

Industrial Requirements for Jobs in Data Science

GOKUL BHOJRAJ
YESWANTH MARKONDA
PÄR ERIKSSON

Agenda

- ▶ Background Study
- ▶ Objective
- ▶ Methodology
- ▶ Results
- ▶ Conclusion

Background

- ▶ Data Science has been established as an important emergent scientific field and paradigm driving research and companies in different fields and markets.
- ▶ It also tackles related new scientific challenges and drives business, ranging from data capture, creation, storage, retrieval, sharing, analysis, optimization, and visualization, to integrative analysis across heterogeneous and interdependent complex resources for better decision-making, collaboration, and, ultimately, value creation.
- ▶ To handle these business challenges companies need people with different competencies and skills

Job Titles in Data Science

Data analyst:

Data analysts sift through data and seek to identify trends. What stories do the numbers tell? What business decisions can be made based on these insights? They may also create visual representations, such as charts and graphs to better showcase what the data reveals.

Data Scientist:

Data scientists are pros at interpreting data, but also tend to have coding and mathematical modelling expertise.

What did we want to find out?

- ▶ Analyse data science job advertisements published in Indeed.com for the year 2018
 - ▶ To find the different categories in Data Science Job market.
 - ▶ To find important requirements for Data Analyst and Data Scientist Jobs
 - ▶ Such as Skills, Tools, Major, Degree.
 - ▶ To visualize the results in an easy understandable manner.
 - ▶ Draw different conclusions from the results that we found.

How did we do it?

- ▶ Dataset 6963 job ads from indeed.com included the attributes:
 - ▶ Position, Company Name, Reviews, Location, and Description about the jobs.
- ▶ We have used Natural Language processing method to process the description about the jobs and found the major requirements from them.
- ▶ a) **Data cleaning:** Removing irrelevant attributes, later based on the data we found the two major job positions in the Data Science Job Market : Data Scientist and Data Analyst.
 - ▶ We categorized the jobs into Data Scientist, Data Analyst, Engineer and Others based on the position title given by the employers and classified the companies into small, medium, large according to the reviews of the company.
 - ▶ After cleaning the data, we found 6953 jobs that were usable and we created new attributes mainly Job category and size of the company.
- ▶ b) **Natural Language processing(NLTK):**
 - ▶ We have used previous works on using natural language processing technique to process the job descriptions.
 - ▶ We used Python libraries for implementing the Natural language processing techniques.

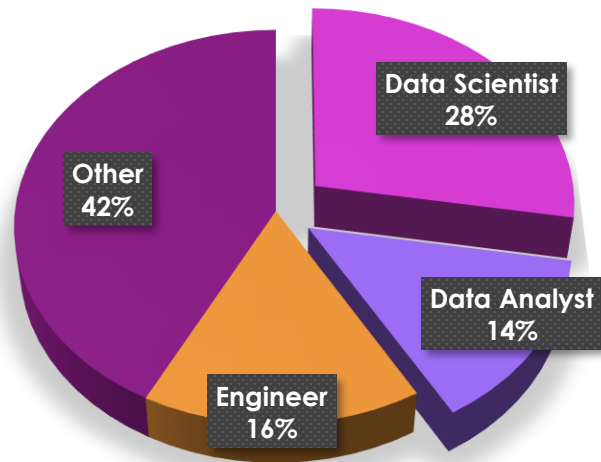
NLTK cont.

- ▶ Such as sentence segmentation, word tokenization, text lemmatization, dependency phrases, named entity recognition (NER).
- ▶ We considered the frequency of each of the requirement to find the importance of that requirement to that job as per the descriptions.
- ▶ Find the summation of that requirement for each of our category to find the importance of it for the job position.
- ▶ Later, we extracted the main requirements from the description namely skills, tools, major and degree for each of the categorized jobs.
- ▶ We consider the highly important requirements and visualize the analysed data in a simpler way using different data representations.

Findings

Share of Job Positions

Job Market

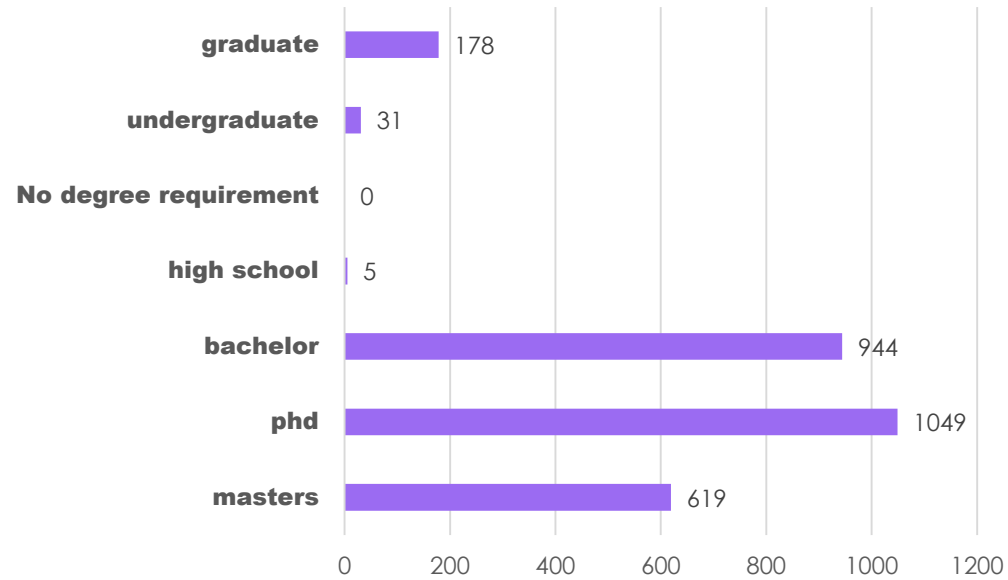


We found that employers not really classify data science jobs into data scientist and data analyst based on any criteria.

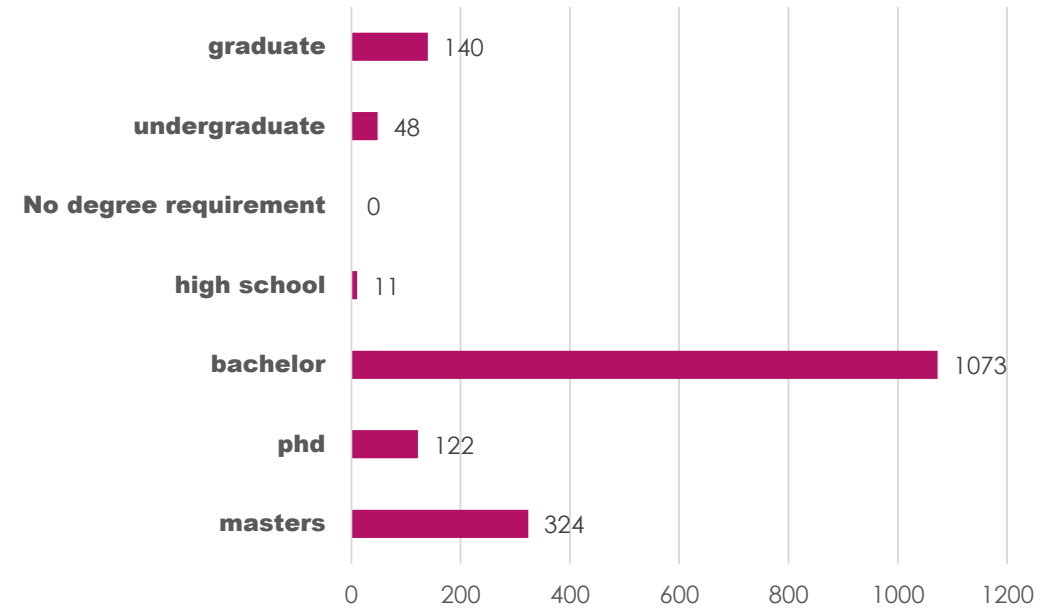
They just give a title which is more suitable for them and there is no specific criteria to divide among the two jobs.

Findings - Degree

Data Scientist

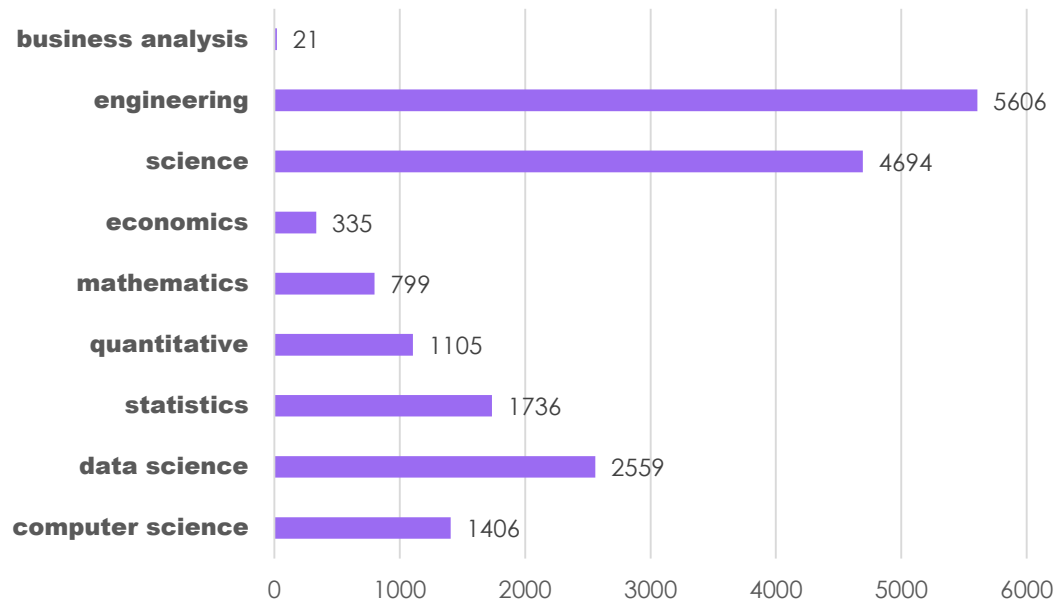


Data Analyst

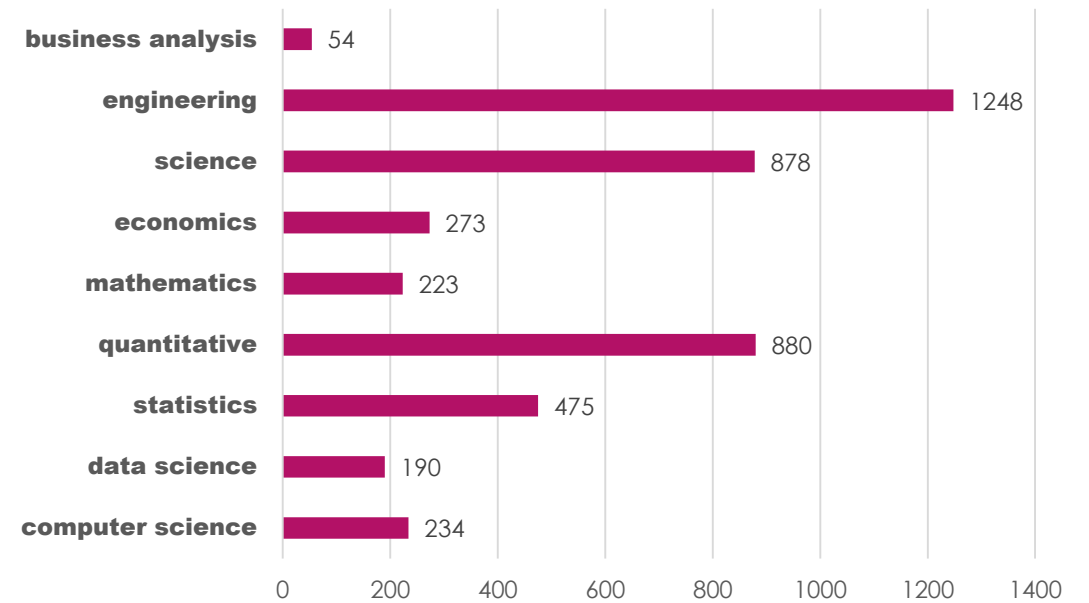


Findings - Major

Data Scientist

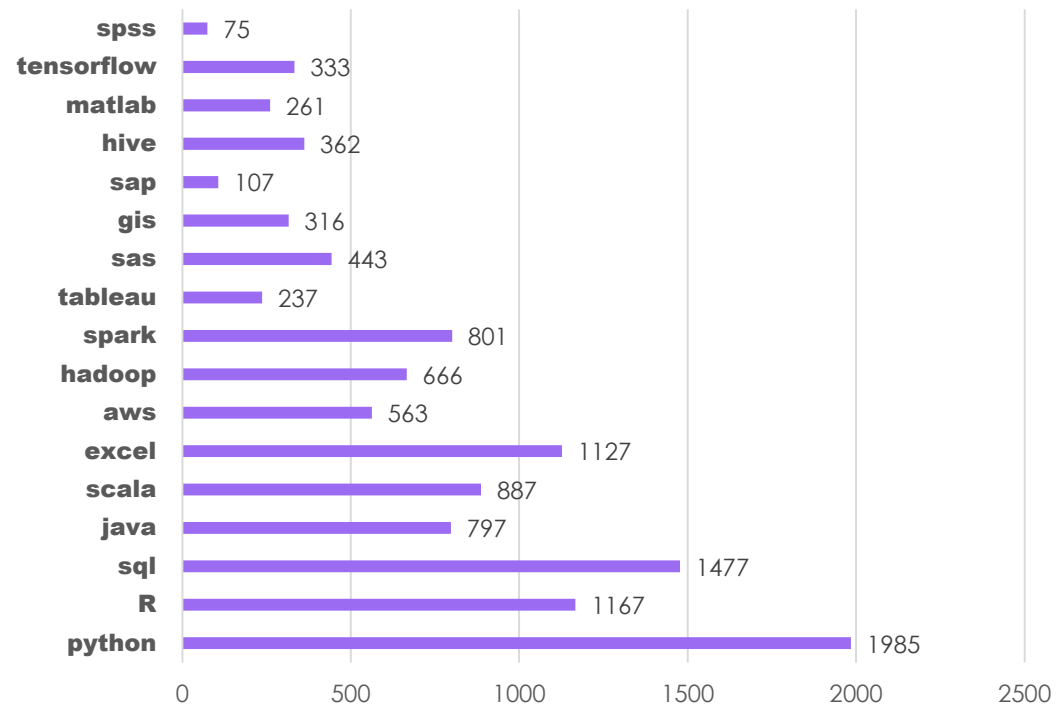


Data Analyst

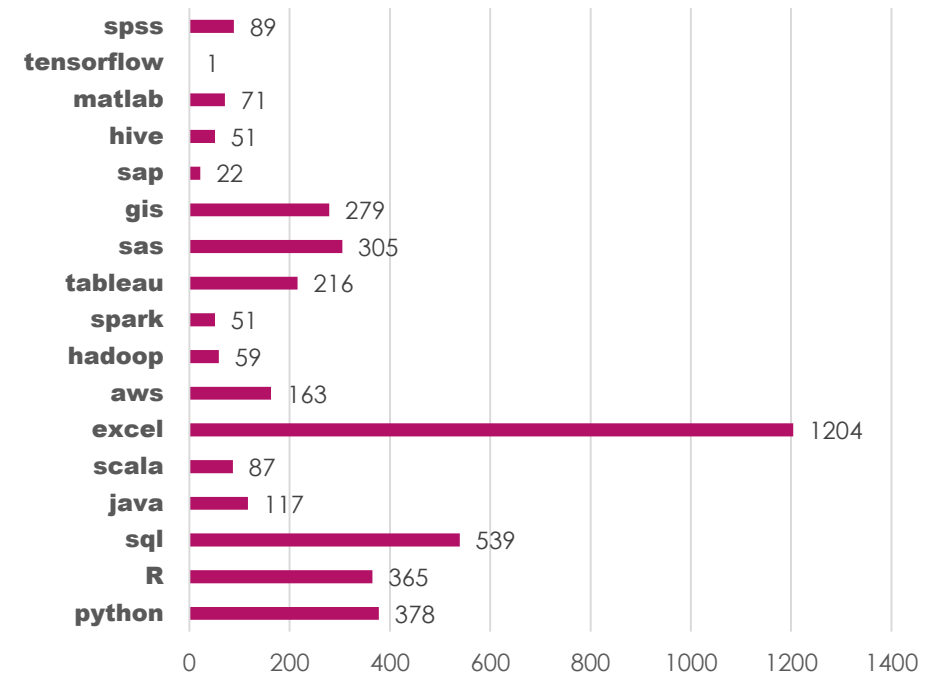


Findings - Tools

Data Scientist

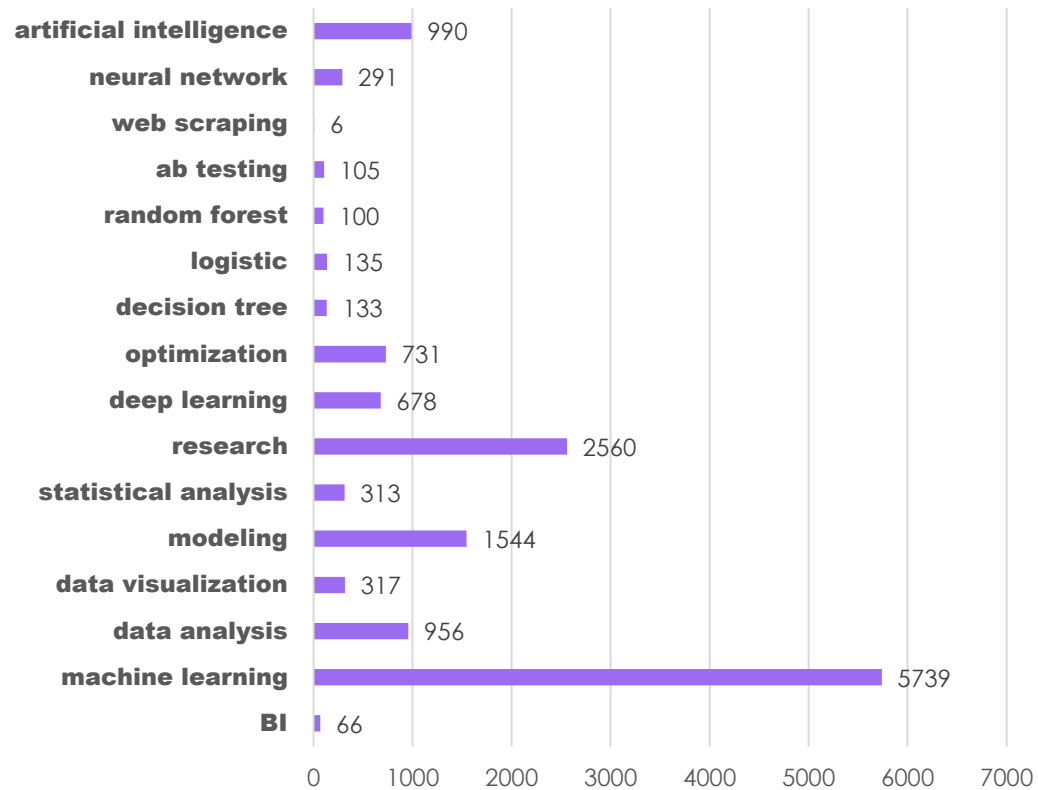


Data Analyst

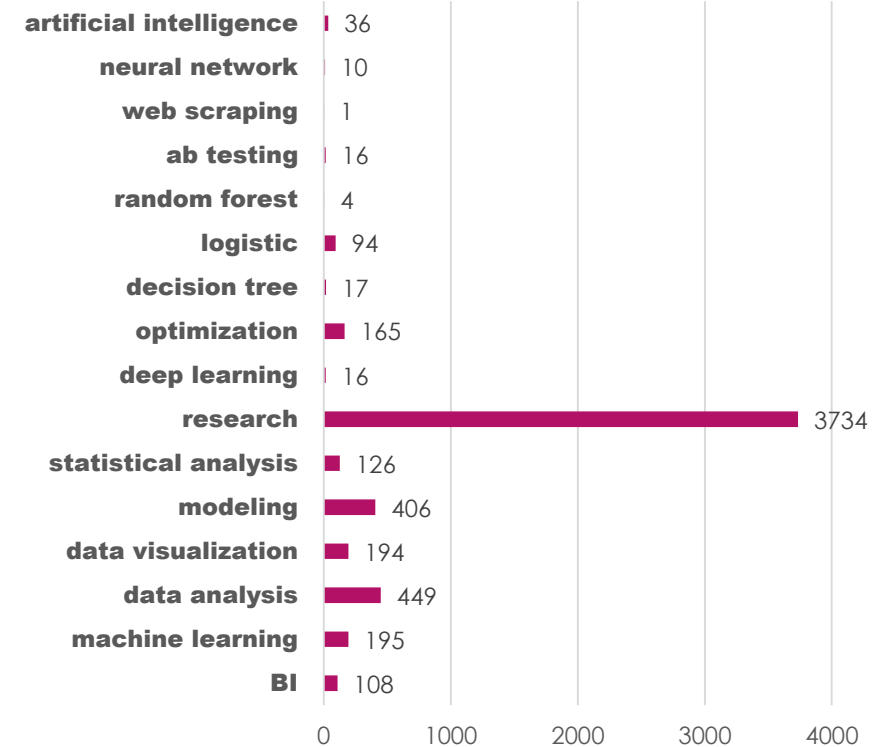


Findings - Skills

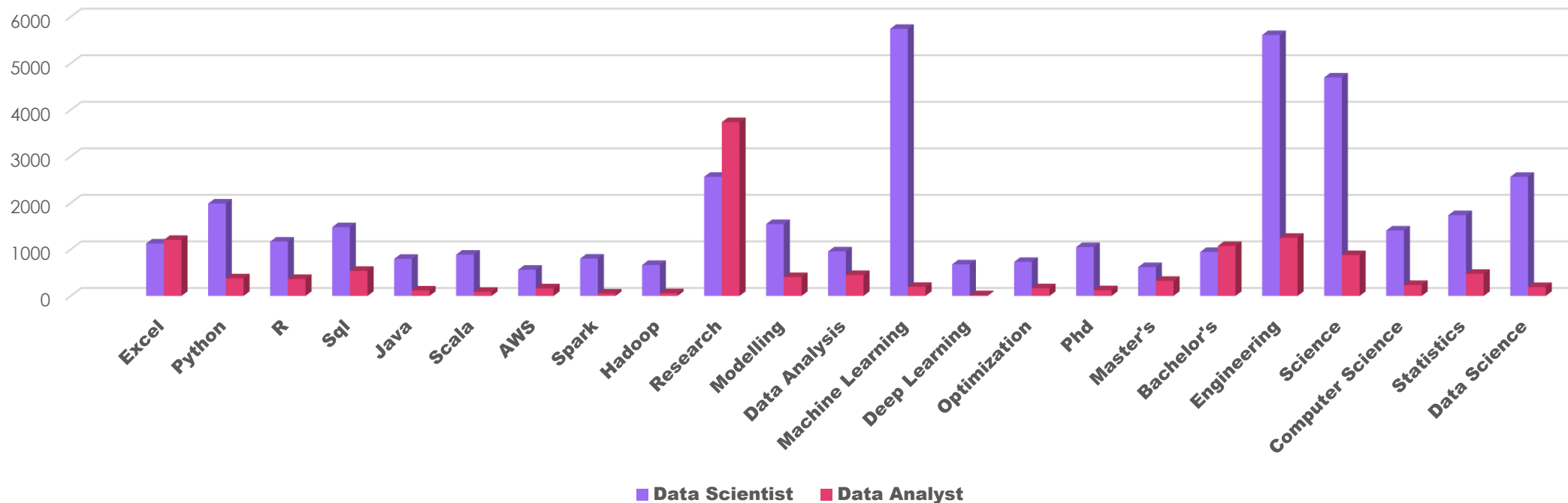
Data Scientist



Data Analyst



Data Science Job Requirements



Conclusion

- ▶ Data science jobs requires high education qualifications and knowledge in various tools and should possess several skills in different domains.
- ▶ In this analysis we have not considered experience required for the particular job position, because experience is just specified by the employer without any exact criteria.
 - ▶ But, however, experience will also be an important parameter during the process of hiring in any company.
 - ▶ Today, most of the employers are looking for hands on experience with various tools, sometimes employers doesn't care for the other requirements if the candidate have a quite good experience in the field

Data Scientist Vs Data Analyst

► ***Requirements for Data Scientist position:***

- ***Degree:*** PhD, Bachelor, Master
- ***Major:*** Engineering, Science, Data Science, Statistics
- ***Tools:*** Python, Sql, R, Excel
- ***Skills:*** Machine Learning, Research, Modelling, Artificial Intelligence

► ***Requirements for Data Analyst position:***

- ***Degree:*** Bachelor, Master
- ***Major:*** Engineering, Quantitative, Science, Statistics
- ***Tools:*** Excel, Sql, Python, R
- ***Skills:*** Research, Data Analysis, Modelling, Machine Learning

Questions?

THANK YOU...