

## ▼ Sesión 3 - Diplomado Data Science Duoc UC

### Módulo: Machine Learning

A continuación, aprenderás a cargar un dataset con el uso de las librerías de Python desde fuentes de datos que se encuentran en la web.

**Recuerda almacenar una copia de este Jupyter Notebook en tu Google Drive para poder ejecutar los bloques de código.**

---

Sigamos con otro clásico dentro del mundo de data science. Trabajaremos con **tips**, otro dataset muy conocido en la academia, que al igual que los anteriormente vistos sirve bastante para comenzar dentro del mundo del Data Science. Para esta sesión se revisarán algunos métodos de la biblioteca Seaborn. Antes de comenzar, demos contexto a todo esto. Tips corresponde a un tipo de dataset para modelos no supervisados donde podemos estimar la propina que dejan los comensales de un restaurant tras ser atendidos.

Las características o features son las siguientes:

- Total\_bill: Total de la cuenta
- Tip: Propina
- Sex: Género del mesero
- Smoker: Dato para diferenciar una mesa de fumador vs una de no fumador.
- Day: Qué día de la semana fue la atención.
- Time: En que horario, si fue almuerzo o cena.
- Size: Tamaño de la mesa.

```
#Cargamos las librerías de Python que nos servirán para todo esto
import pandas as pd
import seaborn as sbn
from seaborn import load_dataset
```

```
#Cargamos el dataset de tips para comenzar a trabajarlo
```

```
tips = load_dataset("tips")
tips.head()
```

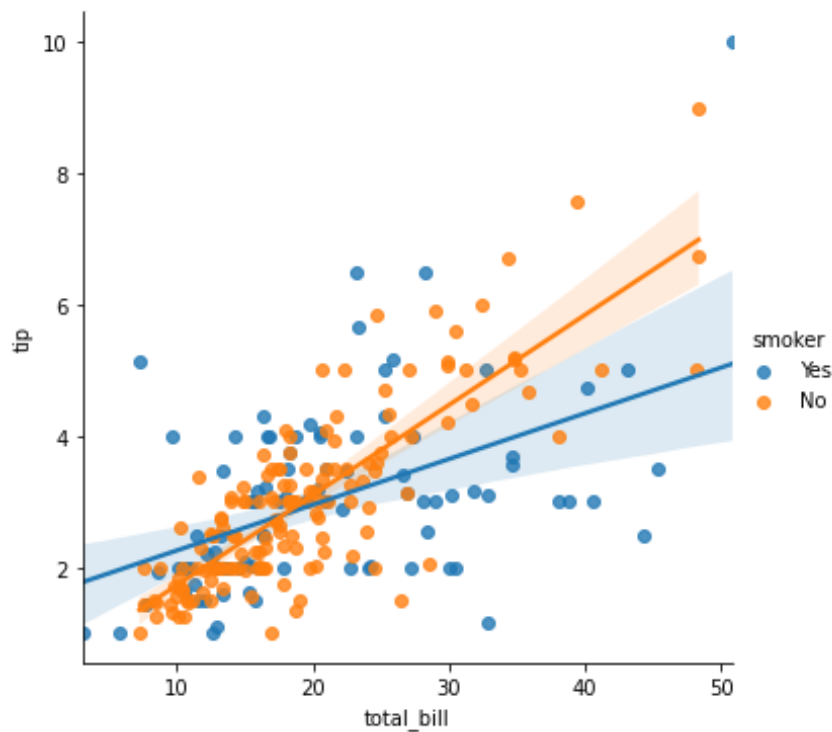
	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3

Añadir qué observaciones hay respecto del comando utilizado

```
3      23.68  3.31   Male    No  Sun  Dinner    2
```

```
from seaborn import lmplot
lmplot(x="total_bill", y="tip", hue="smoker", data=tips)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f9727f92da0>
```



Añadir qué observaciones hay respecto del comando utilizado

```
from seaborn import boxplot
boxplot(x="sex", y="total_bill", data=tips)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9727f928d0>
```

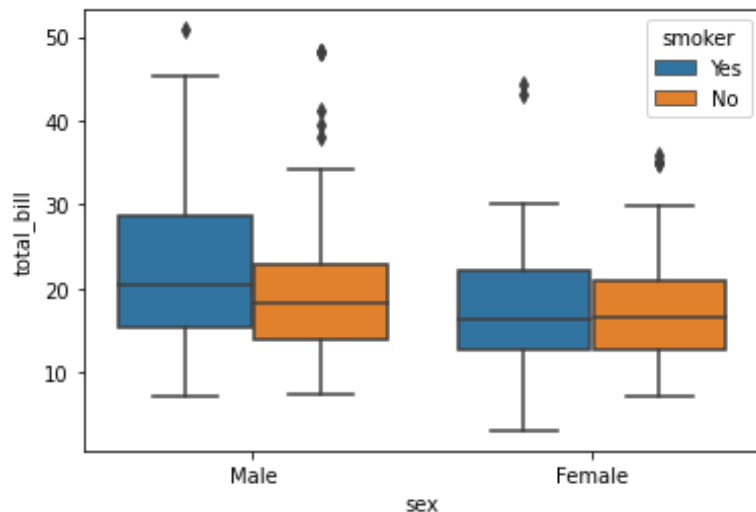


Añadir qué observaciones hay respecto del comando utilizado

```
bill 30 |
```

```
boxplot(x="sex", y="total_bill", hue="smoker", data=tips)
```

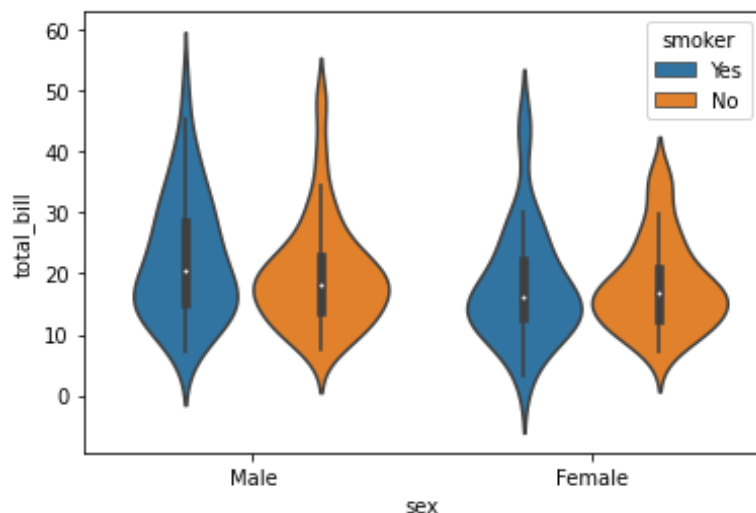
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f972514c828>
```



Añadir qué observaciones hay respecto del comando utilizado

```
from seaborn import violinplot
violinplot(x="sex", y="total_bill", hue="smoker", data=tips)
```

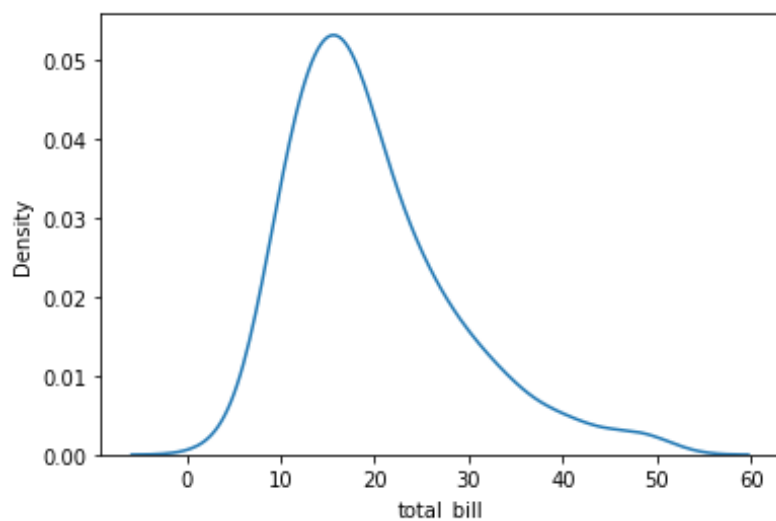
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9725607518>
```



Añadir qué observaciones hay respecto del comando utilizado

```
from seaborn import kdeplot
kdeplot(tips.total_bill)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f972456c2e8>



Añadir qué observaciones hay respecto del comando utilizado

## ▼ Actividad práctica:

```
#Importamos las librerías necesarias
import pandas as pd
import seaborn as sbn
from seaborn import load_dataset
```

```
#Cargamos el dataset de titanic para comenzar a trabajarlo
```

```
titanic = load_dataset("titanic")
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult.
0	0	3	male	22.0	1	0	7.2500	S	Third	man	
1	1	1	female	38.0	1	0	71.2833	C	First	woman	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	
3	1	1	female	35.0	1	0	53.1000	S	First	woman	
4	0	3	male	35.0	0	0	8.0500	S	Third	man	

Proponga y ejecute acá una rutina de análisis exploratorio de datos.

Recuerde hacer uso de todo lo aprendido hasta el momento