



Tecnológico de Monterrey

Unidad de formación:

Gestión de proyectos de plataformas tecnológicas (Gpo 201)

Actividad 4 (Despliegue)

Profesor:

Alfredo García Suarez

Alumna:

Samantha Aja | A01738365

Fecha de entrega:

20 de octubre de 2025

Introducción

El proyecto tuvo como objetivo construir un dashboard interactivo en Streamlit que integrara análisis exploratorio, regresiones lineales y no lineales, así como un explorador automático de modelos predictivos.

El propósito fue crear una herramienta que permita evaluar relaciones entre variables, probar distintos tipos de regresión y determinar automáticamente el mejor modelo según métricas de desempeño como RMSE, MAE y R^2 .

Fases de desarrollo

1. Limpieza y Preparación de Datos

Se partió de un dataset de Airbnb (NY/MX) con columnas numéricas, categóricas y booleanas.

Se aplicó limpieza con:

- Reemplazo de valores t/f por 0/1.
- Conversión de texto a valores numéricos (price, beds, bathrooms, etc.).
- Eliminación de nulos y atípicos.

Se construyó una versión filtrada del Data Frame (dff) que sirve como fuente principal para los gráficos y modelos.

2. Diseño del Dashboard (estructura principal)

El diseño general se desarrolló en Streamlit, con componentes organizados en secciones:

- Sidebar: filtros, selección de vistas y controles del modelo.
- Main View: paneles interactivos con gráficos Plotly.
- Se estableció una paleta visual para todos los gráficos y el mismo tipo de letra para todo el dashboard.

3. Regresión Lineal

Objetivo: Explorar la relación entre variables numéricas y categóricas y construir modelos de predicción lineales.

Principales componentes:

Selección de variable dependiente y variables independientes (numéricas y categóricas).

Opciones para estandarización, partición train/test y transformación logarítmica.

Cálculo de métricas:

RMSE, MAE, R^2 en conjuntos de entrenamiento y prueba.

Gráficos generados:

- Matriz de correlación (heatmap)
- Dispersión Predicho vs Real
- Histograma de errores
- Residuos vs Predicción
- Importancia de variables (coeficientes)
- Curva de validación (Ridge)
- Curva de aprendizaje

Filtros dinámicos para elegir la variable dependiente y las features del modelo.

4. Regresión No Lineal

Objetivo: Capturar relaciones más complejas entre variables que no pueden representarse con una línea recta.

Modelos implementados:

- Lineal simple (referencia)
- Polinomial (grados configurables)
- Exponencial
- Logarítmico
- Sigmoide (curva logística)

Evaluación: Cada modelo calcula automáticamente: RMSE, MAE y R^2

Se construye una tabla comparativa (leaderboard) con los resultados ordenados.

Se destaca el mejor modelo mediante un criterio mixto (menor RMSE + mayor R^2).

Gráficos incluidos:

- Dispersión con curvas de todos los modelos superpuestos.
- Curva RMSE vs grado polinomial.
- Histograma de residuos del mejor modelo.
- Mapa de calor de residuos.

Resultado Final

El proceso de desarrollo permitió integrar conceptos estadísticos, machine learning y visualización interactiva en un solo entorno.