In [ ]:
```
Name=Yash Gunjal
Rollno=724
PRN=202201040106
```

assignment 5

In [4]:
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

In [5]:
```
df=pd.read_csv("sales_data_sample.csv")
df
```

Out[5]:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE
0	10107	30	95.70	2	2871.00	2/24/2003 0:00
1	10121	34	81.35	5	2765.90	5/7/2003 0:00
2	10134	41	94.74	2	3884.34	7/1/2003 0:00
3	10145	45	83.26	6	3746.70	8/25/2003 0:00
4	10159	49	100.00	14	5205.27	10/10/2003 0:00
...	...	...	...	...	...	...
2818	10350	20	100.00	15	2244.40	12/2/2004 0:00
2819	10373	29	100.00	1	3978.51	1/31/2005 0:00
2820	10386	43	100.00	4	5417.57	3/1/2005 0:00
2821	10397	34	62.24	1	2116.16	3/28/2005 0:00
2822	10414	47	65.52	9	3079.44	5/6/2005 0:00

2823 rows × 16 columns

In [6]:
```
df.columns
```

Out[6]:
```
Index(['ORDERNUMBER', 'QUANTITYORDERED', 'PRICEEACH', 'ORDERLINENUMBER',
       'SALES', 'ORDERDATE', 'STATUS', 'QTR_ID', 'MONTH_ID', 'YEAR_ID',
```

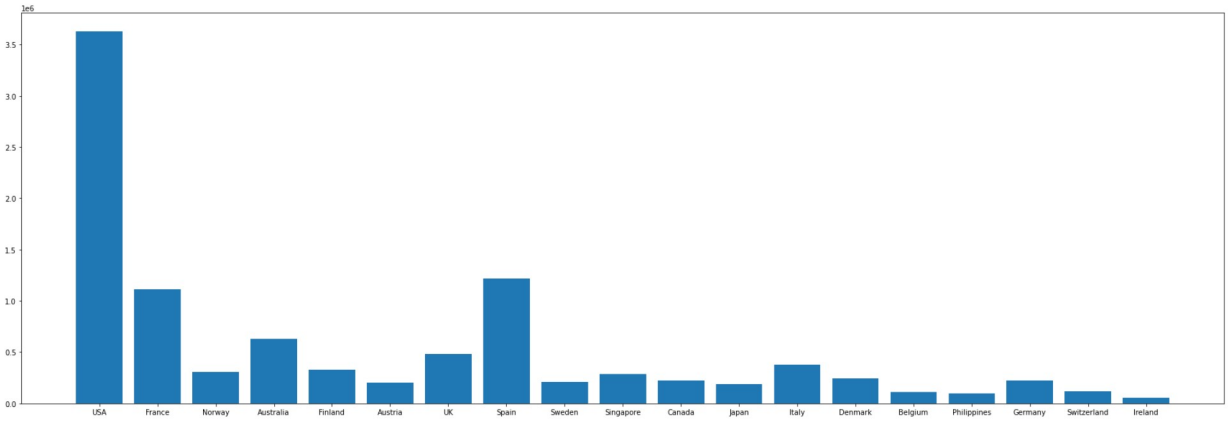
```
       'PRODUCTLINE', 'MSRP', 'PRODUCTCODE', 'CUSTOMERNAME', 'COUNTRY',
       'DEALSIZE'],
```

In [7]:
```
newdf=df.groupby('COUNTRY')
country=df['COUNTRY'].unique()
sum(newdf.get_group('USA')['SALES'])
```

Out[7]: 3627982.83

In [8]:
```
newdf=df.groupby('COUNTRY')
country=df['COUNTRY'].unique()
sales=[]
for cname in country:
    sales.append(sum(newdf.get_group(cname)['SALES']))

f = plt.figure()
f.set_figwidth(30)
f.set_figheight(10)

font1 = {'family':'serif','colour':'blue','size':20}
font1 = {'family':'serif','colour':'darkred','size':15}
plt.bar(country,sales,label="country wise sales")
```

Out[8]: <BarContainer object of 19 artists>



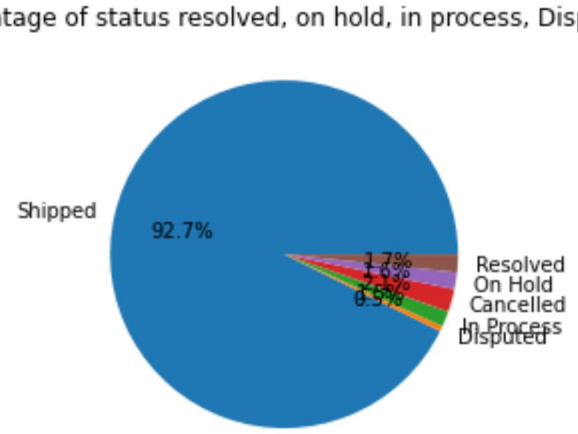
In [9]:
```
dsize=df['DEALSIZE'].unique()
deal=[]
newdf=df.groupby('DEALSIZE')
for dname in dsize:
    deal.append(newdf.get_group(dname)['DEALSIZE'].count())

plt.bar(df['DEALSIZE'].unique(),deal, label="DEALSIZE")
plt.legend(loc="best")
```

Out[9]: <matplotlib.legend.Legend at 0x1f84a2aea30>

In [10]:
```
newdf=df.groupby('STATUS')
tot=df['STATUS'].count()
status=df['STATUS'].unique()
percent=[]
for sname in status:
    percent.append(newdf.get_group(sname)['STATUS'].count()*100/tot)
```

In [11]:
```
plt.pie(percent, labels=status,autopct='%1.1f%%')
plt.title('Percentage of status resolved, on hold, in process, Disputed')
```

Out[11]: Text(0.5, 1.0, 'Percentage of status resolved, on hold, in process, Disputed')



In [13]:
```
plt.scatter(df['QUANTITYORDERED'],df['SALES'])
```

Out[13]: <matplotlib.collections.PathCollection at 0x1f84a5c81f0>