นาย ณภัทรนันท์ ศิลปะ 6421600069

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

In [ ]:
```
sal = pd.read_csv('ds_salaries.csv',encoding = 'iso-8859-1')
```

In [ ]:
```
sal.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
```

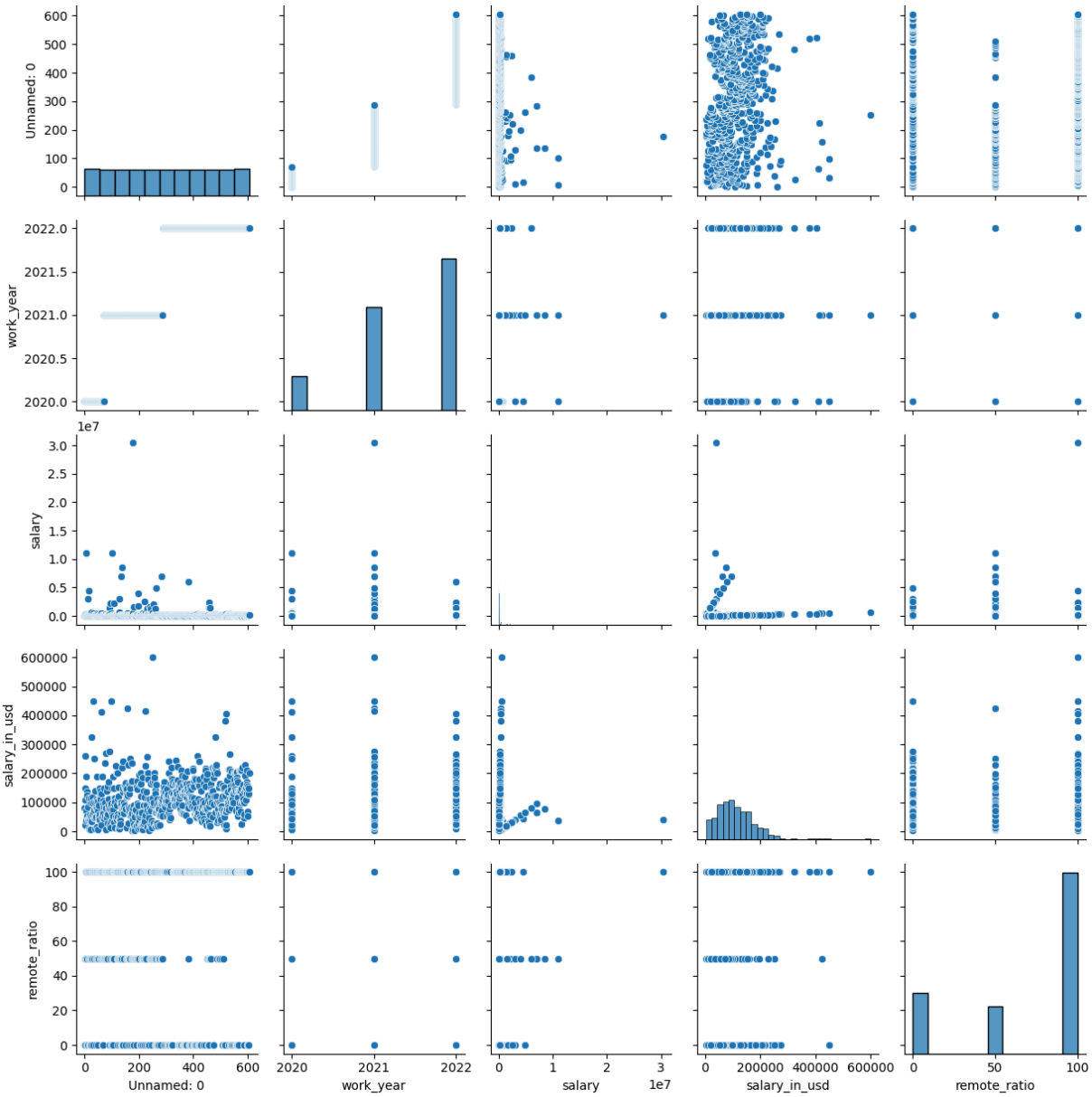
In [ ]:
```
sal.head()
```

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	

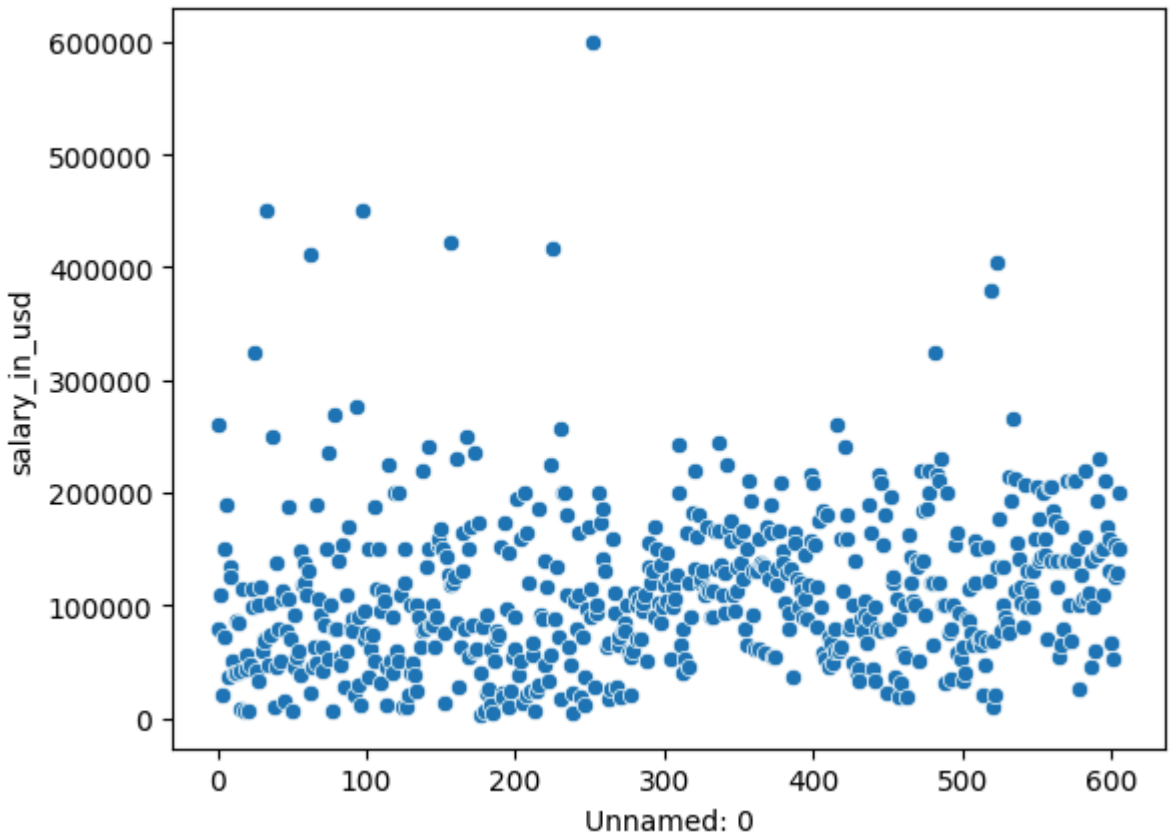
In [ ]:
```
sns.pairplot(data=sal)
```

Out[ ]:
```
<seaborn.axisgrid.PairGrid at 0x1f30923b650>
```



In [ ]:
```
sal2 = sal[['Unnamed: 0','salary_in_usd']].dropna()
# df2 = df[['displacement','horsepower']].dropna()
```

In [ ]:
```
sns.scatterplot(data=sal2 , x='Unnamed: 0',y='salary_in_usd')
plt.show()
```



In [ ]:
```
model = KMeans(n_clusters=4,random_state=0)
model.fit(sal2)
```

Out[ ]:

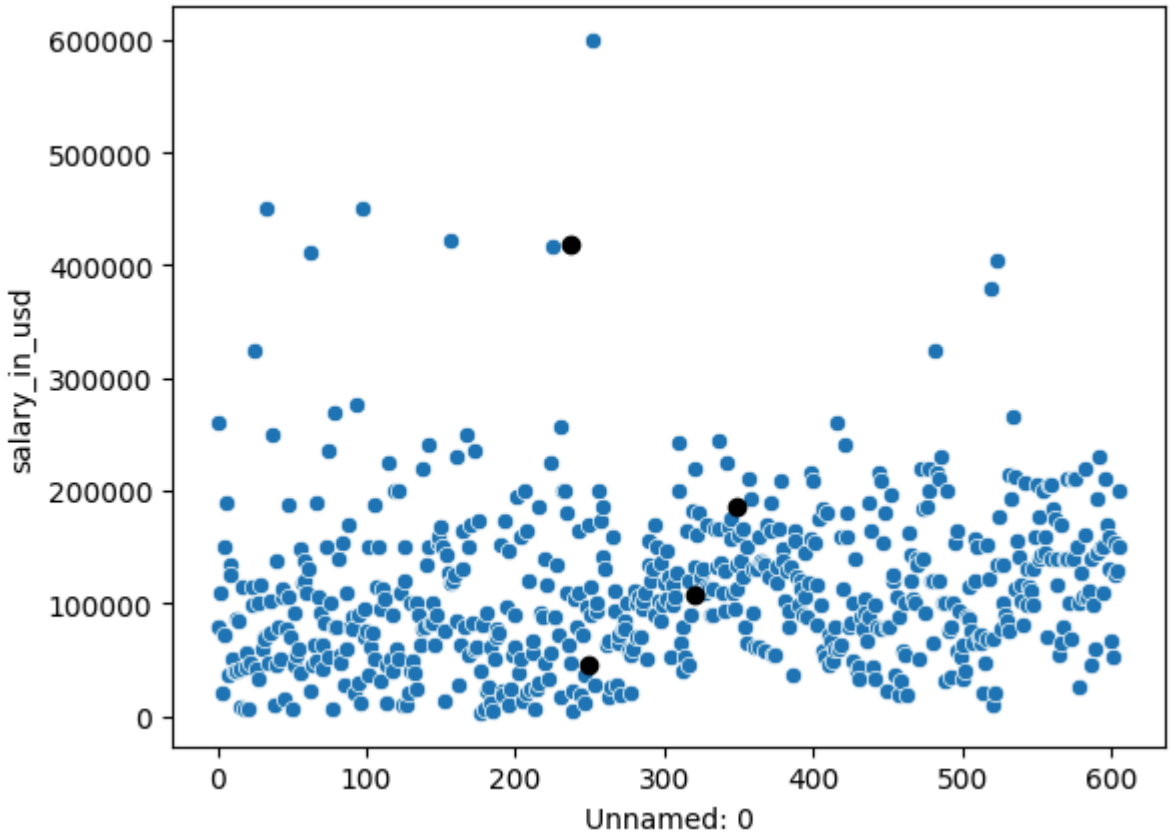
KMeans

KMeans(n_clusters=4, random_state=0)

In [ ]:
```
model.cluster_centers_
```

Out[ ]:
```
array([[3.19751004e+02, 1.08264912e+05],
       [3.48960526e+02, 1.86111546e+05],
       [2.37600000e+02, 4.18500000e+05],
       [2.49413265e+02, 4.45555561e+04]])
```

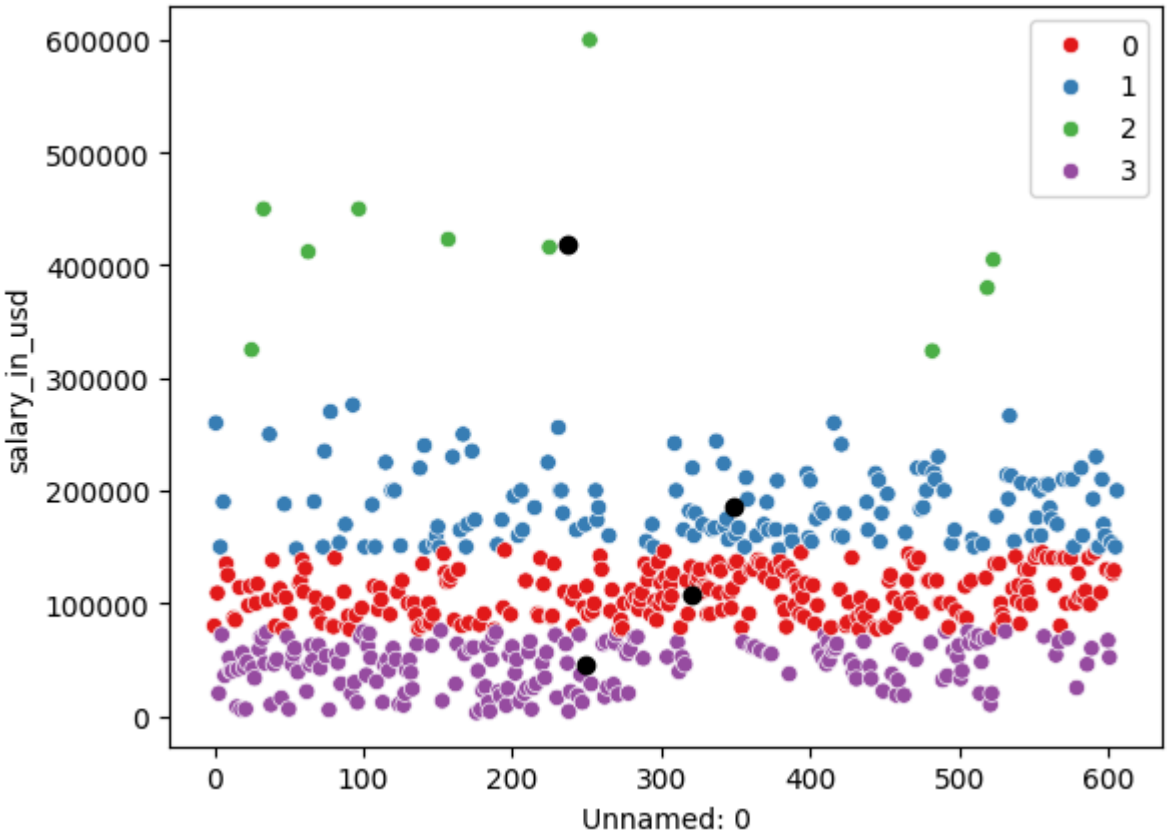
In [ ]:
```
sns.scatterplot(data=sal2 , x='Unnamed: 0',y='salary_in_usd')
plt.scatter(model.cluster_centers_[:,0],model.cluster_centers_[:,1],color='#000')
plt.show()
```



In [ ]:
```
model.labels_
```

Out[ ]:
```
array([0, 1, 0, 3, 1, 3, 1, 3, 0, 0, 3, 3, 3, 0, 0, 3, 3, 0, 3, 3, 3, 3,
       3, 0, 0, 2, 3, 3, 0, 0, 3, 3, 3, 2, 3, 3, 0, 1, 3, 0, 3, 0, 3, 0,
       0, 3, 0, 1, 0, 3, 3, 0, 3, 3, 3, 1, 3, 0, 0, 0, 0, 0, 3, 2, 3, 3,
       3, 1, 0, 0, 3, 3, 0, 1, 1, 3, 0, 3, 1, 0, 0, 0, 3, 3, 1, 3, 3, 0,
       1, 0, 0, 0, 3, 1, 3, 0, 3, 2, 3, 0, 3, 1, 3, 3, 3, 3, 1, 0, 1, 3,
       0, 0, 0, 3, 3, 1, 3, 3, 0, 1, 3, 1, 3, 0, 3, 1, 0, 3, 3, 3, 0, 3,
       3, 3, 0, 0, 3, 0, 1, 0, 0, 1, 1, 0, 0, 0, 3, 0, 1, 1, 1, 1, 3, 3,
       0, 0, 0, 2, 0, 0, 1, 0, 3, 3, 0, 1, 0, 1, 3, 1, 1, 0, 3, 1, 3, 1,
       3, 3, 0, 3, 3, 0, 3, 3, 3, 3, 3, 3, 0, 3, 1, 3, 3, 1, 0, 0, 3, 3,
       3, 0, 3, 1, 3, 3, 1, 3, 1, 1, 3, 0, 3, 3, 3, 3, 3, 1, 3, 0, 0, 0,
       3, 0, 3, 3, 1, 2, 0, 0, 0, 3, 3, 1, 1, 1, 1, 0, 3, 3, 3, 3, 0, 0,
       0, 1, 3, 3, 3, 3, 0, 1, 0, 0, 2, 3, 0, 0, 1, 1, 1, 0, 0, 3, 3, 3,
       3, 1, 0, 0, 3, 3, 3, 3, 0, 0, 0, 0, 3, 3, 3, 3, 0, 0, 3, 0, 3, 0,
       0, 0, 3, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0,
       0, 1, 1, 3, 3, 0, 3, 1, 3, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0,
       0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0,
       1, 0, 0, 3, 1, 1, 1, 0, 0, 3, 0, 3, 1, 0, 0, 3, 0, 1, 0, 1, 1, 3,
       3, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 3, 1, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 3, 1, 3, 1, 3, 3, 3, 3, 0, 3, 1, 3,
       3, 1, 0, 1, 1, 1, 0, 0, 0, 3, 0, 3, 3, 3, 0, 0, 0, 0, 3, 0, 1, 1,
       3, 3, 0, 0, 1, 0, 1, 1, 1, 3, 0, 0, 1, 0, 0, 3, 0, 0, 3, 3, 3, 0,
       3, 3, 1, 0, 0, 0, 0, 0, 0, 3, 1, 0, 1, 0, 1, 1, 1, 0, 0, 3, 2, 1,
       1, 0, 1, 0, 0, 3, 1, 3, 3, 0, 0, 1, 3, 1, 0, 3, 3, 3, 3, 0, 0, 0,
       3, 3, 0, 1, 1, 3, 3, 3, 3, 3, 1, 3, 0, 2, 3, 3, 3, 2, 0, 1, 0, 0,
       0, 0, 0, 3, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
       1, 0, 1, 0, 1, 1, 0, 3, 1, 0, 1, 1, 1, 0, 0, 3, 1, 3, 0, 0, 1, 0,
       0, 3, 1, 0, 1, 1, 0, 3, 0, 0, 1, 1, 0, 0, 3, 0, 0, 3, 1, 0, 1, 1,
       1, 0, 1, 1, 1, 0, 3, 3, 1, 0, 0, 1, 1])
```

In [ ]:
```
sns.scatterplot(data=sal2 , x='Unnamed: 0',y='salary_in_usd',hue=model.labels_ ,pal
plt.scatter(model.cluster_centers_[:,0],model.cluster_centers_[:,1],color='#000')
plt.show()
```


600000
500000
400000
300000
200000
100000
0
salary_in_usd
0
100
200
300
400
500
600
Unnamed: 0
0
1
2
3