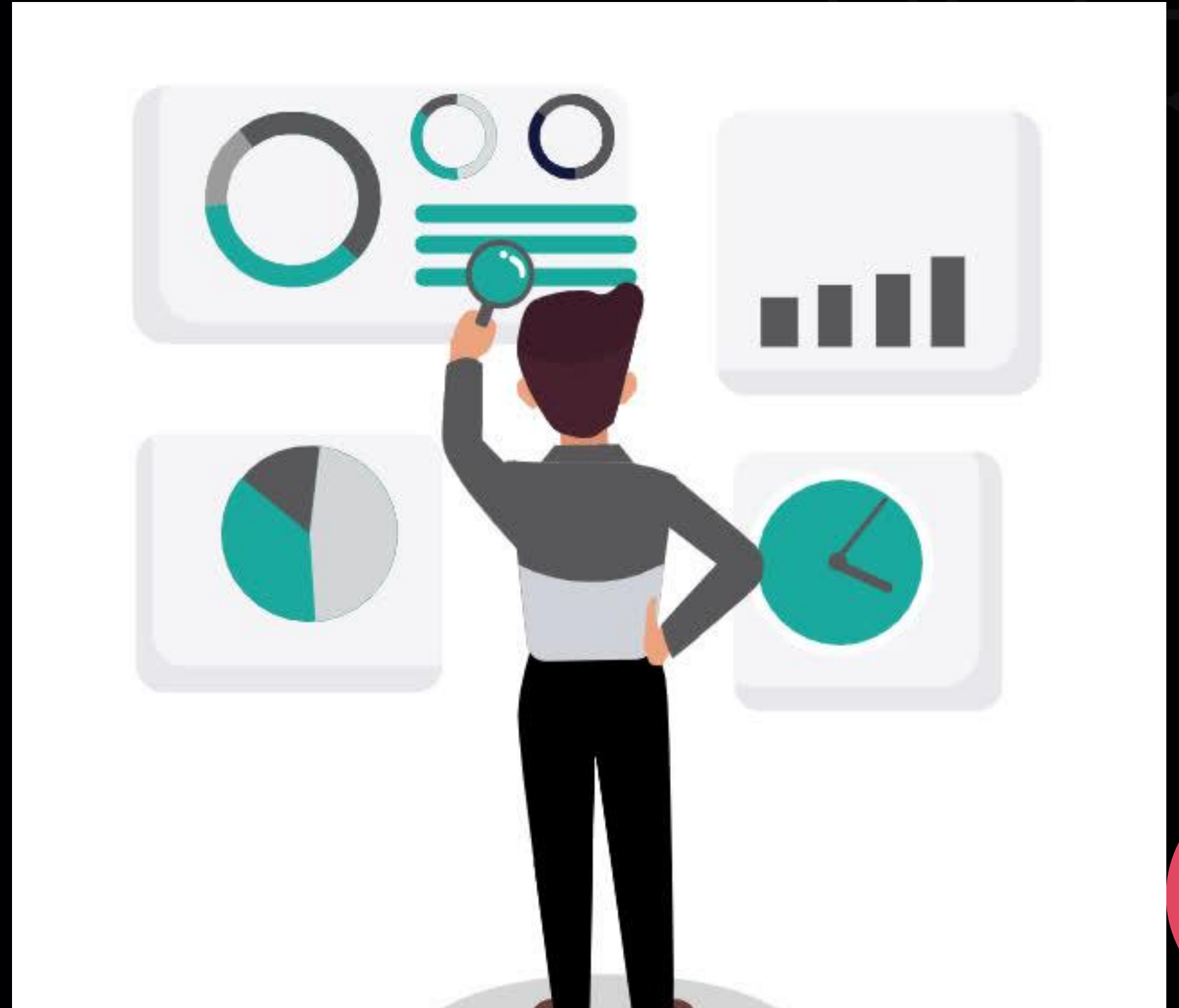# EDA ON DATA SCIENTIST SALARY
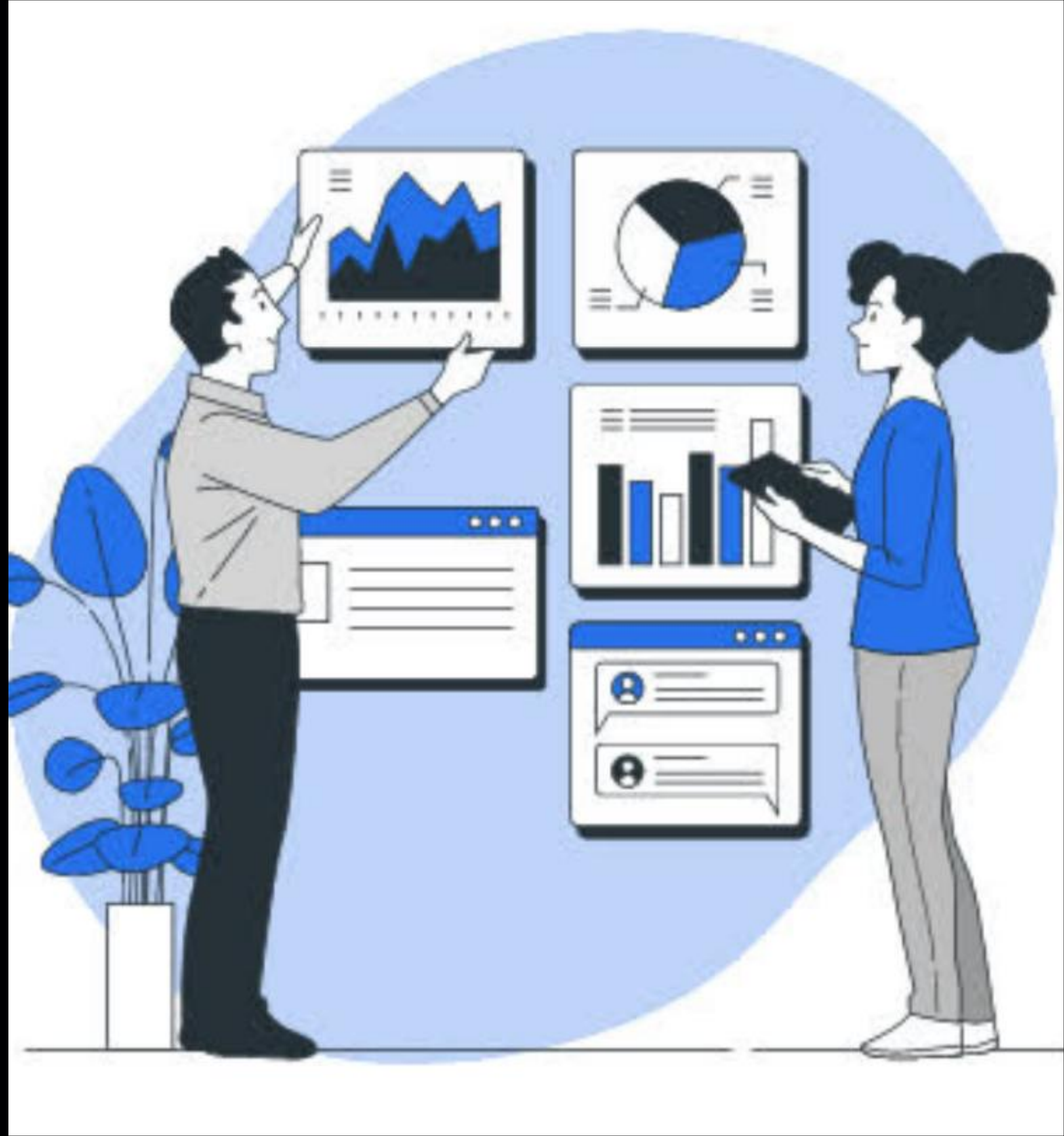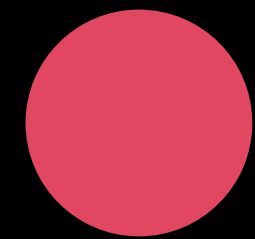
AGENDA

INTRODUCTION

- Data Scientist determine the questions their team should be asking and figure out how to answer those questions using data

  A data scientist might do the following tasks on a day-to-day basis: Find patterns and trends in datasets to uncover insights.

ED A

my data set consist of 607 rows and 12 column

work_year

experience_level

employment_type

job_title

salary

salary_currency

salary_in_usd

employee_resident

remote_ratio

company_location

company_size

SYSTEM DESIGN

Business Analysis → Data collection → Data Preprocessing → Implementation

Conclusion ← Data Visualization

IMPLEMENTION

Import required library

```
1  import numpy as np
2  import pandas as pd
3  import seaborn as sns
4  import matplotlib.pyplot as plt
5  import warnings
6  import matplotlib.pyplot as plt
7
```

Read file from external source

```
In [4]:  1  ds=pd.read_csv('ds_Salaries.csv')

In [5]:  1  ds
```

Shape is used to get the shape of data set

info() function is used to get the information of data set


```
1  ds.shape
```

(607, 12)


```
In [9]:  1  ds.info()

<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
```

DATA PREPROCESSING

Data Cleani

```
In [11]:    1  ds.isnull().sum()

Out[11]:  Unnamed: 0          0
          work_year           0
          experience_level    0
          employment_type     0
          job_title           0
          salary              0
          salary_currency     0
          salary_in_usd       0
          employee_residence  0
          remote_ratio        0
          company_location    0
          company_size        0
          dtype: int64
```
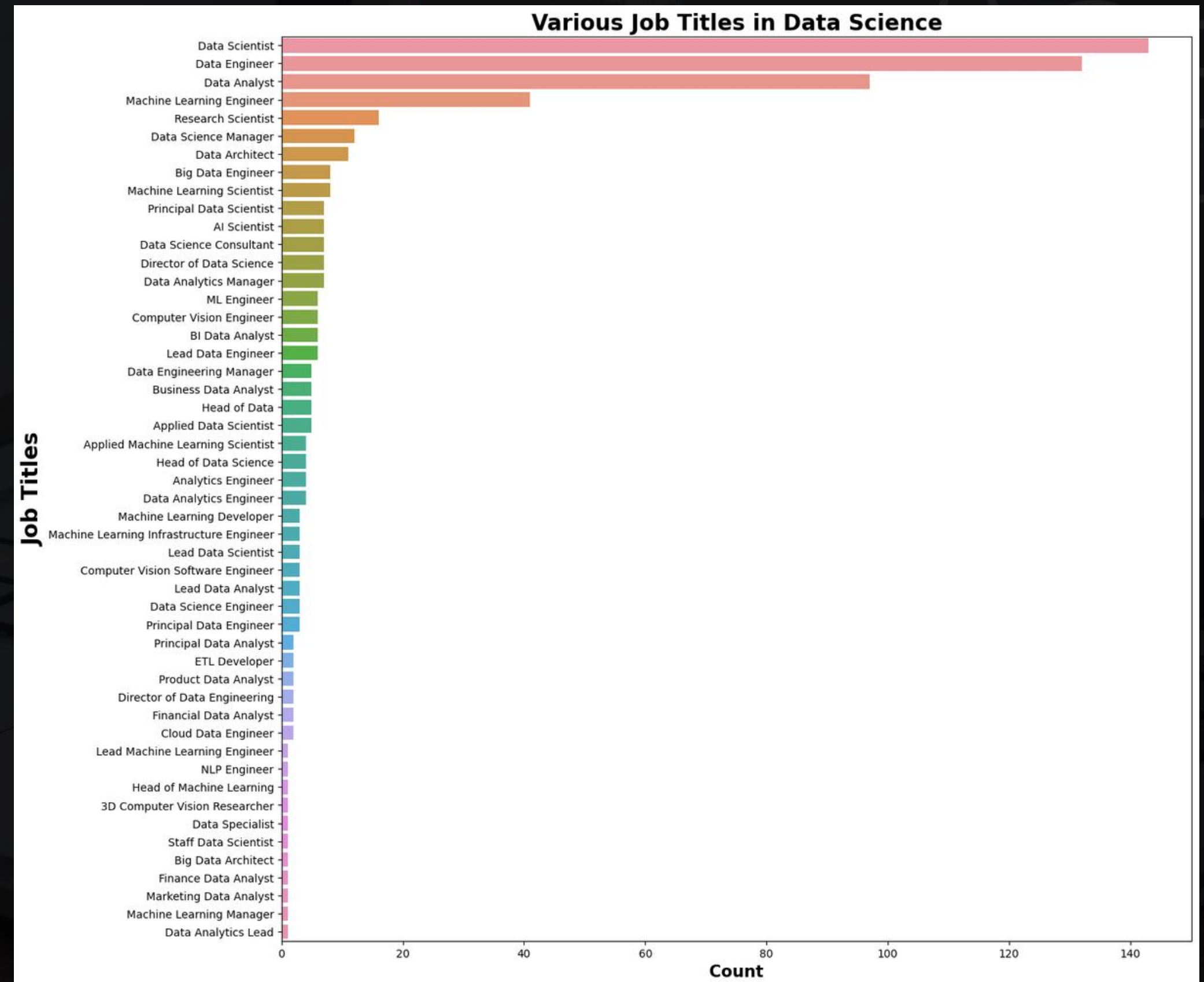
1.JOB TITLES

The top 4 job titles include Data Scientist, Data Engineer, Data Analyst and Machine Learning Engineer are more compared to Other job titles have less than 20 people in the respective posistion.

2.EXPERIENCE LEVEL

A large amount of people are of Data Science Professionals occupy Senior-level/Expert as well as Mid-level/Intermediate positions. However, Executive-level/Directors in Data Science are few, only 26 of the dataset
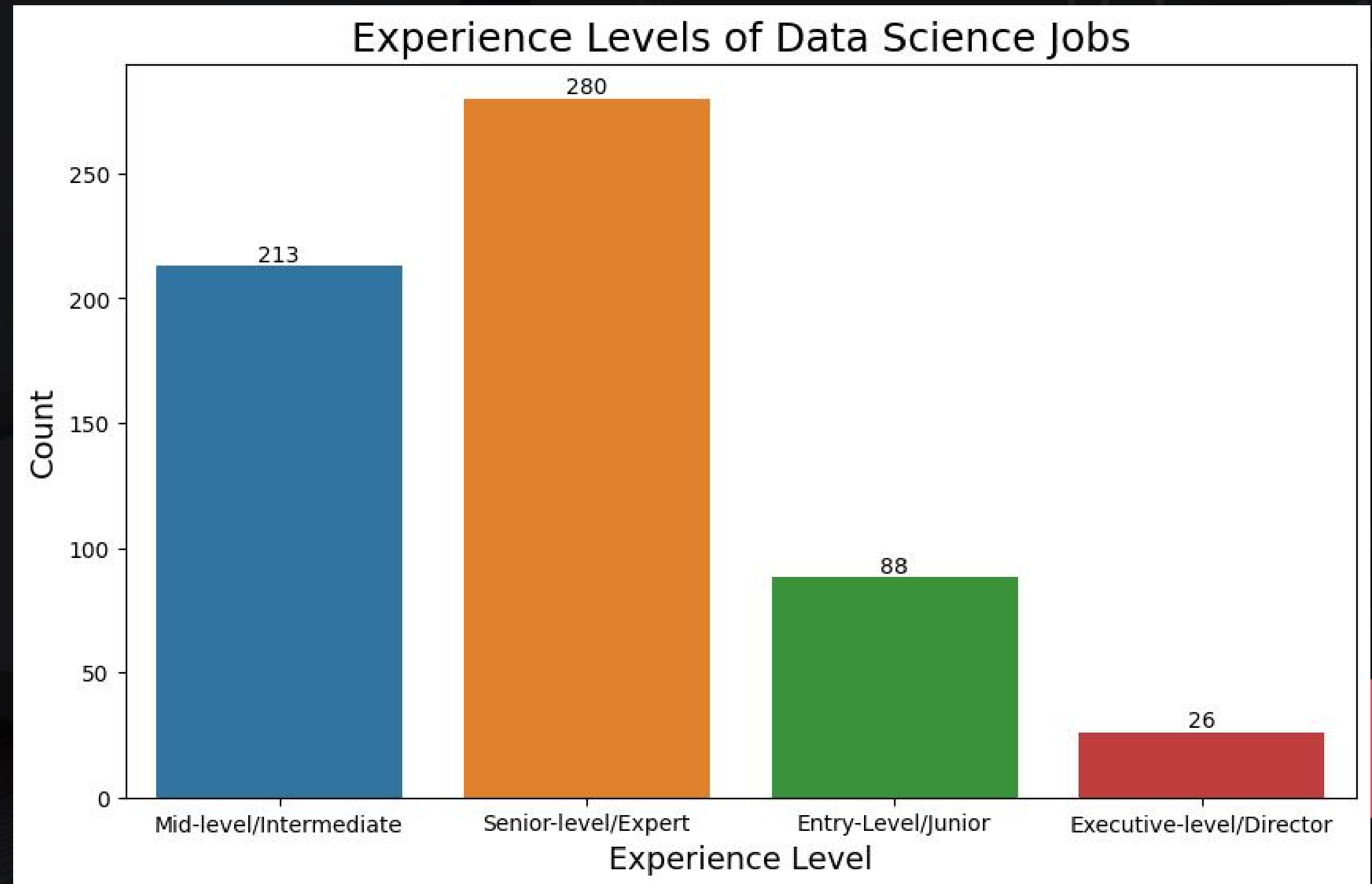
3.EMPLOYMENT TYPE

Full time employment is the most common employment Type in dataset and the other employment are a total of 19. Freelance is the least with only 4 .

Employment Type of Data Science Jobs
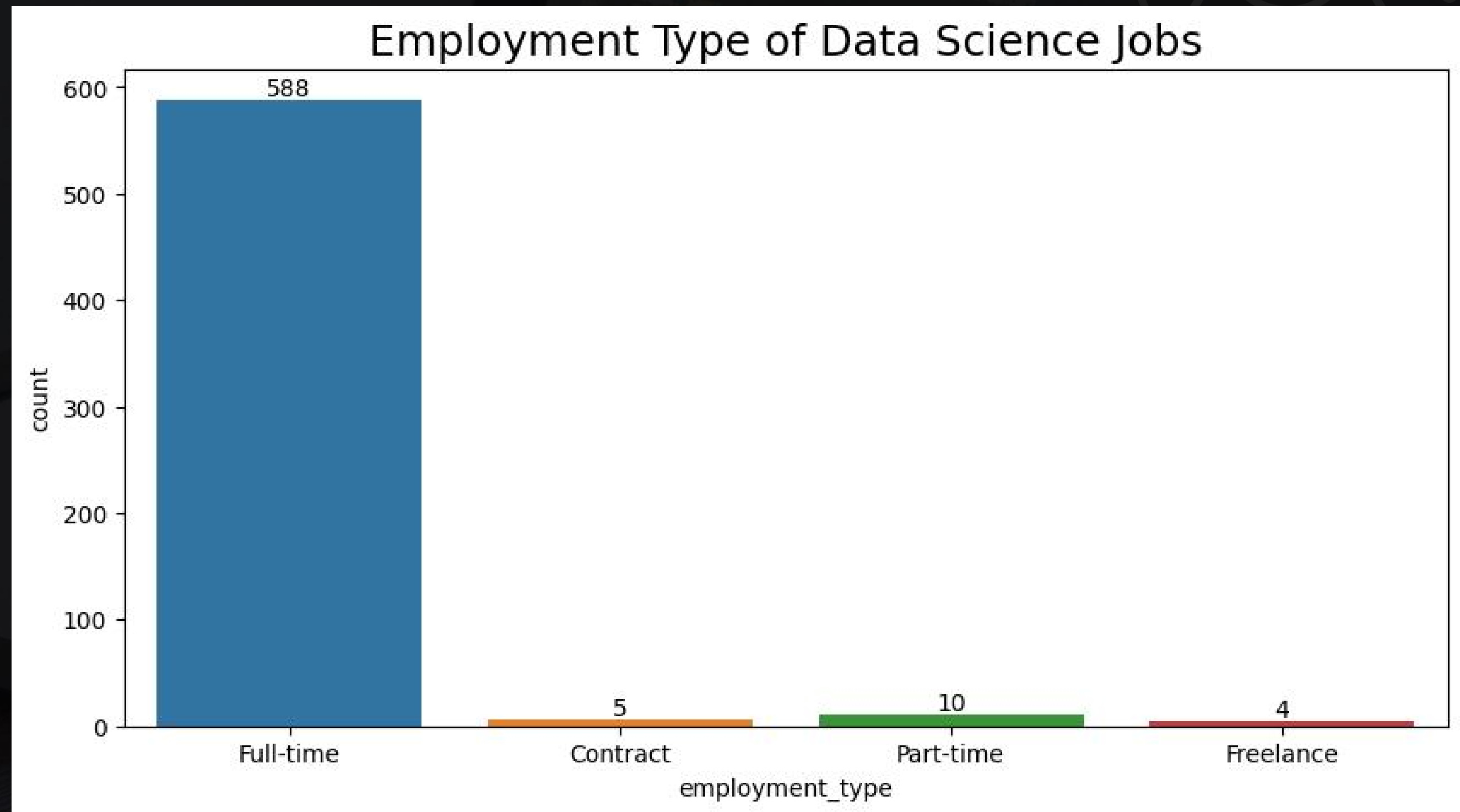
4.Employee Residence & Company

For Employee Residence and Company location, the US is the highest in terms of frequency, followed by Great Britain. A total of over 300 Employees and Companies are resident in the US while other locations record values below 50 for both Employee residence and company location.
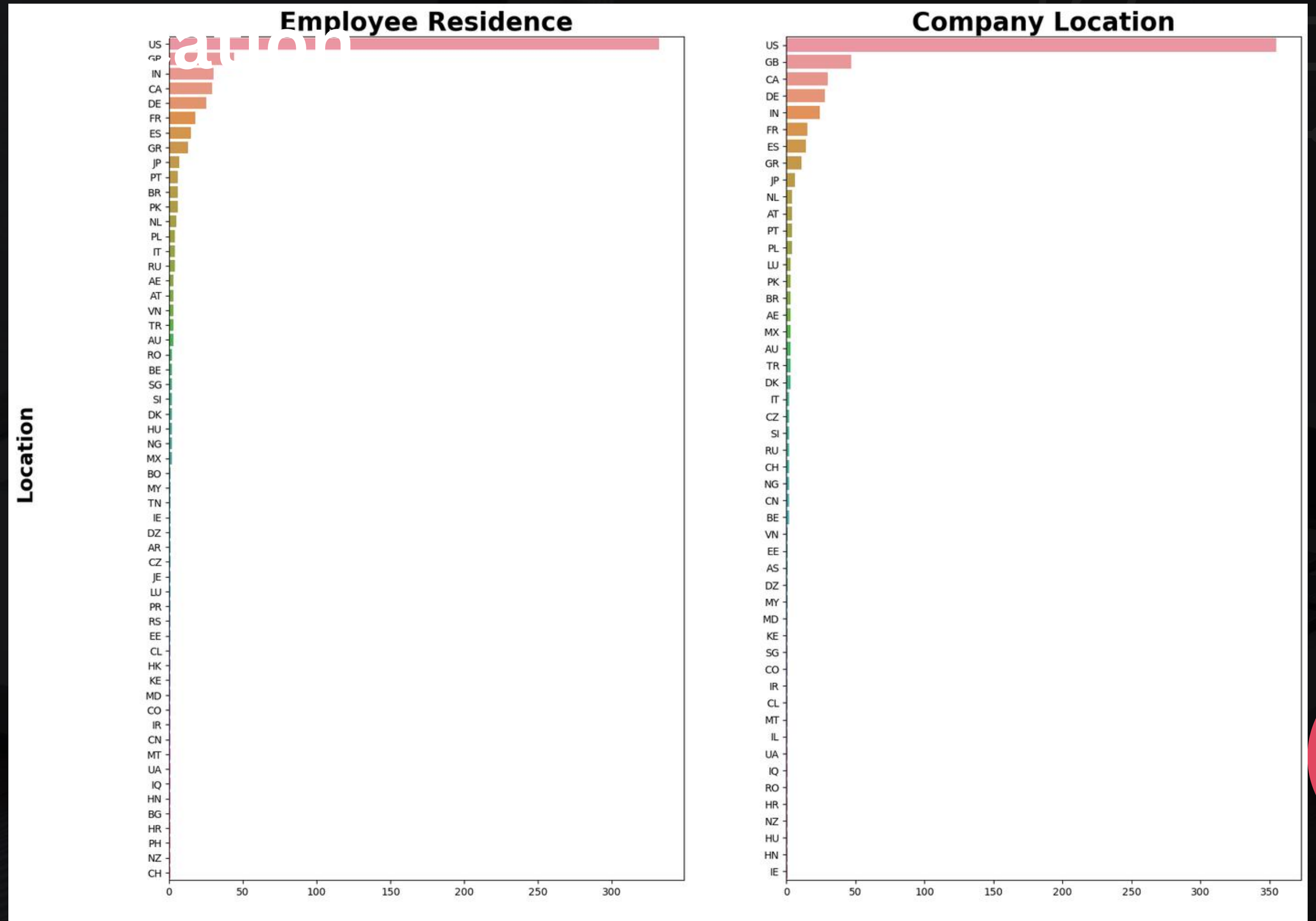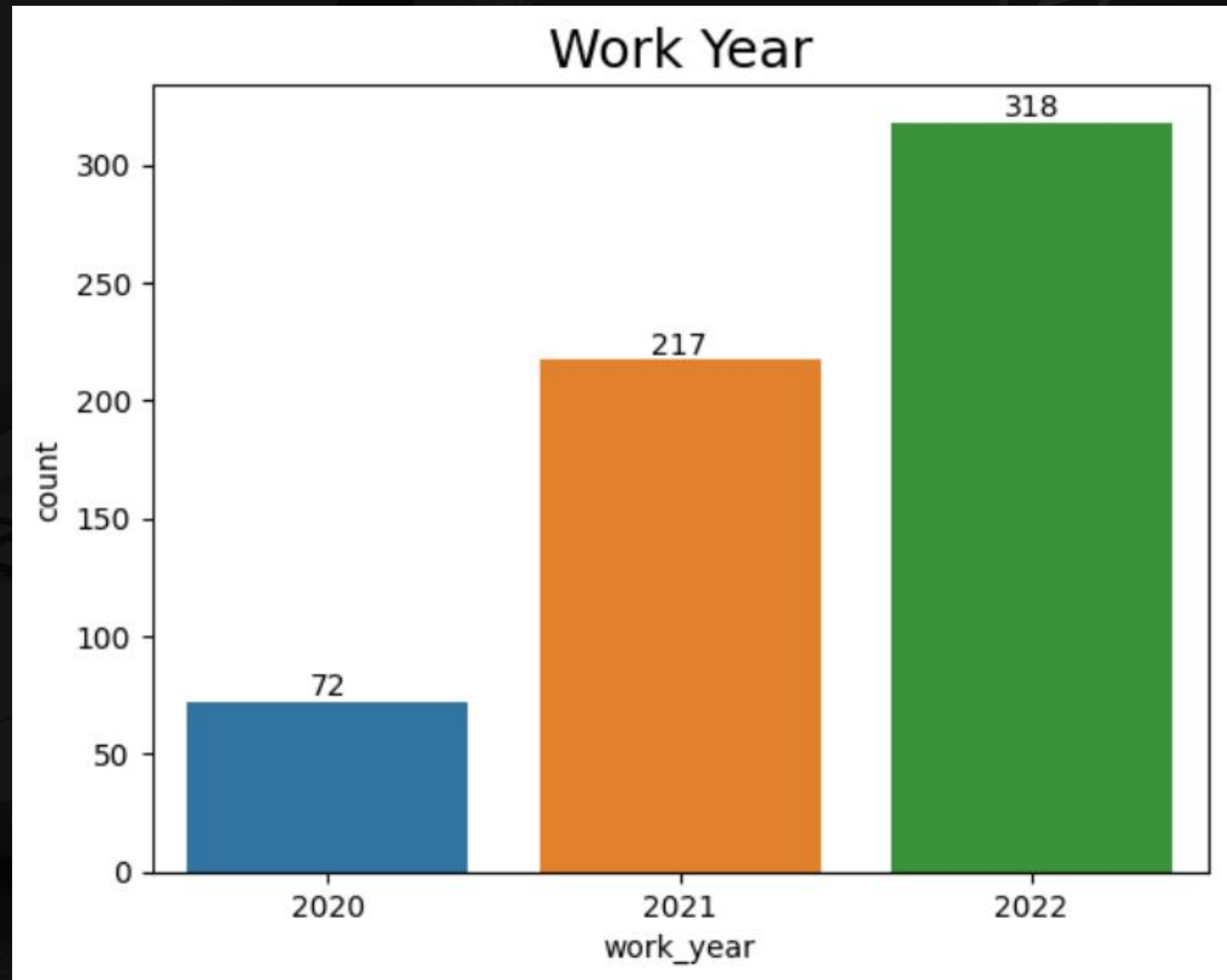
5.Company size

Most companies with Data Science Professionals are Medium Sized companies. Small companies have the least number of Data Science jobs.

Distribution of Company Size

- Small — 13.67%
- Medium — 53.71%
- Large — 32.62%
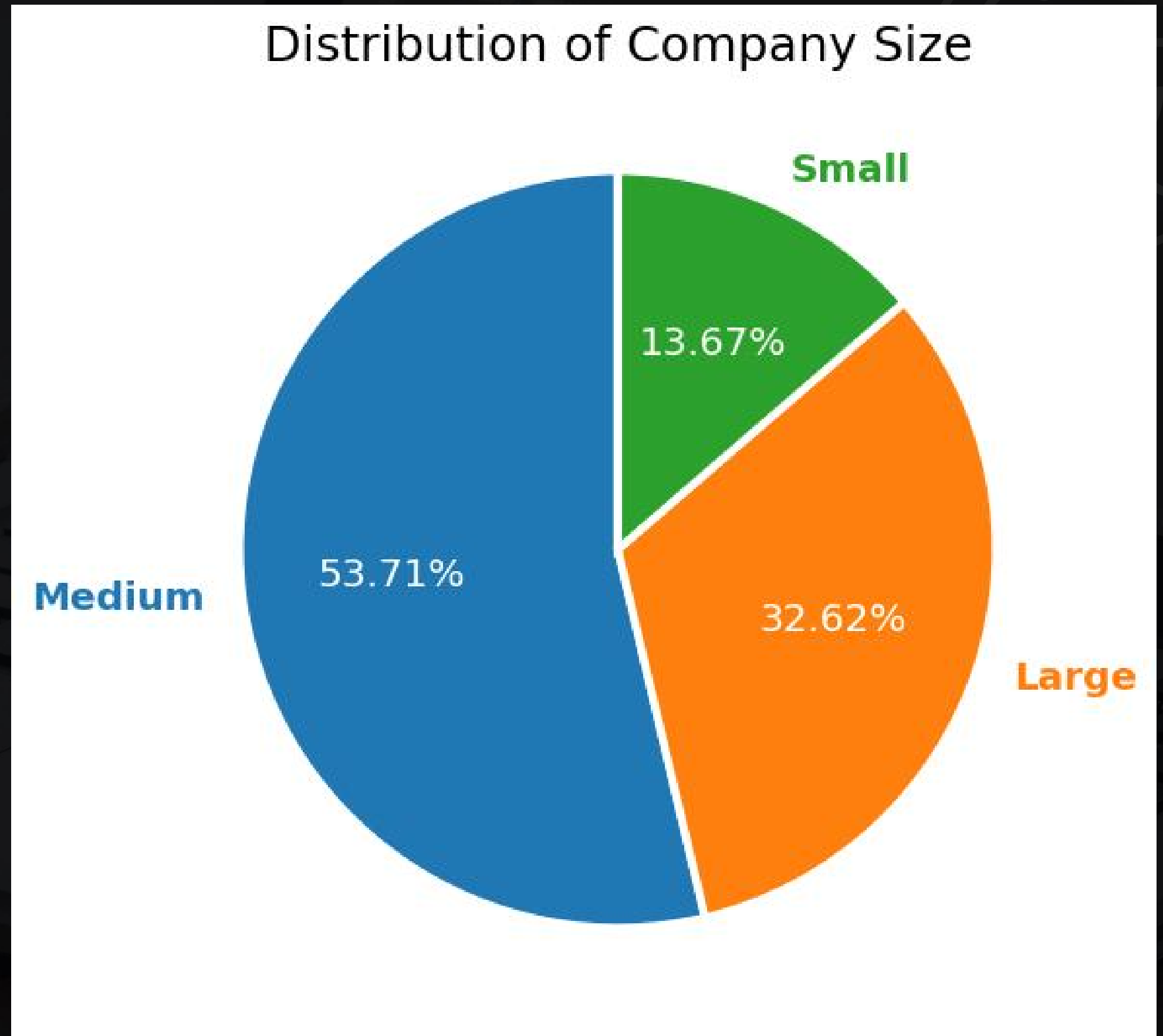
6.Work Year

The Dataset contains values from 2020 - 2022, here we can 2022 has highest Employee compared to other years.

7.Remote ratio by work year

Here the plot show the remote ratio over the year, we can say that in the year almost in 2020 it was no remote work that means working in diffrent place examole work from home,similary partially remote and fully remote in year 2022 the ratio of fully remote was more.
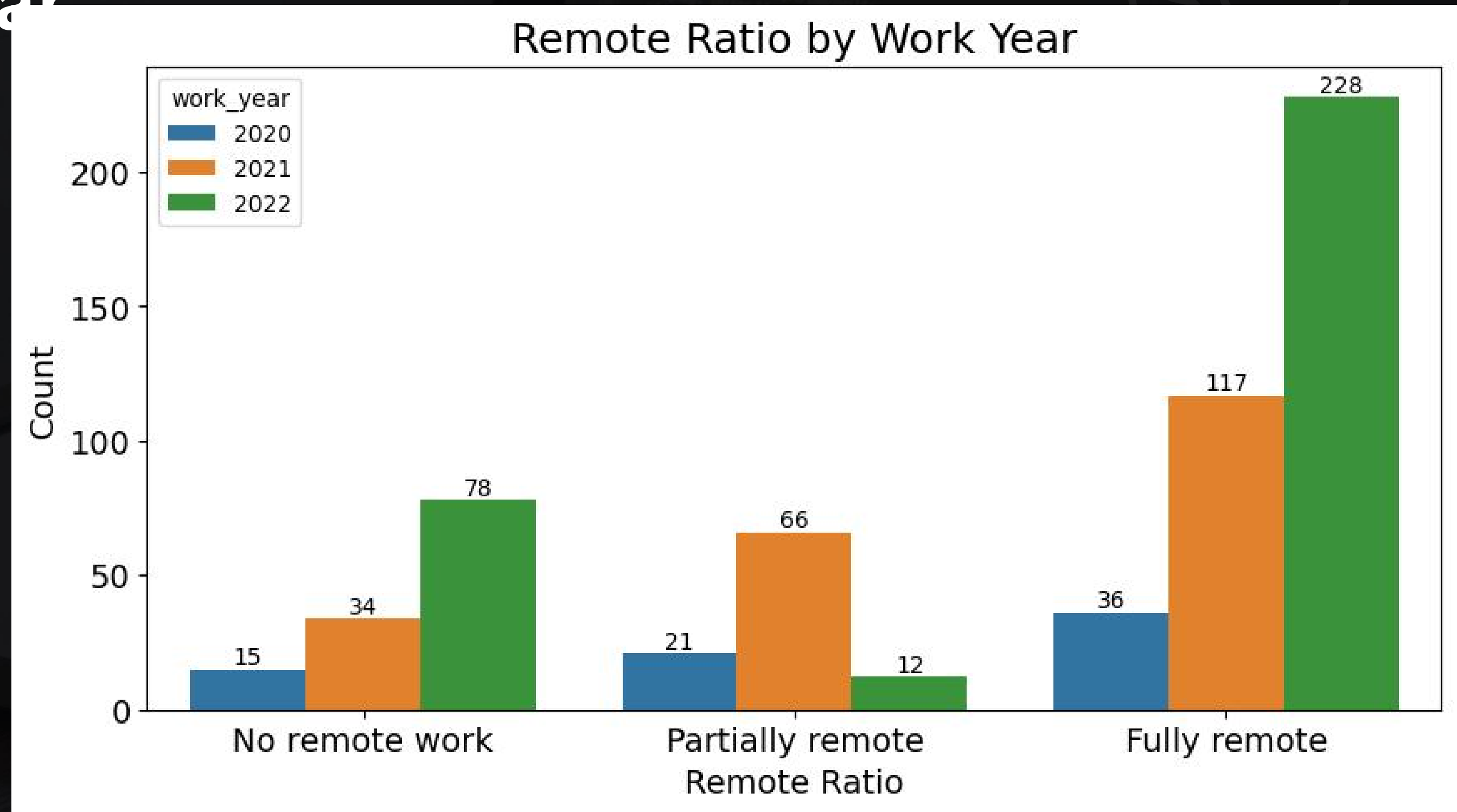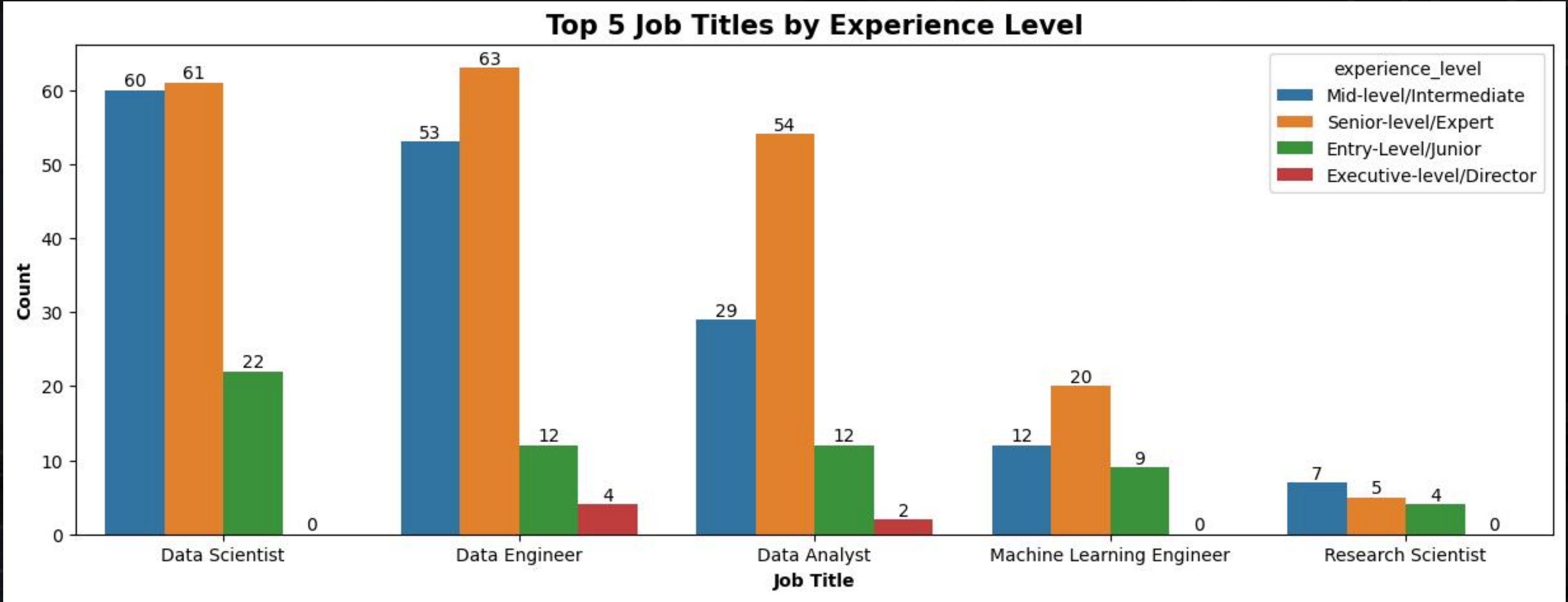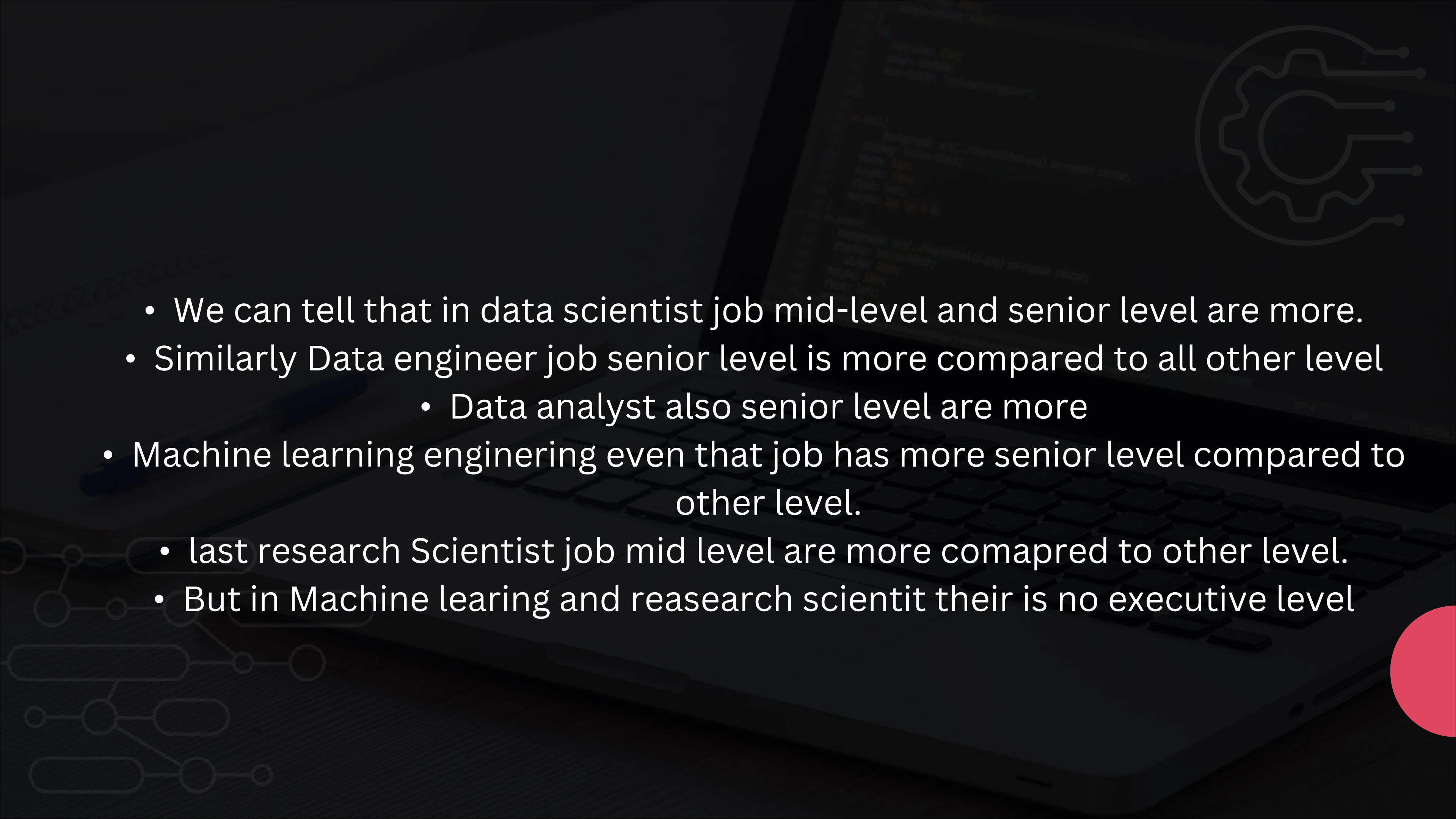
8.Employment Type by Experience Level

The plot shows the Employment type by experience level,here we can tell that full-time employee has all the experience level compared to other level like contract,part-time and freelance.
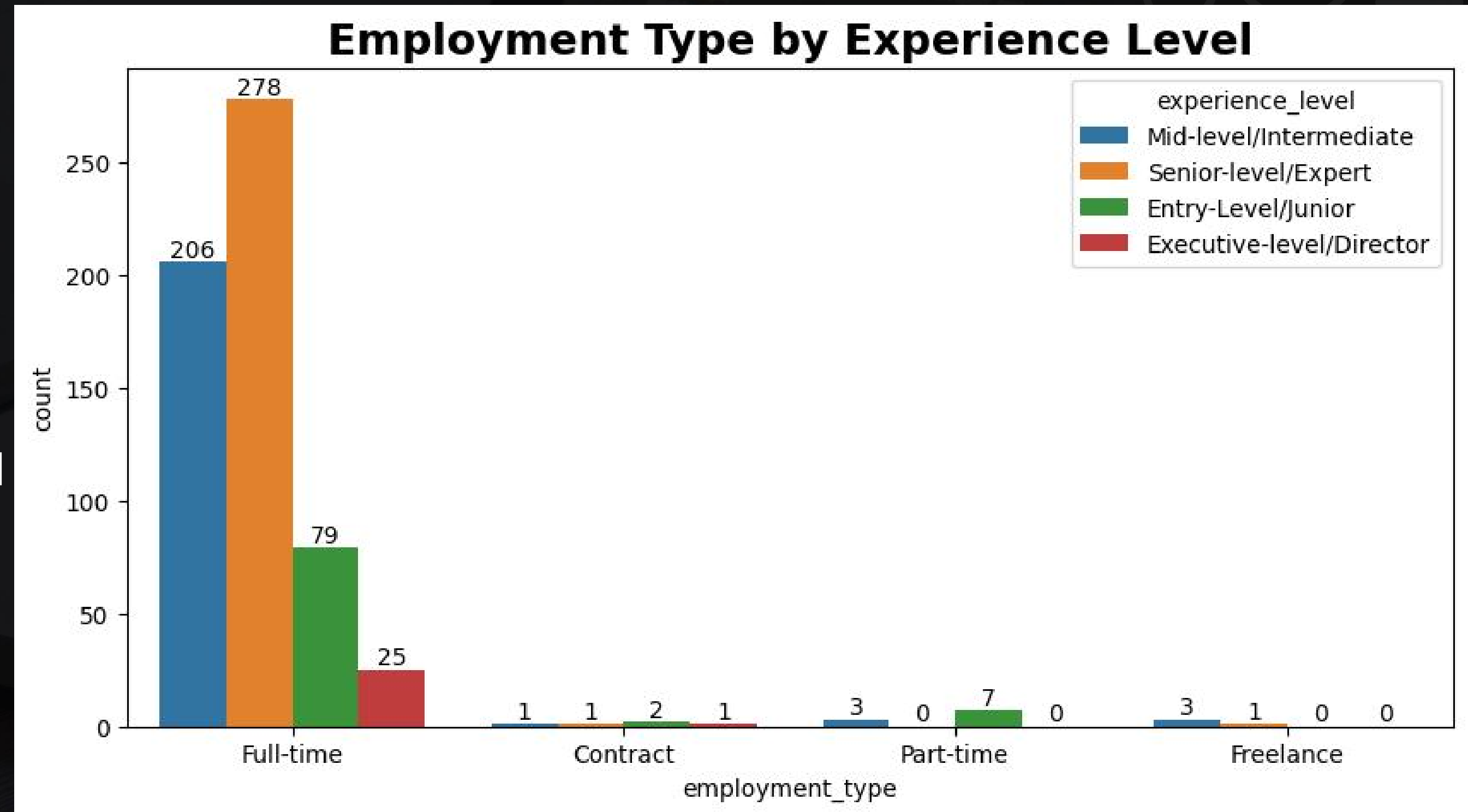
9.Top 5 Job Title by Experience Level

- We can tell that in data scientist job mid-level and senior level are more.
- Similarly Data engineer job senior level is more compared to all other level
- Data analyst also senior level are more
- Machine learning enginering even that job has more senior level compared to other level.
- last research Scientist job mid level are more comapred to other level.
- But in Machine learing and reasearch scientit their is no executive level

10.Distribution of salary by work year
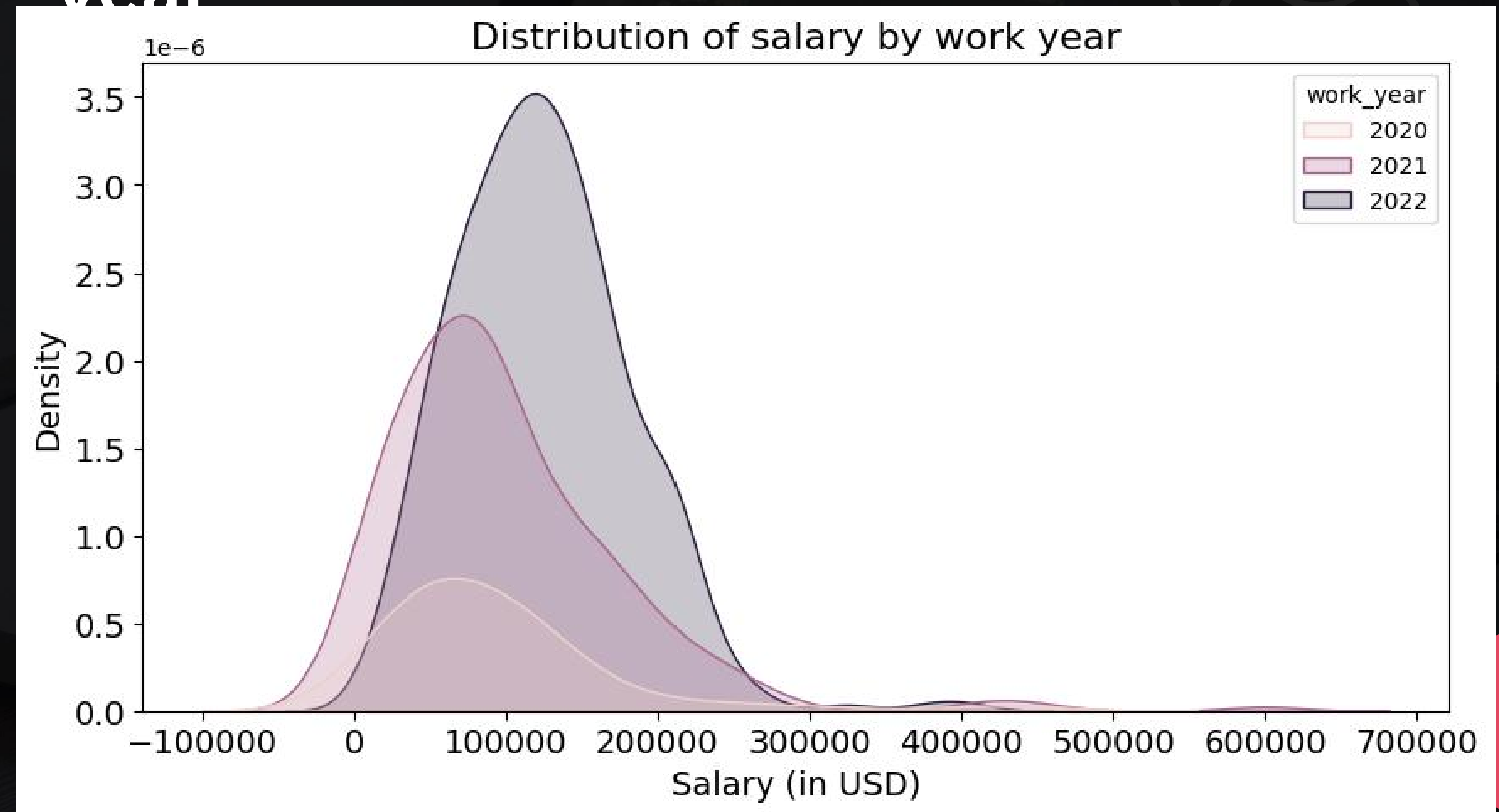
The plot shows that the distribution of salary over the work year ,in 2022 the salary was increased compared to other years.

Distribution of salary by work year
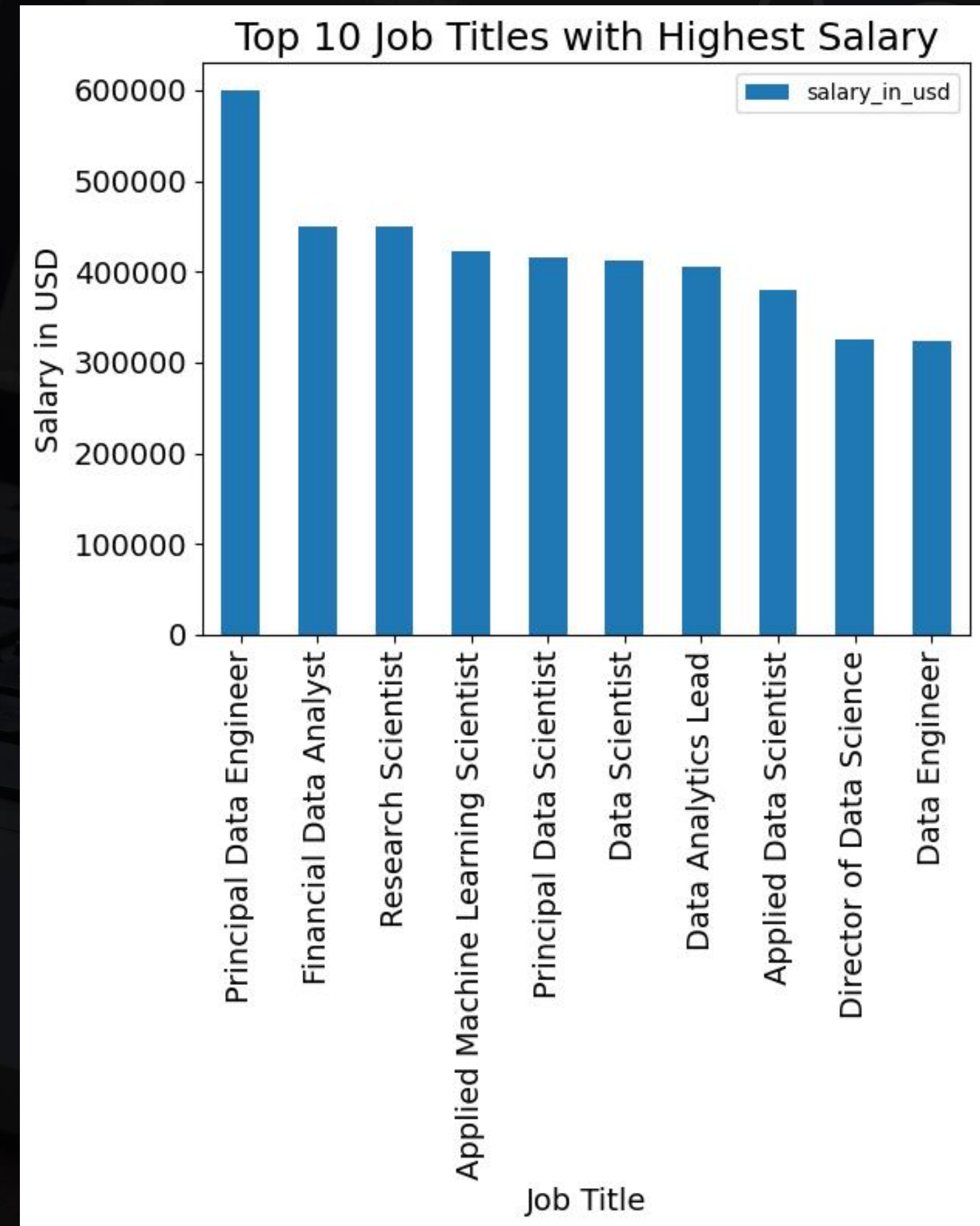
11.Salary by Experience Level

The average salary is higher for a higher experience level. Executive-level/Directors are paid the highest salaries of all the experience levels on average, and Entry-level are the least earners
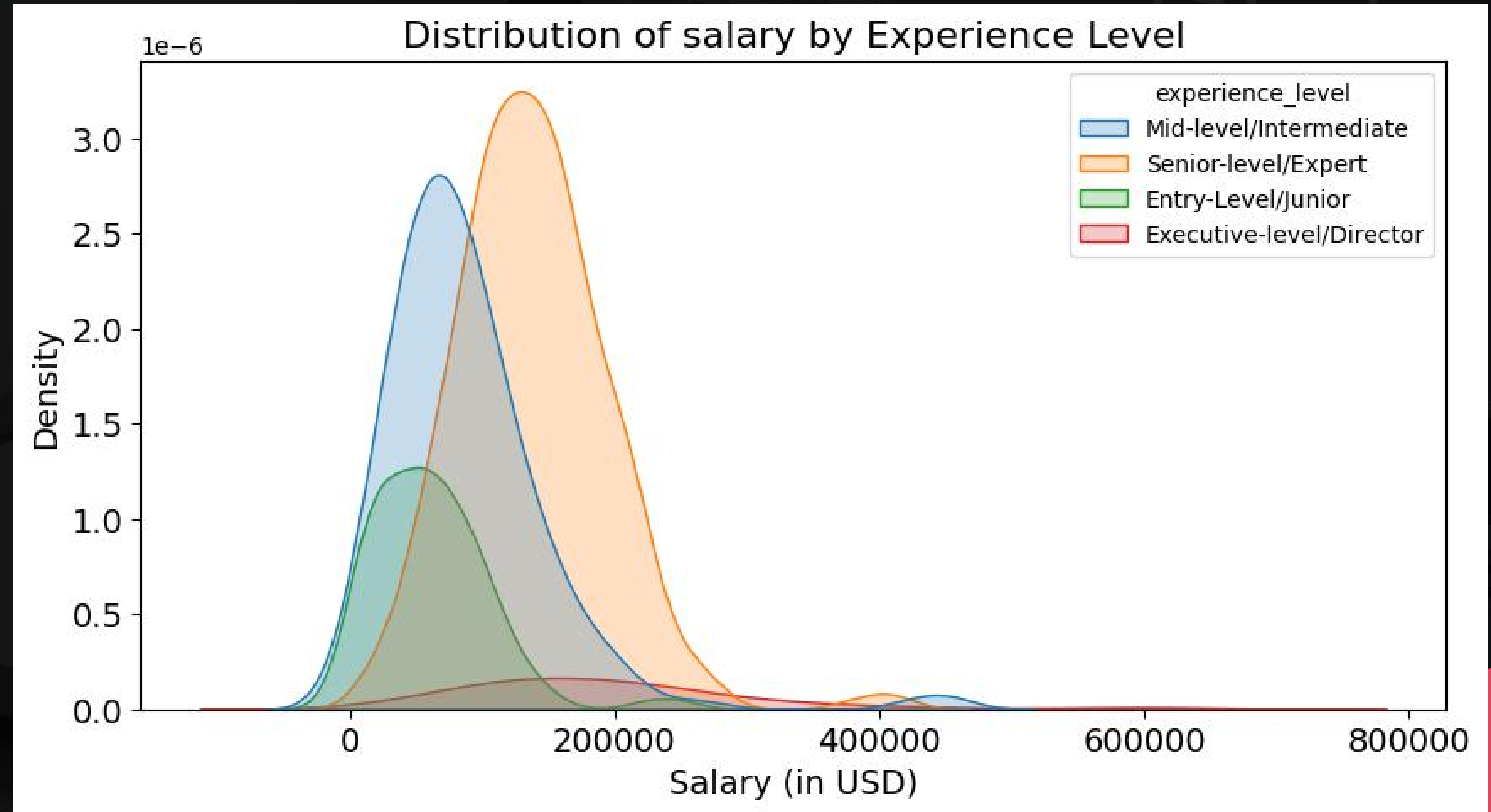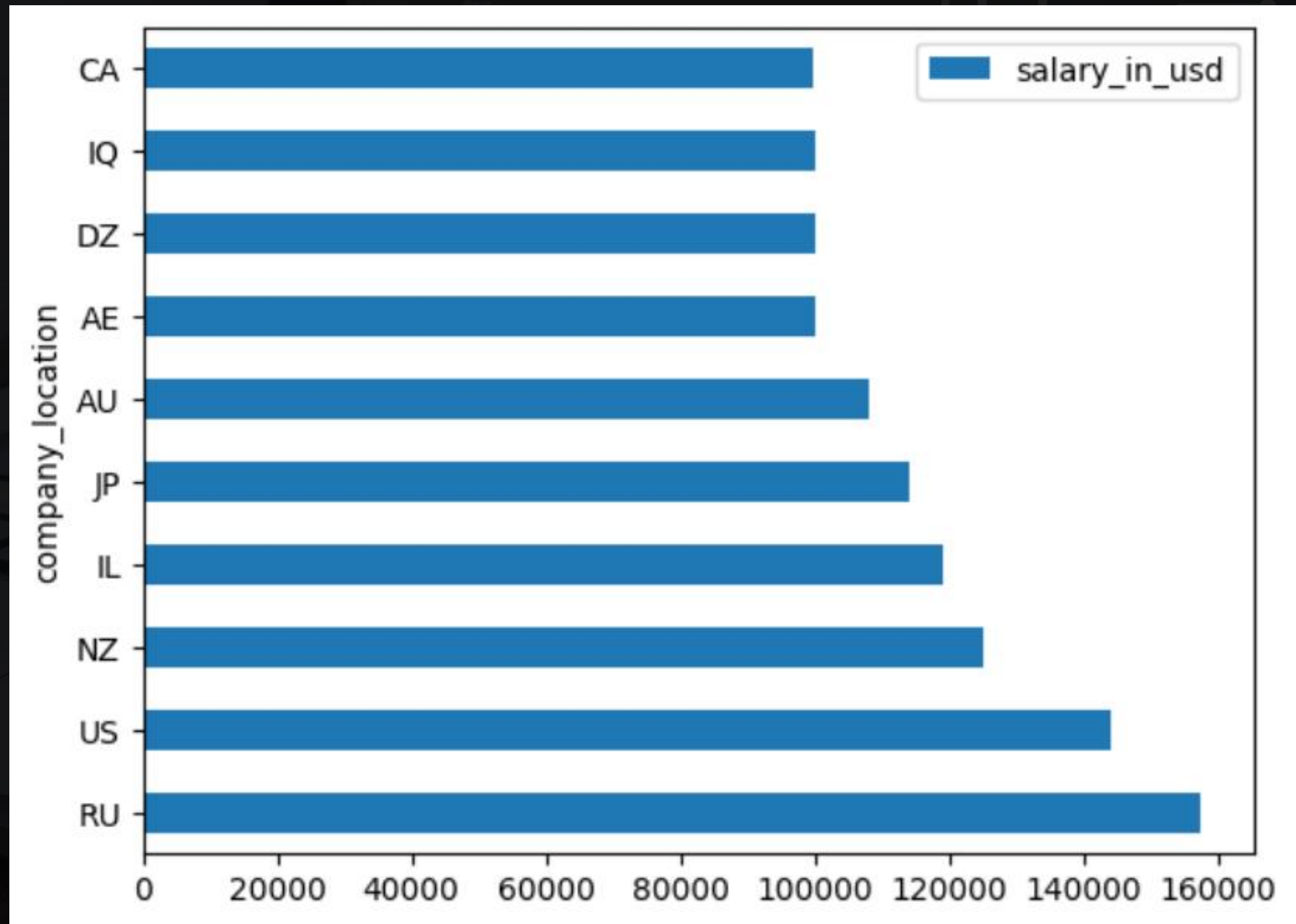
12.Highest salaries by job title

Here the plot shows the highest salary paid for jobs,principal data engineer has hishest salary,than financial data analystand research scientist have same pacakges and others follows
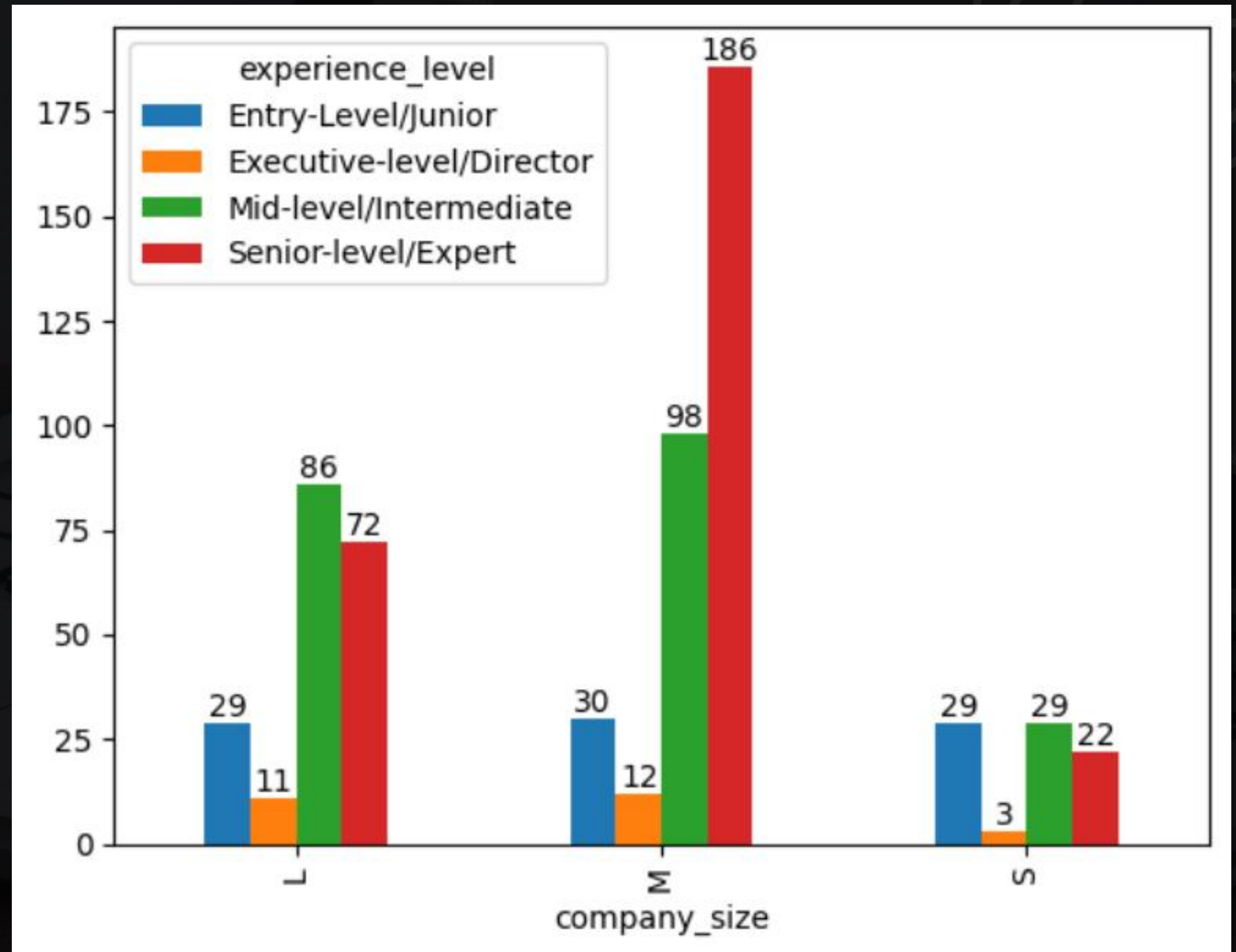
13.Average salary for jobs

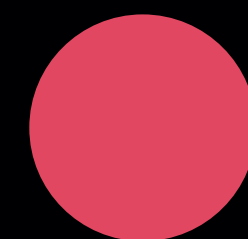**The average salary for Jobs in USD is 100000 to 120000 Usd.**

14.Size of company by Experience level

From the plot we can say that mid level are more in large similarly senior level are more in medium and small level, all experience level are less compared to other level

CONCLUSION

- Most Data Science jobs require Senior-level/Expert experience level and very few positions are available for Executive-level/Directors.
- The most common job titles in the Data Science field are Data Scientist, Data Engineer, Data Analyst and Machine Learning Engineer.
- Most positions are Full-time employment.
- The Number of Data Science jobs and the salary are increasing with each year and experience-level.
- The average salary for Data Science Jobs in USD is 100000usd-120000usd

THANK YOU