EMPLOYEE SALARIES FOR DIFFERENT JOB ROLLS

Load the dataset

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('/content/ds_salaries.csv')
```

Display the first few rows of the dataset

```
df.head(5)
```

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
0	0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	DE	L
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	JP	S
2	2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	GB	M
					Product Data							

Basic information

```
print(df.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
None
```

Descriptive statistics

```
print(df.describe())
```

```
       Unnamed: 0    work_year        salary  salary_in_usd  remote_ratio
count  607.000000   607.000000  6.070000e+02     607.000000     607.00000
mean   303.000000  2021.405272  3.240001e+05  112297.869852      70.92257
std    175.370085     0.692133  1.544357e+06   70957.259411      40.70913
min      0.000000  2020.000000  4.000000e+03    2859.000000       0.00000
25%    151.500000  2021.000000  7.000000e+04   62726.000000      50.00000
50%    303.000000  2022.000000  1.150000e+05  101570.000000     100.00000
75%    454.500000  2022.000000  1.650000e+05  150000.000000     100.00000
max    606.000000  2022.000000  3.040000e+07  600000.000000     100.00000
```

Analyze the distribution of salary by experience level

```
print(df.groupby('experience_level')['salary_in_usd'].mean())
```

```
experience_level
EN     61643.318182
EX    199392.038462
MI     87996.056338
SE    138617.292857
Name: salary_in_usd, dtype: float64
```

Double-click (or enter) to edit

Analyze the distribution of salary by job title

```
print(df.groupby('job_title')['salary_in_usd'].mean())
```

```
job_title
3D Computer Vision Researcher                5409.000000
AI Scientist                                66135.571429
Analytics Engineer                         175000.000000
Applied Data Scientist                     175655.000000
Applied Machine Learning Scientist         142068.750000
BI Data Analyst                             74755.166667
Big Data Architect                          99703.000000
Big Data Engineer                           51974.000000
Business Data Analyst                       76691.200000
Cloud Data Engineer                        124647.000000
Computer Vision Engineer                    44419.333333
Computer Vision Software Engineer          105248.666667
Data Analyst                                92893.061856
Data Analytics Engineer                     64799.250000
Data Analytics Lead                        405000.000000
Data Analytics Manager                     127134.285714
```

```
Data Architect                            177873.909091
Data Engineer                             112725.000000
Data Engineering Manager                  123227.200000
Data Science Consultant                    69420.714286
Data Science Engineer                      75803.333333
Data Science Manager                      158328.500000
Data Scientist                            108187.832168
Data Specialist                           165000.000000
Director of Data Engineering              156738.000000
Director of Data Science                  195074.000000
ETL Developer                              54957.000000
Finance Data Analyst                       61896.000000
Financial Data Analyst                    275000.000000
Head of Data                              160162.600000
Head of Data Science                      146718.750000
Head of Machine Learning                   79039.000000
Lead Data Analyst                          92203.000000
Lead Data Engineer                        139724.500000
Lead Data Scientist                       115190.000000
Lead Machine Learning Engineer             87932.000000
ML Engineer                               117504.000000
Machine Learning Developer                 85860.666667
Machine Learning Engineer                 104880.146341
Machine Learning Infrastructure Engineer  101145.000000
Machine Learning Manager                  117104.000000
Machine Learning Scientist                158412.500000
Marketing Data Analyst                     88654.000000
NLP Engineer                               37236.000000
Principal Data Analyst                    122500.000000
Principal Data Engineer                   328333.333333
Principal Data Scientist                  215242.428571
Product Data Analyst                       13036.000000
Research Scientist                        109019.500000
Staff Data Scientist                      105000.000000
Name: salary_in_usd, dtype: float64
```

Analyze the distribution of salary by company location

```
print(df.groupby('company_location')['salary_in_usd'].mean())
```

```
company_location
AE    100000.000000
AS     18053.000000
AT     72920.750000
AU    108042.666667
BE     85699.000000
BR     18602.666667
CA     99823.733333
CH     64114.000000
CL     40038.000000
CN     71665.500000
CO     21844.000000
CZ     50937.000000
DE     81887.214286
DK     54386.333333
DZ    100000.000000
EE     32974.000000
ES     53060.142857
FR     63970.666667
GB     81583.042553
```

```
GR     52293.090909
HN     20000.000000
HR     45618.000000
HU     35735.000000
IE     71444.000000
IL    119059.000000
IN     28581.750000
IQ    100000.000000
IR      4000.000000
IT     36366.500000
JP    114127.333333
KE      9272.000000
LU     43942.666667
MD     18000.000000
MT     28369.000000
MX     32123.333333
MY     40000.000000
NG     30000.000000
NL     54945.750000
NZ    125000.000000
PK     13333.333333
PL     66082.500000
PT     47793.750000
RO     60000.000000
RU    157500.000000
SG     89294.000000
SI     63831.000000
TR     20096.666667
UA     13400.000000
US    144055.261972
VN      4000.000000
Name: salary_in_usd, dtype: float64
```

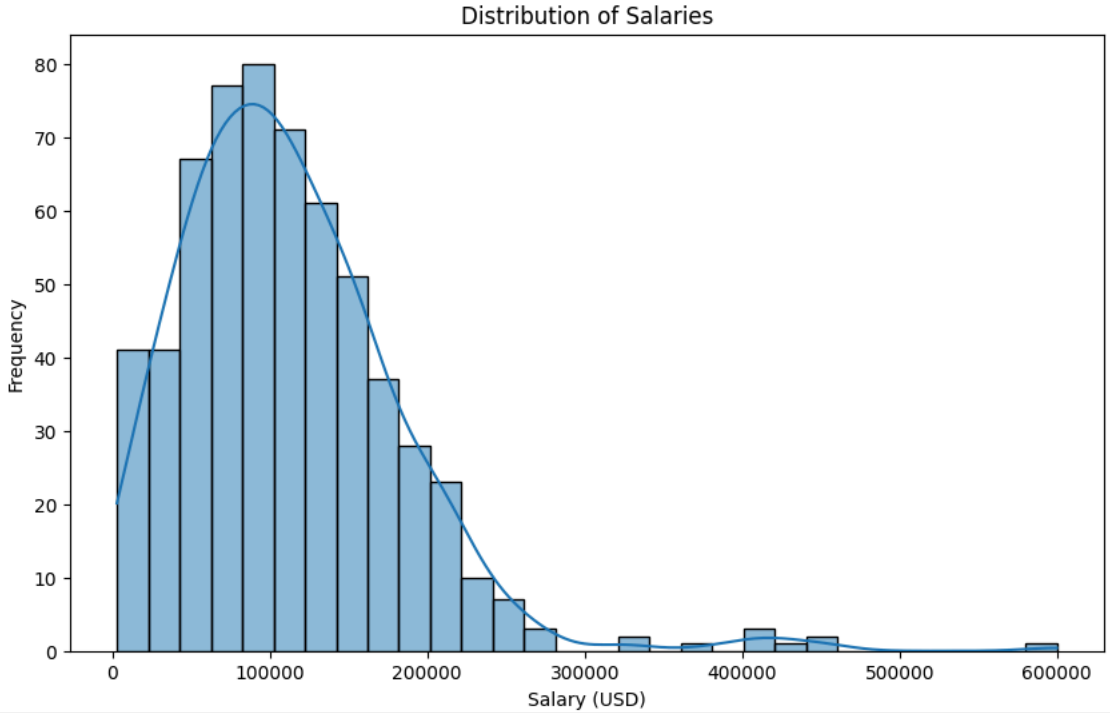
Analyze the relationship between salary and company size

```
print(df.groupby('company_size')['salary_in_usd'].mean())
```

```
company_size
L    119242.994949
M    116905.466258
S     77632.674699
Name: salary_in_usd, dtype: float64
```

Visualize the distribution of salaries

```
plt.figure(figsize=(10, 6))
sns.histplot(df['salary_in_usd'], bins=30, kde=True)
plt.title('Distribution of Salaries')
plt.xlabel('Salary (USD)')
plt.ylabel('Frequency')
plt.show()
```

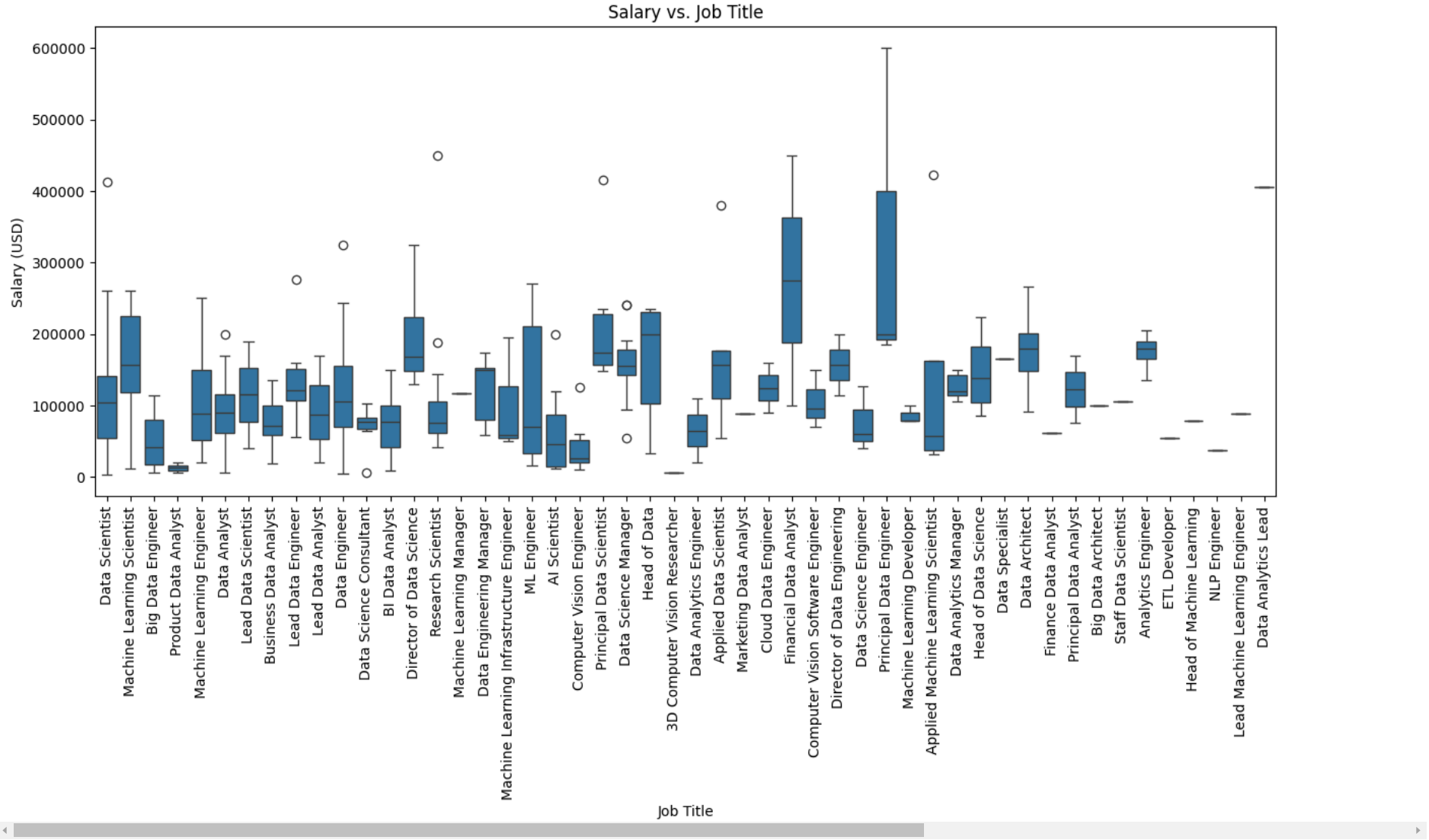


Analyze the relationship between salary and experience level

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='experience_level', y='salary_in_usd', data=df)
plt.title('Salary vs. Experience Level')
plt.xlabel('Experience Level')
plt.ylabel('Salary (USD)')
plt.show()
```

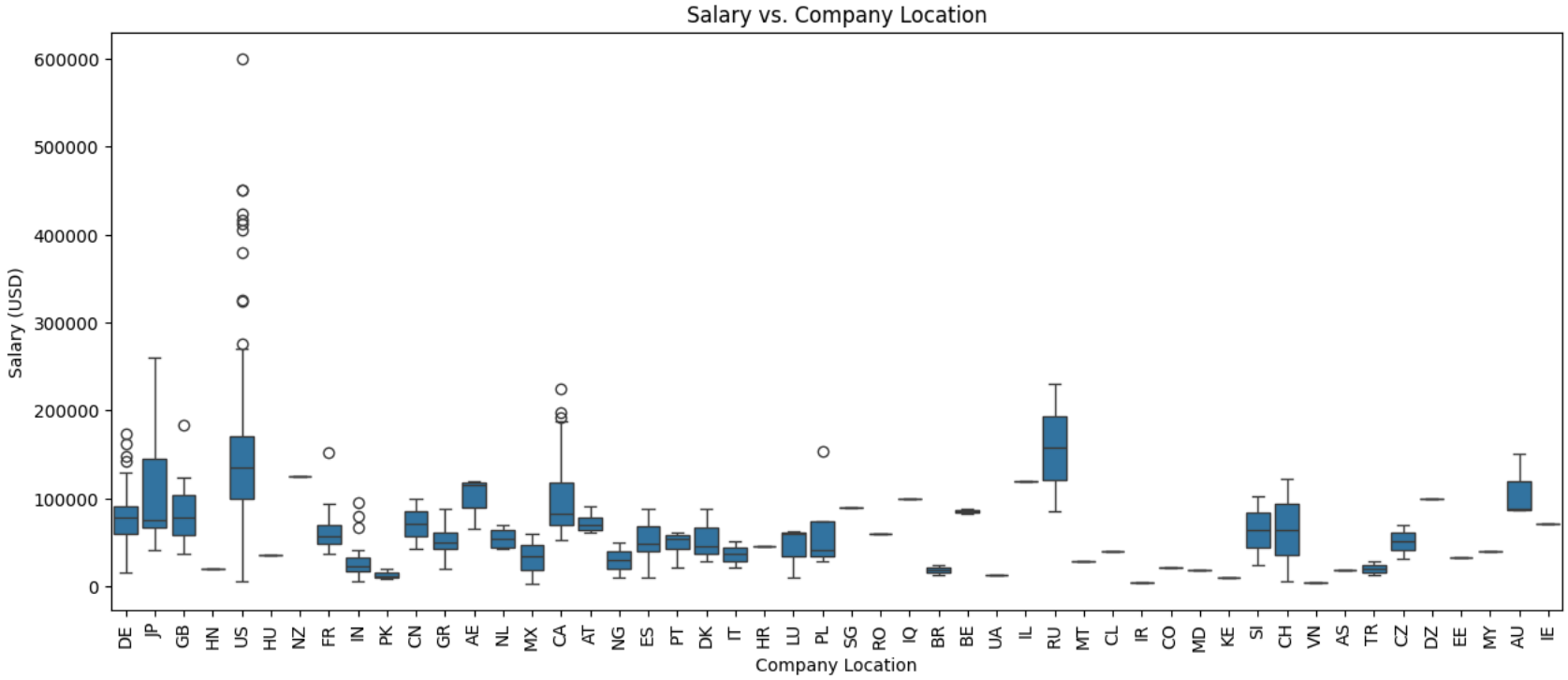
Analyze the relationship between salary and job title

```
plt.figure(figsize=(15, 6))
sns.boxplot(x='job_title', y='salary_in_usd', data=df)
plt.title('Salary vs. Job Title')
plt.xlabel('Job Title')
plt.ylabel('Salary (USD)')
plt.xticks(rotation=90)
plt.show()
```



Analyze the relationship between salary and company location

```
plt.figure(figsize=(15, 6))
sns.boxplot(x='company_location', y='salary_in_usd', data=df)
plt.title('Salary vs. Company Location')
plt.xlabel('Company Location')
plt.ylabel('Salary (USD)')
plt.xticks(rotation=90)
plt.show()
```

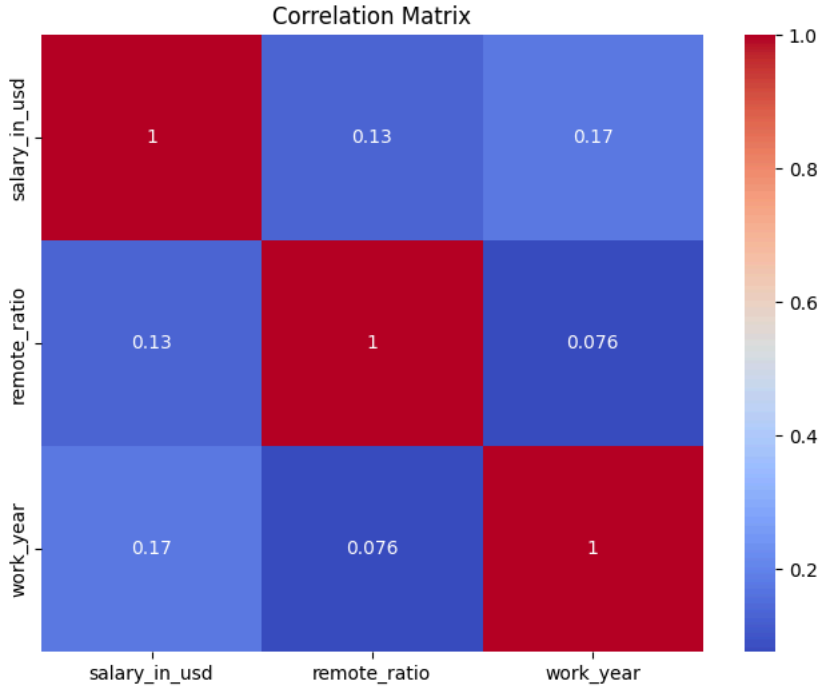


Analyze the relationship between salary and company size

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='company_size', y='salary_in_usd', data=df)
plt.title('Salary vs. Company Size')
plt.xlabel('Company Size')
plt.ylabel('Salary (USD)')
plt.show()
```

Correlation Analysis

```
correlation_matrix = df[['salary_in_usd', 'remote_ratio', 'work_year']].corr()
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



Analyzing the distribution of salaries within specific job roles

```
job_title_salary_distribution = df.groupby('job_title')['salary_in_usd'].describe()
print(job_title_salary_distribution)

plt.figure(figsize=(15, 6))
sns.boxplot(x='job_title', y='salary_in_usd', data=df)
plt.title('Salary Distribution by Job Role')
plt.xlabel('Job Title')
plt.ylabel('Salary (USD)')
plt.xticks(rotation=90)
plt.show()
```

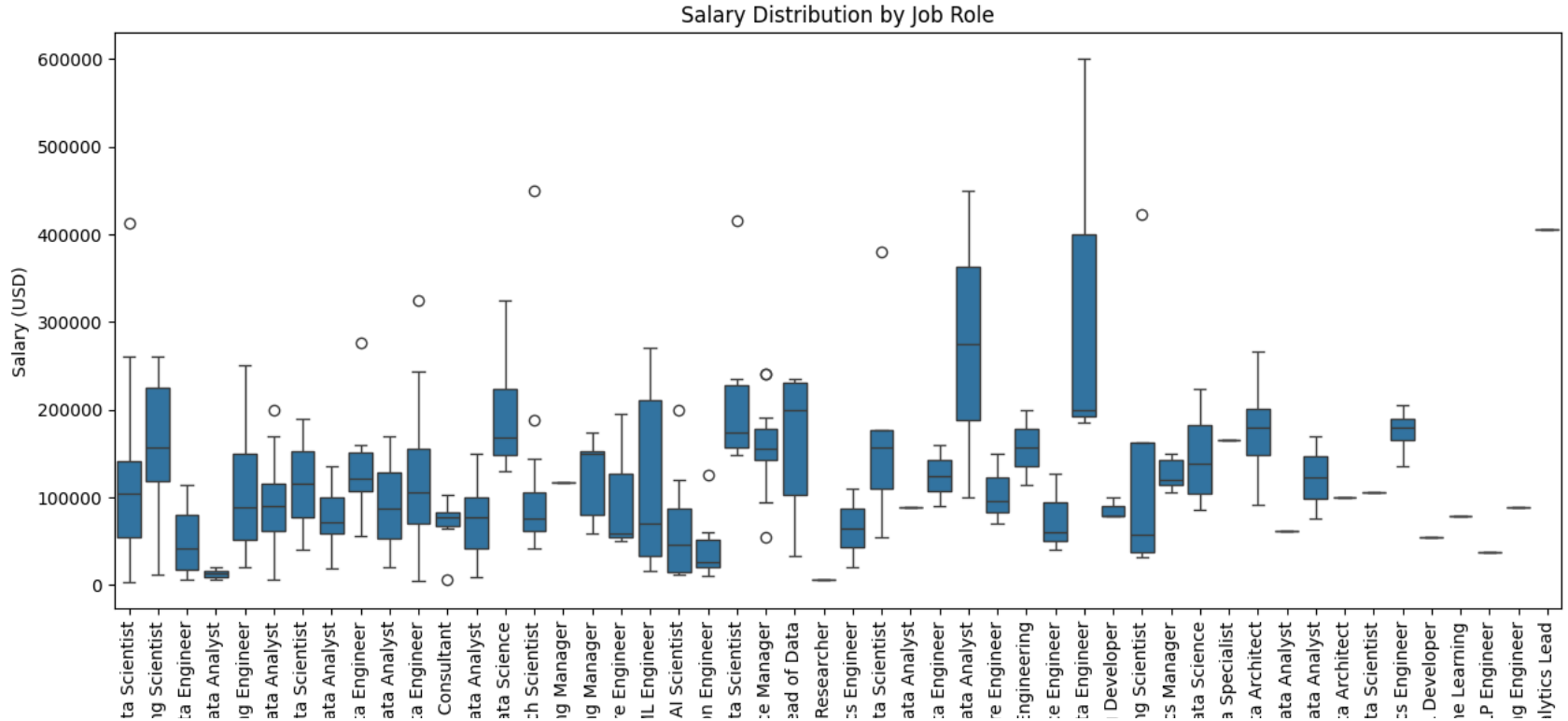
```
                                          count           mean            std  \
job_title
3D Computer Vision Researcher               1.0    5409.000000            NaN
AI Scientist                                7.0   66135.571429   70152.164338
Analytics Engineer                          4.0  175000.000000   29508.077990
Applied Data Scientist                      5.0  175655.000000  123647.186470
Applied Machine Learning Scientist          4.0  142068.750000  188246.696933
BI Data Analyst                             6.0   74755.166667   50989.558548
Big Data Architect                          1.0   99703.000000            NaN
Big Data Engineer                           8.0   51974.000000   42954.761263
Business Data Analyst                       5.0   76691.200000   43814.851149
Cloud Data Engineer                         2.0  124647.000000   49996.692071
Computer Vision Engineer                    6.0   44419.333333   42984.849627
Computer Vision Software Engineer           3.0  105248.666667   40837.795060
Data Analyst                               97.0   92893.061856   39961.075848
Data Analytics Engineer                     4.0   64799.250000   38628.114143
Data Analytics Lead                         1.0  405000.000000            NaN
Data Analytics Manager                      7.0  127134.285714   17853.383299
Data Architect                             11.0  177873.909091   45714.721424
Data Engineer                             132.0  112725.000000   57629.082107
Data Engineering Manager                    5.0  123227.200000   50374.695033
Data Science Consultant                     7.0   69420.714286   30961.654417
Data Science Engineer                       3.0   75803.333333   45617.444737
Data Science Manager                       12.0  158328.500000   52589.180497
Data Scientist                            143.0  108187.832168   64112.840519
Data Specialist                             1.0  165000.000000            NaN
Director of Data Engineering                2.0  156738.000000   61181.707135
Director of Data Science                    7.0  195074.000000   70015.787970
ETL Developer                               2.0   54957.000000       0.000000
Finance Data Analyst                        1.0   61896.000000            NaN
Financial Data Analyst                      2.0  275000.000000  247487.373415
Head of Data                                5.0  160162.600000   88780.420706
Head of Data Science                        4.0  146718.750000   62122.055178
Head of Machine Learning                    1.0   79039.000000            NaN
Lead Data Analyst                           3.0   92203.000000   75330.383160
Lead Data Engineer                          6.0  139724.500000   74845.891227
Lead Data Scientist                         3.0  115190.000000   74715.181188
Lead Machine Learning Engineer              1.0   87932.000000            NaN
ML Engineer                                 6.0  117504.000000  115217.451253
Machine Learning Developer                  3.0   85860.666667   12245.021859
Machine Learning Engineer                  41.0  104880.146341   63218.626439
Machine Learning Infrastructure Engineer    3.0  101145.000000   81381.030806
Machine Learning Manager                    1.0  117104.000000            NaN
Machine Learning Scientist                  8.0  158412.500000   79514.714496
Marketing Data Analyst                      1.0   88654.000000            NaN
NLP Engineer                                1.0   37236.000000            NaN
Principal Data Analyst                      2.0  122500.000000   67175.144213
Principal Data Engineer                     3.0  328333.333333  235389.747724
Principal Data Scientist                    7.0  215242.428571   94705.038268
Product Data Analyst                        2.0   13036.000000    9848.583248
Research Scientist                         16.0  109019.500000   98542.763968
Staff Data Scientist                        1.0  105000.000000            NaN

                                               min        25%       50%  \
job_title
3D Computer Vision Researcher               5409.0    5409.00    5409.0
AI Scientist                               12000.0   15026.50   45896.0
Analytics Engineer                        135000.0  165000.00  179850.0
Applied Data Scientist                     54238.0  110037.00  157000.0
Applied Machine Learning Scientist         31875.0   36768.75   56700.0
BI Data Analyst                             9272.0   40944.25   76500.0
Big Data Architect                         99703.0   99703.00   99703.0
Big Data Engineer                           5882.0   17557.00   41305.5
Business Data Analyst                      18442.0   59102.00   70912.0
```

job_title			
Business Data Analyst	18442.0	99162.00	70912.0
Cloud Data Engineer	89294.0	106970.50	124647.0
Computer Vision Engineer	10000.0	20180.25	26304.5
Computer Vision Software Engineer	70000.0	82873.00	95746.0
Data Analyst	6072.0	62000.00	90320.0
Data Analytics Engineer	20000.0	42500.00	64598.5
Data Analytics Lead	405000.0	405000.00	405000.0
Data Analytics Manager	105400.0	114640.00	120000.0
Data Architect	90700.0	148900.00	180000.0
Data Engineer	4000.0	70039.50	105500.0
Data Engineering Manager	59303.0	79833.00	150000.0
Data Science Consultant	5707.0	66786.00	76833.0
Data Science Engineer	40189.0	50094.50	60000.0
Data Science Manager	54094.0	142285.25	155750.0
Data Scientist	2859.0	54724.00	103691.0
Data Specialist	165000.0	165000.00	165000.0
Director of Data Engineering	113476.0	135107.00	156738.0
Director of Data Science	130026.0	147756.50	168000.0
ETL Developer	54957.0	54957.00	54957.0
Finance Data Analyst	61896.0	61896.00	61896.0
Financial Data Analyst	100000.0	187500.00	275000.0
Head of Data	32974.0	102839.00	200000.0
Head of Data Science	85000.0	103750.00	138937.5
Head of Machine Learning	79039.0	79039.00	79039.0
Lead Data Analyst	19609.0	53304.50	87000.0
Lead Data Engineer	56000.0	106916.75	121593.5
Lead Data Scientist	40570.0	77785.00	115000.0
Lead Machine Learning Engineer	87932.0	87932.00	87932.0
ML Engineer	15966.0	32415.00	70537.5
Machine Learning Developer	78791.0	78791.00	78791.0
Machine Learning Engineer	20000.0	51064.00	87932.0
Machine Learning Infrastructure Engineer	50180.0	54217.50	58255.0
Machine Learning Manager	117104.0	117104.00	117104.0
Machine Learning Scientist	12000.0	118075.00	156500.0
Marketing Data Analyst	88654.0	88654.00	88654.0
NLP Engineer	37236.0	37236.00	37236.0
Principal Data Analyst	75000.0	98750.00	122500.0
Principal Data Engineer	185000.0	192500.00	200000.0
Principal Data Scientist	148261.0	156837.00	173762.0
Product Data Analyst	6072.0	9554.00	13036.0
Research Scientist	42000.0	62176.00	76263.5
Staff Data Scientist	105000.0	105000.00	105000.0

job_title	75%	max
3D Computer Vision Researcher	5409.00	5409.0
AI Scientist	87500.00	200000.0
Analytics Engineer	189850.00	205300.0
Applied Data Scientist	177000.00	380000.0
Applied Machine Learning Scientist	162000.00	423000.0
BI Data Analyst	99500.00	150000.0
Big Data Architect	99703.00	99703.0
Big Data Engineer	79756.00	114047.0
Business Data Analyst	100000.00	135000.0
Cloud Data Engineer	142323.50	160000.0
Computer Vision Engineer	52152.25	125000.0
Computer Vision Software Engineer	122873.00	150000.0
Data Analyst	116150.00	200000.0
Data Analytics Engineer	86897.75	110000.0
Data Analytics Lead	405000.00	405000.0
Data Analytics Manager	142500.00	150260.0
Data Architect	200669.50	266400.0
Data Engineer	154600.00	324000.0
Data Engineering Manager	153000.00	174000.0
Data Science Consultant	83416.50	103000.0

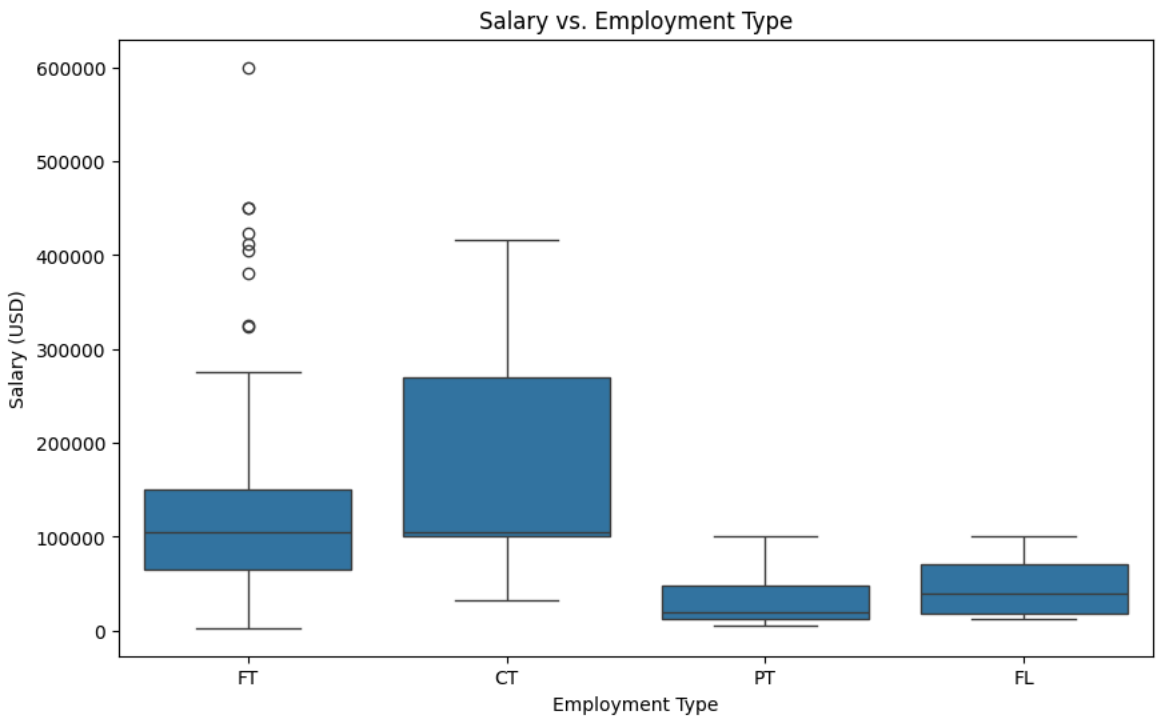
Data Science Engineer	93610.50	127221.0
Data Science Manager	178050.00	241000.0
Data Scientist	140850.00	412000.0
Data Specialist	165000.00	165000.0
Director of Data Engineering	178369.00	200000.0
Director of Data Science	223489.50	325000.0
ETL Developer	54957.00	54957.0
Finance Data Analyst	61896.00	61896.0
Financial Data Analyst	362500.00	450000.0
Head of Data	230000.00	235000.0
Head of Data Science	181906.25	224000.0
Head of Machine Learning	79039.00	79039.0
Lead Data Analyst	128500.00	170000.0
Lead Data Engineer	151250.00	276000.0
Lead Data Scientist	152500.00	190000.0
Lead Machine Learning Engineer	87932.00	87932.0
ML Engineer	211341.00	270000.0
Machine Learning Developer	89395.50	100000.0
Machine Learning Engineer	150000.00	250000.0
Machine Learning Infrastructure Engineer	126627.50	195000.0
Machine Learning Manager	117104.00	117104.0
Machine Learning Scientist	225000.00	260000.0
Marketing Data Analyst	88654.00	88654.0
NLP Engineer	37236.00	37236.0
Principal Data Analyst	146250.00	170000.0
Principal Data Engineer	400000.00	600000.0
Principal Data Scientist	227500.00	416000.0
Product Data Analyst	16518.00	20000.0
Research Scientist	105000.00	450000.0
Staff Data Scientist	105000.00	105000.0



Da
Machine Learnir
Big Dal
Product D
Machine Learnir
D
Lead Da
Business D
Lead Dal
Lead D
Dal
Data Science
BI D
Director of Di
Researc
Machine Learnir
Data Engineerir
Machine Learning Infrastructu
N
Computer Visic
Principal Da
Data Scienc
H
3D Computer Vision
Data Analytic
Applied Da
Marketing D
Cloud Dal
Financial D
Computer Vision Softwa
Director of Data E
Data Scienc
Principal Dal
Machine Learning
Applied Machine Learnir
Data Analytic
Head of Di
Dat
Dal
Finance D
Principal D
Big Dal
Staff Da
Analytic
ETL
Head of Machir
NL
Lead Machine Learnir
Data Ana
Job Title

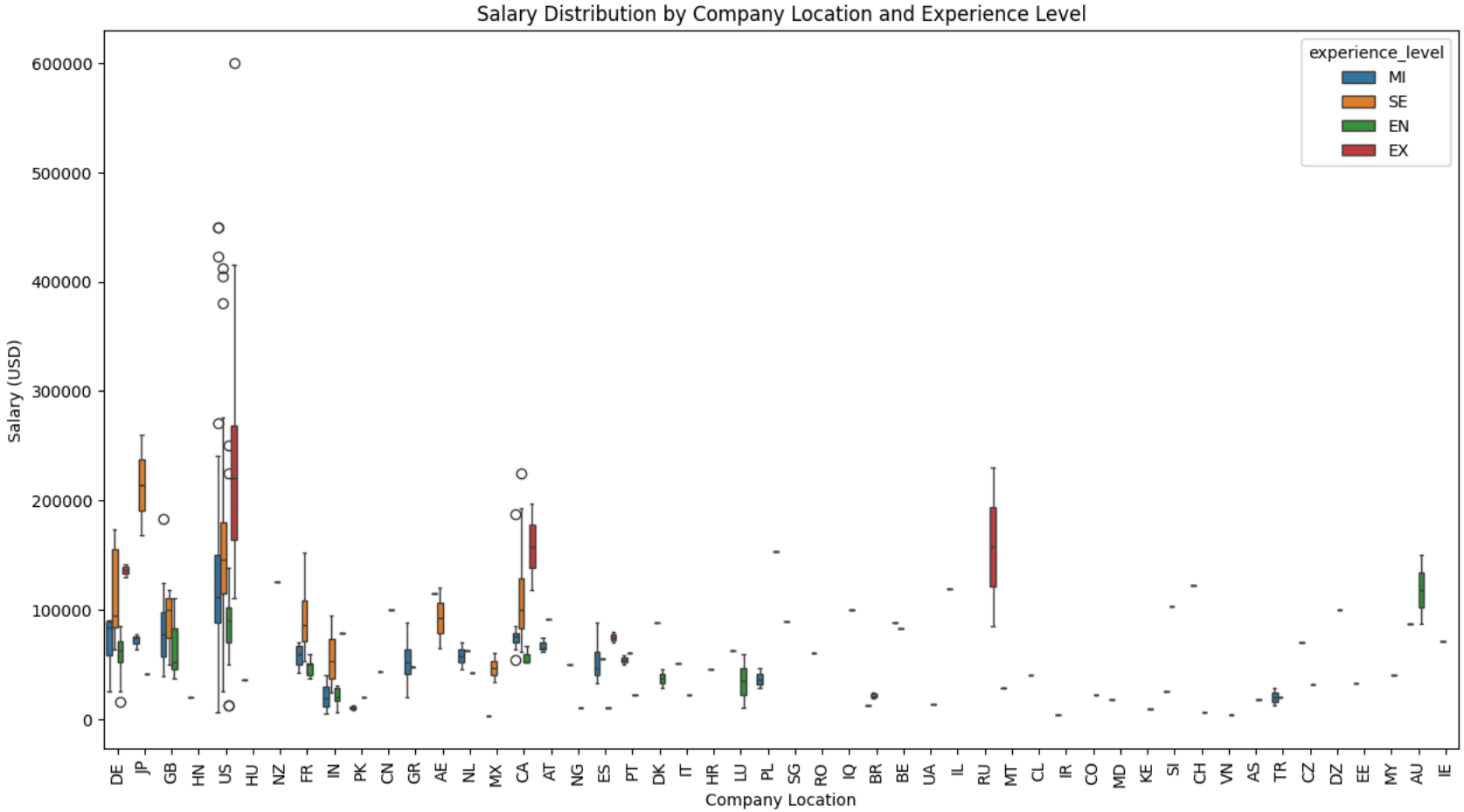
Analyze the relationship between salary and employment type

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='employment_type', y='salary_in_usd', data=df)
plt.title('Salary vs. Employment Type')
plt.xlabel('Employment Type')
plt.ylabel('Salary (USD)')
plt.show()
```



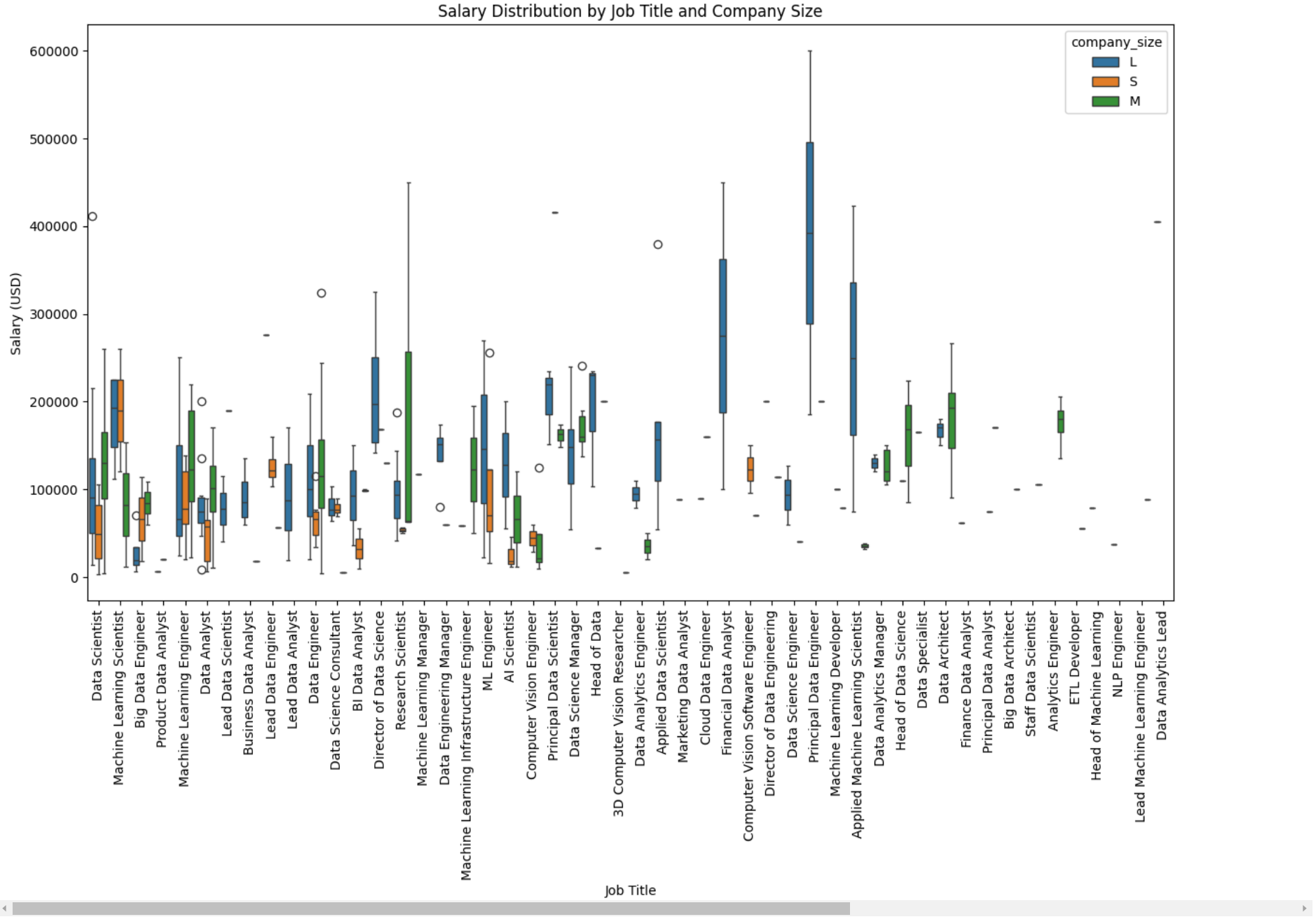
Analyze the distribution of salaries across different company locations and experience levels

```
plt.figure(figsize=(15, 8))
sns.boxplot(x='company_location', y='salary_in_usd', hue='experience_level', data=df)
plt.title('Salary Distribution by Company Location and Experience Level')
plt.xlabel('Company Location')
plt.ylabel('Salary (USD)')
plt.xticks(rotation=90)
plt.show()
```



Analyze the distribution of salaries for different job titles and company sizes

```
plt.figure(figsize=(15, 8))
sns.boxplot(x='job_title', y='salary_in_usd', hue='company_size', data=df)
plt.title('Salary Distribution by Job Title and Company Size')
plt.xlabel('Job Title')
plt.ylabel('Salary (USD)')
plt.xticks(rotation=90)
plt.show()
```



Analyze the relationship between salary and remote ratio