

## Activity 2 <https://github.com/Heli-Aguilar/retoAnalitica.git>

```
from google.colab import files #Cargar archivo a colab
uploaded = files.upload()
```

Elegir archivos student\_hab...rmance.csv

- **student\_habits\_performance.csv**(text/csv) - 73663 bytes, last modified: 6/5/2025 - 100% done  
Saving student\_habits\_performance.csv to student\_habits\_performance.csv

```
import pandas as pd
```

```
df = pd.read_csv('student_habits_performance.csv') # Reemplaza con el nombre de tu archivo
print(df.head(4)) # Muestra las primeras filas
```

```
student_id  age  gender  study_hours_per_day  social_media_hours \
0      S1000   23  Female                0.0                1.2
1      S1001   20  Female                6.9                2.8
2      S1002   21   Male                1.4                3.1
3      S1003   23  Female                1.0                3.9

netflix_hours  part_time_job  attendance_percentage  sleep_hours \
0             1.1           No                 85.0           8.0
1             2.3           No                 97.3           4.6
2             1.3           No                 94.8           8.0
3             1.0           No                 71.0           9.2

diet_quality  exercise_frequency  parental_education_level  internet_quality \
0         Fair                  6                Master      Average
1         Good                  6                High School    Average
2         Poor                  1                High School    Poor
3         Poor                  4                Master      Good

mental_health_rating  extracurricular_participation  exam_score
0                   8                             Yes      56.2
1                   8                             No     100.0
2                   1                             No      34.3
3                   1                             Yes      26.8
```

```
valor = df.iloc[2, 3]
print(valor)
```

```
1.4
```

```
cantidad_objetos = len(df)
print("Número de objetos:", cantidad_objetos)
```

```
Número de objetos: 1000
```

```
valor_variable2_objeto15 = df.iloc[14, 1] # columna 1 → variable 2
print("Valor de la variable 2 del objeto 15:", valor_variable2_objeto15)
```

```
Valor de la variable 2 del objeto 15: 24
```

```
minimo = df.iloc[:, 2].min()
maximo = df.iloc[:, 2].max()
print("Mínimo de la variable 3:", minimo)
print("Máximo de la variable 3:", maximo)
```

```
Mínimo de la variable 3: Female
Máximo de la variable 3: Other
```

```
resumen = df.describe()
print(resumen)
```

```
count      age  study_hours_per_day  social_media_hours  netflix_hours \
count  1000.00000      1000.00000      1000.000000      1000.000000
mean    20.49800       3.55010       2.505500       1.819700
std      2.30810       1.46889       1.172422       1.075118
min     17.00000       0.00000       0.000000       0.000000
25%     18.75000       2.60000       1.700000       1.000000
50%     20.00000       3.50000       2.500000       1.800000
75%     23.00000       4.50000       3.300000       2.525000
max     24.00000       8.30000       7.200000       5.400000
```

	attendance_percentage	sleep_hours	exercise_frequency \
count	1000.000000	1000.000000	1000.000000
mean	84.131700	6.470100	3.042000
std	9.399246	1.226377	2.025423
min	56.000000	3.200000	0.000000
25%	78.000000	5.600000	1.000000
50%	84.400000	6.500000	3.000000
75%	91.025000	7.300000	5.000000
max	100.000000	10.000000	6.000000

	mental_health_rating	exam_score
count	1000.000000	1000.000000
mean	5.438000	69.601500
std	2.847501	16.888564
min	1.000000	18.400000
25%	3.000000	58.475000
50%	5.000000	70.500000
75%	8.000000	81.325000
max	10.000000	100.000000

## ✓ Sin usar Pandas

```
import csv
```

```
# Abrir el archivo CSV
with open('student_habits_performance.csv', newline='') as csvfile:
    lector = csv.reader(csvfile)
    filas = list(lector)
```

```
celda = filas[2][3]
print("Valor en fila 2, columna 3:", celda)
```

➡ Valor en fila 2, columna 3: 6.9

```
import csv
```

```
n = 5 # Número de filas a mostrar
```

```
with open('student_habits_performance.csv', newline='') as csvfile:
    lector = csv.reader(csvfile)

    # Leer y mostrar las primeras n filas (incluyendo el encabezado si está presente)
    for i, fila in enumerate(lector):
        print(fila)
        if i + 1 >= n:
            break
```

➡ ['student\_id', 'age', 'gender', 'study\_hours\_per\_day', 'social\_media\_hours', 'netflix\_hours', 'part\_time\_job', 'attendance\_percentage',  
['S1000', '23', 'Female', '0.0', '1.2', '1.1', 'No', '85.0', '8.0', 'Fair', '6', 'Master', 'Average', '8', 'Yes', '56.2']  
['S1001', '20', 'Female', '6.9', '2.8', '2.3', 'No', '97.3', '4.6', 'Good', '6', 'High School', 'Average', '8', 'No', '100.0']  
['S1002', '21', 'Male', '1.4', '3.1', '1.3', 'No', '94.8', '8.0', 'Poor', '1', 'High School', 'Poor', '1', 'No', '34.3']  
['S1003', '23', 'Female', '1.0', '3.9', '1.0', 'No', '71.0', '9.2', 'Poor', '4', 'Master', 'Good', '1', 'Yes', '26.8']

```
# Revisar encabezado para saber el índice correcto
print("Encabezado:", filas[0])
```

```
# Suponiendo que la variable numérica correcta está en la columna 3 (índice 3)
valores_variable3 = []
```

```
for fila in datos:
    try:
        valor = float(fila[3]) # Cambia el índice según corresponda
        valores_variable3.append(valor)
    except ValueError:
        pass # Ignora filas con valores no numéricos (por si hay errores o valores vacíos)
```

```
minimo = min(valores_variable3)
maximo = max(valores_variable3)
print("Mínimo de la variable 3:", minimo)
print("Máximo de la variable 3:", maximo)
```

```

Encabezado: ['student_id', 'age', 'gender', 'study_hours_per_day', 'social_media_hours', 'netflix_hours', 'part_time_job', 'attendance_p
Mínimo de la variable 3: 0.0
Máximo de la variable 3: 8.3

```

## Basic statistical data visualization

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

```

```

df = pd.read_csv('student_habits_performance.csv')
df.head()
print(df.columns)

```

```

plt.figure(figsize=(6, 4))
sns.boxplot(data=df, y='study_hours_per_day') # cambia el nombre de columna si es diferente
plt.title('Boxplot de Study Hours')
plt.show()

```

```

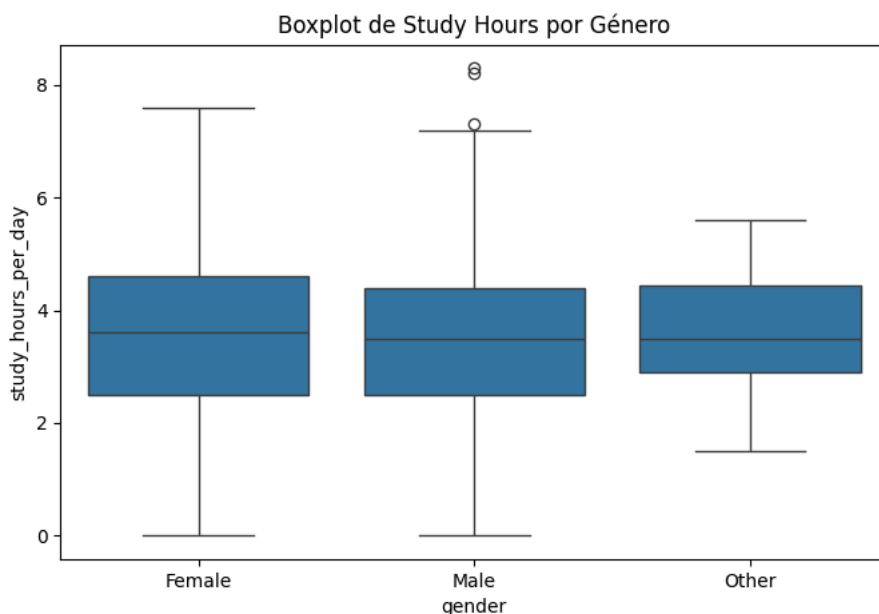
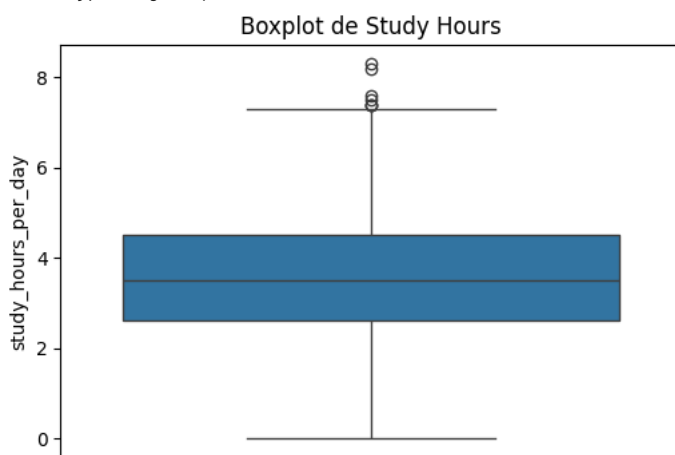
plt.figure(figsize=(8, 5))
sns.boxplot(data=df, x='gender', y='study_hours_per_day') # asegúrate de que los nombres coincidan
plt.title('Boxplot de Study Hours por Género')
plt.show()

```

```

Index(['student_id', 'age', 'gender', 'study_hours_per_day',
      'social_media_hours', 'netflix_hours', 'part_time_job',
      'attendance_percentage', 'sleep_hours', 'diet_quality',
      'exercise_frequency', 'parental_education_level', 'internet_quality',
      'mental_health_rating', 'extracurricular_participation', 'exam_score'],
      dtype='object')

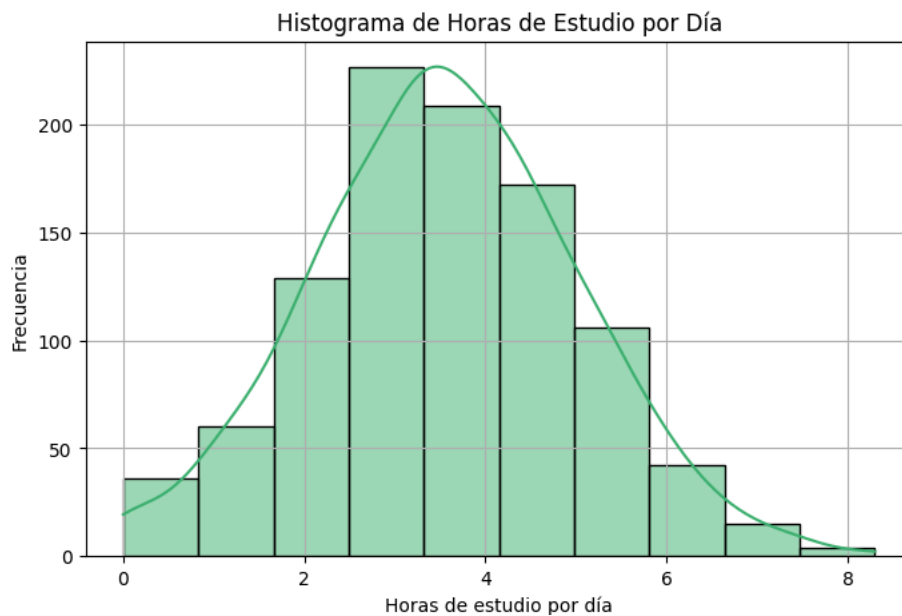
```



```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Cargar los datos
df = pd.read_csv('student_habits_performance.csv')

# Crear el histograma
plt.figure(figsize=(8, 5))
sns.histplot(data=df, x='study_hours_per_day', bins=10, kde=True, color='mediumseagreen')
plt.title('Histograma de Horas de Estudio por Día')
plt.xlabel('Horas de estudio por día')
plt.ylabel('Frecuencia')
plt.grid(True)
plt.show()
```



```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Cargar el DataFrame
df = pd.read_csv('student_habits_performance.csv')

# Calcular la matriz de correlación (solo numéricas)
correlation_matrix = df.corr(numeric_only=True)

# Crear el heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix,
            annot=True,          # Muestra los valores numéricos
            cmap='coolwarm',    # Colores para distinguir positivos y negativos
            fmt=".2f",          # Formato de los números
            linewidths=0.5,     # Líneas entre cuadros
            square=True         # Hacerlo cuadrado
            )
plt.title('Mapa de Calor de Correlación entre Variables Numéricas')
plt.show()
```

