**Explicació detallada del funcionament del codi de *Escales Vaixells.py***



Inicialment es carreguen les llibreries i funcions que s’utilitzaran.

Amb randint i sleep crearem la funció s(t1,t2) que ens permetrà fer descansar el programa un temps aleatori de t1 segons de mitjana, i t2 segons de variància, per donar-li l’aparença de comportament no robòtic.

Com que el nostre scraper el controlarem amb la llibreria selenium en necessitem les propietats de webdriver, Service i By.

També utilitzem WebDriverWait i expected\_conditions per esperar a que es carregui la informació quan la necessitem.

Amb Options li donarem el nom que volem al user-agent.

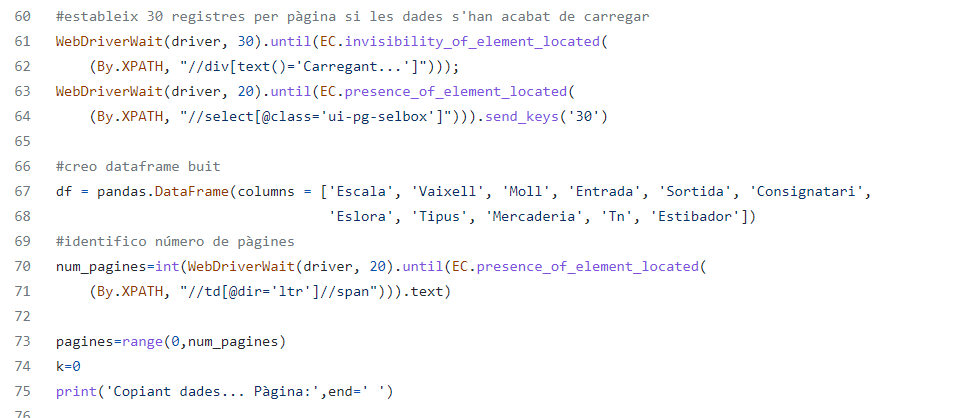


Li posem el nom humà de *Real Human Beep Bop* a l’user-agent, iniciem el chrome.driver i establim l’enllaç de la pàgina i les dates que ens interessen. Finalment hi ha la comprovació (que tenim comentada) del nom de l’user-agent.

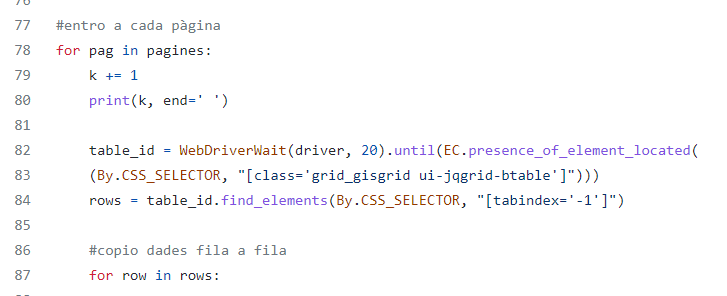


A partir d’ara ja és quan treballa el scraper i hem d’anar amb compte de que no vagi massa ràpid i no doni temps de carregar les pàgines. Per això utilitzem el WebDriverWait amb 20 segons d’espera fins a que els elements que busquem estiguin disponibles per a ésser clicats o detectats, en funció de cada acció que vulguem prendre.

Hem anat a històrics, li hem passat les dates d’interès i hem clicat a Veure. Li donem uns segons de marge per a que aparegui el text *Carregant*  abans de continuar.



Quan l’element *Carregant* desaparegui ja tindrem les dades carregades. Aquí és on creem la base de dades buida (però amb els noms de columnes) i identifiquem el número de pàgines que el scraper haurà de navegar. A partir d’aquí començarà el procés iteratiu.

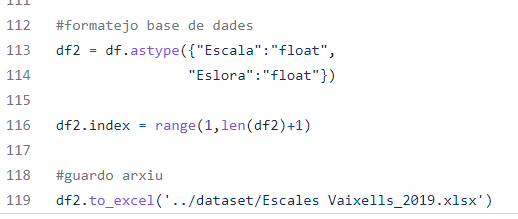


A cada pàgina identificarà la taula i n’agafarà la llista de totes les files. Dins de cada pàgina hi haurà un altre procés iteratiu per a cada fila de la taula.



Per a cada fila detectarà cada valor que ens interessa: escala, vaixell, moll, etc. Ho posa tot en un DataFrame temporal d’una fila, corregeix el format dels números (eliminar els punts de milers) i ho concatena a la base de dades.

En acabar la iteració de totes les files de la pàgina, clica al botó de pàgina següent i continua el procés a la pàgina següent. Quan arriba a l’última pàgina clica al botó però no té efecte.



Finalment es formata les columnes numèriques – a un nou DataFrame per no perdre les dades originals en cas que hi hagi algun error –, es dona a cada registre el número d’índex pertinent i es guarda l’arxiu en format .xlsx a la carpeta dataset.