COVID_19

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv('Covid_19 Patient Dataset.csv')
df
```

```
    Patient_ID  Patient_Name  Age  Gender Infected_Location
Infection_Level  \
0         P001  Rajesh Kumar   45    Male           Chennai
Mild
1         P002  Priya Sharma   32  Female      Gummidipundi
Mild
2         P003    Amit Patel   58    Male            Mumbai
Moderate
3         P004    Neha Singh   28  Female             Delhi
Mild
4         P005  Vikram Reddy   52    Male         Bengaluru
Severe
..         ...           ...  ...     ...               ...
...
495       P496     Divya Das   36  Female           Chennai
Mild
496       P497    Ravi Patel   54    Male         Bengaluru
Mild
497       P498   Shalini Rao   27  Female             Delhi
Mild
498       P499  Nikhil Kumar   61    Male            Mumbai
Moderate
499       P500   Meera Singh   41  Female         Hyderabad
Mild

    Date_Diagnosed     Status Recovery_Date Contact_Traced
Infection_Source
0        3/15/2020  Recovered     3/27/2020            Yes    Family
Contact
1        3/16/2020  Recovered     3/28/2020            Yes
Workplace
2        3/17/2020     Active           NaN            Yes  Public
Transport
3        3/18/2020  Recovered     3/30/2020            Yes      Market
Visit
4        3/19/2020  Recovered     4/13/2020            Yes    Hospital
Visit
..             ...        ...           ...            ...
...
```

```
495      7/23/2021  Recovered      8/4/2021            Yes
Restaurant
496      7/24/2021  Recovered      8/5/2021            Yes
Workplace
497      7/25/2021  Recovered      8/6/2021            Yes      Market
Visit
498      7/26/2021     Active           NaN            Yes    Family
Contact
499      7/27/2021  Recovered      8/8/2021            Yes  Friend
Gathering

[500 rows x 11 columns]
```

```
print(df.head())
```

```
  Patient_ID  Patient_Name  Age  Gender Infected_Location
Infection_Level  \
0       P001  Rajesh Kumar   45    Male           Chennai
Mild
1       P002  Priya Sharma   32  Female      Gummidipundi
Mild
2       P003    Amit Patel   58    Male            Mumbai
Moderate
3       P004    Neha Singh   28  Female             Delhi
Mild
4       P005  Vikram Reddy   52    Male         Bengaluru
Severe

  Date_Diagnosed     Status Recovery_Date Contact_Traced
Infection_Source
0      3/15/2020  Recovered     3/27/2020            Yes    Family
Contact
1      3/16/2020  Recovered     3/28/2020            Yes
Workplace
2      3/17/2020     Active           NaN            Yes  Public
Transport
3      3/18/2020  Recovered     3/30/2020            Yes      Market
Visit
4      3/19/2020  Recovered     4/13/2020            Yes    Hospital
Visit
```

```
print(df.tail(8))
```

```
    Patient_ID  Patient_Name  Age  Gender Infected_Location
Infection_Level  \
492       P493    Vikram Rao   58    Male       Maharashtra
Mild
493       P494  Anjali Kumar   29  Female        Tamil Nadu
Mild
494       P495    Hari Singh   63    Male      Gummidipundi
```

```
Moderate
495       P496     Divya Das   36  Female           Chennai
Mild
496       P497    Ravi Patel   54    Male         Bengaluru
Mild
497       P498   Shalini Rao   27  Female             Delhi
Mild
498       P499  Nikhil Kumar   61    Male            Mumbai
Moderate
499       P500   Meera Singh   41  Female         Hyderabad
Mild

    Date_Diagnosed     Status Recovery_Date Contact_Traced  \
492      7/20/2021  Recovered      8/1/2021            Yes
493      7/21/2021  Recovered      8/2/2021            Yes
494      7/22/2021     Active           NaN            Yes
495      7/23/2021  Recovered      8/4/2021            Yes
496      7/24/2021  Recovered      8/5/2021            Yes
497      7/25/2021  Recovered      8/6/2021            Yes
498      7/26/2021     Active           NaN            Yes
499      7/27/2021  Recovered      8/8/2021            Yes

        Infection_Source
492     Public Transport
493  Religious Gathering
494       Hospital Visit
495           Restaurant
496            Workplace
497         Market Visit
498       Family Contact
499     Friend Gathering

df.columns

Index(['Patient_ID', 'Patient_Name', 'Age', 'Gender',
'Infected_Location',
       'Infection_Level', 'Date_Diagnosed', 'Status', 'Recovery_Date',
       'Contact_Traced', 'Infection_Source'],
      dtype='object')

df.dtypes

Patient_ID           object
Patient_Name         object
Age                   int64
Gender               object
Infected_Location    object
Infection_Level      object
Date_Diagnosed       object
Status               object
```

```
Recovery_Date        object
Contact_Traced       object
Infection_Source     object
dtype: object
```

```
df.info
```

```
<bound method DataFrame.info of     Patient_ID  Patient_Name  Age
Gender Infected_Location Infection_Level  \
0         P001  Rajesh Kumar   45    Male           Chennai
Mild
1         P002  Priya Sharma   32  Female      Gummidipundi
Mild
2         P003    Amit Patel   58    Male            Mumbai
Moderate
3         P004    Neha Singh   28  Female             Delhi
Mild
4         P005  Vikram Reddy   52    Male         Bengaluru
Severe
..         ...           ...  ...     ...               ...
...
495       P496     Divya Das   36  Female           Chennai
Mild
496       P497    Ravi Patel   54    Male         Bengaluru
Mild
497       P498   Shalini Rao   27  Female             Delhi
Mild
498       P499  Nikhil Kumar   61    Male            Mumbai
Moderate
499       P500   Meera Singh   41  Female         Hyderabad
Mild

    Date_Diagnosed     Status Recovery_Date Contact_Traced
Infection_Source
0        3/15/2020  Recovered     3/27/2020            Yes    Family
Contact
1        3/16/2020  Recovered     3/28/2020            Yes
Workplace
2        3/17/2020     Active           NaN            Yes  Public
Transport
3        3/18/2020  Recovered     3/30/2020            Yes      Market
Visit
4        3/19/2020  Recovered     4/13/2020            Yes    Hospital
Visit
..             ...        ...           ...            ...
...
495      7/23/2021  Recovered      8/4/2021            Yes
Restaurant
496      7/24/2021  Recovered      8/5/2021            Yes
Workplace
```

```
497      7/25/2021  Recovered      8/6/2021             Yes      Market
Visit
498      7/26/2021     Active           NaN             Yes    Family
Contact
499      7/27/2021  Recovered      8/8/2021             Yes  Friend
Gathering

[500 rows x 11 columns]>
```

```
df.nunique()
```

```
Patient_ID           500
Patient_Name         175
Age                   46
Gender                 2
Infected_Location     13
Infection_Level        3
Date_Diagnosed       500
Status                 2
Recovery_Date        373
Contact_Traced         1
Infection_Source       8
dtype: int64
```

```
df.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
       ...
495    False
496    False
497    False
498    False
499    False
Length: 500, dtype: bool
```

```
df.isnull()
```

```
     Patient_ID  Patient_Name    Age  Gender  Infected_Location  \
0         False         False  False   False              False
1         False         False  False   False              False
2         False         False  False   False              False
3         False         False  False   False              False
4         False         False  False   False              False
..          ...           ...    ...     ...                ...
495       False         False  False   False              False
496       False         False  False   False              False
497       False         False  False   False              False
```

```
498       False          False  False   False              False
499       False          False  False   False              False

     Infection_Level  Date_Diagnosed  Status  Recovery_Date
Contact_Traced  \
0              False           False   False          False
False
1              False           False   False          False
False
2              False           False   False           True
False
3              False           False   False          False
False
4              False           False   False          False
False
..               ...             ...     ...            ...
...
495            False           False   False          False
False
496            False           False   False          False
False
497            False           False   False          False
False
498            False           False   False           True
False
499            False           False   False          False
False

     Infection_Source
0               False
1               False
2               False
3               False
4               False
..                ...
495             False
496             False
497             False
498             False
499             False

[500 rows x 11 columns]

print(df.isnull().sum())

Patient_ID             0
Patient_Name           0
Age                    0
Gender                 0
Infected_Location      0
```

```
Infection_Level        0
Date_Diagnosed         0
Status                 0
Recovery_Date        126
Contact_Traced         0
Infection_Source       0
dtype: int64
```

```
df["Recovery_Date"]
```

```
0      3/27/2020
1      3/28/2020
2            NaN
3      3/30/2020
4      4/13/2020
         ...
495     8/4/2021
496     8/5/2021
497     8/6/2021
498          NaN
499     8/8/2021
Name: Recovery_Date, Length: 500, dtype: object
```

```
df["Recovery_Date"].fillna("8/10/2021",inplace=True)
df
```

```
    Patient_ID  Patient_Name  Age  Gender Infected_Location
Infection_Level  \
0         P001  Rajesh Kumar   45    Male           Chennai
Mild
1         P002  Priya Sharma   32  Female      Gummidipundi
Mild
2         P003    Amit Patel   58    Male            Mumbai
Moderate
3         P004    Neha Singh   28  Female             Delhi
Mild
4         P005  Vikram Reddy   52    Male         Bengaluru
Severe
..         ...           ...  ...     ...               ...
...
495       P496     Divya Das   36  Female           Chennai
Mild
496       P497    Ravi Patel   54    Male         Bengaluru
Mild
497       P498   Shalini Rao   27  Female             Delhi
Mild
498       P499  Nikhil Kumar   61    Male            Mumbai
Moderate
499       P500   Meera Singh   41  Female         Hyderabad
Mild
```

```
    Date_Diagnosed     Status Recovery_Date Contact_Traced
Infection_Source
0        3/15/2020  Recovered     3/27/2020            Yes    Family
Contact
1        3/16/2020  Recovered     3/28/2020            Yes
Workplace
2        3/17/2020     Active     8/10/2021            Yes  Public
Transport
3        3/18/2020  Recovered     3/30/2020            Yes      Market
Visit
4        3/19/2020  Recovered     4/13/2020            Yes    Hospital
Visit
..             ...        ...           ...            ...
...
495      7/23/2021  Recovered      8/4/2021            Yes
Restaurant
496      7/24/2021  Recovered      8/5/2021            Yes
Workplace
497      7/25/2021  Recovered      8/6/2021            Yes      Market
Visit
498      7/26/2021     Active     8/10/2021            Yes    Family
Contact
499      7/27/2021  Recovered      8/8/2021            Yes  Friend
Gathering

[500 rows x 11 columns]
```

```
grouped_by=df.groupby("Age")
grouped_count=grouped_by["Age"].count()
grouped_count
```

```
Age
23     1
24     2
25    10
26    17
27    15
28    19
29    18
30    12
31    15
32    12
33     3
34     3
35     4
36    19
37    20
38    22
39    16
```

```
40    14
41    11
42    10
43     4
44     4
45     2
46     2
47     2
48     3
49     3
50     3
51    12
52    17
53     9
54    17
55    17
56    16
57    10
58    19
59    10
60    16
61    21
62    18
63    17
64    15
65     8
66     1
67    10
68     1
Name: Age, dtype: int64
```

```
grouped_by=df.groupby("Infection_Source")
grouped_count=grouped_by["Infection_Source"].count()
grouped_count
```

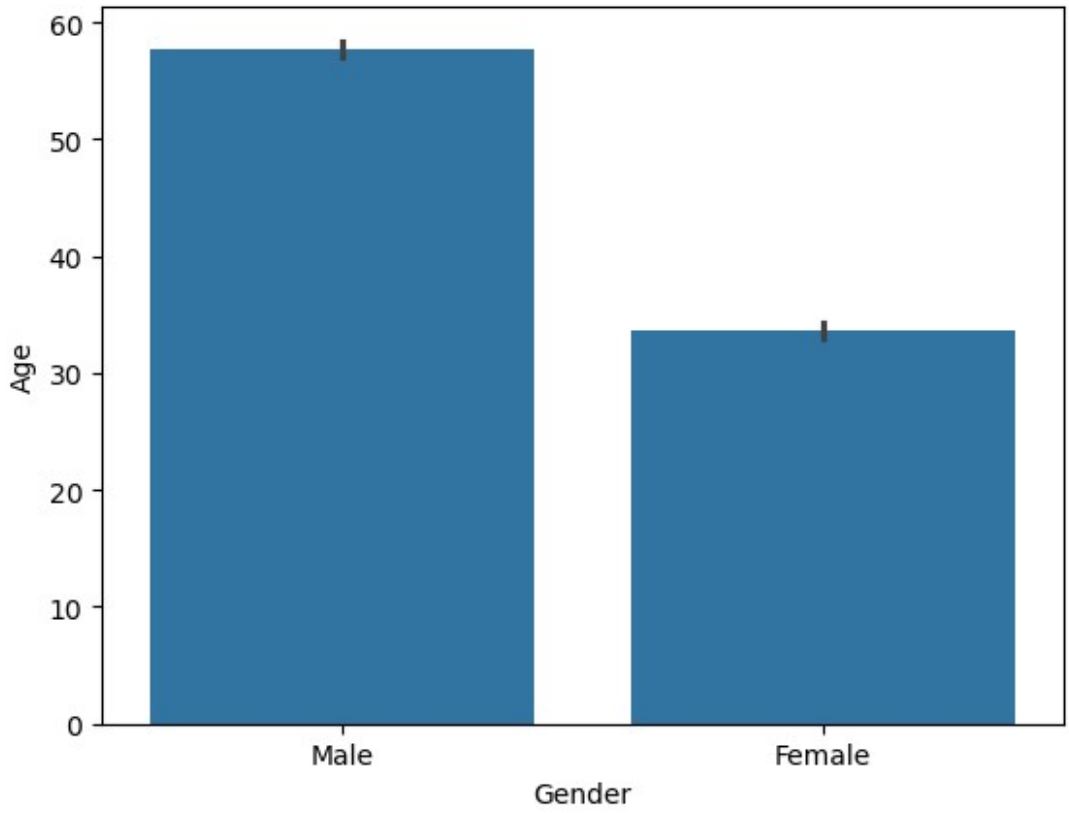
```
Infection_Source
Family Contact         64
Friend Gathering       63
Hospital Visit         61
Market Visit           63
Public Transport       63
Religious Gathering    61
Restaurant             62
Workplace              63
Name: Infection_Source, dtype: int64
```

```
grouped_by=df.groupby("Infected_Location")
grouped_count=grouped_by["Infected_Location"].count()
grouped_count
```

```
Infected_Location
Bengaluru        39
Chennai          39
Delhi            39
Gujarat          38
Gummidipundi     39
Hyderabad        39
Kerala           38
Kolkata          38
Maharashtra      38
Mumbai           39
Pune             38
Tamil Nadu       38
Uttar Pradesh    38
Name: Infected_Location, dtype: int64
```

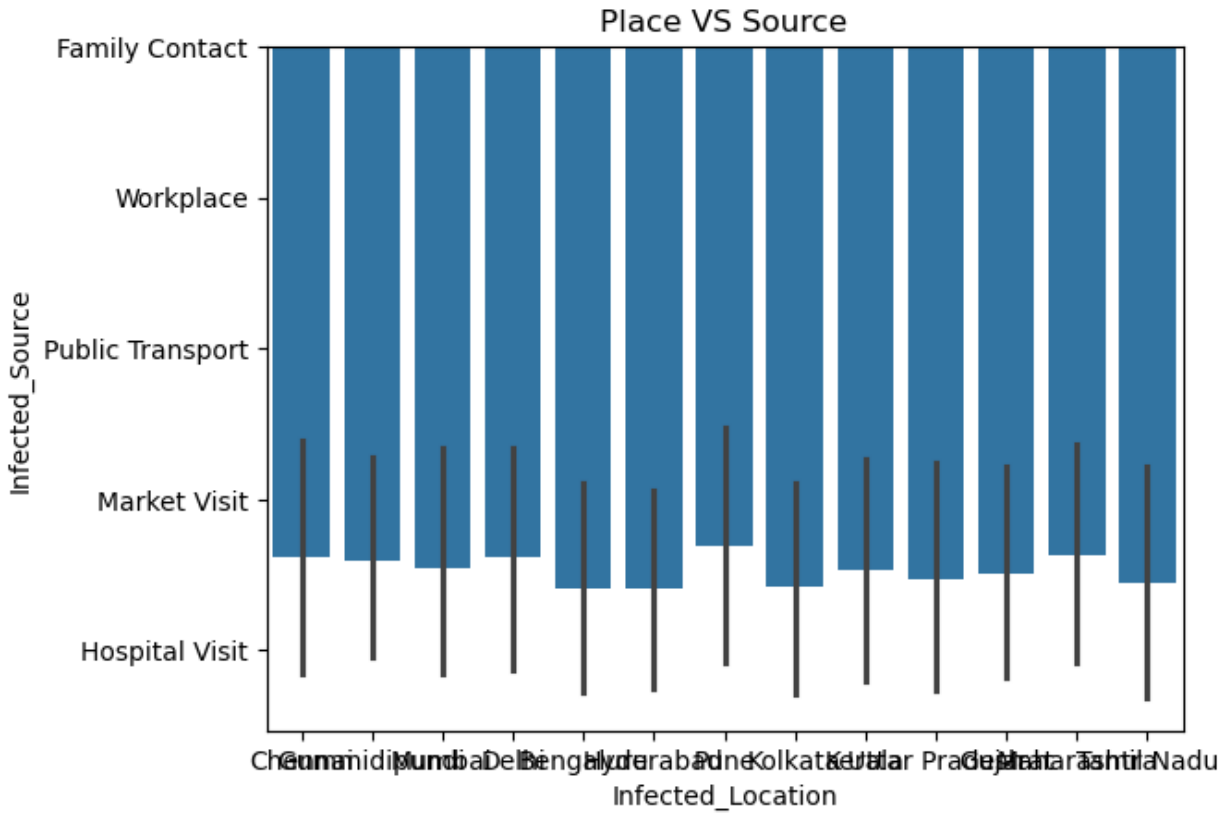
```
sns.barplot(y='Age', x='Gender',data=df)
```

```
<Axes: xlabel='Gender', ylabel='Age'>
```



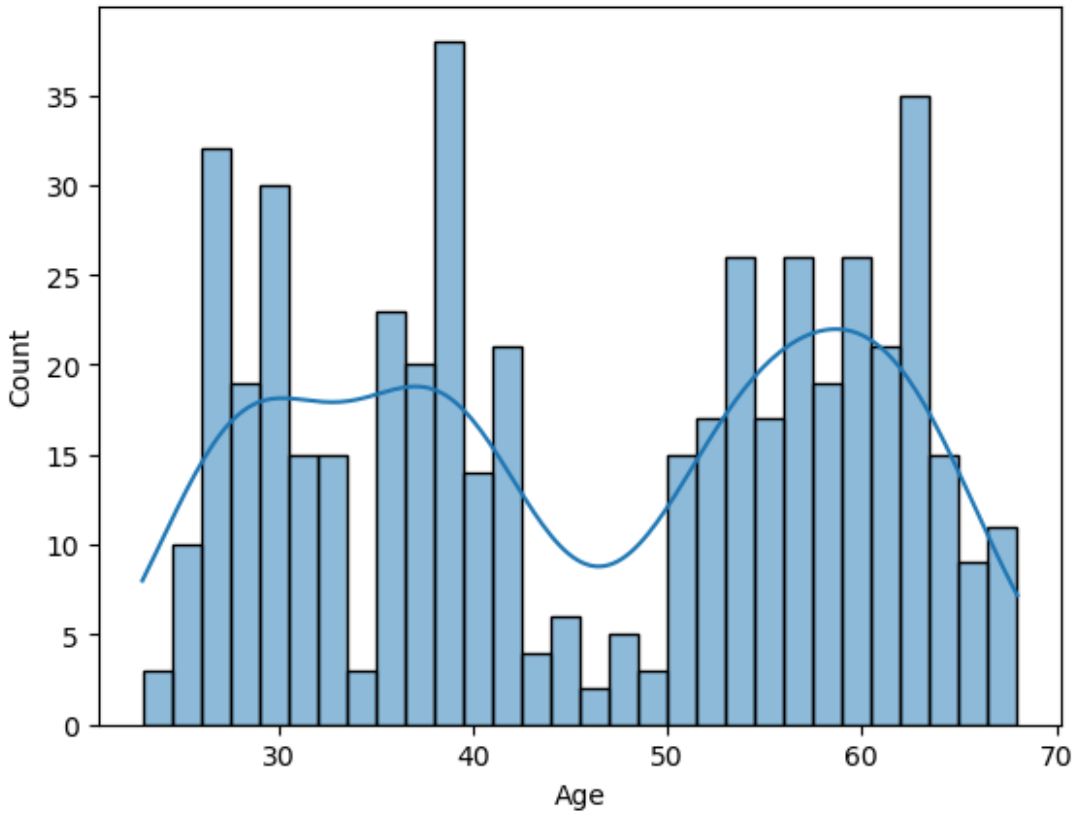

```
sns.barplot(x='Infected_Location',y='Infection_Source',data=df)
plt.title("Place VS Source")
plt.xlabel("Infected_Location")
plt.ylabel("Infected_Source")
```

```
Text(0, 0.5, 'Infected_Source')
```



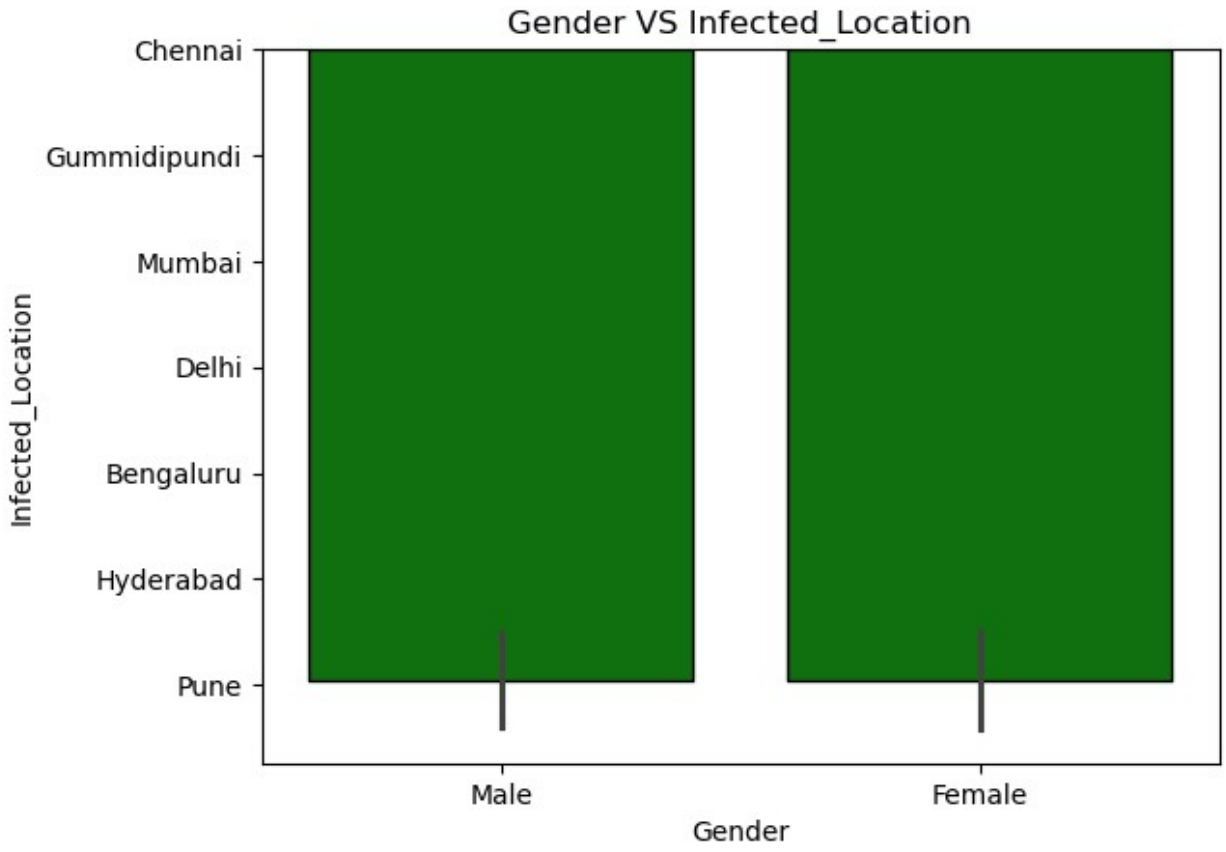

```
sns.histplot(df['Age'],bins=30,kde=True)
```

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




```
sns.barplot(x='Gender',y='Infected_Location',color='Green',edgecolor='
Black',data=df)
plt.title("Gender VS Infected_Location")
plt.xlabel("Gender")
plt.ylabel("Infected_Location")
```

```
Text(0, 0.5, 'Infected_Location')
```


Gender VS Infected_Location
Chennai
Gummidipundi
Mumbai
Delhi
Bengaluru
Hyderabad
Pune
Infected_Location
Male
Female
Gender