Our Project.

DATA SCIENCE CAREER & SALARY SCOPE

Dr/Sara Ashraf

PROBLEM STATEMENT

The Challenge:

- Salaries in the tech industry vary a lot by job title, experience level, and company location.
- Companies need insights to offer fair and competitive salaries.
- Employees want to know if they are being paid fairly.

Our Goal:

- Build a Machine Learning model to predict salaries based on job and company features.

ABOUT DATASET

1. job_title:

- The job role (e.g., Data Scientist, Data Engineer, ML Engineer).

2. experience_level:

- Level of experience (Junior, Mid-level, Senior, Executive).

3. employment_type:

- Type of employment (Full-time, Part-time, Contract, Freelance).

4. work_models:

- Work model (Remote, Hybrid, On-site).

CONTINUE...

5. work_year:

- Year when the salary was recorded (e.g., 2020, 2021, 2022).

6. Employee_residence:

- Country of residence of the employee.

7. Company_location:

- Country where the company is located.

8. Company_size:

- Size of the company (Small, Medium, Large).

Target:

9. Salary:

- Salary in the original currency.

DATA PREPROCESSING

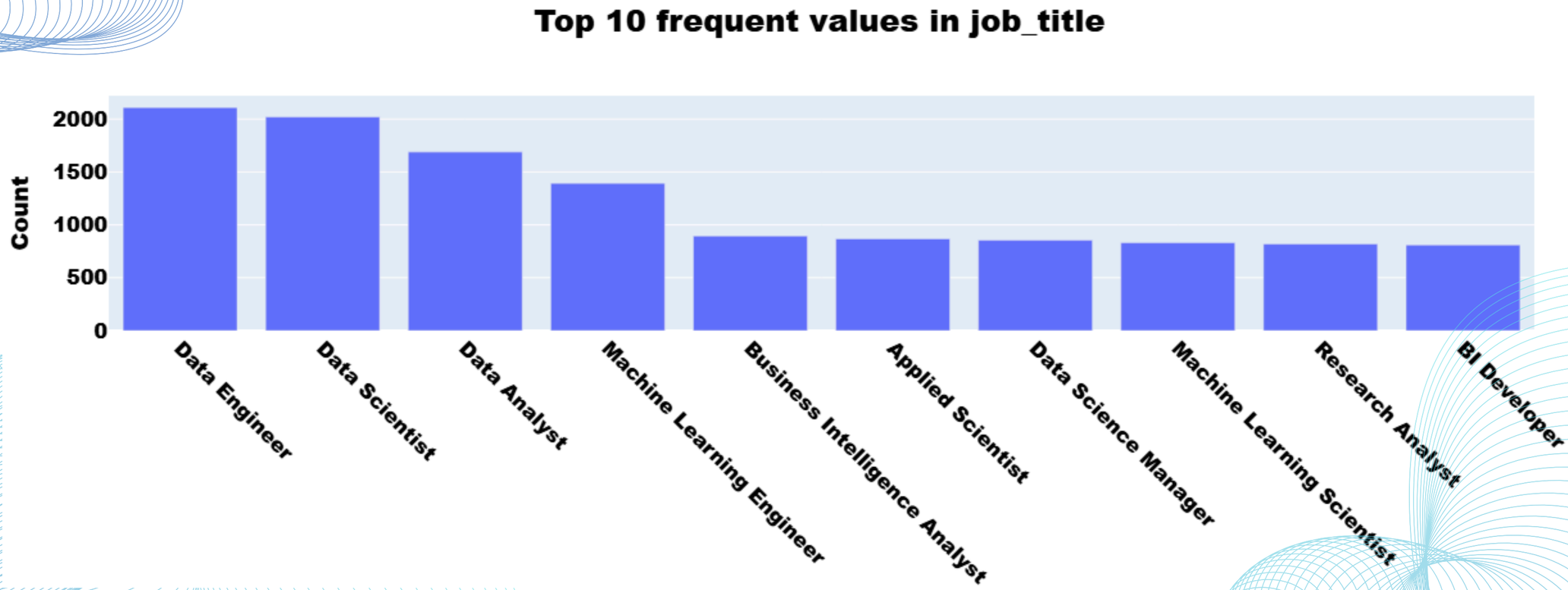
DATA CLEANING & PREPROCESSING

- **Checked for missing values → none found.**
- **Checked for duplicate records → no duplicates detected.**

DATA VISUALIZATION

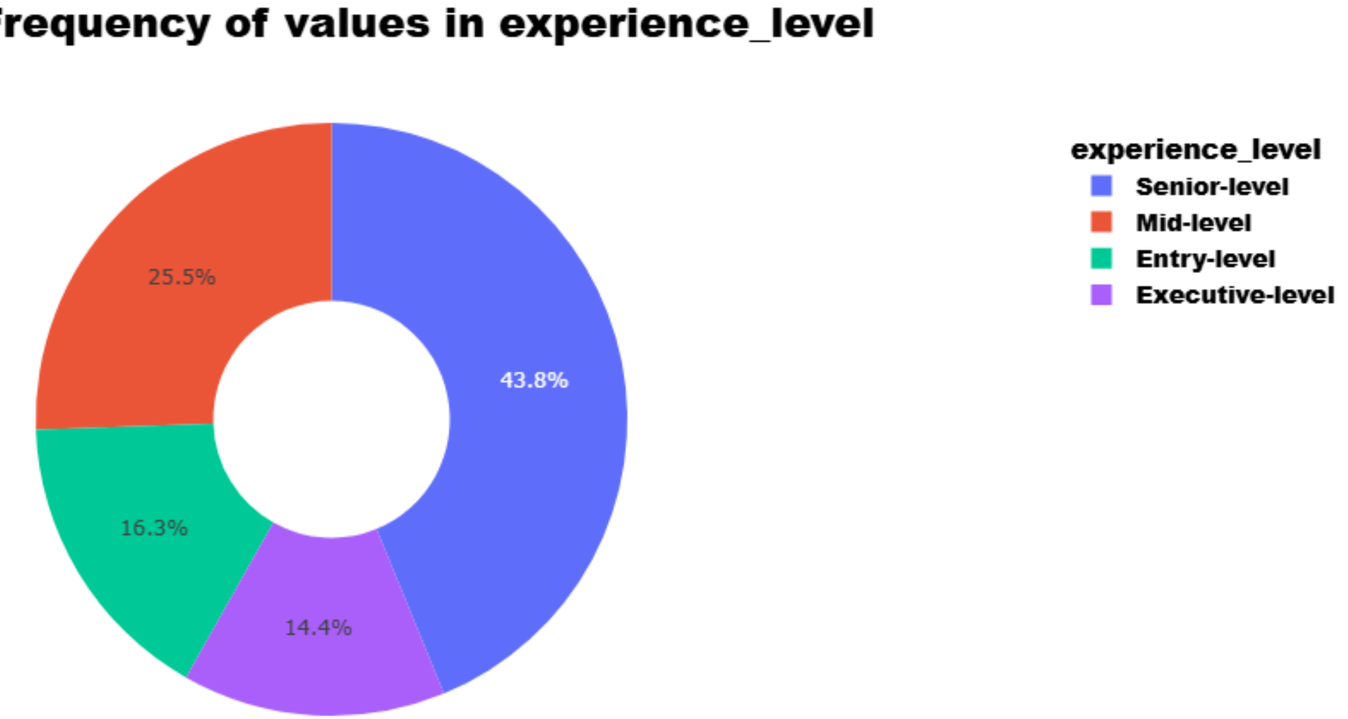
EXPLORATORY DATA ANALYSIS (EDA)

Most frequent job titles are Data Engineer >> Data Scientist then Data Analyst, followed by roles like ML Engineer and BI Analyst.



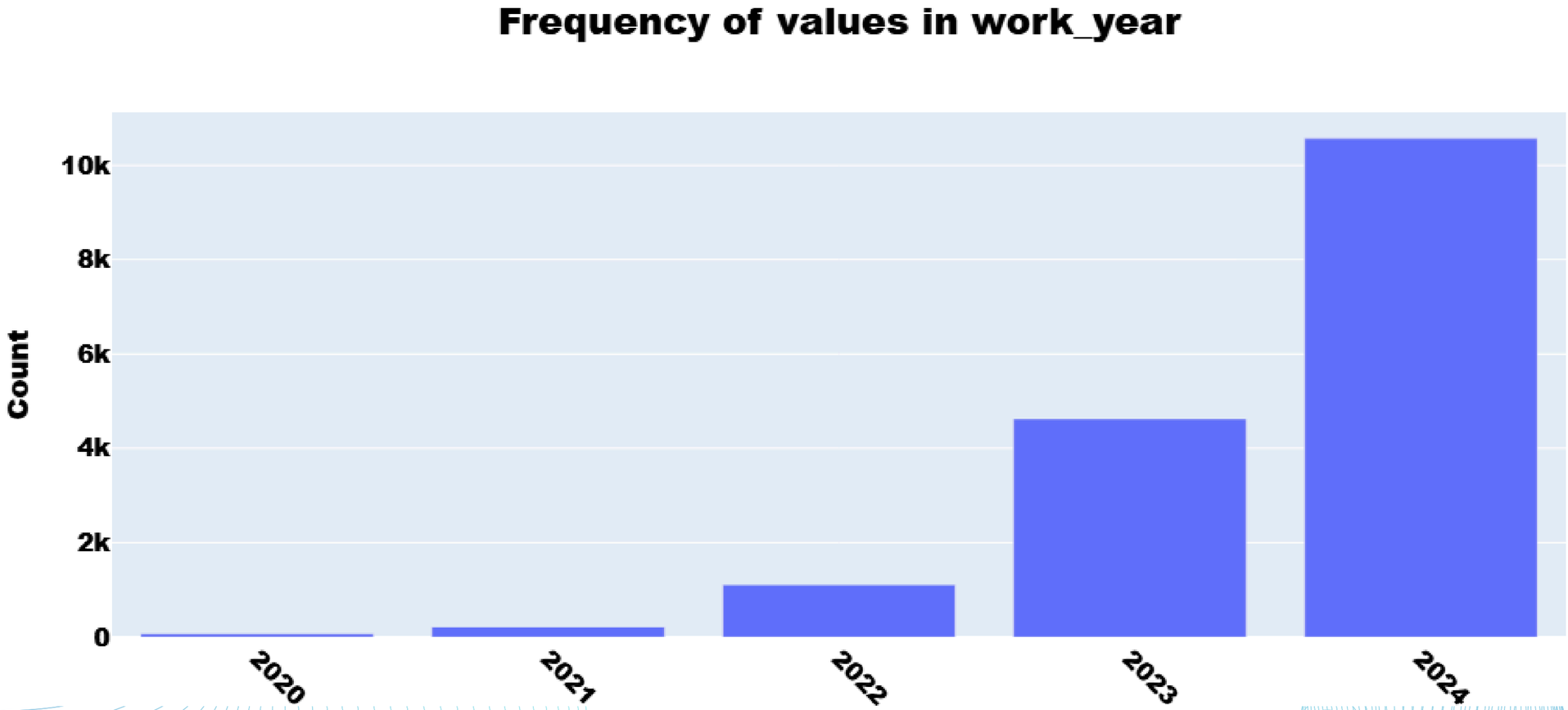
EXPLORATORY DATA ANALYSIS (EDA)

Senior-level is the largest group (43.8%) , Followed by Mid-level (25.5%) Entry-level (16.3%), and Executive-level (14.4%).



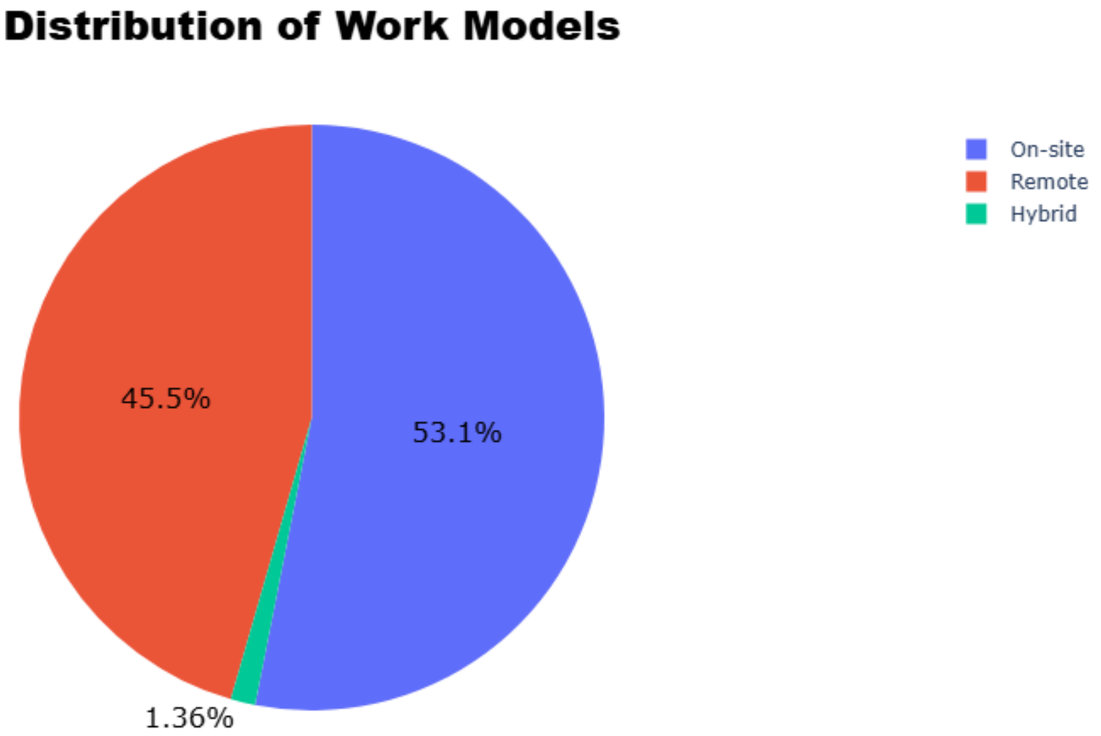
EXPLORATORY DATA ANALYSIS (EDA)

Jobs are increasing year by year, as new and more diverse roles continue to emerge



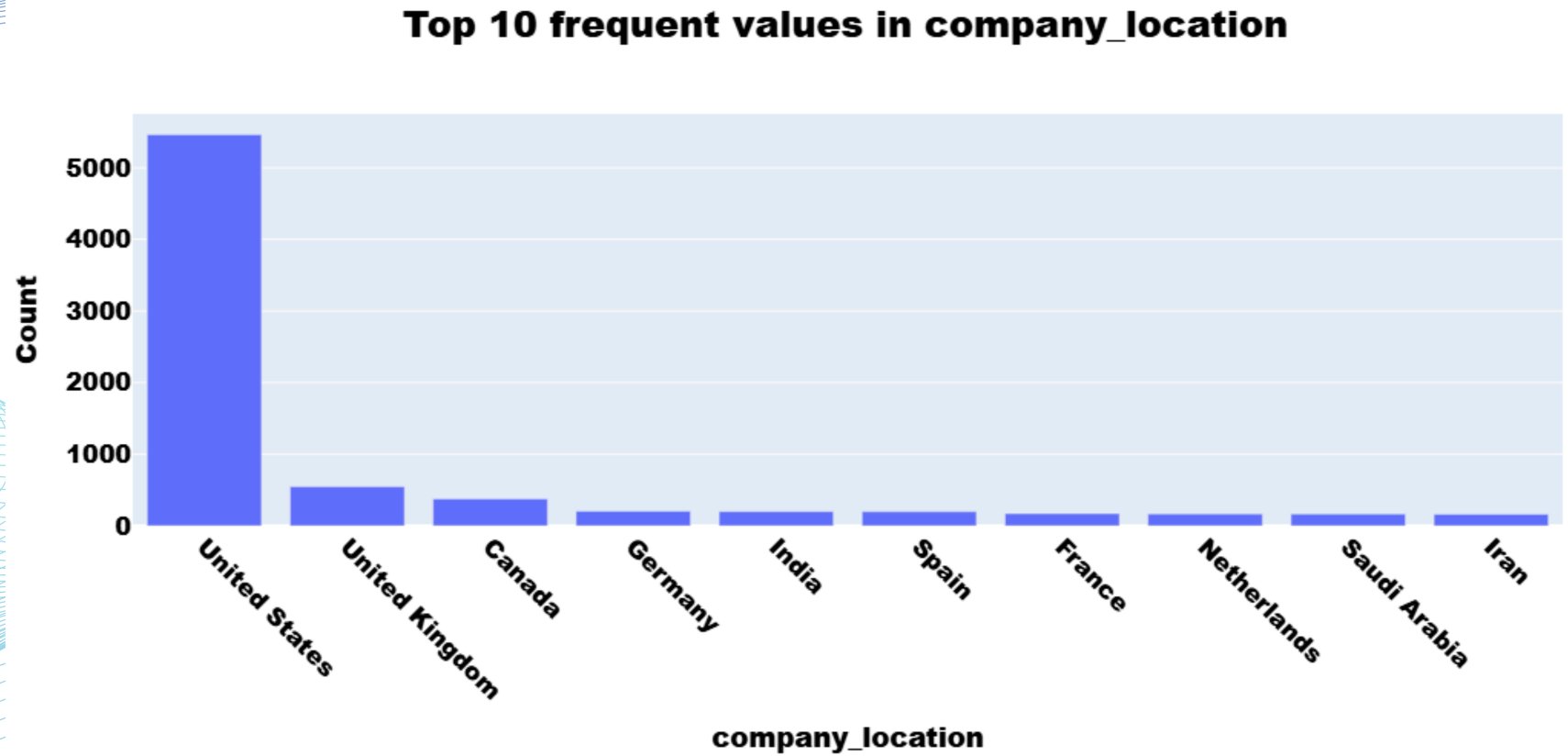
EXPLORATORY DATA ANALYSIS (EDA)

On-site work remains the most common model, followed closely by remote work, while the hybrid model remains marginal.



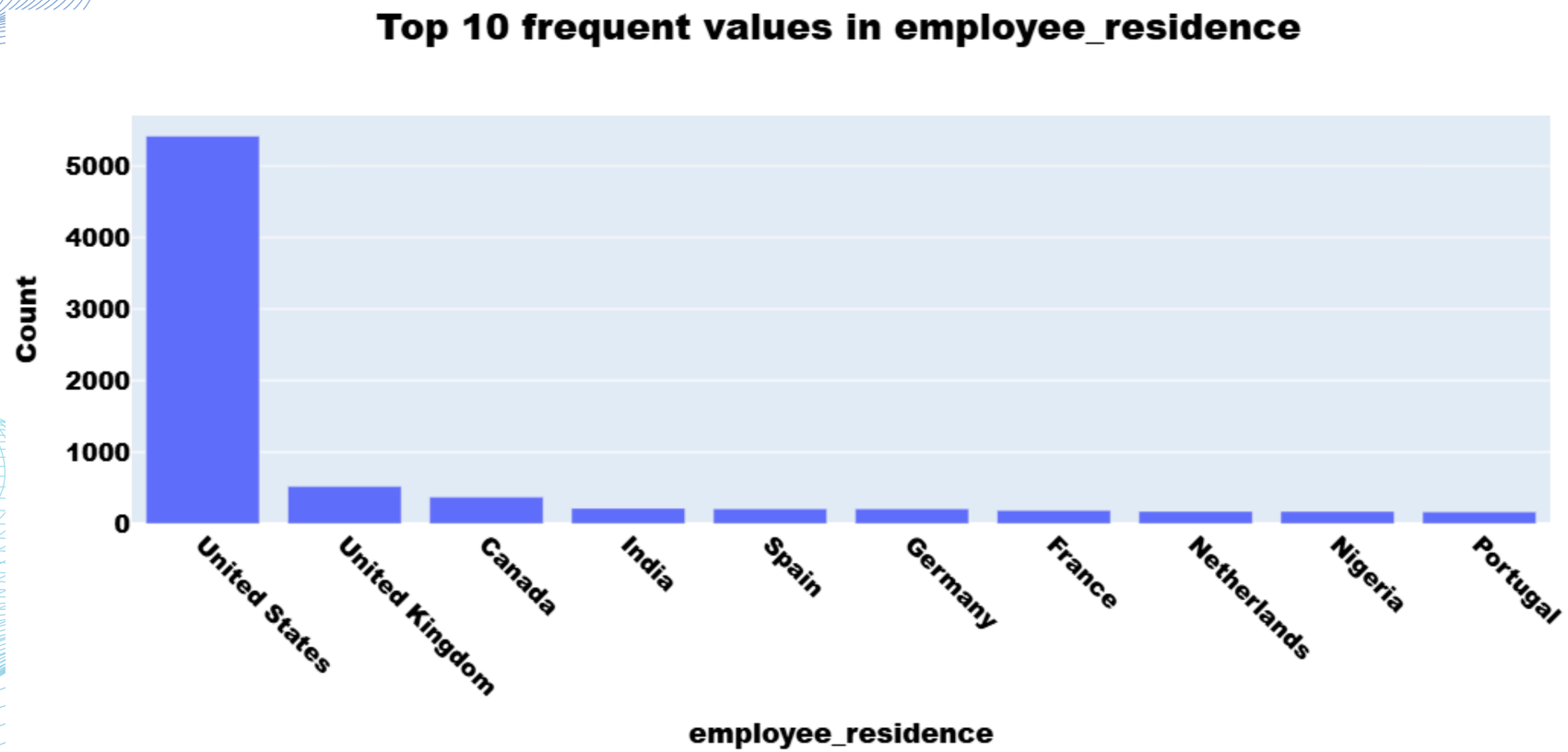
EXPLORATORY DATA ANALYSIS (EDA)

In terms of company location, the United States dominates with a massive presence of around 5,000 companies, followed by the United Kingdom with about 500. Other locations such as Canada, Germany, India, Spain, France, the Netherlands, Brazil, and Iran show significantly lower counts.



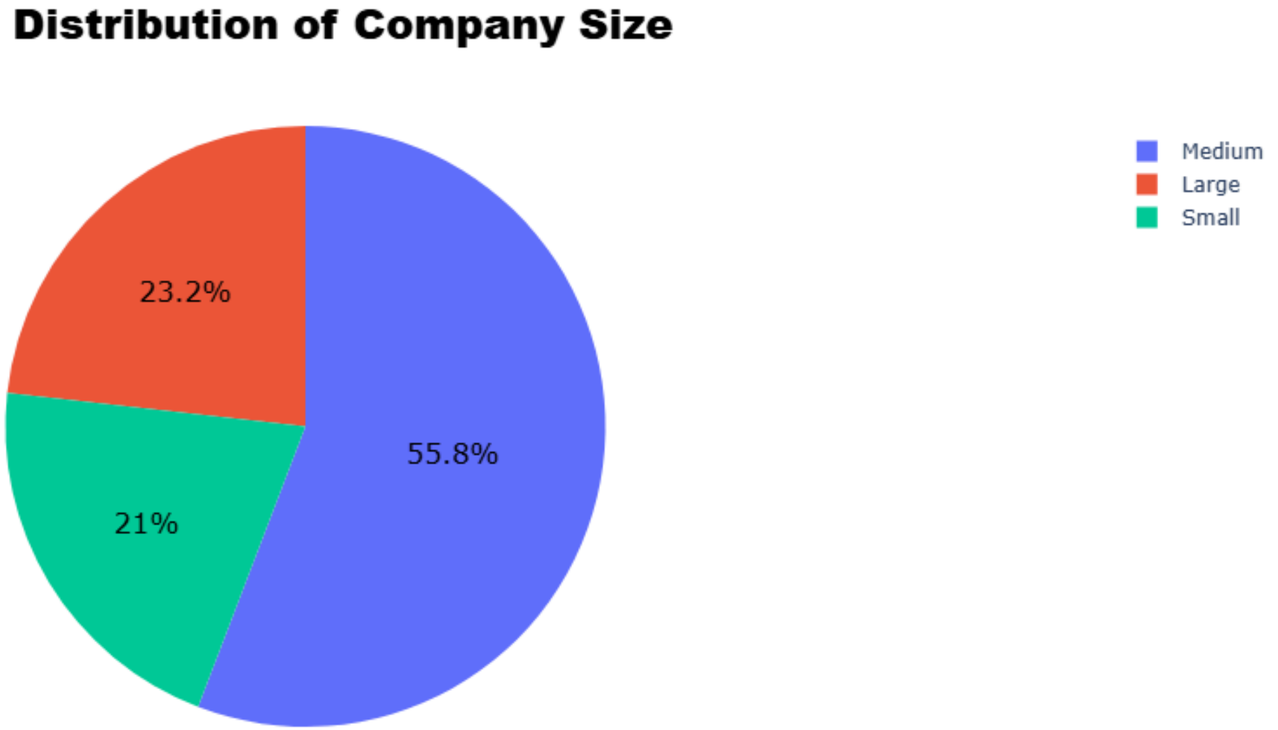
EXPLORATORY DATA ANALYSIS (EDA)

In terms of employee residence, the United States dominates with around 5,000 employees, followed by the United Kingdom with about 500. Other locations such as Canada, India, Spain, Germany, France, the Netherlands, Nigeria, and Portugal show much lower representation.



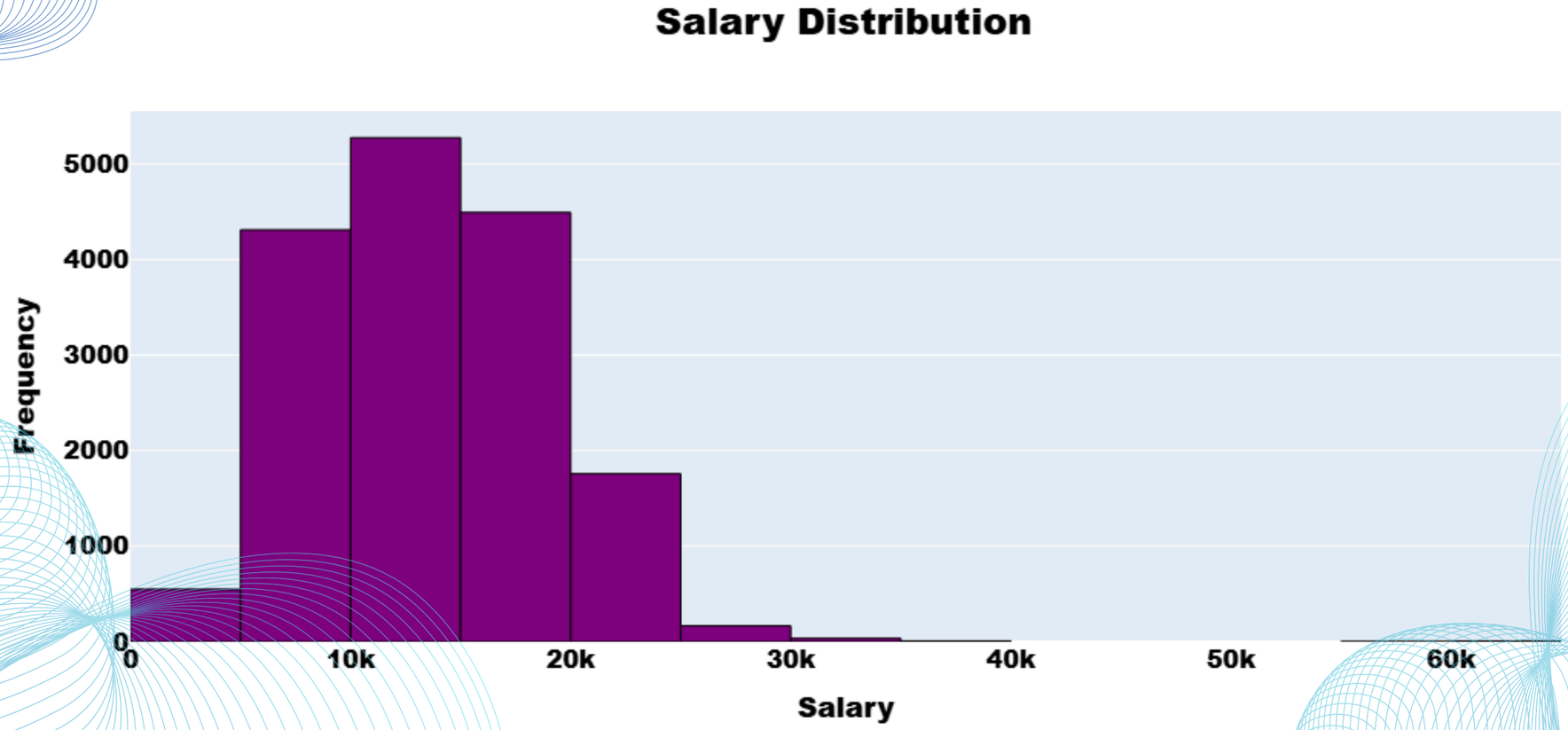
EXPLORATORY DATA ANALYSIS (EDA)

In terms of company size, medium-sized companies represent the majority, followed by large companies, while small companies account for the smallest share.



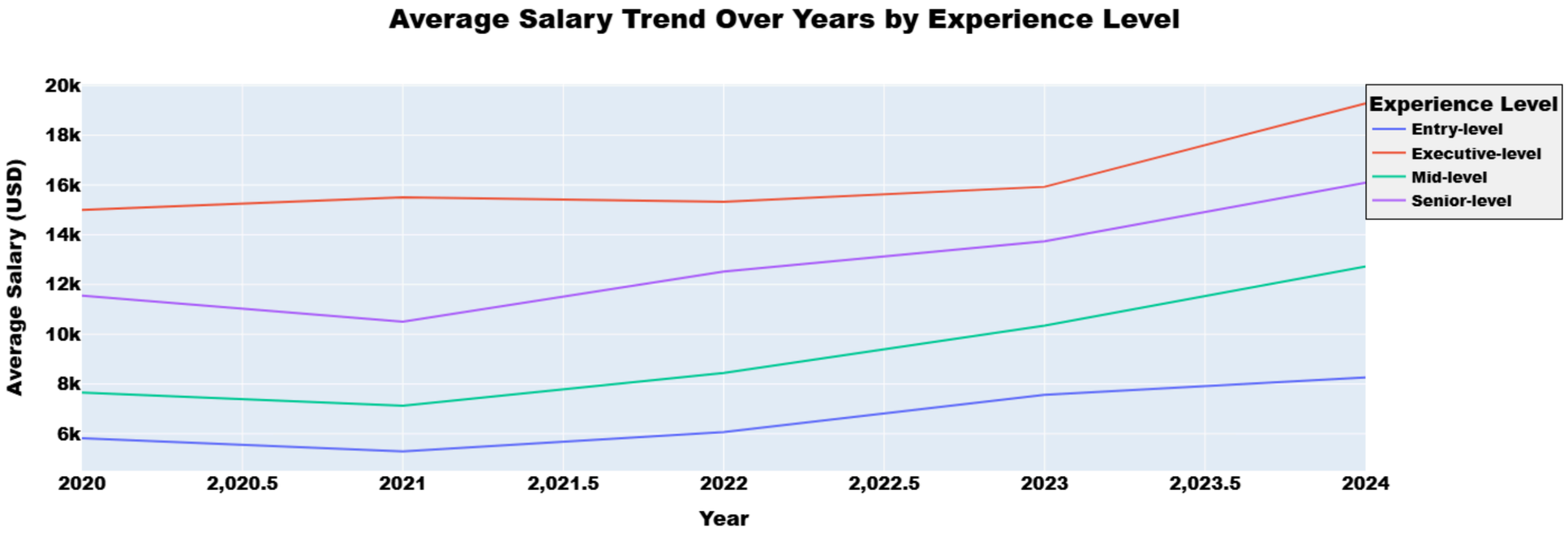
EXPLORATORY DATA ANALYSIS (EDA)

Most salaries concentrated in 5K-20K rangePeak around 10K-15K (~5000+ employees)Gradual decline with higher salariesSalaries above 30K very rare



EXPLORATORY DATA ANALYSIS (EDA)

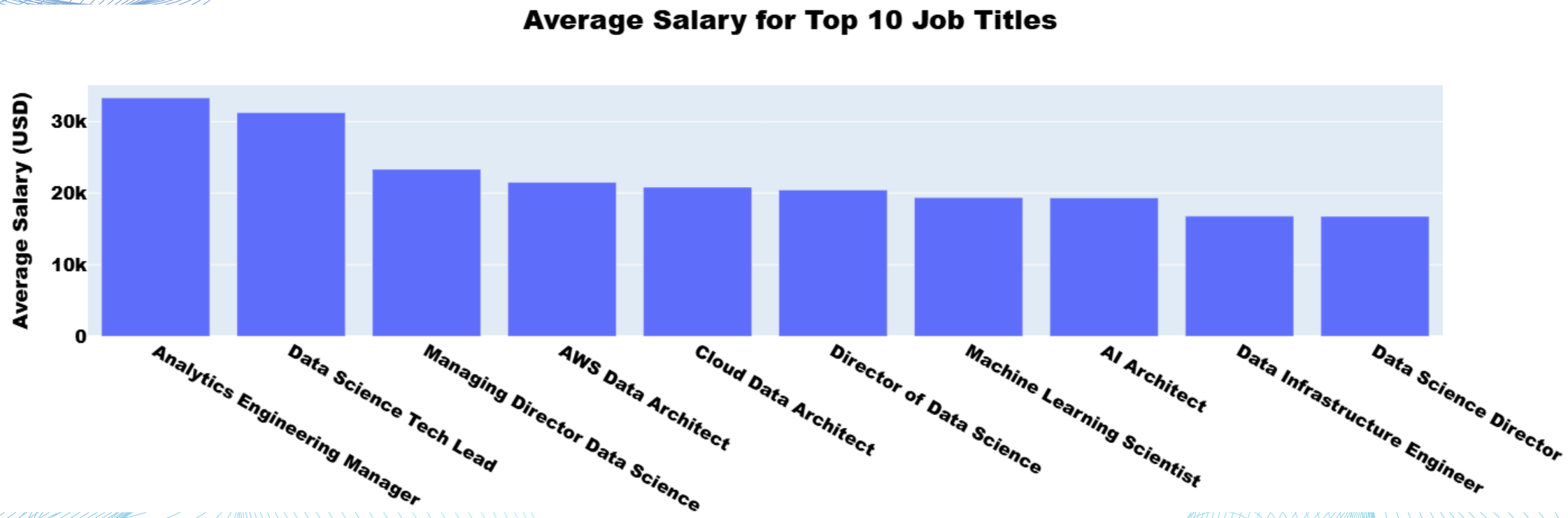
Executive-level (red): Highest salaries (~16-20K) with steady growthMid-level (purple): Second highest (~12-16K) moderate growthEntry-level (blue): Lower salaries (~8-10K) slow growthSenior-level (green): Lowest shown (~6-12K) gradual increase



EXPLORATORY DATA ANALYSIS (EDA)

Analytics Engineering Manager (~35K) - highest

Data Science Tech Lead (~25K) Other roles range 18-22K:AWS Engineer, Cloud Data Engineer, Director of Data Science



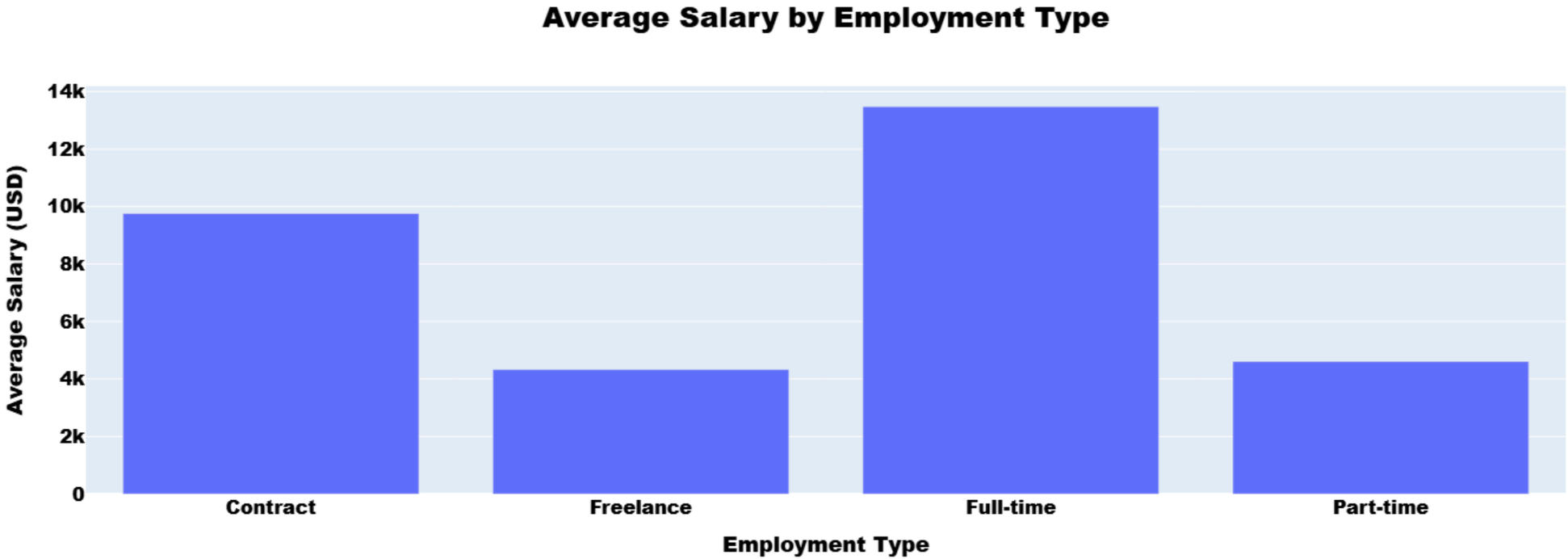
EXPLORATORY DATA ANALYSIS (EDA)

Full-time: ~14K (highest)

Contract: ~10K

Part-time: ~4.5K

Freelance: ~4K (lowest)

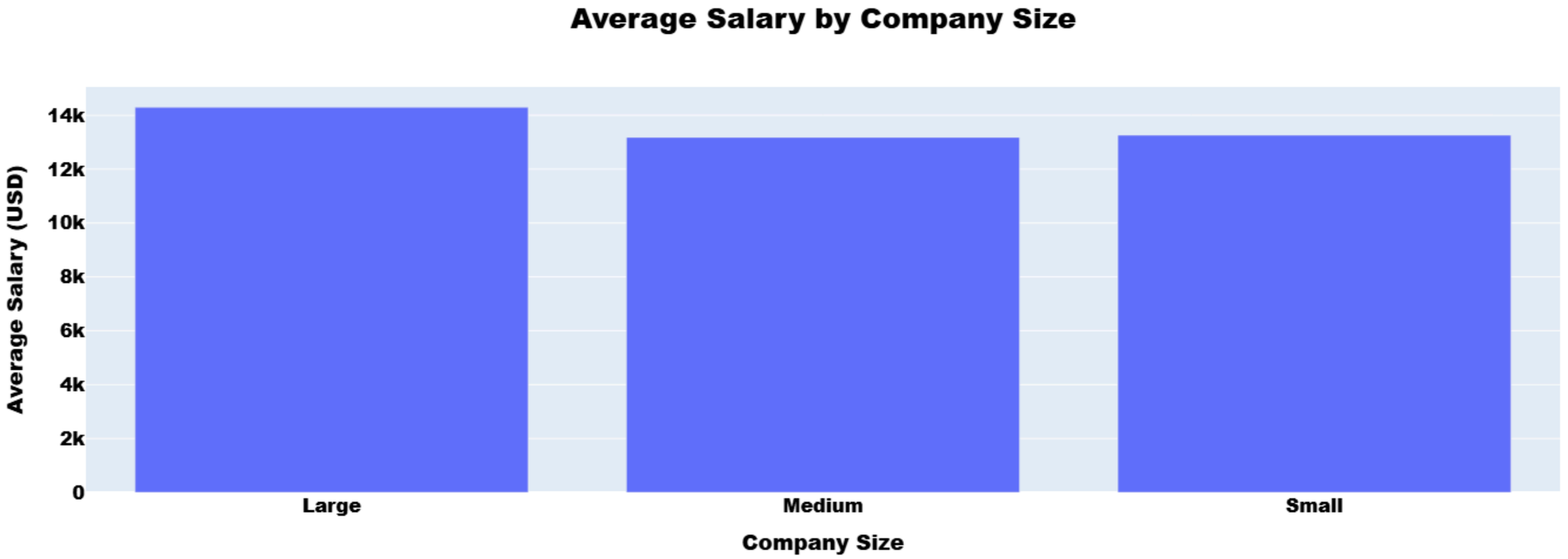


EXPLORATORY DATA ANALYSIS (EDA)

Large companies: ~14K (slightly highest)

Medium companies: ~13K

Small companies: ~13K



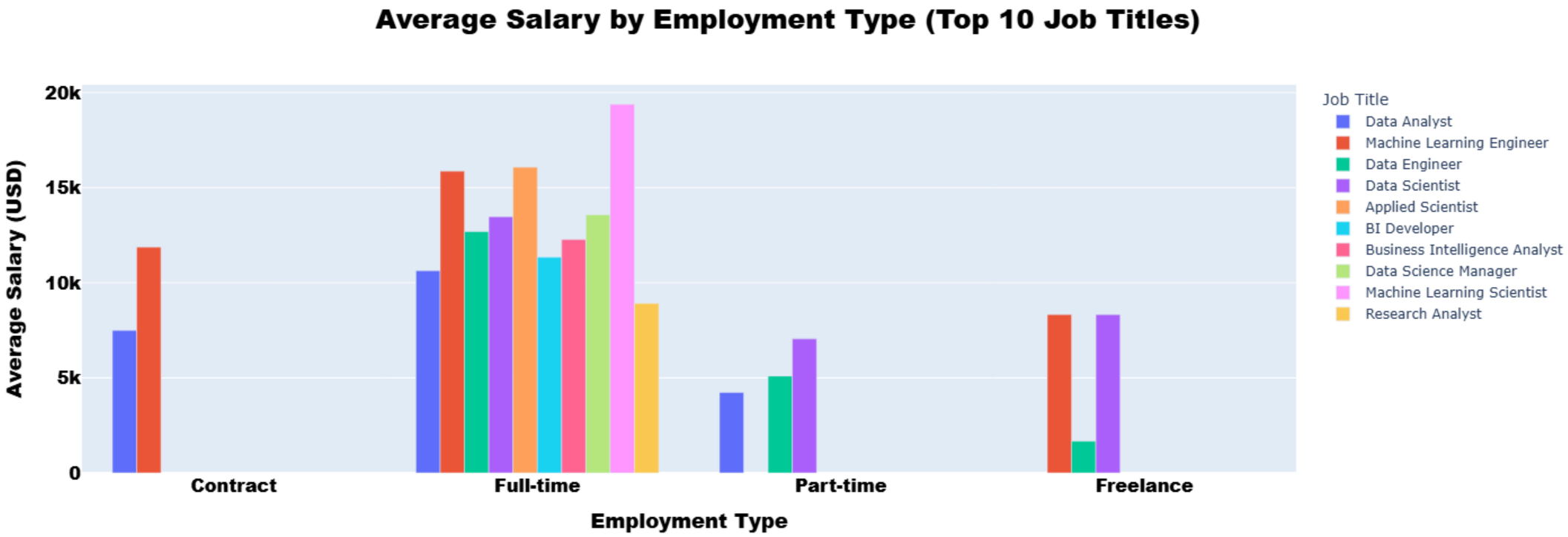
EXPLORATORY DATA ANALYSIS (EDA)

Full-time: Highest & most varied salaries (~10-20K)

Contract: Moderate pay (~5-13K)

Part-time: Lower salaries (~3-7K)

Freelance: Lowest (~8-9K only)



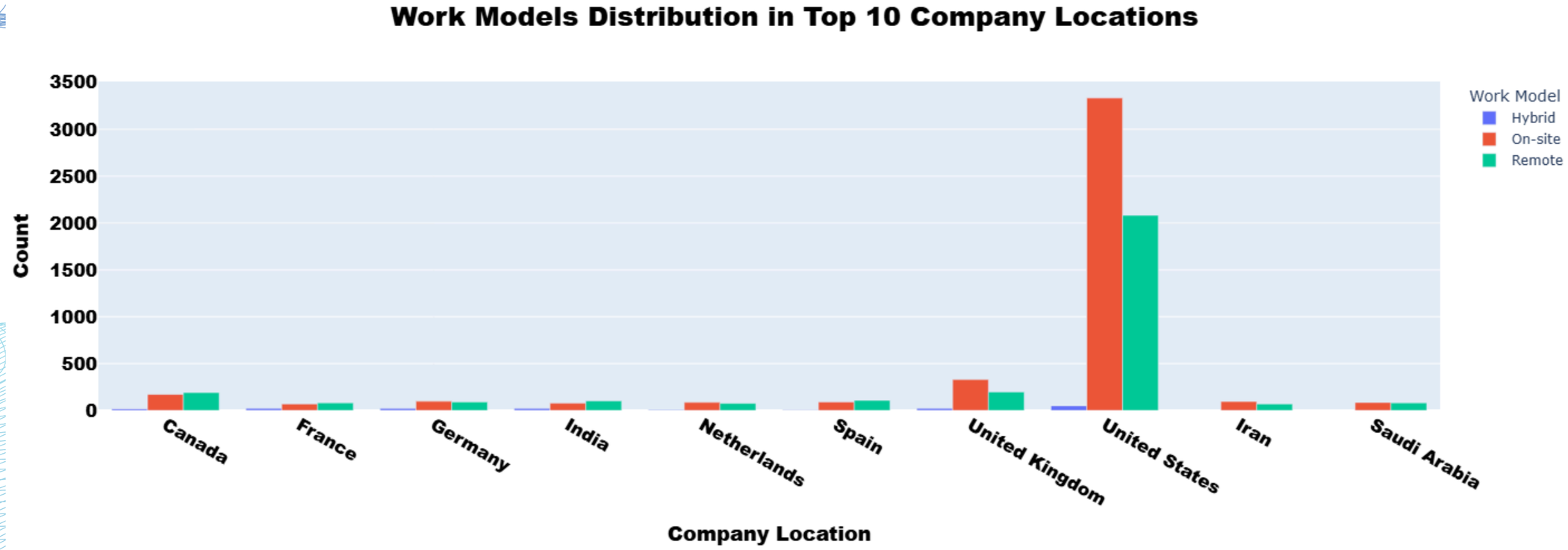
EXPLORATORY DATA ANALYSIS (EDA)

On-site (red): ~3200 companies

Remote (green): ~2000 companies

Hybrid (blue): very minimal

Other countries: UK ~300-400



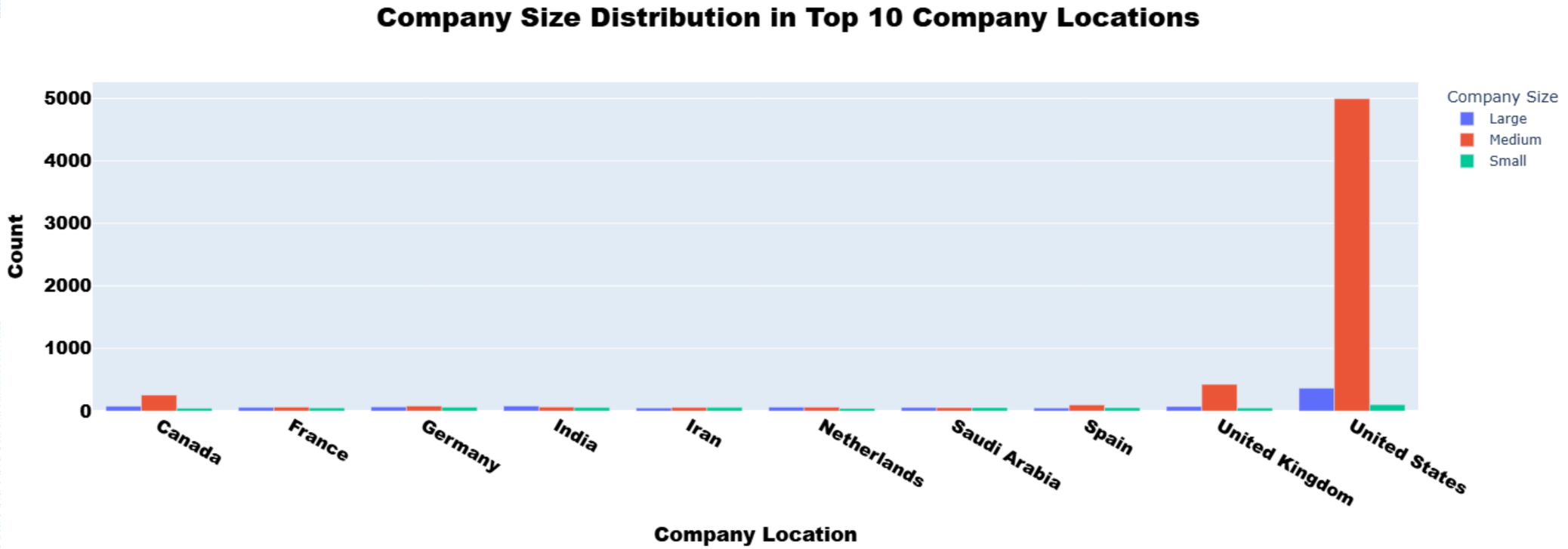
EXPLORATORY DATA ANALYSIS (EDA)

Large companies (red): ~5000

Medium (blue): ~500

Small (green): ~3000

ther countries: UK ~600 total

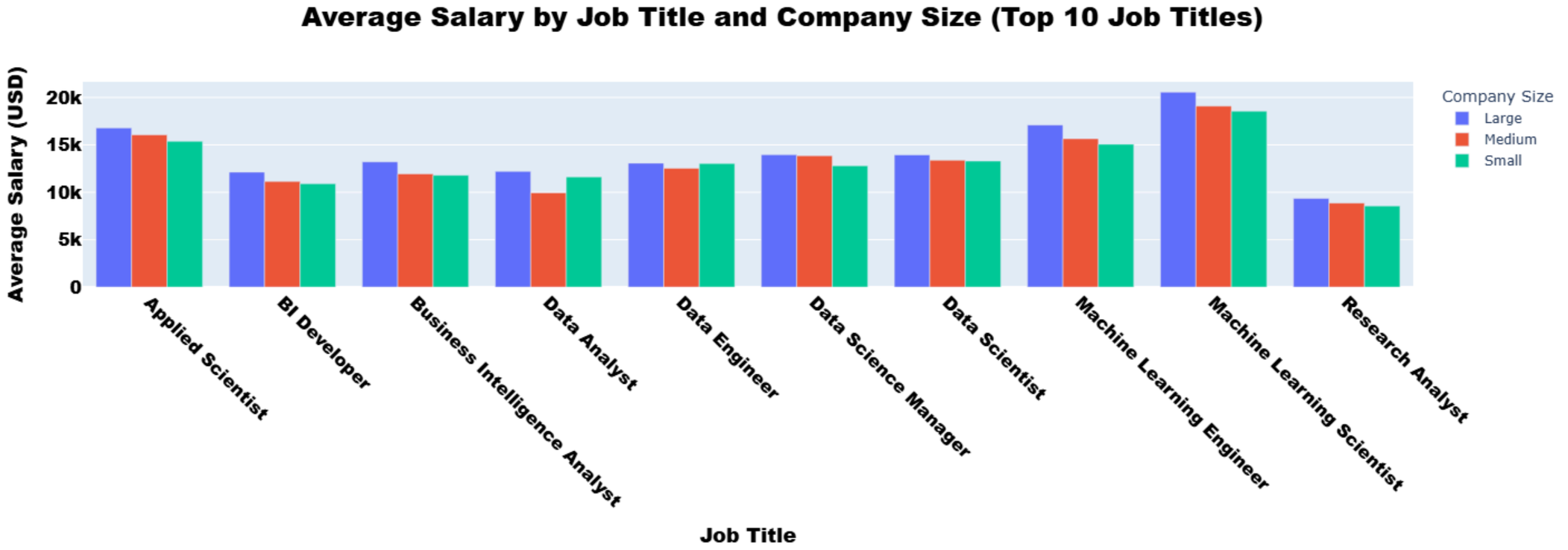


EXPLORATORY DATA ANALYSIS (EDA)

Analytics Engineering Manager: ~18K (similar across all company sizes)

Data Science Tech Lead: ~ 15K

Key insights: Minimal salary differences between company sizes for most roles



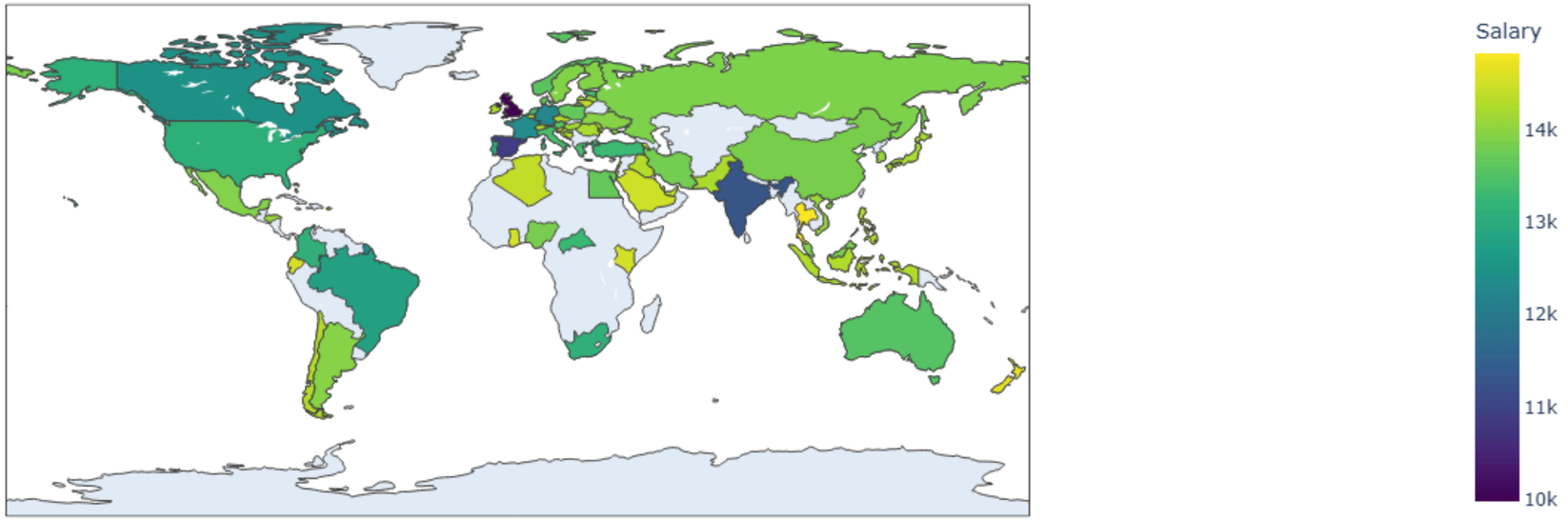
EXPLORATORY DATA ANALYSIS (EDA)

Light yellow (14K+): Highest - US leads

Light green (12-13K): High-medium - Canada, parts of Europe

Medium green (10-11K): Medium - Western Europe, AustraliaDark green (8-9K): Low-medium - Some Asian/South American countries

Dark blue (<6K): Lowest - Parts of Asia/Africa

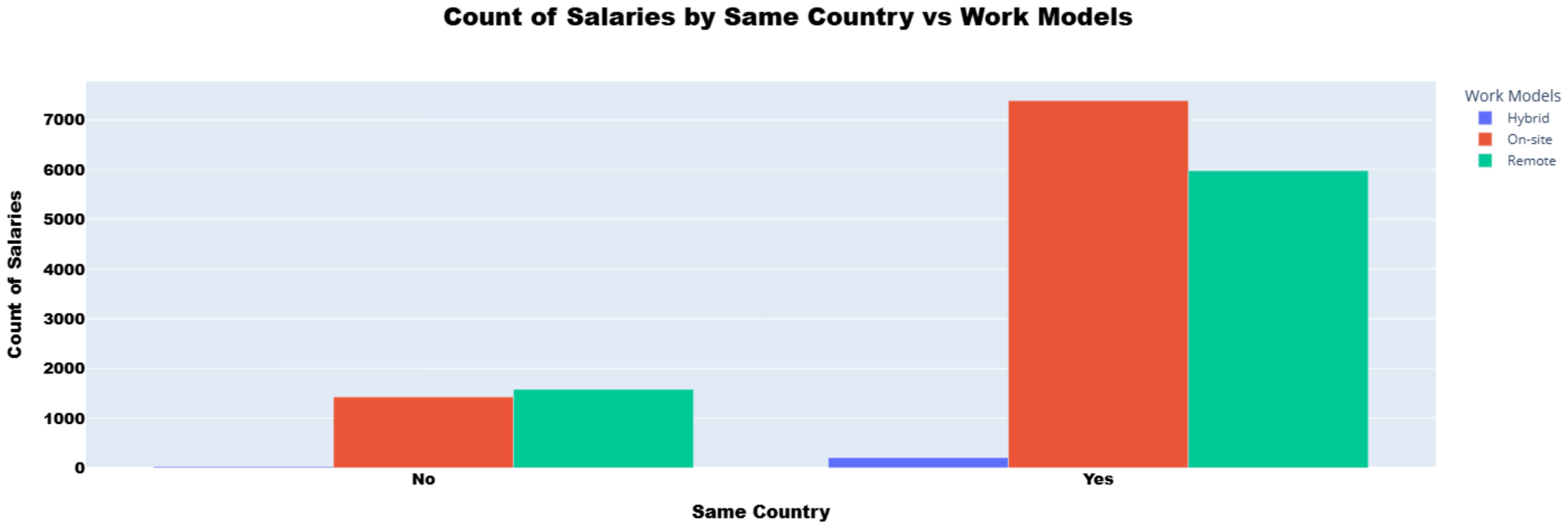


EXPLORATORY DATA ANALYSIS (EDA)

Different countries (No): ~1,500 each for on-site/remote

Same country (Yes): ~7,000 on-site, ~6,000 remote

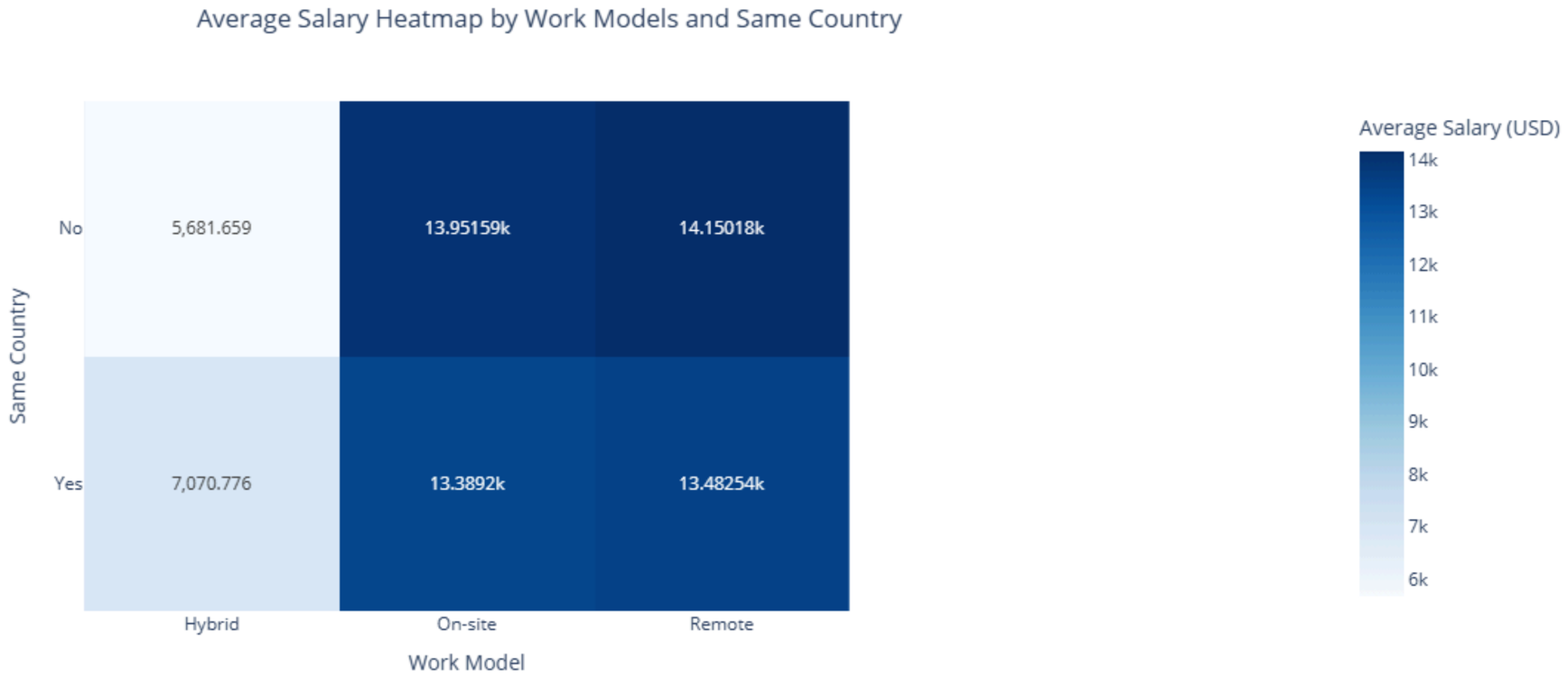
Hybrid minimal in both cases



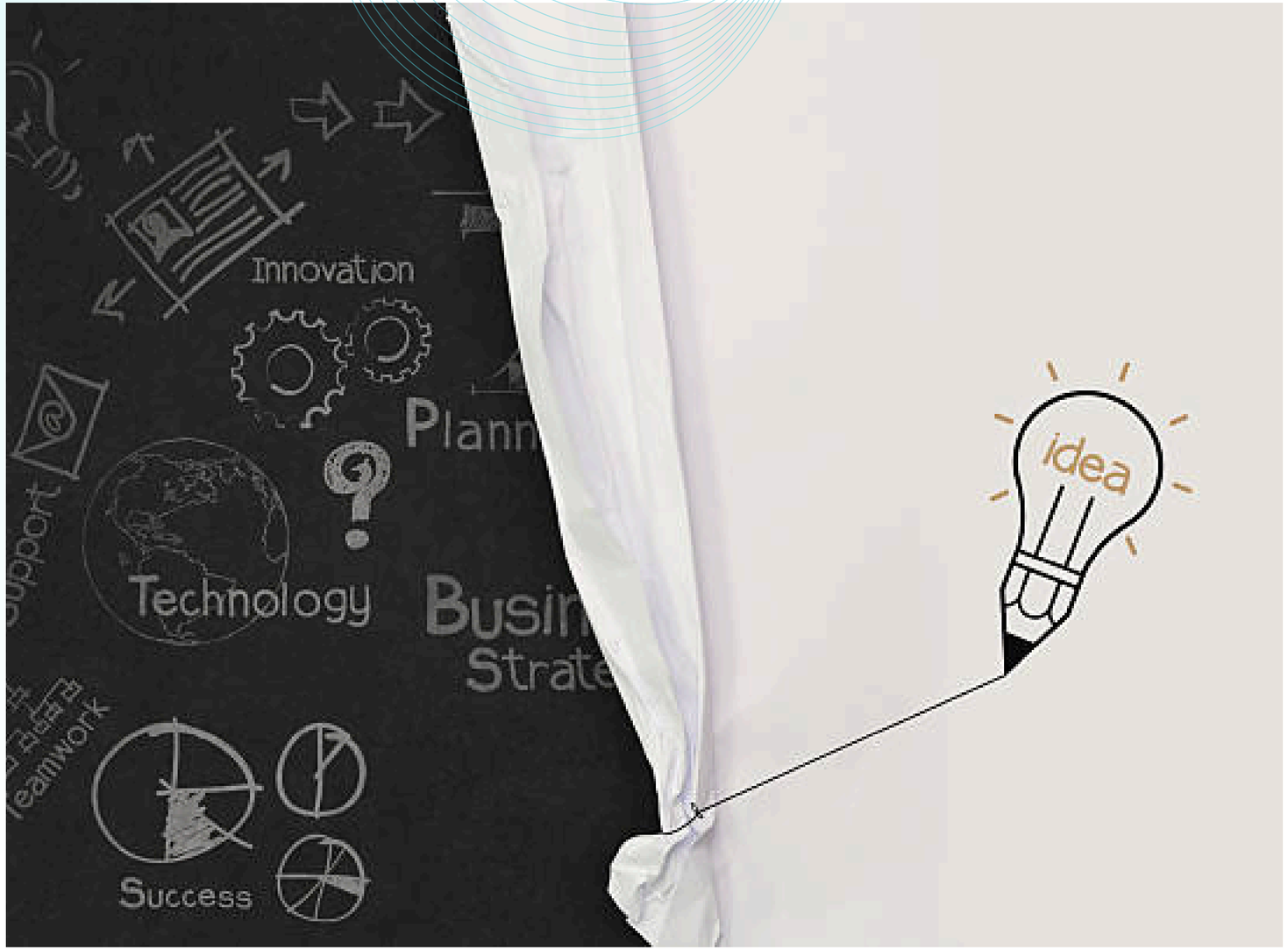
EXPLORATORY DATA ANALYSIS (EDA)

Different cultures (No): Remote ~6K, Hybrid/On-site ~14K

Similar cultures (Yes): Remote ~8K, Hybrid/On-site ~13K



CONFUSION MATRIX



FEATURE ENGINEERING

Columns Names: experience_level, employment_type, work_models, company_size

Encoding Used: Label Encoding

Why? These columns have low cardinality (few unique values).

Benefit: Simple numeric representation that preserves categories without creating too many features.

Columns Names: job_title, employee_residence, company_location

Encoding Used: Target Encoding

Why? These columns have high cardinality (many unique values),

Benefit: Each category is replaced with the mean of the target variable, making the model capture the relationship with the target efficiently.

MODELING BUILDING & EVALUATION

MODELING BUILDING & EVALUATION

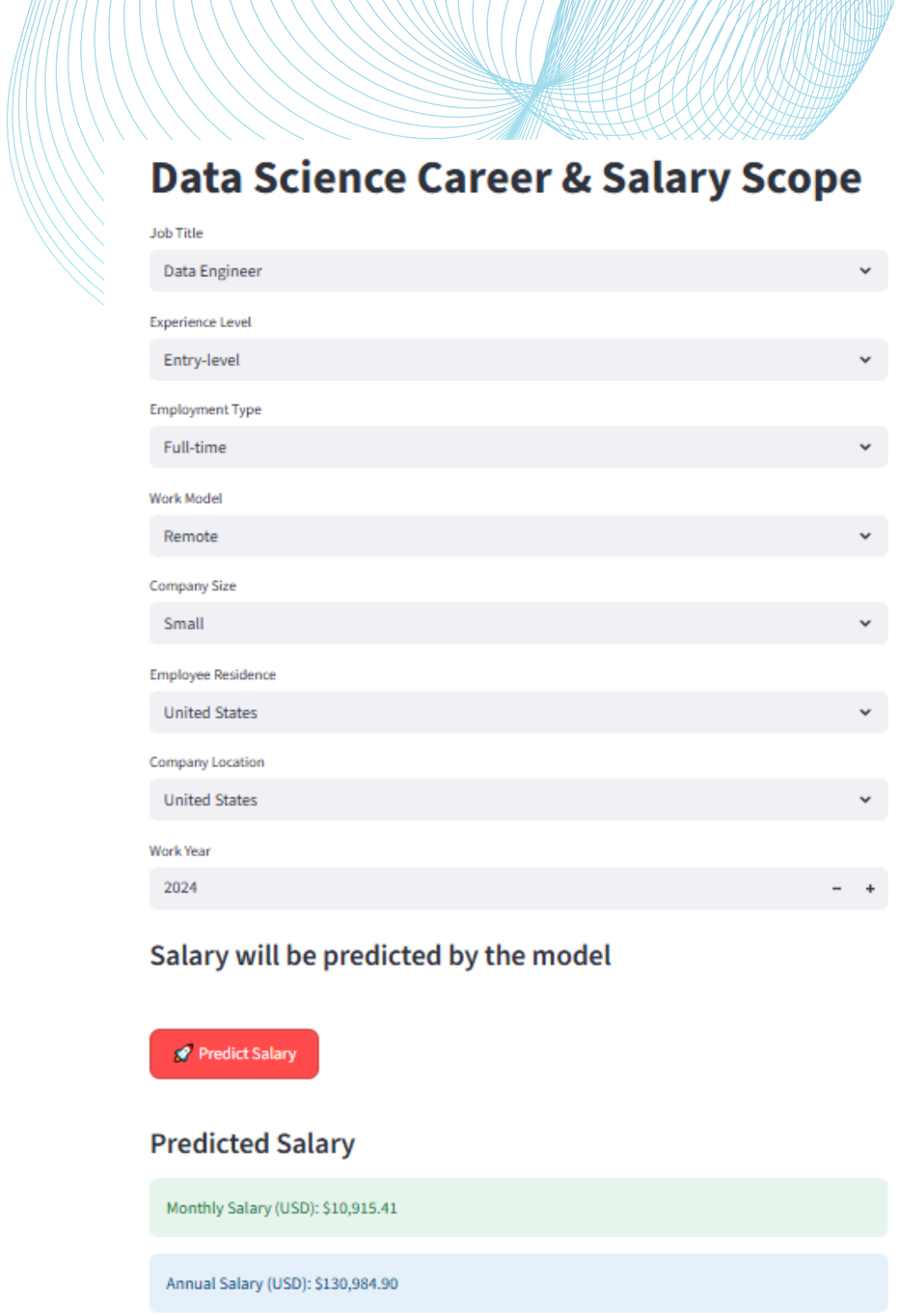
We experimented with a variety of machine learning models, including Linear Regression, KNN, Decision Tree, Random Forest, LightGBM, and XGBoost, in addition to several Artificial Neural Networks (ANN) with different architectures.

Among all, XGBoost Regressor delivered the best performance, achieving:

- **Low error values (MAE ≈ 1.6K, MAPE ≈ 15%)**
- **Reasonable accuracy with $R^2 \approx 0.66$**

This confirms that XGBoost was the most reliable model for predicting salary trends compared to both traditional ML algorithms and neural networks.

DEPLOYMENT


Data Science Career & Salary Scope
Job Title
Data Engineer
Experience Level
Entry-level
Employment Type
Full-time
Work Model
Remote
Company Size
Small
Employee Residence
United States
Company Location
United States
Work Year
2024
Salary will be predicted by the model
Predict Salary
Predicted Salary
Monthly Salary (USD): $10,915.41
Annual Salary (USD): $130,984.90

DEPLOYMENT

We developed an interactive Streamlit web app that predicts salaries based on job title, experience, work model, and company details and other Features.

This tool allows users to explore career and salary trends in data science dynamically.

This deployment not only makes our model accessible but also sets the foundation for scaling it into a professional career guidance platform.

OUR TEAM

Ahmed Hassan

Ahmed Fathy

Nada Mohamed

Ahmed Hassan

Ahmed Fathy

Nada Mohamed

THANKS