



**Universidad Nacional Autónoma
de México**

Facultad de ciencias



Manejo de datos

Profesor(es):

Jessica Santizo Galicia

Sergio Alejandro Chávez Molotla

Integrantes:

González Robles Sofía Quetzalli

Manríquez Rangel Armando Daniel

Mariano Martínez Kevin

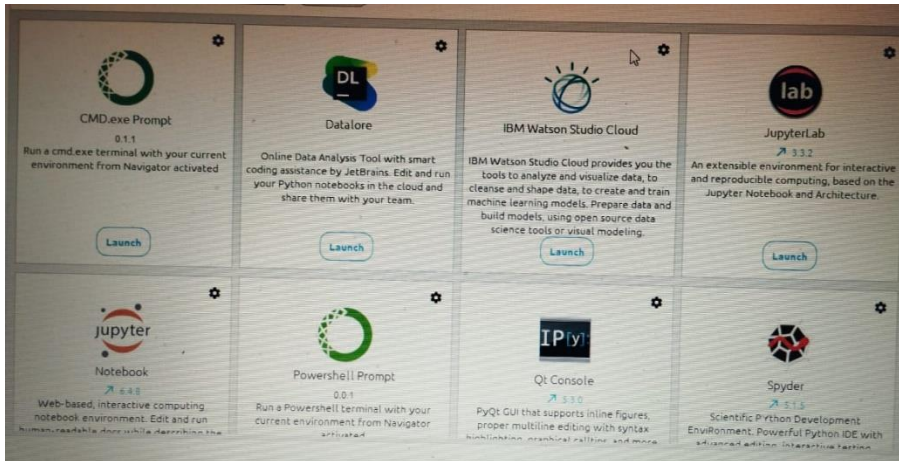
Serralde Salinas Alejandro

Semestre 2023-1

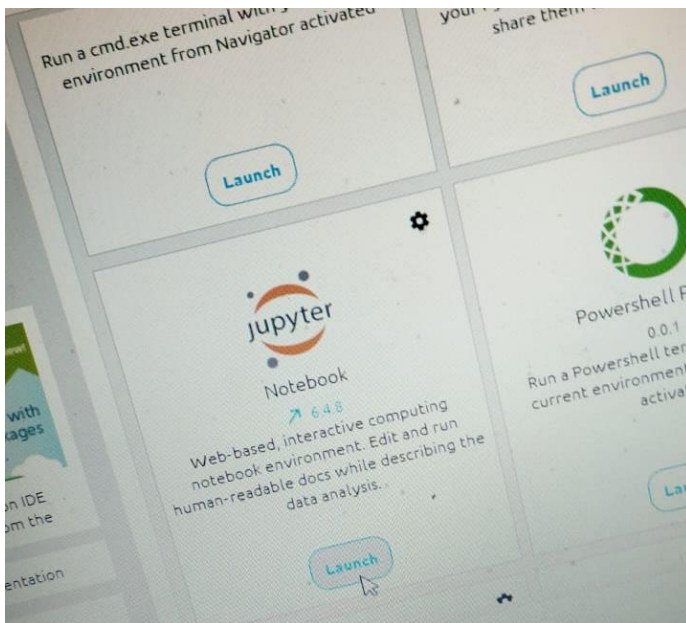
Readme.md

Para que puedan abrir nuestro web scrapper lo primero que necesitamos es que descarguen el archivo llamado Armando boutique .jpynb

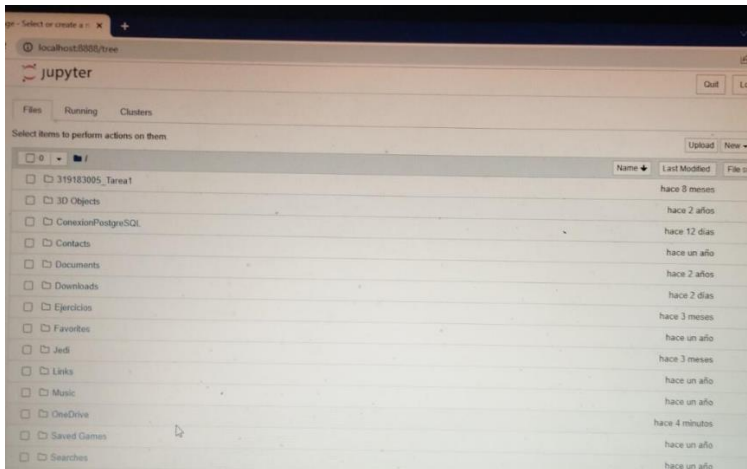
- Una vez con Anaconda instalado, en el buscador buscamos Anaconda Navigator, nos abrirá esto si no instalamos anaconda pasos en classroom para la profa y Sergio



- Abrimos el jupyter notebook para un mejor manejo

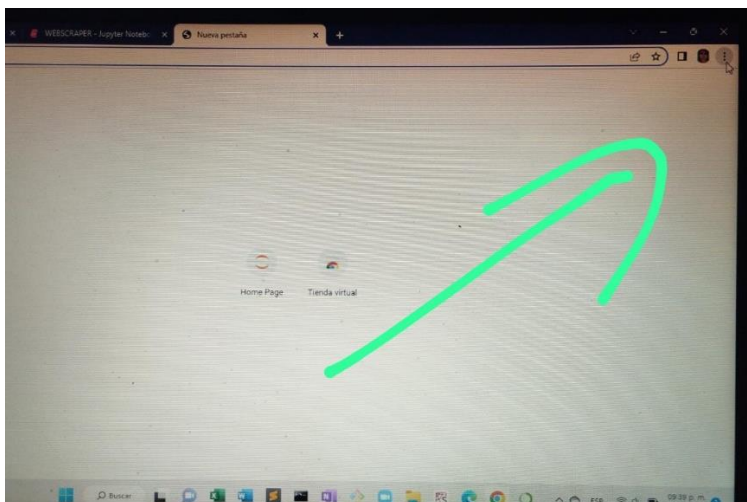


- Y nos va a mandar aquí y buscamos el archivo de Armando Boutique en la ubicación donde lo descargamos

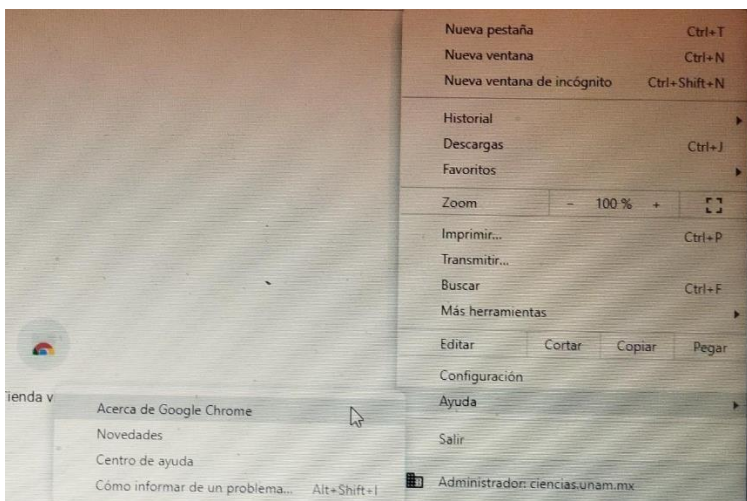


Y una vez con esto, nos falta instalar el web driver, abriremos nuestro Google Chrome y le damos en especificaciones para nuestras especificaciones

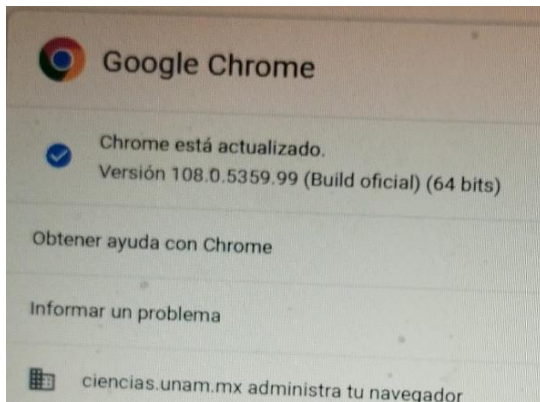
Nota: si ustedes ya tiene el Chrome Driver se puede saltar estos pasos (:



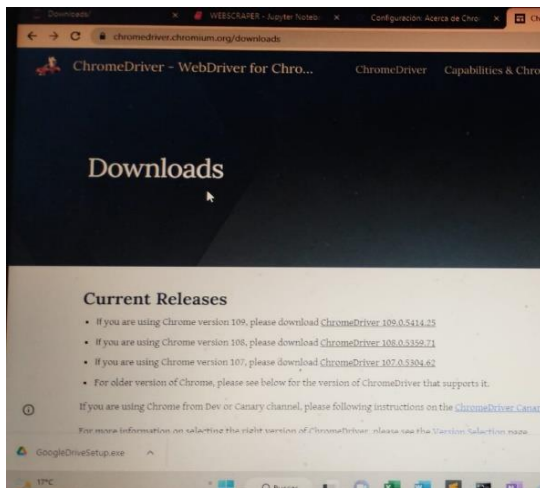
Le dan en ayuda y en la ventana que despliega le damos en acerca de Google Chrome



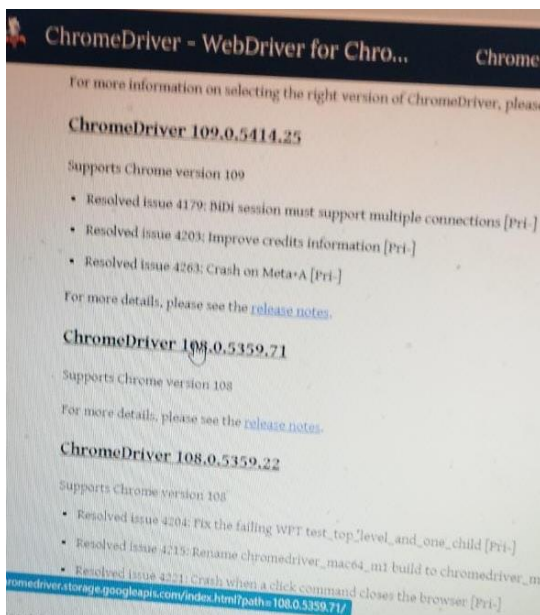
Y ya con las especificaciones y versión de nuestro Chrome (Ejemplo)



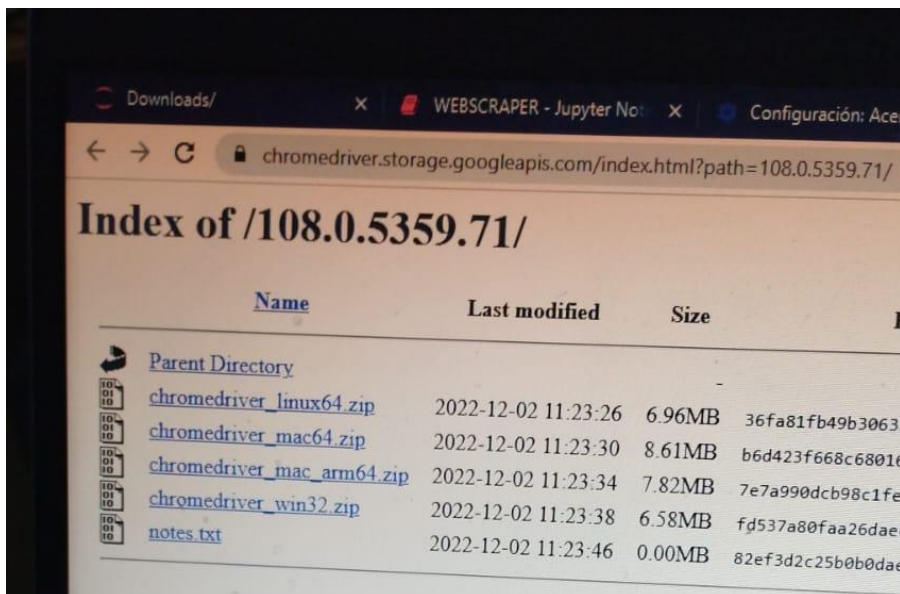
Buscamos en el buscador Chrome driver



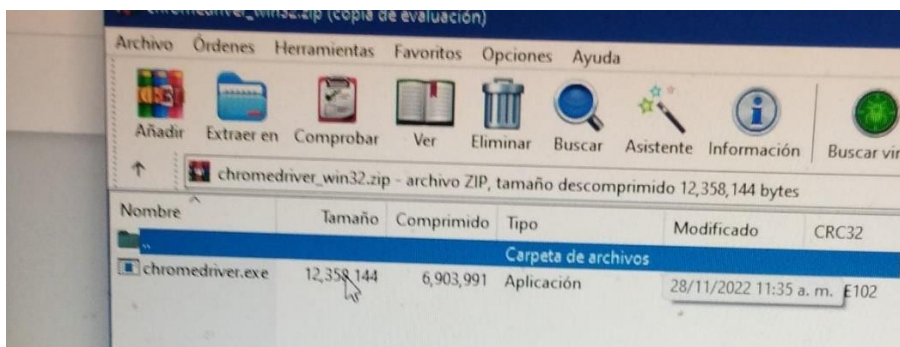
y si no aparece nuestra versión ponemos la que mas se acerca



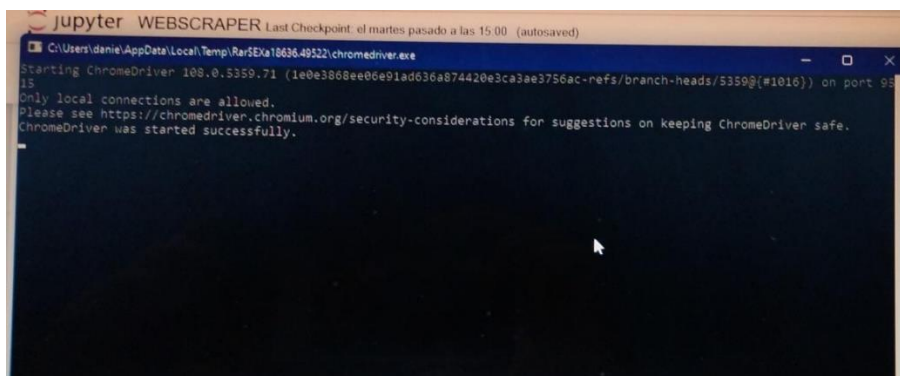
Y escogemos la opción de nuestro sistema operativo



Lo descargamos y descomprimos



Y estará listo cuando nos aparezca esta ventana



Continuando con el código para poder correr deben de:

Para la parte de graficas descargamos: Selenium, Pandasql

E importamos: Beautiful Soup, pandas, Requests, (y básicamente todas las que están al inicio del Notebook de Jupyter que dicen Import)

Después cambiar la ubicación del path mostrado a la ubicación donde guardaron su Chrome driver.

```
def CCP(producto):  
    path = "C:\\webdriver\\chromedriver.exe"
```

 ruta a cambiar por la suya

Y en nuestro caso el time sleep es para que en el tiempo de 8 seg pueda cargar por completo la pagina y no tener complicaciones esto puede variar del equipo de cada persona.

Pero en este caso esperaremos hasta que se carguen las paginas y no tengamos ningún problema

```
driver=webdriver.Chrome(path)  
time.sleep(8)
```

El paso siguiente seria darle en correr paso por paso, ya que esta es la diferencia en jupyter pero es para que tenga una mejor presentacion de cada parte de nuestro web scrapper y estara lista cuando aparezca la palomita dependiendo la versión.

```
def CCP(producto):  
    path = "C:\\webdriver\\chromedriver.exe"
```

nota: una vez corriendo aplicamos los 3 productos

esperamos a que se carguen las paginas ya que de lo contrario toda nuestra informacion no se cargara por completo y veremos que ya esta lista nuestra informacion porque las ventas del Driver se abren solas y no movemos nada hasta que se cierran solas

Una vez dando a correr a cada paso hasta el data frame como se ve a continuacion

```
[ ] ##Aplicamos a 3 productos  
df_zara = pd.DataFrame()  
for productos in ["vestido", "chamarras", "pantalones"]:  
    aux_zara_producto(productos)  
    df_zara = pd.concat([df_zara, aux])
```

```
[ ] ##mostramos el resultado  
df_zara
```

	Fecha	Tienda	Producto	Nombre	URL	PrecioFinal	PrecioAnterior
0	14/12/22	ZARA	vestido	VESTIDO SKATER PUNTO	https://www.zara.com/mx/es/vestido-skater-punt...	1.199,00 MXN	NaN
1	14/12/22	ZARA	vestido	VESTIDO PLISADO CINTURÓN	https://www.zara.com/mx/es/vestido-plisado-cin...	699,00 MXN	1.299,00 MXN
2	14/12/22	ZARA	vestido	VESTIDO CRUZADO SATINADO	https://www.zara.com/mx/es/vestido-cruzado-sat...	899,00 MXN	NaN
3	14/12/22	ZARA	vestido	VESTIDO MIDI JACQUARD	https://www.zara.com/mx/es/vestido-midi-jacqu...	1.299,00 MXN	NaN
4	14/12/22	ZARA	vestido	VESTIDO CORTO ESTAMPADO	https://www.zara.com/mx/es/vestido-corto-estam...	1.199,00 MXN	NaN
...
25	14/12/22	ZARA	pantalón	PANTALÓN FLARE TERCIOPELO	https://www.zara.com/mx/es/pantal%C3%B3n-flare...	1.299,00 MXN	NaN
26	14/12/22	ZARA	pantalón	PANTALÓN ZW THE MARINE STRAIGHT EFECTO PIEL	https://www.zara.com/mx/es/pantal%C3%B3n-zw-8h...	899,00 MXN	NaN
27	14/12/22	ZARA	pantalón	PANTALÓN SATINADO RECTO	https://www.zara.com/mx/es/pantal%C3%B3n-satin...	1.299,00 MXN	NaN
28	14/12/22	ZARA	pantalón	PANTALÓN JOGGER SATINADO EFECTO ARRUGADO	https://www.zara.com/mx/es/pantal%C3%B3n-jogge...	749,00 MXN	NaN
29	14/12/22	ZARA	pantalón	PANTALÓN ABERTURAS LATERALES	https://www.zara.com/mx/es/pantal%C3%B3n-abert...	899,00 MXN	NaN

90 rows x 7 columns

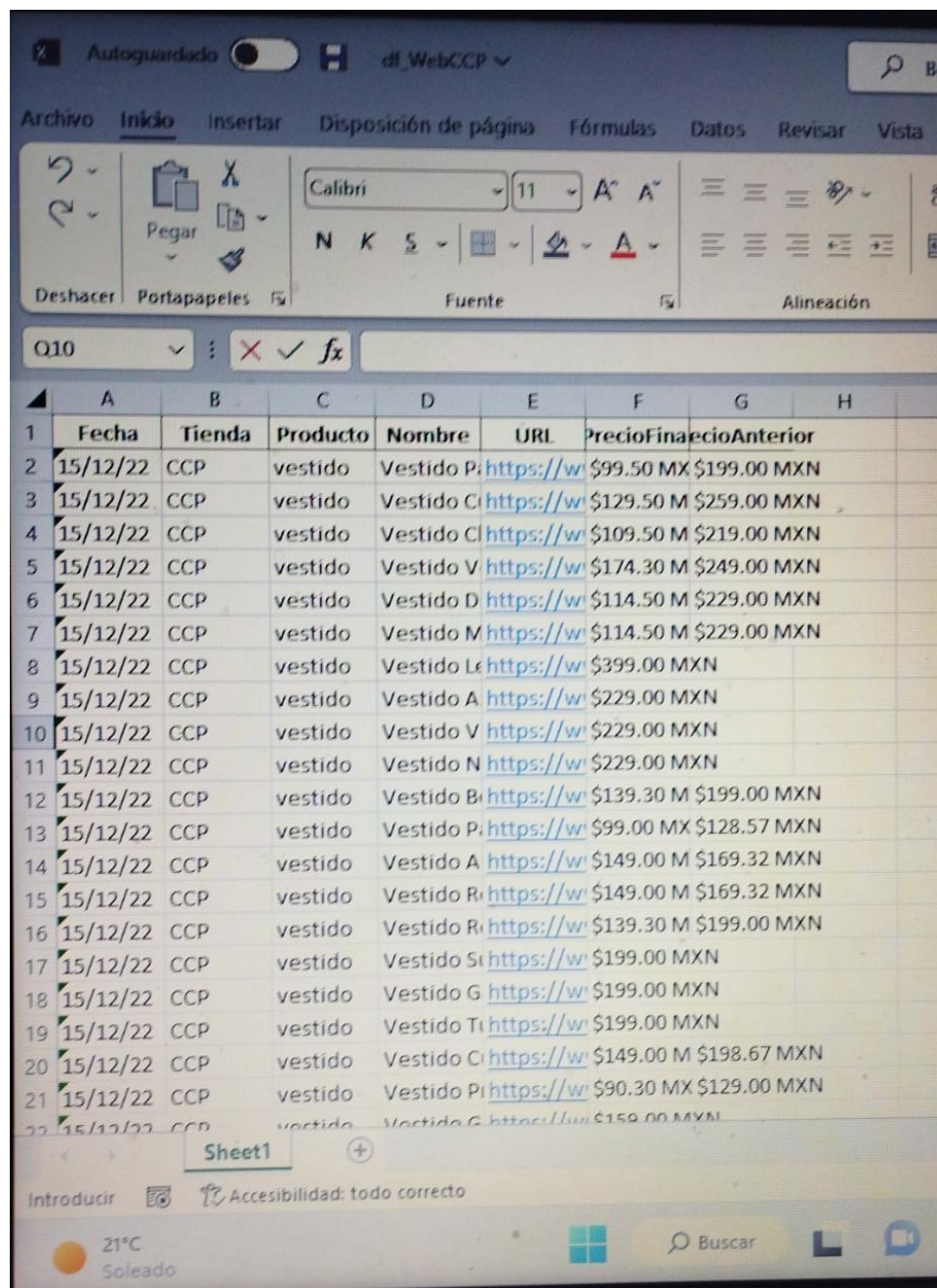
Ahora lo importaremos a excel

```
###Pasamos a el excel  
df_zara.to_excel("df_WebZara.xlsx",index=False)
```

Este paso lo hacemos con cada tienda y al ultimo los concatenamos para comparar

Pero ya con este paso lo podemos buscar nuestro excel creado en nuestro explorador de archivo

Anexamos foto del ejemplo de un tienda, el cual al importarlo nos debe de salir en nuestro explorador de archivos



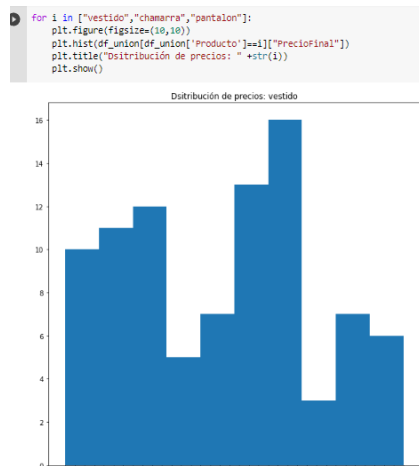
	A	B	C	D	E	F	G	H
1	Fecha	Tienda	Producto	Nombre	URL	PrecioFin	PrecioAnterior	
2	15/12/22	CCP	vestido	Vestido P	https://w	\$99.50 MX	\$199.00 MXN	
3	15/12/22	CCP	vestido	Vestido C	https://w	\$129.50 M	\$259.00 MXN	
4	15/12/22	CCP	vestido	Vestido C	https://w	\$109.50 M	\$219.00 MXN	
5	15/12/22	CCP	vestido	Vestido V	https://w	\$174.30 M	\$249.00 MXN	
6	15/12/22	CCP	vestido	Vestido D	https://w	\$114.50 M	\$229.00 MXN	
7	15/12/22	CCP	vestido	Vestido M	https://w	\$114.50 M	\$229.00 MXN	
8	15/12/22	CCP	vestido	Vestido L	https://w	\$399.00 MXN		
9	15/12/22	CCP	vestido	Vestido A	https://w	\$229.00 MXN		
10	15/12/22	CCP	vestido	Vestido V	https://w	\$229.00 MXN		
11	15/12/22	CCP	vestido	Vestido N	https://w	\$229.00 MXN		
12	15/12/22	CCP	vestido	Vestido B	https://w	\$139.30 M	\$199.00 MXN	
13	15/12/22	CCP	vestido	Vestido P	https://w	\$99.00 MX	\$128.57 MXN	
14	15/12/22	CCP	vestido	Vestido A	https://w	\$149.00 M	\$169.32 MXN	
15	15/12/22	CCP	vestido	Vestido R	https://w	\$149.00 M	\$169.32 MXN	
16	15/12/22	CCP	vestido	Vestido R	https://w	\$139.30 M	\$199.00 MXN	
17	15/12/22	CCP	vestido	Vestido S	https://w	\$199.00 MXN		
18	15/12/22	CCP	vestido	Vestido G	https://w	\$199.00 MXN		
19	15/12/22	CCP	vestido	Vestido T	https://w	\$199.00 MXN		
20	15/12/22	CCP	vestido	Vestido C	https://w	\$149.00 M	\$198.67 MXN	
21	15/12/22	CCP	vestido	Vestido P	https://w	\$90.30 MX	\$129.00 MXN	
22	15/12/22	CCP	vestido	Vestido G	https://w	\$159.00 MXN		

Corremos las consultas nuevamente haciendo paso a paso cada una verificando lo aprendido visto en las clases de introducción a SQL (Ejemplo de cómo les aparecerán)

```
###Consulta donde muestra el nombre y precios de los productos con algun descuento
ps.sqldf("select Nombre,PrecioFinal,PrecioAnterior from df_zara where PrecioAnterior <= 5000")
```

	Nombre	PrecioFinal	PrecioAnterior
0	VESTIDO PLISADO CINTURÓN	699.0	1299.0
1	CAZADORA ACOLCHADA CAPUCHA	959.0	1599.0
2	PANTALÓN ANCHO PINZAS	479.0	799.0

Por último damos run a las gráficas



Y siguiendo estos pasos no deberían tener problema alguno en revisar nuestro web scraper

Gracias