In [ ]:
```
import pandas as pd
import numpy as np
import country_converter
import pickle
```

In [ ]:
```
df = pd.read_csv('Data Science Job Salaries.csv')
```

In [ ]:
```
df.head(1)
```

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

In [ ]:
```
df.salary_in_usd.nunique()
```

In [ ]:
```
for i in df.columns:
    if df[i].nunique()>20:
        print(i, df[i].nunique())
    else:
        print(i," -> ", df[i].unique())
```

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

In [ ]:
```
# df.drop(['salary', 'salary_currency'], axis=1, inplace=True)
```

In [ ]:
```
df_cleaned =df.drop(['salary', 'salary_currency'], axis=1).copy()
```

In [ ]:
```
df_cleaned['job_title'] = df_cleaned['job_title'].str.lower()
```

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

In [ ]:
```
df.duplicated().sum()
```

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

Fixing country names

In [ ]:
```
cc = country_converter.CountryConverter()
```

In [ ]:
```
df_cleaned['company_location'] = cc.convert(df['company_location'], to='name_short'
```

In [ ]:
```
df_cleaned['employee_residence'] = cc.convert(df_cleaned['employee_residence'], to=
```

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

Mapping

In [ ]:
```
df_cleaned['experience_level'] = df_cleaned['experience_level'].map({
    'MI' : 'Mid',
    'SE' : 'Senior',
    'EN': "Entry",
    'EX' : 'Executive'
})
```

In [ ]:
```
df_cleaned['remote_ratio']  = df_cleaned['remote_ratio'].map({
    0: 'Onsite',
    50 : 'Hybrid',
    100 : 'Remote'
})
```

In [ ]:
```
df_cleaned.rename(columns={'remote_ratio' : 'job_type'}, inplace=True)
```

In [ ]:
```
df_cleaned['company_size'] = df_cleaned['company_size'].map({'L':'Large','M':'Mediu
```

In [ ]:
```
df_cleaned['employment_type'] = df_cleaned['employment_type'].map({'FT':'Full-time'
```

In [ ]:
```
df_cleaned
```

In [ ]:
```
df_cleaned.value_counts('employment_type')
```

In [ ]:
```
pd.to_pickle(df_cleaned, 'df_cleaned')
```

In [ ]:
```

```