data(1).csv 🡪 data yang udah di-tokenize dan udah free dari duplicate

cleaned\_data.csv 🡪 belum free dari duplicate

data 🡪 udh gaada kata berimbuhannya lagi, jadi semuanya kata dasar tapi pemilu nya diubah jadi milu jadi kurang recommended

data (3) 🡪 data yang udh di sort jadi ternyata ada 1 kolom text yang null jadi gw hapus, trus ini udh gaada “[]” dan “,” trus juga ini ga dihapus imbuhannya

data(2) 🡪 ga di-sort, yang na ga di buang

data (4) 🡪 sumber daya, geografi, dan demografi di augmented 500