Gaps in the Data Science Job Market: A Comprehensive Review of Two Year Data Science Education.*

Atticus Ginsborg, Sarah Song, Kehan Zhao

December 2nd, 2024

Abstract

At the 2018 Two Year College Data Science Summit hosted by the American Statistical Association Robert Gould (2018) they found 18 data science or analytics programs at the two year level. Since then there has been tremendous growth across data science industries as well as in education at the 4 year university and post-graduate levels. However, the growth of data science at the two year level has been much less documented. The authors seek to better understand the state of data science education across community college's in the United States. To do so we created the first comprehensive dataset of all community college data science and analytics programs in the United States. We found significant growth since the aforementioned 2018 report. We also analyzed the skills required by current data science job postings through LinkedIn data. Our work reaffirms the demand for a data science workforce with appropriate education and skill-set.

KEYWORDS: Data science, job market, two-year college

1 Introduction

Data Science has continued to grow rapidly over the past decade as the field emerges. As the field becomes more established, it has entrenched itself across many industries creating a new demand for data science jobs. These jobs encompass traditional analytical roles that require modern techniques, database design and engineering roles, and increasingly machine learning and artificial intelligence related roles. These jobs require a new set of skills Börner et al. (2018) including technical and soft skills. Our work will add to the previous literature by helping to understand what is required of the modern data scientist. We re-analyzed the education and skill-set requirement of the 2024 data science job market to compare and contrast with previous studies.

The growth of data science in the job market has also been reflected in academia, where there has been a tremendous development of new data science education programs at the four-year university and graduate levels. However, the level of growth at the two year college level has been much less clear Robert Gould (2018). This challenge is likely influenced, at least in part, by

*Authors: Ginsborg: University of Chicago. E-mail: aginsborg@uchicago.edu. Song: University of Chicago. E-mail: sasong@uchicago.edu. Zhao: University of Chicago. E-mail: kzhao1@uchicago.edu

the unique demands of integrating programs into community colleges, which have different demands than four-year universities. Community colleges serve a diverse range of students, including those planning to transfer to four-year institutions, individuals seeking job training to reenter the workforce, and students pursuing two-year degrees to transition directly into employment Robert Gould (2018)National Academies of Sciences and Medicine (2018). These institutions play a key role in expanding access to data science education, as they are often able to impact a more diverse student demographic compared to four-year universities Baumer and Horton (2023).

In this paper, we provide the first comprehensive review of the data science programs currently available at the two-year college level. In doing so, we established a dataset that will be made publicly available for anyone interested in understanding the state of data science education at the community college level. Additionally, we attempted to, where possible, track the type of program (certificate vs. degree) and skills covered (i.e. Excel, R, etc.) by these programs. This is particularly important for evaluating whether these programs adequately equip two-year college students with skills needed to succeed in the job market.

2 Methodology

2.1 Data Collection

2.1.1 Job Market Data

We collected 100 job postings under the "data scientist" search term in the U.S. (including full-time and internship) from LinkedIn. Duplications are removed, although one company with multiple distinct roles are included.

For each job, if the information was listed, the education/background requirement, skills requirement and responsibilities section are recorded. We then conducted natural language processing on the columns for summarization.

For the education/background column, we recorded whether the role mentioned no education, Bachelor level education, Master's level or PhD. For the skills requirement and responsibilities column, we removed common stop-words, tokenized data and created a data science specific dictionary so that specific terms could be recognized. We then conducted frequency analysis on the tokens to extract common key words, based on different n-grams.

For topic extraction, we used Latent Dirichlet Allocation to find the key themes of skills required and role responsibilities. For the skills analysis, we aimed to extract the most common technical (hard) and

soft skills required. For the responsibilities columns, we extracted top 5 key tasks data scientists are expected to perform.

2.1.2 Community college data

To create a dataset[1], we began with Wikipedia's list of all community colleges in the United States Wikipedia contributors (2024), compiling a list for each state. For each college, we recorded the city (choosing the main campus if specified) and state. Where available, we also collected information on the name of the program, whether the college offered a data analytics degree or certificate, whether the college offered a data science degree or certificate, whether the college offered an intro data science class, the department the program was housed in, the year the program started, and key languages taught by the program[2].

We chose these criteria based on the data collected and analyzed for data science programs in 2018 Robert Gould (2018). To collect this information, we Googled each college alongside the keywords "data science" or "data analytics". If these returned a relevant hit, we recorded the information in the database. Additionally, we would search the college's catalog for "data science" or "data analytics"[3]. Programs were excluded if they were massive open online courses (MOOCs) being advertised by the community college, or lacked at least two of the following main subject areas—statistics, computation and data management, mathematics, communications—subjects identified as core to data science in Robert Gould (2018).

2.2 Analysis

We used the collected data to create a data story exploring trends in the shifting data science job and education landscape.[4]

2.2.1 Key Job Market Findings

We found that in accordance to 2018 data Börner et al. (2018), Python, R and SQL are still prominent, appearing in 40% of job listings. However, demand for machine learning (59% of listings), deep learning (19%), and AI (14%) has grown significantly. Cloud computing and big data tools like AWS and Spark also increased in importance.

Soft skills focused on communication, teamwork and stakeholder management, similar to 2018. Topic modeling of responsibilities identified 4 main themes: 1. Build-

[1]Here is a link to the full dataset with longitude and latitude added

[2]We did this based off descriptions in the catalog. We did not infer anything, only recording programming languages explicitly mentioned, which has likely led to an under count. In particular languages like Python and R or Excel were often implied, but unclear

[3]This would miss some AI programs and searching for artificial intelligence could improve the dataset. This oversight was realized too late in the process to rectify it but should be an area of further research

[4]The data story can be found here

ing machine learning models, 2. Insight discovery, 3. Cross-team collaboration, and 4. Reporting and delivering impact. We also found that over 60% of postings mention a university degree, 49% mentioned Master's level study and 30% mentioned PhD. However, it is worth noting that for some postings, education level is not a compulsory requirement and can be substituted with equivalent years of working experience.

2.2.2 Community College Landscape

We analyzed 973 institutions, including two-year colleges and four-year colleges with significant two-year programs. [5] There was at least one community college in every state analyzed. Information on the number of community colleges is summarized by program type in table 1, and by state in table 2. Across the nation, we found a clear expansion in data science educational offerings at the community college level, but they remain unevenly distributed across the country. Figure A1 shows that some states (particularly small ones) offer data science programs at all colleges, while other lack programs entirely.

Additionally, we looked at the computational skills being taught by two-year college programs. As seen in figure A2, the most common skills required are SQL, Python and R, which aligns with what was found in our analysis of data science jobs on LinkedIn. This suggests that while four-year and graduate programs may develop additional intrinsic skills, companies focused on specific technical skills can increasingly find qualified candidates from community colleges.

[5] For large college systems like CUNY we treated each campus as unique colleges

3 Conclusion

From a refreshed view of job market data analysis, education institutions should prioritize offering cutting-edge courses in machine learning, cloud computing and big data to meet growing job market demands. Soft skills development such as stakeholder management and team work could should also be fostered through leadership courses or cross-discipline projects. While more research is needed to better understand how community colleges should adapt their offerings to this shifting data science landscape, progress is evident, with programs increasing from 18 nationwide in 2018 to 177 today.

References

Baumer, B. S. and Horton, N. J.: 2023, Data science transfer pathways from associate's to bachelor's programs, *Harvard Data Science Review* **5**(1).
URL: *http://dx.doi.org/10.1162/99608f92.e2720e81*

Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., Wu, L. and Evans, J. A.: 2018, Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy, *Proceedings of the National Academy of Sciences* **115**(50), 12630–12637.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.1804247115*

National Academies of Sciences, E. and Medicine: 2018, *Data Science for Undergraduates: Opportunities and Options*, The National Academic Press.
URL: *https://doi.org/10.17226/25104*

Robert Gould, Roxy Peck, J. H. N. H. B. K. K. K. J. M.-S. M. R. B. T. M. D. W. R. W.: 2018, The two-year college data science summit, *Technical report*, National Science Foundation.

Wikipedia contributors: 2024, List of community colleges — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=List_of_community_colleges&oldid=1260018413. [Online; accessed 2-December-2024].

4 APPENDIX

Tables

Data Analytics degree/cert Offered	Data Science degree/cert Offered	Offers Certificate	Offers Degree	Number of Programs
False	False	False	False	777
False	True	False	True	25
False	True	True	False	14
False	True	True	True	18
True	False	False	False	2
True	False	False	True	19
True	False	True	False	60
True	False	True	True	30
True	True	False	True	6
True	True	True	False	3
True	True	True	True	19
Total				973

Table 1: Number of Degree and Certificate Programs offered

State	Number of Schools Offering Certificates	Number of Schools Offering Degree	Number of Programs Offering an Intro DS Course	Total Number of Colleges
Alabama	1	0	0	25
Alaska	1	0	0	12
Arizona	8	3	0	22
Arkansas	0	3	4	22
California	35	17	43	115
Colorado	2	0	3	15
Connecticut	4	7	5	12
Delaware	0	0	0	1
Florida	4	6	2	28
Georgia	3	0	0	23
Hawaii	1	0	3	7
Idaho	0	0	0	4
Illinois	1	0	4	48
Indiana	5	6	0	20
Iowa	3	0	1	18
Kansas	2	0	3	20
Kentucky	0	0	0	16
Louisiana	0	0	0	12
Maine	0	0	0	7
Maryland	8	9	8	16
Massachusetts	3	3	3	15
Michigan	4	2	3	29
Minnesota	3	5	3	31
Mississippi	1	1	0	15
Missouri	1	1	0	14
Montana	0	0	0	17
Nebraska	1	1	1	8
Nevada	1	1	1	3
New Hampshire	3	1	1	7
New Jersey	3	7	8	18
New Mexico	1	0	1	12
New York	4	1	3	35
North Carolina	6	9	5	58
North Dakota	0	1	1	10
Ohio	5	4	4	22
Oklahoma	0	1	1	12
Oregon	0	1	1	17
Pennsylvania	3	3	3	14
Rhode Island	1	0	0	1
South Carolina	5	1	1	16
South Dakota	0	1	0	6
Tennessee	0	0	0	13
Texas	4	6	3	60
Utah	0	0	1	3
Vermont	1	0	0	1
Virginia	2	1	0	23
Washington	6	7	5	33
West Virginia	0	0	0	10
Wisconsin	7	7	0	18
Wyoming	0	1	0	8

Table 2: Number of Programs Offering Intro Data Science Courses, Related Certificates and Degrees in Various States

5 Figures

Figure A1: Chloropleth Map of Data Science Program Offerings

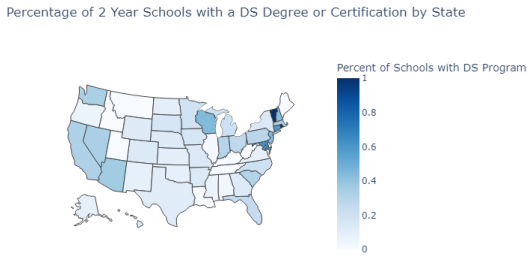

Figure A2: Bar Chart of Skills in Data Science Program Offerings