In [38]:

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

In [39]:

```
df=pd.read_csv("C:/Users/DELL/Desktop/covid.csv",parse_dates=(["date"]))
df.head()
```

Out[39]:

| | location | date | variant | num_sequences | perc_sequences | num_sequences_total |
|---|---|---|---|---|---|---|
| **0** | Angola | 2020-07-06 | Alpha | 0 | 0.0 | 3 |
| **1** | Angola | 2020-07-06 | B.1.1.277 | 0 | 0.0 | 3 |
| **2** | Angola | 2020-07-06 | B.1.1.302 | 0 | 0.0 | 3 |
| **3** | Angola | 2020-07-06 | B.1.1.519 | 0 | 0.0 | 3 |
| **4** | Angola | 2020-07-06 | B.1.160 | 0 | 0.0 | 3 |

In [40]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100416 entries, 0 to 100415
Data columns (total 6 columns):
 #   Column               Non-Null Count   Dtype         
---  ------               --------------   -----         
 0   location             100416 non-null  object        
 1   date                 100416 non-null  datetime64[ns]
 2   variant              100416 non-null  object        
 3   num_sequences        100416 non-null  int64         
 4   perc_sequences       100416 non-null  float64       
 5   num_sequences_total  100416 non-null  int64         
dtypes: datetime64[ns](1), float64(1), int64(2), object(2)
memory usage: 4.6+ MB
```

In [41]:

```
df.describe()
```

Out[41]:

| | num_sequences | perc_sequences | num_sequences_total |
|---|---|---|---|
| **count** | 100416.000000 | 100416.000000 | 100416.000000 |
| **mean** | 72.171676 | 6.154355 | 1509.582457 |
| **std** | 1669.262169 | 21.898989 | 8445.291772 |
| **min** | 0.000000 | -0.010000 | 1.000000 |
| **25%** | 0.000000 | 0.000000 | 12.000000 |
| **50%** | 0.000000 | 0.000000 | 59.000000 |
| **75%** | 0.000000 | 0.000000 | 394.000000 |
| **max** | 142280.000000 | 100.000000 | 146170.000000 |

In [42]:

```
df["location"].value_counts()
```

Out[42]:

```
Bangladesh        1080
Belgium           1080
United States     1080
United Kingdom    1080
France            1080
                  ...
Montenegro         384
Monaco             360
Fiji               336
Benin              336
Brunei             240
Name: location, Length: 121, dtype: int64
```

In [43]:

```
df["location"].unique()
```

Out[43]:

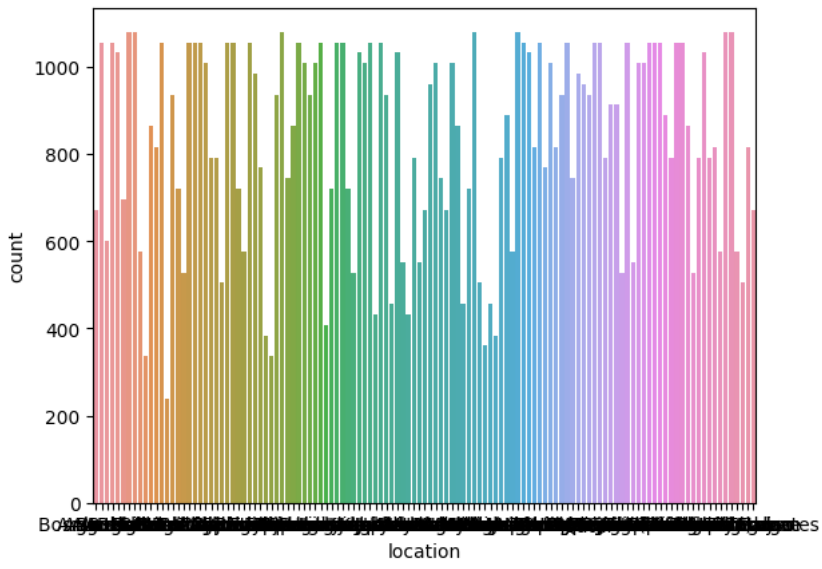
```
array(['Angola', 'Argentina', 'Aruba', 'Australia', 'Austria', 'Bahrain',
       'Bangladesh', 'Belgium', 'Belize', 'Benin',
       'Bosnia and Herzegovina', 'Botswana', 'Brazil', 'Brunei',
       'Bulgaria', 'Cambodia', 'Cameroon', 'Canada', 'Chile', 'Colombia',
       'Costa Rica', 'Croatia', 'Curacao', 'Cyprus', 'Czechia', 'Denmark',
       'Djibouti', 'Dominican Republic', 'Ecuador', 'Egypt', 'Estonia',
       'Ethiopia', 'Fiji', 'Finland', 'France', 'Gambia', 'Georgia',
       'Germany', 'Ghana', 'Greece', 'Guatemala', 'Hong Kong', 'Hungary',
       'Iceland', 'India', 'Indonesia', 'Iran', 'Iraq', 'Ireland',
       'Israel', 'Italy', 'Jamaica', 'Japan', 'Jordan', 'Kazakhstan',
       'Kenya', 'Kosovo', 'Kuwait', 'Latvia', 'Lebanon', 'Liechtenstein',
       'Lithuania', 'Luxembourg', 'Madagascar', 'Malawi', 'Malaysia',
       'Maldives', 'Malta', 'Mauritius', 'Mexico', 'Moldova', 'Monaco',
       'Mongolia', 'Montenegro', 'Morocco', 'Mozambique', 'Nepal',
       'Netherlands', 'New Zealand', 'Nigeria', 'North Macedonia',
       'Norway', 'Oman', 'Pakistan', 'Papua New Guinea', 'Paraguay',
       'Peru', 'Philippines', 'Poland', 'Portugal', 'Qatar', 'Romania',
       'Russia', 'Rwanda', 'Senegal', 'Serbia', 'Seychelles', 'Singapore',
       'Sint Maarten (Dutch part)', 'Slovakia', 'Slovenia',
       'South Africa', 'South Korea', 'Spain', 'Sri Lanka', 'Suriname',
       'Sweden', 'Switzerland', 'Thailand', 'Togo', 'Trinidad and Tobago',
       'Turkey', 'Uganda', 'Ukraine', 'United Arab Emirates',
       'United Kingdom', 'United States', 'Uruguay', 'Vietnam', 'Zambia',
       'Zimbabwe'], dtype=object)
```

In [44]:

```
figsize=(20,20)
sns.countplot(x=df["location"],data=df)
```

Out[44]:

```
<AxesSubplot: xlabel='location', ylabel='count'>
```



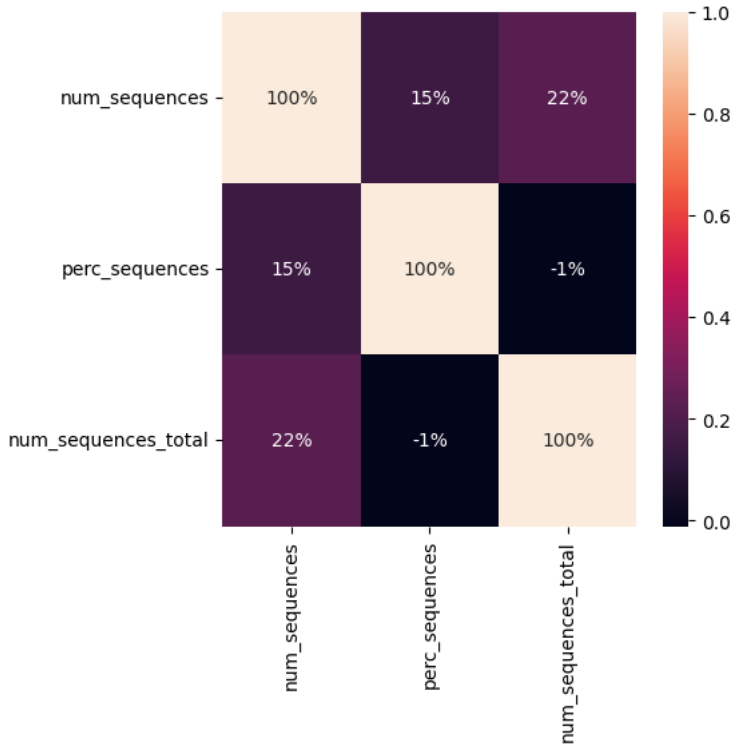
In [45]:

```
plt.figure(figsize=(5,5))
sns.heatmap(df.corr(),annot=True,fmt=".0%")
```

```
C:\Users\DELL\AppData\Local\Temp\ipykernel_11228\3529239389.py:2: FutureWarning: The default value of numeric_only in DataF
rame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of n
umeric_only to silence this warning.
  sns.heatmap(df.corr(),annot=True,fmt=".0%")
```

Out[45]:

```
<AxesSubplot: >
```




In [46]:

```
a=df[df["location"]=="India"].count()
a
```

Out[46]:

```
location              1056
date                  1056
variant               1056
num_sequences         1056
perc_sequences        1056
num_sequences_total   1056
dtype: int64
```

In [47]:

```
a1=df[df["location"]=="India"]["variant"].value_counts()
a1
```

Out[47]:

```
Alpha             44
B.1.1.277         44
others            44
S:677P.Pelican    44
S:677H.Robin1     44
Omicron           44
Mu                44
Lambda            44
Kappa             44
Iota              44
Gamma             44
Eta               44
Epsilon           44
Delta             44
Beta              44
B.1.620           44
B.1.367           44
B.1.258           44
B.1.221           44
B.1.177           44
B.1.160           44
B.1.1.519         44
B.1.1.302         44
non_who           44
Name: variant, dtype: int64
```

In [48]:

```
a2=df[df["location"]=="Sri Lanka"]["variant"].value_counts()
a2
```

Out[48]:

```
Alpha             37
B.1.1.277         37
others            37
S:677P.Pelican    37
S:677H.Robin1     37
Omicron           37
Mu                37
Lambda            37
Kappa             37
Iota              37
Gamma             37
Eta               37
Epsilon           37
Delta             37
Beta              37
B.1.620           37
B.1.367           37
B.1.258           37
B.1.221           37
B.1.177           37
B.1.160           37
B.1.1.519         37
B.1.1.302         37
non_who           37
Name: variant, dtype: int64
```

In [49]:

```
a1=df[df["location"]=="India"]
```

In [51]:

```
c=df.groupby(df["date"].dt.month_name())["variant"].value_counts()
c
```

Out[51]:

```
date       variant
April      Alpha             215
           B.1.1.277         215
           B.1.1.302         215
           B.1.1.519         215
           B.1.160           215
                            ...
September  Omicron           367
           S:677H.Robin1     367
           S:677P.Pelican    367
           non_who           367
           others            367
Name: variant, Length: 288, dtype: int64
```

In [60]:

```
dataframe=pd.DataFrame(c)
dataframe
```

Out[60]:

| | | variant |
|---|---|---|
| **date** | **variant** | |
| **April** | **Alpha** | 215 |
| | **B.1.1.277** | 215 |
| | **B.1.1.302** | 215 |
| | **B.1.1.519** | 215 |
| | **B.1.160** | 215 |
| **...** | **...** | ... |
| **September** | **Omicron** | 367 |
| | **S:677H.Robin1** | 367 |
| | **S:677P.Pelican** | 367 |
| | **non_who** | 367 |
| | **others** | 367 |

288 rows × 1 columns

In [61]:

```
df.head()
```

Out[61]:

| | location | date | variant | num_sequences | perc_sequences | num_sequences_total |
|---|---|---|---|---|---|---|
| **0** | Angola | 2020-07-06 | Alpha | 0 | 0.0 | 3 |
| **1** | Angola | 2020-07-06 | B.1.1.277 | 0 | 0.0 | 3 |
| **2** | Angola | 2020-07-06 | B.1.1.302 | 0 | 0.0 | 3 |
| **3** | Angola | 2020-07-06 | B.1.1.519 | 0 | 0.0 | 3 |
| **4** | Angola | 2020-07-06 | B.1.160 | 0 | 0.0 | 3 |

In [66]:

```
df["month"]=df["date"].dt.month_name()
```

In [67]:

```
df.head()
```

Out[67]:

| | location | date | variant | num_sequences | perc_sequences | num_sequences_total | month |
|---|---|---|---|---|---|---|---|
| **0** | Angola | 2020-07-06 | Alpha | 0 | 0.0 | 3 | July |
| **1** | Angola | 2020-07-06 | B.1.1.277 | 0 | 0.0 | 3 | July |
| **2** | Angola | 2020-07-06 | B.1.1.302 | 0 | 0.0 | 3 | July |
| **3** | Angola | 2020-07-06 | B.1.1.519 | 0 | 0.0 | 3 | July |
| **4** | Angola | 2020-07-06 | B.1.160 | 0 | 0.0 | 3 | July |

In [69]:

```
df["weekday"]=df["date"].dt.weekday
```

In [71]:

```
df.tail()
```

Out[71]:

| | location | date | variant | num_sequences | perc_sequences | num_sequences_total | month | weekday |
|---|---|---|---|---|---|---|---|---|
| **100411** | Zimbabwe | 2021-11-01 | Omicron | 0 | 0.0 | 6 | November | 0 |
| **100412** | Zimbabwe | 2021-11-01 | S:677H.Robin1 | 0 | 0.0 | 6 | November | 0 |
| **100413** | Zimbabwe | 2021-11-01 | S:677P.Pelican | 0 | 0.0 | 6 | November | 0 |
| **100414** | Zimbabwe | 2021-11-01 | others | 0 | 0.0 | 6 | November | 0 |
| **100415** | Zimbabwe | 2021-11-01 | non_who | 0 | 0.0 | 6 | November | 0 |

In [77]:

```
df["weekday"].unique()
```

Out[77]:

```
array([0, 2], dtype=int64)
```

In [104]:

```
month_wise_India=df[df["location"]=="India"]["month"].value_counts()
month_wise_India=pd.DataFrame(month_wise_India)
month_wise_India
```

Out[104]:

| | month |
|---|---|
| **May** | 120 |
| **August** | 120 |
| **November** | 120 |
| **June** | 96 |
| **July** | 96 |
| **September** | 96 |
| **October** | 96 |
| **December** | 96 |
| **January** | 72 |
| **February** | 48 |
| **March** | 48 |
| **April** | 48 |

In [105]:

```
month_wise_India["Index"]=month_wise_India.index
```

In [106]:

```
month_wise_India.head()
```

Out[106]:

| | month | Index |
|---|---|---|
| **May** | 120 | May |
| **August** | 120 | August |
| **November** | 120 | November |
| **June** | 96 | June |
| **July** | 96 | July |

In [ ]:

```

```