Sprint 16 tasca 2 : Web Scapping

**Context**

L’objectiu principal de la practica es la d’explorar l’utilització de diferents llibreries de python per l’extracció de dades i amb elles generar un dataset. Les llibreries utilitzades han estat BeautifulSoup i Selenium i la pagina web d’on hem extret les dades la borsa de Madrid .

* BeautifulSoup

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

* Selenium

<https://www.selenium.dev/>

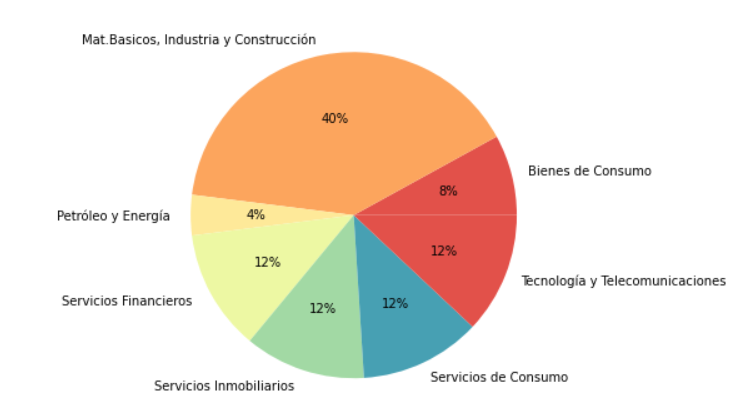
* Borsa de Madrid

<https://www.bolsamadrid.es/esp/aspx/Portada/Portada.aspx>

**Descripció**

El dataset conte 2 csv amb informació rellevant sobre la borsa de Madrid.En el csv anomenat Lista\_Empresas\_pag1.csv conte la primera pagina de les empreses que participen a la borsa .En el segon csv anomenat Resumen\_indices.csv conté la situació dels index actual dels diferents index.

**Representació Gràfica**



**Contingut**

El dataset Lista de Empresas conté diferents pàrametres que ens ajuden a extreure informació rellevant de cada empresa i per alte banda el dataset Resumen índices conté diferents atributs que conformen l’evolució del index en un periode concret.

* Lista\_Empresas\_pag1.csv
  + Nombre: correspon al nom de l’empresa
  + Sector : correspon al sector general on l’empresa desenvolupa la seva activitat
  + Subsector: és el sector especific que ocupa l’empresa dintre dels diferents sectors
  + Mercado: és el mercat on l’empresa cotitza dintre de la borsa
  + Índices: es l’index que li correspon dintre del mercat bursátil
* Resumen\_indices.csv
  + Nombre: nom de l’index
  + Anterior: preu anterior a la obertura dels mercats
  + Último: preu de tancament en els mercats
  + %Dif: diferencia percentual entre el valor anterior i ultim
  + Màximo: valor maxim del periode
  + Mínimo: valor minim del periode
  + Fecha: data de la lectura
  + Hora: hora de la lectura
  + % Dif.\nAño 2021: la diferencia percentual respecte l’any 2021

**Inspiració**

Crec que aquest dataset pot servir d’inspiració per començar a fer els primers pasos en el webscrapping i les diferents llibreries que ofereix Python.A més , pot servir per entendre com es pot extreure informació rellevant de pagines web.

**Codi i Dataset:**

https://github.com/AdriaMoreno1994/Machine-Learning-avan-at