

Informe ejecutivo y técnico



HORIZONTE
INMOBILIARIA
Tu Hogar, Tu Futuro

Horizonte Inmobiliaria

Pipeline de datos y modelo de pricing
Secciones para CEO y equipo de Ciencia de Datos

Índice

1. Resumen ejecutivo
2. Problemática de negocio
3. Solución funcional (pipeline y UI)
4. Solución técnica (backend y modelo)
5. Métricas y operación
6. Próximos pasos

1. Resumen ejecutivo

- Contexto: panel web para ejecutar el pipeline de limpieza y generar dataset de pricing.
- Objetivo: datos consistentes y modelo de precios para publicar inmuebles con respaldo analítico.
- Rol del CEO: habilitar uso operativo y medir impacto en velocidad de venta y desvío de precio.

2. Problemática de negocio

- Decisiones de precio subjetivas (percepción o pedido del propietario).
- Inventario sobrevaluado o subvaluado → rotación ineficiente y riesgo de margen.
- Fuentes de datos heterogéneas y con calidad dispar (nulos, outliers, monedas).

3. Solución funcional (pipeline y UI)

- UI React con pestañas: Pipeline, Pricing, Dashboard, Chatbot, Documentación.
- Carga de CSV (propiedades, dólar) y ejecución de etapas: limpieza, conversión, dataset de pricing.
- Dashboard Power BI embebido para seguimiento de inventario y demanda.
- Chatbot embebido (PowerApps) para consultas rápidas sobre pipeline y datos.

Flujo de uso

- Subir fuentes o usar las existentes en backend/uploads.
- Ejecutar Pipeline completo (o etapas) → genera PropiedadesLimpio_v4.csv.
- Crear dataset de pricing → pricing_dataset.csv y entrenar modelo.
- Consultar Dashboard y simulador para definir precio de publicación.

4. Solución técnica (backend y modelo)

- Backend Flask + Socket.IO: server.py orquesta etapas y emite estado en tiempo real.
- Limpieza: normaliza columnas, corrige lat/long, filtra nulos y outliers, convierte moneda a USD.
- Dataset de pricing: deriva precio USD/m², ratio cubierta y cruza índice MI_DAN_AX03 por periodo.
- Modelo: GradientBoostingRegressor (sklearn) con split 80/20, imputación y one-hot; métricas MAE/MAPE.

Artefactos clave

- PropiedadesLimpio_v4.csv (salida pipeline).
- pricing_dataset.csv (features para modelo).
- pricing_model.pkl (pipeline entrenado, por operación).

5. Métricas y operación

- MAE (USD) y MAPE (%) visibles en UI tras reentrenar; recalcular cuando cambien datos.
- KPI de negocio: desvío de precio publicado vs recomendado y rotación del inventario.
- Alertas en UI si falta MI_DAN_AX03 o si se usan fuentes personalizadas sin validar.

Gobierno y seguridad

- Fuentes subidas se guardan en backend/uploads; se recomienda versionado/backup.
- Permisos: validar acceso a Power BI y PowerApps para Dashboard y Chatbot.

6. Próximos pasos

- Automatizar despliegue (CI/CD) y backups de CSV/modelo.
- Monitoreo de métricas del modelo y drift de datos (mix de propiedades, zonas).
- Explorar features adicionales: amenities, antigüedad, piso/orientación y densidad comercial.
- Documentar SLA operativo: quién ejecuta pipeline, frecuencia y responsable de revisión de métricas.