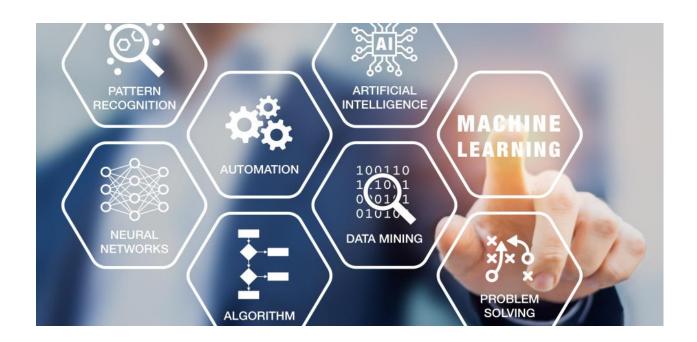


ROYAUME DU MAROC UNIVERSITÉ ABDELMALEK ESSAADI FACULTÉ DES SCIENCES ET TECHNIQUE



TANGER

Rapport de projet de fin de module NoSql



Fait par:

Encadré par:

-> SOUFIANI Assia

-> Prof . EL ACHAK Lotfi

-> MECHHEDAN Insafe

I. Introduction:

Le fait d'utiliser des images afin de comprendre des données existe depuis des siècles, allant des cartes et diagrammes au 17e siècle jusqu'à l'invention du graphique circulaire au début du 19e siècle. Plusieurs décennies plus tard, Charles Minard marqua l'histoire des graphiques statistiques lorsqu'il cartographia l'invasion de la Russie par Napoléon. Sa carte représentait la taille de l'armée et le chemin emprunté par Napoléon pour se retirer de Moscou, ainsi qu'une échelle des températures accompagnées de leur date, pour une compréhension plus immersive des évènements.

Pourtant c'est la technologie qui a véritablement propulsé en avant la visualisation de données. Les ordinateurs ont rendu possible le traitement de grandes quantités de données à une vitesse éclair. Aujourd'hui, la visualisation de données (ou Dataviz) est un savant mélange de science et d'art, qui évolue rapidement et va assurément bouleverser le paysage des entreprises dans les années à venir.

La visualisation de données est la présentation de données sous forme graphique. Elle permet de présenter les données analytiques de manière visuelle, afin de mieux comprendre des concepts complexes ou d'identifier de nouvelles tendances. Grâce à la visualisation interactive, il est possible de pousser le concept encore plus loin en utilisant cette technologie pour explorer en profondeur les tableaux et graphiques, en modifiant interactivement les données que vous voyez et la façon dont elles sont traitées.\

II. L'objectif du projet:

L'objectif principal du notre projet est la réalisation d'une plateforme web de type single page application , pour les opperations d'un ETL " Extract, Transform and Load " et Data Visualisation / Analysis depuis une base de donnees NoSql (Not Only SQL).

III. La description du projet:

Dans notre projet nous avons tout d'abord collecter les donnees d'apres les deux sites Amazon et Flipkart.

L'etape suivante c'est le pre-processing et pre-traitement des données .

Puis nous avons commencer a travailler sur le backend en reliant django,mongodb,elasticsearch..

Et en fin l'etape de visualisation de données.

IV. Les outils utilisés:



Langage de programation utilise dans le backend.



Django est un framework Python de haut niveau, permettant un développement rapide de sites internet, sécurisés, et maintenables



Angular est une plateforme de développement, construite sur TypeScript



permet de créer un objet Python représentant un document HTML. Une manière de créer un tel objet est de passer simplement le code HTML.



permet de créer un objet Python représentant un document HTML. Une manière de créer un tel objet est de passer simplement le code HTML.



MongoDB est une base de données NoSQL orientée documents.le code HTML.



Elasticsearch est un moteur de recherche et d'analyse RESTful distribué, conçu pour répondre à une multitude de cas d'utilisation



GraphQL est un langage de requête et un environnement d'exécution côté serveur pour les interfaces de programmation d'application (API).

V. Scraping des données:

Nous avons pu scraper les donnees a l'aide beautifulsoup.

Scraping Flipkart:

```
import requests
import pandas as pd
from bs4 import BeautifulSoup
import csv

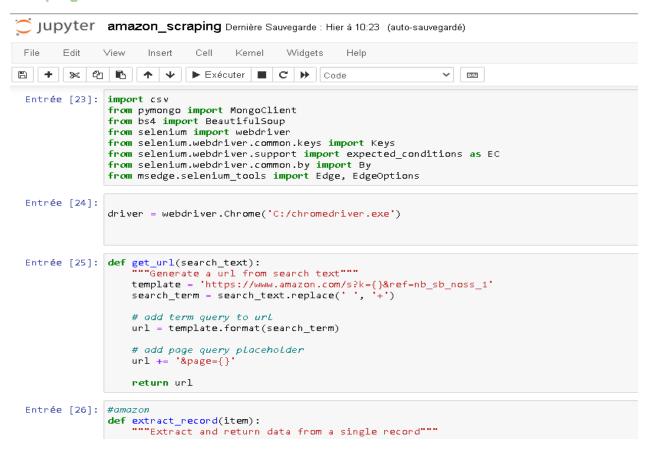
cfrom itertools import zip_longest

print("type what you want to search")
subject = input()
links_title = []
description_title = []
Prix = []
Links = []
products = []
num_page = 1

while True:
resultat = requests.get(
f"https://www.flipkart.com/search?q={subject}&otracker=search&otracker1=search&marketplace=FLIPKART&as-sh
src = resultat.content
srcs = resultats.content
srcs = resultats.content
soup = BeautifulSoup(src, "lxml")
```

Manufacturer;Model Name;Category;Screen Size;Screen;CPU;RAM;Storage;GPU;Operating System;;Weight;Price (Euros) Apple; Macbook Air; Ultrabook; "13.3"""; 1440x900; Intel Core is 1.8GHz; 8GB; 128GB Flash Storage; Intel HD Graphics 6000; macOS; HP;250 G6;Notebook;"15.6""";Full HD 1920x1080;Intel Core i5 7200U 2.5GHz;8GB;25GGB SSD;Intel HD Graphics 620;No OS;;1.86 Apple; MacBook Pro; Ultrabook; "15.4"""; IPS Panel Retina Display 2880x1800; Intel Core i7 2.76Hz; 166B; 512GB SSD; AMD Radeon F Acer; Aspire 3; Notebook; "15.6" "; 1366x768; AMD A9-Series 9420 3GHz; 4GB; 500GB HDD; AMD Radeon R5; Windows; ; 2.1kg; 400 Apple; Macbook Air; Ultrabook; "13.3""; 1440x900; Intel Core i5 1.86Hz; 86B; 256GB Flash Storage; Intel HD Graphics 6000; macOS; Asus; ZenBook UX430UN; Ultrabook; "14.0"""; Full HD 1920x1080; Intel Core i7 8550U 1.8GHz; 16GB; 512GB SSD; Nvidia GeForce MX150 Acer; Swift 3; Ultrabook; "14.0"""; IPS Panel Full HD 1920x1080; Intel Core i5 8250U 1.6GHz; 8GB; 256GB SSD; Intel UHD Graphics HP;250 66;Notebook;"15.6""";1366x768;Intel Core i5 7200U 2.5GHz;4GB;500GB HDD;Intel HD Graphics 620;No 0S;;1.86kg;393,9 HP;250 G6;Notebook;"15.6""";Full HD 1920x1080;Intel Core i3 6006U 2GHz;4GB;500GB HDD;Intel HD Graphics 520;No OS;;1.86kg Apple; MacBook Pro; Ultrabook; "15.4"""; IPS Panel Retina Display 2880x1800; Intel Core i7 2.8GHz; 16GB; 256GB SSD; AMD Radeon F Dell;Inspiron 3567;Notebook;"15.6""";Full HD 1920x1080;Intel Core i7 7500U 2.7GHz;8GB;256GB SSD;AMD Radeon R5 M430;Windo Apple; MacBook Pro; Ultrabook; "15.4"""; IPS Panel Retina Display 2880x1800; Intel Core i7 2.96Hz; 16GB; 512GB SSD; AMD Radeon F Lenovo; IdeaPad 320-151KB; Notebook; "15.6"""; Full HD 1920x1080; Intel Core i3 7100U 2.4GHz; 8GB; 1TB HDD; Nvidia GeForce 940MX Dell;XPS 13;Ultrabook;"13.3""";IPS Panel Full HD / Touchscreen 1920x1080;Intel Core i5 8250U 1.6GHz;8GB;128GB SSD;Intel Asus; Vivobook E200HA; Netbook; "11.6""; 1366x768; Intel Atom x5-Z8350 1.44GHz; 2GB; 32GB Flash Storage; Intel HD Graphics 400; Lenovo:Legion Y520-15IKBN;Gaming;"15.6""";IPS Panel Full HD 1920x1080;Intel Core i5 7300HD 2.5GHz;8GB;128GB SSD + 1TB HD

Scraping Amazon:



```
('SGIN Laptop 15.6 Inch 12GB DDR4 512GB SSD, Windows 11 Laptops with Intel Celeron N4500, FHD 1920x1080, Dual Band WiFi, 2xUSB 3.0, Up to 2.8Ghz, Bluetooth 4.2, Supports 512GB TF Card Expansion, Gray', '$419.99', '', '', 'https://www.amazon.com/gp/slredirect/picassoRedirect.html/ref=pa_sp_atf_aps_sr_pgl_l?ie=UTF8&adId=A04805432ZB92100YNL11&url=%2F5GIN-1920x10 80-Bluetooth-Supports-Expansion%Zefp%ZeF809XQZMS64%Zeref%3Dsr_l_1_sspa%3Fkeywords%3Dlaptop%26qid%3D1654334600%26sr%3D8-1-spons%26psc%3Dl&qualifier=1654334600&id=140499896444953&widgetName=sp_atf')
('Acer Nitro 5 AN515-57-79TD Gaming Laptop | Intel Core i7-11800H | NVIDIA Geforce RTX 3050 Ti Laptop GPU | 15.6" FHD 14 4Hz IPS Display | 8GB DDR4 | 512GB NVMe SSD | Killer Wi-Fi 6 | Backlit Keyboard', '$998.82', '', '', 'https://www.amazon.com/gp/slredirect/picassoRedirect.html/ref=pa_sp_atf_aps_sr_pg1_l?ie=UTF8&adId=A05951653GBLL6ZRENUL4&url=%2FAcer-AN515-57-79TD-i7-11800H-GeForce-Keyboard%2Fdp%2FB09R65RN43%2Fref%3Dsr_1_2_sspa%3Fkeywords%3Dlaptop%26qid%3D1654334600%26sr%3D8-2-spons%26psc%3Dl&qualifier=1654334600&id=1404998964440953&widgetName=sp_atf')
('ASUS ProArt StudioBook Pro 16 OLED Laptop, 16" 3840x2400 OLED Display, Intel Xeon W-11955M, 64GB DDR4, 4TB PCIe SSD, N vidia RTX A5000, Windows 11 Pro, W7600H5A-XH99, Star Black', '$4,999.99', '', 'https://www.amazon.com/ASUS-StudioBook-3840x2400-W-11955M-W7600H5A-XH99/dp/B09NWFSDVN/ref=sr_1_3*keywords=laptop&qid=1654334600&sr=8-3')
('2022 Newest HP 15.6" HD Laptop Computer, Intel Celeron Quad-Core N4120(up to 2.6GHz), 8GB DDR4 RAM, 128GB SSD, HDMI, B luetooth, Webcam, USB-C, R345 Ethernet, Windows 11s, Silver, JVQ Mousepad', '$349.00', '', '', 'https://www.amazon.com/H P-Laptop-Computer-quad-Core-Bluetooth/dp/809YCSS5K4/ref=sr_1_4*keywords=laptop&qid=1654334600&sr=8-4')
('MSI Raider GE66 15.6" UHD 44 120Hz Gaming Laptop Intel Core i9-12900HK RTX3080TI 32GBDDR5 1TB NVMe SSD Win11', '$3,79 9.00', '', '', 'https://www.amazon.com/MSI-Raider-GE66-i9-12900HK-RTX3080TI/dp/809RB91BBT/ref=sr_1_5*keywords=laptop&qid=165
```

VI. Preprocessing des données:

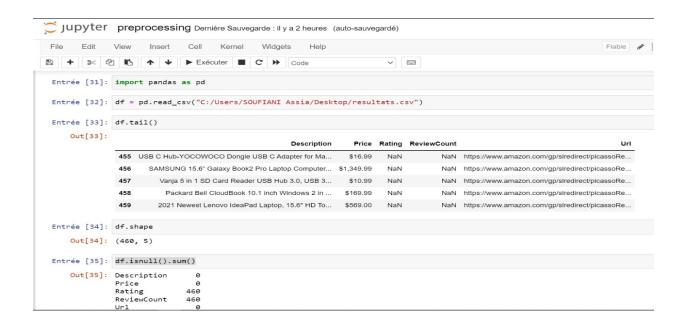


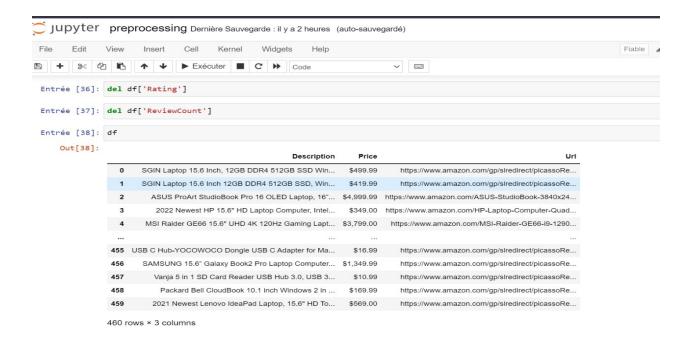
Dans notre cas nous avons fait:

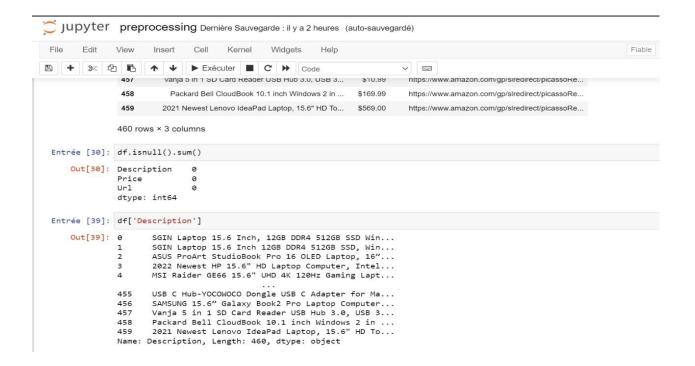
Nettoyage des données : trouvez les « Null », les valeurs manquantes et les données dupliquées. Il faut remplacer les « Null » et les valeurs manquantes par d'autres valeurs (ou les supprimer) .

Décomposition des données : les colonnes de texte contiennent parfois plus d'une information ; nous devons donc les diviser en autant de colonnes dédiées que nécessaire. Si certaines colonnes représentent des catégories, convertissez-les en colonnes de type catégorie.

Exemples:





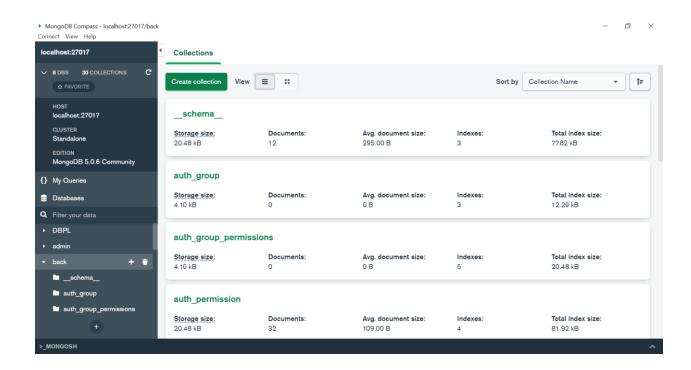


VII. Structure de notre base de données

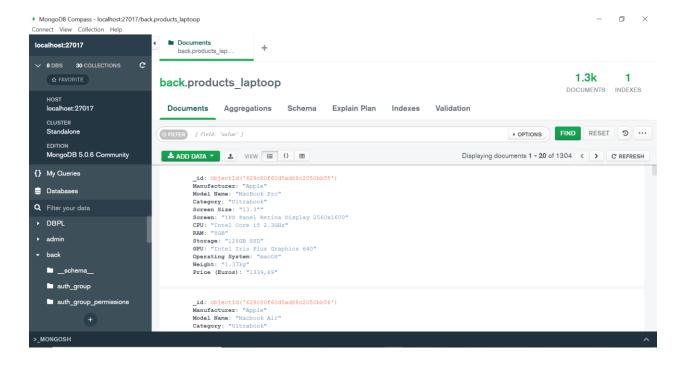
Mise en place de la migration :

Pour se faire, on exécute les commandes suivantes :

- python manage.py makemigrations
- > python manage.py migrate



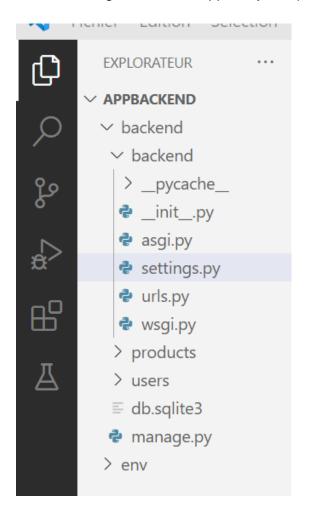
Voilà notre collection back qui contient les données scrapeés



VIII. Strcture de backend

- La première chose que nous avons fait, c'est de créer et activer un environnement virtuel avec les commandes virtualenv env et source env/script/activate.
- Ensuite on installe django avec pip install django et on crée un projet avec : django-admin startproject backend
- Et maintenant on crée notre application avec : python manage.py startapp product.

Et dans settings -> installed app on ajoute 'products'



```
INSTALLED_APPS = [
   'django.contrib.admin',
   'django.contrib.auth',
   'django.contrib.contenttypes',
   'django.contrib.sessions',
   'django.contrib.messages',
   'django.contrib.staticfiles',
   'products',
   'users',
   'graphene_django',
   'django_elasticsearch_dsl',
]
```

IX. GraphQL

Installation de graphql :

On met tous les 'requirements' qu'on aura besoin dans un

fichier.txt et on execute la commande :

Pip install -r requirements.txt

• Maintenant on crée un schéma :

```
settings.py
            managers.py
fields = ('id','title')
 11
 12
 13 v class LaptoopType(DjangoObjectType):
        class Meta:
| model = Laptoop
 14 ∨
 15
            fields = (
 17
               'id',
               'Manufacturer',
 18
               'Model_Name',
'Category',
 19
 20
               'Screen_Size',
 22
               'Screen',
               'CPU',
 23
 25
               'Storage',
               'GPU',
 26
               'Operating_System',
 28
               'Weight',
 29
 30
 31
           )
 33
 35 \vee class Query(graphene.ObjectType):
 36
        categories = graphene.List(CategoryType)
        laptops = graphene.List(LaptoopType)
 38
```

Et voici le modèle :

```
      ◆ schema.py ...\products
      ◆ admin.py
      ◆ models.py ...\products
      X
      ◆ managers.py
      ◆ mc D × □ ···

                schema.py ...\users
backend > products > @ models.py > 😭 Laptoop
                                                                                                                                     from django.db import models
       # Create your models here.
       class Category(models.Model):
           title = models.CharField(max_length=255)
  8
           class Meta:
 10
               verbose_name_plural = 'Categories'
           def __str__(self):
 11
 12
               return self.title
 13
       class Laptoop(models.Model):
           Manufacturer = models.CharField(max_length=300)
 14
           Model_Name = models.CharField(max_length=50)
           Category = models.CharField(max_length=150)
           Screen_Size = models.CharField(max_length=150)
 17
           Screen = models.CharField(max_length=100)
 18
           CPU = models.CharField(max_length=100)
 20
           RAM = models.CharField(max_length=100)
 21
           Storage = models.CharField(max_length=100)
           GPU = models.CharField(max_length=100)
 22
           Operating_System = models.CharField(max_length=100)
 24
           Weight = models.CharField(max_length=100)
           Price = models.CharField(max_length=100)
 25
 26
 27
           def __str__(self):
               return self.Manufacturer
 29
 30
```

```
• Setup d'URL de GraphQL:
    On navigue vers:
    src->products -> urls.py

from django.contrib import admin
from django.urls import path
from graphene_django.views import GraphQLView
from products.schema import schema

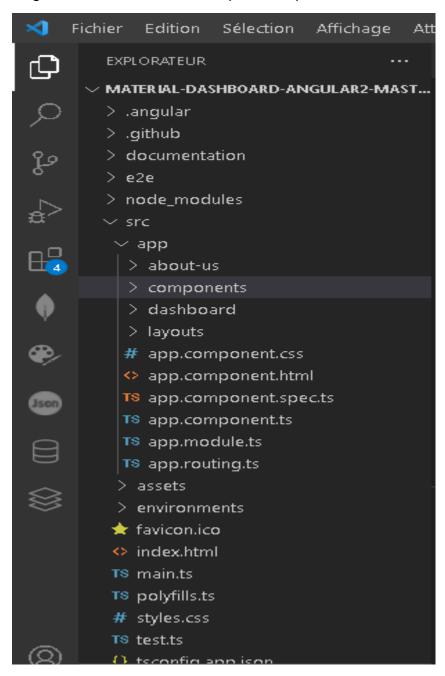
urlpatterns = [
    path('admin/', admin.site.urls),
    path("graphql", GraphQLView.as_view(graphiql=True, schema=schema)),
]
```

Et on lance http://localhost:8000/graphql:

X. Structure de Frontend:

Pour le frontend nous avons travaille avec angular 13

La premiere chose a faire c'est d'installer NodeJS et AngularCli, puis de créer un projet Angular, et la creation des components qu'on a besoin.



XI. Visualisation des données:

