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

In [ ]:
```
df = pd.read_csv('ds_salaries.csv')
```

In [ ]:
```
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
```

In [ ]:
```
df.head()
```

Out[ ]:

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_curr
0	0	2020	MI	FT	Data Scientist	70000	
1	1	2020	SE	FT	Machine Learning Scientist	260000	
2	2	2020	SE	FT	Big Data Engineer	85000	
3	3	2020	MI	FT	Product Data Analyst	20000	
4	4	2020	SE	FT	Machine Learning Engineer	150000	

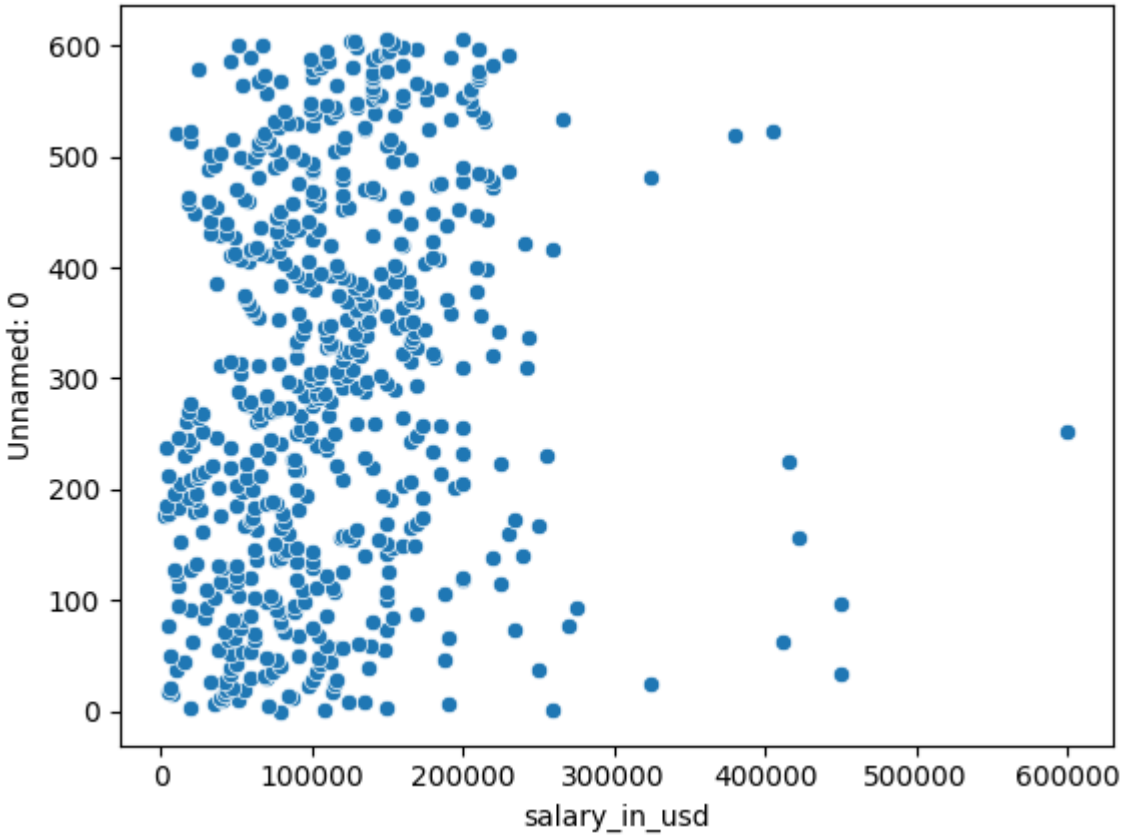
In [ ]:
```
df2 = df [['salary_in_usd','Unnamed: 0']].dropna()
df2.head()
```

Out[ ]:

	salary_in_usd	Unnamed: 0
0	79833	0
1	260000	1
2	109024	2
3	20000	3
4	150000	4

In [ ]:
```
sns.scatterplot(data = df , x = 'salary_in_usd', y = 'Unnamed: 0')
```

Out[ ]:
```
<Axes: xlabel='salary_in_usd', ylabel='Unnamed: 0'>
```



In [ ]:
```
df2 = df [['salary_in_usd','Unnamed: 0']].dropna()
df2.head()
```

Out[ ]:

	salary_in_usd	Unnamed: 0
0	79833	0
1	260000	1
2	109024	2
3	20000	3
4	150000	4

In [ ]:
```
from sklearn.cluster import KMeans
```

In [ ]:
```
model = KMeans(n_clusters=4,random_state=0)
model.fit(df2)
```

Out[ ]:
```
KMeans(n_clusters=4, random_state=0)
```

In [ ]:
```
sns.scatterplot(data = df2 , x = 'salary_in_usd', y = 'Unnamed: 0'
                ,hue=model.labels_,palette='Set2')
plt.scatter(model.cluster_centers_[:,0], model.cluster_centers_[:,1]
            ,color = 'k',marker='*')
```

Out[ ]:
```
<matplotlib.collections.PathCollection at 0x28a57fc60d0>
```