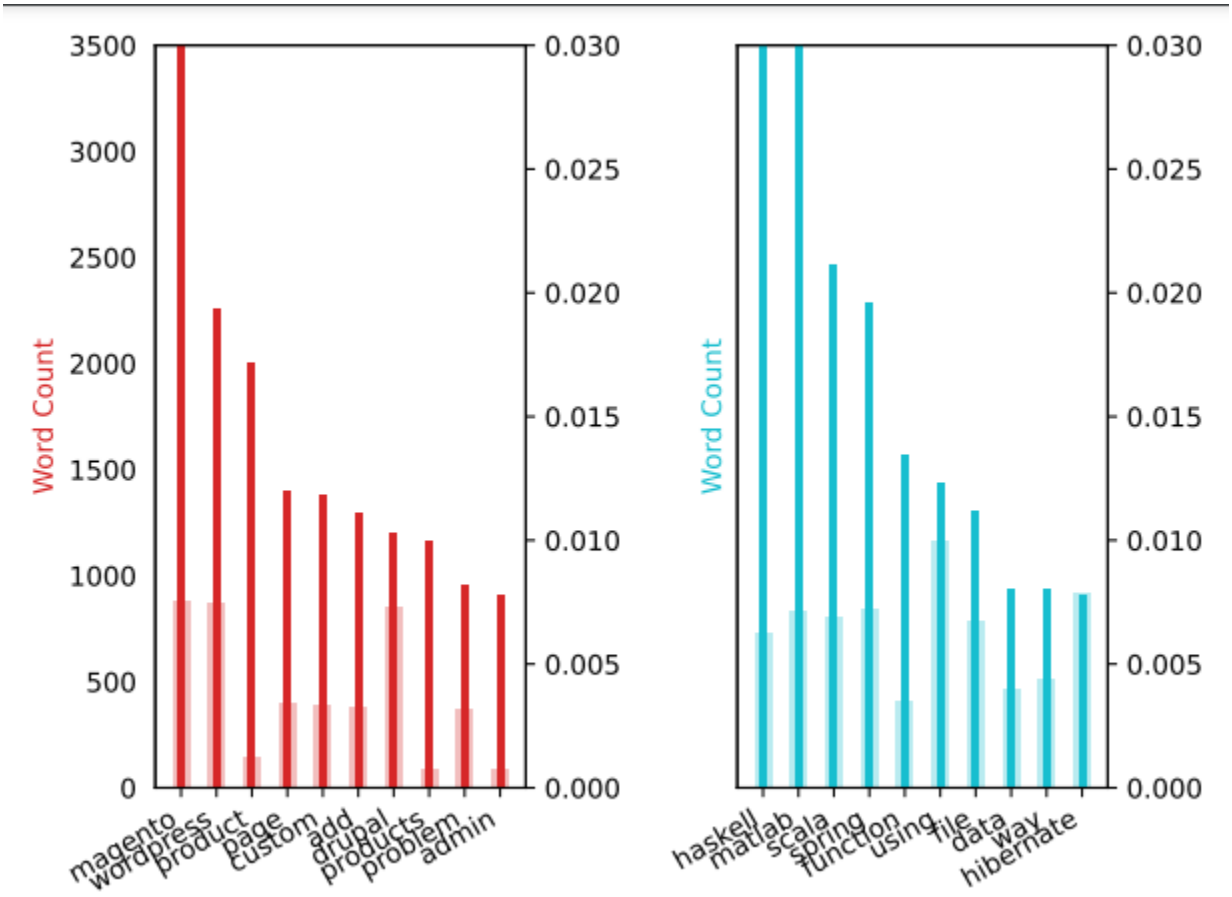
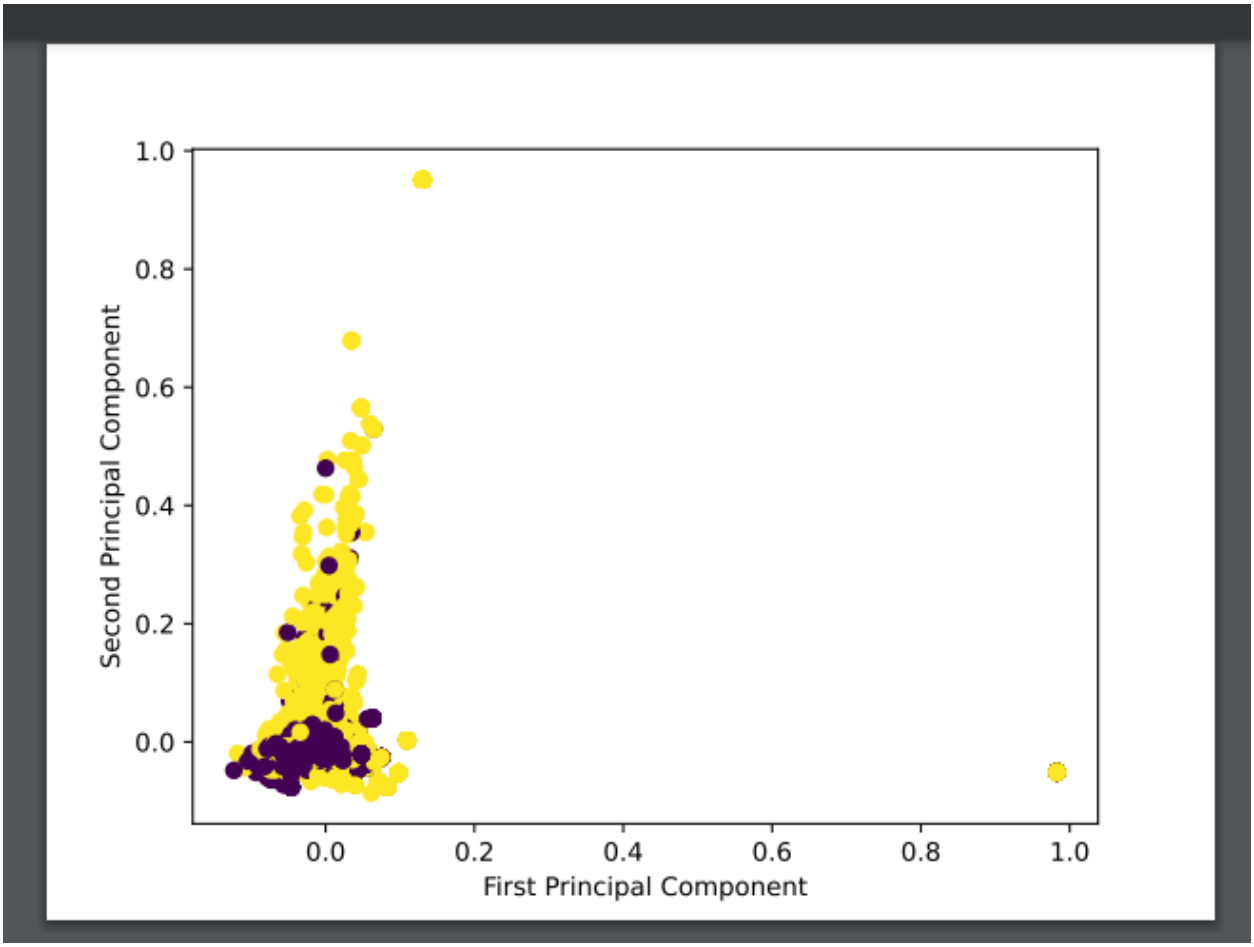
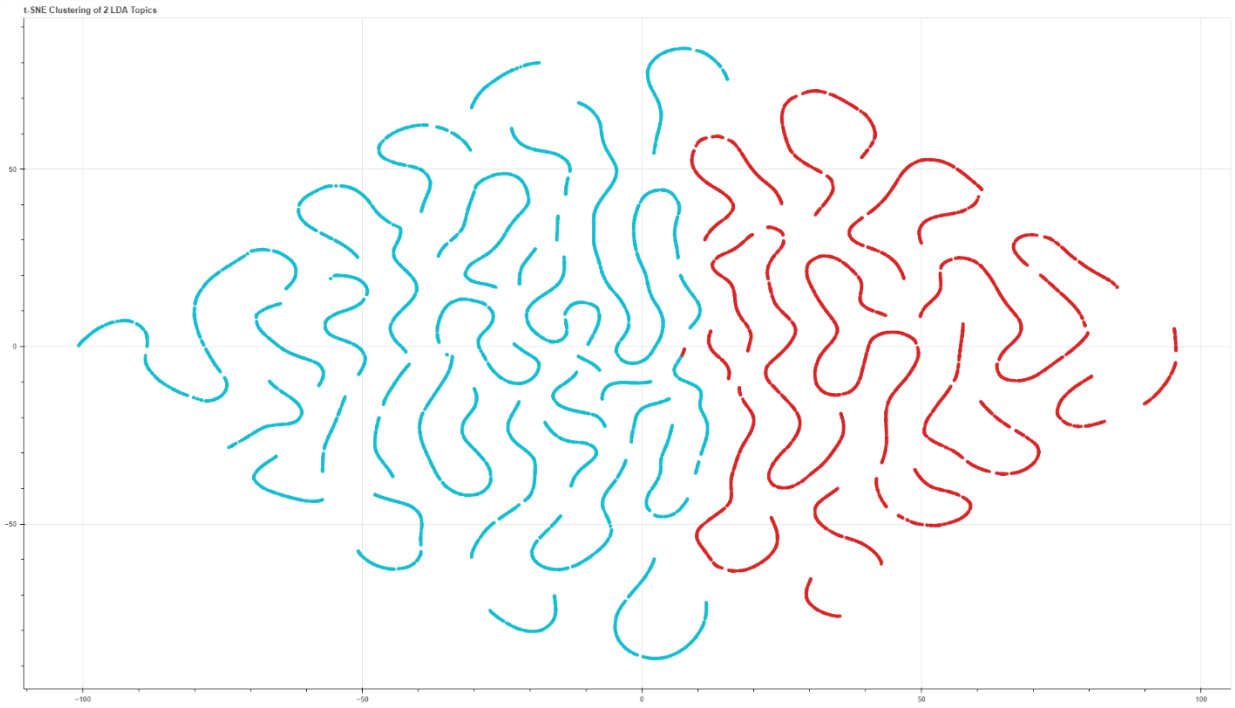


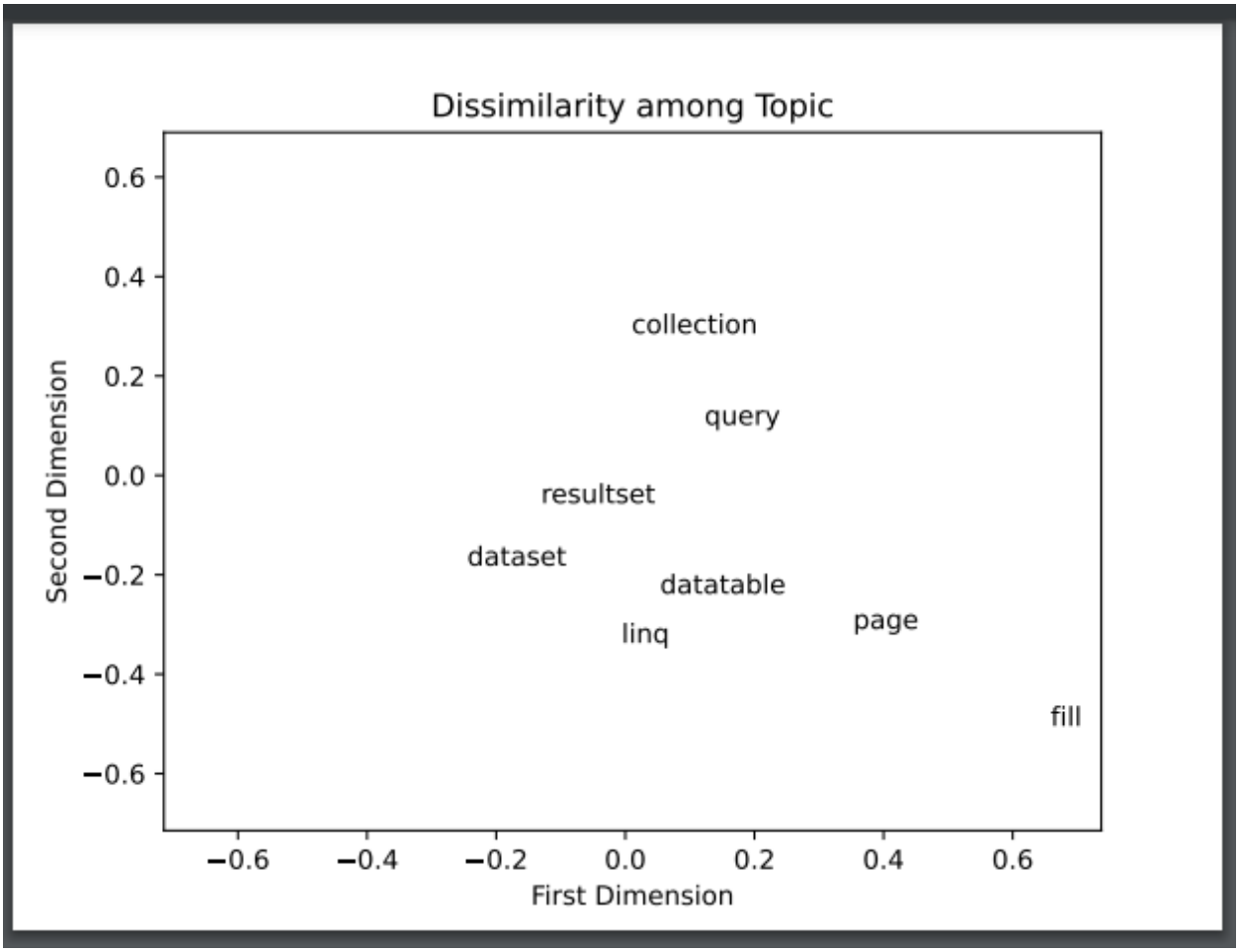
In LDA models, each document is composed of multiple topics. But typically, only one of the topics is dominant. Above figure show the dominant topic for each sentence and the weight of the topic and the keywords. When it comes to the keywords in the topics, the importance (weights) of the keywords matters. Along with that, how frequently the words have appeared in the documents is also interesting to look. We can infer from above figure is first topic is about a sales page. Second topic is about computer science. We can see frequency and weight do not match the reason for that is the more a word appears in another topic it reduces its weight in this topic. So it makes it less important.



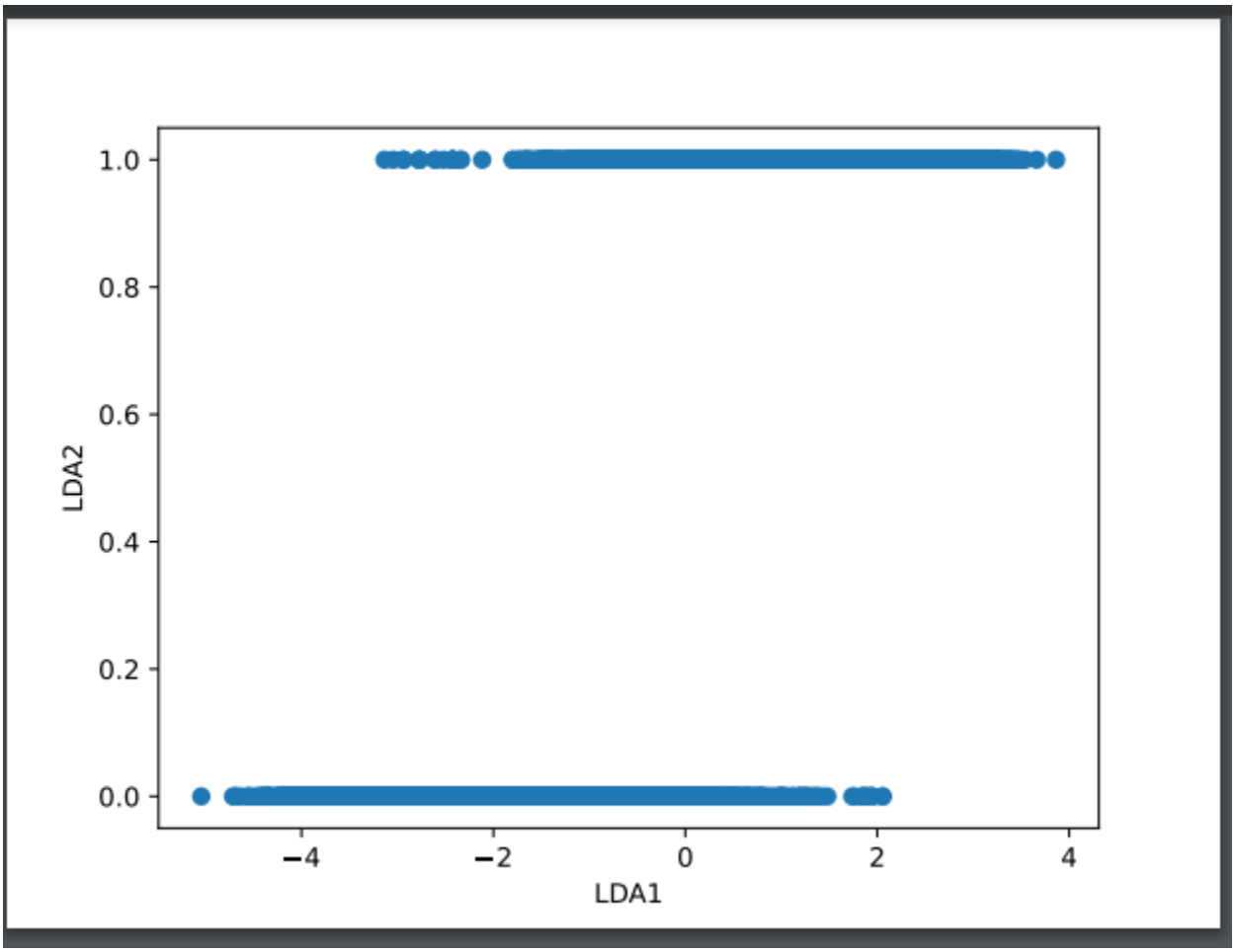
From this figure we can infer first component is not a differentiation factor in this data set. Second component seems effect but near the 0 it contains both types. We can only use it to say it is not purple in higher values.



From this figure we can see x axis is the determining factor. If x > 0 it is on the red topic else it is on the blue topic.



From the figure above we can say dataset and result set are good substitute for each other. Whereas fill lies in a desolate area this means it has no substitute. Other empty areas represent the possible needs.



This directly show us that lda2 decides what is in the first or second topic.

Unfortunately I couldn’t finish dmm part of the homework so there is no visualization for that part.