REDUCCIÓN DE VARIABLES

1. Eliminación de columnas con poca correlación con el precio o de análisis complejo (amenities, name, description)
2. Cluster de barrios en función de la media de **log\_price** (con **k-means**).
3. Creación de una nueva variable **mean\_log\_price** como la media de los pares property\_type-accommodates



# PREPROCESAMIENTO DE DATOS

## CORRELACIONES VARIABLES NUMÉRICAS



## SELECCIÓN DE VARIABLES

### Categorías (One Hot Encoding)

- room\_type
- cancellation\_policy
- city

### Numéricas

- neighbourhood\_cluster
- bedrooms
- bathrooms
- mean\_log\_price

# MEJORES MODELOS

Random Forest Regressor

**RMSE  
0.442**

XGBRegressor

**RMSE  
0.417**

GradientBoostingRegressor

**RMSE  
0.428**

CatBoostRegressor

**RMSE  
0.408**

LGBMRegressor

**RMSE  
0.410**

# AUTOML (H2O.AI)



H2O.ai

GBM\_2\_AutoML\_10\_20231128\_11700

**RMSE**  
**0.389**

# MUCHAS GRACIAS

