Assignment

1. Scrape House Prices from the link: <https://www.magicbricks.com/property-for-sale/residential-real-estate?proptype=Multistorey-Apartment,Builder-Floor-Apartment,Penthouse,Studio-Apartment&cityName=Hyderabad> and map the localities to a unique set of zip-codes in that city. Python modules are available to solve the latter half of the task.
2. Write a code to match names from list1 (provided in the file: name matching -list1.xlsx) to list2 (name matching – list2.xlsx). The names in list1 have multiple variations of a name and map to exactly one name in list2. Also, provide an accuracy measure of the names matched.
3. Given is a list of business descriptions for around 24,000 companies (provided in the file company descriptions). There are 2 types of descriptions – long and short. Use your discretion while using either. Perform the following tasks using the file:

* Classify these companies based on their business descriptions to a only of the industries from the industry labels given (provided in the file: industry labels.xlsx).
* Given the same set of the companies with their business descriptions, cluster these companies, without using any information from the industry labels. Also, identify ways to calculate the accuracy of the clusters generated.

Link to the files required for completing the tasks: <https://isbhydmoh-my.sharepoint.com/:f:/g/personal/bhoomi_thakkar_isb_edu/EuA8UKvZGqBAtQEdfGtb4JUBpbGKQ61cXXc_W68ldPt25Q?e=olN92t>