

```
import pandas as pd
```

Read the dataset

```
dv=pd.read_excel("DoctorVisits (2).xlsx")
dv.head()
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	931.0	0.55	1	4	1	yes	no	no	no
1	2	1	female	0.19	0.45	1	2	1	yes	no	no	no
2	3	1	male	0.19	0.90	3	0	0	no	no	no	no
3	4	1	male	0.19	0.15	1	0	0	no	no	no	no
4	5	1	male	0.19	0.45	2	5	1	no	no	no	yes

Display all the columns of the dataset where datatypes,column name,count and overall memory

```
dv=pd.read_excel("DoctorVisits (2).xlsx")
dv.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5190 entries, 0 to 5189
Data columns (total 13 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   Unnamed: 0   5190 non-null     int64   
1   visits       5190 non-null     int64   
2   gender       5190 non-null     object  
3   age          5190 non-null     float64 
4   income       5190 non-null     float64 
5   illness      5190 non-null     int64   
6   reduced      5190 non-null     int64   
7   health       5190 non-null     int64   
8   private      5190 non-null     object  
9   freepoor     5190 non-null     object  
10  freerepat    5190 non-null     object  
11  nchronic     5190 non-null     object  
12  lchronic     5190 non-null     object  
dtypes: float64(2), int64(5), object(6)
memory usage: 527.2+ KB
```

Find the total no of people based on their count age,income,gender

```
dv["age"]>dv["age"].value_counts()
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	931.0	0.55	1	4	1	yes	no	no	no
1	2	1	female	931.0	0.45	1	2	1	yes	no	no	no
2	3	1	male	931.0	0.90	3	0	0	no	no	no	no
3	4	1	male	931.0	0.15	1	0	0	no	no	no	no
4	5	1	male	931.0	0.45	2	5	1	no	no	no	yes

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
5185	5186	0	female	1078.0	0.55	0	0	0	no	no	no	no
5186	5187	0	male	1323.0	1.30	0	0	1	no	no	no	no
5187	5188	0	female	1813.0	0.25	1	0	1	no	no	yes	no
5188	5189	0	female	2548.0	0.65	0	0	0	no	no	no	no
5189	5190	0	male	3528.0	0.25	0	0	0	no	no	yes	no

5190 rows × 13 columns

```
dv["income"]>dv["income"].value_counts()
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	931.0	0.55	1	4	1	yes	no	no	no
1	2	1	female	931.0	0.45	1	2	1	yes	no	no	no
2	3	1	male	931.0	0.90	3	0	0	no	no	no	no
3	4	1	male	931.0	0.15	1	0	0	no	no	no	no
4	5	1	male	931.0	0.45	2	5	1	no	no	no	yes

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
5185	5186	0	female	1078.0	0.55	0	0	0	no	no	no	no
5186	5187	0	male	1323.0	1.30	0	0	1	no	no	no	no
5187	5188	0	female	1813.0	0.25	1	0	1	no	no	yes	no
5188	5189	0	female	2548.0	0.65	0	0	0	no	no	no	no
5189	5190	0	male	3528.0	0.25	0	0	0	no	no	yes	no

5190 rows × 13 columns

```
dv["gender"].value_counts()
```

```
Female: 2792
Male: 2488
Name: gender, dtype: int64
```

Find the value count of different data types

```
dv["age"].value_counts()
```

```
0      4241
1       182
2       174
3        39
4        24
7        12
6        12
5         9
8         5
9         1
Name: age, dtype: int64
```

```
dv["illness"].value_counts()
```

```
1      1638
0      1554
2       948
3       542
4       274
5       236
Name: illness, dtype: int64
```

Describing the info of the datatypes

```
import pandas as pd
```

```
dv.describe()
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
count	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000
mean	2595.500000	0.301734	1991.288247	8.747399e+07	1.431985	0.861850	1.217534					
std	1498.368279	0.798134	1003.430942	5.533600e+07	1.384152	0.876728	2.124266					
min	1.000000	0.000000	931.000000	0.000000e+00	0.000000	0.000000	0.000000					
25%	1298.250000	0.000000	1078.000000	3.750000e+07	0.000000	0.000000	0.000000					
50%	2595.500000	0.000000	1568.000000	8.250000e+07	1.000000	0.000000	0.000000					
75%	3892.750000	0.000000	3038.000000	1.350000e+08	2.000000	0.000000	2.000000					
max	5190.000000	9.000000	3528.000000	2.250000e+08	5.000000	14.000000	12.000000					

```
dv=pd.read_excel("DoctorVisits (2).xlsx")
dv.dropna(axis=0)
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	0.19	0.55	1	4	1	yes	no	no	no
1	2	1	female	0.19	0.45	1	2	1	yes	no	no	no
2	3	1	male	0.19	0.90	3	0	0	no	no	no	no
3	4	1	male	0.19	0.15	1	0	0	no	no	no	no
4	5	1	male	0.19	0.45	2	5	1	no	no	no	yes

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
5185	5186	0	female	0.22	0.55	0	0	0	no	no	no	no
5186	5187	0	male	0.27	1.30	0	0	1	no	no	no	no
5187	5188	0	female	0.37	0.25	1	0	1	no	no	yes	no
5188	5189	0	female	0.52	0.65	0	0	0	no	no	no	no
5189	5190	0	male	0.72	0.25	0	0	0	no	no	yes	no

5190 rows × 13 columns

```
dv=pd.read_excel("DoctorVisits (2).xlsx")
dv.fillna("28")
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	0.19	0.55	1	4	1	yes	no	no	no
1	2	1	female	0.19	0.45	1	2	1	yes	no	no	no
2	3	1	male	0.19	0.90	3	0	0	no	no	no	no
3	4	1	male	0.19	0.15	1	0	0	no	no	no	no
4	5	1	male	0.19	0.45	2	5	1	no	no	no	yes

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
5185	5186	0	female	0.22	0.55	0	0	0	no	no	no	no
5186	5187	0	male	0.27	1.30	0	0	1	no	no	no	no
5187	5188	0	female	0.37	0.25	1	0	1	no	no	yes	no
5188	5189	0	female	0.52	0.65	0	0	0	no	no	no	no
5189	5190	0	male	0.72	0.25	0	0	0	no	no	yes	no

5190 rows × 13 columns

```
dv.isna().sum()
```

```
Unnamed: 0    0
visits        0
gender        0
age           0
income        0
illness       0
reduced       0
health        0
private       0
freepoor      0
freerepat     0
nchronic      0
lchronic      0
dtype: int64
```

Analyzing the variables

```
import pandas as pd
dv=pd.read_excel("DoctorVisits (2).xlsx")
dv.info()
```

```
array([1, 2, 3, 4, 8, 5, 7, 6, 9, 0], dtype=int64)
```

```
dv.gender.unique()
```

```
array(['female', 'male'], dtype=object)
```

```
dv.freerepat.unique()
```

```
array(['no', 'yes'], dtype=object)
```

```
dv.private.unique()
```

```
array(['yes', 'no'], dtype=object)
```

```
dv.nchronic.unique()
```

```
array(['no', 'yes'], dtype=object)
```

```
dv.age.unique()
```

```
array([0.19, 0.22, 0.27, 0.32, 0.37, 0.42, 0.47, 0.52, 0.57, 0.62, 0.67, 0.72])
```

```
dv.income.unique()
```

```
array([0.55, 0.45, 0.9 , 0.15, 0.35, 0.65, 0.25, 0. , 0.06, 1.1 , 0.75, 0.01, 1.3 , 1.5 ])
```

```
dv.nunique()
```

```
Unnamed: 0    5190
visits        10
gender         2
age           12
income        14
illness        6
reduced       15
health        13
private        2
freepoor       2
freerepat     2
nchronic       2
lchronic       2
dtype: int64
```

Exploring and plotting the data

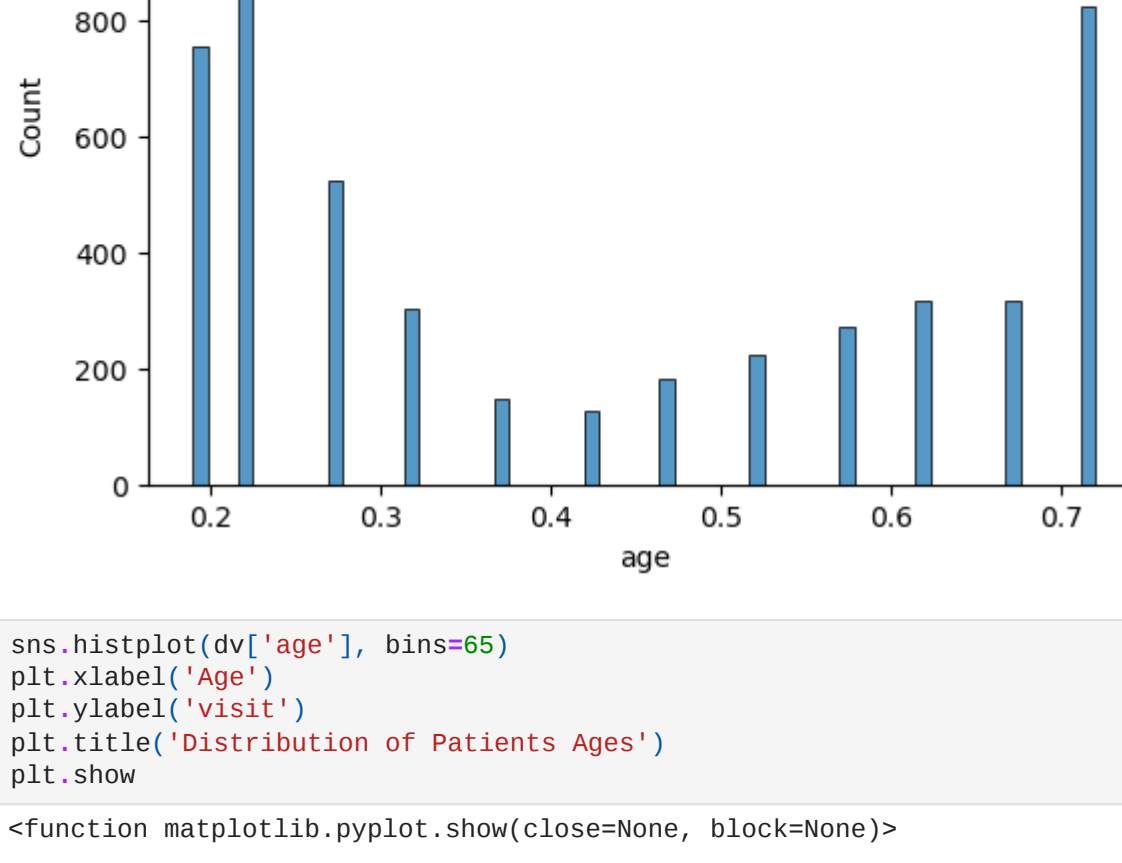
```
import matplotlib.pyplot as plt
sns.set()
```

```
dv=pd.read_excel("DoctorVisits (2).xlsx")
dv.info()
```

Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	0.19	0.55	1	4	1	yes	no	no	no
1	2	1	female	0.19	0.45	1	2	1	yes	no	no	no
2	3	1	male	0.19	0.90	3	0	0	no	no	no	no
3	4	1	male	0.19	0.15	1	0	0	no	no	no	no
4	5	1	male	0.19	0.45	2	5	1	no	no	no	yes

```
sns.histplot(dv["age"], bins=60)
```

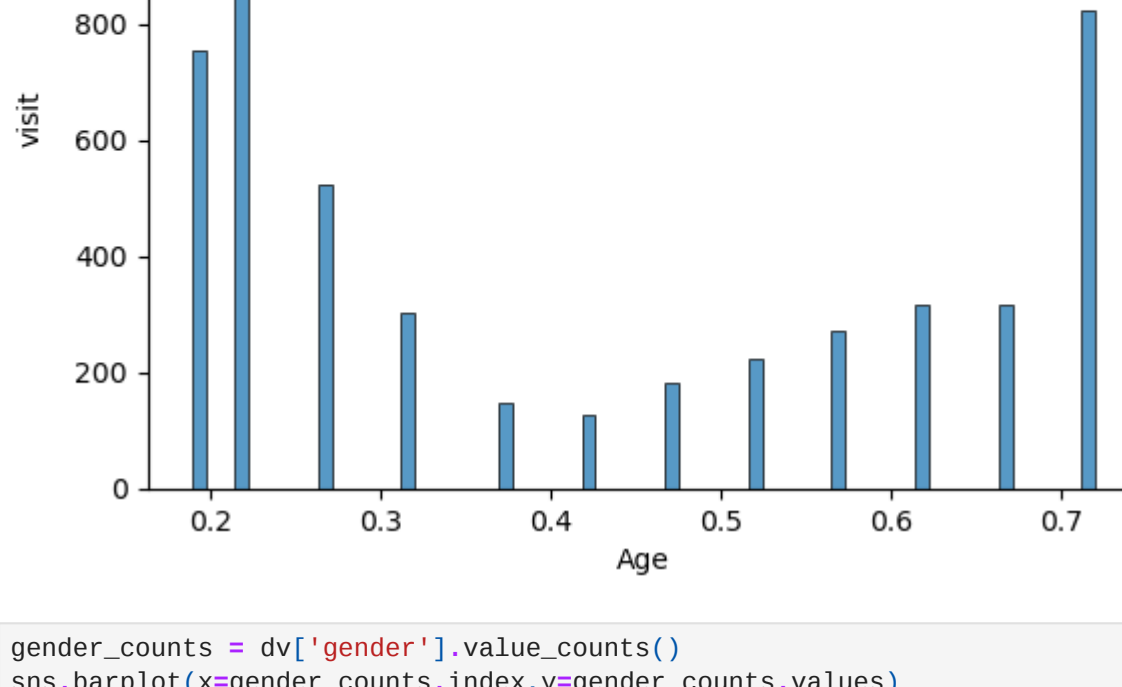
```
<Axes: xlabel='age', ylabel='Count'>
```



```
sns.histplot(dv["age"], bins=65)
```

```
plt.xlabel('age')
plt.ylabel('visit')
plt.title('Distribution of Patients Ages')
plt.show()
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



```
gender_counts = dv["gender"].value_counts()
sns.barplot(x=gender_counts.index, y=gender_counts.values)
```

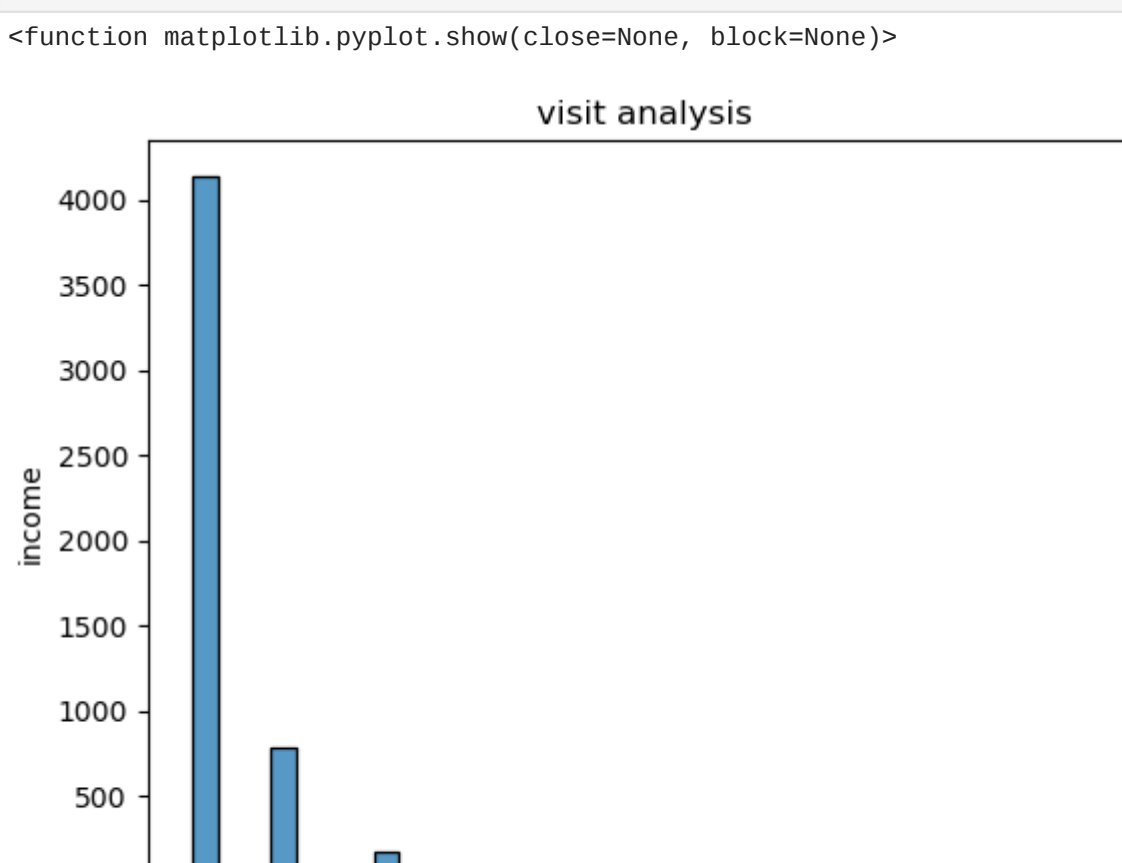
```
plt.xlabel('gender')
plt.ylabel('visit')
plt.title('Distribution of Patient Gender')
plt.show()
```



```
sns.histplot(dv["visits"], bins=35)
```

```
plt.xlabel('visits')
plt.ylabel('income')
plt.title('Income Analysis')
plt.show()
```

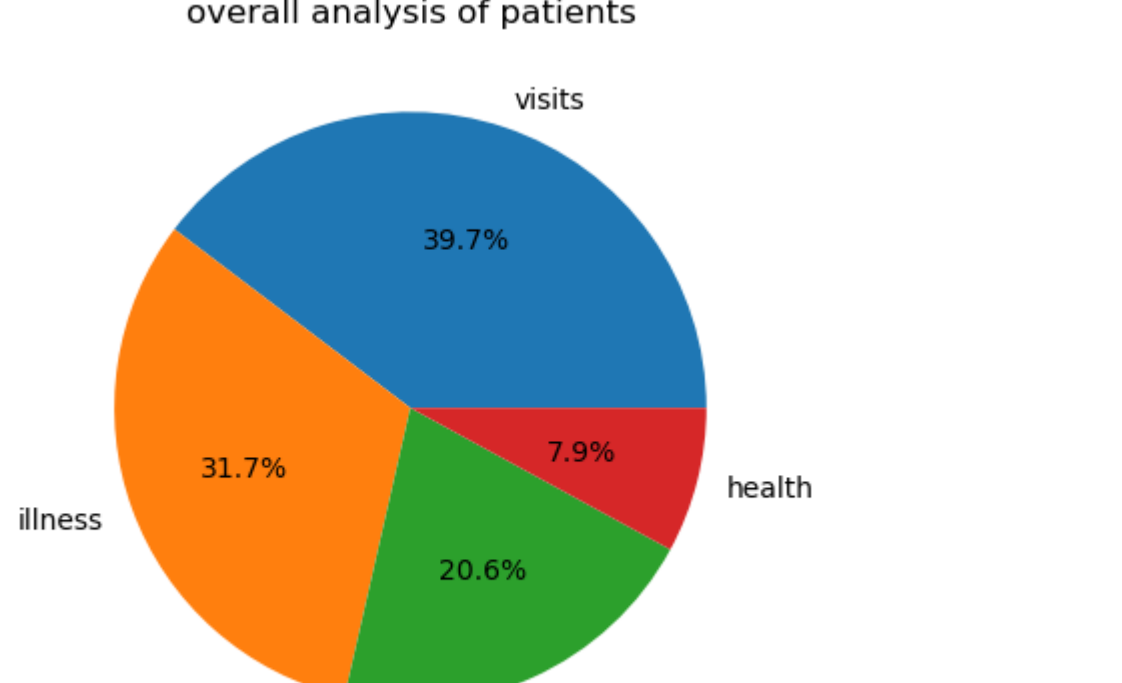
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Observations

```
labels=["visits","illness","reduced","health"]
sizes=[25,20,13,5]
```

```
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
plt.title('overall analysis of patients')
plt.show()
```



```
x = [16,14,12,18,20]
```

```
y = [20,12,14,9,7]
```

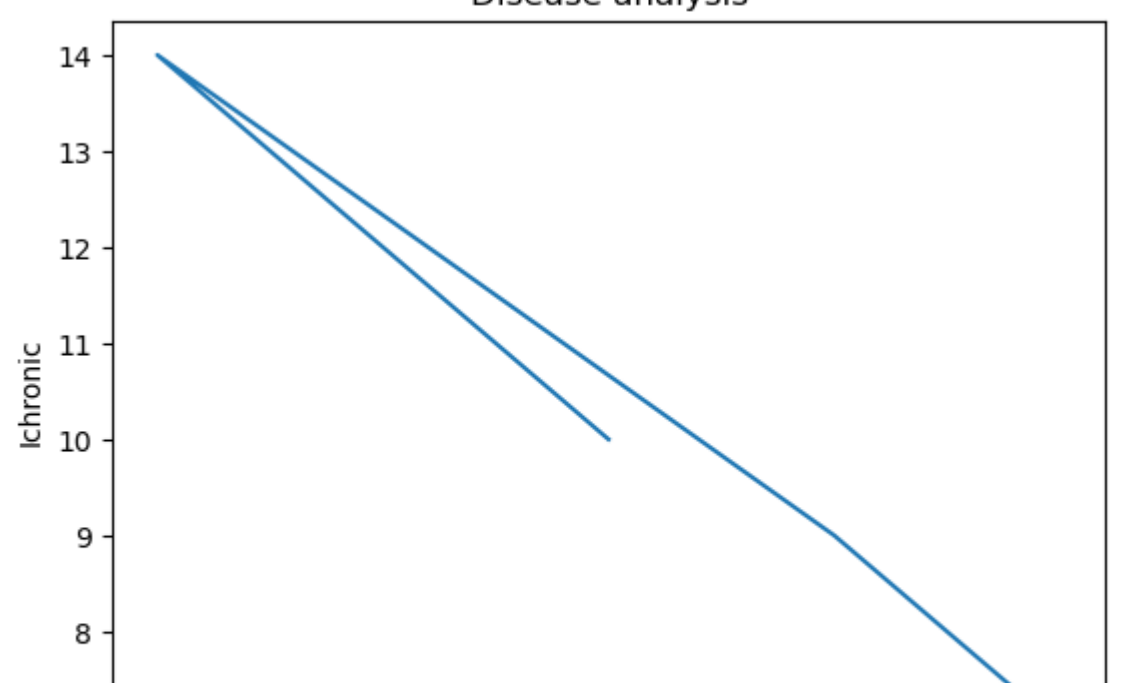
```
plt.plot(x,y)
```

```
plt.xlabel('nchronic')
```

```
plt.ylabel('lchronic')
```

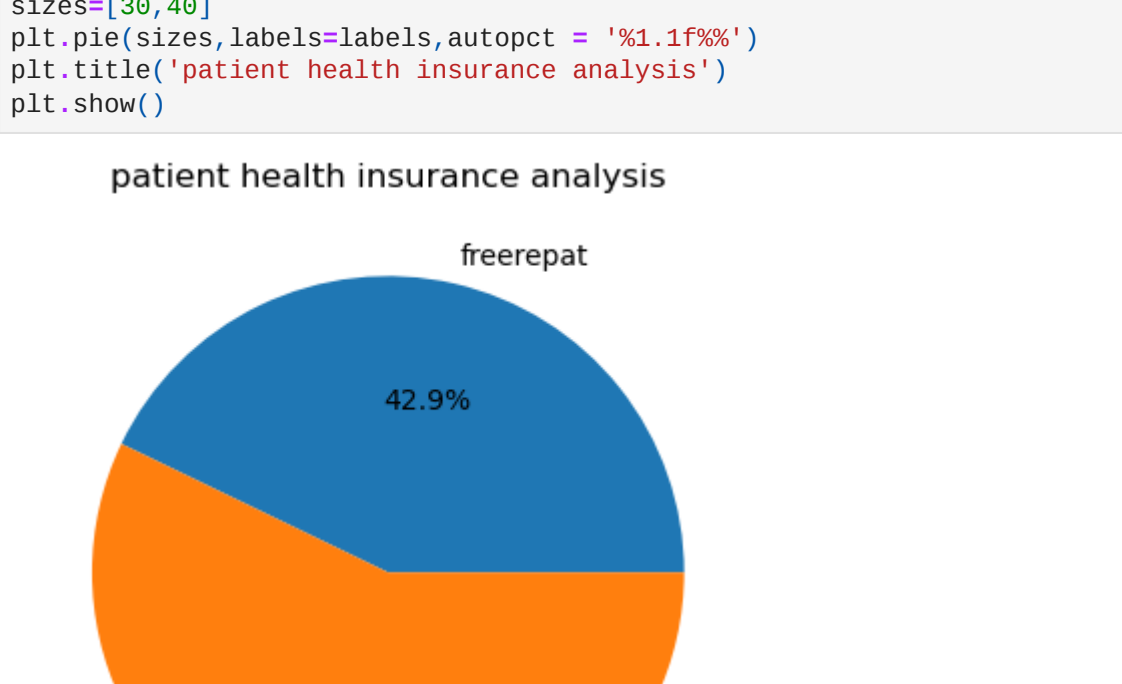
```
plt.title('Disease analysis')
```

```
plt.show()
```



```
labels=["freerepat","freepoor"]
sizes=[30,40]
```

```
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
plt.title('patient health insurance analysis')
```

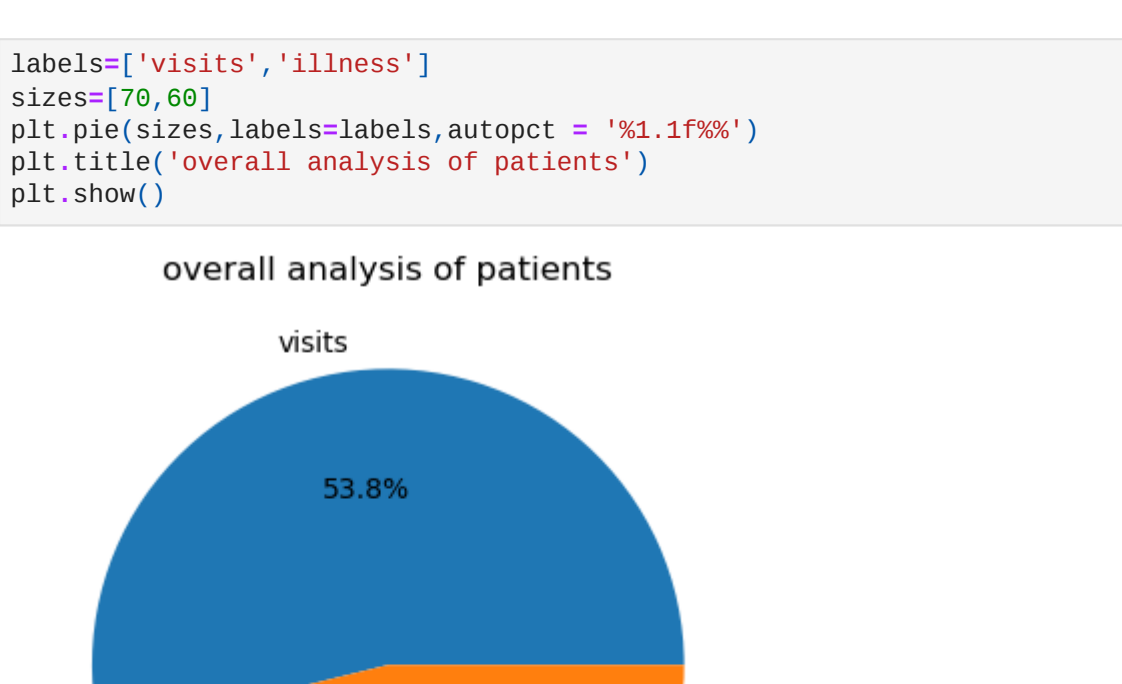


```
labels=["visits","illness"]
sizes=[70,60]
```

```
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
```

```
plt.title('overall analysis of patients')
```

```
plt.show()
```



```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

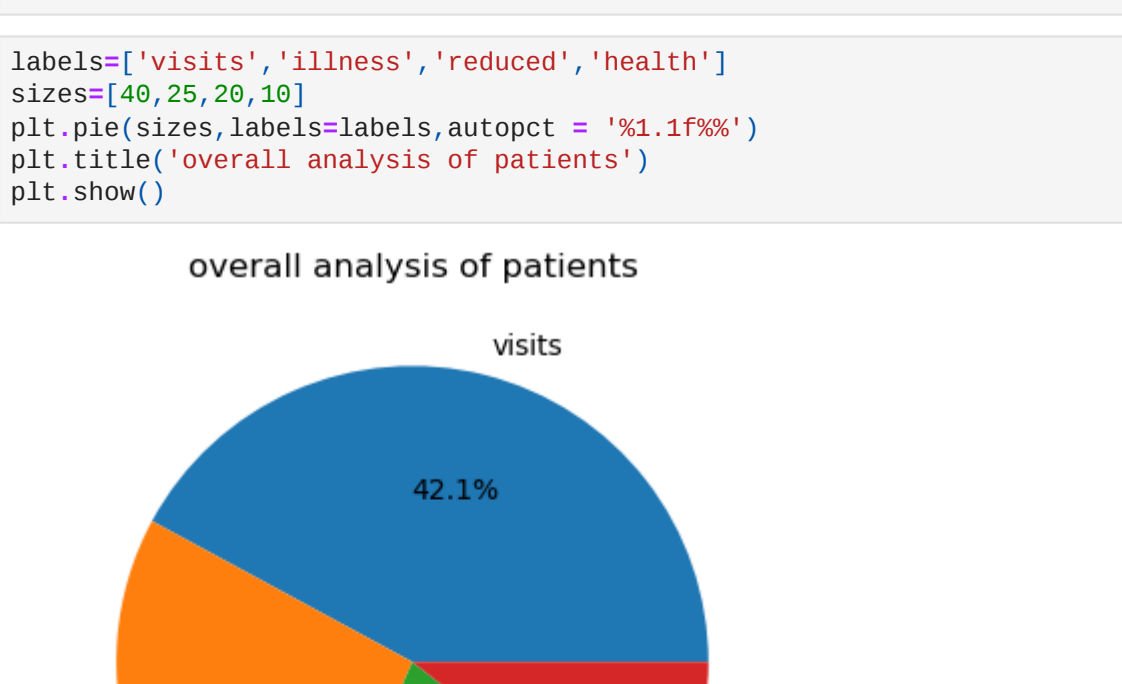
```
labels=["visits","illness","reduced","health"]
```

```
sizes=[25,20,13,5]
```

```
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
```

```
plt.title('overall analysis of patients')
```

```
plt.show()
```



```
x = [20,50,66,99,10]
```

```
y = [30,30,46,80,55]
```

```
plt.scatter(x,y)
```

```
plt.xlabel('freepoor')
```

```
plt.ylabel('private')
```

```
plt.title('insurance analysis')
```

```
plt.show()
```