

Opgave 3



Indlæs de to rds-filer samt csv-filen. Lav evt et samplet udvalg på 300 anmeldelser hvis din maskine ikke kan trække det fulde datasæt.

Lav et subset af csv-filen så den kun har tillægsord. Find de 20 hyppigst forekommende navneord i alle anmeldelser.

Generer bigrams fra anmeldelserne (evt nøjes med dit test-set) og udvælg nu bigrams hvor det første ord forekommer i dit subset med adjektiver og det andet i din top-20 og tæl hyppigheden.

Tilføj en score-kolonne hvor du ganger hyppigheden af bigrammet med sentimentet for adjektivet så du f.eks har:

word1	word2	n	X5	score
god	service	14	3	42
hurtig	levering	12	2	24