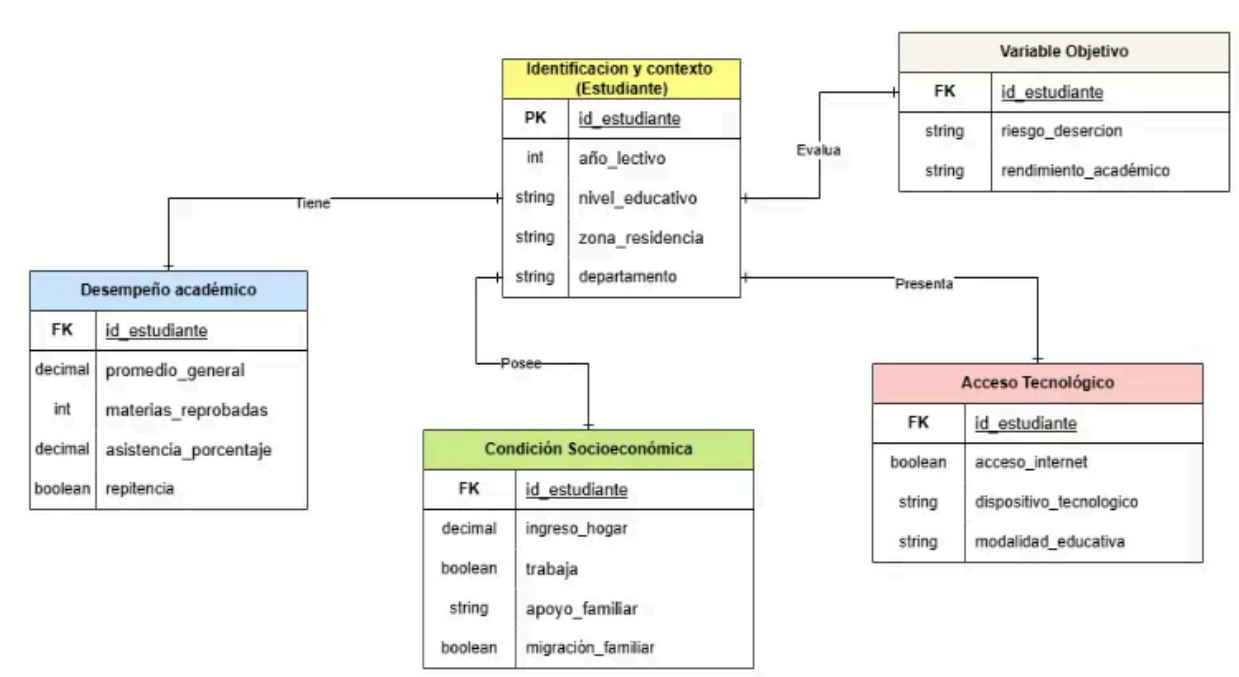


Desarrollo del Proyecto

1.1 Recopilación de datos

Debido a la imposibilidad para conseguir una muestra de población real, se incorporaron datasets simulados que replican comportamientos de un entorno académico reales basados en el análisis de los datos de la *Encuesta de Hogares de Propósitos Múltiples* y estudios como *Pobreza y desigualdad en El Salvador: Panorama general de las tendencias actuales* del Banco Mundial.



[Diagrama de los datos](#)

[Datos utilizados](#)

2.1 Limpieza y preparación de los datos

Para garantizar la calidad de la información, se utilizó Google Sheets en la etapa de procesamiento. El trabajo consistió en la eliminación de registros duplicados y la unificación de diversas tablas, logrando así una base de datos coherente y lista para ser procesada por las herramientas de visualización.

[Base de datos en Google Sheets](#)

3.1 Diseño de la arquitectura visual (Dashboard)

El proceso de diseño se ejecutó en dos niveles. Primero, se creó un prototipo funcional en Figma para definir la disposición de los elementos. Una vez validada la estructura, se

procedió a implementar el estilo visual definitivo en Looker Studio, enfocándose en que la interfaz fuera clara y fácil de navegar.

[Diseño en Figma](#)

4.1 Definición de KPIs y filtros operativos

Se identificaron los indicadores clave de desempeño (KPI) esenciales para el monitoreo educativo. Estos se integraron en el dashboard junto con filtros dinámicos que permiten segmentar la información por diferentes categorías, facilitando una lectura personalizada de los datos según la necesidad del usuario.

[Dashboard en Looker Studio](#)

5.1 Análisis predictivo con Python y API de Gemini

La parte técnica avanzada se resolvió mediante un script de Python que conecta el dataset con la API de Gemini. Este algoritmo automatiza el análisis de los datos, utilizando inteligencia artificial para identificar patrones de comportamiento. Gracias a esto, el sistema genera hipótesis automáticas que permiten predecir qué alumnos tienen mayor riesgo de abandonar sus estudios y quiénes presentan una tendencia de éxito académico.

La API del modelo recibirá un dataset de estudiantes que han desertado y verá cuales son los patrones comunes en sus historias. También recibirá un dataset de estudiantes que aún cursan la educación media, con distintos niveles de rendimiento. Las especificaciones de la consulta son las siguientes:

Prompt:

1. Cumple el rol de un científico de datos que debe generar hipótesis a partir de los patrones observados en el dataset de desertores
2. Observa los patrones que se cumplen en las personas que dejan de estudiar
3. Crea categorías de riesgo: ALTO, MODERADO y MEDIO
4. Calcula la probabilidad de que los estudiante del dataset de personas que aún siguen estudiando dejen la escuela en base a los patrones que reflejan sus datos
5. Agrega una columna con su categoría de riesgo y genera un archivo .csv y un archivo .json con los datos de los estudiantes y una

Reglas:

1. Enfócate en predecir sólo la deserción escolar, ignora el analisis predictivo de cualquier otra variable (Ejemplo: predecir si vive en una Zona Rural o Urbana, predecir si tiene acceso a internet o no, etc.)

