Analysis of the Danish Data Science Job Market

Barbara Plank
Department of Computer Science
bplank@itu.dk

IT UNIVERSITY OF COPENHAGEN

Technical report, December 14, 2018

Abstract

This study analyzes the need of data scientists in Denmark on the basis of a 2018 job post analysis ranging over data from a 2.5 month period. Our analysis provides strong evidence for a high demand for data scientists in Denmark, providing additional insights on geographical dispersion, language considerations and types of sectors.

1 Introduction

Data scientists are in high demand globally and demand outpaces availability [1, 2, 3]. In this report, we seek to provide evidence for the high demand of data scientists in Denmark across three dimensions, by providing empirical results of a job posting analysis.

Specifically, we provide empirical evidence for:

i) the immediate open job market for data scientists in Denmark with respect to its *geographical distribution*,
ii) a quantification of job postings for data scientists in Denmark in respect to *language considerations* and
iii) the demand of types of *employers* (public/private etc) seeking data science competencies in Denmark.

Our study on a dataset of job postings collected over a period of 2.5 month shows that data science competencies are of high demand locally. This complements recent findings from an earlier empirical study in Denmark [4] and evidence at a global scale, e.g., the LinkedIn workforce report [1].

Contributions Our study shows:

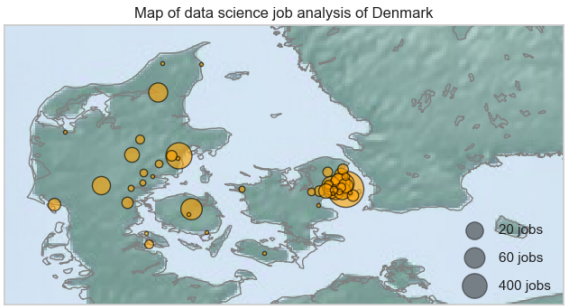
- Data scientists are needed throughout Denmark; the largest proportion of job are sought after in the main capital area (40% in the municipality of Copenhagen).
- The job market is looking for English-speaking employees, as evidenced by the fact that 66.3% of the job postings are in English.
- Data science competencies are needed in many sectors, i.e., the public sector, companies and universities.

2 Method

2.1 Data

We analyze job postings collected in the time frame of 2.5 months, in particular, from September 20 to December 3, 2018. We follow the setup of [4] and use their data-science-related terms (e.g., ‘data science‘, ‘predictive analytics‘, ‘machine learning‘, ‘data mining‘, ‘data engineer‘ details in the

(a) Denmark Map



(b) Most frequent cities

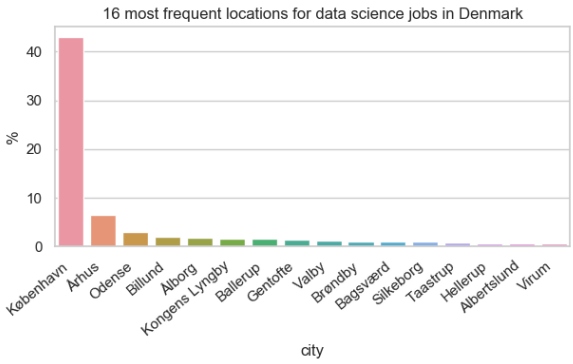


Figure 1: Locations of data science jobs in Denmark.

reference) to query for jobs posted in Denmark.[1].

After duplicate removal (outlined in Section 2.2), the resulting data set consists of 1,390 job advertisements. For each job we have access to job title, company, language of the post (as provided by the data provider) and a short job snippet. We also scrape the entire job posting for further analysis, and extract the job description section of the job post (using `BeautifulSoup` for localization of the job description and for boilerplate removal).

An example job posting is provided below.[2]

```
Lead Data Scientist for R&D Division
Company: KMD
Location: Ballerup
Snippet:
Do you want to help making the world a
better place, using machine learning
and data mining? You can develop
data science knowledge utilizing
multiple sources...
```

2.2 Preprocessing

The first step consists of the removal of duplicates or near-duplicates. To do so, we use a simple bag-of-word analysis and compare similarity of the entire job postings using cosine similarity on a a term frequency (TF)-inverse document frequency (TF-IDF) representation for the documents. By manual inspection, we set the similarity threshold t to $t = 0.7$. This resulted in the final data set of $n = 1,390$ job posts (out of 1,900 job posts collected initially), where highly overlapping or near-duplicates were removed.

We further enrich each job post by geolocation information (longitude and latitude).

Results

The job analysis provides three take-aways.

Geographical distribution As illustrated in Figure 1 (a) and (b), there is a high demand for jobs in the main capital region (53% of the job posts are around the greater Copenhagen area[3]). In fact, the highest density of jobs can be found in København, followed by Jutland (Aarhus) and Funen (Odense).

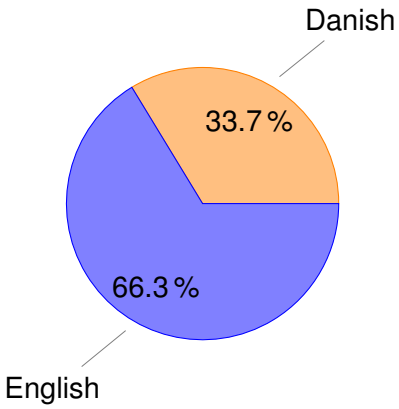
Language considerations The analysis provides strong empirical evidence that English is important for the IT job market in Data Science. As shown

[1]From `http://dk.indeed.com/`.

[2]The data is available at: `https://www.itu.dk/people/bapl/dsjobsdk/`

[3]Taking into account København, Frederiksberg, Valby, Brøndby, Bagsværd, Taastrup, Gentofte, Hellerup, Albertslund, Virum and Kongens Lyngby.

(a) Language of job posting



(b) Employers

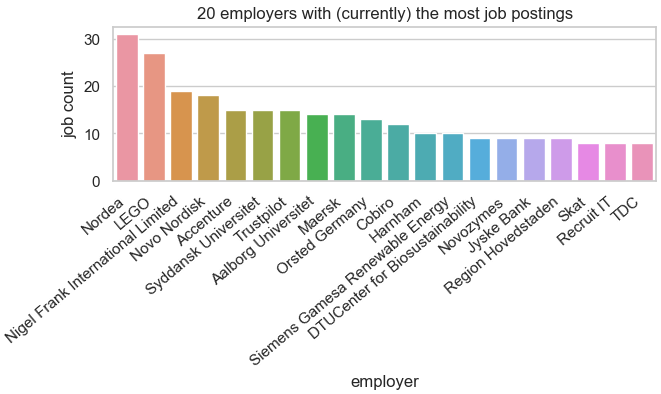


Figure 2: Job profile analysis: language of job posts, top employers (from our sample of 1,390 jobs).

in Figure 2 (a), the majority of job posts (66.3%) has been published in English. Moreover, out of these 921 English job postings, a simple keyword search reveals that 66% of the jobs explicitly require English language skills.[4] This further supports a global market.

Competencies across sectors Figure 2 (b) shows the most frequent employers. Amongst the 20 employers that seek the most data science competencies in our sample, we can find the Danish public sector (e.g., Skat, Region Hovedstaden), the banking sector (e.g., Nordea Jyske Bank), international cooperations (LEGO, Trustpilot, Maersk, Novo Nordisk, Novozymes, TDC) and universities (Syddank Universitet and Aalborg Universitet). This closer look further supports the need for data science competencies across industries, public agencies and universities, as has been identified at a global scale [1] and as shown here, is the case for Denmark as well.

[4]By searching for (in a case-insensitive manner) for the term "(in) English" in the full job descriptions, we found statements such as 'Fluent in English', 'Excellent oral and written communication in English', 'academic level in Danish or English', 'possess good communications skills in English'.

References

[1] August LinkedIn Workforce Report: Data Science Skills are in High Demand Across Industries. https://news.linkedin.com/2018/8/linkedin-workforce-report-august-2018.

[2] The Figure Eight 2017 Data Scientist Report. https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport.pdf.

[3] The Figure Eight 2018 Data Scientist Report. https://www.figure-eight.com/figure-eight-2018-data-scientist-report.

[4] Natalie Schluter and Rasmus Pagh. An empirical analysis of the current Data Science Job Market in Denmark. Technical report, IT University of Denmark, 2016.