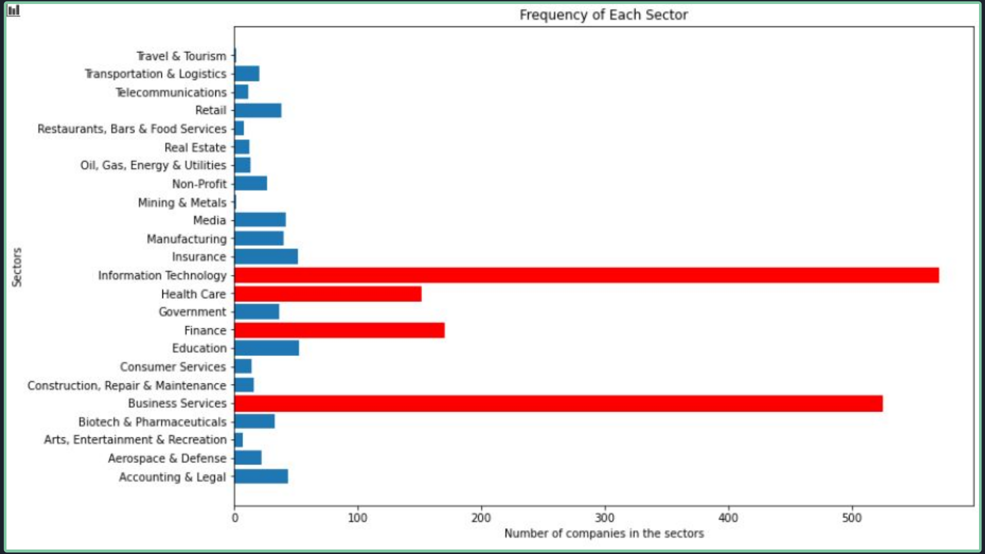
Hackathon 2

Group 11

Data Preparation

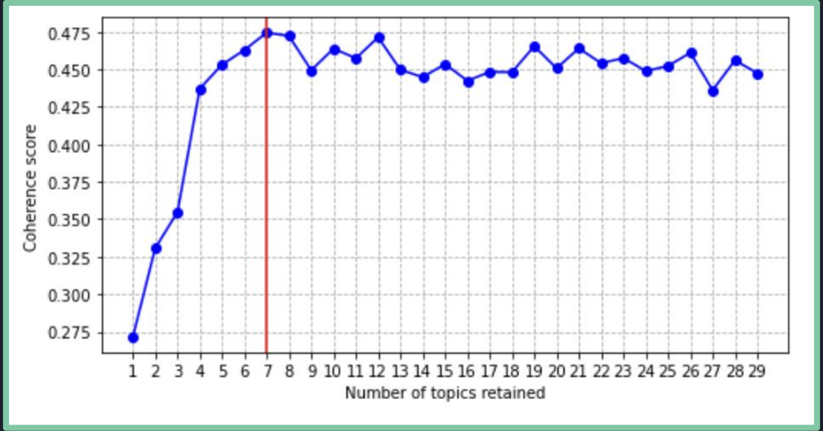


1. Group by Sectors and count the number of jobs for each sector
2. Most predominant sectors: Business Services, Information Technology, Finance, and Healthcare.
3. Data cleaning in text corpus “Job Description”
4. Tokenization of job description corpus
5. Remove stop words

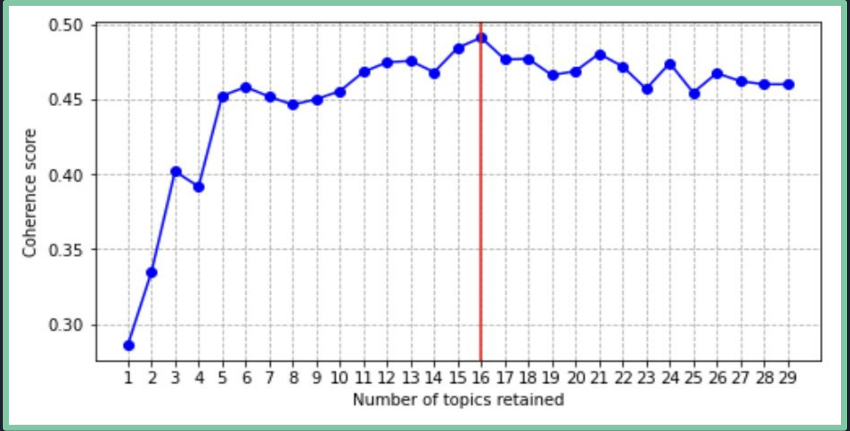
Topic modelling

Method: using coherence score to tune the optimal number of topic

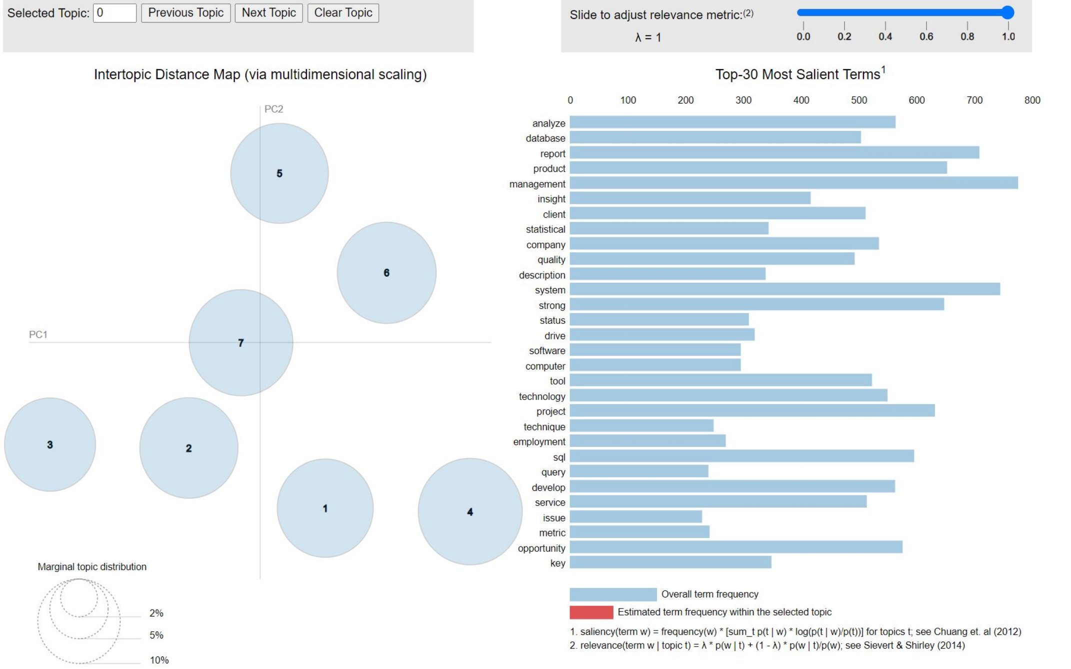
- Optimal number of topics for IT sector: 7



- Optional number of topics for Business Services Sector: 16



Results for the IT sector



Selected Topic: 0

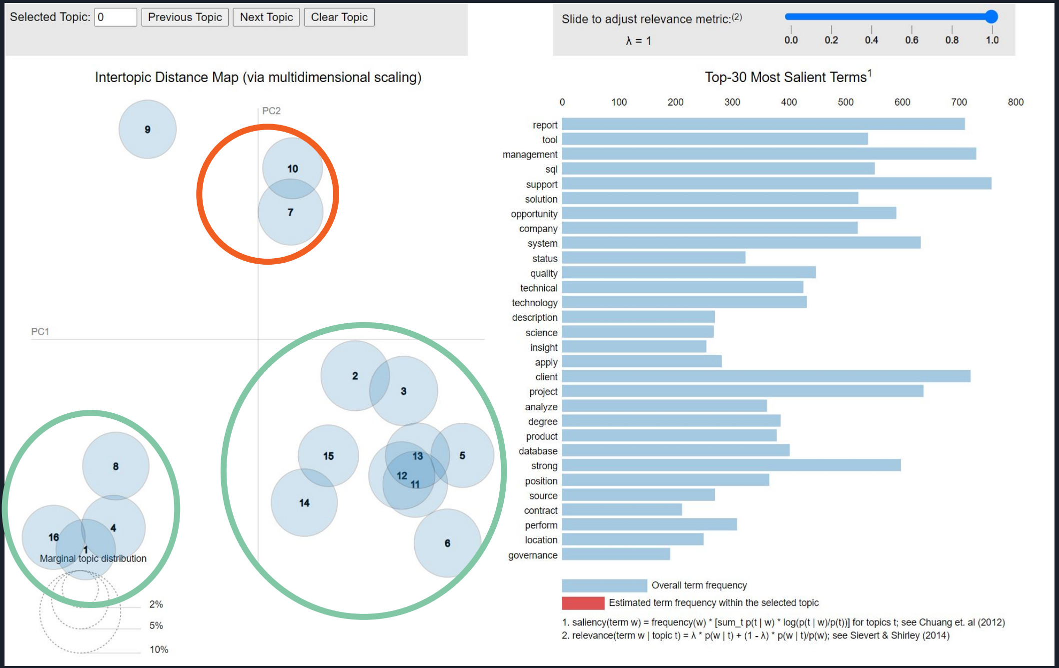
Slide to adjust relevance metric:(2)

λ = 1

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

- Topic 5 refer to the required qualifications background for the job
- Topic 4 refers to the culture and workplace values
- Topic 2 and 3 refers to the tasks and responsibilities of the position
- Topic 1 and 7 refers to the technical and soft skills required for the position

Results for the Business Service sector



Selected Topic: 0

Slide to adjust relevance metric:[2]

λ = 1

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

- Topic 9 refers to the basic informations for the position such as title job, location, salary, etc ...
- Cluster at the right bottom refers to the required education level , competencies, responsibilities, etc…
- Topic 16, 1, 4, 8 refer to the equality and diversity rights of applicants
- Topic 7 and 10 refer to the technical and soft skills

Comparison of Results

- In IT sector, topic 7 is the largest as compared to the others Analytical and computing skills are important for a data scientist in the current market.
- In Business Service sector, soft skills such as communication and interpersonal skills make this sector distinctive from other sectors
- Well defined topics for the IT sector as compared to the Business sector
- The topics in the Business Service sector are overlapping as these topics are correlated to each other indicating they are semantically related