```
In [1]: #importing lib
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv("E:/Data Analytics Project/New folder (5)/New folder (4)/ds_salaries.csv")
print(data)
```

```
     Unnamed: 0  work_year experience_level employment_type  \
0             0       2020               MI              FT
1             1       2020               SE               FT
2             2       2020               SE               FT
3             3       2020               MI              FT
4             4       2020               SE               FT
..          ...        ...              ...              ...
602         602       2022               SE               FT
603         603       2022               SE               FT
604         604       2022               SE               FT
605         605       2022               SE               FT
606         606       2022               MI              FT

                    job_title  salary salary_currency  salary_in_usd  \
0              Data Scientist   70000             EUR          79833
1    Machine Learning Scientist  260000             USD         260000
2             Big Data Engineer   85000             GBP         109024
3           Product Data Analyst   20000             USD          20000
4     Machine Learning Engineer  150000             USD         150000
..                        ...     ...             ...            ...
602              Data Engineer  154000             USD         154000
603              Data Engineer  126000             USD         126000
604               Data Analyst  129000             USD         129000
605               Data Analyst  150000             USD         150000
606               AI Scientist  200000             USD         200000

     employee_residence  remote_ratio company_location company_size
0                    DE             0               DE            L
1                    JP             0               JP            S
2                    GB            50               GB            M
3                    HN             0               HN            S
4                    US            50               US            L
..                  ...           ...              ...          ...
602                  US           100               US            M
603                  US           100               US            M
604                  US             0               US            M
605                  US           100               US            M
606                  IN           100               US            L

[607 rows x 12 columns]
```

```
In [2]: data.head()
```

Out[2]:

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residenc
0	0	2020	MI	FT	Data Scientist	70000	EUR	79833	DI
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JI
2	2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GI
3	3	2020	MI	FT	Product Data Analyst	20000	USD	20000	HI
4	4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	U!

```
In [3]: data.dtypes
```

```
Out[3]:  Unnamed: 0              int64
         work_year              int64
         experience_level       object
         employment_type        object
         job_title              object
         salary                 int64
         salary_currency        object
         salary_in_usd          int64
         employee_residence     object
         remote_ratio           int64
         company_location       object
         company_size           object
         dtype: object
```

In [4]: `data.describe()`

Out[4]:

	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

In [5]: `print(data.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
```

In [10]:
```python
df = pd.DataFrame(data)
# print(df)
gp = data.groupby("job_title").agg({"salary_in_usd":"count"})
print(gp)
```

```
                                   salary_in_usd
job_title
3D Computer Vision Researcher                  1
AI Scientist                                   7
Analytics Engineer                             4
Applied Data Scientist                         5
Applied Machine Learning Scientist             4
BI Data Analyst                                6
Big Data Architect                             1
Big Data Engineer                              8
Business Data Analyst                          5
Cloud Data Engineer                            2
Computer Vision Engineer                       6
Computer Vision Software Engineer              3
Data Analyst                                  97
Data Analytics Engineer                        4
Data Analytics Lead                            1
Data Analytics Manager                         7
Data Architect                                11
Data Engineer                                132
Data Engineering Manager                       5
Data Science Consultant                        7
Data Science Engineer                          3
Data Science Manager                          12
Data Scientist                               143
Data Specialist                                1
Director of Data Engineering                   2
Director of Data Science                       7
ETL Developer                                  2
Finance Data Analyst                           1
Financial Data Analyst                         2
Head of Data                                   5
Head of Data Science                           4
Head of Machine Learning                       1
Lead Data Analyst                              3
Lead Data Engineer                             6
Lead Data Scientist                            3
Lead Machine Learning Engineer                 1
ML Engineer                                    6
Machine Learning Developer                     3
Machine Learning Engineer                     41
Machine Learning Infrastructure Engineer       3
Machine Learning Manager                       1
Machine Learning Scientist                     8
Marketing Data Analyst                         1
NLP Engineer                                   1
Principal Data Analyst                         2
Principal Data Engineer                        3
Principal Data Scientist                       7
Product Data Analyst                           2
Research Scientist                            16
Staff Data Scientist                           1
```

In [15]:
```python
print(pd.melt(df, id_vars = ["job_title"], value_vars =["experience_level","company_size"], var_name = "experien
```

```
                      job_title  experience_level  values
0                Data Scientist  experience_level      MI
1     Machine Learning Scientist  experience_level      SE
2             Big Data Engineer  experience_level      SE
3           Product Data Analyst  experience_level      MI
4      Machine Learning Engineer  experience_level      SE
...                          ...               ...     ...
1209              Data Engineer       company_size       M
1210              Data Engineer       company_size       M
1211               Data Analyst       company_size       M
1212               Data Analyst       company_size       M
1213               AI Scientist       company_size       L

[1214 rows x 3 columns]
```
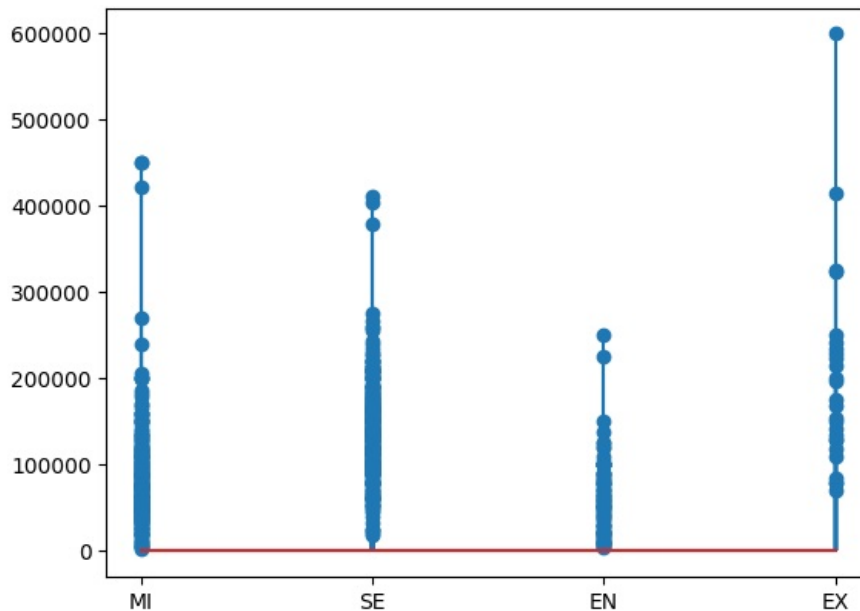
In [191...
```python
a=pd.melt(df, id_vars = ["job_title"], value_vars =["experience_level","company_size"], var_name = "experience_
plt.figure(figsize=(10, 7))
plt.hist(a, orientation = 'horizontal' )
plt.savefig("hist.png", transparent = True)
```
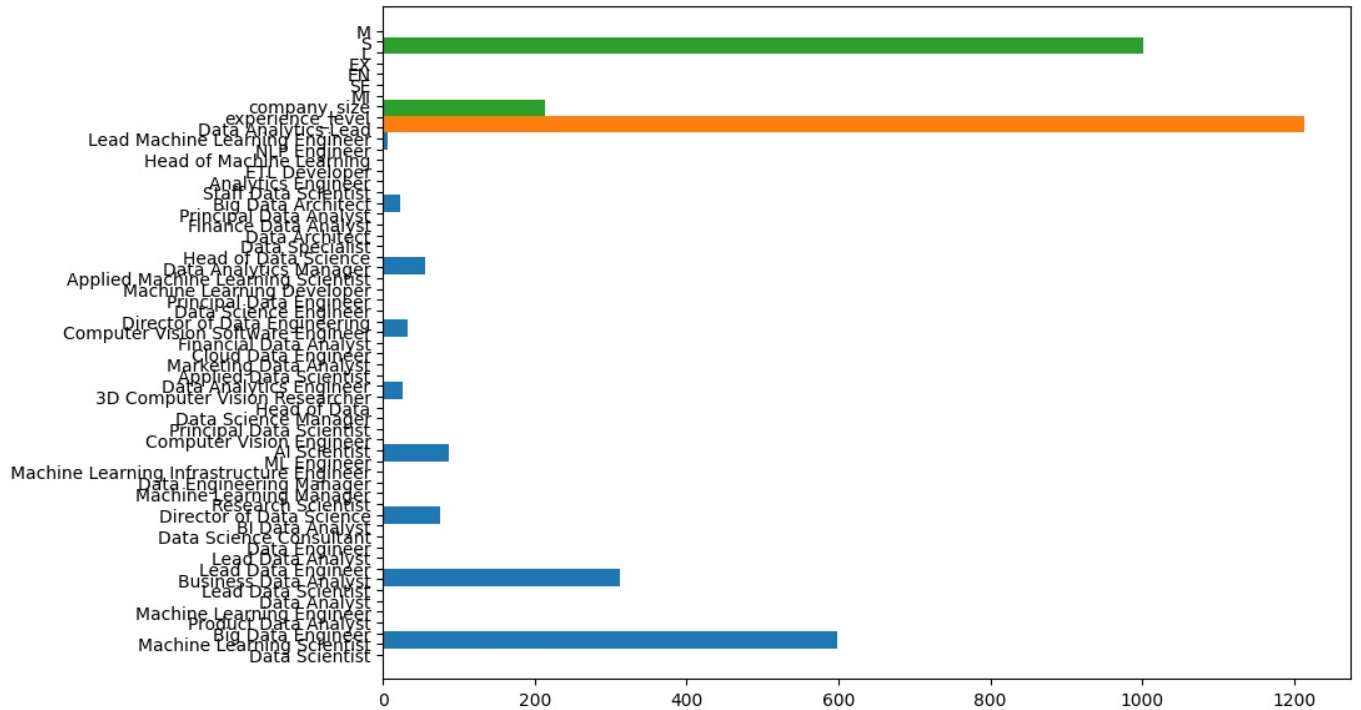
```python
plt.stem(df["experience_level"], df["salary_in_usd"])
plt.savefig("stem.png")
plt.show()
```
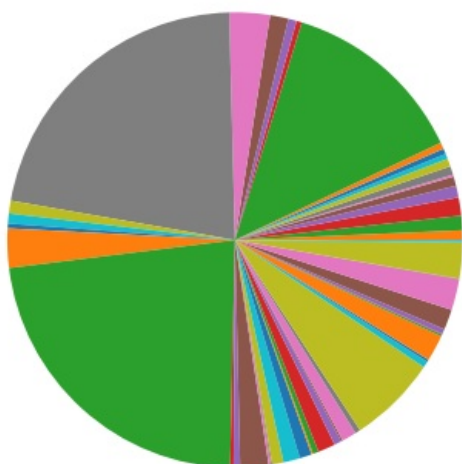
```python
plt.plot(df["salary_in_usd"], df["work_year"], label = "2020")
plt.plot(df["salary_in_usd"], df["work_year"], label = "2021")
plt.plot(df["salary_in_usd"], df["work_year"], label = "2022")
#plt.legend()
plt.show()
plt.savefig("plot.png", transparent = True)
```

```
<Figure size 640x480 with 0 Axes>
```
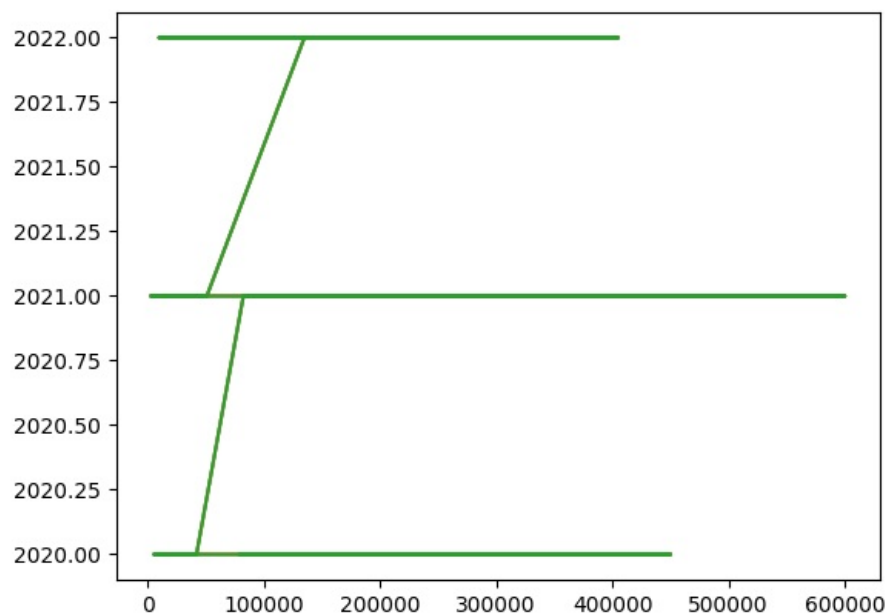
In [189…
```python
grouped_by = df.groupby("job_title")["salary_in_usd"].sum()
#print(grouped_by)
plt.pie(grouped_by.values)
plt.show()
plt.savefig("pie.png", transparent = True)
```



```
<Figure size 640x480 with 0 Axes>
```

In [188…
```python
grouped_by = df.groupby(["job_title","experience_level"])["salary_in_usd"].sum()
print(grouped_by)
plt.plot(grouped_by.values)
plt.show()
plt.savefig("plot.png" , transparent = True)
```

```
job_title                   experience_level
3D Computer Vision Researcher  MI                   5409
AI Scientist                   EN                  87949
                               MI                 320000
                               SE                  55000
Analytics Engineer             EX                 310000
                                                     ...
Product Data Analyst           MI                  26072
Research Scientist             EN                 344528
                               MI                 955487
                               SE                 444297
Staff Data Scientist           SE                 105000
Name: salary_in_usd, Length: 105, dtype: int64
```
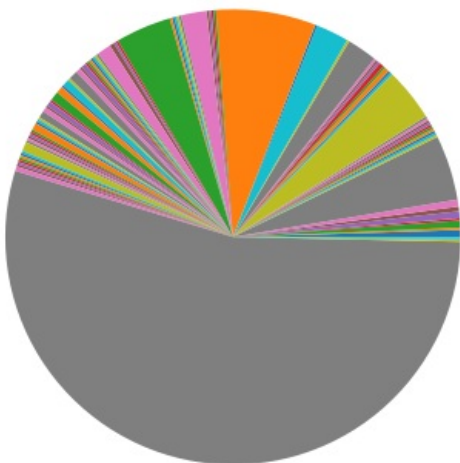
```
le6
```

<figure: line plot with y-axis "1e6" from 0 to ~9, x-axis from 0 to 100>

```
<Figure size 640x480 with 0 Axes>
```
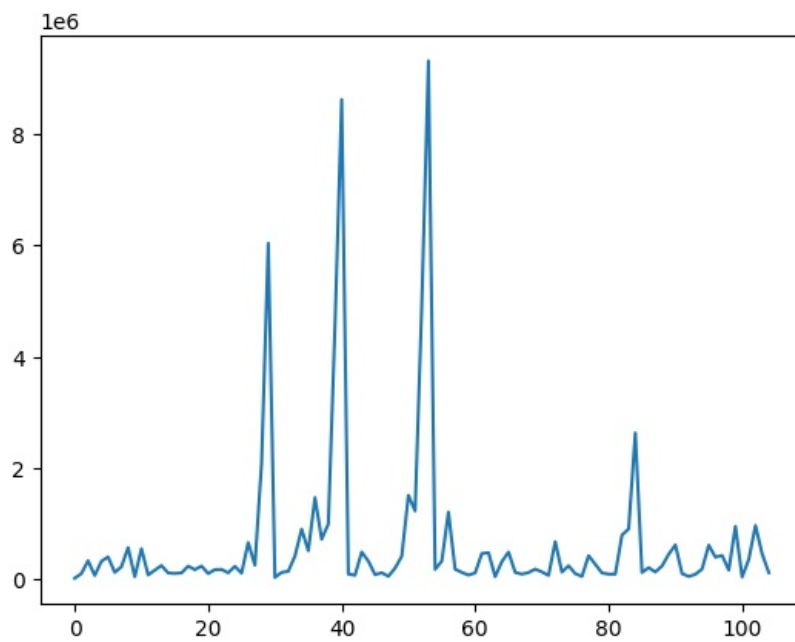
```
In [187…  gp1 = df.groupby(["company_location","employee_residence"])["remote_ratio"].count()
          print(gp1)
          plt.pie(gp1)
          plt.show()
          plt.savefig("pie.png", transparent = True)
```

```
company_location  employee_residence
AE                AE                     3
AS                IN                     1
AT                AT                     3
                  DE                     1
AU                AU                     3
                                       ...
US                RO                     1
                  RU                     2
                  US                   330
                  VN                     1
VN                VN                     1
Name: remote_ratio, Length: 90, dtype: int64
```



```
<Figure size 640x480 with 0 Axes>
```

```
In [115…  gp2 = df.groupby("job_title").agg({"work_year":"count","experience_level":"count"})
          print(gp2)
          sns.histplot(gp2, kde = True)
```

```
                                    work_year  experience_level
job_title
3D Computer Vision Researcher              1                 1
AI Scientist                               7                 7
Analytics Engineer                         4                 4
Applied Data Scientist                     5                 5
Applied Machine Learning Scientist         4                 4
BI Data Analyst                            6                 6
Big Data Architect                         1                 1
Big Data Engineer                          8                 8
Business Data Analyst                      5                 5
Cloud Data Engineer                        2                 2
Computer Vision Engineer                   6                 6
Computer Vision Software Engineer          3                 3
Data Analyst                              97                97
Data Analytics Engineer                    4                 4
Data Analytics Lead                        1                 1
Data Analytics Manager                     7                 7
Data Architect                            11                11
Data Engineer                            132               132
Data Engineering Manager                   5                 5
Data Science Consultant                    7                 7
Data Science Engineer                      3                 3
Data Science Manager                      12                12
Data Scientist                           143               143
Data Specialist                            1                 1
Director of Data Engineering               2                 2
Director of Data Science                   7                 7
ETL Developer                              2                 2
Finance Data Analyst                       1                 1
Financial Data Analyst                     2                 2
Head of Data                               5                 5
Head of Data Science                       4                 4
Head of Machine Learning                   1                 1
Lead Data Analyst                          3                 3
Lead Data Engineer                         6                 6
Lead Data Scientist                        3                 3
Lead Machine Learning Engineer             1                 1
ML Engineer                                6                 6
Machine Learning Developer                 3                 3
Machine Learning Engineer                 41                41
Machine Learning Infrastructure Engineer   3                 3
Machine Learning Manager                   1                 1
Machine Learning Scientist                 8                 8
Marketing Data Analyst                     1                 1
NLP Engineer                               1                 1
Principal Data Analyst                     2                 2
Principal Data Engineer                    3                 3
Principal Data Scientist                   7                 7
Product Data Analyst                       2                 2
Research Scientist                        16                16
Staff Data Scientist                       1                 1
```

Out[115… <Axes: ylabel='Count'>
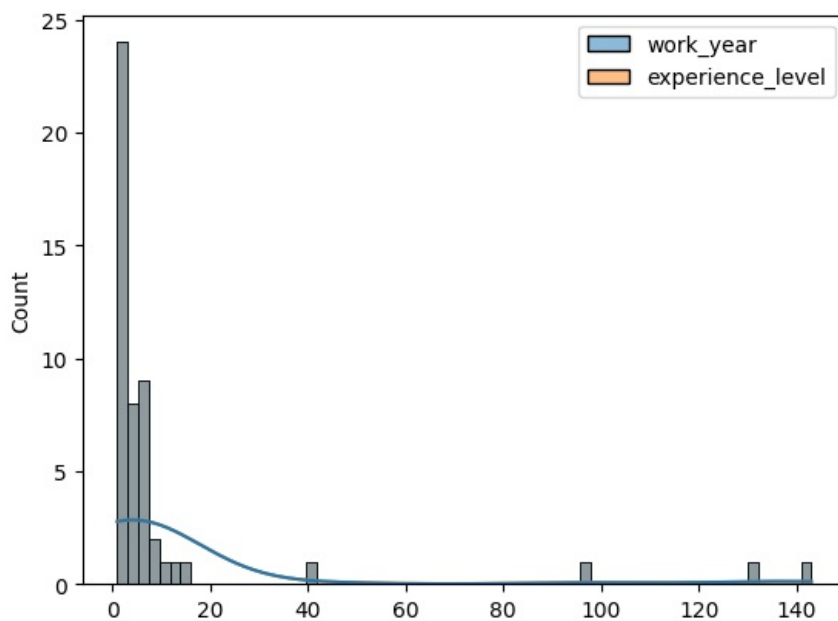

```
In [183…  b = (pd.melt(df,id_vars = ["job_title"],value_vars = ["experience_level"]))
          print(b)
          plt.xlabel("job_title")
          plt.hist(b)
```

```
plt.savefig('hist.png', transparent = True)
```

```
                       job_title            variable value
0                 Data Scientist    experience_level    MI
1     Machine Learning Scientist    experience_level    SE
2              Big Data Engineer    experience_level    SE
3            Product Data Analyst    experience_level    MI
4     Machine Learning Engineer    experience_level    SE
..                         ...                 ...   ...
602                Data Engineer    experience_level    SE
603                Data Engineer    experience_level    SE
604                 Data Analyst    experience_level    SE
605                 Data Analyst    experience_level    SE
606                  AI Scientist    experience_level    MI

[607 rows x 3 columns]
```
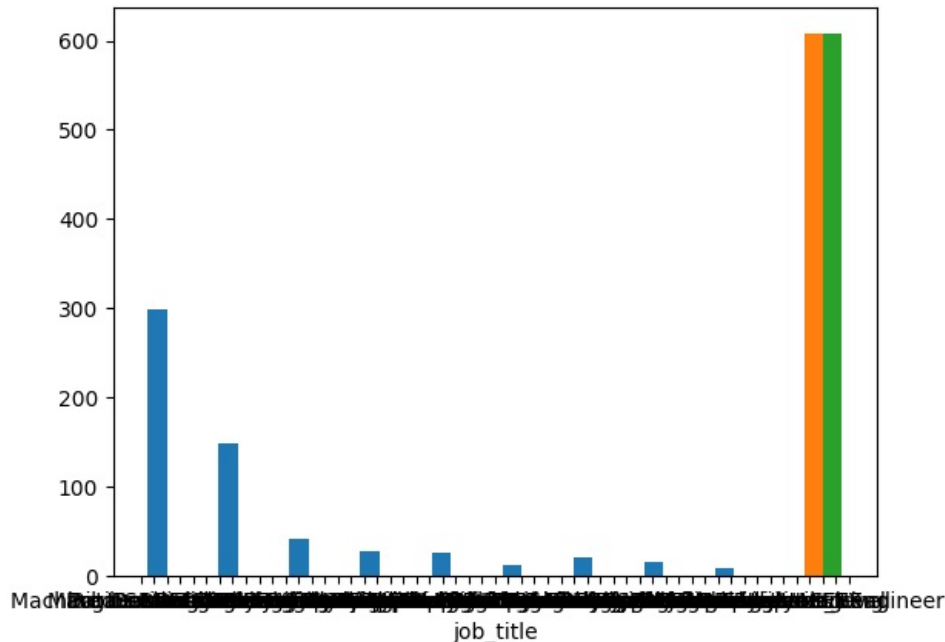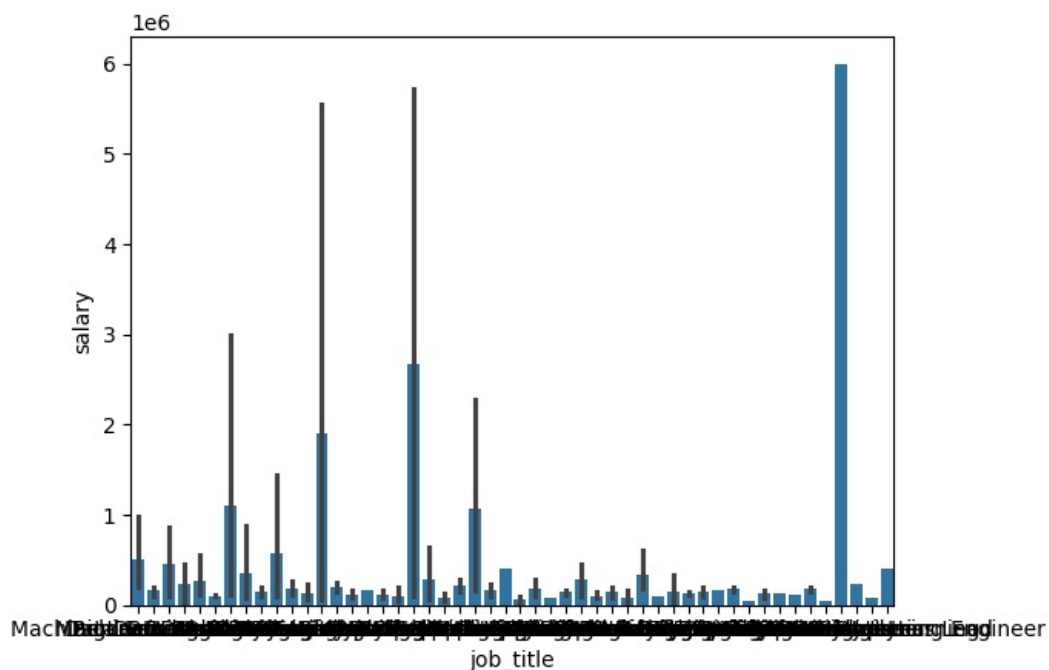

In [152...
```python
sns.barplot(data = data , x = "job_title", y = "salary")

#plt.legend(bbox_to_anchor = (0.2,0,1.2,1.2))
```

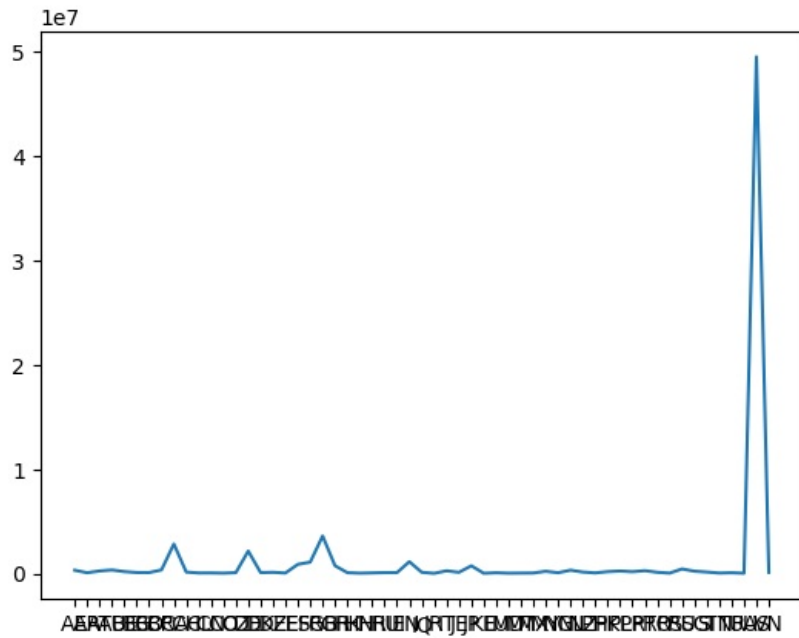Out[152... <Axes: xlabel='job_title', ylabel='salary'>

In [186...
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

R = df.groupby("employee_residence")["salary_in_usd"].sum()
#print(R
plt.plot(R)
```
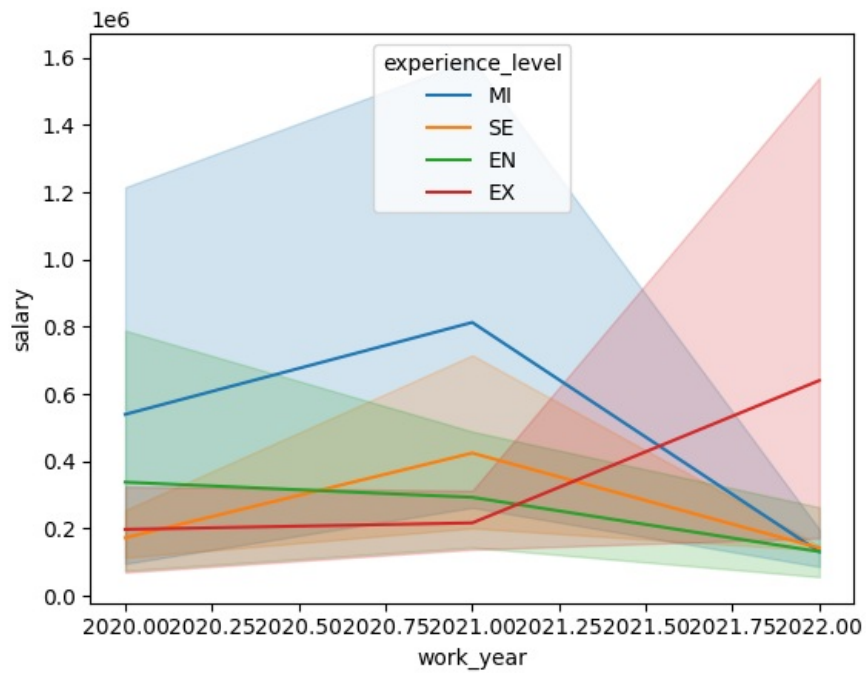
```
plt.savefig("plot.png", transparent = True)
```



```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.lineplot(data = data, x = "work_year", y = "salary", hue = "experience_level") #style = "Bronzers")
plt.show()
```