In [1]:
```
import numpy as np
import pandas as pd
```

In [4]:
```
df=pd.read_csv("Bank_Customer_Churn.csv",encoding='latin1')
```

In [5]:
```
df.head()
```

Out[5]:

	Unnamed: 0	age	gender	dependents	occupation	city	churn
0	0	66	Male	0.0	self_employed	187.0	0
1	1	35	Male	0.0	self_employed	NaN	0
2	2	31	Male	0.0	salaried	146.0	0
3	3	90	NaN	NaN	self_employed	1020.0	1
4	4	42	Male	2.0	self_employed	1494.0	1

In [7]:
```
df.shape
```

Out[7]:
```
(28382, 7)
```

In [8]:
```
df.tail
```

Out[8]:
```
<bound method NDFrame.tail of        Unnamed: 0  age  gender  dependents
occupation    city  churn
0                0   66    Male         0.0  self_employed   187.0      0
1                1   35    Male         0.0  self_employed     NaN      0
2                2   31    Male         0.0       salaried   146.0      0
3                3   90     NaN         NaN  self_employed  1020.0      1
4                4   42    Male         2.0  self_employed  1494.0      1
...            ...  ...     ...         ...            ...     ...    ...
28377        28377   10  Female         0.0        student  1020.0      0
28378        28378   34  Female         0.0  self_employed  1046.0      0
28379        28379   47    Male         0.0       salaried  1096.0      1
28380        28380   50    Male         3.0  self_employed  1219.0      0
28381        28381   18    Male         0.0        student  1232.0      1

[28382 rows x 7 columns]>
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
Unnamed: 0       0
age              0
gender         525
dependents    2463
occupation      80
city           803
churn            0
dtype: int64
```

In [10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28382 entries, 0 to 28381
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   Unnamed: 0  28382 non-null  int64  
 1   age         28382 non-null  int64  
 2   gender      27857 non-null  object 
 3   dependents  25919 non-null  float64
 4   occupation  28302 non-null  object 
 5   city        27579 non-null  float64
 6   churn       28382 non-null  int64  
dtypes: float64(2), int64(3), object(2)
memory usage: 1.5+ MB
```

In [11]:
```
city_mean=df['city'].mean()
```

In [12]:
```
df['city'].replace(np.NaN,city_mean,inplace=True)
```

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
Unnamed: 0       0
age              0
gender         525
dependents    2463
occupation      80
city             0
churn            0
dtype: int64
```

In [20]:
```
df['occupation']=df['occupation'].dropna()
```

In [21]:
```
df.isnull().sum()
```

Out[21]:
```
Unnamed: 0       0
age              0
gender         525
dependents    2463
occupation       0
city             0
churn            0
dtype: int64
```

In [22]:
```
gender_mode=df['gender'].mode()
```

In [23]:
```
gender_mode
```

Out[23]:
```
0    Male
Name: gender, dtype: object
```

In [24]:
```
df['gender']=df['gender'].fillna(value='gender_mode')
```

In [25]:
```
df.isnull().sum()
```

Out[25]:
```
Unnamed: 0       0
age              0
gender           0
dependents    2463
occupation       0
city             0
churn            0
dtype: int64
```

In [26]:
```
dep_mean=df['dependents'].mean()
```

In [27]:
```
df['dependents']=df['dependents'].fillna(value='dep_mean')
```

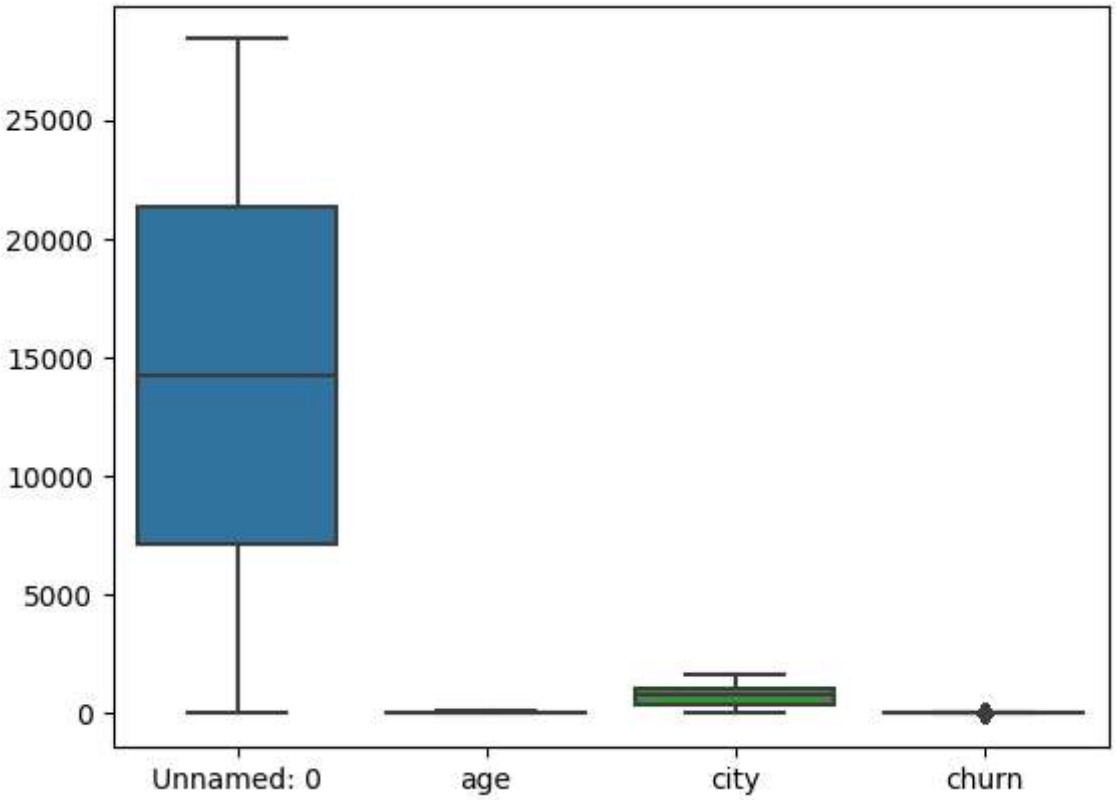
In [28]:
```
df.isnull().sum()
```

Out[28]:
```
Unnamed: 0    0
age           0
gender        0
dependents    0
occupation    0
city          0
churn         0
dtype: int64
```

In [29]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

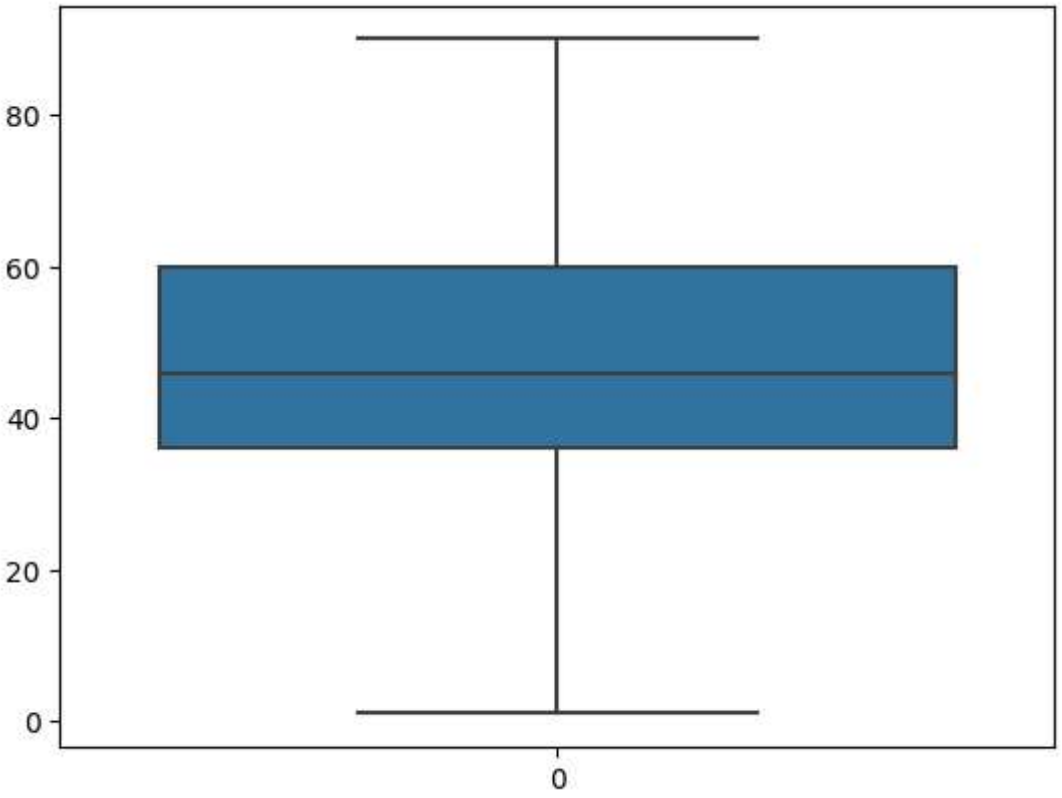
In [30]:
```
sns.boxplot(df)
```

Out[30]: <AxesSubplot: >



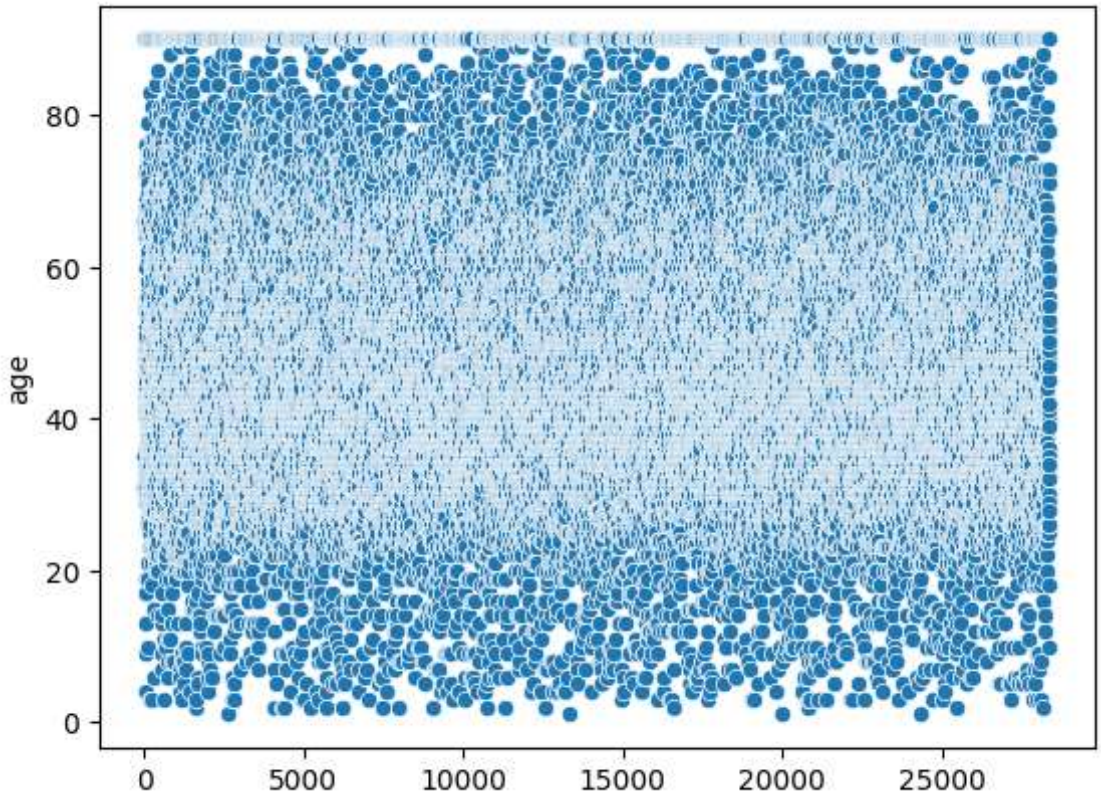
In [32]:
```
sns.boxplot(df['age'])
```

Out[32]: <AxesSubplot: >



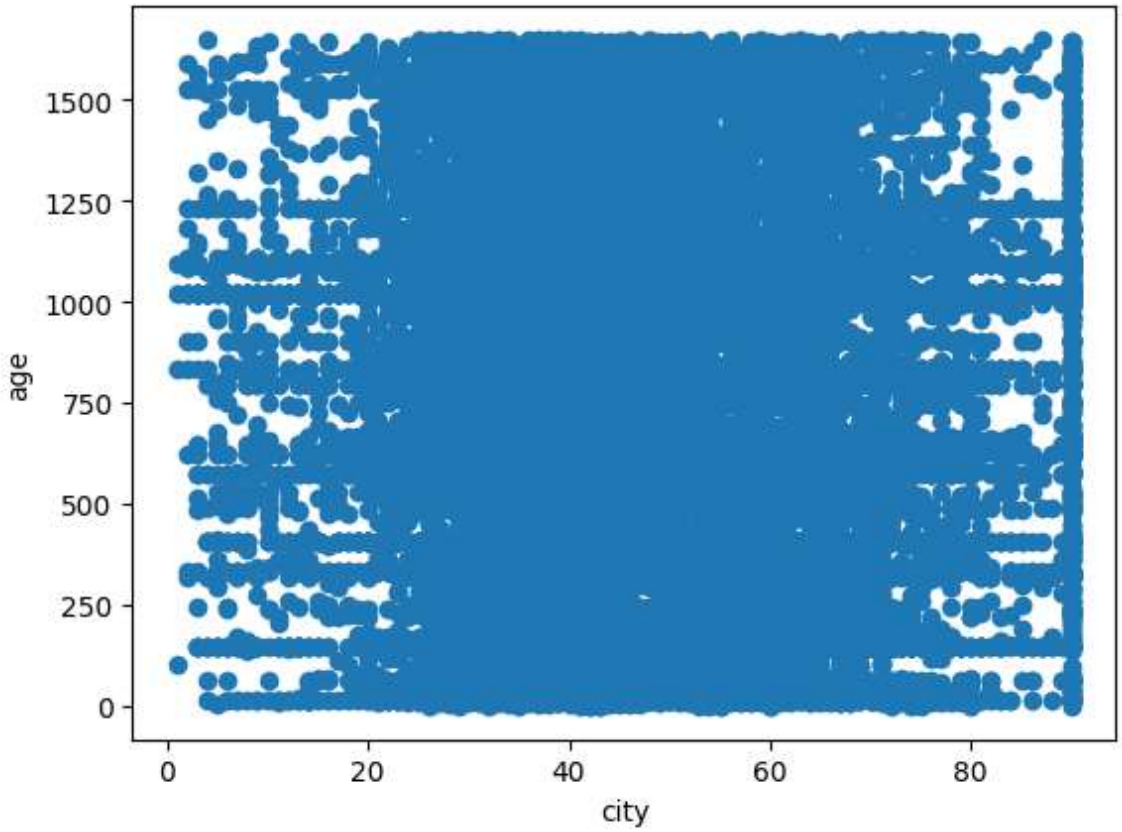
In [33]:
```
sns.scatterplot(df['age'])
```

Out[33]:
```
<AxesSubplot: ylabel='age'>
```



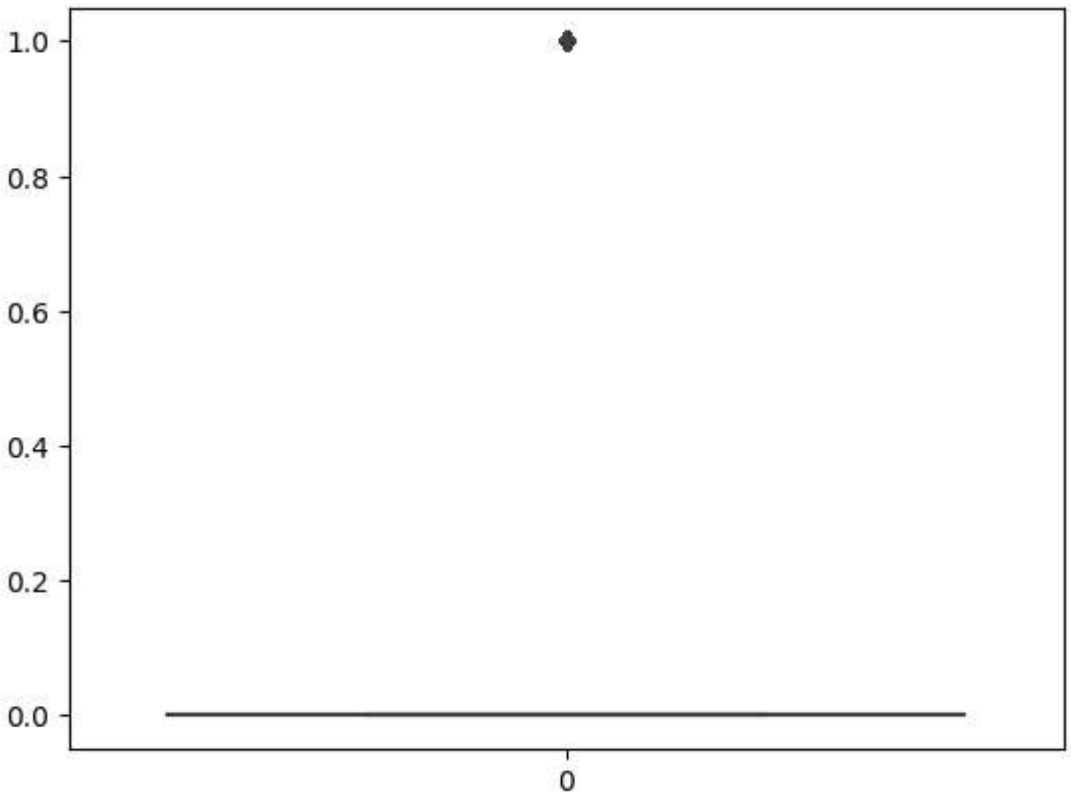
In [35]:
```
plt.scatter(y=df['city'],x=df['age'])
plt.xlabel("city")
plt.ylabel("age")
```

Out[35]:
```
Text(0, 0.5, 'age')
```



In [36]:
```
sns.boxplot(df['churn'])
```

Out[36]: <AxesSubplot: >



In [39]:
```
Q1=np.percentile(df['age'],25)
Q3=np.percentile(df['age'],75)
```

In [40]:
```
Q1,Q3
```

Out[40]: (36.0, 60.0)

In [41]:
```
IQR=Q3-Q1
```

In [42]:
```
IQR
```

Out[42]: 24.0

In [43]:
```
lower_bound=Q1-1.5*IQR
upper_bound=Q3+1.5*IQR
```

In [44]:
```
lower_bound,upper_bound
```

Out[44]: (0.0, 96.0)

In [45]:
```
outliers_rows=df[(df['age'] < lower_bound) | (df['age'] > upper_bound)]
```

In [46]:
```
outliers_rows.shape
```

Out[46]: (0, 7)

In [48]:
```
outliers_rows
```

Out[48]:

	Unnamed: 0	age	gender	dependents	occupation	city	churn

In [49]:
```
df['age']=np.where(
    df['age']>upper_bound,
    upper_bound,np.where(
    df['age']<lower_bound,
        lower_bound,
        df['age']
    )
)
```

In [50]:
```
outliers_rows=df[(df['age'] < lower_bound) | (df['age'] > upper_bound)]
```

In [51]:
```
outliers_rows
```

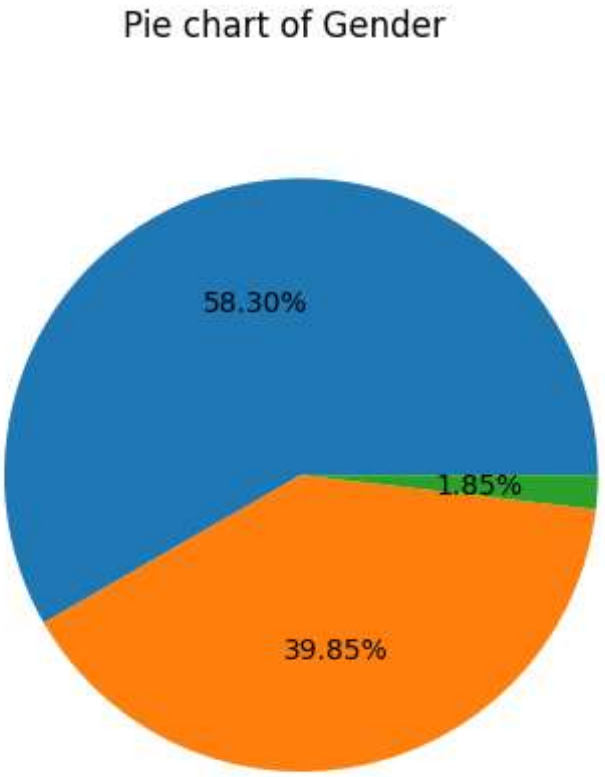
Out[51]:

	Unnamed: 0	age	gender	dependents	occupation	city	churn

In [52]:
```
df['gender'].value_counts()
```

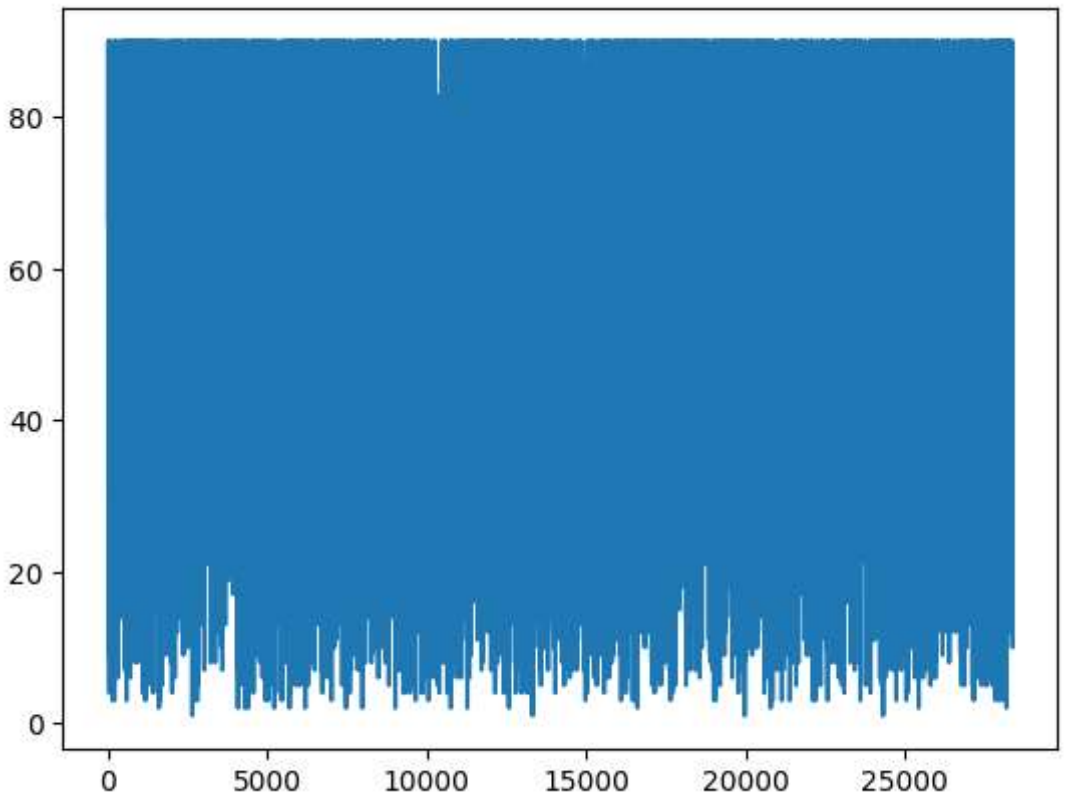
Out[52]:
```
Male           16548
Female         11309
gender_mode      525
Name: gender, dtype: int64
```

In [53]:
```
plt.pie(df['gender'].value_counts(),autopct="%1.2f%%")
plt.suptitle("Pie chart of Gender")
plt.show()
```



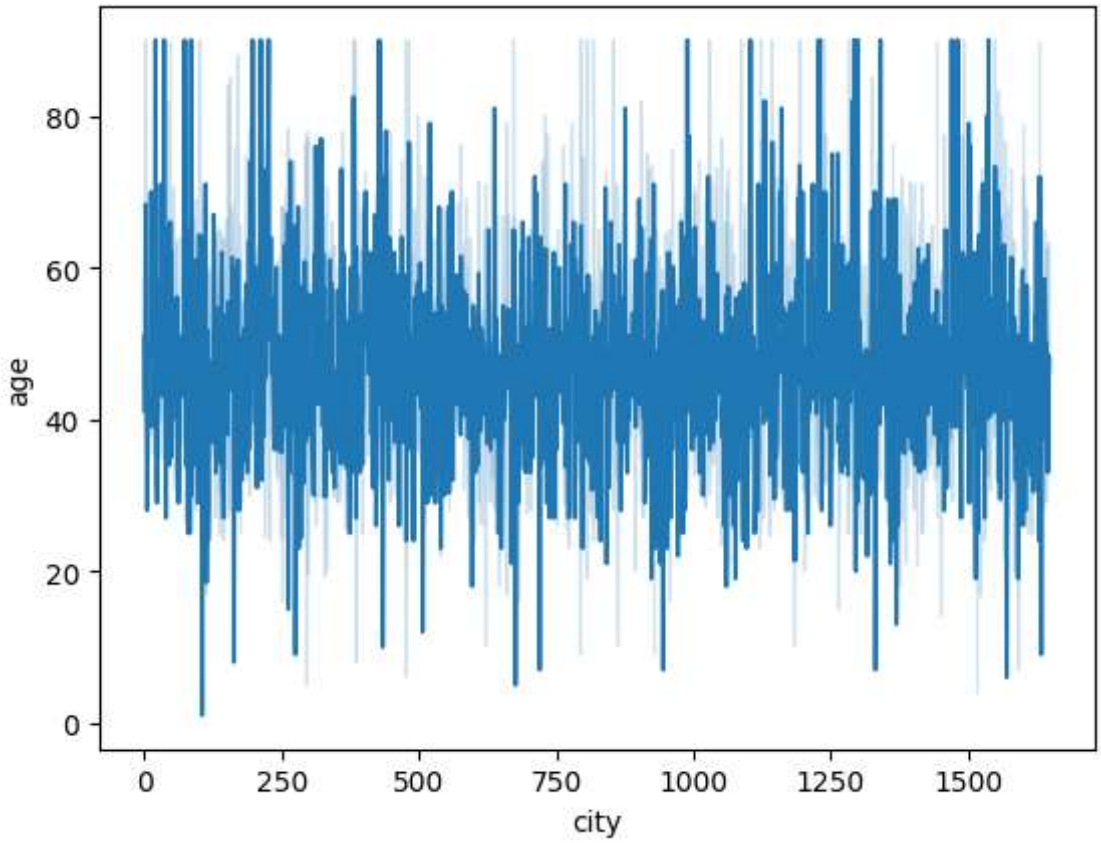
In [54]:
```
plt.plot(df['age'])
```

Out[54]:
```
[<matplotlib.lines.Line2D at 0x22a63dd2530>]
```

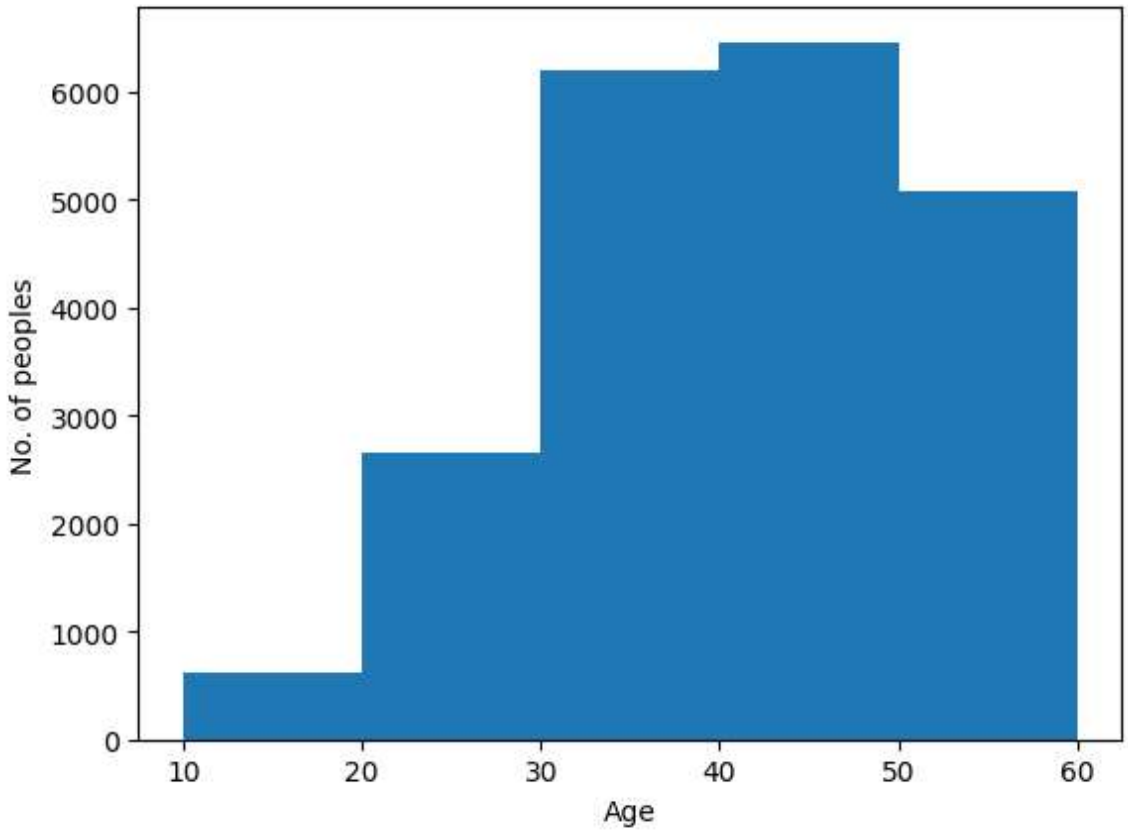


In [55]:
```
sns.lineplot(y=df['age'],x=df['city'])
```

Out[55]:
```
<AxesSubplot: xlabel='city', ylabel='age'>
```

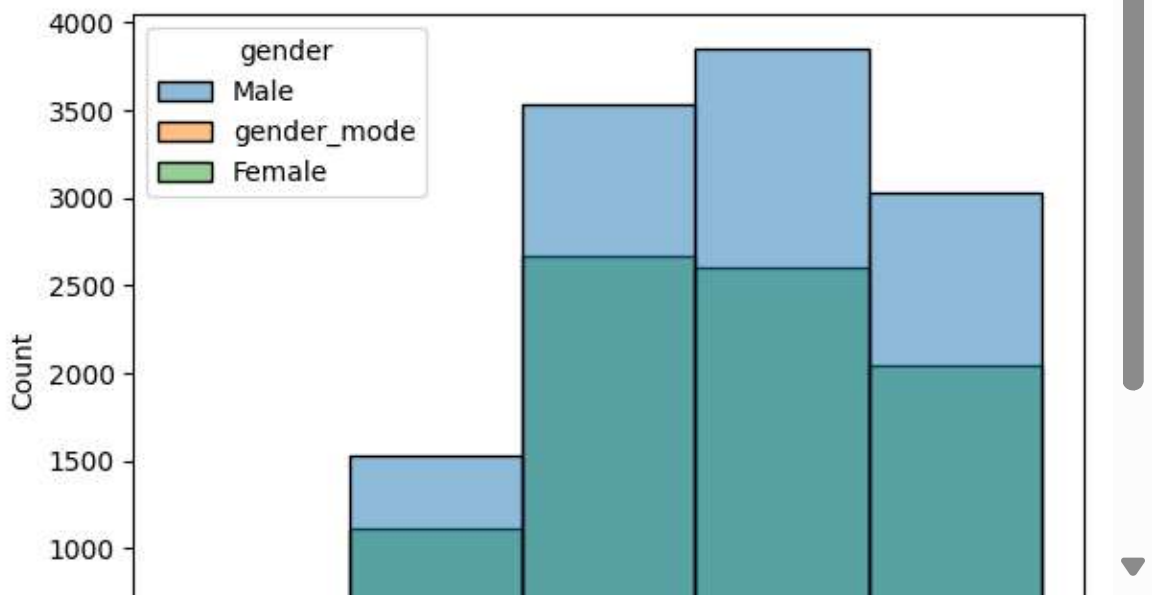


In [56]:
```
bins=[10,20,30,40,50,60]
plt.hist(x=df["age"],bins=bins)
plt.xlabel("Age")
plt.ylabel("No. of peoples")
plt.show()
```



In [57]:
```
sns.histplot(x=df['age'],bins=bins,hue=df['gender'])
```

Out[57]:
```
<AxesSubplot: xlabel='age', ylabel='Count'>
```



In [ ]: