Job openings in Data Science

(COMP3125 Individual Project)

Ryan Kennedy   
*Wentworth Istitute of Technology – School of Computing and Data Science*

*Abstract*—This research paper will analyze and review date pertaining to job openings in the data science field. By examining data from various sources, we will draw logical conclusions on if the number of job openings are increasing or decreasing, desired skills in these jobs, and where in the world has the highest number of openings.

Keywords—BLS (Bureau of Labor Statistics), Data Science, Python

# Introduction

Data Science is a vast field that, at its core, compiles and dissects great amounts of data and transforms it into comprehensive graphs or diagrams. In common conversations, it is often described as an ever-growing field that has hundreds of thousands of openings just in the US alone [6]. However, it is also fairly common to see people argue that the field of data science is oversaturated and even a bit of a dead-end. This feels like a contradiction, so in order to come to a more concrete answer I will be compiling data in order to see if data science is as much of a growing field as people claim. This is especially prevalent to myself and my peers, as many of us will likely be searching for an entry level data analyst position after college.

# Datasets

## Sources of data

For this analysis, I have compiled data from a variety of websites. Most notably, I located this data from Statista[12], the US Bureau of Statistics (BLS)[6], and a LinkedIn article relevant to this subject[10]. Statista is a well known and trusted source of information that is often used by data analysts as a reliable source of information. The BLS is also considered a reliable source of information that is managed by the government. As for the LinkedIn article, it sites multiple sources to back up its findings, leading me to believe that it is reliable as well. Not only that, but LinkedIn is a trusted website in networking and employment, which ties directly into the topic of this paper.

To compile this dataset, I used Excel to store and categorize each dataset in a clear spreadsheet. This dataset compiles information from three different sources (Medium, 365dataScience, Statista) to compare their findings on popular skills in the Data Science field. I needed to clean this data by making sure that Python was able to recognize when a source was missing a piece of information. In Excel, I had placed an N/A. This data was able to compile without issue once I had python recognize the N/A symbols and simply leave that portion blank. Below is a recreation of this dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| Skills/Sources | Statista | 365 | Medium |
| Python | 76.13% | 56.7% | 78.9% |
| R | 57.92% | 33% | 78.9% |
| SQL | 58.87% | 30.4% | 55.3% |
| Unix | 48.55% | N/A | 21.1% |
| AWS | 38.17% | 19.7% | N/A |
| C/C++ | 24.31% | .6% | 73.5% |

# Methodology

## Statista analysis and plotting

By exploring the website Statista and examining relevant articles and data sets, I was able to find a reasonable amount of data on trends in technology for data science skills. By comparing these findings to other sources such as Medium[3] and 365DataScience[5], we can conclude which of the following skills is most sought after. For this portion, I utilized Python, NumPy, pandas, seaborn, and matplotlib to craft a functional graph. This methodology, while effective in displaying trends, does suffer from the drawback of sample size. As there is only so much data on this, the graph and data size for it are small.

## Plotting US Bureau of Statistics data

The US Bureau of Statistics is a famous government run website that collects, calculates, and analyzes data on certain jobs and their positions. By analyzing and plotting this data, while also comparing it to another source[8], we can begin to have a better understanding of trends in job openings in Data Science. Not unlike the Statista analysis, I utilized Python, NumPy, pandas, seaborn, and matplotlib to craft a functional graph. This methodology, while effective in displaying trends, does suffer from the drawback of sample size. As there is only so much data on this, the graph and data size for it are small.

# Results

## Growth of Data Science

A graph with a line going up

Description automatically generated

Figure : Graph demonstrating the growth in Data Science as a whole.

By compiling the data found on BLS and comparing it with other sources such as LinkedIn, this graph was able to be produced. It seems as though Data Science truly is a growing field, with a projected 26% growth rate between 2023 and 2033.

## Sought after skills in Data Science

By examining various websites (such as Statista and Medium) and compiling their data into one excel sheet, I was able to produce a graph detailing some of the most sought-after skills in Data Science.

A graph of different colored rectangular shapes

Description automatically generated

Figure : graph of various Data Science skills and applications.

Throughout various sites, Python was sited as one of the most integral tools an upcoming data analyst should have in their repertoire. This is mostly due to its flexibility and overall capabilities when it comes to managing data.

## Increases and Decreases in Data Science

While difficult to display with the gathered data and graph that have been created, Statista provided insight into which specific fields in Data Science are excelling more than others[11]. Namely, data management and storage is increasing exponentially, while things such as cloud storage, although still rising, is rising much slower.

## Where in the world are the most openings?

While similarly difficult to C in displaying this in graph, sites such as LinkedIn are able to provide insight on where the most job openings in Data Science are [1]. There are some expected answers, such as the United States, with some more surprising results, such as Switzerland being the number one location in the world for Data Science positions.

## More Job openings in bigger or smaller companies?

## Despite extensive reseach into answering this question, I have not been able to find a conclusive answer to this question. As such, I am unable to form any kind of graph or examine any data. While there is a general consensus that there are more openings in larger companies than in smaller, there is no real data provided to back this claim up [7]. It can be assumed that this conclusion is derived from the general rule that smaller startup companies (companies with 150 people or less) don’t really have a need for a data analyst yet.

# Discussion

Despite some initial success in locating sources for this endeavor, it became increasingly clear that this topic was not as well documented as I had once thought. While there are indeed sources (BLS, LinkedIn, Statista) that touch or hint at how data science as a career is growing, they do not provide clear indications of where this information was gathered from or how it was concluded. More crucially, none of these sites provided any way to compare current growth of the field to previous years, allowing for little way to demonstrate progression over time. This process only became more tricky once I lost reliable access to Statista (many datasets became unavailable to me off campus), making it more difficult to search for relevant data. If I was to do this again, I would need to expand my search to data science trends as opposed to job openings specifically. Similarly, I would need to double check that some of these questions are more feasible to answer (ex: making sure I can truly answer questions like letter E, comparing larger companies to smaller startups).

# Conclusion

Data Science, as a field, is a vast and ever-growing field. Claims that it is constantly growing and expanding seem to be backed up by the data. If myself or one of my peers are looking to break into this field, it seems integral that we make sure we have a solid understanding of programming languages such as python and R, to ensure we have the best chances possible for breaking into the field.

##### References

1. Academy, Walter Shields Data. “The Top 5 Countries Offering the Highest Pay to Data Scientists.” *LinkedIn*, 13 Aug. 2024, www.linkedin.com/pulse/top-5-countries-offering-highest-pay-data-9cove/.
2. *Amsterdam Tech*, amsterdam.tech/6-most-sought-after-data-science-skills-to-pick-up/. Accessed 8 Dec. 2024.
3. AnalytixLabs. “Data Science Skills Survey Report 2024 - A Comprehensive Market Study of Most in-Demand Skills.” *Medium*, Medium, 3 Apr. 2024, medium.com/@byanalytixlabs/data-science-skills-survey-report-2024-a-comprehensive-market-study-of-most-in-demand-skills-3d9baf772272#:~:text=Top%20Data%20Science%20Tools%20to,important%20when%20hiring%20data%20scientists.
4. “Computer and Information Research Scientists.” *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, 29 Aug. 2024, www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm#tab-1.
5. “Data Scientist Job Market 2024: Analysis, Trends, Opportunities.” *365 Data Science*, 10 Oct. 2024, 365datascience.com/career-advice/data-scientist-job-market/.
6. “Data Scientists.” *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, 29 Aug. 2024, www.bls.gov/ooh/math/data-scientists.htm.
7. Dragova, Niya. “Data Science Careers - Does Company Size Matter?” *Candor*, Candor, 12 Apr. 2020, candor.co/articles/career-paths/data-science-careers-does-company-size-matter.
8. “National Employment Trends: 15-2051.00 - Data Scientists.” *O\*NET OnLine*, www.onetonline.org/link/localtrends/15-2051.00?st=. Accessed 8 Dec. 2024.
9. Published by                                    Statista Research Department, and Aug 8. “Top 20 Fields with the Highest Number of Internship Positions in the U.S. 2016.” *Statista*, 8 Aug. 2024, www.statista.com/statistics/673316/top-fields-for-internship-positions/.
10. Share      LinkedIn      Facebook      Twitter. “November Workforce Report 2024.” *November Workforce Report 2024*, economicgraph.linkedin.com/resources/linkedin-workforce-report-november-2024. Accessed 8 Dec. 2024.
11. Sherif, Ahmed. “Data Scientists Skill Proficiency Worldwide 2024.” *Statista*, 3 Sept. 2024, www.statista.com/statistics/1490020/average-skill-proficiency-of-data-scientists/.
12. Sherif, Ahmed. “Tech Companies Increasing Technical Jobs Worldwide 2023.” *Statista*, 6 June 2024, www.statista.com/statistics/1418121/tech-companies-increasing-technical-position-globally/.