Report on Wids Project

———— Hari Hara Teja

# First three weeks ….

We started the first few weeks with an introduction to some languages like Python, HTML, CSS, JavaScript and many libraries for Python, such as pandas,numpy, matplotlib, for plotting graphs using the data files, etc. Those were very interesting to do. After that, we were given one week to complete the assignments and complete the incomplete work.

Now comes the final project,

The problem statement is to write a code, which automatically stores the specified content in the specified website.

We imported libraries like requests, selenium,beautifulsoup4,

We used this link <“https://www.imdb.com/search/title/?groups=top\_100&sort=user\_rating,desc"> top movies in imdb

Our goal is to store the data of the top 25 movies such as name,

Duration, rating, director, etc

Using requests, we sent a request to the IMDb website and then using beautiful soup and html5lib, we parsed the HTML code

And by using XPath and CSS selectors, we stored the links of the top 25 movies in a list and in a movie\_url JSON file

Now we have got a list of URLs by running a loop over the list and accessing every movie URL, and then collecting the data, through xpath

And appending them to the .json file