

ESTADÍSTICA DE
NEGOCIOS



TAREA SEMANA 1

ANÁLISIS DESCRIPTIVO

Tarea semana 1: Análisis descriptivo

Problema Stats Health:

- La empresa Stats Health está desarrollando un análisis de [usabilidad](#) de un nuevo sistema informático de tele-rehabilitación para pacientes de artroplastia de cadera.
 - Vea: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8606428>
- Stats Health ha recopilado un dataset acerca de la usabilidad de la interfaz del sistema que contiene:
 - Variables de uso de la aplicación: Time, y la escala de usabilidad dada por el usuario: SUS.
 - Variables demográficas del usuario, Género (Gender), Edad (Age), entre otras.

Archivo de datos:

- [UX_data.csv](#)
- Lea el archivo csv y ordene las variables categóricas de una manera adecuada.

```
import pandas as pd
import seaborn as sns

data = pd.read_csv("../data/UX_data.csv") # si la estructura es la del repositorio de GitHub

# si está trabajando en local y el archivo de datos y el notebook están en el mismo directorio
# data = pd.read_csv("UX_data.csv")

# También puede leer el archivo desde el repositorio de GitHub.
#data = pd.read_csv("https://raw.githubusercontent.com/marsgr6/EN-online/8a1cee296279c274d8f05f3905f97c1f7636a2/data/UX_data.csv")

data['Age'] = pd.Categorical(data['Age'],
                             categories=['twenty', 'thirty', 'forty'],
                             ordered=True)

data.head()
```

- Encabezado de nuestro dataset.

	SUS	Time	ID	Gender	Age	Kinect	NUI	Games
0	42.5	179.3	student	M	twenty	high	average	high
1	92.5	153.3	student	M	twenty	high	high	high
2	95	135.7	student	M	twenty	average	average	high
3	92.5	148.3	student	F	twenty	average	high	high
4	80	148	student	M	twenty	average	average	high

Variables:

- SUS (System Usability Scale): Escala de Usabilidad del Sistema.
 - Un valor entre 0 y 100, dado por el usuario al evaluar el sistema informático.
 - Ver: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>
- Time: Tiempo en que el usuario realizó la prueba del sistema.
- Gender: M (Male, masculino), F (Female, Femenino).
- Age: twenty (veinte años), thirty (treinta años), forty (cuarenta años).

Análisis a realizar:

A partir de los datos **UX_data.csv** realice un análisis descriptivo de:

1. (1 pt) Variable numéricas:

- Escriba el código Python para describir las **variables numéricas** del conjunto de datos (vea el resultado esperado en la tabla a continuación).
- Interprete los cuartiles para cada variable.

	SUS	Time
count	30	30
mean	76.0833	157.5
std	14.617	22.3994
min	40	112.7
25%	68.125	140.025
50%	78.75	159.7
75%	87.5	177.05
max	95	188.3

2. (1 pt) Variables categóricas:

- Escriba el código Python para describir las variables categóricas del conjunto de datos (vea el resultado esperado en la tabla a continuación para **ID, Gender y Age**).

ID		Gender		Age	
employee	15	M	20	twenty	17
student	15	F	10	thirty	7
				forty	6

3. (1 pt) Descripción de variables numéricas agrupadas:

- Escriba el código Python para describir las **variables numéricas** del dataset agrupadas por "**Gender**".
- Interprete las medidas de centralidad y dispersión para cada variable agrupadas por género, es decir compare las clases F (Female) vs M (Male).

		F	M
SUS	count	10	20
	mean	74	77.125
	std	17.3285	13.4329
	min	40	42.5
	25%	66.25	69.375
	50%	77.5	80
	75%	87.5	87.5
	max	92.5	95
Time	count	10	20
	mean	172.19	150.155
	std	15.6563	21.8936
	min	144	112.7
	25%	165.975	135.6
	50%	175.5	148.65
	75%	184.75	167.3
	max	188.3	185.3

4. (1 pt) Descripción de variables numéricas agrupadas por "**Age**":

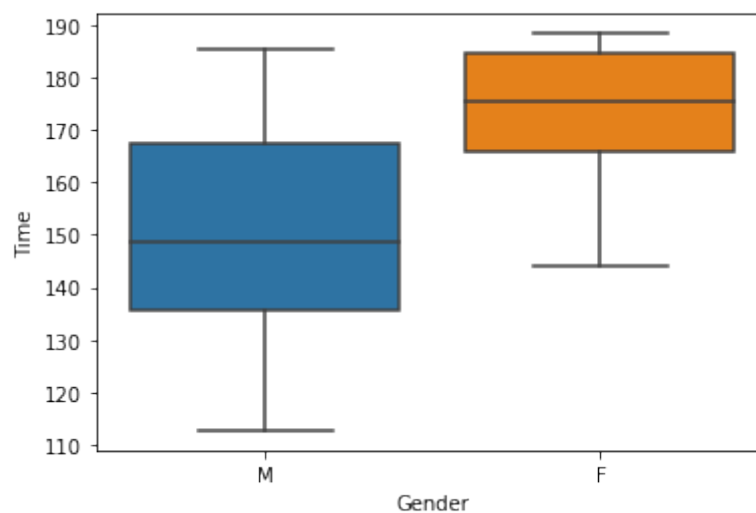
- Interprete las medidas de centralidad y dispersión para cada rango de edad.

		twenty	thirty	forty
SUS	count	17	7	6
	mean	78.0882	74.6429	72.0833
	std	15.8983	12.5357	14.3542
	min	40	57.5	52.5
	25%	75	66.25	63.125
	50%	80	75	71.25
	75%	87.5	83.75	85
	max	95	90	87.5
Time	count	17	7	6
	mean	146.835	166.214	177.55
	std	20.2663	22.1184	5.76637
	min	112.7	128.3	171
	25%	135.7	154.65	173.65
	50%	148	176.3	176
	75%	157.7	181.8	182.325
	max	188.3	186	185

A partir del análisis descriptivo agrupado que ha realizado, construya los siguientes gráficos:

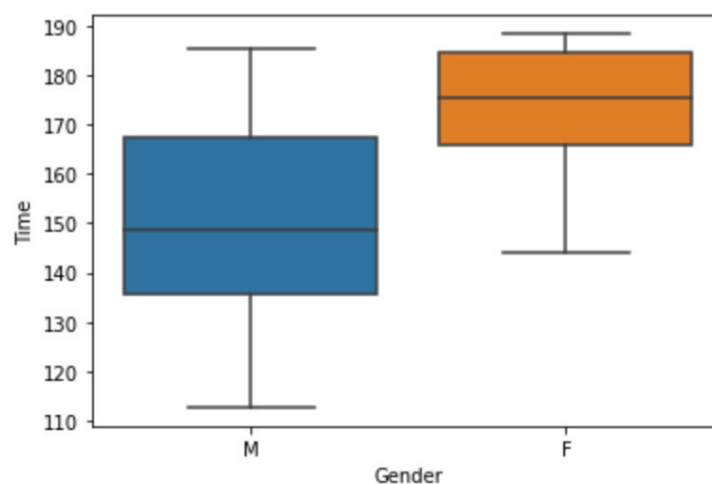
5. (1 pt) Diagrama de cajas **Gender vs SUS:**

- Escriba el código Python para construir un diagrama de cajas que muestre el género (**Gender**) en x y la evaluación de usabilidad (**SUS**) en y.
- **Discuta:** Observe diferencia en la evaluación de usabilidad del sistema (SUS) realizada por los usuarios de acuerdo al género (Gender). Realice el análisis con respecto a la información mostrada en el diagrama de cajas (resumen de 5 puntos, posición (de los cuartiles), dispersión (IQR, bigotes), etc.).



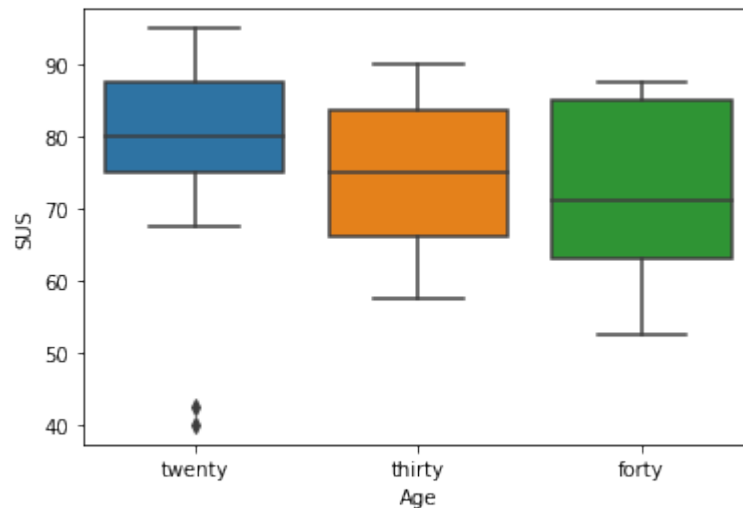
6. (1 pt) Diagrama de cajas **Gender vs Time:**

- Escriba el código Python para construir un diagrama de cajas que muestre el género (**Gender**) en X y el tiempo tomado en realizar el ejercicio (**Time**) en Y.
- **Discuta:** Observe diferencia en el tiempo en que realizaron la tarea en el sistema los usuarios de acuerdo al género. Realice el análisis con respecto a la información mostrada en el diagrama de cajas (resumen de 5 puntos, posición (de los cuartiles), dispersión (IQR, bigotes), etc.).



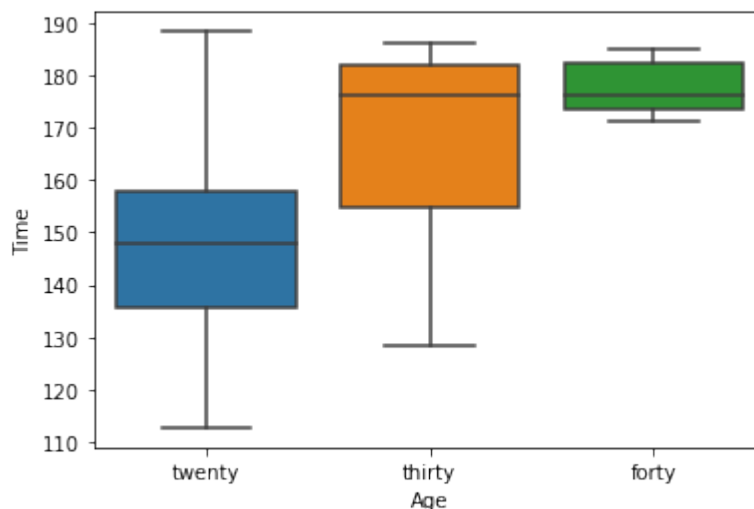
7. (1 pt) Diagrama de cajas **Age vs SUS**:

- Escriba el código Python para construir un diagrama de cajas que muestre el rango de edad (**Age**) en x y la evaluación de usabilidad (**SUS**) en y.
- **Discuta:** Observe la diferencia en la evaluación de usabilidad del sistema (SUS) de acuerdo al rango de edad (Age). Realice el análisis con respecto a la información mostrada en el diagrama de cajas (resumen de 5 puntos, posición (de los cuartiles), dispersión (IQR, bigotes), etc.).



8. (1 pt) Diagrama de cajas **Age vs Time**:

- Escriba el código Python para construir un diagrama de cajas que muestre el rango de edad (**Age**) en x y el tiempo tomado en realizar el ejercicio (**Time**) en y.
- **Discuta:** Observe la diferencia en el tiempo en que realizaron la tarea en el sistema (Time) los usuarios de acuerdo a la edad (Age). Realice el análisis con respecto a la información mostrada en el diagrama de cajas (resumen de 5 puntos, posición (de los cuartiles), dispersión (IQR, bigotes), etc.).



9. (2 pts) Diagrama de dispersión **SUS vs Time** agrupados por **Gender**:

■ **Discuta:**

- Escriba el código Python para construir un diagrama de dispersión que compare las variables evaluación de usabilidad (SUS) en x y el tiempo tomado en realizar el ejercicio (**Time**) en y, agrupados por género (**Gender**).
- Qué relación observa entre las variables SUS y Time.
- Discuta la relación de estas variables de acuerdo al género.
- Se corresponde este análisis con el observado en los diagramas de caja.
- Justifique su respuesta.

