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

In [2]:
```
visa_df=pd.read_csv(r'C:\Users\ADMIN\Documents\Naresh it\EDA with python\data fi
visa_df.head(2)
```

Out[2]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	Asia	High School	N	N
1	EZYV02	Asia	Master's	Y	N

In [4]:
```
visa_df.select_dtypes(include='object').columns
```

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 [6]:
```
visa_df['continent'].unique()
```

Out[6]:
```
array(['Asia', 'Africa', 'North America', 'Europe', 'South America',
       'Oceania'], dtype=object)
```

In [7]:
```
d={'Africa':0,'Asia':1,'Europe':2,'North America':3,'Oceania':4,'South America':
d
```

Out[7]:
```
{'Africa': 0,
 'Asia': 1,
 'Europe': 2,
 'North America': 3,
 'Oceania': 4,
 'South America': 5}
```

In [8]:
```
visa_df['continent'].map(d)
```

Out[8]:
```
0        1
1        1
2        1
3        1
4        0
        ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: continent, Length: 25480, dtype: int64
```

In [9]:
```
visa_df['continent']=visa_df['continent'].map(d)
```

In [11]:
```
visa_df.head()
```

Out[11]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	1	High School	N	N
1	EZYV02	1	Master's	Y	N
2	EZYV03	1	Bachelor's	N	Y
3	EZYV04	1	Bachelor's	N	N
4	EZYV05	0	Master's	Y	N

In [13]:
```
visa_df=pd.read_csv(r'C:\Users\ADMIN\Documents\Naresh it\EDA with python\data fi
visa_df.head()
```

Out[13]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	Asia	High School	N	N
1	EZYV02	Asia	Master's	Y	N
2	EZYV03	Asia	Bachelor's	N	Y
3	EZYV04	Asia	Bachelor's	N	N
4	EZYV05	Africa	Master's	Y	N

In [12]:
```
visa_df['has_job_experience'].unique()
```

Out[12]:
```
array(['N', 'Y'], dtype=object)
```

In [14]:
```
d1={'N':0,'Y':1}
d1
```

Out[14]:
```
{'N': 0, 'Y': 1}
```

In [15]:
```
visa_df['has_job_experience']=visa_df['has_job_experience'].map(d1)
```

In [17]:
```
visa_df.head()
```

Out[17]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	Asia	High School	0	N
1	EZYV02	Asia	Master's	1	N
2	EZYV03	Asia	Bachelor's	0	Y
3	EZYV04	Asia	Bachelor's	0	N
4	EZYV05	Africa	Master's	1	N

In [18]:
```
visa_df=pd.read_csv(r'C:\Users\ADMIN\Documents\Naresh it\EDA with python\data fi
visa_df.head(2)
```

Out[18]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	Asia	High School	N	N
1	EZYV02	Asia	Master's	Y	N

In [20]:
```
cols=visa_df.select_dtypes(include='object').columns
for i in cols:
    labels=list(visa_df[i].unique())
    values=[i for i in range(len(labels))]
    d=dict(zip(labels,values))
    visa_df[i]=visa_df[i].map(d)

visa_df
```

Out[20]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_trai
0	0	0	0	0	
1	1	0	1	1	
2	2	0	2	0	
3	3	0	2	0	
4	4	1	1	1	
...	...	...	...	...	
25475	25475	0	2	1	
25476	25476	0	0	1	
25477	25477	0	1	1	
25478	25478	0	1	1	
25479	25479	0	2	1	

25480 rows × 12 columns

In [21]:
```
visa_df=pd.read_csv(r'C:\Users\ADMIN\Documents\Naresh it\EDA with python\data fi
visa_df.head(2)
```

Out[21]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	Asia	High School	N	N
1	EZYV02	Asia	Master's	Y	N

In [22]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
visa_df['continent']=le.fit_transform(visa_df['continent'])
visa_df.head()
```

Out[22]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	EZYV01	1	High School	N	N
1	EZYV02	1	Master's	Y	N
2	EZYV03	1	Bachelor's	N	Y
3	EZYV04	1	Bachelor's	N	N
4	EZYV05	0	Master's	Y	N

In [24]:
```
cols=visa_df.select_dtypes(include='object').columns
for i in cols:
    visa_df[i]=le.fit_transform(visa_df[i])

visa_df.head()
```

Out[24]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training
0	0	1	2	0	0
1	1	1	3	1	0
2	2	1	0	0	1
3	3	1	0	0	0
4	4	0	3	1	0

In [31]:
```
visa_df=pd.read_csv(r'C:\Users\ADMIN\Documents\Naresh it\EDA with python\data fi
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()

visa_df['case_status']=le.fit_transform(visa_df['case_status'])
print(visa_df['case_status'].values)
print(le.inverse_transform(visa_df['case_status']))
```

```
[1 0 1 ... 0 0 0]
['Denied' 'Certified' 'Denied' ... 'Certified' 'Certified' 'Certified']
```

In [32]:
```
visa_df=pd.read_csv(r'C:\Users\ADMIN\Documents\Naresh it\EDA with python\data fi
con=visa_df['case_status']=='Certified'
visa_df['case_status']=np.where(con,0,1)
visa_df
```

Out[32]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_t
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 [ ]: