

**NANYANG TECHNOLOGICAL UNIVERSITY**

**EE5087 Living with Mathematics**

**REPORT**

**Topic of Project:**

**Income affecting Birth Rates**

Group Number: 7

Group Members:

Augustine Jeow Kai Loon (U1920288K)

Muhammad Azfar Nasri Bin Azman (U2222289H)

Singhal Raghav (U2023945J)

Tan Keng Kai Luke (U1921535H)

Tsai Pei Yu (U2223001F)

Table of Contents

[Introduction 3](#_Toc130923617)

[Data Collection 3](#_Toc130923618)

[Exploratory and Data Analysis 3](#_Toc130923619)

[Average Birth Rate 4](#_Toc130923620)

[Basic Mean Salary 4](#_Toc130923621)

[Average Birth Rate vs Basic Mean Salary 5](#_Toc130923622)

[Correlation between Average Birth Rate and Basic Mean Salary 6](#_Toc130923623)

[Mathematical Derivations 7](#_Toc130923624)

[Linear Regression 7](#_Toc130923625)

[Result of Linear Regression 7](#_Toc130923626)

[Observation from Data Analysis 9](#_Toc130923627)

[Additional Observations 10](#_Toc130923628)

[Birth Rate of Ethnic Groups vs Income Salary 10](#_Toc130923629)

[Birth rates 10](#_Toc130923630)

[Ethnic Birth rates and Income Salary 11](#_Toc130923631)

[Gross Salary Increment vs birth rates 12](#_Toc130923632)

[Conclusion 14](#_Toc130923633)

[References 15](#_Toc130923634)

# Introduction

In recent years, Singapore suffered a decline in birth rate. This can be due to various factors including employment related matters. Thus, this report aims to find a connection between newly graduates starting salary to the birth rate of Singapore. Linear Regression will be used to predict the birth rates Singapore might observe soon. Additionally, this report also aims to find other factors that might contribute to the decline of birth rate.

# Data Collection

There will be 2 datasets used for this project. Both are retrieved from https://data.gov.sg/. The mean of both Basic and Gross Salary will be compared with each other to determine if there is an increase of workload. The salary is then compared against the birth rate of that year to determine the correlation between them. We will also compare the different racial groups birth rates and compare it to the salary rate to further understand the relationship from the different factors.

# Exploratory and Data Analysis

One dataset contains data collected from the graduate employment survey over the years by the universities in Singapore. There are variables like employment rate by different degrees and universities, mean monthly salary and more. Another dataset contains the crude birth and death rate by ethnic groups over the years in Singapore.

Exploratory Data Analysis (EDA) is done on both the graduate employment survey dataset and the crude birth and death rate dataset. We will be comparing the data from the year 2013 to 2021. All findings will be done with Python and Machine learning, using packages like ‘numpy’, ‘pandas’, ‘seaborn’ and ‘matplotlib’.

## Average Birth Rate

This graph compares the average crude birth rate across all racial groups from 2013 to 2021. This demonstrates that Singapore has been having a decline in birth rates, which is accelerating. (Authority, 2022)

Chart, line chart

Description automatically generated

Figure 1: Average Birth Rate from 2013 to 2021

## Basic Mean Salary

From 2013 to 2021, this graph compares the average Basic Mean Salary of all graduates who participated in the graduate employment survey. This graph shows that the Basic Mean Salary has been gradually increasing from 2013 to 2021. (Education, 2022)

Chart, line chart

Description automatically generated

Figure 2: Basic Mean Salary from 2013 to 2021

## Average Birth Rate vs Basic Mean Salary

When comparing the graphs above, we can see that while the Basic Mean Salary increases, the birth rate falls from 2013 to 2021. Even during the year 2013-2014, when wages were declining, the fertility rate increased. This shows that there is some correlation between the two outcomes. (Authority, 2022) (Education, 2022)

Chart, line chart

Description automatically generated

Figure 3: Average Birth Rate Vs Basic Salary Mean

## Correlation between Average Birth Rate and Basic Mean Salary

To identify the strength of the relationship between the two outcomes, a correlation table is created. There is a significant inverse correlation between the two outcomes, with the correlation = -0.961479. This demonstrates the strong relationship between financial stability and the declining birth rates.

|  |  |
| --- | --- |
| **Correlation between Outcomes** | Basic Mean Salary |
| Average Crude Birth Rate | -0.961479 |

A picture containing website

Description automatically generated

Figure 4: Correlation Table between Average Birth Rate and Basic Mean Salary

# Mathematical Derivations

## Linear Regression

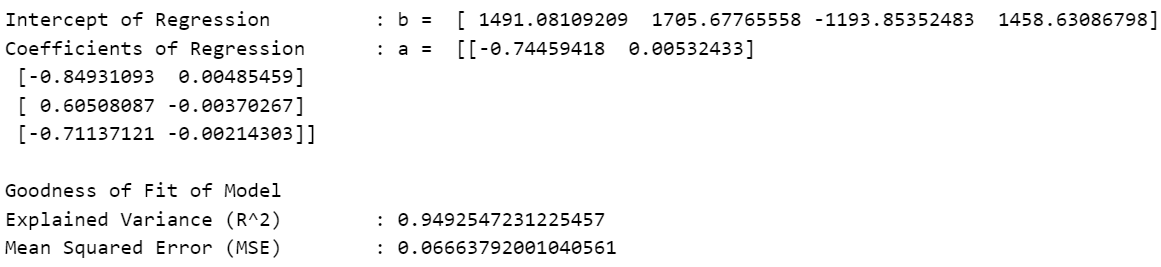
Linear regression analysis is used to predict the value of a variable based on the value of another variable. The formation of the regression equation is like the equation of the linear graph, , where is the dependent variable, is the gradient of the graph, is the independent variable, and is the -intercept of the graph. In the regression equation, is the estimated dependent variable score, is the regression coefficient, is the score on the independent variable, and is the regression intercept. Therefore, we can train the dataset to predict the value using the value. The type of linear regression we used in this project is the single linear regression with random train set and test set split.

To assess the difference between the observed and predicted values, we need to determine the regression line average error which is the Mean Squared Error (). To find out the , we take the observed values, , subtract which are the predicted values and square the differences. Then, sum all the squared values and divide by the number of observations, . As a result, a linear regression model with a lower Mean Square Error () is preferred because the smaller the , the closer we are to determine the best fit line. The formula for Mean Squared Error () is .

We also included another calculation to determine the goodness of fit of the linear regression model which is the Explained Variance (). This is like the purpose of calculating Mean Square Error () which is to measure the discrepancy between a model and actual data. A stronger strength of correlation is indicated by higher explained variance percentages. The formula for Explained Variance () is .

According to the EDA shown above, the basic salary of graduates influences the birth rate in Singapore. As a result, we utilised a linear regression model to forecast the dependent variable (birth rate) based on the independent variable (graduates' basic salary).

### Result of Linear Regression



The high Explained Variance () value and the low Mean Squared Error () value of the linear regression model indicates that there is a strong connection between the Singapore Birth Rate and the Basic Salary of the Graduates in Singapore.

# Observation from Data Analysis

Based on the analysis we have done; we can conclude that there is a steady decrease in the Average Birth Rate in Singapore and a steady increase in the Basic Mean Salary in Singapore. In addition, we identified that there is a strong inverse correlation between the Average Birth Rate and the Basic Mean Salary in Singapore, as seen by the correlation coefficient of -0.96. In this section, we aim to identify reasons to account for these trends.

Firstly, this inverse correlation could be explained by the increase in the proportion of highly educated Singaporeans. Based on a report released by the Singapore Department of Statistics, the number of Singaporeans aged 25 years and older with the highest education qualification of University increased from slightly over 750,000 in 2015 to over a million in 2021 (Singapore D. o., 2023).This increase in proportion of highly educated Singaporeans results in an increase in the expected salary of Singaporeans. This trend also causes Singaporeans to be getting married at a later age, as most would choose to complete their education before tying the knot. This leads to an increase in the average childbearing age in Singapore and consequently, fewer children in younger families.

In addition, rising inflation and the recent increase in GST from 7% to 8% has resulted in a rising cost of living in Singapore. Based on a survey done by Human Resources Online, following the rise in inflation in recent years, almost half (48.95%) will ask their employers for a pay rise, and a similar number (47.25%) will search for a better-paying job this year. (Sofiah, 2023) These pressures on employers will consequently result in an increase in the average salary of Singaporeans. Rising costs of living also make it more difficult and costly for families to have children. According to NUS economists in 2018, the cost to raise a child in Singapore is estimated to be between S$280,000 and S$560,000, depending on household income. (Singapore S. , 2021)With rising inflation, these numbers are expected to increase and would act as a further deterrent for families to have children.

In general, we can summarise by saying that this inverse correlation between the Average Birth Rate and the Basic Mean Salary in Singapore can be explained by the increase in the proportion of highly educated Singaporeans and the rising cost of living in Singapore.

# Additional Observations

## Birth Rate of Ethnic Groups vs Income Salary

### Birth rates

Chart, line chart

Description automatically generated

Figure 5: Birth Rates of different Ethnic Groups

We investigate the birth rates of various racial groups in Singapore using the idea of how higher income salary inversely relates to declining birth rates. According to the graph, all races, except for Malays, are witnessing a decrease in birth rates.

### Ethnic Birth rates and Income Salary

Chart, line chart

Description automatically generated

Figure 6: Birth Rates of different Ethnic Groups vs Basic Salary Mean

According to the graph above, the Malay ethnic group is the only one with an increasing trend in birth rates. Using our original analysis of how the increase in basic salary mean is inversely proportional to birth rates, it is reasonable to infer that this ethnic group earns less than the other ethnic groups. Another belief is that fewer Malay ethnic groups attend university to further their education. These two assumptions are also limited by cultural differences and behaviours that may influence ethnic group birth rates.

## Gross Salary Increment vs birth rates

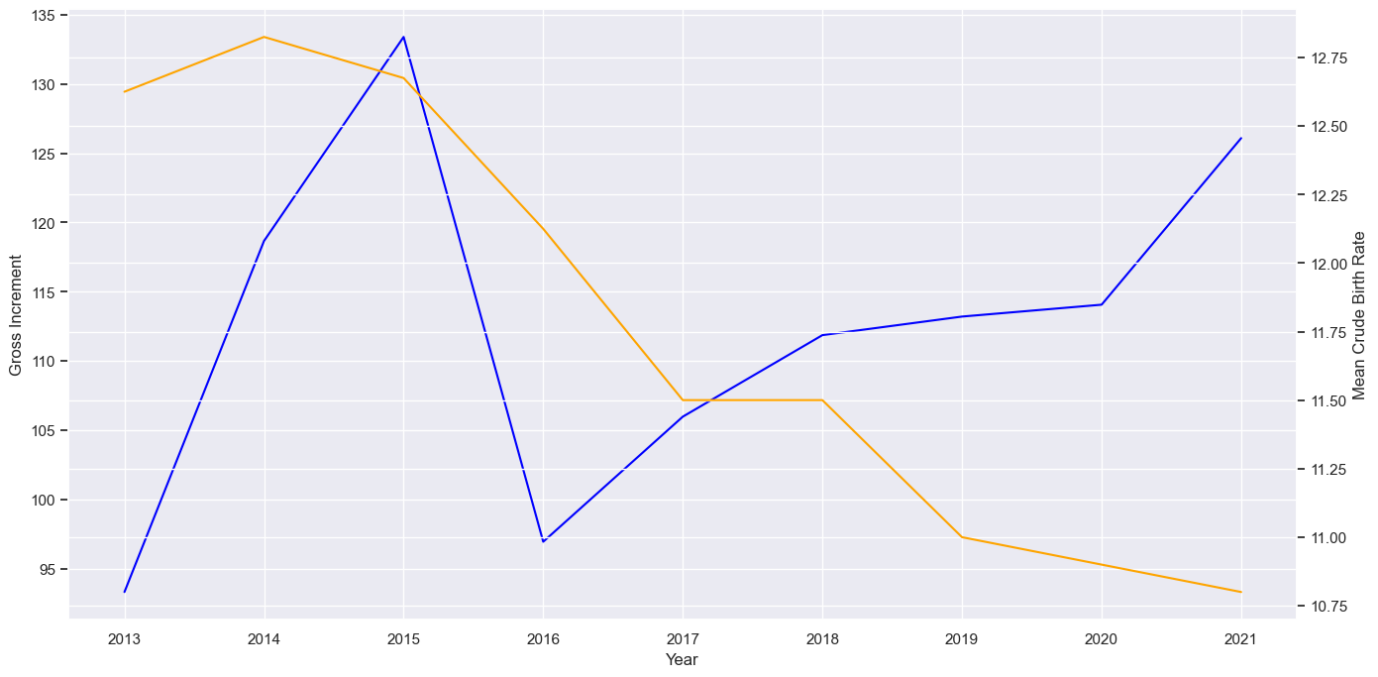


Figure 7: Gross Salary Increment vs Mean Crude Birth Rate

A change in salary is just one of several variables that may affect people's decisions about family planning; birth rates are influenced by a wide range of social, cultural, economic, and political factors.

According to the graphs, in comparison between gross salary increment and birth rates, there is a correlation that birth rates are decreasing as when the gross salary are increasing. In 2015, Singapore’s GDP growth was the slowest since 2009, and the world economy was slowing down, which may cause the reduction in gross salary increment.

While considering whether to have children, people in Singapore may consider a wide range of considerations, including career objectives, educational aspirations, family values, and financial security. Although it might influence their choices, pay increases are unlikely to be the only factor influencing birth rates.

Moreover, a fall in overall income levels may not always follow a decline in pay increases. For instance, a person may have a stable work and a generally high income even while their wage increase drops one year. Singapore also boasts a very robust social safety net, with laws and initiatives supporting families and kids, including financial aid for daycare, healthcare, and education. These measures might lessen the effect on birth rates of brief economic downturns.

According to local studies, many single people want to get married someday but prefer to focus on their school or careers over dating. Most married couples have children; however, the majority only have one or two due to the high costs associated with raising children and the desire to devote more time to each child. Couples who might normally want kids express worries about the morals of a demanding childhood and upbringing or that they wouldn't have the energy or skills to support their kids well enough to compete.

While wage increment may have an impact on Singapore's birth rates, it is unlikely to be the only factor, and it might not have much of an effect on birth rates by itself. When examining trends in birth rates and their determinants, it is crucial to take a variety of factors into account.

# Conclusion

In conclusion, although the relationship between the two, income and birth rate, is complicated and influenced by a variety of circumstances, wealth can have a major impact on birth rates. Although having more money might give people and families the means to support having children, wealth is not the sole factor that affects birth rates. Birth rates can also be significantly influenced by cultural perspectives on family planning, access to healthcare and education, and work-life balance. Likewise, various racial or socioeconomic groups may have distinct effects of economic downturns or income inequality on birth rates. It is crucial to consider a variety of policies and programs that address these variables and foster a supportive environment for families and kids to address declining birth rates and encourage healthy population growth.

# References

Authority, M. o.-I. (22 August, 2022). *Crude Birth, Death & Natural Increase Rates by Ethnic Group*. Retrieved from data.gov.sg: https://data.gov.sg/dataset/crude-death-rates-by-ethnic-group-from-1971-onwards?view\_id=61a6e3af-03dc-4e66-88a9-8726d372925f&resource\_id=1c8ea8a5-3f1f-43be-ad36-3e510adbd2d3

Education, M. o. (21 November, 2022). *Graduate Employment Survey - NTU, NUS, SIT, SMU, SUSS & SUTD*. Retrieved from data.gov.sg: https://data.gov.sg/dataset/graduate-employment-survey-ntu-nus-sit-smu-suss-sutd

Singapore, D. o. (28 March, 2023). *Department of Statistics Singapore*. Retrieved from Education and Literacy: https://www.singstat.gov.sg/publications/reference/ebook/population/education-and-literacy

Singapore, S. (23 November, 2021). *How Much Does It Cost To Raise A Child In Singapore?* Retrieved from Syfe: https://www.syfe.com/magazine/how-much-does-it-cost-to-raise-a-child-in-singapore/

Sofiah, A. (13 January, 2023). *Human Reseources Online.net*. Retrieved from Inflation likely to push about 49% of Singapore workers surveyed to ask for a salary increase: https://www.humanresourcesonline.net/inflation-likely-to-push-about-49-of-singapore-workers-surveyed-to-ask-for-a-salary-increase

# Appendix

# Essential Libraries

|  |
| --- |
| **import** numpy **as** np **import** pandas **as** pd **import** seaborn **as** sb  **import** matplotlib.pyplot **as** plt *# we only need pyplot* sb**.**set() |

In [21]:

|  |
| --- |
| birth\_death\_rate\_data **=** pd**.**read\_csv('crude-birth-death-natural-increase-rates-by-et graduate\_salary\_data **=** pd**.**read\_csv('graduate-employment-survey-ntu-nus-sit-smu-suss |

In [22]:

# Singapore Birth Rate over the years

|  |
| --- |
| birth\_rate **=** pd**.**DataFrame(birth\_death\_rate\_data[['crude\_birth\_rate','year']]) birth\_rate\_above\_2012 **=** birth\_rate[birth\_rate['year'] **>=** 2013]  mean\_birth\_rate\_per\_year **=** {} **for** x **in** birth\_rate\_above\_2012['year']:  mean\_birth\_rate **=** birth\_rate\_above\_2012**.**query('year == ' **+** str(x))['crude\_birth mean\_birth\_rate\_per\_year[x] **=** mean\_birth\_rate df\_mean\_birth\_rate\_per\_year **=** pd**.**DataFrame(mean\_birth\_rate\_per\_year**.**items(), column  f **=** plt**.**figure(figsize**=**(16, 8))  sb**.**lineplot(data **=** df\_mean\_birth\_rate\_per\_year, x **=** 'Year', y **=** 'Mean Crude Birth R |

In [23]:

<AxesSubplot: xlabel='Year', ylabel='Mean Crude Birth Rate'> Out[23]:

Chart, line chart

Description automatically generated

# Fresh Graduate Salary over the Years

|  |
| --- |
| basic\_salary\_rate **=** pd**.**DataFrame(graduate\_salary\_data[['year','basic\_monthly\_mean' basic\_salary\_rate **=** basic\_salary\_rate[basic\_salary\_rate["basic\_monthly\_mean"]**.**str**.**  mean\_basic\_salary\_per\_year **=** {} **for** x **in** basic\_salary\_rate['year']:  mean\_basic\_salary\_rate **=** basic\_salary\_rate**.**query('year == ' **+** str(x))['basic\_mo mean\_basic\_salary\_per\_year[x] **=** mean\_basic\_salary\_rate print(mean\_basic\_salary\_per\_year) |

In [24]:

c

df\_mean\_basic\_salary\_per\_year **=** pd**.**DataFrame(mean\_basic\_salary\_per\_year**.**items(), co

f **=** plt**.**figure(figsize**=**(16, 8)) sb**.**lineplot(data **=** df\_mean\_basic\_salary\_per\_year, x **=** 'Year', y **=** 'Mean Basic Salar

{2013: 3254.883116883117, 2014: 3250.233009708738, 2015: 3358.0091743119265, 2016:

3431.119266055046, 2017: 3508.278260869565, 2018: 3567.198347107438, 2019: 3744.5

2, 2020: 3803.517985611511, 2021: 3934.6716417910447}

<AxesSubplot: xlabel='Year', ylabel='Mean Basic Salary'> Out[24]:

Chart, line chart

Description automatically generated

# Displaying both Birth Rates and Graduate Salary from 2013 to 2021

|  |
| --- |
| fig**=**plt**.**figure(figsize**=**(16,8)) ax1 **=** fig**.**add\_subplot(111) ax2 **=** ax1**.**twinx()  jointDF\_birthRate\_basicSalary **=** pd**.**merge(df\_mean\_birth\_rate\_per\_year,df\_mean\_basic\_  *# changes here*  sb**.**lineplot(x **=** 'Year', y **=** 'Mean Crude Birth Rate',data**=**jointDF\_birthRate\_basicSal sb**.**lineplot(x **=** 'Year', y **=** 'Mean Basic Salary' ,data**=**jointDF\_birthRate\_basicSalary  *# and here*  plt**.**xticks(rotation**=**60) |

In [25]:

(array([2012., 2013., 2014., 2015., 2016., 2017., 2018., 2019., 2020., Out[25]:

2021., 2022.]),

[Text(2012.0, 0, '2012'), Text(2013.0, 0, '2013'),

Text(2014.0, 0, '2014'),

Text(2015.0, 0, '2015'), Text(2016.0, 0, '2016'),

Text(2017.0, 0, '2017'),

Text(2018.0, 0, '2018'),

Text(2019.0, 0, '2019'), Text(2020.0, 0, '2020'),

Text(2021.0, 0, '2021'),

Text(2022.0, 0, '2022')])

Chart, line chart

Description automatically generated

# Correlation Between Graduate Salary vs Birth Rate

In [26]: fig**=**plt**.**figure(figsize**=**(16,8)) sb**.**heatmap(jointDF\_birthRate\_basicSalary**.**corr(), vmin **=** **-**1, vmax **=** 1, annot **=** **True**

<AxesSubplot: > Out[26]:

A picture containing website

Description automatically generated

In [27]: jointDF\_birthRate\_basicSalary**.**corr()

Out[27]: **Year Mean Crude Birth Rate Mean Basic Salary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | 1.000000 | -0.962903 | 0.986401 |
| **Mean Crude Birth Rate** | -0.962903 | 1.000000 | -0.961479 |
| **Mean Basic Salary** | 0.986401 | -0.961479 | 1.000000 |

# Increment in Graduate Gross Salary

This would mean that there is more overtime pay over the years.

In [28]:

c

i

|  |
| --- |
| gross\_salary\_rate **=** pd**.**DataFrame(graduate\_salary\_data[['year','gross\_monthly\_mean' gross\_salary\_rate **=** gross\_salary\_rate[gross\_salary\_rate["gross\_monthly\_mean"]**.**str**.**  mean\_gross\_salary\_per\_year **=** {} **for** x **in** gross\_salary\_rate['year']:  mean\_gross\_salary\_rate **=** gross\_salary\_rate**.**query('year == ' **+** str(x))['gross\_m mean\_gross\_salary\_per\_year[x] **=** mean\_gross\_salary\_rate  *#print( mean\_gross\_salary\_per\_year)*  gross\_increment\_salary\_per\_year **=** {} **for** x **in** mean\_gross\_salary\_per\_year:  gross\_increment **=** ( mean\_gross\_salary\_per\_year[x] **-** mean\_basic\_salary\_per\_year gross\_increment\_salary\_per\_year[x] **=** gross\_increment  *#print( gross\_increment\_salary\_per\_year)* df\_gross\_increment\_salary\_per\_year **=** pd**.**DataFrame(gross\_increment\_salary\_per\_year**.**  fig**=**plt**.**figure(figsize**=**(16,8)) ax1 **=** fig**.**add\_subplot(111) ax2 **=** ax1**.**twinx()  jointDF\_grossIncrement\_meanSalary **=** pd**.**merge(df\_gross\_increment\_salary\_per\_year,  *# changes here*  sb**.**lineplot(x **=** 'Year', y **=** 'Gross Increment',data**=**jointDF\_grossIncrement\_meanSalar sb**.**lineplot(x **=** 'Year', y **=** 'Mean Basic Salary' ,data**=**jointDF\_grossIncrement\_meanSa  *# and here*  plt**.**xticks(rotation**=**60) |

df\_

(array([2012., 2013., 2014., 2015., 2016., 2017., 2018., 2019., 2020., Out[28]:

2021., 2022.]),

[Text(2012.0, 0, '2012'), Text(2013.0, 0, '2013'),

Text(2014.0, 0, '2014'),

Text(2015.0, 0, '2015'), Text(2016.0, 0, '2016'),

Text(2017.0, 0, '2017'),

Text(2018.0, 0, '2018'),

Text(2019.0, 0, '2019'), Text(2020.0, 0, '2020'),

Text(2021.0, 0, '2021'),

Text(2022.0, 0, '2022')])

Chart, line chart

Description automatically generated

|  |
| --- |
| jointDF\_grossIncrement\_birthRate **=** pd**.**merge(df\_mean\_birth\_rate\_per\_year, df\_gross\_i jointDF\_grossIncrement\_birthRate**.**corr() |

In [29]:

Out[29]: **Year Mean Crude Birth Rate Gross Increment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | 1.000000 | -0.962903 | 0.326838 |
| **Mean Crude Birth Rate** | -0.962903 | 1.000000 | -0.116502 |
| **Gross Increment** | 0.326838 | -0.116502 | 1.000000 |

# Birth Rate by Ethnic Groups from 2013 to 2021

In [30]: race\_birth\_rate **=** pd**.**DataFrame(birth\_death\_rate\_data[['year', 'ethnic\_group', 'crud race\_birth\_rate\_above\_2012 **=** race\_birth\_rate[race\_birth\_rate['year'] **>=** 2013]

df\_race\_birth\_rate\_above\_2012 **=** race\_birth\_rate\_above\_2012**.**pivot(index**=**'year', colu print(df\_race\_birth\_rate\_above\_2012) df\_race\_birth\_rate\_above\_2012**.**plot(color**=**['red', 'green', 'blue', 'orange'], figsiz

ethnic\_group Chinese Indians Malays Others year 2013 8.2 11.0 11.7 19.6

1. 8.8 10.8 12.4 19.3
2. 8.6 10.9 13.3 17.9
3. 8.2 9.9 13.5 16.9 2017 7.7 9.5 13.7 15.1
4. 7.5 9.2 14.1 15.2
5. 7.6 8.7 14.0 13.7
6. 7.2 8.3 14.2 13.9
7. 7.3 8.7 14.4 12.8

<AxesSubplot: xlabel='year'> Out[30]:

Chart, line chart

Description automatically generated

# Birth Rate by Ethnic Groups vs Graduate Salary

|  |
| --- |
| df2\_mean\_basic\_salary\_per\_year **=** pd**.**DataFrame(mean\_basic\_salary\_per\_year**.**items(),  *# changes here*  fig**=**plt**.**figure(figsize**=**(16,8)) ax1 **=** fig**.**add\_subplot(111) ax2 **=** ax1**.**twinx() jointDF\_birthRate\_meanSalary **=** pd**.**merge(race\_birth\_rate\_above\_2012,df2\_mean\_basic\_s |

In [31]:c sb**.**lineplot(x **=** 'year', y **=** 'Mean Basic Salary' ,data**=**jointDF\_birthRate\_meanSalary, sb**.**lineplot(x **=** 'year', y **=** 'crude\_birth\_rate',data**=**jointDF\_birthRate\_meanSalary,ax

<AxesSubplot: xlabel='year', ylabel='crude\_birth\_rate'> Out[31]:

Chart, line chart

Description automatically generated

# Linear Regression of Birth Rate

|  |
| --- |
| *# Import essential models and functions from sklearn* **from** sklearn.linear\_model **import** LinearRegression **from** sklearn.model\_selection **import** train\_test\_split **from** sklearn.metrics **import** mean\_squared\_error basicSalary\_train, basicSalary\_test, birthRate\_train, birthRate\_test **=** train\_test\_s  *# Linear Regression using Train Data*  linreg **=** LinearRegression() *# create the linear regression object* linreg**.**fit(basicSalary\_train, birthRate\_train) *# train the linear regression*  *# Coefficients of the Linear Regression line*  print('Intercept of Regression \t: b = ', linreg**.**intercept\_) print('Coefficients of Regression \t: a = ', linreg**.**coef\_) print()  *# Predict Birth Rate values corresponding to Basic Salary* birthRate\_train\_pred **=** linreg**.**predict(basicSalary\_train)  *# Check the Goodness of Fit (on Train Data)* print("Goodness of Fit of Model")  print("Explained Variance (R^2) \t:", linreg**.**score(basicSalary\_train, birthRate\_tra print("Mean Squared Error (MSE) \t:", mean\_squared\_error(birthRate\_train, birthRate print() |

In [61]:

|  |  |
| --- | --- |
| Intercept of Regression  83 1458.63086798] | : b = [ 1491.08109209 1705.67765558 -1193.853524 |
| Coefficients of Regression  [-0.84931093 0.00485459]  [ 0.60508087 -0.00370267]  [-0.71137121 -0.00214303]] Goodness of Fit of Model | : a = [[-0.74459418 0.00532433] |
| Explained Variance (R^2) | : 0.9492547231225457 |
| Mean Squared Error (MSE) | : 0.06663792001040561 |