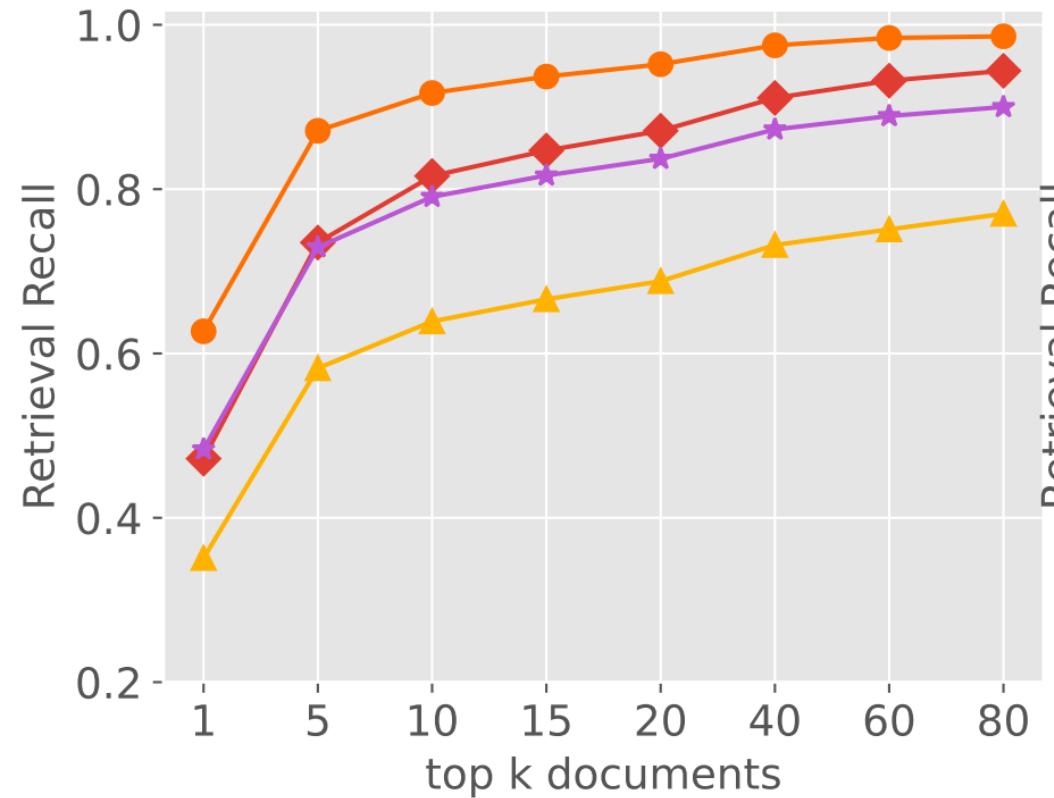
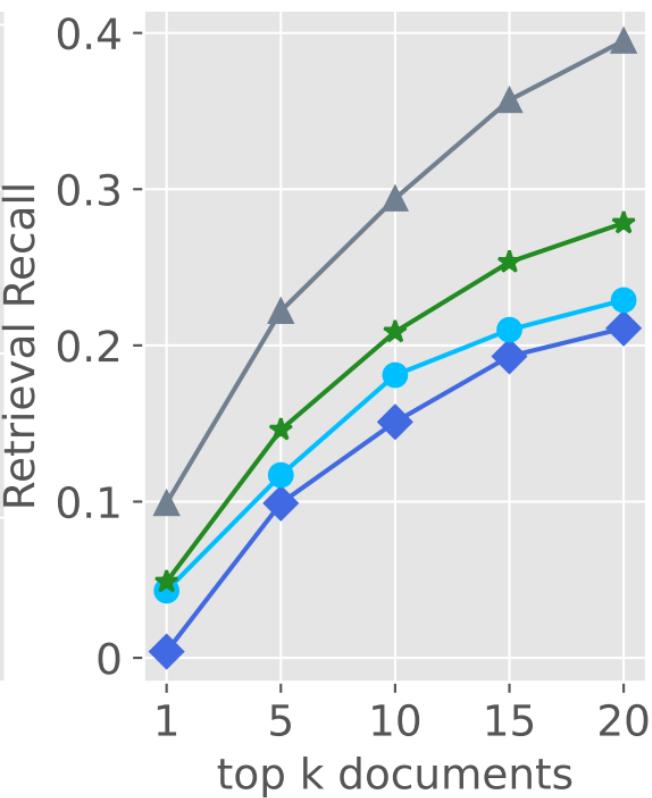


QA Datasets



Code Generation Datasets



Legend:

- NQ (Red Diamond)
- TriviaQA (Orange Circle)
- hotpotQA (Yellow Triangle)
- qa avg (Purple Star)
- conala (Blue Diamond)
- DS1000 (Blue Circle)
- pandas\_numpy\_eval (Grey Triangle)
- code avg (Green Star)