

# DATA SCIENCE PROJECT ANALYZING COVID-19 karnataka DATASET

In [6]:

```
!pip install pandas
!pip install numpy
!pip install seaborn
```

```
Requirement already satisfied: pandas in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (1.3.0)
Requirement already satisfied: numpy>=1.17.3 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from pandas) (1.20.3)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from pandas) (2.8.1)
Requirement already satisfied: pytz>=2017.3 in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (from pandas) (2021.1)
Requirement already satisfied: six>=1.5 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
Requirement already satisfied: numpy in c:\users\invest 360\appdata\roaming\python\python39\site-packages (1.20.3)
Requirement already satisfied: seaborn in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (0.11.1)
Requirement already satisfied: numpy>=1.15 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from seaborn) (1.20.3)
Requirement already satisfied: matplotlib>=2.2 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from seaborn) (3.4.2)
Requirement already satisfied: pandas>=0.23 in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (from seaborn) (1.3.0)
Requirement already satisfied: scipy>=1.0 in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (from seaborn) (1.7.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from matplotlib>=2.2->seaborn) (1.3.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (from matplotlib>=2.2->seaborn) (8.2.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from matplotlib>=2.2->seaborn) (2.4.7)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from matplotlib>=2.2->seaborn) (2.8.1)
```

Requirement already satisfied: cyclr>=0.10 in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from matplotlib>=2.2->seaborn) (0.10.0)  
Requirement already satisfied: six in c:\users\invest 360\appdata\roaming\python\python39\site-packages (from cyclr>=0.10->matplotlib>=2.2->seaborn) (1.16.0)  
Requirement already satisfied: pytz>=2017.3 in c:\users\invest 360\appdata\local\programs\python\python39-32\lib\site-packages (from pandas>=0.23->seaborn) (2021.1)

```
In [7]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from pandas import read_csv
```

```
In [8]: data = pd.read_csv('Karnataka_data_covid-19.csv')
```

```
In [9]: data.head()
```

```
Out[9]:
```

	Daily samples collected for Testing	Daily samples reported as negative	Daily Samples Positive for COVID-19	People in Observation	KIA-Passengers	Mangalore-Passengers	Seaport M and K-passengers
0	579	460	6	NaN	72726.0	27172.0	5394.0
1	60	80	0	1345.0	2728.0	615.0	NaN
2	92	50	0	1657.0	3004.0	791.0	5439.0
3	131	19	0	1877.0	2376.0	358.0	5458.0
4	33	86	1	2221.0	2025.0	748.0	5505.0

```
In [10]: data.columns
```

```
Out[10]: Index(['Daily samples collected for Testing',
'Daily samples reported as negative',
'Daily Samples Positive for COVID-19', 'People in Observation',
'KIA-Passengers', 'Mangalore-Passengers',
'Seaport M and K- passengers'],
dtype='object')
```

```
In [11]: data.tail()
```

```
Out[11]:
```

	Daily samples collected for Testing	Daily samples reported as negative	Daily Samples Positive for COVID-19	People in Observation	KIA-Passengers	Mangalore-Passengers	Seaport M and K-passengers
112	16210	14470	1502	44950.0	0.0	NaN	NaN

<b>113</b>	18307	16290	1696	82082.0	0.0	NaN	NaN
<b>114</b>	17592	15294	1839	56341.0	0.0	NaN	NaN
<b>115</b>	16899	14649	1925	53803.0	0.0	NaN	NaN
<b>116</b>	15880	13742	1843	56927.0	0.0	NaN	NaN

In [12]: `data.describe()`

Out[12]:

	Daily samples collected for Testing	Daily samples reported as negative	Daily Samples Positive for COVID-19	People in Observation	KIA-Passengers	Mangalore-Passengers	Seaport M and K-passengers
<b>count</b>	117.000000	117.000000	117.000000	116.000000	57.000000	11.000000	12.000000
<b>mean</b>	6207.897436	5733.068376	216.393162	26875.879310	1794.929825	2957.454545	5186.416667
<b>std</b>	5340.258431	4978.231949	408.136386	13363.960008	9593.477499	8032.555687	1630.926422
<b>min</b>	10.000000	19.000000	0.000000	1345.000000	0.000000	271.000000	46.000000
<b>25%</b>	585.000000	558.000000	11.000000	19217.250000	0.000000	441.000000	5453.250000
<b>50%</b>	4892.000000	4717.000000	36.000000	25549.500000	231.000000	566.000000	5587.500000
<b>75%</b>	10177.000000	9813.000000	239.000000	35611.750000	531.000000	681.500000	5775.750000
<b>max</b>	18307.000000	16290.000000	1925.000000	82082.000000	72726.000000	27172.000000	6022.000000

In [13]: `data.columns`

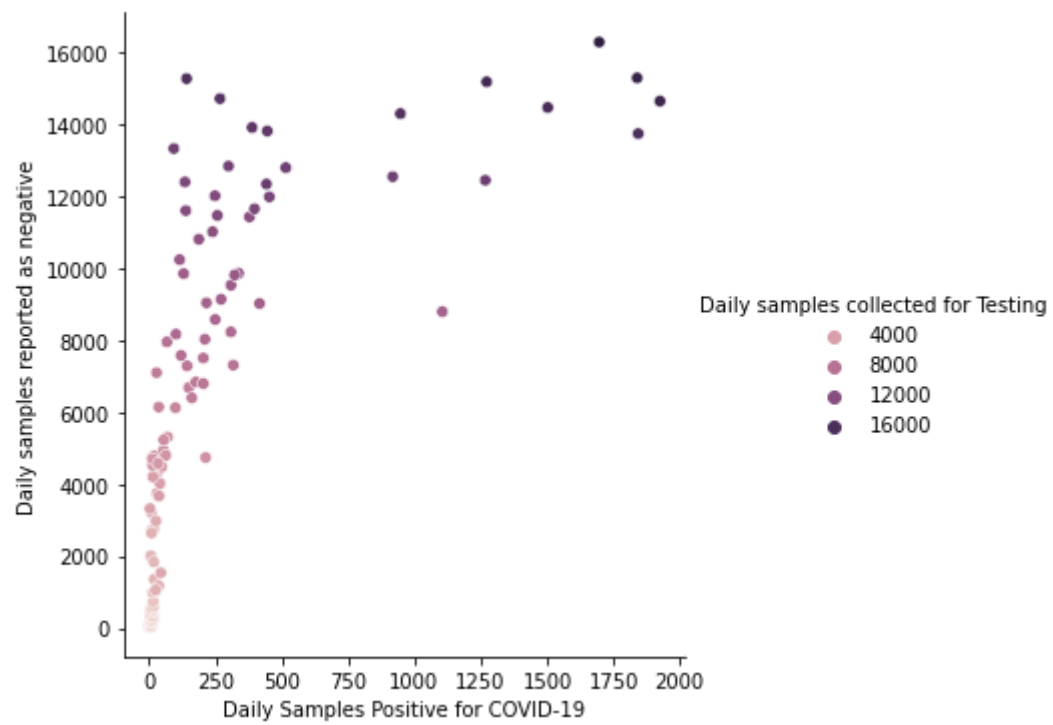
Out[13]: Index(['Daily samples collected for Testing',  
'Daily samples reported as negative',  
'Daily Samples Positive for COVID-19', 'People in Observation',  
'KIA-Passengers', 'Mangalore-Passengers',  
'Seaport M and K- passengers'],  
dtype='object')

`data.isnull().sum()`

## Relating the variable with scatterplots

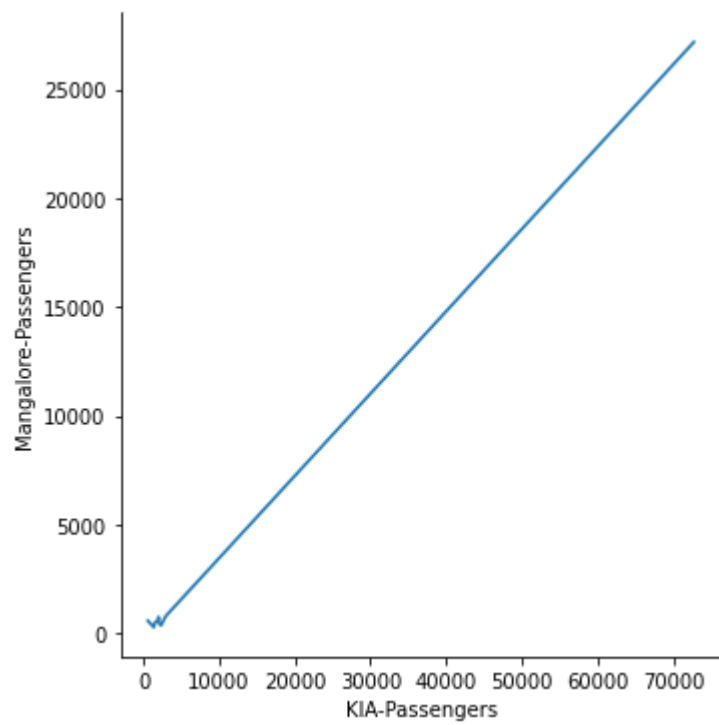
In [14]: `sns.relplot(x="Daily Samples Positive for COVID-19" , y="Daily samples reported as negative" ,hue='Daily samples collected for`

Out[14]: `<seaborn.axisgrid.FacetGrid at 0x5854d18>`



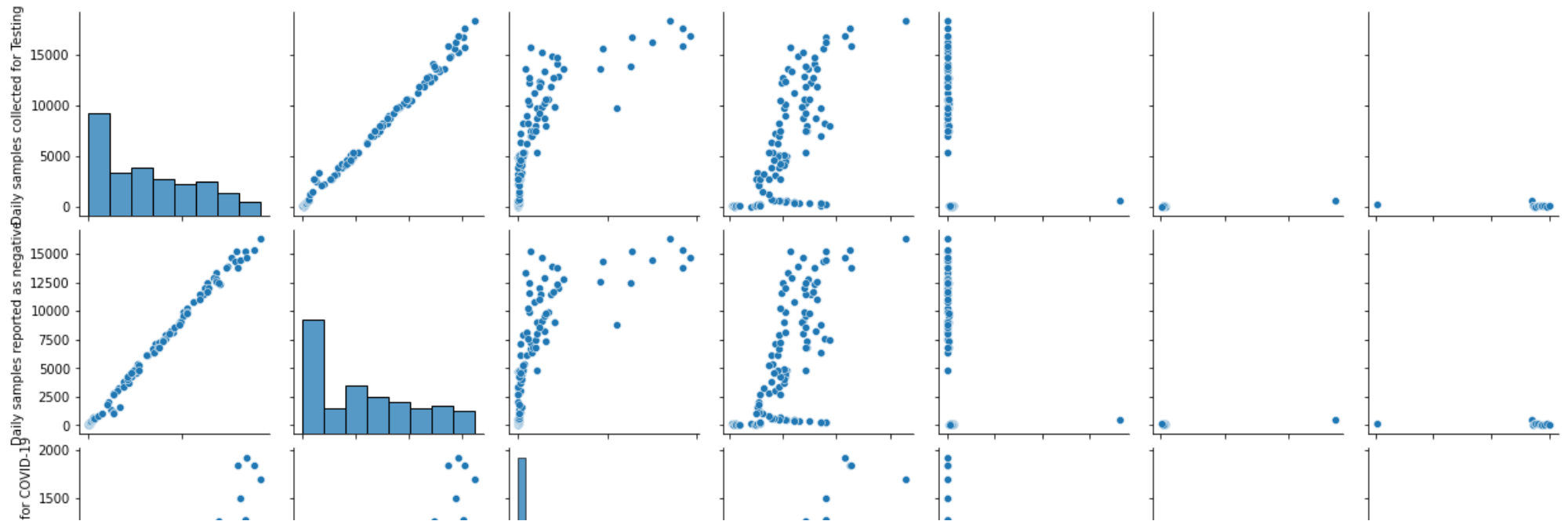
```
In [15]: sns.relplot(x="KIA-Passengers" , y="Mangalore-Passengers",kind='line' , data=data)
```

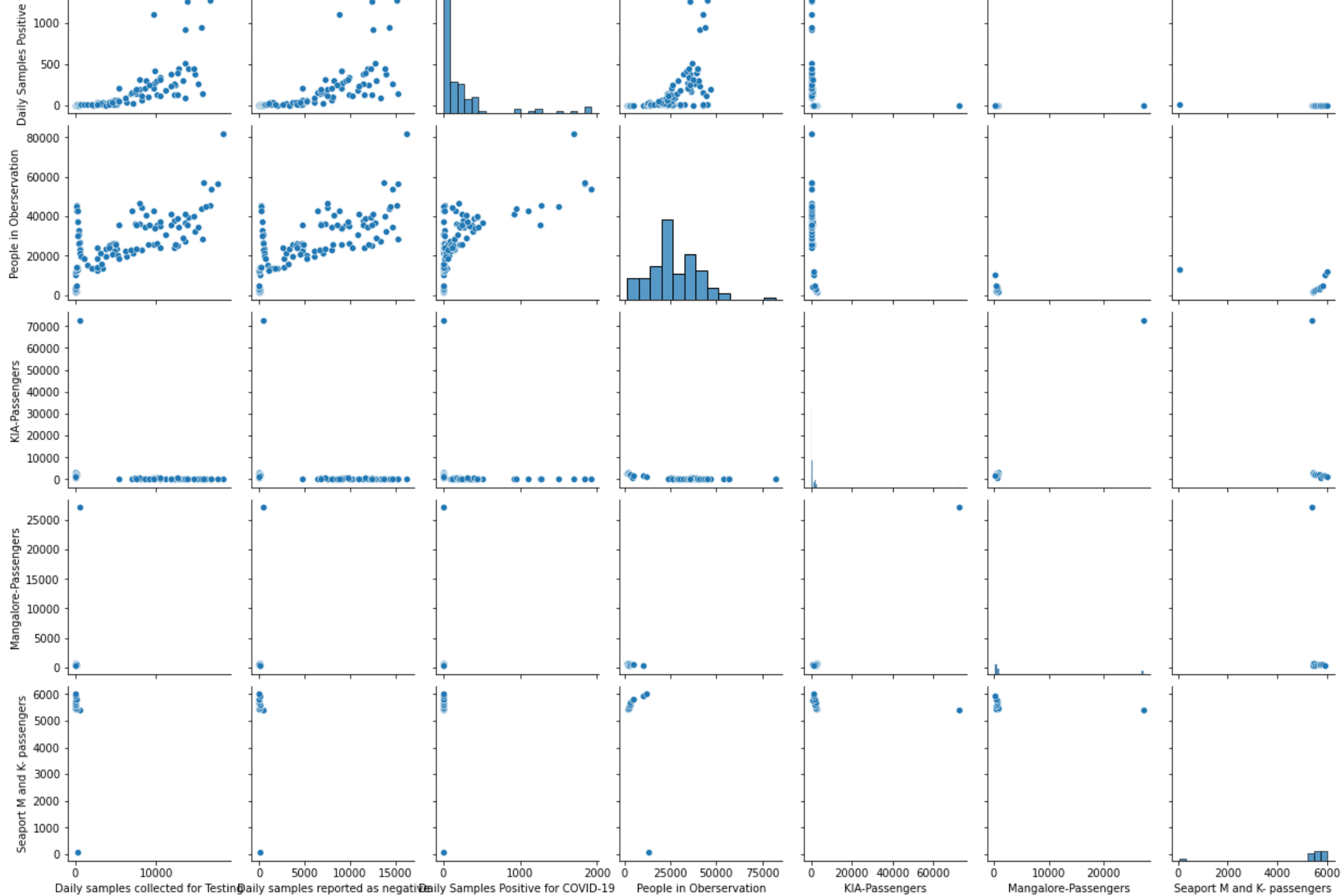
```
Out[15]: <seaborn.axisgrid.FacetGrid at 0x1403c0b8>
```



```
In [16]: sns.pairplot(data)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x143749b8>
```





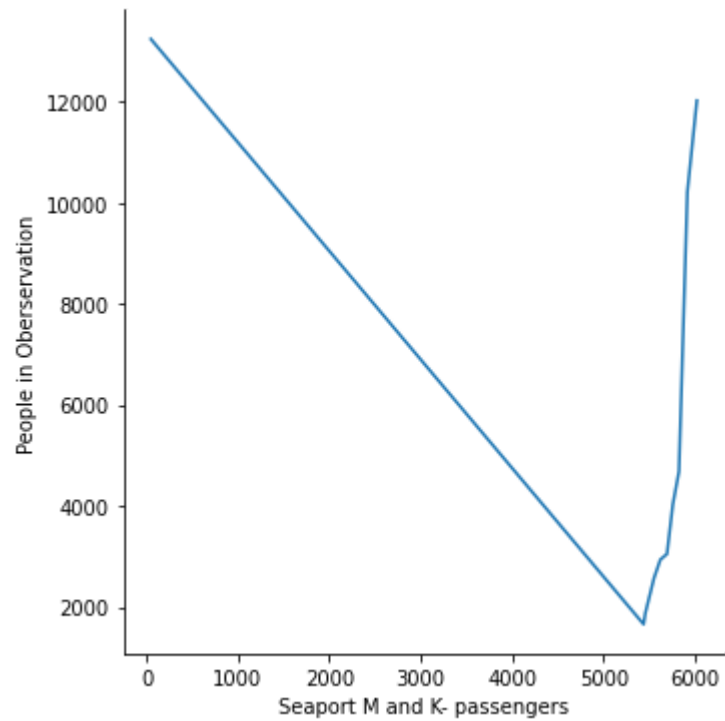
```
In [17]: data.columns
```

```
Out[17]: Index(['Daily samples collected for Testing',
               'Daily samples reported as negative',
               'Daily Samples Positive for COVID-19', 'People in Observation',
```

```
'KIA-Passengers', 'Mangalore-Passengers',  
'Seaport M and K- passengers'],  
dtype='object')
```

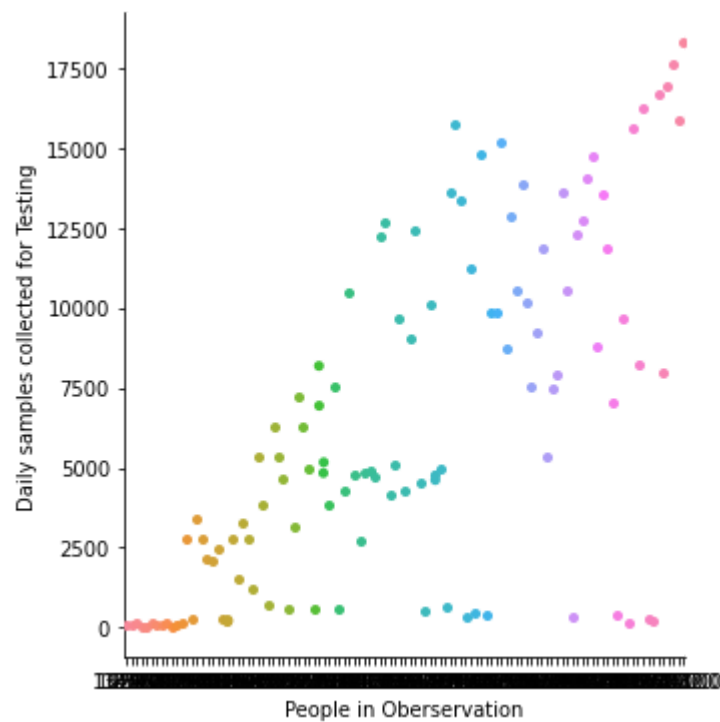
```
In [18]: sns.relplot(x="Seaport M and K- passengers" , y="People in Observation" ,kind ='line', data=data)
```

```
Out[18]: <seaborn.axisgrid.FacetGrid at 0x154bed48>
```



```
In [19]: sns.catplot(x="People in Observation",y="Daily samples collected for Testing",data=data)
```

```
Out[19]: <seaborn.axisgrid.FacetGrid at 0x15501550>
```



```
In [20]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from pandas import read_csv
```

```
In [21]: data = pd.read_csv('Karnataka_covid-19.csv')
```

```
In [22]: data.head()
```

```
Out[22]:
```

	Case	Date	Age	Sex	City	State	Cluster	Reason	Nationality	RD	P	C	RE
0	1	09-Mar	41.0	Male	Bangalore-Urban	Karnataka	From USA	Texas US	India	C	0	1	NaN
1	2	10-Mar	0.0	Female	Bangalore-Urban	Karnataka	From USA	Spouse	India	C	1	0	NaN
2	3	10-Mar	13.0	Female	Bangalore-Urban	Karnataka	From USA	Daughter	India	C	1	0	NaN
3	4	NaN	0.0	NaN	Bangalore-Urban	Karnataka	From United Kingdom	London	India	C	0	1	NaN
4	5	13-Mar	26.0	Male	Bangalore-Urban	Karnataka	From the rest of Europe	Greece	India	C	0	0	NaN



In [23]: data.tail()

Out[23]:

	Case	Date	Age	Sex	City	State		Cluster	Reason	Nationality	RD	P	C	RE	
	274	275	15-Apr	38.0	Female	Vijayapura	Karnataka		Severe Acute Respiratory Infection	NaN	India	C	221	0	NaN
	275	276	15-Apr	25.0	Male	Vijayapura	Karnataka		Severe Acute Respiratory Infection	NaN	India	C	221	0	NaN
	276	277	15-Apr	32.0	Female	Bangalore-Urban	Karnataka		Severe Acute Respiratory Infection	NaN	India	C	252	0	NaN
	277	278	15-Apr	28.0	Female	Vijayapura	Karnataka		Severe Acute Respiratory Infection	NaN	India	C	221	0	NaN
	278	279	15-Apr	80.0	Female	Belgavi	Karnataka	TJ Congregation from 13th to 18th March in Delhi		NaN	India	D	224	0	NaN

In [24]: data.describe()

Out[24]:

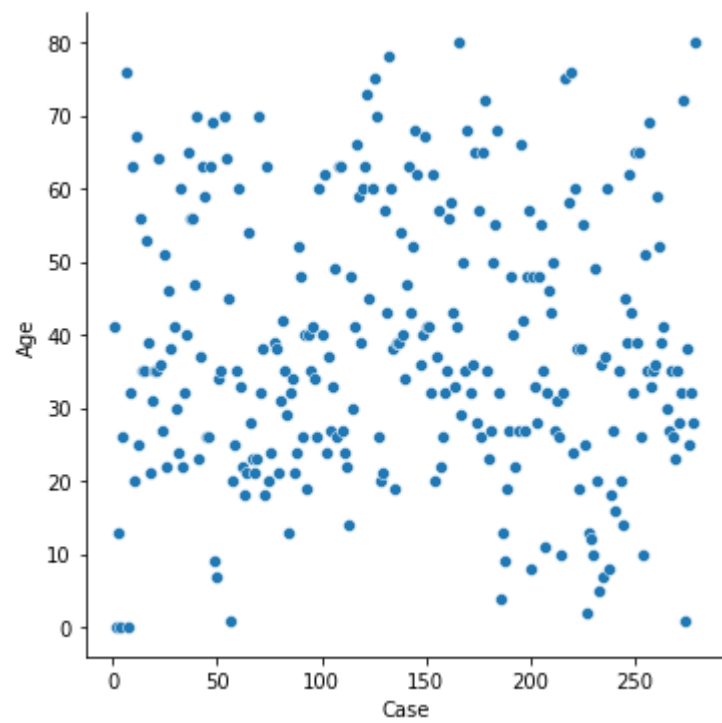
	Case	Age	P	C
count	279.000000	279.000000	279.000000	279.000000
mean	140.000000	38.354122	45.329749	0.272401
std	80.684571	18.154102	70.055468	0.445995
min	1.000000	0.000000	0.000000	0.000000
25%	70.500000	26.000000	0.000000	0.000000
50%	140.000000	35.000000	0.000000	0.000000
75%	209.500000	52.000000	88.000000	1.000000
max	279.000000	80.000000	252.000000	1.000000

In [29]: data.columns

Out[29]: Index(['Case', 'Date', 'Age', 'Sex', 'City', 'State', 'Cluster', 'Reason', 'Nationality', 'RD', 'P', 'C', 'RE'], dtype='object')

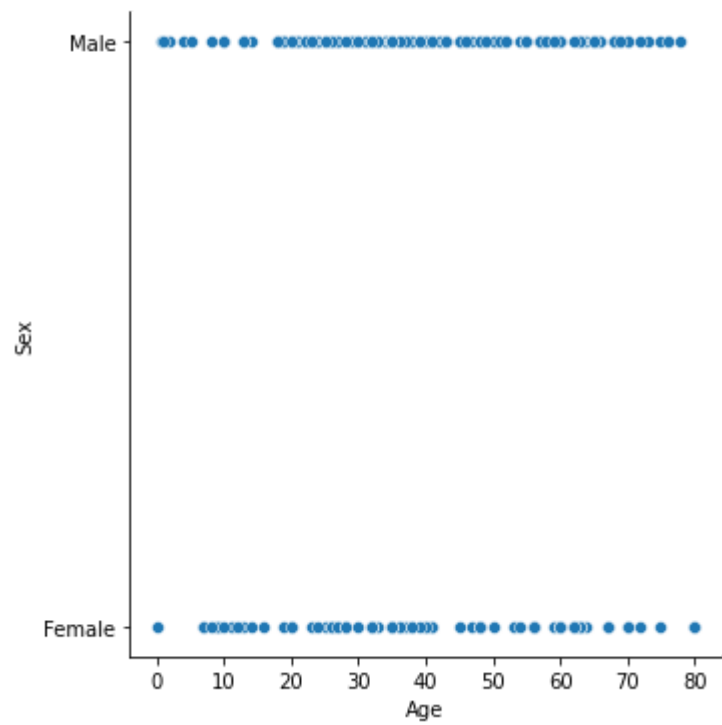
In [25]: sns.relplot(x="Case" , y="Age", data=data)

Out[25]: <seaborn.axisgrid.FacetGrid at 0x1551fd48>



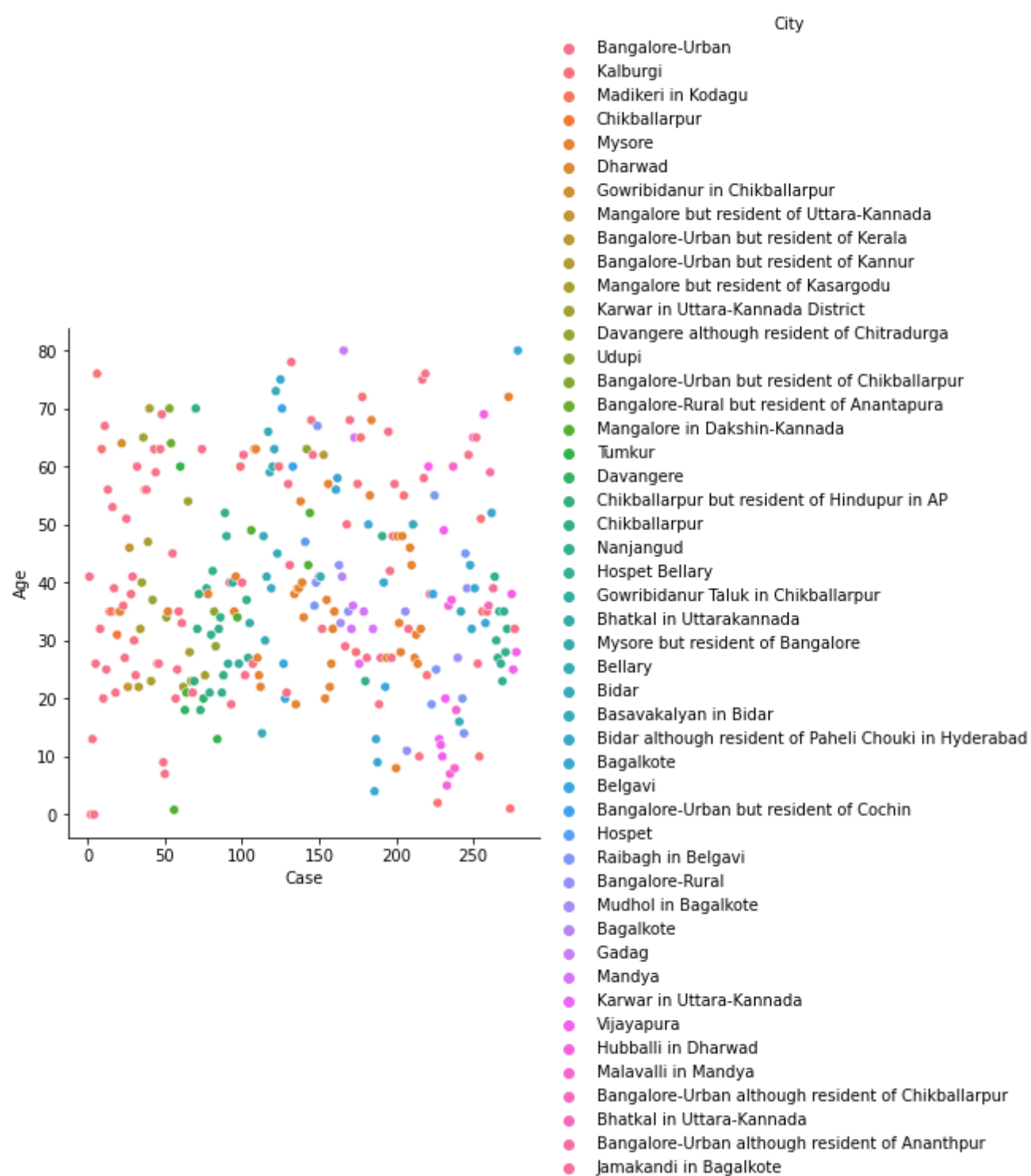
```
In [26]: sns.relplot(x="Age" , y="Sex", data=data)
```

```
Out[26]: <seaborn.axisgrid.FacetGrid at 0x16067238>
```



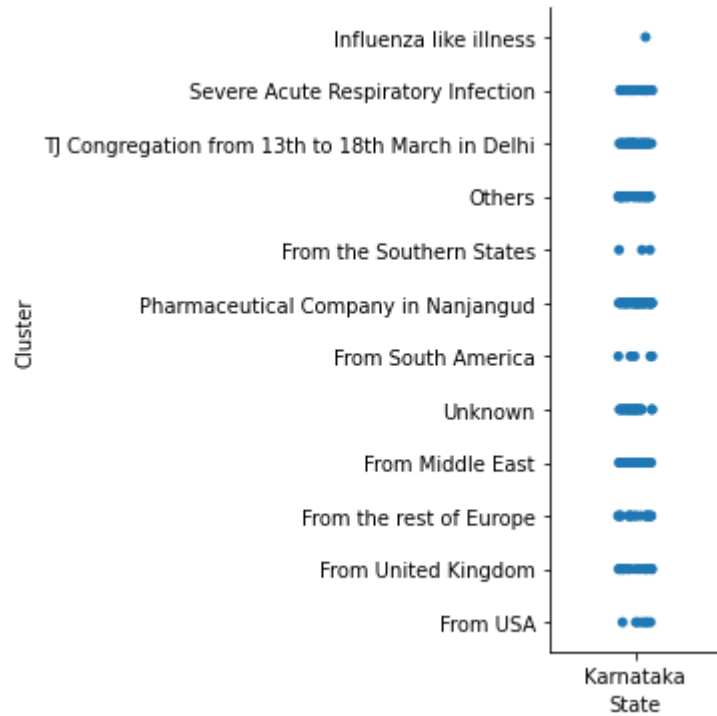
```
In [27]: sns.relplot(x="Case" , y="Age",hue='City' ,data=data)
```

```
Out[27]: <seaborn.axisgrid.FacetGrid at 0x1609ff88>
```



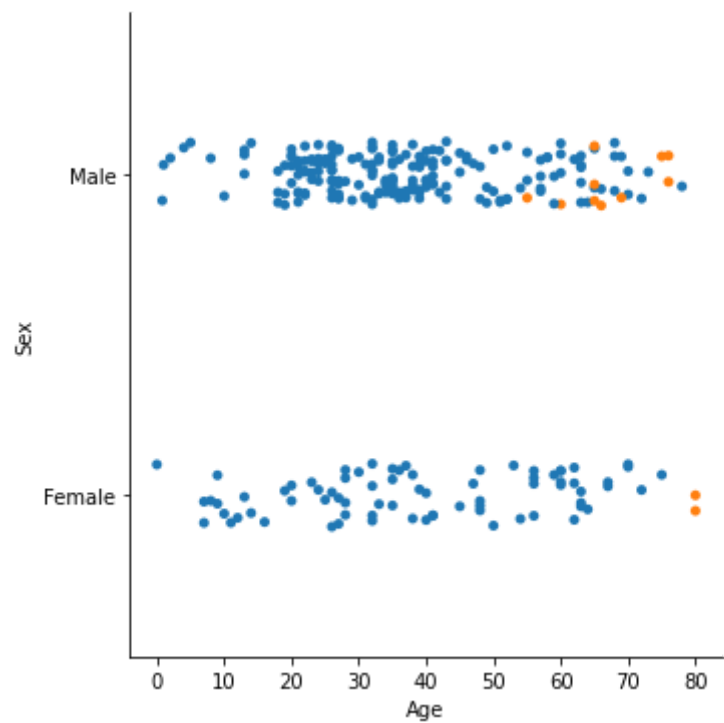
```
sns.catplot(x="State", y="Cluster", data=data)
```

Out[28]: <seaborn.axisgrid.FacetGrid at 0x154e8418>



```
In [30]: sns.catplot(x="Age", y="Sex", hue='RD', data=data)
```

Out[30]: <seaborn.axisgrid.FacetGrid at 0x12f769a0>



In [ ]:

In [ ]:

In [ ]: