Thanks, Suvarna

So, we start with importing the necessary libraries.

We have scraped the data form the website : <https://songmeanings.com/>

Songmeanings.com contains millions of song’s lyrics and comments. To obtain song lyrics/comments, we need to obtain the URLs to that page.

This is a premade website index info file were

**Letter column** contains All the artist whose names starts with letter A

**pages column** Total pages corresponding to that letter

So, in the below cell we have looped the letter and the number of pages to construct it into a URL structure to access the data.

Below is the sample photo of one of the pages which contains all the artist starting with A.

Similarly, after constructing all the URLs, we have started scraping the data from each page using beautiful soup and finding the table tag and the contents of the table body.

Here our main aim is to scrap the Artist name and his URL(profile URL)

For example, we have scraped the URL '//songmeanings.com/artist/view/songs/6024/' for the Artist 'A' as shown below from the anchor tag which is inside the table body tag.

For each artist we have found the URL and made it into the data frame for future use.

After getting the URL of each artist we are scraping the song name, song URL and number of comments for each song by using beautiful soup and finding the table tag and the contents of the table body.

For a Particular Artist we are scraping the list of songs and its corresponding link and the number of comments

For each artist , song and its URL and the number of comments we have made it in a data frame

So, as we are also been working on comments so we will scrap only those songs where there will be at least 1 comment.

With the help of beautiful soup and have scraped those songs which have at least one comment

For example, in the below image we can see the lyrics and comment on the same page we have scraped for lyrics and comments into two separate data frames.

Total Songs Scraped : 8650

Total Comments Scraped : 31612

# Going ahead with Data Cleaning and Pre-Processing part

We have dropped all those records if there are no lyrics or comments present and checked for null values.

## Now EDA part will be explained by Abhishek