Name:Ritika Choudhari

Roll No:730-G2

PRN:202201090148

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

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

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

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

In [*]:
```
#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}
font1 = {'family':'serif','color':'darkred','size':15}
plt.bar(country, sales, label="Country wise sales")
plt.legend(loc="best")
```

In [*]:
```
#2.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")
```

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

In [*]:
```
#Identify Relationship Between Quality ordered and sales
plt.scatter(df['QUANTITYORDERED'],df['SALES'])
```

In [*]:
```
#identify yearwise sales
newdf=df.groupby('YEAR_ID')
year=df['YEAR_ID'].unique()
sales=[]
for yr in year:
    sales.append(sum(newdf.get_group(yr)['SALES']))


plt.plot(year.astype(str),sales)
plt.xlabel('YEAR')
plt.ylabel('SALES')
plt.title('Yearwise Sales')
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: