

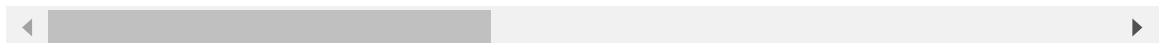
```
In [51]: import pandas as pd
import numpy as np
```

```
In [2]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadataset\visadf")
```

```
Out[2]:
```

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
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



Converting case_id into numerical data

```
In [11]: from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
visadf['case_id'] = le.fit_transform(visadf['case_id'])

le.inverse_transform(visadf['case_id'])

print(visadf['case_id'].values)
print(le.inverse_transform(visadf['case_id']))
```

```
[ 0  1  2 ... 17206 17207 17209]
[ 0  1  2 ... 17206 17207 17209]
```

Converting continent into numerical data

```
In [42]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadataset\visadf")

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
visadf['continent'] = le.fit_transform(visadf['continent'])

valuelist = visadf['continent'].values
```

```
labels = le.inverse_transform(visadf['continent'])

dictContinent=dict(zip(labels,valueList))

dictContinent
```

```
Out[42]: {'Asia': 1,
         'Africa': 0,
         'North America': 3,
         'Europe': 2,
         'South America': 5,
         'Oceania': 4}
```

```
In [43]: visadfContinent = visadf['continent']
visadfContinent
```

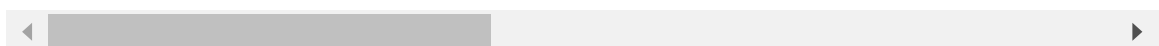
```
Out[43]: 0      1
         1      1
         2      1
         3      1
         4      0
         ..
25475    1
25476    1
25477    1
25478    1
25479    1
Name: continent, Length: 25480, dtype: int32
```

```
In [44]: visadf ## converted continent into numeric data
```

```
Out[44]:
```

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
0	EZYV01	1	High School	N	
1	EZYV02	1	Master's	Y	
2	EZYV03	1	Bachelor's	N	
3	EZYV04	1	Bachelor's	N	
4	EZYV05	0	Master's	Y	
...
25475	EZYV25476	1	Bachelor's	Y	
25476	EZYV25477	1	High School	Y	
25477	EZYV25478	1	Master's	Y	
25478	EZYV25479	1	Master's	Y	
25479	EZYV25480	1	Bachelor's	Y	

25480 rows × 12 columns



Converting education_of_employee into numerical data

```
In [45]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadataset")

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
visadf['education_of_employee'] = le.fit_transform(visadf['education_of_employee'])

valuelist = visadf['education_of_employee'].values

labels = le.inverse_transform(visadf['education_of_employee'])

dictEducation_of_employee = dict(zip(labels, valuelist))

dictEducation_of_employee
```

```
Out[45]: {'High School': 2, "Master's": 3, "Bachelor's": 0, 'Doctorate': 1}
```

```
In [47]: visadfEduOfEmployee = visadf['education_of_employee']
visadfEduOfEmployee
```

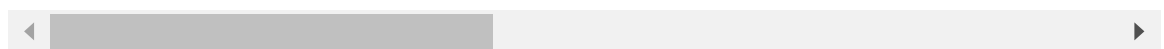
```
Out[47]: 0      2
1      3
2      0
3      0
4      3
..
25475   0
25476   2
25477   3
25478   3
25479   0
Name: education_of_employee, Length: 25480, dtype: int32
```

```
In [48]: visadf ## converted education_of_employee into numeric data
```

Out[48]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
0	EZYV01	Asia		2	N
1	EZYV02	Asia		3	Y
2	EZYV03	Asia		0	N
3	EZYV04	Asia		0	N
4	EZYV05	Africa		3	Y
...
25475	EZYV25476	Asia		0	Y
25476	EZYV25477	Asia		2	Y
25477	EZYV25478	Asia		3	Y
25478	EZYV25479	Asia		3	Y
25479	EZYV25480	Asia		0	Y

25480 rows × 12 columns



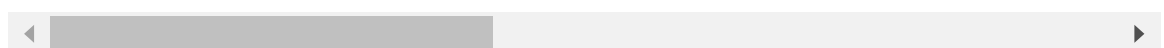
Converting has_job_experience into numerical data

```
In [52]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadataset")
con=visadf['has_job_experience']=='Y'
visadf['has_job_experience']=np.where(con,0,1)
visadf #converted has_job_experience column to numeric
```

Out[52]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
0	EZYV01	Asia	High School	1	
1	EZYV02	Asia	Master's	0	
2	EZYV03	Asia	Bachelor's	1	
3	EZYV04	Asia	Bachelor's	1	
4	EZYV05	Africa	Master's	0	
...
25475	EZYV25476	Asia	Bachelor's	0	
25476	EZYV25477	Asia	High School	0	
25477	EZYV25478	Asia	Master's	0	
25478	EZYV25479	Asia	Master's	0	
25479	EZYV25480	Asia	Bachelor's	0	

25480 rows × 12 columns



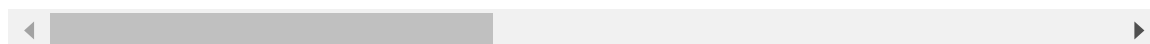
Converting requires_job_training into numerical data

```
In [55]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadataset")
con=visadf['requires_job_training']=='Y'
visadf['requires_job_training']=np.where(con,0,1)
visadf #converted requires_job_training column to numeric
```

```
Out[55]:
```

	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



Converting region_of_employment into numerical data

```
In [56]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadataset")
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
visadf['region_of_employment']=le.fit_transform(visadf['region_of_employment'])

valueList = visadf['region_of_employment'].values

labels = le.inverse_transform(visadf['region_of_employment'])

dictregion=dict(zip(labels,valueList))

dictregion
```

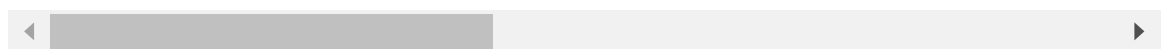
```
Out[56]: {'West': 4, 'Northeast': 2, 'South': 3, 'Midwest': 1, 'Island': 0}
```

```
In [57]: visadf #converted requires_job_training column to numeric
```

Out[57]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
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



Converting unit_of_wage into numerical data

```
In [60]: visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadatabase
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
visadf['unit_of_wage']=le.fit_transform(visadf['unit_of_wage'])

valueList = visadf['unit_of_wage'].values

labels = le.inverse_transform(visadf['unit_of_wage'])

dictUnitofEWage=dict(zip(labels,valueList))

dictUnitofEWage
```

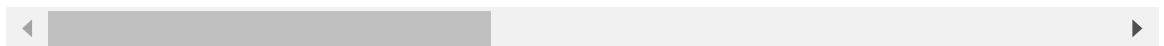
Out[60]: {'Hour': 0, 'Year': 3, 'Week': 2, 'Month': 1}

```
In [61]: visadf      # #converted unit_of_wage column to numeric
```

Out[61]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
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

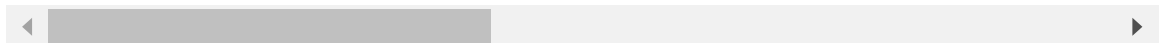


In [62]: `visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadatabase
con=visadf['full_time_position']=='Y'
visadf['full_time_position']=np.where(con,0,1)
visadf #converted requires_job_training column to numeric`

Out[62]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
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 [63]: `visadf = pd.read_csv(r"C:\Users\Anuja_PC\OneDrive\Documents\dataFiles\Visadatabase
con=visadf['full_time_position']=='Y'`

```
visadf['full_time_position']=np.where(con,0,1)
visadf #converted full_time_position column to numeric
```

Out[63]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_1
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