

Alcance y Descripción Detallada del Proyecto

Objetivo del Proyecto

El objetivo principal de este proyecto consiste en analizar la relación entre los valores evaluados de propiedades inmobiliarias (Assessed Value) y sus precios reales de venta (Sale Amount) en distintas localidades. Para lograr esto, se utilizarán datos históricos recopilados en diferentes años, lo cual permitirá obtener insights relevantes para la toma de decisiones en inversiones inmobiliarias, valoraciones fiscales y estudios de mercado inmobiliario.

Preguntas clave del Proyecto

- ¿Existe una correlación significativa entre el valor fiscal evaluado y el precio real de venta de las propiedades?
- ¿Qué localidades presentan mayores discrepancias entre los valores fiscales y los precios reales del mercado?
- ¿Es posible predecir con precisión el precio de venta basándose únicamente en el valor fiscal evaluado?

Justificación del Modelo Seleccionado

Se optó por utilizar un modelo de regresión lineal simple debido a su simplicidad interpretativa, rapidez de implementación y efectividad demostrada en análisis iniciales, donde se observó una fuerte correlación entre la variable independiente (Assessed Value) y la dependiente (Sale Amount).

Estrategias ante diferentes escenarios

Incremento de volumen de datos (100x):

Si el volumen de datos aumentara considerablemente, se implementaría una solución basada en tecnologías como Apache Spark o AWS Glue para procesamiento distribuido, permitiendo manejar eficientemente grandes volúmenes mediante procesamiento paralelo.

Ejecución diaria en ventana específica:

Para la ejecución diaria del pipeline en una ventana de tiempo específica, se utilizaría Apache Airflow o cron jobs gestionados mediante AWS Lambda o AWS Step Functions, garantizando así la automatización y el control preciso de la ejecución.

Base de datos con acceso simultáneo de más de 100 usuarios:

En caso de requerirse soporte para más de 100 usuarios simultáneos, se migraría a una base de datos escalable como PostgreSQL en Amazon RDS o una solución NoSQL como MongoDB Atlas, garantizando robustez, escalabilidad y rendimiento adecuado.

Realización de analítica avanzada:

Para la implementación de analítica avanzada, se incorporaría una infraestructura en la nube utilizando AWS SageMaker o Databricks, permitiendo ejecutar análisis predictivos avanzados, machine learning y visualización interactiva de resultados mediante plataformas como Power BI o Tableau.