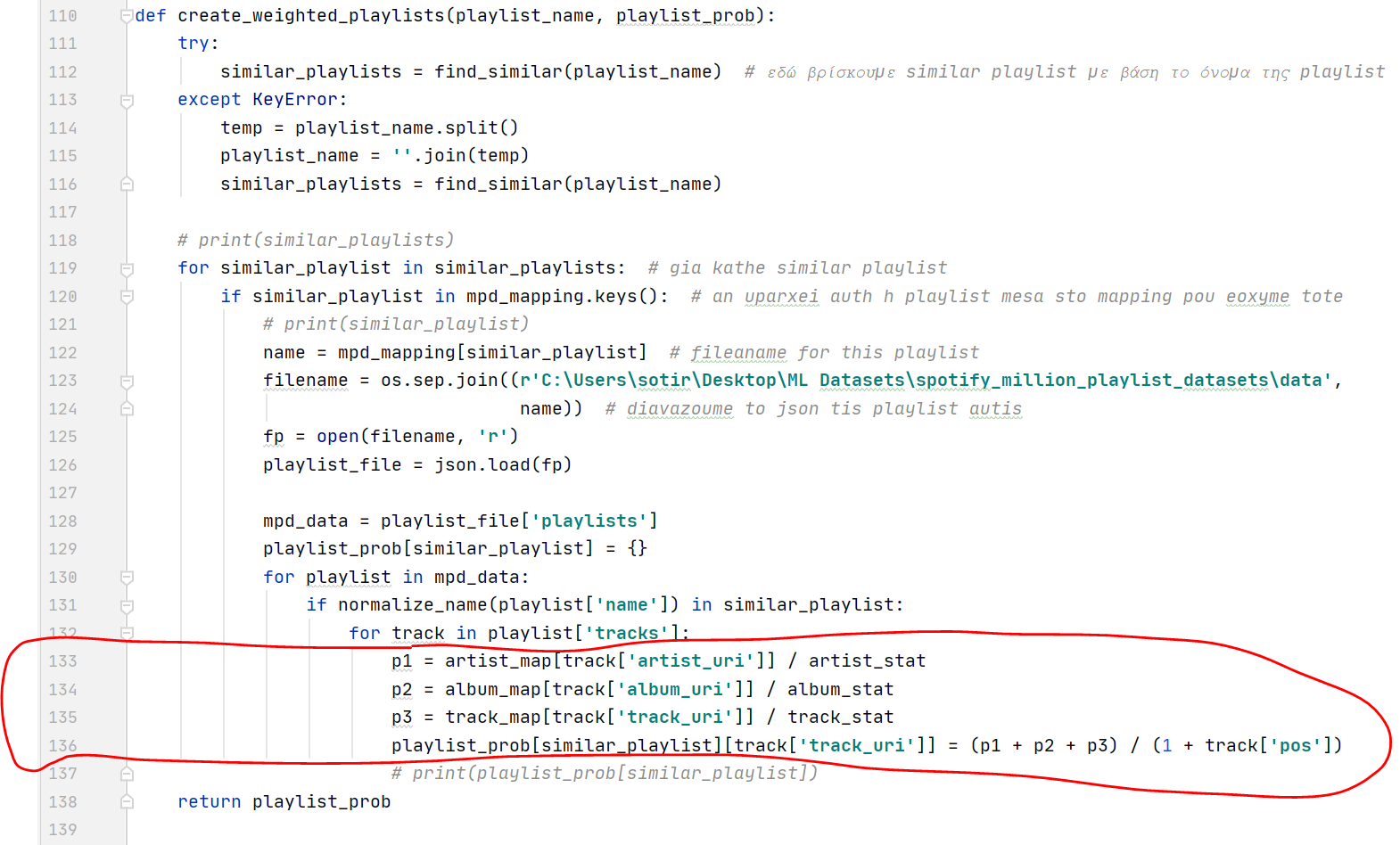
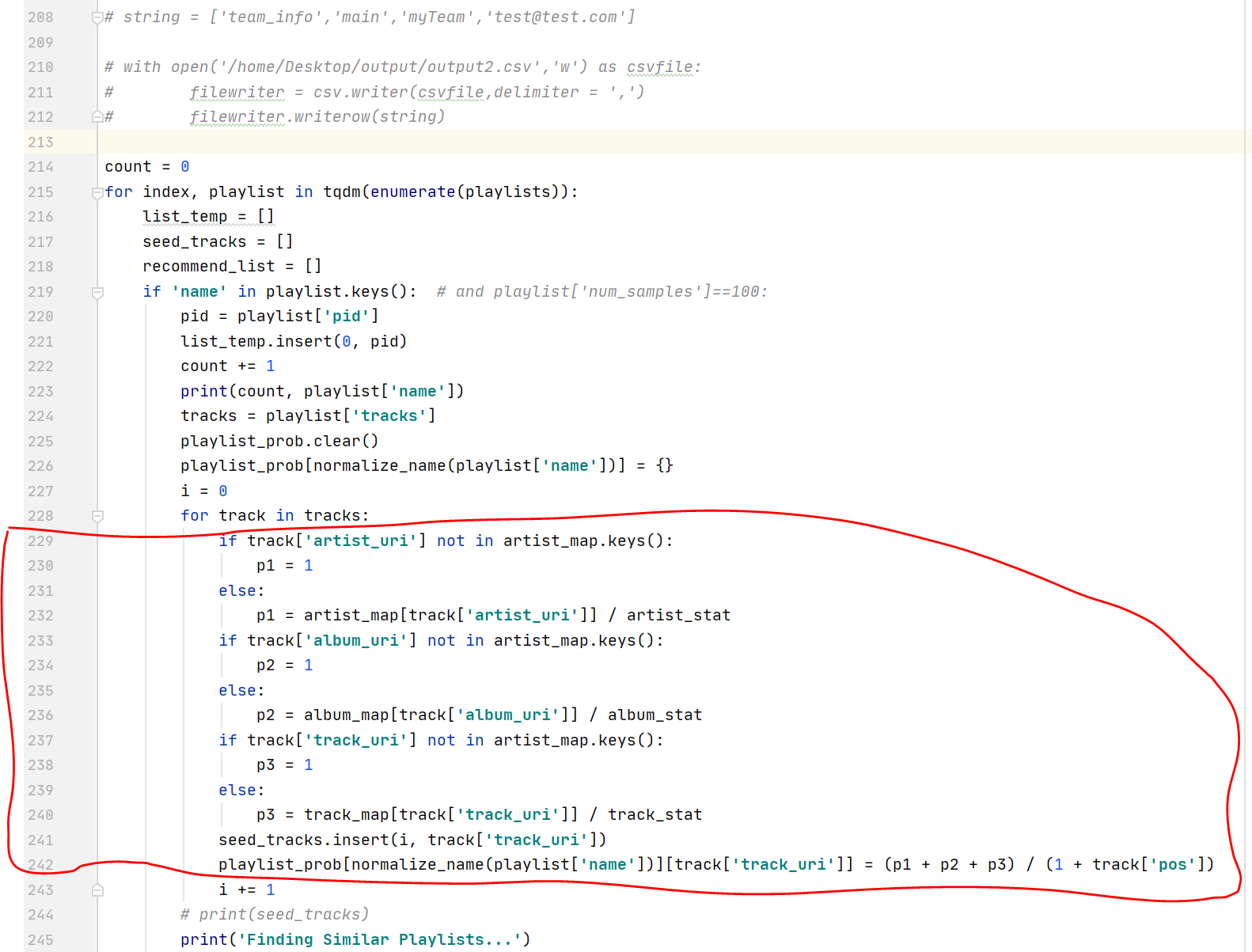
I work on a project about a recommendation system at Spotify where I try to find similarities between playlists from the challenge\_set(training\_set.json) and the million playlist dataset(training\_set). I want you to write a program in python language that uses the existing code from the file DIPLOMATIKI2B.ipynb. Please delete the lines  133-136(I have flagged these lines in the first screenshot below) in DIPLOMATIKI2B.ipynb file and substitute these lines by calculating the Jaccard similarity between these playlists that take into account track\_uri, artist\_uri and album\_uri saving the results in the playlist\_prob  variable using the following formula:

(1+ 50% \* Jaccard(track\_uri) + 25% \* Jaccard(artist\_uri) + 25% \* Jaccard(album\_uri))  / (1+ (Jaccard(name)).

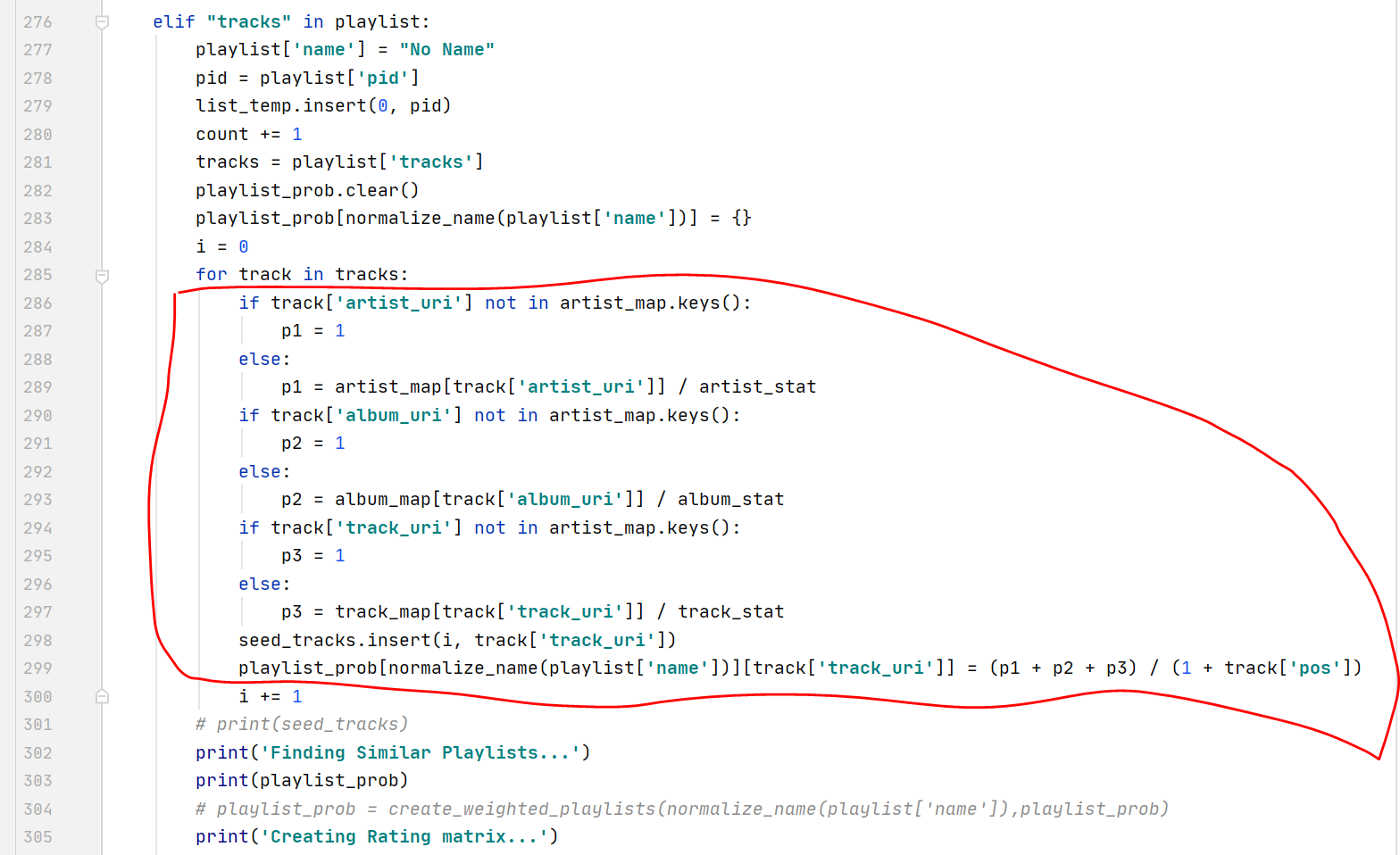
I also want to do the same for the lines 228-242 using the same formula (I have flagged these lines in the second screenshot below) saving the results in the playlist\_prob  variable.





In addition I want to delete the lines 286-299 using the slighlty different formula (I flagged these lines in the screenshot below) saving the results in the playlist\_prob  variable:

(1+ 50% \* Jaccard(track\_uri) + 25% \* Jaccard(artist\_uri) + 25% \* Jaccard(album\_uri)).



Please after you finish with this,  I want to write a second program that does exactly the same  thing with the same code as I described before  except that in the last notebook cell uses firstly

if **"tracks"** in playlist:

 and secondly

elif **'name'** in playlist.keys():

 Make sure that the program builds the final submission.csv file correctly in the last notebook cell using the verify\_submission.ipynb file  to verify it.

Please write the code as simple as possible and write some comments to understand it better and I am sorry for the big message but I only want to describe the problem as clear as possible.

The webpage of this project is located in the following link:   <https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge>

The million playlist dataset(training set) and the challenge set(test set that contains incomplete data) are located in the following link:   <https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge/dataset_files>