```
In [33]: import pandas as pd
         import matplotlib.pyplot as plt
         import numpy as np
         import seaborn as sns
```

```
In [34]: path= r"C:\Users\tanma\DATASCIENCE\string & list\Visadataset.xlsx"
         visa_df =pd.read_excel(path)
         visa_df.head(5)
```

Out[34]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training	no_
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	

◀ ▬▬▬▬▬▬▬▬▬▬ ▶

readingaspeci f iccolumn

```
In [35]: visa_df['education_of_employee']
```

```
Out[35]: 0        High School
         1           Master's
         2         Bachelor's
         3         Bachelor's
         4           Master's
                    ...
         25475     Bachelor's
         25476     High School
         25477       Master's
         25478       Master's
         25479     Bachelor's
         Name: education_of_employee, Length: 25480, dtype: object
```

```
In [36]: visa_df[['education_of_employee']]    #dataframe
```

Out[36]:

	education_of_employee
0	High School
1	Master's
2	Bachelor's
3	Bachelor's
4	Master's
...	...
25475	Bachelor's
25476	High School
25477	Master's
25478	Master's
25479	Bachelor's

25480 rows × 1 columns

```
In [37]: visa_df.education_of_employee   #series
```

```
Out[37]: 0        High School
         1           Master's
         2         Bachelor's
         3         Bachelor's
         4           Master's
                     ...
         25475     Bachelor's
         25476    High School
         25477       Master's
         25478       Master's
         25479     Bachelor's
         Name: education_of_employee, Length: 25480, dtype: object
```

```
In [38]: visa_df.columns
```

```
Out[38]: Index(['case_id', 'continent', 'education_of_employee', 'has_job_experienc
         e',
                'requires_job_training', 'no_of_employees', 'yr_of_estab',
                'region_of_employment', 'prevailing_wage', 'unit_of_wage',
                'full_time_position', 'case_status'],
               dtype='object')
```

```
In [ ]:
```

```
In [ ]:
```

```
In [39]: cols=['education_of_employee','has_job_experience','requires_job_training']
         visa_df[cols]
```

Out[39]:

	education_of_employee	has_job_experience	requires_job_training
0	High School	N	N
1	Master's	Y	N
2	Bachelor's	N	Y
3	Bachelor's	N	N
4	Master's	Y	N
...	...	...	...
25475	Bachelor's	Y	Y
25476	High School	Y	N
25477	Master's	Y	N
25478	Master's	Y	Y
25479	Bachelor's	Y	N

25480 rows × 3 columns

```
In [40]: visa_df.values
```

Out[40]:
```
array([['EZYV01', 'Asia', 'High School', ..., 'Hour', 'Y', 'Denied'],
       ['EZYV02', 'Asia', "Master's", ..., 'Year', 'Y', 'Certified'],
       ['EZYV03', 'Asia', "Bachelor's", ..., 'Year', 'Y', 'Denied'],
       ...,
       ['EZYV25478', 'Asia', "Master's", ..., 'Year', 'N', 'Certified'],
       ['EZYV25479', 'Asia', "Master's", ..., 'Year', 'Y', 'Certified'],
       ['EZYV25480', 'Asia', "Bachelor's", ..., 'Year', 'Y', 'Certifie
d']],
      dtype=object)
```

$educationofemployee$

```
In [41]: col=['education_of_employee']
         visa_df[col]
```

Out[41]:

	education_of_employee
0	High School
1	Master's
2	Bachelor's
3	Bachelor's
4	Master's
...	...
25475	Bachelor's
25476	High School
25477	Master's
25478	Master's
25479	Bachelor's

25480 rows × 1 columns

unique

```
In [42]: visa_df['education_of_employee'].unique()
```

Out[42]: array(['High School', "Master's", "Bachelor's", 'Doctorate'], dtype=objec
         t)

```
In [43]: set(visa_df['education_of_employee'].values)
```

Out[43]: {"Bachelor's", 'Doctorate', 'High School', "Master's"}

```
In [44]: visa_df['education_of_employee'].unique()
```

Out[44]: array(['High School', "Master's", "Bachelor's", 'Doctorate'], dtype=objec
         t)

```
In [ ]:
```

```
In [45]:    ## Q1)out of total observations How many bachelor's degree observations are

            con=visa_df['education_of_employee']=="Bachelor's"
            visa_df[con]
```

Out[45]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_trainii
2	EZYV03	Asia	Bachelor's	N	
3	EZYV04	Asia	Bachelor's	N	
6	EZYV07	Asia	Bachelor's	N	
7	EZYV08	North America	Bachelor's	Y	
8	EZYV09	Asia	Bachelor's	N	
...	...	...	...	...	...
25466	EZYV25467	Europe	Bachelor's	Y	
25468	EZYV25469	Asia	Bachelor's	N	
25473	EZYV25474	Asia	Bachelor's	Y	
25475	EZYV25476	Asia	Bachelor's	Y	
25479	EZYV25480	Asia	Bachelor's	Y	

10234 rows × 12 columns

```
In [46]:    con=visa_df['education_of_employee']=="Bachelor's"
            len(visa_df[con])
```

Out[46]: 10234

```
In [47]:    con=visa_df['education_of_employee']=="Master's"
            len(visa_df[con])
```

Out[47]: 9634

```
In [48]:    con=visa_df['education_of_employee']=="High school"
            len(visa_df[con])
```

Out[48]: 0

```
In [49]:    unique_labels=visa_df['education_of_employee'].unique()
            for i in unique_labels:
                con=visa_df['education_of_employee']==i
                print(i,":",len(visa_df['education_of_employee']))

            High School : 25480
            Master's : 25480
            Bachelor's : 25480
            Doctorate : 25480
```

```
In [ ]:
```

frequencytable

```
In [50]: unique_labels=visa_df['education_of_employee'].unique()
         count=[]
         for i in unique_labels:
             con=visa_df['education_of_employee']==i
             count.append(len(visa_df[con]))

         education_of_employee_df=pd.DataFrame(zip(unique_labels,count),
                                           columns=['education_of_employee','coun
         education_of_employee_df.to_csv('education_of_employee_df.excel',index=False
```

```
In [51]: education_of_employee_df
```

Out[51]:

	education_of_employee	count
0	High School	3420
1	Master's	9634
2	Bachelor's	10234
3	Doctorate	2192

```
In [ ]:
```

valuecounts

```
In [55]: education_of_employee_vc=visa_df['education_of_employee'].value_counts()
         education_of_employee_vc
```

```
Out[55]: education_of_employee
         Bachelor's      10234
         Master's         9634
         High School      3420
         Doctorate        2192
         Name: count, dtype: int64
```

```
In [59]: education_of_employee_vc.keys()
```

```
Out[59]: Index(['Bachelor's', 'Master's', 'High School', 'Doctorate'], dtype='objec
         t', name='education_of_employee')
```

```
In [58]: education_of_employee_vc.values
```

```
Out[58]: array([10234,  9634,  3420,  2192], dtype=int64)
```

```
In [62]: education_of_employee_vc=visa_df['education_of_employee'].value_counts()
         l1=education_of_employee_vc.keys()
         l2=education_of_employee_vc.values
         education_of_employee_vc_df=pd.DataFrame(zip(l1,l2),
                                        columns=['education_of_employee','co
         education_of_employee_df
```

Out[62]:

	education_of_employee	count
0	High School	3420
1	Master's	9634
2	Bachelor's	10234
3	Doctorate	2192

```
In [63]: education_of_employee_vc
```

Out[63]:
```
education_of_employee
Bachelor's    10234
Master's       9634
High School    3420
Doctorate      2192
Name: count, dtype: int64
```

```
In [64]: education_of_employee_df
```

Out[64]:

	education_of_employee	count
0	High School	3420
1	Master's	9634
2	Bachelor's	10234
3	Doctorate	2192

barchart

```
In [65]: education_of_employee_vc_df
```

Out[65]:

	education_of_employee	count
0	Bachelor's	10234
1	Master's	9634
2	High School	3420
3	Doctorate	2192

```
In [68]: plt.figure(figsize=(10,4))
         plt.bar('education_of_employee','count',
                 data=education_of_employee_vc_df)
         plt.xlable('education_of_employee')
         plt.ylable('count')
         plt.titlebar('bar chart')
         plt.savefig('education_of_employee.jpg')
         plt.show()
```
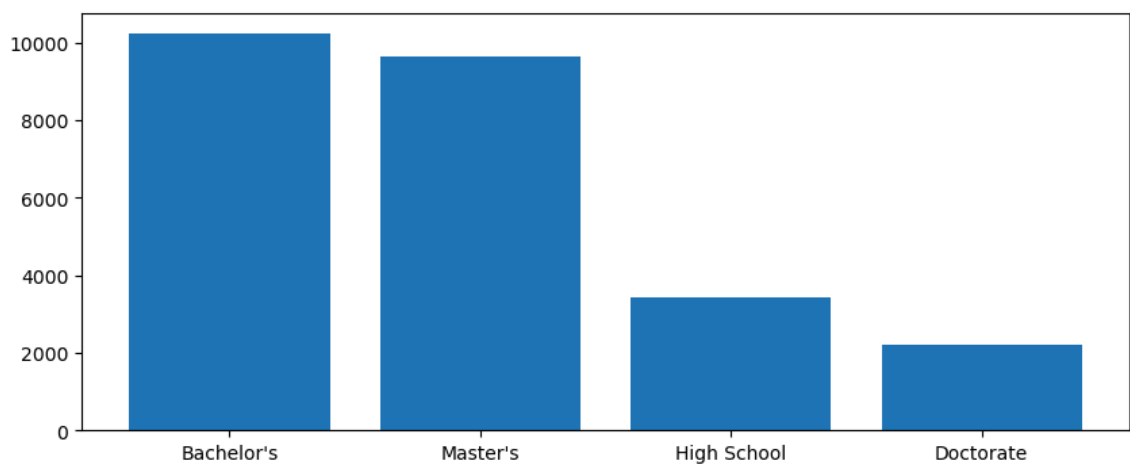
```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Cell In[68], line 4
      1 plt.figure(figsize=(10,4))
      2 plt.bar('education_of_employee','count',
      3         data=education_of_employee_vc_df)
----> 4 plt.xlable('education_of_employee')
      5 plt.ylable('count')
      6 plt.titlebar('bar chart')

AttributeError: module 'matplotlib.pyplot' has no attribute 'xlable'
```
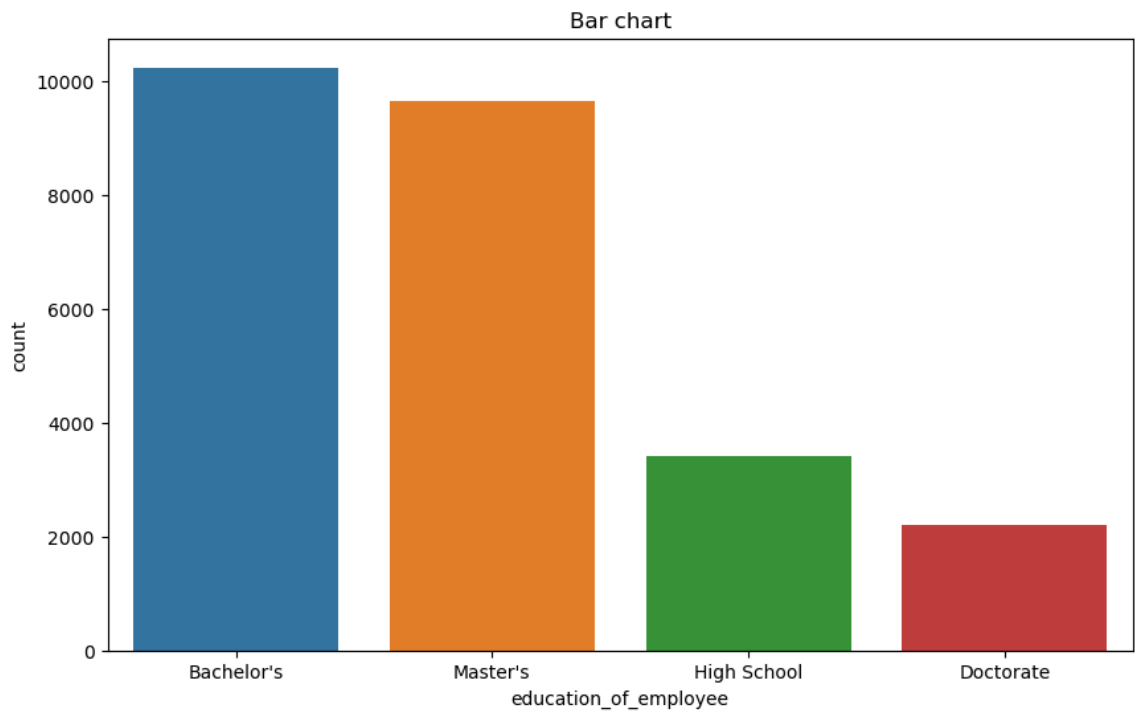
```
In [69]:  plt.figure(figsize=(10,6))
          l=visa_df['education_of_employee'].value_counts().keys()
          sns.countplot(data=visa_df,
          x='education_of_employee',
          order=l)
          plt.xlabel("education_of_employee") # x-axis name
          plt.ylabel('count') # y-axis name
          plt.title("Bar chart") # title of the chart
          plt.savefig('education_of_employee_bar.jpg')
          plt.show()
```

Bar chart

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: