

INTRODUCCIÓN AL BIG DATA Y BUSINESS INTELLIGENCE (IFCT153PO)

Tamara del Estal Hernández

Examen

La prueba consiste en cuatro actividades relacionadas con los contenidos de la acción formativa.

Cada actividad pondera 25 puntos, siendo necesario para calificar la prueba como apta un total de 50 puntos.

Actividad 1 - Python - Mongo

En un archivo de Word, Google Docs o Microsoft Word online realiza las siguientes actividades. (muy breve, únicamente lo relevante para demostrar que has realizado los items solicitados). Añade capturas de pantalla o recortes para aportar tu solución. Convierte el archivo a PDF para facilitar su lectura.

Bases de datos NoSQL.- Mongo. (10 puntos)

En Mongo estamos trabajando con base de datos en la nube, en concreto con MongoAtlas.

Explica paso a paso cómo puedes acceder a tu entorno Mongo Atlas (oculta la contraseña).

Prueba a acceder a las colecciones y crea una base de datos llamada examen1.

En esta base de datos, crea una colección llamada productos.

Añade dos productos con los siguientes datos.

id : 1

nombre : camisa

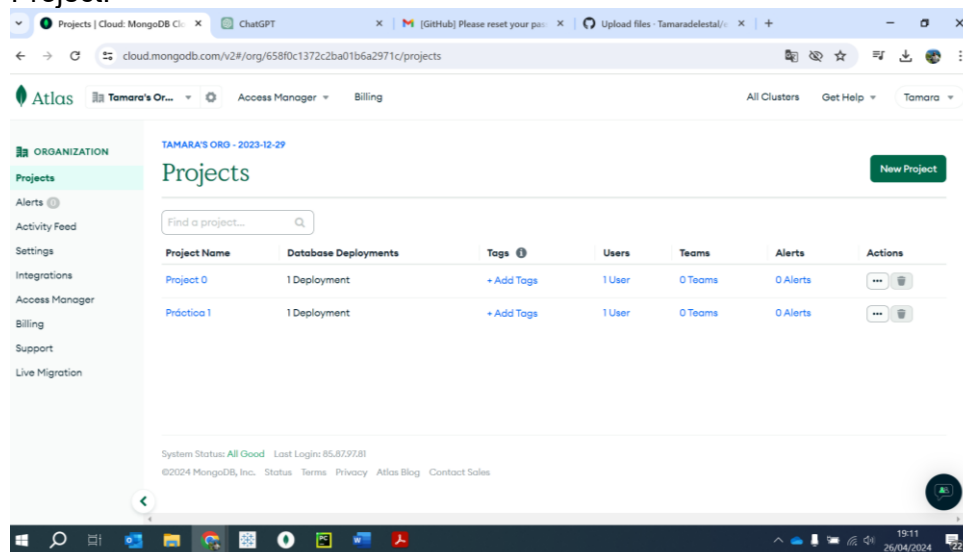
precio : 15

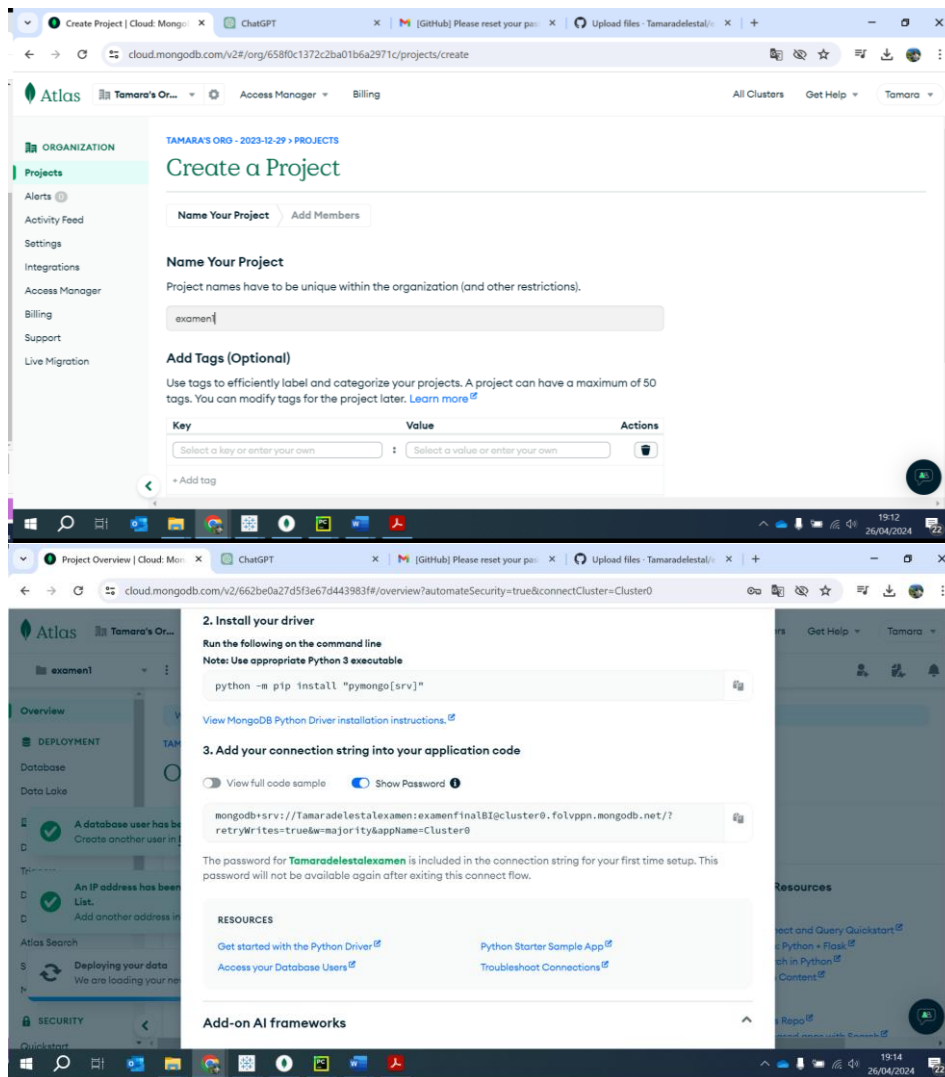
id:2

nombre : chaqueta

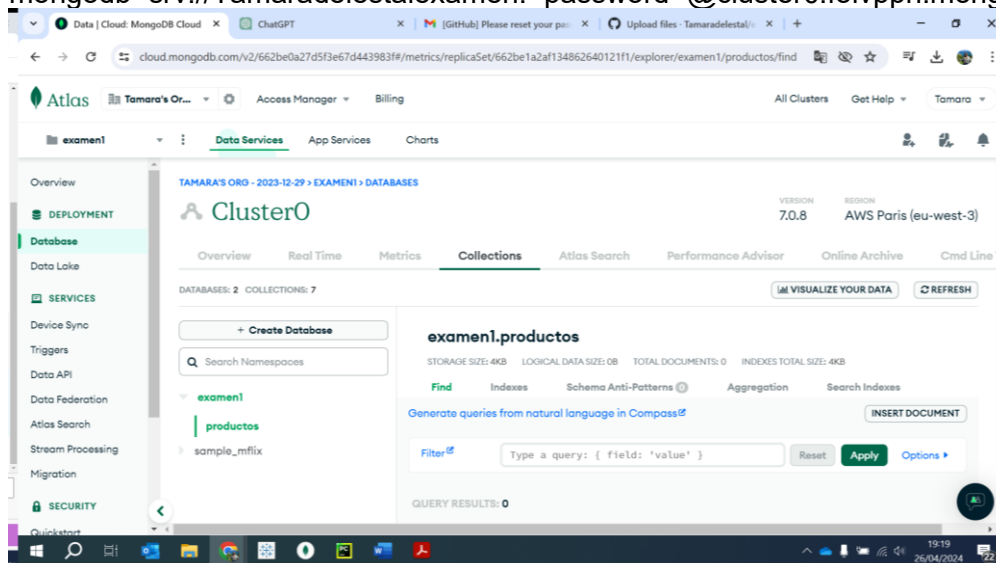
precio : 20

Acceso a través de la web de Mongo, y para crear una nueva DB se pincha en New Project.

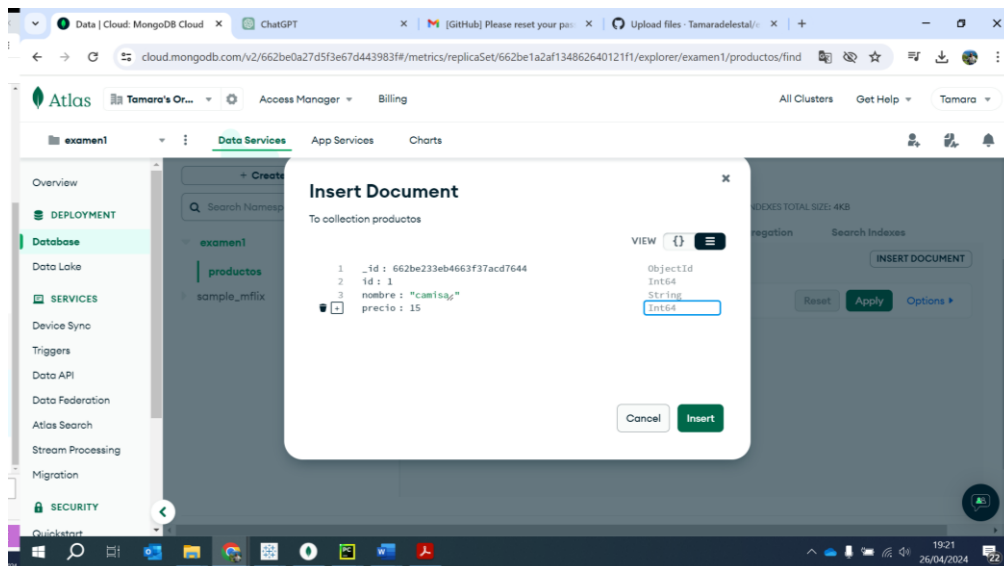




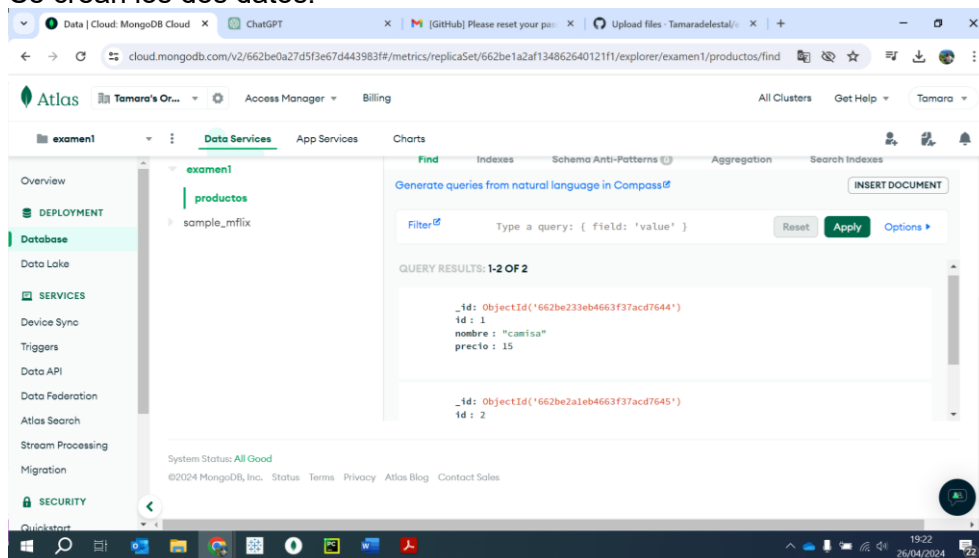
Se crea la nueva contraseña y usuario de la nueva DB y se copia la cadena de conexión para usarla posteriormente.
 mongodb+srv://Tamaradelestalexamen:<password>@cluster0.folvppn.mongodb.net/



Una vez creada la colección “productos” se insertan datos (insert document)

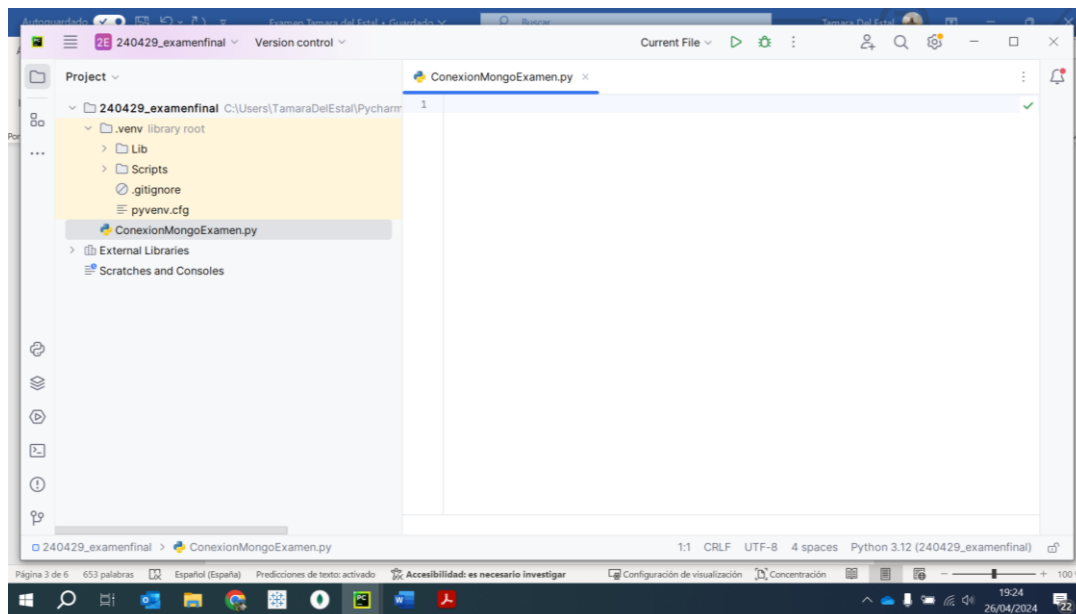


Se crean los dos datos.



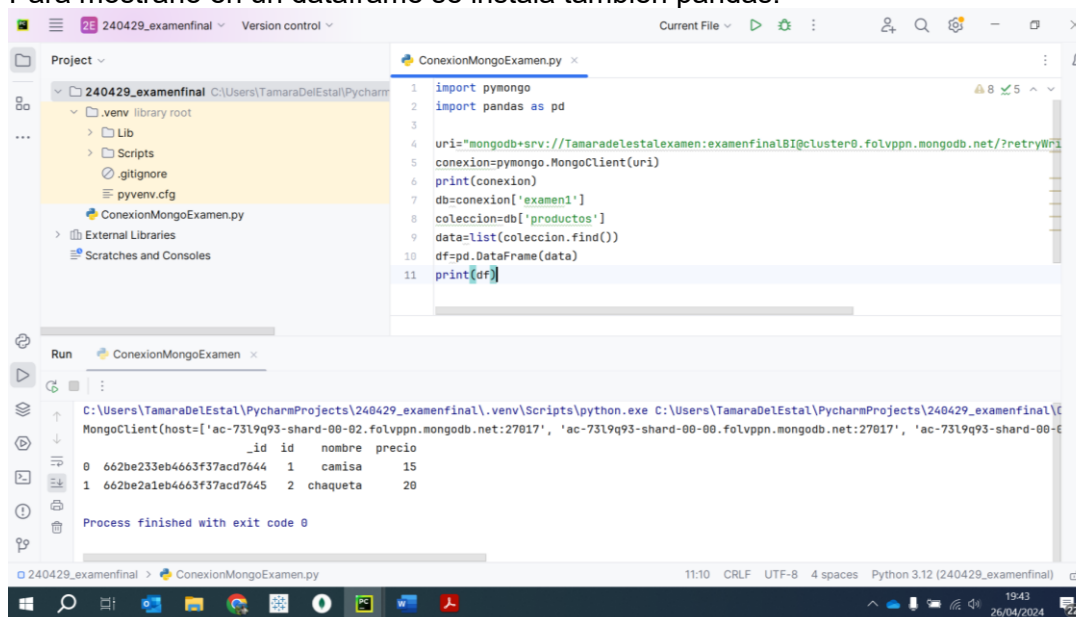
Python. Conexión con Mongo (10 puntos)

Para escribir scripts en Python puedes utilizar editores de código como Visual Studio Code o entornos de desarrollo como PyCharm, o incluso Anaconda Jupyter. Accede a los datos de la colección productos de la base de datos examen1 creada en el punto anterior. A través de un DataFrame muestra los datos en consola.



Se instala pymongo para poder acceder a MongoDB a través de la cadena de conexión.

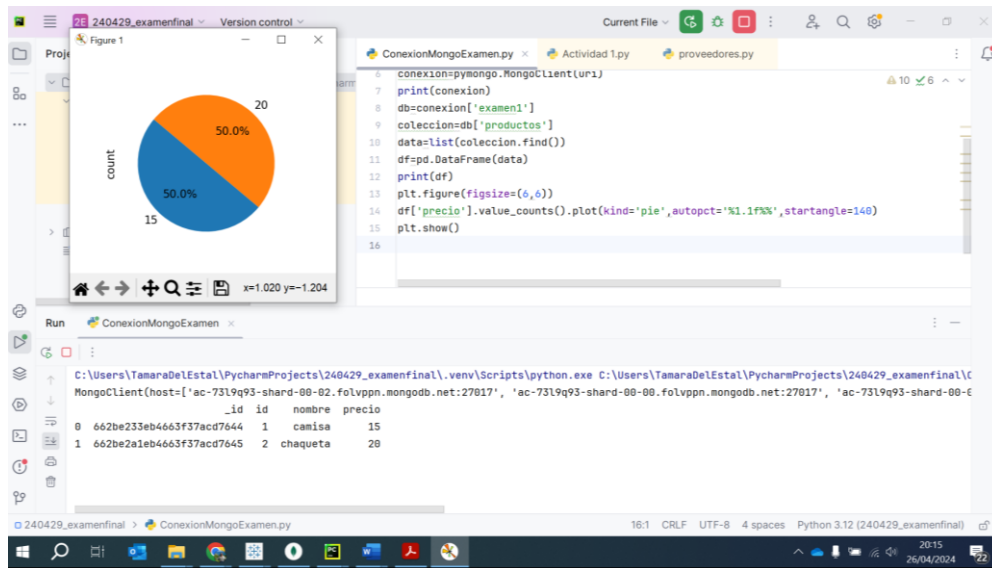
Para mostrarlo en un dataframe se instala también pandas.



Diseño de gráfico en Python. (5 puntos)

Diseña un gráfico a tu elección utilizando alguna librería en Python para mostrar los datos de la colección productos.

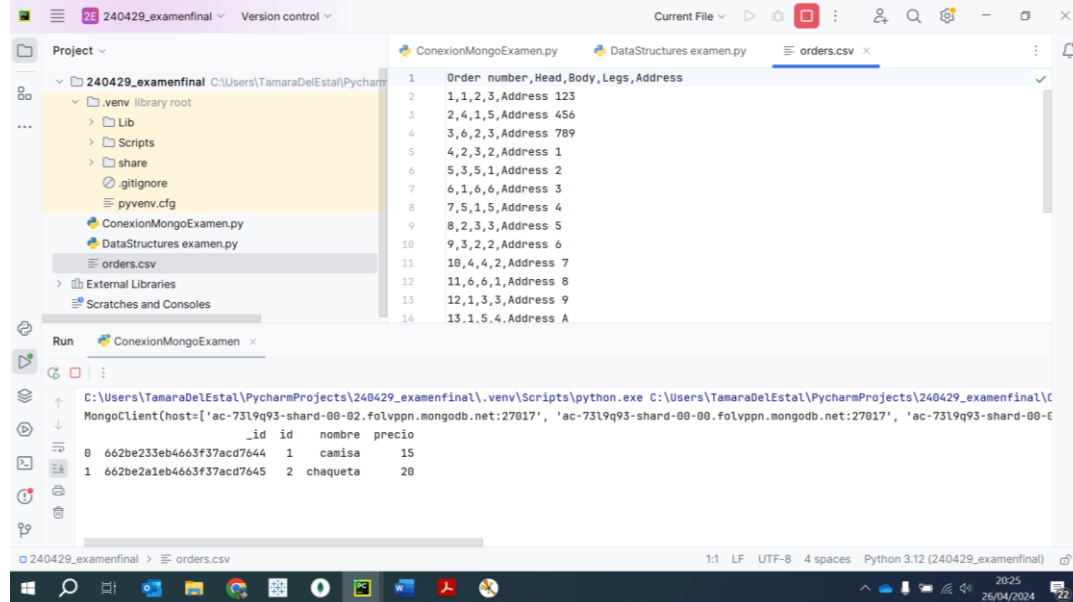
Se instala la librería matplotlib.



Actividad 2 - Python - DataStructures

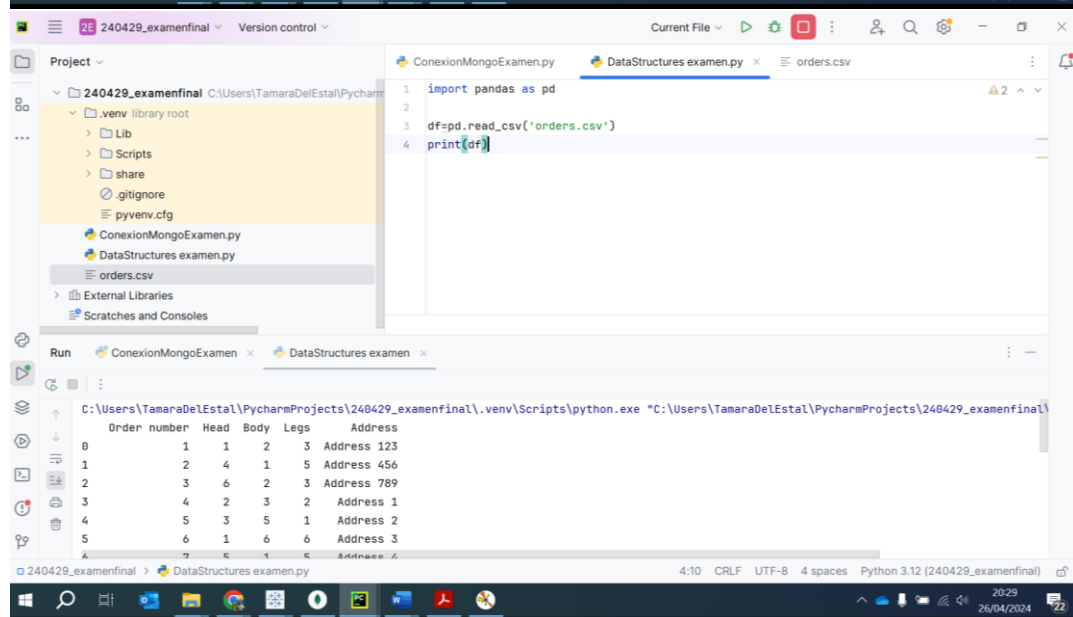
Acceso a datos(10 puntos)

En la sección Documentación / Archivos dispones del archivo orders.csv.
Desde Python, accede al contenido del archivo y muestra su contenido en consola



The screenshot shows the PyCharm IDE with the 'orders.csv' file open. The file contains the following data:

Order	number	Head	Body	Legs	Address
1	1,1,2,3	Address	123		
2	2,4,1,5	Address	456		
3	3,6,2,3	Address	789		
4	4,2,3,2	Address	1		
5	5,3,5,1	Address	2		
6	6,1,6,6	Address	3		
7	7,5,1,5	Address	4		
8	8,2,3,3	Address	5		
9	9,3,2,2	Address	6		
10	10,4,4,2	Address	7		
11	11,6,6,1	Address	8		
12	12,1,3,3	Address	9		
13	13,1,5,4	Address	A		



The screenshot shows the PyCharm IDE with the 'DataStructures examen.py' file open. The code in the file is:

```
1 import pandas as pd
2
3 df=pd.read_csv('orders.csv')
4 print(df)
```

The console output shows the following data:

	Order	number	Head	Body	Legs	Address
0	1	1,1,2,3	Address	123		
1	2	2,4,1,5	Address	456		
2	3	3,6,2,3	Address	789		
3	4	4,2,3,2	Address	1		
4	5	5,3,5,1	Address	2		
5	6	6,1,6,6	Address	3		
6	7	7,5,1,5	Address	4		
7	8	8,2,3,3	Address	5		
8	9	9,3,2,2	Address	6		
9	10	10,4,4,2	Address	7		
10	11	11,6,6,1	Address	8		
11	12	12,1,3,3	Address	9		
12	13	13,1,5,4	Address	A		

Manipulación de datos(10 puntos)

Diseña una función que permita ordenar los datos por una de las columnas a tu elección.

Diseña otra función que permita filtrar los datos por un dato a tu elección.

Muestra el funcionamiento correcto del script.

The screenshot shows a Jupyter Notebook with a file explorer on the left displaying a project structure. The main area contains a code cell with the following Python code:

```
1 import pandas as pd
2
3 df=pd.read_csv('orders.csv')
4 print(df)
5 df_trabajo=df[['Order number','Head','Body','Legs']]
6 df_sorted=df.sort_values(by='Body')
7 print(df_sorted)
```

The output of the code is displayed in the console below the code cell, showing a DataFrame with columns: Order number, Head, Body, Legs, and Address. The data is sorted by the 'Body' column.

	Order number	Head	Body	Legs	Address
19	20	1	1	5	Address H
19	20	1	1	5	Address H
1	2	4	1	5	Address 456
13	14	3	1	5	Address B
6	7	5	1	5	Address 4
14	15	3	2	6	Address C
8	9	3	2	2	Address 6

Con la función sorted se ordena.

Se instala filtro para poder filtrar por un item que se desee. En este caso por nombre de la calle.

The screenshot shows a Jupyter Notebook with a file explorer on the left displaying a project structure. The main area contains a code cell with the following Python code:

```
1 import filtro
2 import pandas as pd
3
4 df=pd.read_csv('orders.csv')
5 print(df)
6 df_trabajo=df[['Order number','Head','Body','Legs']]
7 df_sorted=df.sort_values(by='Body')
8 print(df_sorted)
9 filtroAddress=df[df['Address']=='Address H']
10 print(filtroAddress)
```

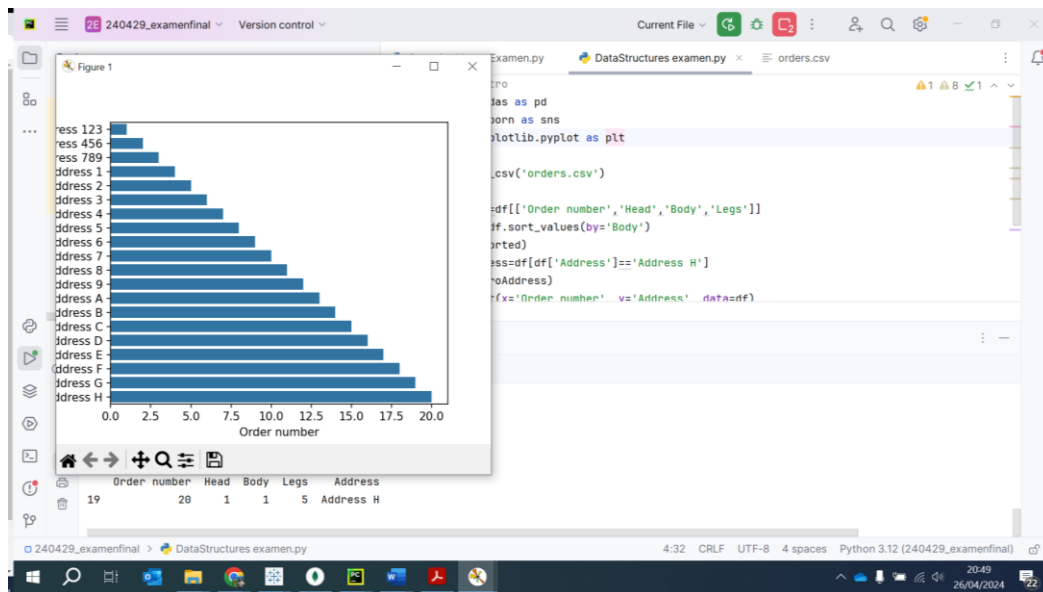
The output of the code is displayed in the console below the code cell, showing a DataFrame with columns: Order number, Head, Body, Legs, and Address. The data is filtered to only include rows where the 'Address' is 'Address H'.

	Order number	Head	Body	Legs	Address
18	19	4	6	2	Address 6
10	11	6	6	1	Address 8
5	6	1	6	6	Address 3
19	20	1	1	5	Address H

Diseño de gráfico en Python. (5 puntos)

Diseña un gráfico con la librería seaborn a tu elección en Python para mostrar los datos del DataFrame que hemos trabajado.

Se instala en primer lugar la librería seaborn.



Enlace de github.

<https://github.com/Tamaradelestal/examenpython.git>

Actividad 3 - Visualización con Tableau

Comparte el enlace a tu dashboard

Aceso a datos(10 puntos)

En la sección Documentación / Archivos dispones de los archivos products tanto en formato csv como json.

Debes cargar el archivo que prefieras en Tableau (web o Desktop) como fuente de datos.

Diseño de Dashboard(10 puntos)

Diseña un dashboard en donde muestras 4 gráficos para analizar el precio

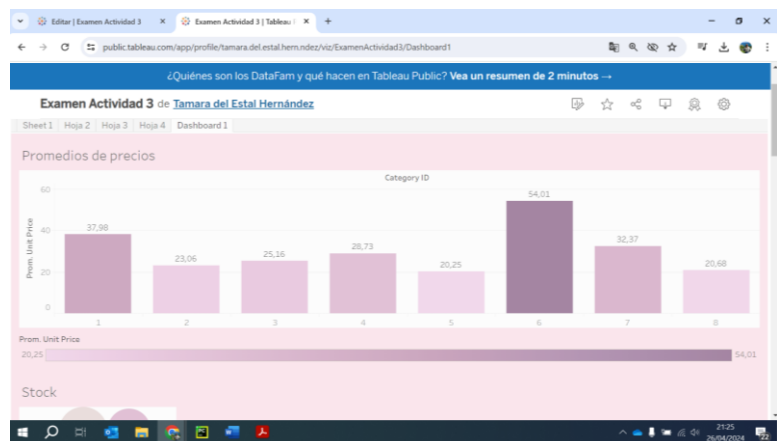
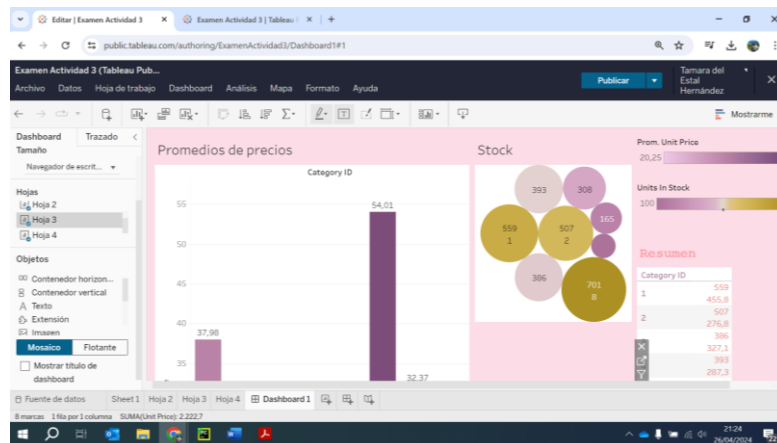
stock

categorías de los productos

Enlace al DashBoard. (5 puntos)

Comparte el enlace del dashboard que has diseñado.

https://public.tableau.com/views/ExamenActividad3/Dashboard1?:language=es-ES&:sid=&:display_count=n&:origin=viz_share_link



Actividad 4 - Visualización con Tableau 2

Comparte el enlace a tu dashboard

Acceso a datos(10 puntos)

En la sección Documentación / Archivos dispones del archivo USDA_activity_data. Debes cargar el archivo USDA_activity_data en Tableau (web o Desktop) como fuente de datos.

Explora los datos que ofrece y comprueba si son numéricos o de texto. También comprueba que puedes tener datos geográficos.

Diseño de Dashboard(10 puntos)

En una hoja diseña un gráfico que utilice datos geográficos.

En otra hoja diseña un gráfico que incorpore alguna especie de filtro.

Enlace al DashBoard. (5 puntos)

Comparte el enlace del dashboard que has diseñado

https://public.tableau.com/views/ExamenActividad4/Dashboard1?:language=es-ES&:sid=&:display_count=n&:origin=viz_share_link