**Members of 17th Lok Sabha**

The idea of the project was to scrape off the list of members of the 17th Lok Sabha from the website, **https://en.m.wikipedia.org/wiki/List\_of\_members\_of\_the\_17th\_Lok\_Sabha**

and store it in a csv file.

To follow it up, the python library, BeautifulSoup was used. The first thing was to install the package bs4 in the terminal using the command, pip install bs4. Then the package was imported using the command **from** bs4 **import** BeautifulSoup.

Afterwards the scraping was done with following code:

url = **'https://en.m.wikipedia.org/wiki/List\_of\_members\_of\_the\_17th\_Lok\_Sabha'**r = requests.get(url)  
**if** r.status\_code == 200: *#checks the status code*

The above line checks the status code is equal to 200, if yes then only the code will proceed else would give the output as “Invalid Source Code.”

After this line, inside the if statement comes the lines of code that parses the url and finds all the data in form the webpage.

soup = BeautifulSoup(r.text, **'html.parser'**) *#parses the webpage and stores its source code* table\_ = soup.find(**'table'**, class\_ = **'wikitable'**) *#finds the first table of the webpage* row = table\_.find\_all(**'tr'**) *#finds the tr tags of the table*

row has all the rows of the table in the form of source code.

Then to store all the entries that are in the row, in a list named y, the below given code was used.

**for** team **in** range(0,len(row)):  
 head = row[team].find\_all(**'td'**) *#finds all the data entries in every row* **for** i **in** head:  
 **if ''** != i.text.strip():  
 y.append(i.text.strip())

when all the data was scraped, it was found that there were a lot of empty spaces in the list, that’s why, the if condition checks if there is any blank space, it skips it and only enters the non-empty values into the list.

In the code, ‘i’ stores the data in the form of source code so to convert it in the form of normal text, i.text is used. But there also gets some unwanted spaces in it, so instead ‘i.text.strip()’ is used.