Shea Durgin
https://github.com/sheaDurgin/NLP23/tree/main/Assignment_4

Q1 - Results:

Average P@1: 0.0070921985815602835
Average Mean Reciprocal Rank: 0.12549970266865101

Q2 - Results:

Average P@1: 0.0
Average Mean Reciprocal Rank: 0.1706378281750524

Q3 - Results:

Average P@1: 0.0
Average Mean Reciprocal Rank: 0.17041136844808843

Extra Credit
Q1 - Results:

Average P@1: 0.010638297872340425
Average Mean Reciprocal Rank: 0.08936049168439619

Q2 - Results:

Average P@1: 0.0
Average Mean Reciprocal Rank: 0.1373780999741669

Q3 - Results:

Average P@1: 0.0
Average Mean Reciprocal Rank: 0.10229937076634055

What I gathered from this assignment was that the fine-tuning I did had essentially no effect as the MRR was only 0.0002 higher on my fine-tuned run on the test-set (Q2) as compared to the no fine-tuning version (Q3). The 10% of my test set had no p@1 scores of 1, but had a higher MRR as compared to testing the whole set.

The extra credit results were all lower compared to the quora for each question I ran it on. However, fine-tuning actually gave a somewhat meaningful increase on MRR as compared to now fine-tuning. Since this is a small sample set, I am unable to determine which scores are outliers and where the true avg increase/decrease would land.