BA 482 Final Project

Kirsten Ballard, Ava Peterson, Amanda Mustaklem, Garrett Neal, Samantha Rasmason, and Danielle Schwartz

BA 482: Business Analytics Research

Dr. Lu Wang

April 28, 2022

Relevance of our project

As seniors in college, we are beginning to look at career opportunities and what options are available for individuals with a business analytics background, making the understanding of the job market a top priority for us. Out of the 6 members in the group, five members are currently looking/in the interview phase of finding a job in the field, while four out of the six members have already found a job or an internship in the field. Through this project, we sought to bring more awareness to the job search process.

Testing Theories

Since many of us have all been looking into jobs and our careers in the last year, we have been encouraged to become knowledgeable about the business analytics and data analytics fields. Everyone who is looking for a job wants the best possible offer they can get for the type of job they want to pursue. We sought to discover if analytical jobs are increasing in demand and salary, aligning with society's notions. These theories are tested based on business analysts, data analysts, and data scientists which may contribute to the demand and change in salary from one position to another.

Data Description

Combined Data Variables

Job_Post_Date <- Date variable with the day that the job posting was uploaded to the website.

Job_Tittle <- Text variable of the posting's job title.

Company_Name <- Text variable of the company name who posted the listing.

**City <- Text variable of the city where the position is being held in.

Company_Rating <- Integer variable with the company rating from 0-5 as reported by past employees.

**State <- Text variable of the state where the position is being held in.

Salary_Lower <- Integer variable of the minimum of the average salary range for the position.

Salary_Upper <- Integer variable of the maximum of the average salary range for the position.

**Salary_Avg <- Integer variable of the average salary for the position.

Salary_Estimated_GD <- Integer variable of the estimated salary for the position on glassdoor.

Salary_Estimated_Company <- Integer variable of the estimated salary for the company.

Full_Time_Job <- Dummy variable for full-time job positions. 0 = No, 1 = Yes.

Part_Time_Job <- Dummy variable for part-time job positions. 0 = No, 1 = Yes.

Temp_Contract_Job <- Dummy variable for either temporary or contract job positions. 0 = No, 1 = Yes.

Size <- Text variable for the company size bracket; the count of employees.

Industry <- Text variable for the industry that the company falls in.

Bac_Req <- Dummy variable for if the job listing requires a bachelor’s degree. 0 = No, 1 = Yes.

Qualifications <- Text variable for the preferred qualifications for applicants; degrees, years of experience, skills, etc.

Full_Job_Description <- Text variable with the entirety of the job description containing the position, compensation, requirements, qualifications, company culture, etc.

Master Requirement Regression (Added Variables)

**Mast_Req -> Dummy Variable detailing if the term Masters appears in either the job qualifications or description. 0 = No, 1= Yes

**Cost of Living -> Continuous Variable from the Cost of Living Data series identifying the cost of living for each of the 50 United States as compared to the overall Cost of living for the US.

Average Ind Income -> Continuous Variable of the average **individual income divided up by the 50 United States

Job Title Variations (Added Variables)

**Dajob <- Dummy variable for job positions containing the keywords of Data Analyst. 0 = No, 1 = Yes.

**Dsjob <- Dummy variable for job positions containing the keywords of Data Scientist. 0 = No, 1 = Yes.

**Bajob <- Dummy variable for job positions containing the keywords of Business Analyst. 0 = No, 1 = Yes.

**Compsize <- Integer variable for the avg between the range of company size in 'Size' variable

*** = Variables most frequently used in hypotheses and research questions*

Cost of Living Analysis (Added Variables)

Jba <- Dummy variable for the position of junior business analyst 1=Yes 0=No

Sba <- Dummy variable for the position of senior business analyst 1=Yes 0=No

Jda <- Dummy variable for the position of junior data analyst 1= Yes 0=No

Sda <- Dummy variable for the position of senior data analyst 1=Yes 0=No

Jds <- Dummy variable for the position of junior data scientist 1=Yes 0=No

Sds <- Dummy variable for the position of senior data scientist 1=Yes 0=No

Livingstate <- Continuous variable that is the cost of living per state (collected from an outside source)

Livingcity <- Continuous variable that is the cost of living per city (collected from an outside source)

Shortcomings of our data

Though there are many conclusions to be made based on the data we have collected, it is still important to consider the shortcomings and areas of weakness within our working data. Given the amount of time we spent collecting and scraping the data, we only have data captured from a few short months of January and February. There are many jobs that have opened up and finished hiring in the time it has taken us to fully complete this project; thus, our conclusions will not fully encapsulate the entire job market picture. Further to that point, the data we collected came only from two different job posting sites. We collected our data from Glassdoor and Indeed, which are the two most common job posting sites; however, we have no data collected from Linkedin or other popular sites that individuals use to look for jobs.

In terms of the actual data and data variables we collected, some variables were too specific while others were too general to be able to draw any definite or relevant conclusions. For example, the variable, industry, was too specific to be able to draw any relevant conclusions without making more general categories. We were apprehensive about categorizing the industries into larger groups in case we accidentally missed an industry or inappropriately categorized one. Another variable that was difficult to work with directly was company ratings. Company ratings are based on indexes; therefore, determining relevant factors that influence company ratings would prove to be very difficult. As a result of these various shortcomings, we were unable to answer some desired hypotheses and research questions such as, "*The medical and science field tends to dominate the job market in areas, making a higher percentage of these types of jobs available*" and "*What factors are typically mentioned/considered in a company rating? Who* *typically gives company ratings (employees, customers,etc.)?"* Nonetheless, there are still many relevant questions and hypotheses we were able to analyze with the data we collected throughout this semester.

Hypotheses

Higher company ratings, larger company sizes, and the title of "data scientist" are positively correlated with average salary rates:

For our research, we were extremely interested in which job title and which company factors would return the highest average salary. After discovering that the data scientist job title relies on heavy coding and has an "average annual salary range between $105,750 and $180,250" (Burnham 2021), we decided to use data scientist in our hypothesis. Additionally, we predicted that a larger company size would yield a higher salary, as well as, higher company ratings. This is due to the fact that larger, more popular, and highly rated companies typically have good benefits and corporate governance.

The first step in testing this hypothesis was to convert the three different job titles into dummy variables. We used SAS to create new columns that marked the job listing as 1 if it contained the keywords, and 0 if it did not. The process created three new variables; dsjob (data scientist), dajob (data analyst), and bajob (business analyst).

The next step in testing the hypothesis was to create the regression. We used STATA as the main tool to create and run the regression. The regression model was as follows:

$$salary\_avg = B_0 + B_1\, company\_rating + B_2\, compsize + B_3\, dajob + B_4\, dsjob + u$$

We decided to keep the business analyst job title out of the model so it is implemented in the constant variable. We chose the business analyst job title over the other two as the constant since their responsibilities are more related than the business analyst position. After running the regression with the following code, *reg salary_avg company_rating compsize dajob dsjob,* we can see that each of the independent variables are significant at the 95% confidence level and are relevant to the model, due to each p-value being less than 0.05. Our estimated equation is:

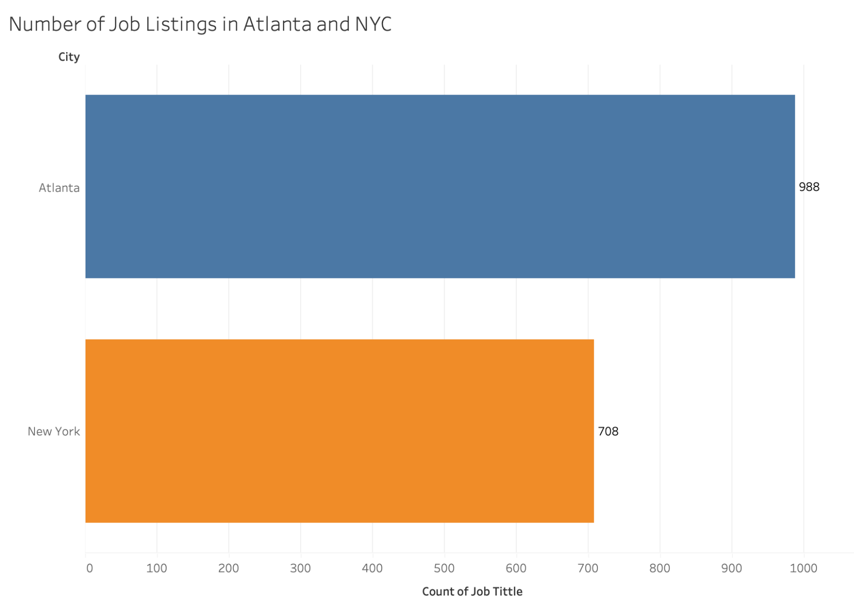
$$salary\_avg = 51843.07 + 5832.3\ company\_rating + 2.06\ compsize - 10861.63\ dajob + 7876\ dsjob$$

```
. reg salary_avg company_rat compsize dajob dsjob

      Source |       SS           df       MS      Number of obs   =       800
-------------+----------------------------------   F(4, 795)       =     11.21
       Model |  4.9847e+10         4  1.2462e+10   Prob > F        =    0.0000
    Residual |  8.8343e+11       795  1.1112e+09   R-squared       =    0.0534
-------------+----------------------------------   Adj R-squared   =    0.0486
       Total |  9.3327e+11       799  1.1681e+09   Root MSE        =     33335

--------------------------------------------------------------------------------
    salary_avg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
company_rating |   5832.295   2091.379     2.79   0.005     1727.018    9937.572
      compsize |   2.059789   .8184836     2.52   0.012     .4531447    3.666433
         dajob |  -10861.63   2779.015    -3.91   0.000     -16316.7    -5406.55
         dsjob |   7875.969   3200.987     2.46   0.014     1592.583    14159.36
         _cons |   51843.07   8209.781     6.31   0.000     35727.66    67958.48
--------------------------------------------------------------------------------
```

From these results, we can see that when company rating increases by 1 star, the salary avg increases by $5,832 if all other variables remain the same. Additionally, when company size increases by 1, the salary avg increases by $2 if all other variables remain the same. As for the job titles, we found that a Data Analyst makes $10,861.63 less than someone with the title Business Analyst if all other variables remained the same. As for our hypothesis, the job title Data Scientist has an average salary $7,876 higher than Business Analyst and $18,738 higher than Data Analyst if all other variables remain the same. Due to our findings, we can conclude that higher company ratings, larger company size, and title of data scientist does have a higher average salary rate.

The ratio of jobs to the population in Atlanta, GA is proportional to the ratio of jobs to the population in New York City, NY:

Two of the largest cities on the East Coast are New York City and Atlanta. Knowing this we figured that both cities would have a relatively similar number of job openings in comparison to the population of each city. Further, we sought to explore cities on the same coast to allow school schedules to align, as many west coast schools are on trimesters and quarter systems whereas the East Coast typically works on semester schedules. This led us to develop our hypothesis: The ratio of jobs to the population in Atlanta, GA is proportional to the ratio of jobs to the population in New York City, New York. We found that the current population in New York City is 8,177,025 (*New York City*, 2022) and the population in Atlanta is 6,013,000 (*Atlanta*, 2022). Using tableau we counted the number of jobs in each city and the results are shown in the table below.



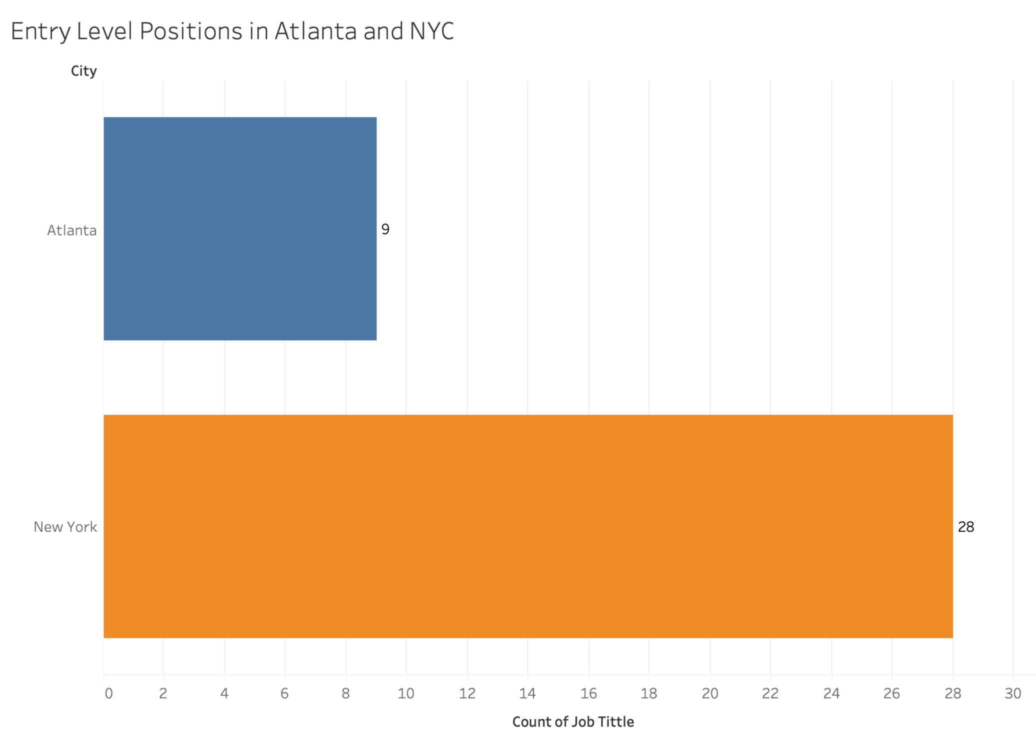
Using these numbers and the formula, ratio = Antecedent/Consequent, we got the following results:

$$\text{Ratio}_{\text{NYC}} = 708/8{,}177{,}025 \text{ or } 0.00008658$$

$$\text{Ratio}_{\text{Atlanta}} = 988/6{,}013{,}000 \text{ or } 0.00016431$$

Seeing that the Ratio for New York City was smaller than the ratio for Atlanta we did some more research to see if there was any explanation for what we found. During the Covid-19 pandemic, there was a lot of talk about how people were choosing to move out of New York City. After further research we found that over 100,000 people had left the Big Apple for one reason or another (Team, 2021). The same article also made a comment that Atlanta was growing very rapidly at the same time the population in New York was decreasing (Team, 2021). In fact, the population in Georgia had increased by over 200,000 people in the same year (Team, 2021). This could explain why the ratio for Atlanta was larger than the ratio for New York City.

After Finding out that the job ratio in Atlanta is higher than the job ratio in New York City, we wanted to see if there would be a difference for the number of entry level positions. Using tableau we filtered the count of jobs to show only entry level positions. The results from using this filter are shown below:

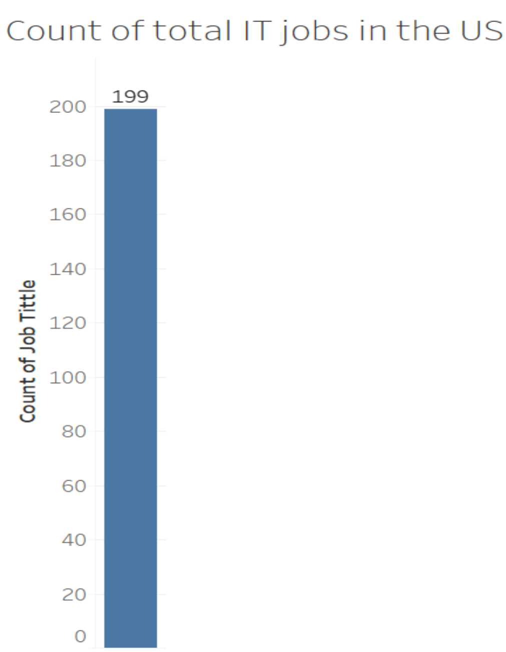


The number of entry-level jobs was much higher for New York than those for Atlanta. When looking into this, we found an article stating that individuals "aged 24 to 32, many of whom are coming for jobs or who are working remotely and want to give city living a try" (Klein, 2021). NYC is opening back up following the pandemic, so more young individuals are now heading back to the city to experience all of the activities that New York has to offer.

There is significant data to suggest there are more data scientist jobs in Texas than in California.

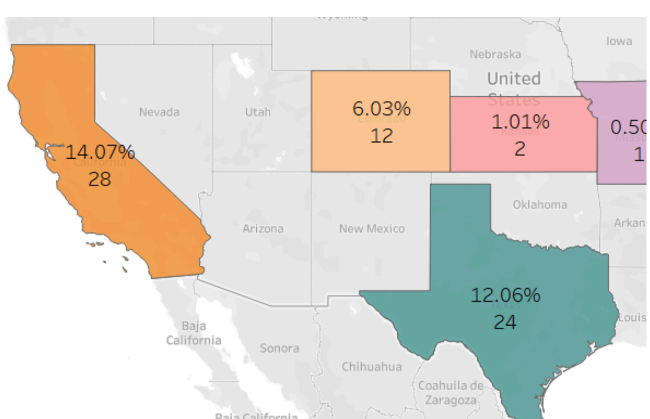
According to a research article published by Folger, Silicon Valley, California has turned into the tech industry's epicenter; however, increasing housing costs, high tax rates, and strict regulations have created challenges for workers to successfully work, live, and do business (*Folger*, 2021). Two large corporations, Oracle and Hewlett Packard have already moved from California to Texas. Information found from Investopedia states, "As of November, 35 companies had relocated to or opened new facilities in the Austin area in 2020 alone, according to data from the Austin Chamber of Commerce". Some of these companies include 8VC, FileTrail, DZS INc. and QuestionPro (*Folger*, 2021). Further, Elon Musk recently built a factory down in Austin and has announced he would be moving out of California to Texas to supervise the facility. These stats suggest companies are favoring the lower cost of living, and more-favorable tax laws the Austin community has to offer (*Folger*, 2021). This motivated our first hypothesis of assuming there are more data scientist jobs in Texas compared to California. Within the three available job types, evidence shown by Techtarget, have alluded to more Indeed posting by employers to hire data scientists (*Holak*, 2019). The site also said, "The January report from Indeed, a top recruiting site, showed a 29% increase in demand for data scientists year over year and a 344% increase since 2013-- a dramatic upswing" (*Holak*, 2019). This upswing made our hypothesis focus on data scientists in the Tech Industry.

There were several steps that went into the development and testing of this hypothesis. We first looked up data scientist jobs available in both Texas and California in the IT Industry. Within this process, filtering had to be made in both job title and industry. In the job title, I had to filter jobs that are data scientists. In industry, I had to filter out all industries other than the information technology industry. Below, shows a bar chart of the total count of jobs available in the USA that are in the information technology industries and that are also data scientists.



.

Next, we sorted out the number of jobs in each state and the percentage that goes along with the type of filter which is, IT industries for Data Scientist jobs.



Focusing only on California and Texas. We ran a one-sided-z-test to see if the difference in jobs was statistically significant. We began by moving the decimals. We validated the assumptions of a one-sided-z-test and found that Texas's proportion is at .1206(p) and California's proportion is at .1407(q) The total number of IT Data Scientists in all 50 states is 199(n).

$H0$: $p \geq 0.1206$

Ha: $p < 0.1206$

	A	B	C
1	Same size	199	
2	test proprtion (p)	0.1206	
3	q(1-p)=	0.8794	
4	sample propion (p̂)=	0.137139	
5	z-test	0.7164202	
6	p-value	0.763134	
7	Significance level	0.05	
8		.7631> .05	

The one-tail z-test had a significance level of 5%. The p-value was .763134. Since .763134 > .05, there is not enough evidence to support the claim that at least 12.06% of IT Data Scientist applications can be found in Texas. In section B5, the formula is a z-statistic test. The P-value was found by using the Norm.s.Dist function. Which returns the cumulative probability for the normal distribution with mean = 0 and standard deviation = 1. The p-value is greater than the significance level, so we fail to reject the null hypothesis. There is no sufficient evidence to truly say that more jobs will be available in Texas rather than California.

In conclusion, even though there are more data scientists in the information technology industry in California, there is a trackable shift with more IT companies moving to the Texas area due to the researcher found by *Folger*. The attractiveness that Texas has with money-saving accommodations seems to be beneficial for IT companies. This is important due to the movement of companies coming and going throughout different states. Due to the early evidence of IT companies moving to texas, we can assume that even if we can't see an effect of change in job applications now growth could still happen in the near future. This hypothesis is important because as our group goes out into the workplace, we now know that even though more companies are moving out of epicenters, action/jobs in data scientists in the IT industry can still be more commonly found in epicenters rather than other states.

Research Qs

Is there statistical significance there is a positive relationship between the state level of cost of living and job salary rate. How does the relationship shift between the data scientist, data analyst, and business analyst career?

There is a positive relationship between the salary average and cost of living per state as well as the city. The city regression offers a more specific measure to define how much the cost of living affects the salary average. The cost of living for city and state has been collected via two sources, Travel Safe Abroad and the Institution of Missouri. To create both of these regressions, variable junior business analyst, junior data analyst, junior data scientist, senior business analyst, senior data analyst, and senior data scientist needed to be created. To create this regression, any data collected that did not have an average salary posted was left out of the analysis. The cost of living of the city regression has fewer observations than the state because some of the jobs were remote and could not give a specific location to the salary they were going to make. On some occasions, it would have the city listed but not the state in which there are many cities that have the same name throughout the United States.

From both regressions, it is found that junior business analysts and junior data analysts have a negative relationship with the salary average. Senior business analysts, senior data analysts, junior data scientists, and senior data scientists all have positive relationships with the salary average. You can tell from both regressions that senior data scientists have the highest average salary as the coefficient in both regressions is the largest. However, we can conclude that this analysis makes sense as the skill set needed for a data scientist position is more demanding than that of a business analyst or data analyst. It also makes sense that junior positions would make less than a senior position. The cost of living per city is the better regression for this analysis as it gives a better understanding and significant coefficient. The coefficient is 1,023.13 in the city of cost of living whereas in the state cost of living it is only 443.74. In this case, it shows a better representation of how the cost of living in large cities can be higher than in the overall state which would affect the average salary. Lastly, you can note that almost all of the variables included in this regression are significant as the p-value is less than .05. This shows that every variable in this analysis is important to the change in the salary average. From our observations, we can conclude that the cost of living of city affects the average salary and should be considered as an important factor when studying any effects on salary.

```
. rename costoflivingperstate livingstate

. reg salary_avg livingstate jba sba jda sda jds sds, cformat (%8.3f)

      Source |       SS           df       MS      Number of obs   =    11,833
-------------+----------------------------------   F(7, 11825)     =    281.84
       Model |  1.8372e+12         7  2.6245e+11   Prob > F        =    0.0000
    Residual |  1.1011e+13    11,825   931204786   R-squared       =    0.1430
-------------+----------------------------------   Adj R-squared   =    0.1425
       Total |  1.2849e+13    11,832  1.0859e+09   Root MSE        =     30516

------------------------------------------------------------------------------
  salary_avg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 livingstate |    443.744     12.698    34.95   0.000      418.853     468.635
         jba |  -1.83e+04   4599.665    -3.98   0.000    -2.73e+04   -9295.450
         sba |   8405.417   1361.396     6.17   0.000     5736.857   11073.978
         jda |  -2.06e+04   4021.319    -5.13   0.000    -2.85e+04   -1.28e+04
         sda |  10325.262   1668.254     6.19   0.000     7055.209   13595.316
         jds |  10434.581   4667.856     2.24   0.025     1284.814   19584.347
         sds |  40484.926   1611.800    25.12   0.000    37325.532   43644.320
       _cons |  38445.909   1462.576    26.29   0.000    35579.019   41312.799
------------------------------------------------------------------------------

. reg salary_avg livingcity jba sba jda sda jds sds, cformat (%8.3f)

      Source |       SS           df       MS      Number of obs   =    11,736
-------------+----------------------------------   F(7, 11728)     =    335.84
       Model |  2.1281e+12         7  3.0401e+11   Prob > F        =    0.0000
    Residual |  1.0617e+13    11,728   905241344   R-squared       =    0.1670
-------------+----------------------------------   Adj R-squared   =    0.1665
       Total |  1.2745e+13    11,735  1.0860e+09   Root MSE        =     30087

------------------------------------------------------------------------------
  salary_avg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  livingcity |   1023.125     25.619    39.94   0.000      972.908    1073.342
         jba |  -1.76e+04   4535.302    -3.89   0.000    -2.65e+04   -8757.673
         sba |   8062.005   1358.658     5.93   0.000     5398.809   10725.201
         jda |  -2.44e+04   3968.205    -6.15   0.000    -3.22e+04   -1.66e+04
         sda |  12646.221   1646.160     7.68   0.000     9419.474   15872.968
         jds |   5167.662   4608.278     1.12   0.262    -3865.329   14200.653
         sds |  38954.642   1593.895    24.44   0.000    35830.343   42078.940
       _cons |  10273.149   1981.712     5.18   0.000     6388.664   14157.633
------------------------------------------------------------------------------
```

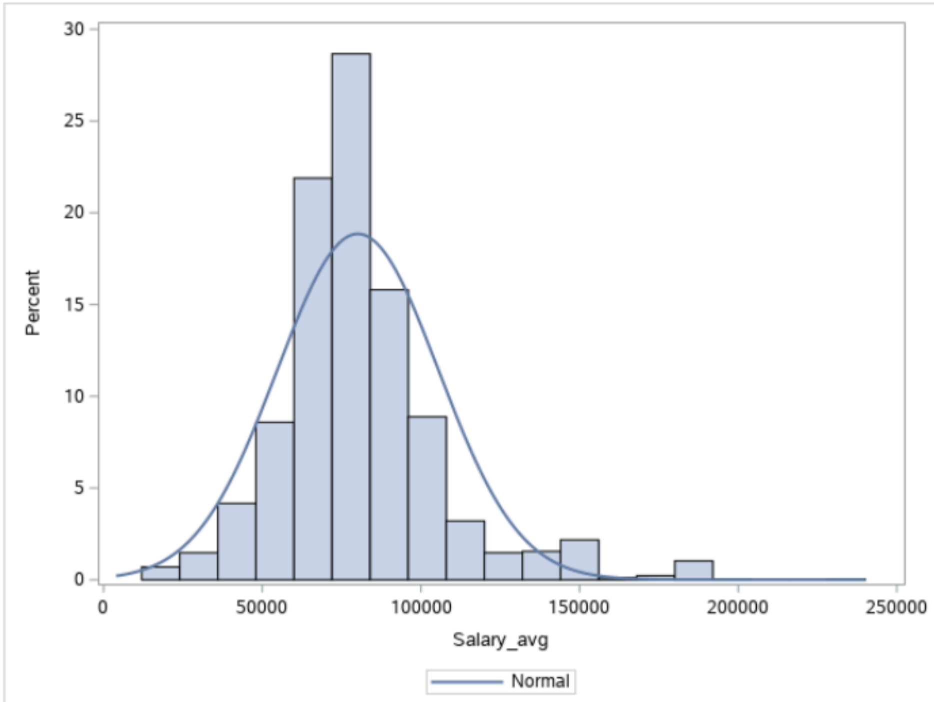
Does the salary of the data scientist and data analyst jobs differ significantly?

While only one word separates the data scientist and data analyst professions, the roles and responsibilities of these professions differ significantly. While a data scientist finds new ways to capture and analyze data, data analysts do the actual analysis. Data analysts make sense of existing data and assist in a company's decision-making process while data scientists are often referred to as part mathematicians, part computer scientists, and part trendspotters. Due to the difference between these professions, we were interested in finding potential trends in the average salaries. We sought to answer whether the average salary of a data scientist and data analyst differ significantly.

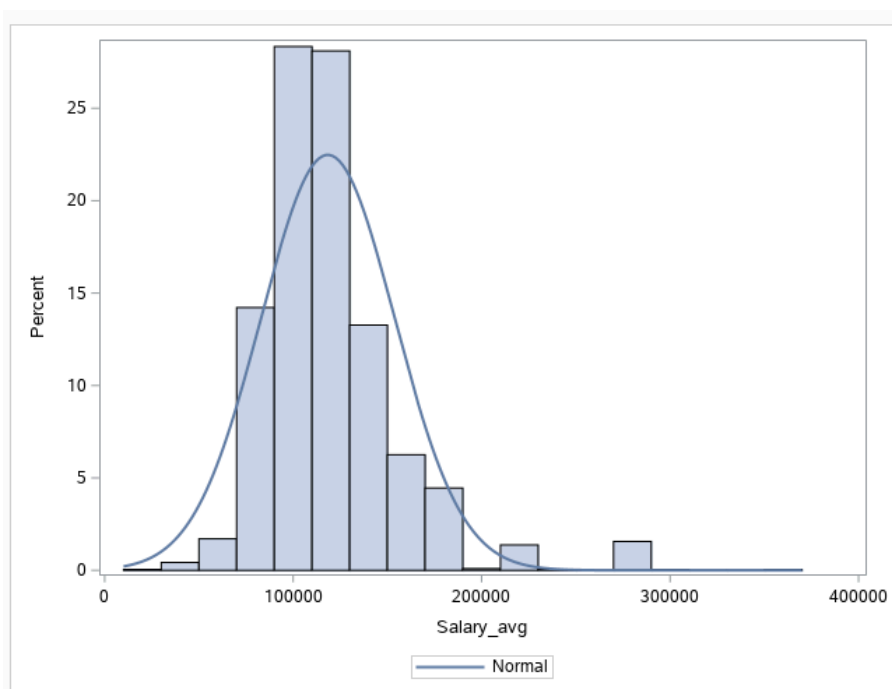
The analysis for this question was completed with Excel, SAS, and a graphing calculator. We started by comparing the sample sizes of the data analyst and data scientist roles in Excel. It was important to ensure the sample sizes of the two groups were similar because a comparison analysis would prove to be insignificant if the sample sizes differed by a significant degree. The sample size of the data scientist was 4,046 while the sample size of the data analyst was 5,654. These sample sizes are relatively similar, but we also decided to look at the sample sizes in comparison with the other profession we analyzed, business analyst. This sample size differed significantly from the other two with a sample size of 10,334. Thus, we figured continuing with a comparative analysis between the data scientist jobs and data analyst jobs would be appropriate.

In SAS, we found the mean average salary of the data analyst to be 62,195.18 while the mean average salary of the data scientist was 89,978.77. Given these numbers, the mean of the average data scientist salary is, on average, more than $20,000 than the average data analyst's salary. Further, a data analyst, on average, is making 70% of what a data scientist is making. However, statistical analysis must be conducted before making any real conclusions about the pattern of salaries.

- Data Analyst Salary



- Data Analyst Salary



-

Both salaries appear to be normally distributed with a slight right skew.

Given the data we have, a 2 sample unpooled t-test was conducted to compare the average salaries. This test was chosen because the variances of these two samples differ and we don't have the population standard deviation or mean. We had to check the other assumptions of this test before continuing. The assumptions of a 2 sample unpooled t-test are the following: data values must be independent, data in each group must be obtained via a random sample from the population, data in each group are normally distributed, and data values are continuous. After checking these assumptions and finding they were all met, we were able to continue with the test.

The hypothesis of this 2 sample unpooled t-test assumed the means of the salaries are equal. With 6630 degrees of freedom, the test statistic was -25.86 producing a P-value of approximately 0. Given this information, we have evidence to assume the average salaries between the two careers are not equal. This aligns with our understanding of the jobs, as the job of a data scientist requires in-depth coding knowledge, something not necessarily required for a data analyst.

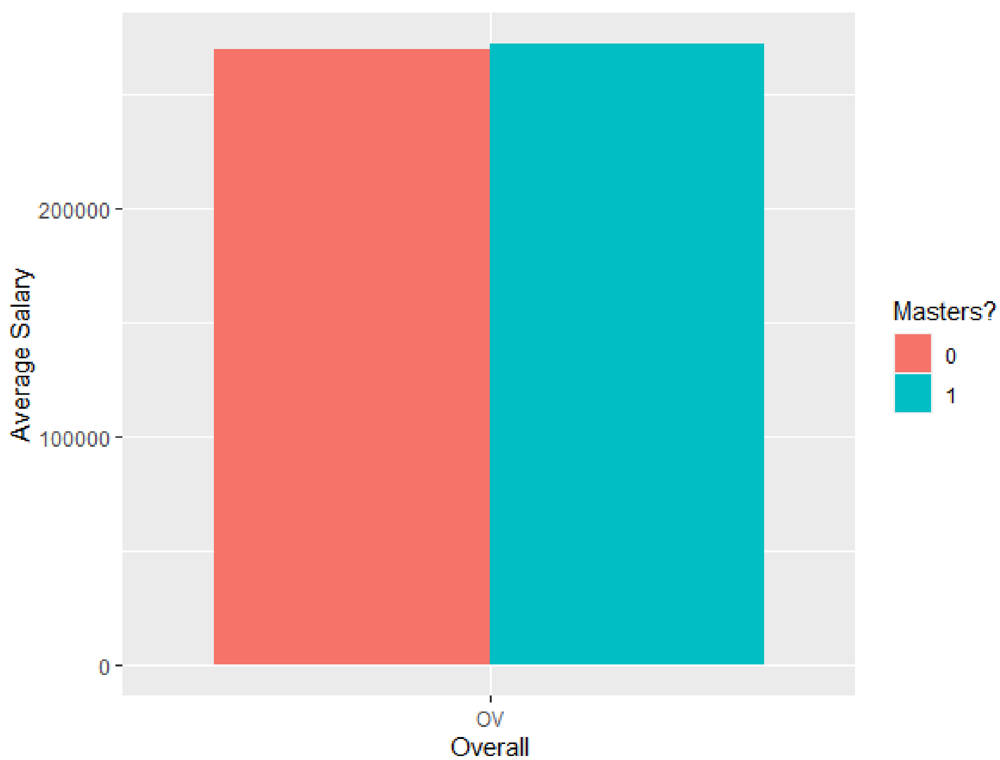
How does the requirements needed for a position affect the salary? By how much does a salary increase if the job requires a master's degree?

One of the things almost all undergraduate students have to consider at some point in their college career is whether or not they intend to further their education and obtain a master's degree. We wanted to take this chance with the data we collected and try to determine the impact that obtaining a master's can have on our future salaries once we break into the workforce. After breaking down our data and determining which postings required a master's degree we found that 21% of our overall data required a master's degree. But to break that down further to examine the frequency of master's requirement among the different job types we found that Data Scientist job positions required Master's degrees more than the other 2 job types. Specifically, 40% of all

Data Scientist positions in our data require a master's degree. While 16% of all Data/Business Analyst positions require a master's. This is likely because the Data Scientist positions require more technical skill and knowledge so they expect more education. Because of this difference though we did however also want to look into the impact of a master's on the average salary for our three different job types.

To start, we first examined whether the estimated coefficient for master's degree would be positive or negative. Based on common assumptions it would make sense for jobs requiring a master's to pay more as those employees are then more qualified but we wanted to get a basic idea if this was the case based on our data. We calculated the overall average salary of all our job postings that were not determined to require a master's degree and compared that to the overall average salary of our job postings requiring a master's degree (see graphic below).

Overall Average Salary Comparison Table



What we determined from this is that while the average overall salary for jobs requiring a master's degree is higher, it is only higher by a very small amount (less than 3000). We still wanted to determine the exact impact on the salary difference. Next, we looked to create a linear multiple regression model using variables from both our original data and other variables from outside our initial data to try and develop an accurate model and determine what the impact of having a master's degree has on average salary and whether its impact is significant. The following table (**Master's Degree Regression Table**) displays our initial regression results broken down into four models based on the different job types. Model 1 is the Data scientist postings, Model 2 is Data Analyst postings, Model 3 is Business Analyst postings, and Model 4 is all job postings combined.

What we found via these results was rather interesting but we will get into that in a moment. First, we need to examine the significance of the model and more specifically the impact of the Master's requirement. On all 4 models, regardless of the specific job type, the Master's dummy variable is statistically significant on at least the 10% level if not the 5% level (based on a 1-tailed test).

What distinguishes the three models though is their estimated coefficient of Mast_Req, because for 3 out of the 4 models that impact is positive, indicating a higher average salary for jobs requiring a master's degree. But when you look specifically at the estimated impact for Data Scientist jobs the estimated impact of the master's degree is negative. Thus, our models indicate overall jobs that require a master's degree do have a higher salary, however, upon distinguishing between different job types, Data Scientist jobs with this requirement have a lower average salary than those that do not require a Master's.

There are some issues with these models that need to be addressed to fully interpret the significance of our findings. These models are not largely accurate (with variance hovering around 10 to 20%), which means there is a significant amount of variance in the average salary that is not determined by our examined variables. Another issue is that like with our original dummy variable for bachelor's requirement, this derived variable comes from examining the job qualifications and descriptions for the term master's and not the terms "need" or ``requires." While we know the master's requirement is a significant variable, its exact impact is not determinable based on our model. Thus, we are able to determine master's degrees have a positive impact on salary overall even if not by a large amount.

Master's Degree Regression Table

Table

	Dependent variable:			
	Salary_avg			
	(1)	(2)	(3)	(4)
Company_Rating	6,827.821***	6,631.689***	4,532.223***	8,114.660***
	p = 0.0002	p = 0.000	p = 0.00000	p = 0.000
Salary_Estimated_GD	-12,620.310***	1,559.940	-6,158.768***	-3,177.586***
	p = 0.00002	p = 0.292	p = 0.00001	p = 0.005
Salary_Estimated_Company	12,632.240***	2,609.698	-503.487	976.515
	p = 0.0004	p = 0.108	p = 0.720	p = 0.442
Full_time_job	-270.084	1,143.770	740.063	-4,003.315***
	p = 0.887	p = 0.273	p = 0.406	p = 0.00000
Part_time_job	2,532.073	-1,698.914	-13,560.140***	-11,254.820***
	p = 0.759	p = 0.647	p = 0.0002	p = 0.0003
Temp_Contract_job	-23,065.850***	141.566	-11,789.870***	-12,396.910***
	p = 0.000	p = 0.947	p = 0.000	p = 0.000
Size.L	7,456.340***	4,364.743***	-2,610.373**	2,089.935*
	p = 0.003	p = 0.006	p = 0.039	p = 0.062
Size.Q	2,234.831	2,593.096	805.275	3,503.083***
	p = 0.412	p = 0.118	p = 0.533	p = 0.003
Size.C	1,722.724	976.501	1,924.488*	1,866.982*
	p = 0.470	p = 0.488	p = 0.092	p = 0.068
Size4	1,879.251	2,307.357*	2,248.024**	2,939.232***
	p = 0.439	p = 0.089	p = 0.044	p = 0.004
Size5	-7,338.881**	1,099.707	990.005	50.357
	p = 0.011	p = 0.481	p = 0.423	p = 0.965
Mast_Req	-3,403.686**	2,068.603*	2,230.164*	7,244.853***
	p = 0.038	p = 0.093	p = 0.068	p = 0.000
`Cost of Living`	317.329***	165.683***	105.910***	182.302***
	p = 0.000	p = 0.000	p = 0.00004	p = 0.000
`Average Ind Income`	0.214	0.207**	0.238***	0.188***
	p = 0.159	p = 0.018	p = 0.002	p = 0.005
Constant	44,798.270***	15,328.700**	43,935.150***	26,467.400***
	p = 0.0002	p = 0.023	p = 0.000	p = 0.00000
Observations	1,155	1,485	2,319	4,959
R^2	0.198	0.090	0.096	0.100
Adjusted R^2	0.188	0.081	0.090	0.098
Residual Std. Error	26,862.500 (df = 1140)	19,102.240 (df = 1470)	19,836.620 (df = 2304)	25,474.130 (df = 4944)
F Statistic	20.096*** (df = 14; 1140)	10.402*** (df = 14; 1470)	17.469*** (df = 14; 2304)	39.270*** (df = 14; 4944)
Note:				*p<0.1; **p<0.05; ***p<0.01

Significance of our findings

As many of us are graduating seniors, we sought to uncover truths behind our future careers through our research project. Our crafted hypotheses and the research questions helped us gain a better understanding of the job market we are entering. With our first hypothesis, we found that although more data scientist jobs in the IT industry are available in the business epicenter of Silicon Valley, California, Information Technology companies seem to be shifting out of California and into Texas. This will be interesting to consider as we search for places to move after graduation day. Beyond physical location, we were also interested in the salaries of our future careers. We were interested in finding that a career entitled, data scientist makes, on average, more money than a data analyst. This is likely due to the heavy coding knowledge required for this career. We also found the cost of living in cities is positively correlated to the average salary with senior titles making the most. Therefore, if we decide to move to a city with a higher cost of living, we will likely be offered a higher salary on paper. Lastly, continuing with the motivation of uncovering the reality behind salary offerings, we discovered jobs requiring a master's degree offer higher average salaries, though the salary difference is actually pretty minuscule. This is an interesting point to consider as we weigh the true benefits of obtaining a graduate degree. This research project helped us to address many of our concerns upon entering the job market in either the data scientist, data analytics, or the business analytics field and give us a better understanding of where we may be in a year from now.

References

Atlanta. Atlanta - Place Explorer - Data Commons. (n.d.). Retrieved April 21, 2022, from https://datacommons.org/place/geoId/1304000

Cost of Living Data Series. Cost of Living Data Series | Missouri Economic Research and Information Center. (n.d.). Retrieved April 26, 2022, from https://meric.mo.gov/data/cost-living-data-series

Cost of Living in United States (2022 updated). Travel Safe- Abroad. (2022, February 11). Retrieved April 26, 2022, from https://www.travelsafe-abroad.com/cost-of-living/united-states/

Klein, M. (2021, August 31). *Young people rushing back to live in NYC after pandemic exodus*. New York Post. Retrieved April 20, 2022, from https://nypost.com/2021/08/28/young-people-flocking-back-to-big-apple/

New York City. New York City - Place Explorer - Data Commons. (n.d.). Retrieved April 21, 2022, from https://datacommons.org/place/geoId/3651000?utm_medium=explore&mprop=count&popt=Person&hl=en

Team, H. (2021, November 3). *Reasons why many New Yorkers are moving to Atlanta*. Haven Lifestyles. Retrieved April 19, 2022, from https://havenlifestyles.com/reasons-why-many-new-yorkers-are-moving-to-atlanta/

Folger, J. (2021, September 10). *Why Silicon Valley companies are moving to Texas*.

Investopedia. Retrieved April 19, 2022, from https://www.investopedia.com/why-silicon-valley-companies-are-moving-to-texas-5092782

Holak, B. (2019, January 31). *Demand for data scientists is booming and will only increase*. SearchBusinessAnalytics. Retrieved April 19, 2022, from https://www.techtarget.com/searchbusinessanalytics/feature/Demand-for-data-scientists-is-booming-and-will-increase

Alternatives to Silicon Valley surprise data scientists. SMU. (2021, January 5). Retrieved April 26, 2022, from https://datascience.smu.edu/blog/comparing-tech-hub-cities-guide/

Galvanize, & *, N. (2020, October 8). *8 top cities to be a data scientist*. Galvanize Blog. Retrieved April 26, 2022, from https://blog.galvanize.com/8-top-cities-to-be-a-data-scientist/

Reuter, D. (2021, December 1). *Tesla's move to Texas from California is officially complete*. Business Insider. Retrieved May 2, 2022, from https://www.businessinsider.com/tesla-texas-headquarters-move-from-california-officially-complete-elon-musk-2021-12#:~:text=It's%20official%3A%20Tesla%20is%20now,Texas%2078725%2C%22%20it%20said.

Cost of living data series. Cost of Living Data Series | Missouri Economic Research and Information Center. (n.d.). Retrieved March 27, 2022, from https://meric.mo.gov/data/cost-living-data-series

Average income by state, plus median and top 1%. DQYDJ. (2022, February 6). Retrieved March 27, 2022, from https://dqydj.com/average-income-by-state-median-top-percentiles/