

First ML Assignment to collect your own data

Problem statement: Parse website makemytrip.com or goibibo.com to collect all the listed hotel details.

Step1: pick any one website mentioned above.

Step 2: analyse and find a pattern in hotel urls with city/ country etc details and also use

- www.makemytrip.com/robots.txt
- www.goibibo.com/robots.txt

To check their sitemap, analyze and parse ONLY allowed links. (DO NOT DO ANY ILLEGAL HACKING - INSTITUTE OR FACULTY WON'T BE RESPONSIBLE FOR THAT)

Step3: generate a pattern in urls by step 2, so that you can parse all the urls in your loop in the code.

Step 4: use BeautifulSoup library to parse obtained links from step 3, and figure out html tags, classes or other property you need to extract out details and make a row to be further added to your csv (result) file.

Data to be extracted from hotel details pages:

Hotel Name	City	Country	Star Rating	Price per night	Amenities (array)	Hotel description	Any 2 reviews
------------	------	---------	-------------	-----------------	-------------------	-------------------	---------------

Note: make sure to go to the hotel details page to extract out all the data (as the hotel list page might not have the desired information).

Step 5: Save the row (collected information of each hotel) in the csv file

Step 6: Group the list in your result by country and then by city. Try adding all the hotels listed in the website - worldwide)