DATA SCIENCE JOB SALARIES ANALYSIS

By Himanshu Nagapure

INTRODUCTION

In today's competitive job market, understanding salary trends in data science is crucial for professionals seeking career growth and employers making hiring decisions.
This report explores salary distributions, employment types, experience levels, job titles, and key trends affecting data professionals globally. Using statistical modeling and machine learning, we provide insights to help individuals make informed career choices.

LETS GET STARTED

EXPLORATORY DATA ANALYSIS

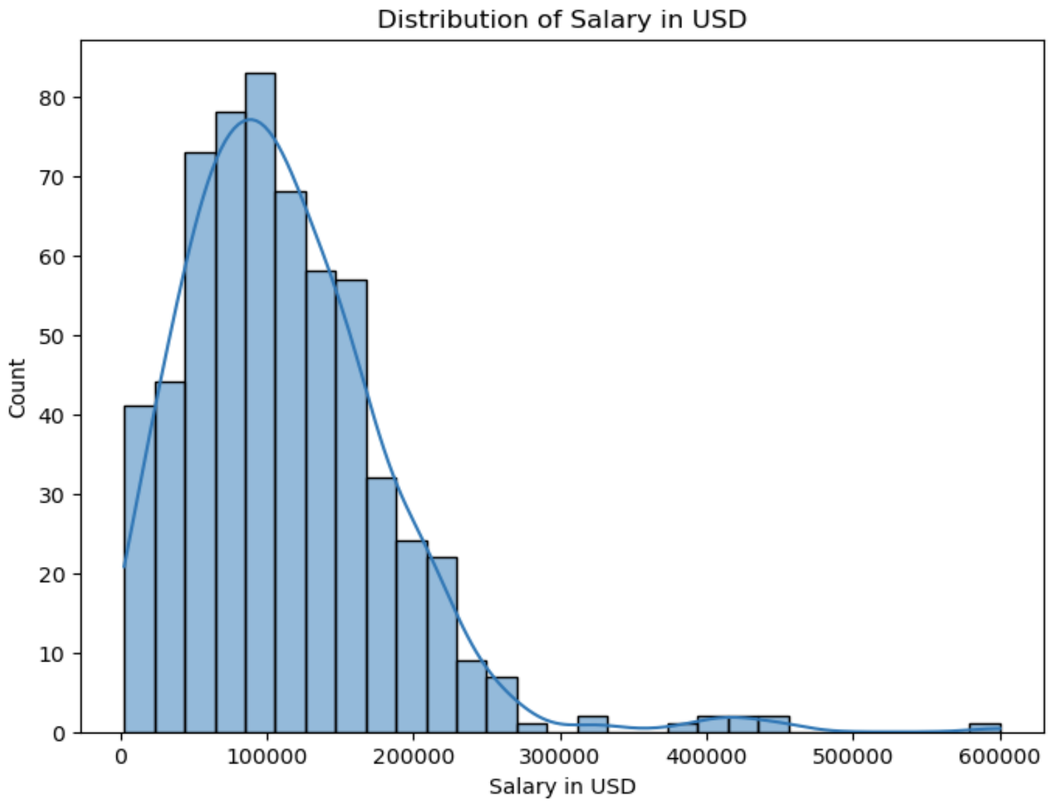
Through Exploratory Data Analysis (EDA), we've examined the structure of our data, identified key variables, outliers, and tested assumptions.
This has provided critical insights, guiding our choice of models for further analysis and ensuring our data is ready for more detailed investigation and reliable results

DATASET DESCRIPTION

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 607 entries, 0 to 606
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   Unnamed: 0          607 non-null    int64 
 1   work_year           607 non-null    int64 
 2   experience_level    607 non-null    object
 3   employment_type     607 non-null    object
 4   job_title           607 non-null    object
 5   salary              607 non-null    int64 
 6   salary_currency     607 non-null    object
 7   salary_in_usd       607 non-null    int64 
 8   employee_residence  607 non-null    object
 9   remote_ratio        607 non-null    int64 
 10  company_location    607 non-null    object
 11  company_size        607 non-null    object
dtypes: int64(5), object(7)
memory usage: 57.0+ KB
```

UNIVARIATE ANALYSIS

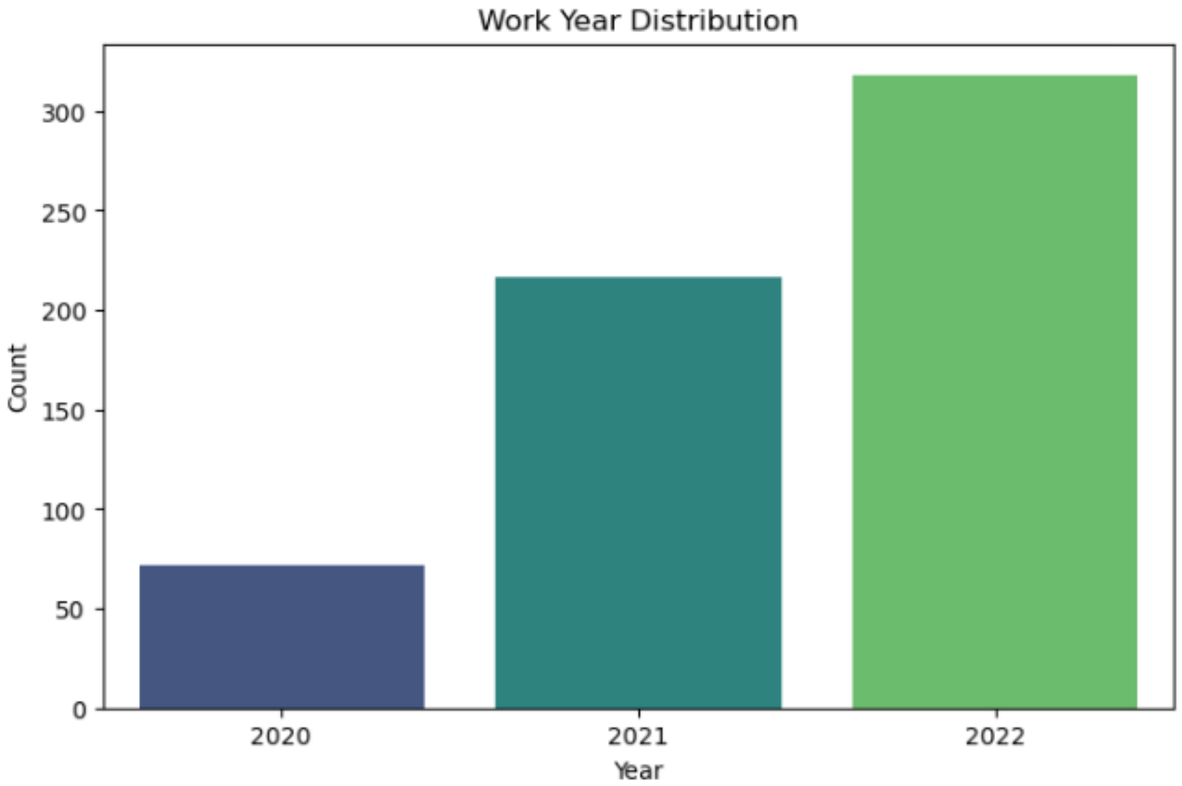
SALARY IN USD



Insights:

• The salary distribution for data science professionals is right-skewed, with a majority of professionals earning lower to mid-range salaries.
• A noticeable peak in the distribution suggests a concentration of professionals within a specific salary range.

WORK YEAR DISTRIBUTION

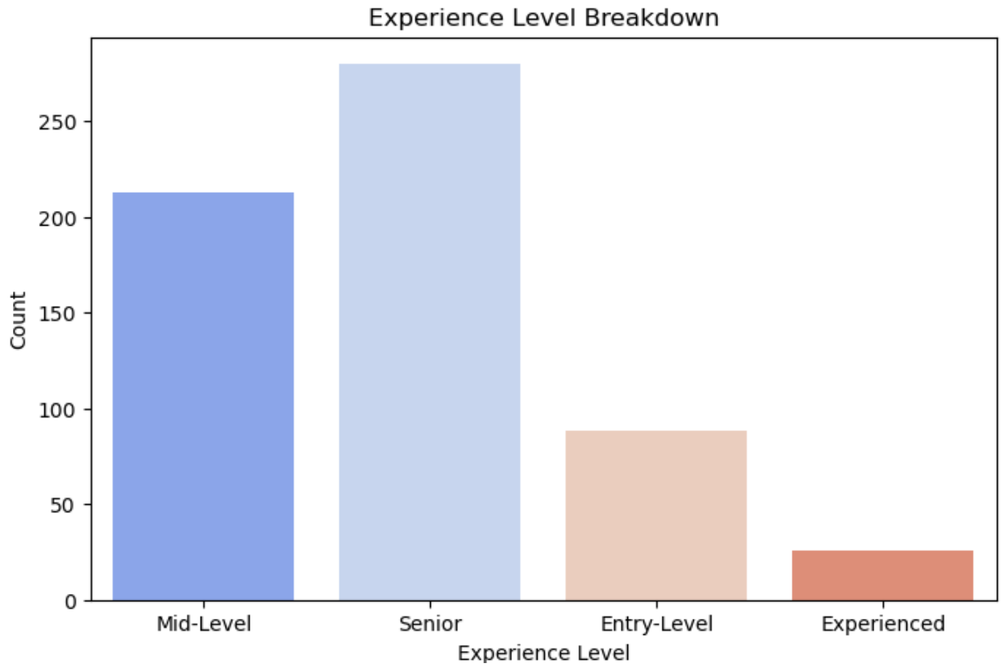


Insights:

• The number of records grows from 2020 to 2022. This suggests more data is available for recent years, possibly due to an increase in data science jobs or better data collection.

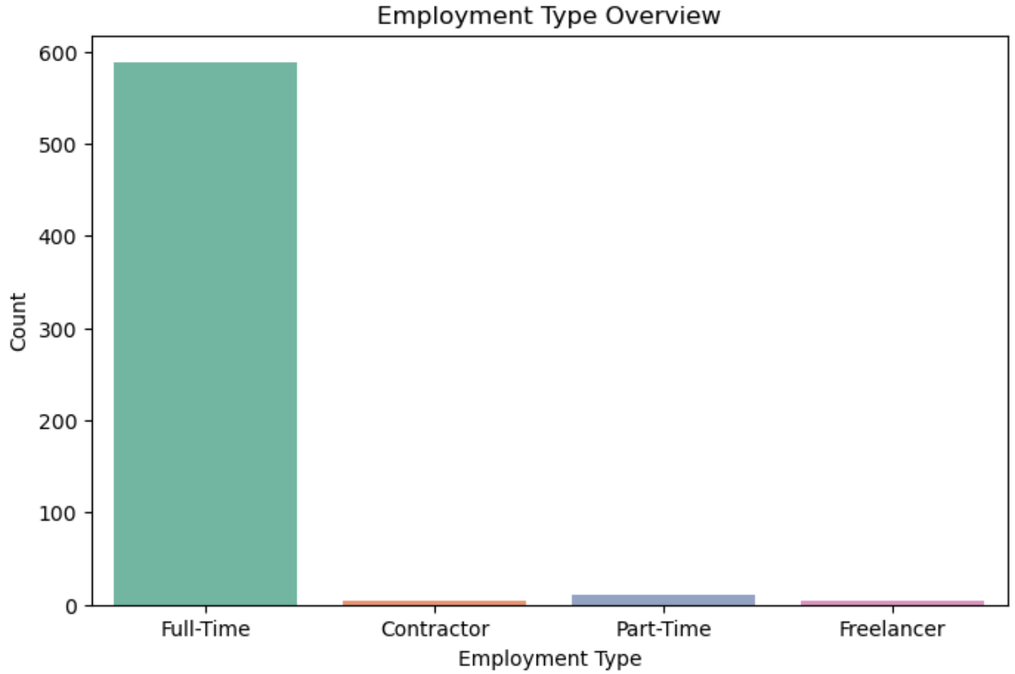
UNIVARIATE ANALYSIS

EXPERIENCE LEVEL



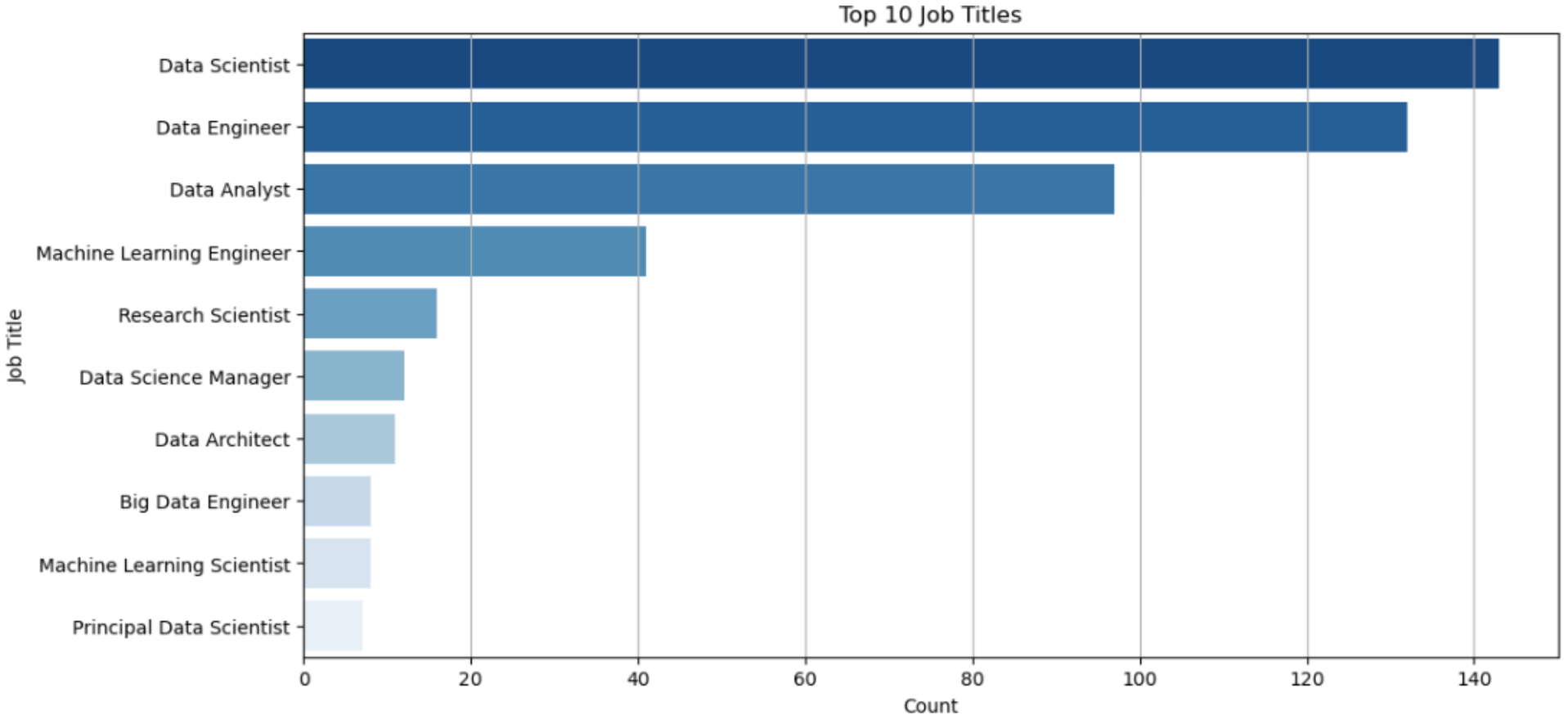
Data Scientist	143
Data Engineer	132
Data Analyst	97
Machine Learning Engineer	41
Research Scientist	16
Data Science Manager	12
Data Architect	11
Big Data Engineer	8
Machine Learning Scientist	8
Principal Data Scientist	7
AI Scientist	7
Data Science Consultant	7
Director of Data Science	7
Data Analytics Manager	7
ML Engineer	6
Computer Vision Engineer	6

EMPLOYMENT TYPE



UNIVARIATE ANALYSIS

TOP 10 JOB TITLES

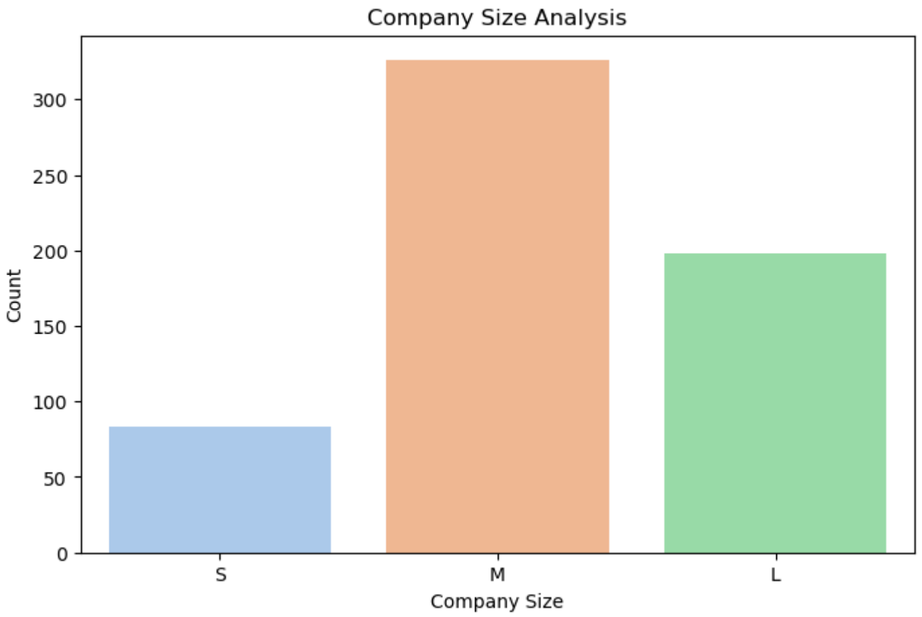


UNIVARIATE ANALYSIS

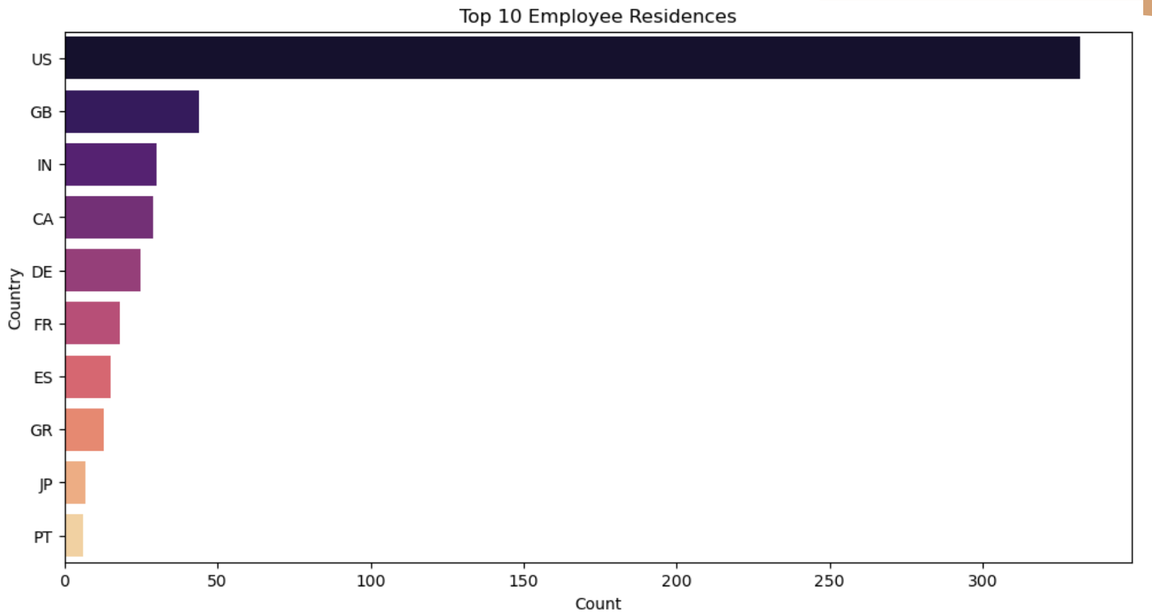
COMPANY SIZE

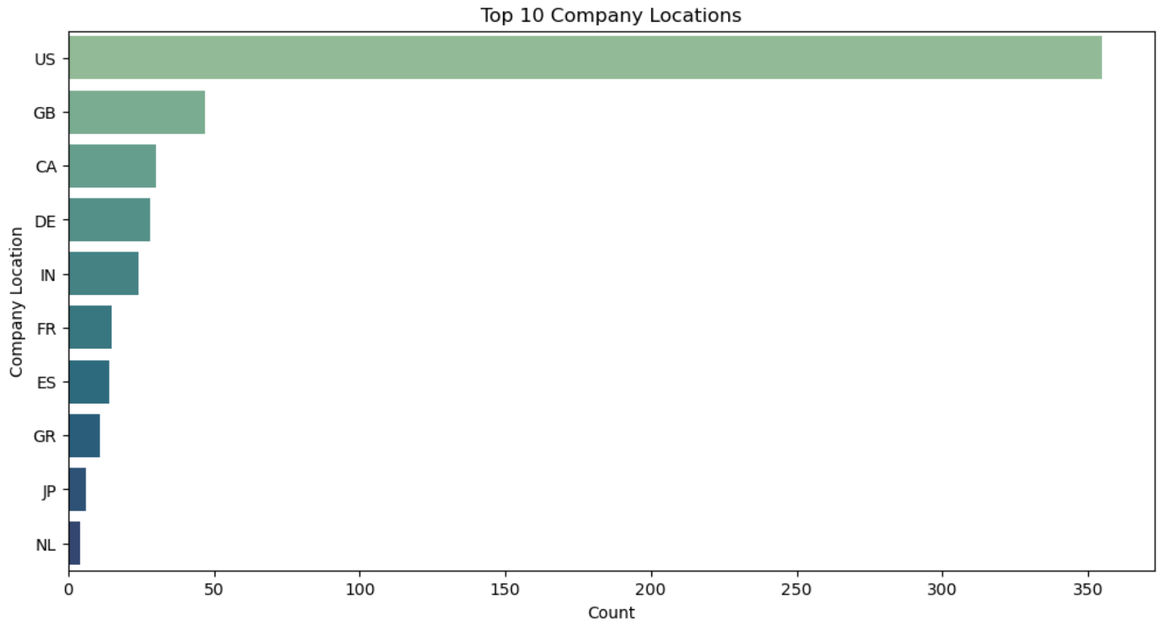
```
M    326
L    198
S     83
Name: company_size, dtype: int64

: Text(0.5, 1.0, ' Distribution of Company Size')
```



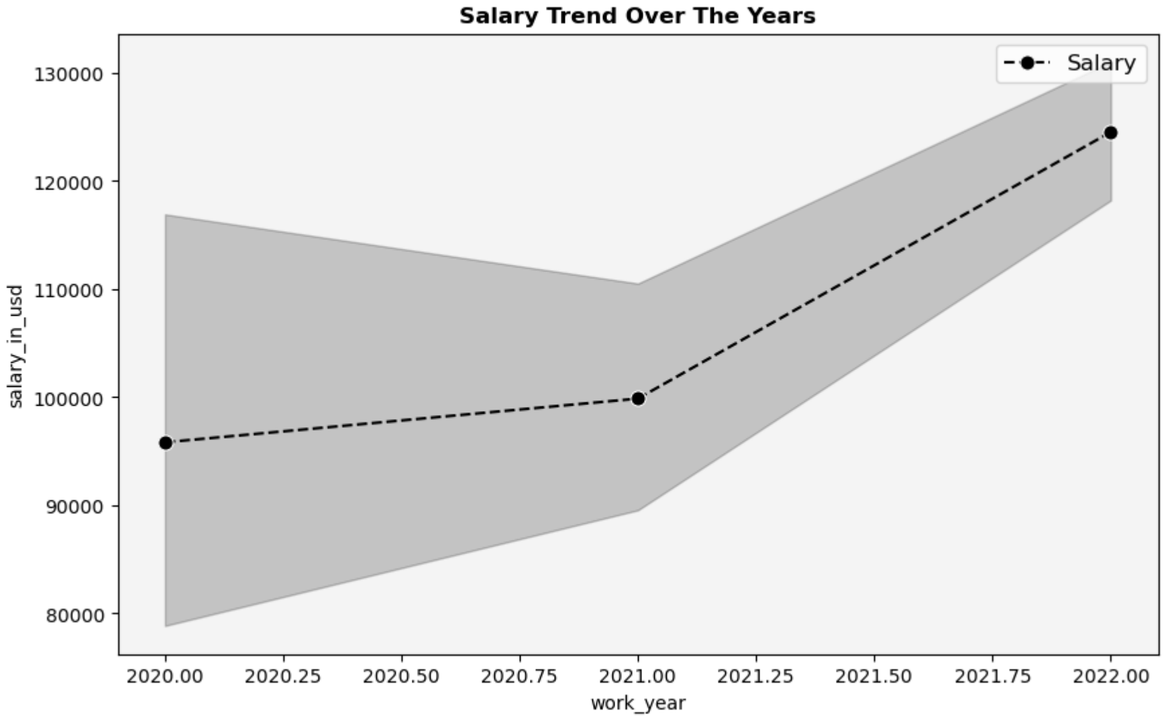
LOCATION





BIVARIATE ANALYSIS

WORK YEAR ANALYSIS



Insights:

The median salary increases from 2020 to 2022, showing a positive salary trend in the data science field.
More high-paying roles are emerging, possibly due to demand for experienced professionals.

BIVARIATE ANALYSIS

SALARY BY EXPERIENCE LEVEL

Insights:

• More experience leads to higher pay, but the variation in salaries suggests that factors like company, location, and industry may also play a role.
• Some professionals earn significantly higher salaries, possibly due to niche expertise or leadership roles.
• Entry-level salaries are relatively low, reinforcing the need for experience and skill growth in the field.

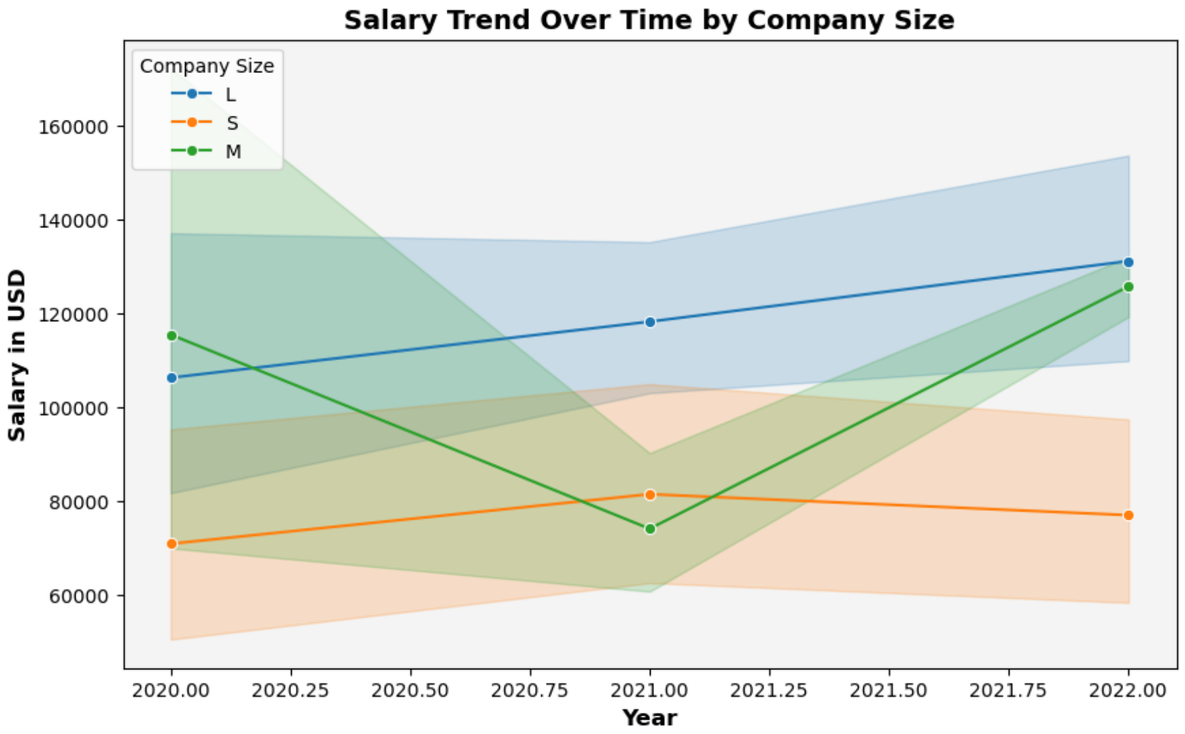
EMPLOYMENT TYPEVS SALARY

Insights:

• Contractors earn the most but have the highest salary variation.
• Full-time employees have stable salaries with people with high-salaries than any others.
• Part-time and freelancers earn the least, with lower salary variability.

BIVARIATE ANALYSIS

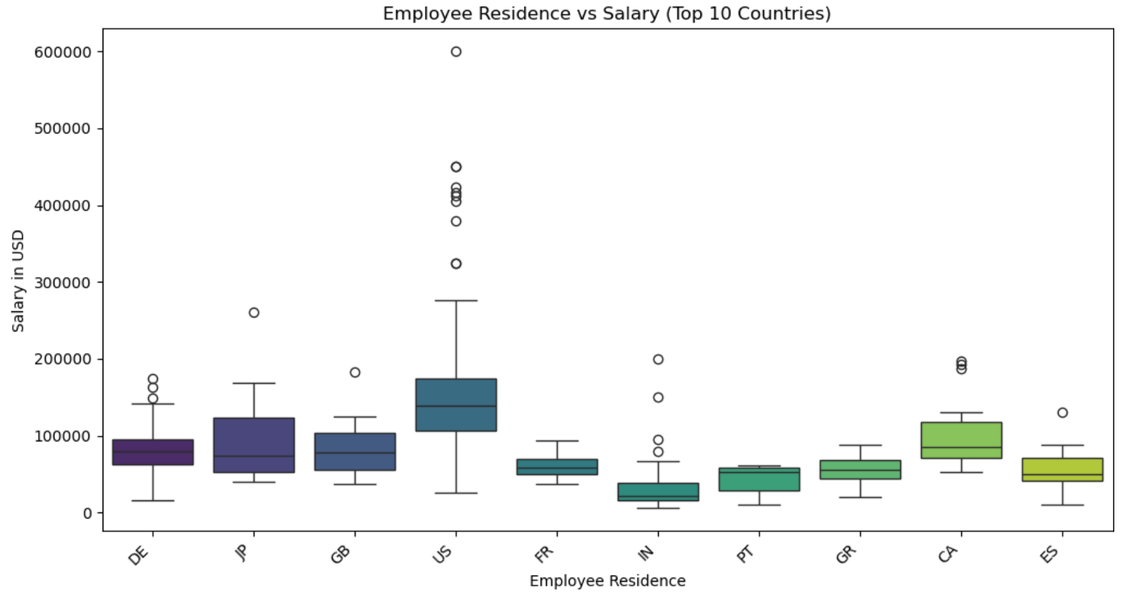
SALARY BY COMPANY SIZE



Insights:

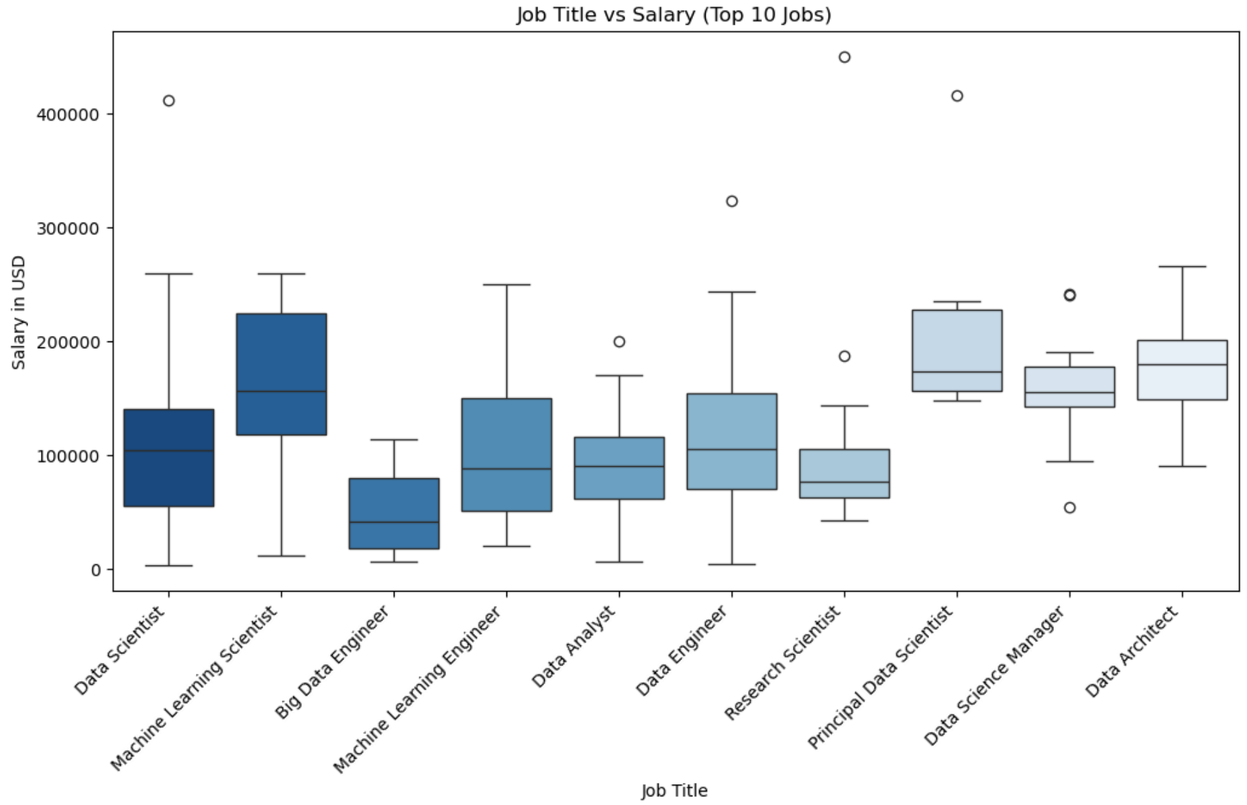
The salary trend in medium-sized companies is increasing compared to small and large companies. The salary line in large companies is not fluctuating as much as in medium-sized companies.

LOCATION

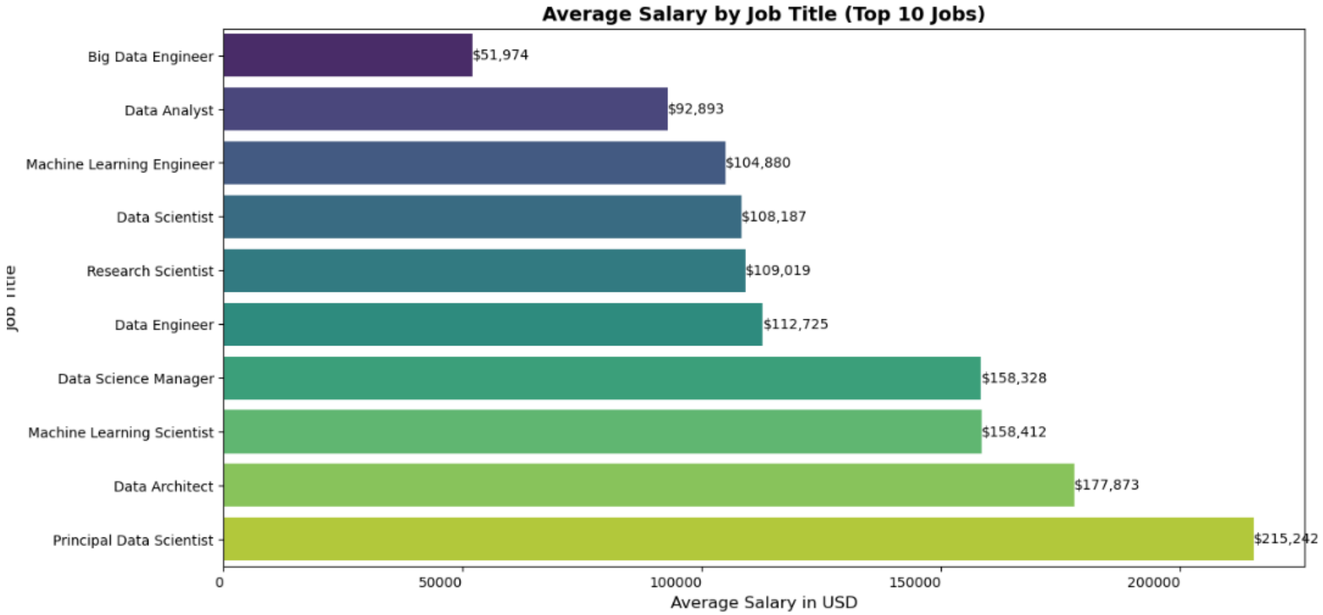


BIVARIATE ANALYSIS

SALARY BY JOB TITLE



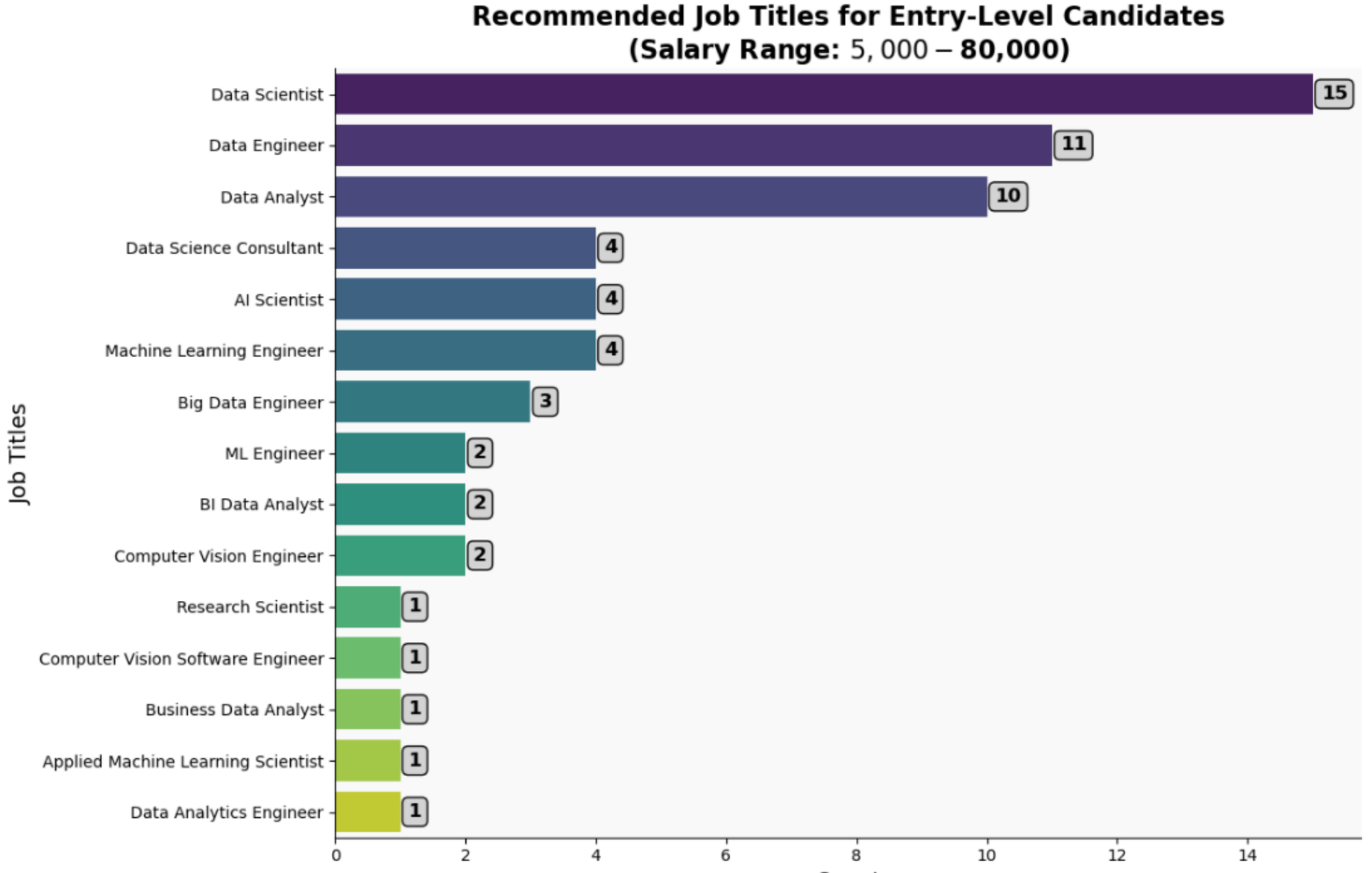
AVERAGE SALARY BY JOB TITLE



Insights:

- Leadership and specialized roles (Data Science Manager, ML Scientist, Principal Data Scientist, Data Architect) have the highest salaries.
- Technical roles (Data Scientist, Data Engineer, ML engineer, Research Scientist) have decent but varied salaries.
- Entry-level roles (Data Analyst, Big Data Engineer) earn less but provide a stepping stone into higher-paying jobs.

BONUS


Recommended Job Titles for Entry-Level Candidates
(Salary Range: 5,000 – 80,000)
Job Titles
Data Scientist
Data Engineer
Data Analyst
Data Science Consultant
AI Scientist
Machine Learning Engineer
Big Data Engineer
ML Engineer
BI Data Analyst
Computer Vision Engineer
Research Scientist
Computer Vision Software Engineer
Business Data Analyst
Applied Machine Learning Scientist
Data Analytics Engineer
15
11
10
4
4
4
3
2
2
2
1
1
1
1
1
0
2
4
6
8
10
12
14

MODELING

	MAE	MSE	RMSE	R2 Score
Linear Regression	54.065140	5361.743853	73.223930	0.521414
Ridge Regression	49.705779	4555.718142	67.496060	0.593359
Lasso Regression	49.834473	4602.111230	67.838862	0.589218
Random Forest	54.286213	5603.642377	74.857480	0.499822
XGBoost	55.235290	5966.747303	77.244723	0.467412

From the results, Ridge Regression performed the best.

- Lowest MAE (49.70) → Most accurate predictions on average.
- Lowest MSE (4555.71) & RMSE (67.49) → Smaller errors overall.
- Highest R^2 Score (0.5933) → Explains ~59.3% of variance in salary.

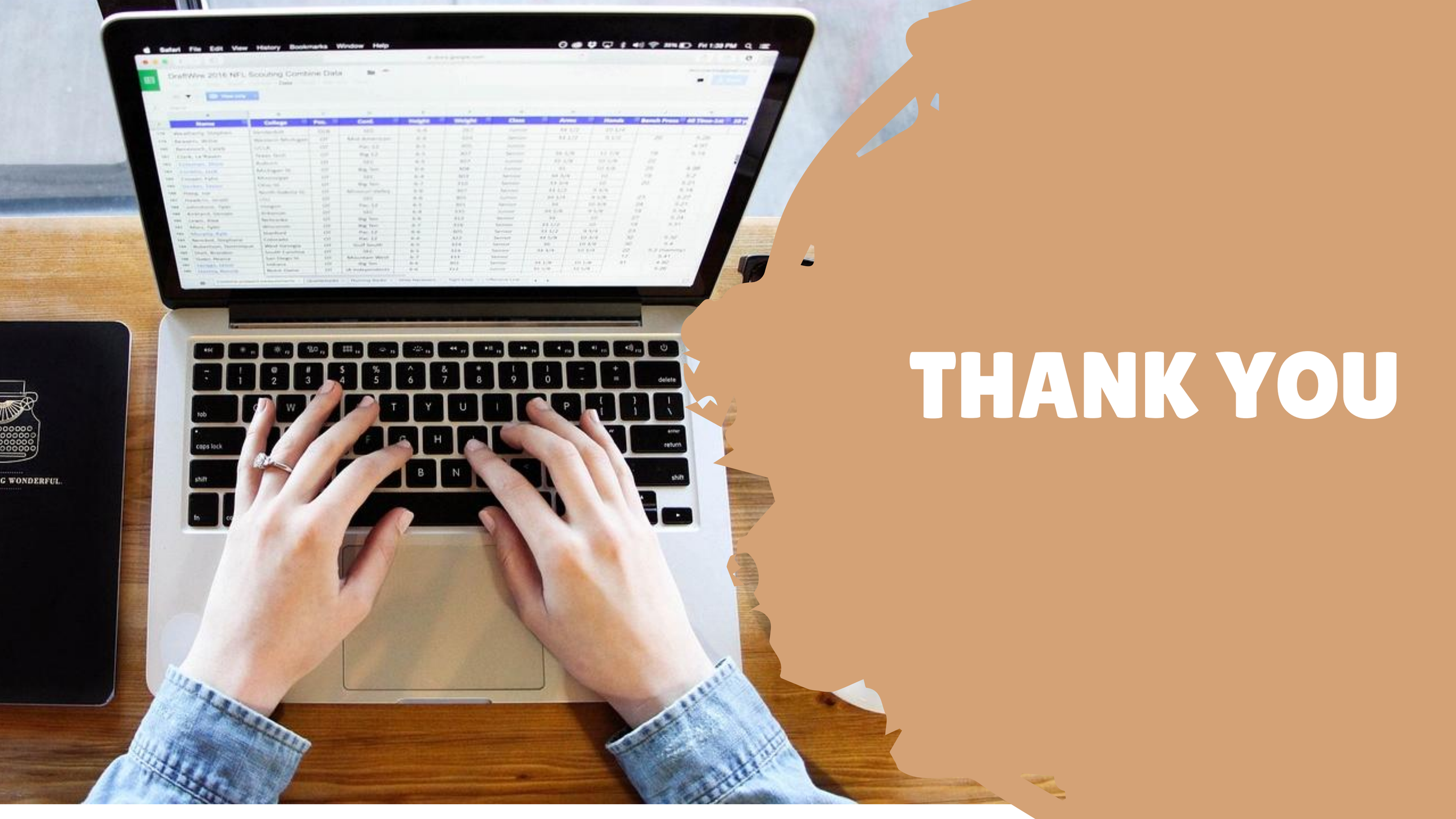
SUMMARY

This analysis provides a clear understanding of salary trends in data science and offers actionable recommendations for professionals at different career stages. By leveraging data-driven insights, individuals can make strategic career moves that maximize their earning potential and job satisfaction.

🚀 **Key Takeaways:**

- Salary growth is influenced by **experience, job role, company size, and location**.
- **Medium-sized companies offer the best salary growth potential**.
- **The US remains the best-paying country**, with Japan and Canada emerging as strong alternatives.
- **Full-time roles provide stable salaries, while contracting offers higher but fluctuating pay**.
- **For optimal salary growth, professionals should upskill in ML, AI, and cloud computing**.

This report provides a **data-backed roadmap** for professionals looking to enhance their careers in data science. By using salary data effectively, candidates can make **informed career decisions** and **negotiate better salaries** with confidence.



THANK YOU