

## ✓ Práctica 1: Uso de aplicaciones colaborativas para el uso de herramientas de inteligencia artificial

### Actividades a realizar

- Selección de una herramienta colaborativa: Google Colab
- Definición del proyecto: Análisis de Rendimiento Académico
- Configuración del entorno: Local y en la nube

**Problemática:** Este proyecto pretende simular datos académicos de estudiantes para analizar cómo las variables denominadas horas de estudio, asistencia y nivel socioeconómico influyen en el promedio final.

- Proceso de instalación de las librerías:
- `pip install Faker`
- `!pip install pygwalker`

```
!pip install Faker
!pip install pygwalker
```


 [Mostrar el resultado oculto](#)

- Proceso para importar las librerías

```
from faker import Faker
import pandas as pd
import numpy as np
import random
import pygwalker as pyg
```

- Proceso para importar el módulo drive de Google Colab y vincular Google Drive

```
from google.colab import drive
drive.mount('/content/drive')
```

 Drive already mounted at /content/drive; to attempt to forcibly remount, call `drive.mount("/content/drive", force_remount=True)`.

- Proceso para generar los datos falsos en idioma español y con formato latino
- Creación de categorías que se usarán para generar registros simulados
- Estructura para generar datos falsos

```
fake = Faker('es')
```

```

provincias = ['Esmeraldas', 'Guayas', 'Manabi', 'Pichincha', 'Tungurahua', 'Loja', 'Chimborazo']
niveles_ed = ['Secundaria', 'Bachillerato']
niveles_se = ['Bajo', 'Medio', 'Alto']

```

```

data = []
for i in range(10000):
    horas = random.randint(0, 40)
    asistencia = round(random.uniform(50, 100), 1)
    promedio = round(0.05*horas + 0.04*asistencia + random.normalvariate(1, 0.5), 2)
    promedio = min(max(promedio, 0), 10)

```

```

data.append({
    'id': i+1,
    'nombre_completo': fake.name(),
    'edad': random.randint(15,25),
    'provincia': random.choice(provincias),
    'nivel_educativo': random.choice(niveles_ed),
    'horas_estudio_semana': horas,
    'asistencia_%': asistencia,
    'internet_en_casa': random.choice([True, False]),
    'nivel_socioeconomico': random.choice(niveles_se),
    'promedio_final': promedio
})

```




```

df = pd.DataFrame(data)
print("Tamaño del DataFrame:", df.shape) # (10000, 10)
print("Cantidad total de filas:", len(df)) # 10000
display(df.head())
display(df.tail())

```



Tamaño del DataFrame: (10000, 10)  
Cantidad total de filas: 10000

	id	nombre_completo	edad	provincia	nivel_educativo	horas_estudio_semana	asistencia_%	internet_en_casa	nivel_socioeconomico	promedio_final	
0	1	Dionisio Lupe Valentín Peiró	25	Chimborazo	Bachillerato	3	57.2	True	Alto	3.69	
1	2	Quirino Otero Pagès	25	Guayas	Bachillerato	21	58.2	True	Medio	5.09	
2	3	Gastón Ferrando Cámara	18	Guayas	Bachillerato	18	58.6	True	Bajo	4.10	
3	4	Plácido Reyes Gálvez	17	Esmeraldas	Bachillerato	36	68.6	True	Bajo	5.17	
4	5	Máxima Conesa	17	Pichincha	Secundaria	5	71.8	True	Alto	3.73	
	id	nombre_completo	edad	provincia	nivel_educativo	horas_estudio_semana	asistencia_%	internet_en_casa	nivel_socioeconomico	promedio_final	
9995	9996	Nico Medina-Vélez	24	Chimborazo	Secundaria	29	50.6	False	Bajo	4.89	
9996	9997	Tere Rosalina Segura Hoyos	24	Tungurahua	Bachillerato	19	88.1	False	Bajo	4.23	
9997	9998	Silvestre Diez Cabañas	16	Esmeraldas	Secundaria	11	99.5	False	Bajo	6.27	
9998	9999	Gustavo Miralles	15	Guayas	Secundaria	23	65.8	True	Medio	5.50	
9999	10000	Roberta Becerra Araujo	25	Guayas	Secundaria	9	61.1	False	Bajo	3.52	

Guarda tu DataFrame df como un archivo CSV con el nombre Byron\_datos.csv

```
df.to_csv('Byron_datos.csv', index=False)
```

Copia el archivo CSV recién creado hacia Google Drive

```
!cp Byron_datos.csv "/content/drive/MyDrive/Colab Notebooks/Practica_1/"
```

Define una variable ruta con la ubicación completa del archivo dentro de Google Drive

```
ruta = "/content/drive/MyDrive/Colab Notebooks/Practica_1/Byron_datos.csv"
```

Carga el archivo CSV desde Google Drive y lo guarda en el nuevo DataFrame datos

```
datos = pd.read_csv(ruta)
```

## ✓ Enriquecimiento del Dataset y Análisis Exploratorio

A partir del conjunto de datos generado con 10.000 registros simulados, se agregaron **cuatro nuevas columnas derivadas** para enriquecer el análisis:

- **categoria\_edad**: clasifica a los estudiantes en tres grupos etarios (Adolescente, Joven, Adulto joven) según su edad.
- **es\_aplicado**: identifica a los estudiantes que dedican más de 20 horas de estudio por semana y tienen más del 80% de asistencia.
- **acceso\_total\_tecnologia**: indica si el estudiante tiene internet en casa y pertenece al nivel socioeconómico Medio o Alto.
- **rendimiento\_academico**: categoriza el promedio final en tres niveles (Bajo, Medio, Alto), según los siguientes rangos:

```
df['categoria_edad'] = pd.cut(df['edad'], bins=[14, 17, 20, 25], labels=['Adolescente', 'Joven', 'Adulto joven'])
df['es_aplicado'] = (df['horas_estudio_semana'] > 20) & (df['asistencia_%'] > 80)
df['acceso_total_tecnologia'] = df.apply(
    lambda row: row['internet_en_casa'] and row['nivel_socioeconomico'] in ['Medio', 'Alto'],
    axis=1
)
df['rendimiento_academico'] = pd.cut(
    df['promedio_final'],
    bins=[-1, 6, 8, 10],
    labels=['Bajo', 'Medio', 'Alto']
)
df.head()
```

	id	nombre_completo	edad	provincia	nivel_educativo	horas_estudio_semana	asistencia_%	internet_en_casa	nivel_socioeconomico	promedio_final	categoria_edad	es_aplicado	acceso_
0	1	Dionisio Lupe Valentín Peiró	25	Chimborazo	Bachillerato	3	57.2	True	Alto	3.69	Adulto joven	False	

1	2	Quirino Otero Pagès	25	Guayas	Bachillerato	21	58.2	True	Medio	5.09	Adulto joven	False
2	3	Gastón Ferrando Cámara	18	Guayas	Bachillerato	18	58.6	True	Bajo	4.10	Joven	False
3	4	Plácido Reyes Gálvez	17	Esmeraldas	Bachillerato	36	68.6	True	Bajo	5.17	Adolescente	False

Próximos pasos:

[Generar código con df](#)

[Ver gráficos recomendados](#)

[New interactive sheet](#)

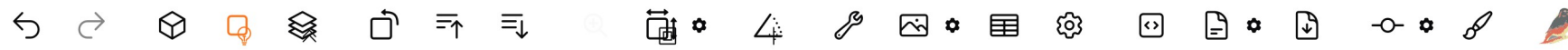
```
walker = pyg.walk(datos)
```

Data Visualization Chat

Chart 1 Chart 2 : + New

What visualization your want to draw from the dataset

Ask



Field List

- nombre\_completo
- edad
- provincia
- nivel\_educativo
- internet\_en\_casa
- nivel\_socioeconomico
- Measure names
- id
- horas\_estudio\_semana
- asistencia\_%
- promedio\_final
- Row count
- Measure values

Filters

Color

nivel\_socioeconomico

Opacity

Drop Field Here

Size

asistencia\_%

Shape

Drop Field Here

Details

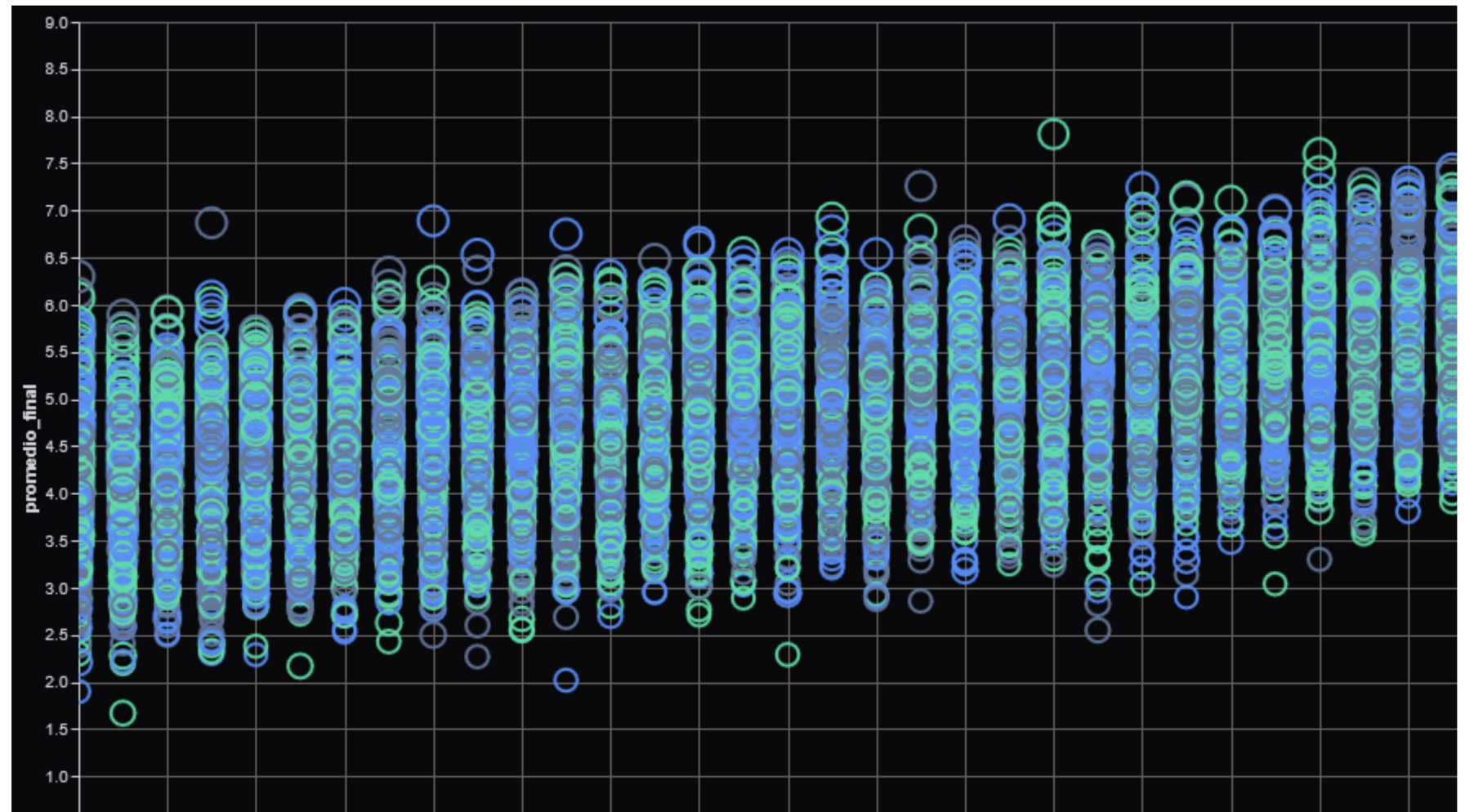
Drop Field Here

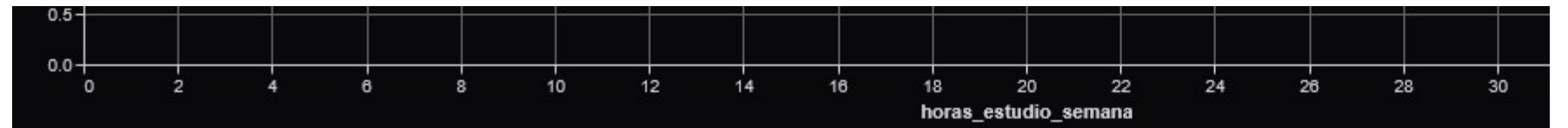
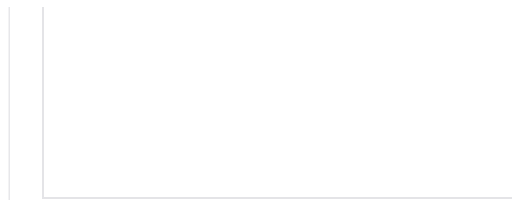
X-Axis

horas\_estudio\_semana

Y-Axis

promedio\_final





Haz doble clic (o ingresa) para editar

Haz doble clic (o ingresa) para editar

