



Tendencias e Innovación en Tecnología Agrícola- TEA (CG2335-223E)

Laboratorio y Problemas 07 Introducción al Análisis de Datos en Python Parte #2 Jupyter Notebooks, Pandas, Matplotlib y Proyecto

Estudiante: Valeria Araujo

Grupo: GOLD HEALTH

Proyecto: NUTRI- ANALYTICS

Identificador de Github: valeriaaraujo03

Fecha: 29 de Octubre del 2022.

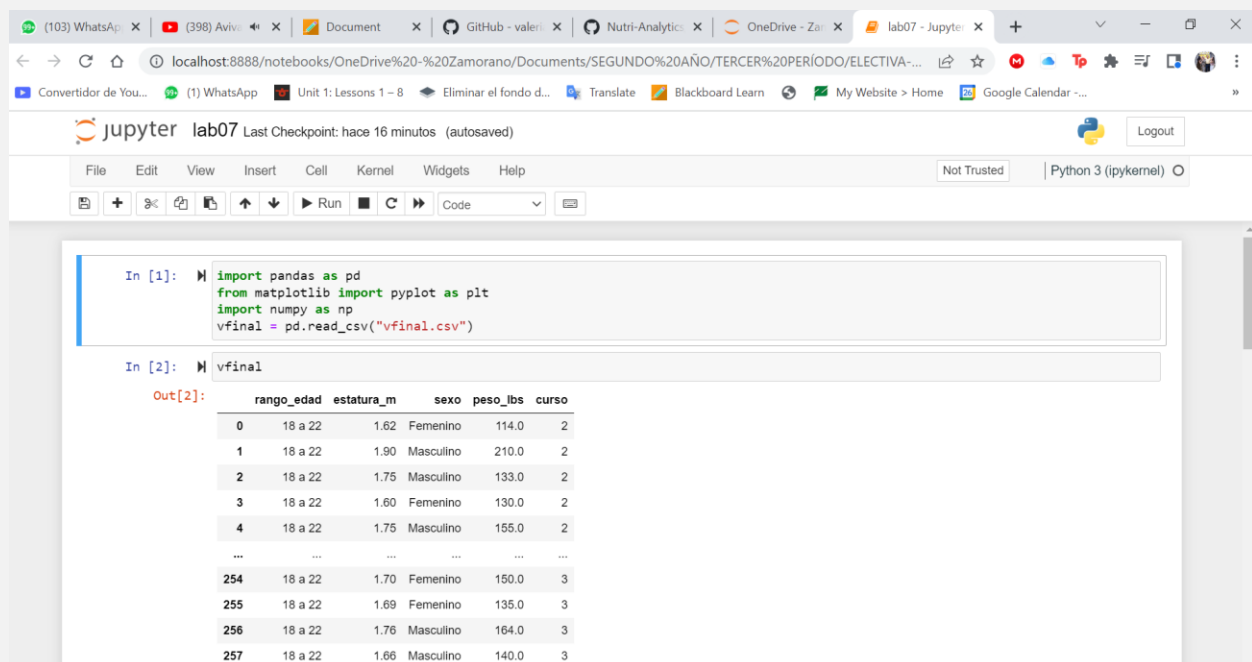
Parte 1: Problema - Jupyter Notebook (<https://jupyter.org/>)

Siga las instrucciones del documento de laboratorio y problemas 07.

¿Cuántos registros/líneas tiene su fuente de datos?

Nuestra fuente de datos tiene 259 registros, es decir, se obtuvieron los datos de 259 estudiantes tanto de primero, segundo y tercer año.

Fuente de Datos del Proyecto (Subiendo la fuente de datos a memoria utilizando Pandas)
Incluya la captura de pantalla acá. La captura de pantalla debe mostrar el código, el contenido del DataFrame, y el Jupyter Notebook corriendo en Jupyter (<https://jupyter.org/try-jupyter/lab/>).



The screenshot shows a Jupyter Notebook interface with the following content:

```
In [1]: import pandas as pd
        from matplotlib import pyplot as plt
        import numpy as np
        vfinal = pd.read_csv("vfinal.csv")
```

In [2]: vfinal

Out[2]:

	rango_edad	estatura_m	sexo	peso_lbs	curso
0	18 a 22	1.62	Femenino	114.0	2
1	18 a 22	1.90	Masculino	210.0	2
2	18 a 22	1.75	Masculino	133.0	2
3	18 a 22	1.60	Femenino	130.0	2
4	18 a 22	1.75	Masculino	155.0	2
...
254	18 a 22	1.70	Femenino	150.0	3
255	18 a 22	1.69	Femenino	135.0	3
256	18 a 22	1.76	Masculino	164.0	3
257	18 a 22	1.66	Masculino	140.0	3

Parte 2: Problema – Análisis de la fuente de datos (Valor: 2 puntos)

¿Revisó la documentación de Matplotlib?

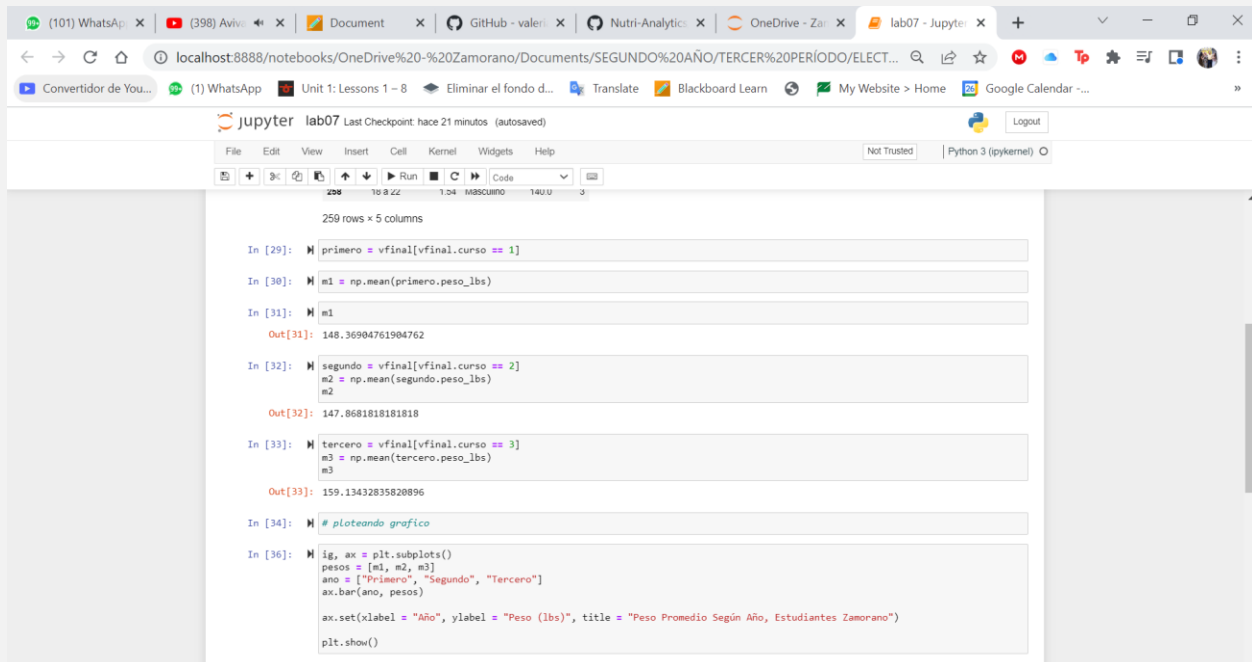
Sí/No

¿Qué tipo de gráfica están utilizando para el análisis de los datos y por qué?

El tipo de gráficas que estamos utilizando para el análisis de los datos obtenidos, son los gráficos de barras, debido a que estos ayudan a comparar dos o más variables de una forma clara de entender. Que en el caso de nuestra investigación, se comparan el peso respecto al año.

Analizando la Fuente de Datos del Proyecto (utilizando Matplotlib)

Incluya la captura de pantalla acá. La captura de pantalla debe mostrar el código, el contenido del DataFrame, la gráfica generada desde Matplotlib, el Jupyter Notebook corriendo en Jupyter (<https://jupyter.org/try-jupyter/lab/>).



```
259 rows x 5 columns

In [29]: m1 = vfinal[vfinal.curso == 1]

In [30]: m1 = np.mean(m1.peso_lbs)

In [31]: m1
Out[31]: 148,36904761904762

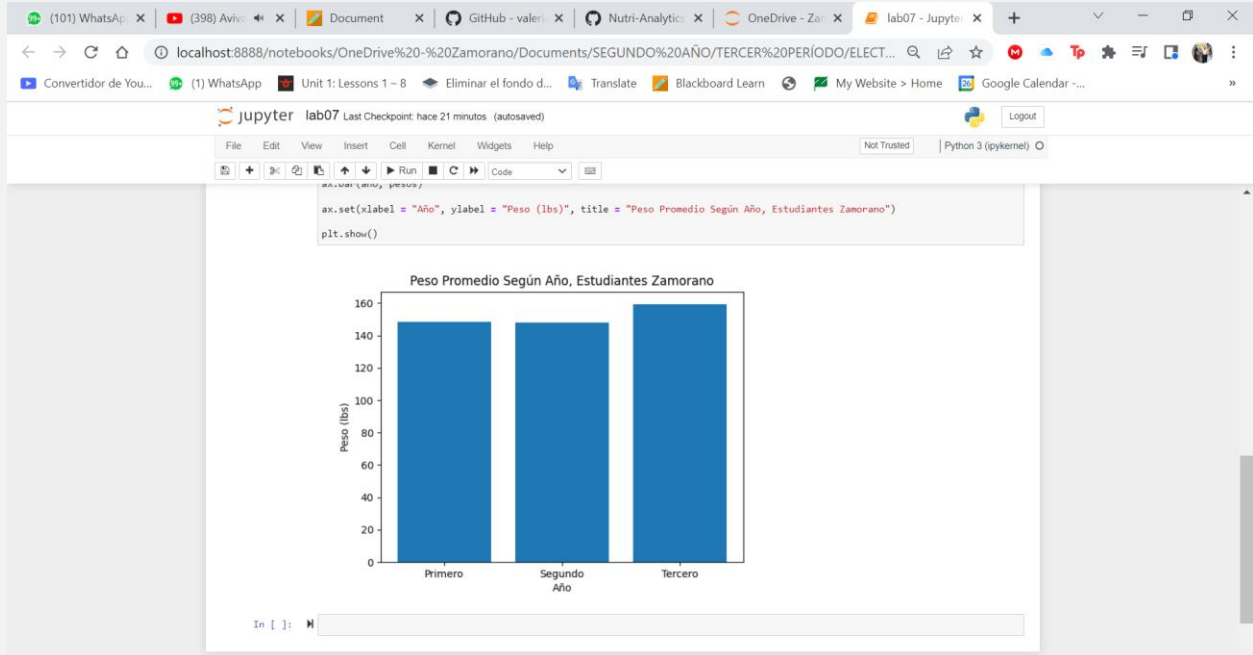
In [32]: m2 = vfinal[vfinal.curso == 2]
m2 = np.mean(m2.peso_lbs)
m2
Out[32]: 147,8681818181818

In [33]: m3 = vfinal[vfinal.curso == 3]
m3 = np.mean(m3.peso_lbs)
m3
Out[33]: 159,13432835828896

In [34]: # plotando grafico

In [36]: fig, ax = plt.subplots()
pesos = [m1, m2, m3]
ano = ["Primero", "Segundo", "Tercero"]
ax.bar(ano, pesos)

ax.set(xlabel = "Año", ylabel = "Peso (lbs)", title = "Peso Promedio Según Año, Estudiantes Zamorano")
plt.show()
```



Parte 3: Laboratorio

¿Revisó, estudió, y analizó los videos de la clase y laboratorio?

Sí/No

(Marque su respuesta)

(opcional) ¿Cuál es el url/link/enlace del directorio creado en Github? ¿A dónde se subieron los archivos de este laboratorio?

<https://github.com/valeriaaraujo03/Tea.git>

Incluya el enlace del Jupyter Notebook subido a Github (opcional):

Ejemplo: <https://github.com/proyecto/>

TODOS deben subir este documento a Blackboard.