

MAPPING THE FUTURE OF DATA SCIENCE CAREERS: A BIG DATA APPROACH TO JOB TREND ANALYSIS AND RECOMMENDATION SYSTEMS

ADVISOR : DR.EDDY ZHANG

AIT-614-001-BIG DATA ESSENTIALS

APRIL - 2024

PRITHAM MAHAJAN

RAKESH SOMUKALAHALLI VENKATAPPA

NITHISH BILASUNUR MANJUNATHA REDDY

SAI SRIRAM DUBSI



Objective of project

- Understanding Salary Dynamics.
- Predictive Modeling for Salary Forecasting.
- Enhancing Career Decision-Making.



OVERVIEW OF DATASET

The dataset used in this project comes from AI Jobs.net. It includes job postings for data science positions. It is noteworthy that this dataset is updated monthly throughout the duration of 2024, ensuring the data's relevance and effectiveness to our analysis.

Additionally we cleaned the dataset by deleting unwanted columns, removing null values and by adding more useful data which was recently.

Overview of

Dataset

- work_year : The year that the data was collected.
- job_title : The job role's title .
- job_category : For simple analysis, the job role is divided into 10 general types.
- salary_in_usa : The gross annual pay in US dollars (USD).
- experience_level : Divides professional experience into four categories: entry-level, mid-level, senior, and executive.
- employment_type : Shows the kind of work: "Full-time," "Part-time," and "Contract."
- work_setting : The type of work environment, such as "Hybrid," "Remote," or "In-person."
- company_location: The region where the business is situated. places the employer company in the small, medium, and big categories.
- company_size : Divides the employer company into three categories: small, medium, and giant.

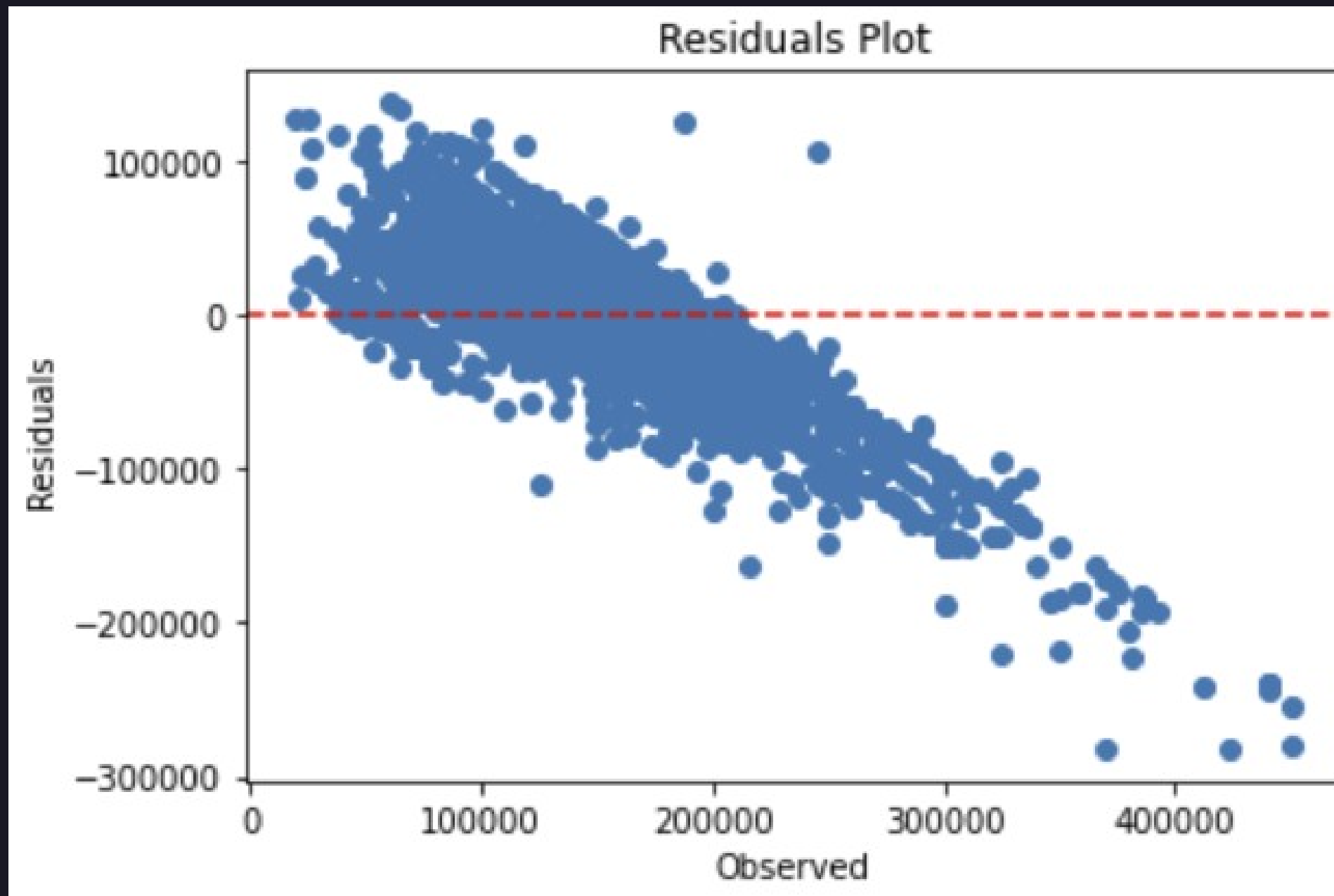
Snapshot of the Dataset

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	work_setting	company_location	company_size	job_category
2024	Entry-level	Freelance	Applied Data Scientist	30000	USD	30000	United Kingdom	Remote	United Kingdom	M	Data Science and Research
2024	Executive	Full-time	Business Intelligence	230000	USD	230000	United States	In-person	United States	M	BI and Visualization
2024	Executive	Full-time	Business Intelligence	176900	USD	176900	United States	In-person	United States	M	BI and Visualization
2024	Senior	Full-time	Data Architect	171210	USD	171210	Canada	In-person	Canada	M	Data Architecture and Modeling
2024	Senior	Full-time	Data Architect	92190	USD	92190	Canada	In-person	Canada	M	Data Architecture and Modeling
2024	Mid-level	Full-time	Data Science	46203	GBP	57753	United Kingdom	In-person	United Kingdom	M	Data Science and Research
2024	Mid-level	Full-time	Data Science	38280	GBP	47850	United Kingdom	In-person	United Kingdom	M	Data Science and Research
2024	Entry-level	Full-time	Insight Analyst	50000	USD	50000	United States	Remote	United States	M	Data Analysis
2024	Entry-level	Full-time	Insight Analyst	40000	USD	40000	United States	Remote	United States	M	Data Analysis
2024	Senior	Full-time	Data Engineer	276000	USD	276000	United States	In-person	United States	M	Data Engineering
2024	Senior	Full-time	Data Engineer	148000	USD	148000	United States	In-person	United States	M	Data Engineering
2024	Senior	Full-time	Research Scientist	234000	USD	234000	United States	In-person	United States	M	Data Science and Research
2024	Senior	Full-time	Research Scientist	146000	USD	146000	United States	In-person	United States	M	Data Science and Research
2024	Entry-level	Full-time	Business Intelligence Analyst	192300	USD	192300	United States	In-person	United States	M	BI and Visualization
2024	Entry-level	Full-time	Business Intelligence Analyst	120200	USD	120200	United States	In-person	United States	M	BI and Visualization
2024	Entry-level	Full-time	Analytics Engineer	132500	USD	132500	United States	In-person	United States	M	Leadership and Management
2024	Entry-level	Full-time	Analytics Engineer	111500	USD	111500	United States	In-person	United States	M	Leadership and Management
2024	Entry-level	Full-time	Data Engineer	234000	USD	234000	United States	In-person	United States	M	Data Engineering
2024	Entry-level	Full-time	Data Engineer	146000	USD	146000	United States	In-person	United States	M	Data Engineering
2024	Senior	Full-time	Data Scientist	322000	USD	322000	United States	In-person	United States	M	Data Science and Research
2024	Senior	Full-time	Data Scientist	168000	USD	168000	United States	In-person	United States	M	Data Science and Research
2024	Mid-level	Full-time	Data Scientist	152000	USD	152000	Canada	In-person	Canada	M	Data Science and Research
2024	Mid-level	Full-time	Data Scientist	98000	USD	98000	Canada	In-person	Canada	M	Data Science and Research
2024	Senior	Full-time	Research Engineer	138000	USD	138000	United States	In-person	United States	M	Data Science and Research
2024	Senior	Full-time	Research Engineer	86000	USD	86000	United States	In-person	United States	M	Data Science and Research
2024	Senior	Full-time	Data Science	268700	USD	268700	United States	In-person	United States	M	Data Science and Research
2024	Senior	Full-time	Data Science	130300	USD	130300	United States	In-person	United States	M	Data Science and Research
2024	Senior	Full-time	BI Developer	60000	GBP	75000	United Kingdom	In-person	United Kingdom	M	BI and Visualization
2024	Senior	Full-time	BI Developer	40000	GBP	50000	United Kingdom	In-person	United Kingdom	M	BI and Visualization
2024	Senior	Full-time	Data Analyst	115000	USD	115000	United States	In-person	United States	M	Data Analysis
2024	Senior	Full-time	Data Analvst	105000	USD	105000	United States	In-person	United States	M	Data Analysis

Salary Prediction

11

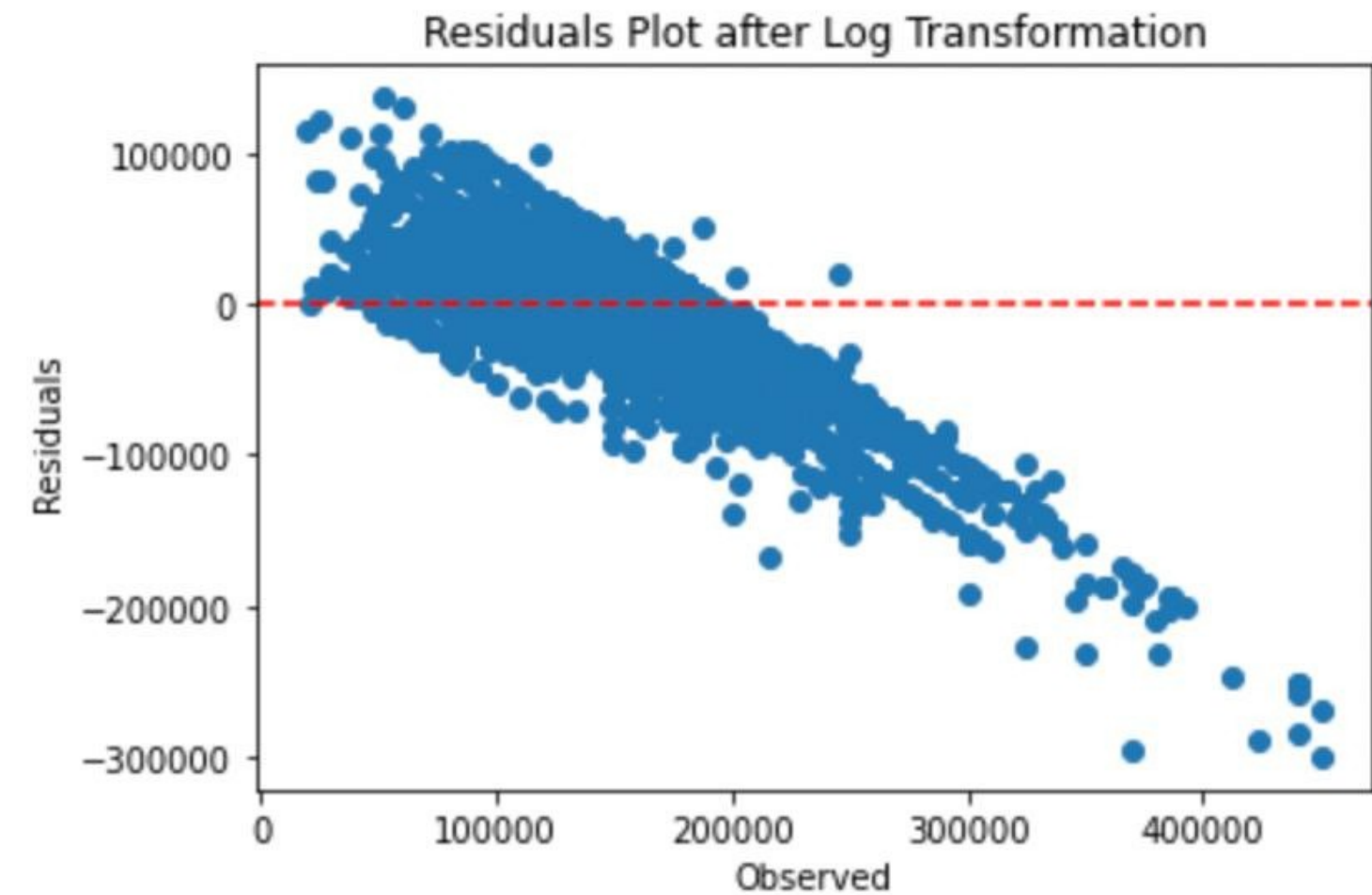
GRADIENT BOOSTING REGRESSOR MODEL



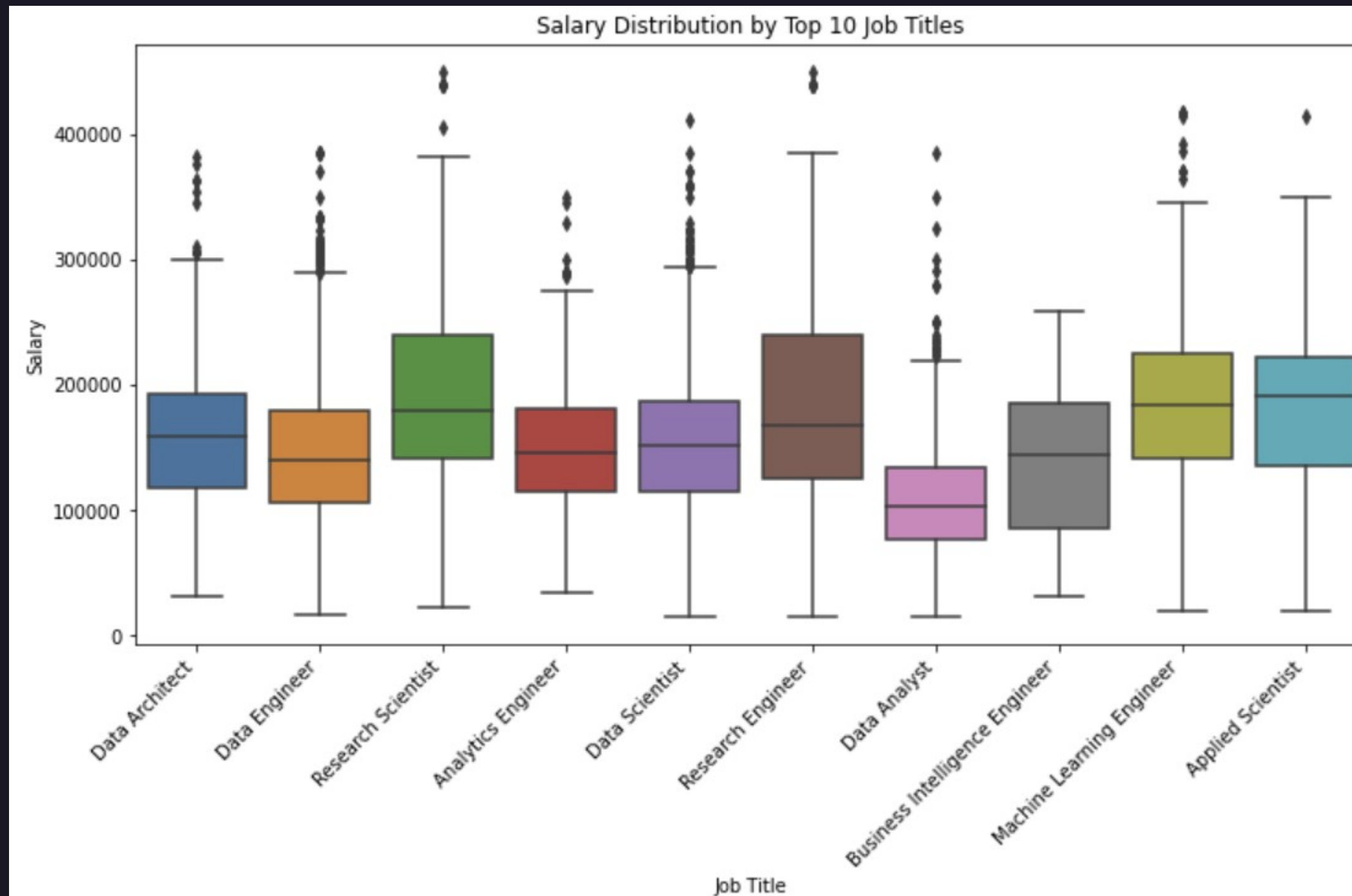
- Predict exact salary figures for professionals in the data industry.
- Root Mean Squared Error (RMSE): \$53,491.
- Residuals plot indicated underprediction of higher salaries and overprediction of lower ones.
- Potential biases or a misfit to the underlying salary distribution.

LOG TRANSFORMATION MODEL

- Normalize the distribution of salaries and reduce the impact of outliers.
- Applied a logarithmic transformation to the target salary variable.
- \$53,997, showing a marginal improvement.
- Residuals plot still shows underprediction at the higher end of the salary spectrum.
- Continued refinement necessary to address persistent prediction challenges at higher salary levels.



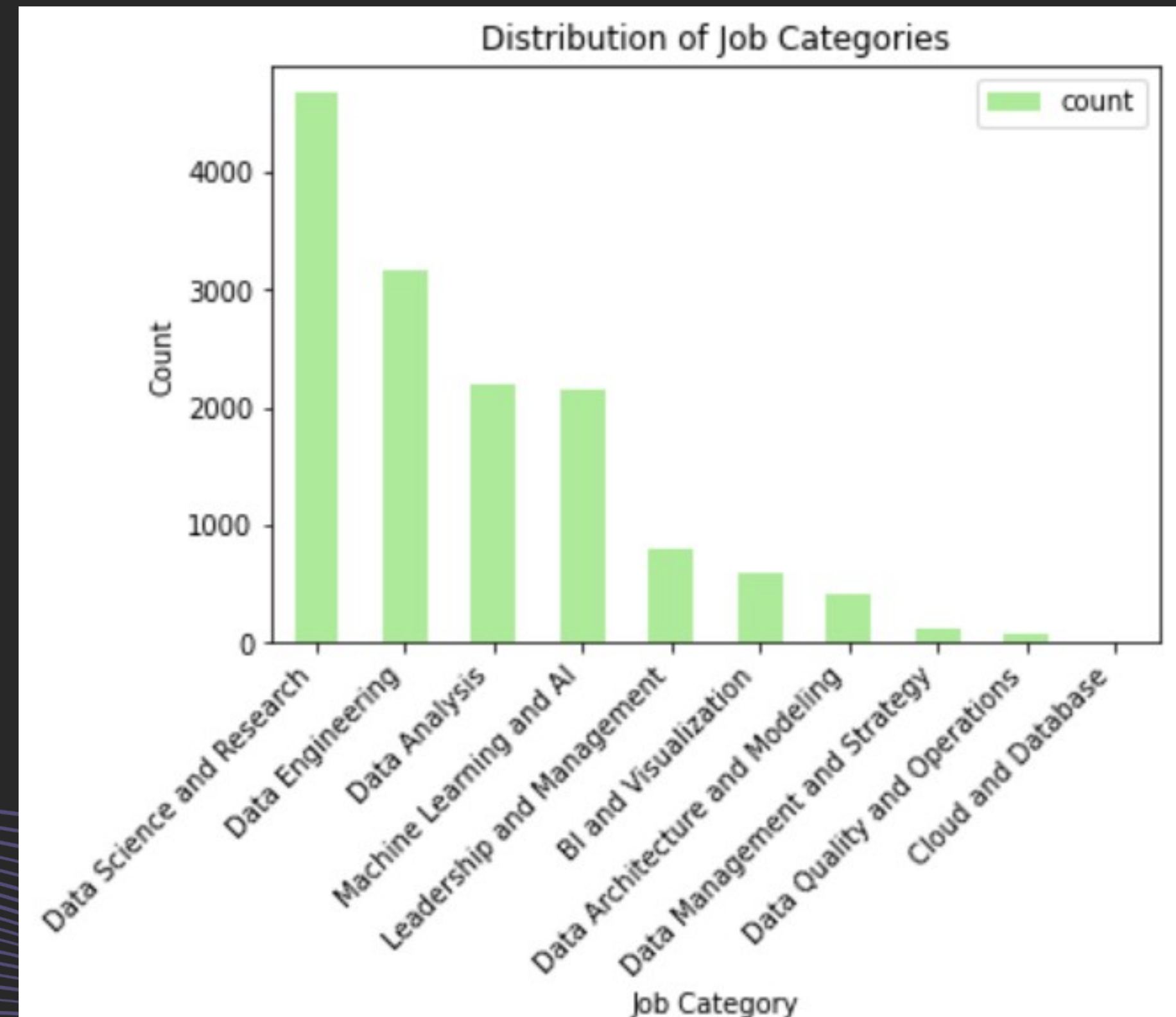
BOX PLOT OF SALARY DISTRIBUTION BY TOP 10 JOB TITLES



- Salary distribution across the top 10 most common job titles extracted from LinkedIn job postings where the salary of the Research Engineer is higher.
- Utilizes a box plot to visually depict the salary ranges for each of the top job titles.
- Variation in salary levels among different job titles within the data science field.
- Seekers, employers, and educators to understand salary dynamics and make informed decisions regarding job searches, hiring practices, and educational programs.

DISTRIBUTION OF JOB CATEGORIES BASED ON THEIR COUNTS

- Each bar represents the count of jobs available in the corresponding category.
- The 'Data Science and Research' category stands out with the highest number of jobs.
- Other categories are also visible, but their counts are lower.



BOX PLOT OF SALARY DISTRIBUTION BY EMPLOYMENT TYPE

- Larger IQR, suggesting wider salary spread.
- Several outliers above the box, indicating high earners.
- Lower salaries with a small interquartile range (IQR).
- One outlier indicates significantly higher earnings for at least one freelancer.
- Similar median salary and IQR to full-time employees.
- Small IQR, less variability in earnings.
- Use the provided box plot with colored boxes for each employment type, Highlight median lines and whiskers.

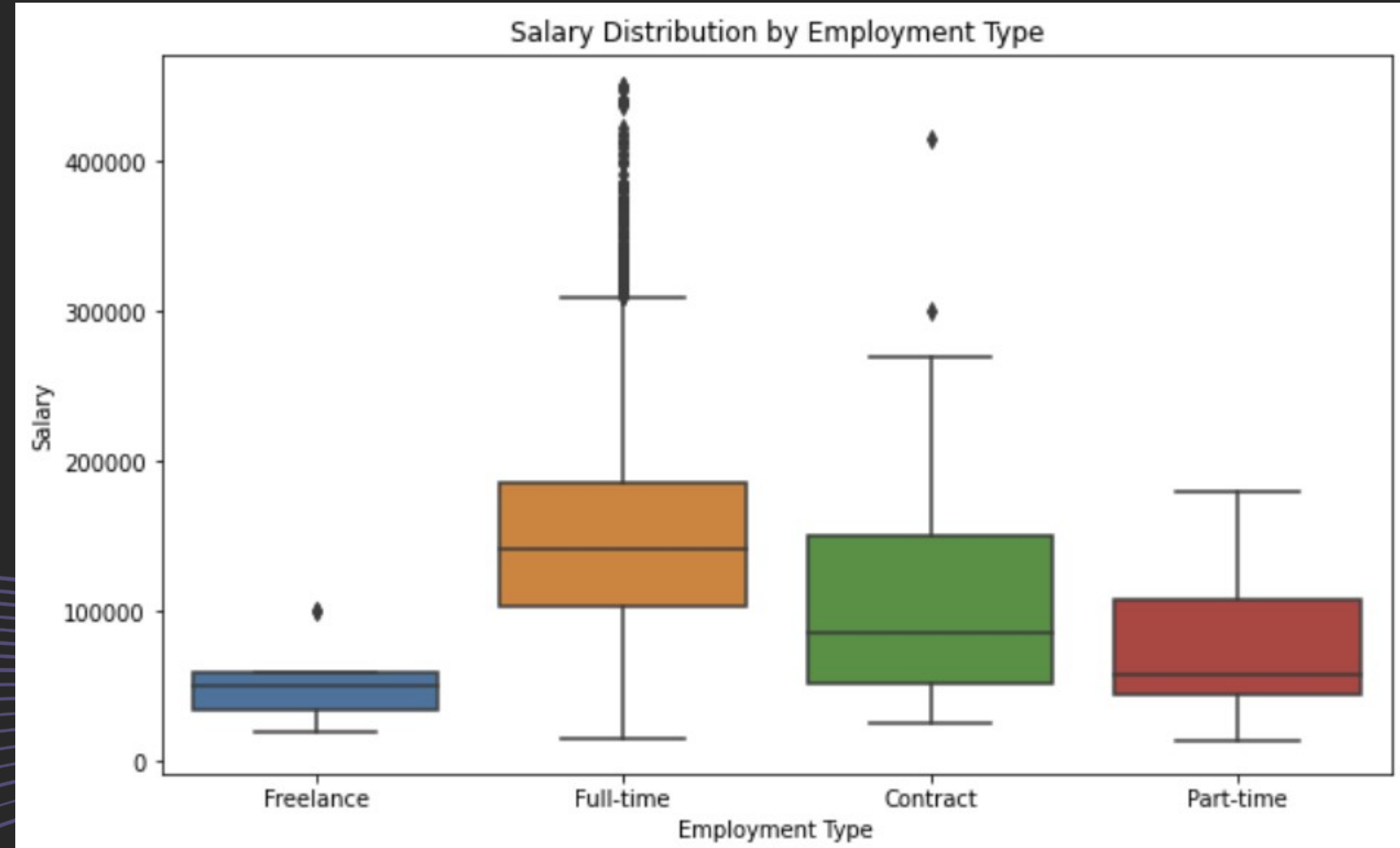
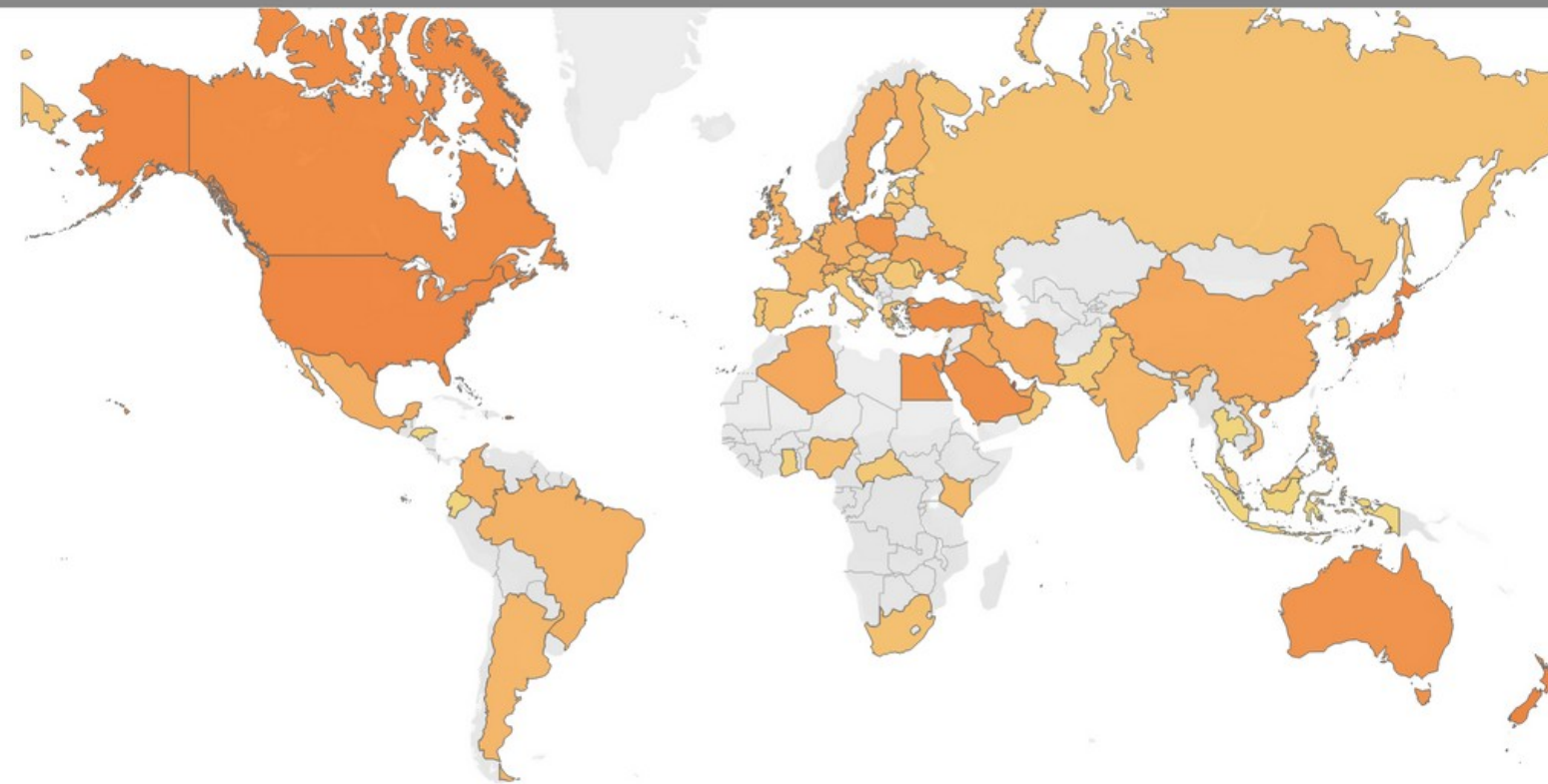


Tableau Dashboard

Geographical Insights



© 2024 Mapbox © OpenStreetMap

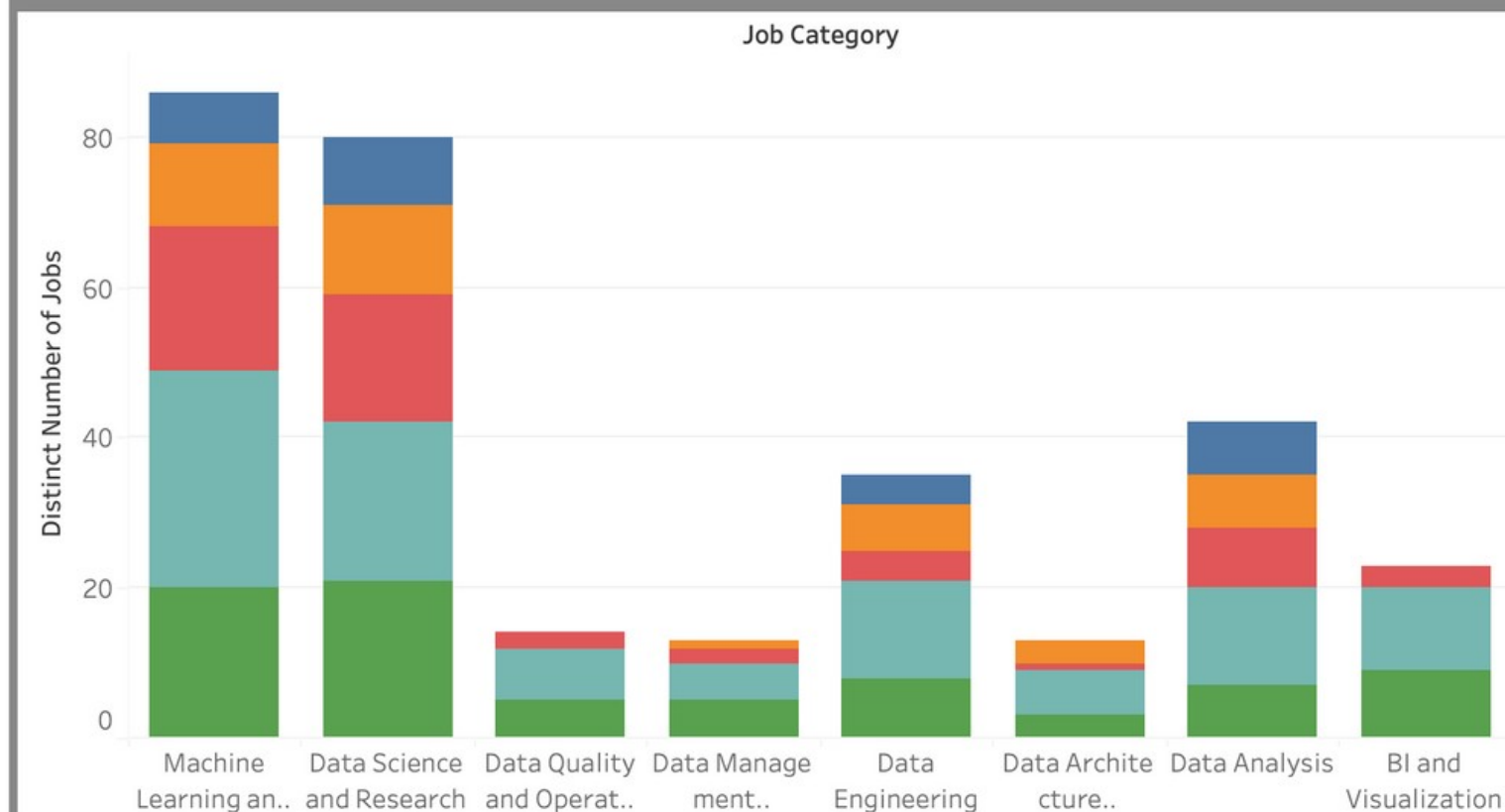
Avg. Salary
15,000 300,000

Job Category
☒ BI and Visualization
☐ Cloud and Database
☒ Data Analysis
☒ Data Architecture and..
☒ Data Engineering
☒ Data Management an..
☒ Data Quality and Oper..
☒ Data Science and Res..
☐ Leadership and Mana..
☒ Machine Learning and..

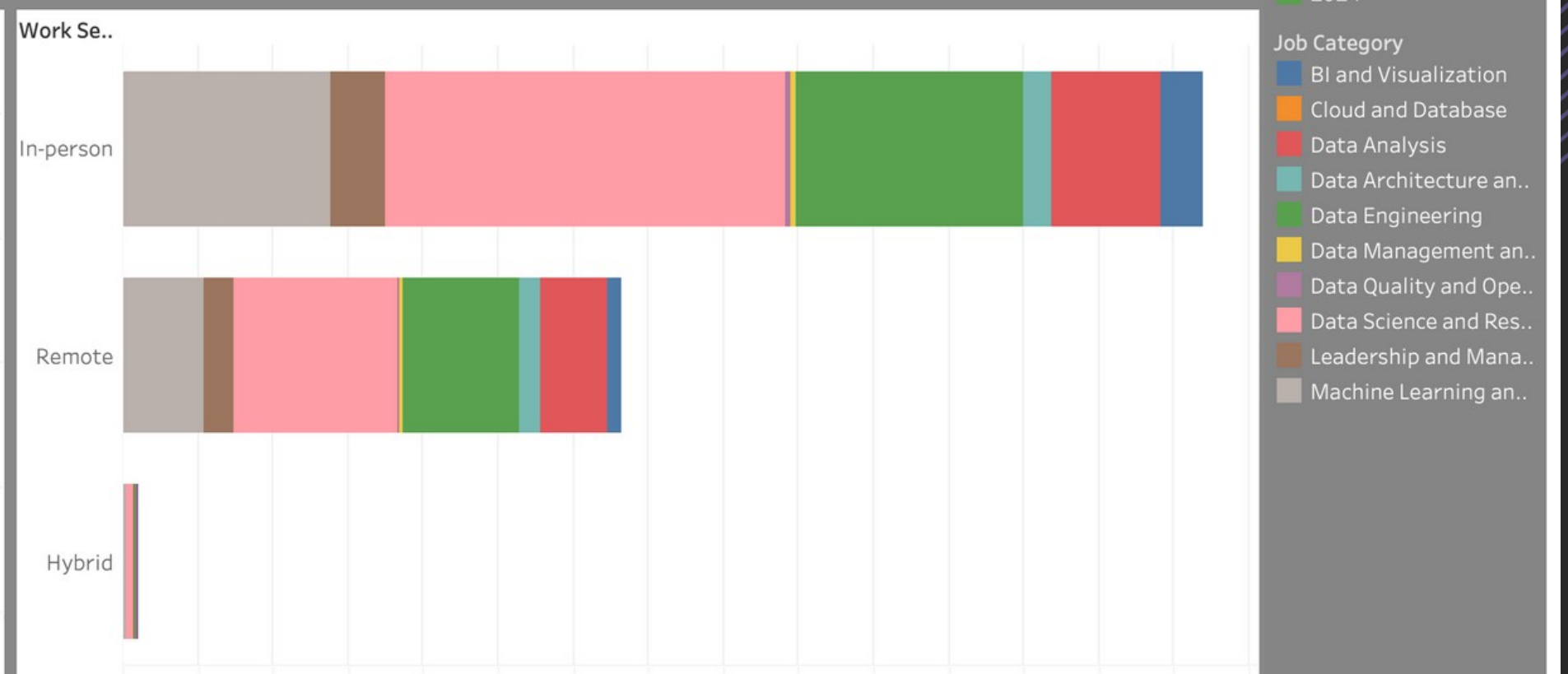
Work Year
2020 to 2024

Work Year
2020
2021
2022
2023
2024

Hiring Trends Over Time

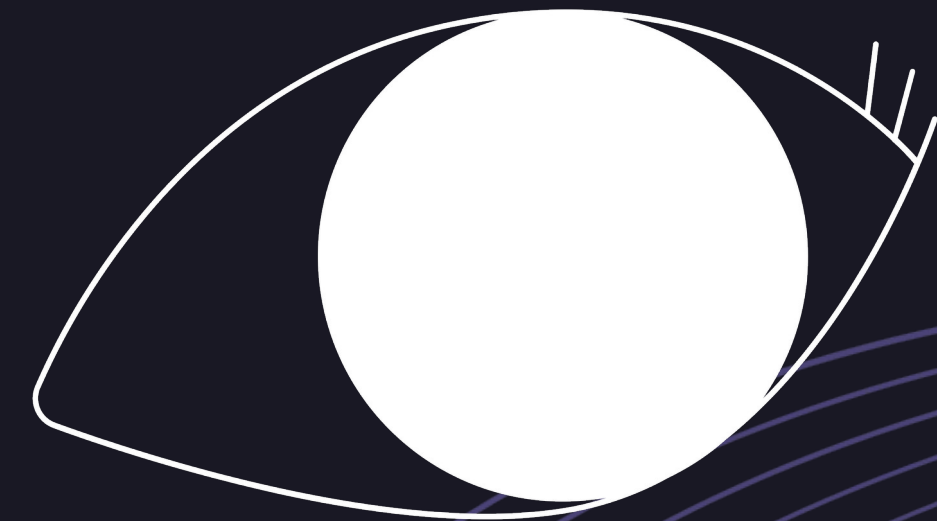


Work Setting Vs Salary



CONCLUSION

- Significant insights on data-related careers and pay scales are revealed by the analysis.
- Experience level has a significant impact on pay in a number of different job areas.
- High-paying jobs have abundant chances for professional advancement.
- In some countries, entry-level data analysts identify the best prospects.
- Making wise professional selections is aided by having a thorough understanding of wage differences.
- Salary comparisons are improved with the inclusion of cost of living statistics.
- Fair representation is ensured by correcting dataset imbalances
- Real-time trend knowledge is ensured by dynamic analysis with weekly updates.



REFERENCES

- The ambiguity of data science team roles and the need for a data science workforce framework. (2017, December 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/8258190>
- GlassDoor Job Description Analytics – Analyzing Data science Professional roles and skills. (2021, April 21). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/9453931>
- Shapiro, J. (2023, September 8). 4 Skills the next generation of data scientists needs to develop. Harvard Business Review. <https://hbr.org/2023/09/4-skills-the-next-generation-of-data-scientists-needs-to-develop>
- Van Loon, R. (2024, January 16). Top 20+ Data Scientist skills You need in 2024. Simplilearn.com. <https://www.simplilearn.com/what-skills-do-i-need-to-become-a-data-scientist-article s>
- The Global AI, ML, Data Science Salary Index for 2024. (n.d.). ai-jobs.net. <https://ai-jobs.net/salaries/2024/>

THANK YOU

Team 7

