

A Data Science Approach to Defining a Data Scientist

Andy Ho¹, An Nguyen¹, Jodi Pafford¹, and Dr. Robert Slater¹

Master of Science in Data Science, Southern Methodist University, Dallas TX 75275
USA {atho,angyuen2, jpafford, rslater}@smu.edu

Abstract. In this paper, we apply Data Science techniques to solve the problem of not having a common definition and list of skillsets for a Data Scientist. As it relates to the field of data science, adoption has spanned nearly all industries and disciplines - an increase in the necessity for a wide range of technical skills, soft skills, experiences, and education to work on applications within data science (i.e. Artificial Intelligence, Machine Learning, Big Data Analytics, Data Visualization, etc.). The result is an overlap and ambiguity of various roles such as data scientist, data engineer, data analyst, software engineer, database administrator, and statistician. To solve the problem, we collect over 8,000 job postings from Indeed.com for the six job titles to form the data set. Each corpus contains text on job qualifications, skills, responsibilities, educational preferences, and requirements. Natural Language Processing (NLP) techniques and Universal Sentence Encoder (USE) cleans/transforms the data for clustering, classification and analysis - using advance Machine Learning (ML). The result is the key findings on the definition and list of skillsets of a Data Scientist. A secondary finding provides confirmation that differentiating Data Scientist job postings from the other five job titles was not decisively clear. The conclusion and common definition: A Data Scientist codes, collaborates, and communicates – transforming data into insights using statistics, analytics, and machine learning.

1 Introduction

Over 2.5 quintillion bytes of data is created every single day and the pace is expected to accelerate with the growth of the Internet of Things (IoT), a society more dependent on data, and more businesses making data-driven decisions. On any given day, 500 million tweets are sent, 294 billion emails delivered, 4 petabytes of data created on Facebook, and IoT products such as driverless cars, wearables, and smart cities will push the daily amount collected to 463 exabytes. This expansion of data has created the need for individuals (i.e. Data Scientists) who are trained, skilled, and educated in the field (and application) of Data Science.

There is no common definition and a list of skillsets for a Data Scientist. This becomes evident when the ‘What is’ or ‘Who is’ a Data Scientist is in question or when a job search (on multiple job sites) returns job title postings different

from the search term. According to Dictionary.com, a data scientist is a person employed to analyze and interpret complex digital data, such as the usage of statistics of a website, especially in order to assist a business in its decision-making. But we postulate that there is more to a data scientist, and a common definition along with a list of skills is within reach using data science techniques.

To apply data science to this problem, our approach is scrap job postings from a job site to form the data set, pre-process the data, explore the data, then transform the data into high-dimensional vectors to cluster, classify, and analyze. The application of NLP, USE, and ML techniques led to the end result of visualizations and explanations of the findings.

Since there is no public data sets available for this problem, we web scraped job postings listed under Data Scientist and other similar roles from multiple job sites to form the training data set. Features such as job title, job description/summary, experience and skill requirements, and location was collected. NLP was performed to read, decipher, understand, and make sense of each job posting. Unsupervised ML within NLP was performed to find document and text similarities between each job posting. Multiple comparative analytics provided additional results and insights to arrive at a common definition and list of skillsets of a data scientist.

The main conclusions are the definition and list of skills of data scientist varies from one job posting to another. We observed that multiple job sites have different algorithms presenting roles that seem like a data scientist but in fact an entirely different role. In applying data science techniques to this problem, the following are a few key findings:

1. Data Scientists are most associated with Data Analyst and Statistician
2. Data Scientists at minimum have programming/coding expertise in Python or R
3. Data Scientists often collaborate with other data scientists in a team environment
4. Data Scientists must be able to communicate their approach, findings, and insights
5. Data Scientists have a background or knowledge of statistics, analytics, and machine learning

In addition to the Abstract and Introduction section, this paper is organized in the following manner: Tutorial Sections to educate readers on the general principles and techniques used in NLP, ML, USE for this research; a Data Set section explaining how data was collected, features selected and visualization from the data exploration analysis; a Methods section on the data science techniques and algorithms performed; a Related Work section discussion other studies already completed; a Results section listing and explaining our experiment; an Analysis section on the results; an Ethics section on the possible ethical ramification of our findings; and a Conclusions section listings multiple conclusions from our analysis.

2 Artificial Intelligence Explained

To understand the data science techniques applied to our research, we want to first provide a baseline understanding for Artificial Intelligence (AI). This term has been associated to robots functioning on their own or a global network of computers rising up against mankind. Though that may be the case in the distant future, present AI is the development of advanced software programs to perform tasks that would normally require human intelligence. Alan Turing's paper established the fundamental goal and vision for AI [7]. Much has developed since Turing's endeavor to simulate human intelligence in machines Table 1 shows a few examples of applied AI today:

Table 1.

1. Translation between languages (i.e. Google Translate)
2. Facial Recognition (i.e. Auto Tag on Facebook)
3. Virtual Assistant (i.e. Amazon Alexa)

The data science techniques applied will be in the realm of Narrow AI where machine learning and natural language processing is used to perform tasks in data collection, analysis, and modeling.

2.1 The Importance of Machine Learning to AI

In short, ML is when computers learn from the data it is provided. Without ML, AI could not provide insights, decisions, and/or actions for the examples of applied AI. There are many ML algorithms, all of which can be stratified into 3 classifications, described in Table 2.

Table 2. Machine Learning Classifications.

Classification	Description
Supervised Learning (SL)	Labeled input/output data is fed into an algorithm multiple times to arrive at a pattern for prediction. Algorithm examples could be Linear/Logistic Regression, K-Nearest Neighbor, Naïve Bayes, Decision Tree.
Unsupervised Learning (UL)	Labeled input is fed into an algorithm multiple times to form clusters for the unlabeled output data. A new input is then added to predict which cluster it is associated with. Algorithm examples could be K-means clustering, Principal Component Analysis.
Reinforcement Learning (RL)	A reward base learning where feedback is provided on the output to improve the prediction accuracy.

In this paper, we will apply both SL and UL algorithms to find clusters and patterns for the data collected from natural language processing.

2.2 The Application of Natural Language Processing

Natural Language Processing (NLP) is a subfield of AI that focuses on how to program computers to process and analyze large amounts of human/natural language data. NLP incorporates both speech and written language. For this research, only written text is used to solve the problem. Using NLP, we are able to web scrape job postings from multiple sources to create our data set to perform lexical, syntactic, and semantic analysis. Throughout this paper, Python packages specializing in NLP are leveraged to obtain text and document similarity scores of the job postings for each job title searched. The python libraries of Beautiful Soup, Natural Language ToolKit (NLTK), and spaCy are used to gather and clean the text.

2.3 Feature Vector Creation using Universal Sentence Encoder

Google's TensorFlow HUB (TF Hub) is an open source library for advanced Machine Learning and for numerical computation. In addition, the library contains an arsenal of algorithms for deep learning for digit classification, image recognition, word/sentence embeddings, recurrent neural networks, and for this paper natural language processing. The Universal Sentence Encoder (USE), which uses TensorFlow library, encodes text into feature vectors for the purpose of text classification, semantic similarity, and other natural language tasks. At the core, USE produces sentence embeddings for transfer learning [10] and is made publicly available on TF Hub. On a high level, a corpus of text is fed into the encoder and a 512-dimensional vector is output for semantic retrieval purposes. The vectors are then used to form clusters and classification on prediction accuracy. An example is below in Fig. 1.

Fig. 1. Classification using a simple binary text classifier

3 Related Work

In a paper from the University of Rome, Mauro et al. presented a classification of job roles and skills in the area of Big Data Analytics. The researchers used web scraping to retrieve job postings from many prominent websites. Natural language processing was then applied to this dataset to discover four essential job groups, most frequent bigrams appearing in the job titles: Business Analysts, Data Scientists, Developers and System Mangers. Then using the Latent Dirichlet Allocation, LDA, classification techniques the authors clusters skills into 9

topics that were generated by human interpretation of the skills. The 9 topics are: Cloud, Coding, Database management, Architecture, Project Management, Systems Management, Distributed Computing, Analytics, Business impact. Finally, the job skill sets are mapped to job roles by a measure of the extent at which each skill set is represented within each job post description [1]. (see Fig. 16)

Fig. 2. Job skill sets are mapped to job roles by a measure of the ‘extent at which each skill set is represented within each job post description.

A second group out of California State University focused entirely on the difference between Business data analytics, Database Administrator and Data Science, DS. Radovilski et al. manually collected job descriptions of Database Administrator and Data Science jobs from job boards. Using the job description, they identify skill sets associated with Business, Analytical, Technical, and Communication knowledge domains. Using text mining approaches, Document Data Matrix, Term Cloud, Singular Vector Decomposition, VARIMAX rotation and Latent Class Analysis the authors found the most frequent BDA and DS terms used, (see Fig. 17)[8].

Fig. 3. Terms used in job descriptions shown as proportion of the whole.

Some published papers to add to the paper at a later time. [4, 5, 2, 6, 7]

4 Data Set Collection

Data is collected from Indeed.com. Indeed.com is the number one job site in the world with over 250 million unique visitors every month [6]. Indeed.com gives users free access to complete job-seeking tasks such as searching for jobs, posting resumes, and researching companies. Globally, Indeed.com has 9.8 jobs added posted every second. Data is not readily available for download from Indeed.com, so it must be scrapped. Web scraping, also referred to as web harvesting or web data extraction, is a process commonly used by Data Scientist to get data from the World Wide Web directly from the respective website (NEED REFERENCE). Web Scraping for this project is done using the python library, BeautifulSoup (beautifulsoup4). BeautifulSoup allows us to scrape Indeed.com for information on several job titles and extract the information. “Beautiful Soup sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree (reference: <https://pypi.org/project/beautifulsoup4/>).” Table 3 illustrates the data that was gathered during the scraping process.

Table 3. Data Set Description.

Data Set	Variables
Job Search Page	Individual Job Posting Page Link
Job Search Page	Job Title
Job Search Page	Location
Separated from Location	City
Separated from Location	State
Separated from Location	Zip
Separated from Location	Country
Job Posting Page	Qualifications
Job Posting Page	Skills
Job Posting Page	Responsibilities
Job Posting Page	Education
Job Posting Page	Requirements
Job Posting Page	Full Description
Job Search Page	Salary (if available)

4.1 Data Set Contents

Data is scraped from Indeed.com for the following 6 job titles: Data Scientist, Data Analyst, Data Engineer, Database Administrator, Software Engineer, and Statistician. Each job title is search in the following 16 cities: Atlanta, GA, Austin, TX, Bellevue, WA, Boston, MA, Chicago, IL, Cupertino, CA, Dallas, TX, Denver, CO, Houston, TX, Los Angeles, CA, Mountain View, CA, New York, NY, Pittsburgh, PA, Seattle, WA, San Francisco, CA, and Washington, DC. By searching this criterion, we have 96 different possible combinations and a potential of over 28,000 job posts. However, many cities did not pull 300 job postings and some of our cities were so close to each other that duplicates are found. After duplicates are eliminated, our total data set contains 8,738 unique job postings. A distribution of all the job postings pulled can be seen in Fig. 4, 5),and 6) below.

4.2 Exploratory Data Analysis

Indeed.com allows companies to post jobs with their own HTML code beyond the generic required information. This means that some companies included salary information, company logo, and/or company rating while others did not. This also means that most of the Job Description Summary sections are different base on how the company posted the job (bold with bullets, one single paragraph, and a myriad of other variation in between). Additionally, running the entire job description text through our analysis means that we include items like company information and non-discrimination clauses. After examining the data, we determine that enough of the job postings have bullet

Job Title Distribution

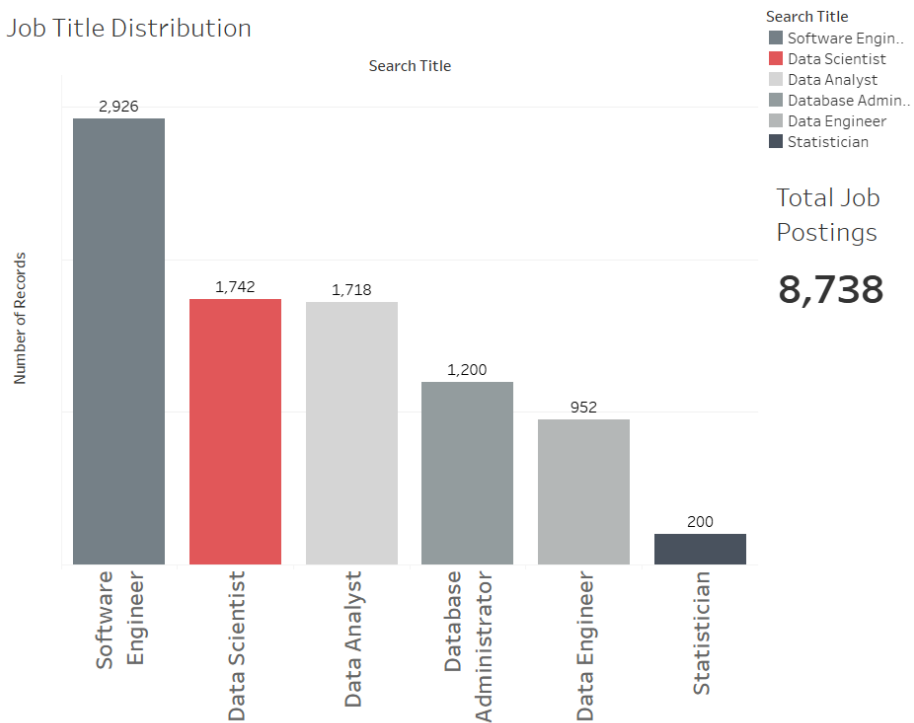


Fig. 4. Count of Job Postings

Job Posting Locations

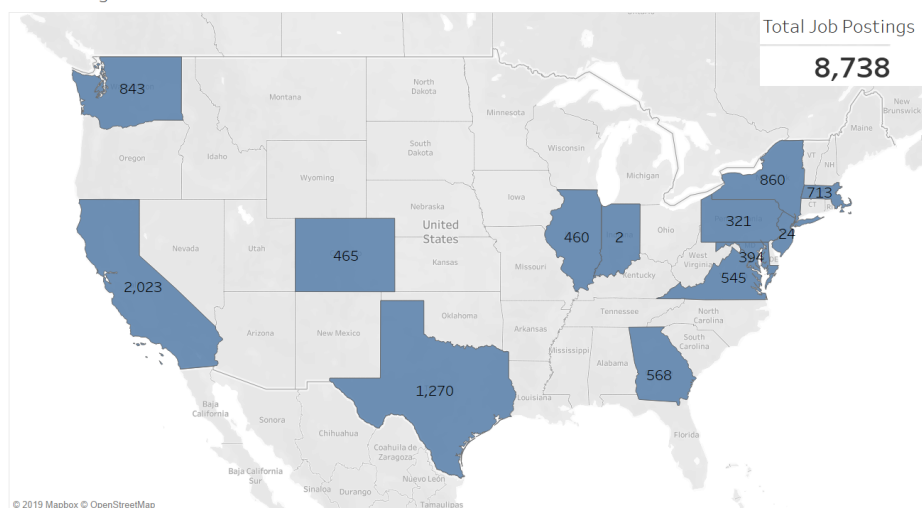


Fig. 5. Distribution of job postings across the United States

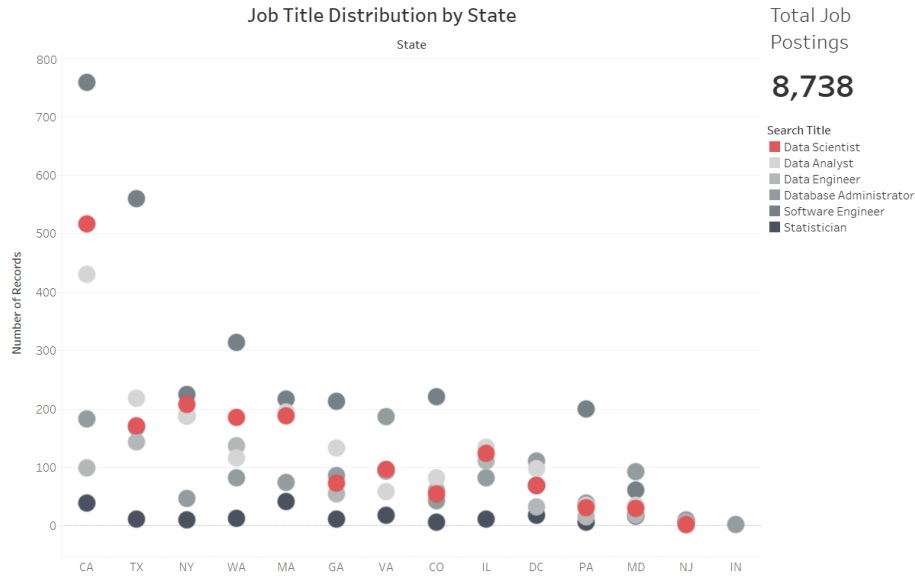


Fig. 6. Count of Job Titles by State

points within the text. This creates a uniform way to pull information for analysis that excludes information we do not want. After removing job postings who did not utilize the bold words, “Education”, “Qualifications”, “Requirements”, “Responsibilities”, and/or “Skills” and bullet point information below, our data set shrunk to a total 4,156 job postings. The final count of jobs used can be found in Fig. 10). It is interesting to note that all of our final counts ended up with a total of two different digits within each total (i.e. 1,311 uses the digits 1 and 3). This does not have any effect on the outcome, but we found it interesting. Fig. 8 shows the breakdown of those jobs throughout the United States. Table (see Fig. 9) shows the breakdown by job title searched.

5 5 Methodology from a Bird’s Eye View

The high-level overview of the data science approach to solving the problem: Unstructured data from Indeed.com was web scrapped, pre-processed to extract relevant text and transformed into feature vectors using USE. The next phase was a continual cycle of clustering and classification for both exploratory and insight purposes. Once an acceptable accuracy score was obtained, the clusters were reviewed to find closes neighbors to the data scientist job postings – closes neighbors being statistician and data analyst. The final step was to find the features of importance for each of the three roles to reverse engineering the results. The process is shown below in Fig. 11

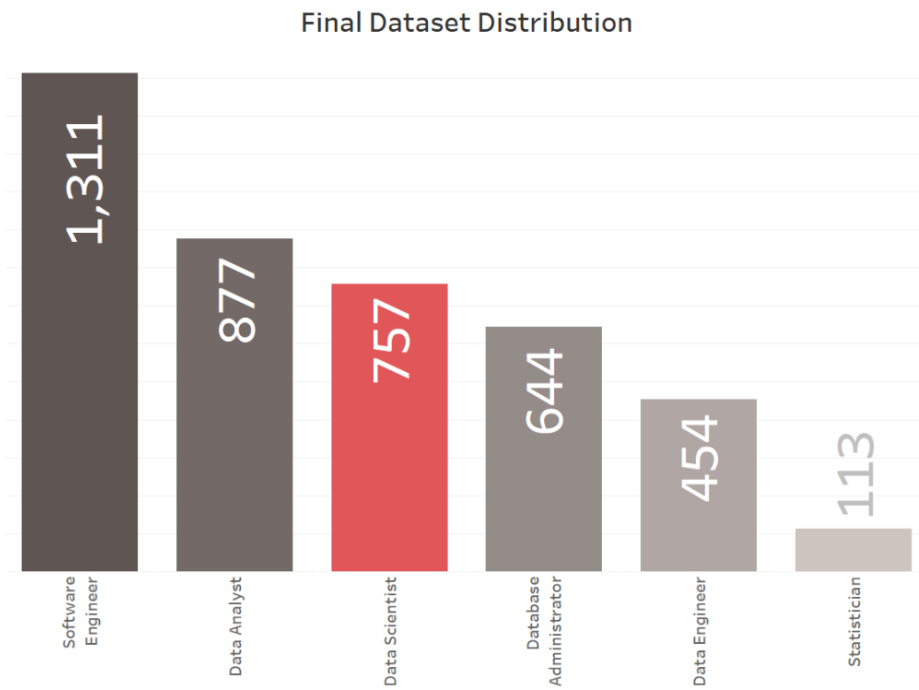


Fig. 7. Count of Jobs in the Final Dataset

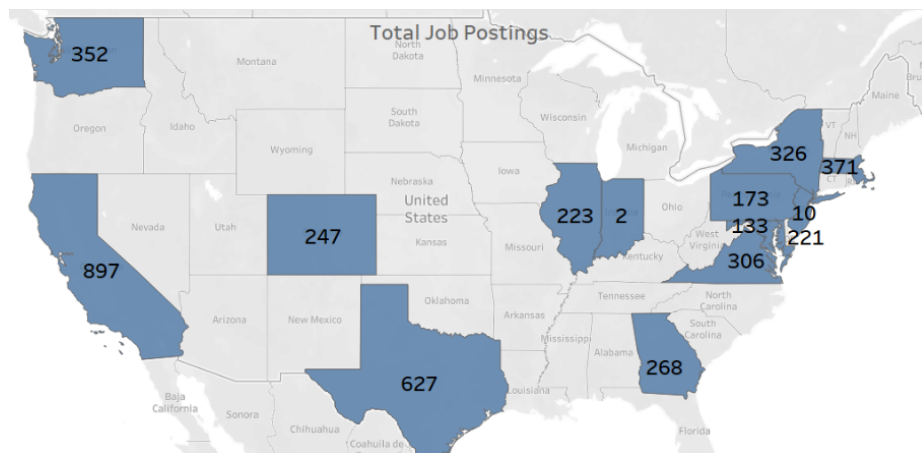


Fig. 8. Final Data Set Counts by State

Job Posting Used in Analysis															
Search Title	State														Grand Total
	CA	CO	DC	GA	IL	IN	MA	MD	NJ	NY	PA	TX	VA	WA	
Data Scientist	205	28	43	42	51		83	8	1	74	14	92	54	62	757
Data Analyst	207	47	52	67	73		99	21		78	18	122	31	62	877
Data Engineer	54	31	22	31	50			7	3	78	7	79	44	48	454
Database Administrator	90	18	68	32	42	2	54	60	2	20	20	72	126	38	644
Software Engineer	326	121	25	93			108	26		70	110	255	43	134	1,311
Statistician	15	2	11	3	7		27	11	4	6	4	7	8	8	113
Grand Total	897	247	221	268	223	2	371	133	10	326	173	627	306	352	4,156

Fig. 9. Distribution of Job Titles Throughout the States

Fig. 10. NEED CAPTION

5.1 5.1 USE for Sentiment Retrieval – 512 Dimensional Feature Vector Transformation

Once the data was collected, it needed to be transformed into something that could be quantified and analyzed. Tensorflow-Hub’s Universal Sentence Encoder (USE) transformed each of the unique job posting corpus into numerical vectors. USE transforms our text by finding semantically similar sentences (job postings) and places it into a 512-dimensional feature vectors for clustering and classification. Tensorflow-hub’s USE is pretrained with a deep averaging network (DAN) encoder and ready for our use ‘out of the box’.

5.2 Cluster Analysis on Key Terms

The first step of analysis is exploratory to find a ground truth which is deciding whether there is a clear market definition of a data scientist. Using the scikit-learn library (insert reference link: <https://scikit-learn.org>) , principal component analysis was performed to reduce the 512 features vector to two dimensions. Each job description is plotted with the centroid of the clusters highlighted in red (Fig. 11). Judging from the initial clusters for each job title, the centroid for each job posting is located closer to the coordinates (0,0) but no decisive/clear insights is drawn.

Figure 12)

KMeans from the same library is the next method attempted. The results were used to form the baseline model when comparing to the neural network clusters. The method of initialization is ‘k-means++’, the number of times the algorithm runs is set at 10 with 300 maximum number of iterations per run and relative tolerance set at 0.0001. The data is not modified and the “elkan” algorithm was used. A total of seven data sets and a range of 1 through 30 clusters separately informs the optimal number of clusters for the data sets. Each job title is cluster individually and all six titles concatenated into one data set is cluster. The number of clusters for each analysis was determine by locating

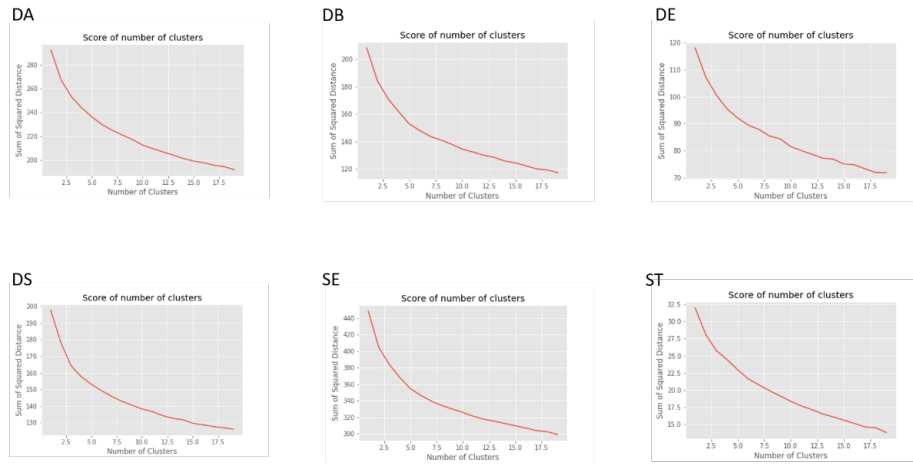


Fig. 11. PCA visualization of each job description as one cluster.

the elbow from plotting the sum of the squared distances of each sample to their closest cluster center (Fig. figure 12 and figure 13

From the plots 3 clusters for the individual job titles and 4 clusters when all job titles are combined into one dataset is chosen. This is an incorrect conclusion as it is known that there should only be one cluster for each individual job titles and 6 clusters for the combined data sets (Fig. 14 and 15). From these clusters, the ground truth was established – there is market confusion and/or ambiguity for the role of data scientist. Figure 13 below is the KMeans clustering of the full dataset when 2 clusters is specified. Because this is unsupervised clustering an arbitrary label of 0 and 1 is assigned by the algorithm to each vector. These labels can be compared to the original datasets 0 is found to contain more vectors from the DS job description and 1 contains all the other job description, non-DS. After determining the label accuracy of the model is determine by the percentage of vector correctly labeled. The accuracy of the binary KMeans clustering algorithm is 55.37% (Fig. 16) - not useful for conclusion. The same procedure but with 6 clusters specified in Fig. 17 below. The Data Scientist job is mapped to group 1, dark blue. Data Analyst job is mapped to group 2, dark green. Statistician is mapped to group 4, light green. Database Administrator is mapped to group 5, yellow. Software Engineer is mapped to both group 0 and 3, purple and green. Data Engineer was not detected by the K- NN algorithm. Data Scientist and Statistician has a large overlap. Data Analyst group is separated from the Data Scientist/Statistician group. Database Administrator is also well separated from Data Scientist/Statistician group. Software Engineer is well separated from the Data Scientist/Statistician group and the Data Analyst group but overlap with the Database Administrator group. The algorithm could not detect the Data Engineer group. A possible explanation is that Data Engineer is too similar the

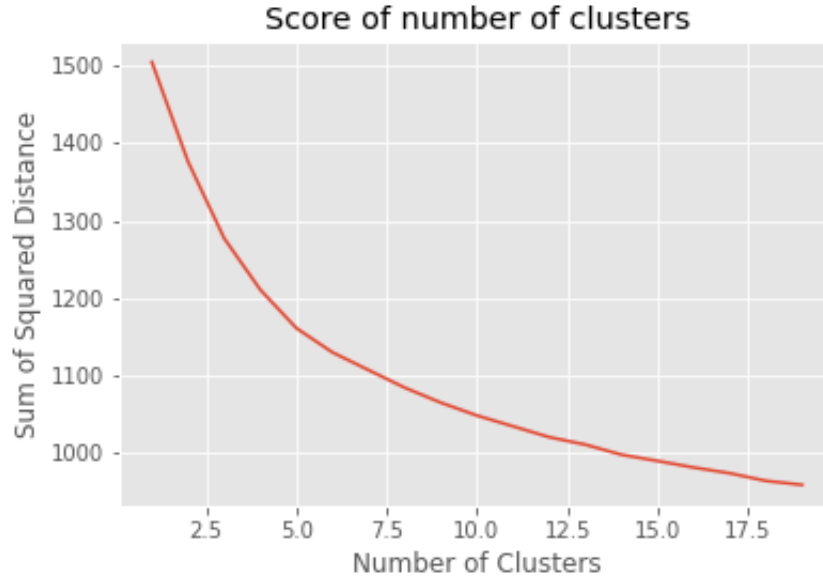


Fig. 12. Sum of Squared Distance for each job title with different number of clusters.

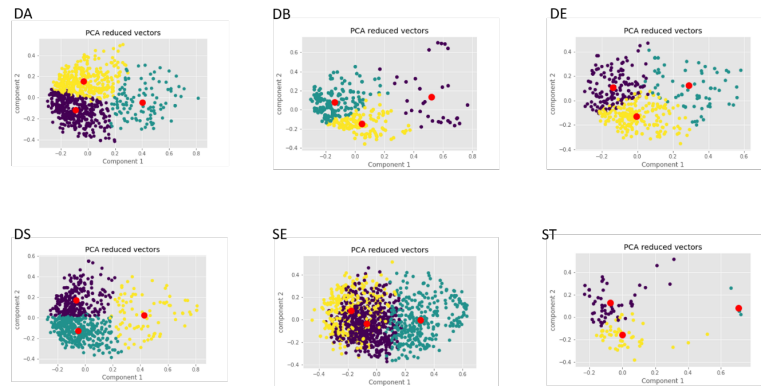


Fig. 13. Sum of Squared Distance for all the job titles combined.

Software Engineer group, observe figure 10. The accuracy of the model is 53.51% - not useful for conclusion.

To increase the model's predictive accuracy, a two layer dense neural network (NN) was used. Table 4 (Will make a table) is the summary of the NNs. The activation function of the input layer for both NNs is 'relu' and 256 dimensions output. The binary NN output layer's activation function is 'sigmoid' and 1-dimension output. The loss function used is 'binary_crossentropy.' To get

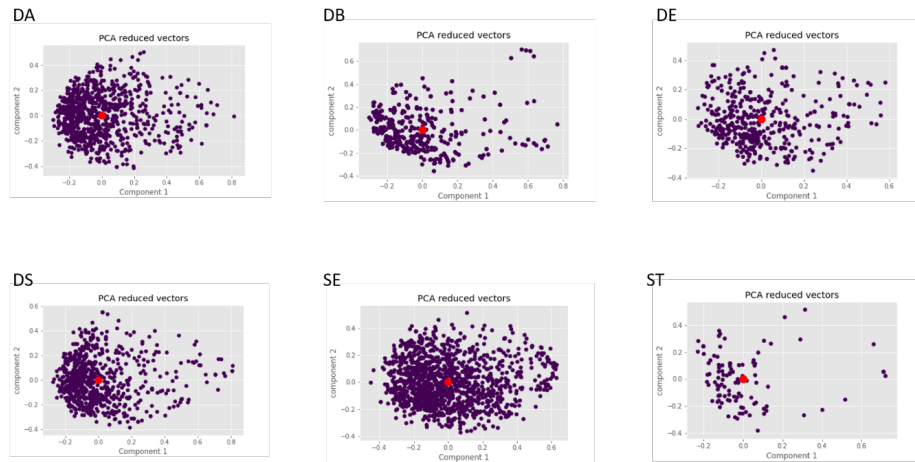


Fig. 14. PCA visualization of each job description as three clusters.

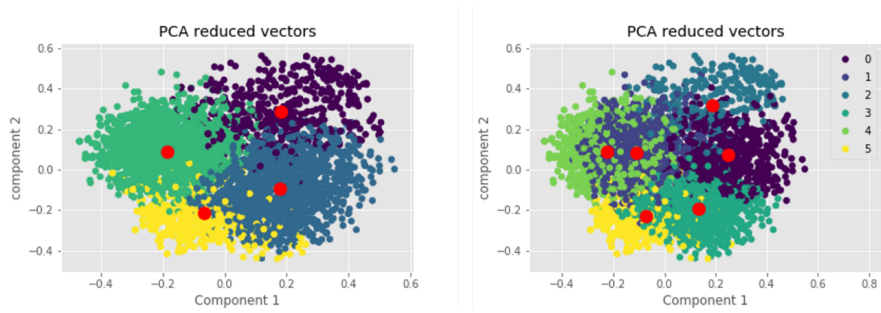


Fig. 15. PCA visualization of the combination of all six job titles with 4 clusters

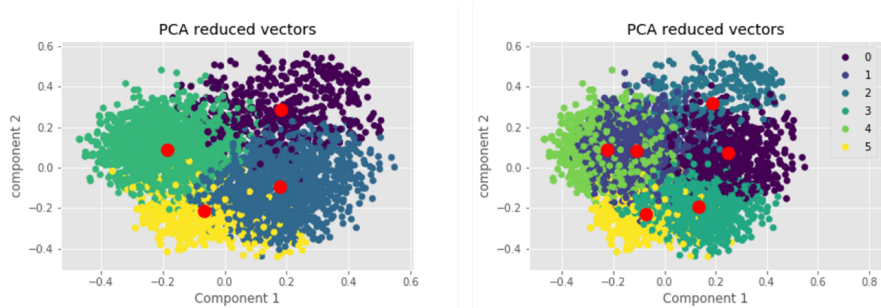


Fig. 16. PCA visualization of the combination of all six job titles with 2 clusters

6 clusters the output layer's activation function is 'softmax' and 6-dimensions output. The loss function is 'sparse_categorical_crossentropy.' For both NN the

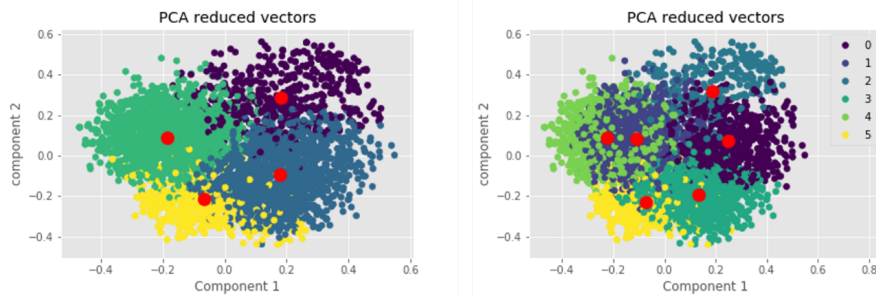


Fig. 17. PCA visualization of the combination of all six job titles with 6 clusters

‘Adam’ optimizer with a learning rate of 0.001. Both NNs has 15 epochs with a batch size of 75. The binary NN’s, looking at DS jobs and non-DS jobs, accuracy is 92.25% (figure 19). While the NN of all 6 job descriptions accuracy is 89.27%, figure 20.

Fig. 18.

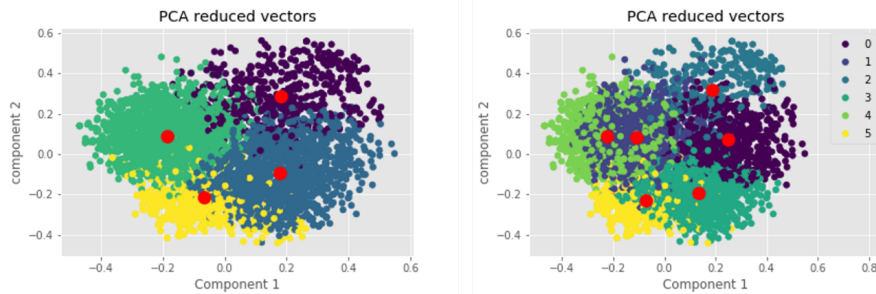


Fig. 19. PCA visualization of the combination of all six job titles with 2 classifications

6 Defining a Data Scientist

The final neural network model shows that Data Scientist is most closely related to Data Analyst and Statistician. With this insight, we look back at our dataset job postings to reverse engineer the features of importance. All duplicate words within each posting as well as some additional stop words that would not inform

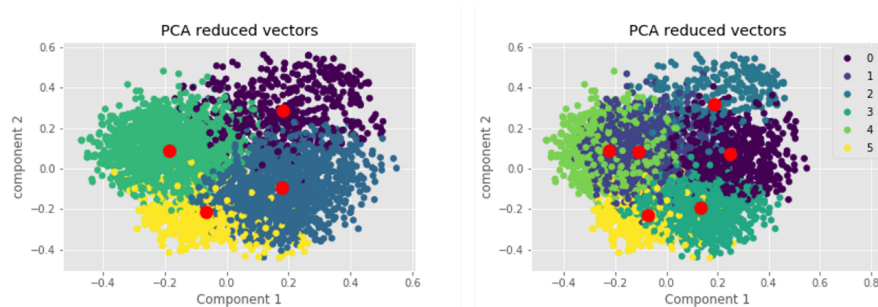


Fig. 20. PCA visualization of the combination of all six job titles with 6 classifications

our results (preferred, quality, field, work, strong, working, etc, large, experience, ability...) are removed. The top (highest word count) 25 words for all 6 job titles are below in Fig 21 .

Fig. 21. Top 25 words found in the job postings in the final data set.

6.1 6.1 Focusing on Data Scientist, Data Analyst, and Statistician

The most common words in the job postings for Data Scientist, Data Analyst, and Statistician are used to create the diagram below in Fig 22. Data Scientist and Data Analyst have more words in common and therefore, Statistician is a smaller circle. Data Scientist, Data Analyst, and Statistician share common words such as Research, Analysis, Statistics, Communication, and Team. Data Scientist and Data Analyst share words such as Python, Tools, Business, Support, sql. Data Scientist and Statistician do not share any top words that are not also shared with Data Analyst. A Data Scientist differs from both a Data Analyst and Statistician by including skills and tools such as Programming, Modeling, Computer Science, Engineering, and Machine Learning.

Fig. 22. Top 25 words found in the job postings in the final data set.

6.2 What is a Data Scientist?

Based on the information gathered in job postings, when an employer is looking for a Data Scientist, they look for the following: A Data Scientist codes, com-

municates, and collaborates – transforming data into insights using statistical, analytical, and machine learning techniques.

7 Analysis

7.1 Discoveries

8 Ethical Considerations

Ethics plays a role in the entire job search and interviewing process. There are many laws and regulations that oversee the process once the interviewing begins, however, there are not many laws and regulations when it comes to the job search process.

8.1 Website Usage

Scraping data from the website must be done with extreme caution. Each website is required to publish a robots.txt file that describes sections of a website that is not allowed to be scrapped. Additionally, a website's terms and conditions may prevent someone from scraping. Falling outside the guidelines and/or company policies can be bad. There are criminal implications such as identity theft and hacking if information is scraped from a website without following the proper protocol. For the novice programmer, it can be easy to make this mistake. According to the 2016 lawsuit, LinkedIn V. Doe Defendants (reference needed), LinkedIn sued 100 people who scraped their website anonymously. The lawsuit was stopped in the U.S. District Court where Judge Edward Chen ruled that LinkedIn couldn't block companies from deploying bots to scrape data from a public website. Though this was not held up in court, it speaks to the breadth of the dangers and risks of web scraping.

8.2 Job Search Ramifications

Many ethical issues related to job searches revolve around the truthful representations of jobs. Employers may try to entice more applicants by displaying the role as more desirable than it is. According to the Society for Human Resource Management (SHRM), creating fake job descriptions is a common way to get more applicants in a pool even though the role does not exist and is not advertised as a pool. SHRM has a code of ethics for the overall human resource profession which addresses recruiting [3]. Inversely, applicants could utilize algorithm or apply data science to falsify or embellish their resumes. The goal would be to trick resume tracking software into classifying the applicant as qualified, competent, and/or a fit for the open position. Not only is this misrepresentation but it prevents other qualified applicants from being interviewed.

8.3 Model Used for Profiling Candidates - aka AI bias

One shared concern regarding Artificial Intelligence is its ability to be fair and neutral. Although there are many advantages of AI (i.e. speed and capacity to process), the software/programs are still written by humans who are biased and judgmental. The application of NLP and ML to identify features or patterns in job postings could also be used by employers to profile applicants. The risk is knowingly removing applicants based on gender, race, creed, religion, and/or sexual orientation.

9 Conclusion and Future Work

The title of 'Data Scientist' is still a new concept that has and will continue to evolve as the role molds to the needs of artificial intelligence and business requirements. Our current definition is the result of the current market's view via job postings. Additional future work is required to capture a higher accuracy. In data science, the insights are only as reliable and accurate as the data itself.

For future work, additional data such as surveys, in person interviews with working data scientists, and web scraping additional websites could provide a more well-rounded (and accurate) prediction on the definition and skillset. One of the flaws of unstructured data is the possibility of selection bias during the pre-processing phase. A possible solution would be to continually collect job postings to better train the model over time.

In conclusion, the common definition and skillset of a data scientist is simplified within the 3Cs and SAM. A data scientist codes, collaborates, and communicates – transforming data into insights using techniques in statistics, analytics, and machine learning.

References

1. Andrea DeMauro, Marco Greco, M.G.P.R.: Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management* **54**(5) (9 2018)
2. Ankita Srivastava, Yogesh Tiwari, H.K.: Attrition and retention of employees in bpo sector. *International Journal of Computer Technology and Applications* **2**(6), 3056–3065 (2011)
3. Bates, S.: Do recruiters need a code of ethics (2019), <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/do-recruiters-need-code-of-ethics.aspx>
4. Brijesh Kishore Goswami, S.J.: Attrition issues and retention challenges of employees. *International Journal of Scientific & Engineering Research* **3**(3) (4 2012)
5. Collins Marfo Agyeman, P.V.P.P.: Employee demographic characteristics and their effects on turnover and retention in msme. *International Journal of Recent Advances in Organizational Behaviour and Decision Sciences* **1**(1) (2014)
6. Kaplan, J.: Artificial Intelligence: What Everyone Needs to Know. *What Everyone Needs To Know®*, Oxford University Press (2016), <https://books.google.com/books?id=wPvmDAAAQBAJ>

7. Touring: Alan Turing: His Work and Impact. Computing Machinery and Intelligence®, Gale Virtual Reference Library (GURL) (2013)
8. Zinovy Radovitsky, Vishwanath Hegde, A.A.U.U.: Skills requirements of business data analytics and data science jobs: A comparative analysis. Journal of Supply Chain and Operations Management **16**(1) (3 2018)