In [3]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df=pd.read_csv("data science salaries.csv")
print(df)
```

```
      work_year experience_level employment_type                 job_title  \
0          2023               SE              FT  Principal Data Scientist   
1          2023               MI              CT               ML Engineer   
2          2023               MI              CT               ML Engineer   
3          2023               SE              FT            Data Scientist   
4          2023               SE              FT            Data Scientist   
...         ...              ...             ...                       ...   
3750       2020               SE              FT            Data Scientist   
3751       2021               MI              FT  Principal Data Scientist   
3752       2020               EN              FT            Data Scientist   
3753       2020               EN              CT     Business Data Analyst   
3754       2021               SE              FT      Data Science Manager   

       salary salary_currency  salary_in_usd employee_residence  remote_ratio  \
0       80000             EUR          85847                 ES           100   
1       30000             USD          30000                 US           100   
2       25500             USD          25500                 US           100   
3      175000             USD         175000                 CA           100   
4      120000             USD         120000                 CA           100   
...       ...             ...            ...                ...           ...   
3750   412000             USD         412000                 US           100   
3751   151000             USD         151000                 US           100   
3752   105000             USD         105000                 US           100   
3753   100000             USD         100000                 US           100   
3754  7000000             INR          94665                 IN            50   

     company_location company_size  
0                  ES            L  
1                  US            S  
2                  US            S  
3                  CA            M  
4                  CA            M  
...               ...          ...  
3750               US            L  
3751               US            L  
3752               US            S  
3753               US            L  
3754               IN            L  

[3755 rows x 11 columns]
```

In [33]:

```
c=df.groupby('employee_residence').get_group('US')
print('No. of employees from USA',len(c))
```

```
No. of employees from USA 3004
```

In [3]:

```
India=df.groupby('employee_residence').get_group('IN')
print(India)
```

```
      work_year experience_level employment_type  \
41         2022               MI              FT   
82         2023               MI              FT   
83         2022               EN              FT   
156        2023               MI              FT   
217        2023               EN              FT   
...         ...              ...             ...   
3689       2020               MI              FT   
3705       2021               EN              FT   
3729       2021               EN              FT   
3734       2021               MI              FT   
3754       2021               SE              FT   

                              job_title   salary salary_currency  \
41            Machine Learning Engineer  1650000             INR   
82    Applied Machine Learning Engineer    65000             EUR   
83                         AI Developer   300000             USD   
156              Applied Data Scientist  1700000             INR   
217                       Data Engineer  1400000             INR   
...                                 ...      ...             ...   
3689               Product Data Analyst   450000             INR   
3705                  Big Data Engineer   435000             INR   
3729                       AI Scientist  1335000             INR   
3734                  Lead Data Analyst  1450000             INR   
3754               Data Science Manager  7000000             INR   

      salary_in_usd employee_residence  remote_ratio company_location  \
41            20984                 IN            50               IN   
82            69751                 IN           100               DE   
83           300000                 IN            50               IN   
156           20670                 IN           100               IN   
217           17022                 IN           100               IN   
...             ...                ...           ...              ...   
3689           6072                 IN           100               IN   
3705           5882                 IN             0               CH   
3729          18053                 IN           100               AS   
3734          19609                 IN           100               IN   
3754          94665                 IN            50               IN   

     company_size  
41              L  
82              S  
83              L  
156             L  
217             L  
...           ...  
3689            L  
3705            L  
3729            S  
3734            L  
3754            L  

[71 rows x 11 columns]
```

In [4]:

```
print("Average salary of Indians in USD is",India['salary_in_usd'].mean())
```

```
Average salary of Indians in USD is 36218.45070422535
```

In [5]:

```
Companies_in_India=df.groupby('company_location').get_group('IN')
Posts_in_India=Companies_in_India['job_title'].tolist()
```

```
Posts_in_India=set(Posts_in_India)
print("Posts available for people in data science in India is")
print(Posts_in_India)
```

```
Posts available for people in data science in India is
{'Power BI Developer', 'Data Engineer', 'Head of Machine Learning', 'Machine Lear
ning Research Engineer', 'Data Scientist', 'Applied Machine Learning Scientist',
'Lead Machine Learning Engineer', 'Applied Data Scientist', 'Data Science Consult
ant', 'Business Data Analyst', 'Computer Vision Engineer', 'Machine Learning Engi
neer', 'Lead Data Scientist', 'BI Data Analyst', 'AI Developer', 'Lead Data Analy
st', 'Data Science Manager', 'Head of Data Science', 'Research Scientist', 'Data
Analyst', 'Principal Data Architect', 'Big Data Engineer', 'Product Data Analys
t', '3D Computer Vision Researcher'}
```

In [6]:

```
companylocation=df['company_location'].tolist()
companylocation=set(companylocation)
companylocation=list(companylocation)
```

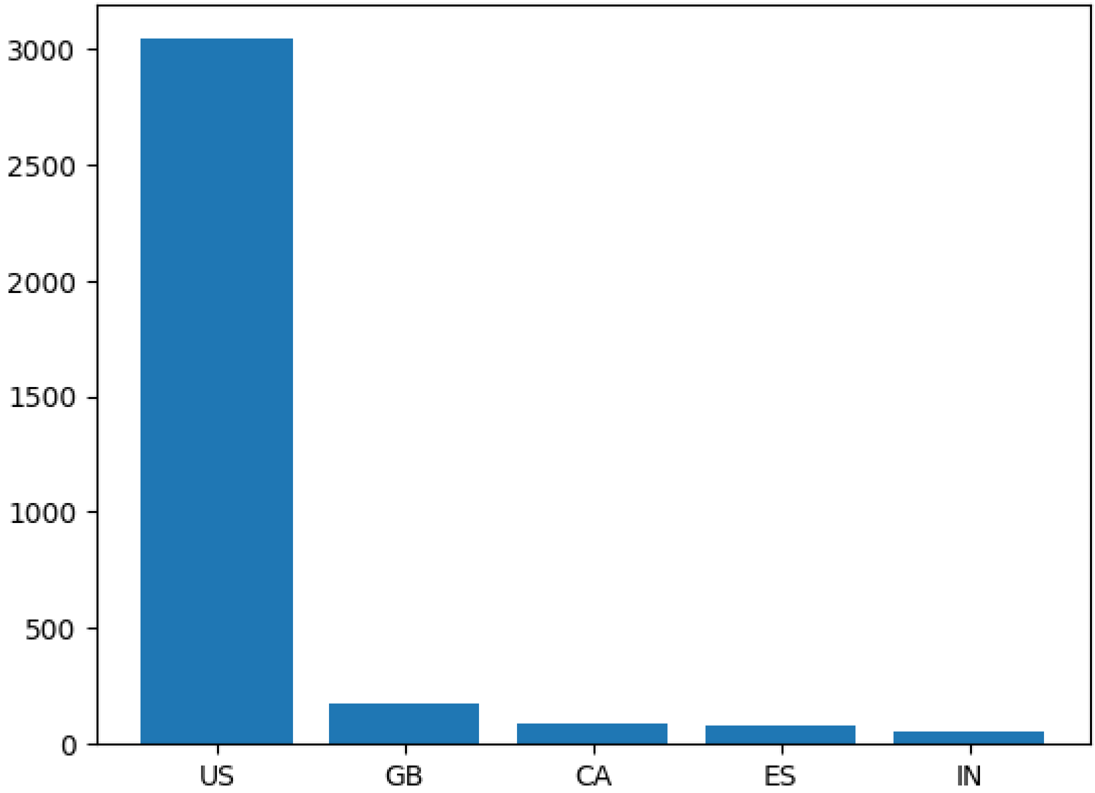
In [7]:

```
no_of_employees=[]
descendingorder=[]
for i in range(len(companylocation)):
    c=df.groupby('company_location').get_group(companylocation[i])
    d=len(c)
    no_of_employees.append(d)
    descendingorder.append(d)
no_of_employees.sort(reverse=True)
favlocation=[]
order=[]
for i in range(5):
    for  j in range(len(companylocation)):
        if(descendingorder[j]==no_of_employees[i]):
            favlocation.append(companylocation[j])
            order.append(descendingorder[j])
print("Top 5 locatons for companies are")
plt.bar(favlocation,order)
```

```
Top 5 locatons for companies are
```

Out[7]:

```
<BarContainer object of 5 artists>
```



In [3]:
```
c=input("enter the name of the post")
d=df.groupby('job_title').get_group(c)
print("The max salary for this post is",d['salary_in_usd'].max())
print("The mean salary for this post is",d['salary_in_usd'].mean())
print("The median salary for this post is",d['salary_in_usd'].median())
```

```
enter the name of the postApplied Scientist
The max salary for this post is 350000
The mean salary for this post is 190264.4827586207
The median salary for this post is 191737.5
```

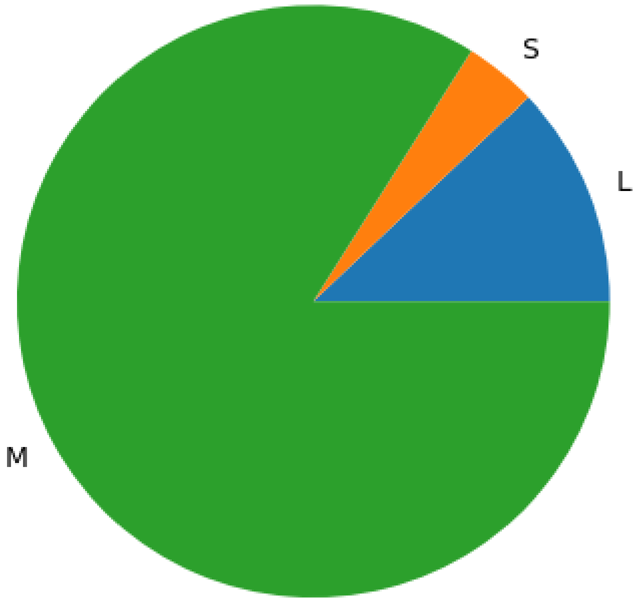
In [5]:
```
c=input("enter the code of country")
d=df.groupby('employee_residence').get_group(c)
print('Max salary offered in the country is',d['salary_in_usd'].max())
print('Mean salary offered in the country is',d['salary_in_usd'].mean())
print('Median salary offered in the country is',d['salary_in_usd'].median())
Companies=df.groupby('company_location').get_group(c)
Posts=Companies['job_title'].tolist()
Posts=set(Posts)
print("Posts available for people in data science are")
print(Posts)
```

```
enter the code of countryJP
Max salary offered in the country is 260000
Mean salary offered in the country is 103537.71428571429
Median salary offered in the country is 74000.0
Posts available for people in data science are
{'Machine Learning Engineer', 'Machine Learning Scientist', 'ML Engineer', 'Data
Engineer', 'Director of Data Science'}
```

In [7]:
```
c=df.groupby('company_size').get_group('L')
l=len(c)
d=df.groupby('company_size').get_group('S')
m=len(d)
e=df.groupby('company_size').get_group('M')
```

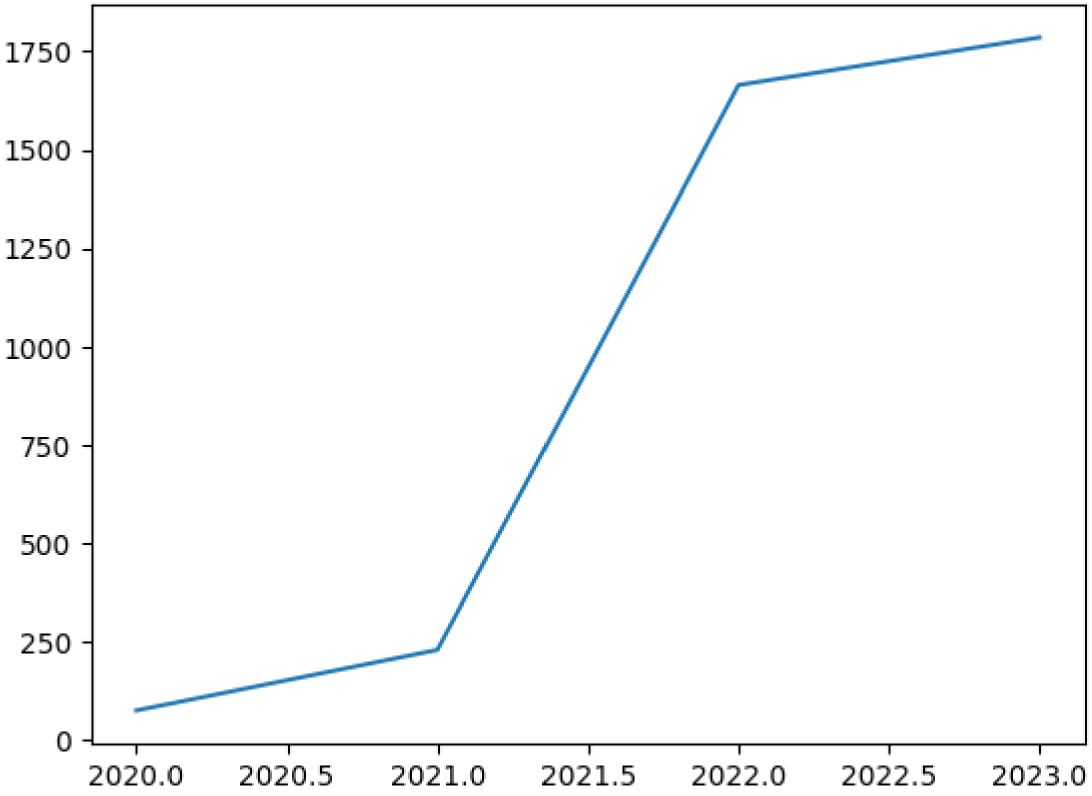
```
n=len(e)
a=np.array([l,m,n])
b=np.array(['L','S','M'])
plt.pie(a,labels=b)
```

```
Out[7]: ([<matplotlib.patches.Wedge at 0x1c40b5ae190>,
          <matplotlib.patches.Wedge at 0x1c40b609b90>,
          <matplotlib.patches.Wedge at 0x1c40b60ac90>],
         [Text(1.0215981211725451, 0.4078446748662119, 'L'),
          Text(0.6978993053093786, 0.850256761013217, 'S'),
          Text(-0.9634048160461234, -0.5308965628247514, 'M')])
```



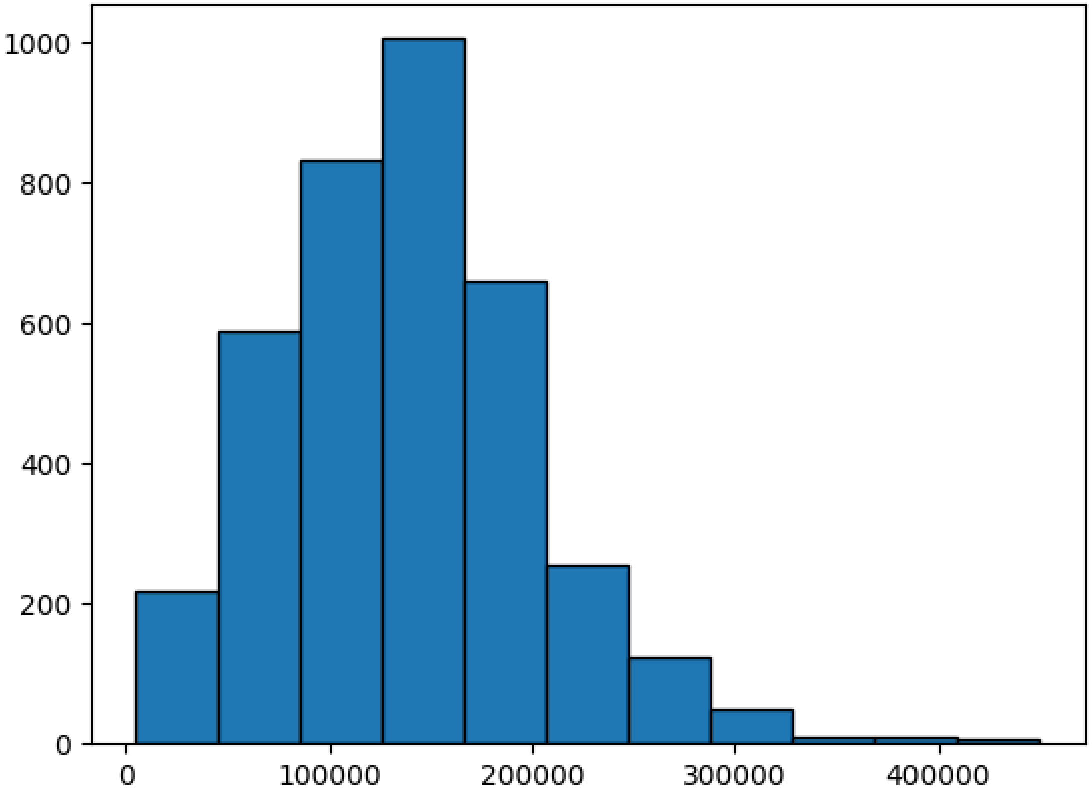
```
In [25]: c=df['work_year'].tolist()
         c=set(c)
         c=list(c)
         workemployee=[]
         for i in range(len(c)):
             d=df.groupby('work_year').get_group(c[i])
             e=len(d)
             workemployee.append(e)
         plt.plot(c,workemployee)
```

```
Out[25]: [<matplotlib.lines.Line2D at 0x1e4933172d0>]
```



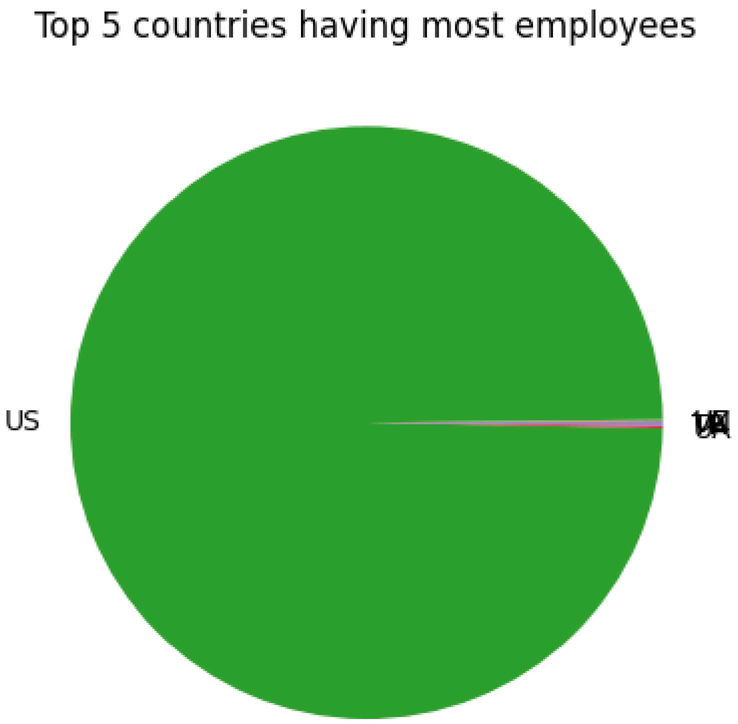
In [7]:
```
d=df['salary_in_usd']
plt.hist(d,bins=11,edgecolor='black')
```

Out[7]:
```
(array([ 218.,  590.,  832., 1005.,  661.,  255.,  123.,   49.,    8.,
           8.,    6.]),
 array([  5132.        ,  45574.54545455,  86017.09090909, 126459.63636364,
        166902.18181818, 207344.72727273, 247787.27272727, 288229.81818182,
        328672.36363636, 369114.90909091, 409557.45454545, 450000.        ]),
 <BarContainer object of 11 artists>)
```



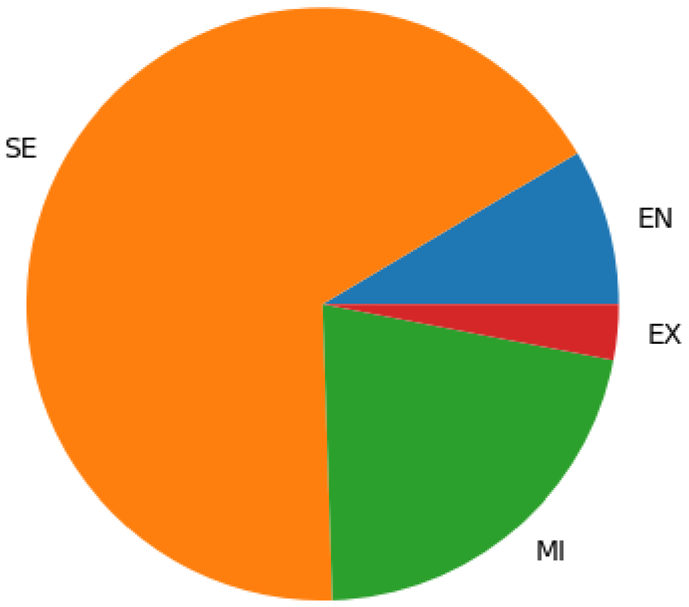
In [12]:
```
c=df['employee_residence'].tolist()
c=set(c)
c=list(c)
country=[]
employee=[]
c.sort(reverse=True)
for i in range(5):
    a=df.groupby('employee_residence').get_group(c[i])
    a=len(a)
    employee.append(a)
    country.append(c[i])
plt.pie(employee,labels=country)
plt.title('Top 5 countries having most employees')
```

Out[12]:
```
Text(0.5, 1.0, 'Top 5 countries having most employees')
```



In [32]:
```
c=df['experience_level']
c=set(c)
c=list(c)
experienceno=[]
for i in range(len(c)):
    a=df.groupby('experience_level').get_group(c[i])
    a=len(a)
    experienceno.append(a)
plt.pie(experienceno,labels=c)
```

Out[32]:
```
([<matplotlib.patches.Wedge at 0x1c40f24e690>,
  <matplotlib.patches.Wedge at 0x1c40f24f3d0>,
  <matplotlib.patches.Wedge at 0x1c40f240890>,
  <matplotlib.patches.Wedge at 0x1c40f241b50>],
 [Text(1.0608125642967916, 0.2909926174837866, 'EN'),
  Text(-0.9647343017648791, 0.5284767989214204, 'SE'),
  Text(0.7141304120336674, -0.8366706368748844, 'MI'),
  Text(1.0950005456880005, -0.10475593034755154, 'EX')])
```



In [24]:

```
a=np.mean(df['salary_in_usd'])
print("Mean salary in data science is",a)
```

```
Mean salary in data science is 137570.38988015978
```

In [25]:

```
b=np.median(df['salary_in_usd'])
print("Median salary in data science is",b)
```

```
Median salary in data science is 135000.0
```

In [34]:

```
c=np.min(df['salary_in_usd'])
print("Minimum salary is",c)
```

```
Minimum salary is 5132
```

In [38]:

```
c=np.std(df['salary_in_usd'])
print("Standard difference in salary is",c)
```

```
Standard difference in salary is 63047.228497405435
```

In [41]:

```
d=np.count_nonzero(df['work_year']==2020)
e=np.count_nonzero(df['work_year']==2021)
f=np.count_nonzero(df['work_year']==2022)
g=np.count_nonzero(df['work_year']==2023)
print(d,e,f,g)
```

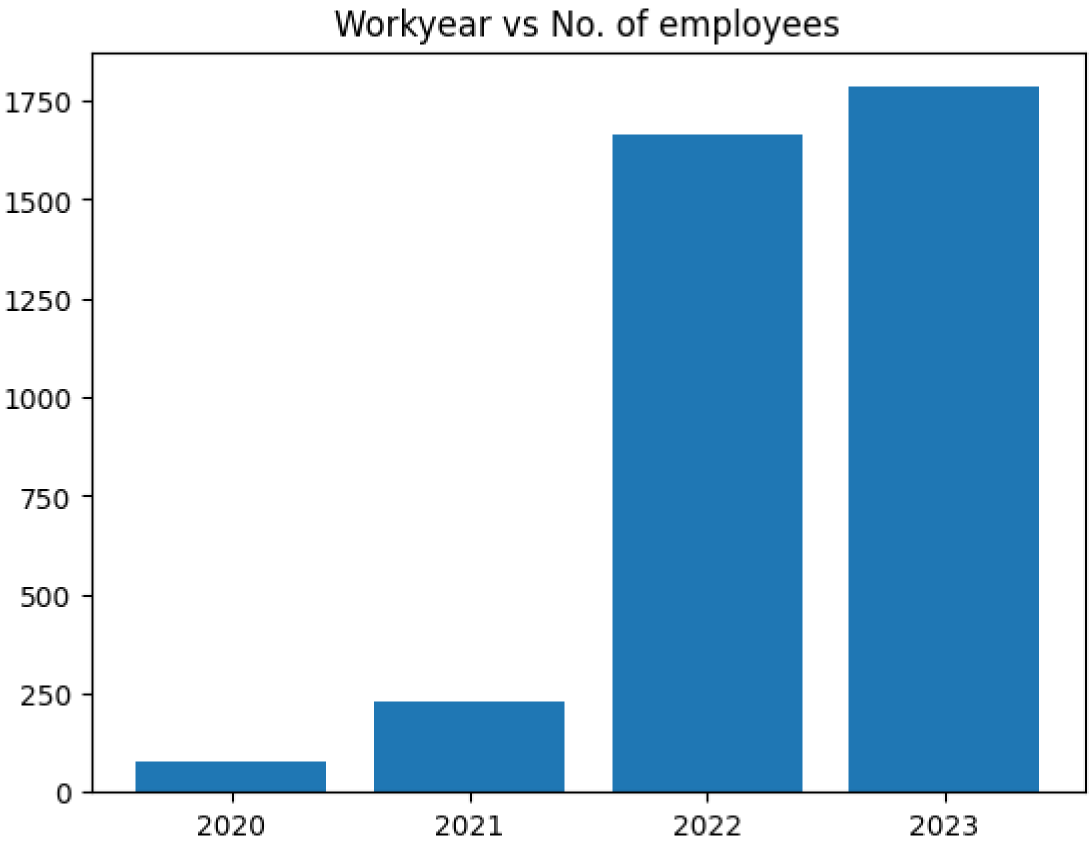
```
76 230 1664 1785
```

In [46]:

```
c=np.array([d,e,f,g])
a=np.array(['2020','2021','2022','2023'])
plt.bar(a,c)
plt.title('Workyear vs No. of employees')
```

Out[46]:

```
Text(0.5, 1.0, 'Workyear vs No. of employees')
```


Workyear vs No. of employees
1750
1500
1250
1000
750
500
250
0
2020
2021
2022
2023

```
In [ ]:
```