

WEB SCRAPING

Pre-presentación Proyecto Grupo 2 PRODUCTO 6

Integrantes:

- FABIAN OMAR FIGUEROA BUENO
- FELIPE GUAICO BARRAZA
- DAVID VALENZUELA PEREZ

Agenda

Tecnología a utilizar
Acciones de navegación
Diagrama del proceso / Pseudo código
Estructura archivo / Output / Docs

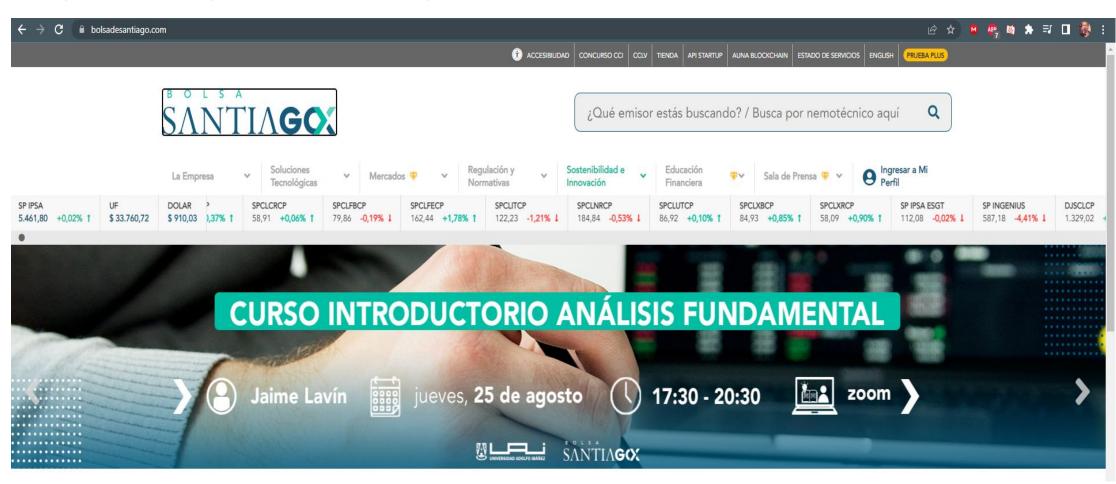
Tecnología a utilizar Proyecto – Grupo 2

Se busca que el proyecto integre todas las tecnologías vistas en clase, para lo cual, desarrollaremos un script que pueda:

- 1. Navegar la página de la bolsa de Santiago (https://bolsadesantiago.com/) con **Selenium**.
 - Todo esto con iteradores y listas para consultar un número controlado de empresas.
- 2. Descargar tablas Excel que posee la página y almacenar en una carpeta específica para cada empresa (Mediante clics con Selenium)
- 3. Descargar información adicional que se refleje en la pagina, con Beautiful Soup 4
- 4. Sobre las lógicas:
 - Idealmente si la Empresa es "nueva" en el proceso de consultas, deberemos crear una carpeta con Selenium y/o OS (Siempre y cuando exista dentro del consultador la bolsa de Santiago.
 - De lo contrario, deberá leerse la carpeta respectiva con los archivos descargados anteriormente.
 - Los puntos a y b de este apartado deberán ser controlados y manipulados con Pandas

Queremos ver la posibilidad de complementar la extracción de la información con el uso de alguna API que pueda existir para la bolsa de Santiago o páginas asociadas que puedan tener más historia para la información.

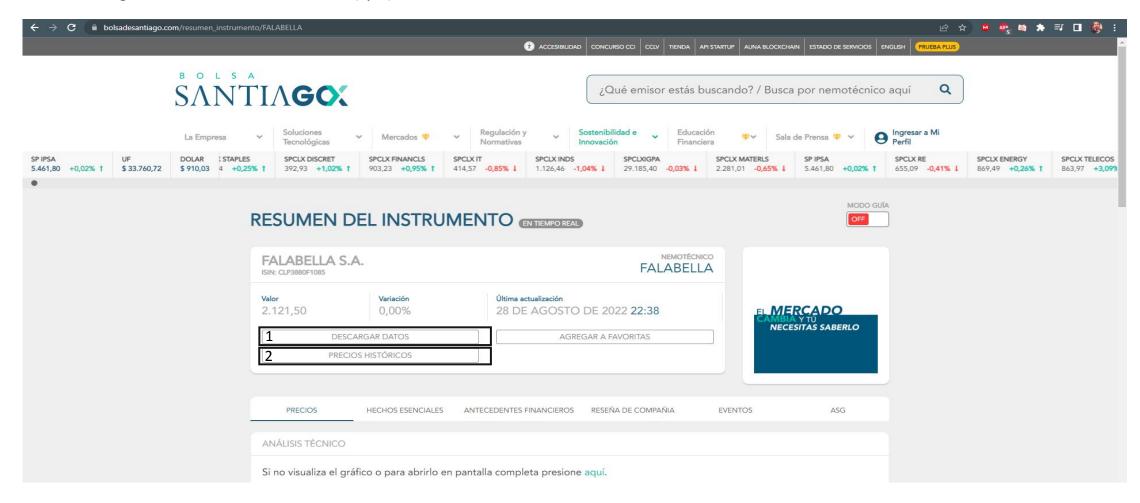
 Extracción de Información de Acciones Navegación: Ingresamos a la Página de la Bolsa de Santiago



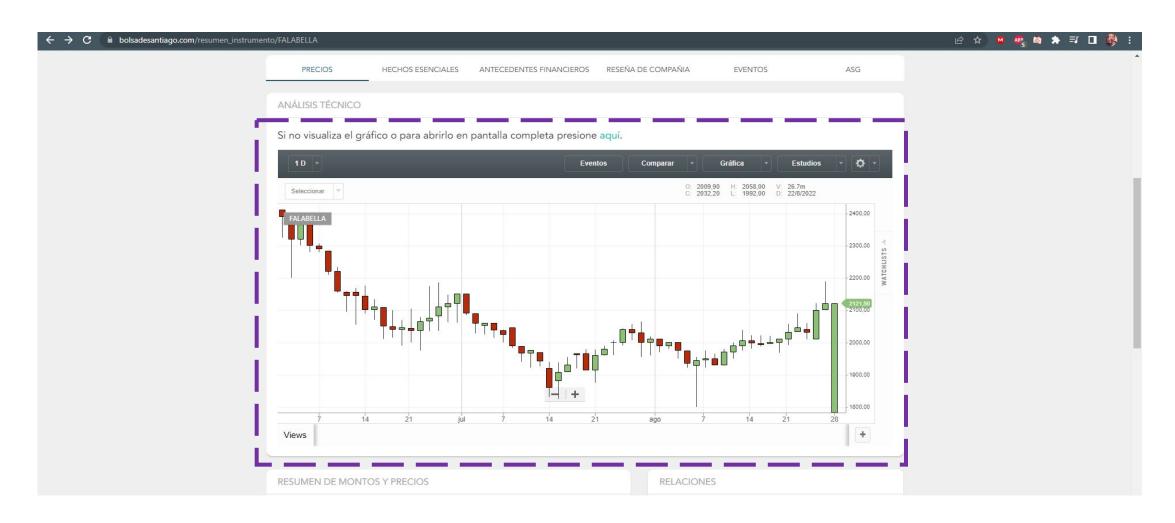
Consultamos una de las Empresas a monitorear



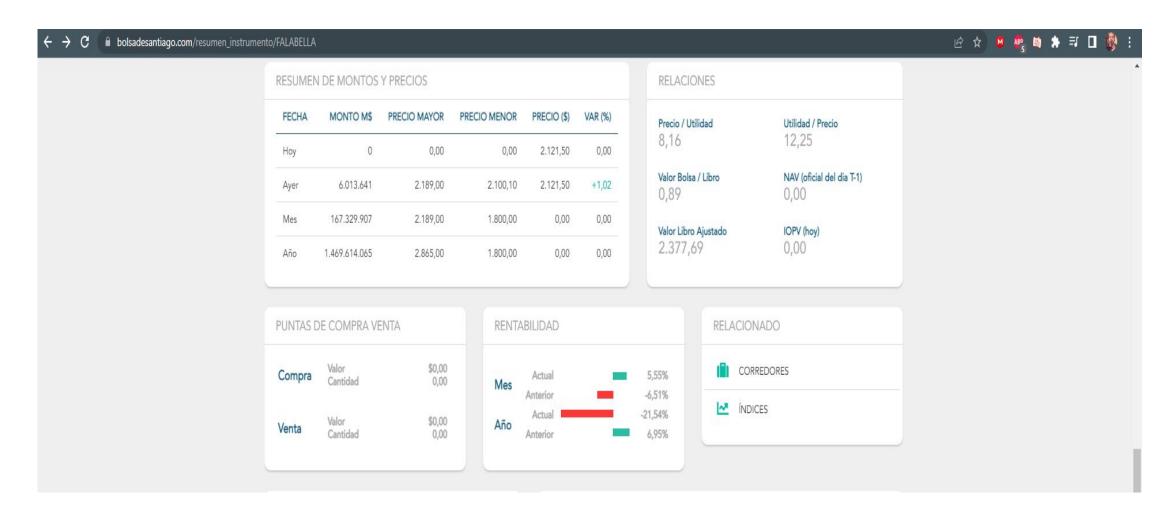
- 3. Si la empresa existe, apretar Enter e ingresar a la misma
- 4. Descargamos los Excel de los Puntos 1) y 2)



5. Abajo, aparece un gráfico que tiene más data histórica, ver si se puede conseguir con una API. – Deseable



6. En la misma página, seguimos bajando y se observan los siguientes indicadores:



7. Descargar con BS4, las últimas variaciones de Capital y respaldar en algún lado:

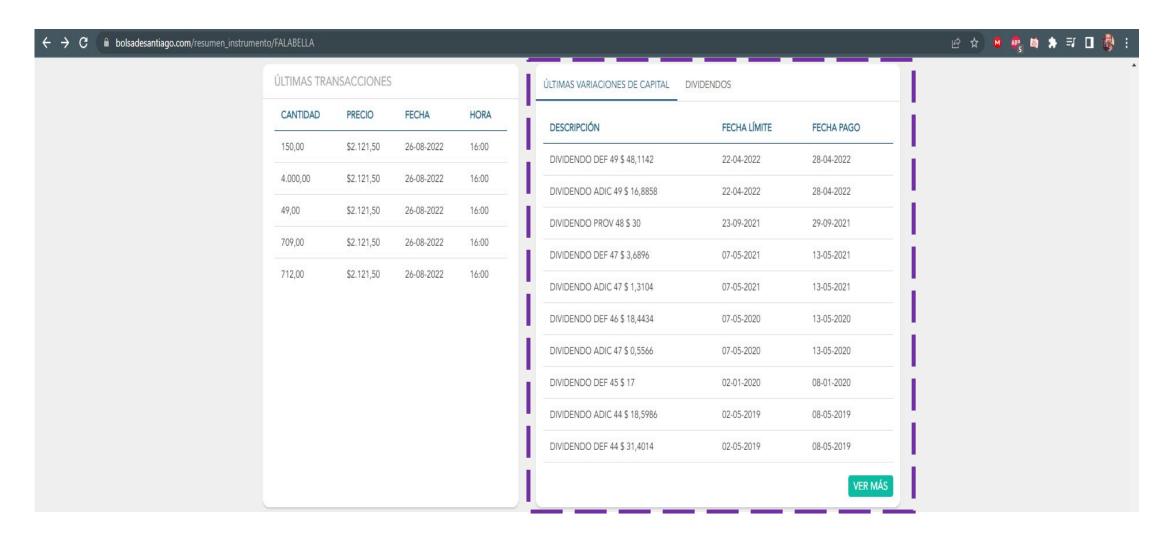


Diagrama del proceso / Pseudo código Proyecto – Grupo 2

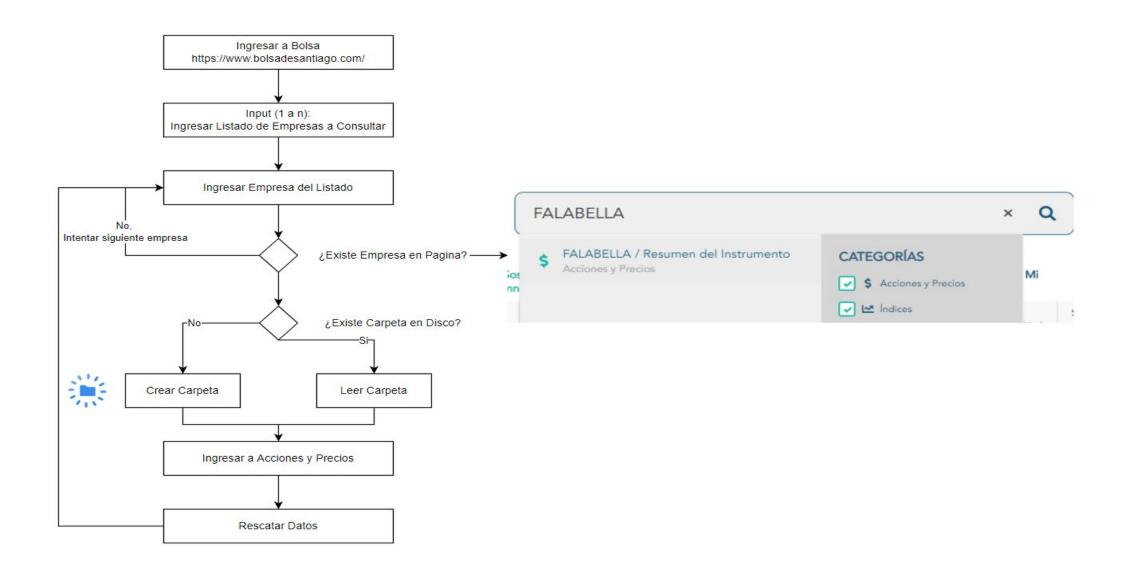
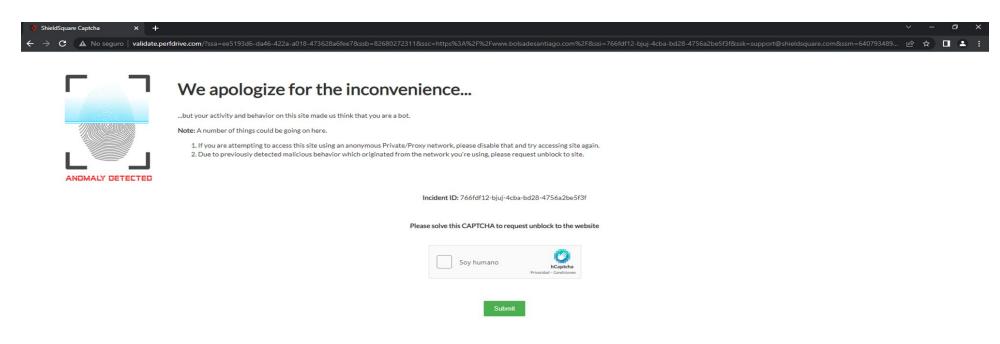


Diagrama del proceso / Pseudo código Proyecto – Grupo 2

Pero...

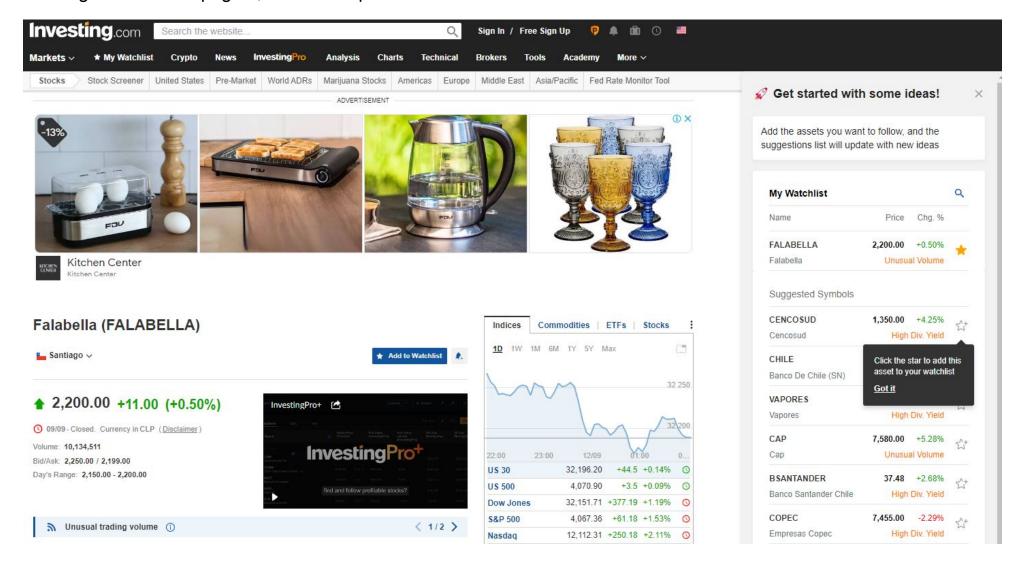


Tecnología a utilizar (Opción 2) Proyecto – Grupo 2

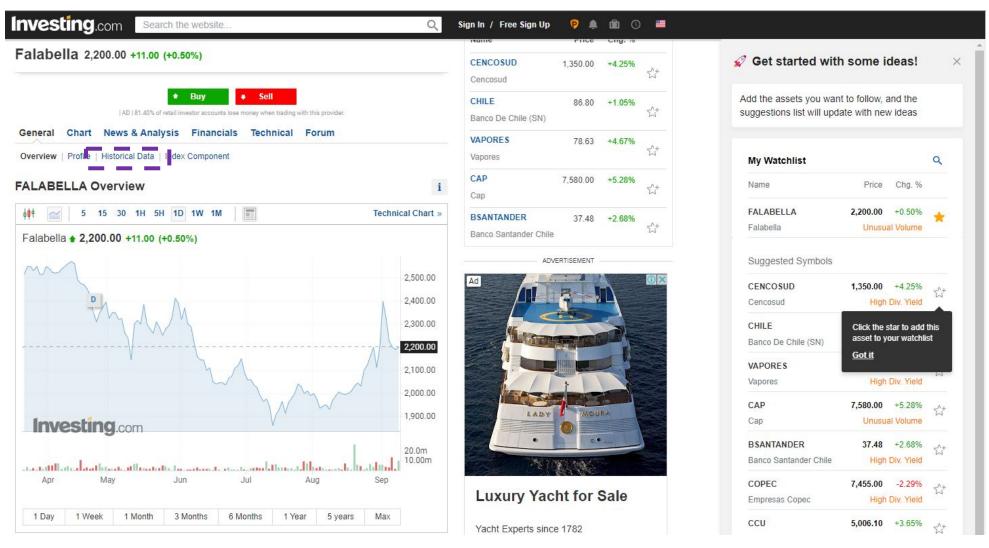
Se busca que el proyecto integre todas las tecnologías vistas en clase, para lo cual, desarrollaremos un script que pueda:

- 1. Revisar si existe la carpeta output, en caso de no existir, la crearemos.
- 2. Navegar la página de Investing (https://www.investing.com/equities/falabella) con **Selenium**.
 - Realizar todas las interacciones necesarias vía navegador para que la página web genere la información que necesitamos.
- 3. Integrar el código fuente obtenido de la interacción de **Selenium** con **Beautiful Soup 4** para poder descargar la información.
- 4. Dar formato la información descargada:
 - Convertir la información a DF.
 - Limpiar y ajustar el formato del DF.
 - Exportar a csv.

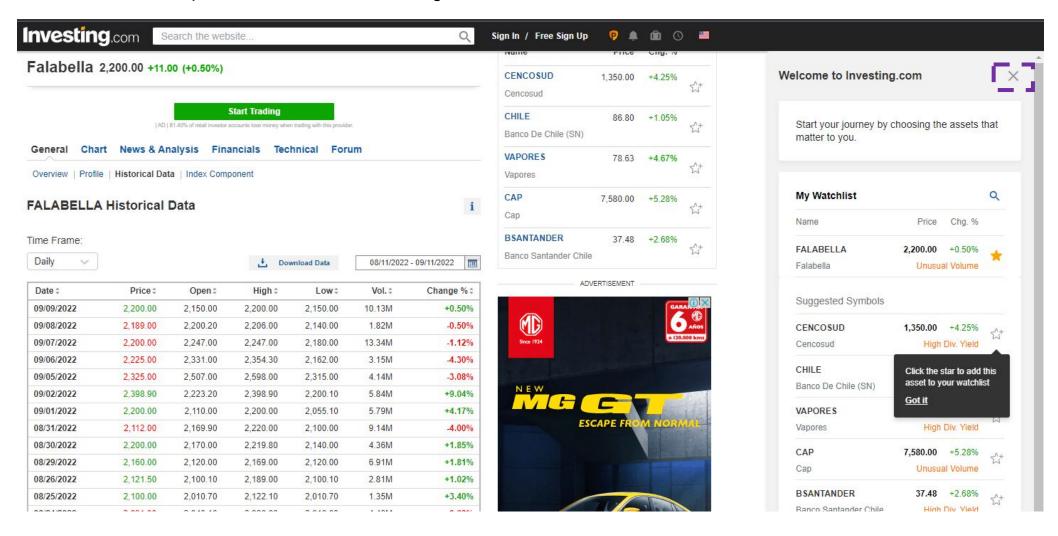
1. Ingresamos a la página, enfocados por ahora en Falabella.



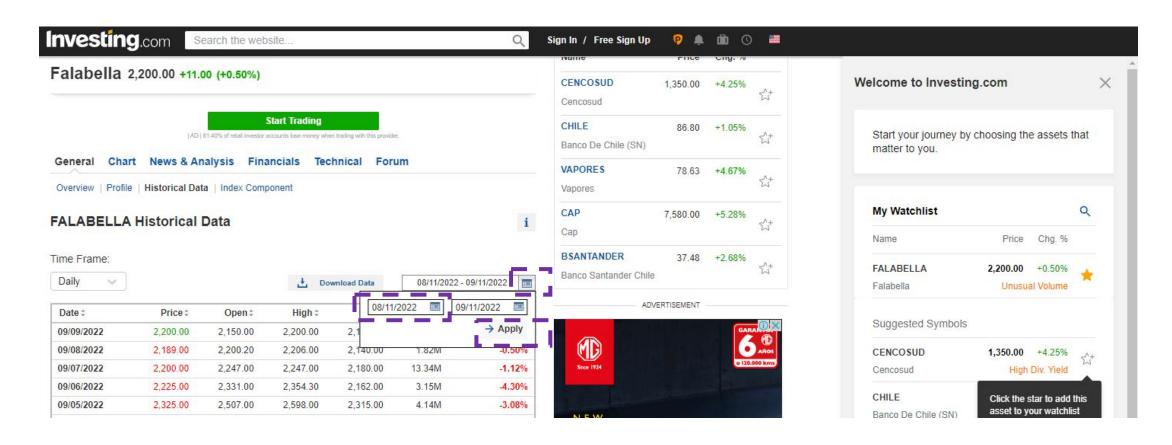
- 2. Hacemos un "AVPÁG" para bajar una pantalla y poder interaccionar con el link "Historical Data"
- Hacemos clic en "Historical Data"



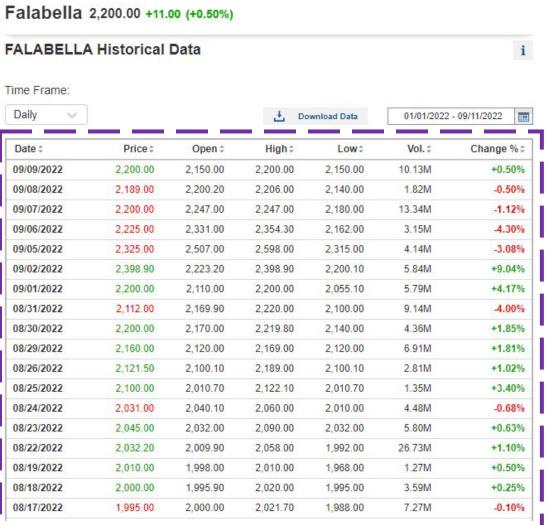
- 4. Hacemos un "AVPÁG" para bajar una pantalla y poder interaccionar con el link "Historical Data"
- 5. Cerramos el PopUP de "Welcome to Investing.com"



- 5. Aperturamos el Selector de Fechas
- 6. Ingresamos la fecha Inicial de "01-01-2022"
- 7. Hacemos clic en "Apply"



8. Descargamos Tabla







- Descargamos la tabla completa
- Convertimos la Tabla a DF con formato
- Escribimos DF en .CSV

Diagrama del proceso / Pseudo código (Opción 2) Proyecto – Grupo 2

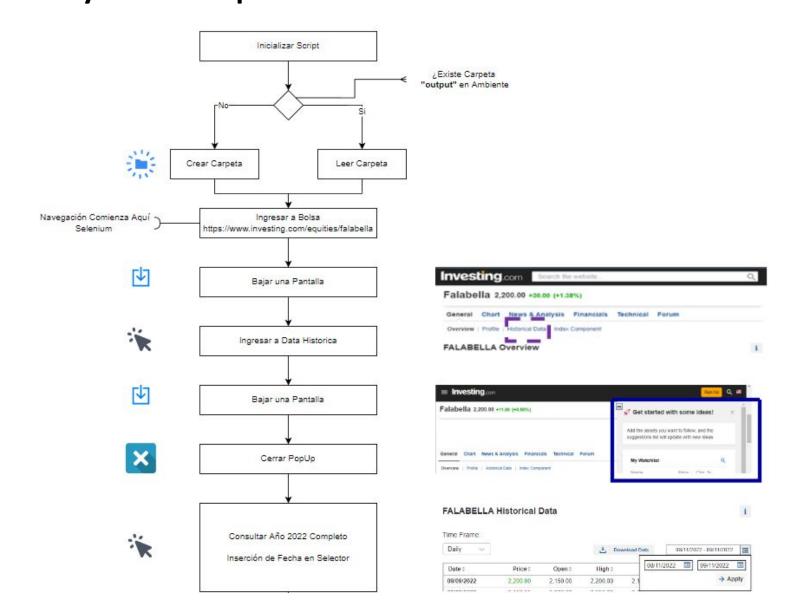
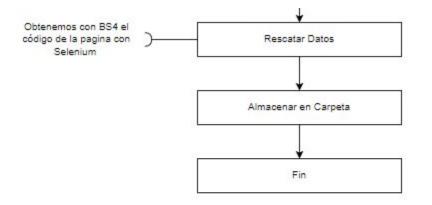


Diagrama del proceso / Pseudo código (Opción 2)

Proyecto – Grupo 2





Estructura archivo / Output / Docs Proyecto – Grupo 2

EJ: Precio de la Acción de Falabella en el Tiempo:

	Date	Price	Open	High	Low	Vol	ChangePercent
0	2022-09-09	2200.0	2150.0	2200.0	2150.0	10.13M	0.50
1	2022-09-08	2189.0	2200.2	2206.0	2140.0	1.82M	-0.50
2	2022-09-07	2200.0	2247.0	2247.0	2180.0	13.34M	-1.12
3	2022-09-06	2225.0	2331.0	2354.3	2162.0	3.15M	-4.30
4	2022-09-05	2325.0	2507.0	2598.0	2315.0	4.14M	-3.08
	22.0	2.0		523		722	
172	2022-01-07	2500.0	2550.0	2571.0	2465.0	4.97M	-1.92
173	2022-01-06	2549.0	2535.0	2574.0	2519.9	2.47M	0.55
174	2022-01-05	2535.0	2575.9	2575.9	2509.9	1.42M	-0.55
175	2022-01-04	2549.0	2651.0	2651.0	2509.0	1.72M	-3.81
176	2022-01-03	2650.0	2774.7	2774.8	2650.0	1.39M	-4.64



Nombre	Tipo	Scrap
Date	Date	SI
Price	Float	SI
Open	Float	SI
High	Float	SI
Low	Float	SI
Vol.	Str	NO
Change_Percent	Float	SI

