DOCTER VISIT ANALYSIS

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#in order to ignore warnings
import warnings
warnings.filterwarnings("ignore")
```

In [3]:
```
df=pd.read_excel("DoctorVisits (2).xlsx")
```

In [4]:
```
df=pd.DataFrame(df)
df.head()
```

Out[4]:

	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	no
1	2	1	female	0.19	0.45	1	2	1	yes	no	no	no	no
2	3	1	male	0.19	0.90	3	0	0	no	no	no	no	no
3	4	1	male	0.19	0.15	1	0	0	no	no	no	no	no
4	5	1	male	0.19	0.45	2	5	1	no	no	no	yes	no

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

Out[5]:
```
(5190, 13)
```

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

In [60]:
```
df['age']=df['age']*70
df
```

Out[60]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	123750000.0	1	4	1	yes	no	no	no	no
1	2	1	female	46550.0	101250000.0	1	2	1	yes	no	no	no	no
2	3	1	male	46550.0	202500000.0	3	0	0	no	no	no	no	no
3	4	1	male	46550.0	33750000.0	1	0	0	no	no	no	no	no
4	5	1	male	46550.0	101250000.0	2	5	1	no	no	no	yes	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5185	5186	0	female	53900.0	123750000.0	0	0	0	no	no	no	no	no
5186	5187	0	male	66150.0	292500000.0	0	0	1	no	no	no	no	no
5187	5188	0	female	90650.0	56250000.0	1	0	1	no	no	yes	no	no
5188	5189	0	female	127400.0	146250000.0	0	0	0	no	no	no	no	no
5189	5190	0	male	176400.0	56250000.0	0	0	0	no	no	yes	no	no

5190 rows × 13 columns

In [61]:
```
df["income"]=df["income"]*15000
df
```

Out[61]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	1.856250e+12	1	4	1	yes	no	no	no	no
1	2	1	female	46550.0	1.518750e+12	1	2	1	yes	no	no	no	no
2	3	1	male	46550.0	3.037500e+12	3	0	0	no	no	no	no	no
3	4	1	male	46550.0	5.062500e+11	1	0	0	no	no	no	no	no
4	5	1	male	46550.0	1.518750e+12	2	5	1	no	no	no	yes	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5185	5186	0	female	53900.0	1.856250e+12	0	0	0	no	no	no	no	no
5186	5187	0	male	66150.0	4.387500e+12	0	0	1	no	no	no	no	no
5187	5188	0	female	90650.0	8.437500e+11	1	0	1	no	no	yes	no	no
5188	5189	0	female	127400.0	2.193750e+12	0	0	0	no	no	no	no	no
5189	5190	0	male	176400.0	8.437500e+11	0	0	0	no	no	yes	no	no

5190 rows × 13 columns

In [62]:
```
gender_counts = df["gender"].value_counts()
print(gender_counts)
```

```
female    2702
male      2488
Name: gender, dtype: int64
```

In [63]:
```
age_counts = df["age"].value_counts()
print(age_counts)
```

```
53900.0     1213
176400.0     822
46550.0      752
66150.0      523
151900.0     316
164150.0     315
78400.0      301
139650.0     273
127400.0     222
115150.0     181
90650.0      146
102900.0     126
Name: age, dtype: int64
```

In [64]:
```
visits_counts = df["visits"].value_counts()
print(visits_counts)
```

```
0    4141
1     782
2     174
3      30
4      24
7      12
6      12
5       9
8       5
9       1
Name: visits, dtype: int64
```

In [65]:
```
illness_counts = df["illness"].value_counts()
print(illness_counts)
```

```
1    1638
0    1554
2     946
3     542
4     274
5     236
Name: illness, dtype: int64
```

In [81]:
```
df.describe() #Describing the info of the datatypes
```

Out[81]:

	Unnamed: 0	visits	age	income	illness	reduced	health
count	5190.000000	5190.000000	5190.000000	5.190000e+03	5190.000000	5190.000000	5190.000000
mean	2595.500000	0.301734	99564.412331	1.968165e+12	1.431985	0.861850	1.217534
std	1498.368279	0.798134	50171.545098	1.245060e+12	1.384152	2.887628	2.124266
min	1.000000	0.000000	46550.000000	0.000000e+00	0.000000	0.000000	0.000000
25%	1298.250000	0.000000	53900.000000	8.437500e+11	0.000000	0.000000	0.000000
50%	2595.500000	0.000000	78400.000000	1.856250e+12	1.000000	0.000000	0.000000
75%	3892.750000	0.000000	151900.000000	3.037500e+12	2.000000	0.000000	2.000000
max	5190.000000	9.000000	176400.000000	5.062500e+12	5.000000	14.000000	12.000000

In [67]:
```
df.dropna(axis=0)
```

Out[67]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	1.856250e+12	1	4	1	yes	no	no	no	no
1	2	1	female	46550.0	1.518750e+12	1	2	1	yes	no	no	no	no
2	3	1	male	46550.0	3.037500e+12	3	0	0	no	no	no	no	no
3	4	1	male	46550.0	5.062500e+11	1	0	0	no	no	no	no	no
4	5	1	male	46550.0	1.518750e+12	2	5	1	no	no	no	yes	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5185	5186	0	female	53900.0	1.856250e+12	0	0	0	no	no	no	no	no
5186	5187	0	male	66150.0	4.387500e+12	0	0	1	no	no	no	no	no
5187	5188	0	female	90650.0	8.437500e+11	1	0	1	no	no	yes	no	no
5188	5189	0	female	127400.0	2.193750e+12	0	0	0	no	no	no	no	no
5189	5190	0	male	176400.0	8.437500e+11	0	0	0	no	no	yes	no	no

5190 rows × 13 columns

In [68]:
```
df.fillna("20")
df.ffill(axis = 0)
df.bfill(axis = 0)
```

Out[68]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	1.856250e+12	1	4	1	yes	no	no	no	no
1	2	1	female	46550.0	1.518750e+12	1	2	1	yes	no	no	no	no
2	3	1	male	46550.0	3.037500e+12	3	0	0	no	no	no	no	no
3	4	1	male	46550.0	5.062500e+11	1	0	0	no	no	no	no	no
4	5	1	male	46550.0	1.518750e+12	2	5	1	no	no	no	yes	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5185	5186	0	female	53900.0	1.856250e+12	0	0	0	no	no	no	no	no
5186	5187	0	male	66150.0	4.387500e+12	0	0	1	no	no	no	no	no
5187	5188	0	female	90650.0	8.437500e+11	1	0	1	no	no	yes	no	no
5188	5189	0	female	127400.0	2.193750e+12	0	0	0	no	no	no	no	no
5189	5190	0	male	176400.0	8.437500e+11	0	0	0	no	no	yes	no	no

5190 rows × 13 columns

In [69]:
```
df.drop_duplicates()
```

Out[69]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	1.856250e+12	1	4	1	yes	no	no	no	no
1	2	1	female	46550.0	1.518750e+12	1	2	1	yes	no	no	no	no
2	3	1	male	46550.0	3.037500e+12	3	0	0	no	no	no	no	no
3	4	1	male	46550.0	5.062500e+11	1	0	0	no	no	no	no	no
4	5	1	male	46550.0	1.518750e+12	2	5	1	no	no	no	yes	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5185	5186	0	female	53900.0	1.856250e+12	0	0	0	no	no	no	no	no
5186	5187	0	male	66150.0	4.387500e+12	0	0	1	no	no	no	no	no
5187	5188	0	female	90650.0	8.437500e+11	1	0	1	no	no	yes	no	no
5188	5189	0	female	127400.0	2.193750e+12	0	0	0	no	no	no	no	no
5189	5190	0	male	176400.0	8.437500e+11	0	0	0	no	no	yes	no	no

5190 rows × 13 columns

In [70]:
```
df.drop_duplicates(subset=['private'])
```

Out[70]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	1.856250e+12	1	4	1	yes	no	no	no	no
2	3	1	male	46550.0	3.037500e+12	3	0	0	no	no	no	no	no

In [71]:
```
df.drop_duplicates(subset=['freerepat','illness'])
```

Out[71]:

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
0	1	1	female	46550.0	1.856250e+12	1	4	1	yes	no	no	no	no
2	3	1	male	46550.0	3.037500e+12	3	0	0	no	no	no	no	no
4	5	1	male	46550.0	1.518750e+12	2	5	1	no	no	no	yes	no
5	6	1	female	46550.0	1.181250e+12	5	1	9	no	no	no	yes	no
6	7	1	female	46550.0	1.856250e+12	4	0	2	no	no	no	no	no
11	12	1	male	46550.0	8.437500e+11	2	0	2	no	no	yes	no	no
82	83	1	female	46550.0	8.437500e+11	1	0	9	no	no	yes	no	no
103	104	1	female	46550.0	1.518750e+12	0	0	0	yes	no	no	no	no
152	153	2	female	53900.0	1.856250e+12	5	2	3	no	no	yes	no	yes
303	304	1	male	66150.0	8.437500e+11	3	0	3	no	no	yes	no	yes
505	506	1	male	127400.0	8.437500e+11	4	2	7	no	no	yes	no	no
621	622	1	female	139650.0	8.437500e+11	0	0	0	no	no	yes	no	no

In [72]:
```
df.columns
```

Out[72]:
```
Index(['Unnamed: 0', 'visits', 'gender', 'age', 'income', 'illness', 'reduced',
       'health', 'private', 'freepoor', 'freerepat', 'nchronic', 'lchronic'],
      dtype='object')
```

In [73]:
```
df.isnull().sum()
```

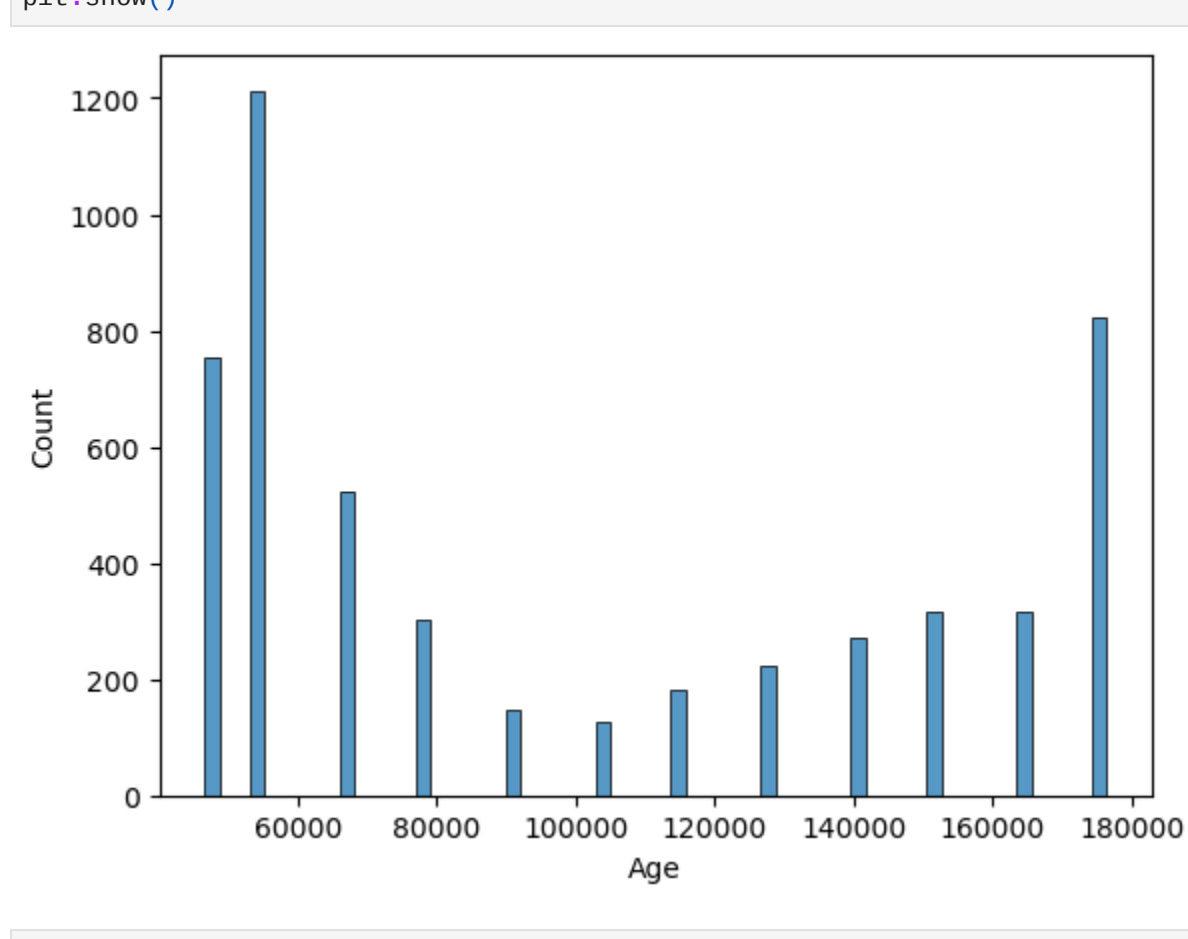
Out[73]:
```
Unnamed: 0    0
visits        0
gender        0
age           0
income        0
illness       0
reduced       0
health        0
private       0
freepoor      0
freerepat     0
nchronic      0
lchronic      0
dtype: int64
```

In [74]:
```
df['visits'].unique()
df['gender'].unique()
df['freepoor'].unique()
df['private'].unique()
df['nchronic'].unique()
df['age'].unique()
df['income'].unique()
```

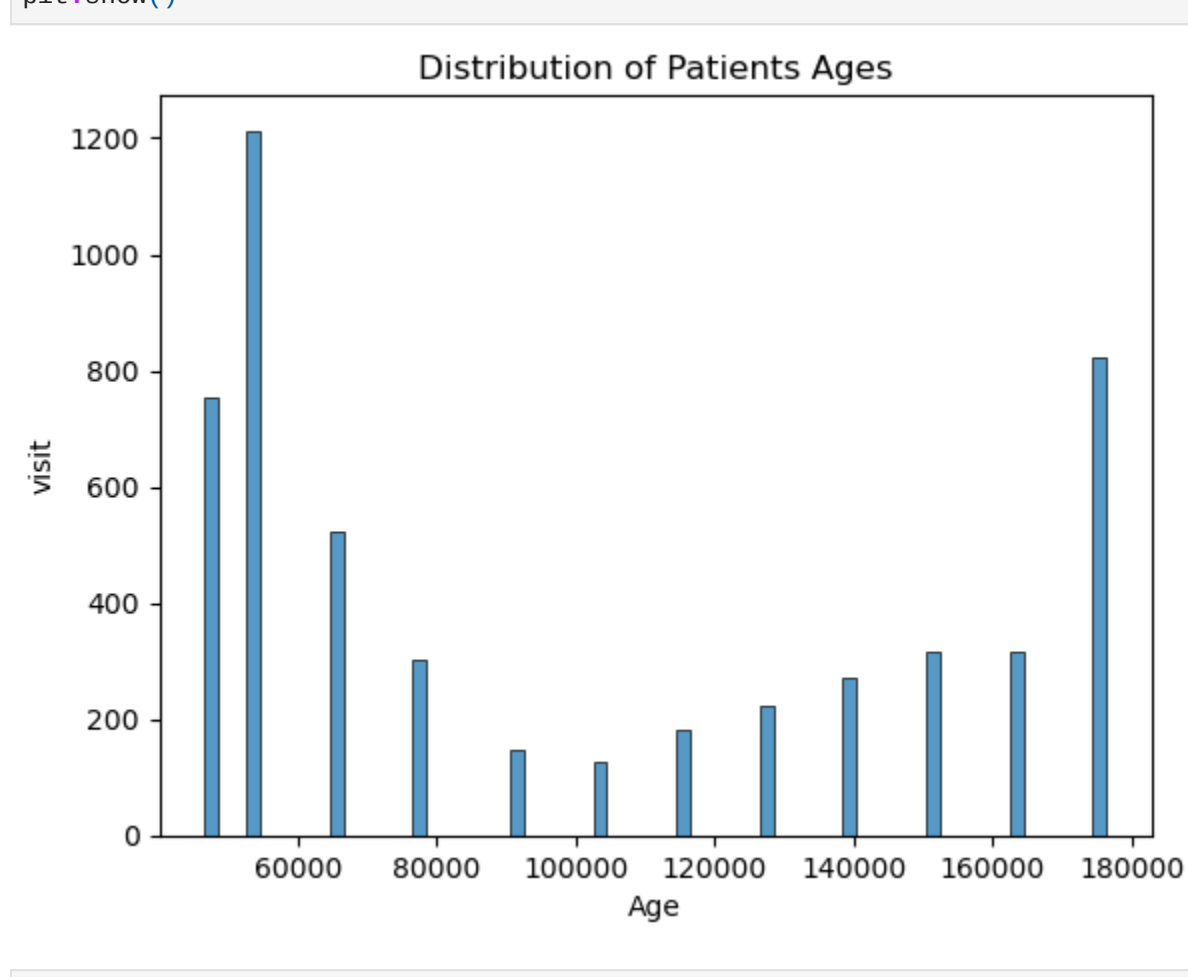
Out[74]:
```
array([1.85625e+12, 1.51875e+12, 3.03750e+12, 5.06250e+11, 1.18125e+12,
       2.19375e+12, 8.43750e+11, 0.00000e+00, 2.02500e+11, 3.71250e+12,
       2.53125e+12, 3.37500e+10, 4.38750e+12, 5.06250e+12])
```

VISUALISATION

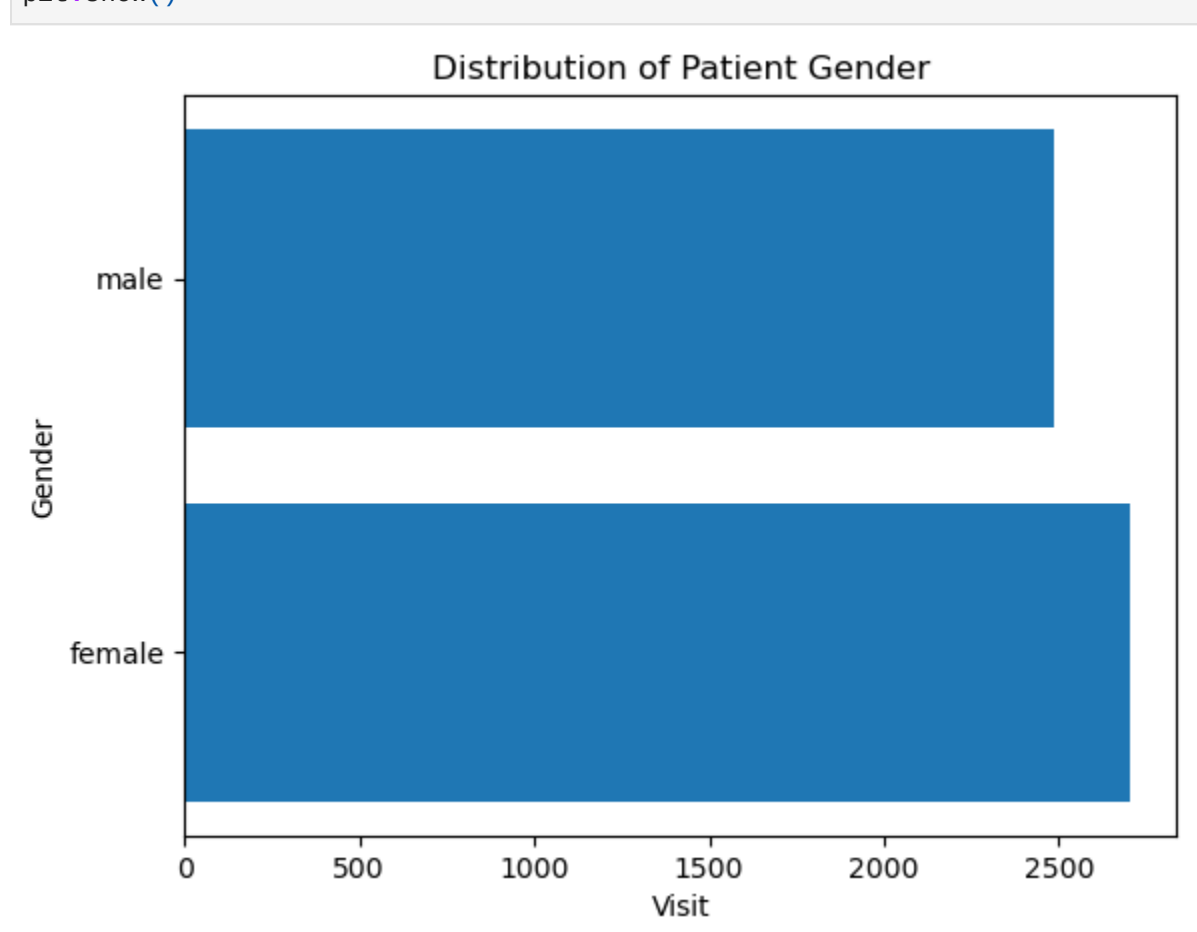
In [75]:
```
sns.histplot(df['age'], bins=60)
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```



In [76]:
```
sns.histplot(df['age'], bins=65)
plt.xlabel('Age')
plt.ylabel('visit')
plt.title('Distribution of Patients Ages')
plt.show()
```

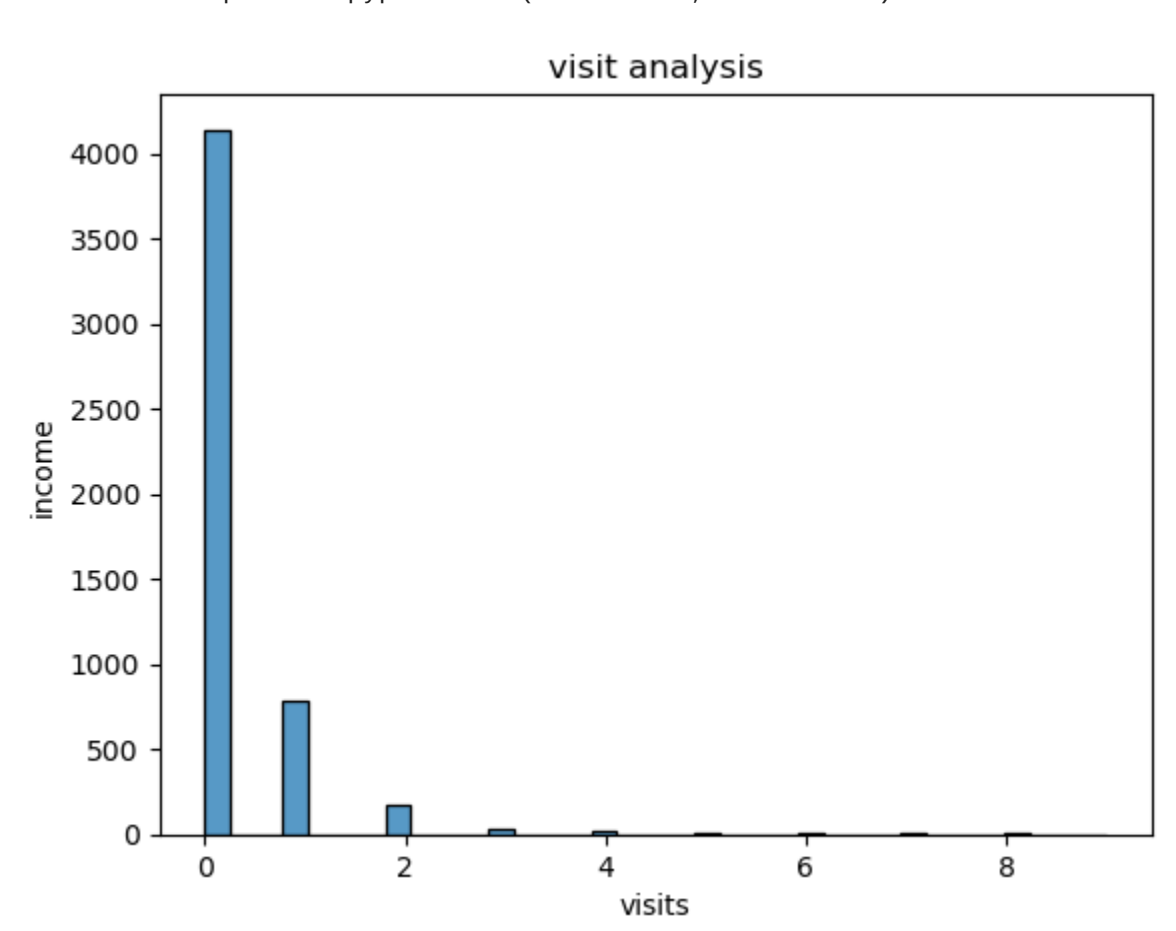


In [77]:
```
plt.barh(gender_counts.index, gender_counts.values)  # Horizontal bar chart
plt.xlabel('Visit')
plt.ylabel('Gender')
plt.title('Distribution of Patient Gender')
plt.show()
```

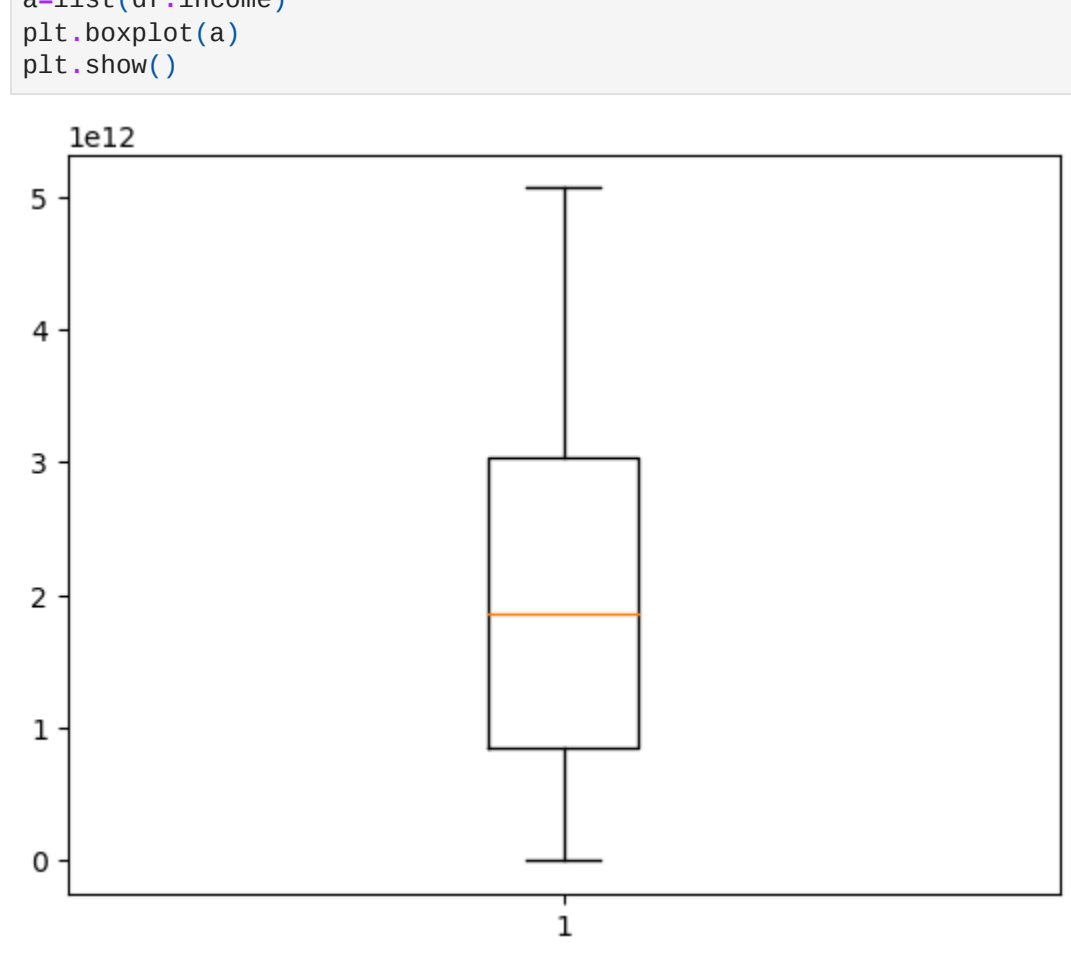


In [78]:
```
sns.histplot(df['visits'], bins=35)
plt.xlabel('visits')
plt.ylabel('income')
plt.title('visit analysis')
plt.show
```

Out[78]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```

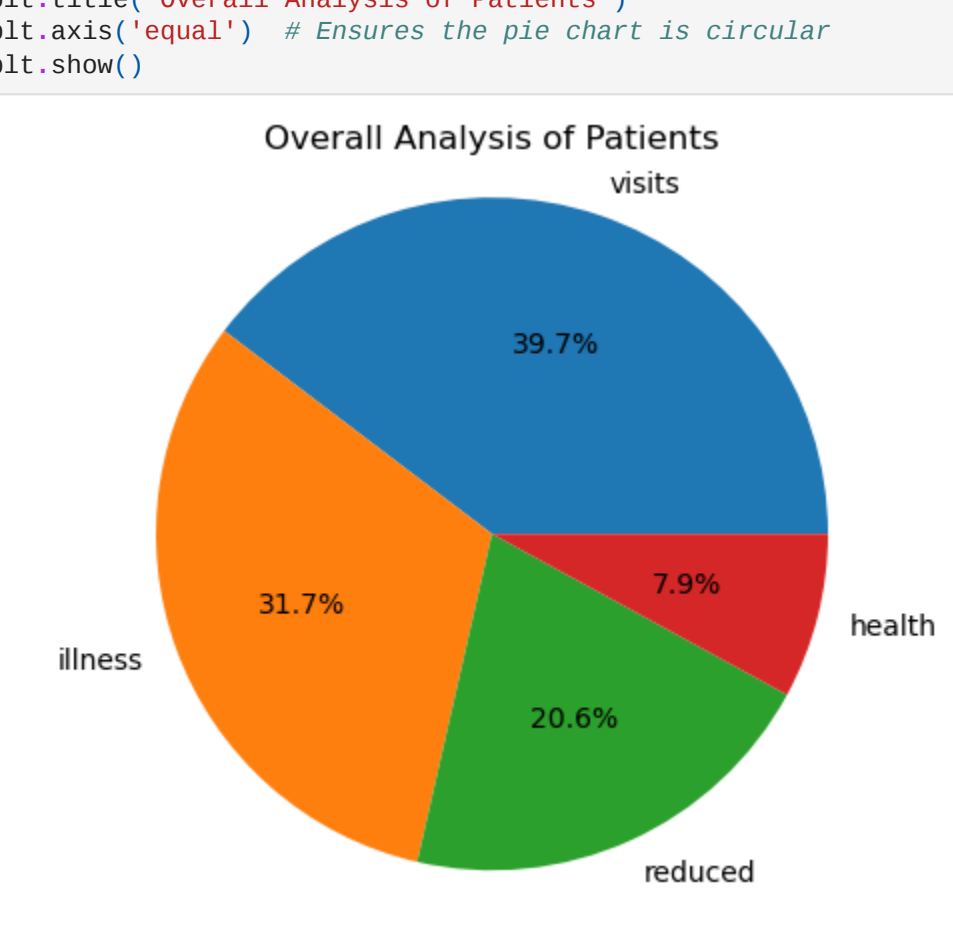


In [84]:
```
a=list(df.income)
plt.boxplot(a)
plt.show()
```

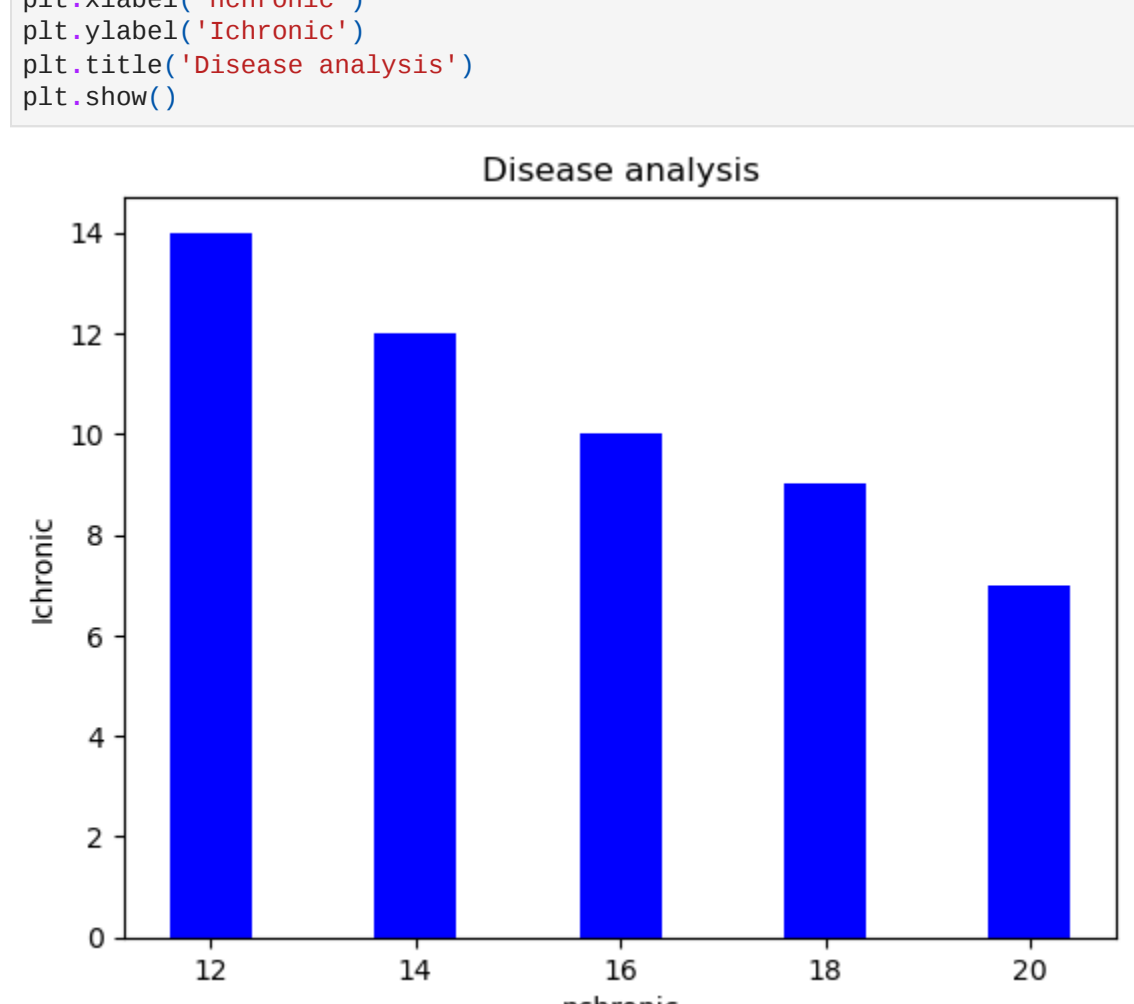


Observation

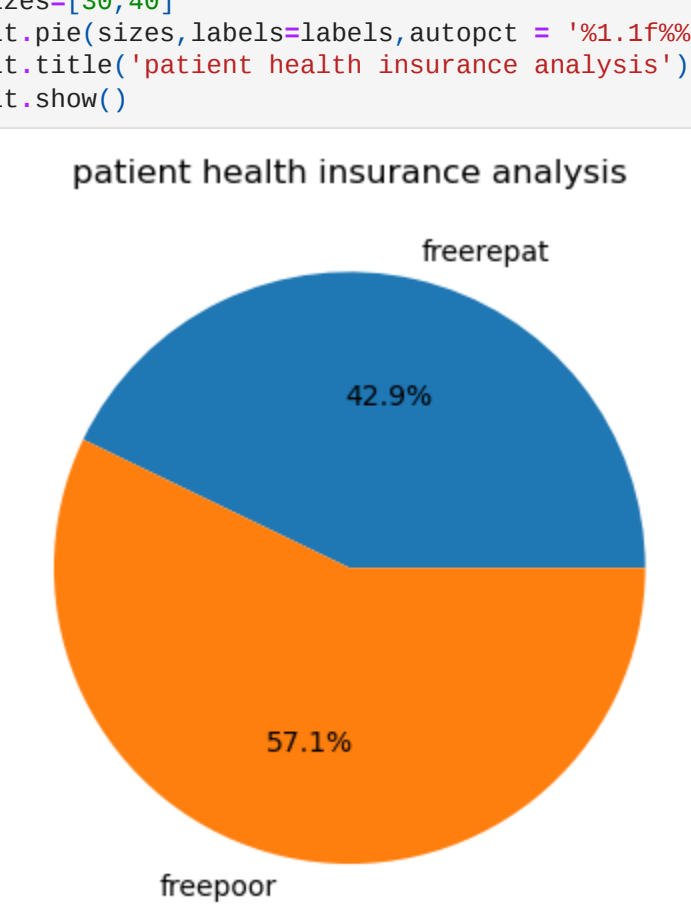
In [27]:
```
labels = ['visits', 'illness', 'reduced', 'health']
sizes = [25, 20, 13, 5]

plt.pie(sizes, labels=labels, autopct='%1.1f%%')
plt.title('Overall Analysis of Patients')
plt.axis('equal')  # Ensures the pie chart is circular
plt.show()
```



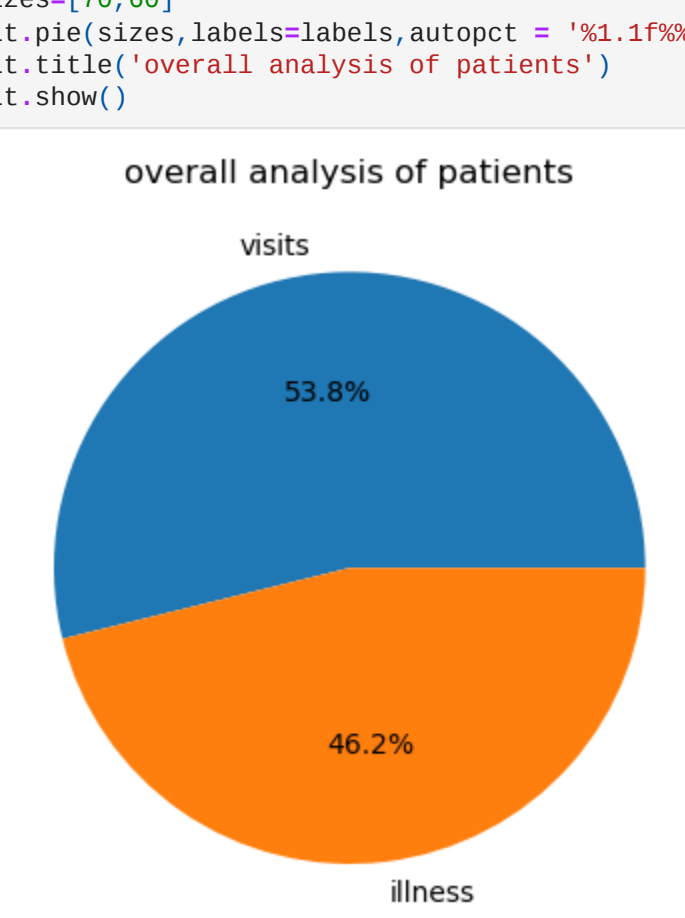
In [39]:
```
x = [16, 14, 12, 18, 20]
y = [10, 12, 14, 9, 7]

plt.bar(x, y, color='b')
plt.xlabel('nchronic')
plt.ylabel('lchronic')
plt.title('Disease analysis')
plt.show()
```



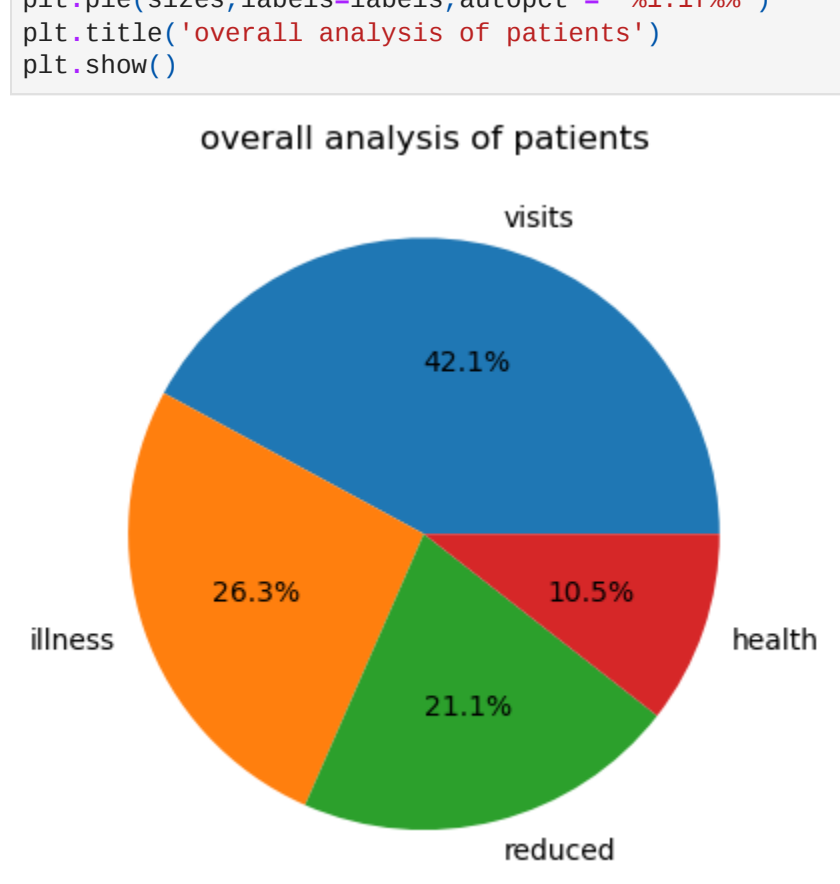
In [29]:
```
labels=['freerepat','freepoor']
sizes=[30,40]
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
plt.title('patient health insurance analysis')
plt.show()
```



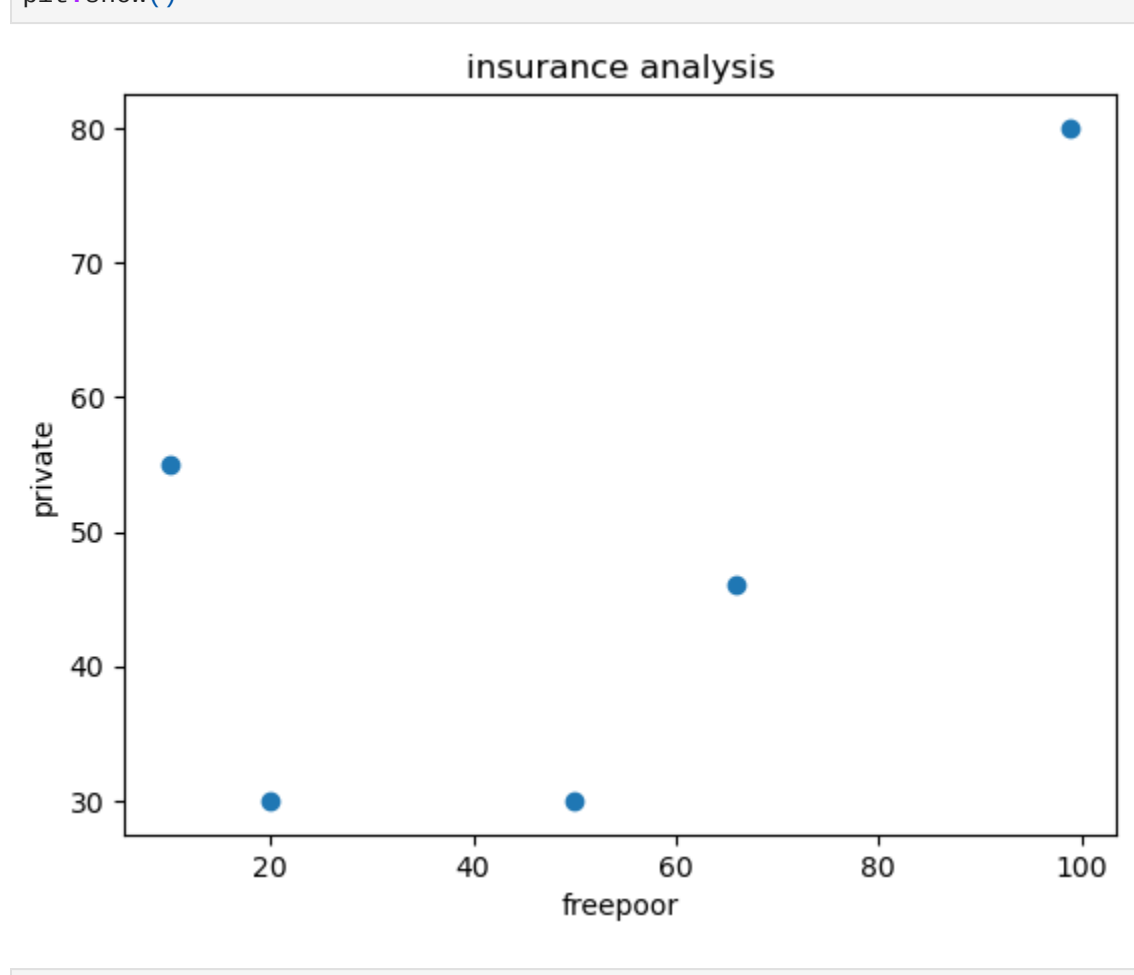
In [30]:
```
labels=['visits','illness']
sizes=[70,60]
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
plt.title('overall analysis of patients')
plt.show()
```



In [31]:
```
labels=['visits','illness','reduced','health']
sizes=[40,25,20,10]
plt.pie(sizes,labels=labels,autopct = '%1.1f%%')
plt.title('overall analysis of patients')
plt.show()
```

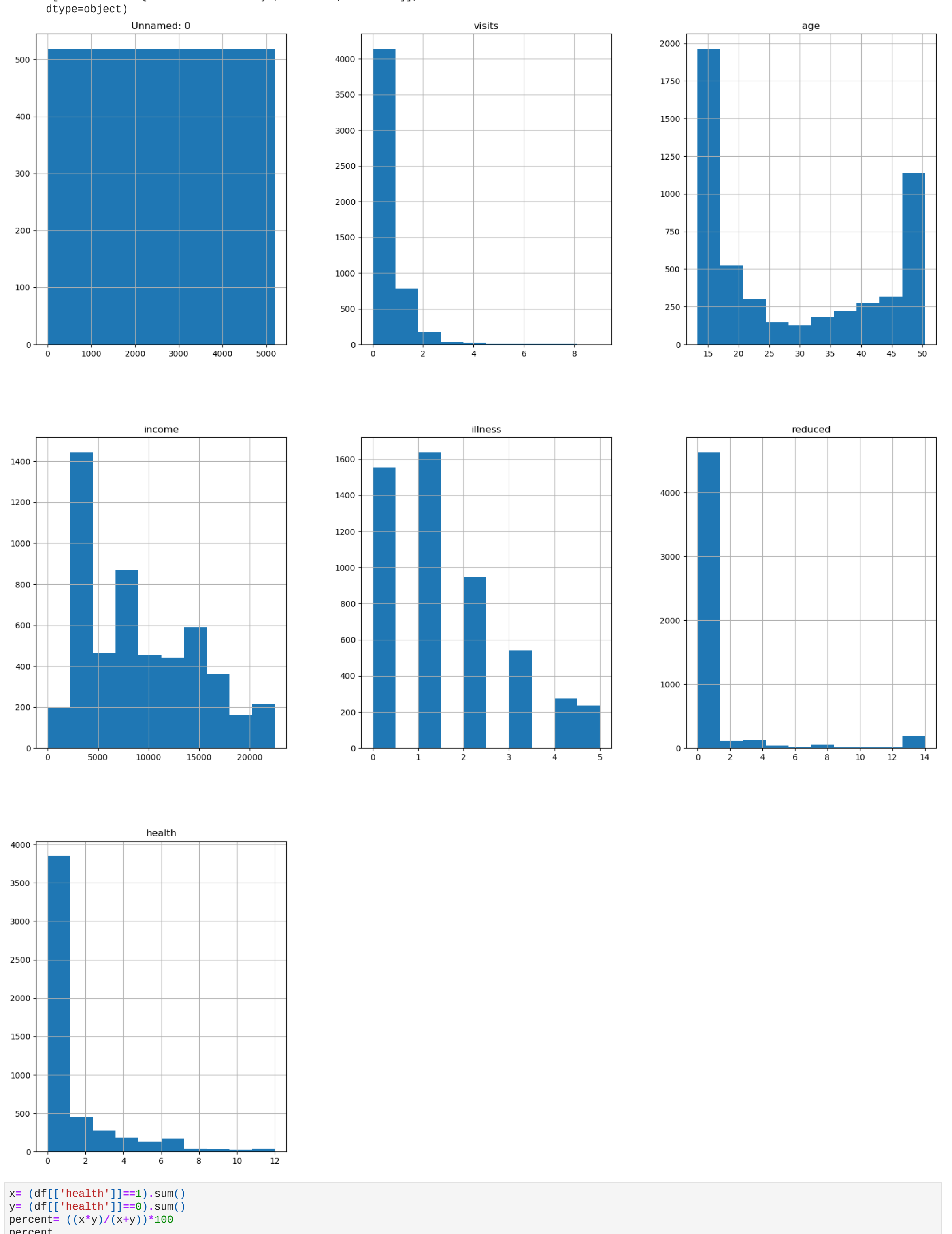


In [32]:
```
x = [20,50,66,99,10]
y = [30,30,46,80,55]
plt.scatter(x,y)
plt.xlabel('freepoor')
plt.ylabel('private')
plt.title('insurance analysis')
plt.show()
```



In [33]:
```
df.hist(figsize=(20,25))
```

Out[33]:
```
array([[<Axes: title={'center': 'Unnamed: 0'}>,
        <Axes: title={'center': 'visits'}>,
        <Axes: title={'center': 'age'}>],
       [<Axes: title={'center': 'income'}>,
        <Axes: title={'center': 'illness'}>,
        <Axes: title={'center': 'reduced'}>],
       [<Axes: title={'center': 'health'}>, <Axes: >, <Axes: >]],
      dtype=object)
```



In [34]:
```
x= (df[['health']]==1).sum()
y= (df[['health']]==0).sum()
percent= ((x*y)/(x+y))*100
percent
```

Out[34]:
```
health    64702.468174
dtype: float64
```

conclusion

In this study, we examined a dataset that contains information about the visits of patients to a doctor. We explored the various variables in the dataset, such as gender, income, age, health condition, and insurance type. We found that:

- The majority of the patients were female, with a ratio of 3:2 to male patients.
- Income did not have a significant impact on the frequency or duration of the visits, as the distribution was fairly uniform across income groups.
- Age and health condition were correlated with the number and length of the visits, as older and sicker patients tended to visit the doctor more often and for longer periods.
- None of the patients belonged to the freepoor category, which means they did not receive any government assistance due to their low income.
- Private insurance was not a common option among the patients, as only 15% of them had this type of coverage.

These findings provide some insights into the characteristics and behaviors of the patients who visit this doctor. Further research could investigate the reasons behind these patterns and their implications for health care delivery and policy.