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

In [2]:
```
file_path=r"C:\Users\Mrityunjay\Desktop\Data science naresh it\Class notes by me\ED
visa_df=pd.read_csv(file_path)
visa_df
```

Out[2]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_trair
0	EZYV01	Asia	High School	N	
1	EZYV02	Asia	Master's	Y	
2	EZYV03	Asia	Bachelor's	N	
3	EZYV04	Asia	Bachelor's	N	
4	EZYV05	Africa	Master's	Y	
...	...	...	...	...	
25475	EZYV25476	Asia	Bachelor's	Y	
25476	EZYV25477	Asia	High School	Y	
25477	EZYV25478	Asia	Master's	Y	
25478	EZYV25479	Asia	Master's	Y	
25479	EZYV25480	Asia	Bachelor's	Y	

25480 rows × 12 columns

In [3]:
```
cat_col=visa_df.select_dtypes(include="object").columns
```

In [4]:
```
cat_col
```

Out[4]:
```
Index(['case_id', 'continent', 'education_of_employee', 'has_job_experience',
       'requires_job_training', 'region_of_employment', 'unit_of_wage',
       'full_time_position', 'case_status'],
      dtype='object')
```

In [5]:
```
col=visa_df["continent"].unique()
col
```

Out[5]:
```
array(['Asia', 'Africa', 'North America', 'Europe', 'South America',
       'Oceania'], dtype=object)
```

In [6]:
```
d={}
for i in range(len(col)):
```

```
    d[col[i]]=i
d
```

Out[6]:
```
{'Asia': 0,
 'Africa': 1,
 'North America': 2,
 'Europe': 3,
 'South America': 4,
 'Oceania': 5}
```

In [7]:
```
c=visa_df["continent"].map(d)
c
```

Out[7]:
```
0        0
1        0
2        0
3        0
4        1
        ..
25475    0
25476    0
25477    0
25478    0
25479    0
Name: continent, Length: 25480, dtype: int64
```

In [8]:
```
col1=visa_df["education_of_employee"].unique()
col1
```

Out[8]:
```
array(['High School', "Master's", "Bachelor's", 'Doctorate'], dtype=object)
```

In [9]:
```
d1={}
for i in range(len(col1)):
    d[col1[i]]=i
d1
```

Out[9]:
```
{}
```

In [10]:
```
e=visa_df["education_of_employee"].map(d)
e
```

Out[10]:
```
0        0
1        1
2        2
3        2
4        1
        ..
25475    2
25476    0
25477    1
25478    1
25479    2
Name: education_of_employee, Length: 25480, dtype: int64
```

In [11]:
```
col=visa_df["has_job_experience"].unique()
```

```
col
```

Out[11]:
```
array(['N', 'Y'], dtype=object)
```

In [12]:
```
d={}
for i in range(len(col)):
    d[col[i]]=i
d
```

Out[12]:
```
{'N': 0, 'Y': 1}
```

In [13]:
```
h=visa_df["has_job_experience"].map(d)
h
```

Out[13]:
```
0        0
1        1
2        0
3        0
4        1
        ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: has_job_experience, Length: 25480, dtype: int64
```

In [14]:
```
col=visa_df["full_time_position"].unique()
col
```

Out[14]:
```
array(['Y', 'N'], dtype=object)
```

In [15]:
```
d={}
for i in range(len(col)):
    d[col[i]]=i
d
```

Out[15]:
```
{'Y': 0, 'N': 1}
```

In [16]:
```
f=visa_df["full_time_position"].map(d)
f
```

Out[16]:
```
0        0
1        0
2        0
3        0
4        0
        ..
25475    0
25476    0
25477    1
25478    0
25479    0
Name: full_time_position, Length: 25480, dtype: int64
```

In [17]:
```
col=visa_df["region_of_employment"].unique()
col
```

Out[17]:
```
array(['West', 'Northeast', 'South', 'Midwest', 'Island'], dtype=object)
```

In [18]:
```
d={}
for i in range(len(col)):
    d[col[i]]=i
d
```

Out[18]:
```
{'West': 0, 'Northeast': 1, 'South': 2, 'Midwest': 3, 'Island': 4}
```

In [19]:
```
rn=visa_df["region_of_employment"].map(d)
rn
```

Out[19]:
```
0        0
1        1
2        0
3        0
4        2
        ..
25475    2
25476    1
25477    2
25478    0
25479    3
Name: region_of_employment, Length: 25480, dtype: int64
```

In [20]:
```
col=visa_df["requires_job_training"].unique()
col
```

Out[20]:
```
array(['N', 'Y'], dtype=object)
```

In [21]:
```
d={}
for i in range(len(col)):
    d[col[i]]=i
d
```

Out[21]:
```
{'N': 0, 'Y': 1}
```

In [22]:
```
rt=visa_df["requires_job_training"].map(d)
rt
```

Out[22]:
```
0        0
1        0
2        1
3        0
4        0
        ..
25475    1
25476    0
25477    0
25478    1
25479    0
Name: requires_job_training, Length: 25480, dtype: int64
```

In [23]:
```
col=visa_df["unit_of_wage"].unique()
col
```

Out[23]:
```
array(['Hour', 'Year', 'Week', 'Month'], dtype=object)
```

In [24]:
```
d={}
for i in range(len(col)):
    d[col[i]]=i
d
```

Out[24]:
```
{'Hour': 0, 'Year': 1, 'Week': 2, 'Month': 3}
```

In [25]:
```
u=visa_df["unit_of_wage"].map(d)
u
```

Out[25]:
```
0        0
1        1
2        1
3        1
4        1
        ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: unit_of_wage, Length: 25480, dtype: int64
```

In [26]:
```
col=visa_df["case_status"].unique()
col
```

Out[26]:
```
array(['Denied', 'Certified'], dtype=object)
```

In [27]:
```
d={}
for i in range(len(col)):
    d[col[i]]=i
d
```

Out[27]:
```
{'Denied': 0, 'Certified': 1}
```

In [28]:
```
cs=visa_df["case_status"].map(d)
```

```
cs
```

Out[28]:
```
0        0
1        1
2        0
3        0
4        1
        ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: case_status, Length: 25480, dtype: int64
```

In [142…
```
col=['continent', 'education_of_employee', 'has_job_experience',
       'requires_job_training', 'region_of_employment', 'unit_of_wage',
       'full_time_position', 'case_status']
df=pd.DataFrame(zip(c,e,h,f,rn,rt,u,cs),columns=col)

df
```

Out[142…

	continent	education_of_employee	has_job_experience	requires_job_training	region
0	0	0	0	0	
1	0	1	1	0	
2	0	2	0	0	
3	0	2	0	0	
4	1	1	1	0	
...	...	...	...	...	
25475	0	2	1	0	
25476	0	0	1	0	
25477	0	1	1	1	
25478	0	1	1	0	
25479	0	2	1	0	

25480 rows × 8 columns

In [62]:
```
cat_col[1:]
```

Out[62]:
```
Index(['continent', 'education_of_employee', 'has_job_experience',
       'requires_job_training', 'region_of_employment', 'unit_of_wage',
       'full_time_position', 'case_status'],
      dtype='object')
```

In [130…

```
l=[]
d={}
for i in cat_col[1:]:
    col=visa_df[i].unique()
    #l.append(col)
    #d={}
    for j in range(len(col)):
        d[col[j]]=j
    n=visa_df[i].map(d)
    l.append(n)
l
```

Out[130… [0 0
1 0
2 0
3 0
4 1
..
25475 0
25476 0
25477 0
25478 0
25479 0
Name: continent, Length: 25480, dtype: int64,
0 0
1 1
2 2
3 2
4 1
..
25475 2
25476 0
25477 1
25478 1
25479 2
Name: education_of_employee, Length: 25480, dtype: int64,
0 0
1 1
2 0
3 0
4 1
..
25475 1
25476 1
25477 1
25478 1
25479 1
Name: has_job_experience, Length: 25480, dtype: int64,
0 0
1 0
2 1
3 0
4 0
..
25475 1
25476 0
25477 0
25478 1
25479 0
Name: requires_job_training, Length: 25480, dtype: int64,
0 0
1 1
2 0
3 0
4 2
..
25475 2
25476 1

```
25477    2
25478    0
25479    3
Name: region_of_employment, Length: 25480, dtype: int64,
0        0
1        1
2        1
3        1
4        1
        ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: unit_of_wage, Length: 25480, dtype: int64,
0        0
1        0
2        0
3        0
4        0
        ..
25475    0
25476    0
25477    1
25478    0
25479    0
Name: full_time_position, Length: 25480, dtype: int64,
0        0
1        1
2        0
3        0
4        1
        ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: case_status, Length: 25480, dtype: int64]
```

In [ ]: