In [2]:
```
import pandas as pd
```

In [3]:
```
data=pd.read_csv("survey lung cancer.csv")
data
```

Out[3]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING
0	M	69	1	2	2	1	1	2	1	2	
1	M	74	2	1	1	1	2	2	2	1	
2	F	59	1	1	1	2	1	2	1	2	
3	M	63	2	2	2	1	1	1	1	1	
4	F	63	1	2	1	1	1	1	1	2	
...	...	...	...	...	...	...	...	...	...	...	...
304	F	56	1	1	1	2	2	2	2	1	
305	M	70	2	1	1	1	1	2	2	2	
306	M	58	2	1	1	1	1	2	2	2	
307	M	67	2	1	2	1	1	2	2	1	
308	M	62	1	1	1	2	1	2	2	2	

309 rows × 16 columns

In [4]:
```
data["AGE"]
```

Out[4]:
```
0      69
1      74
2      59
3      63
4      63
       ..
304    56
305    70
306    58
307    67
308    62
Name: AGE, Length: 309, dtype: int64
```

In [5]:
```
data["GENDER"]
```

Out[5]:
```
0      M
1      M
2      F
3      M
4      F
      ..
304    F
305    M
306    M
307    M
308    M
Name: GENDER, Length: 309, dtype: object
```

In [6]:
```
data.info
```

Out[6]:
```
<bound method DataFrame.info of     GENDER  AGE  SMOKING  YELLOW_FINGERS  ANXIETY  PEER_PRESSURE  \
0        M   69        1               2        2              1   
1        M   74        2               1        1              1   
2        F   59        1               1        1              2   
3        M   63        2               2        2              1   
4        F   63        1               2        1              1   
..     ...  ...      ...             ...      ...            ...   
304      F   56        1               1        1              2   
305      M   70        2               1        1              1   
306      M   58        2               1        1              1   
307      M   67        2               1        2              1   
308      M   62        1               1        1              2   

     CHRONIC DISEASE  FATIGUE   ALLERGY   WHEEZING  ALCOHOL CONSUMING  \
0                  1         2         1         2                  2   
1                  2         2         2         1                  1   
2                  1         2         1         2                  1   
3                  1         1         1         1                  2   
4                  1         1         1         2                  1   
..               ...       ...       ...       ...                ...   
304                2         2         2         2                  2   
305                1         2         2         2                  2   
306                1         1         2         2                  2   
307                1         2         2         1                  2   
308                1         2         2         2                  2   

     COUGHING  SHORTNESS OF BREATH  SWALLOWING DIFFICULTY  CHEST PAIN  \
0           2                    2                      2           2   
1           1                    2                      2           2   
2           2                    2                      1           2   
3           1                    1                      2           2   
4           2                    2                      1           1   
..        ...                  ...                    ...         ...   
304         2                    2                      2           1   
305         2                    2                      2           2   
306         2                    1                      1           2   
307         2                    2                      1           2   
308         1                    1                      2           1   

    LUNG_CANCER  
0           YES  
1           YES  
2            NO  
3            NO  
4            NO  
..          ...  
304         YES  
305         YES  
306         YES  
307         YES  
308         YES  

[309 rows x 16 columns]>
```

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

Out[7]:
```
(309, 16)
```

In [8]:
```
data.columns
```

Out[8]:
```
Index(['GENDER', 'AGE', 'SMOKING', 'YELLOW_FINGERS', 'ANXIETY',
       'PEER_PRESSURE', 'CHRONIC DISEASE', 'FATIGUE ', 'ALLERGY ', 'WHEEZING',
       'ALCOHOL CONSUMING', 'COUGHING', 'SHORTNESS OF BREATH',
       'SWALLOWING DIFFICULTY', 'CHEST PAIN', 'LUNG_CANCER'],
      dtype="object")
```

In [10]:
```
from matplotlib import pyplot as plt
```

In [11]:
```
data.isnull().sum()
```

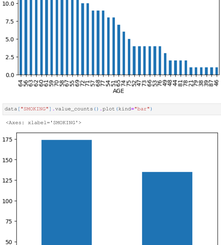
Out[11]:
```
GENDER                   0
AGE                      0
SMOKING                  0
YELLOW_FINGERS           0
ANXIETY                  0
PEER_PRESSURE            0
CHRONIC DISEASE          0
FATIGUE                  0
ALLERGY                  0
WHEEZING                 0
ALCOHOL CONSUMING        0
COUGHING                 0
SHORTNESS OF BREATH      0
SWALLOWING DIFFICULTY    0
CHEST PAIN               0
LUNG_CANCER              0
dtype: int64
```

In [14]:
```
data["GENDER"].value_counts().plot(kind="bar")
plt.title("Male percentage is higher compared to female percentage in smoking")
data
```

Out[14]:

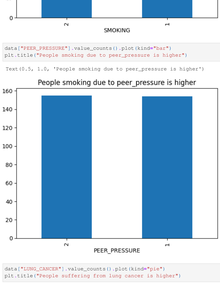
	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING
0	M	69	1	2	2	1	1	2	1	2	
1	M	74	2	1	1	1	2	2	2	1	
2	F	59	1	1	1	2	1	2	1	2	
3	M	63	2	2	2	1	1	1	1	1	
4	F	63	1	2	1	1	1	1	1	2	
...	...	...	...	...	...	...	...	...	...	...	...
304	F	56	1	1	1	2	2	2	2	1	
305	M	70	2	1	1	1	1	2	2	2	
306	M	58	2	1	1	1	1	2	2	2	
307	M	67	2	1	2	1	1	2	2	1	
308	M	62	1	1	1	2	1	2	2	2	

309 rows × 16 columns



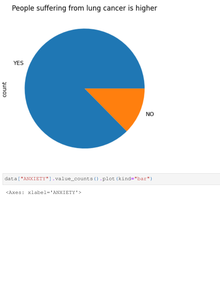
In [19]:
```
data["AGE"].value_counts().plot(kind="bar")
plt.title("Age graph")
```

Out[19]:
```
Text(0.5, 1.0, 'Age graph')
```



In [23]:
```
data["SMOKING"].value_counts().plot(kind="bar")
```

Out[23]:
```
<Axes: xlabel='SMOKING'>
```

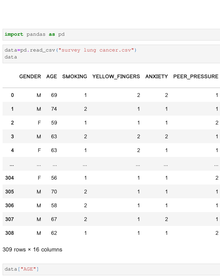


In [33]:
```
data["PEER_PRESSURE"].value_counts().plot(kind="bar")
plt.title("People smoking due to peer_pressure is higher")
```

Out[33]:
```
Text(0.5, 1.0, 'People smoking due to peer_pressure is higher')
```

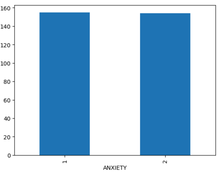

In [26]:
```
data["LUNG_CANCER"].value_counts().plot(kind="pie")
plt.title("People suffering from lung cancer is higher")
```

Out[26]:
```
Text(0.5, 1.0, 'People suffering from lung cancer is higher')
```



In [28]:
```
data["ANXIETY"].value_counts().plot(kind="bar")
```

Out[28]:
```
<Axes: xlabel='ANXIETY'>
```



In [ ]: