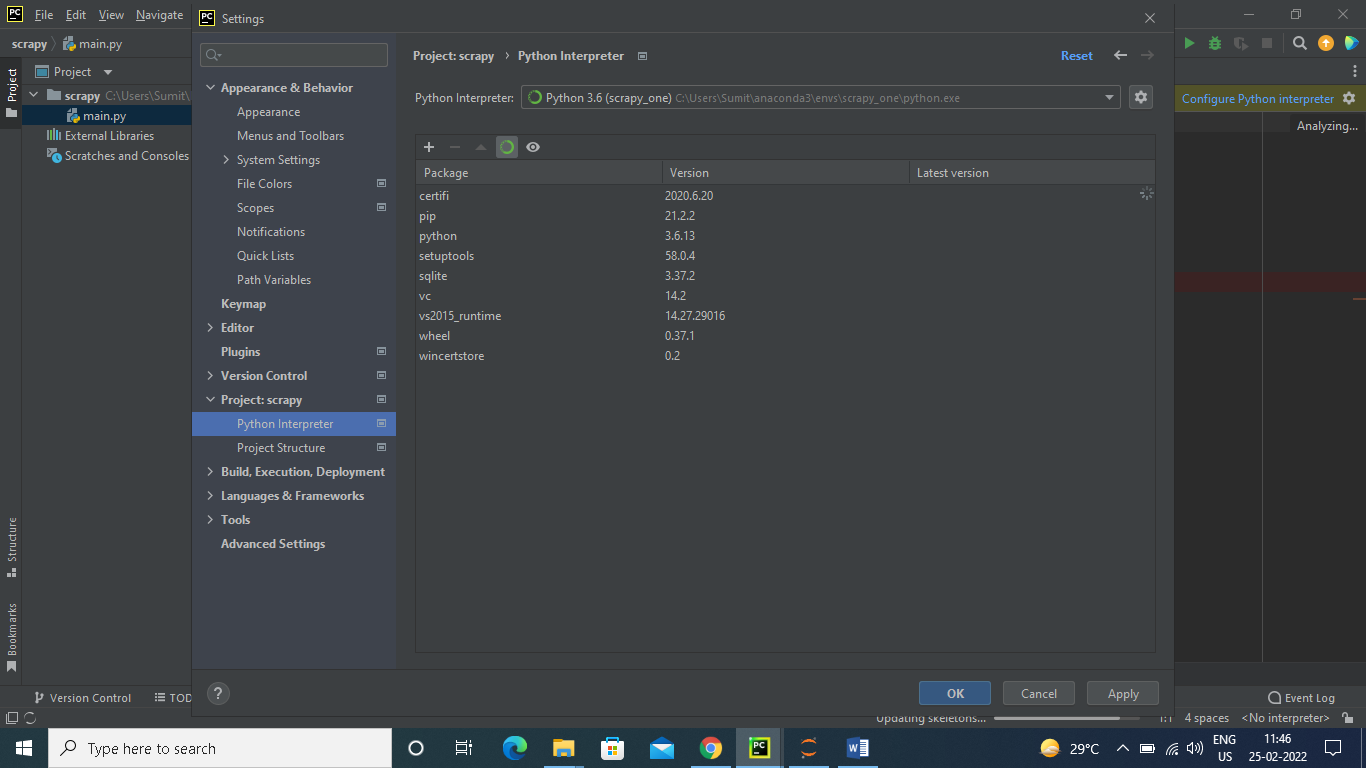
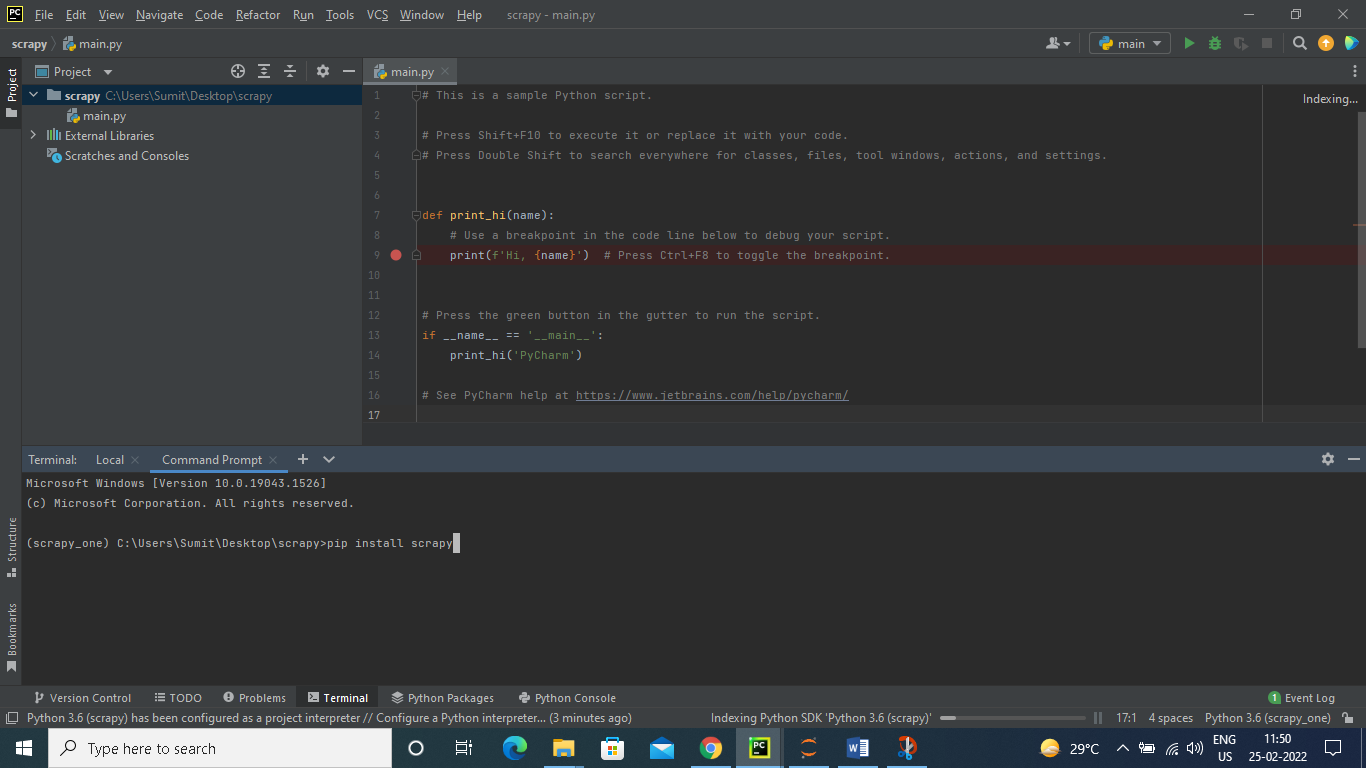
**Software Engineer Assignment Details**

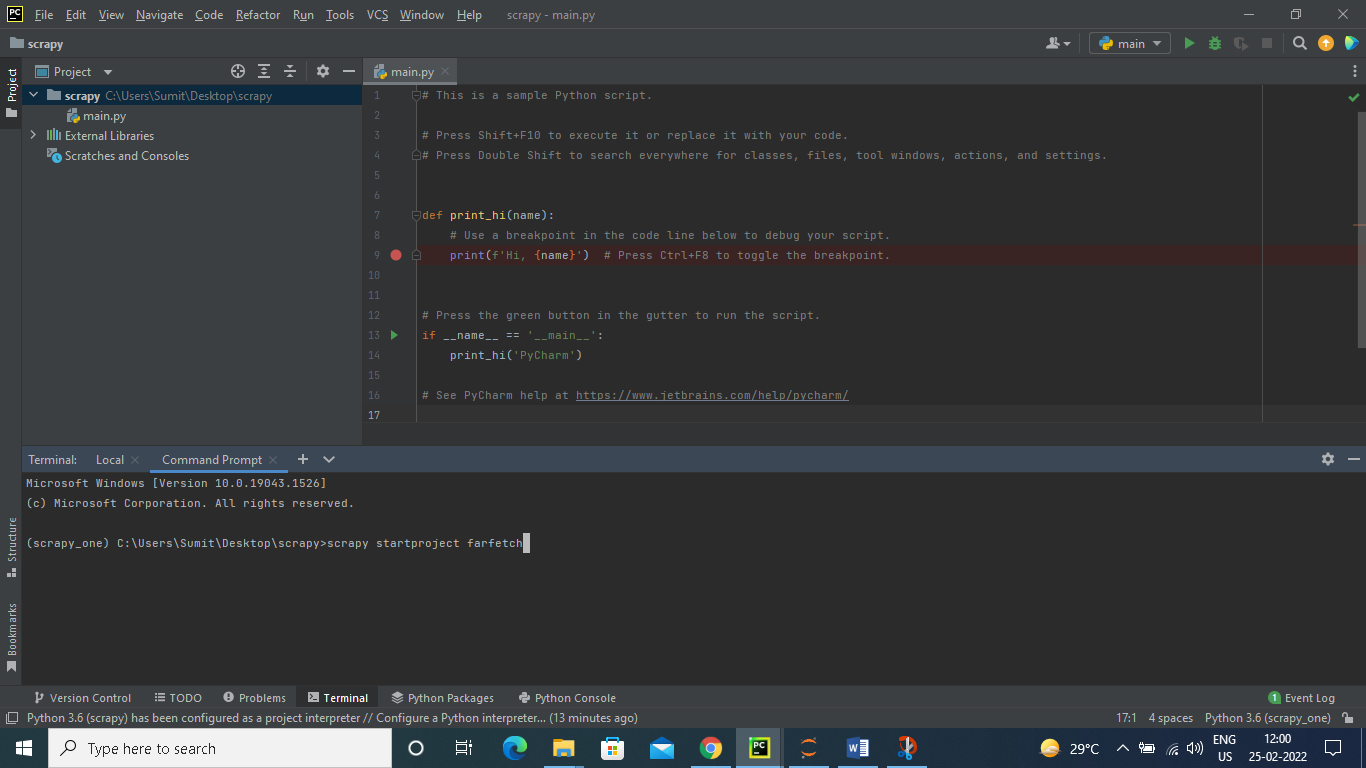
I am using pycharm for this project assignment.

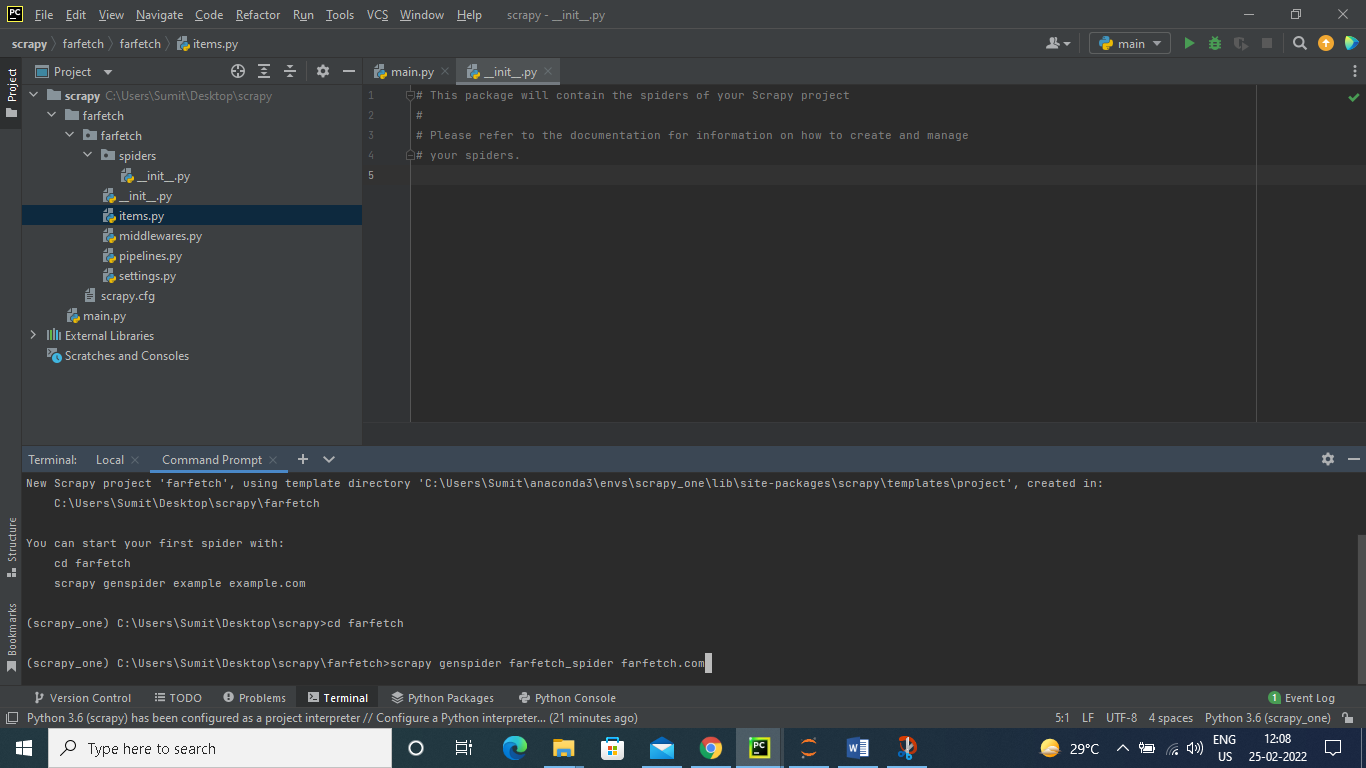
Step 1 : creating new environment for this project – scrapy\_one



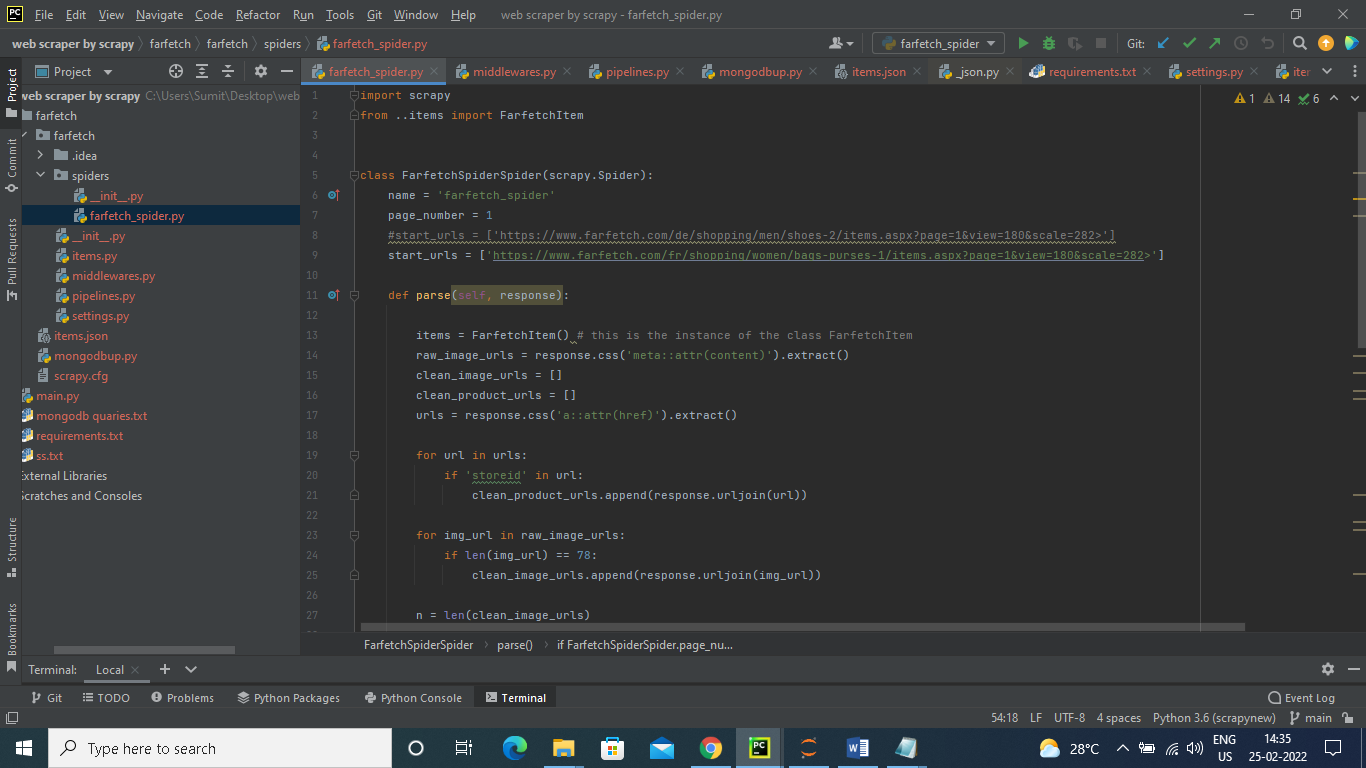
Step 2: install scrapy in this environment



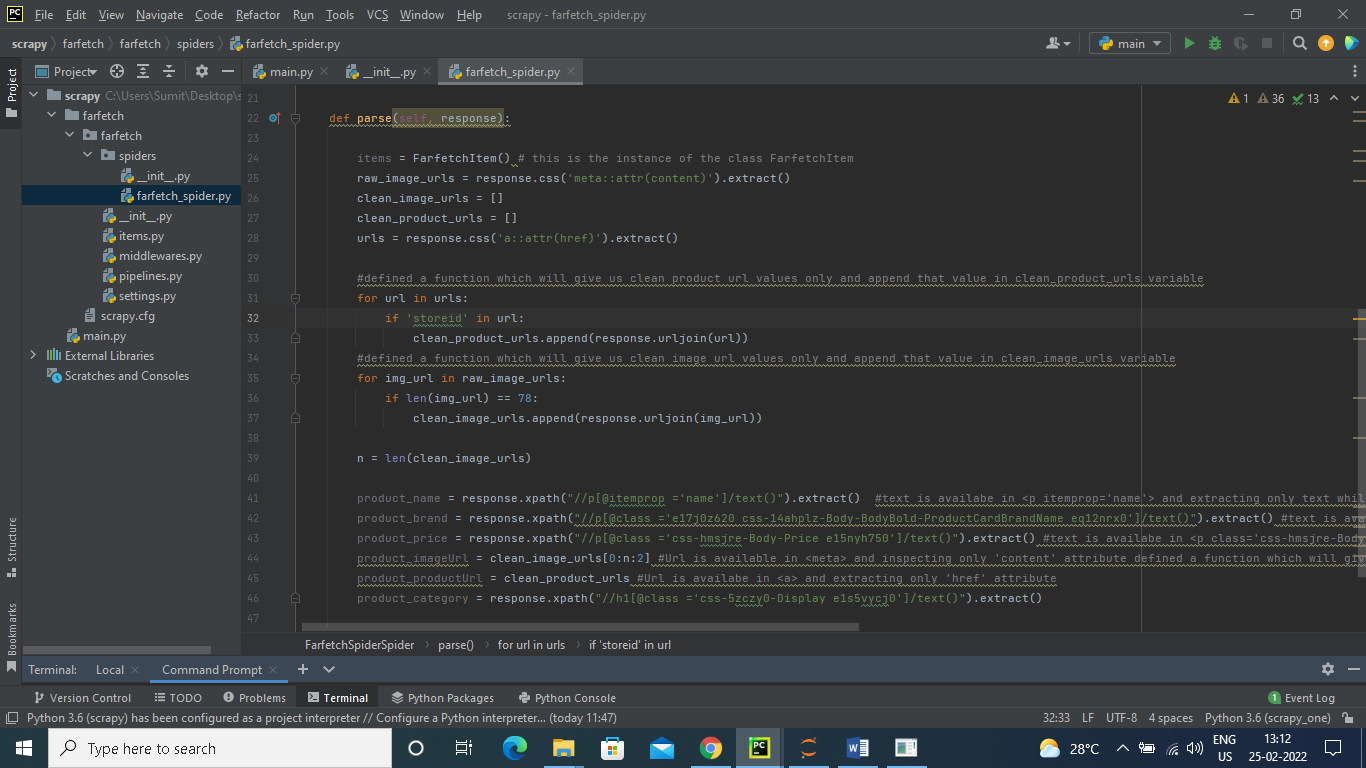
Step 3: start creating new project 

Step 4 : create new spider farfetch\_spider

Step 5: there are two items to extract 1) shoes 2) Bags and Purses, so to select these two urls we have defined a function which will select urls one by one



Step 6: setup parse function: scraping data from web site



Product\_name = text is available in <p itemprop='name'> and extracting only text while inspecting

Product\_brand = text is available in <p class='e17j0z620 css-14ahplz-Body-BodyBold-ProductCardBrandName eq12nrx0'> and extracting only text while inspecting

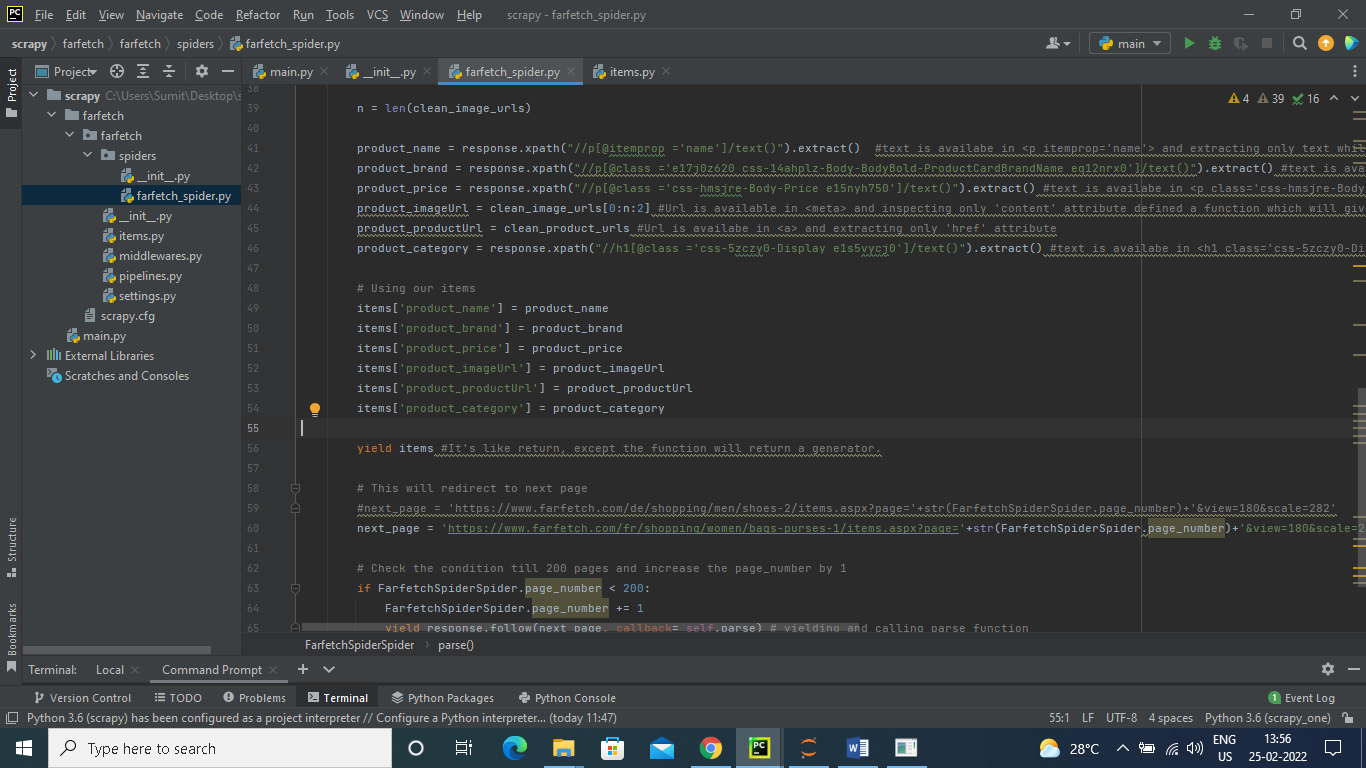
Product\_price = text is available in <p class='css-hmsjre-Body-Price e15nyh750'> and extracting only text while inspecting

Product\_imageurl = Url is available in <meta> and inspecting only 'content' attribute used for loop which will give us clean product url values only and append that value in clean\_product\_urls variable

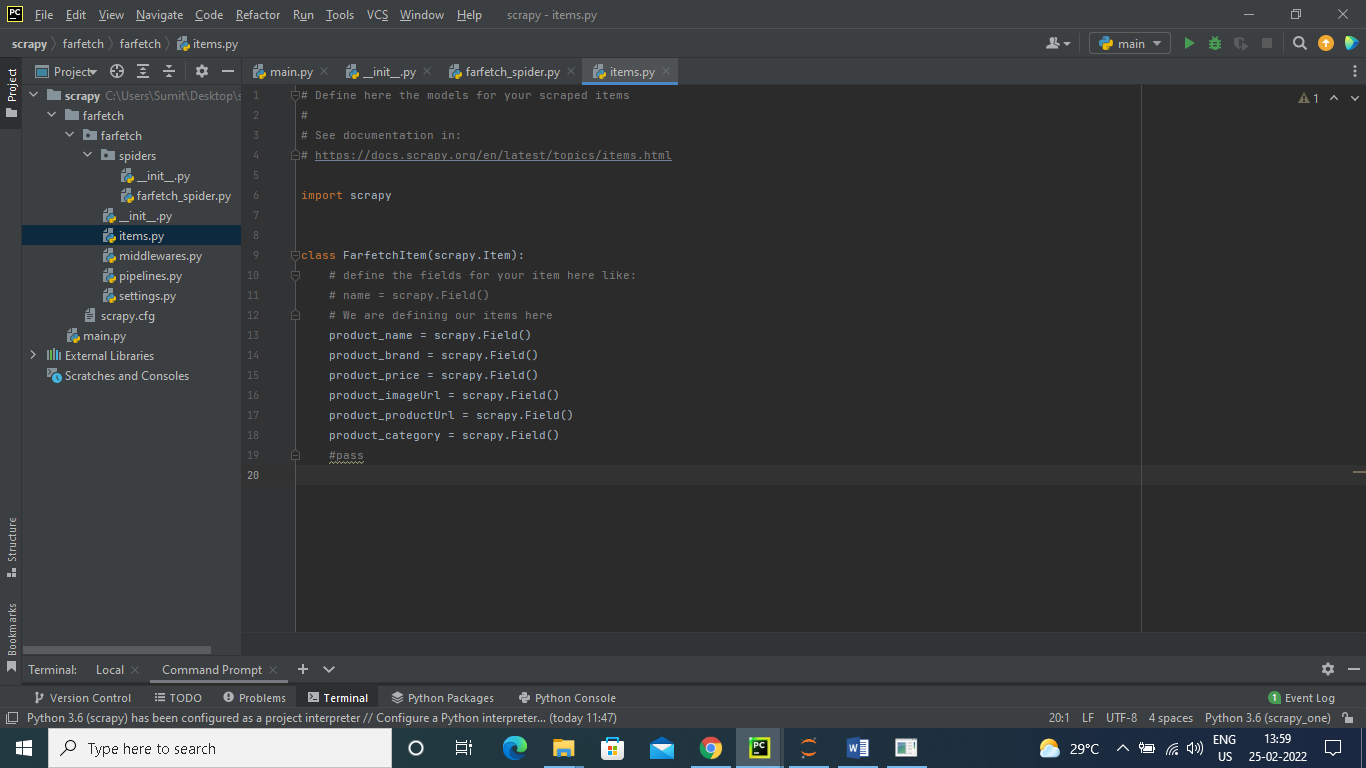
Product\_producturl = Url is available in <a> and extracting only 'href' attribute used for loop which will give us clean product url values only and append that value in clean\_product\_urls variable

Product\_category = text is available in <h1 class='css-5zczy0-Display e1s5vycj0'> and extracting only text while inspecting

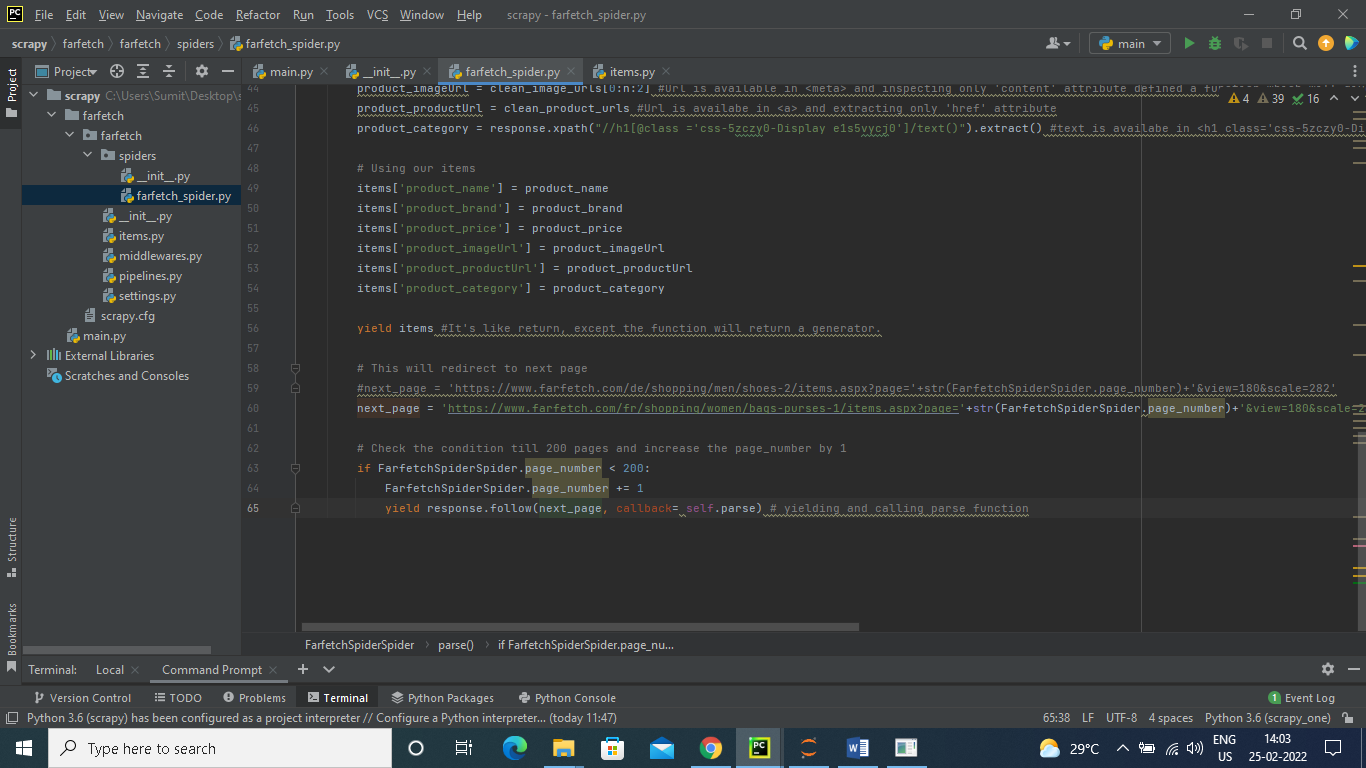
Step 7: saving all scraped data in items



Step 8: setting up items.py file by defining our item

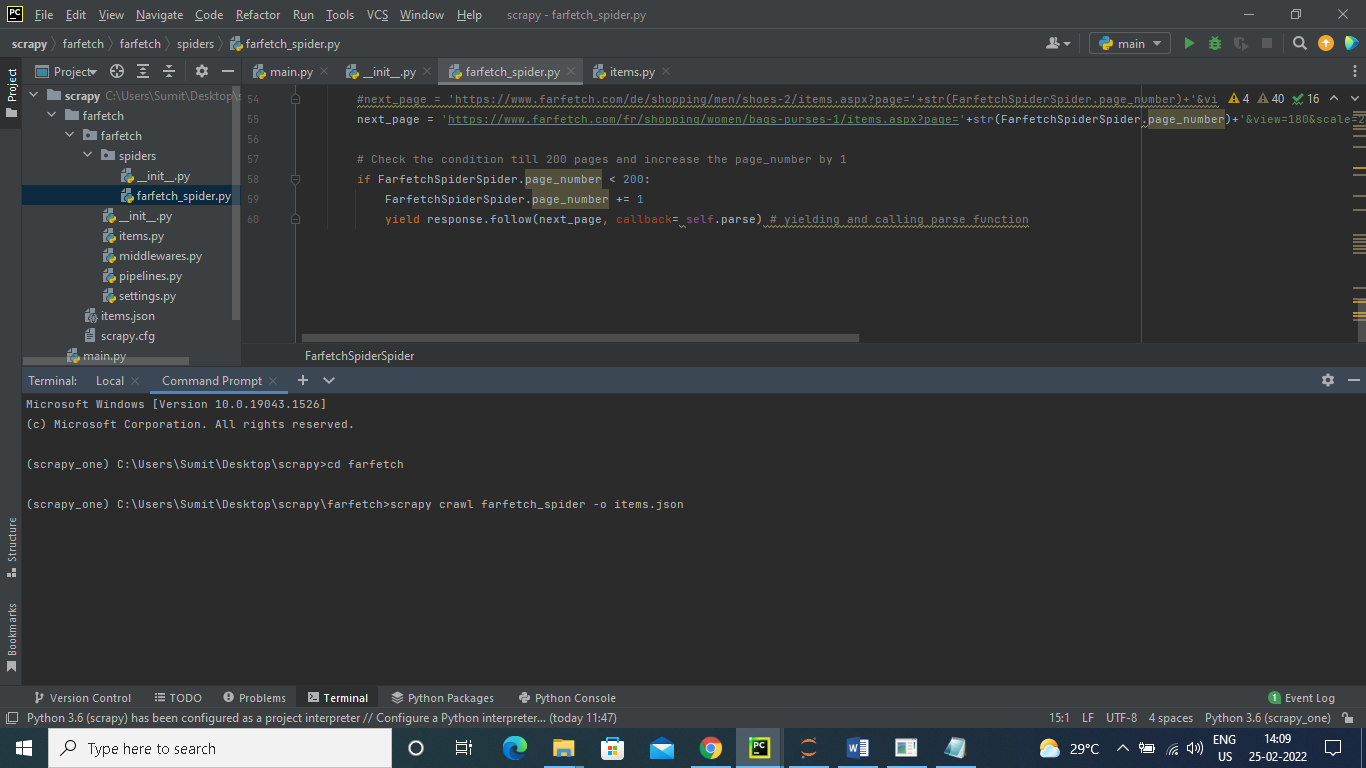


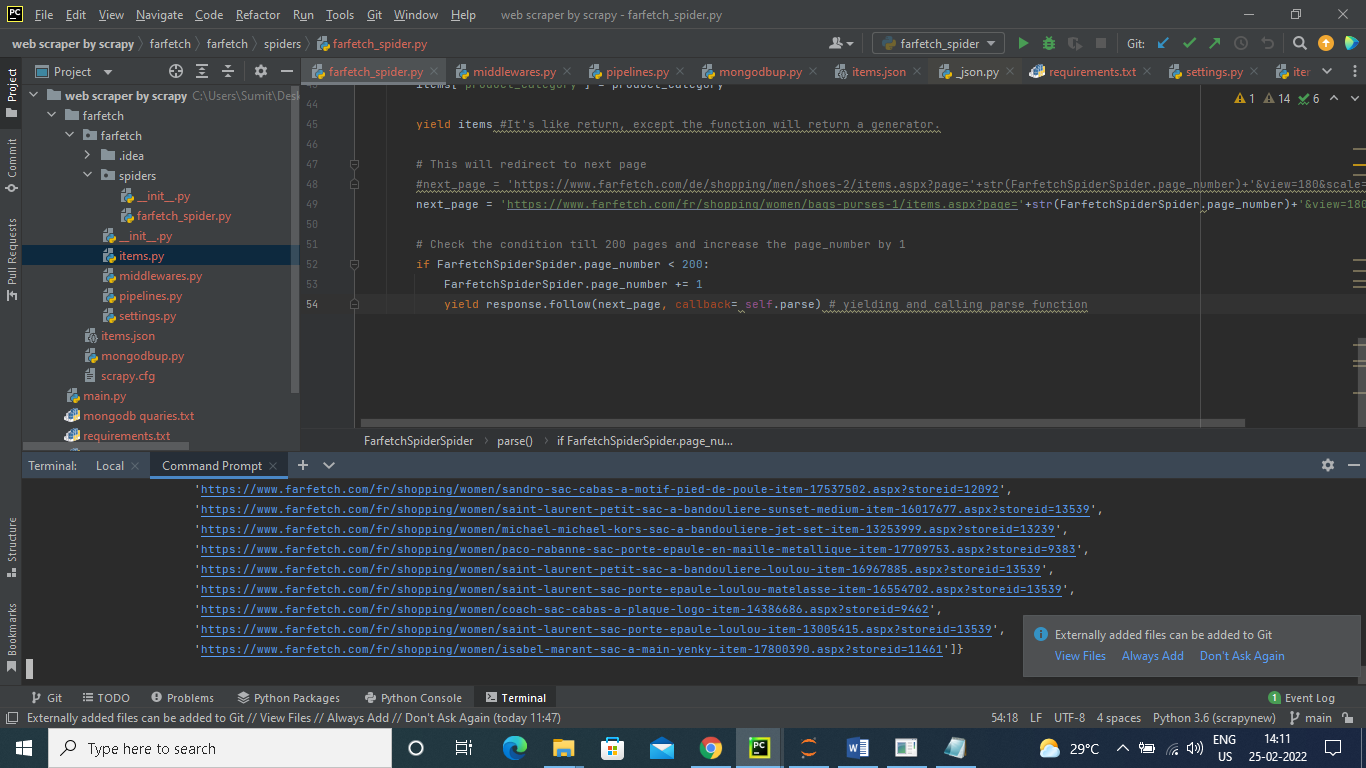
Step 9: here we have to extract 200 pages so for that we have defined a condition here to redirect to next page



Step 10: scraping data from web site and saving that data in item.json using command

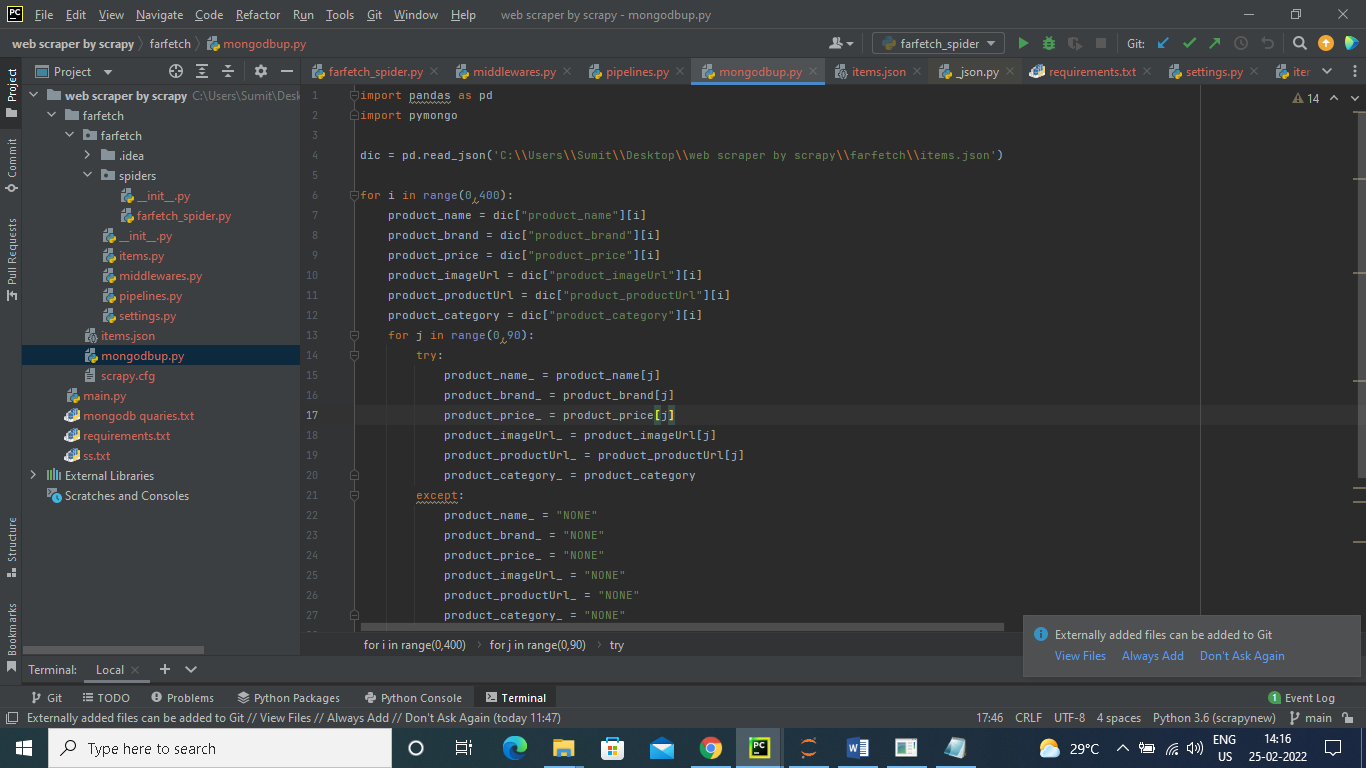
scrapy crawl farfetch\_spider -o items.json



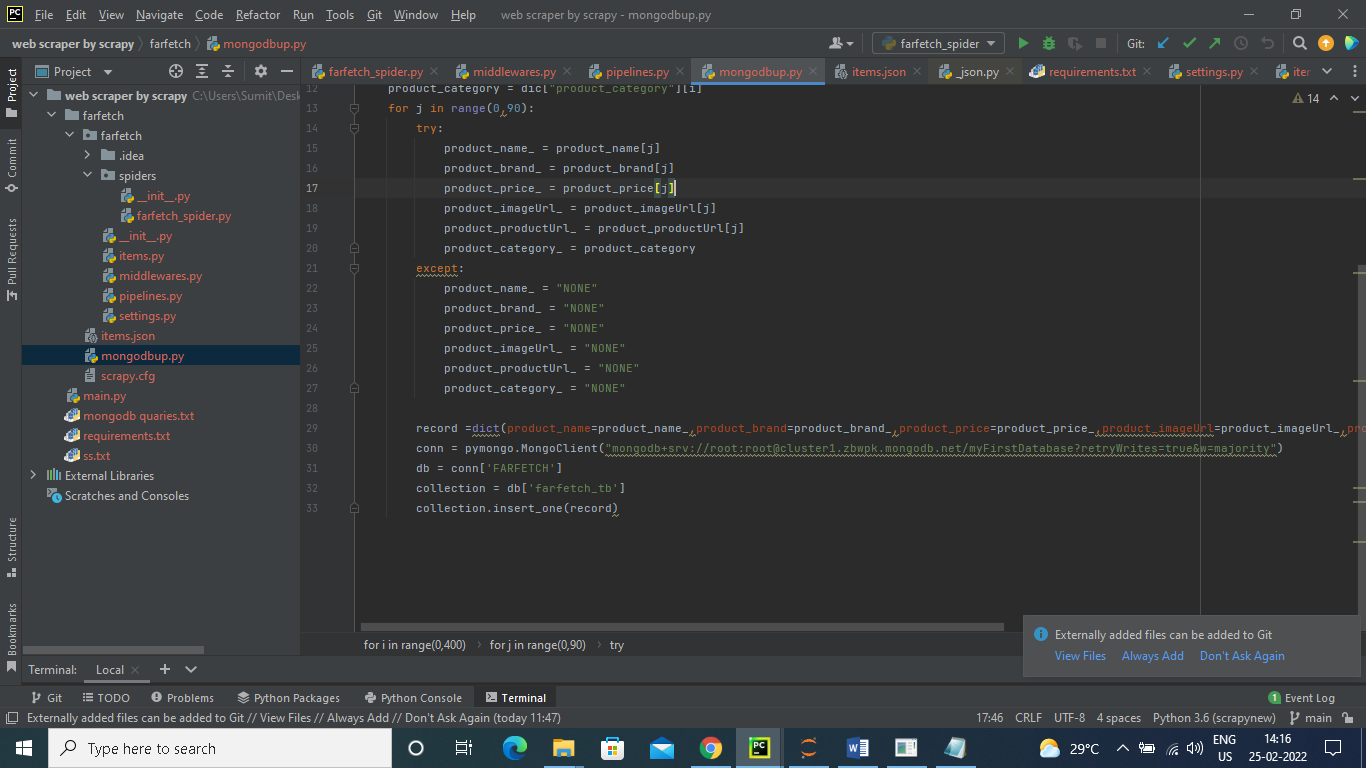


Step 11:creating new file mongoup.py for upload this data to mongodb atlas database

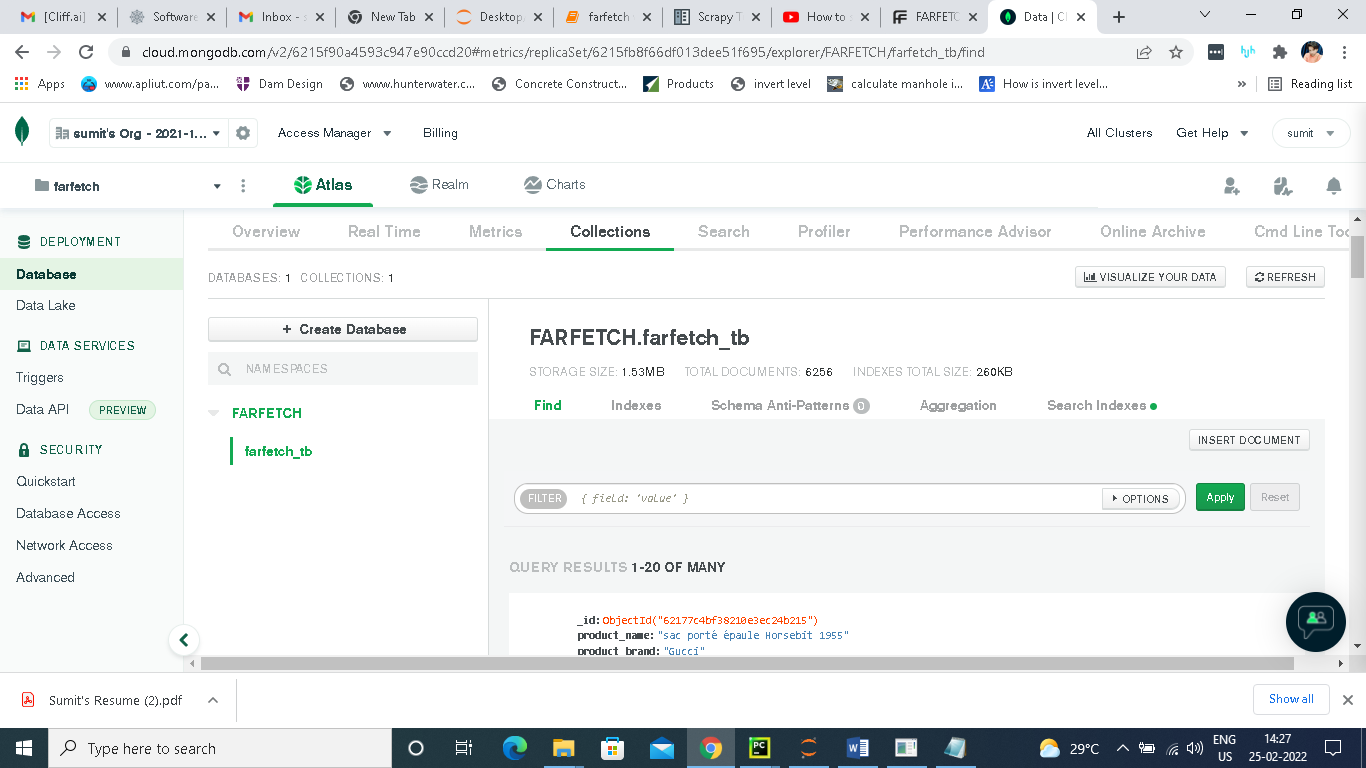
# here we can also use scrapy pipelines.py to setup automatic file upload but in our case it will take the page wise data but we want item wise so we will create mongodb.py file to resolve this problem



Here I have created two for loop it will take saved product items in item.json file one by one. it will take first value of first item from json file and upload that bunch of product\_name,product\_brand,product\_price,product\_imageUrl,product\_productUrl,product\_category on mongodb atlas database.



Mongodb database Image



MongoDB Queries

1. How many products did you scrape?

Query - db.farfetch\_tb.count()

2. How many products have a discount on them?

Query - No discount available

3. How many Bags and Purses products don't have any discount on them?

Query - No discount available

4. How many unique brands are present in the collection?

Query - db.farfetch\_tb.aggregate(

[

{$group: {\_id: "$product\_brand"}},

{$count: "product\_brand"}

]

)

5. What is the count of discounted products for each brand?

Query - No discount available

6. How many products have `shoe` in their name?

Query - db.farfetch\_tb.find({product\_name : { $regex : /Shoe /}}).pretty()

7. How many products have offer price greater than 300?

Query - db.farfetch\_tb.find({product\_price : {$gt : 300}})

8. How many products have discount % greater than 30%?

Query - No discount available

9. How many Footwear products have a 50% discount?

Query - No discount available

10. Which brand in Footwear section is selling the most number of products?