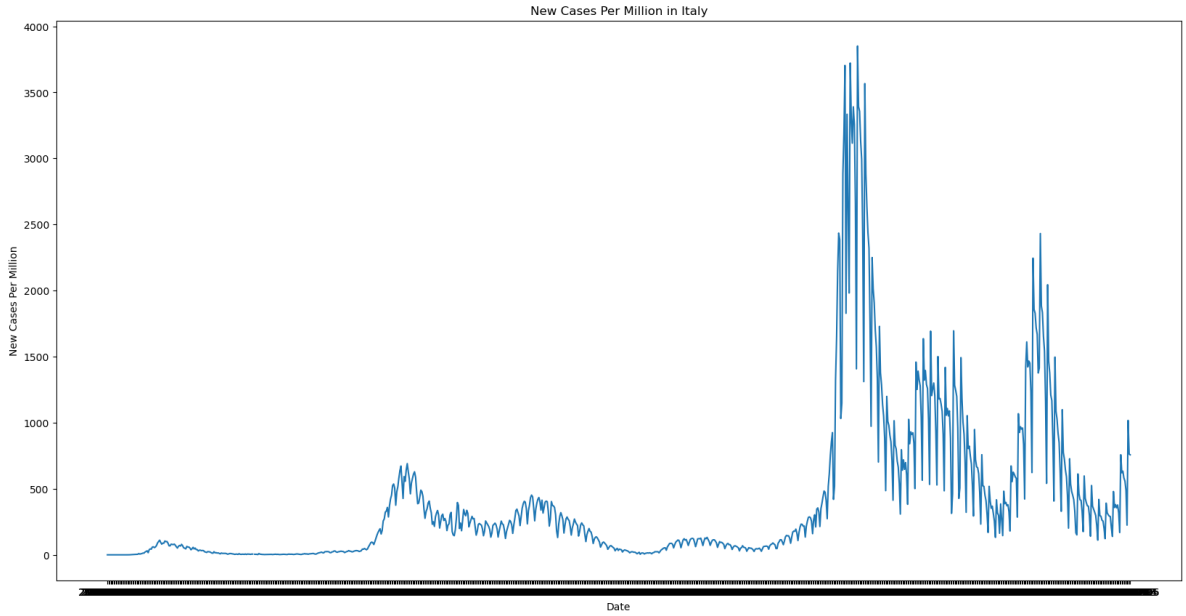
In [18]:

```
'''import owid_covid_data.csv'''

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

# Read the data
df = pd.read_csv('owid_covid_data.csv')

# Check the data
df.head()

'''all data from Italy must be in a separate dataframe'''

# Create a new dataframe with only Italy data
df_italy = df[df['location'] == 'Italy']


'''filter collumns from df_italy except date, total_cases, new_cases, total_deaths, new_dea

df_italy = df_italy[['date', 'total_cases', 'new_cases', 'total_deaths', 'new_deaths', 'tot
print(df_italy.head())

'''visualize the df_italy data by plotting a line graph with date being the x-axis and new_

# Plot the data
plt.figure(figsize=(20,10))
plt.plot(df_italy['date'], df_italy['new_cases_per_million'])
plt.xlabel('Date')
plt.ylabel('New Cases Per Million')
plt.title('New Cases Per Million in Italy')
plt.show()

# show dimensions of the df_italy dataframe
print(df_italy.shape)
```

```
            date  total_cases  new_cases  total_deaths  new_deaths  \
98722  2020-01-31          2.0        2.0           NaN         NaN
98723  2020-02-01          2.0        0.0           NaN         NaN
98724  2020-02-02          2.0        0.0           NaN         NaN
98725  2020-02-03          2.0        0.0           NaN         NaN
98726  2020-02-04          2.0        0.0           NaN         NaN

       total_cases_per_million  new_cases_per_million  \
98722                    0.034                  0.034
98723                    0.034                  0.000
98724                    0.034                  0.000
98725                    0.034                  0.000
98726                    0.034                  0.000

       total_deaths_per_million  new_deaths_per_million  total_vaccination
s  \
98722                       NaN                     NaN                 Na
N
```

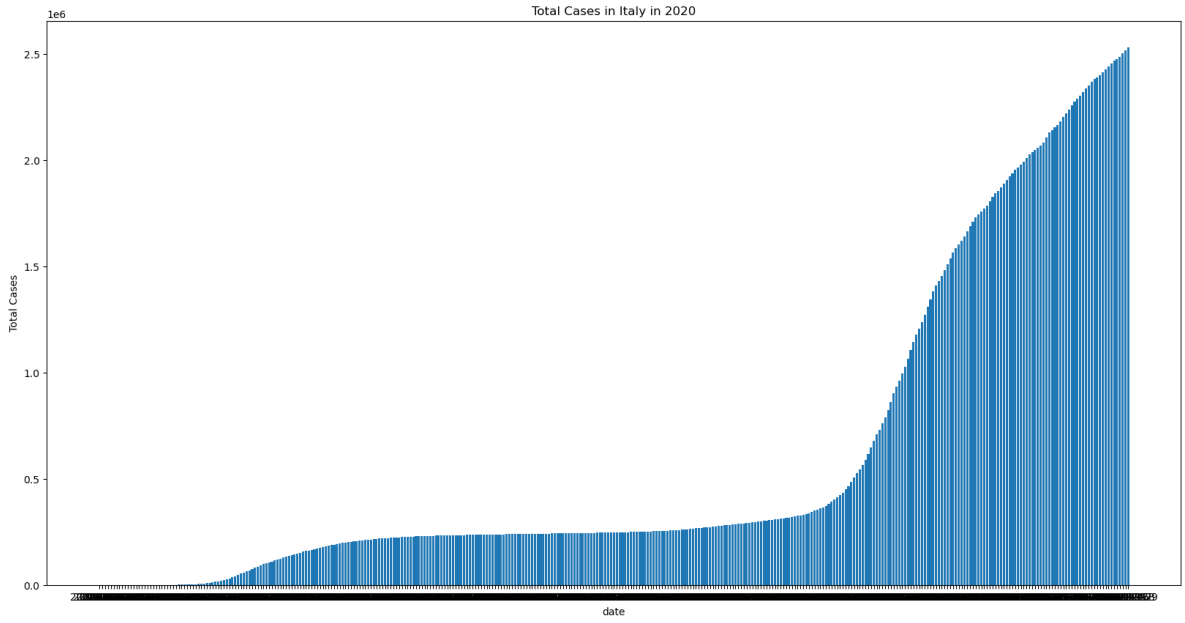
```
98723                      NaN                      NaN                   Na
N
98724                      NaN                      NaN                   Na
N
98725                      NaN                      NaN                   Na
N
98726                      NaN                      NaN                   Na
N

       people_fully_vaccinated_per_hundred
98722                                  NaN
98723                                  NaN
98724                                  NaN
98725                                  NaN
98726                                  NaN
```



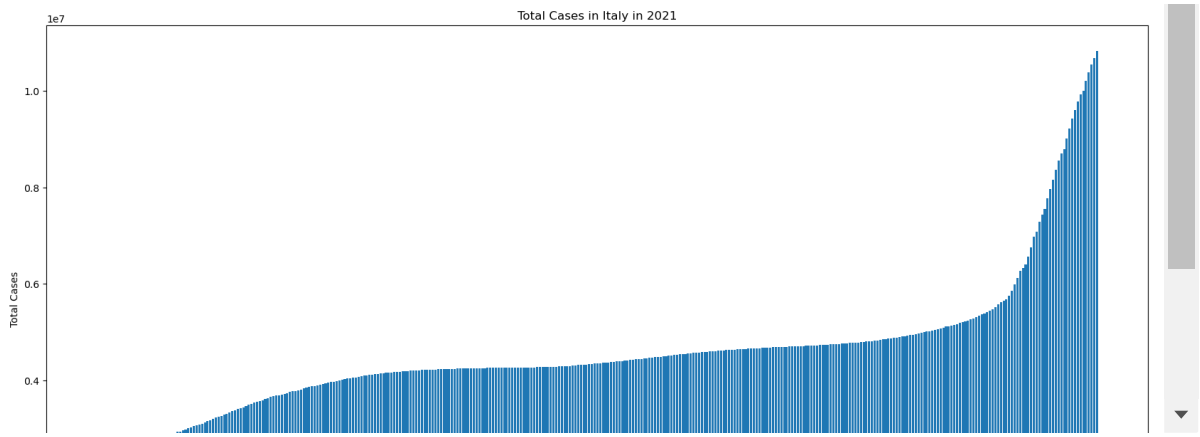

```
(980, 11)
```

In [19]:

```
'''second subquestion: how many covid cases were there in Italy in 2020 and 2021?'''

# visualize the total number of cases in Italy in 2020
plt.figure(figsize=(20,10))
plt.bar(df_italy['date'][0:365], df_italy['total_cases'][0:365])
plt.xlabel('date')
plt.ylabel('Total Cases')
plt.title('Total Cases in Italy in 2020')
plt.show()

# create a new dataframe with only total cases in 2020 and dates
df_italy_2020 = df_italy[['date', 'total_cases']][0:365]

# pick last date of year 2020 and show total cases value
print('Total number of new cases in Italy in 2020: ', df_italy_2020.iloc[364, 1])

# visualize the total number of cases in Italy in 2021
plt.figure(figsize=(20,10))
plt.bar(df_italy['date'][365:365+365], df_italy['total_cases'][365:365+365])
plt.xlabel('date')
plt.ylabel('Total Cases')
plt.title('Total Cases in Italy in 2021')
plt.show()

# create a new dataframe with only total cases in 2021 and dates
df_italy_2021 = df_italy[['date', 'total_cases']][365:365+365]

# pick last date of year 2021 and show total cases value and suberact the total cases value
print('Total number of new cases in Italy in 2021: ', df_italy_2021.iloc[364, 1]-df_italy_2
```




```
Total number of new cases in Italy in 2020:  2529070.0
```




```
Total number of new cases in Italy in 2021:  8292305.0
```

Type *Markdown* and LaTeX: α^2

In [25]:

```
# subuestion 3: was there an asending or descending trend in covid cases in Italy in the mo

# create a new dataframe with only total cases in July 2020 and dates
df_italy_july_2020 = df_italy[['date', 'total_cases']][183:183+31]

# print the new dataframe
print(df_italy_july_2020)

#print size of the dataframe df_italy_july_2020
print(df_italy_july_2020.shape)

# visualize the total number of cases in Italy in July 2020
plt.figure(figsize=(20,10))
plt.bar(df_italy_july_2020['date'], df_italy_july_2020['total_cases'])
plt.xlabel('date')
plt.ylabel('Total Cases')
plt.title('Total Cases in Italy in July 2020')
plt.show()

# create a new dataframe with only total cases in July 2021 and dates
df_italy_july_2021 = df_italy[['date', 'total_cases']][548:548+31]

# print the new dataframe
print(df_italy_july_2021)

#print size of the dataframe df_italy_july_2021
print(df_italy_july_2021.shape)

# visualize the total number of cases in Italy in July 2021
plt.figure(figsize=(20,10))
plt.bar(df_italy_july_2021['date'], df_italy_july_2021['total_cases'])
# y axis scale is to big, so make it smaller
plt.ylim(4000000, 5000000)
plt.xlabel('date')
plt.ylabel('Total Cases')
plt.title('Total Cases in Italy in July 2021')
plt.show()

#calculate the total number of new cases in July 2020
print('Total number of new cases in Italy in July 2020: ', df_italy_july_2020.iloc[30, 1]-d

#calculate the total number of new cases in July 2021
print('Total number of new cases in Italy in July 2021: ', df_italy_july_2021.iloc[30, 1]-d
```

```
            date  total_cases
98905  2020-08-01     247832.0
98906  2020-08-02     248070.0
98907  2020-08-03     248229.0
98908  2020-08-04     248419.0
98909  2020-08-05     248803.0
98910  2020-08-06     249204.0
98911  2020-08-07     249756.0
98912  2020-08-08     250103.0
98913  2020-08-09     250566.0
98914  2020-08-10     250825.0
98915  2020-08-11     251237.0
98916  2020-08-12     251713.0
98917  2020-08-13     252235.0
```

```
98918  2020-08-14     252809.0
98919  2020-08-15     253438.0
98920  2020-08-16     253915.0
98921  2020-08-17     254235.0
98922  2020-08-18     254636.0
```