

REPORTE FINAL COMPLETO – Airbnb NYC

Análisis, Limpieza, EDA, Modelo KNN y Dashboard

1. Introducción

Este proyecto analiza el mercado de Airbnb en NYC mediante limpieza, EDA, visualización, modelo KNN optimizado y dashboard interactivo.

2. Objetivos

- Objetivo de negocio: recomendaciones de pricing basadas en datos.
- Objetivo analítico: construir un modelo capaz de predecir precios y explicar sus factores.

3. Dataset

Dataset de Kaggle: Airbnb NYC Open Data (2019), 48,895 registros y 16 columnas.

4. Limpieza

Se corrigieron NaN, valores corruptos ("Auto%#") y tipos incorrectos sin eliminar registros. Se normalizaron columnas confillna(), replace() y astype().

5. EDA

Se analizaron patrones de precio por borough, tipo de habitación, reviews, disponibilidad y ubicación. Factores turísticos, inseguridad y trabajos temporales influyen en la demanda.

6. Modelo KNN

Se aplicó LabelEncoding, train/test split y GridSearchCV para optimizar "k". Resultados:

- MAE: 18–29 USD
- RMSE: 40–60 USD
- R²: 0.35–0.50

7. Dashboard Streamlit

Incluye selección de variables, optimización automática de "k", visualizaciones, mapa, predicciones y tabla de vecinos.

8. Conclusiones

El proyecto integró limpieza, análisis, predicción y despliegue interactivo con resultados coherentes.

9. Futuro

Incorporar datos del NYPD, modelos avanzados y un análisis temporal completo.