SID:Renhao Wu 1074372
Bowen Fan 1035162
Xinyuan Qiao 1137963
Jeffrey Mak 1086144

Introduction

Gross State Product(GSP) is said to have a positive and significant influence on living standards, which include life expectancy, infant mortality rate and literacy(Kakwani, 1993). By improving the GSP of a state, the sustainability and livability of a state can therefore be improved.

An increase in population growth indicates an increase in the labour force, which results in higher productivity and lower unemployment rate in the state and thus a higher GSP. However, an increase in population growth also means that more resources, such as food and medical care are required for sustaining the community, especially for elderly since they will not be in the labour force due to retirement or health problems.

There have been studies investigating the relationship between population and economic growth of certain countries, while some suggest GDP is linearly dependent on population growth(Dao, 2014) and there is a positive correlation(Thuku, Paul & Almadi, 2013), other suggest the relationship between population and economic growth changed due to Industrial Revolution, which results in a negative correlation between population growth and economic growth(Weil, 2012). There are also studies stating that there is no long-run relationship between population growth and economic growth(Thornton, 2001).

Population growth acts as a double edged sword in economic development of a country. We believe that while population growth has both positive and negative impact on the economy and there exist many other factors affecting the growth of the economy(e.g. Unemployment rate, personal earning), population has a positive correlation with GSP in victoria.

Objective

This project aims to identify the relationship between population and economy of Victoria using the theme "*Using open data for understanding the liveability, inclusiveness, health and sustainability of communities in Victoria*".

We will be investigating the sustainability of communities in Victoria using the following research question:

> "*How does population growth affects economic growth in Victoria*"

By investigating this question, we hope policy makers and the government make use of the result of these projects and take the relationship between population and economy into account when considering new policy, and improve a community with greater sustainability and livability.

Our hypothesis in this paper is that there will be a positive correlation between population growth and economic growth in victoria. Economic growth has positive correlation with GSP, personal earning, and negative correlation with unemployment rate.

Dataset

The datasets used in this project are from.au, and there are a couple reasons we chose these datasets. Firstly, it is an authoritative website, which means that the source of the data is reliable and accurate. Secondly, by choosing datasets from the same website, they are more likely to have a similar format so we can spend less time on wrangling and pre-processing the datasets, and instead spend more time analysing and producing more meaningful graphs. Moreover, datasets from the same website often treat missing or invalid data the same way, thus we can have a uniform format across datasets even if there is missing or invalid data, which achieves consistency in our analysis.

Here's the datasets we are going to use:
1. [Australian National Accounts: State Accounts: Table1 - Gross state product, chain volume measures and current prices](#)
2. [National, state and territory population: Population - Victoria](#)
3. [Average weekly earnings, Victoria (dollars) - original](#)
4. [Australia Unemployment Rate](#)

Fortunately, the datasets have no error entries so no data cleaning method is required.

Wrangling

File Type

Both datasets are of xls file type, so it can be easily read into dataframe in python pandas.

Linking

Both datasets have similar format, and both have date as the first column, so we treated it as a common id for linking purposes. However, the data in the population datasets start from 1971 while the data in the GSP datasets start from 1991, since there is no GSP data from 1971-1990, we treat it as missing data and trim the 1971-1990 population data, to match the date in both datasets.

Data Type

In order to do calculations such as percentage change, pearson correlation and regression, the default data type (String) of the data needs to be converted to numeric data. We achieved this by using .astype and .to_datetime operations.

Analysis

First, we will look at the changes in population growth and GSP in 1991-2020
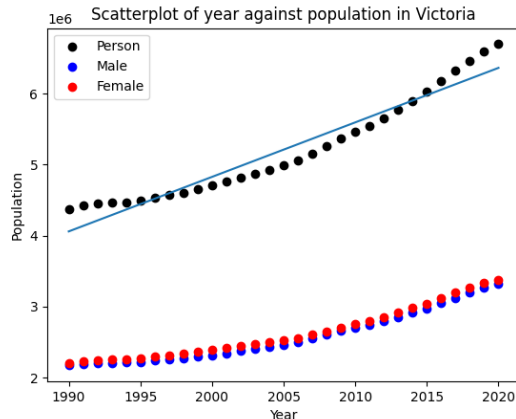
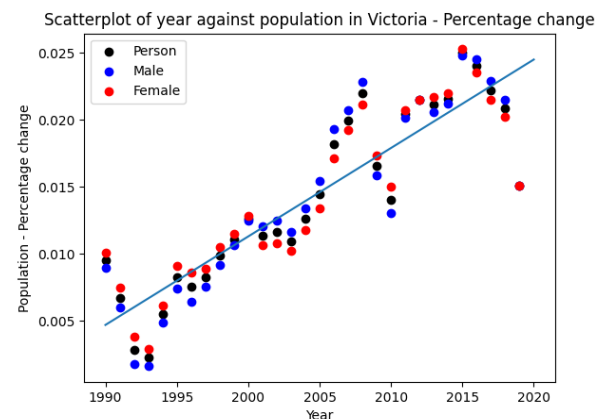# Population growth over time

Figure.1a

Figure.1b

Figure.1a and Figure.1b shows the population growth over the past years, we can see that the population growth is increasing at a higher rate starting from 2005 with a couple dips in 2010s, but overall the population is growing steadily from 1990 to 2020, and the best-fit line predicts that the population will continue to grow at a rapid rate in the future as well.

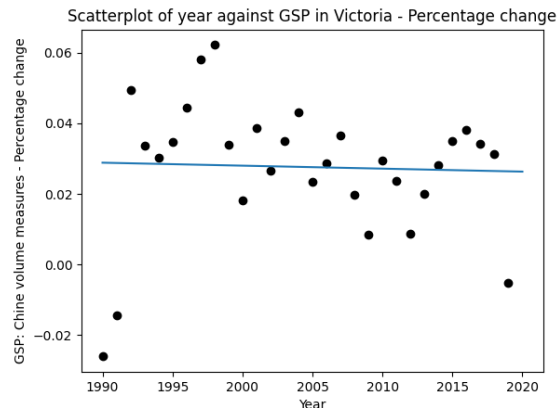# GSP changes over time

Figure.2a

Figure.2b

From Figure.2a we can see that the GSP increases steadily with no sudden drop or increase throughout the years, there is also not much change in GSP percentage change as shown in Figure.2b. In fact, the trend for percentage change of the GSP in Victoria is declining as represented by the best-fit line.
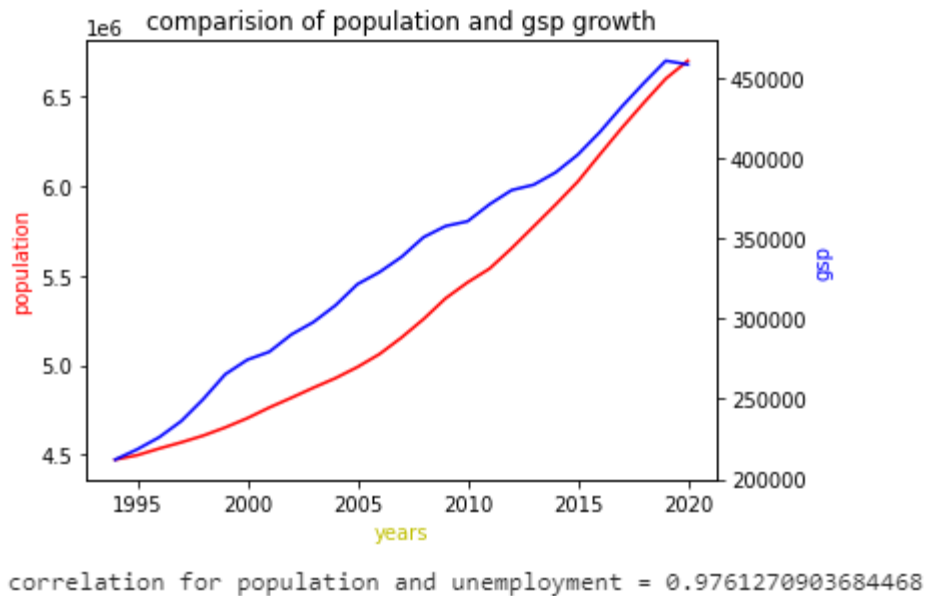
Population against GSP

correlation for population and unemployment = 0.9761270903684468

Figure 3a

Figure 3a shows that as population increases, GSP increases as well, and the pearson agrees with this assessment as well. The pearson correlation is calculated by numpy.corrcoef() function with population and GSP as input, the results shows 0.9761 which indicates a positive correlation, meaning the population does have a positive effect on GSP in Victoria.
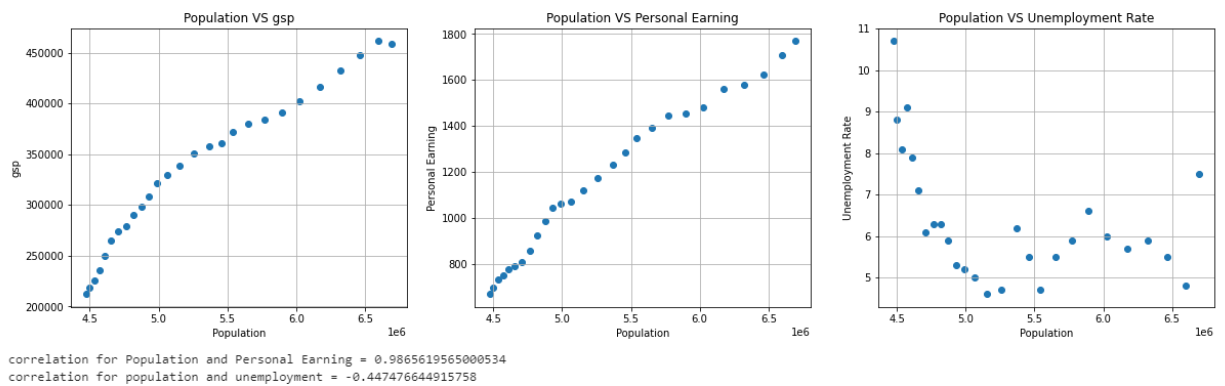

correlation for Population and Personal Earning = 0.9865619565000534
correlation for population and unemployment = -0.447476644915758

*Figure 4a (through plt.scatter() and .corr())*

Figure 4b

However, a good economy is not equal to high GSP. The personal earning is also an important factor to define whether an economic system is healthy, In Figure 4a, a strong correlation(0.986) between population and personal earning has been shown.Apart from that, the relationship between unemployment rate and population can also prove that population growth can stimulate the growth of the economy. As Alan S noted, a high unemployment rate can harm not only economy but also slow down the rate of economic growth. In this situation, a society with a healthy economy always comes up with a low unemployment rate. It can be seen from the 4a, as the population increases, the unemployment rate significantly decreases. Meanwhile, if we look at Figure 4a and 4b, which is a box plot produced by plt.boxplot(). it is easy to see that almost all the high unemployment rate(top 25 percentile) appear when population is lower than 5.0*10^6.

In order to make the data more reliable, we use machine learning to verify the relationship between gsp with population, unemployment rate and personal earning.

Firstly, We use linear regression to prove the predictable of these data.
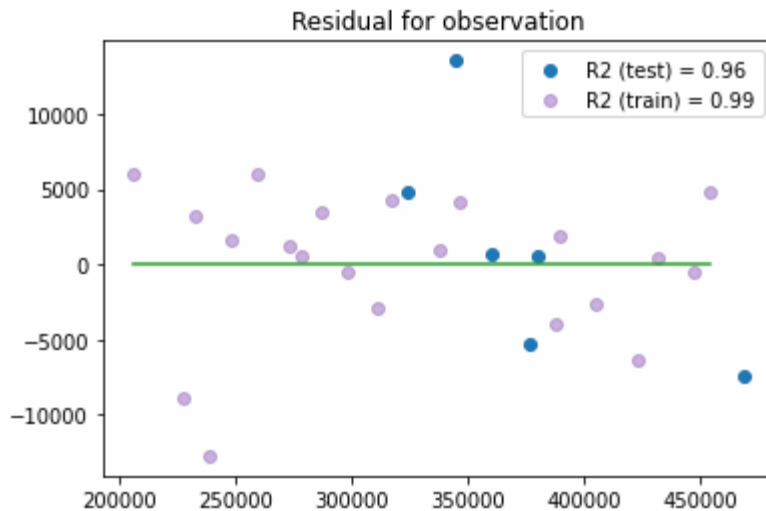
Figure 5a(using plt.scatter() to plot point x-axis is predict value y-axis stand for error value calculated by real value - predict value and use plt.plot() to add the green line)

As the graph above shows, the distribution of trained data is concentrated and linear(coefficient= 0.99). Which proves it is predictable.

```
actual gsp values of the first 5 test data:
        Economy
2012    380422
2011    371607
2006    329216
2019    461248
2009    357970

predicted gsp values of the first 5 test data:
[[379872.11918642]
 [376881.97666839]
 [324472.33646492]
 [468604.92382254]
 [344428.13848809]]

Coefficients:  [[ 6.38894463e-02 -9.29872774e+03  6.92432009e+01]]

Intercept:  [-26444.18645353]

Coefficient of determination (test): 0.97
Coefficient of determination (training): 1.00
```

Figure 5b

Then we can start to predict the gsp value and compare it with the actual value, predict value is produced by linear_model.LinearRegression() function from library sklearn can be seen from the

Figure 5b that the predicted gsp is very approaches to the actual value(mean error ±1.67%). In this situation, the reliability of data was confirmed and it can be used to predict the future data.
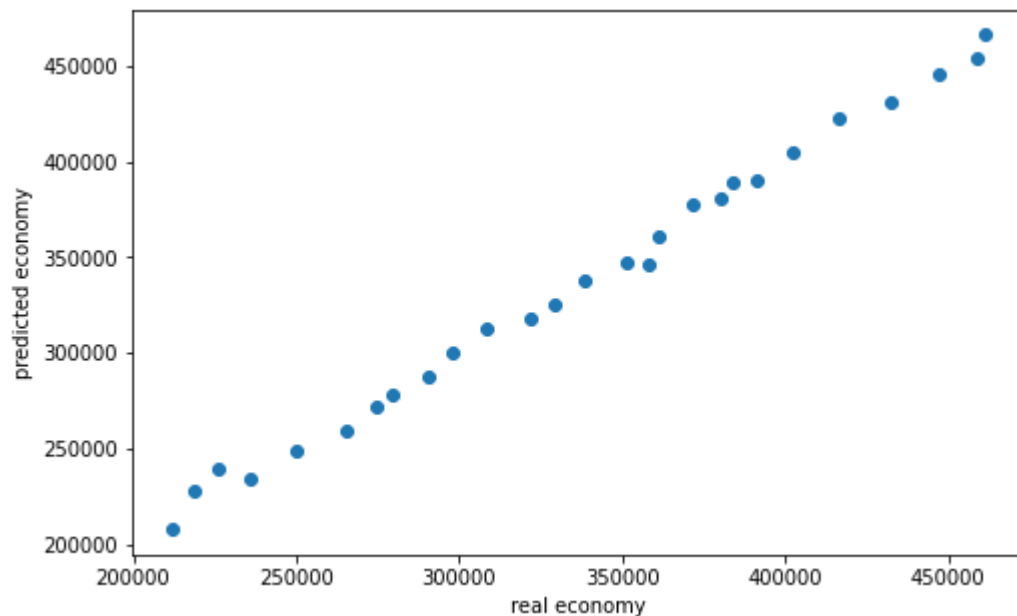

Figure 5c(real economy vs predicted economy)

Result

The results show that there exists a positive relationship between population and economic growth in Victoria. The positive correlation confirms that as long as the population increases, the GSP will most likely increase as well.

While percentage change in population and economic growth may contradict the result as the economic growth is declining and the population growth is increasing, this can be explained by other factors that affect economic growth, such as political issues or global issues. One such example is the covid-19, we can see from Figure 2b. that the last dot in the graph has a sudden drop, which represents the impact of covid-19 in the GSP of Victoria.

This result implies that given that there are enough resources to sustain the community, policy makers should not interfere with fertility rate and migration rate if they want improvement in the economic growth, sustainability and livability in Victoria.

Limitation

Economic growth is affected by many other factors, analysing solely the relationship between population and GSP, personal earning, unemployment rate may not be sufficient to produce a result that is applicable to the real world. This can be improved via analysing other factors such as political issues and technological development.

The result of this project is only applicable in Victoria since the source of the data is from Victoria, this not only means the limited usage of this project's result, but it also means that there is no other supporting data or paper from others, this can be resolved via analysing the relationship between population and economic growth in a larger scope.

Reference List

N. Kakwani (1993). Performance in living standards: An international comparison. Journal of
    Development Economics, Volume 41, Issue 2, pp. 307-336.
        doi:https://doi.org/10.1016/0304-3878(93)90061-Q

Dao MQ (2014), Population and Economic Growth in Developing Countries. Studies in
    Economics and Econometrics, Vol. 38, Issue 1, pp. 75-85.
        doi: https://doi.org/10.1080/10800379.2014.12097264

Thuku GK, Paul G, and Almadi O (2013), The Impact of Population Change on Economic
    Growth in Kenya, International Journal of Economics and Management Sciences,
    Vol. 2, No. 6, pp. 43-60.
        doi:10.4172/2162-6359.1000137

Weil, D.(2012).Economic Growth. Prentice Hall. Pearson

Thornton J (2001), Population Growth and Economic Growth: Long-run Evidence from Latin
    America, Southern Economic Journal, Vol. 62(2), pp. 464-468
        doi: 10.2307/1061606
Blinder, Alan S. "The Challenge of High Unemployment." From The American Economic
Review, Vol. 78, No. 2, pp. 1-15, (May 1988).