Methodology of Economic and Statistical Studies - Final project

US job market : Trending sectors and most in-demand skills of 2019

Arthur BARTOLI
Albane MOEBS

16 janvier 2020

Call for tender

The academic administration of an American university in New York state wants to adapt the courses it will offer its student for the upcoming year in order to make them as competitive as possible on the job market. Indeed, recent evolutions in the economy and the development of cutting edge technology has made certain types of jobs obsolete, while others are created each day. As the image and ranking of the university depend greatly on the employability of its graduates, it wants to teach classes as innovative and up-to-date as possible. They ask for a report on the most in-demand qualifications, skills and positions currently in the US and/or New York state.

Our interest in the subject

We are two students in data science having received a formation in economics. As such, your call for tender is particularly interesting. It gives us an opportunity to apply what we have learned in both subjects of our studies and apply the methods we have seen in class. The NYC job market is very competitive and therefore very specific in their requirements for employment. As a diverse and global economic center, it has cutting edge market job offers that reflect the trend toward which the economy is headed. It is a great starting point to a study of the state of the labor market in the US, but also in developed countries.
Because of automation, the job market is quickly evolving and universities have to constantly update the set of skills they teach in order to keep their students competitive.

"The fear of artificial intelligence is real. The McKinsey Global Institute estimates AI can automate 50% of all paid tasks today. Our primitive machine learning AI can easily complete repetitive human tasks better than any human could, including recognizing speech, context, shapes, and images. These narrow AI can also complete tasks like navigating an unpredictable field of obstacles, play an instrument, analyze large amounts of unorganized data and more." - Sylvain Rochon, "The Job Market in Year 2040"

However, we think it is not just about making bots to keep robots from taking your job, but a more complex big picture and ecosystem being created. It would be interesting to find out if only menial tasks would likely be automated (transport, manufacturing, basic services..), or also more reflection-based tasks (manager jobs, accounting, banking, health services, administration..). All in all, which qualifications would likely ensure the the long-lasting success of your students, given the uncertainty around the subject ? The aim is to identify the different sets of skills and fields of study that are the most promising in terms of access to employment.

To answer this question, we will collect great amounts of data on job postings from which we will try and extract relevant information. By analysing their title, description, field of work, location and required skills, we hope to pinpoint the major trends in the job market at a national level.

Table of contents

I. Literature review : state of the art

A preliminary task to our US fastest growing employment sectors study is to gather information on the subject by finding similar studies which have been conducted in the past. We found several relevant information sources and articles, which we will now briefly present.

The first article which had interesting findings in our opinion was "The most promising jobs of 2019", published in January 2019 by the professional social network LinkedIn on its Official Blog. It presents the results of a study, conducted on the websites' own data, which focuses on the most in-demand skills and positions. A sector is considered "promising" if it has known a considerable growth in the past few years, if it has had a large number of job openings and if it offers high salaries and career opportunities.

To conduct this study, LinkedIn pooled data from millions of member profiles and job openings and grouped the job listings by the keywords found in their titles. Once the listings were classified by sector and position, they were able to attribute a score to each group based on the number of openings in the US, growth, salary, career advancement, and location.

The 15 positions and/or sectors which obtained the best ranking using this methodology were : data scientist, site reliability engineer, enterprise account executive, product designer, product owner, customer success manager, engagement manager, solutions architect, information technology lead, scrum master, cloud architect, product marketing manager, solutions consultant, product manager, machine learning engineer.

LinkedIn also listed the 5 most in-demand hard skills from its job postings : cloud computing, artificial intelligence, analytical reasoning, people management and user experience design.

What is immediately striking about these results is that the vast majority of the promising positions are related to information technology and computer science. We should investigate specifically on these sectors when conducting our own study.

A second information source is the US Bureau of Labor Statistics, which published the results of a study conducted on the fastest growing occupations in the US in 2018 and makes projections for the decade to come. They collected the data from their own Occupational Employment Statistics program. They obtained the following results (numbers are in thousands) :

2018 National Employment Matrix title and code		Employment		Change, 2018–28		Median annual wage, 2018(1)
		2018	2028	Number	Percent	
Total, all occupations	00-0000	161,037.7	169,435.9	8,398.1	5.2	$38,640
Solar photovoltaic installers	47-2231	9.7	15.8	6.1	63.3	$42,680
Wind turbine service technicians	49-9081	6.6	10.3	3.8	56.9	$54,370
Home health aides	31-1011	831.8	1,136.6	304.8	36.6	$24,200
Personal care aides	39-9021	2,421.2	3,302.1	881.0	36.4	$24,020
Occupational therapy assistants	31-2011	43.8	58.3	14.5	33.1	$60,220
Information security analysts	15-1122	112.3	147.7	35.5	31.6	$98,350
Physician assistants	29-1071	118.8	155.7	37.0	31.1	$108,610
Statisticians	15-2041	44.4	58.0	13.6	30.7	$87,780
Nurse practitioners	29-1171	189.1	242.4	53.3	28.2	$107,030
Speech-language pathologists	29-1127	153.7	195.6	41.9	27.3	$77,510
Physical therapist assistants	31-2021	98.4	125.0	26.7	27.1	$58,040
Genetic counselors	29-9092	3.0	3.8	0.8	27.0	$80,370
Mathematicians	15-2021	2.9	3.6	0.8	26.0	$101,900
Operations research analysts	15-2031	109.7	137.9	28.1	25.6	$83,390
Software developers, applications	15-1132	944.2	1,185.7	241.5	25.6	$103,620
Forest fire inspectors and prevention specialists	33-2022	2.2	2.8	0.5	24.1	$39,600
Health specialties teachers, postsecondary	25-1071	254.8	313.9	59.1	23.2	$97,370
Phlebotomists	31-9097	128.3	157.8	29.5	23.0	$34,480
Physical therapist aides	31-2022	49.8	61.2	11.3	22.8	$26,240
Medical assistants	31-9092	686.6	841.5	154.9	22.6	$33,610
Substance abuse, behavioral disorder, and mental health counselors	21-1018	304.5	373.1	68.5	22.5	$44,630
Marriage and family therapists	21-1013	55.3	67.7	12.3	22.3	$50,090
Massage therapists	31-9011	159.8	195.2	35.4	22.2	$41,420
Cooks, restaurant	35-2014	1,362.3	1,661.3	299.0	21.9	$26,530
Physical therapists	29-1123	247.7	301.9	54.2	21.9	$87,930
Respiratory therapists	29-1126	134.0	162.0	27.9	20.8	$60,280
Market research analysts and marketing specialists	13-1161	681.9	821.1	139.2	20.4	$63,120
Actuaries	15-2011	25.0	30.0	5.0	20.1	$102,880
Computer numerically controlled machine tool programmers, metal and plastic	51-4012	24.3	29.2	4.9	20.0	$53,190
Nursing instructors and teachers, postsecondary	25-1072	69.0	82.8	13.8	20.0	$73,490

According to the Bureau of Labor Statistics, the occupations which will know the highest growth, in number of job openings, are : personal care aide or home health aide, software or application developer, medical assistant, restaurant cook, market research analyst or marketing specialist.

The paramedical sector is largely represented, with a projected growth of 1 431 500 job openings before 2028. The medical field and computer science/information technology are also among the fastest projected growing sectors.

The Bureau of Labor Statistics published another study, called “Monthly Labor Review, Projections overview and highlights, 2018-2028”. They used data from the Census Bureau to make projections in four domains : population and labor force, aggregate demand, industry output and employment, and occupational employment. Those projections are made for each semester of the study. Each following estimation is based on the previous ones.

Assumptions are made about immigration and their effect on the labor market, but the article reminds that this factor is very uncertain. The macroeconomic model is licensed from Macroeconomic Advisers by IHS Markit.

Findings : the BLS has projected a growth by 8.4 million jobs to 169.4 million jobs from 2018 to 2028. However the participation rate is expected to decrease, mainly because of the aging of the population. The consequence is an inflated job market related to healthcare, welfare and more specifically personal care (more than 40% of the additional jobs over the studied period).

The rest of this growth is expected to come from the service economy, especially food preparation and serving, as well as computer and statistics-related job market.

Three labor groups is expected to shrink: all production related labor market, menial clerk work and sales.

In conclusion, the various studies conducted had similar findings : the fastest growing employment sectors, in 2018 and projected on the decade to come, are computer science and information technologies, the medical and paramedical sectors, and the service industry.

We expect to find similar results conducting our own study on US online job posting data.

II. Primary analysis : descriptive statistics

To start off our study, we looked for reliable sources online about recent job offers in the US. We reckoned our ideal dataset would meet the following criteria : have as much observations as possible, be transparent on the origin of the data, and contain as features the location of the offered position, its title, a description of the missions to perform and a list of the required skills.

We first selected a dataset on Kaggle.com, which is available at this address :
https://www.kaggle.com/promptcloud/monster-usa-job-postings-dataset

It was pre-processed by PromptCloud's in-house web scraping solution. It is a subset of 20 000 observations extracted from a larger dataset.

After filtering the data by country and keeping only the observations from the US, it contained information on 13 859 online job postings published on the leading job board Monster.com, between June and September 2019. We specifically wanted to work on the latest available data in order for our results to be as up to date as possible.

Among the 34 available attributes, the features we are going to be interested in are :

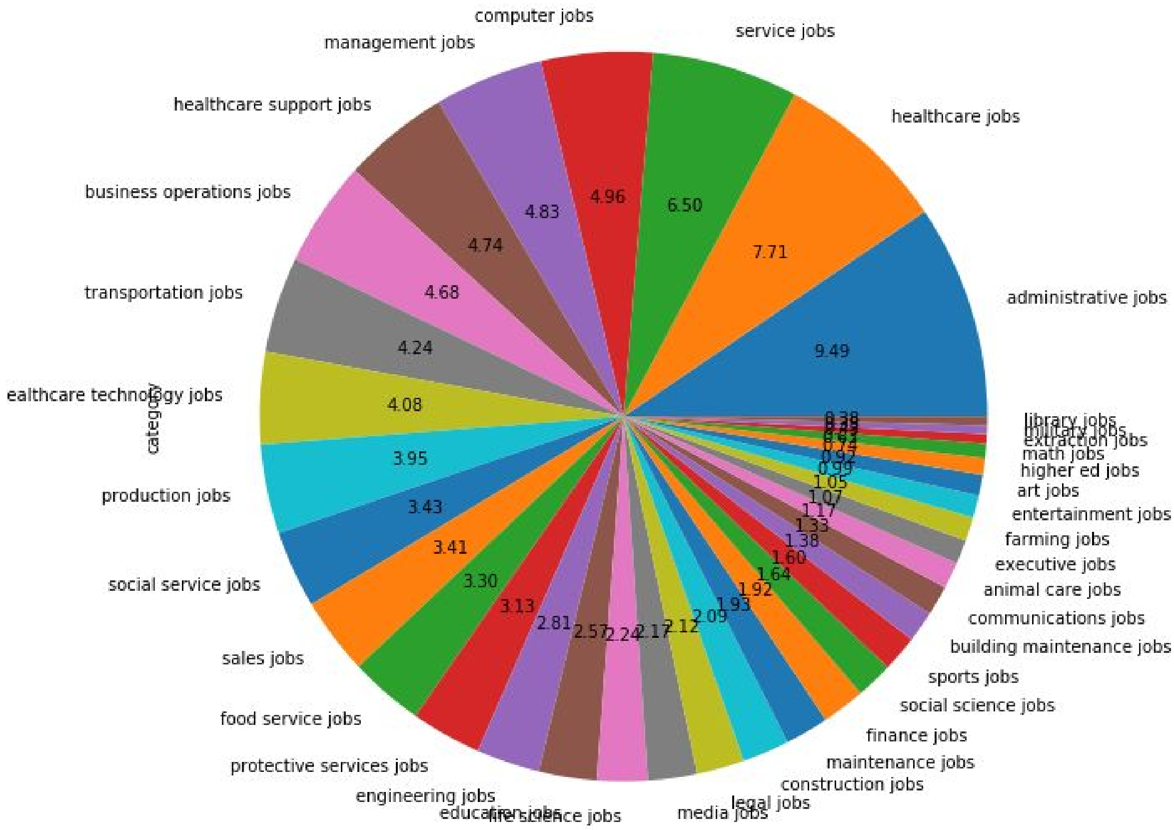
- Category : the general field of work
- Job title
- Job description
- State

It also had a "required skills" and a "salary offered" columns but these were not exploitable because they contained mostly incomplete or irrelevant data.

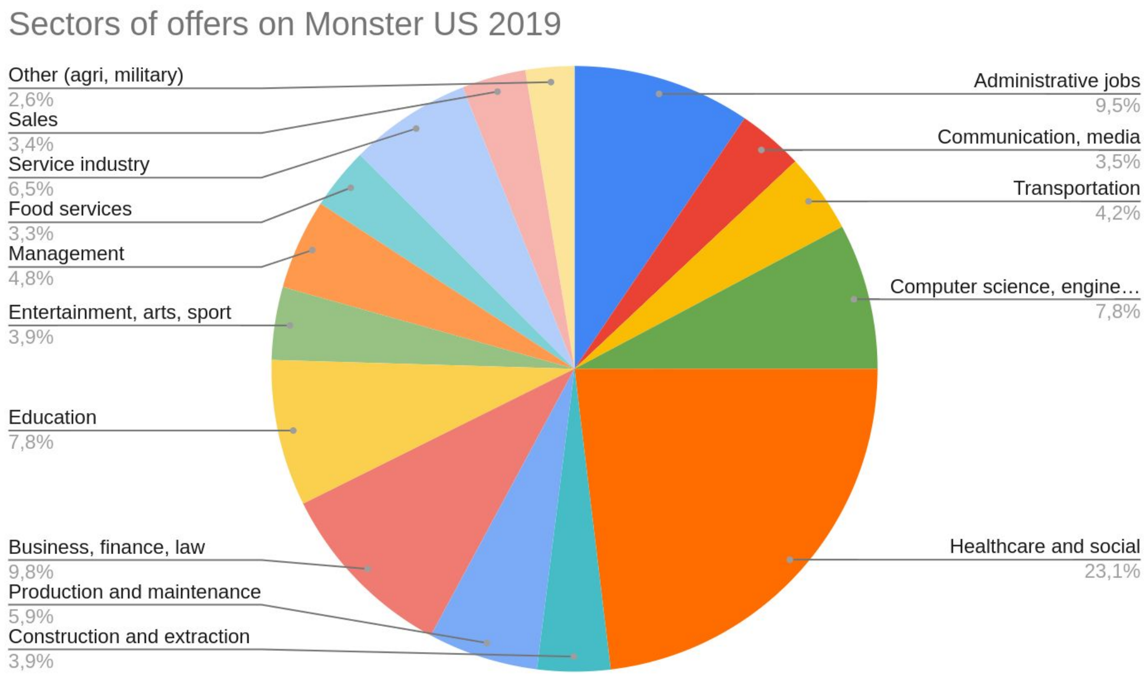
The easiest approach to know which sectors are looking to hire new workers the most is to look at the distribution of the "job categories" attribute in our dataset.

Here are the top 15 most frequent categories for job listings posted on Monster.com between June and September 2019 :

1	administrative jobs	9.49%
2	healthcare jobs	7.71%
3	service jobs	6.50%
4	computer jobs	4.96%
5	management jobs	4.83%
6	healthcare support jobs	4.74%
7	business operations jobs	4.68%
8	transportation jobs	4.24%
9	healthcare technology jobs	4.08%
10	production jobs	3.95%
11	social service jobs	3.43%
12	sales jobs	3.41%
13	food service jobs	3.30%
14	protective services jobs	3.13%
15	engineering jobs	2.81%



Above is a pie chart of all sectors. It represents the exact distribution as it is in our dataset, but is not easily readable. To get a better overview, we grouped similar categories together and obtained the following graph :



From this representation we can conclude that the most dynamic sectors in terms of employment in our dataset are the healthcare and social field (23.1% of observations), business-finance-law (9.8%), administrative tasks (9.5%) and computer science / engineering (7.8%). Now that we have grouped all medical-related categories, it is apparent that healthcare is by far the most prevalent sector in our dataset. This is consistent with the findings of studies mentioned in our literature review. However, the fact that business-finance-law is ranked second comes as a surprise, but it may be due to our arbitrary grouping : this category is probably too broad. It is also surprising that offers for administrative positions outnumber those targeting computer scientists and engineers.

The results slightly vary when we filter the dataset to keep only entries for jobs in New York state : the top two categories are still administrative and healthcare jobs, but business operations positions come third with 6.17% (compared to 4.68% at country level).

1	administrative jobs	7.32%
2	healthcare jobs	7.04%
3	business operations jobs	6.17%
4	service jobs	5.46%
5	computer jobs	5.46%
6	management jobs	5.31%
7	media jobs	4.60%

8	healthcare support jobs	4.17%
9	legal jobs	4.02%
10	engineering jobs	4.02%

To avoid distorting the results with our categorization method, we then tried using a more specific and objective approach. Inspired be LinkedIn’s study of the keywords contained in offers’ titles, we conducted a bag-of-words analysis to identify the most recurrent terms in each category of our dataset. To do this, we first filtered the raw text from job titles to remove punctuation, stop words (transition words which carry no specific meaning) and words shorter than 4 letters. Then, we extracted the top 20 most frequent words for each category.

Here is an excerpt of our results on the top 4 categories :

Sector	**administrative**	**healthcare jobs**	**service jobs**	**computer jobs**
Frequency %	9.49	7.71	6.50	4.96
Nb of offers	1315	1069	901	687
Top 20 keywords, number of occurrences	[('clerk', 190), ('assistant', 143), ('office', 125), ('time', 86), ('manager', 83), ('administrative', 72) ('driver', 64), ('part', 63), ('receptionist', 61), ('courier', 56), ('specialist', 55) [...]	[('nurse', 294), ('therapist', 145), ('registered', 123), ('travel', 99), ('physician', 96), ('assistant', 79), ('clinical', 73), ('medical', 68), ('time', 66), ('practitioner', 63), ('dietitian', 62), ('care', 59) [...]	[('needed', 73), ('home', 67), ('assistant', 67), ('funeral', 65), ('time', 65), ('personal', 56), ('retail', 54), ('nurse', 47), ('part', 44), ('careers', 44), ('manager', 44), ('travel', 42) [...]	[('developer', 165), ('engineer', 146), ('software', 129), ('analyst', 76), ('network', 68), ('senior', 60), ('data', 58), ('technician', 57), ('computer', 52), ('specialist', 51), ('application', 49), ('security', 39), ('database', 37) [...]

From this descriptive analysis, we can conclude that the most frequently offered specific positions are :

- Nurse (294 occurrences in Healthcare jobs category)
- Manager (291 in Management jobs)
- Driver (239 in Transportation jobs)
- Healthcare assistant (211 in Healthcare support jobs)
- Engineer (198 in Engineering jobs)
- Teacher (196 in Education jobs)
- Administrative clerk (190 occurrences)
- Developer (165 in Computer jobs)

This is still more or less consistent with the findings of various online sources. The most hiring sectors according to them were supposed to be healthcare, medical support, computer/IT and services. The high ranks of *manager* and *driver* positions in our dataset were not expected.

We start to get the sense that this dataset might be biased, and that the offers posted on Monster.com may not accurately represent the US job market as a whole.

To address the sample representativity issue, we will first try and gather data from other sources to get more observations, and then redress the sample according to the official employment distribution by sectors if needed.

III. Getting more varied and exhaustive exploitable data through classification techniques

Our goal is to get different relevant datasets which contain job offers from other sources than Monster.com. The problem we came across is that, when the occupational sector was mentioned at all, their classification system did not match this of our Monster.com dataset. Aggregating the data in one large exploitable set required classifying all additional offers into the same categories as our first dataset. To do this, we tried two machine learning methods : clustering using natural language processing (NLP) techniques on the one hand, and neural networks on the other hand.

1) NLP classifier

This method uses the frequency of each word in job descriptions from Monster.com to learn how to classify offers. To do this, we first add each word contained in one or more descriptions into a vocabulary vector. This will be the columns of our input table. Only words longer than three letters which carry a specific meaning are taken into consideration. We filter out any punctuation, capitalization and numbers.
Then, each row will represent a job description for which we will write the word counts in columns. The last column will contain the correct class label given by Monster.com.
Here is an illustration :

Job Descriptions from Monster.com	**Vocabulary vector** (extract)				**True class label**
	developer	clerk	assistant	practitioner	
Looking for a new web developer ...	2	0	0	0	computer jobs
… assistant to our main practitioner...	0	0	1	1	healthcare
Air Force pilot needed from	0	0	0	0	military

For instance, the word “developer” appears twice in the first job offer’s description.
In reality, the vocabulary vector has several thousand unique words in it.

We feed this input table, containing 80% of the observations of the Monster.com dataset, to the NLP algorithm which will learn to associate some recurrent words to their matching job category. When a word is used in many categories, then it is not a good predictor to perform the classification and it will get a low coefficient for every class (job sector). On the contrary, when a word appears only in one category, then it is a good way to distinguish categories from one another and will get a high coefficient in the corresponding class.

Once the algorithm has attributed coefficients to each word for each category, it has learned a classification method. We use a sub sample from our Monster.com dataset (20% of observations), on which the algorithm has not been trained, to test its performance. It will predict a class to each row (offer) and check with the true class provided by Monster.com to see if it guessed correctly.

With our data, however, the algorithm was not able to show satisfactory performances. It only classified about 28.7% of job offers right in the testing sample of the Monster.com dataset. This score is the average performance over 4 distinct cross-validations. Had it been accurate, we could have used this algorithm to assign sectors to unclassified offers from other datasets. We should also add that the process was rather slow for that performance and would take 40 minutes to classify all the dataset on a standard CPU (2.7 GHz double-core). We will have to find another method.

2) Neural network classifier

Our go-to method was to build a dense neural network and use it to classify each job offer into the right sector. Architecture of the neural network :

LAYER	Number of parameters	ACTIVATION FUNCTION
Input Layer	448576	-
Dense Layer (1)	8320	ReLU
Dense Layer (2)	33024	ReLU
Dense Layer (3)	131584	ReLU
Output Layer	18468	Softmax

- 3 densely connected layers of 128, 256 and 512 neurons, all with a ‘ReLU’ activation function because of its ability to avoid the ‘vanishing gradient’ issue
- All densely connected layers are affected by a dropout rate of 25%, which means that at each batch training 25% random weights are ignored to avoid neurons to be correlated to each other.
- A final layers of 36 neurons each activated by a softmax function in order to classify each job description into one of 36 sectors
- The loss function used for training is categorical cross-entropy.
- A total of 639,972 parameters.

First, we needed to vectorize the data into a format that the neural network could compute. We used the regex module to pre-process the text to remove special symbols, blank spaces, isolated characters and put the whole text in lower case. Then we fitted a *tfidf-vectorizer* provided by the scikit-learn module in Python and transformed the data.
TF-IDF (term frequency–inverse document frequency) is a standard vectorization technique in the industry which computes the ratio between frequency of the word within a job description and the frequency of the word among all job descriptions. The result is a matrix of size 9285 x 85067. Since this matrix is very sparse it has been converted in a sparse matrix *scipy* format containing only non-zero values and their coordinates.

We trained the network for 100 epochs using 20-80 cross-validation on the training dataset and a batch size of 32 but the results were meager (**about 50%**) on the validation dataset. It took one hour of computation using Google's colaboratory tool, and therefore GPUs of their server farm. Google would occasionally lock us out of the server and switch to a new environment because we overflowed the RAM, erasing our progress. This was too computationally-intensive for us to try different settings so we decided to use the job title instead of its full description. The job title contained by definition the essentials to understand what the work was about and was naturally shorter than the description. The shape of the final input matrix was reduce compared to the original : 9285 x 7008.

This method provided very good results for a shorter computing-time. We trained the neural network for 150 epochs using a batch size of 16. This resulted in a **69.8% accuracy** on the validation dataset in about 3 minutes, measured by a 20-80 cross-validation.

Once the network was trained we were finally able to classify each job offer from different sources into the same categories.

3) Results : sectors distribution on more exhaustive data

Thanks to our neural network classifier, we were able to merge data from our various sources containing much more observations (87 872) than any individual table we could find.

The additional data we used was also posted on Kaggle. We selected two tables :
- one very similar to the Monster.com dataset, but collected on another job posting website : Dice.com. It contains 19 942 observations worldwide and provides the same essential features (job title, description, location…), except for the sector of activity
- a dataset created by JobsPikr, which is a platform specialized in job data delivery. It extracted information on job offers directly from more than 300 companies' websites, using machine learning techniques. It has 21668 observations.

We cleaned the data, filtered only the positions offered in the US and ran our machine learning algorithm to assign a sector to each new observation before merging them into one large dataset.

It is easily exploitable as the sector of employment of each offer has been harmonized according to our simple classification.

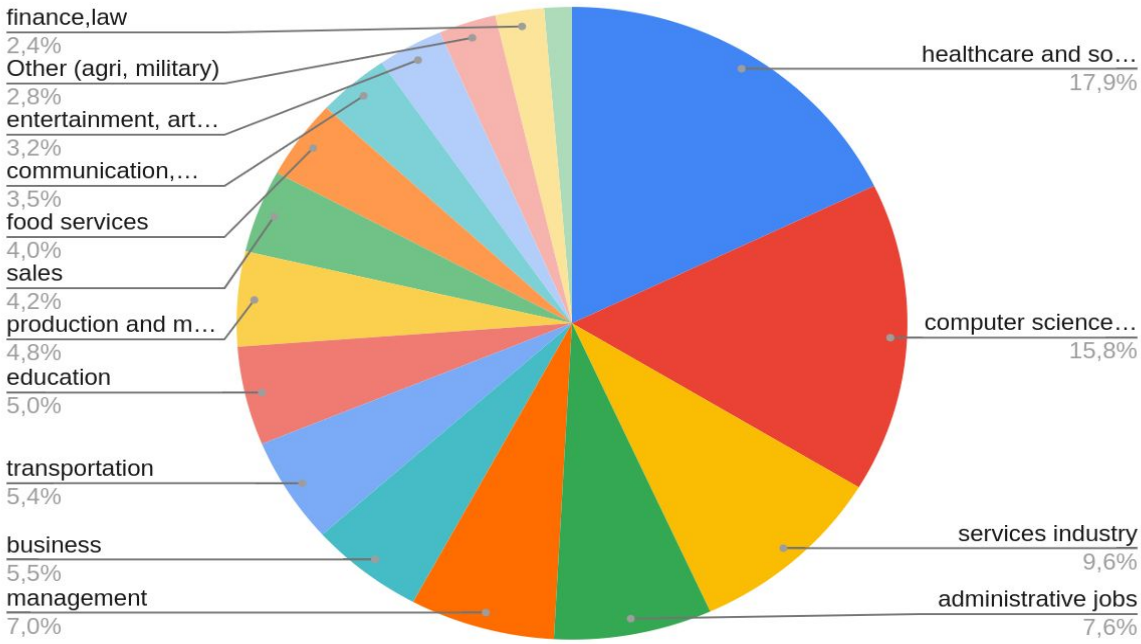
Here is an excerpt showing some of the accurately classified job offers :

Job Title	Assigned Sector	Location	JobText
Dialysis Technician	healthcare technology jobs	Stafford	The Dialysis Technician functions under the di...
Specialty Pharmaceutical Sales	sales jobs	Scranton	TITLE: Specialty Pharmaceutical Sales Rep/Cli...
NLP Research Scientist II/Senior NLP Research ...	higher ed jobs	Lakewood	Overview \nACT is a nonprofit organization hel...
Machine Learning Software Engineering Manager	computer jobs	Seattle	Seattle, Washington\n\nSkills : Machine Learn...
Counselor	social service jobs	New Orleans	Overview \n Behavioral Health Group (BHG), a ...
Associate Corporate Counsel	legal jobs	Mahwah	The Associate Corporate Counsel will serve as ...
Class-A CDL Truck Driving Positions	transportation jobs	Pittsburgh	At Dick Lavy Trucking, you'll find pay and per...
Security Guard	protective services jobs	Philadelphia	Allied Universal is seeking Professional Secu...
Registered Nurse - RN	healthcare jobs	York	The RN - Nurse Supervisor is responsible for s...
Sales Director	sales jobs	Saint Louis	Who We AreJost Chemical, www.jostchemical.com,...

We then computed descriptive statistics on the “assigned sectors” vector to answer our main question : which sectors offered the highest demand for workers in the US in 2019 ?

Sector	**% of offers**
computer jobs	12.50
services	9.56
healthcare	7.69
administrative	7.62
management	6.99
transportation	5.41
sales	4.23
business operations	4.08
food services	4.03
engineering	3.32
healthcare support	3.18
production	2.67
healthcare technology	2.62
media	2.60
protective services	2.56
social service	1.80
education	1.59
executive	1.43
finance	1.34
life science	1.30
maintenance	1.25
farming	1.21
construction	1.15
sports	1.15
social science	1.10
animal care	1.02
law	1.01
art	0.99
building maintenance	0.93
communication	0.88
entertainment	0.80
math	0.66
military	0.54
higher education	0.39
library jobs	0.21
extraction	0.18

Above is the complete distribution of sectors across our dataset. Here is a clearer overview after having grouped related sectors together using the same classification as we did earlier:

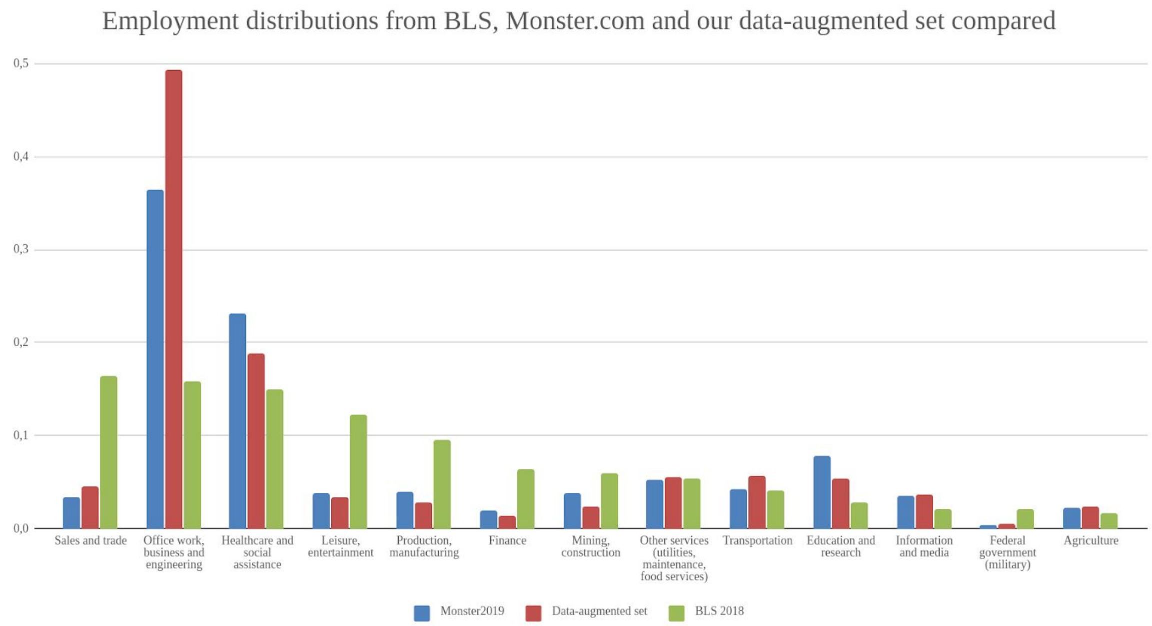


Now the repartition of our collected offers between occupational sectors seems a lot more realistic. It even matches the findings from articles in our literature review : the most dynamic sectors in terms of employment are healthcare & social work, computer science and the services industry. Finding additional data and making it usable for our study was a successful strategy.

V. Comparison with current employment distribution from official sources

To check for representativity in our sample, we decided to compare our findings to the official employment distribution by sector from the US government agency Bureau of Labor Statistics (BLS).

To do this, we first had to match as precisely as we could our sector classes to these used by the BLS classification. Some sectors did not exactly fit in any category from the BLS table, so we had to make arbitrary choices which might affect the accuracy of this comparison.



It is obvious that our data is biased if we take the BLS data as a benchmark. This lack of representativity of our sample might be due to the nature of our data : we collected job listings online from specialized websites. These platforms are not used by every single employer in the country, and this phenomenon is not homogeneous among sectors. For instance, the federal government has no reason to post job offers on third-party platform, it will use its own media. This explains why we find almost no position in this sector in our data, unlike the BLS. Similarly, employers in computer science, business and engineering are very likely to target the right candidates when posting job offers online (young, fully-connected workers). This may be why these sectors are over-represented in our data.

Another explanation for these differences in sectors distribution is that our study looked at vacancies : positions that are not occupied yet and for which a new worker is needed. On the other hand, the BLS studied positions that are occupied at the moment. So sectors which used to hire a lot of workers but are now declining, like manufacturing which is getting more and more automatized, are expected to be under-represented in our data compared to BLS. This is not a bias, it is only linked to our topic of interest : ongoing trends in employment.

All in all, we could redress our sample to make our merged dataset sectors distribution match this of the Bureau of Labor Statistics. To do this, we would have to use a weight matrix and apply coefficients to our sectors to reflect the official distribution.

However, it would not really make sense to apply this technique to our case because the employment sectors is our variable of interest. It would only distort our results.

VI. Detailed analysis : most sought-after skills in the tech industry

Since we want to make recommendations to the university board on which skills their students should acquire in order be more competitive on the market, we decided to carry a more detailed analysis focusing on the IT and computer science sector. We chose this category because it is among the top-hiring sectors and the offers frequently mention very specific required skills.

The dataset we use in this section is freely available on Kaggle :
https://www.kaggle.com/PromptCloudHQ/us-technology-jobs-on-dicecom/data

It contains information about 22 000 online job postings from the US specialized in the technology sector. The variables of interest we will use are job title, job description, and required skills. The goal of this analysis is to identify the most in-demand skills in the IT/computer science sector of employment, which we found has been one of the most hiring field in the last few months.

We gathered all the skills which were required by employers and counted the number of occurrences for each of them. The most in-demand skills in the tech industry, are, according to this dataset :

	Skill name	**Count** (nb obs = 20 000)
1	java	2131
2	sql	1811
3	development	1770
4	management	1564
5	javascript	1488
6	linux	1073
7	project	990
8	agile	950
9	testing	918
10	c#	904

11	html	892
12	python	868
13	security	857
14	analysis	820
16	css	772
17	oracle	711
19	manager	632
20	windows	629
21	developer	599
22	architecture	591
23	http	589
24	unix	579
25	project management	535
26	networking	528
27	database	528
28	programming	519
29	jquery	493
30	sql server	478

The most useful languages/environments to master in order to enter the job market seem to be : Java/JavaScript/jQuery, SQL/SQL Server, Linux/Unix, C#, HTML, Python, CSS, Oracle, Windows.

Employers tend to look for skilled individuals in the following fields : development, (project) management, Agile method, testing, security, analysis, architecture, databases, and programming.

VII. Limitations and scope for improvement

The amount of relevant, exploitable data we could find to answer the question properly was limited. We intended to also carry detailed studies on the other trending sectors, but the abilities required by students aspiring to work in the healthcare, administrative and social service sectors were either never mentioned, or excessively vague. We were looking for specific, teachable skills and ignored all character-related qualities.

Furthermore, we would have liked the job listings to explicitly mention the salaries offered, but we understand that this is not the standard procedure because wages are often up for negotiation. They also usually depend on individual qualifications. Otherwise, we could have built an econometric model to explain salaries by factors such as sector, position, location, skills needed. This would have allowed us to make recommendations not only based on the number of vacancies for a type of job, but also the specific niche markets for students to target to maximise their expected salary.

VIII. Conclusion and recommendations

Our findings are similar to what we found in the literature when it comes to the high demand for skilled workforce in healthcare support (7,7% of offers) and computer science / information technologies (12,5%). In these sectors, the positions which are the most often offered in the US are nurse, therapist and medical assistant for the former and (software) developer/engineer and data analyst for the latter. The services industry is also well represented (9,6%), as expected, and the demand for personal aids is especially high.
What is more surprising was the strong demand for administrative workers, which accounts for almost 8% of all offers in our dataset. Most of the listings are offering positions of administrative clerks, assistants, managers and receptionists.

Our analysis focusing on the state of New York shows that the sectors which hire the most workers are generally the same as in the whole country with the exceptions of business operations, media and legal jobs which ranked a little higher than expected.

By conducting a more in-depth analysis of the technology sector, we learnt that the most sought-after specific skills in this field are experience in project management, programming obviously and databases management. The use of Java, SQL, Linux, C#, HTML, Python, CSS and Oracle are required by most of the employers.

We would advise your university to invest in the medical, administrative and engineering/IT departments, with a focus on the training of nurses, medical assistants, personal aids, administrative clerks, software developers and data analysts. For these last two specialties, programming courses on the languages and softwares mentioned in the paragraph above should be particularly useful to improve the employability of your students.

Bibliography

- Kumaresh Pattabiraman, January 10 2019, *LinkedIn's Most Promising Jobs of 2019*, LinkedIn Official Blog, accessed 30/10/19, <https://blog.linkedin.com/2019/january/10/linkedins-most-promising-jobs-of-2019>

- Nick Bunker, September 17 2019, *These Jobs Have the Brightest Futures — and Pay Well Now*, Indeed Hiring Lab US, accessed 30/10/19, <https://www.hiringlab.org/2019/09/17/bright-futures-pay-well-jobs/>

- Bureau of Labor Statistics, October 1 2019, *Monthly Labor Review, Projections overview and highlights, 2018–28*, https://www.bls.gov/opub/mlr/2019/article/projections-overview-and-highlights-2018-28.htm

- Bureau of Labor Statistics, September 4 2019, *Fastest growing occupations*, United States Department of Labor, accessed 30/10/19, <https://www.bls.gov/emp/tables/fastest-growing-occupations.htm>

- Sylvain Rochon, January 30 2019, *The Job Market in Year 2040*, Data Driven Investor, accessed 15/01/20, <https://medium.com/datadriveninvestor/the-job-market-in-year-2040-d17f7743c0fd>