

1. Harsha thulasi
2. I spent about 5 hours on this lab. It took a lot longer than I anticipated. I ran into some issues with the EMR cluster getting terminated here and there. In addition, my notebooks kept crashing here and there which also didn't help. I got different values compared to my Lab6 (for the "dead KING garden!" query) and trying to identify the reason took some of my time as well. Same with trying to match with the example in your lab6.txt. Ultimately I tried not to match the tfidf values perfectly, just tried match the docids.
3. I completed the lab and I believe the solution is correct but that is for you to evaluate. I don't know which part is extra credit but I'm assuming it's the same as lab6. Most of it was pretty straightforward as I just had to find the dataframe equivalent to what I did with RDDs.