

Equipo 6; Visualización



1869524 Luis Said Martinez Hernandez
1687417 Francisco Javier Gonzales Olivares
1838148 Raúl Flores Robles
1817336 Edmundo Isai Ibarra Cavazos

¿Que es la Visualización?





Visualización

“La visualización de datos es la representación gráfica de información y datos. Al utilizar elementos visuales como cuadros, gráficos y mapas, las herramientas de visualización de datos proporcionan una manera accesible de ver y comprender tendencias, valores atípicos y patrones en los datos”

La visualización de datos puede:

- Identificar áreas que necesitan atención o mejoras.
- Esclarecer qué factores influyen el comportamiento de los clientes.
- Ayudarle a entender qué productos colocar en qué lugar.
- Predecir volúmenes de ventas.



Ejercicio; Visualización de datos de alumnos que consumen alcohol en la universidad

Ejercicio de Visualización

Equipo 6

1869524 Luis Said Martinez Hernandez

1687417 Francisco Javier Gonzales Olivares

1817336 Edmundo Isai Ibarra cavazos

1838148 Raúl Flores Robles

Carga De Base De Datos

```
[2] import pandas as pd
pd.plotting.register_matplotlib_converters()
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

from google.colab import files #Linea para cargar el documento con las bases de datos
files.upload()
```

Seleccionar archivos student-mat.csv

• **student-mat.csv**(text/csv) - 41983 bytes, last modified: n/a - 100% done

Saving student-mat.csv to student-mat.csv

{'student-mat.csv': b'school,sex,age,address,famsize,Pstatus,Medu,Fedu,Mjob,Fjob,reason,guardian,traveltime,studytime,failures,schoolsup,f

Grafica de lineas

permiten visualizar los cambios a lo largo de un rango continuo, como el tiempo o la distancia

```
[3] # Primero de debe crear una ruta d archivo para leer la grafica de lineas
fam_rel = "student-mat.csv"
```

```
# Despues leemos el archivo en una variable fam_data
fam_data = pd.read_csv(fam_rel, index_col="famrel", parse_dates=True)
# Par confirmar que se hizo correctamente se leen las primeras 5 filas
fam_data.head()
```

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup
famrel																
4	GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes
5	GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no
4	GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes
3	GP	F	15	U	GT3	T	4	2	health	services	home	mother	1	3	0	no
4	GP	F	16	U	GT3	T	3	3	other	other	home	father	1	2	0	no

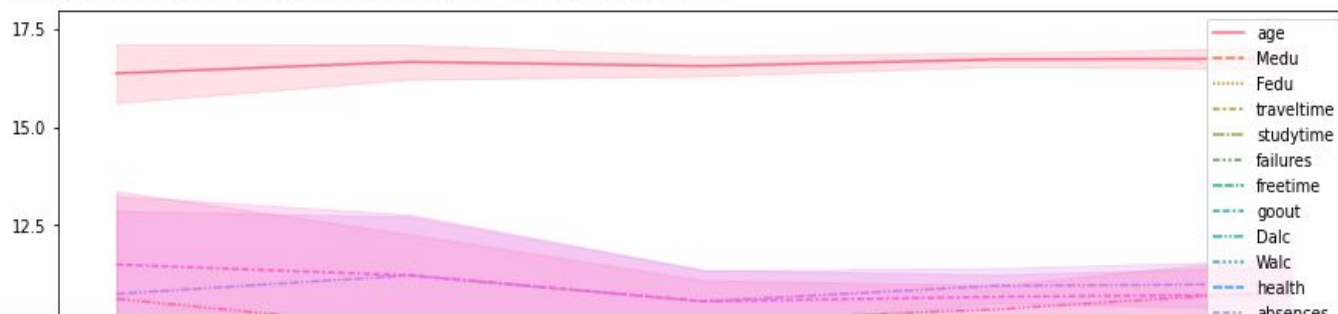
```
[4] # De igual forma se imprimen las ultimas 5 filas para asegurarnos( Opcional )
fam_data.tail()
```

✓ 0 s [4] # De igual forma se imprimen las ultimas 5 filas para asegurarnos(Opcional)
fam_data.tail()

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famrel
5	MS	M	20	U	LE3	A	2	2	services	services	course	other	1	2	2	no	
2	MS	M	17	U	LE3	T	3	1	services	services	course	mother	2	1	0	no	
5	MS	M	21	R	GT3	T	1	1	other	other	course	other	1	1	3	no	
4	MS	M	18	R	LE3	T	3	2	services	other	course	mother	3	1	0	no	
3	MS	M	19	U	LE3	T	1	1	other	at_home	course	father	1	1	0	no	

✓ 3 s [5] # Luego leemos la grafica de lineas con respecto a la columna famrel
plt.figure(figsize=(15,8)) #En esta linea configuramos el tamaño de nuestra grafica que deseamos tener
sns.lineplot(data=fam_data) #Codigo para crear la grafica

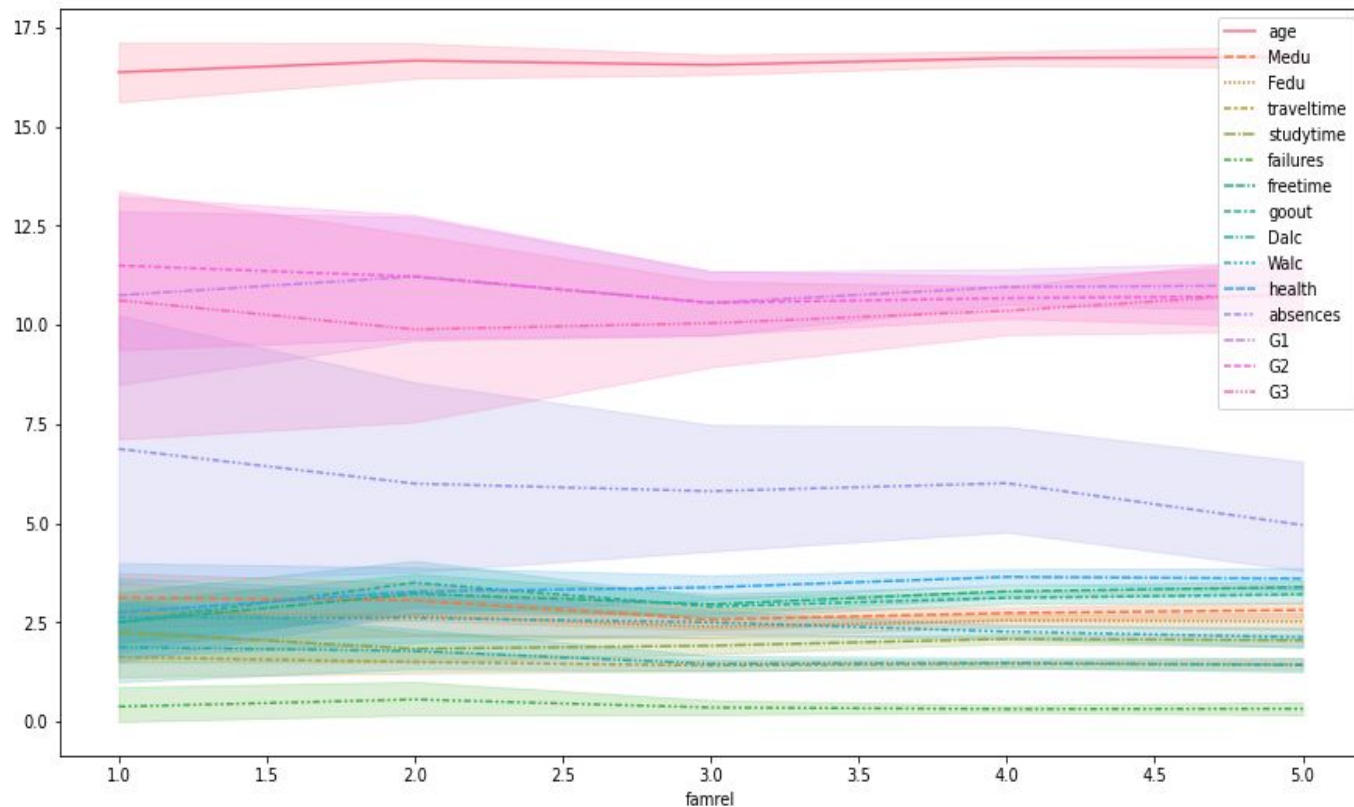
<matplotlib.axes._subplots.AxesSubplot at 0x7f01a6db7290>



✓ 1 s se ejecutó 23:39

```
[5] # Luego leemos la grafica de lineas con respecto a la columna famrel
plt.figure(figsize=(15,8)) #En esta linea configuramos el tamaño de nuestra grafica que deseamos tener
sns.lineplot(data=fam_data) #Codigo para crear la grafica
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f01a6db7290>




```
[6] # Ahora que tenemos la grafica con respecto a famrel, podemos acortar la grafica a las columnas que nos interesen visualizar mejor
# Escogemos el tamaño (Opcional)
plt.figure(figsize=(10,5))

# Añadimos titulo
plt.title("Numero de integrantes en la familia de los Medu y Fedu")

# Grafico de lineas con respecto a famrel que muestra los flujos de "Medu"
sns.lineplot(data=fam_data['Medu'], label="Medu")

# Grafico de lineas con respecto a famrel que muestra los flujos de "Fedu"
sns.lineplot(data=fam_data['Fedu'], label="Fedu")
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f01a5917810>

Numero de integrantes en la familia de los Medu y Fedu

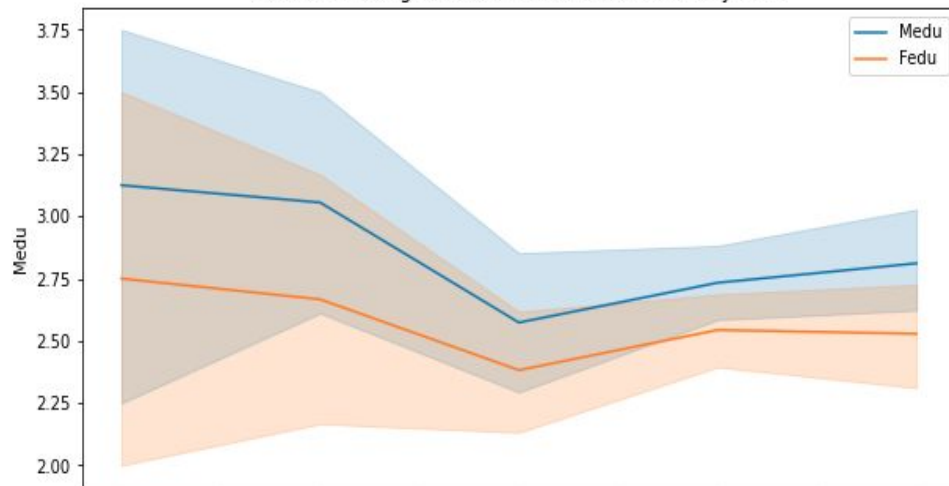


Grafico De Barras

Las gráficas de barras se utilizan para comparar dos o más valores. Las barras pueden ser horizontales o verticales

```
[7] students_filepath="student-mat.csv"
student_data=pd.read_csv(students_filepath)
student_data.head()
```

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	fams
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	
1	GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	y
2	GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes	
3	GP	F	15	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	y
4	GP	F	16	U	GT3	T	3	3	other	other	home	father	1	2	0	no	y

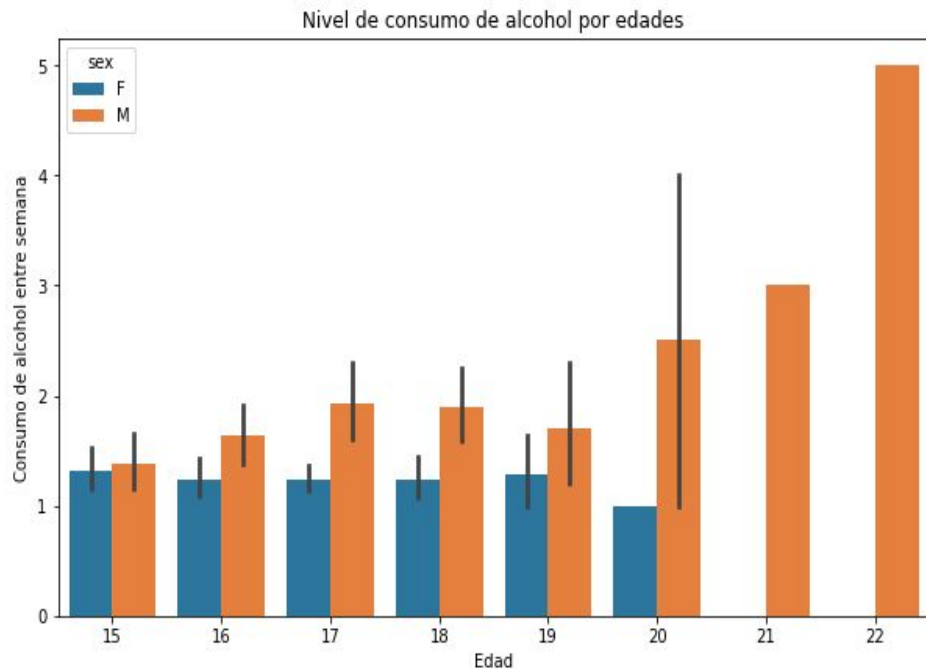
Describe la relación entre la cantidad de alcohol consumido entre semana por los jóvenes y su edad,ademas de separarlos por género

```
[8] plt.figure(figsize=(10,6))
plt.title("Nivel de consumo de alcohol por edades")
sns.barplot(x=student_data['age'],y=student_data['Dalc'],hue=student_data['sex'])
sns.set_style("dark")
plt.xlabel("Edad")
plt.ylabel("Consumo de alcohol entre semana")
```

Describe la relación entre la cantidad de alcohol consumido entre semana por los jóvenes y su edad, además de separarlos por género

```
[8] plt.figure(figsize=(10,6))  
plt.title("Nivel de consumo de alcohol por edades")  
sns.barplot(x=student_data['age'],y=student_data['Dalc'],hue=student_data['sex'])  
sns.set_style("dark")  
plt.xlabel("Edad")  
plt.ylabel("Consumo de alcohol entre semana")
```

```
Text(0, 0.5, 'Consumo de alcohol entre semana')
```



Grafica de dispección

Tipo de diagrama matematico que utiliza las cordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos

```
[9] # Ruta Para leer archivo
insurance_filepath = "student-mat.csv"

# leer el archivo en una variable insurance_data
insurance_data = pd.read_csv(insurance_filepath)
```

```
[ ] #Confirmar los datos imprimiendo los primeros 5
insurance_data.head()
```

ddress	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	travelttime	studytime	failures	schoolsup	famsup	paid	activities	n
U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no	no	
U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no	no	
U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes	no	yes	no	
U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes	yes	yes	
U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes	yes	no	

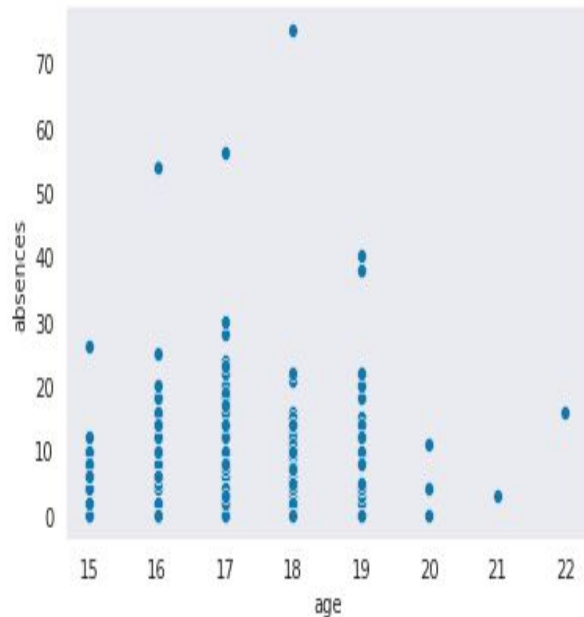
Grafica de dispersión

se usan para averiguar la intensidad de la relación entre dos variables numéricas (o 3 variables segun sea el caso)



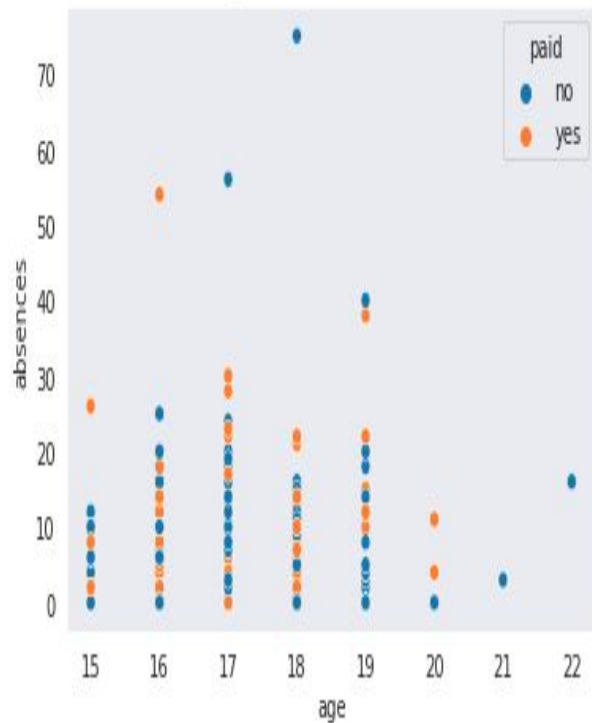
```
sns.scatterplot(x=insurance_data['age'], y=insurance_data['absences'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f0198c07fd0>



```
[11] #Cambio de color( 3 variables)  
sns.scatterplot(x=insurance_data['age'], y=insurance_data['absences'], hue=insurance_data['paid'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f01a3bbfcd0>



Graficos De Densidad

Índice

visualiza la distribución de datos en un intervalo o período de tiempo continuo

Histograma

indican la frecuencia de un hecho mediante una distribución de los datos

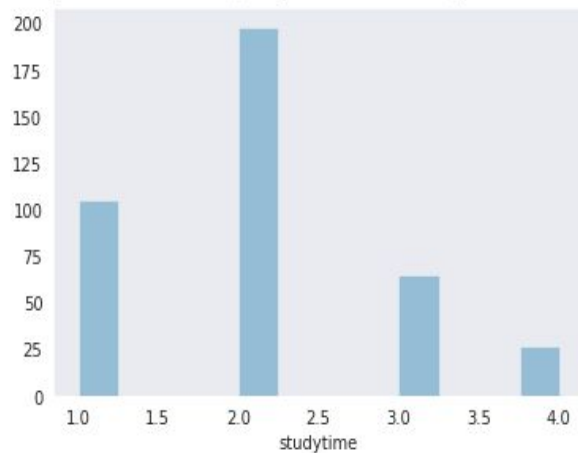
```
[12] # Path of the file to read
iris_filepath = "student-mat.csv"
# Read the file into a variable iris_data
iris_data = pd.read_csv(iris_filepath, index_col="age")
# Print the first 5 rows of the data
iris_data.head()
```

	school	sex	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famsup
age																
18	GP	F	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no
17	GP	F	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes
15	GP	F	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes	no
15	GP	F	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes
16	GP	F	U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes

```
[13] # Histogram
sns.distplot(a=iris_data['studytime'], kde=False)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Use `displot` instead.
warnings.warn(msg, FutureWarning)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f01a3a27510>
```



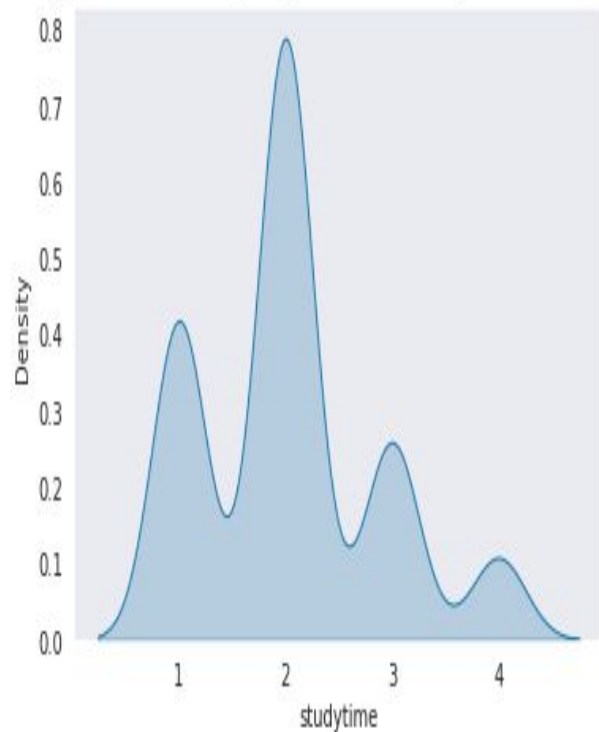
```
[14] # KDE plot
sns.kdeplot(data=iris_data['studytime'], shade=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f01a395efd0>
```




```
[14] # KDE plot  
sns.kdeplot(data=iris_data['studytime'], shade=True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f01a395efd0>

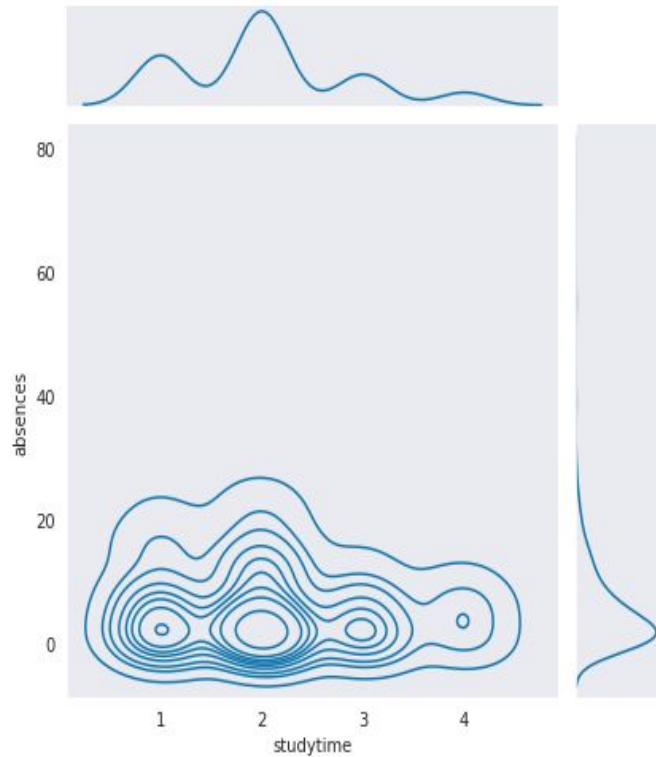


Graficos 2D KDE Plots

2D KDE plot

```
sns.jointplot(x=iris_data['studytime'], y=iris_data['absences'], kind="kde")
```

<seaborn.axisgrid.JointGrid at 0x7f01a18dc610>





Referencias;

<https://www.kaggle.com/learn/data-visualization>

<https://www.tableau.com/es-mx/learn/articles/data-visualization>