**As mentioned in the question .The task was to crawl the website and perform analysis on the data obtained**.

**ASSIGNMENT OF DATA CRAWLING AND DATA ANALYSIS**

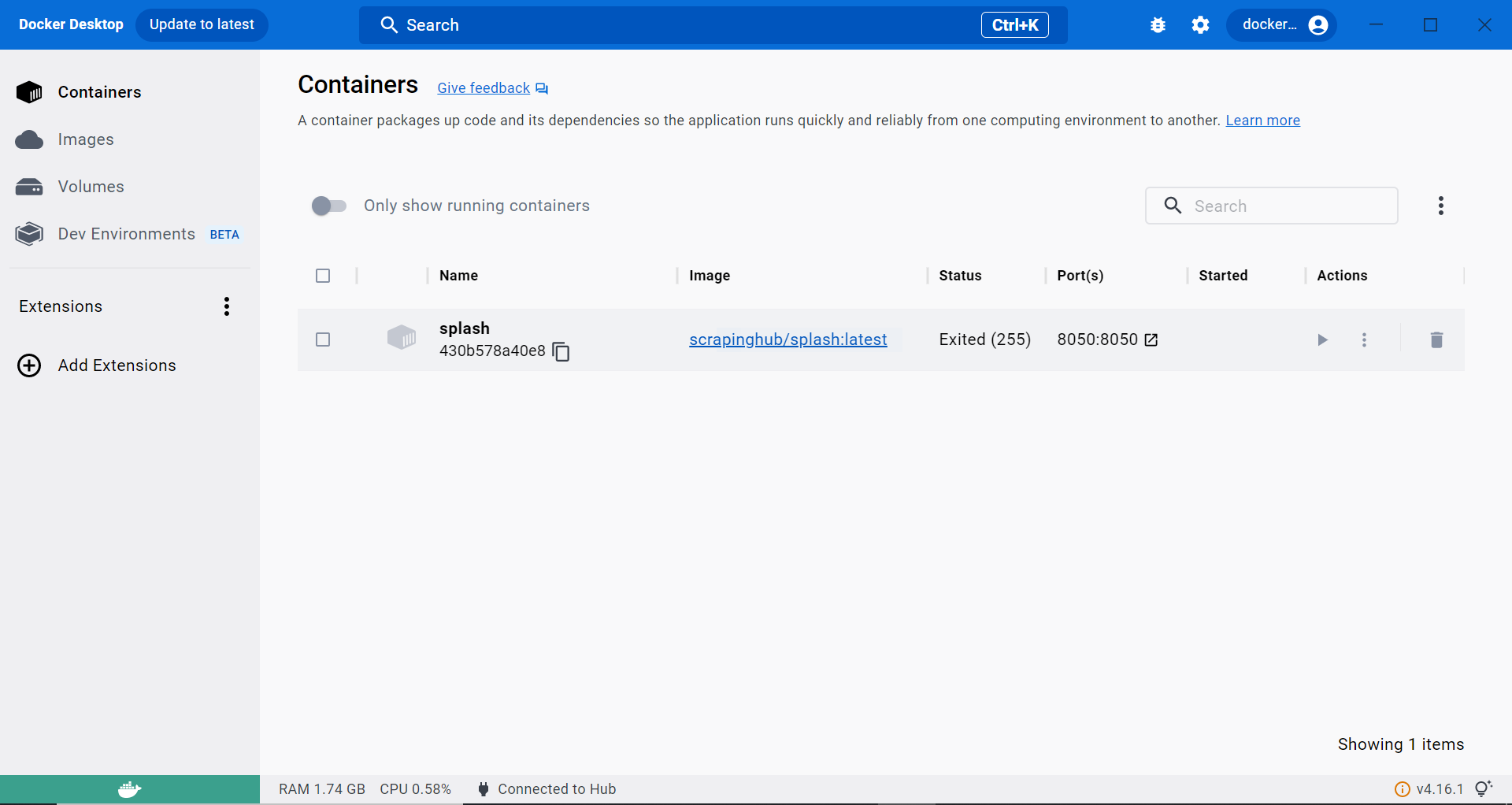
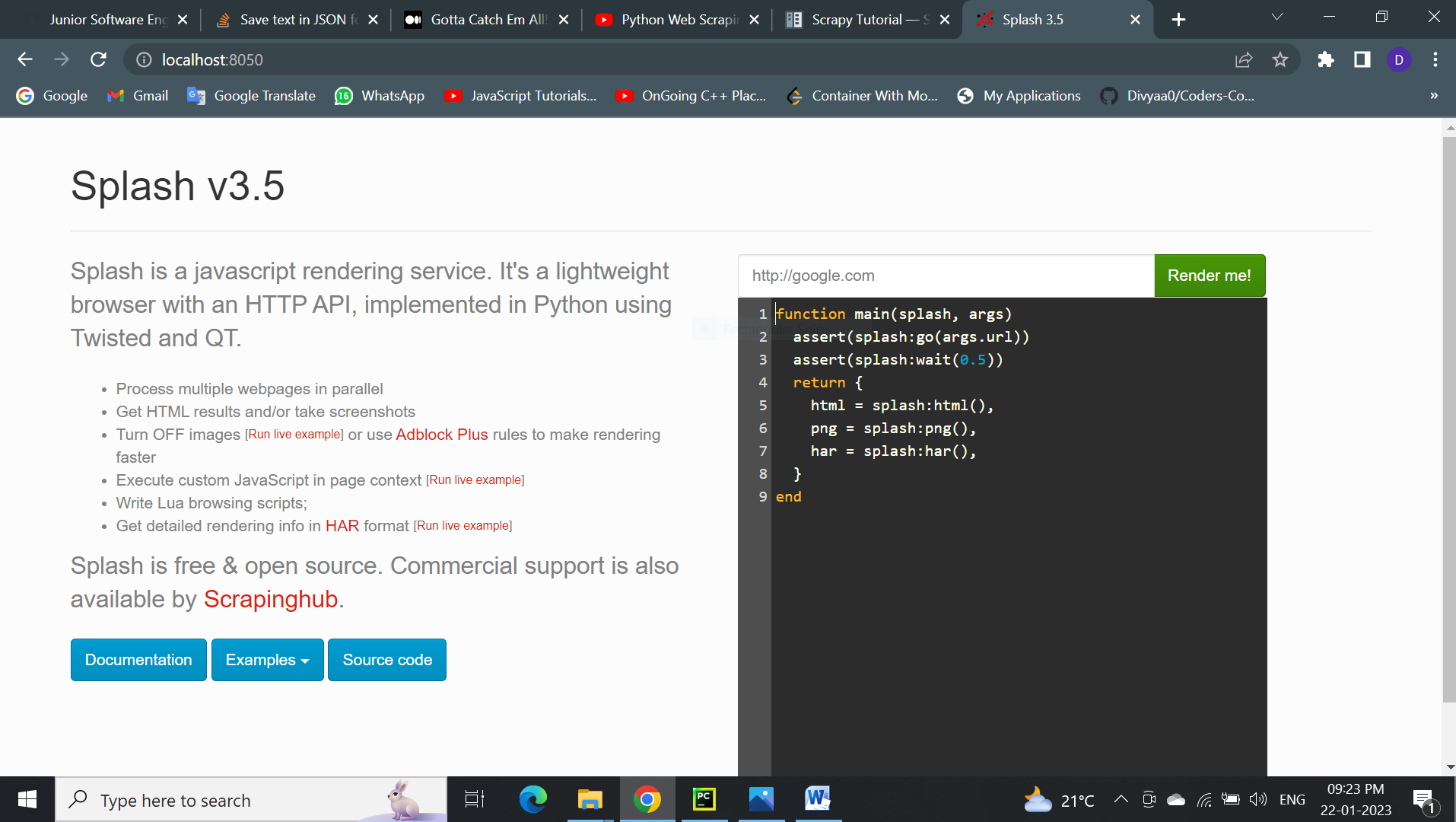
As an **absolute Beginner** of web scrapping , I tried to learn every concept from scratch .How web scraping is performed and what is the need of web scraping.

We were asked to perform web scraping with the the help of **SCRAPY FRAMEWORK .** I started with learning the basics of web scraping using scrapy i.e

about Project Structure (spider,items.py,settings.py,pipelines.py,middlewares.py),Extracting data with CSS SELECTORS and XPATH, about the pipelines .

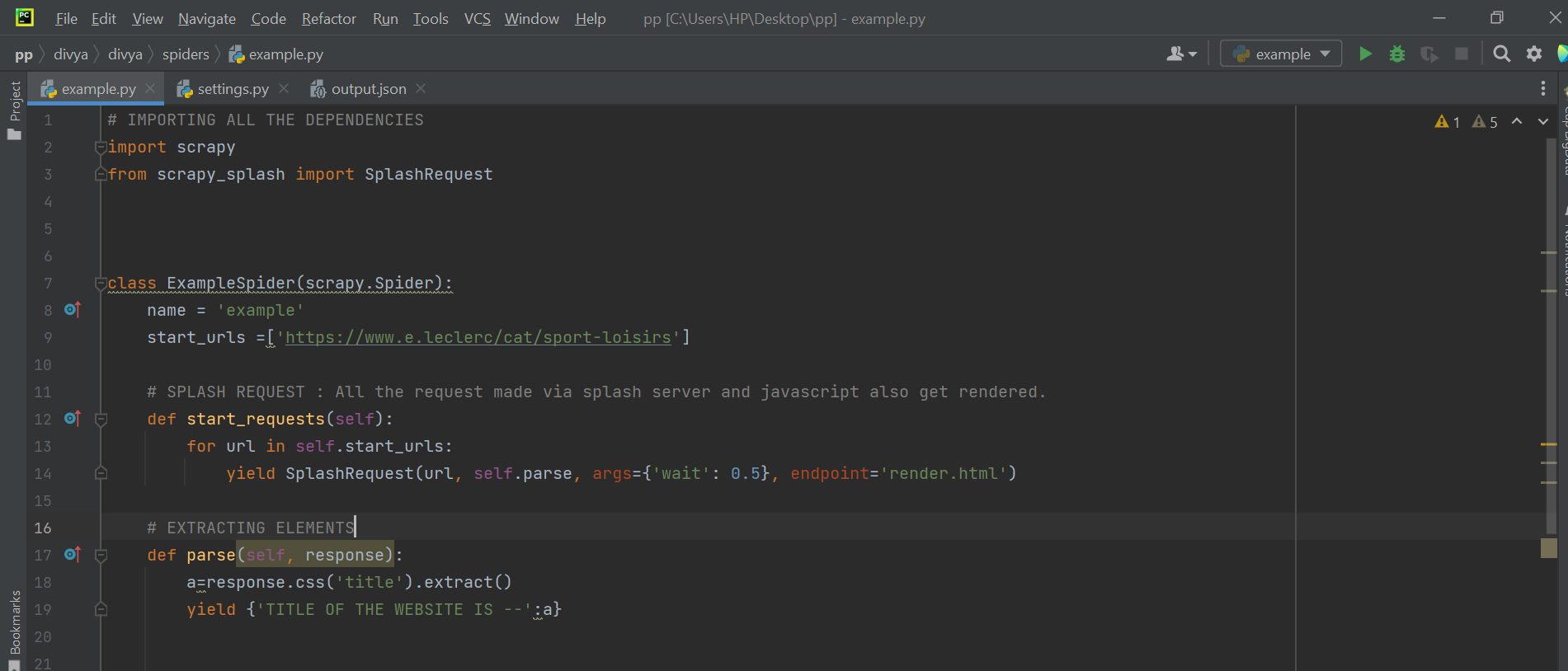
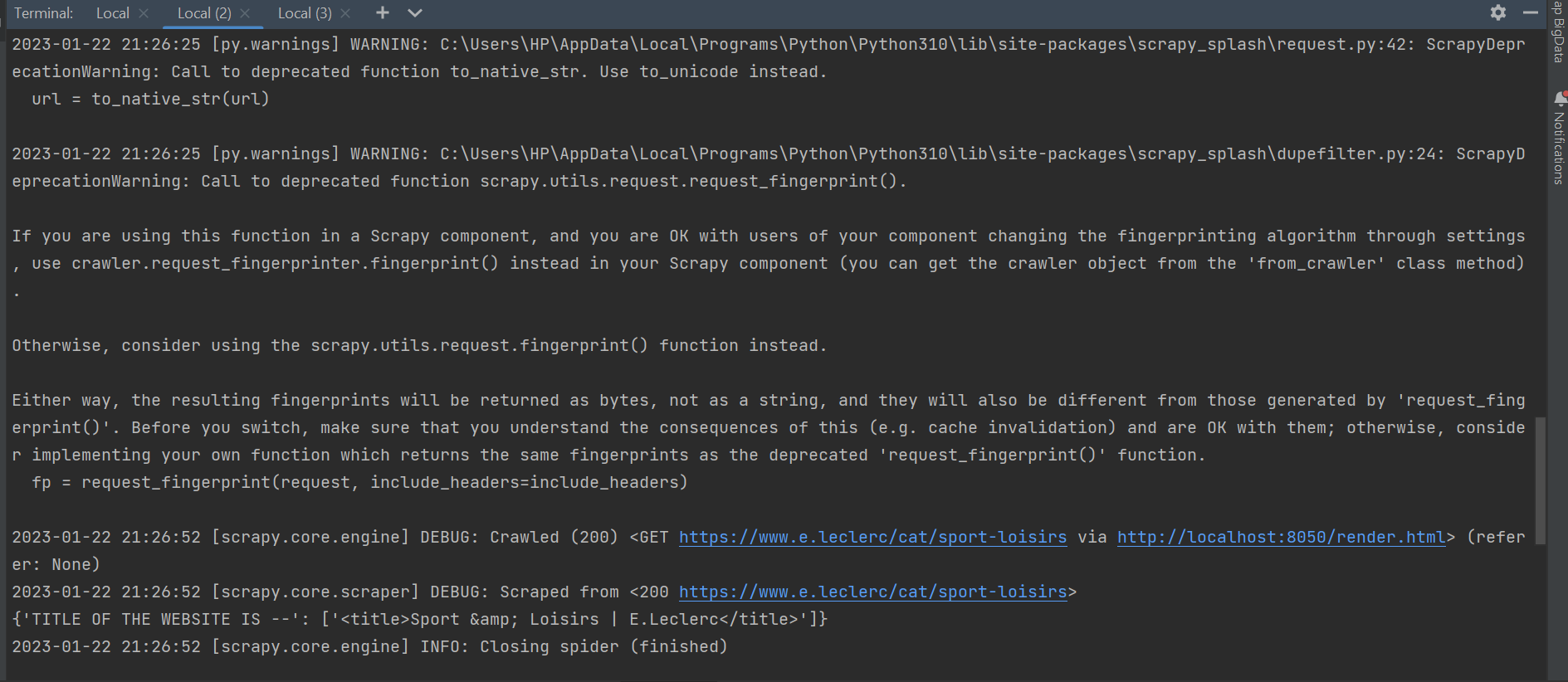
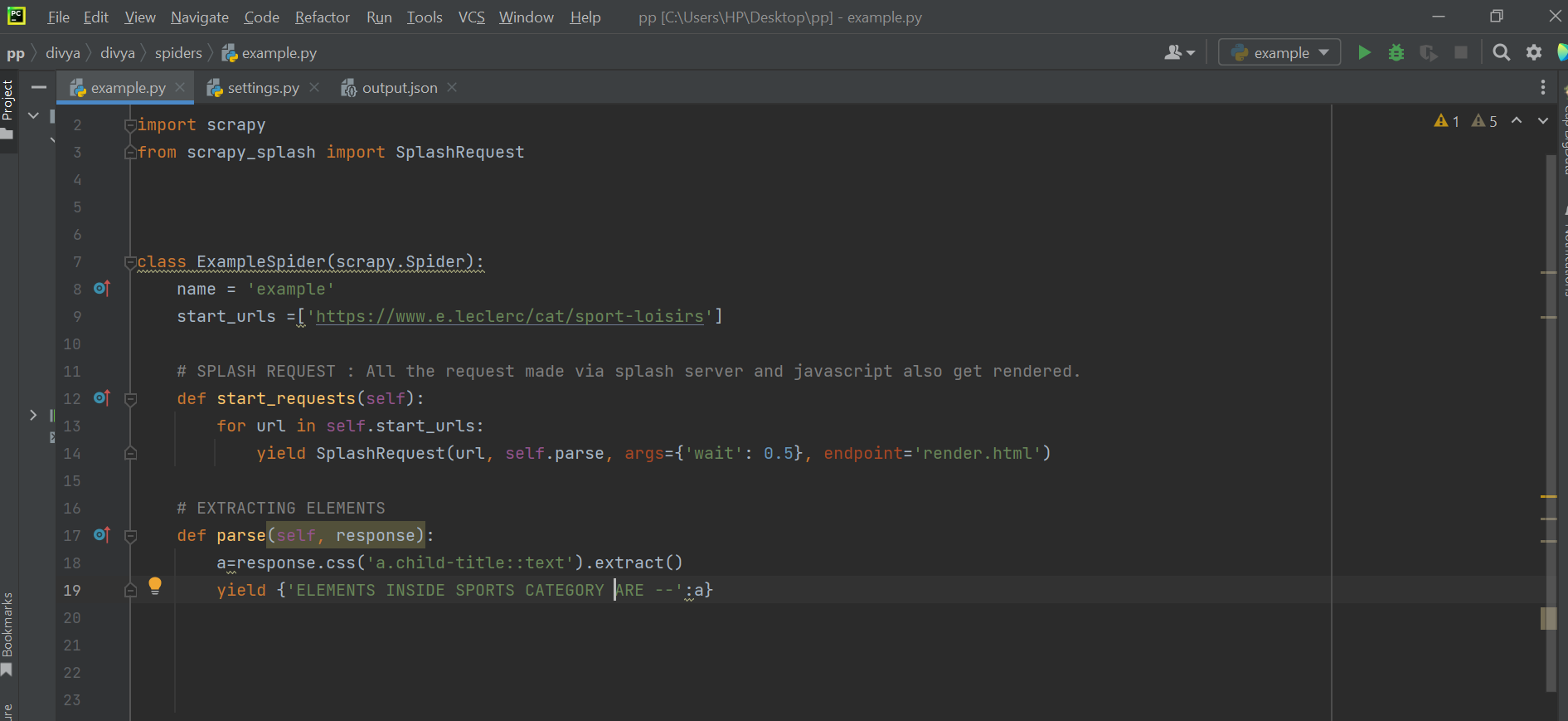
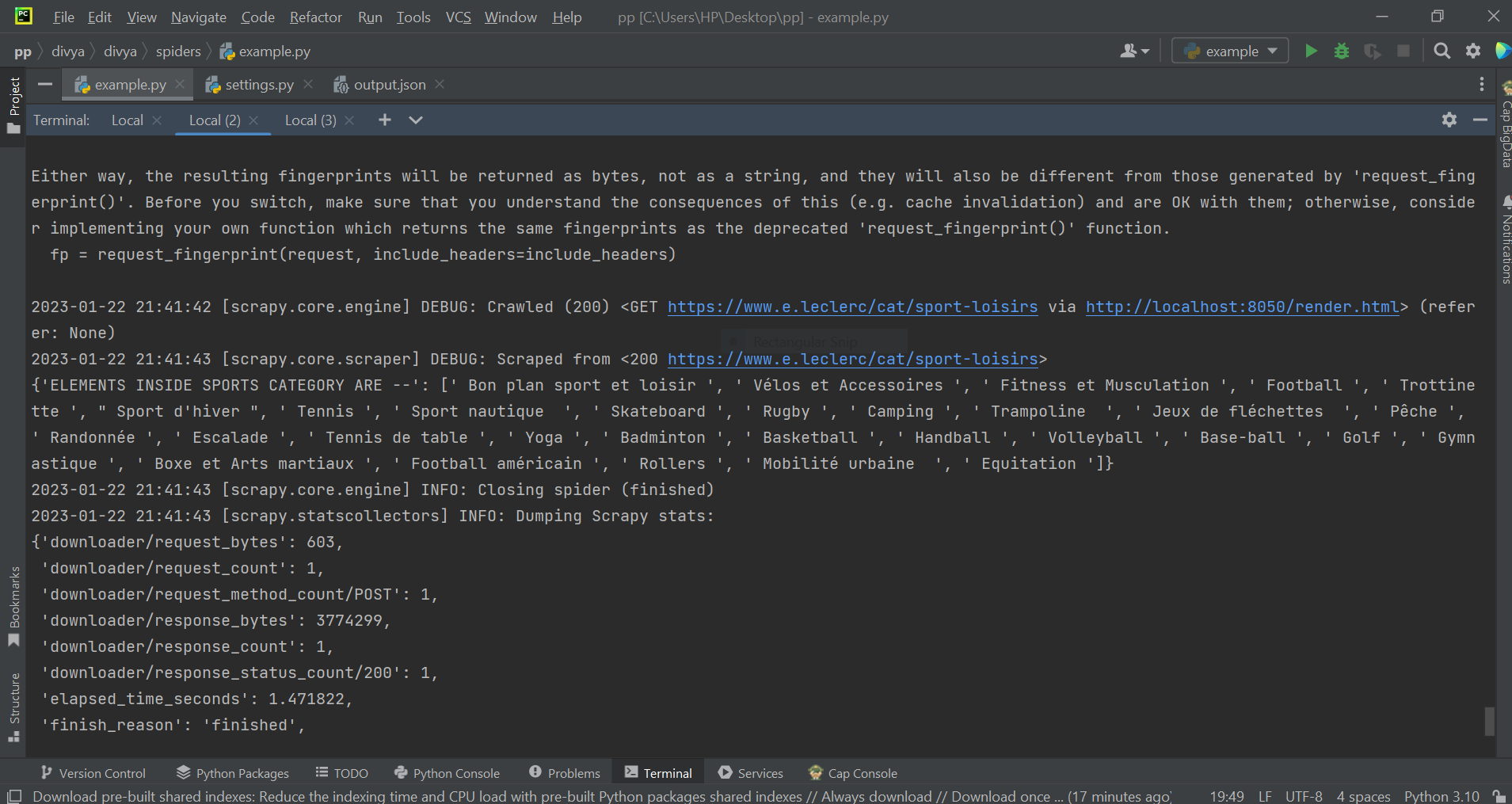
But then got to know that Scrapy alone can only be used with static website. For this **Assignment WEBSITE** which is completely dynamic website that makes the use of Javascript . It is required to use Scrapy –Splash .

For The Scrapy Splash , I installed Docker in my PC.

* Pulled The image of scrapy splash.
* 
* Splash was successfully running on localhost.
* 

SCRAPY FOLDER STRUCTURE :

**PP >> divya(folder project name) >> example (spider)**

* Implementing title extraction
* 
* **OUTPUT**
* 
* **ACCESSING THE ITEMS UNDER SPORTS CATEGORY**
* 
* 

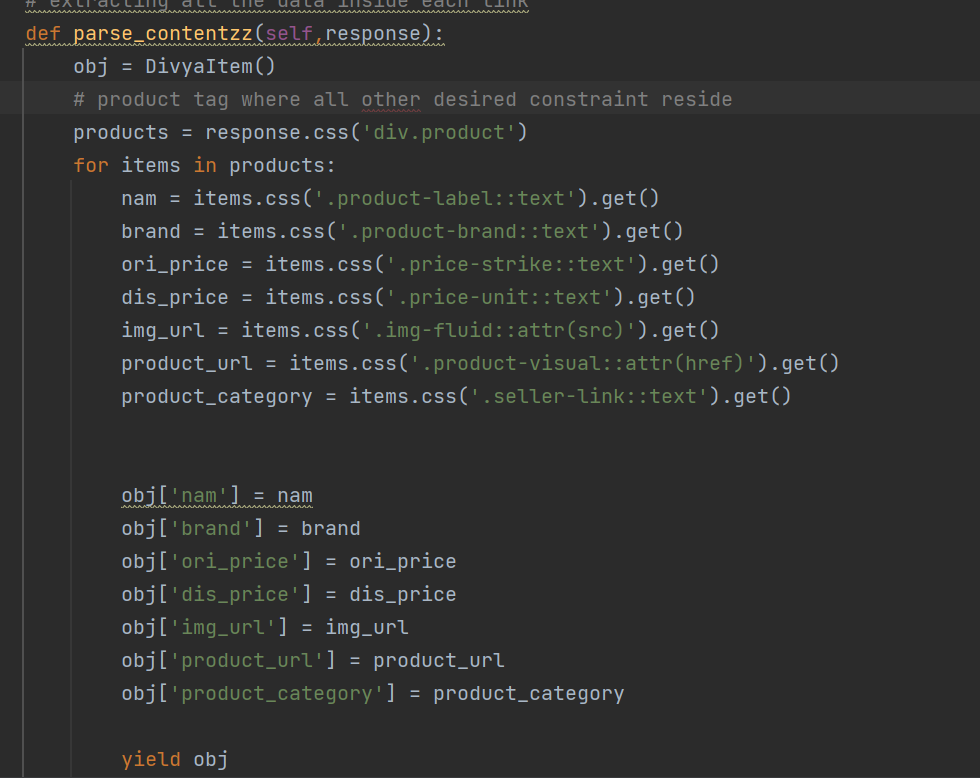
**APPROACH OF SCRAPING:-**

1. Firstly , I extracted all the links of sports category items
2. Started a loop on all the links

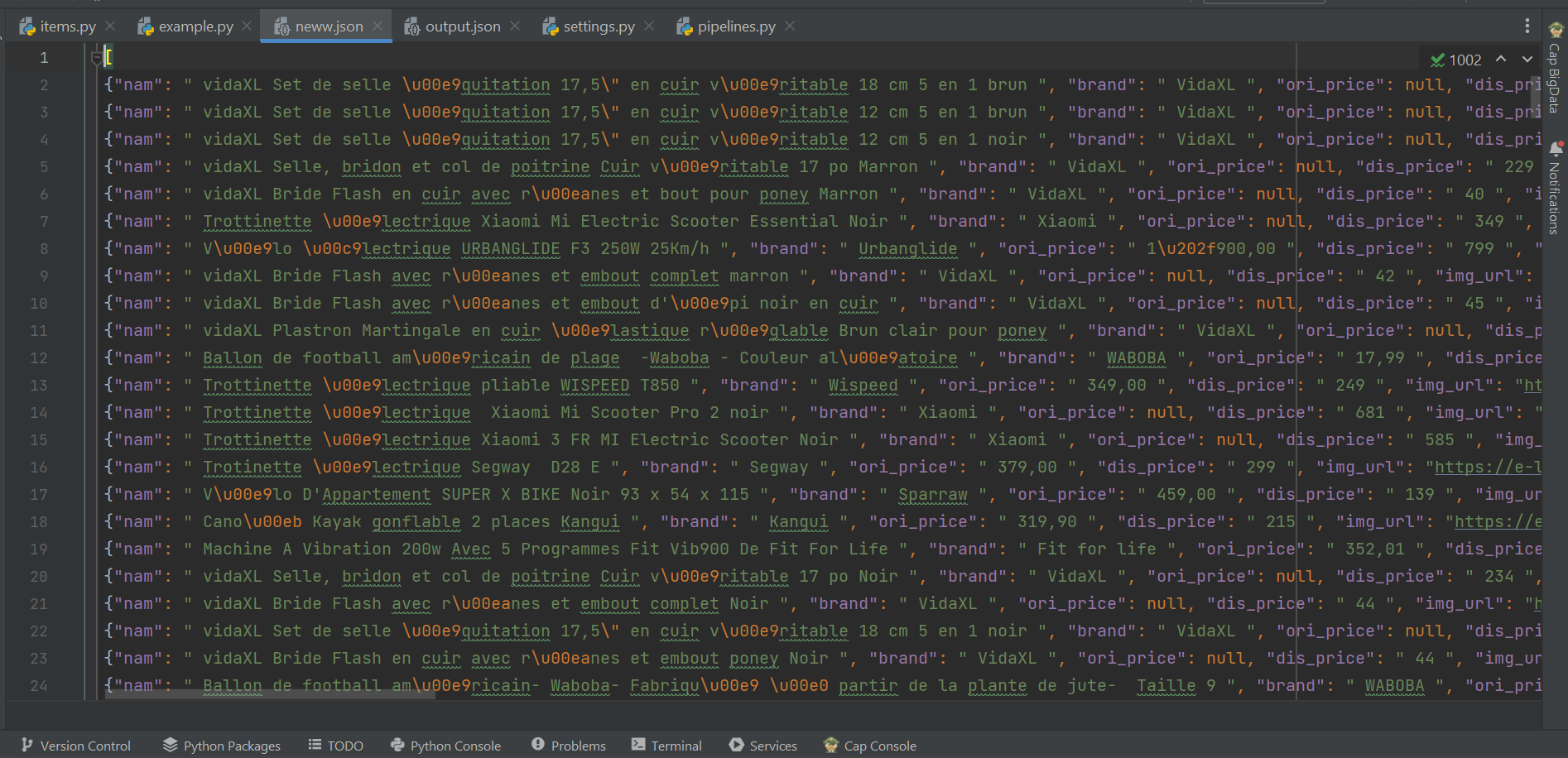
* Inside loop >> link will be opened
* Call a function (parse\_contentzz) where all the products items will be extracted one after another by using CSS selector of class name
* After successfully parsing page of every sport category item . Same process will be repeated forJwellery page

**SPIDER –example.py**

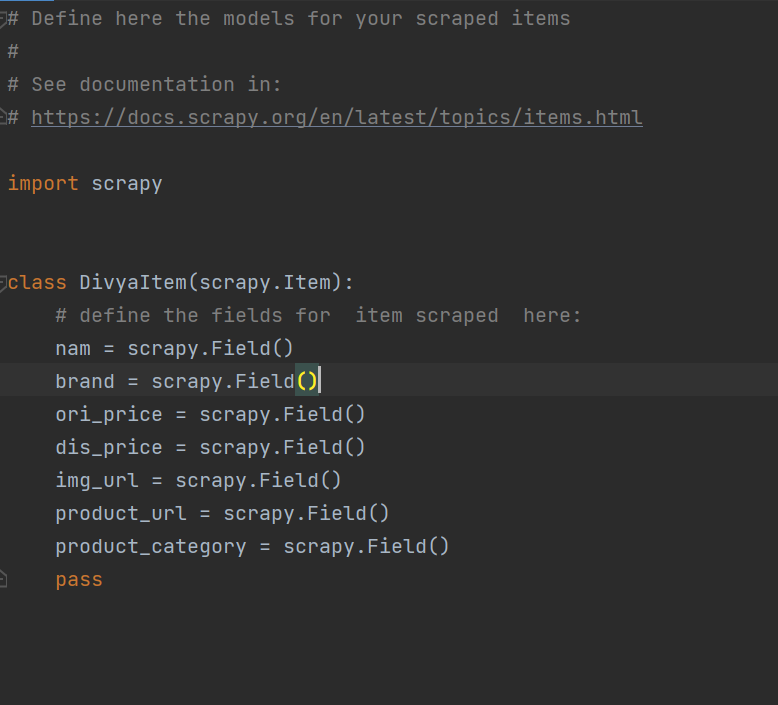




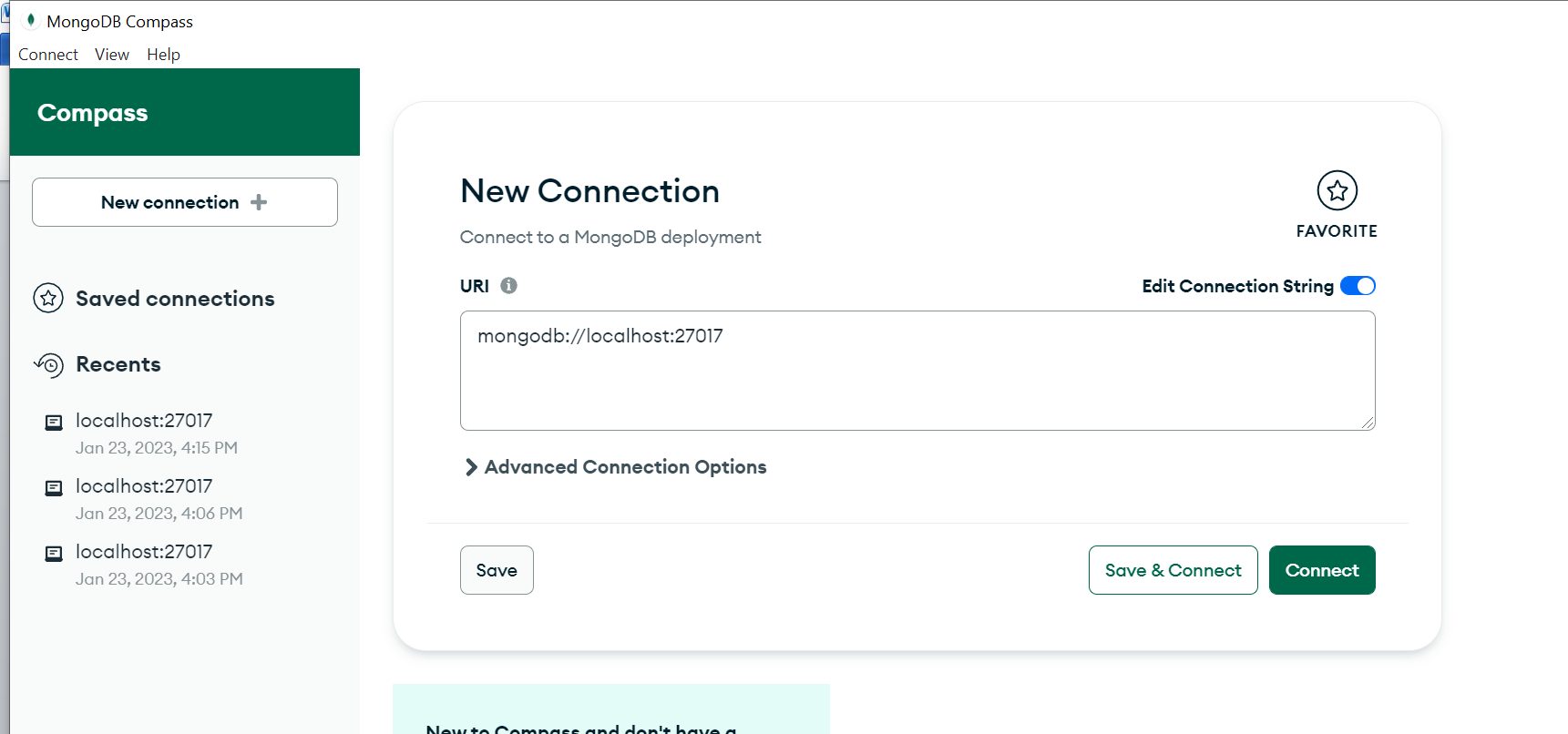
**Then,Stored all the data in a json file named as neww.json**



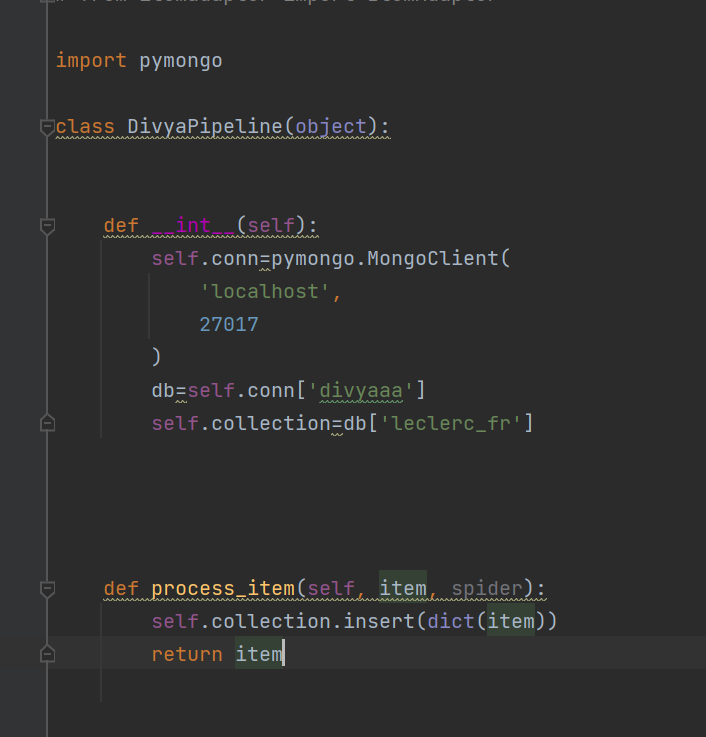
**After That , made changes in items.py file i.e created the container for the data being extracted in spider .**



**Next is Storing The Data in container , To insert in database with mongoDB with the help of pymongo**

* **Installed mongoDB in PC**
* 

**PIPELINE.PY FILE**



**BUT IT IS GIVING ME ERROR {ATTRIBUTE ERROR}**



**I HAVE TRIED ALL POSSIBLE ASPECTS TO SOLVE THIS ISSUE BUT COULD NOT SUCCEED DUE TO LACK OF TIME BUT I WILL FOR SURE TRY TO LEARN FROM THIS ERROR AND SOLVE IT .**

**I HAVE WRITTEN THE CODE FOR THE CONNECTION OF DATABASE AND COLLECTION BUT ITS SHOWING ERROR (ATTRIBUTE ERROR) BUT I WILL RESOLVE THIS ERROR BY STRENGTHENING**

**MY KNOWLEDGE ABOUT EVERY CONCEPT IMPLEMENTED .**