In [2]:
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

In [4]:
```
df=pd.read_csv("sales_data_sample.csv")
df
```

Out[4]:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDER
0	10107	30	95.70	2	2871.00	2/2
1	10121	34	81.35	5	2765.90	5/7/200
2	10134	41	94.74	2	3884.34	7/1/200
3	10145	45	83.26	6	3746.70	8/2
4	10159	49	100.00	14	5205.27	10/1
...	...	...	...	...	...	
2818	10350	20	100.00	15	2244.40	12/
2819	10373	29	100.00	1	3978.51	1/3
2820	10386	43	100.00	4	5417.57	3/1/200
2821	10397	34	62.24	1	2116.16	3/2
2822	10414	47	65.52	9	3079.44	5/6/200

2823 rows × 16 columns

In [5]:
```
df.columns
```

Out[5]:
```
Index(['ORDERNUMBER', 'QUANTITYORDERED', 'PRICEEACH', 'ORDERLINENUMBER',
       'SALES', 'ORDERDATE', 'STATUS', 'QTR_ID', 'MONTH_ID', 'YEAR_ID',
       'PRODUCTLINE', 'MSRP', 'PRODUCTCODE', 'CUSTOMERNAME', 'COUNTRY',
       'DEALSIZE'],
      dtype='object')
```

In [10]:
```
newdf=df.groupby('COUNTRY')
country=df['COUNTRY'].unique()
sum(newdf.get_group('USA')['SALES'])
```

Out[10]:
```
3627982.83
```

In [22]:
```
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)
```

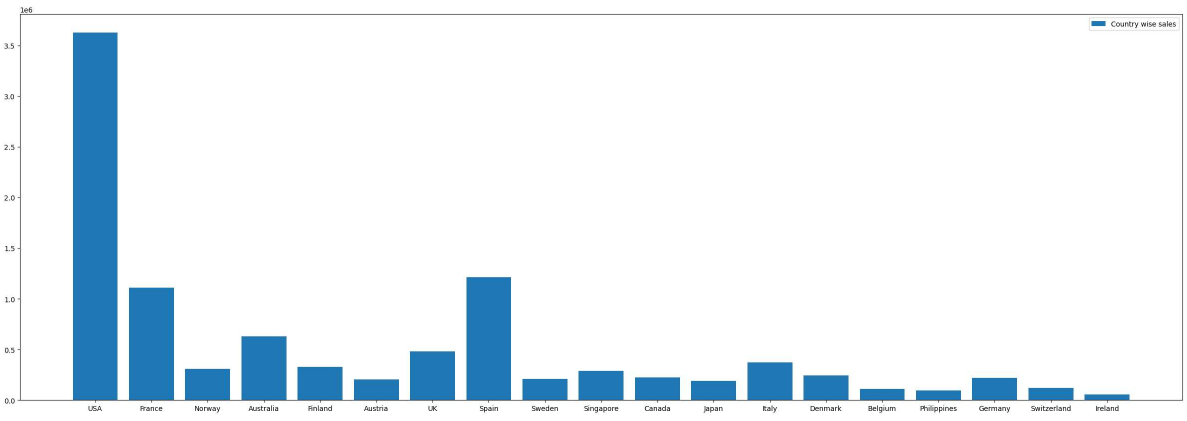
```
<Figure size 3000x1000 with 0 Axes>
```

In [23]:
```
#1. Identify Country wise sales
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','color':'blue','size':20}
font2 = {'family':'serif','color':'darked','size':15}
plt.bar(country, sales, label="Country wise sales")
plt.legend(loc="best")
```

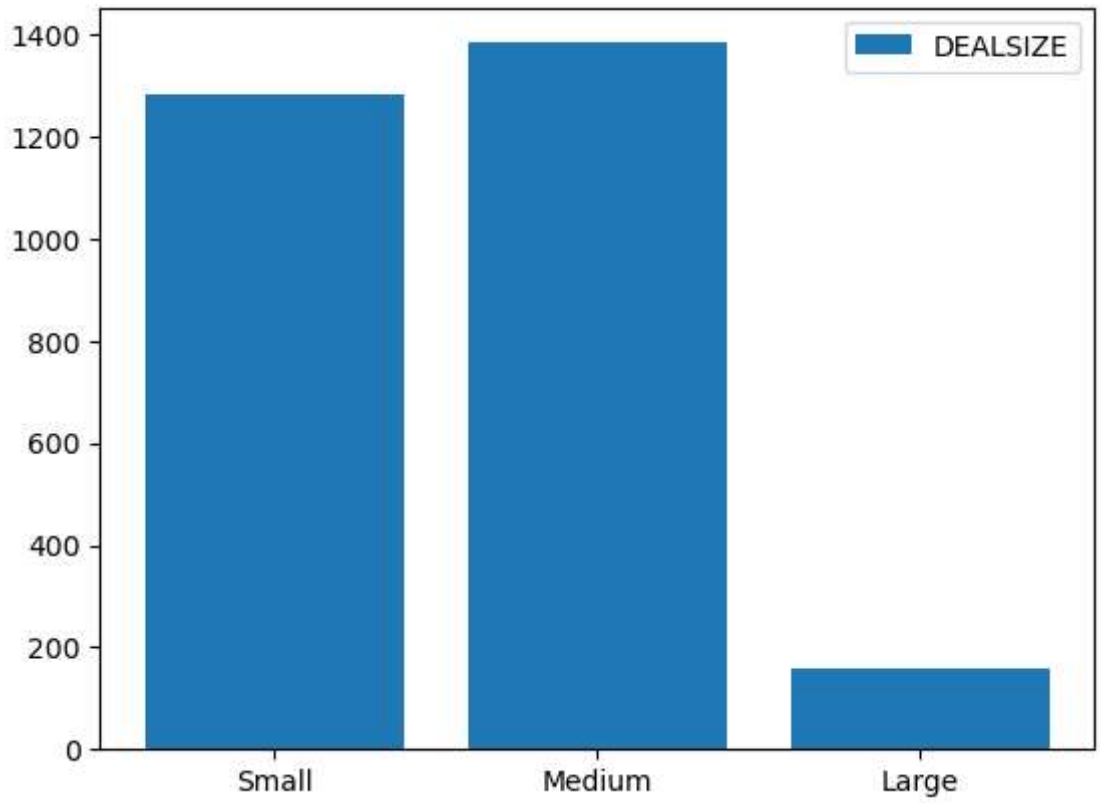
Out[23]:
```
<matplotlib.legend.Legend at 0x24830a10e50>
```



In [30]:
```
#Identify the most common DEALSIZE
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")
```

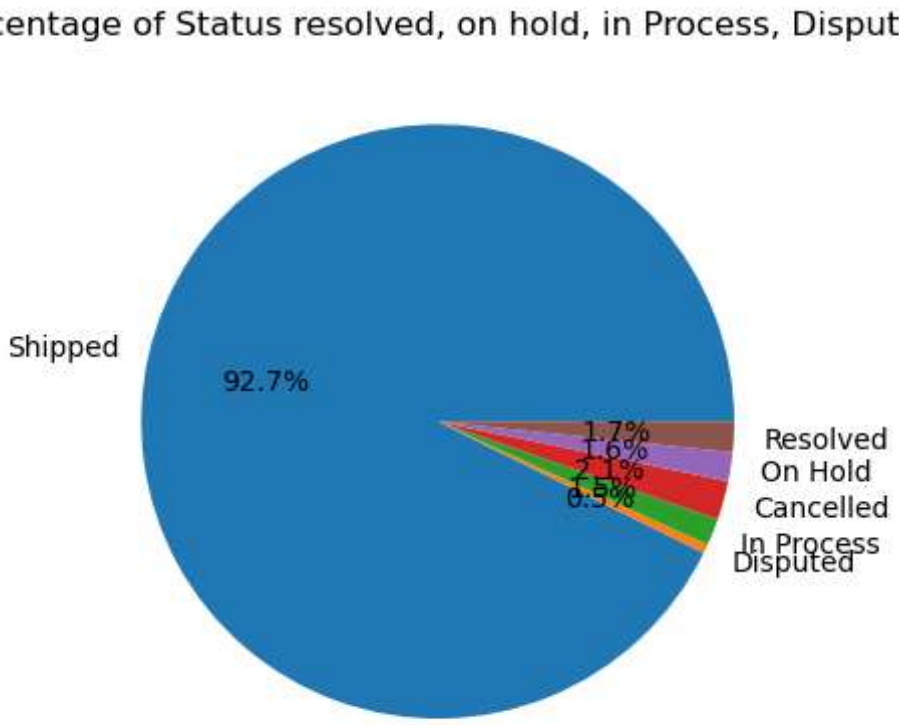
Out[30]:
```
<matplotlib.legend.Legend at 0x24832579060>
```



In [31]:
```
#3.find percentage of status resolved, on hold, in Process ,Disputed
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)
```

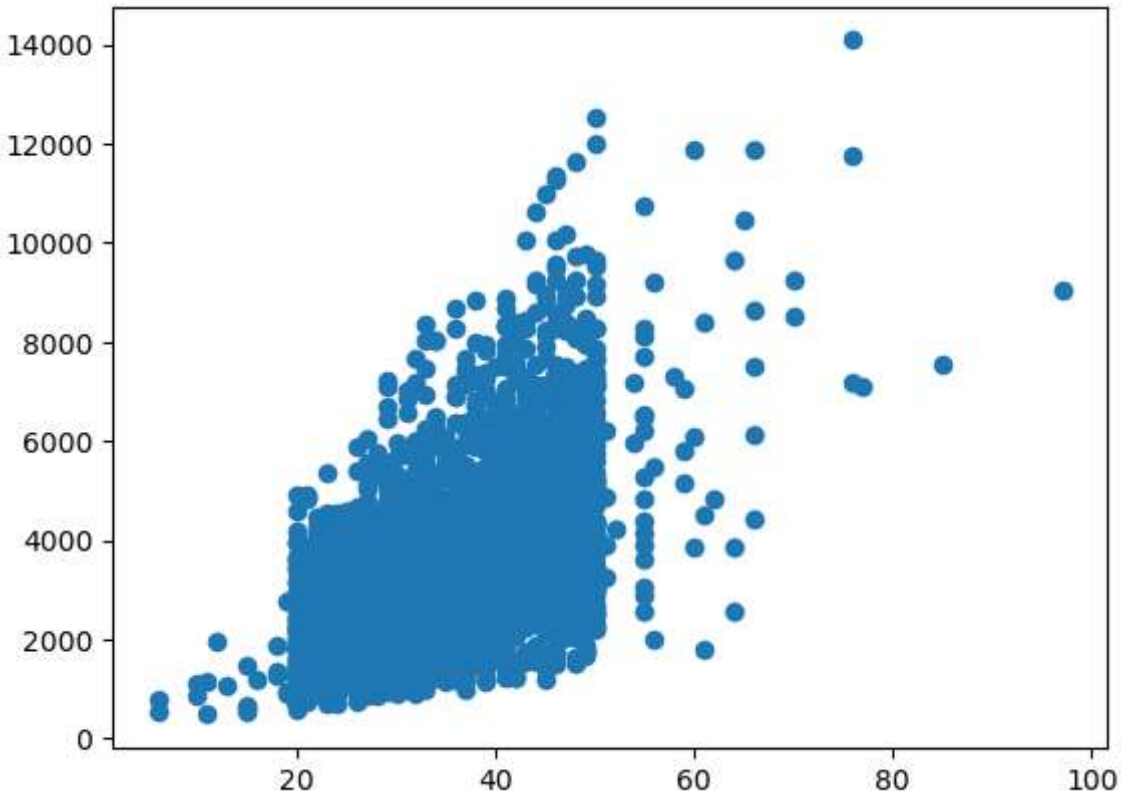
In [32]:
```
plt.pie(percent, labels=status,autopct='%1.1f%%')
plt.title('Percentage of Status resolved, on hold, in Process, Disputed')
```

Out[32]:
```
Text(0.5, 1.0, 'Percentage of Status resolved, on hold, in Process,\xa0Disput
ed')
```



In [34]:
```
#4. Identify Relationship between Qquantity Ordered and Sales
plt.scatter(df['QUANTITYORDERED'],df['SALES'])
```

Out[34]: <matplotlib.collections.PathCollection at 0x2482f91fc10>



In [ ]: