Please Note:

1. All the tasks are independent of each other.
2. I have repeated the code and saved the results and using those results in the next task.
3. You need to have glove embedding on your google drive (It's fast that way). Else these can be loaded to memory with command !wget -P /root/input/ -c "https://nlp.stanford.edu/data/glove.6B.zip" . This has to be unzipped then !unzip /root/input/glove.6B.zip
These commands and how to get embedding is shown in task 1 towards the end for reference.

Instructions:
1. Task 1 is completely independent.
2. Task 2 gets embedding, train data and validation data from google drive . That's it. After this the code should run without any requirements. This will also save validation and train cosine in google drive.
3. Task 3 again takes embedding, train data and validation data from google drive. This time the cosine dataframe for train (changes the above) is saved to google drive.
4. Task 4, use the above saved dataframe to do under sampling. So, just the validation cosine dataframe from Task 2 and training dataframe with cosine from task 3 is required in google drive.