



# PROYECTO DE **TOMA DE DECISIONES I**

Realizado por Castillo, Gutiérrez, Martínez y Ramos

# INTRODUCCIÓN

## ÁREA DE INTERÉS

El rendimiento académico de los estudiantes es un indicador clave del éxito educativo y un reflejo del entorno socioeconómico, las políticas escolares y los hábitos individuales de estudio. Entender qué factores afectan significativamente las calificaciones puede ayudar a diseñar estrategias más efectivas para mejorar los resultados educativos. En particular, el análisis de datos puede revelar patrones que de otro modo serían difíciles de identificar, como el impacto del apoyo familiar, los hábitos de consumo de alcohol, y el acceso a recursos como internet o apoyo escolar.

Para este proyecto, utilizamos el conjunto de datos "Student Performance" del repositorio de UCI Machine Learning, que proporciona información detallada sobre estudiantes de secundaria en Portugal. Este conjunto de datos contiene variables relacionadas con características demográficas, familiares, escolares y personales de los estudiantes.



# DESCRIPCIÓN DEL CONJUNTO DE DATOS SELECCIONADO

El conjunto de datos utilizado en este proyecto proviene de la base de datos "Student Performance" disponible en el repositorio de UCI Machine Learning. Este conjunto de datos incluye información sobre el desempeño académico de estudiantes de secundaria en Portugal, específicamente en las asignaturas de Matemáticas y Lengua Portuguesa. Contiene 32 variables y 649 registros, con detalles sobre características demográficas, familiares, escolares y personales. Las variables clave incluyen:

- Notas: G1 (parcial 1), G2 (parcial 2), G3 (nota final).
- Apoyo escolar y familiar: schoolsup, famsup.
- Hábitos de estudio: studytime, failures
- Estas variables serán analizadas para identificar relaciones significativas con el rendimiento académico.



# METODOLOGIA

1

## Elección del área de interés y conjunto de datos

- Se seleccionó el rendimiento académico como área de interés, debido a su relevancia en la toma de decisiones educativas y en la mejora de estrategias pedagógicas.
- El conjunto de datos elegido fue "Student Performance", disponible en el repositorio de UCI Machine Learning, el cual proporciona información sobre estudiantes de secundaria en Portugal.

2

## Creación de la base de datos

- Los datos fueron organizados y almacenados en una base de datos SQLite para garantizar un manejo eficiente y estructurado.
- Se diseñó una tabla que reflejara las variables del conjunto de datos, permitiendo un acceso sencillo y organizado a la información.

3

## Análisis exploratorio de datos (EDA)

- Usando pandas, los datos fueron importados desde la base de datos SQLite.
- Se realizó un análisis estadístico básico, calculando métricas como la media, mediana, moda y desviación estándar para identificar tendencias generales.
- Se verificó la calidad de los datos, identificando valores faltantes o atípicos. Estos valores se trataron de manera adecuada para evitar sesgos en el análisis



# METODOLOGIA

4

## Transformaciones y filtrado de datos

- Se realizaron filtrados y normalizaciones necesarias para preparar las variables clave para el análisis, como las calificaciones (G1, G2, G3), horas de estudio (studytime) y el apoyo escolar (schoolsup).
- Las variables categóricas fueron codificadas para facilitar su análisis e interpretación.

5

## Visualización de datos

- Se generaron gráficos relevantes con matplotlib y seaborn, como gráficos de barras, dispersión y pastel, para visualizar las relaciones entre las variables clave.
- Cada visualización incluyó títulos, etiquetas y leyendas para facilitar su comprensión.

6

## Desarrollo de una aplicación interactiva

- Utilizando Streamlit, se diseñó una aplicación web interactiva que permite a los usuarios:
  - Explorar estadísticas descriptivas y gráficos interactivos.
  - Filtrar los datos por categorías relevantes, como colegio, género, o apoyo escolar.
  - Obtener insights personalizados para la toma de decisiones.



# BASE DE DATOS

Nombre de la variable	Role	Tipo	Demográfico	Descripción	Unidades	Valores faltantes
escuela	Característica	Categorico		escuela del estudiante (binario: 'GP' - Gabriel Pereira o 'MS' - Mousinho da Silveira)		No
sexo	Característica	Binario	Sexo	sexo del estudiante (binario: 'F' - femenino o 'M' - masculino)		No
edad	Característica	Entero	Edad	Edad del estudiante (numérica: de 15 a 22)		No
DIRECCIÓN	Característica	Categorico		Tipo de dirección del domicilio del estudiante (binario: 'U' - urbano o 'R' - rural)		No
tamaño de fam	Característica	Categorico	Otro	tamaño de la familia (binario: 'LE3' - menor o igual a 3 o 'GT3' - mayor que 3)		No
Estado de P	Característica	Categorico	Otro	estado de cohabitación de los padres (binario: 'T' - viven juntos o 'A' - separados)		No
Medio	Característica	Entero	Nivel de educación	Educación de la madre (numérico: 0 - ninguna, 1 - educación primaria (4º grado), 2 - 5º a 9º grado, 3 - educación secundaria o 4 - educación superior)		No
Fedú	Característica	Entero	Nivel de educación	Educación del padre (numérico: 0 - ninguna, 1 - educación primaria (4º grado), 2 - 5º a 9º grado, 3 - educación secundaria o 4 - educación superior)		No
M-job	Característica	Categorico	Ocupación	trabajo de la madre (nominal: 'maestra', relacionado con la atención 'sanitaria', 'servicios' civiles (por ejemplo, administrativos o policiales), 'en casa' u 'otros')		No
Trabajo	Característica	Categorico	Ocupación	trabajo del padre (nominal: 'maestro', relacionado con la atención 'sanitaria', 'servicios' civiles (por ejemplo, administrativos o policiales), 'en casa' u 'otros')		No

Nombre de la variable	Role	Tipo	Demográfico	Descripción	Unidades	Valores faltantes
escuela	Característica	Categorico		escuela del estudiante (binario: 'GP' - Gabriel Pereira o 'MS' - Mousinho da Silveira)		No
sexo	Característica	Binario	Sexo	sexo del estudiante (binario: 'F' - femenino o 'M' - masculino)		No
edad	Característica	Entero	Edad	Edad del estudiante (numérica: de 15 a 22)		No
Granjero	Característica	Entero		Calidad de las relaciones familiares (numérica: de 1 - muy mala a 5 - excelente)		No
tiempo libre	Característica	Entero		Tiempo libre después de la escuela (numérico: de 1 - muy bajo a 5 - muy alto)		No
salir	Característica	Entero		Salir con amigos (numérico: de 1 - muy bajo a 5 - muy alto)		No
Dalc	Característica	Entero		Consumo de alcohol en jornada laboral (numérico: de 1 - muy bajo a 5 - muy alto)		No
Walc	Característica	Entero		Consumo de alcohol durante el fin de semana (numérico: de 1 - muy bajo a 5 - muy alto)		No
salud	Característica	Entero		Estado de salud actual (numérico: de 1 - muy malo a 5 - muy bueno)		No
ausencias	Característica	Entero		Número de ausencias escolares (numérico: de 0 a 93)		No

# INFORMACIÓN



GitHub



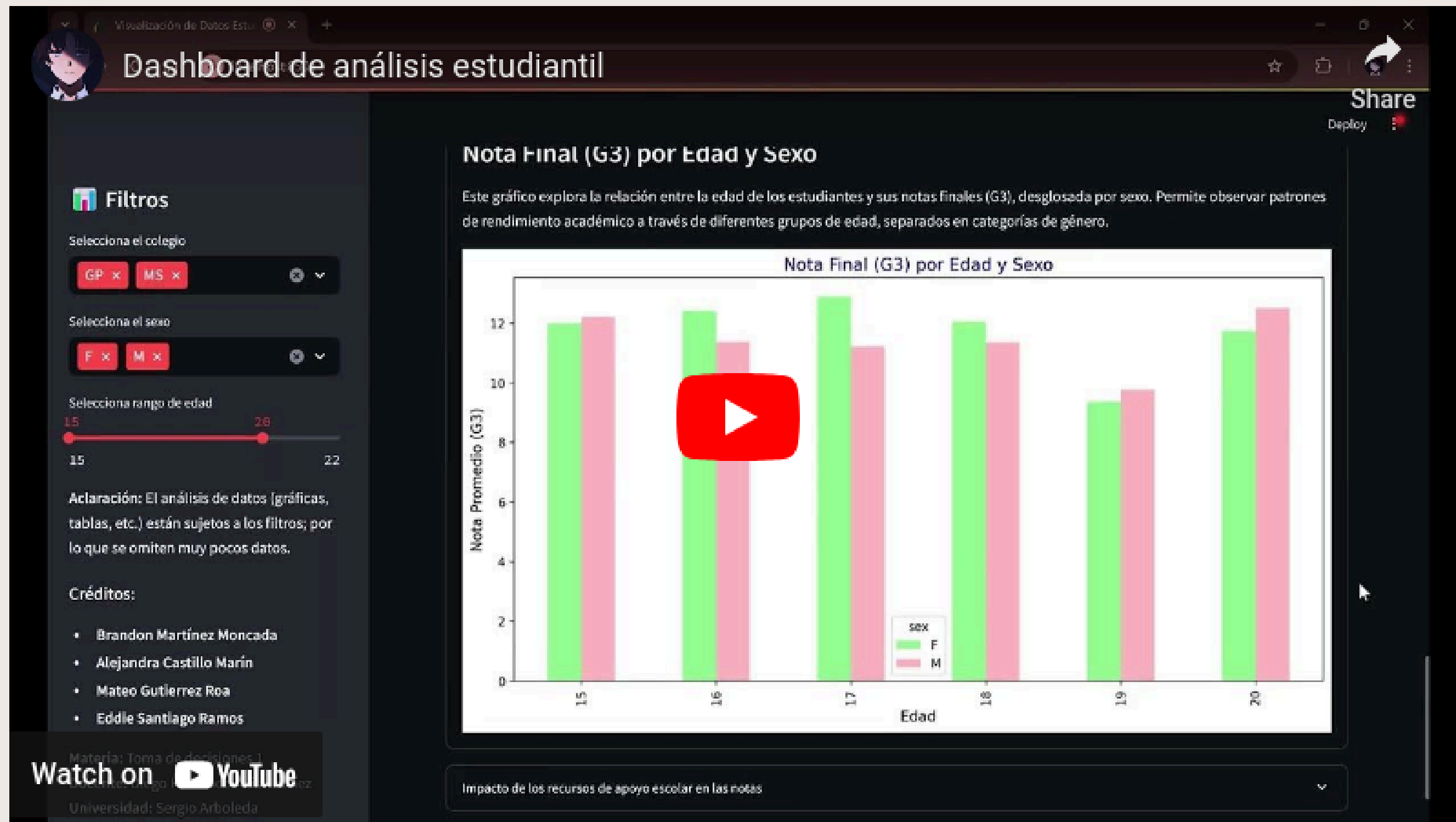
Colab 1



video



# DASHBOARD







**GRACIAS**