

Top Data Science Skills by Job Postings

Bridget Boakye, Wilson Ng, Neil Hodgkinson,
Marjete Vucinaj



Approach

Problem statement: what are the most valued skills in data science?

- We use labor demand as proxy for value of skills
- Indeed, for many people, a skill's value is measured by its demand in the job market which inevitably means how fast and for what salary they can get a job





**Data Source: Kaggle, Jeff
Hale, 2018**





Step 1: Data Cleaning & Analysis

- Pulled data in and creating data frames (skills = SKDF, software = SFDF)
- Ensure all data types were correct. (Chr, num, etc)
- Removed Columns & rows
- Created a “total” column
- Exported data frames out as .csv



Step 2a: Data Analysis, Software

Description: df [37 × 5]

Keyword <chr>	linkedin_proportion <dbl>	indeed_proportion <dbl>	simplyhired_proportion <dbl>	monster_proportion <dbl>
Python	0.163	0.139	0.136	0.131
R	0.117	0.113	0.113	0.122
SQL	0.100	0.096	0.097	0.095
Spark	0.056	0.056	0.055	0.055
Hadoop	0.055	0.057	0.055	0.062
Java	0.050	0.050	0.050	0.052
SAS	0.044	0.041	0.043	0.051
Tableau	0.031	0.037	0.037	0.038
Hive	0.030	0.030	0.030	0.032
Scala	0.027	0.027	0.028	0.027

1-10 of 37 rows

Previous 1 2 3 4 Next

Step 2a: Data Analysis, Software

Description: df [37 x 2]

Keyword <chr>	mentions_proportion <dbl>
Python	0.146
R	0.116
SQL	0.097
Spark	0.056
Hadoop	0.057
Java	0.050
SAS	0.044
Tableau	0.035
Hive	0.031
Scala	0.027

1-10 of 37 rows

Previous 1 2 3 4 Next



Step 2b: Data Analysis, Skills

1. Find proportion of each skill for each website

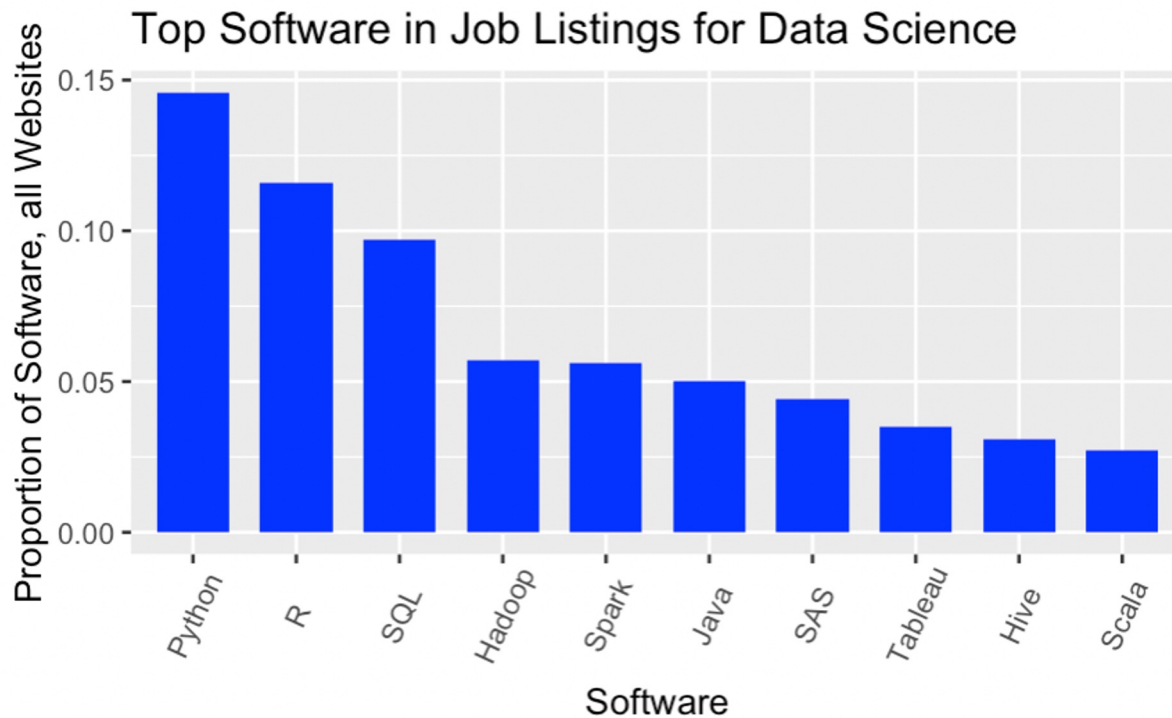
Keyword <chr>	linkedin_proportion <dbl>	indeed_proportion <dbl>
machine learning	0.163	0.148
analysis	0.147	0.151
statistics	0.140	0.129
computer science	0.129	0.118
communication	0.097	0.101
mathematics	0.074	0.085
visualization	0.054	0.061
AI composite	0.045	0.048
deep learning	0.037	0.042
NLP composite	0.035	0.039

2. Find proportion of skills across all websites

Keyword <chr>	mentions_proportion <dbl>
machine learning	0.148
analysis	0.154
statistics	0.133
computer science	0.118
communication	0.101
mathematics	0.083
visualization	0.059
AI composite	0.044
deep learning	0.038
NLP composite	0.035

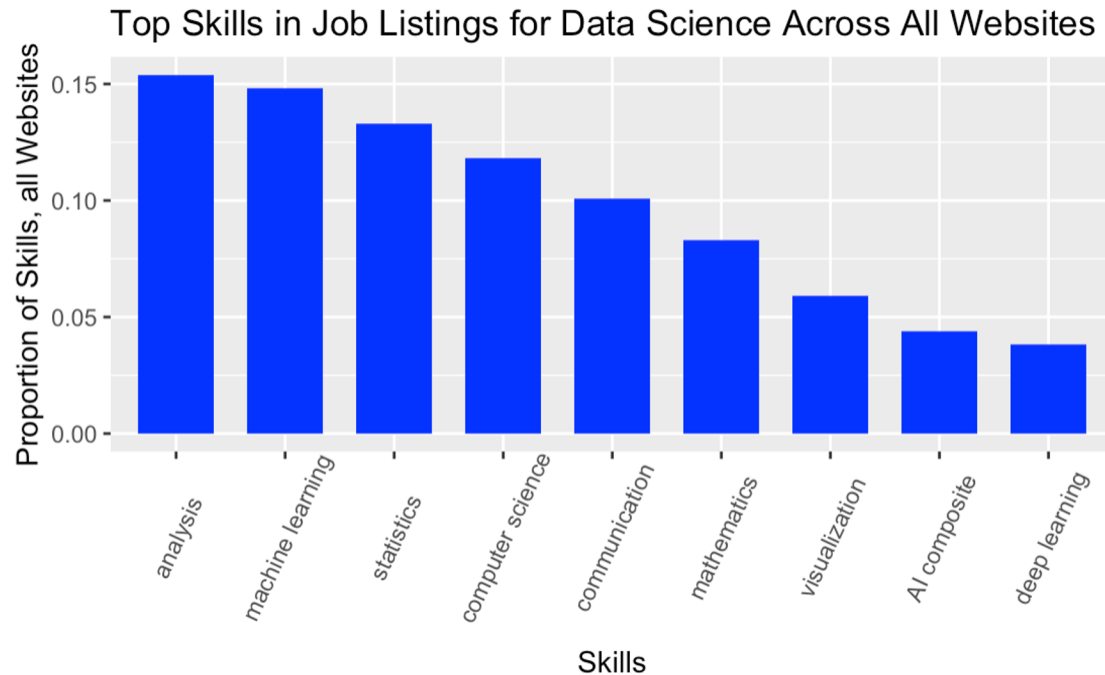


Step 3a: Visualisations, Software Across All Websites

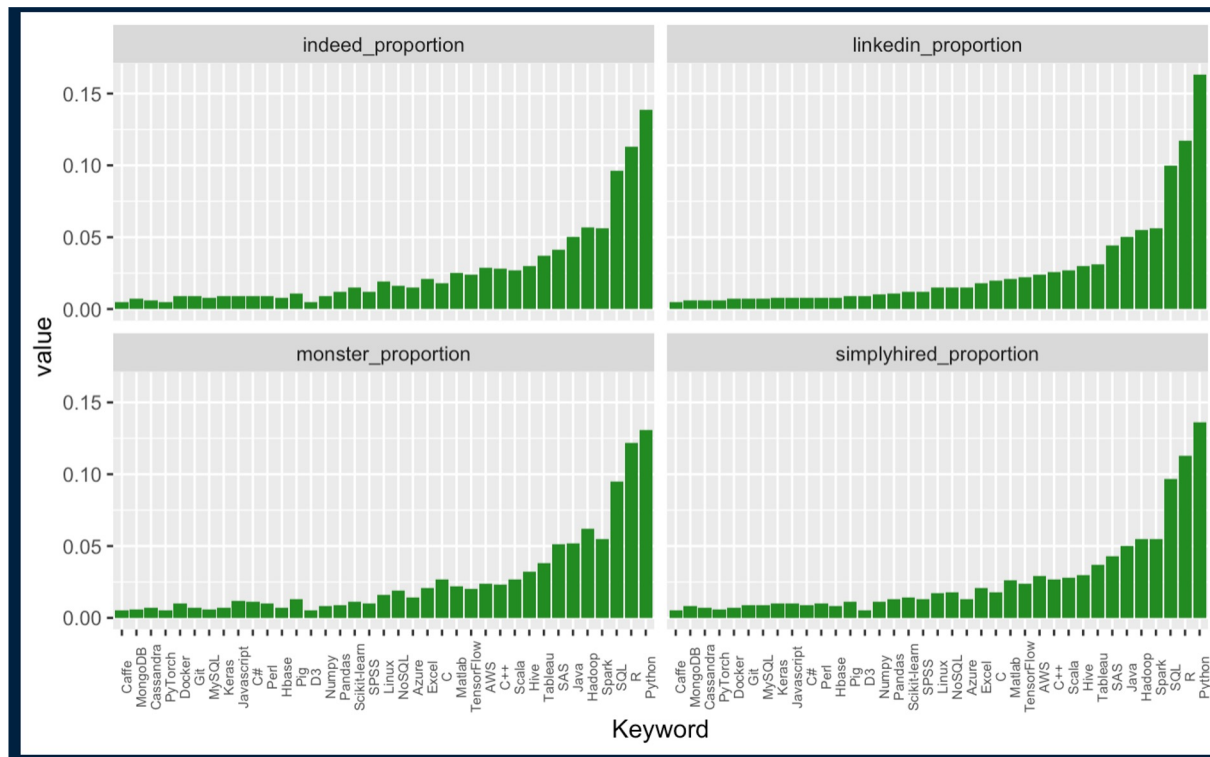




Step 3b: Visualisations, Skills Across Websites

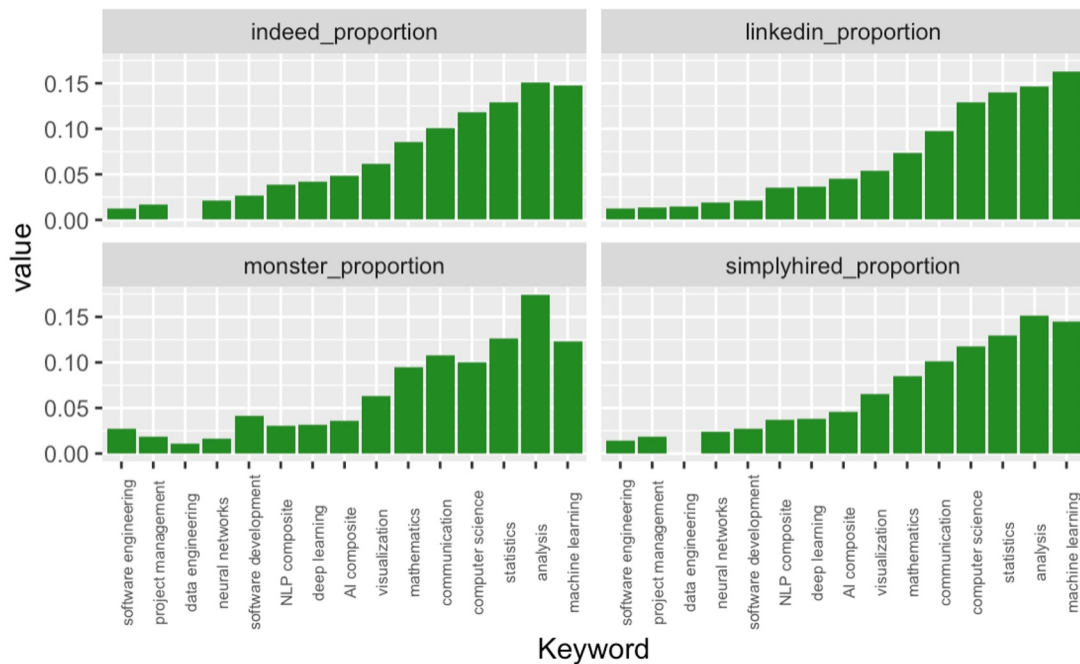


Step 3c: Visualisations, Software For Each Website





Step 3d: Visualisations, Skills For Each Website



Take-Aways:

According to employer demand, the most valued skills in data science are: Python and Analysis

However, there are some differences across sites, specifically for skills, that job seekers should pay attention to maximise their position in the job market.

Future Work

- Update dataset for comparison - Jeff's dataset is from 2018
- Map skills to salary to further assess value in market
- Look at requirements for promotion within the field of Data Science
- Investigate other websites geared towards tech roles, such as angel.co and built.in

