

Salaries DataSet

In [91]:

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

df = pd.read_csv('ds_salaries.csv')
df.dropna(inplace=True)

df.head(10)
```

Out[91]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employ
0	2023	SE	FT	Principal Data Scientist	80000	EUR	85847	
1	2023	MI	CT	ML Engineer	30000	USD	30000	
2	2023	MI	CT	ML Engineer	25500	USD	25500	
3	2023	SE	FT	Data Scientist	175000	USD	175000	
4	2023	SE	FT	Data Scientist	120000	USD	120000	
5	2023	SE	FT	Applied Scientist	222200	USD	222200	

Оставим только компании, расположенные в России.

In [92]:

```
df[df['company_location'] == 'RU']
```

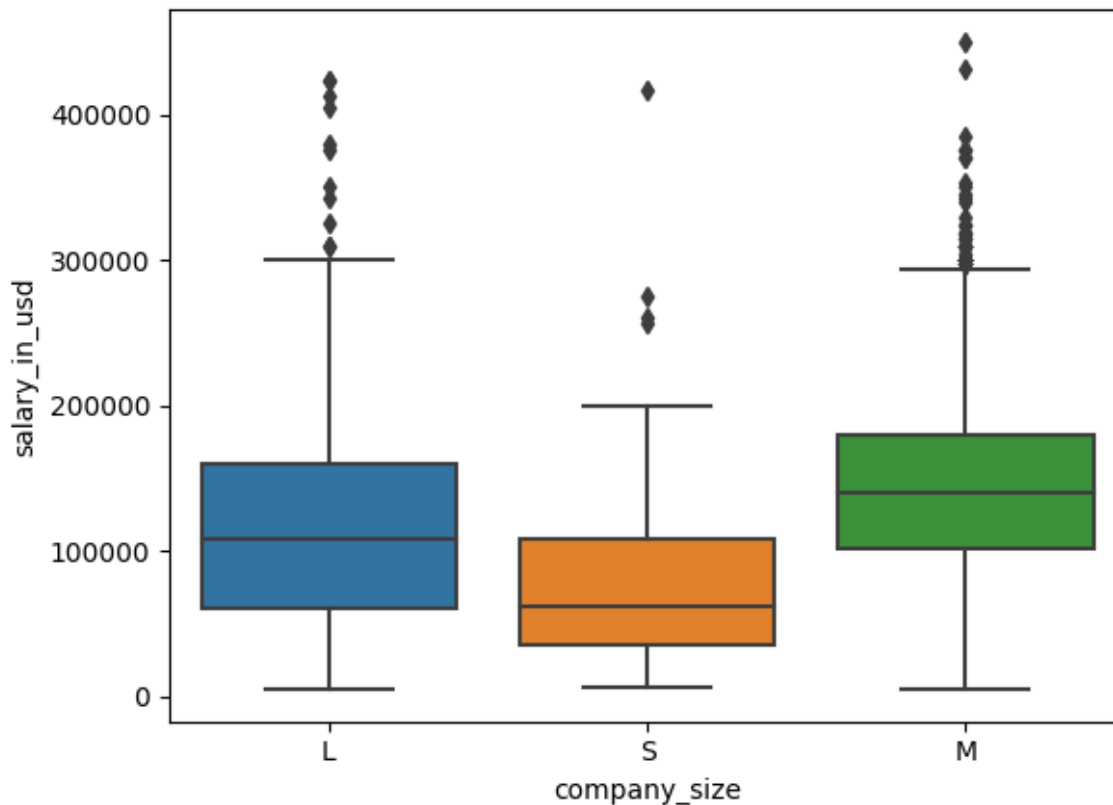
Out[92]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salar
974	2022	EX	FT	Data Science Manager	106000	USD	
3518	2021	EX	FT	Head of Data Science	85000	USD	
3519	2021	EX	FT	Head of Data	230000	USD	

Проследим за зависимостью зарплат от размера компании (данные уже по всем странам). Как видим, зарплаты в маленьких компаниях намного ниже, чем в средних и больших. Во вторых и третьих они примерно равны (средняя зп немного выше даже в средних).

In [93]:

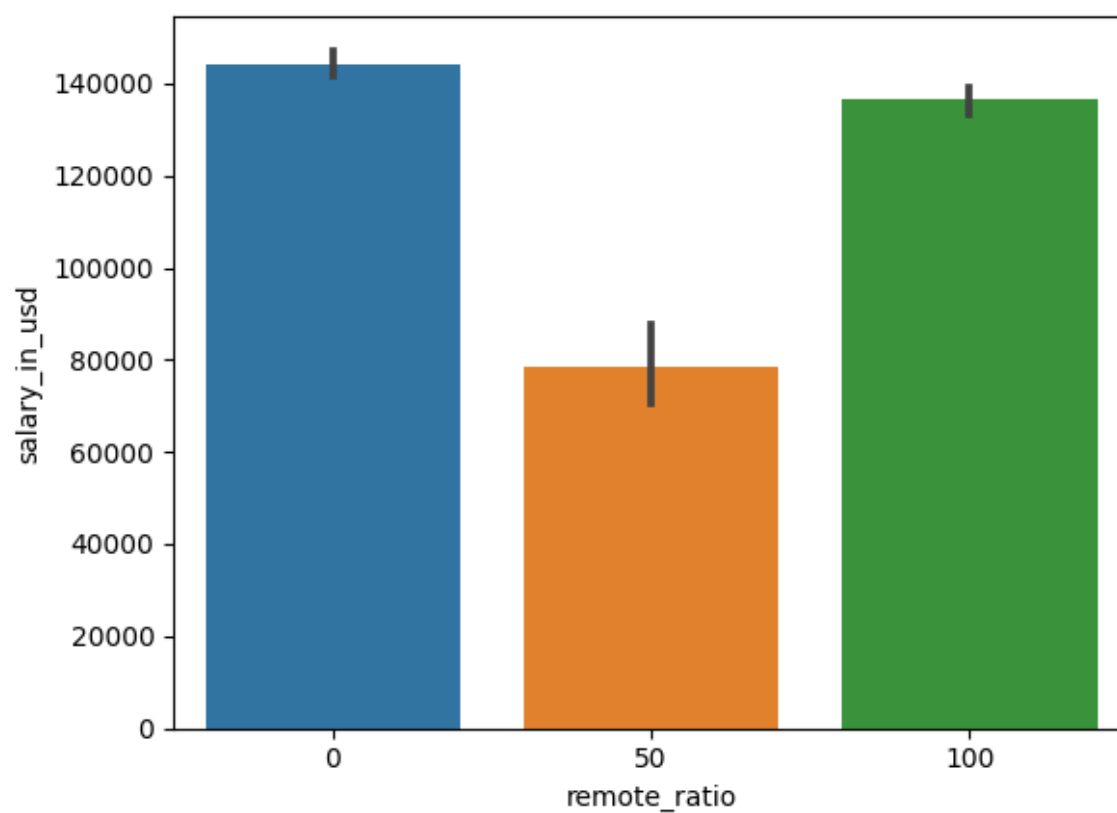
```
sns.boxplot(x='company_size', y='salary_in_usd', data=df)
plt.show()
```



Найдем среднюю зарплату в США Data Scientist-истов, зависящую от того, насколько работа удаленная (0%, 50% или 100%). Как видим, работа на полу-удаленной основе наиболее плохо оплачивается.

In [103]:

```
df.company_location='US'  
df.job_title='Data Scientist'  
sns.barplot(data=df, x="remote_ratio", y="salary_in_usd")  
plt.show()
```



Интересные графики для различных типов работы по "удаленности" (можно заметить, например, что в 2022 году из-за карантина зп на удаленке достигла зп на очке!):

In [83]:

```
sns.pairplot(df, hue="remote_ratio", palette='rainbow')  
plt.show()
```

