ANALYZING DATA IN THE FORM OF

.CSV FILE THROUGH PYTHON

Gabrian Granito Yosia

03411942000009

08 March 2020

GEOPHYSICAL ENGINEERING DEPARTMENT

FACULTY OF CIVIL, PLANNING, AND EARTH ENGINEERING

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

```
In [1]: 1 import pandas as pd # importing pandas as pd
```

```
In [2]: 1 a = pd.read_csv("Birth,Death,Natural_Increase.csv") # Creating the dataframe
        2 a # Print the dataframe
```

Out[2]:

	Period	Births	Deaths	Natural_Increase
0	2000	56604	26658	29943
1	2001	55800	27825	27972
2	2002	54021	28065	25956
3	2003	56136	28011	29125
4	2004	58074	28419	29655
5	2005	57744	27033	30711
6	2006	59193	28245	30949
7	2007	64044	28521	35520
8	2008	64341	29187	35154
9	2009	62541	28965	33579
10	2010	63897	28437	35457
11	2011	61404	30081	31320
12	2012	61179	30099	31080
13	2013	58719	29568	29148
14	2014	57243	31062	26181
15	2015	61038	31608	29430
16	2016	59430	31179	28251
17	2017	59610	33339	26268
18	2018	58020	33225	24795
19	2019	59637	34260	25377

Figure1. Importing dataframe

```
In [3]: 1 a.head() #view first 5 rows of dataframe
```

Out[3]:

	Period	Births	Deaths	Natural_Increase
0	2000	56604	26658	29943
1	2001	55800	27825	27972
2	2002	54021	28065	25956
3	2003	56136	28011	29125
4	2004	58074	28419	29655

```
In [4]: 1 a.tail() #view last 5 rows of dataframe
```

Out[4]:

	Period	Births	Deaths	Natural_Increase
15	2015	61038	31608	29430
16	2016	59430	31179	28251
17	2017	59610	33339	26268
18	2018	58020	33225	24795
19	2019	59637	34260	25377

Figure2. Dataframe shown with ‘head’ and ‘tail’ function

```
In [5]: 1 a.describe() #view statistical parameters
```

Out[5]:

	Period	Births	Deaths	Natural_Increase
count	20.00000	20.000000	20.000000	20.000000
mean	2009.50000	59433.750000	29689.350000	29793.550000
std	5.91608	2883.169124	2144.992045	3288.943671
min	2000.00000	54021.000000	26658.000000	24795.000000
25%	2004.75000	57618.750000	28200.000000	27546.000000
50%	2009.50000	59311.500000	29076.000000	29542.500000
75%	2014.25000	61235.250000	31091.250000	31140.000000
max	2019.00000	64341.000000	34260.000000	35520.000000

Figure3. Viewing statistical parameters of dataframe by ‘describe’ function

```
In [6]: 1 a.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 4 columns):
Period              20 non-null int64
Births              20 non-null int64
Deaths              20 non-null int64
Natural_Increase    20 non-null int64
dtypes: int64(4)
memory usage: 768.0 bytes
```

Figure4. Information on dataframe viewed with ‘info’ function

```
In [7]: 1 import matplotlib.pyplot as plt #importing matplotlib.pyplot as plt
```

```
In [8]: 1 plt.figure(figsize=(10,8)) #adjust size of figure
        2 plt.plot(a['Births'], label='Births') #to plot 'Births' and give description to line
        3 plt.plot(a['Deaths'], label='Deaths') #to plot 'Deaths' and give description to line
        4 plt.plot(a['Natural_Increase'], label='Natural_Increase') #to plot 'Natural_Increase' and give description to line
        5 plt.xlabel('Period') #give description to x-axis
        6 plt.ylabel('Data_Count') #give description to y-axis
        7 plt.title('Population Statistics') #give title of the figure
        8 plt.legend() #give description of each line
        9
```

Figure5. Plotting dataframe

```
Out[8]: <matplotlib.legend.Legend at 0x18ebcf58c48>
```

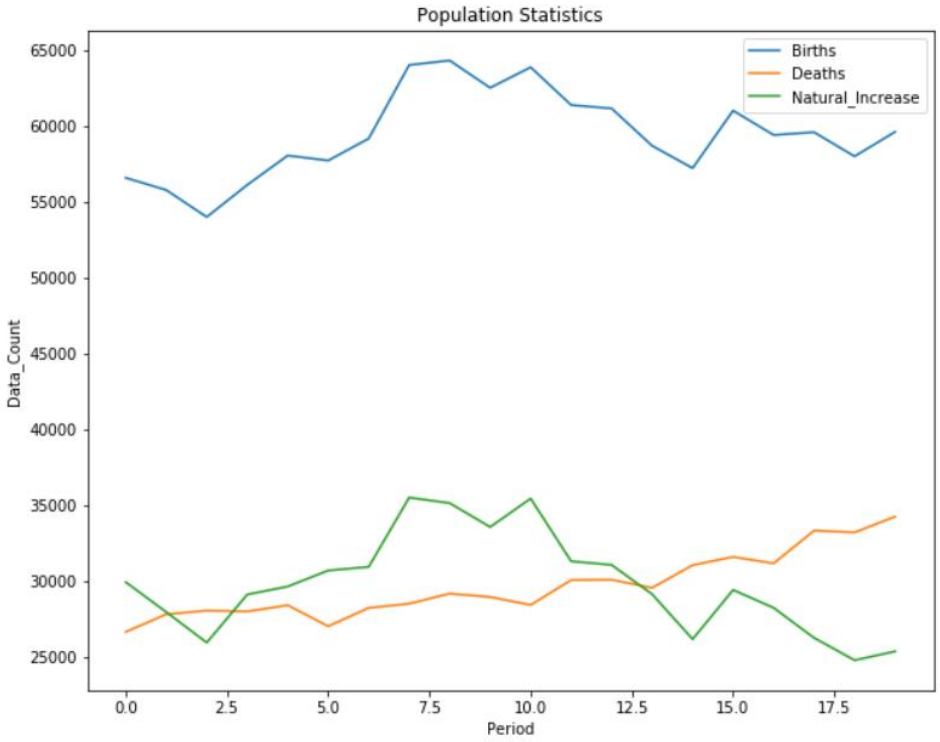

Figure6. Result of dataframe after being plotted

The data was taken from https://www.stats.govt.nz/large-datasets/csv-files-for-download/ which contains the latest data from Infoshare taken by census. The data was taken and then processed in order to analyse and visualize growth in population which took place in New Zealand in almost two decades in the period of 2000 until 2019 with births, deaths, and natural increases of human population as the variables. Graph on Figure6. shows the result of processing the data which have been previously taken in order to easily visualize and compare the following variables.

It could be seen from Figure1. The average birth rate over the periods are generally higher than the average death rate thus making the natural increases in a surplus at around 25,000 to 35,000 growth in a period. Figure6. visualizes the difference between birth rate and death rate along with natural increases.

It could be seen in Figure1. and Figure6. that over the first three periods starting from 2000, the birth rate is decreasing up until around 54,000 people which then significantly rise and gradually increases until 2007 where the growth suddenly spikes up to around 64,000 people. The birth rate stays stable until 2011 on which it drops down and gradually decreases up to around 57,000 then followed with a rise of 2,000 on the next period. The birth rate goes up and down for the rest of the periods up to the latest 2019.

The death rate over the periods of 2000 to 2019 didn't have many significant rise or fall. It had times of when it increases or decreases, yet it is in a constant rate of increasing over the years which could easily be seen on Figure6.

The natural increase is the difference between birth rate and death rate recorder over the periods. It naturally follows the birth rate and death rate. Along with the increase in birth rate and decrease in death rate, the natural increase increases. And vice versa, with the decrease in birth rate and increase in death rate, the natural increase decreases. In Figure6. it could be seen that the natural increases generally follow the birth rate. The natural increase increases when the birth rate increases, which could be seen in 2007, and decreases when the birth rate decreases, which could be seen in 2002 and 2014.

Going through Figure6. which shows the high discrepancies between the birth rate which is greatly higher than the death rate, it could be concluded that New Zealand have

had positive rate of natural increase or growth in population over the years, which rises at 2007-2010, yet currently underdoing a decrease gradually up until 2019. Going by the latest trend, it could be assumed that the natural increase would continue to decrease if the birth rate continues to decrease and the death rate continues to increase.