****

**ABSTRACT**

In the twenty-first century, data DNA, which supplies crucial knowledge, insights, and capacities, has become an important aspect of all data-based organisms, ushering in the era of big data and data economies. For an in-depth comprehension of data and the DNA of its residents, an examination of new data science fields and their pillars is required. Although there is much debate on whether big data is merely hype, data science is still in its infancy, with many new challenges and possibilities arising, or data science is driven by research, innovation, business, profession, and academia. This article presents a thorough examination of the relationship between the title of data scientist and the number of data scientists employed. Salary, education, skills, company, and location were all factors considered in this study. This information will help anticipate if data analysts' salaries will rise over time, which is becoming more important in today's environment, as well as gender equality.

Data science encompasses a set of principles, problem definitions, algorithms, and processes for extracting non-obvious and useful patterns from large data sets. Many of the elements of data science have been developed in related fields such as machine learning and data mining. Data mining generally deals with the analysis of structured data and often implies an emphasis on commercial applications. Data science takes all of these considerations into account but also takes up other challenges, such as the capturing, cleaning, and transforming of unstructured social media and web data; the use of big-data technologies to store and process big, unstructured data sets; and questions related to data ethics and regulation.

**Keywords: Data, Big data, Innovation, Data Scientists**

**INTRODUCTION**

The technique of analyzing large volumes of unstructured and structured data to uncover patterns and extract relevant information is known as data science. Data science is an interdisciplinary method to gaining relevant insights from today's businesses' massive and expanding data. To uncover patterns and make educated decisions for shareholders, data scientists must prepare data for analysis and processing, examine advanced data, and give outcomes. Cleaning, gathering, and altering data to prepare it for a certain sort of processing is known as data preparation.

Algorithms, analysis, and artificial intelligence models must all be developed and used in the analysis process. It's backed up by software that separates data from patterns and translates those patterns into estimations to assist businesses in making better decisions. These estimations' correctness must be checked by well-prepared tests and experiments.

Furthermore, survey findings should be made public by utilizing data visualization tools to uncover patterns and trends. How do they pull it off? These are just a few of the questions we think about from time to time. In practice, projects like this are impossible to do without data. Data science is all about using data to solve issues.

The problem probably is one of decision-making, such as figuring out which emails are spam and which aren't. Maybe a product recommendation, like which product will sell the most? Or making predictions, such as who will be President of the United States in the future? As a result, the fundamental role of a data scientist is to interpret data, extract useful information from it, and apply it to problem-solving.

In this study, we're looking at the titles of data analysts and scientists in connection to their incomes, education, skills, firm, and location in India to determine whether there's a link between these factors and the differences we detect. Key factors that are driving the data science platform market growth include the growing inclination of organizations towards data-intensive business strategies and the rising adoption of advanced technologies in creating opportunities for the data science platform vendors.

**Review of Literature**

In this section, we will analyze existing literature sources related to (a) probability to get a job in a Data science job in top companies, (b) Skills required for a Data Science job, and (c) What skills the Data science industry expects from schools. We have identified a few papers that describe various aspects of Data science jobs like skills required for Data science jobs, the probability of getting a job in a company.

**Predicting the Probability and Salary to Get Data Science Job in Top Companies**

This article is about the research done to Predict the probability of getting a data science job in Fortune 500 companies through predictive analysis. In this research, they have followed the Cross-Industry Standard Process for Data Mining with the first five phases. The phases are Business Understanding, Date understanding, Data preparation, Modeling, and Evaluation. The target in Business understanding research is to predict a candidate's probability to get a job with the expected salary in different company categories and to analyze the significance of variables and find out which variable affects probability most. In Data understanding, date features like username, Education, current and past job, salary information, job description information were collected from LinkedIn and Glassdoor. After the data extraction, the data was cleaned, sorted, and stored for analysis. The data were modeled with the Gradient Boosting Decision Tree (GBDT) classifier from Python sklearn package. The overall model accuracy range was between 78% to 94%. The result after the analysis shows that the data science skills include Python, Data Mining, R, Statistics, etc. play a crucial role to enter the top company.

The limitation of this research was that.

1. Limited sample size: Only 3000 profiles in total were collected, and only 180 people worked or were working in the top companies out of the data collected
2. Their research and analysis were based on LinkedIn and Glassdoor. There are many people who work in top companies who are not on LinkedIn and Glassdoor.
3. There was an assumption that all the people tried to get into the top companies when doing the analysis. In fact, some of them never tried to apply for these companies.

**Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis.**

During the last several years, we have observed an exponential increase in the demand for Data Scientists in the job market. As a result, training, courses, books, and university educational programs (both at undergraduate, graduate, and postgraduate levels) have been labeled as “Big data” or “Data Science”. In this research paper, Business Data Analytics and Data Science domain and skills were identified, and a comparative analysis was don’t to understand the skills required for these jobs. By reviewing the literature sources, they develop a model to systemize the expected knowledge domains and skills for these two fields. Data for business data analytics and data science jobs were collected from online job-related websites, and the data was analyzed using text mining methods. The result from the text mining showed some similarities as well as some differences between the required knowledge domains and groups of skills for these two job profiles. These results provide insights that are vital in designing curriculum and training in the evolving business analytics and data sciences areas and enable professionals to sharpen their skills that are aligned with job market requirements. The question addressed in this research paper is: What are the groups of knowledge and skills required for Business data analytics and Data science professions? What are common and distinguishing groups of skills between the two professions? And What are important insights that can be revealed to business organizations and educational programs in terms of Business data analytics?

In this paper, a research study was done to know the knowledge domains and groups of skills relevant to Business data analytics and Data science, and these data were compared based on the job-related skill requirements and qualifications for these professions. They analyzed the Business data analytics and data science skills in conjunction with several important job-related characteristics and attributes, i.e., years of experience (e.g., 1-3 years, 3-5 years, 5+years), position level (entry, middle, and senior-level), an education level (bachelor, masters, Ph.D.), job location. The entire data list of knowledge domains for Business data analytics and Data science was extended to four main domains, including Technical, Analytical, Business, and Communication. They developed a new research model that incorporates these domains along with the groups of skills. Top-ranked skills were identified and titled together as a group of skills. The result obtained after the research shows that the group of skills in business data analytics and the Data science profession is useful to Employees and specialists for better understanding and for sharpening their skills in Business data analytics and data science. Employers developing job requirements for big data job positions and Academic and professional education for developing new or improving existing analytics and data science programs and certificates. Most companies hiring data scientists are in California, Washington, New York, govt offices, and so on.

**Beyond Data Scientists: A review of Big Data Skills and Job Families**

This article is about the research done on a large number of job posts published online. The questions addressed in this research paper are to identify the most important ‘job families’ related to Big Data and recognize homogeneous groups of skill sets that are most sought after by companies and characterize each job family with the appropriate level of competence required within each Big Data skillset.

In design, methodology, and approach, they implemented a semi-automated, fully reproducible, analytical methodology that can cope with the significant amount of job posts obtained by scraping some of the most popular job search online portals. Job families were identified through the expert evaluation of the most important keywords appearing in job posts’ titles. Skillsets were obtained by using Latent Dirichlet Allocation. Unsupervised machine learning algorithm used for text classification. Job families were characterized through a measure of the relative importance of each skill set. This study represents one of the first attempts to classify jobs in families and describe them in terms of skill requirements by means of a large-scale, semiautomated job post analysis, based on machine learning algorithms. To do this, they propose an original combination of various analytical techniques, which are widely established in previous scientific works.

The result obtained brings clarity to the multifaceted nature of Big Data competency requirements and job role types. Their results can concretely help business leaders and HR managers create clearer strategies for the procurement of the right skills needed to leverage Big Data at best. The structured classification of job families and skill sets will help establish a common language to be used within the job market, through which supply and demand can more effectively meet.

**Stem Jobs: 2017 Update**

This is an article about jobs related to STEM. STEM workers have a high impact on competitiveness, economic growth, and overall standard of living. STEM workers drive innovation and they have the flexible skills needed for the modern economy. They have the ability to adapt to new circumstances and processes which makes them highly sought after. STEM jobs have lower employment rates and have high wages, regardless of educational background or other factors. Businesses and workers are both benefited from STEM jobs. STEM jobs can provide benefits to both businesses and workers.

In this article, data is produced by the U.S. Census Bureau and the Bureau of Labor Statistics on many aspects of the STEM economy. The data from these agencies, the Census Bureau’s Current Population Survey (CPS) and American Community Survey (ACS), gave a close idea of the trends in STEM and helped in performing a detailed analysis of wages and employment. Finding from the research is “STEM: Good Jobs Now and for the future.”The data collected for the research is gender, race, and ethnicity in the STEM economy. This research will use that data to provide a look at the geography of STEM jobs in the United States and the skills needs of the current and future STEM workforce.

Analysis of data from the U.S. The Census Bureau’s American Community Survey and Current Population Survey gave new insights into the growing STEM workforce that makes our economy strong. This report takes a wide-angle approach to STEM jobs and STEM degrees to provide an overview of their impact.

**The Data Scientist: What companies expect from recent graduates and the role of Business Schools**

This research aims to grasp if universities are preparing students to a tolerable degree to cater to companies’ current data science needs, particularly in terms of skills. To do so, the study reviews data science programs in Portuguese business schools and follows a mixed-method approach, by interviewing data science managers and surveying recent data scientists online. it had been concluded that there's room to enhance the skill sets of graduates and suggestions are provided to business schools, who can have a strategic role in bridging the skill gap for future graduates, in what's consensually perceived as strategic, fast-growing professional occupation.

How can Business Schools help in bridging those gaps? - is answered, within the kind of recommendations: ultimately, there's room for business schools to play a job in bridging the talents gap expressed, which managers made clear and survey respondents reflected in their need for added training and not agreeing above a norm of seven that universities prepared them to a tolerable degree. Additionally, from the revision of Portuguese business schools’ data science programs, it appears that the curricula in situ are on the correct track to bridge the gap, with the bulk of them incorporating and addressing most of the requirements and concerns reported within the interviews with managers and from the survey. the perfect curriculum must have: technical quantitative background thoroughly covered; programming taught beyond the fundamentals, promoting efficiency in coding and specializing in practical problems, Python and SQL should be the necessity, the foundations of coding must be a priority to permit them to be easily transferable; must provide 25 business understanding and know-how, particularly in terms of translating data terms into nontechnical terms; must develop student’s soft skills, mainly resilience, problem-solving, critical and analytical thinking, and must bring awareness to ethics with reference to data.

**Methodology/Procedure**

To make informed decisions for data science jobs analysis we have used data analysis methods to strategize our steps. The data analysis process is a collection of steps required to make sense of the available data. Each step is equally important to ensure that the data is analyzed correctly and provides valuable and actionable information. Below are the five essential steps:



***Fig 1- Data Analysis Process***

**Why you need data analysis:**

Data analysis is essential in business to understand problems facing an organization and explore data in meaningful ways. Data in itself is merely facts and figures. Data analysis organizes, interprets, structures, and presents the data into useful information that provides context for the data. Decision-makers can then use this context to take action to enhance productivity and business gain.

**Data Collection:**

For this research paper, we selected a dataset regarding Data Science Jobs from the year 2017 to 2021 it contains information on jobs around the world. The data has been selected from www.kaggle.com. The dataset contains information regarding the data scientist jobs worldwide (ex: California, Texas, India, etc.).

The dataset contains 2.5K records with 29 columns, which contains information such as Company name, Job level (ex: Entry-level, Senior, Manager, etc.), Job location, Employee year of experience, Employee base Salary, Employee education level, etc.,

**Below is a few data columns description:**

| **Column Name** | **Description** | **Data Type** |
| --- | --- | --- |
| Company | Specify the company name regarding the job opportunity. | Text |
| Level | Describes the employee level for which Job is being considered. | Text |
| TotalYearlyCompensation. | Describes the total yearly compensation for an employee. | Number |
| location | Specify the job location of an employee. | Text |
| baseSalary | Specify the base salary for the job. | Number |
| stockGrantValue | Specify the stock value for an employee for that company. | Number |
| gender | Provides information regarding employee gender. | Text |
| Education | Provides information on employees’ education levels. | Text |

**Data Cleaning:**

Once data is collected from all the necessary sources, your data team will be tasked with cleaning and sorting through it. Data cleaning is essential during the data analysis process, simply because not all data is good data.

This dataset used for this analysis contained a few null values, and a few of the data records were blank/missing values so we removed those records and we filtered the job list for a data scientist as we were focusing on the data scientist jobs. Also, we removed the unwanted columns and separated the date column into days, months, and years to analyze data based on the years. These are the few steps we took to clean our data. These will improve the data quality and help us to focus on the important points. There are many ways we can clean the data, we cleaned the data manually using an excel sheet, and once the data was clean, we used the cleaned dataset for our analysis.

**Analyzing the data:**

In this step, we have sliced and diced our dataset to extract meaningful insights from it. Using the techniques and methods of data analysis, we will be looking for hidden patterns and relationships and finding insights and predictions.

We will be using data analysis methods such as descriptive analysis and predictive analysis for this research paper. Descriptive Analysis is the type of analysis of data that helps describe, show or summarize data points in a constructive way such that patterns might emerge that fulfill every condition of the data. For predictive analysis, we will be finding the pattern and trying to predict the outcome of the data to help the user understand what conditions will be used to identify the pattern and analyze the future outcomes.

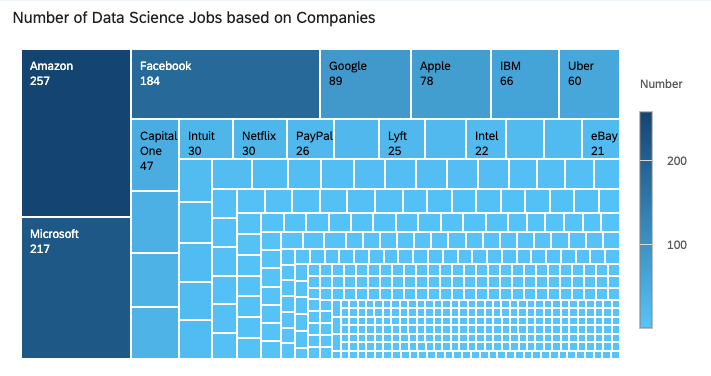
**Interpret the results and apply them:**

The final step is interpreting the results from the data analysis. This part is essential because it’s how we will gain actual value from the previous four steps.

Data visualization helps tell stories by curating data into a form that is easier to understand, highlighting trends and outliers. A good visualization tells a story, removing the noise from data and highlighting useful information. Using visuals to represent the data will help us understand the patterns better, and we can focus on the important points. We will be using a bar graph, geo graphs, and a few others to represent our data for this dataset.

**Research Questions**

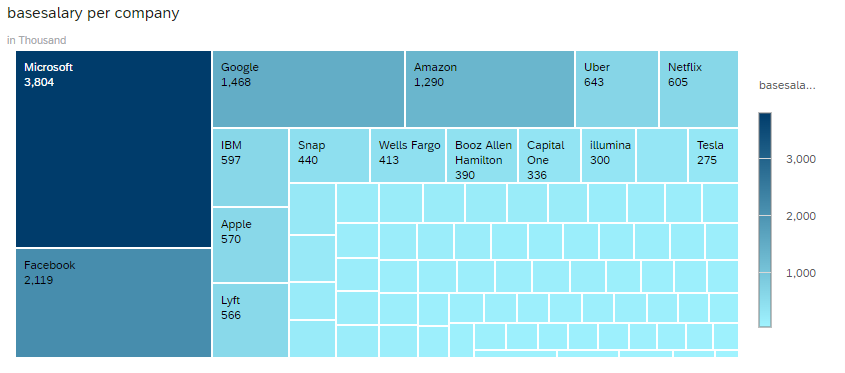
1. **What are the data scientist job opportunities based on the companies?**



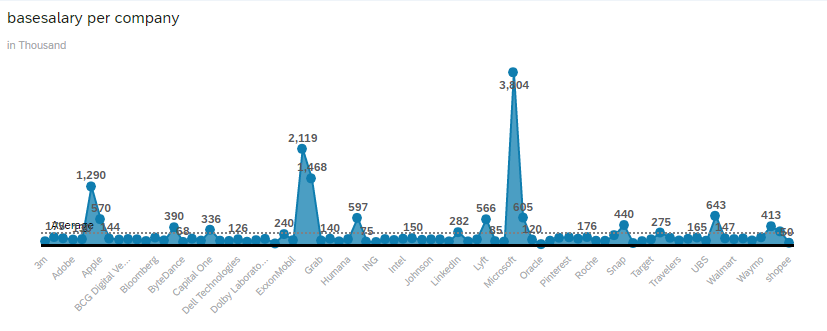
***Figure 2 - TreeMap of Number of Data Science Jobs based on Companies***

The above graph (Fig 2) represents the number of jobs in the data science field based on the companies. The darker shade represents the highest number of jobs, and the lighter shade represents the lowest number of job openings. As we know, data science jobs have increased each year vividly, and from the above analysis, we can observe that many product-based companies have a high demand for data science positions. Google has the highest opening for data science jobs, followed by Microsoft, Facebook, and so on. Also, from the analysis, we can see that a lot of companies have jobs based and data analysis. The table representation of the above graph data is shown in table 1.

1. **What is the base salary offered for a Data scientist job for freshers (with zero years of experience) in different companies?**



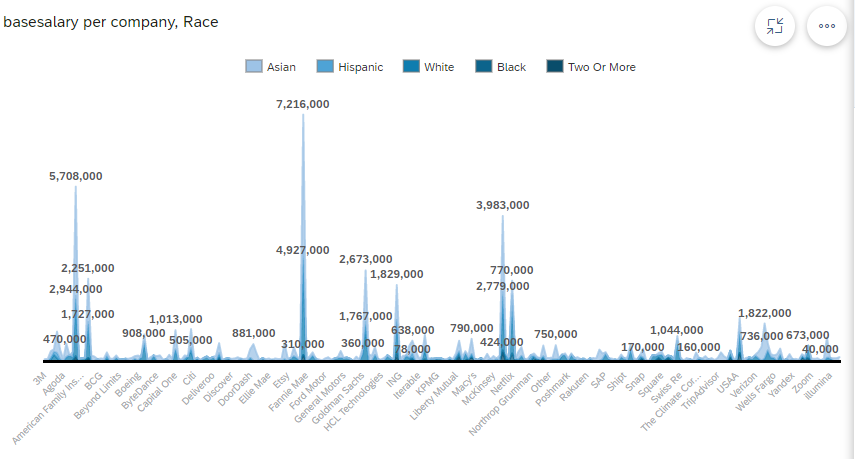
***Figure 3 - TreeMap of Base salary offered by different companies for employees with zero years of experience***



***Figure 4 - Line graph of Base salary offered by different companies***

The above two figures show the base salary offered by different companies for freshers with zero years of experience. The data for the above two graphs is shown in table 2. The darker shade represents the highest salary offered and the lightest shade represents the lowest salary offered. From the above graph, it is shown that Microsoft offers the highest base salary, showing 3538 thousand deviations from the average. Facebook offers the second highest salary which is 2,119 thousand higher than the average. The top ten companies that offer high salaries are Microsoft, Facebook, Google, Amazon, Uber, Netflix, IBM, Apple, and Lyft. Companies that offer lower salaries are ING, Shopee, Societe Generale, Dream 11 and Oracle offer the lowest salary.

1. **What is the base salary offered for different races offered by a different company?**

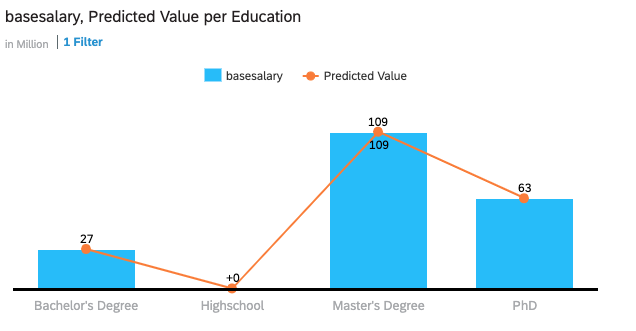


***Figure 5 - Line Graph of Base salary offered by different companies for different Race***

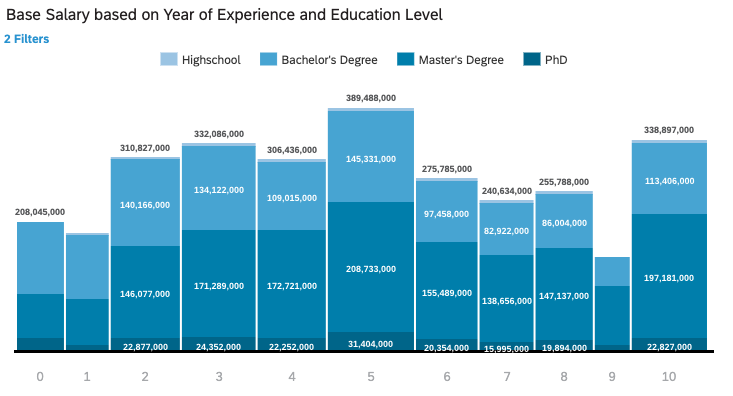
The above graph shows the base salary offered by different companies for different races. The data for the above two graphs is shown in table 3. Each shade from the above graph represents different races. Darker shade represents two or more mixed races and lighter shade represents Asian race. From the above graph, it is shown that the highest salary is offered to the Asian race. Facebook+Asian has the highest salary showing 6,741 thousand deviations from the average. Amazon+ Asian have the second-highest salary showing a 5,708 thousand deviations from the average. From this analysis, we see that Asians and White races are offered the highest job opportunity compared to other races.

**Data Analysis**

1. **Predict if the salary will be increased based on the employee’s education.**



***Figure 6 - Base Salary based on Education Level***

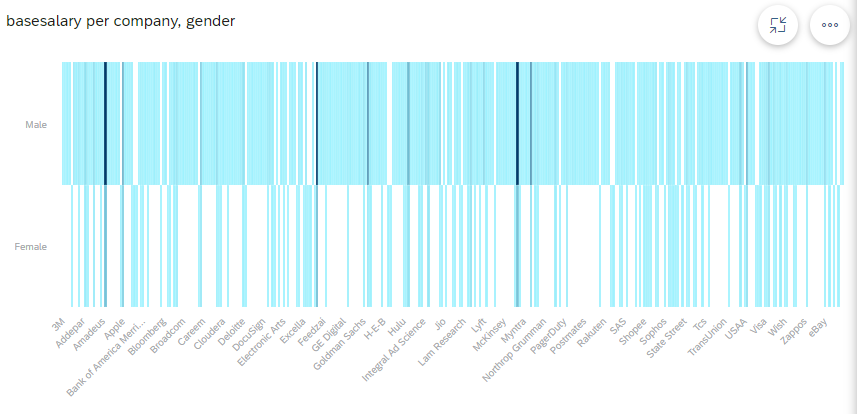
******

***Figure 7 - Base Salary based on Education Level and Year of Experience***

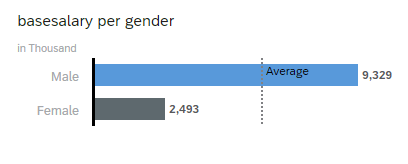
The above figure represents the base salary of an employee based on their year of experience and education level. The X-axis represents the year of experience, and the y-axis represents the base salary based on education level. The legend is based on employee education level (e.x: light blue represents High school, and darkest blue represents the Ph.D.).

From the above graph, we can analyze that employees with five years of experience have the highest base salary followed by employees with ten years of work experience and that employees with Master’s Degree have the highest base salary with 200K. And also, an employee with a master’s degree has the highest (51% of total) base salary compared to other education levels, followed by Bachelor’s degree, Ph.D., and high school. So from the analysis, we can predict that an employee with a master’s degree is highly likely to get a better base salary compared to others. The table representation of the above graph data is shown in table 4.

1. **What is the gender difference in salaries offered by different companies?**

****

***Figure 8 - Heat map representing the salary based on gender***

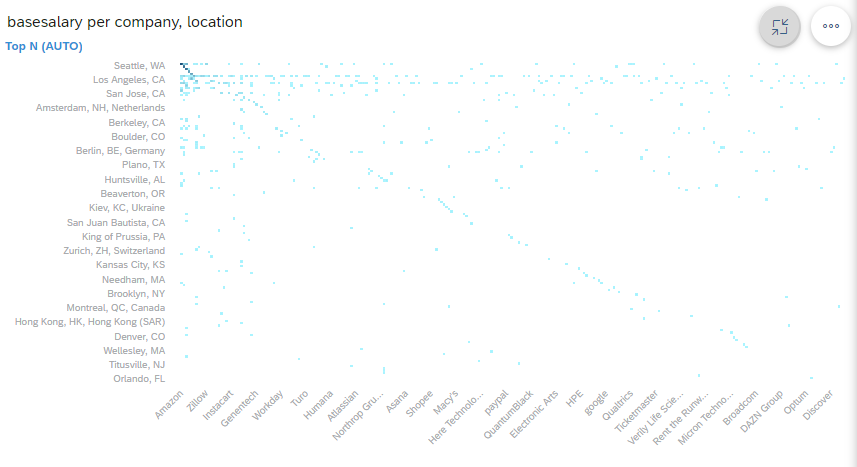


***Figure 9 - Bar graph representing the base salary per gender***

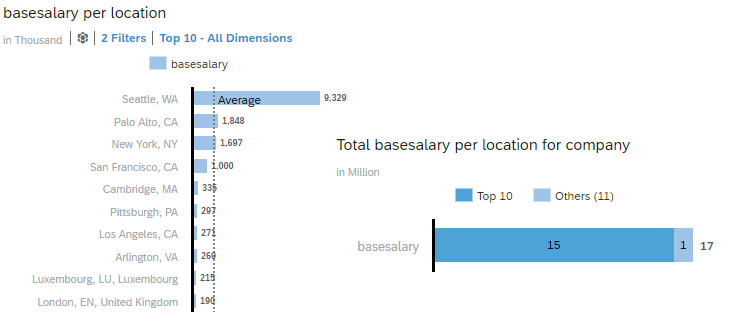
Figure 8 is the graphical representation of the salary offered to different genders in different companies. The X-axis represents the base salary based on the company and the y-axis represents gender. Figure 9 is the bar graph of the base salary offered to different genders. The data for the above two graphs are shown in table 5.

As per the analysis, it is shown that the male gender has the highest salary compared to the female gender. Male is the top gender contributor, which is 58% above the average. From the analysis, it is predicted that the highest salaries offered by top companies are to the male gender.

1. **What is the salary offered by different companies based on the location?**

****

***Figure 10 - Heat map representing the salary based on Location***

****

***Figure 11 - Graphs representing the salary based on Top locations***

Figure 10 is the Heat map of the salary offered by different locations by different companies. The X-axis represents the base salary based on the company and the y-axis represents the Location. The darker shade represents the highest salary and the lighter shade represents the lower salary. Figure 8 is the graphical representation of the base salary offered to the top 10 locations. The data for the above two graphs are shown in table 6.

As per the analysis, it is shown that Seattle is the top contributor compared to other locations. The top 10 highest paying locations are Seattle, Palo Alto, New York, San Francisco, Cambridge, Pittsburgh, Los Angeles, Arlington, Luxembourg, London. The salary offered in Seattle is 9329 higher than the average salary. From the analysis, it is predicted that different locations have different salaries and location is one of the criteria to consider when applying for jobs.

**Conclusion**

From our research, we can see that jobs based on data science have increased over time. The U.S. Bureau of Labor Statistics sees strong growth for data science jobs skills in its prediction that the data science field will grow about 28% through 2026. Also, as technology improves, companies have been able to increase the sophistication of their data operations and analysis.

For our research analysis, we have used a dataset related to Data Science jobs salary and STEM. We reviewed the literature on the same topic to get a clear understanding and learn more about our topic and later after cleaning the dataset and using the methodology step we used SAP Cloud Analytics to create the visual representation of our data so we can have clear insight and focus on the import factors. The data science job has openings in product-based companies and the base salary varies based on the employee's year of experience and education level, location. To conclude Data-driven decision-making is the simple solution to be a successful company in the 21st century you have to use data to your advantage.

**References**

Data Science Salon (2020): *What Indeed’s Job Market Data can tell us about Demographics and Trends in Data Science*; https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma991072192971802901&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=10

Gigaom Blog (2014): *Where the data science jobs are, by sector and by state*; https://www.proquest.com/docview/1619241908?pq-origsite=primo

*Where the data science jobs are, by sector and by state (2014)*; https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi\_proquest\_miscellaneous\_1619241908&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=30

Hira, Ron (2009); *U.S. workers in a global job market*; https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi\_gale\_infotraccpiq\_197666876&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=30

*Bridging the demand and the offer in data science*; URL: https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplaydocid=cdi\_crossref\_primary\_10\_1002\_cpe\_5200&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=0; Date published: December 05, 2021

Tech for Hire: *Data Science-Related Jobs Signal Economic Growth*

; URL: https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi\_ieee\_primary\_9096765&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=0/; Date published**:** December 05, 2021.

*Establishing a Classification Model using Data Science-related Jobs*; URL: https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi\_acm\_primary\_3402291&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=0; Date published**:** June 25, 2021.

*ProQuest | Better research, better learning, better insights*: https://www.proquest.com/docview/1951120146?accountid=10352&pq-origsite=primo; Date published: December 5, 2021

*Over 364,000 data science jobs to be created in India by 2020: IIT-M*: https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi\_proquest\_reports\_2307171873&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=0; Date published: December 5, 2021

Li, Guoyan (2021); *Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis*; https://csu-la.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi\_crossref\_primary\_10\_1016\_j\_jmsy\_2021\_07\_007&context=PC&vid=01CALS\_ULA:01CALS\_ULA&lang=en&search\_scope=MyInst\_and\_CI&adaptor=Primo%20Central&tab=MyInst\_and\_CI&query=any,contains,data%20science%20jobs&offset=30

Situ, Wangming; Zheng, Lei; Yu, Xiaozhou; Daneshmand, Mahmoud (2017); *Predicting the Probability and Salary to Get Data Science Job in Top Companies*; https://www.proquest.com/docview/1951120146?fromopenview=true&pq-origsite=gscholar

Zinovy Radovilsky (2018); *Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis*; https://www.csupom.com/uploads/1/1/4/8/114895679/v16n1p5.pdf

Andrea De Mauro(2016); *Beyond Data Scientists: a Review of Big Data Skills and Job Families*; https://www.researchgate.net/publication/305109030\_Beyond\_Data\_Scientists\_a\_Review\_of\_Big\_Data\_Skills\_and\_Job\_Families

Ryan Noonan(2017); *STEM Jobs: 2017 Update*; https://files.eric.ed.gov/fulltext/ED594354.pdf

Monteiro, Carolina Lourenço (2021); *The data scientist: what companies expect from recent graduates and the role of business schools*; https://run.unl.pt/handle/10362/123030

Carlos Orellana Fantoni (2020); *Tech for Hire: Data Science-Related Jobs Signal Economic Growth*; https://ieeexplore.ieee.org/abstract/document/9096765

Andrea De Mauro (2021); *A Systematic Review of Data Analytics Job Requirements and Online-Courses*; https://www.tandfonline.com/doi/abs/10.1080/08874417.2021.1971579

John King and Roger Magoulas (2015); *2015 DATA SCIENCE SALARY SURVEY*; http://www.eli.sdsu.edu/courses/fall16/cs696/2015-data-science-salary-survey.pdf

Hsia-Ching Chang, Chen-Ya Wang, Suliman Hawamdeh (2018); *Emerging trends in data analytics and knowledge management job market*; https://www.emerald.com/insight/content/doi/10.1108/JKM-02-2018-0088/full/html

Filipe Baumeister(2020); *What Is Required to Be a Data Scientist?: Analyzing Job Descriptions With Centering Resonance Analysis;* https://www.igi-global.com/article/what-is-required-to-be-a-data-scientist/259946

**Tables**

**Tables 1: Sample Data Representation of Figure 1 Number of Data Science Jobs based on Company**

| **Measures** | **Number** |
| --- | --- |
| 3M | 1 |
| 3m | 1 |
| 7-Eleven | 1 |
| ABB | 1 |
| ADP | 3 |
| AMAZON | 2 |
| AMazon | 1 |
| ASML | 1 |
| AT&T | 9 |
| AWS | 1 |
| Abbott | 1 |
| Accenture | 14 |
| Acorns | 2 |
| Addepar | 1 |
| Adobe | 18 |
| Aetna | 10 |
| Affirm | 3 |

**Tables 2: Sample Data Representation of Figure 2 base salary offered for a Data scientist job for freshers (with zero years of experience) in different companies**

| **Company** | **Base Salary** | **Years of Experience** |
| --- | --- | --- |
| Apple | 125000 | 0 |
| Lyft | 130000 | 0 |
| Microsoft | 109000 | 0 |
| Facebook | 115000 | 0 |
| Twitter | 137000 | 0 |
| Microsoft | 142000 | 0 |
| Microsoft | 117000 | 0 |
| Affirm | 130000 | 0 |
| Microsoft | 120000 | 0 |
| Amazon | 105000 | 0 |
| Microsoft | 117000 | 0 |
| Microsoft | 109000 | 0 |
| Uber | 134000 | 0 |
| Uber | 127000 | 0 |
| Facebook | 150000 | 0 |
| Schlumberger | 93000 | 0 |
| Netflix | 350000 | 0 |
| Microsoft | 122000 | 0 |
| Facebook | 140000 | 0 |
| Microsoft | 122000 | 0 |
| Google | 137000 | 0 |
| Capital One | 90000 | 0 |
| Wells Fargo | 150000 | 0 |
| Target | 125000 | 0 |
| Integral Ad Science | 110000 | 0 |
| AT&T | 92000 | 0 |
| Walmart Labs | 145000 | 0 |
| Google | 137000 | 0 |
| Microsoft | 130000 | 0 |
| Facebook | 150000 | 0 |
| Microsoft | 143000 | 0 |
| Microsoft | 141000 | 0 |
| Microsoft | 155000 | 0 |
| Microsoft | 127000 | 0 |
| Google | 137000 | 0 |
| Amazon | 110000 | 0 |
| Microsoft | 118000 | 0 |
| Microsoft | 131000 | 0 |
| Microsoft | 118000 | 0 |
| Facebook | 120000 | 0 |

**Tables 3: Sample Data Representation of Figure 4 the job opportunity for different races offered by a different company**

| **Company** | **Base Salary** | **Race** |
| --- | --- | --- |
| IBM | 71000 | Black |
| Google | 170000 | White |
| Facebook | 150000 | Asian |
| Microsoft | 150000 | White |
| PayPal | 160000 | Asian |
| Amazon | 185000 | White |
| AT&T | 83000 | White |
| Intel | 121000 | Asian |
| Expedia | 115000 | Asian |
| PayPal | 207000 | Asian |
| GoDaddy | 115000 | Asian |
| Apple | 187000 | White |
| Amazon | 140000 | Asian |
| Capital One | 92000 | Asian |
| Apple | 225000 | Asian |
| Capgemini | 40000 | White |
| Swiss Re | 100000 | White |
| Google | 170000 | Asian |
| Grab | 45000 | Asian |
| Asana | 147000 | Asian |
| Dropbox | 120000 | Asian |
| IBM | 130000 | Asian |
| Google | 172000 | White |
| Wayfair | 120000 | White |
| Amazon | 140000 | White |
| Apple | 145000 | White |
| JUUL Labs | 160000 | Asian |
| Next-door | 143000 | White |
| Facebook | 120000 | Asian |
| Yandex | 22000 | White |
| Oracle | 117000 | Black |
| Ernst and Young | 98000 | Asian |
| Morgan Stanley | 90000 | Black |
| GoDaddy | 112000 | Asian |
| Intuit | 128000 | Asian |
| Apple | 130000 | Asian |
| Boeing | 80000 | Asian |
| Amazon | 200000 | Asian |
| Kaiser Permanente | 159000 | White |
| Workday | 127000 | Asian |
| Apple | 170000 | White |
| Walmart Labs | 95000 | Asian |
| Amazon | 33000 | Asian |
| Amazon | 180000 | Asian |
| Apple | 195000 | Asian |
| Netflix | 400000 | White |
| Netflix | 450000 | Asian |
| Google | 172000 | Asian |
| Aetna | 200000 | Asian |
| Airbnb | 145000 | Asian |

**Tables 4: Table Representation of Figure 5 - Base Salary, Predicted Value per Education**

| **Measures** | **Base Salary** | **Predicted Value** | |
| --- | --- | --- | --- |
| Bachelor's Degree | 27471000 | 27681839 |  |
| Highschool | 402000 | 402726 |  |
| Master's Degree | 108724000 | 108750215 |  |
| PhD | 62891000 | 62755636 |  |

**Tables 5: Sample Data Representation of Figure 7 - The gender difference in salaries offered by different companies**

| **Company** | **Base Salary** | **Gender** |
| --- | --- | --- |
| LinkedIn | 162000 | Male |
| Microsoft | 165000 | Male |
| Twitter | 200000 | Male |
| Facebook | 190000 | Male |
| Amazon | 150000 | Male |
| Microsoft | 200000 | Male |
| Google | 240000 | Male |
| Netflix | 600000 | Male |
| Tesla | 118000 | Male |
| Grub Hub | 150000 | Male |
| Cisco | 156000 | Female |
| Sandia National Labs | 103000 | Male |
| Airbnb | 165000 | Male |
| Amazon | 300000 | Male |
| Microsoft | 126000 | Male |
| Microsoft | 109000 | Male |
| Feedzai | 100000 | Male |
| Tesla | 120000 | Male |
| Microsoft | 157000 | Male |
| TheClimate Corporation | 125000 | Male |
| Walmart | 130000 | Male |
| Airbnb | 210000 | Male |
| Amazon | 135000 | Male |
| Walmart Labs | 200000 | Male |
| Electronic Arts | 88000 | Male |
| Electronic Arts | 88000 | Male |
| Apple | 176000 | Male |
| Startup | 110000 | Male |
| Facebook | 115000 | Male |
| Twitter | 137000 | Male |
| Capital One | 121000 | Male |
| LinkedIn | 165000 | Male |
| Capital One | 180000 | Male |
| Amazon | 125000 | Male |
| Microsoft | 142000 | Male |
| Juniper Networks | 103000 | Male |
| Microsoft | 230000 | Male |
| Microsoft | 117000 | Male |
| Airbnb | 165000 | Male |
| Cisco | 150000 | Male |
| Intuit | 148000 | Male |
| HERE Technologies | 97000 | Female |
| Affirm | 130000 | Male |
| Intel | 120000 | Male |
| eBay | 140000 | Male |
| Applied Materials | 130000 | Male |
| Amazon | 145000 | Male |
| Capital One | 124000 | Female |
| eBay | 300000 | Male |

**Tables 6: Sample Data Representation of Figure 9 - The salary offered by different companies based on the location**

| **Company** | **Base Salary** | **Location** |
| --- | --- | --- |
| LinkedIn | 162000 | San Francisco, CA |
| Microsoft | 165000 | Seattle, WA |
| Twitter | 200000 | San Francisco, CA |
| Facebook | 190000 | Seattle, WA |
| Amazon | 150000 | Seattle, WA |
| Microsoft | 200000 | Bellevue, WA |
| Google | 240000 | Kirkland, WA |
| Netflix | 600000 | Los Gatos, CA |
| Tesla | 118000 | Palo Alto, CA |
| Grub Hub | 150000 | New York, NY |
| Facebook | 155000 | Menlo Park, CA |
| Qualcomm | 154000 | San Diego, CA |
| Capital One | 107000 | Mc Lean, VA |
| Microsoft | 137000 | Redmond, WA |
| Workday | 115000 | San Francisco, CA |
| Microsoft | 203000 | Redmond, WA |
| Uber | 132000 | San Francisco, CA |
| Intuit | 144000 | Mountain View, CA |
| Airbnb | 146000 | San Francisco, CA |
| Microsoft | 205000 | Redmond, WA |
| eBay | 189000 | Seattle, WA |
| Facebook | 150000 | Menlo Park, CA |
| Apple | 125000 | Cupertino, CA |
| Google | 221000 | Seattle, WA |
| Microsoft | 154000 | Redmond, WA |
| Amazon | 160000 | Palo Alto, CA |
| Facebook | 171000 | Seattle, WA |
| Amazon | 180000 | Sunnyvale, CA |
| Oath | 130000 | New York, NY |
| PayPal | 150000 | San Jose, CA |
| Google | 132000 | Mountain View, CA |
| Google | 132000 | Mountain View, CA |
| Apple | 125000 | Cupertino, CA |
| Microsoft | 160000 | Redmond, WA |
| Salesforce | 225000 | San Francisco, CA |
| Uber | 140000 | San Francisco, CA |
| Microsoft | 166000 | Redmond, WA |
| Apple | 176000 | Cupertino, CA |
| Amazon | 118000 | Seattle, WA |
| Indeed | 85000 | Austin, TX |
| Indeed | 97000 | New York, NY |
| Indeed | 104000 | New York, NY |
| Lyft | 130000 | San Francisco, CA |
| Microsoft | 157000 | Boston, MA |
| Intuit | 200000 | Mountain View, CA |
| Amazon | 127000 | Seattle, WA |