FOR TASK 1: Have to upload the tsv to colab or have it in the curr dir

TASK 2.1:
There are 4 sections. In the first one, the embeddings dim is 384. In the second, it is 200 (but no embeddings are generated, just like tutorial). In the third, it is also 384. And finally, in the 4th, the transformer, it is also 384.

TASK 2.2:
For K-means, we define it, and the tutorial uses 5 so I used 5. IT took 24.20102000236512seconds to run, and the time complexity of k-means is O(n^2)

For the agglomerative clustering, I ran into OOM errors. This is likely due to the O(n^2) memory requirement, which is also its runtime complexity btw. I had tried reducing the distance constraint to be more flexible. In the end, just to see, I took the first 10k entries and clustered. It took 31.65395450592041 seconds to run, and returned 1709 clusters.

For more details, check the python notebook included.

TASK 2.3:
I defined my new vector dim to be 120. Again, 5 for kmeans, and this time 1614 for agglomerative clustering. Kmeans took 10.465742349624634 seconds, and agglomerative took 11.293874740600586 (but, again, this is on a subset of 10k entries; else agglomerative was running into an OOM error again). In addition to what I tried in task 2.3, I also tweaked

the dims, from 128 -> 120 -> 50 -> 10 -> 5. More details are in the notebook as well.

TASK 2.4:
For this one, I used the recommended fast_clustering solution provided on the HW7 doc. The running time for this algorithm(https://github.com/UKPLab/sentence-transformers/blob/master/sentence_transformers/util.py#L346) was 0.2782862186431885 secs. Normalized mutual info score was 0.8701767165285461, and adjusted rand score was 0.6595287776974524