ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 13 (2023) No. xx, pp. xxx-xxx
DOI:10.33168/JSMS.2023.0101

Data Science Job Salaries: Exploring Compensation Trends and Factors using Tableau

Ray Anthony Pranoto [1], Christopher Abie Diaz Doviano [2], and Christian Ken [3]

[1, 2, 3] Major of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia,

Abstract: Development of big data has rapidly increased gradually, creating an urgent need for data science in Indonesia. The increase in demand is not in line with the availability of adequate data science education and environments, putting pressure on the development of data science jobs in Indonesia. It is important to understand the underlying factors about data science job information such as salary trends, job distribution, and the impact of experience and geographical location on salary, affecting employment opportunities in the data science field. This reveals the gaps in understanding the distribution of salaries globally, which is important in planning a career and managing expectations in this industry. Using secondary data analysis and visualization techniques using Tableau, this research illustrates salary trends by job type, currency and geographic location. Results show the highest paying jobs are Data Analyst Lead (USD 405,000), Principal Data Engineer (USD 328,333), and Financial Data Analyst (USD 275,000), with India accounting for 28.34% of salaries in INR. The analysis shows senior positions have the highest average salary (USD 38,612,842) and entry-level is lower (USD 5,424,612). Salaries show significant growth, especially for senior positions. Full-time jobs dominate recruitment, with the United States, particularly New York, having the highest salary distribution. Remote working trends highlight Data Engineer as the job with the highest opportunities. This research provides a comprehensive overview of the data science job market, illustrating significant differences in salaries by job type, experience, and geography, as well as the growing demand for data science professionals and remote work opportunities.

Keywords: big data, data science, salary distribution, visualization, Tableau

1. Introduction

In this era of technology, "big data" is no longer an unfamiliar term. Big data is often used to refer to a variety of concepts, from the collection and aggregation of large data to the application of various digital techniques to identify the patterns of human behavior (Favaretto et al., 2020). There are two literatures that are relevant to this matter. The first one is "Research Challenges of Big Data," classifies big data based on 3V (Volume, Velocity, Variety) and 5V (Volume, Velocity, Variety, Veracity, and Value) characteristics (Younas, 2019). Meanwhile, the second journal is "Data Architecture (Second Edition): A Primer for the Data Scientist," defines big data as a large amount of data that is economically stored, managed by the "Roman census" method, and stored in an unstructured format (Inmon et al., 2019). Big data technology may collect massive volumes of data from numerous sources in order to gain information by displaying trends or disclosing knowledge about past, present, and future occurrences at a rapid pace. Data science approach is used to filter, select, and prepare a large amount of data for processing and analysis. (Nainggolan, 2019). Therefore, big data can be managed by a data scientist using data science methodology.

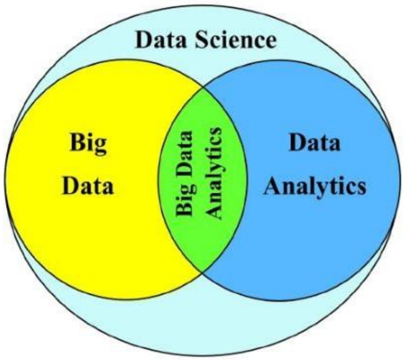


Fig. 1: Relationship between Data Science and Big Data (Lavasani et al., 2021)

As a data science expert, it is important to have the skills and knowledge needed to engage in this field (Meyer, 2019). This is reflected in a Venn diagram that places big data in the scope of data science, with data analytics skills, as well as substantial expertise (Wickham et al., 2023). According to Hadley Wickham (Chang & Grady, 2019), data science is an important science for analyzing data. Additionally, as noted by the National Institute of Standards and Technology's Big Data Working Group, data scientists have a role in extracting knowledge from data to generate meaningful action (Manshur, 2021). Data science is a discipline that requires a comprehensive grasp of probability, distribution, hypothesis testing, and multivariate analysis, as well as knowledge of data structures, algorithms, and database systems (Hadoop) and the ability to articulate issues to find successful answers. (Mount & Zumel, 2019).

Data Science is increasingly needed from time to time (Himanen et al., 2019). along with various industriesthat are starting to use big data (Wali et al., 2023). Many companies are using big data today because it canbring processed information and deep insights to achieve significant business benefits (Ananta, 2023). In fact, Harvard University explains that Data Science is included in "The Best Looking Job of the 21st Century" because data science is becoming the basic needs of every industry, be it civil or technical (Northumbria, 2024). Based on data from DQLab as shown in Figure 2, it states that interest in data scientists is increasing every year and data scientists are the number 1 desired profession in Indonesia (DQLab, 2022). Obviously, this demand has caused pressure on higher education institutions at the national level recently to adjust their curriculum to meet market demands because the use of data science in Indonesia is still low (Priambudi et al., 2021). Despite these educational efforts, there is not yet an adequate data environment to support the development of data science jobs in Indonesia (Hindrayani, 2022).

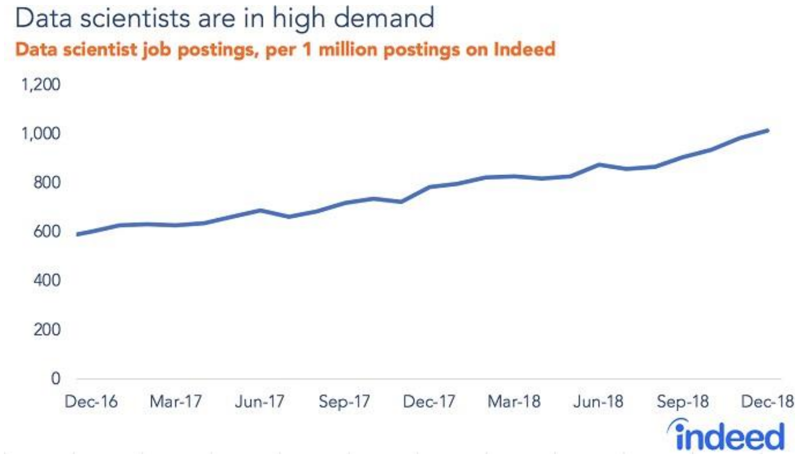


Fig. 2: Graph of increasing demand for Data Science [11]

Therefore, this study was designed to provide a comprehensive overview to data science graduates who arelooking for information regarding job opportunities in various countries (Bhatia, 2022). The expected overview includes the distribution of salaries, the relationship between salaries and work experience, and the trend of data science job growth each year. In addition, the visualization results in this study also lead to encouragement for the development of the data environment in Indonesia, so that it can provide wider support for data science practitioners. This research uses an Exploratory data Analysis (EDA) approach to understand datasets collected from various data sources (Arora et al., 2020). Exploratory Data Analysis (EDA) is an approach that aims to identify the main characteristics by visualizing with the right representation this also includes modeling, hypothesis testing, handling missing values and variable transformation (Sahoo et al., 2019). This approach is used because it can provide descriptive visualization and cleansing stages that have been recorded by data (Lester et al., 2020). This research is expected to make a significant contribution in formulating educational policies and developing human resources in the field of data science in Indonesia.

2. Related Work

2.1. Big Data

Big Data refers to an ever-increasing accumulation of data in numerous types, including structured, unstructured, and semi-structured. Big Data's complexity necessitates the use of advanced technologies and algorithms. Hence, typical static Business Intelligence tools are no longer efficient in Big Data applications (Oussous et al., 2019). McKinsey describes Big Data as data collections that transcend the capabilities of typical database software for capture, storage, administration, and analysis (Manyika et al., 2021). Gartner introduced the concept of Big Data using the '3V' framework: Big Data encompasses information assets with large volumes, high speed in collectionand processing, and high variety and diversity from structured to unstructured. It requires innovative and cost-effective information processing approaches to improve insights and retrieval(Pastorino et al., 2019). Effective use of rapid and large-scale. Big data can influence firms' decision-making approaches (Shamim et al., 2019).

2.2. Big Data Processing

Big Data processing involves the use of technologies such as large-scale data distribution and storage systems, and parallel processing algorithms to handle high data volumes and velocities (Kurniawan et al., 2024). The challenges of Big Data processing include volume, velocity, variety, and veracity, commonly referred to as "4Vs" (Vaidya & Kshirsagar, 2020). These challenges drive the development of specialized techniques and methodologies to effectively extract valuable insights. Moreover, techniques such as data streaming and parallel processing facilitate real-time or near real-time analysis of streaming data and parallel execution of tasks across distributed nodes. Machine learning algorithms and artificial intelligence techniques are increasingly being applied to Big Data for tasks such as predictive analysis and anomaly detection, further enriching data processing capabilities (Nassif et al., 2021). These techniques are applied in various fields, including business, healthcare, finance, smart cities, and the Internet of Things (IoT), driving data-driven decision making, innovation, and operational efficiency (Faridoon & Imran, 2021).

2.3. Data Scientist

Data Science is a combination of several skills from various fields such as statistics, math, and programming to extract information from data. A data scientist uses methodologies such as data collection, data analysis, and statistical modeling to understand patterns and trends in data, and provide recommendations based on these findings (Syamsu & Widodo, 2021). The data transformation that occurs in industry 4.0 will certainly not be possible without collaboration between humans and machines, here the role of data science and data scientists is needed to organize and manage data and information that continues to grow, so the existence of Data Science will become a link in various fields of job skills in the hope of processing and aiming to translate data both written and spoken (Ho et al., 2019). Data scientists are essential for companies to effectively utilize big data, bring structure and insight to complex problems, and provide advice to executives regarding implications for products, processes, and decisions (Davenport & Patil, 2021). Data scientists perform the data analysis process using various methods and algorithms to combine several pairs of data to reveal a pattern. The discovery of patterns can be used as a reference to predict the business movement of a product.

2.4. Data Analysis

Data analysis includes the process of analyzing, cleaning, interpreting, and drawing conclusions from large sets of data to understand patterns, trends, and relationships within them. The goal of data analysis is to gain insights that can be used for better decision-making. The use of tools in this analysis is largely determined by the right acumen and accuracy in inferring (Alem, 2020). Data analysis is an important part of research that makes the study results more effective, which is a supporting factor for researchers to reach a conclusion (Bhatia, 2019). Today, no business can survive without analyzing the available data (Islam, 2020). Data analysis is an important part of both scientific research and business, where the demand for data-driven decision-making has increased in recent years.

2.5. Data Visualization

Data visualization is the graphical representation of data and information. Its main purpose is to present information visually so that it can be easily understood and used to make inferences. Data visualization can be in the form of graphs, charts, maps, and other various types of diagrams. The origins of the phrase "data visualization" can be traced back to the second century AD. Drawings and other visual aids were employed in ancient societies to both record historical events and conduct globe exploration. Throughout human history, data visualization has significantly aided in discoveries and creations (Crapo, Waisel, Wallace, & Willemain, 2000). The way data is visually represented has changed significantly since the development of computer technology. Data analysts now use computer graphical data visualization to analyze data more quickly and accurately. Research in a wide range of disciplines, including computer vision, human perception, animation, and algorithms, now heavily relies on data visualization (Unwin, 2020). Data visualization is the process of displaying data through graphical displays. A scatterplot shows each data point individually, whereas a histogram shows aggregate statistics (Zhou, 2023). The most effective method of communicating with people is increasingly data visualization. This paper will explore the importance of datavisualization and the current use of data visualization (Balaji et al., 2021).

2.6. Literature Review

A research made by Oetama et al. (2020) used clustering analysis with the K-means algorithm andTableau to reveal differences in drug crime rates in DKI Jakarta, finding North Jakarta, Central Jakarta, and Thousand Islands as areas with the highest crime rates. This study also analyzed the characteristics of the perpetrators and the distribution patterns of drug crimes. Tambe et al. (2020) in their research revealed that companies using new technology can attract more productive IT workers with the same salary, affecting retention, salary strategies, labor mobility, and diversity of IT companies. This study shows a positive relationship between the use of new information technology and the wages desired by IT workers.

In 2020, Balaji et al. examined how Tableau and Splunk can improve the ease of visualizing data,especially for inexperienced interns. Using visualization techniques such as heatmaps, bar charts, line graphs, and tree maps, this research shows how information can be presented more clearly andintuitively. The results of this study emphasize that visualization tools such as Tableau can significantly simplify the data analysis process. Furthermore, a study conducted by Tee and Raheem (2022) showed that polynomial regression models are more accurate in predicting salaries in the data science field than linear regression, especially in the case of non-linear relationships between salary and job position level. This study also highlighted the complexity in salary prediction with the use of multiple input variables that can decrease accuracy.

Siregar et al. examined the introduction of data science and the data scientist profession at SMA Pramita Tangerang, showing a visualization of the high demand for data scientists that is not balanced with the availability of human resources in Indonesia, especially in the context of big data mining (Bakti et al., 2022). Research by Ariandi and Puteri (2022) also shows that the use of Tableau Public for data visualization of Kertapati sub-district is very helpful in presenting population information comprehensively, using various types of graphs such as histograms, line charts, and barplots. The results show that this approach accelerates the decision-making process at the Kertapati sub-district office which previously took a lot of time.

Afikah et al. (2022) in their research implemented Business Intelligence using the Tableau platformto analyze data on corona virus cases in Indonesia. They managed to produce a dashboard that includes the number of confirmed cases, deaths, and recoveries in various provinces, which supports decision making related to handling the Covid-19 pandemic in Indonesia. Meanwhile, research by Natasuwarna (2021) highlights the important role of data science as a sought-after job in the era of the Industrial Revolution 4.0. They point out that a deep understanding of consumer needs is increasingly important, and data science is an effective tool to respond to this challenge, integrating concepts such as data mining and big data with the era of Industrial Revolution 4.0 and society 5.0.

3. Research Methodology

A secondary data source was used to collect data for this study, data obtained from the external platform Kaggle, a platform that was the primary source for the secondary data used in this study. The dataset, titled "Data Science Job Salaries," is taken from the Kaggle website [18] and focuses on job salary data in the data science field. This dataset contains important information about salaries and other attributes associated with jobs in data science.Kaggle, which serves as a repository for datasets and a hub for data-related competitions, makes it easy to obtain a variety of datasets, including from overseas sources. This dataset was taken from Ruchi Gupta, and provides useful insights into the salaries of jobs in data science. It was published on Kaggle on June 15, 2022 and includes important variables such as job title, location, education, experience, and salary. Then this research also uses a qualitative approach. Aiming to understand more deeply the complexity of factors affecting salaries in the data science industry, the method helped the research explore in more detail the social, cultural, and organizational contexts in which salary decisions are made.

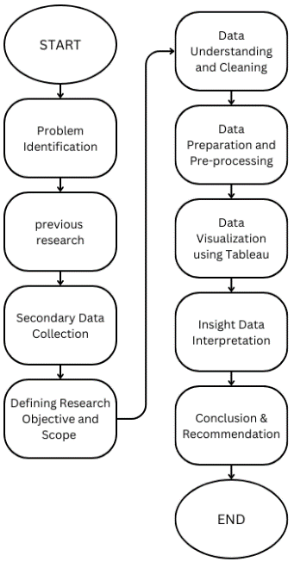

START
Problem Identification
previous research
Secondary Data Collection
Defining Research Objective and Scope
Data Understanding and Cleaning
Data Preparation and Pre-processing
Data Visualization using Tableau
Insight Data Interpretation
Conclusion & Recommendation
END

Fig. 4: Research Flow

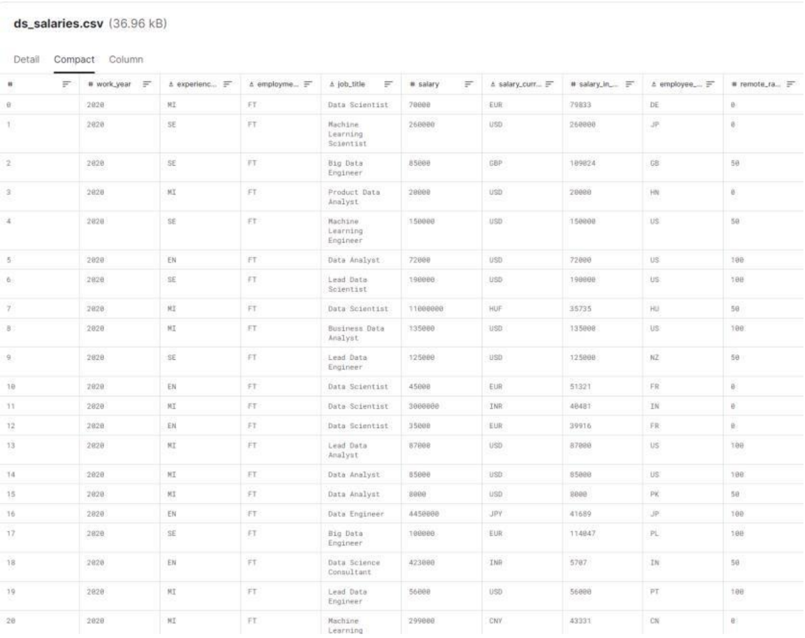
- **Problem Identification**

  Problem identification is carried out to find out what problems we will examine so that research can focus on solving problems. Based on the research topic, the problem is that the increasing number of requests regarding data science is causing pressure on higher education institutions at the national level to adjust their curriculum to meet market demands because the use of Data Science in Indonesia is still low. Although efforts in the education sector have reached this stage, there is not yet an adequate data environment to support the development of data science jobs in Indonesia.

- **Previous Research**

  Furthermore, previous research is an important stage in this research, which provides references regarding the research to be carried out. The research provides information and suggestions to increase the chances of researchers making more meaningful research. Previous research was takenfrom scopus journals, kc.umn.ac.id website (UMN journals), google scholar, and ResearchGate with details of 3 international journals, 3 national journals, and 2 UMN journals.

- **Secondary Data Collection**



ds_salaries.csv (36.96 kB)

#	work_year	experienc...	employme...	job_title	salary	salary_curr...	salary_in_...	employee_...	remote_ra...
0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50
3	2020	MI	FT	Product Data Analyst	20000	USD	20000	HN	0
4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	US	50
5	2020	EN	FT	Data Analyst	72000	USD	72000	US	100
6	2020	SE	FT	Lead Data Scientist	190000	USD	190000	US	100
7	2020	MI	FT	Data Scientist	11000000	HUF	35735	HU	50
8	2020	MI	FT	Business Data Analyst	135000	USD	135000	US	100
9	2020	SE	FT	Lead Data Engineer	125000	USD	125000	NZ	50
10	2020	EN	FT	Data Scientist	45000	EUR	51321	FR	0
11	2020	MI	FT	Data Scientist	3000000	INR	40481	IN	0
12	2020	EN	FT	Data Scientist	35000	EUR	39916	FR	0
13	2020	MI	FT	Lead Data Analyst	87000	USD	87000	US	100
14	2020	MI	FT	Data Analyst	85000	USD	85000	US	100
15	2020	MI	FT	Data Analyst	8000	USD	8000	PK	50
16	2020	EN	FT	Data Engineer	4450000	JPY	41689	JP	100
17	2020	SE	FT	Big Data Engineer	100000	EUR	114047	PL	100
18	2020	EN	FT	Data Science Consultant	423000	INR	5707	IN	50
19	2020	MI	FT	Lead Data Engineer	56000	USD	56000	PT	100
20	2020	MI	FT	Machine Learning	299000	CNY	43331	CN	0

Fig. 5: Dataset ‘Data Science Job Salaries’

Based on the research method using secondary data it can help regarding the researcher's view of obtaining a pre-existing data source, Secondary data based on this research uses data taken from Kaggle entitled 'Data Science Job Salaries' which contains the following columns: year salary paid, worker experience level, type of job (part-time, full-time, contract, freelance), job role or title, gross salary amount, salary currency, salary in USD, employee's country of primary residence, percentage of work performed remotely, country where the company's main office is located, and company size by number of employees.

- **Defining Research Objective and Scope**

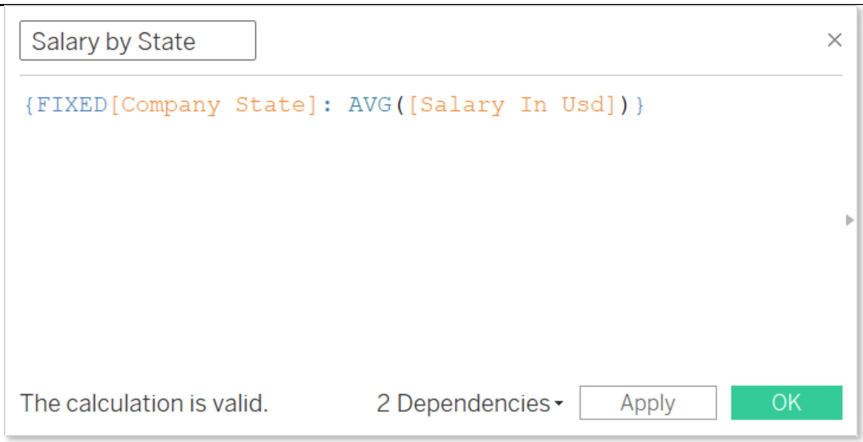
  It is useful to help researchers to know what they want to achieve with their study and helps to keepthe research focused and not too broad. Based on the research topic, the objective of this study aimsto investigate several important aspects related to data science jobs. Firstly, it will investigate the salary distribution of data science jobs by company location, with the aim of understanding how geographical location affects the compensation offered. Secondly, an analysis will be conducted to examine the relationship between salary and level of work experience, in order to evaluate the extent to which work experience contributes to the amount of salary received by data science professionals. Thirdly, the research also aims to identify the development of employment trends in the data science field over the years, with a particular focus on the accumulated salaries offered by different companies around the world, so as to provide insights into the dynamics of the job market in this sector. Finally, the research will determine the percentage of data science jobs that offer remote work options and identify the best types of jobs that meet remote work criteria based on location to salary. The focus is to gain a deep understanding of how experience level, job type, location and other variables contribute to salary in the context of the data science industry. The scope of the study will include a comprehensive analysis of compensation in data science jobs, taking into account various variables such as experience level, job type, and location. This study will use a dataset obtained from Kaggle, focusing on salary data in Seattle to investigate salary patterns in the data science industry.

- **Data Understanding and Cleaning**

  Before the data can be processed further, the first step required is to understand the data well and perform initial processing on the data. This part aims to understand the meaning of the data and determine which columns can be processed further. Before proceeding to the next stage, it is necessary to clean the data first. This is necessary because there are still some data that have null values and outliers that are far from the range of data in the "Data Science Job Salaries" dataset. By doing this cleaning process, we can ensure the cleanliness and quality of the data before enteringthe further analysis stage. Based on the analysis results, there are no null values in each column inthe "Data Science Job Salaries" dataset, ensuring that the data is ready for further processing without any lack of values.

- **Data Preparation and Preprocessing**

  After the data has been processed, a further process is needed to ensure that the resulting data visualization does not cause ambiguity or confusion. This step is very important so that the designed data visualization can be more efficient and informative in conveying the value of the data. Based on the data, there are two columns that need adjustment. First, the date column needs to be converted into a date data type so that it can be used appropriately in the visualization. Second, the country column needs to be changed to region or country (region/country) so that the data can be visualized in the form of a map. By making these adjustments, the data will be better equipped to be visualized clearly and accurately, and facilitate deeper analysis of the values contained therein.



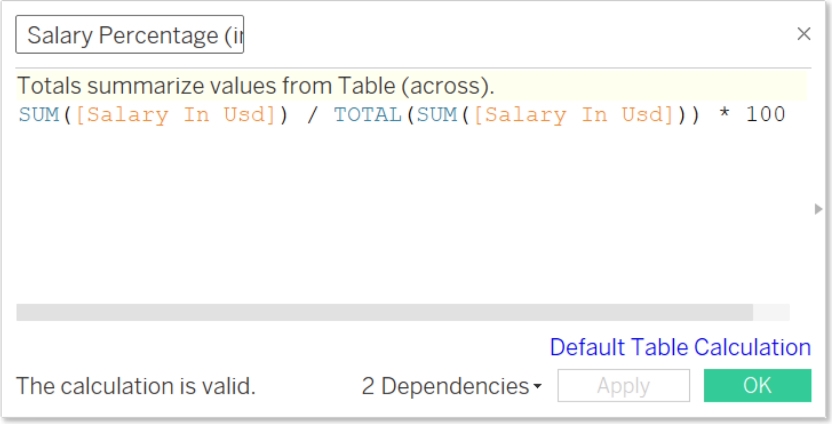
Salary by State

```
{FIXED[Company State]: AVG([Salary In Usd])}
```

The calculation is valid.

Fig. 6: LOD Calculation for Salaries based on States

Then in addition there are several calculated fields in the preparation process. Figure 6 above represents a calculation that uses the Level of Detail (LOD) feature. LOD allows users to calculate values in different levels of data, regardless of the general level of aggregation specified by the view. In this expression, [Company State] is the dimension used to determine the LOD level. That is, the calculation will be performed for each company state separately. Whereas, AVG([Salary In Usd]) is the calculation that will be run, which is the average of the salaries in USD. Thus, overall, the expression produces the average salary in USD for each state of the company separately, regardless of the other aggregation levels in the view. The result of the code is named "Salary by State".



Salary Percentage (i

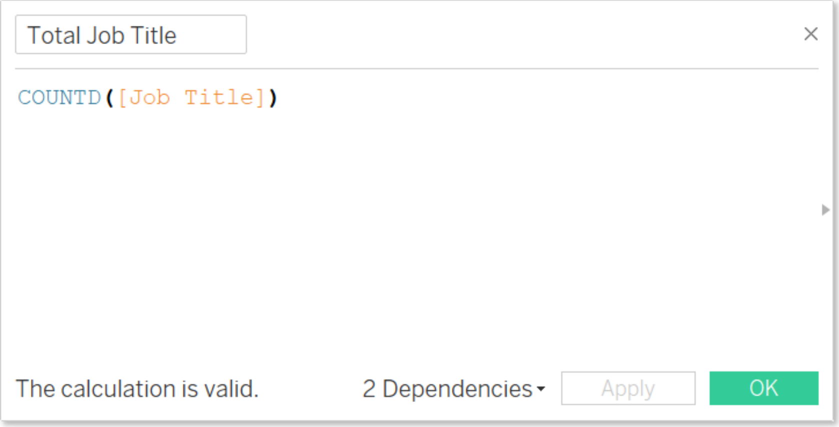
Totals summarize values from Table (across).

```
SUM([Salary In Usd]) / TOTAL(SUM([Salary In Usd])) * 100
```

The calculation is valid.

Fig. 7: Number Calculation for Salary percentage from Total Salary

Also in Figure 7 above is the number calculation 'Salary Percentage' to calculate the percentage ofsalary. In this calculation, SUM([Salary In Usd]) is used to calculate the total salary in USD currency for each row or entry in the data. Meanwhile, TOTAL(SUM([Salary In Usd])) calculatesthe overall total salary in USD without regard to dimensions or splits in the dataset. Through the formula (SUM([Salary In Usd]) / TOTAL(SUM([Salary In Usd]))) * 100, the salary of each job type is divided by the total salary, and then the result is multiplied by 100 to convert it into a percentage. As such, this number calculation provides information on how much each entity's salary contributes to the overall total salary, allowing for a deeper analysis of the salary distribution in the dataset.



Total Job Title

```
COUNTD([Job Title])
```

The calculation is valid.

Fig. 8: Calculated Field for the column ‘Job Title’

Figure 8 is the code used to count the number of 'Job Titles' in each category of work experience. The function sums all the 'Job Title' values, where the COUNTD (Count Distinct) function is usually used in data analysis to count the number of unique values in a data set. This function is useful when you want to know the number of distinct or unique entities in a variable.

- **Data Visualization Using Tableau**

The data visualization methodology using Tableau began with a detailed analysis of the "Data Science Job Salaries" dataset. First of all, the average salary by job type was visualized using a bar chart, where Data Analyst Lead, Principal Data Engineer, and Financial Data Analyst emerged as the highest paid positions. A color scale from gold to green is used to depict the salary range from lowest to highest, providing an immediate visual understanding of the salary comparison between jobs. Furthermore, a pie chart is used to show the distribution of salaries by currency, providing insight into the company's country of origin and the currency used. Level of Detail (LOD) is also applied to calculate the average salary by state, and salary distribution is displayed using map visualization to provide a clearer picture of salary geography. In the creation of the dashboard, visualizations were incorporated to provide an overview of the salary distribution, number of jobs by experience level, and salary growth per year by experience. In addition, the number of jobs by job type and job type, and the percentage ratio of remote work were analyzed. Polynomial and clustering models were used to explore trends and visualize the distribution of jobs with the highest remote rates. The final dashboard created combines all the previous visualizations to provide a comprehensive overview of the salary distribution, trend line, and distribution of jobs with the highest remote rate, which overall provides a deep understanding of the salary characteristics, job distribution, and remote work trends in the data science industry.

- **Insight Data Interpretation**

This step involves interpreting the results of the data analysis and visualizations that have been made in depth. This process involves drawing careful conclusions, identifying patterns or trends that emerge from the data, and understanding the implications of the findings. From the results of this analysis, readers will be able to see a clear picture of the dynamics in the data analyzed, whether it is related to the distribution of salaries by company location, the relationship between salary and level of work experience, the development of job trends in the data science field over time, the accumulated salaries offered by various companies around the world, or the percentage of jobs that offer remote work options. An in-depth interpretation of the results of this analysis will allow readers to gain a better insight into the characteristics and dynamics of the job market in the data science industry, as well as its implications in a broader context.

- **Conclusion & Recommendation**

  This final stage involves drawing conclusions from the research and providing recommendations for future action based on the findings. This process combines the research findings into one comprehensive conclusion, and evaluates the implications and consequences of the findings in a broader context. The summary of the research findings highlights the salary distribution of data science jobs by company location, the relationship between salary and level of work experience, the development trend of data science jobs over the years, the accumulated salaries offered by different companies in the world, as well as the percentage of jobs that offer remote work options and the best types of jobs for remote options. The implications of the findings are clearly outlined,providing a deeper understanding of how certain factors such as location, experience, and market trends affect compensation and career options in the data science industry. Recommendations for further action are based on the research conclusions, offering concrete suggestions for individuals,organizations, and policies that can be implemented to make effective use of the research findings.As such, this stage provides a solid closure to the research, emphasizing the importance of the findings and providing direction for next steps in supporting developments and advancements in the data science field.

4. Result and Discussion

4.1. Average Distribution of Data Science Salary by Company Location

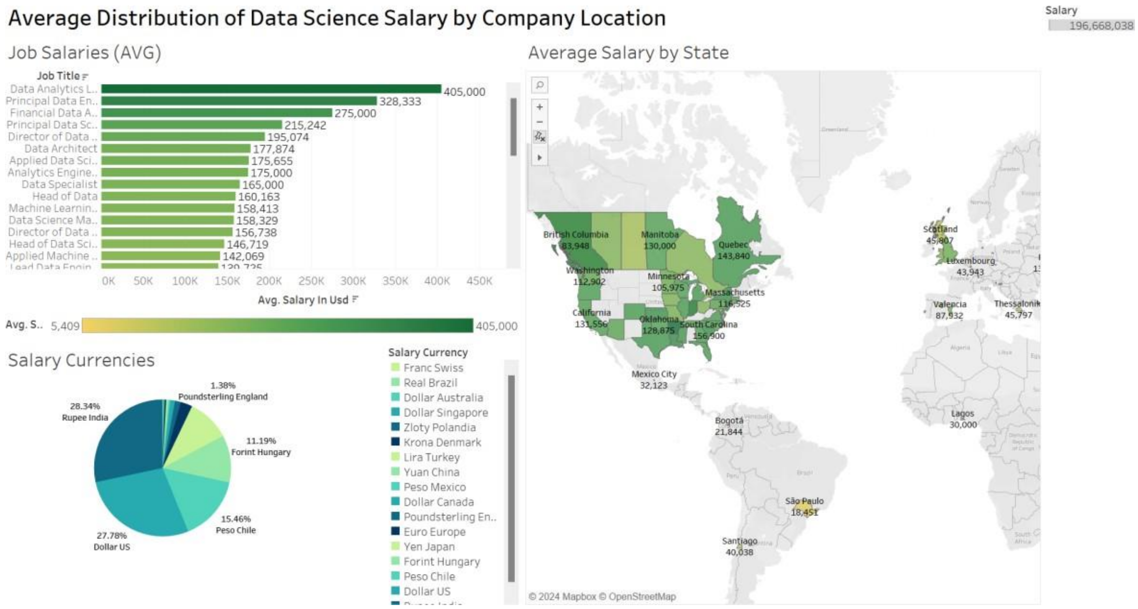


Fig. 9: Average Distribution of Data Science Salary by Company Location Dashboard

As shown in Figure 9, the dashboard above is an overview of several visualizations about the distribution of average data science salaries by company location. There are 3 visualizations on this dashboard, consisting of the first visualization, which illustrates the salaries of various jobs, with the three highest salary jobs (in USD) occupied by Data Analyst Lead (405,000 USD), Principal Data Engineer (328,333 USD), then Financial Data Analyst (275,000 USD). The second visualization uses a pie chart type of visualization, which analyzes salaries by currency in various countries, with the conclusion that the most common currency is INR or Indian Rupee, which means that India contributes greatly to the provision of jobs in the field of data science. The third visualization displays a map visualization using the LOD of income by state.

4.2. Relationship between Data Science Salary and Work Experience

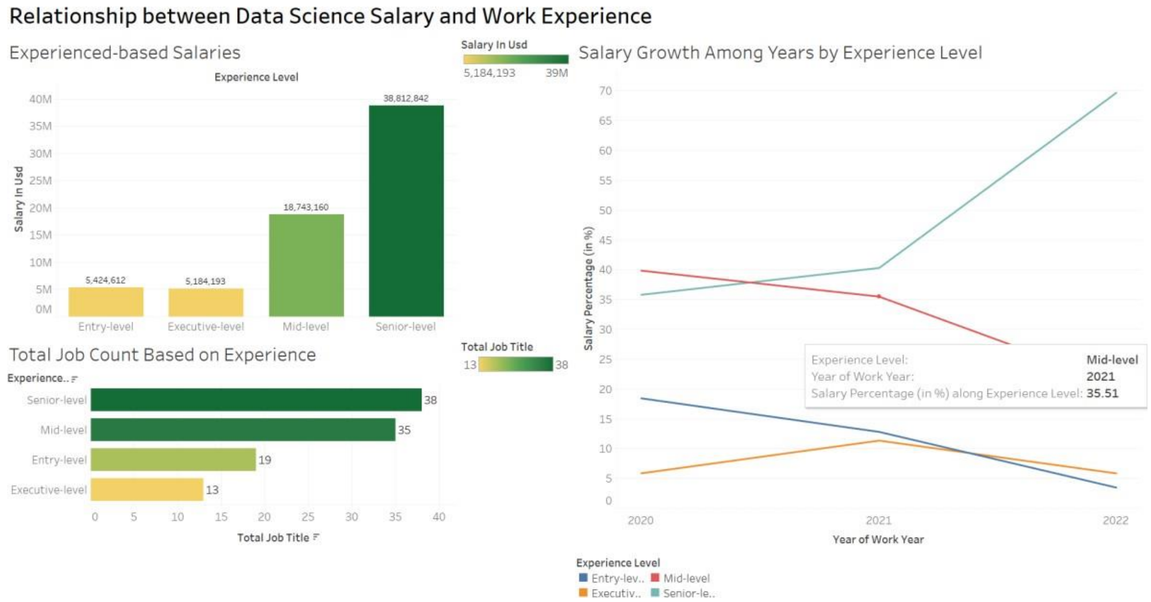
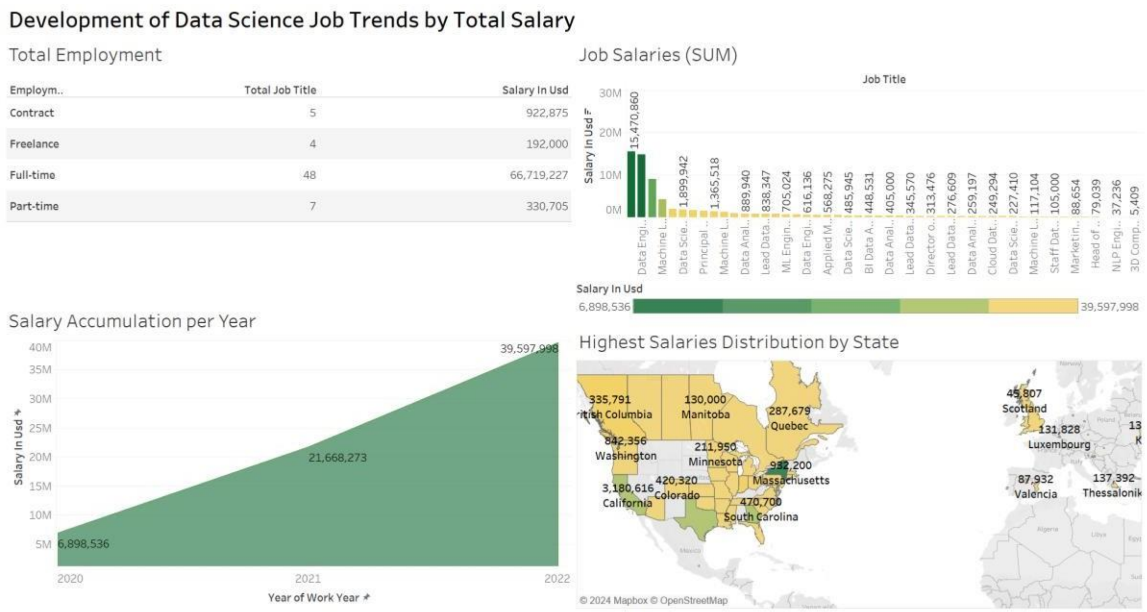


Fig. 10: Salary Growth Dashboard by Experience Level

Figure 10 is a dashboard consisting of various graphs explaining the growth of data science salariesby job experience level. The first visualization is a comparison of salaries by experience level, withthe Senior experience level showing a drastic spike in salaries. Then, the second visualization explains the salary growth over the year showing a significant increase for the senior experience level by 2022, while the entry level shows minimal growth. The last visualization is a bar graph visualization of the total number of jobs by experience level, with the senior category having the highest number and the lowest category being executives. These three visualizations combined in one dashboard can be used as a reference of how experience levels affect job demand each year.

4.3. Development of Data Science Job Trends by Total Salary



Development of Data Science Job Trends by Total Salary

Total Employment

Employm..	Total Job Title	Salary In Usd
Contract	5	922,875
Freelance	4	192,000
Full-time	48	66,719,227
Part-time	7	330,705

Job Salaries (SUM)

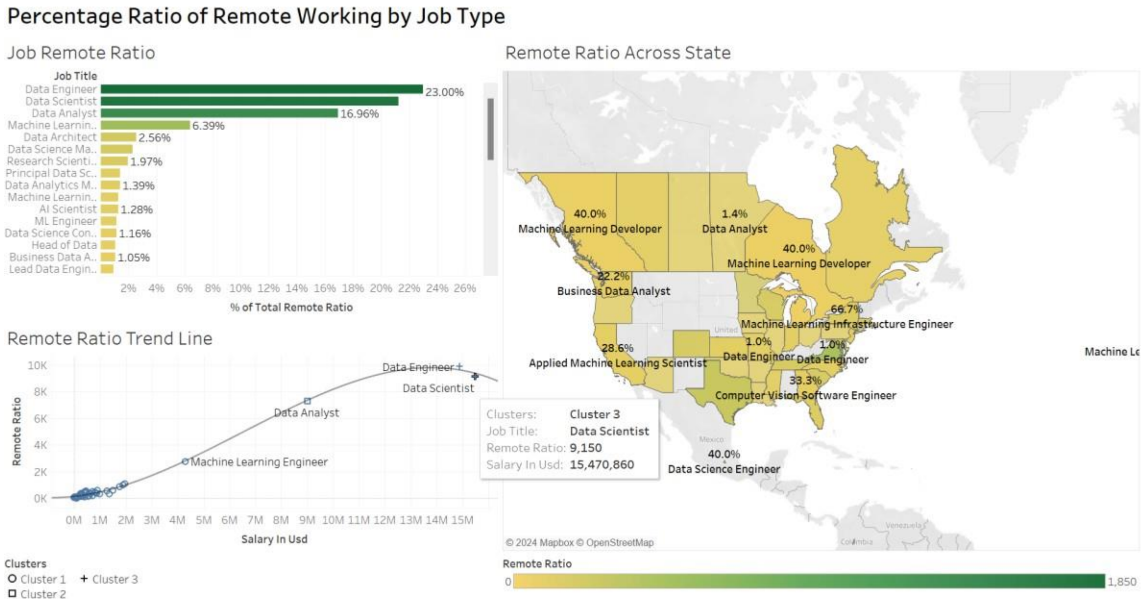
Salary Accumulation per Year

Highest Salaries Distribution by State

Fig. 11: Data Science Job Trend Development Dashboard by Salary Amount

The dashboard in Figure 11 is a collection of four visualizations that provide an overview of the development of data science job trends. The first visualization in this dashboard is to illustrate thedistribution of the total number of jobs by job type marked with the codes CT (Contract), FL (Freelance), FT (Full Time), and PT (Part Time), with the accumulated salary of data science jobs. It can be seen that data science jobs with Full-Time job types are more in demand with a total salary of 66,719,227 USD over 3 years. The second visualization in this dashboard is to illustrate the amount of income in each job sorted by the largest income to the smallest. The third visualization in this dashboard illustrates the increasing trend of data science salaries per year. In the visualization we can see that there is a significant increase from 2021 to 2022, and also a slight increase from 2020 to 2021. The fourth visualization in the dashboard is a visualization of the distribution of salaries across countries. This visualization uses a form of map visualization to more clearly illustrate the location of these countries. With this visualization in the dashboard, data science graduates can know that New York is the state with the highest annual salary accumulation.

4.4. Percentage Ratio of Remote by Job Type



Clusters: Cluster 3
Job Title: Data Scientist
Remote Ratio: 9,150
Salary In Usd: 15,470,860

Fig. 12: Dashboard Percentage Ratio of Remote Working by Job Type

Figure 12 is a dashboard that displays three visualizations, namely the percentage of remote work,a trend line of job types using clusters, and a map visualization for the distribution of jobs with thehighest remote rate in a state. The first visualization displays the percentage ratio of remote work in various types of jobs, with Data Engineer, Data Scientist, Data Analyst, Machine Learning Engineer, and Data Architect as the five jobs with the highest remote work ratio. The second visualization represents the level of remote work ratio using a polynomial model-based trend line and clustering to group job types by salary, with an R-squared value of 0.99627 and a P-value of less than 0.0001, indicating that the polynomial model is almost perfect in predicting the trend. The final visualization depicts the distribution of specific job types with the highest remote ratio in the 3 states of New York (66.7%), Wales (50%), and Mexico City (40%). The end result of this dashboard is very useful to see what data science jobs have the best remote ratio with adequate salary quality.

5. Conclusions

Based on the results of the research and process conducted previously, it can be concluded that there are several jobs in the field of data science that are very promising, especially the Data Analyst Lead job, whichis the job with the highest average salary (405,000 USD), followed by Principal Data Engineering (328,333USD) and also Financial Data Analyst (275,000 USD). After that, it can be seen that the largest currency distribution on the data science job list comes from India at 28.34%, which shows that India contributes greatly to the provision of jobs in the field of data science. The salary comparison between worker experience is also very significant. Workers in the senior category earn a salary far above the other categories at 38,612,842 million USD. In addition, this study found that the salary development of each category of work experience of senior workers has the most significant salary increase compared to other categories. Furthermore, full-time workers get the most hires, amounting to 48 hires, when compared to part-time, contract, and freelance workers. The salaries earned by these workers also increase constantly from year to year which makes the job in the field of data science a job with excellent prospects. In addition, the total salary of Data Scientist is the largest total salary compared to other jobs in the field of data science. It can also be seen that the distribution of the highest salaries of workers in the field of data science is in the United States, precisely in the city of New York, therefore the US is one of the destination countries for workers in the field of data science. In the end, the ratio of jobs that are mostly done remotely is data engineer jobs, with the state of New York at 66.7%, which is the state with the highest remote ratio.

References

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PloS One*, 15(2), e0228987. https://doi.org/10.1371/journal.pone.0228987

Younas, M. (2019). Research challenges of big data. *Service-oriented Computing and Applications*, 13(2), 105–107. https://doi.org/10.1007/s11761-019-00265-x

Inmon, W., Linstedt, D., & Levins, M. (2019). A brief history of big data. *Elsevier eBooks* (pp. 67–71). https://doi.org/10.1016/b978-0-12-816916-2.00010-3

Nainggolan, D. R. M. (2019). Sains Data, Big Data, Dan Analisis Prediktif: Sebuah Landasan Untuk Kecerdasan Keamanan Siber. *Jurnal Pertahanan Dan Bela Negara*, 7(2). 139-154. https://doi.org/10.33172/jpbh.v7i2.187

Lavasani, M. S., Ardali, N. R., Sotudeh-Gharebagh, R., Zarghami, R., Abonyi, J., & Mostoufi, N. (2021). Big data analytics opportunities for applications in process engineering. *Reviews in Chemical Engineering*, 39(3), 479–511. https://doi.org/10.1515/revce-2020-0054

Meyer, M. A. (2019). Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings. *Journal of the American Medical Informatics Association*, 26(5), 383–391. https://doi.org/10.1093/jamia/ocy181

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. " *O'Reilly Media, Inc*, 2, 30-32.

Chang, W. L., & Grady, N. (2019). NIST Big Data Interoperability Framework: Volume 2, Big Data Taxonomies [Version 2]. *NIST*. https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-2-big-data-taxonomies-version-2

Manshur, A. (2021). Satu Data, Big Data dan Analitika Data: Urgensi Pelembagaan, Pembiasaan dan Pembudayaan. *Bappenas Working Papers*, 4(1), 30–46. https://doi.org/10.47266/bwp.v4i1.82

Mount, J., & Zumel, N. (2019). *Practical data science with R. Manning Publications*.

DQLab. (2022). Intip 4 Fakta Menarik Tentang Karir Data Scientist. *DQLab*, vol. 4, no. 15

Himanen, L., Geurts, A., Foster, A. S., & Rinke, P. (2019). Data-Driven Materials Science: Status, challenges, and Perspectives. *Advanced Science*, 6(21). https://doi.org/10.1002/advs.201900808

Wali, M., Efitra, Sudipa, I. G. I., Heryani, A., ... & Sepriano (2023). Penerapan & Implementasi Big Data di Berbagai Sektor (Pembangunan Berkelanjutan Era Industri 4.0 dan Society 5.0). *PT. Sonpedia Publishing Indonesia.*

Ananta, A. (2023). Pemodelan Data dan Pengelolaan Data: Perspektif Big Data. *ResearchGate*. https://www.researchgate.net/publication/376781605_Pemodelan_Data_dan_Pengelolaan_Data_Perspektif_Big_Data

Northumbria University London. (2024). MSC Big Data and Data Science Technology | *Northumbria London*. https://london.northumbria.ac.uk/course/msc-big-data-and-data-science-technology/

Priambudi, B. N., Ariani, N. M., Wijaya, M. I. H., & Pradana, B., (2021). Eksplorasi Pentingnya Penggunaan Data Science Dalam Perencanaan Pemodelan Transportasi Perkotaan. *Jurnal Sistem dan Teknologi*, 5(3). https://journal.itk.ac.id/index.php/sjt/article/view/375

Hindrayani, K. M. (2022). Analisa Pekerjaan Sains Data di Australia menggunakan Pendekatan Exploratory Data Analysis. https://prosiding-senada.upnjatim.ac.id/index.php/senada/article/view/51

Bhatia, R., (2022). Data Science Job Salaries, *Kaggle*, https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries

Arora, A. S., Rajput, H., & Changotra, R. (2020). Current perspective of COVID-19 spread across South Korea: exploratory data analysis and containment of the pandemic. *Environment, Development and Sustainability*, 23(5), 6553–6563. https://doi.org/10.1007/s10668-020-00883-y

Sahoo, K., Samal, A. K., Pramanik, J., Pani, S. K., (2019). Exploratory data analysis using Python, *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727-4735. https://doi.org/10.35940/ijitee.L3591.1081219

Lester, J. N., Cho, Y., & Lochmiller, C. R. (2020). Learning to do qualitative data analysis: a starting point. Human Resource Development Review, 19(1), 94–106. https://doi.org/10.1177/1534484320903890

Oetama, R. S., Heng, T. T., & Tjahjana, D. (2020). Sebuah Pola Cluster Geospatial Eksplorasi Kejahatan Narkoba di DKI Jakarta. Ultima InfoSys, 11(1), 57–62. https://doi.org/10.31937/si.v9i1.1514

Tambe, P., Ye, X., & Cappelli, P. (2020). Paying to program? engineering brand and High-Tech wages. Management Science, 66(7), 3010–3028. https://doi.org/10.1287/mnsc.2019.3343

Balaji, N., Pai, B. H. K., Bhat, B., & Praveen, B. (2021). Data Visualization in Splunk and Tableau: A case study demonstration. Journal of Physics. Conference Series, 1767(1), 012008. https://doi.org/10.1088/1742-6596/1767/1/012008

Tee, Z., & Raheem, M. (2022). Salary prediction in data science field using specialized skills and job benefits. ResearchGate.

https://www.researchgate.net/publication/362280362_Salary_Prediction_in_Data_Science_Field_Using_Specialized_Skills_and_Job_Benefits_-A_Literature_Review

Siregar, B., Pangruruk, F. A., Siridion, S. T., Immanuel, K. R., … Gani, J. I., (2022). Pengenalan Data Science dan Profesi Data Scientist di SMA Pramita Tangerang, *Jurnal Pengabdian Masyarakat Bestari (JPMB)*, 1(2), 87-96. https://dx.doi.org/10.55927/jpmb.v1i3.620

Ariandi, M., Puteri, S. R., (2022). Analisis Visualisasi Data Kecamatan Kertapati menggunakan Tableau Public, *Jurnal Penelitian Ilmu dan Teknologi Komputer (JUPITER)*, 14(2), https://jurnal.polsri.ac.id/index.php/jupiter/article/view/5141

Afikah, P., Affandi, I. R., & Hasan, F. N. (2022). Implementasi Business Intelligence Untuk Menganalisis Data Kasus Virus Corona di Indonesia Menggunakan Platform Tableau. *Jurnal Pseudocode*, 9(1), 25–32. https://doi.org/10.33369/pseudocode.9.1.25-32

Natasuwarna, A. P., Alfaqy, M. H., & Simaremare, T. J. (2021). Data Science Sebagai Pekerjaan Baru Yang Banyak Dicari Era Revolusi Industri 4.0. *Semnas Corisindo*, 1(1), 132-138, https://ejournal.raharja.ac.id/index.php/corisindo/article/view/2027

Oussous, A., Benjelloun, F., Lahcen, A. A., & Belfkih, S. (2019). Big Data technologies: A survey. *Journal of King Saud University. Computer and Information Sciences*, 30(4), 431–448. https://doi.org/10.1016/j.jksuci.2017.06.001

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2021). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., & Boccia, S. (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *European Journal of Public Health*, 29(3), 23–27. https://doi.org/10.1093/eurpub/ckz168

Shamim, S., Zeng, J., Shariq, S. M., & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management*, 56(6), 103-135. https://doi.org/10.1016/j.im.2018.12.003

Kurniawan, S. D., Widiastuti, R. Y., Candrasari, D. M., …, Judijanto, L., (2024). Big Data: Mengenal Big Data & Implementasinya di Berbagai Bidang. *Google Books*,. 1(1), 16

Vaidya, G. M., & Kshirsagar, M. M., (2020). A Survey of Algorithms, Technologies and Issues in Big Data Analytics and Applications, *2020 4th International Conference on Intelligent Computing and Control Systems (ICICC)*, 347-350. https://doi.org/10.1109/ICICCS48265.2020.91210641

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M., (2021). Machine Learning for Anomaly Detection: A Systematic Review. *in IEEE Access*, 9, 78658-78700. https://doi.org/10.1109/ACCESS.2021.3083060

Faridoon, A., & Imran, M. (2021). Big data storage tools using NoSQL databases and their applications in various domains: A Systematic Review. *Computing and Informatics*, 40(3), 489–521. https://doi.org/10.31577/cai_2021_3_489

Syamsu, M., & Widodo, W. (2021). Peran Data Science dan Data Scientist Untuk Mentransformasi Data Dalam Industri 4.0. *Jutech*, 2(1), 27–36. https://doi.org/10.32546/jutech.v2i1.1540

Ho, A., Nguyen, A., Pafford, J. L., & Slater, R. (2019). A Data Science Approach to Defining a Data Scientist. *SMU Data Science Review*. 2(3) https://scholar.smu.edu/datasciencereview/vol2/iss3/4/

Davenport, T. H., & Patil, D. (2021). Data Scientist: The sexiest job of the 21st century. *Harvard Business Review*. 90(10), 70-76. https://www.researchgate.net/publication/232279315_Data_Scientist_The_Sexiest_Job_of_the_21st_Century

Alem, D. D. (2020). An Overview of Data Analysis and Interpretations in Research, *International Journal of Academic Research in Education and Review*, 8(1), 1-27. https://doi.org/10.14662/IJARER2020.015

Bhatia, M. K., (2019). Data analysis and its importance. *International Research Journal of Advanced Engineering and Science*, 2(1), 166-168. https://irjaes.com/wp-content/uploads/2020/10/IRJAES-V2N1P58Y17.pdf

Islam, M. (2020). Data Analysis: Types, Process, Methods, Techniques and Tools. *International Journal on Data Science and Technology*, 6(1), 10. https://doi.org/10.11648/j.ijdst.20200601.12

Unwin, A., (2020). Why is Data Visualization Important? What is Important in Data Visualization?. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.8ae4d525

Zhou, L. (2023). An introduction to data visualization. *Highlights in Science, Engineering and Technology*, 31, 60–63. https://doi.org/10.54097/hset.v31i.4813