

Documento de Visión y Alcance: NYC Property Price Predictor (Proyecto Express - 5 Días)

Versión: 1.0

Fecha: 2025-05-01

1. Introducción

- **1.1 Propósito:** Este documento define la visión, los objetivos, el alcance y las limitaciones de un proyecto express de 5 días para desarrollar un prototipo de herramienta de análisis predictivo. El objetivo es estimar el precio de venta de propiedades en la ciudad de Nueva York utilizando técnicas de Inteligencia Artificial (IA) y aprendizaje automático (ML).
- **1.2 Equipo:** El proyecto será ejecutado por un equipo de **5** personas: **3** Analistas de Datos/ML (con experiencia en Python), **1** Desarrollador Front-end y **1** Desarrollador Back-end.
- **1.3 Plazo:** 5 días hábiles.
- **1.4 Contexto Original vs. Adaptación:** La solicitud inicial mencionaba la predicción de demanda de productos basada en datos históricos, estacionalidad y tendencias de mercado. Dado que el dataset proporcionado ("NYC Property Sales") contiene registros de transacciones inmobiliarias y no ventas de productos, **el enfoque de este proyecto se adapta para predecir el PRECIO DE VENTA (SALE PRICE) de las propiedades en NYC**, utilizando características de las propiedades y la fecha de venta para inferir posibles patrones estacionales o tendencias observables dentro del período del dataset.

2. Oportunidad de Negocio / Problema

- El mercado inmobiliario de Nueva York es complejo y dinámico. Comprender los factores que influyen en los precios de venta y poder estimarlos basándose en datos históricos puede ofrecer información valiosa.
- Este proyecto busca explorar la viabilidad de crear rápidamente un modelo predictivo básico para estimar precios de propiedades, sirviendo como prueba de concepto para futuras herramientas más robustas.

3. Visión del Producto

- Crear un prototipo funcional que:
 - Analice el dataset histórico de ventas de propiedades en NYC.
 - Implemente un modelo de Machine Learning capaz de predecir el precio de venta (SALE PRICE) de una propiedad basándose en sus características (ej. ubicación, tamaño, tipo, antigüedad).
 - Ofrezca una interfaz de usuario sencilla para ingresar las características de una propiedad y obtener una predicción de precio estimada.
 - Explore visualmente patrones básicos o tendencias en los datos (ej. precios por distrito, influencia del tamaño, posible estacionalidad simple basada en la fecha de venta).

4. Alcance del Proyecto

- **4.1 Funcionalidades Incluidas (In-Scope):**

- **Análisis de Datos (Analistas Python):**

- Carga y limpieza básica del dataset "NYC Property Sales" (manejo de valores nulos, formatos, eliminación de ventas no representativas como las de \$0).
- Ingeniería de características básicas (ej. extraer mes/año de `SALE DATE`, convertir variables categóricas).
- Análisis Exploratorio de Datos (EDA) para identificar relaciones entre características y `SALE PRICE` (usando Pandas, Matplotlib).
- Visualizar distribuciones, correlaciones básicas, precios medios por `BOROUGH` o `BUILDING CLASS CATEGORY`.
- Entrenamiento y evaluación de *al menos un* modelo de regresión (ej. Regresión Lineal, Random Forest Regressor) para predecir `SALE PRICE` usando Scikit-learn, ya sea en Google Colab o en otro software.
- Evaluación del modelo usando métricas estándar (ej. MAE, RMSE, R^2).

- **Desarrollo Back-end (Desarrollador Back-end):**

- Creación de una API REST simple (ej. con Flask o FastAPI) que cargue el modelo entrenado y exponga un endpoint para recibir características de una propiedad y devolver la predicción del precio.

- **Desarrollo Front-end (Desarrollador Front-end):**

- Creación de una interfaz de usuario web básica (HTML/CSS/JS simple) que permita al usuario ingresar manualmente las características clave de una propiedad (ej. metros cuadrados, distrito, tipo de edificio, año de construcción).
- La interfaz enviará estos datos a la API del back-end y mostrará la predicción del precio de venta recibida.

- **Herramientas:** Python (Pandas, Scikit-learn, Matplotlib), Google Colab, Git (para control de versiones básico), Framework de API Python (Flask/FastAPI), HTML/CSS/JavaScript básico.
- **4.2 Funcionalidades Excluidas (Out-of-Scope):**
 - Integración de **datos externos** para tendencias de mercado (ej. tasas de interés, indicadores económicos de NYC). El análisis de "tendencias" se limitará a lo observable *dentro* del dataset de 1 año.
 - Análisis de series temporales complejo o predicción de **volumen** de ventas futuro. El foco es el **precio** de venta individual.
 - Desarrollo de múltiples modelos o afinamiento exhaustivo de hiperparámetros (se buscará un modelo funcional, no necesariamente el óptimo).
 - Interfaz de usuario compleja, diseño visual avanzado, autenticación de usuarios o gestión de cuentas.
 - Despliegue en un entorno de producción robusto o escalable (el prototipo funcionará localmente o en entornos de desarrollo como Colab/Heroku free tier si el tiempo lo permite).
 - Manejo exhaustivo de todos los casos borde o problemas de calidad de datos profundos en el dataset.
 - Predicción de otros aspectos como la probabilidad de venta o el tiempo en el mercado.
 - Capacidades de predicción en tiempo real o actualizaciones automáticas del modelo.

5. Criterios de Éxito (Proyecto Express)

- Entregar un prototipo funcional que cumpla con las funcionalidades *In-Scope* al final de los 5 días.
- El modelo predictivo demuestra una capacidad de predicción razonable (mejor que una predicción aleatoria o basada solo en la media).
- La interfaz de usuario permite ingresar datos y visualizar la predicción.
- El código está documentado y disponible en un repositorio compartido (ej. GitHub, Colab Notebooks compartidos).
- Realizar una breve demostración interna del prototipo al finalizar el proyecto.

6. Supuestos y Restricciones

- **Supuestos:**
 - El dataset "NYC Property Sales" proporcionado es suficiente para entrenar un modelo predictivo básico, con alrededor de 84000 muestras.
 - Las herramientas gratuitas listadas (Colab, Python libs) son adecuadas y accesibles para todos.
 - La colaboración y comunicación dentro del equipo serán eficientes para cumplir con el plazo ajustado.
 - Es aceptable un nivel de precisión del modelo "demostrativo" en lugar de "óptimo".
- **Restricciones:**
 - Tiempo: Estrictamente 5 días hábiles.
 - Recursos: Limitados al equipo de 5 personas y herramientas gratuitas.
 - Alcance: Estrictamente limitado a lo definido en la sección 4.1.
 - Dataset: Limitado al dataset provisto y sus características inherentes.

Nota: El objetivo principal es construir un prototipo rápido como ejercicio o prueba de concepto, dada la limitación de tiempo. La profundidad del análisis y la sofisticación del modelo y la aplicación serán necesariamente limitadas. La comunicación constante dentro del equipo será clave para ajustar tareas y asegurar la entrega del prototipo funcional.