Prevalence_rate

December 6, 2020

```
[12]: from IPython.display import Image
      Image("../Images/Logo.jpg")
```
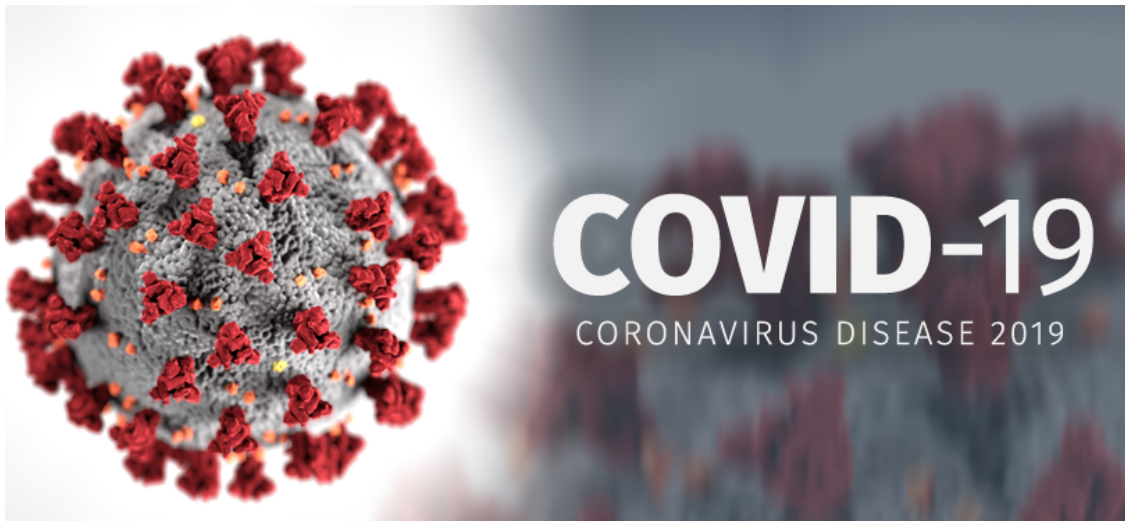
[12]:

#

Graduate Project ENEL 698

Github Link

```
[13]: Image("../Images/Covid-19.png")
```

[13]:

0.1 This Notebook constitues of analysis of Infection prevalence rate and Infection fatality rate of COVID-19 Canada Specific data.

```
[1]: import pandas as pd
     import numpy as np

     import matplotlib.pyplot as plt
     import seaborn as sns
     %matplotlib inline
```

1 Data Loading

```
[2]: df = pd.read_csv("https://health-infobase.canada.ca/src/data/covidLive/
     ↪covid19-download.csv")
```

- Data Source (https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavir

```
[3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4107 entries, 0 to 4106
Data columns (total 35 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   pruid              4107 non-null    int64
 1   prname             4107 non-null    object
 2   prnameFR           4107 non-null    object
 3   date               4107 non-null    object
 4   numconf            4107 non-null    int64
 5   numprob            4107 non-null    int64
 6   numdeaths          3988 non-null    float64
 7   numtotal           4107 non-null    int64
 8   numtested          4050 non-null    float64
 9   numrecover         3583 non-null    float64
 10  percentrecover     3379 non-null    float64
 11  ratetested         3768 non-null    float64
 12  numtoday           4107 non-null    int64
 13  percentoday        4107 non-null    float64
 14  ratetotal          3837 non-null    float64
 15  ratedeaths         3837 non-null    float64
 16  numdeathstoday     3988 non-null    float64
 17  percentdeath       3710 non-null    float64
 18  numtestedtoday     4050 non-null    float64
 19  numrecoveredtoday  3583 non-null    float64
 20  percentactive      3710 non-null    float64
 21  numactive          3710 non-null    float64
 22  rateactive         3837 non-null    float64
```

```
23   numtotal_last14      3855 non-null   float64
24   ratetotal_last14     3598 non-null   float64
25   numdeaths_last14     3855 non-null   float64
26   ratedeaths_last14    3598 non-null   float64
27   numtotal_last7       3855 non-null   float64
28   ratetotal_last7      3598 non-null   float64
29   numdeaths_last7      3855 non-null   float64
30   ratedeaths_last7     3598 non-null   float64
31   avgtotal_last7       3855 non-null   float64
32   avgincidence_last7   3598 non-null   float64
33   avgdeaths_last7      3855 non-null   float64
34   avgratedeaths_last7  3598 non-null   float64
dtypes: float64(27), int64(5), object(3)
memory usage: 1.1+ MB
```

1.0.1 Visualizing Null values in the dataset

```
[4]: sns.heatmap(df.isnull())
```

```
[4]: <matplotlib.axes._subplots.AxesSubplot at 0x21435154c08>
```
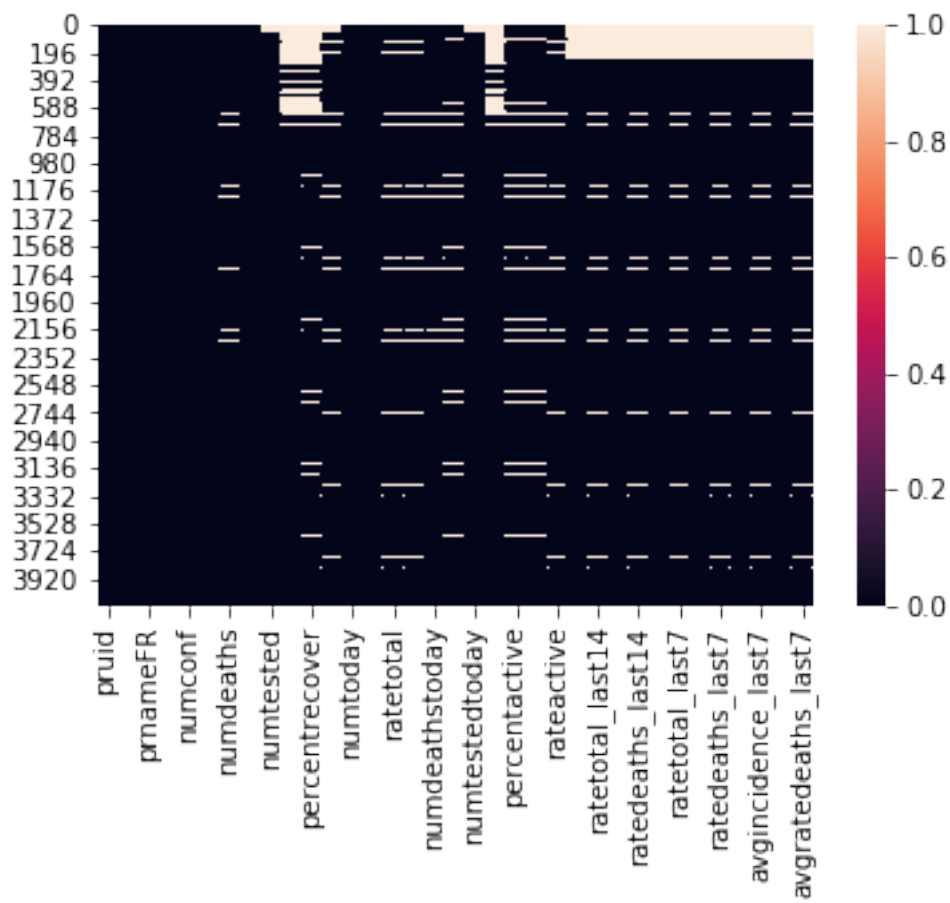
```
[5]:  #### Getting the canada specific data(not the provincial data)

      df = df[df['prname']=='Canada']
      #df.head()
```

```
[6]:  df['date'] = pd.to_datetime(df['date'])
      df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 286 entries, 2 to 4106
Data columns (total 35 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   pruid              286 non-null    int64
 1   prname             286 non-null    object
 2   prnameFR           286 non-null    object
 3   date               286 non-null    datetime64[ns]
 4   numconf            286 non-null    int64
 5   numprob            286 non-null    int64
 6   numdeaths          286 non-null    float64
 7   numtotal           286 non-null    int64
 8   numtested          270 non-null    float64
 9   numrecover         242 non-null    float64
 10  percentrecover     242 non-null    float64
 11  ratetested         270 non-null    float64
 12  numtoday           286 non-null    int64
 13  percentoday        286 non-null    float64
 14  ratetotal          286 non-null    float64
 15  ratedeaths         286 non-null    float64
 16  numdeathstoday     286 non-null    float64
 17  percentdeath       286 non-null    float64
 18  numtestedtoday     270 non-null    float64
 19  numrecoveredtoday  242 non-null    float64
 20  percentactive      286 non-null    float64
 21  numactive          286 non-null    float64
 22  rateactive         286 non-null    float64
 23  numtotal_last14    257 non-null    float64
 24  ratetotal_last14   257 non-null    float64
 25  numdeaths_last14   257 non-null    float64
 26  ratedeaths_last14  257 non-null    float64
 27  numtotal_last7     257 non-null    float64
 28  ratetotal_last7    257 non-null    float64
 29  numdeaths_last7    257 non-null    float64
 30  ratedeaths_last7   257 non-null    float64
 31  avgtotal_last7     257 non-null    float64
 32  avgincidence_last7 257 non-null    float64
```

```
 33  avgdeaths_last7      257 non-null     float64
 34  avgratedeaths_last7  257 non-null     float64
dtypes: datetime64[ns](1), float64(27), int64(5), object(2)
memory usage: 80.4+ KB
```

- So our data is nicely formatted and we have converted the date column to datetime formating now we will calculate the infection prevalence rate, and fatality rate

2 Prevalence

source (https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html) ### Definition of prevalence

Prevalence, sometimes referred to as prevalence rate,is the proportion of persons in a population who have a particular disease or attitude at a specified point in time or over a specified period of time.

- Point Prevalence refers to the prevalence measured at a particular point in time. It is the proportion of persons with a particular disease or attribute on a particular disease or attribute on a particular date.
- Period Prevalence refers to prevalence measured over an interval of time. It is the proportion of persons with a particular disease or attribute at any time during the interval.

Method of Calculating prevalence of disease (All new and pre-existing cases during a given time period/Population during the same time period)*100

3 Infection Prevalence rate

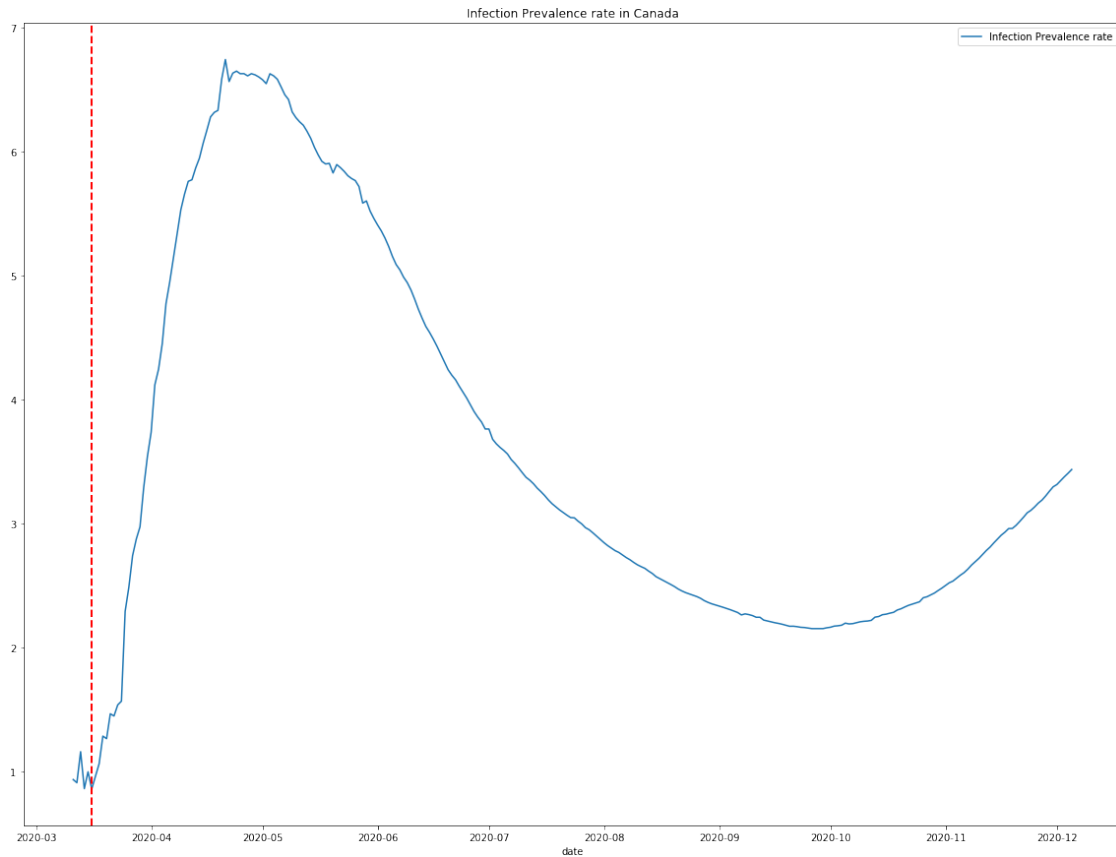- The rate is calculated on number of tested people only.

```
[10]: plt.figure(figsize=(20,15))
      ax = sns.lineplot(x=df['date'],y=(df['numconf']*100/(df['numtested'])),
       →data=df, label='Infection Prevalence rate')
      ax.axvline(pd.to_datetime('2020-03-16'),color='r', linestyle='--', lw=2, label
       →= 'Lockdown Announced')
      plt.title('Infection Prevalence rate in Canada')
```

```
[10]: Text(0.5, 1.0, 'Infection Prevalence rate in Canada')
```

Infection Prevalence rate

- From this graph we can see a high prevalance rate of 6.5% in the month of april, which dropped to around 2% by october.

4   Infection fatality prevalence rate

- The rate is calculated on number of tested people only.

```
[8]:  plt.figure(figsize=(20,15))
      ax = sns.lineplot(x=df['date'],y=(df['numdeaths']*100/df['numtested']),
      ↪data=df, label='Fatality Prevalence rate')
      ax.axvline(pd.to_datetime('2020-03-16'),color='r', linestyle='--', lw=2, label
      ↪= 'Lockdown Announced')
      plt.title('Fatality Prevalence rate in Canada')
```

```
[8]:  Text(0.5, 1.0, 'Fatality Prevalence rate in Canada')
```

Fatality Prevalence rate in Canada
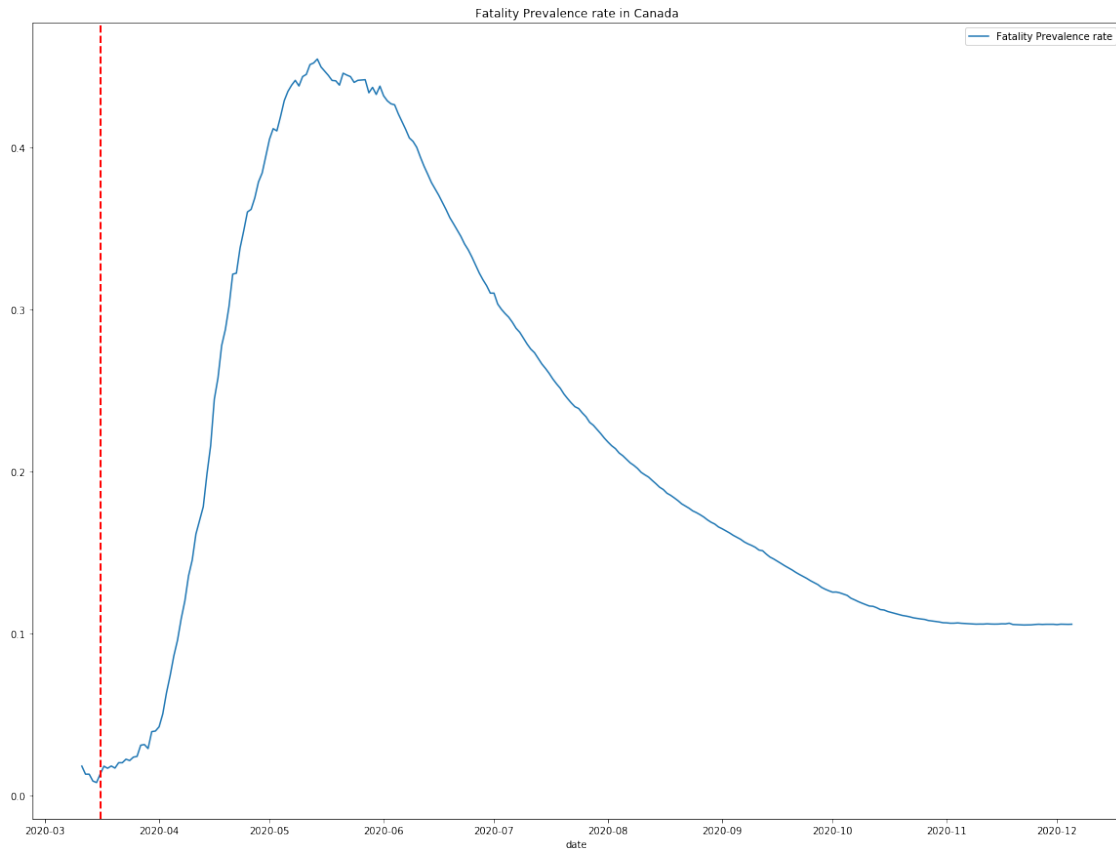
- From this graph we can see that fatality prevalence rate also dropped to around 0.1% from 0.4%.