



PREDICCIÓN INTELIGENTE DE PRECIOS EN AIRBNB MADRID

Predice, compara y encuentra el precio justo

MÁSTER EN DATA SCIENCE

Presentado por:
Katherine López Ramírez

C O N T E N I D O

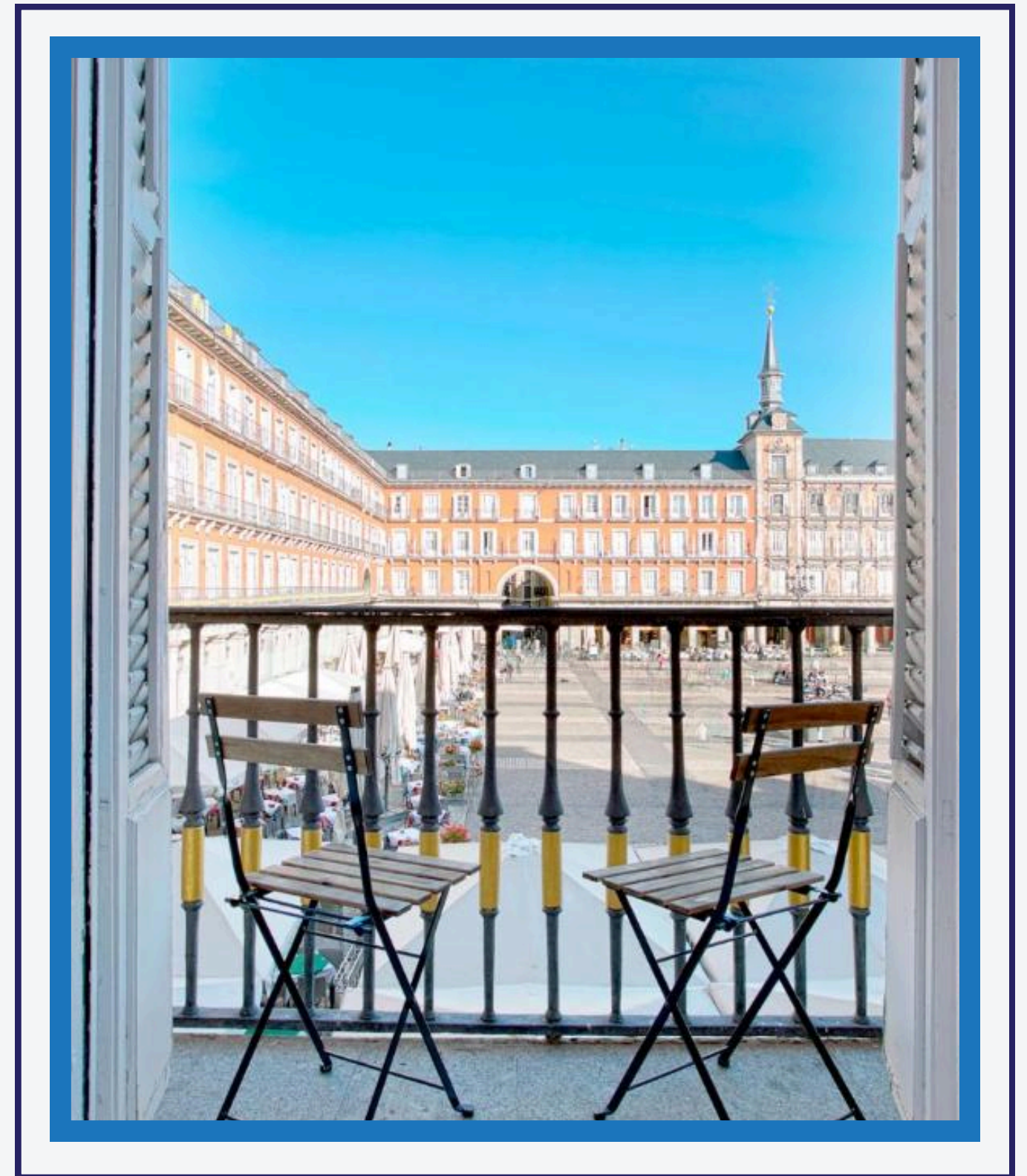
Contenidos del Proyecto

01. Contexto y justificación	3
02. Objetivo del proyecto	4
03. análisis de datos	5 - 6
04. Comparación de Modelos	7 - 9
05. Visualización y aplicación	10 -11
06. Tecnologías utilizadas	12
07. Conclusiones	13
08. Futuras mejoras	14

CONTEXTO Y JUSTIFICACIÓN

Bienvenidos al Proyecto

- Crecimiento del turismo en Madrid y auge de Airbnb.
- Problema: fijar precios óptimos en un mercado competitivo.
- Oportunidad: un modelo de predicción de precios que ayuda a anfitriones a ajustar tarifas según zona y estacionalidad, a turistas a encontrar mejores opciones de alojamiento y a inversores a tomar decisiones estratégicas.



OBJETIVO DEL PROYECTO

General: Analizar y predecir los precios de alojamientos en Airbnb Madrid según ubicación y estacionalidad, usando Machine Learning.

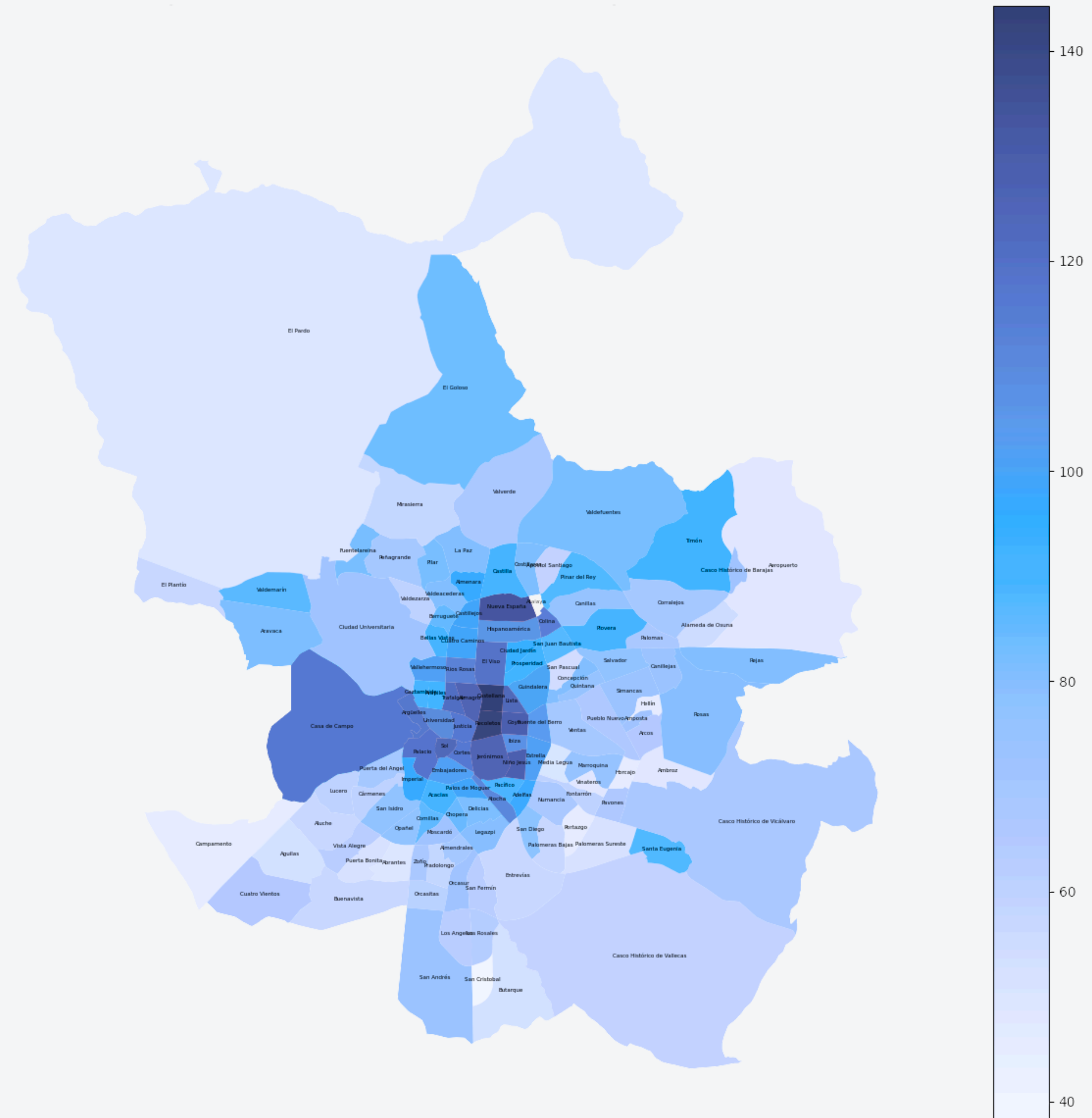
Específicos:

- Desarrollar un modelo de Machine Learning que prediga precios con precisión
- Facilitar comparaciones por meses y barrios
- Generar insights útiles para anfitriones, viajeros e inversores

ANÁLISIS DE DATOS

Contenidos del Proyecto

- Fuente: Inside Airbnb
- 25.000 alojamientos, 9M registros de calendario
- Procesamiento: limpieza, detección de outliers, integración de estacionalidad.
- Variables clave: ubicación, tipo de alojamiento, reseñas, disponibilidad, precio mensual promedio

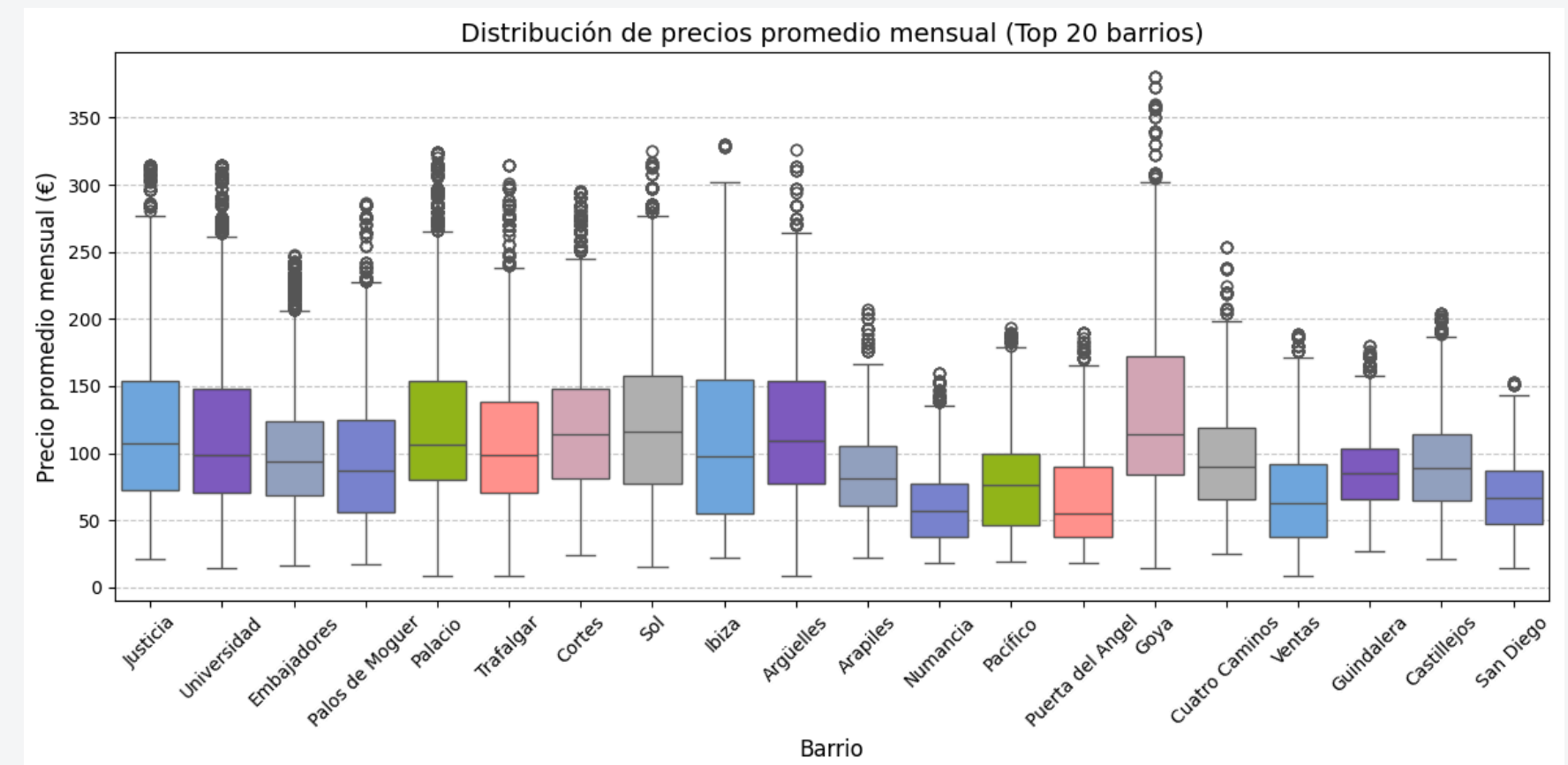
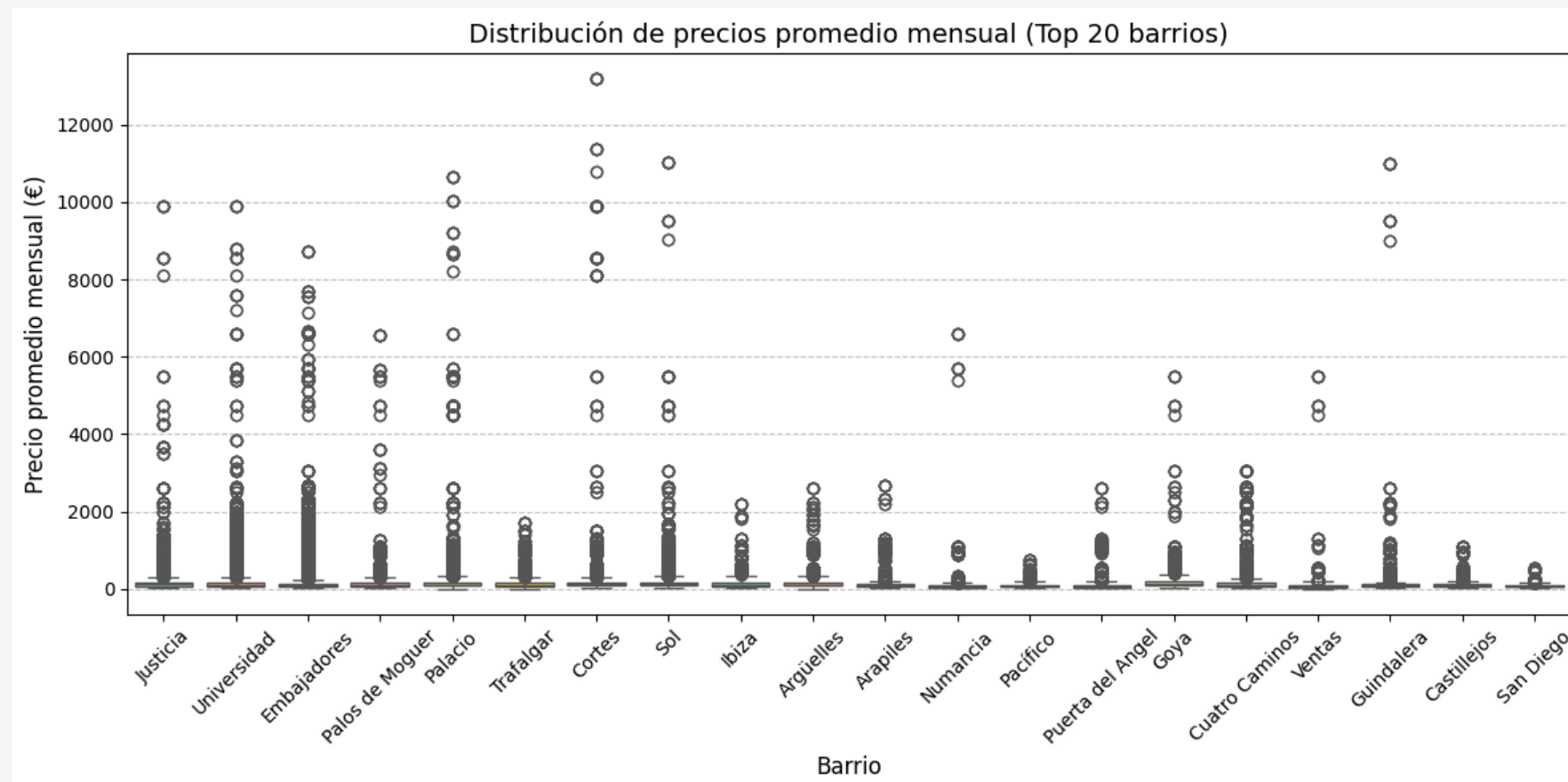


LIMPIEZA Y PREPARACIÓN

- Eliminación de nulos y outliers extremos (> 12.000 €)
- Normalización de variables del host y reseñas
- Creación de la variable de estacionalidad (precio promedio mensual por barrio)

Antes

Despues



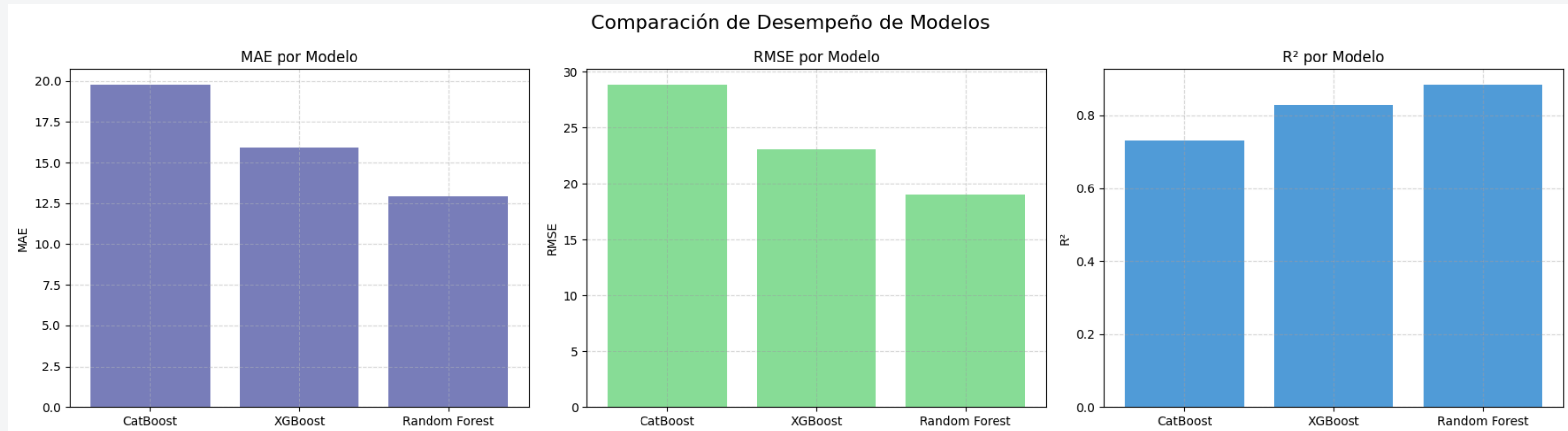
COMPARACIÓN DE MODELOS

- **CatBoost** → MAE: 19,75 €, RMSE: 28,81 €, R^2 : 0,731
- **XGBoost** → MAE: 15,91 €, RMSE: 23,04 €, R^2 : 0,828
- **Random Forest** → **MAE: 12,90 €, RMSE: 18,99 €, R^2 : 0,883**

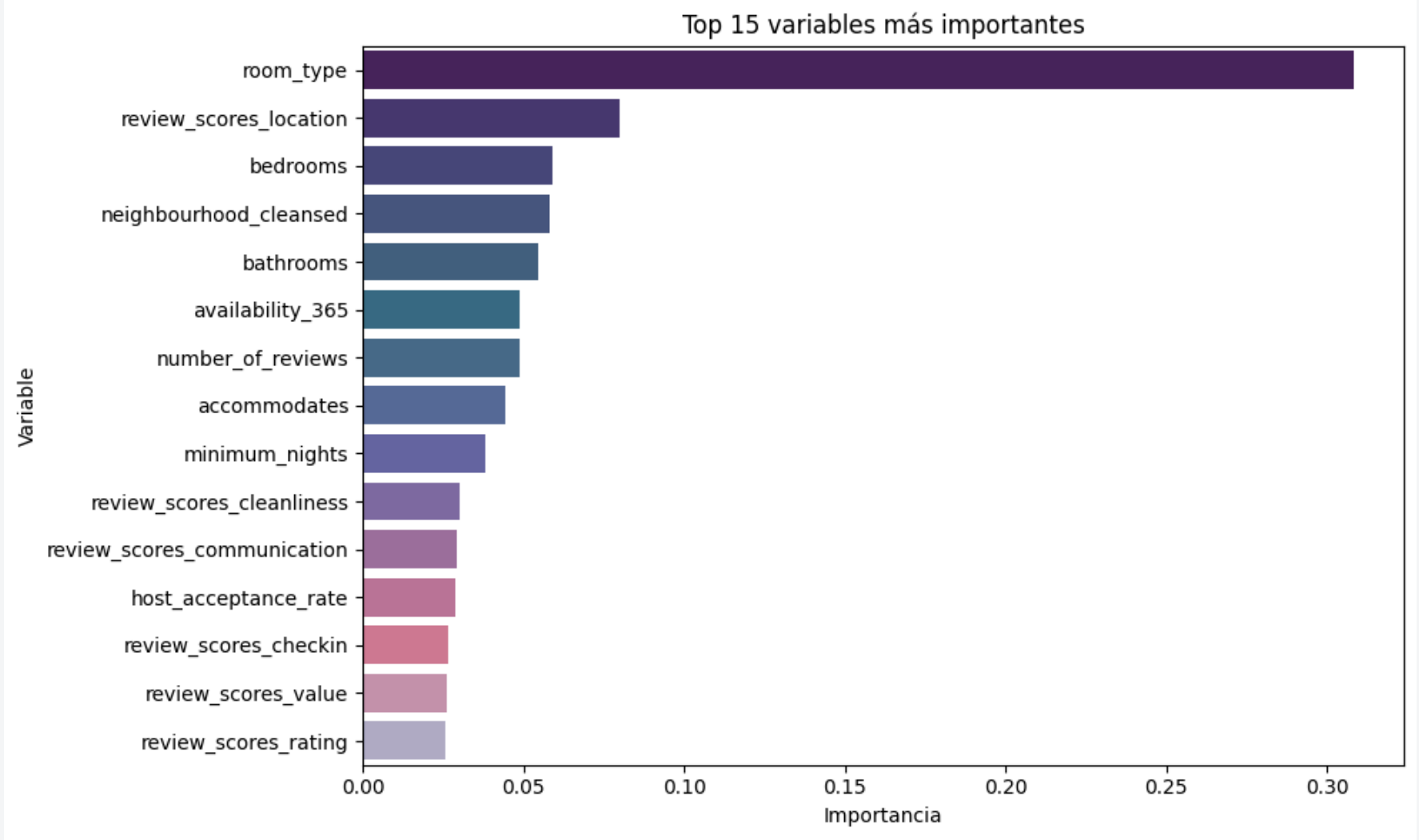
Encodings aplicados

- **CatBoost:** maneja categorías de forma nativa.
- **XGBoost:** requiere Target Encoding para convertir categorías en valores numéricos representativos.
- **Random Forest:** funciona mejor con Ordinal Encoding, evitando la alta cardinalidad y manteniendo buen rendimiento.

Conclusión: Cada encoding se eligió según las necesidades del algoritmo. Entre todos, Random Forest destaca por ofrecer el mejor rendimiento, estabilidad y robustez, siendo la opción final para el proyecto.

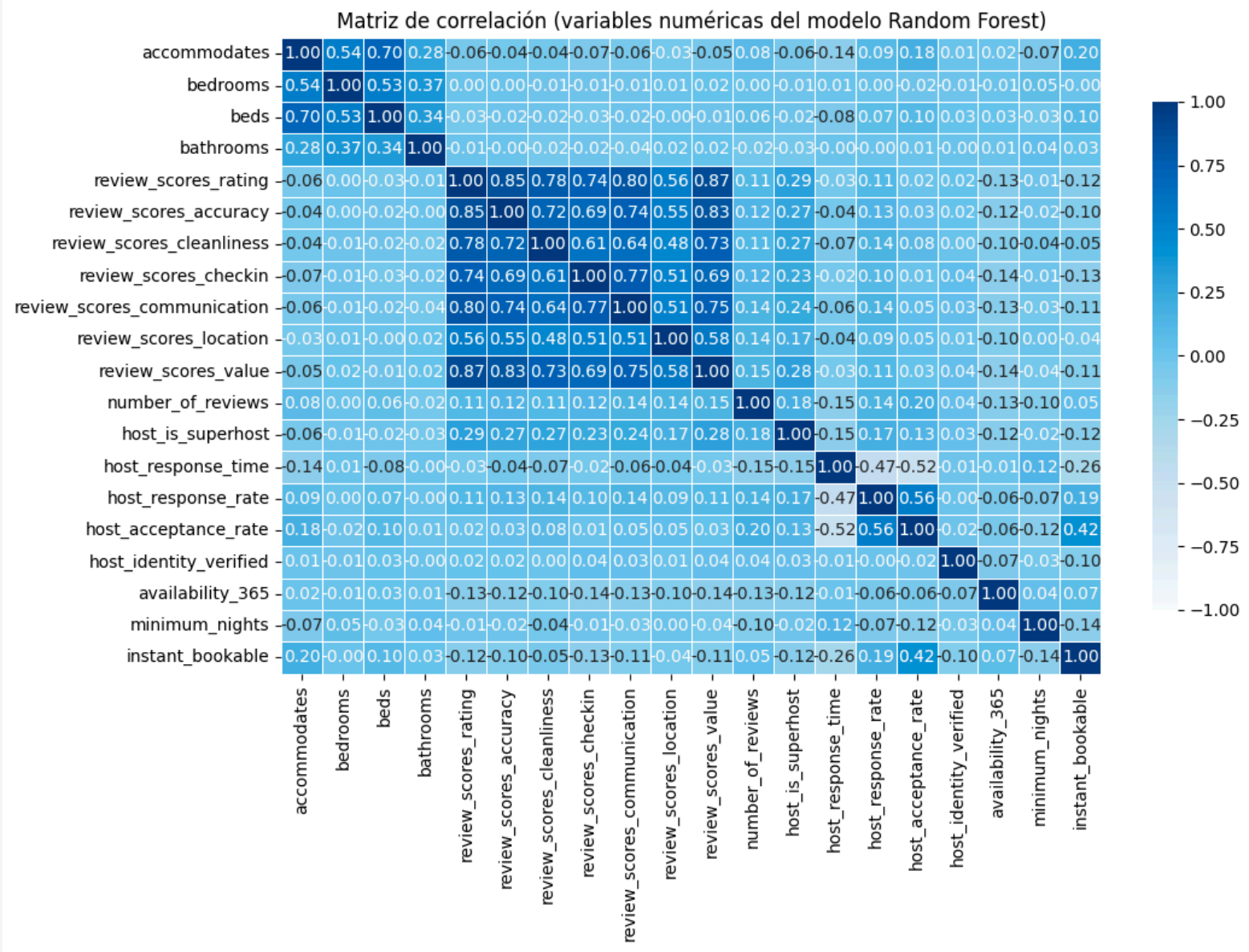


MATRIZ DE CORRELACIÓN Y VARIABLES IMPORTANTES



Importancia de las variables

Confirma que tipo de alojamiento, ubicación y reputación son claves en el mercado Airbnb Madrid.



Matriz de correlación

- Reseñas: alta correlación entre sí.
- Capacidad: relación fuerte (accommodates, bedrooms, beds).
- Otras (mínimo noches, disponibilidad): baja correlación, info extra.
- Random Forest gestiona redundancias → modelo robusto.

RESULTADOS

- Predicción por mes: precios ajustados a la estacionalidad
- Comparativa de barrios: identifica zonas más económicas o caras respecto al elegido
- Métricas finales:
 - Train: R^2 0.895, MAE 11.9 €
 - Test: R^2 0.883, MAE 12.9 €

Interpretación

- **MAE \approx 13 €** → predicciones muy precisas.
- **$R^2 \approx 0.90$** → explica el 90% de la variabilidad.
- **Métricas train/test similares** → sin sobreajuste, buen equilibrio.

VISUALIZACIÓN Y APLICACIÓN PRÁCTICA

- Comparativa de precios por meses (estacionalidad).
- Alternativas de barrios más caros y más baratos.
- Mapa interactivo con predicciones por zona.

Barrio

Argüelles

Tipo de alojamiento

Apartamento entero

Personas

3

Habitaciones

2

Baños

2

Mes

mayo

☒ Comparar con otros meses

Meses a comparar

enero

marzo

junio

septiembre

Precio estimado en mayo

208 €

Comparativa por meses

Mes	Precio (estimado)	seleccionado
mayo	208 €	<input checked="" type="checkbox"/>
enero	207 €	<input type="checkbox"/>
marzo	207 €	<input type="checkbox"/>
junio	207 €	<input type="checkbox"/>
septiembre	206 €	<input type="checkbox"/>
diciembre	206 €	<input type="checkbox"/>

Ningún mes cae en ≤ 50 €. Sugerencia: **septiembre** (~206 €).

> Mín/Máx estimado (estos meses)

Deploy

Predicción de precio • Airbnb Madrid

Predice, compara y encuentra el precio justo en Airbnb Madrid

Alternativas por barrio

Precios más cercanos al barrio elegido

Barrio	Precio (estimado)	Δ vs elegido
Abrantes	208 €	+0 €
Buenavista	208 €	+0 €

Alternativas más baratas

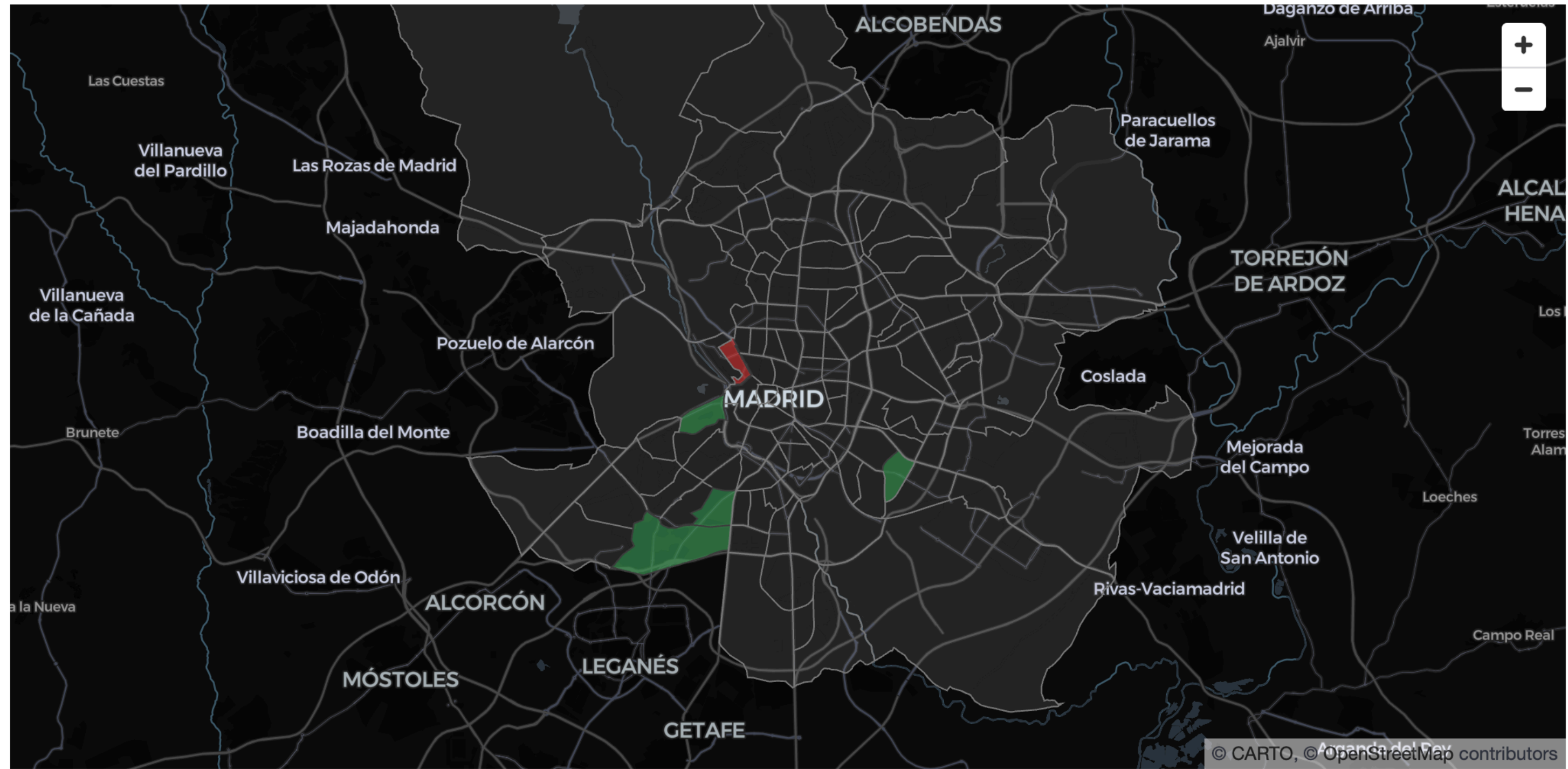
Barrio	Precio (estimado)	Δ vs elegido
Puerta del Angel	184 €	-25 €
Portazgo	184 €	-25 €

Resumen de busqueda

Tus preferencias:

Barrio: Argüelles | Tipo: Apartamento entero | Personas: 3 | Habitaciones: 2 | Baños: 2 | Mes: mayo | Precio estimado: 208 €

MAPA DE BARRIOS



TECNOLOGÍAS UTILIZADAS



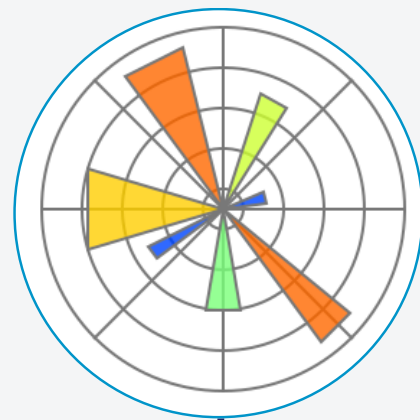
PYTHON

Lenguaje principal de desarrollo.



PANDAS & NUMPY

Limpieza, unión y manipulación de datos.



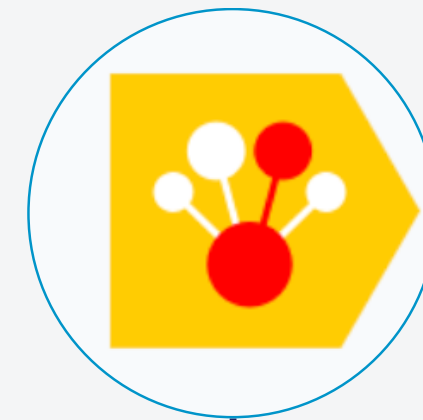
MATPLOTLIB & SEABORN

Visualización de datos y análisis exploratorio.



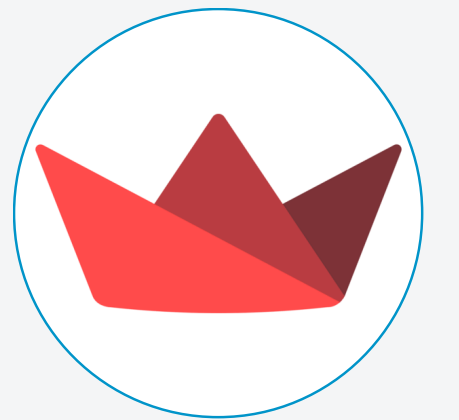
SCIKIT-LEARN

Preprocesamiento, métricas y validación de modelos.



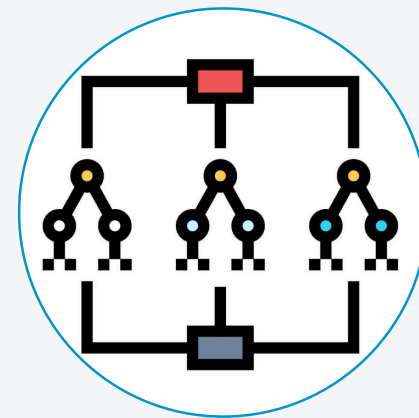
**CATBOOST,
XGBOOST,
RANDOM FOREST**

Modelos de predicción y comparación.



STREAMLIT


Interfaz interactiva para mostrar resultados de forma práctica.



CONCLUSIONES

- 1 El modelo ofrece predicciones sólidas y útiles
- 2 Apoya decisiones de anfitriones, viajeros e inversores
- 3 Es replicable en otras ciudades → potencial de escalabilidad
- 4 Proyecto con potencial de integrarse en soluciones comerciales (ej. PriceLabs, AirDNA)

FUTURAS MEJORAS

- 1 Integración con APIs de datos en tiempo real.
 - 2 Optimización de hiperparámetros con técnicas avanzadas.
 - 3 Extender análisis a otros mercados turísticos.
- 



CIERRE

**“CONVERTIR DATOS EN DECISIONES
INTELIGENTES PARA UN TURISMO MÁS
COMPETITIVO”**



MUCHAS
GRACIAS