

**Master's Degree in Data Sciences** | EAFIT University

**Course:** Fundamentals of Data Science

**Group:** Sara Martinez Rendon (smartiner4@eafit.edu.co) | Heider Zapata (hzapata1@eafit.edu.co) | Yeison Londoño (ylondon4@eafit.edu.co)

**Date:** 04/02/2026

## Taller 4: Definición del negocio/problema

Seleccionar un negocio para analizar y responder las siguientes preguntas:

### 1. Determinar los Objetivos de Negocio

**¿Cuál es el objetivo principal del proyecto?**

(Ej: "Queremos reducir la pérdida de clientes en un 10%"). Justificar.

Optimizar el proceso de programación académica de la Universidad EAFIT prediciendo la cantidad de grupos que será necesario ofertar por asignatura (de pregrado) en el siguiente ciclo lectivo (2026-1) a partir de información histórica anonimizada (2022-2026). Se espera que este modelo permita reducir la brecha entre los cursos que se ofertan por materia y los que terminan siendo demandados realmente, evitando reprocesos que hasta ahora se realizan de manera manual en la Oficina de Registro Académico.

**Meta:** alcanzar una tasa desajuste entre los cursos esperados (ofertados) y los demandados que sea al menos un 50% inferior a la tasa de desajuste promedio de los últimos 8 semestres

**¿Qué procesos de negocio se ven afectados por este problema? (ventas, marketing, infraestructura, logística, etc). Justificar.**

- Apoyo Financiero -> reliquidar estudiantes
- Registro Académico -> cerrar grupos innecesarios, procesos manuales
- Contact Center -> atención al cliente (correos/whatsapp)
- Marketing -> imagen de la universidad / favorabilidad entre estudiantes / factor reputacional / Confianza
- Infraestructura -> distribución de aulas y espacios de aprendizaje

**¿Cómo se toma la decisión hoy en día? (¿Es por intuición, por reglas manuales o no se hace nada?).**

Actualmente, la planeación de la oferta académica se basa exclusivamente en el pre-registro. Sin embargo, la experiencia demuestra que este predictor es insuficiente; un modelo basado en datos históricos podría aportar precisión y minimizar pérdidas. Al ejecutarse de manera manual y reactiva, el proceso actual resulta lento, tedioso y propenso a ineficiencias operativas.

**¿Quiénes son los stakeholders o interesados principales? (clientes, sponsors, gerencia, coordinadores, etc). Justificar.**

Los principales interesados en este desarrollo son: Oficina de Registro Academico, Vicerrectoría de Aprendizaje, Servicios Financieros, área de Mercadeo y la comunidad universitaria.

---

## 2. Evaluar la Situación Actual

**¿De qué datos disponemos actualmente? (¿Están en una base de datos SQL, archivos Excel, o hay que recolectarlos?).**

Disponemos de 2 datasets en formato EXCEL (.xlsx) con 9 tablas cada uno. Contienen información de la programación académica y los datos de pre-matricula para el periodo 2022-2026.

Entre las variables de interés se encuentran las asignaturas, grupos por asignatura, horarios, capacidad de los grupos, estudiantes por asignatura, entre otras.

***Hacer un sondeo inicial del estado del dataset:***

**Existen espacios en blanco? Hay información duplicada? Hay inconsistencias? Hay textos mal escritos o incompletos? Hay columnas que están íntimamente relacionadas?**

- Hay missings (NAN)
- Hay Duplicados
- Sin inconsistencia
- Con algunas variables íntimamente relacionadas (pre-matricula y Matricula)

**Definir las desventajas del modelo de solución actual y ventajas del modelo de solución nuevo.**

El modelo de solución actual presenta limitaciones críticas al basar la proyección exclusivamente en la prematricula, omitiendo variables clave como el comportamiento histórico, tasas de aprobación y oferta académica. Esta visión parcial genera ineficiencias

operativas y un alto volumen de procesos manuales para ajustar la oferta al inicio del semestre.

En contraste, el nuevo modelo propone un enfoque analítico que busca garantizar una predicción más precisa de los grupos necesarios. Sus beneficios principales incluyen la optimización de recursos al reducir el cierre de grupos innecesarios (o la apertura de aquellos que no fueron proyectados), así como la automatización de tareas, permitiendo que el personal administrativo migre de labores operativas tediosas hacia actividades de mayor impacto estratégico para la institución.

---

### 3. Determinar los Objetivos del Modelo de Aprendizaje

#### ¿Qué tipo de problema de datos es?

Modelo de regresión

#### ¿Qué queremos que el modelo prediga exactamente?

Predecir la cantidad de estudiantes, y con esto, el número de grupos que se deben abrir por cada asignatura el próximo semestre

#### ¿Qué nivel de precisión o error es aceptable para que el modelo sea útil?

Una tasa desajuste entre los cursos esperados (ofertados) y los demandados que sea al menos un 50% inferior a la tasa de desajuste promedio de los últimos 8 semestres.

---