**Q1)** See Q1.py

**Q2)** See Q2.py. The Dataframe is stored in Q2\_Data.pkl

**2c.** The company with the most number of job ads in the past 2 years is ArmenTel CJSC.

**2d.** The month with the largest number of job ads over the years is March.

**Q3)** See Q3.py. Related Dataframe is stored in Q3\_Data.pkl where the document cluster assignment is in the ‘classes’ column.

**3b.** The top group consists of high frequency bigram/trigram such as

[('software developer', 496), ('software engineer', 370), ('senior software', 291), ('qa engineer', 196), ('java developer', 187)],

and relates to IT related roles.

The second group consists of

[('project coordinator', 74), ('project manager', 61), ('monitoring evaluation', 59), ('project assistant', 54), ('team leader', 43)],

and mostly relates to project manager roles.

The third group consists of

[('sale manager', 251), ('marketing manager', 100), ('marketing specialist', 90), ('sale consultant', 68), ('brand manager', 53)],

and mostly relates to sales and marketing roles.

The fourth group consists of

[('chief accountant', 278), ('senior accountant', 46), ('deputy chief', 38), ('deputy chief accountant', 35), ('procurement specialist', 33)],

and mostly relates to accounting and finance roles.

**Q4)** See Q4.py. Related Dataframe is stored in Q4\_Data.pkl.

**4c.** 44 out of the 116 best-selling music artists are from the 1990s.