

Use Data Science to Boost Job Searching

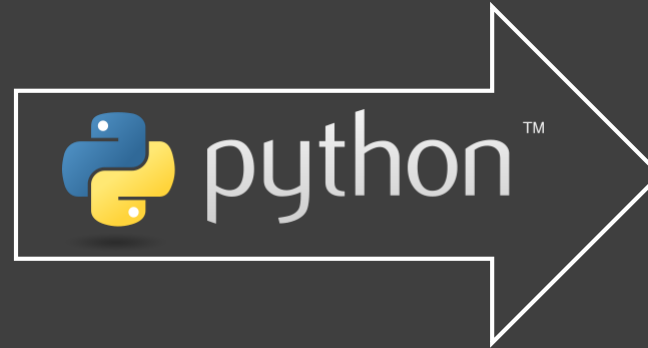
Jonathan Liu

Job Searching is A Full Time Job

- Learn about job market
 - Industries
 - Employers
 - Skills
- Resume polishing
 - Relevant Keywords



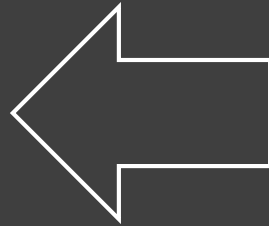
LinkedIn
Job Listing



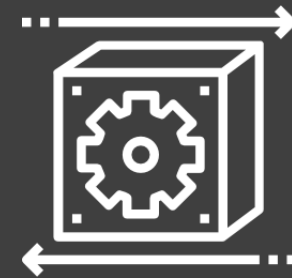
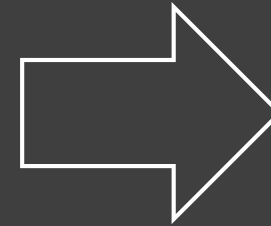
Job Description



Feature Extraction



Job Detail



function(){document.addEventListener('flushReady',function(){window.__LI_JOBS_DS__.flushModule("sha
script><code id="decoratedJobPostingModule"><!--{"decoratedJobPosting":{"jobPosting":{"listDate":1470
"],"companyName":"Quintiles","jobState":"LISTED","postalCode":"14850","description":{"rawText":"PURPO
Scientist builds on past experience, skills and personal attributes of an accomplished bioanalytical
the ability to act independently with little or no guidance. The position involves applying analytical
development and implementation of profit-generating biomarker and biotherapeutic assays. An advanced
chromatography and modern quadrupole or orbitrap mass spectrometry instrumentation, development/valid
biological sample preparation, interpreting of data, interacting with clients, and working to objectiv
to this position.\xc2\xa0RESPONSIBILITIESWork under the general guidance of a Scientific Director to
bioanalytical approaches pertaining specifically to biomarker and biotherapeutic assays for disease m
and support of drug development studiesWork with key clients to implement this service in accordance
expectationsDevelop methods including sample preparation, chromatography and mass spectrometry approa
quantitation of identified biomarker compoundsValidate associated methodologies to a fit-for-purpose
application of validated methods to sample analysisManage bioanalytical data in Watson LIMSCommunicat
progress on scientific and operational objectivesWork with management to set work plans, goals and me
servicePrepare and presents project data and supporting information. Contributes to and reviews techn
scientific meetings and client visits to promote Q2\'s LCMS Biologics businessTrain less experienced
when necessary, to support growth of LCMS Biologics serviceValidate methods to regulatory (GLP) stand
application of the validated methods to routine sample analysisOversee daily management of intra- and
relationships to ensure timely completion of projects in accordance with customer specificationProvid
of project status with clientsPrepare and make external presentations of Q2\'s LCMS Biologics capabil
prospective clientsApply analytical skills to assist in the development and implementation of profit-
:[{"attribute":[],"endIdx":2356,"type":"BR","startIdx":2356},{ "attribute":[],"endIdx":2356,"type":"BR
ute":[],"endIdx":2356,"type":"BR","startIdx":2356} {"attribute":[],"endIdx":2356,"type":"SPAN", "start
"endIdx":2257,"type":"BR", "startIdx":2257},{ "attribute":[],"endIdx":2356,"type":"LI", "startIdx":2257}
257,"type":"BR", "startIdx":2257},{ "attribute":[],"endIdx":2257,"type":"BR", "startIdx":2257},{ "attribu


```
ipt><code id="decoratedJobPostingModule"><!--{"decoratedJobPosting":{"jobPosting":{"listDate":1470  
"companyName":"Quintiles","jobState":"LISTED","postalCode":"14850","description":{"rawText":"PURPO
```

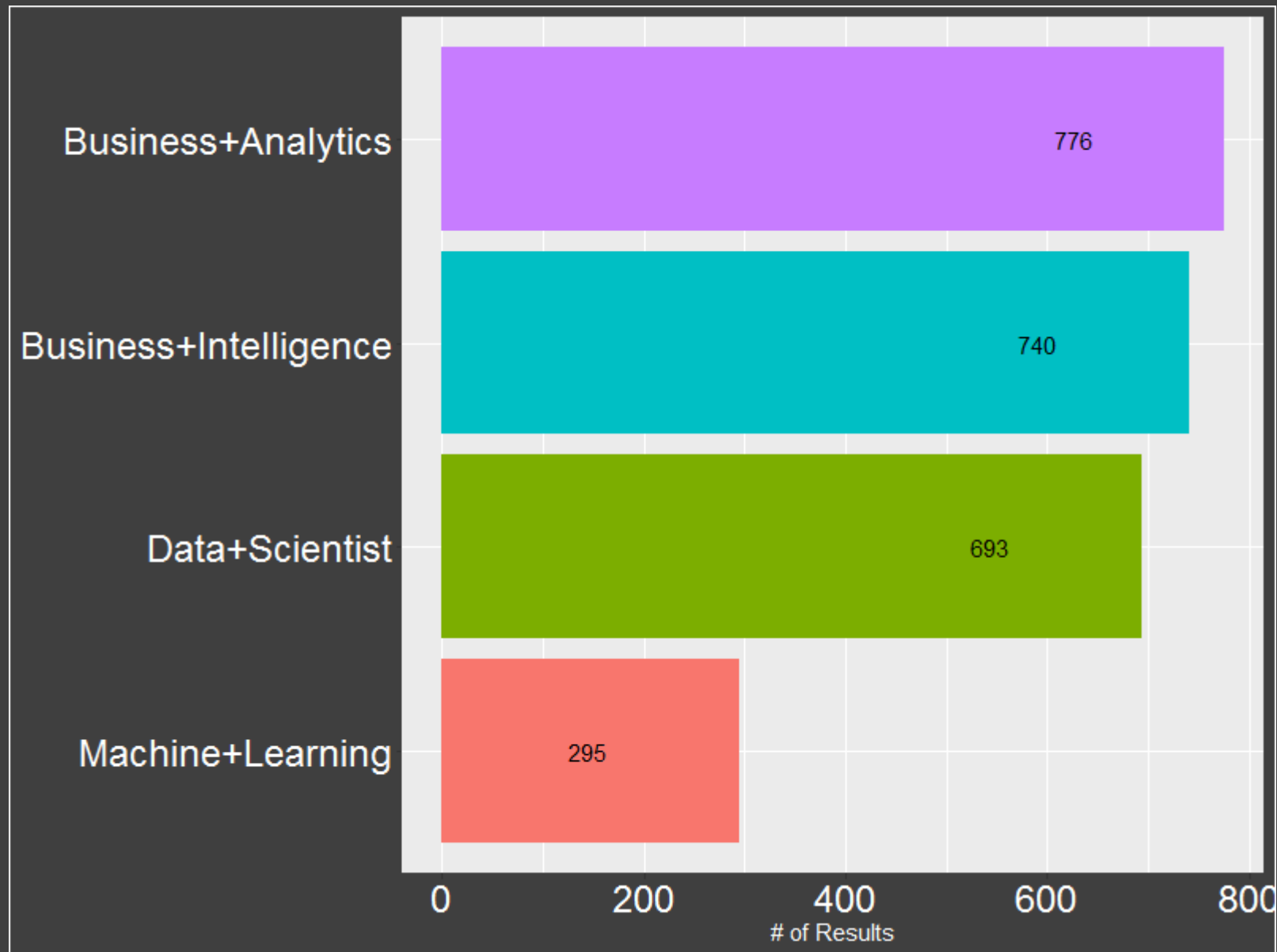
Job Function
Company Industry
Post Date
Description



Quick Insights



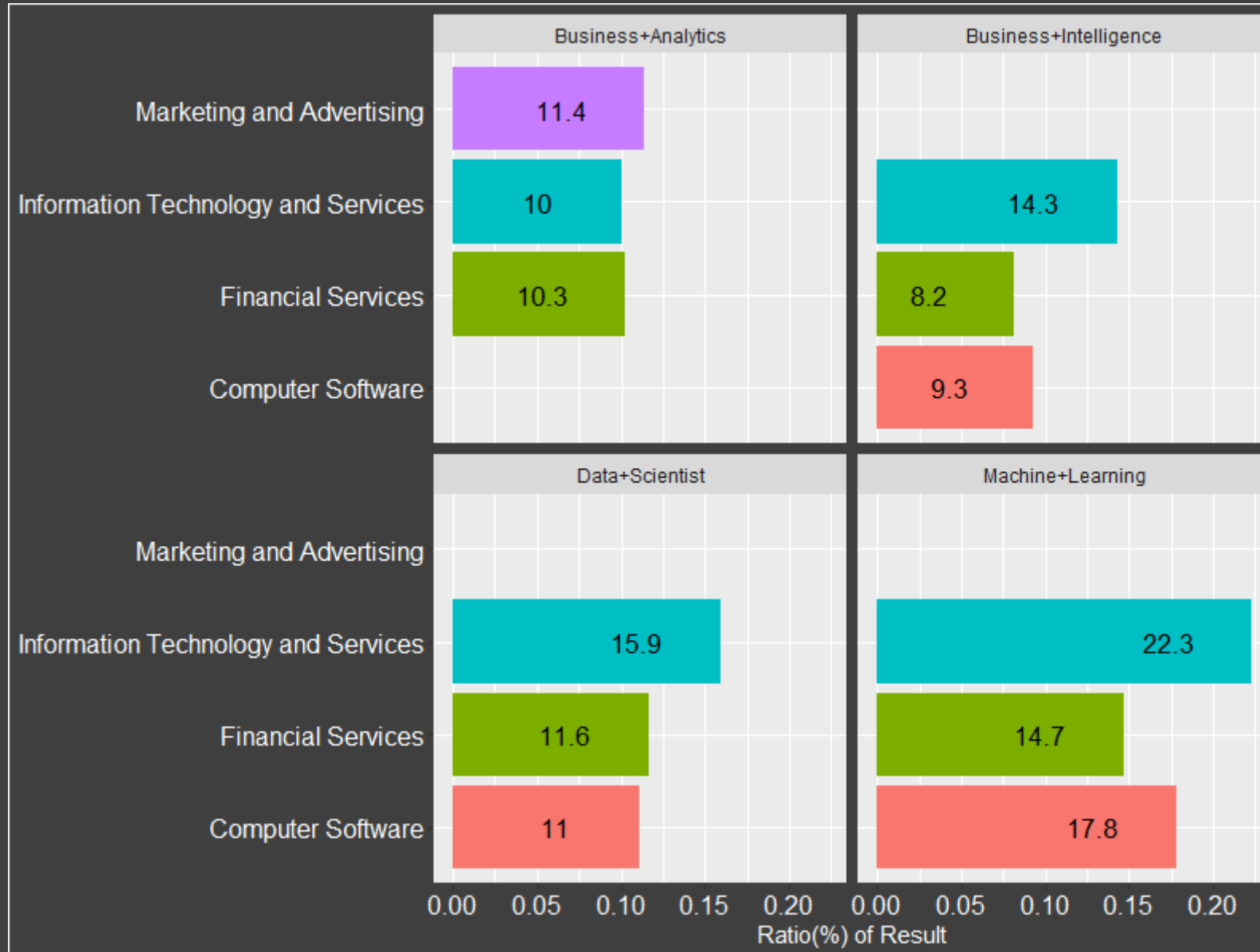
Search Results by Key Words



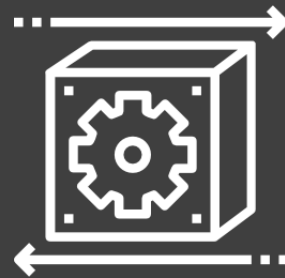
Job Functions



Top 3 Industries by Search Term



What about
skills?



Qualifications:

Technical:

- 4-6 years of big data analytics manipulating large structured or unstructured datasets for analysis
- 2+ years experience developing data science driven products in a corporate environment, ideally for a publisher / newsroom
- Proficient in scripting languages (e.g., SQL, Python, Ruby, Perl, Bash) and working with APIs, writing web services
- Abreast of the latest technologies and techniques in Data Science

```
\u003cli\\u003eStrong program  
e", "skillsDescription": null}-  
s__.flushModule("jobDescripti  
isCrawler": false. "isMobilePho
```

N-grams – Business Analytics

| | Uni | Bi | Tri |
|---|-----------------|---------------------------|----------------------------------|
| 0 | (busi, 3349) | ((year, experi), 417) | ((u002d, u002d, u002d), 146) |
| 1 | (manag, 3306) | ((commun, skill), 314) | ((equal, opportun, employ), 130) |
| 2 | (data, 3217) | ((new, york), 279) | ((without, regard, race), 111) |
| 3 | (analyt, 3028) | ((e, g), 248) | ((regard, race, color), 101) |
| 4 | (team, 2928) | ((bachelor, degree), 242) | ((race, color, religion), 93) |
| 5 | (experi, 2886) | ((cross, function), 235) | ((cross, function, team), 92) |
| 6 | (market, 2643) | ((work, close), 232) | ((sexual, orient, gender), 91) |
| 7 | (develop, 2499) | ((skill, abil), 229) | ((orient, gender, ident), 91) |
| 8 | (work, 2421) | ((project, manag), 228) | ((5, year, experi), 84) |
| 9 | (client, 2199) | ((problem, solv), 220) | ((employ, without, regard), 81) |

N-grams – Data Scientist

| | Uni | Bi | Tri |
|---|-----------------|--------------------------|---------------------------------|
| 0 | (data, 5057) | ((big, data), 454) | ((equal, opportun, employ), 93) |
| 1 | (experi, 2823) | ((year, experi), 359) | ((new, york, citi), 89) |
| 2 | (work, 2213) | ((data, scienc), 319) | ((5, year, experi), 85) |
| 3 | (team, 1938) | ((new, york), 291) | ((without, regard, race), 78) |
| 4 | (busi, 1912) | ((comput, scienc), 289) | ((degre, comput, scienc), 78) |
| 5 | (develop, 1821) | ((machin, learn), 275) | ((race, color, religion), 75) |
| 6 | (manag, 1751) | ((commun, skill), 251) | ((regard, race, color), 74) |
| 7 | (analyt, 1639) | ((e, g), 201) | ((2, year, experi), 70) |
| 8 | (skill, 1208) | ((data, scientist), 194) | ((sexual, orient, gender), 69) |
| 9 | (product, 1146) | ((experi, work), 189) | ((orient, gender, ident), 69) |

Given my skill set, what is the most relevant keyword (skill) mentioned in job descriptions?

Word2Vec

- Scan through documents, n words each time, build shallow neural networks to learn relationship between words “in the window” and “outside of the window”

```
w2v_model.most_similar( [ 'sql', 'python', 'r', 'shiny',  
                           'pandas', 'numpy', 'matplotlib',  
                           'regression', 'lasso', 'ridge',  
                           'mongodb', 'bash', 'aws', 'pca',  
                           'forest', 'knn', 'tableau', 'dataiku',  
                           ], topn=20)
```

```
[(u'vebe', 0.9707914590835571),  
 (u'spotfire', 0.960374116897583),  
 (u'basics', 0.9601113796234131),  
 (u'hibernate', 0.9597258567810059),  
 (u'pythonfamiliarity', 0.9562673568725586),  
 (u'redis', 0.9556588530540466),  
 (u'bootstrap', 0.9553016424179077),  
 (u'spring', 0.9551326036453247),  
 (u'redshift', 0.9542033672332764),  
 (u'matlab', 0.9531645774841309),  
 (u'angularjs', 0.9530525207519531),  
 (u'glacier', 0.9529001712799072),  
 (u'jenkins', 0.9517754316329956),  
 (u'similarfamiliarity', 0.9513939619064331),  
 (u'vertica', 0.9506332874298096),  
 (u'mapreduce', 0.9505549669265747),  
 (u'pig', 0.9502567052841187),  
 (u'emr', 0.9494842290878296),  
 (u'powerview', 0.9493334293365479),  
 (u'qualtrics', 0.9469818472862244)]
```


Q & A