Tommy Tan
Lab 4 EDA report

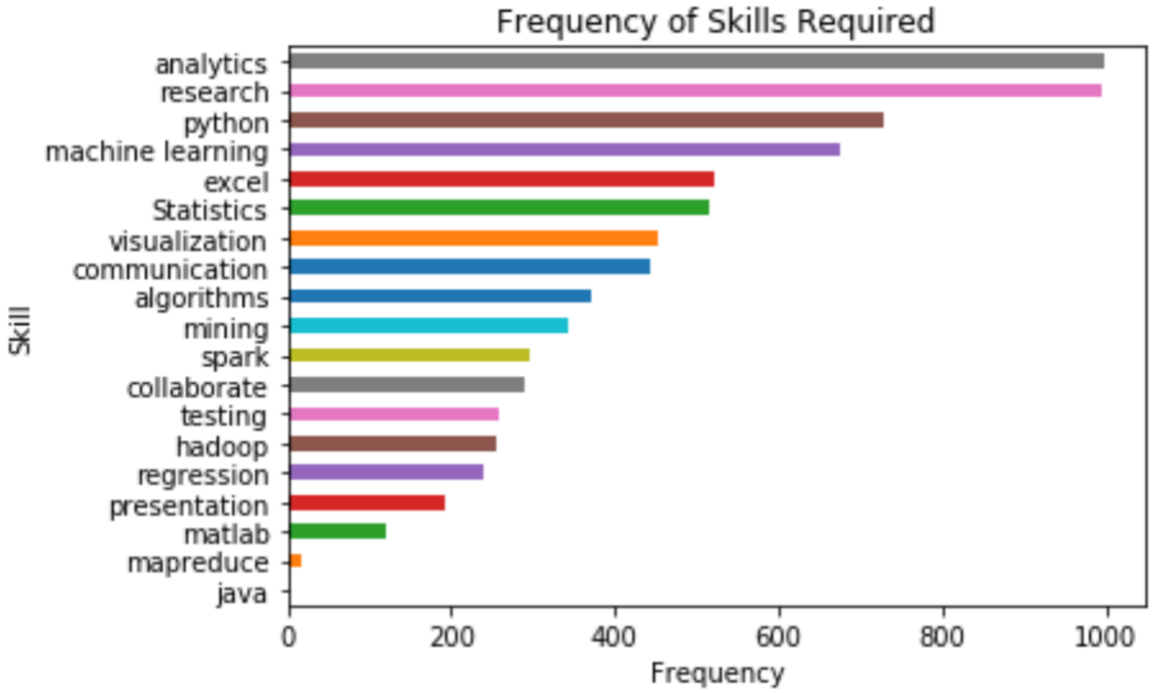


Figure 1

For this lab, we are tasked with scraping data from indeed.com that queries in jobs for data scientists. I scraped the webpages for elements for job titles, job names, job location and 20 skills that are crucial for the workforce and is specifically listed on the webpage. From figure 1, one can deduct that there are a couple of skills that are deemed 'popular' by job recruiters as per job posting. Analytics, research and python are amongst the top few candidates that recruiters are looking for when it comes to working in the data science field. The less popular and rarer skills, Java and mapreduce, are not frequently desired in many companies and thus it becomes redundant to have those skills in a resume or cover letter.

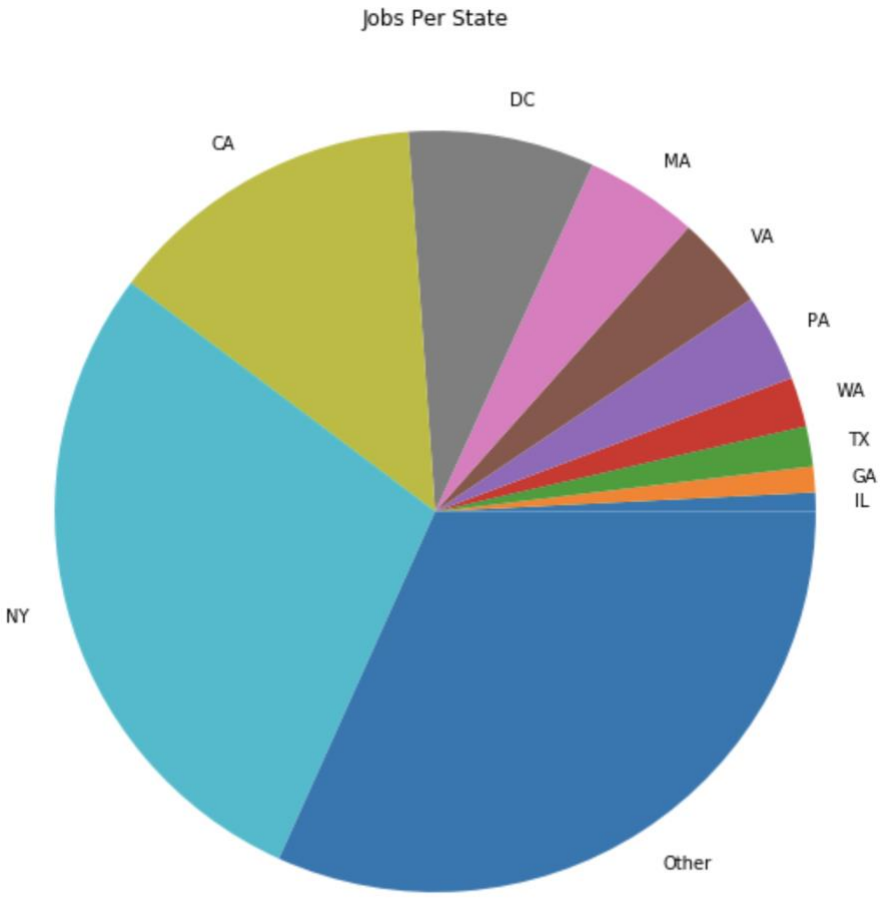


Figure 2

For this second figure, what I've done is that I've organized and sorted the locations of every job posting that is seeking for a data scientist. I've only labelled a couple of states in this chart because those states are known for the gigantic corporate entities that reside in it. One can conclude from the pie chart that NY and California make up roughly 2/3 of the job postings from indeed.com. The 'other' category are labelled as such because they aren't properly classified under the popular states; from the dataset, these 40 other states each have a small, minute slice in the pie that it isn't sensible to post each slice in the pie chart as the labels will overlap each other.