Python Project 3

**Scraping data from portexaminer.com**

# **REQUIREMENTS**

The customers are often manufacturing companies. To analyse these supply chains, we must first identify who their suppliers are. One of the ways to identify suppliers is to look at the shipping data about the imports by these companies. The website that we will be using for this assignment is <https://portexaminer.com/>.

Here's the description from the website itself:

## [Trade Data from U.S. Customs](https://portexaminer.com/)

Port Examiner lets you examine **US customs import records** for millions of real ocean shipments.

You can search for the names of overseas suppliers or US importers to **see who is importing what – and from whom**.

You can also look up import records for **specific commodities or products**.

Access to our database of customs import records is **absolutely free** – simply type your search query in the box above to get started today.

You will be collecting the data about the below companies:

1. Ford Motor Company
2. Toyota
3. Tesla Motors
4. Mercedes-Benz
5. General Motors

You can see the data about their shipments by following the below steps:

1. Go to: <https://portexaminer.com/>
2. Change the dropdown option from ‘Supplier Name’ to ‘Importer Name’.
3. Search the name of the company. Eg. Ford Motor Company.   
   You will notice that after searching, the url changes to: <https://portexaminer.com/search.php?search-field-1=consignee&search-term-1=Ford+Motor+Company>  
   Here you will see the first of many pages of the shipment data.
4. To see more data about a particular shipment, you can click on its link. Eg. <https://portexaminer.com/trade-data/zhaoqing-hongteo-accurate-ford-motor-company/hlcuca4120851626/>  
   These links should be available in the html contents of the search results.

You are expected to automate the process of getting this data by performing the following steps:

1. For each company (searched as an importer name), get the detailed data on the first page of the search results.
2. Parse the data from all links on the first page and get the output in the following format:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Searched Importer** | **Shipper** | **Consignee** | **Notify Party** | **Bill of Lading No.** |
| Ford Motor Company | TRW AUTOMOTIVE GB | Ford Motor Company | UTI ROMULUS GSDB | HLCUCA4120851626 |
| Ford Motor Company | SUMMIT PLASTICS NANJING CO.,LTD. | Ford Motor Company | EXEL CONTRACT LOGISTICS | APLU062548853 |
| Toyota | Sinotrans Guangdong Imp & Exp C | Toyota Tsusho America, Inc. |  | EXDO612482590 |

You only need to extract the fields shown in the above table. You can identify and extract more fields if time permits.