Exploring the world of Data Science and the related jobs in the filed

Brought to you by the team of:

Hao, Yan, Escoffrey

SMU DATA SCIENCE BOOTCAMP FEB2019

July 27. 2019

**Background**

Over the last six months we immense ourselves into the Universe of BIG Data aka “Data” as we journey into the world of Data Science on the Python Shuttle.

Our mission: to gain an in-depth understanding and increase knowledge of Data Science

Acquiring the skills and mastering the techniques to extract, transform and load our data

Performing various analyses to obtain valuable information, knowledge to provide recommendations and using Machine Learning to make predictions.

Give life to our results through dynamic visualization.

Now we are ready to launch our careers in Data Science, but, first we have to perform a job search, apply for the job, demonstrate our knowledge and skills in a convincing way that our choice company will offer us the job.

**Introduction**

Searching for a job can be a daunting task and giving that we are new to the field of Data Science will not make it any easier. As such we have embarked on our mission to explore and determine the state of the US Data Science job market as our final project.

Data Science is fairly new and growing profession that offers exciting careers, but what are the job options, what are the duties, what are the requirements and valued skills, how much does these jobs pay, where are the jobs and which companies are hiring.

As such, we decided to tackle these questions to assist us and other Data Science enthusiasts in performing a data driven job search.

**Data Source**

We obtained the following data file from the following:

Kaggle.com: Data Scientist Jobs in the US as scraped from Indeed website (August 2018) and downloaded as a csv file. The file contains Company Name, Job Title, Location, Job Description and Number of Company Reviews.

Kaggle.com: 2018 Survey results of Data Science jobs in the US download as a csv file. This file includes information such as Salary, Education, Gender, Experience, Employment Status and Job Satisfaction.

**Methods and Dependencies**

The downloaded csv data files were read into pandas using Juypter Notebook to perform data transformation

Python was used for writing scripts for data manipulation, analysis, and visualization.

JavaScript library and CSS were used for making custom interactive web-based data visualizations

**Questions and Results of Analysis**

**What is Data Science?**

Data science is a multi-disciplinary field that combines several disciplines, including statistics, data analysis, machine learning, and computer science. Data science uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data

What are the top job titles in the field of Data Science?

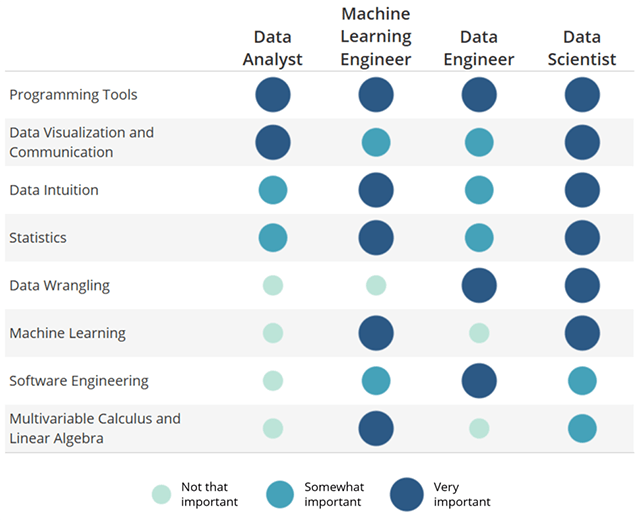
Data Scientist: An analytical data professional with a high degree of technical skill and knowledge, usually with expertise in programming languages such as R and Python. Data scientists help businesses collect, compile, interpret, format, model, make predictions about, and manipulate all kinds of data in all manner of ways. They’re experts at both construction and deconstruction of data.

Data Analyst: An interpreter of data who typically specializes in identifying trends. They’re similar to data scientists, sans the coding experience. One way to think about data analysts is that they’re junior data scientists on their way to becoming full-fledged data scientists.

Data Engineer: Anyone who designs, QAs, and maintains the systems that data scientists employ daily. Whereas a data scientist might be focused on data analysis, a data engineer focuses more on data preparedness.

Machine Learning Engineer: A data scientist does the statistical analysis required to determine which machine learning approach to use, then they model the algorithm and prototype it for testing. At that point, a machine learning engineer takes the prototyped model and makes it work in a production environment at scale. A machine learning engineer isn’t necessarily expected to understand the predictive models and their underlying mathematics the way a data scientist is. A machine learning engineer is, however, expected to master the software tools that make these models usable.

We used the scrapped data of job descriptions to determine the top 4 data science job titles:



How many data science jobs are available in the market?

There 2,213 job science job vacancies

Which are the most in demand jobs?

What qualification do you need?

What skill sets employers most desires when hiring?

Python, SQL and R are the most widely valued skills for data related jobs, followed by Java, Hadoop, Spark and Excel.

How much do data science jobs pay?

Where are these jobs located in the US?

Which are the leading industries for Data Science job?

Which are the top companies for Data Science jobs?

Based on the number of employed and the number of related job openings.

Observations and Future Trends

The future belongs to the companies that figure out how to collect and decipher data successfully. Google, Amazon, Facebook, Netflix, and LinkedIn have all tapped their data networks and made that the core of their success. Now small businesses are following their path whether it’s mining social media data, recommending products based on a user’s purchase history, or studying the URLs

The current trends in Data science which will continue into the future include self-service analytics, crowdsourced, open-source tools, data security regulations, and more organizational structure around data science career growth.

**Challenges and Limitations**

Data Science job descriptions often used a blanket job title such as “Data Scientist” to describe jobs that are drastically different and requires different data skill sets.

Difficulty in obtaining one data set with all the relevant criteria and inconsistencies in the data type, style and format as a result of the different sources from which the data was obtained.

Obtaining the

Close API

**Opportunities for Future Analysis and Improvements**

We could use a machine learning module where a job searcher input their skill sets and have return a list of recommend jobs to apply for?

Based on selected criteria: location, salary, company size, industry, etc. will we be able to predict job satisfaction rating?

Clean the descriptions to particularly outline which responsibilities and qualifications are required for each position. This would help to remove company related descriptions and focus on what actually is required of each position.

Compare job posting from sites such as Glassdoor, Indeed and Linkedin

**Conclusion**

The current and future demands for Data Science professionals remain very due to the explosive growth in the [volume of data](https://hub.packtpub.com/highest-paying-data-science-jobs-2017/) and its’ impact in every business sector. This is driving up the number of data science professionals required to help companies uncover the insights they need to stay competitive. These jobs often prove to be challenging, takes an average of five days longer to close the market average, causing employers to pay premium data science salaries for qualified professionals

Data science is changing the way we work and opening up new frontiers leading to new fields of prediction and data visualization through artificial intelligence and machine learning. Although in the future data science will be increasingly automated as a result, the need for human judgment to interpret the data will not disappear any time soon given the number of job vacancies. As such, the data science skills will continue to pay huge dividends in our increasingly data-driven world.

Data science is, indeed, one of the most important things to understand in the coming years.

The limitation and increase processing time of H1B visas.

Based on the results of our analysis, Data

Per the 2019 Salary Guide release by Robert Half, 43%

**References**

<https://www.springboard.com/blog/data-science-salaries/>

<https://data.bls.gov/oes/#/home>

<https://www.springboard.com/blog/data-science-terms/>