What Technical Skills Would Help to Generate a Higher Pay for Data Analysts?

Research Question: What technical skills would help to generate a higher pay for data analysts?

It's important yet challenging to answer this question as many factors affect the wide range of salaries in Data science, especially for Data Analyst. One of the factors is the wide variety of jobs in the market. Data analyst jobs are available in various sectors, including healthcare, finance, information technology, and business services. Different businesses may value Specific abilities or traits differently, making it difficult to identify universal causes for greater compensation. A data analyst in the business sector may have a salary influenced by their ability to analyze what drives sales and optimize inventory, which could be different from the considerations in other industries. Identifying universal causes for greater compensation across these varied job markets is challenging due to the industry-specific nuances. Each sector values different abilities and traits in data analysts, making it essential to consider the specific requirements and expectations of the industry when determining salary ranges.

Knowledge of technical skills may be more helpful with higher salary range jobs. Someone with greater knowledge of technical skills can apply to a broader range of employment, increasing the probability of landing them. Technical skills mainly consist of coding languages like Python, R, SQL, etc., and software like Tableau, Power BI, Excel, Alteryx, etc. Knowledge and experience of technical skills are the basis to fit in a higher-paying job position. While skills can be narrowed down, experience with those specific skills is difficult to measure.

In addition, changing industry trends may affect the wide range of salaries in data science. The field of data analysis is ever-changing, with new tools, methodologies, and technologies being developed regularly. Furthermore, data analyst salaries might vary significantly based on region. The cost of living, the need for data analysts, and the local economic conditions all play a factor. For example, we compare the salaries of Data Analysts in Silicon Valley with those in a smaller city.

It is also crucial to note that seniority and experience are some of the factors considered by most employers. Experience is frequently associated with more excellent compensation. While technical abilities are essential for data analysts, soft skills such as communication, problem-solving, and teamwork

also impact earning potential. In some cases, the nature of roles and obligations may also dictate the possible salaries for data analysts. The specific tasks and responsibilities that fall under the purview of data analyst roles vary greatly.

Lastly, access to accurate and comprehensive salary data is essential for setting competitive compensation packages. Limited or outdated data can result in underpayment or overpayment of Data Analysts, affecting employee satisfaction and retention. Collecting complete data on salaries and the specific skills and factors that influence them might be difficult. Many salary surveys and job advertising are too general.

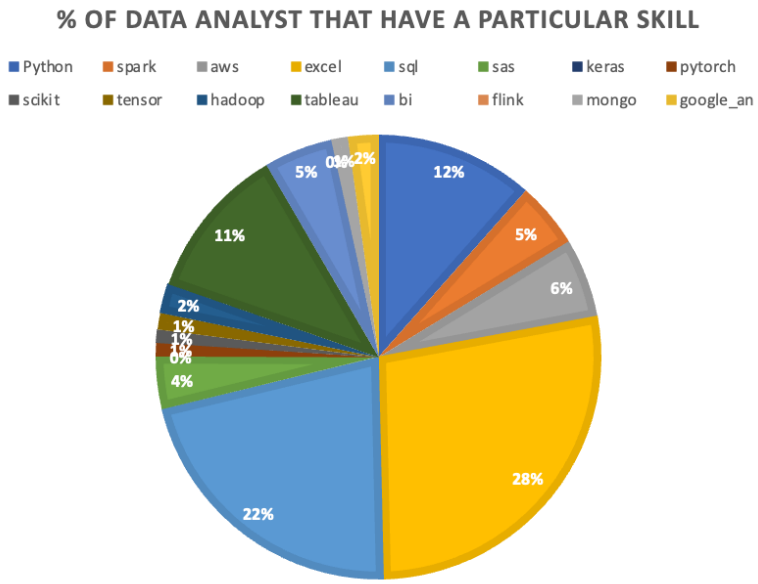
Data

We were able to find and utilize this data set from Kaggle. We had a question about looking into our and our peers' futures in the workforce and seeing what we should expect regarding salary. We decided to use Kaggle over collecting data ourselves for two reasons: we wanted a large data set (more than 30 observations), and we inherently do not have access to the data for the question we tried to answer from the current or past MSBA students. Kaggle is easy to use, and most of the data is already cleaned and ready for statistical analysis in a .csv format.

We found a data set initially that showed over 600 observations of time series data looking at data analytics realm roles (2021-2023) with variables such as the year that the salary was paid, position title, experience level, employment level (full-time or part-time), salary in US dollars, company location, and company size, but our question was either too complicated or too general so we decided to go with the current data set we are using. Our current data set has 742 observations of data showing data analytic realm roles in 2021. We were given a generous amount of variables at 42, but we decided to steer our question toward what skills affect average salary. We utilized the average salary dependent variable and our skills independent variables, including the different software and hard skills such as Python, SQL, Tableau, etc.

Statistical Analysis

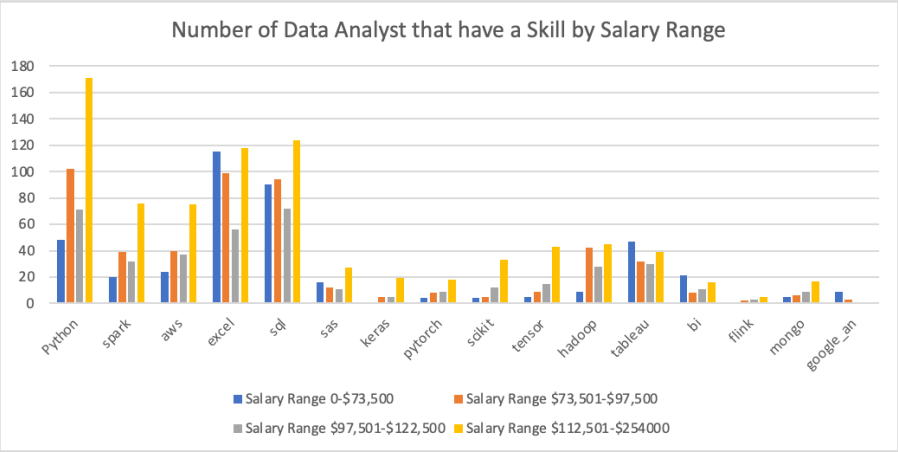
We started our analysis process by using Excel to get a general understanding of the data. From the pie chart below, out of 742 data analysts in our dataset, it is evident that Excel (28%), Python (22%), and SQL (12%) are the top three most common and desired skills for data analysts.



The salary range was determined using the quartile function, then VLOOKUP to assign each average salary into the appropriate salary range. The chart below shows the number of data analysts with technical skills within each salary range. For most of the skills, the higher the salary range, there are more data analysts that possess the technical skill. The high salary range of $112,501-$254,000, has the highest count for almost all the technical skills listed, but Python, Excel, SQL, Spark, and AWS are the most

apparent ones. This can suggest that the possession of these skills may impact a data analyst's salary range positively.

Furthermore, we would also be interested to know what technical skills will be necessary if we wanted to work in a certain sector. We've picked out the most common sectors for data analysts: information technology, business services, insurance, healthcare, and finance. As you can see below, Python, Spark, AWS, Excel, and SQL are still the most common skills to have. However, the information technology, business services, and healthcare sectors can be seen to require more technical skills than the



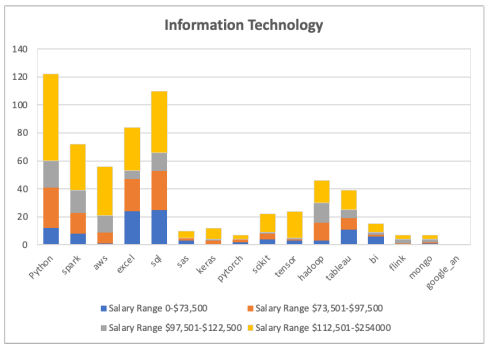
insurance and finance sectors. The insurance and finance sector doesn't have that many data analysts with Spark and AWS skills, whereas the other three sectors do.

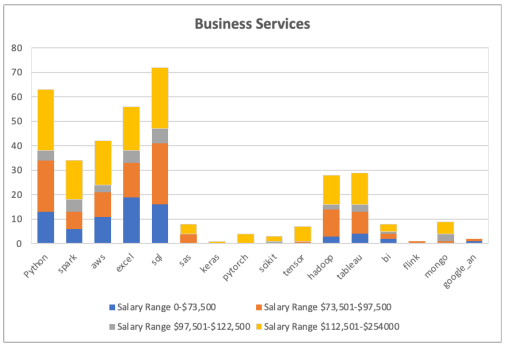
Moreover, data analysts' earnings that fall under the higher salary range of $112,501-$254,000 for all of these sectors, except healthcare and finance, have clear knowledge of and acquired those top five

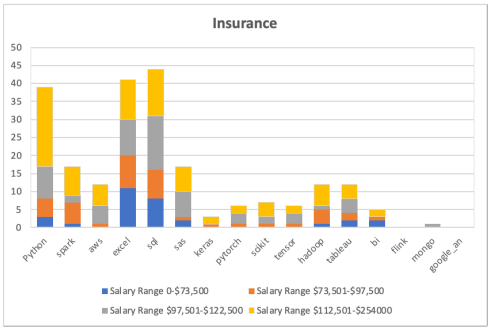
skills. These skills are crucial for data analysts with a higher average salary working in the information technology, business services, and insurance sectors.

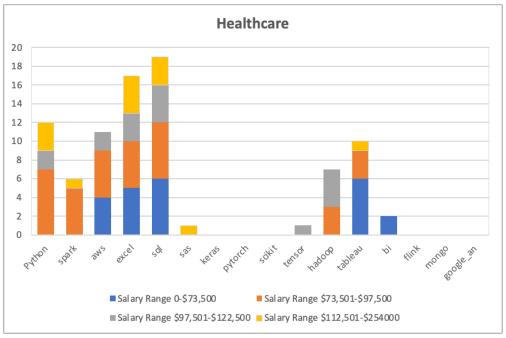
Regression Analysis

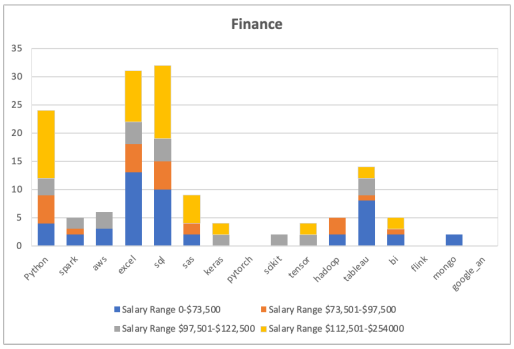
As part of our statistical analysis, we conducted a multiple regression analysis in order to establish a











relationship between a set of independent variables and the dependent variable. A multiple regression analysis aims to find an explanation for some unexplained variations. Our group sought to explain, if any, the relationship between specific technical skills such as Python, SQL, Excel, and others and the average income and figured out whether the regression model can explain any relationship.

Our first step in our multiple regression analysis was to understand if the independent variables in the model can explain the variation of the dependent variable. We look at the Multiple Coefficient of

Determination (R Square) for that. While our Coefficient of Determination is greater than zero, implying that our regression model can explain some variation in average income, a higher Coefficient of Determination will indicate a more robust regression model. While there is some explanation for the variation observed in the sample, we now seek to understand whether these observations in our linear regression model are statistically significant.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.430[a]	.185	.167	34.20965

a. Predictors: (Constant), Google=1.0, Scikit=1.0, Mango=1.0, Bi=1.0, Flink=1.0, Sas=1.0, Excel=0.0, Hadoop=1.0, AWS=1.0, Python=1.0, Pytorch=1.0, Tableau=1.0, Keras=1.0, SQL=1.0, Spark=1.0, Tenspr=1.0

b. Dependent Variable: AVG_Salary

Our second step was to test for significance. The essence of this test is to explain that while our Coefficient of Determination is greater than zero, is it statistically significant from zero at a five percent significance level? We conducted a hypothesis test where our null hypothesis (Ho) is as follows: The regression model does not explain any of the total variation. If we reject the null hypothesis, we can prove that our Coefficient of Determination is statistically significant from zero. We first look at our F-statistic value and compare that to the F-critical value, and we reject the null hypothesis if the F-statistic value is greater than the F-critical value. As shown in the ANOVA table, our F-statistic equals 10.285, and the F-critical value equals 1.5705. In this test, we reject the null hypothesis and conclude that the regression model does explain a proportion of the total variation.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	192588.486	16	12036.780	10.285	<.001[b]
	Residual	848467.585	725	1170.300		
	Total	1041056.072	741			

a. Dependent Variable: AVG_Salary

b. Predictors: (Constant), Google=1.0, Scikit=1.0, Mango=1.0, Bi=1.0, Flink=1.0, Sas=1.0, Excel=0.0, Hadoop=1.0, AWS=1.0, Python=1.0, Pytorch=1.0, Tableau=1.0, Keras=1.0, SQL=1.0, Spark=1.0, Tenspr=1.0

Conclusion

With 742 data analysts from our dataset, we analyzed the insight of the skills used throughout the different levels of salary range and sectors. Firstly, we can easily find that the most common skills data analysts require are Python, Excel, SQL, Spark, and AWS in higher salaries pool (refers to Chart #2) and within all the sectors (refers to Chart #3). This indicates that for those who want to become a data analyst, no matter what sector, these five skills are what you want to be a master with. In addition to these five most commonly required skills, if you are interested in pursuing a career in information technology, business services, or healthcare sectors, becoming a master in visualization skills such as Tableau or BI will be a plus.

Furthermore, we built in between skills and salaries from the linear regression model. We found that skills do contribute to higher salary; however, this regression model also gives us a high residual value and a low coefficient correlation, which means besides skills, there are other factors that contribute to salary that are undefined in our model. This makes sense because other variables can impact the salary, such as region, years of working, etc., but our regression model is to analyze between salary and skills.

Works Cited

Bhathi, N. (2021, December 28). Data Scientist Salary. Retrieved from https://www.kaggle.com/datasets/nikhilbhathi/data-scientist-salary-us-glassdoor

Anderson, D., Sweeney, D, (2019). Statistics for Business Economics 14e. Cengage