Job posting Dataset

computing\_desc\_job\_posting.csv

- only computing related job postings using TF-IDF

- tokenize, lemmatize

- NER for extraction of skills related

large\_cleanned\_desc\_job\_posting

- 5.5 hour code for tokenization on whole dataset

- mostly is hospice, care patient and healthcare. Hence decided to perform further EDA and focus on computing

-

postings\_large.csv

- original (124k rows)

postings\_small.csv

- small test dataset (1k rows)