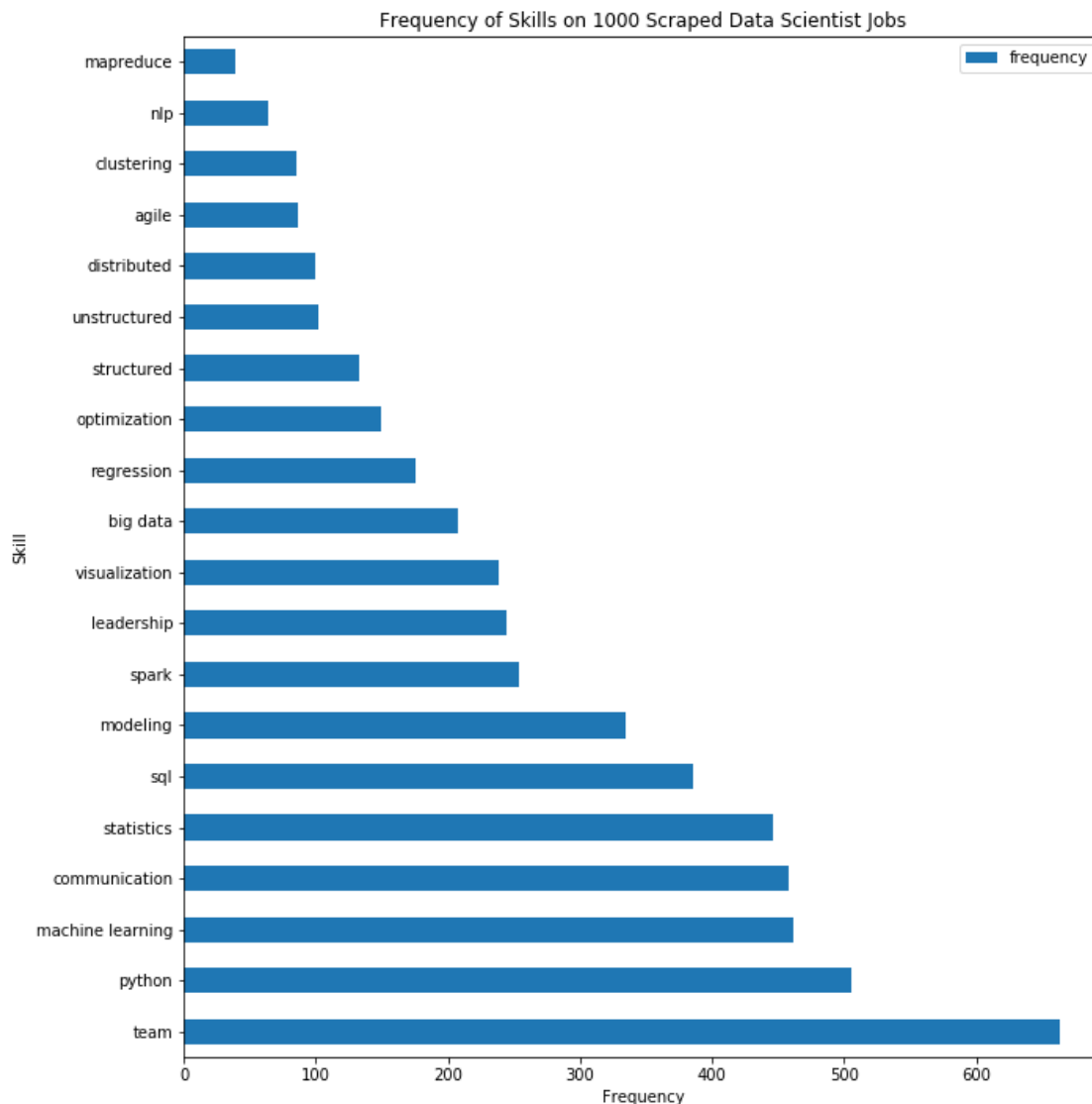


Data Science Job Posting Analysis

Exploratory Data Analysis on Scraped Data by Sam Cook

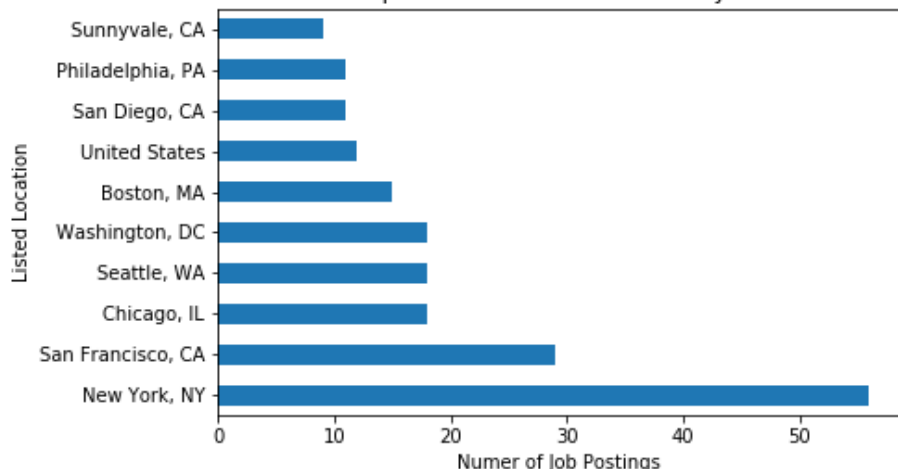
The goal of this project was to scrape a job posting website for data science jobs, parse them for keywords, and undertake some exploratory data analysis of the scraped data. I scraped Indeed.com for 1000 jobs, taking note of the role, company, location, and whether the job posting contained certain keywords. Below, you can see what the keywords are and their frequency in the job postings.



The most common keywords to appear included “team”, “communication”, “python”, “machine learning”, “statistics”, and “modeling”. These results were not surprising, but what was surprising was how low “agile” ranked—17/20—although that may be my software engineer bias.

I thought it would also be interesting to analyze the most common titles, companies, and cities where data scientist jobs were. Below, you can see the top 10 job titles, cities, and companies.

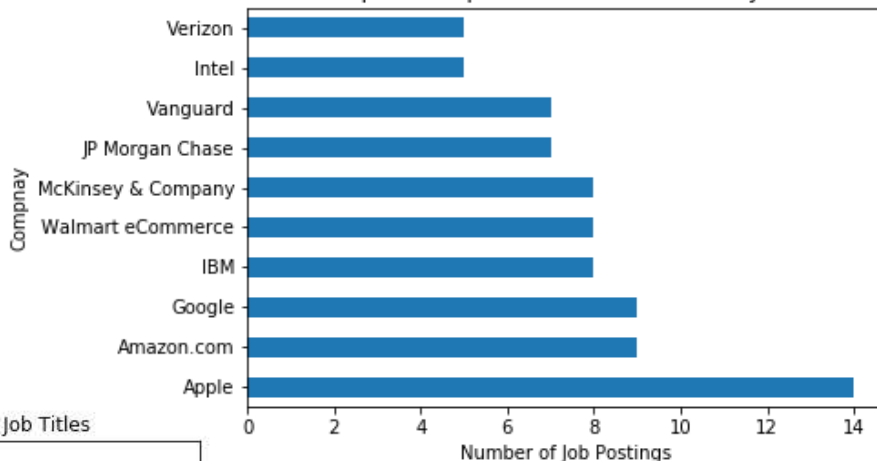
Top 10 Cities with Data Science Jobs



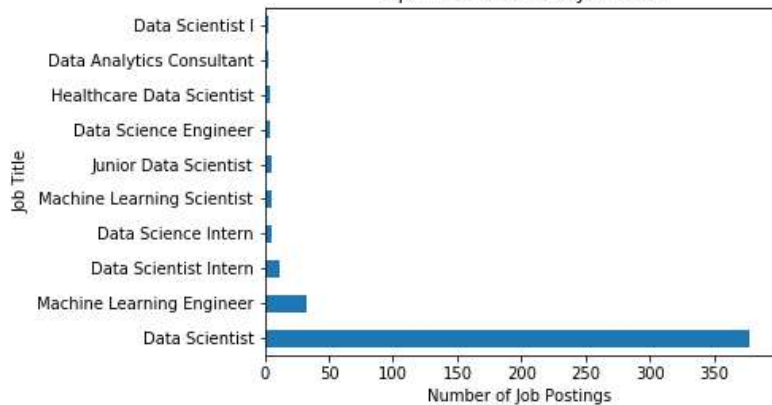
Looking at the top cities, it is slightly surprising that New York is the city with the most data science roles. My guess for this is that all the data science jobs are in New York proper, while many jobs in Silicon Valley are not in SF but rather Palo Alto, Santa Clara, etc. Support for this comes from the fact that Sunnyvale is the 10th city on the list. I know that the same situation exists for Philadelphia—many jobs are outside of the city in suburbs like Malvern, Wayne, etc.

Looking at the top companies hiring data scientists, it is unsurprisingly lead by revolutionary technology companies and large, non-tech, corporations. Many of the technology companies are leading in data science, and many non-technology companies have mass amounts of data they wish to analyze.

Top 10 Companies with Data Science Jobs



Top 10 Data Science Job Titles



Unsurprisingly, the most common job title was “Data Scientist”, by an incredibly large margin. The rest of the job titles were gradations on the Data Scientist job title, like Junior Data Scientist, or specializations, like Healthcare Data Scientist, while a few were for internships.

Apple

Philadelphia

mapreduce	1	mapreduce	0
spark	5	spark	3
visualization	4	visualization	4
python	8	python	7
clustering	0	clustering	2
regression	4	regression	3
big data	2	big data	3
unstructured	3	unstructured	3
structured	3	structured	3
distributed	2	distributed	2
sql	5	sql	4
optimization	1	optimization	4
machine learning	7	machine learning	7
statistics	6	statistics	6
nlp	2	nlp	1
team	9	team	9
modeling	3	modeling	4
leadership	2	leadership	3
communication	5	communication	9
agile	2	agile	1

Let's take a closer look at the skills that Apple, the company with the most postings, is hiring for. This is interesting, but let's compare these skills to the skills that are in-demand in Philadelphia.

There are 14 jobs from Apple and 11 jobs from Philadelphia. They are both roughly the same, but it seems that Philadelphia jobs involve more clustering, optimization, and communication.

We can analyze the skills required of the top 10 cities with data science jobs in to estimate where our skills would be most valuable.

Individual Skill Count in Top 10 Cities

	New York, NY	San Francisco, CA	Chicago, IL	Seattle, WA	Washington, DC	Boston, MA	United States	San Diego, CA	Philadelphia, PA	Sunnyvale, CA
mapreduce	5	2	0	1	0	1	2	0	0	0
spark	22	11	3	5	6	3	4	5	3	3
visualization	14	13	5	7	5	5	2	3	4	1
python	30	18	9	10	14	12	8	7	7	4
clustering	2	1	2	3	1	0	2	2	2	1
regression	7	5	4	8	3	1	2	3	3	2
big data	16	11	2	6	6	7	4	3	3	2
unstructured	7	2	2	0	4	4	2	1	3	1
structured	10	3	2	1	6	4	3	1	3	1
distributed	6	3	5	2	3	2	1	0	2	0
sql	22	20	7	7	10	8	4	3	4	4
optimization	5	8	3	3	3	4	3	2	4	4
machine learning	29	16	8	3	13	10	9	7	7	5
statistics	26	15	9	7	12	9	5	3	6	4
nlp	5	2	0	0	3	1	1	0	1	0
team	42	24	14	13	16	13	10	8	9	4
modeling	21	12	5	5	12	6	4	5	4	4
leadership	18	9	4	5	3	5	4	0	3	2
communication	33	19	11	9	8	8	6	7	9	4
agile	5	4	1	2	2	1	1	1	1	1

Looking ahead, it would be nice to be able to group jobs based on Metropolitan Area like Silicon Valley and Philadelphia as opposed to only the city. It would also be very interesting to plot the jobs on a map to get a sense of the data science hubs nationwide.