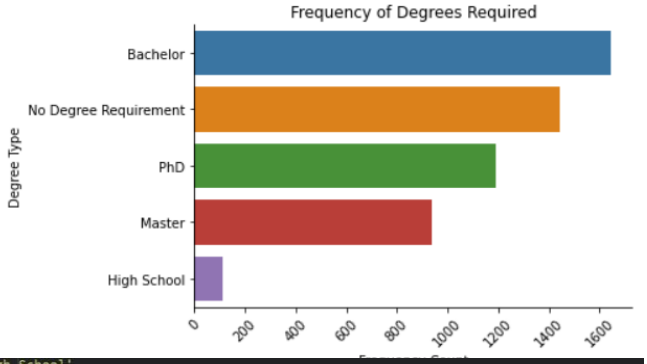
Group: Bar-Coders

- Janek, Arasan, Emmanuel, Elin, Mohit -

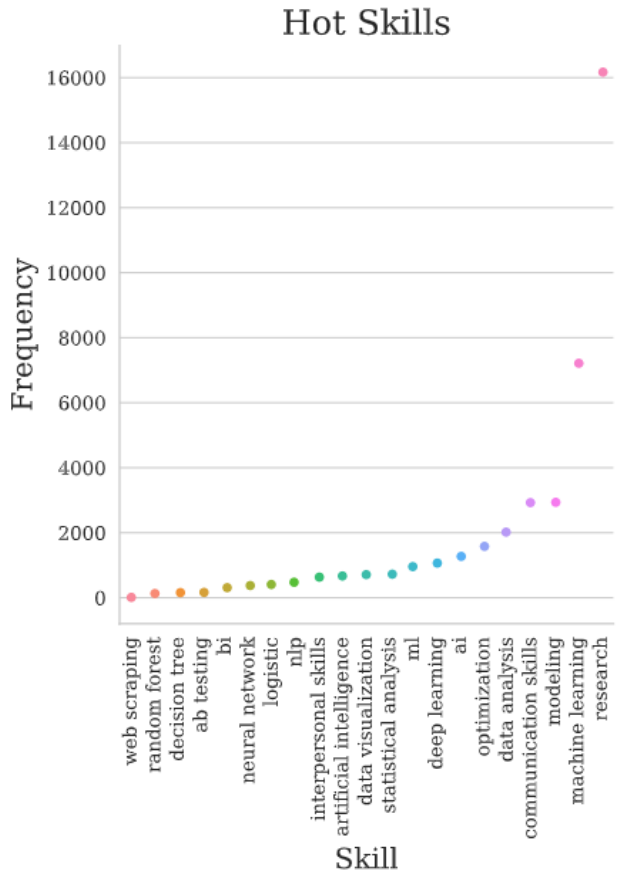
Desired Degrees

- Applicants need for over two Thirds of jobs an academic degree of some sort.
- Most employers want a candidate with at least a Bachelor degree (ca. 30% of all job listings / ca. 45 % of all academic degrees)
- A sizable amount of job listings do not require any degree (ca. 27 % of all jobs)




```
data['degree'] = np.where(data['description'].str.contains('high school'),'High School',
                 np.where(data['description'].str.contains('bachelor'),'Bachelor',
                 np.where(data['description'].str.contains("bachelor's"),'Bachelor',
                 np.where(data['description'].str.contains('master'),'Master',
                 np.where(data['description'].str.contains("master's"),'Master',
                 np.where(data['description'].str.contains('phd'),'PhD',
                 np.where(data['description'].str.contains('ph.d'),'PhD','No Degree Requirement')))))))
```

Image: Example of code to determine which degrees are desired by companies



Hot Skills for Data Science

- Most desired skills in the data were “research”, “machine learning”, “modeling”, “communication skills”, “data analysis” and “optimization”.
- The word “research” was mentioned most (over 16,000 times), which is three times the number of all job postings.
- Among many hard skills, employers also value people who have good communication and interpersonal skills.

Review of work

➔ We struggled to find easy ways to extract information out of columns with long texts without including common words such as “a”, “for” and “the”.

Limitations in the data set

➔ The structure of the scraped job descriptions was not standardized which made it hard to easily identify top skills and conduct other relevant data point frequency analysis.

Commentary

➔ Many models were made and deciding which ones were the most interesting to display was difficult.

➔ We have learnt that working with web scraped data can be challenging.

➔ The quality of raw data determines the amount of work needed to clean and wrangle the data.