Job Position Analysis

We analysed roughly 7,000 jobs offers related to data scraped from the web. The aim of this study is to offer insight to Data Analytics students in regards with landing a job in the data industry. For this, we accounted for the most demanded tools, skills and backgrounds that we might (hopefully) master in the bootcamp.

1 Data

We picked data scraped from the Indeed website containing information of 7,000 data scientist jobs around the U.S. on August 3rd 2018. The information that we worked on is structured in: Company Name, Position Name, Location, Job Description, and Number of Reviews of the Company (Download the dataset here).

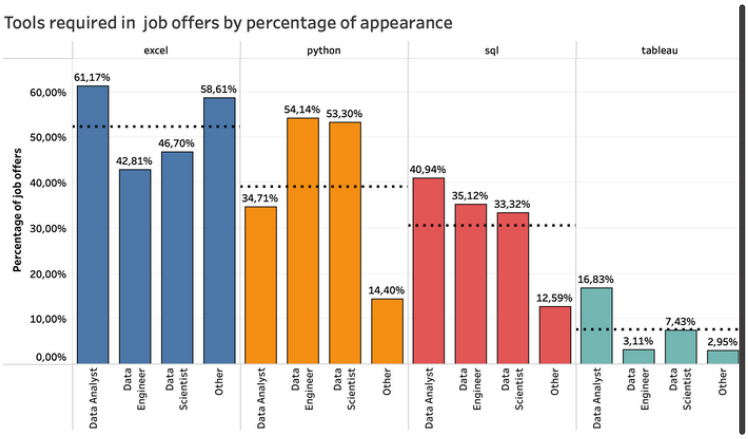
2 Data processing

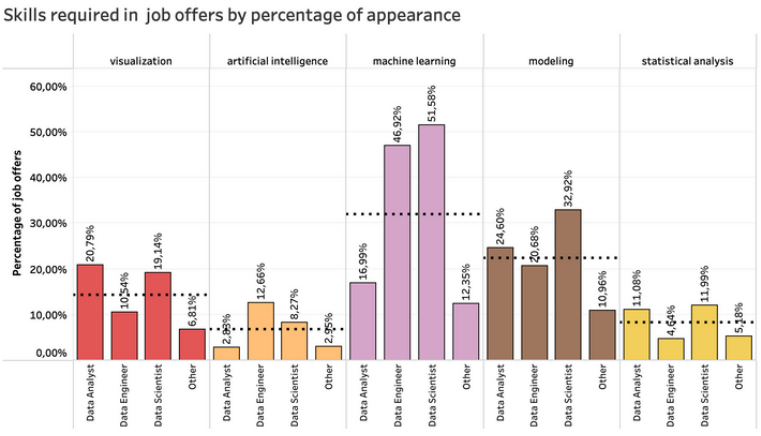
First, we did some basic cleaning and deleted the location columns since they didn't contribute to our analysis. Then, we standardized the different position names in 4 categories: [Data Analyst, Data Scientist, Data Engineer and other]. After that, we cleaned the description column so that we only considered the more relevant words for each of the positions. We did that by tokenizing the words and removing stop words. Finally, we iterated through all the rows in search of keywords from 4 different domains: tools, skills, education level and majors. Keywords can be found in the plots below. We accounted for just one appearance of the word per job description, so no keyword was overrepresented.

3 Analysis

Tools: Excel is still very popular, specially for the data analyst position, possibly because it is not as technical as the others. As for Tableau, it is most demanded for Data Analysts (the percentage would probably be higher if we take into consideration some similar expressions such as visualization tools, Power BI, etc.). Also, from the graph we can infer that our position assignation was quite effective, since the percentages for the 'other' category are significantly lower except for Excel, the least specialized tool.

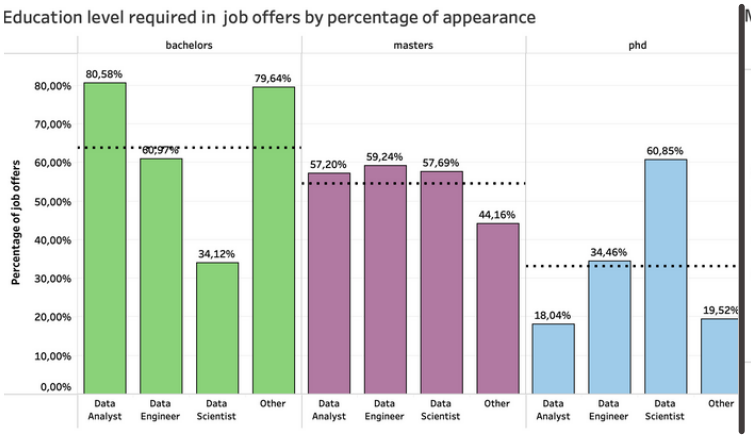
Skills: Machine learning is very demanded for data scientist and data engineers, but not so much for analyst, the same as for AI.

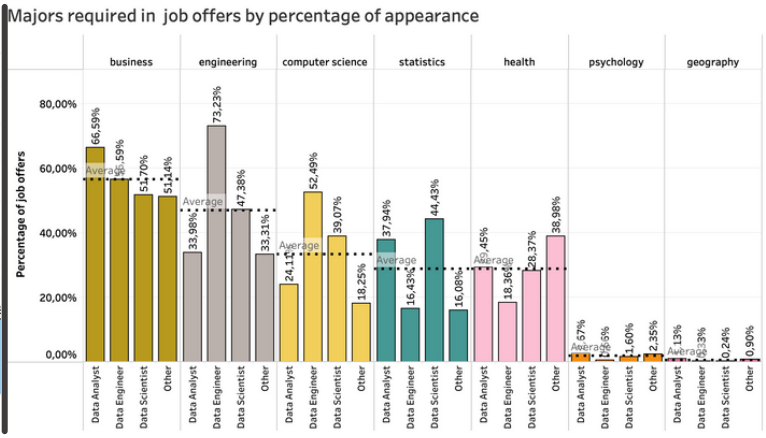




Education level: Most Data Analyst jobs require either a Bachelor's or Master's. For Data Scientist, on the other hand, we find higher education requirements: less percentage of job positions ask for Bachelor's, but a significant amount of them ask for a PhD, so we presume recruiters don't even list a Bachelor's in the job offer.

Majors: According to our results, business is the most common or demanded background. This can be biased by the popularity of the word itself. Very specialized majors (in the sense that they differ significantly from the data path), such as geography or psychology, have marginal niche demand.





Limitations of the analysis:
On one hand, we did not manage the overlapping of terms. For instance, when looking for some specific majors, such as business, we may have counted the word when it was used in another context (i.e. experience in business consultancy). On the other, there are multiple ways of mentioning the same skill or tool (i.e. visualization, tableau, power BI, presentation tools, etc.). Also, the standardization of the jobs' positions, having been done en masse, is necessarily biased, so we surely have different proportions compared to the original intention of the employers.

4 Conclusions

- Job positions overlap frequently, and roles in the data industry are not well-defined. Therefore, don't let job positions names differing from data analyst stop you from applying.
- You're okay with a bachelor's degree to access most of the jobs in terms of education level requirements.
- The most demanded tools in the industry for the data analyst role are excel and SQL.