

Propuesta de Proyecto (etapa inicial)

Ciencia de datos aplicada: Primer entregable

1. Información general del proyecto	
Nombre del proyecto	Cifra
Responsables	Ignacio Bruzone, Christian Tomas Iijas, Felix Lopez Menardi
Fecha de inicio	08/2025
Organización/Compañía involucrada	Mercado Inmobiliario

2. Contexto y motivación	
Área o unidad de negocio involucrada	Tasación y fijación de precios de alquiler.
Stakeholders del proyecto <small>(Interesados en el proyecto, tomadores de decisiones)</small>	Agentes inmobiliarios, propietarios, inquilinos, inmobiliarias, analistas de datos del sector.
Problema o desafío actual del negocio	Tasación hecha de forma manual y subjetiva, dependiendo de cada agente. Precios pueden presentar gran variabilidad incluso dentro del mismo barrio. Riesgo de sobreprecio/subprecio.
Oportunidades detectadas	Datos históricos de publicaciones de alquiler (2021-2022) para crear modelo predictivo de precios ajustado a pesos constantes (por la inflación). Potencial de generar un sistema de apoyo para las decisiones de agentes y propietarios a la hora de tasar.

3. Objetivos del negocio	
Objetivo principal del proyecto	Desarrollar un modelo de ML que prediga precios de alquiler en AMBA con precisión ≥90%, reduciendo el tiempo de tasación en 85% y la variabilidad entre tasadores en 70%.

Subobjetivos específicos	<ol style="list-style-type: none">1. Realizar un análisis exploratorio de datos (EDA) para identificar relaciones entre variables (ubicación, tamaño, amenities) y precio2. Detectar y tratar problemas de calidad de datos: valores faltantes, inconsistencias, duplicados y outliers3. Generar visualizaciones geográficas que muestren la distribución espacial de precios en el AMBA4. Entrenar un modelo de Machine Learning para predecir el precio en pesos constantes5. Validar el modelo con métricas de error y establecer criterios de éxito
KPIs de éxito	<ul style="list-style-type: none">• Root Mean Absolute Error: raíz cuadrada del error promedio absoluto de predicción debe ser menor o igual al 10% del valor real• R^2 (Coeficiente de determinación) ≥ 0.75 como medida de explicación de la varianza• % de predicciones dentro de $\pm 10\%$ del precio real $\geq 80\%$

4. Beneficios esperados	
Impacto en procesos internos	Reducción del tiempo necesario para tasar una propiedad, procesos de fijación de precios más estandarizados y menos dependientes de la intuición (mayor objetividad)
Impacto en la toma de decisiones	Mayor precisión y objetividad en la fijación de precios, información clara y transparente para propietarios e inquilinos
Posibles retornos de inversión	Disminución del tiempo promedio de publicación antes de alquilar, incremento de la satisfacción de clientes (propietarios e inquilinos), posible monetización como herramienta de tasación

5. Restricciones y riesgos

Restricciones de datos (privacidad, regulaciones, calidad de datos)	<ul style="list-style-type: none">• Dataset limitado a dos años (2021–2022) puede no reflejar dinámicas más recientes (cambios regulatorios)• Posible sesgo si existe mayor densidad de datos en barrios céntricos que en zonas periféricas
Limitaciones tecnológicas	<ul style="list-style-type: none">• Procesamiento y modelado de un dataset con muchas variables requiere capacidad de cómputo adecuada• Integración con mapas interactivos para visualizaciones puede requerir librerías adicionales
Riesgos potenciales	<ul style="list-style-type: none">• Cambios macroeconómicos (inflación, regulación de alquileres) pueden alterar la validez del modelo• Riesgo de sobreajuste (modelo demasiado adaptado a los datos de 2021–2022 y con menor generalización futura)• Overfitting

6. Próximos pasos

Acciones inmediatas a tomar (Ejemplo: Recolectar datos, seleccionar técnicas, etc.)	<ol style="list-style-type: none">1. Adquisición y carga del dataset2. Exploración inicial de variables: identificar tipos de datos, distribuciones y correlaciones con el precio3. Limpieza de datos: tratamiento de valores faltantes, detección y eliminación/corrección de outliers, codificación de variables categóricas (ej: barrio, tipo de propiedad)4. Visualización geográfica: mapa del AMBA con precios promedio por polígono5. Selección de técnicas de modelado: regresión lineal múltiple, Random Forest, Gradient Boosting, etc.6. Entrenamiento y validación del modelo
--	--

	<p>usando métricas de error (MAE, RMSE, R²).</p> <p>7. Documentación en Jupyter Notebook con análisis exploratorio, limpieza y primeros modelos</p>
Cronograma estimado de etapas	<p>El proyecto se desarrollará en tres etapas: primero el análisis exploratorio y preparación de datos con limpieza, ajuste por inflación y visualizaciones geográficas del dataset de alquileres AMBA; luego el modelado y evaluación implementando desde regresión lineal hasta modelos más complejos con optimización de hiperparámetros; y finalmente la evaluación exhaustiva y entrega con análisis de interpretabilidad y documentación.</p> <ul style="list-style-type: none">- Etapa 1: Análisis Exploratorio y Preparación de Datos<ul style="list-style-type: none">- Adquisición y carga del dataset- Análisis exploratorio inicial- Limpieza profunda de datos- Feature engineering básico- Etapa 2: Modelado y Evaluación<ul style="list-style-type: none">- Preparación para modelado- Modelos baseline- Cálculo de métricas- Etapa 3: Evaluación final y entrega<ul style="list-style-type: none">- Evaluación exhaustiva y correcciones- Interpretabilidad del modelo- Producción del entregable final