In [1]:
```
# import the packages
# read the data
# divide into numerical and categorical
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

file_path=r'C:\Users\omkar\OneDrive\Documents\Gen_AI\Data_files\Visadataset.csv'
visa_df=pd.read_csv(file_path)

cat=visa_df.select_dtypes(include='object').columns
num=visa_df.select_dtypes(exclude='object').columns
```

prevailing_Wage

In [4]:
```
visa_df['prevailing_wage']
```

Out[4]:
```
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
```

- count
- min
- mean
- medain
- max
- 25p
- 50p
- 75p

In [11]:
```
wage_data=visa_df['prevailing_wage']
Count=len(wage_data)
Min=round(np.min(wage_data),2)
Max=round(np.max(wage_data),2)
Mean=round(np.mean(wage_data),2)
Median=round(np.median(wage_data),2)
index=['Count','Min','Mean','Median','Max']
```

```
values=[Count,Min,Mean,Median,Max]
pd.DataFrame(values,index=index,columns=['prevailing_wage'])
```

Out[11]:

	prevailing_wage
Count	25480.00
Min	2.14
Mean	74455.81
Median	70308.21
Max	319210.27

In [13]:
```
wage_data=visa_df['prevailing_wage']
Count=len(wage_data)
Min=round(wage_data.min(),2)
Max=round(wage_data.max(),2)
Mean=round(wage_data.mean(),2)
Median=round(wage_data.median(),2)
index=['Count','Min','Mean','Median','Max']
values=[Count,Min,Mean,Median,Max]
pd.DataFrame(values,index=index,columns=['prevailing_wage'])
```

Out[13]:

	prevailing_wage
Count	25480.00
Min	2.14
Mean	74455.81
Median	70308.21
Max	319210.27

percentile-quantile

In [16]:
```
np.percentile(wage_data,50)
```

Out[16]: 70308.20999999999

In [18]:
```
np.quantile(wage_data,0.5)
```

Out[18]: 70308.20999999999

In [20]:
```
np.median(wage_data)
```

Out[20]: 70308.20999999999

- 50p=70308.20
- 50 perentage of total data has less than 70308.20
- 50 percentage of 25480 has less than 70309.20
- 12740 applicants has wages less than 70309.20

In [28]:
```
wage_data=visa_df['prevailing_wage']
p_50=np.percentile(wage_data,50)
con=wage_data<p_50
len(visa_df[con])
```

Out[28]: 12740

In [33]:
```
wage_data=visa_df['prevailing_wage']
p_25=np.percentile(wage_data,25)
con=wage_data<p_25
len(visa_df[con])==25*25480/100
```

Out[33]: True

In [35]:
```
wage_data=visa_df['prevailing_wage']
p_75=np.percentile(wage_data,75)
con=wage_data<p_75
len(visa_df[con])==75*25480/100
```

Out[35]: True

In [37]:
```
wage_data=visa_df['prevailing_wage']
Count=len(wage_data)
Min=round(np.min(wage_data),2)
Mean=round(np.mean(wage_data),2)
Median=round(np.median(wage_data),2)
p_25=round(np.percentile(wage_data,25),2)
p_50=round(np.percentile(wage_data,50),2)
p_75=round(np.percentile(wage_data,75),2)
Max=round(np.max(wage_data),2)
index=['Count','Min','Mean','Median',
       '25P','50p','75P','Max']
values=[Count,Min,Mean,Median,p_25,p_50,p_75,Max]
pd.DataFrame(values,index=index,
             columns=['prevailing_wage'])
```

Out[37]:

	prevailing_wage
Count	25480.00
Min	2.14
Mean	74455.81
Median	70308.21
25P	34015.48
50p	70308.21
75P	107735.51
Max	319210.27

In [39]:
```
l=[]
for i in num:
    data=visa_df[i]
    Count=len(data)
```

```
    Min=round(np.min(data),2)
    Mean=round(np.mean(data),2)
    Median=round(np.median(data),2)
    p_25=round(np.percentile(data,25),2)
    p_50=round(np.percentile(data,50),2)
    p_75=round(np.percentile(data,75),2)
    Max=round(np.max(data),2)
    index=['Count','Min','Mean','Median',
           '25P','50p','75P','Max']
    values=[Count,Min,Mean,Median,p_25,p_50,p_75,Max]
    l.append(values)
```

In [41]:
```
l
```

Out[41]:
```
[[25480, -26, 5667.04, 2109.0, 1022.0, 2109.0, 3504.0, 602069],
 [25480, 1800, 1979.41, 1997.0, 1976.0, 1997.0, 2005.0, 2016],
 [25480, 2.14, 74455.81, 70308.21, 34015.48, 70308.21, 107735.51, 319210.27]]
```

In [51]:
```
pd.DataFrame(l,
             index=num,
             columns=index).T
```

Out[51]:

	no_of_employees	yr_of_estab	prevailing_wage
Count	25480.00	25480.00	25480.00
Min	-26.00	1800.00	2.14
Mean	5667.04	1979.41	74455.81
Median	2109.00	1997.00	70308.21
25P	1022.00	1976.00	34015.48
50p	2109.00	1997.00	70308.21
75P	3504.00	2005.00	107735.51
Max	602069.00	2016.00	319210.27

In [53]:
```
visa_df.describe()
```

Out[53]:

	no_of_employees	yr_of_estab	prevailing_wage
count	25480.000000	25480.000000	25480.000000
mean	5667.043210	1979.409929	74455.814592
std	22877.928848	42.366929	52815.942327
min	-26.000000	1800.000000	2.136700
25%	1022.000000	1976.000000	34015.480000
50%	2109.000000	1997.000000	70308.210000
75%	3504.000000	2005.000000	107735.512500
max	602069.000000	2016.000000	319210.270000

In [ ]: