In [1]:
```
import pandas as pd
import numpy as np
```

In [7]:
```
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[7]:

	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 [11]:
```
#standardization
p_wage=visa_df["prevailing_wage"]
p_wage
```

Out[11]:
```
0           592.2029
1         83425.6500
2        122996.8600
3         83434.0300
4        149907.3900
             ...
25475     77092.5700
25476    279174.7900
25477    146298.8500
25478     86154.7700
25479     70876.9100
Name: prevailing_wage, Length: 25480, dtype: float64
```

In [15]:
```
p_mean=np.mean(visa_df["prevailing_wage"])
p_mean
```

Out[15]: 74455.81459209183

In [17]:
```
p_std=np.std(visa_df["prevailing_wage"])
p_std
```

Out[17]: 52814.90589711402

In [21]:
```
z_score=(p_wage-p_mean)/p_std
z_score
```

Out[21]:
```
0       -1.398537
1        0.169835
2        0.919079
3        0.169994
4        1.428604
           ...
25475    0.049924
25476    3.876159
25477    1.360280
25478    0.221509
25479   -0.067763
Name: prevailing_wage, Length: 25480, dtype: float64
```

In [23]:
```
#notalization
p_wage=visa_df["prevailing_wage"]
p_wage,min(p_wage),max(p_wage)
```

Out[23]:
```
(0           592.2029
 1         83425.6500
 2        122996.8600
 3         83434.0300
 4        149907.3900
             ...
 25475     77092.5700
 25476    279174.7900
 25477    146298.8500
 25478     86154.7700
 25479     70876.9100
 Name: prevailing_wage, Length: 25480, dtype: float64,
 2.1367,
 319210.27)
```

In [29]:
```
min_max=(p_wage-min(p_wage))/(max(p_wage)-min(p_wage))
min_max
```

Out[29]:
```
0        0.001849
1        0.261345
2        0.385312
3        0.261371
4        0.469616
           ...
25475    0.241505
25476    0.874579
25477    0.458311
25478    0.269895
25479    0.222033
Name: prevailing_wage, Length: 25480, dtype: float64
```

In [31]:
```
num_columns=visa_df.select_dtypes(exclude="object").columns
num_columns
```

Out[31]:
```
Index(['no_of_employees', 'yr_of_estab', 'prevailing_wage'], dtype='object')
```

In [39]:
```
for i in num_columns:
    min_max=(visa_df[i]-min(visa_df[i]))/(max(visa_df[i])-min(visa_df[i]))
    print(min_max)
```

```
0        0.024147
1        0.004049
2        0.073859
3        0.000206
4        0.001840
           ...
25475    0.004363
25476    0.005481
25477    0.001905
25478    0.003229
25479    0.005350
Name: no_of_employees, Length: 25480, dtype: float64
0        0.958333
1        0.935185
2        0.962963
3        0.449074
4        0.949074
           ...
25475    0.962963
25476    0.953704
25477    0.509259
25478    0.402778
25479    0.740741
Name: yr_of_estab, Length: 25480, dtype: float64
0        0.001849
1        0.261345
2        0.385312
3        0.261371
4        0.469616
           ...
25475    0.241505
25476    0.874579
25477    0.458311
25478    0.269895
25479    0.222033
Name: prevailing_wage, Length: 25480, dtype: float64
```

In [41]:
```
for i in num_columns:

    z_score=(visa_df[i]-visa_df[i].mean())/visa_df[i].std()
    print(z_score)
```

```
0        0.386659
1       -0.142279
2        1.694950
3       -0.243424
4       -0.200413
           ...
25475   -0.134018
25476   -0.104601
25477   -0.198709
25478   -0.163872
25479   -0.108054
Name: no_of_employees, Length: 25480, dtype: float64
0        0.651217
1        0.533201
2        0.674820
3       -1.945148
4        0.604011
           ...
25475    0.674820
25476    0.627614
25477   -1.638304
25478   -2.181181
25479   -0.458139
Name: yr_of_estab, Length: 25480, dtype: float64
0       -1.398510
1        0.169832
2        0.919060
3        0.169991
4        1.428576
           ...
25475    0.049923
25476    3.876083
25477    1.360253
25478    0.221504
25479   -0.067762
Name: prevailing_wage, Length: 25480, dtype: float64
```

In [ ]: