Name:-Rutuja Udanshiv

Roll no:- 285 PRN:-202201040120

Division:-B Batch:-B4

Assignment 1

Problem statement : **Take/Prepare any text files for any real-life application. For Ex. "Stud.txt", "Placement.csv" and "Result. csv" files for result Analysis. Combine into "StudentDetails.csv". Perform all statistical analysis (Average, Max, Min, Count, Sum, Percentage) on it**

Source code:

In [1]:
```
#merge 2 data sets
f1 = open("emp.csv","r")
f2 = open("sal.csv","r")
f3 = open("emp_sal.csv","w")

contents1 = f1.read()
contents2 = f2.read()

nm = []
sal = []

lines1 = contents1.split("\n")
lines2 = contents2.split("\n")
for l1 in lines1:
    words1 = l1.split(",")

    for l2 in lines2:
        words2 = l2.split(",")
        if(words1[0]==words2[0]):
            l1 = l1 + "," + words2[1]+","+words2[2] +"\n"
            f3.write(l1)

            nm.append(words1[1])
            sal.append(int(words2[2]))

f1.close()
f2.close()
f3.close()

print(nm)
print(sal)
print("Largest salary is :",max(sal))
print("Least salary is :",min(sal))
print("Average salary is :",sum(sal)/len(sal))
```

```
#for finding index sal.index(max(sal))

hs = nm[sal.index(max(sal))]
print(hs ,"has the highest salary")

desgn = ["Manager","Sr. Manager","Ast. Manager","Supervisor","Employee"]
print("The employee with designation is Sr. Manager is :")
for i in range(len(desgn)):
    if desgn[i]=="Sr. Manager" or desgn[i]=="sr. Manager":
        print(nm[i],end=" ")
```

Output:

```
['Sanvi', 'Mrunmayee', 'Jayesh', 'Gouri', 'Mahesh']
[100000, 150000, 90500, 100500, 85000]
Largest salary is : 150000
Least salary is : 85000
Average salary is : 105200.0
Mrunmayee has the highest salary
The employee with designation is Sr. Manager is :
Mrunmayee
```

```
Product_details=[]
Supplier_details={}
Customer_details=[]
gender={}

f1=open('/content/Sales.csv','r')

while(True):
   data=f1.readline()
   if not data:
      break;
   data=data.replace("\n","")
   temp=data.split(",")
   print(temp)
   Product_details.append(temp[1])
   Customer_details.append(temp[3])
   Supplier_details.update({temp[0]:temp[2]})
   gender.update({temp[3]:temp[4]})

f1.close()

Customer_details=tuple(Customer_details)
print(type(Customer_details))
```

```
['Product ID', 'Product details', 'Supplier Details', 'Customer
Details', 'Gender']
['P00001', 'Lenovo Laptop', 'Raka Ele.', 'Kaustubh Mahajan', 'Male']
['P00002', 'Samsung M31', 'Vijay Sales', 'Siddhi Kiwale', 'Female']
['P00003', 'Realmi 10pro', 'Gada Ele.', 'Sanket Kandalkar', 'Male']
['P00004', 'Oppo F21', 'Surya Ele.', 'Yash Mali', 'Male']
['P00005', 'Lenovo Laptop', 'Raka Ele.', 'Yash Bagul', 'Male']
['P00006', 'Samsung M31', 'Gada Ele.', 'Siddhi Kiwale', 'Female']
['P00007', '"LG TV 32"""', 'Vijay Sales', 'Sanket Kandalkar', 'Male']
['P00008', 'Oppo F21', 'Surya Ele.', 'Kaustubh Mahajan', 'Male']
['P00009', 'Lenovo Laptop', 'Raka Ele.', 'Yash Mali', 'Male']
['P00010', 'Samsung M31', 'Gada Ele.', 'Siddhi Kiwale', 'Female']
['P00011', '"LG TV 32"""', 'Surya Ele.', 'Sanket Kandalkar', 'Male']
['P00012', 'Lenovo Laptop', 'Raka Ele.', 'Kaustubh Mahajan', 'Male']
['P00013', 'Samsung M31', 'Surya Ele.', 'Yash Mali', 'Male']
['P00014', 'Realmi 10pro', 'Raka Ele.', 'Siddhi Kiwale', 'Female']
['P00015', 'Lenovo Laptop', 'Gada Ele.', 'Tanuja Mali', 'Female']
['P00016', 'Oppo F21', 'Vijay Sales', 'Kaustubh Mahajan', 'Male']
['P00017', '"LG TV 32"""', 'Deshmukh sales', 'Sanket Kandalkar',
'Male']
['P00018', 'Lenovo Laptop', 'Raka Ele.', 'Siddhi Kiwale', 'Female']
['P00019', 'Samsung M31', 'Deshmukh sales', 'Kaustubh Mahajan',
'Male']
['P00020', '"LG TV 32"""', 'Gada Ele.', 'Yash Mali', 'Male']
<class 'tuple'>
```

```
print("\nProduct_details\n",Product_details,end="")
print("\n\ncustomer_details\n",Customer_details,end="")
print("\n\nSupplier_details\n",Supplier_details,end="")
print("\n\ngender_details\n",gender,end="")
```

```
Product_details
 ['Product details', 'Lenovo Laptop', 'Samsung M31', 'Realmi 10pro',
'Oppo F21', 'Lenovo Laptop', 'Samsung M31', '"LG TV 32"""', 'Oppo
F21', 'Lenovo Laptop', 'Samsung M31', '"LG TV 32"""', 'Lenovo Laptop',
'Samsung M31', 'Realmi 10pro', 'Lenovo Laptop', 'Oppo F21', '"LG TV
32"""', 'Lenovo Laptop', 'Samsung M31', '"LG TV 32"""']

customer_details
 ('Customer Details', 'Kaustubh Mahajan', 'Siddhi Kiwale', 'Sanket
Kandalkar', 'Yash Mali', 'Yash Bagul', 'Siddhi Kiwale', 'Sanket
Kandalkar', 'Kaustubh Mahajan', 'Yash Mali', 'Siddhi Kiwale', 'Sanket
Kandalkar', 'Kaustubh Mahajan', 'Yash Mali', 'Siddhi Kiwale', 'Tanuja
Mali', 'Kaustubh Mahajan', 'Sanket Kandalkar', 'Siddhi Kiwale',
'Kaustubh Mahajan', 'Yash Mali')

Supplier_details
 {'Product ID': 'Supplier Details', 'P00001': 'Raka Ele.', 'P00002':
'Vijay Sales', 'P00003': 'Gada Ele.', 'P00004': 'Surya Ele.',
'P00005': 'Raka Ele.', 'P00006': 'Gada Ele.', 'P00007': 'Vijay Sales',
'P00008': 'Surya Ele.', 'P00009': 'Raka Ele.', 'P00010': 'Gada Ele.',
'P00011': 'Surya Ele.', 'P00012': 'Raka Ele.', 'P00013': 'Surya Ele.',
'P00014': 'Raka Ele.', 'P00015': 'Gada Ele.', 'P00016': 'Vijay Sales',
'P00017': 'Deshmukh sales', 'P00018': 'Raka Ele.', 'P00019': 'Deshmukh
sales', 'P00020': 'Gada Ele.'}

gender_details
 {'Customer Details': 'Gender', 'Kaustubh Mahajan': 'Male', 'Siddhi
Kiwale': 'Female', 'Sanket Kandalkar': 'Male', 'Yash Mali': 'Male',
'Yash Bagul': 'Male', 'Tanuja Mali': 'Female'}
```

```
 # The most popular product

 def most_frequent(Product_details):
     counter = 0
     num = Product_details[0]

     for i in Product_details:
         curr_frequency = Product_details.count(i)
         if (curr_frequency> counter):
            counter = curr_frequency
            num = i
     return num
print(most_frequent(Product_details))
```

Lenovo Laptop

```
#The most popular supplier
frequency = {}
 #iterating over the last
for item in Supplier_details.values():
   # Checking the element in dictionary
   if item in frequency:
     #incrementing the counter
     frequency[item] += 1
   else:
    # initializing the count
    frequency[item] = 1
#printing the frequency
print(frequency)

marklist = sorted(frequency.items(),key=lambda x:x[1],reverse=True)
sortdict = dict(marklist)
print(sortdict)
print("the most popular Supplier for sales",list(sortdict.keys())
[0],"sold",list(sortdict.values())[0],"Items")
```

```
{'Supplier Details': 1, 'Raka Ele.': 6, 'Vijay Sales': 3, 'Gada Ele.':
5, 'Surya Ele.': 4, 'Deshmukh sales': 2}
{'Raka Ele.': 6, 'Gada Ele.': 5, 'Surya Ele.': 4, 'Vijay Sales': 3,
'Deshmukh sales': 2, 'Supplier Details': 1}
the most popular Supplier for sales Raka Ele. sold 6 Items
```

```
# The Customer who buys most of the products

frequency = {}
#iterating over the list
for item in Customer_details:
  #checking the elements in dictionary
  if item in frequency:
    #incrementing the counter
    frequency[item] += 1
  else:
     #initalizing the count
     frequency[item] = 1
#printing the frequency
print("Frequency is as given below: \n ",frequency)
marklist = sorted(frequency.items(), key=lambda x:x[1],reverse=True)
sortlist = dict(marklist)
print("\nSorted Dict is as below;\n", sortdict)
print("\n\nThe customer who buys most of the products",
list(sortdict.keys())[0]," buy",list(sortdict.values())[0], "Items")
```

```
Frequency is as given below:
  {'Customer Details': 1, 'Kaustubh Mahajan': 5, 'Siddhi Kiwale': 5,
'Sanket Kandalkar': 4, 'Yash Mali': 4, 'Yash Bagul': 1, 'Tanuja Mali':
1}

Sorted Dict is as below;
 {'Raka Ele.': 6, 'Gada Ele.': 5, 'Surya Ele.': 4, 'Vijay Sales': 3,
'Deshmukh sales': 2, 'Supplier Details': 1}


The customer who buys most of the products Raka Ele.  buy 6 Items
```

```python
# No. customer who are females
from collections import Counter
counter = dict(Counter(Customer_details))
names=list(counter.keys())
print(names)
male=0
female=0

for name in names:
  if gender[name] == "Male":
    male += 1
  if gender[name] == "Female":
    female += 1

print("Total no of male =",male)
print("Total no of Female =",female)
```

```
['Customer Details', 'Kaustubh Mahajan', 'Siddhi Kiwale', 'Sanket
Kandalkar', 'Yash Mali', 'Yash Bagul', 'Tanuja Mali']
Total no of male = 4
Total no of Female = 2
```

```
import numpy as np
array3=
np.loadtxt("/content/testmarks1.csv",delimiter=',',dtype=str,ski
prows=1)
print(array3)
Rollno=[]
Eds=[]
son=[]
Dt=[]
Et=[]
for i in array3:
 Eds.append(float(i[1]))
 son.append(float(i[2]))
 Dt.append(float(i[3]))
 Et.append(float(i[4]))
print(Eds)
print(son)
print(Dt)
print(Et)
m=max(Edsarr)
mi=min(sonarr)
Edsarr=np.array(Eds)
sonarr=np.array(son)
Dtarr=np.array(Dt)
Etarr=np.array(Et)
std=np.std(Dtarr)
med=np.median(Etarr)
var=np.var(Edsarr)
mean=np.mean(sonarr)
sort=np.sort(Dtarr)
search = np.where(sonarr == 26.16)
print(dt)
print("The min of son",mi)
print("The max of eds",m)
print("The std of Dtarr",std)
print("The med of Etarr",med)
print("The var of Edsarr",var)
print("The mean ofsonarr ",mean)
print("The sortedc arr of Dtarr ",sort)
print("The search arr of sonarr ",search)
```

OUTPUT

```
['801' '43.05' '27.79' '28.7' '27.79']
 ['802' '43.47' '28.52' '28.98' '27.89']
 ['803' '42.24' '28.16' '28.16' '25.63']
 ['804' '39.24' '26.16' '26.16' '26.16']
 ['805' '40.9' '26.03' '27.27' '25.65']
 ['806' '39.47' '26.31' '26.31' '25.21']
 ['807' '41.68' '25.63' '27.79' '25.46']
 ['808' '42.19' '27.61' '28.13' '26.21']
 ['809' '44.75' '28.35' '29.83' '28.21']
 ['810' '46.95' '28.88' '31.3' '28.53']]
[43.05, 43.47, 42.24, 39.24, 40.9, 39.47, 41.68, 42.19, 44.75, 46.95]
[27.79, 28.52, 28.16, 26.16, 26.03, 26.31, 25.63, 27.61, 28.35, 28.88]
[28.7, 28.98, 28.16, 26.16, 27.27, 26.31, 27.79, 28.13, 29.83, 31.3]
[27.79, 27.89, 25.63, 26.16, 25.65, 25.21, 25.46, 26.21, 28.21, 28.53]
The min of son 25.63
The max of eds 46.95
The std of Dtarr 1.4784725225718605
The med of Etarr 26.185000000000002
The var of Edsarr 4.920064000000002
The mean ofsonarr  27.344
The sortedc arr of Dtarr  [26.16 26.31 27.27 27.79 28.13 28.16 28.7  28.98
29.83 31.3 ]
The search arr of sonarr  (array([3]),)
```

```
import pandas as pd
import numpy as np
f1 = open("F:\grainsales.csv","r")
data = pd.read_csv(f1)
df = pd.DataFrame(data)
maindata = df
df['Sales'].describe()
df=df.groupby('Months').sum()
df=df.sort_values (by= ['Sales'], ascending=False) df.head(1)
print("Best Month for Sales: July")
print("Revenue Earned was: 16000000")
df
maindata
df = df.groupby("GrainName").sum()
df = df.sort_values(by=["Sales"], ascending = False)
df.head (1)
print("Most Sold Grain is: Wheat")
print("The Best Month for sales is July and this product has occured in July
so this is most sold product with highest sales")
df
maindata
df= df.groupby("City").sum()
df = df.sort_values (by = ['Sales'], ascending= False)
df.head (1)
print("'Asansole' Has sold highest no. of products")
df
maindata
df = df.groupby('State').sum()
df = df.sort_values (by = ['Sales'], ascending = False) print("West
Bengol has highest sales")
```

```
Best Month for Sales: July
Revenue Earned was: 16000000
```

Most Sold Grain is: Wheat

The Best Month for sales is July and this product has occured in July so this is most sold product with highest sales

'Asansole' Has sold highest no. of products

West Bengol has highest sales.

Name -Rutuja Udanshiv

Roll no-285

Batch-B4

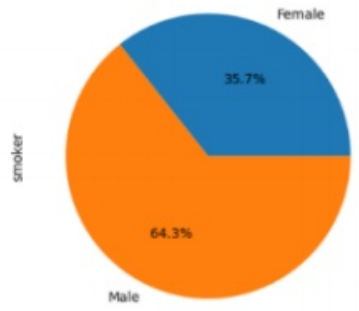
Prn no-202201040120

ASSIGNMENT -5

```
[ ] import pandas as pd
    import matplotlib.pyplot as plt
    d = pd.read_csv('/content/tips.csv')
    print(d)
```

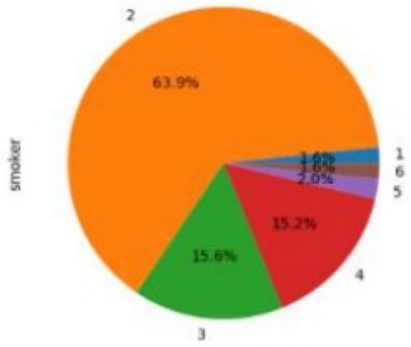
```
     total_bill   tip     sex smoker   day    time  size
0         16.99  1.01  Female     No   Sun  Dinner     2
1         10.34  1.66    Male     No   Sun  Dinner     3
2         21.01  3.50    Male     No   Sun  Dinner     3
3         23.68  3.31    Male     No   Sun  Dinner     2
4         24.59  3.61  Female     No   Sun  Dinner     4
..          ...   ...     ...    ...   ...     ...   ...
239       29.03  5.92    Male     No   Sat  Dinner     3
240       27.18  2.00  Female    Yes   Sat  Dinner     2
241       22.67  2.00    Male    Yes   Sat  Dinner     2
242       17.82  1.75    Male     No   Sat  Dinner     2
243       18.78  3.00  Female     No  Thur  Dinner     2

[244 rows x 7 columns]
```

```
import pandas as pd
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
t1 = d.groupby("sex").count()
t1["smoker"].plot(kind = "pie",autopct = '%1.1f%%')
plt.show()
print(t1)
```



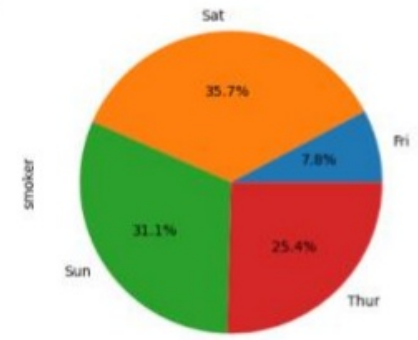

```
        total_bill  tip  smoker  day  time  size
sex
Female          87   87      87   87    87    87
Male           157  157     157  157   157   157
```

```
import pandas as pd
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
t1 = d.groupby("size").count()
t1["smoker"].plot(kind = "pie",autopct = '%1.1f%%')
plt.show()
print(t1)
```



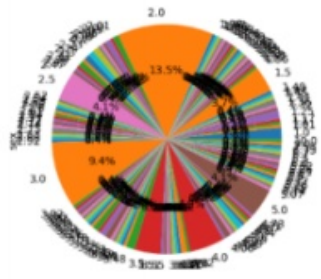
size	total_bill	tip	sex	smoker	day	time
1	4	4	4	4	4	4
2	156	156	156	156	156	156
3	38	38	38	38	38	38
4	37	37	37	37	37	37
5	5	5	5	5	5	5
6	4	4	4	4	4	4

```
import pandas as pd
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
t1 = d.groupby("day").count()
t1["smoker"].plot(kind = "pie",autopct = '%1.1f%%')
plt.show()
print(t1)
```



day	total_bill	tip	sex	smoker	time	size
Fri	19	19	19	19	19	19
Sat	87	87	87	87	87	87
Sun	76	76	76	76	76	76
Thur	62	62	62	62	62	62

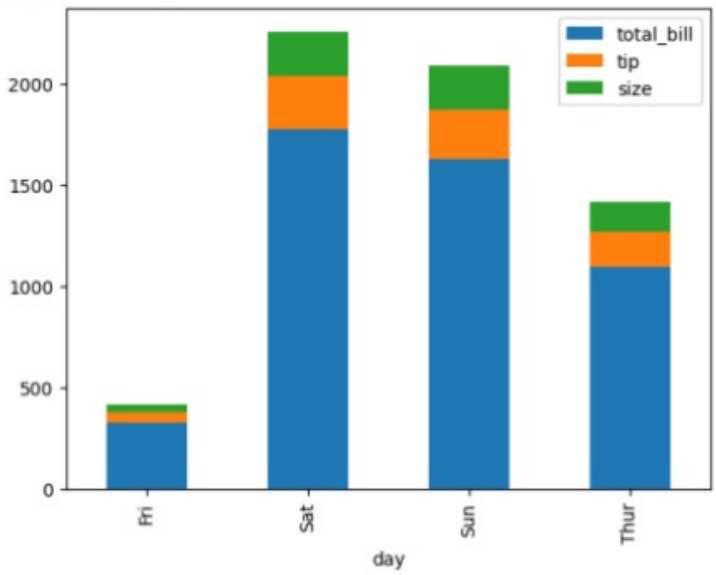
```
import pandas as pd
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
t1 = d.groupby("tip").count()
t1["sex"].plot(kind = "pie",autopct = '%1.1f%%')
plt.show()
print(t1)
```




```
     total_bill  sex  smoker  day  time  size
tip
1.00          4    4       4    4     4     4
1.01          1    1       1    1     1     1
1.10          1    1       1    1     1     1
1.17          1    1       1    1     1     1
1.25          3    3       3    3     3     3
...         ...  ...     ...  ...   ...   ...
6.70          1    1       1    1     1     1
6.73          1    1       1    1     1     1
7.58          1    1       1    1     1     1
9.00          1    1       1    1     1     1
10.00         1    1       1    1     1     1

[123 rows x 6 columns]
```

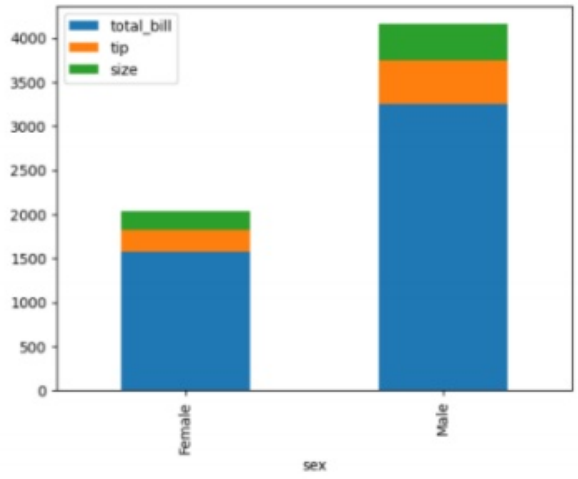
```
t2 = d.groupby("day").sum("total_bill")
t2.plot(kind = "bar",stacked = True)
```

```
<Axes: xlabel='day'>
```



```
t2 = d.groupby("sex").sum("smoker")
t2.plot(kind = "bar",stacked = True)
```

```
<Axes: xlabel='sex'>
```

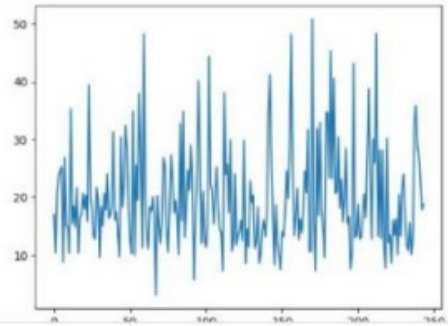



```
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
d1 = d.groupby("total_bill").count()
print(d1)
```

```
             tip  sex  smoker  day  time  size
total_bill
3.07           1    1       1    1     1     1
5.75           1    1       1    1     1     1
7.25           2    2       2    2     2     2
7.51           1    1       1    1     1     1
7.56           1    1       1    1     1     1
...          ...  ...     ...  ...   ...   ...
45.35          1    1       1    1     1     1
48.17          1    1       1    1     1     1
48.27          1    1       1    1     1     1
48.33          1    1       1    1     1     1
50.81          1    1       1    1     1     1

[229 rows x 6 columns]
```

```
d["total_bill"].plot()
```

```
<Axes: >
```

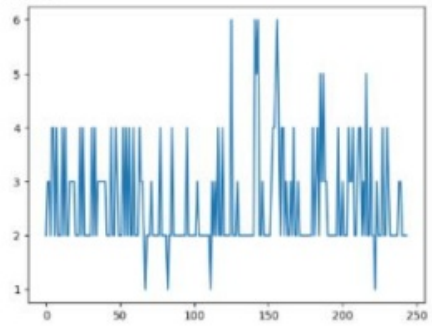


```
import pandas as pd
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
d1 = d.groupby("size").count()
print(d1)
```

```
      total_bill  tip  sex  smoker  day  time
size
1              4    4    4       4    4     4
2            156  156  156     156  156   156
3             38   38   38      38   38    38
4             37   37   37      37   37    37
5              5    5    5       5    5     5
6              4    4    4       4    4     4
```

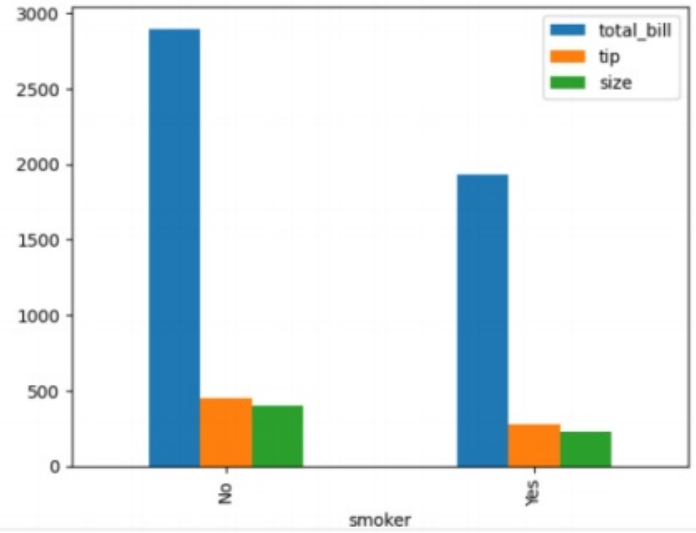
```
d["size"].plot()
```

```
<Axes: >
```



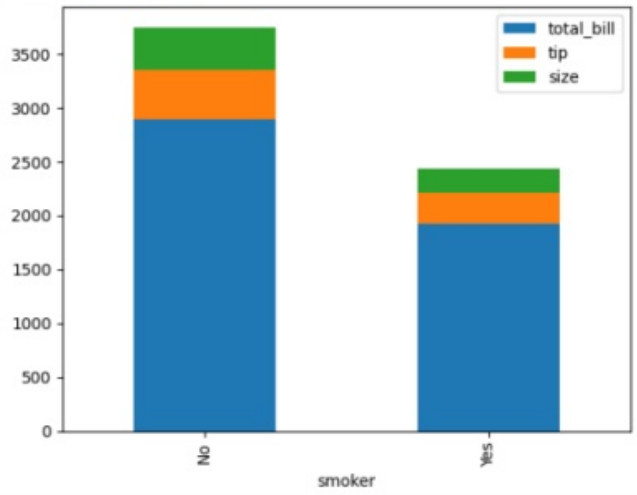

```
t2 = d.groupby("smoker").sum("tip")
t2.plot(kind = "bar",stacked = False)
```

```
<Axes: xlabel='smoker'>
```

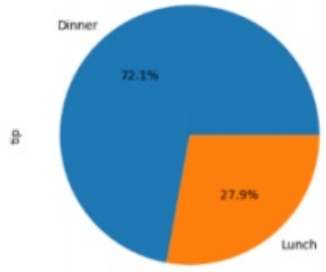


```
t2 = d.groupby("smoker").sum("time")
t2.plot(kind = "bar",stacked = True)
```

```
<Axes: xlabel='smoker'>
```




```
import pandas as pd
import matplotlib.pyplot as plt
d = pd.read_csv('/content/tips.csv')
#print(d)
t1 = d.groupby("time").count()
t1["tip"].plot(kind = "pie",autopct = '%1.1f%%')
plt.show()
print(t1)
```




```
        total_bill  tip  sex  smoker  day  size
time
Dinner         176  176  176     176  176   176
Lunch           68   68   68      68   68    68
```