

```
In [1]: # import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

from plotly.offline import init_notebook_mode
init_notebook_mode(connected = True)

import warnings
warnings.filterwarnings("ignore")
```

```
In [4]: #reading the dataset to a dataframe
data = pd.read_excel(r"C:\Users\Chinenye Claire\Desktop\Applied Epidemiology and Public Health\Written Assignment 1\City of Edinb
data.head()
```

Out[4]:

	Date	Location	Daily positive cases	Daily deaths	Daily positive tests	Daily Negative Tests	TotalTests	NHS (pillar 1) positive tests	NHS (Pillar 1) total tests	UK Gov (Pillar 2) positive tests	...	Lateral flow positive tests	TotalILFD	Unnamed: 13	Unnamed: 14	Unnamed: 15	Unnam
0	2020-03-04	City of Edinburgh	1	0	1	24	25	1.0	25	0.0	...	NaN	NaN	Total Number of Positive Cases	NaN	Average Positive Case per day	N
1	2020-03-05	City of Edinburgh	0	0	1	22	23	1.0	23	0.0	...	NaN	NaN	183951	NaN	224.604396	N
2	2020-03-06	City of Edinburgh	1	0	2	21	23	2.0	23	0.0	...	NaN	NaN	NaN	NaN	NaN	N
3	2020-03-07	City of Edinburgh	2	0	3	24	27	3.0	27	0.0	...	NaN	NaN	Total Tests	NaN	NaN	N
4	2020-03-08	City of Edinburgh	2	0	2	16	18	2.0	18	0.0	...	NaN	NaN	2625595	NaN	NaN	N

5 rows × 21 columns

```
In [5]: len(data)
```

Out[5]: 819

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 819 entries, 0 to 818
Data columns (total 21 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Date                                     819 non-null    datetime64[ns]
1   Location                                819 non-null    object
2   Daily positive cases                    819 non-null    int64
3   Daily deaths                            819 non-null    int64
4   Daily positive tests                    819 non-null    int64
5   Daily Negative Tests                    819 non-null    int64
6   TotalTests                             819 non-null    int64
7   NHS (pillar 1 ) positive tests          747 non-null    float64
8   NHS (Pillar 1) total tests              819 non-null    int64
9   UK Gov (Pillar 2) positive tests        747 non-null    float64
10  UK Gov (Pillar 2) total tests            819 non-null    int64
11  Lateral flow positive tests              147 non-null    float64
12  TotalILFD                               147 non-null    float64
13  Unnamed: 13                             11 non-null    object
14  Unnamed: 14                             0 non-null     float64
15  Unnamed: 15                             8 non-null     object
16  Unnamed: 16                             0 non-null     float64
17  Unnamed: 17                             4 non-null     object
18  Unnamed: 18                             6 non-null     object
19  Unnamed: 19                             4 non-null     object
20  Unnamed: 20                             4 non-null     object
dtypes: datetime64[ns](1), float64(6), int64(7), object(7)
memory usage: 134.5+ KB
```

In [7]: data.sample(10)

Out[7]:

	Date	Location	Daily positive cases	Daily deaths	Daily positive tests	Daily Negative Tests	TotalTests	NHS (pillar 1) positive tests	NHS (Pillar 1) total tests	UK Gov (Pillar 2) positive tests	...	Lateral flow positive tests	TotalLFD	Unnamed: 13	Unnamed: 14	Unnamed: 15	Unna
785	2022-04-28	City of Edinburgh	221	1	380	3355	3735	10.0	578	149.0	...	221.0	2775.0	NaN	NaN	NaN	
6	2020-03-10	City of Edinburgh	1	0	5	95	100	5.0	100	0.0	...	NaN	NaN	Incidence Rate	NaN	Total number of Deaths	
254	2020-11-13	City of Edinburgh	48	1	60	1371	1431	14.0	666	46.0	...	NaN	NaN	NaN	NaN	NaN	
105	2020-06-17	City of Edinburgh	4	0	7	588	595	7.0	335	0.0	...	NaN	NaN	NaN	NaN	NaN	
146	2020-07-28	City of Edinburgh	3	0	0	1335	1335	NaN	411	NaN	...	NaN	NaN	NaN	NaN	NaN	
432	2021-05-10	City of Edinburgh	19	0	20	2056	2076	5.0	1406	15.0	...	NaN	NaN	NaN	NaN	NaN	
282	2020-12-11	City of Edinburgh	69	3	71	1704	1775	12.0	674	59.0	...	NaN	NaN	NaN	NaN	NaN	
267	2020-11-26	City of Edinburgh	45	2	53	1605	1658	11.0	478	42.0	...	NaN	NaN	NaN	NaN	NaN	
207	2020-09-27	City of Edinburgh	76	0	79	694	773	6.0	232	73.0	...	NaN	NaN	NaN	NaN	NaN	
773	2022-04-16	City of Edinburgh	334	3	547	3631	4178	17.0	368	165.0	...	365.0	3401.0	NaN	NaN	NaN	

10 rows × 21 columns

In [8]: format_dict={'Date': '{:%Y}', 'NHS (pillar 1) positive tests ': '{0:,.2f}', 'UK Gov (Pillar 2) positive tests': '{0:,.2f}', 'Lateral f

In [9]: data.head().style.format(format_dict)

Out[9]:

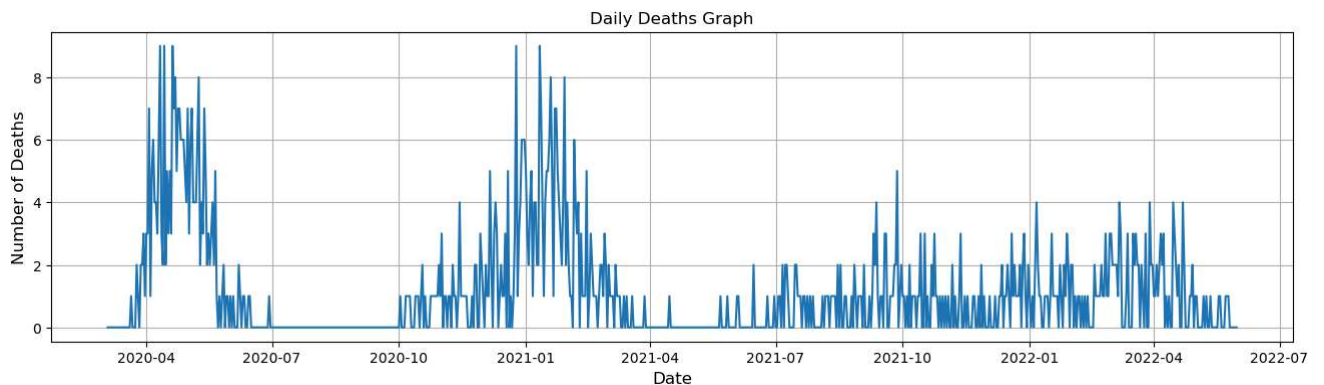
	Date	Location	Daily positive cases	Daily deaths	Daily positive tests	Daily Negative Tests	TotalTests	NHS (pillar 1) positive tests	NHS (Pillar 1) total tests	UK Gov (Pillar 2) positive tests	UK Gov (Pillar 2) total tests	Lateral flow positive tests	TotalLFD	Unnamed: 13	Unnamed: 14	Unnamed: 15	Unr
0	2020	City of Edinburgh	1	0	1	24	25	1.00	25	0.00	0	nan	nan	Total Number of Positive Cases	nan	Average Positive Case per day	
1	2020	City of Edinburgh	0	0	1	22	23	1.00	23	0.00	0	nan	nan	183951	nan	224.604396	
2	2020	City of Edinburgh	1	0	2	21	23	2.00	23	0.00	0	nan	nan	nan	nan	nan	
3	2020	City of Edinburgh	2	0	3	24	27	3.00	27	0.00	0	nan	nan	Total Tests	nan	nan	
4	2020	City of Edinburgh	2	0	2	16	18	2.00	18	0.00	0	nan	nan	2625595	nan	nan	

```
In [10]: #getting the descriptive statistics foe each numerical column
data.describe().T
```

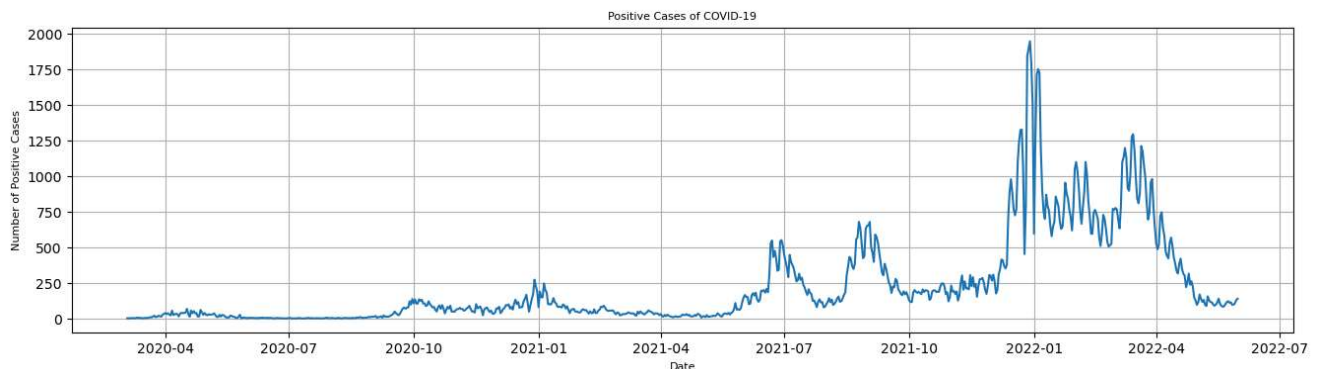
```
Out[10]:
```

	count	mean	std	min	25%	50%	75%	max
Daily positive cases	819.0	224.604396	320.794109	0.0	24.5	89.0	267.0	1947.0
Daily deaths	819.0	1.124542	1.696140	0.0	0.0	1.0	1.0	9.0
Daily positive tests	819.0	280.610501	415.904559	0.0	28.0	104.0	296.0	2037.0
Daily Negative Tests	819.0	2925.244200	3147.042579	16.0	983.0	1950.0	3411.0	18164.0
TotalTests	819.0	3205.854701	3508.436323	18.0	1012.0	2035.0	3659.0	20182.0
NHS (pillar 1) positive tests	747.0	16.657296	16.336372	0.0	5.5	12.0	21.0	129.0
NHS (Pillar 1) total tests	819.0	662.477411	395.545840	18.0	339.5	556.0	985.0	1695.0
UK Gov (Pillar 2) positive tests	747.0	190.433735	266.910683	0.0	21.0	87.0	266.5	1981.0
UK Gov (Pillar 2) total tests	819.0	1315.347985	1220.897855	0.0	487.0	1017.0	1707.0	6445.0
Lateral flow positive tests	147.0	509.190476	280.988674	100.0	303.5	506.0	667.5	1183.0
TotalLFD	147.0	6841.877551	3815.509885	1050.0	2853.5	7695.0	9959.5	12743.0
Unnamed: 14	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Unnamed: 16	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN

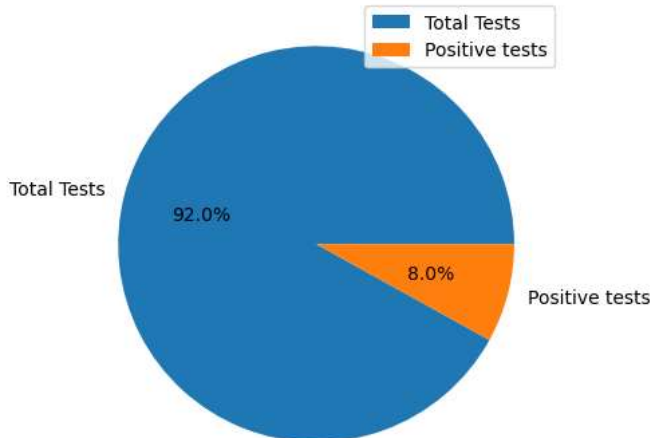
```
In [13]: gn=data.groupby("Date")["Daily deaths "].sum()
plt.figure(figsize=(16, 4))
plt.plot(gn)
plt.title('Daily Deaths Graph', fontsize=12)
plt.xlabel('Date', fontsize=12)
plt.ylabel('Number of Deaths', fontsize=12)
plt.grid(True)
plt.show()
```



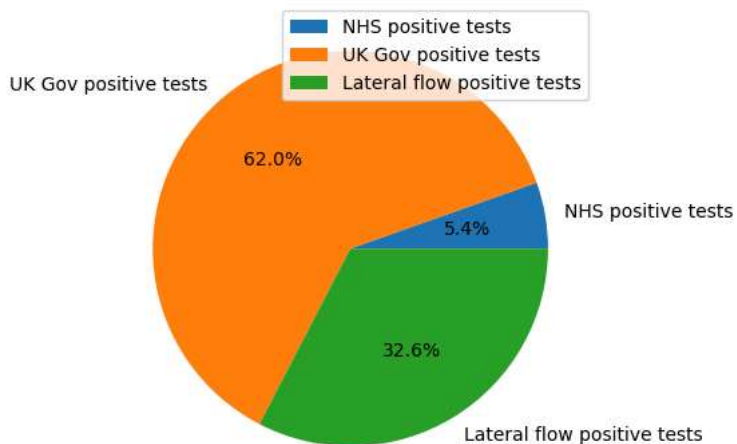
```
In [14]: go=data.groupby("Date")["Daily positive cases "].sum()
plt.figure(figsize=(16, 4))
plt.plot(go)
plt.title('Positive Cases of COVID-19', fontsize=8)
plt.xlabel('Date', fontsize=8)
plt.ylabel('Number of Positive Cases', fontsize=8)
plt.grid(True)
plt.show()
```



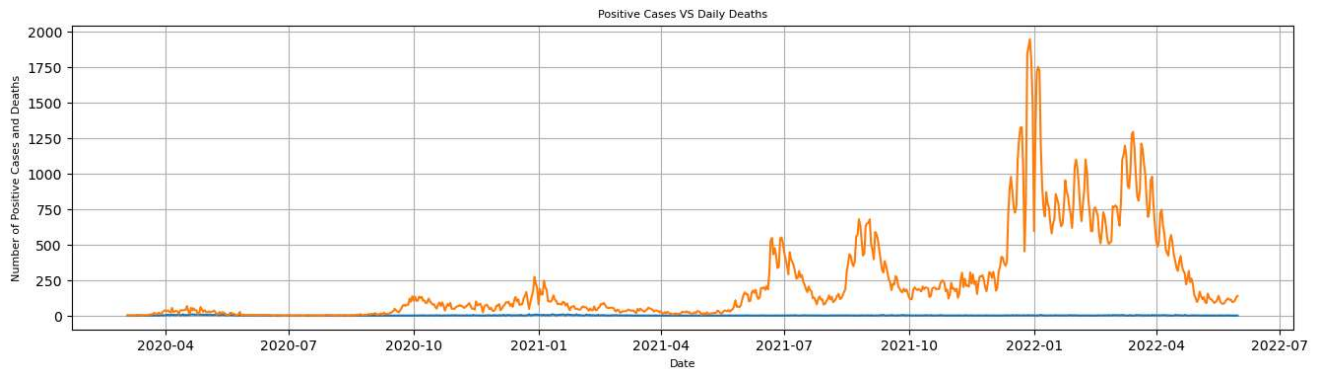
```
In [15]: ga=data[["TotalTests", "Daily positive tests "]].sum()
y=data[["TotalTests", "Daily positive tests "]].sum()
mylabels=["Total Tests", "Positive tests"]
plt.pie(y, labels=mylabels, autopct='%1.1f%%')
plt.legend()
plt.show()
```



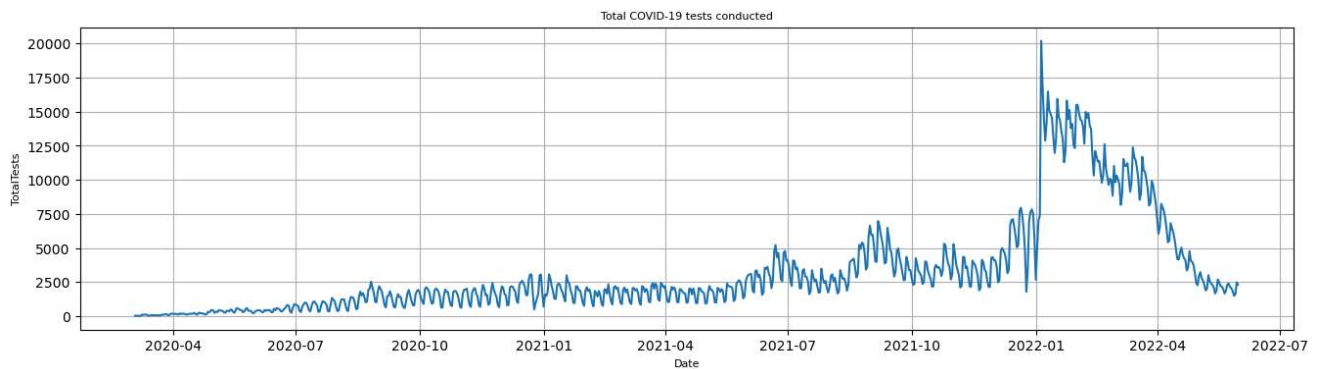
```
In [18]: y=data[["NHS (pillar 1 ) positive tests ", "UK Gov (Pillar 2) positive tests", "Lateral flow positive tests"]].sum()
mylabels=["NHS positive tests", "UK Gov positive tests", "Lateral flow positive tests"]
plt.pie(y, labels=mylabels, autopct='%1.1f%%')
plt.legend()
plt.show()
```



```
In [19]: tab=data.groupby("Date")["Daily deaths ", "Daily positive cases "].sum()
plt.figure(figsize=(16, 4))
plt.plot(tab)
plt.title('Positive Cases VS Daily Deaths', fontsize=8)
plt.xlabel('Date', fontsize=8)
plt.ylabel('Number of Positive Cases and Deaths', fontsize=8)
plt.grid(True)
plt.show()
```



```
In [20]: Cases=data.groupby("Date")["TotalTests"].sum()
plt.figure(figsize=(16, 4))
plt.plot(Cases)
plt.title('Total COVID-19 tests conducted', fontsize=8)
plt.xlabel('Date', fontsize=8)
plt.ylabel('TotalTests', fontsize=8)
plt.grid(True)
plt.show()
```



```
In [21]: corr = data.corr()
plt.figure(figsize=[20, 8])
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)
plt.title('Correlation Matrix', fontsize=16)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show();
```

