

# JOBs AND SALARIES IN DATA-RELATED CAREERS (2024)

December 2024

---

DUY DAT NGUYEN

# TABLE OF CONTENT

---

- 01 Purpose of The Analysis
- 02 Data Organization
- 03 Data Cleaning
- 04 Data Analysis and Visualization
- 05 Summary



# PURPOSE OF THE ANALYSIS

# PURPOSE OF THE ANALYSIS

The exponential growth of data has driven a significant demand for data science professionals, while simultaneously creating substantial fluctuations in salary structures. In a context where information about salaries in this field is limited and often outdated, this project focuses on examining the evolution of data science salaries from 2020 to 2024, it places a particular focus on trends observed in 2024.

This project not only helps professionals better understand the labor market landscape but also provides practical data to make informed career and recruitment decisions. This analysis aims to offer a comprehensive and realistic view of salary trends, thereby supporting the sustainable growth of the data science industry.



PERSONAL PROJECT

---

# DATA ORGANIZATION

---

# DATA ORGANIZATION

 PERSONAL PROJECT

01

## Data Source

The data is located at *ai-jobs.net* and entitled "**The Global AI, ML, Data Science Salary Index for 2024** 💰📊'" which is based on internal data survey submissions and jobs with open salaries. The dataset is updated every week. You can contribute to them by submitting your salary info. The project collects salary information anonymously from professionals and employers all over the world and make it publicly available for anyone to use, share, and play around with.

# DATA ORGANIZATION

 PERSONAL PROJECT

## 02 Data Description

**work\_year**

The year the salary was paid.

**experience\_level**

The experience level in the job during the year with the following possible values:

**EN** Entry-level / Junior

**MI** Mid-level / Intermediate

**SE** Senior-level / Expert

**EX** Executive-level / Director

**employment\_type**

The type of employment for the role:

**PT** Part-time

**FT** Full-time

**CT** Contract

**FL** Freelance

# DATA ORGANIZATION

 PERSONAL PROJECT

## 02 Data Description

**job\_title**

The role worked in during the year.

**salary**

The total gross salary amount paid.

**salary\_currency**

The currency of the salary paid as an ISO 4217 currency code.

**salary\_in\_usd**

The salary in USD (FX rate divided by avg. USD rate of the respective year) via statistical data from the [BIS and central banks](#).

**employee\_residence**

Employee's primary country of residence in during the work year as an ISO 3166 country code.

**remote\_ratio**

The overall amount of work done remotely, possible values are as follows:

**0** No remote work (less than 20%)

**50** Partially remote/hybrid

**100** Fully remote (more than 80%)

# DATA ORGANIZATION

 PERSONAL PROJECT

## 02 Data Description

**company\_location** The country of the employer's main office or contracting branch as an ISO 3166 country code.

**company\_size** The average number of people that worked for the company during the year:

**S** less than 50 employees (small)

**M** 50 to 250 employees (medium)

**L** more than 250 employees (large)

# DATA ORGANIZATION

PERSONAL PROJECT

03

## Data Overview

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70796 entries, 0 to 70795
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   work_year        70796 non-null   int64  
 1   experience_level 70796 non-null   object  
 2   employment_type   70796 non-null   object  
 3   job_title         70796 non-null   object  
 4   salary            70796 non-null   int64  
 5   salary_currency   70796 non-null   object  
 6   salary_in_usd    70796 non-null   int64  
 7   employee_residence 70796 non-null   object  
 8   remote_ratio      70796 non-null   int64  
 9   company_location  70796 non-null   object  
 10  company_size      70796 non-null   object  
dtypes: int64(4), object(7)
memory usage: 5.9+ MB
```



PERSONAL PROJECT

---

# DATA CLEANING

---

# DATA CLEANING

01

Summarize counts, nulls, unique, duplicates, and percentage of duplicated rows

	Count	Null	Null %	Unique	Duplicated
work_year	70796	0	0.0	5	37988
experience_level	70796	0	0.0	4	37988
employment_type	70796	0	0.0	4	37988
job_title	70796	0	0.0	286	37988
salary	70796	0	0.0	6904	37988
salary_currency	70796	0	0.0	25	37988
salary_in_usd	70796	0	0.0	7483	37988
employee_residence	70796	0	0.0	93	37988
remote_ratio	70796	0	0.0	3	37988
company_location	70796	0	0.0	86	37988
company_size	70796	0	0.0	3	37988

- Number of duplicated rows: **37988**
- Percentage of duplicated rows: **53.7**

**Note:** The data set contains **37988** duplicates, representing **53.7%** of the data set. I have chosen to retain these duplicates, as they appear to represent different entries from the survey with identical responses. This is plausible, as specific positions/roles may have identical work settings and salaries within a given country.

# DATA CLEANING

02

## Categorization and Labeling

This section focuses on enhancing the clarity and usability of the dataset by creating and applying more informative labels. This involves:

- **Generating job category labels** by mapping job titles to broader categories.
- **Generating work model labels** by mapping remote ratios to descriptive work models.
- **Renaming and mapping labels** for experience levels, employment types, company sizes, company locations, and employee residences to be more descriptive.

These steps will facilitate more effective analysis and interpretation of the data by organizing it into logical groupings.

# DATA CLEANING

02

## Categorization and Labeling

	work_year	experience_level	employment_type	job_title	salary	salary_currency
0	2024	Mid-Level	Contract	Data Management Specialist	99000	USD
1	2024	Mid-Level	Contract	Data Management Specialist	81000	USD
2	2024	Junior	Full-Time	Data Analyst	85000	USD
3	2024	Junior	Full-Time	Data Analyst	45000	USD
4	2024	Junior	Full-Time	Data Analyst	72467	USD
...	...	...	...	...	...	...
70791	2020	Senior	Full-Time	Data Scientist	412000	USD
70792	2021	Mid-Level	Full-Time	Principal Data Scientist	151000	USD
70793	2020	Junior	Full-Time	Data Scientist	105000	USD
70794	2020	Junior	Contract	Business Data Analyst	100000	USD
70795	2021	Senior	Full-Time	Data Scientist	7000000	INR

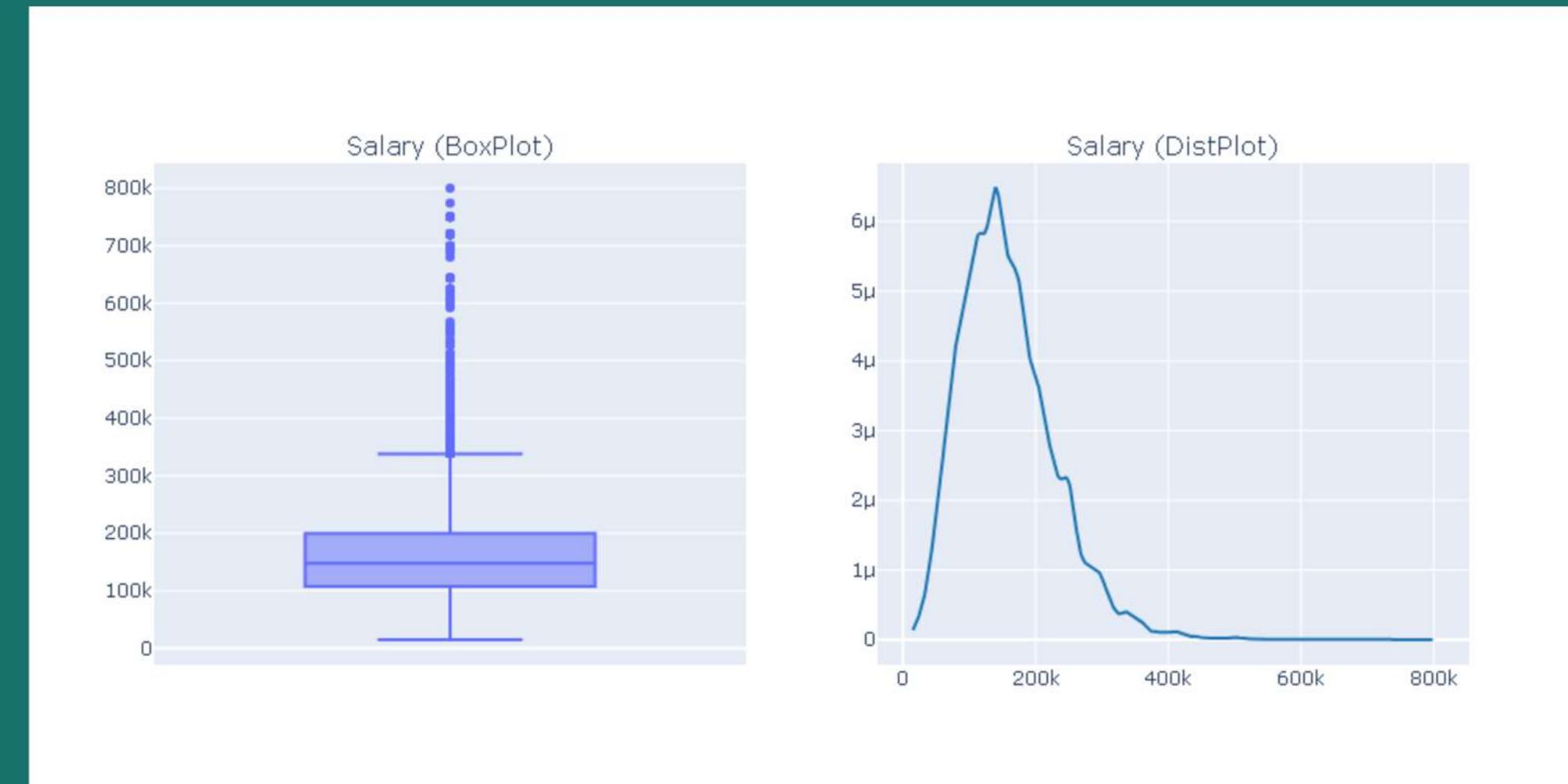
	salary_in_usd	employee_residence	remote_ratio	company_location	company_size	job_category	work_model
99000	United States	0	United States	Medium	Data Scientist	Onsite	
81000	United States	0	United States	Medium	Data Scientist	Onsite	
85000	United States	0	United States	Medium	Data Analysis	Onsite	
45000	United States	0	United States	Medium	Data Analysis	Onsite	
72467	United States	0	United States	Medium	Data Analysis	Onsite	
...	...	...	...	...	...	...	...
412000	United States	100	United States	Large	Data Scientist	Remote	
151000	United States	100	United States	Large	Data Scientist	Remote	
105000	United States	100	United States	Small	Data Scientist	Remote	
100000	United States	100	United States	Large	Data Analysis	Remote	
94665	India	50	India	Large	Data Scientist	Hybrid	

# DATA CLEANING

03

## Detecting and handing outliers

We will visualize salary data using box plots and distribution plot.



The box plot and distribution plot both highlight a right-skewed salary distribution. The box plot shows the concentration of salaries within the IQR, with many outliers at higher values. The distribution plot confirms this skew, with a peak at lower salaries and a long tail indicating the presence of higher outliers.

# DATA CLEANING

03

## Detecting and handing outliers

We calculate Q1, Q2, Q2, and the IQR to determine the lower and upper bounds for outlier detection. Then, we identify the specific data points that fall outside these bounds.

```
Q1 = 107730.0, Q2 = 148000.0, Q3 = 200000.0
```

```
IQR = 92270.0
```

```
MIN = -30675.0, MAX = 338405.0
```

Number of outliers: 1327

We will remove rows containing outliers and then regenerate the plots.

# DATA CLEANING

03

## Detecting and handing outliers



The new plots indicate the positive impact of the outlier removal process. These showcase a more representative and cleaner distribution of salaries, which is now more suitable for further analysis or model training. This distribution is still right skewed but to a far smaller extent.



PERSONAL PROJECT

---

# DATA ANALYSIS AND VISUALIZATION

---

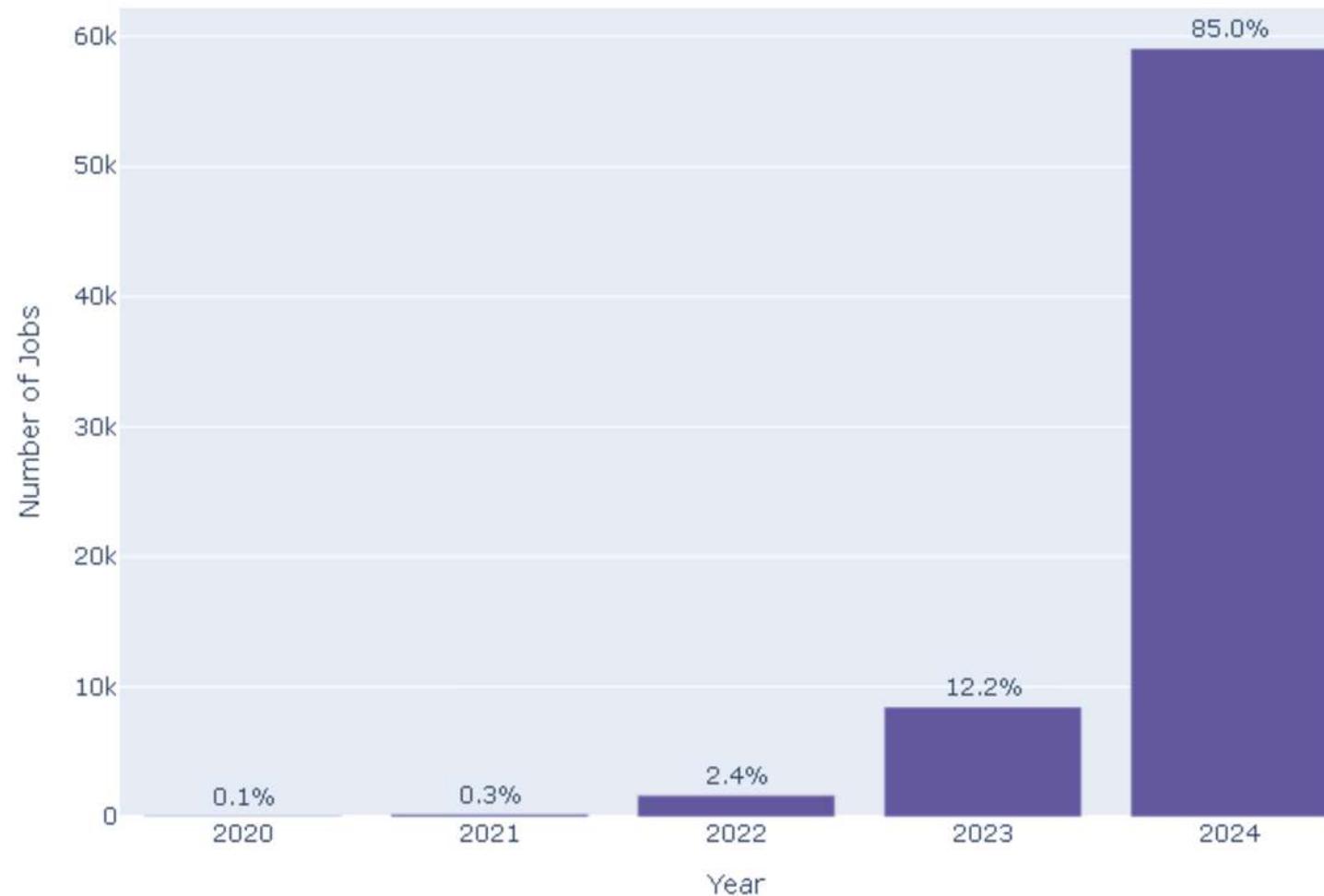
# DATA ANALYSIS AND VISUALIZATION

01

## Data Overview and Analysis

- Analyzing *work\_year*:

Job Postings Over Years



### Insights:

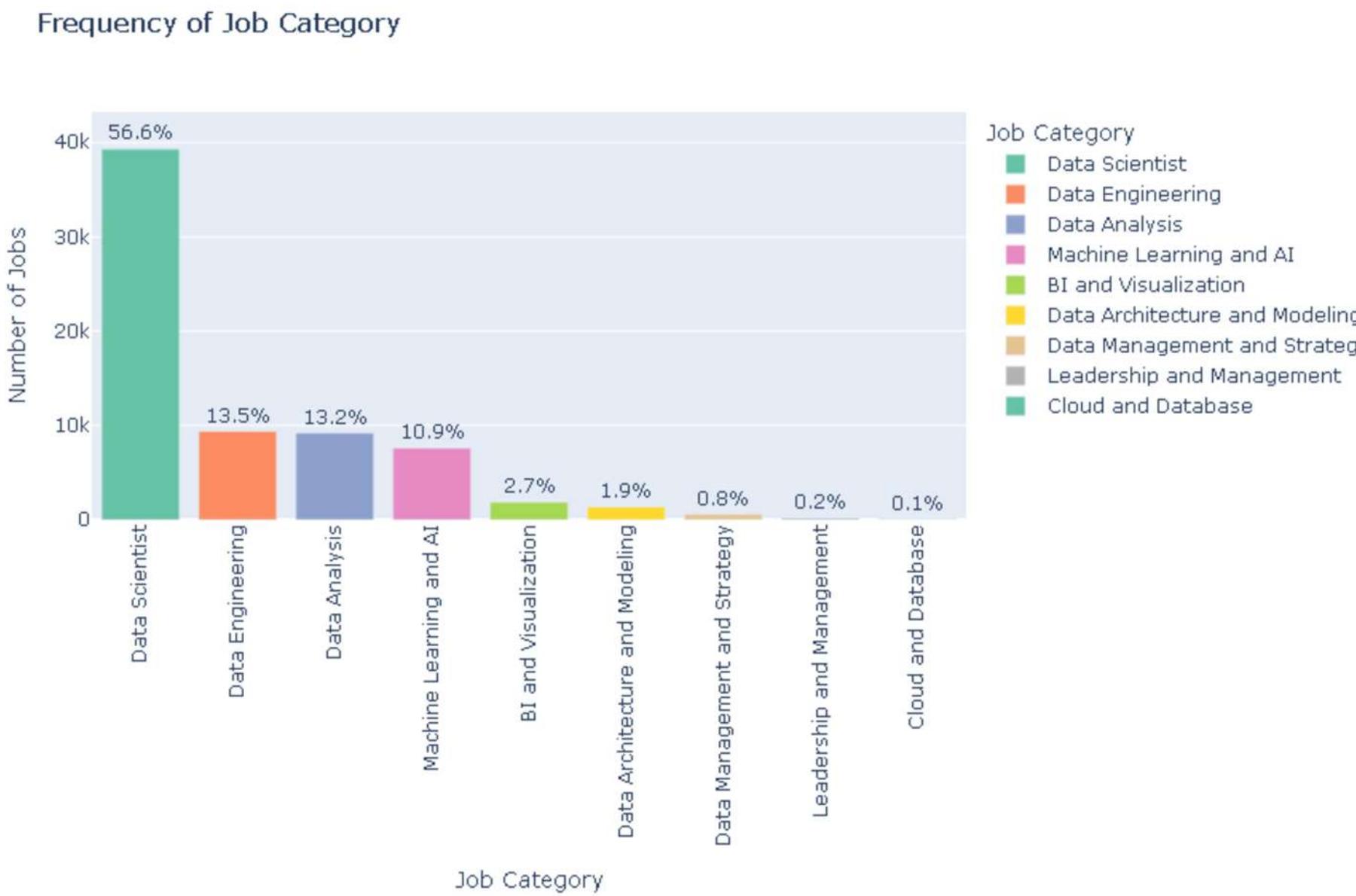
- There will be a boom in demand in 2024, with 85 percent of job postings in data science, showing strong investments by businesses in data-driven solutions.
- Career opportunities are abundant in 2024 with high demand for skills in Machine Learning, AI, and Data Engineering.
- Industry trends show rapid growth in Data Science, especially in sectors like finance, technology, and healthcare.

# DATA ANALYSIS AND VISUALIZATION

01

## Data Overview and Analysis

- Analyzing *job\_category*:



### Insights:

- Data Scientist Lead:** Data Scientist positions are the most in-demand, comprising over half (**56.6%**) of all job postings, indicating a strong market focus.
- Secondary Demand:** Data Engineering (**13.5%**) and Data Analysis (**13.2%**) are the next most prevalent roles.
- Niche & Specialized Roles:** Other categories like Machine Learning/AI (**10.9%**) and BI/Visualization (**2.7%**) have a smaller but noteworthy presence, indicating specialization within Data Science. The rest (Data Architecture, Data Management, Leadership, and Cloud) have relatively little representation.
- Lower Frequency Roles:** The remaining categories ("BI and Visualization", "Data Architecture and Modeling", "Data Management and Strategy", "Leadership and Management", and "Cloud and Database") have substantially lower job counts, each with less than **3%** and represented by the shorter bars.

# DATA ANALYSIS AND VISUALIZATION

01

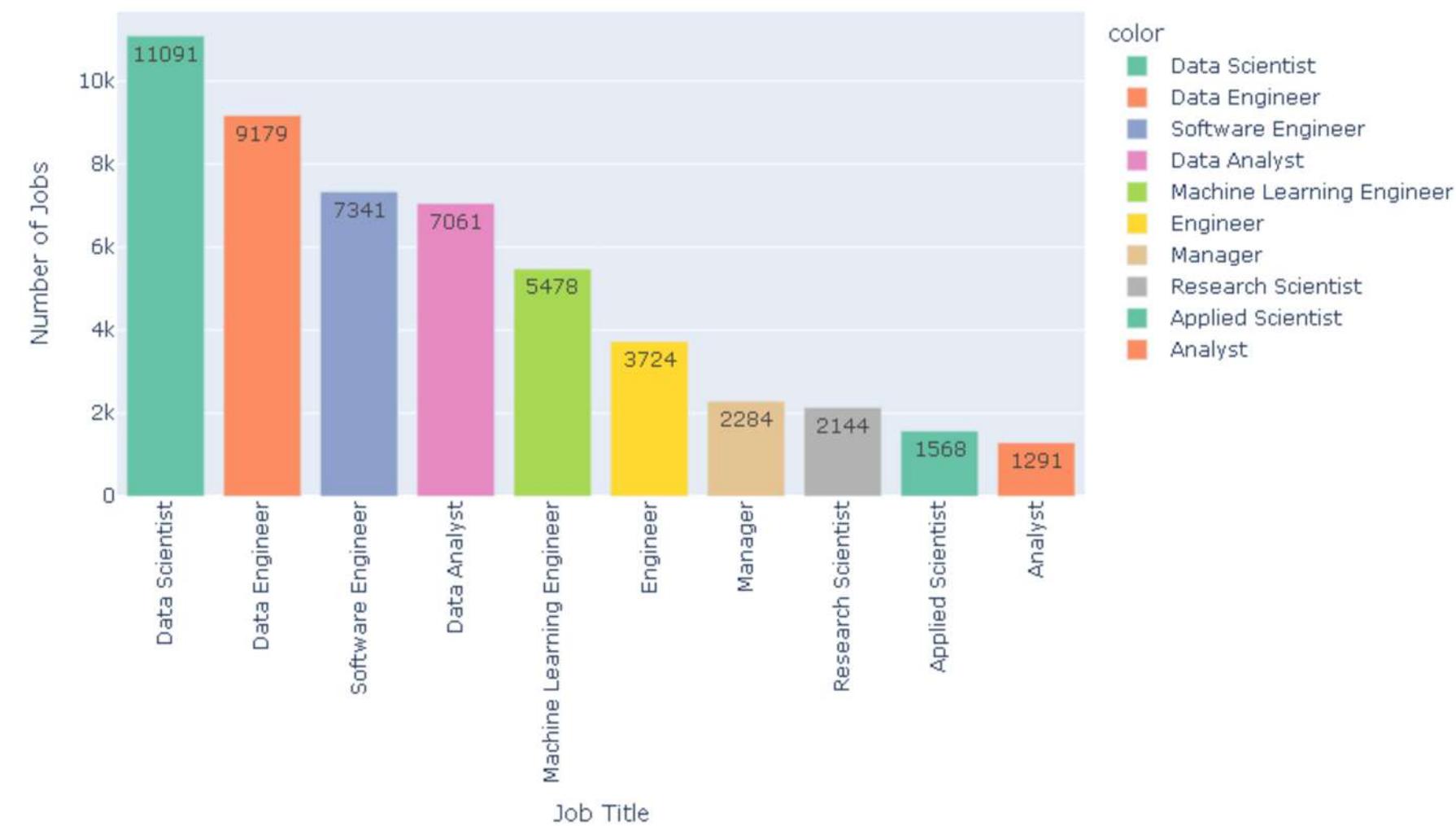
## Data Overview and Analysis

- Analyzing *job\_title*:

### Insights:

- Data Scientist Core:** Data Scientist is the most in-demand role, highlighting the need for strong analytical skills.
- Infrastructure Matters:** High demands for Data Engineers underscores the importance of data infrastructure.
- Tech and Analysis Balance:** Software & Data Analysts are crucial, showing a need for both implementation and interpretation.
- AI Growth:** Machine Learning Engineers are in high demand, reflecting the growth in AI.
- Top Tier Demand:** There is a clear demand divide between the top 5 jobs in the list and the rest.

Job Postings by Job Titles (Top 10)



# DATA ANALYSIS AND VISUALIZATION

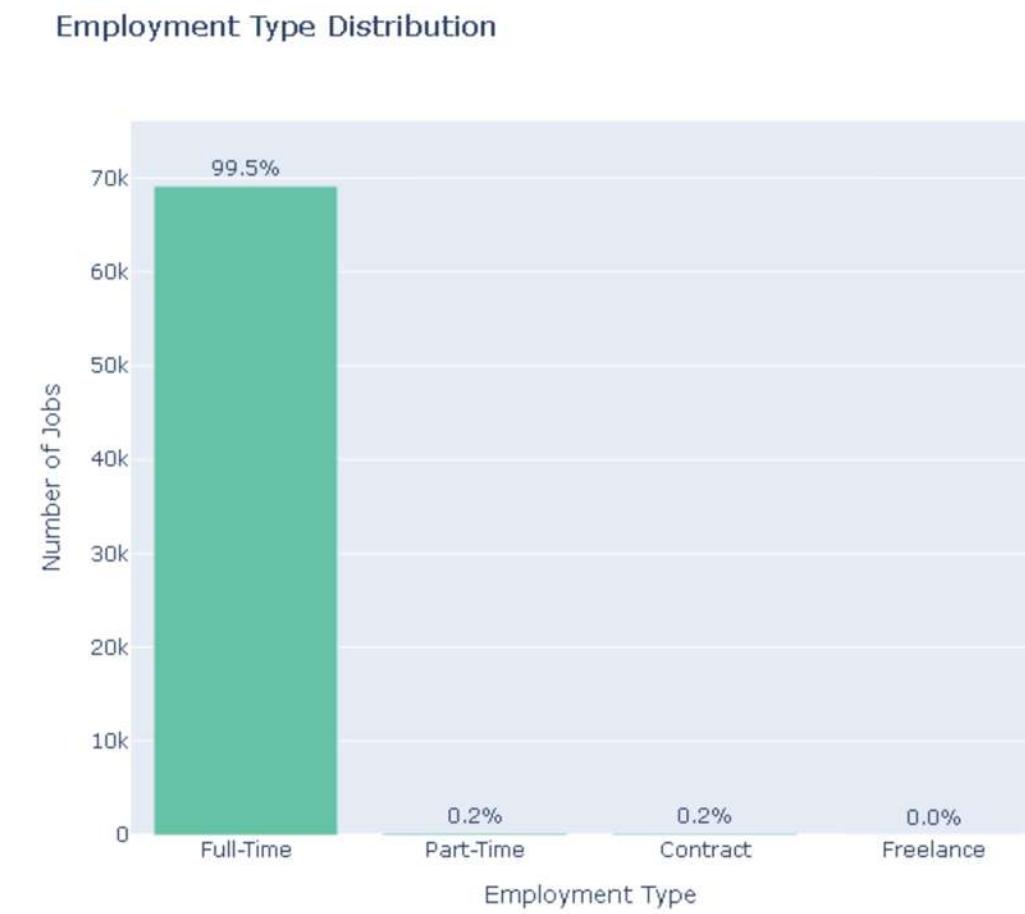
01

## Data Overview and Analysis

- Analyzing *experience\_level*:



- Analyzing *employment\_type*:



### Insights:

The data industry primarily hires for stable, full-time positions, with a strong emphasis on experienced professionals, particularly at the senior level. This structure reveals a clear career path for those moving from junior to mid-level and ultimately to senior roles, yet also indicates a competitive environment both for employers seeking talent and for data professionals pursuing career advancement.

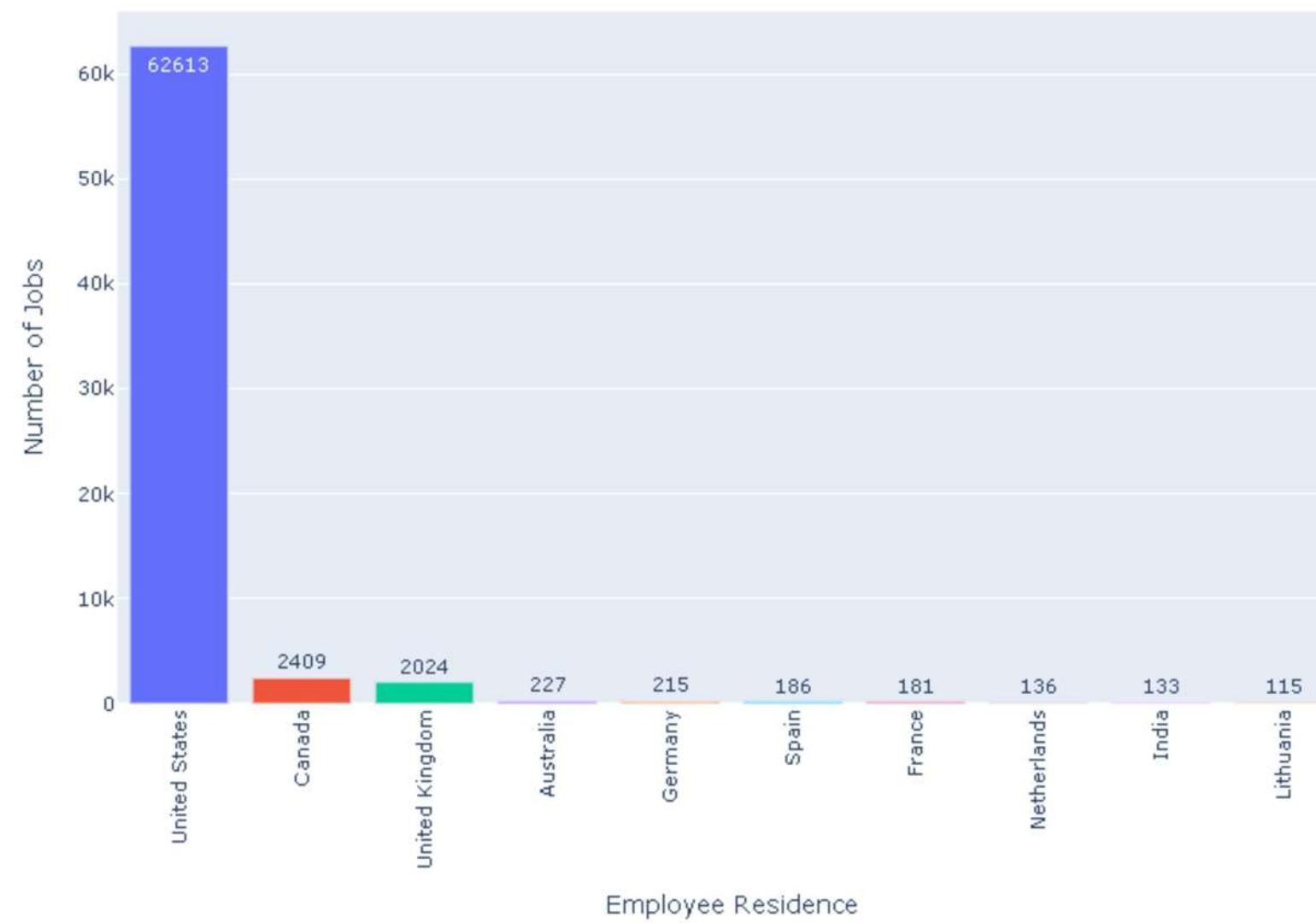
# DATA ANALYSIS AND VISUALIZATION

01

## Data Overview and Analysis

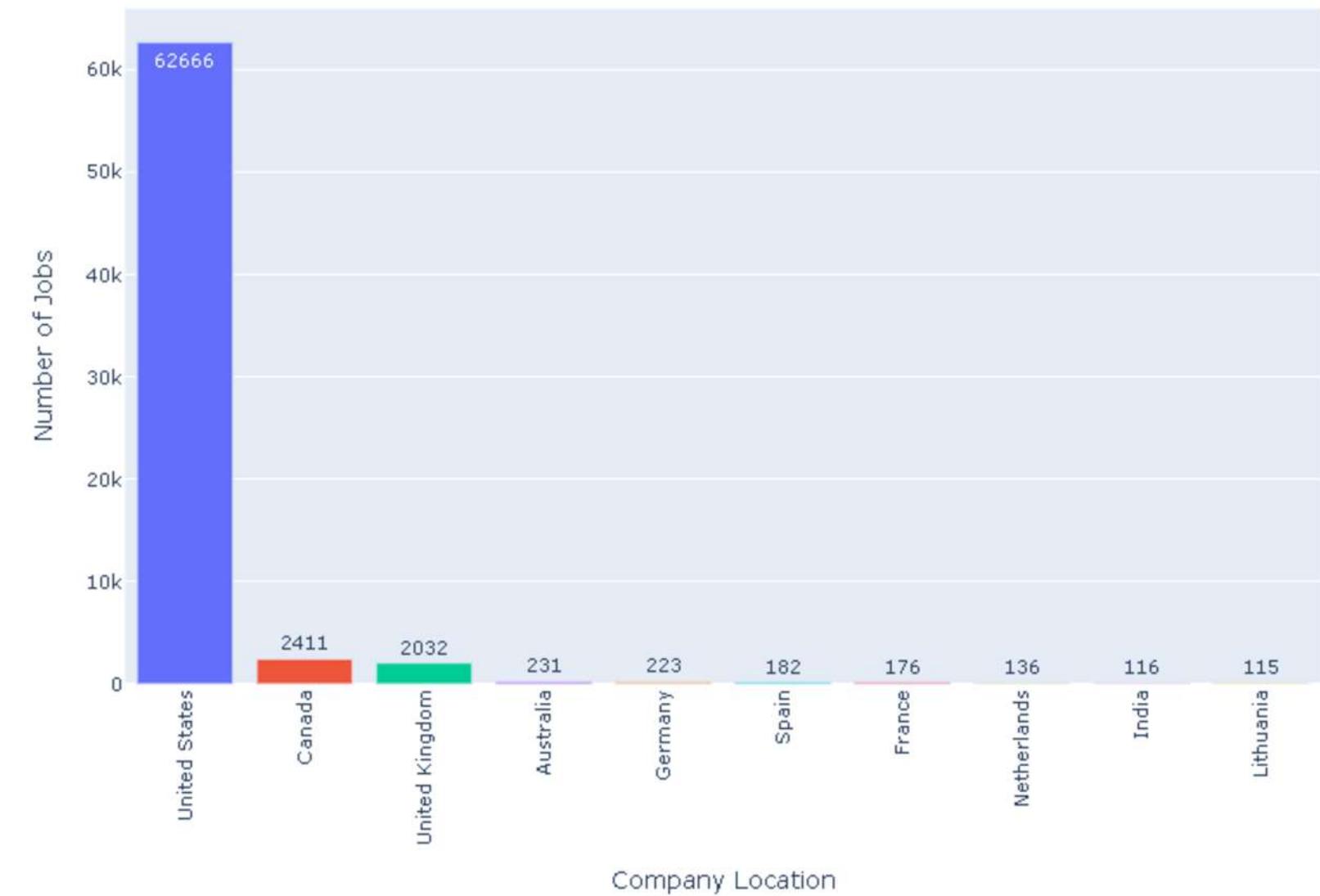
- Analyzing *employee\_residence*:

Job Postings by Employee Residence (Top 10)



- Analyzing *company\_location*:

Job Postings by Company Location (Top 10)



# DATA ANALYSIS AND VISUALIZATION



The US is the dominant hub for the data industry, with significantly more job postings and employees than any other country; secondary hubs include Canada and the UK.

## Overwhelming Dominance of the United States

- Both charts show the United States as the dominant location for both job postings and employee residence, with a significantly larger number of postings than all other countries combined.
- This suggests the US is the major hub for the Data industry. It's the primary driver of demand and likely experts the strongest influence on global salary benchmarks.

## Secondary Hubs: Canada and the United Kingdom

- Both Canada and the UK follow the US as secondary locations for job postings and employee residences but at a much lower volume.
- These represent significant, though smaller, hubs for the Data industry, potentially reflecting strong tech sectors or international business centers.

## Other European Countries

- Other European countries, such as Germany, Spain, France and the Netherlands, have moderate representation for both job postings and employee residence.
- This likely reflects the presence of growing technology sectors within these regions.

# DATA ANALYSIS AND VISUALIZATION

PERSONAL PROJECT

## Employees Working in Own Country



The chart shows that most employees do not seek job opportunities abroad. This could be due to domestic salaries meeting their living needs, as well as cultural factor, where individuals prefer to work close to their families and communities.

**99.8%**

**Working in Own Country**

**0.02%**

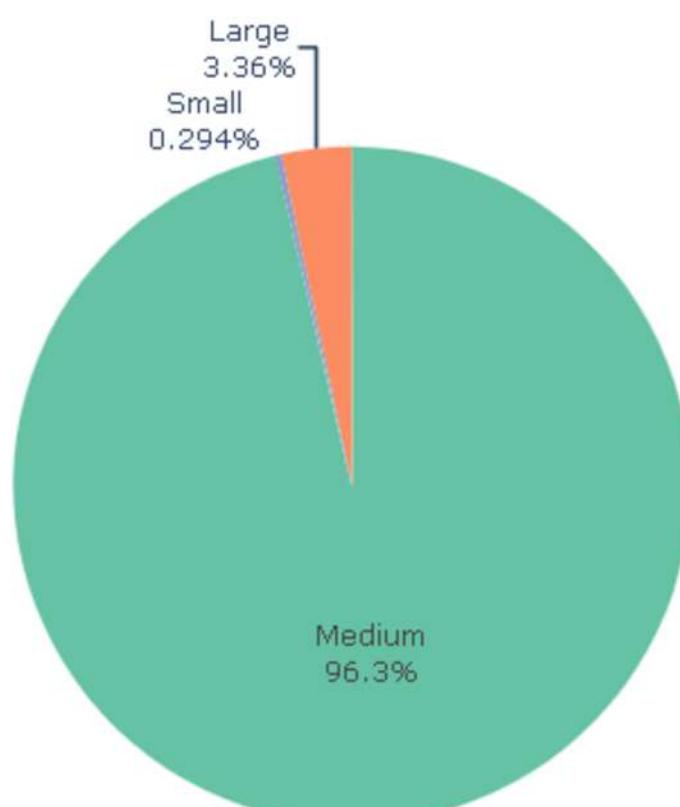
**Working Outside Own Country**

# DATA ANALYSIS AND VISUALIZATION

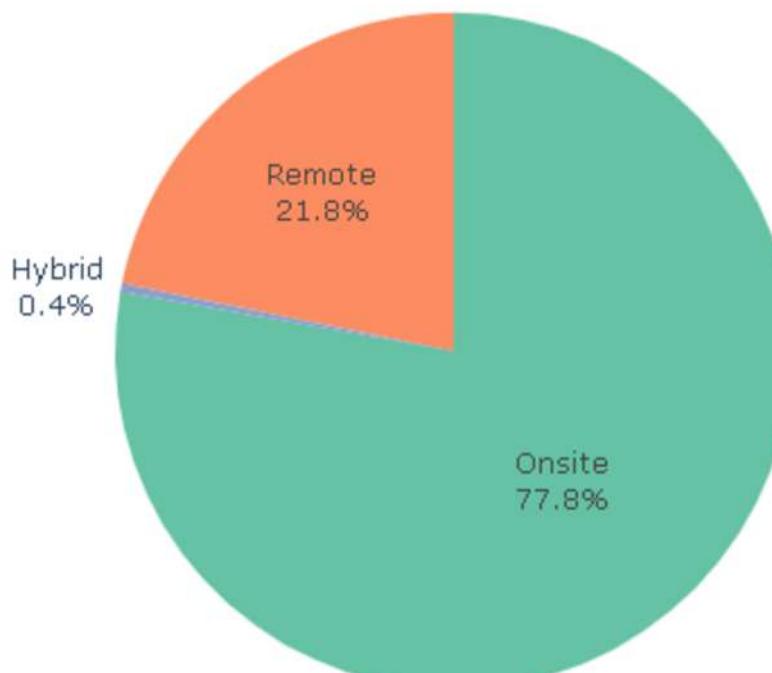
01

## Data Overview and Analysis

- Analyzing *company\_size* and *work\_model*:



Company Size Distribution



Work Model Distribution

### Insights:

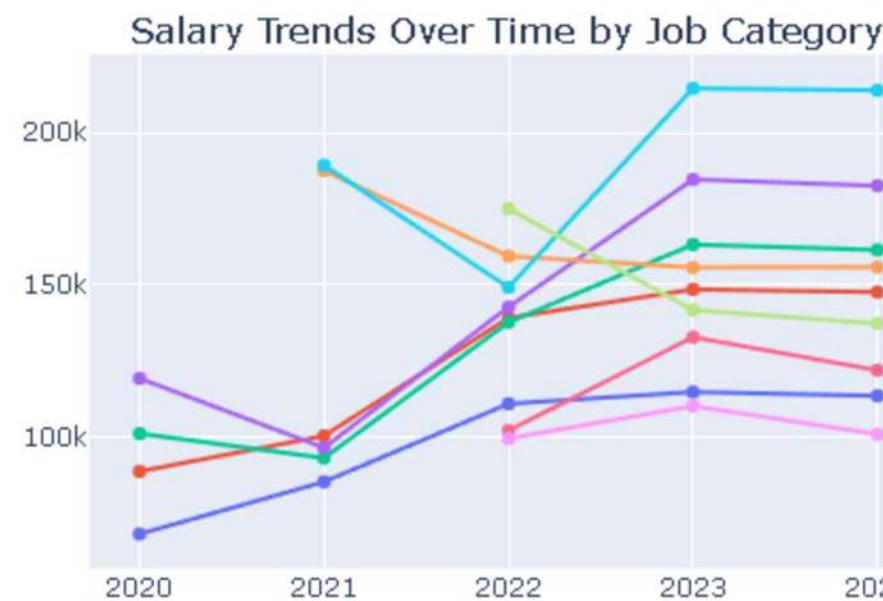
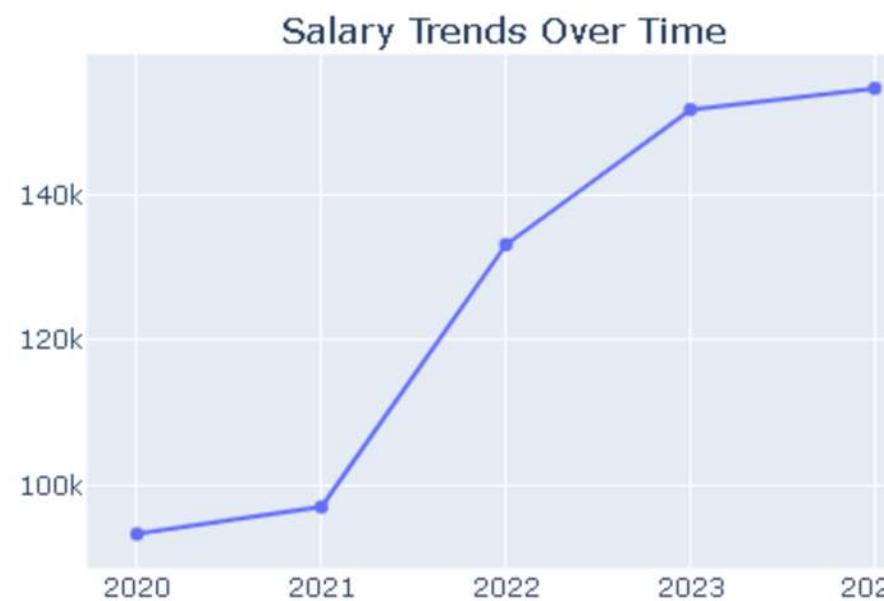
- Medium-sized companies employ the vast majority of data professionals (**96.3%**), compared to the very small number at small (**0.294%**) and large (**3.36%**) firms.
- This could mean that medium-sized companies likely have the resources and growth stage that drive the demand for dedicated data teams. Smaller companies may have limited budgets and needs, and larger companies may focus on internal promotion, which decreases the need for external talent.
- Onsite positions are the most common in the data industry (**77.8%**), while remote work has a substantial presence (**21.8%**), and hybrid work is less frequent (**0.4%**).
- The preference for onsite work in the data industry persists due to collaboration needs, team dynamics, security issues, and other factors related to physical location.

# DATA ANALYSIS AND VISUALIZATION

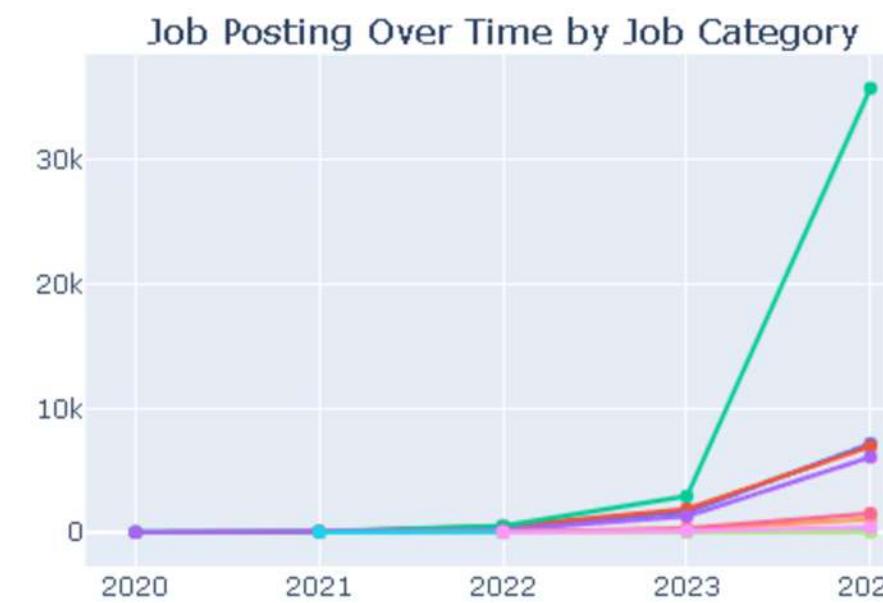
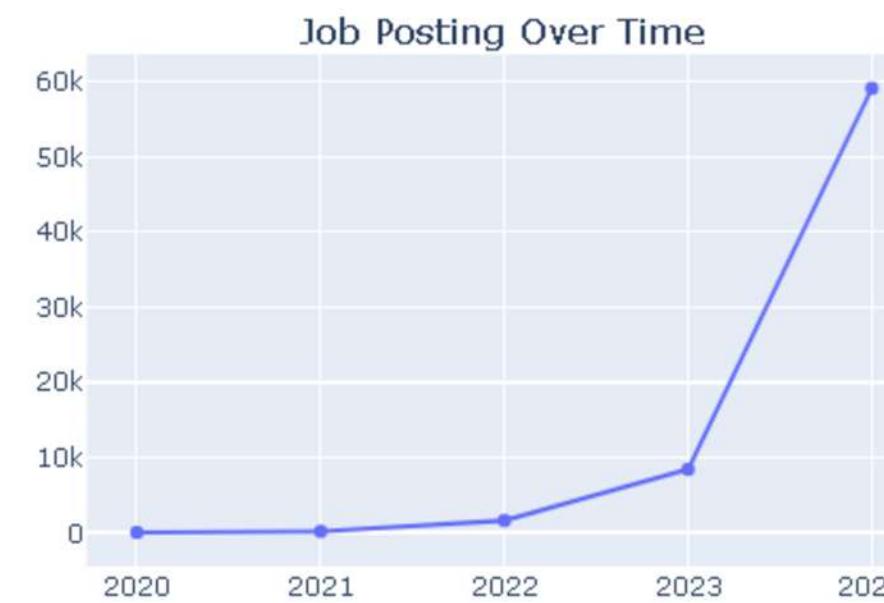
02

## Exploratory Data Analysis

Overview: Salary and Job Postings Trends



- Data Analysis
- Data Analysis
- Data Engineering
- Data Engineering
- Data Scientist
- Data Scientist
- Machine Learning and AI
- Machine Learning and AI
- Data Architecture and Modeling
- Data Architecture and Modeling
- Leadership and Management
- Leadership and Management
- BI and Visualization
- BI and Visualization
- Cloud and Database
- Cloud and Database
- Data Management and Strategy
- Data Management and Strategy

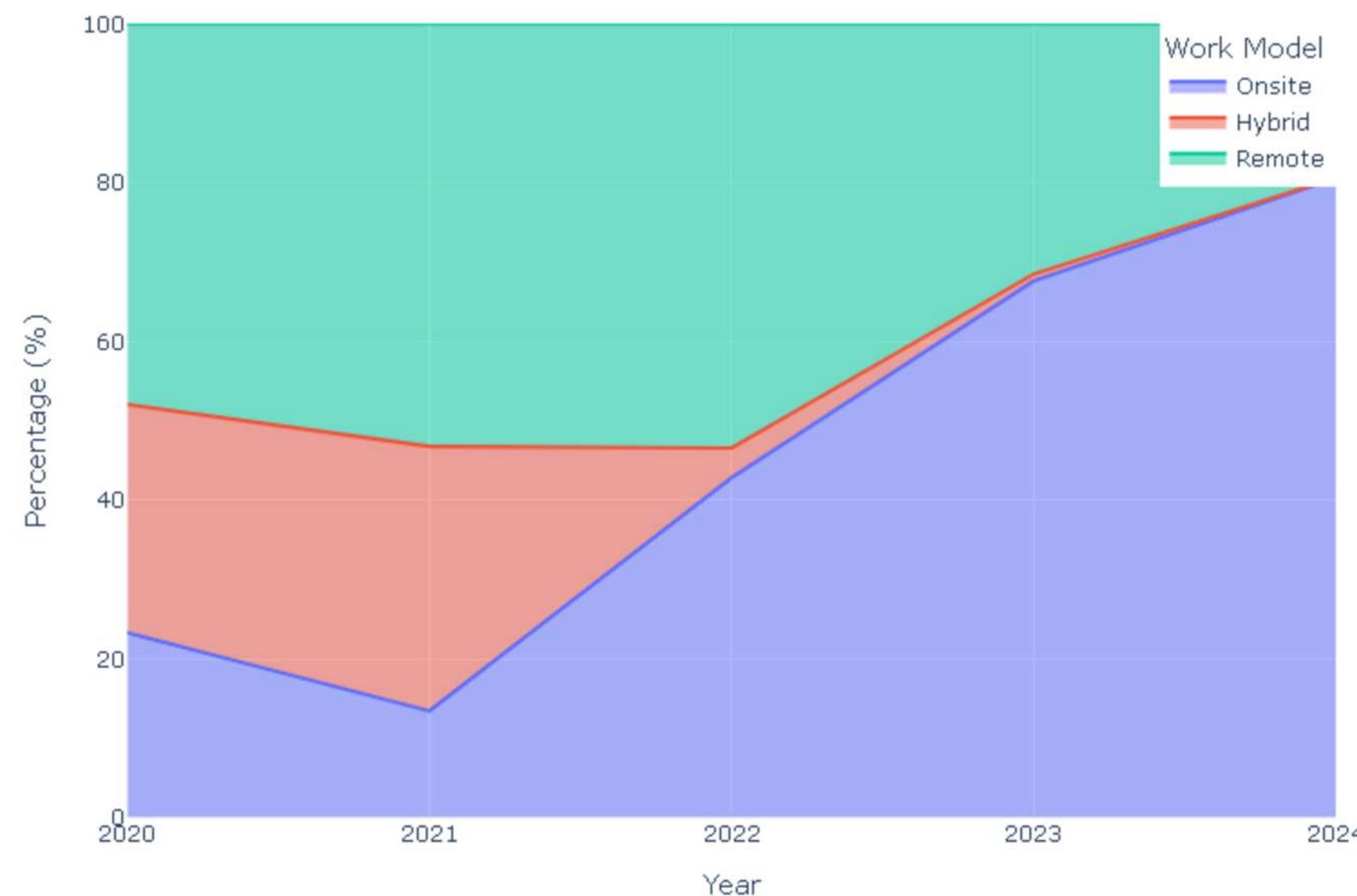


# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

Remote Work Trends Over Years



### Insights:

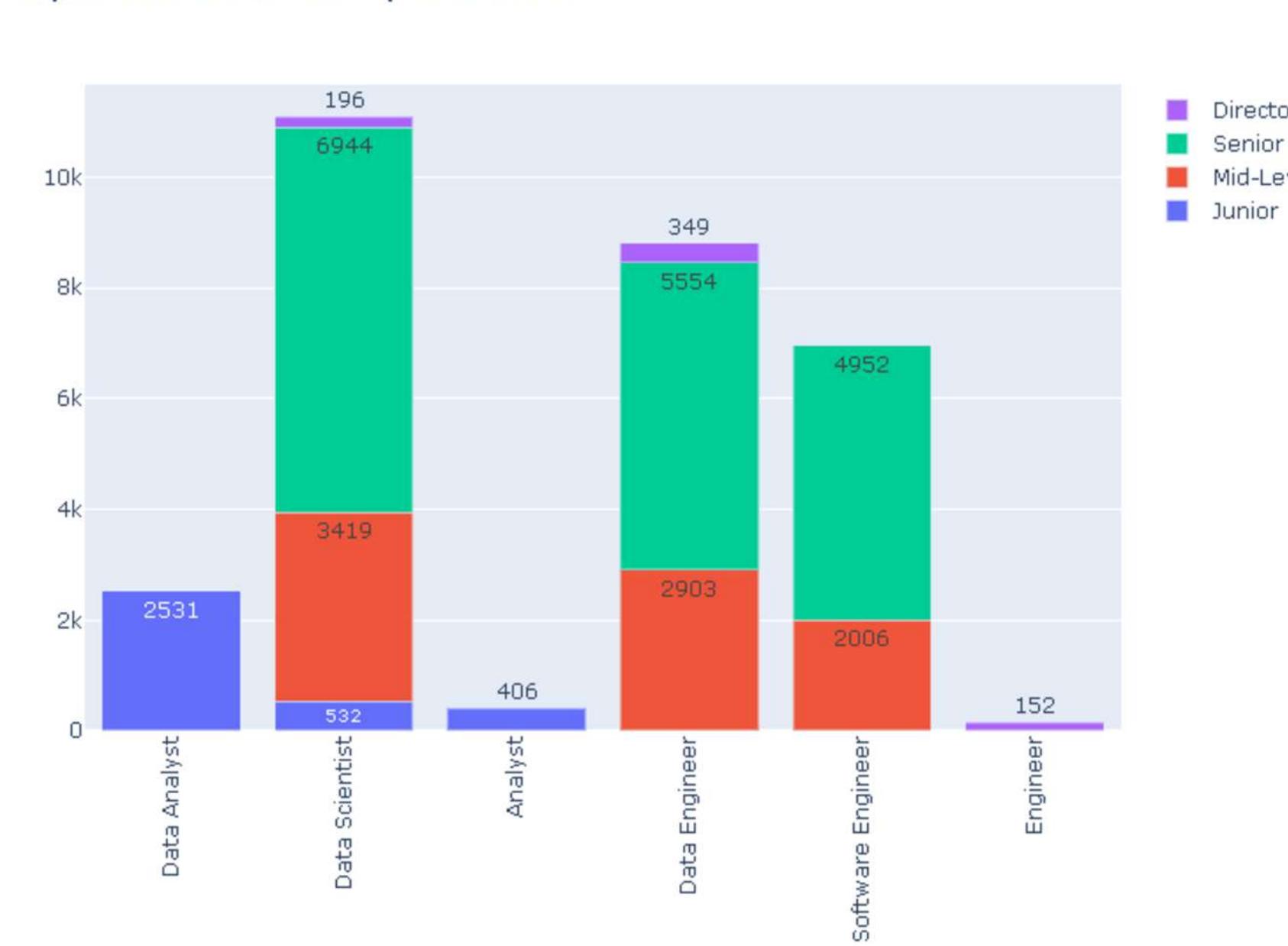
- Remote Work Trends:** The hybrid work model was notable in 2020, but disappeared by 2024. Remote work increased and then declined, with onsite work becoming dominant again by 2024.
- Overall Salary Growth:** There has been a consistent upward trend in salaries from 2020 to 2024, with a notable jump between 2022 and 2023 and then again in 2024.
- Salary Trends by Category:** Leadership/Management and Data Science roles command the highest salaries, with notable growth also in Data Engineering roles.
- Job Posting Growth:** Job postings have steadily increased from 2020 to 2023, and saw a huge jump in 2024.
- Job Posting by Category:** Data Scientist, Data Analysis, Data Engineering, and Machine Learning/AI roles have seen the largest growth in job postings.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

Experience Level with Top 3 Job Titles



### Insights:

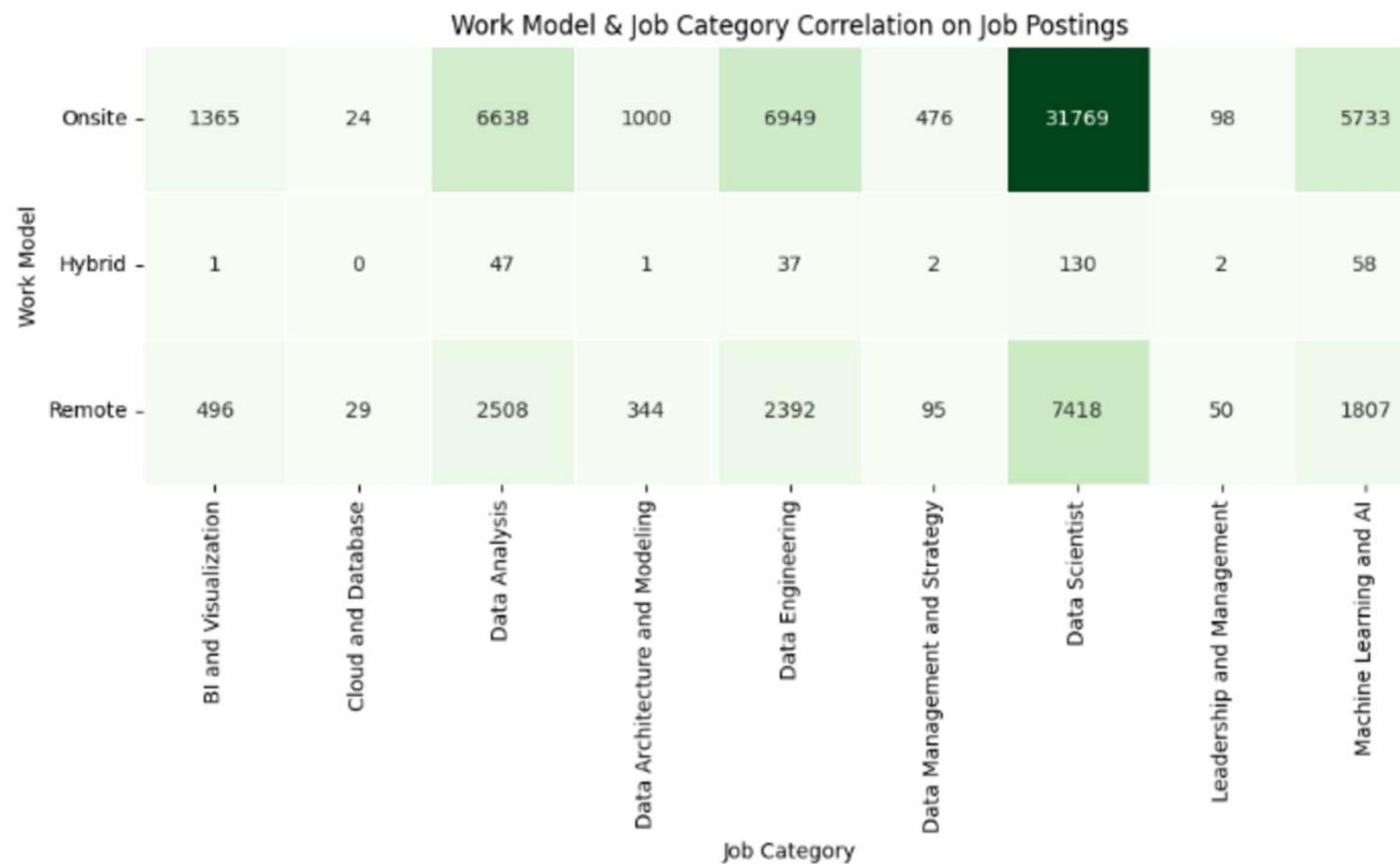
- **Data Analyst:** Mainly junior-level, due to lower entry requirements.
- **Data Scientists & Data Engineer:** Focus on experienced (Senior and Mid-Level) professionals, due to specialized skill needs.
- **Analyst:** All junior-level, likely an entry position.
- **Software Engineer:** Mostly Senior and Mid-level employees, due to the complexity of the work.
- **Engineer:** Few Directors, potentially due to a career path or the company's maturity.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- **Work Model & Job Category** Correlation on by Job Postings



### Insights:

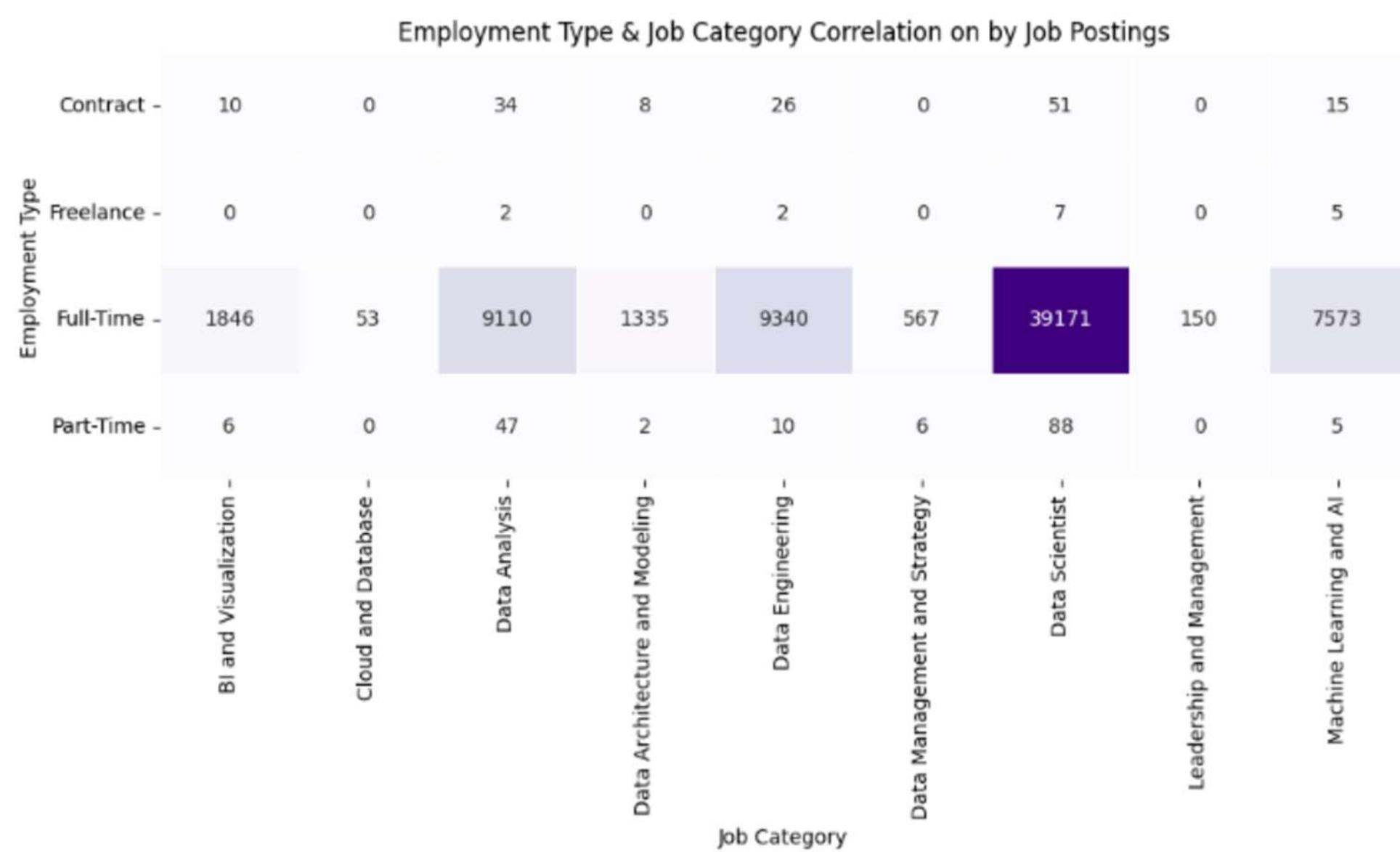
- Data Scientists are prioritized for hiring, especially for onsite positions.
- There are a significant number of remote positions for Data Scientists, Data Analysts, and Data Engineers, demonstrating that remote work remains a popular trend.
- The hybrid work model appears to be utilized less by companies than the main three.
- There is a balance between the number of Data Analyst, Data Engineer, and Data Scientist positions being hired for both onsite and remote work.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- **Employment Type & Job Category** Correlation on by Job Postings



### Insights:

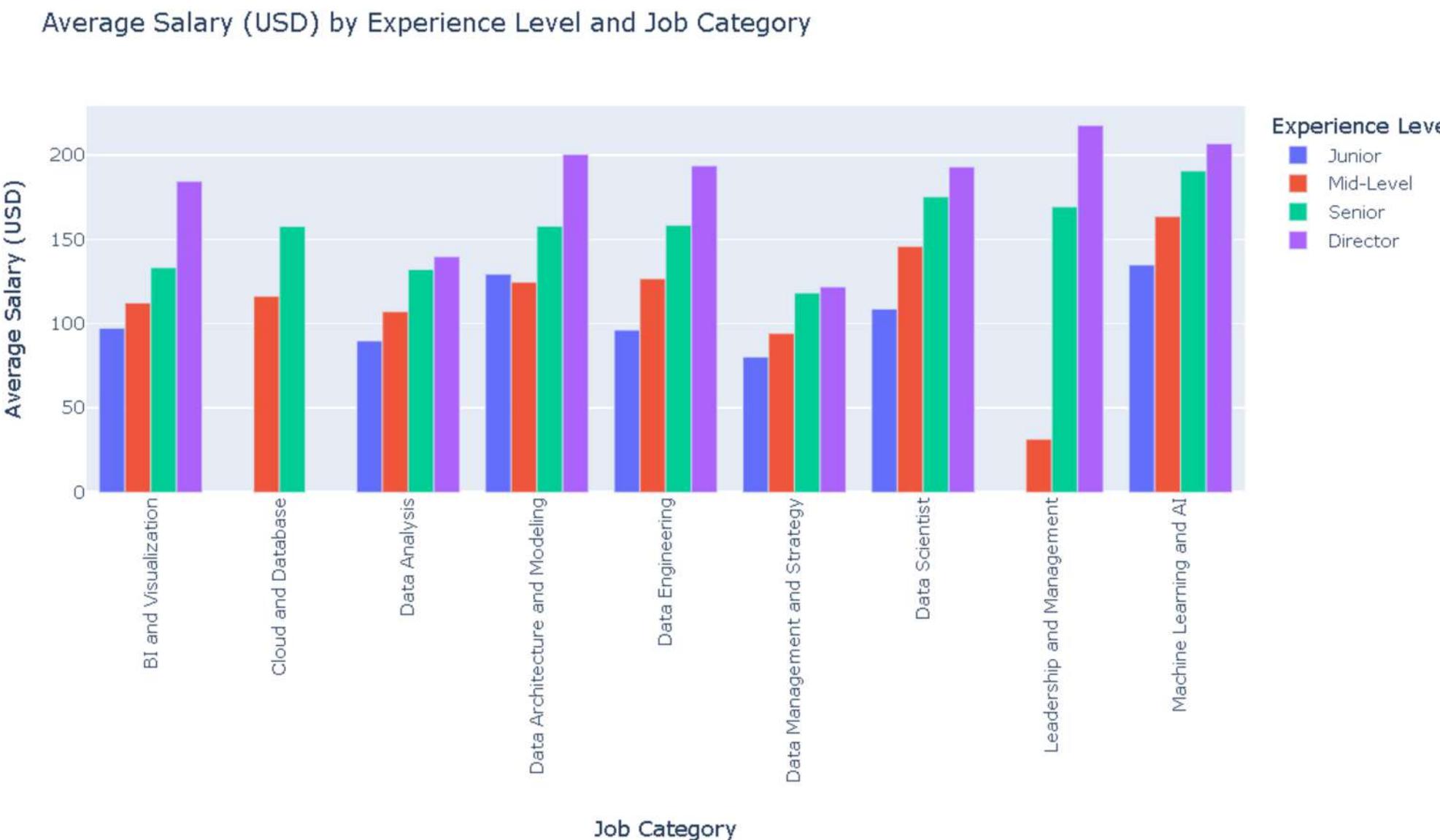
- Full-time hiring is overwhelmingly dominant compared to other employment types.
- The Data Scientist position is the most hired role, especially in full-time positions.
- Freelancer and part-time roles are not common in this field.
- Some contract positions still exist, indicating a demand for short-term projects.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- Average Salary (USD) by Experience Level and Job Category



### Insights:

- Salaries generally increase with experience and vary significantly across job categories, with directors typically earning the highest.
- Fields like AI and Machine Learning command the highest salaries, while Data Management tends to have lower salaries.
- Market demand, experience skills, along with location and company size all influence salary levels, and salary hierarchies are usually consistent with experience.
- The current job market highly values technology-related roles, especially those in AI.

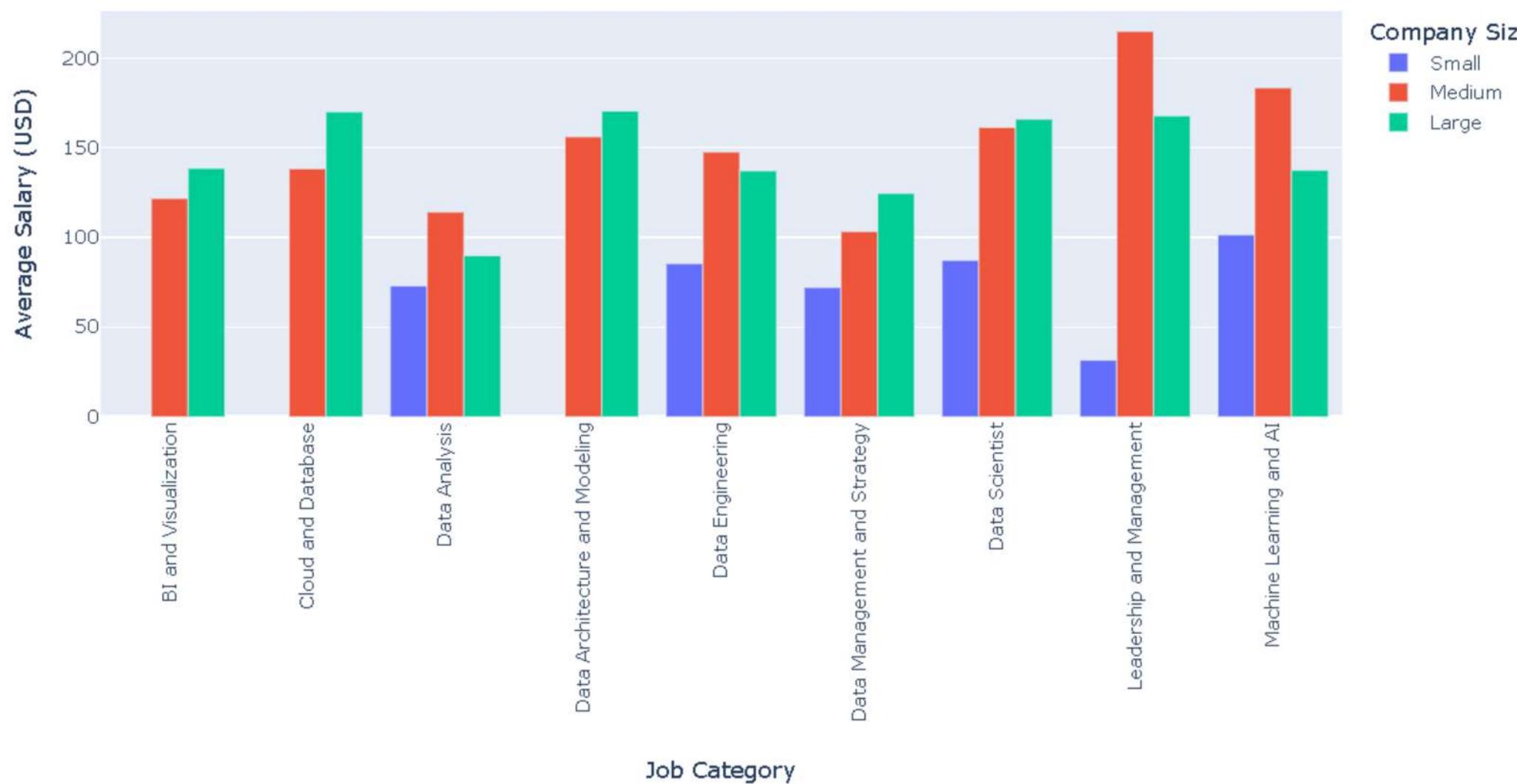
# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- Average Salary (USD) by Company Size and Job Category

Average Salary (USD) by Experience Level and Job Category



### Insights:

- Large companies generally offer higher average salaries across most job categories, likely due to greater resources.
- Medium companies are the exception and often offer better pay for Data Analysis, and Leadership & Management roles.
- Small companies tend to provide the lowest average salaries compared to medium and large organizations.
- There is a significant variance in pay among different job categories.

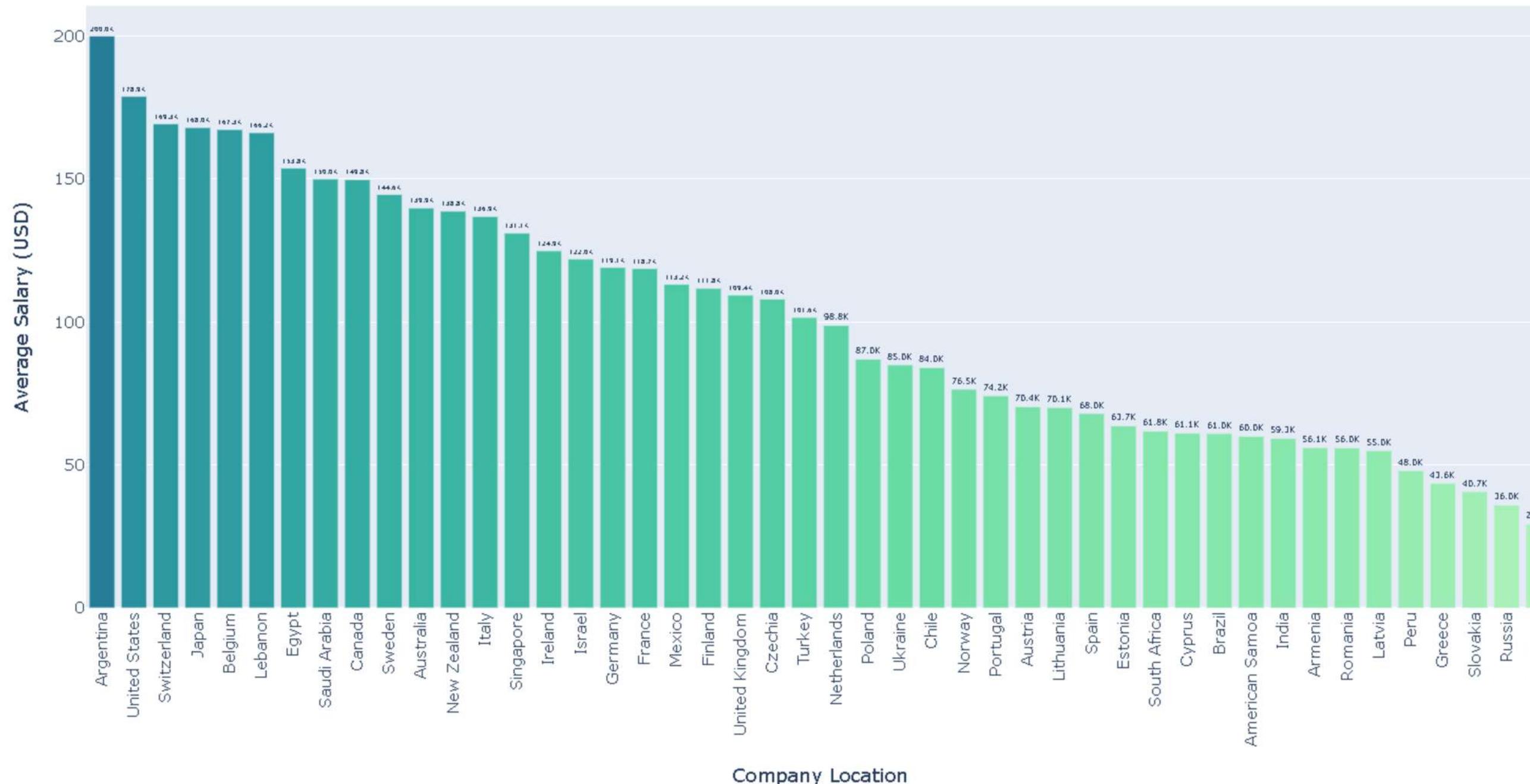
# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- Average Salary (USD) for Senior Data Scientists by Company Location

Senior Data Scientist average salary by company location

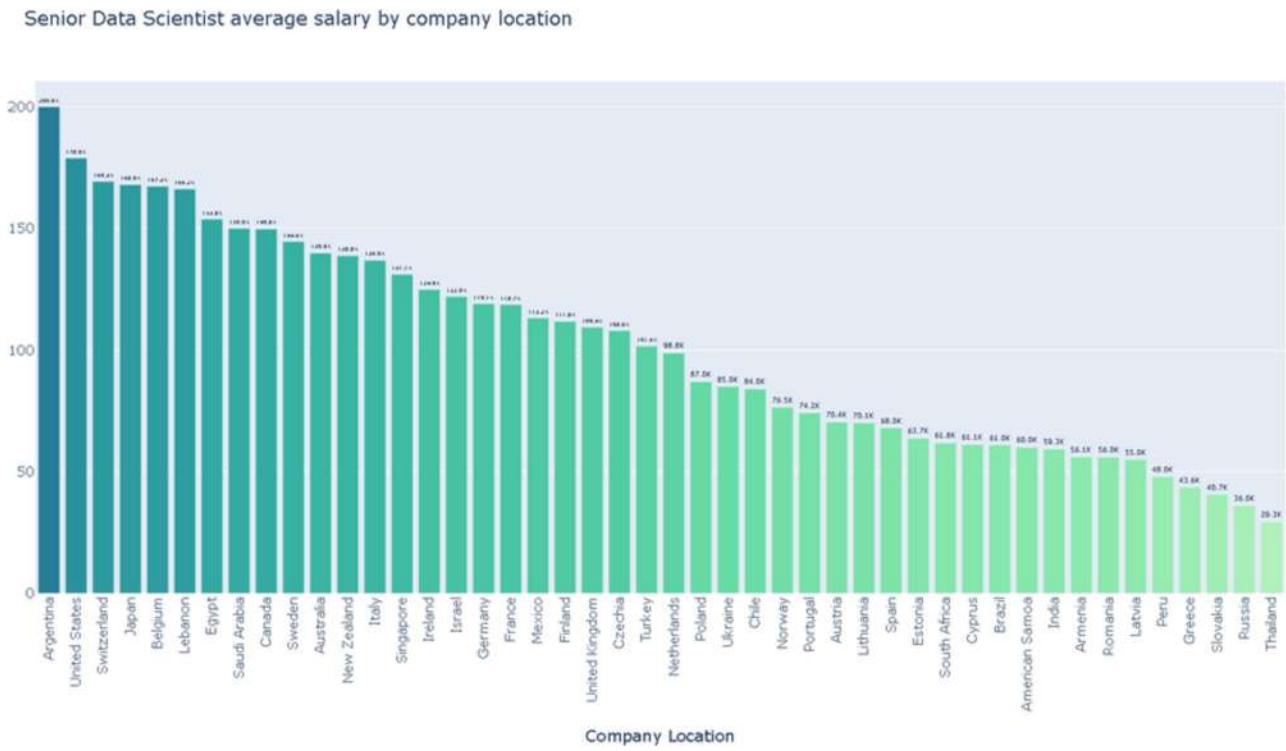


# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- Average Salary (USD) for Senior Data Scientists by Company Location



### Insights:

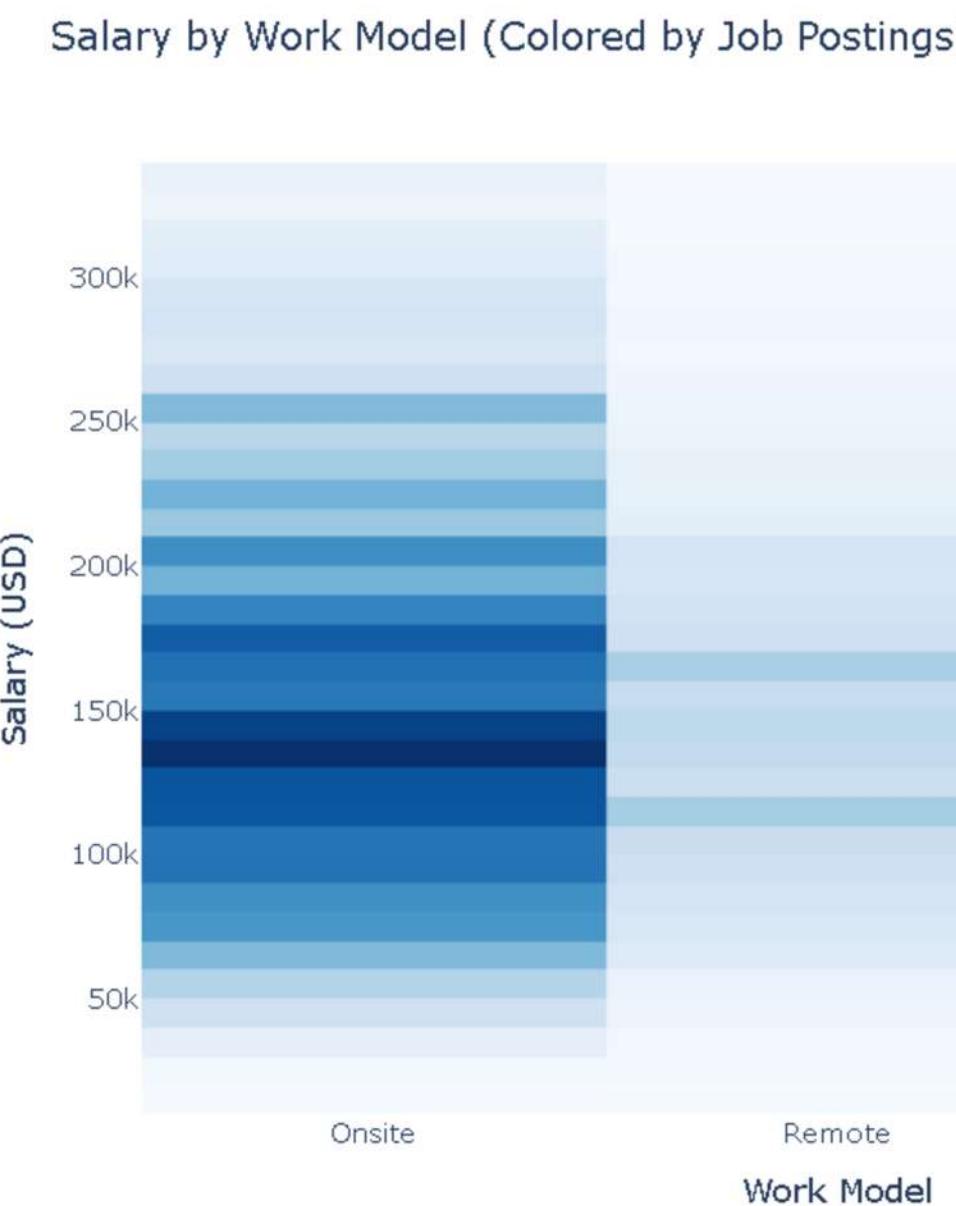
- It's clear that where you as a Senior Data Scientist really matter when it comes to your paycheck. Average salaries for this role vary greatly from one country to another.
- Argentina, the United States, and Switzerland are where you'll likely find the highest compensation for Senior Data Science positions, likely because of the strong demand and higher cost of living in those areas.
- On the flip side, countries like Thailand and Russia tend to have the lowest average salaries for Senior Data Scientists. This might be due to differences in how well-established the Data Science field is in those countries, and the lower cost of living there.
- The differences in salaries can be attributed to things like: how expensive it is to live there, how much demand there is for these types of jobs in that location, the general economic state of the country, and the currency rates between the local currency and the dollar.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- Average Salary & Work Model Comparison by Job Postings



### Insights:

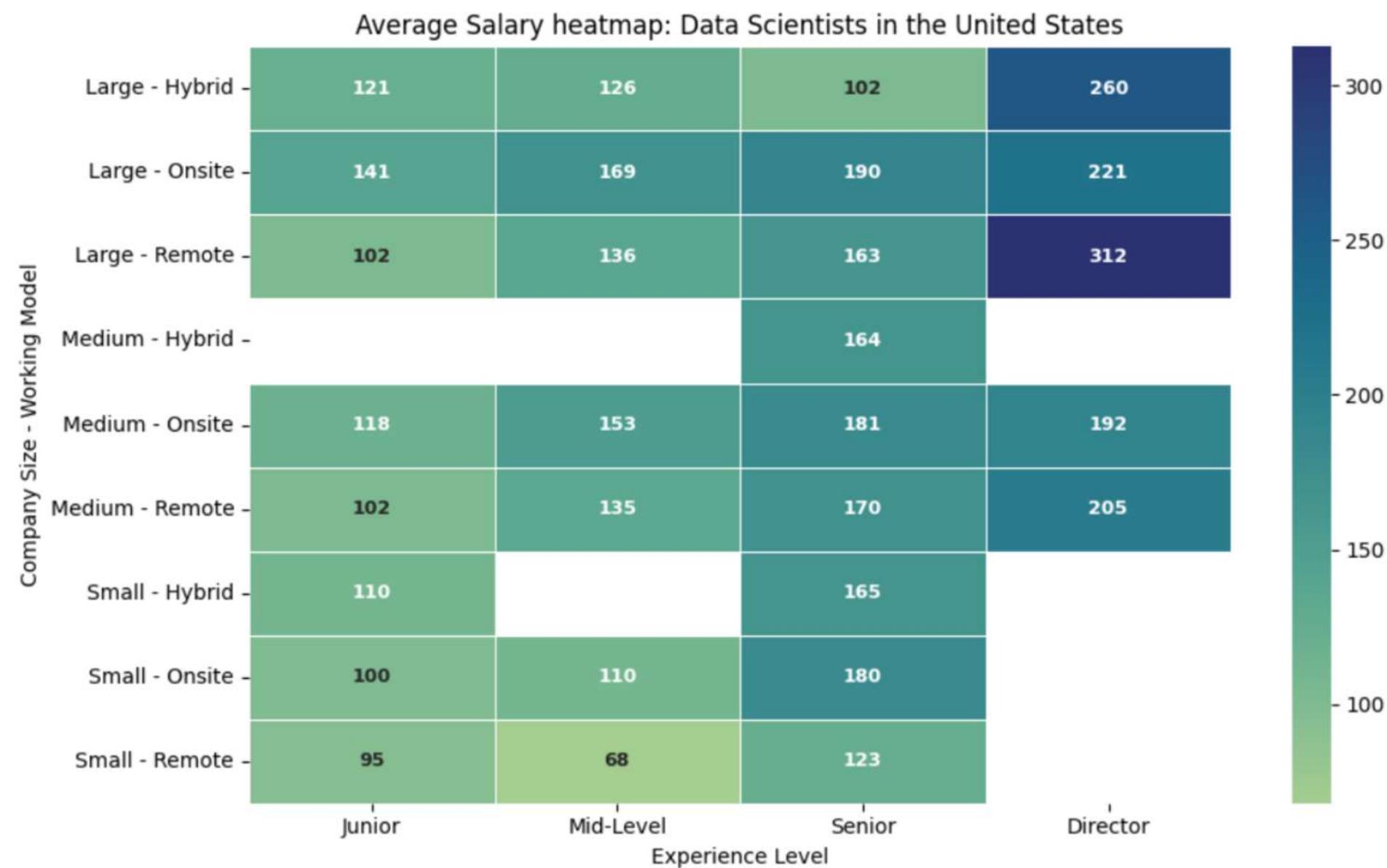
- Onsite work remains the most common model, especially for jobs at mid-range and higher salary levels.
- The majority of job opportunities are concentrated within the 100k to 200k dollar salary range.
- Remote job posting are less numerous compared to onsite options, and they tend to cluster around the 100k to 175k dollar salary range.
- The hybrid work model is the least prevalent among the three, and it exhibits a narrower salary range.
- The overall salary level does not seem to be greatly influenced by the work model chosen.
- Hybrid models may be utilized by companies to retain talent that desires flexibility while being familiar with onsite work.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- Average Salary heatmap: **Data Scientists in the United States**



### Insights:

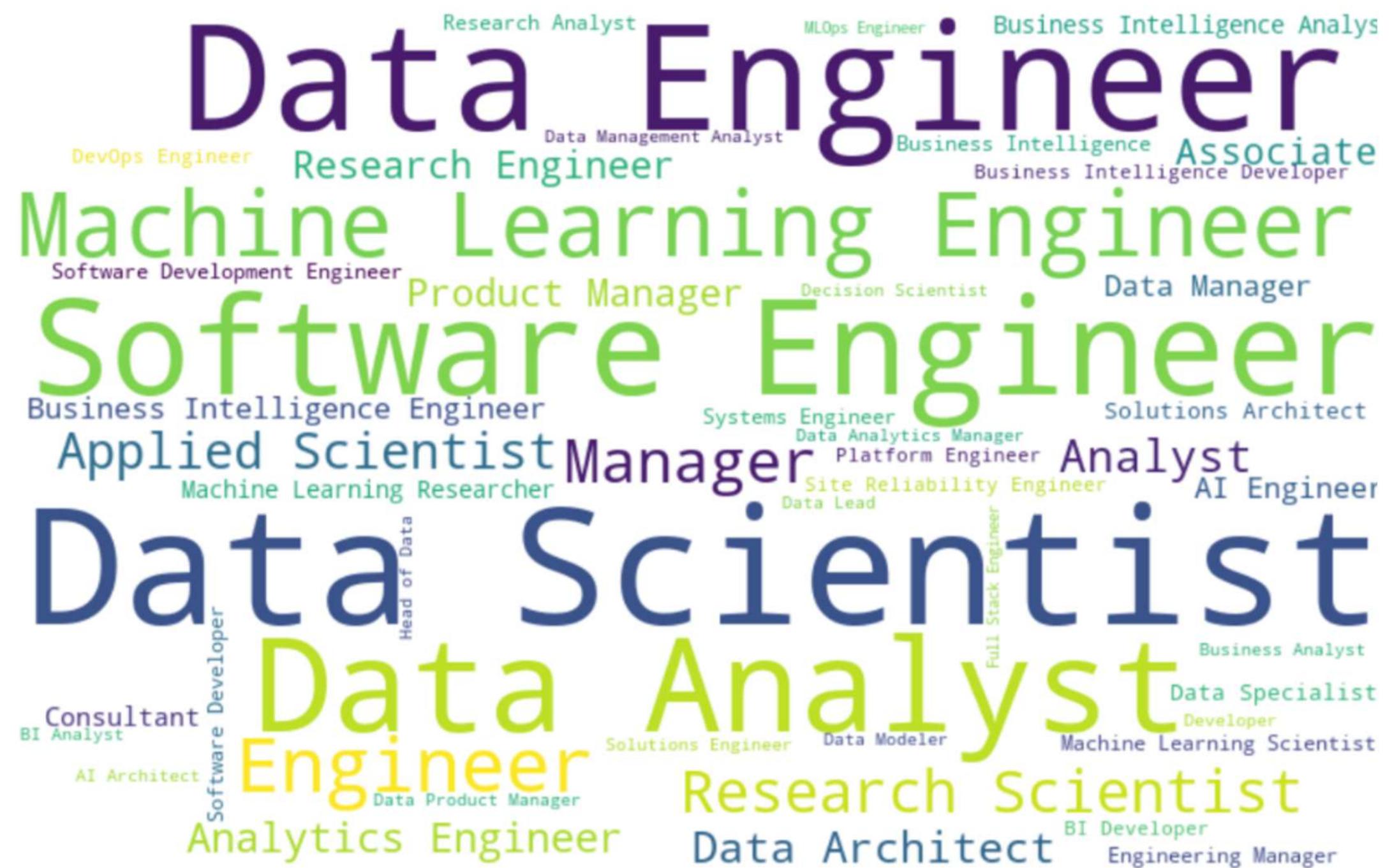
- Director-level positions consistently have the highest average salaries, especially in large and remote companies.
- Large companies generally offer higher pay compared to medium and small companies, particularly for leadership roles.
- Salaries for Data Scientists increase as experience progresses from junior to mid-level and then to senior.
- Small companies typically provide lower average salaries compared to larger companies.
- Remote work models can lead to a pay increase at high-level positions, primarily in large organizations.
- Experience plays a vital role in determining the salary level of Data Scientists.

# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- **WordCloud:** Job title Distribution by Job Postings



# DATA ANALYSIS AND VISUALIZATION

02

## Exploratory Data Analysis

- **WordCloud:** Job title Distribution by Job Postings



### Insights:

The current job market is dominated by tech and data roles. Positions related to data processing, analysis, and management are highly prevalent, highlighting the digitalization trend across various industries. To meet this demand, both technical skills in system building and analytical skills for decision-making based on data are crucial. Additionally, the variety of specialized roles showcases increasing specialization in the tech market and the ever-growing importance of leadership in complex technology-based projects.



PERSONAL PROJECT

---

# SUMMARY

---

# SUMMARY

## *Summary*

The Data Science job market is experiencing strong growth, especially in 2024, with the highest demand focused on Data Scientist positions. The US is the main hub of this market, and medium-sized companies are the most active recruiters. While onsite work remains prevalent, remote work options also hold significant importance. Salaries tend to increase, particularly for experienced professionals, management roles, and those in AI/Machine Learning fields. Factors such as experience, position, company size, and location all have a significant impact on salary.

# KEY INSIGHTS

## *Key Insights*

- **High Demand for Data Scientists:** Data Scientist positions are the most in-demand, comprising over half of all job postings, indicating a strong market focus.
- **Strong Growth in the Data Science Market:** 2024 is seeing a surge in hiring demand in Data Science, especially for roles in Machine Learning, AI, and Data Engineering.
- **The US is the Main Hub:** The majority of job postings and employees in Data Science are located in the United States, making it the most important market.
- **Medium-sized Companies Dominate Hiring:** Medium-sized companies tend to hire more Data Science professionals compared to large or small companies.
- **Onsite Work Remains Prevalent:** Despite the growth of remote work, onsite work is still the most common model in Data Science.
- **Steady Salary Growth:** Salaries in Data Science tend to increase steadily, particularly in management, Data Science, and Data Engineering roles.

# KEY INSIGHTS

## *Key Insights*

- **Experience and Position Greatly Influence Salary:** Salaries increase with experience and there is a significant difference between positions, with management positions typically having the highest salaries.
- **Salaries Vary by Country:** Salaries for Senior Data Scientists vary greatly between countries, reflecting differences in demand, cost of living, and other economic factors.
- **Roles of Technology and Analytics:** Positions related to technology and data analysis are dominating the job market, highlighting the importance of technical and analytical skills.
- **Variety in Work Models:** Although onsite work is still the most common, there is a significant number of remote positions, especially for Data Scientists, Data Analysts, and Data Engineers.

# RECOMMENDATIONS

## *Recommendations*

- **For Job Seekers:** Focus on Data Scientist, Data Engineering, and Machine Learning/AI positions. Develop technical and data analysis skills. Consider opportunities in the US and at medium-sized companies.
- **For Employers:** Prioritize finding and attracting experienced Data Scientist professionals. Offer competitive salaries and flexible work models to attract top talent. Consider expanding remote recruitment to access a wider pool of potential candidates.
- **For Those Seeking Career Advancement:** Invest in developing specialized skills, particularly in AI/Machine Learning. Consider transitioning into management positions for higher compensation.



# THANK YOU!



December 2024

---

DUY DAT NGUYEN