# Annex F – Data Preparation & Modelling Report (20%)

**(To be completed individually)**

|  |  |
| --- | --- |
| **Module Group:** | ITBW21-02 |
| **Name:** | **Chee Bo Yu** |

#### Data Preparation & Modelling Processes

#### Scope of datasets

#### List down all possible data sources with description, including those that have been studied and explain why a dataset is included for further analysis or excluded

|  |  |
| --- | --- |
| **Name and Link** | **Reason** |
| **Students and Teachers under MOE**  (The total number of students and education officers under the Ministry of Education.)**Database Name:**Primary\_Students\_and\_TeachersSecondary\_Students\_and\_Teachers**Link:**[**Students and Teachers under MOE**](https://data.gov.sg/dataset/students-and-teachers-in-schools?view_id=f1cd1180-d064-43ce-a10a-67c9beac7900&resource_id=22600569-848a-40af-af4c-b6f60c35c711) | To find out if the government is giving **students** the possibility for **a better future**.I only use **primary and secondary** data as I am only interested in finding the number of teachers in primary and secondary as it is the **minimum education** a student needs. |
| **Percentage of N-level Cohort that Progressed to Post-Secondary Education**(Percentage of N-Level Cohort progress to post-secondary that are classified by Race and Year)**Database Name:**Percentage\_of\_N-level\_Cohort**Link:**[**N-Level Cohort Progress to Post Secondary Education**](https://data.gov.sg/dataset/percentage-of-n-level-cohort-that-progressed-to-post-secondary-education) | To find out if Singapore's education system brings out the best in every **N-Level student** by looking at the **percentage per year**. |

|  |  |
| --- | --- |
| **Percentage of O-Level Cohort that Progressed to Post-Secondary Education** (Percentage O-Level Cohort progress to post-secondary that are classified by Race and Year)**Database Name:**Percentage\_of\_O-level\_Cohort**Link:**[**Percentage of O-Level Cohort that Progressed to Post-Secondary Education-Data.gov.sg**](https://data.gov.sg/dataset/percentage-of-o-level-cohort-that-progressed-to-post-secondary-education) | To find out if Singapore's education system brings out the best in every **O-Level student** by looking at the **percentage per year**. |
| **Intake, Enrolment and Graduates by Institutions** (The number of full-time intakes, enrolment and graduates in Post-Secondary Education Institutions.)**Database Name:** Enrolment\_by\_institutions**Link:**[**Intake, Enrolment and Graduates by Institutions**](https://data.gov.sg/dataset/intake-enrolment-and-graduates-by-institutions) | It is requested by our target user to find out the **number of enrolments by institutions** and it will also show if the student of that year prefers which pathways.I only use **enrolment** data as I am only interested in finding the **enrolment of the students** in each institution. |
| **Enrolment for Kindergartens** (The number of enrolments in kindergartens.)**Database Name:** Enrolment\_moe\_kindergartens**Link:** [**Enrolment for Kindergartens**](https://data.gov.sg/dataset/enrolment-moe-kindergartens?view_id=7cddea4b-df86-40bf-a685-a7d9c5573d5f&resource_id=4ad866a7-c43a-4645-87fd-fc961c9de78a) | To find out **kindergartens intake** and if parents are willing to let their children study in Singapore by looking at the **enrolment per year**. |

|  |  |
| --- | --- |
| **Enrolment for Secondary School**(The number of enrolments in Secondary School.)**Database Name:**Enrolment\_secondary\_by\_level\_and\_ course**Link:** [**Enrolment for Secondary School**](https://data.gov.sg/dataset/enrolment-secondary-by-level?view_id=9fc5a148-53cd-4cf6-8eee-add13338e1d6&resource_id=e30f699f-a246-4016-8bba-d83366da46c6) | To find out each **stream** in secondary school intake for the student studying in Singapore by looking at the **enrolment per year**. |
| **Enrolment for Primary School**(The number of enrolments in primary school)**Database Name:**Enrolment\_primary**Link:**[**Enrolment for Primary School**](https://data.gov.sg/dataset/enrolment-primary-by-level?view_id=8995c187-610a-4493-af8c-f7afad9158bb&resource_id=b4b01dcd-9262-44e6-bf60-6d0126b7b2bb) | To find out about **Primary school enrolment** for the student studying in Singapore by looking at the **enrolment per year**.I only use **primary one** data as I am only interested in finding the enrolment of the student entering primary school at the age of **7** |
| **Enrolment for Pre-University**(The number of enrolments in Pre- University School)**Database Name:**Enrolment\_pre-university**Link:**[**Enrolment for Pre-University**](https://data.gov.sg/dataset/enrolment-pre-university-by-level) | To find out about the **number of intakes of Pre-University schools** compared to **Polytechnic.** |
| **Enrolment for University**(The number of enrolments in University School)**Database Name:**Universities-enrolment**Link:****[Universities - Intake, Enrolment and Graduates by Course-Data.gov.sg](https://data.gov.sg/dataset/universities-intake-enrolment-and-graduates-by-course)** | To find out about the **number of intakes of university schools** in Singapore by looking at the **enrolment per year**. |
| **Graduate Employment Survey for Polytechnic Student**(The statistics show Polytechnic student’s employment rates and gross monthly salary)**Database Name:**Polytechnic\_graduate\_survey**Link:**[**Graduate Employment Survey for Polytechnic Student**](https://www.todayonline.com/singapore/over-90-polytechnic-graduates-seeking-work-got-jobs-within-6-months-graduating-2021-survey-1789956) | To find out about the **employment rates and gross monthly salary** for a polytechnic student. |
| **Graduate Employment Survey for University Student**(The statistics show University student’s employment rates and gross monthly salary)**Database Name:**Graduate\_employment\_survey\_ntu-nus-sit-smu-suss-sutd**Link:** [**Graduate Employment Survey for University Student**](https://data.gov.sg/dataset/graduate-employment-survey-ntu-nus-sit-smu-suss-sutd) | To find out about the **employment rates, gross monthly salary** for a university student. |

#### Is the dataset shared or common?’

|  |
| --- |
| This dataset is **not shared** as each of my team members is doing an individual component. |

#### How is the dataset being used? e.g., selected fields or all fields?

|  |
| --- |
| For **Enrolment\_secondary\_by\_level\_and\_ course**, I have selected **course,enrolment\_secondary** and **year columns** to create the visualization in power bi.For **Enrolment\_primary**, I have selected **enrolment\_pri1** and **year columns** to create the visualization in power bi.For **Enrolment\_by\_institutions**, I have selected **ite and year columns** to create the visualization in power bi.For **Polytechnics\_information,** I have selected **course, enrolment, graduates, intake and year columns** to create the visualization in power bi.For **Enrolment\_pre-university**, I have selected **enrolment\_preu, level and year columns** to create the visualization in power bi.For **Universities-enrolment,** I have selected **course, enrolment and year columns** to create the visualization in power bi.For **Percentage\_of\_N-level\_Cohort,** I have selected **all the fields** to create the visualization in power bi.For **Percentage\_of\_O-level\_Cohort,** I have selected **all the fields** to create the visualization in power bi.For **Primary\_Students\_and\_Teachers,** I have selected **all the fields** except **sex and student\_pri columns** to create the visualization in power bi.For **Secondary\_Students\_and\_Teachers,** I have selected **all the fields** except **sex and student\_sec columns** to create the visualization in power bi.For **Polytechnic\_graduate\_survey,** I have selected **all the fields** to create the visualization in power bi.For **Graduate\_employment\_survey\_ntu-nus-sit-smu-suss-sutd,** I have selected **all the fields** expect **basic\_monthly\_median, employment\_rate\_overall, gross\_monthly 25 and 75\_percentile, gross\_monthly\_mean and median** to create the visualization in power bi.My **datasets** include Enrolment\_by\_institutions, Enrolment\_moe\_kindergartens, Enrolment\_pre-university, Enrolment\_primary, Enrolment\_secondary\_by\_level\_and\_ course, Graduate\_employment\_survey\_ntu-nus-sit-smu-suss-sutd, Percentage\_of\_N-level\_Cohort, Percentage\_of\_O-level\_Cohort, Polytechnic\_graduate\_survey, Polytechnics\_information, Primary\_Students\_and\_Teachers, Secondary\_Students\_and\_Teachers and Universities-enrolment |

#### Data Preparation

#### Describe the data cleaning and transformation done

|  |  |
| --- | --- |
| **Old Database Name** | **New Database Name** |
| enrolment-by-institutions | Enrolment\_by\_institutions |
| enrolment-moe-kindergartens | Enrolment\_moe\_kindergartens |
| enrolment-primary-one | Enrolment\_primary |
| enrolment-secondary-by-level-and-course | Enrolment\_secondary\_by\_level\_and\_ course |
| enrolment-pre-university-by-level-and-course | Enrolment\_pre-university |
| universities-intake-enrolment-and-graduates-by-course | Universities-enrolment |
| graduate-employment-survey-ntu-nus-sit-smu-suss-sutd | Graduate\_employment\_survey\_ntu-nus-sit-smu-suss-sutd |
| polytechnics-intake-enrolment-and-graduates-by-course | Polytechnics\_information |
| intake-by-institutions | Intake\_by\_institutions |
| universities-intake-enrolment-and-graduates-by-course | Universities-enrolment |
| Percentage of N-level Cohort | Percentage\_of\_N-level\_Cohort |
| Percentage of O-Level Cohort | Percentage\_of\_O-Level\_Cohort |
| Poly\_graduate\_survey | Polytechnic\_graduate\_survey |
| students-and-teachers-primary-schools | Primary\_Students\_and\_Teachers |
| Students and Teachers under MOE | Secondary\_Students\_and\_Teachers |
| students-and-teachers-pre-university | Pre-university\_students\_and\_teachers |

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **New Column Name** | **Customization** | **Database Name** |
|  |  | Sorting Year Column by descending | Enrolment\_secondary\_by\_level\_and\_ course |
|  |  | Filter year from 2016 to 2020 | Enrolment\_secondary\_by\_level\_and\_ course |
|  |  | Filter year from 2016 to 2020 | Polytechnics\_intake\_enrolment\_and\_graduates-by-course |
|  |  | Filter year from 2016 to 2020 | Enrolment\_primary |
|  |  | Sorting Year Column by descending | Enrolment\_ by\_ institutions |
|  |  | Filter year from 2016 to 2020 | Enrolment\_ by\_ institutions |
|  |  | Sorting Year Column by descending | Enrolment\_ moe\_ kindergartens |
|  |  | Filter year from 2018 to 2022 | Enrolment\_ moe\_ kindergartens |
|  |  | Filter year from 2016 to 2020 | Enrolment\_pre-university |
|  |  | Remove null value in enrolment\_preu | Enrolment\_pre-university |
|  |  | Filter year from 2016 to 2020 | Universities-enrolment |
|  |  | Sorting Year Column by descending | Graduate\_ employment\_survey\_ntu-nus-sit-smu-suss-sutd |

|  |  |  |  |
| --- | --- | --- | --- |
| employment\_rate\_ft\_permen | employment\_ rate\_ full\_ time\_ permanent | Renaming Column Name | Graduate\_ employment\_survey\_ntu-nus-sit-smu-suss-sutd |
|  |  | Round to 2 Decimal Place inemployment\_rate\_full\_time\_permenant | Graduate\_ employment\_survey\_ntu-nus-sit-smu-suss-sutd |
| gross\_ mthly\_25\_percentile | gross\_ monthly\_25\_ percentile | Renaming Column Name | Graduate\_ employment\_survey\_ntu-nus-sit-smu-suss-sutd |
| gross\_ mthly\_75\_percentile | gross\_ monthly\_75\_ percentile | Renaming Column Name | Graduate\_ employment\_survey\_ntu-nus-sit-smu-suss-sutd |
|  |  | Filter year from 2015 to 2019 | Graduate\_employment\_survey\_ntu-nus-sit-smu-suss-sutd |
|  |  | Sort year column by descending | Primary\_ Students\_ and\_Teachers |
|  |  | Sort year column by descending | ite\_intake\_ enrolment\_ and\_ graduates-by\_course |
|  |  | Filter year from 2015 to 2019 | Percentage\_of\_N-level\_Cohort |
| percentage\_prog\_ postsec | percentage\_progress\_postsec | Renaming Column Name | Percentage\_of\_N-level\_Cohort |
|  |  | Round to 2 Decimal Place for percentage\_progress\_postsec | Percentage\_of\_N-level\_Cohort |
|  |  | Round to 2 Decimal Place for percentage\_progress\_postsec | Percentage\_of\_O-level\_Cohort |
|  |  | Filter year from 2016 to 2020 | Percentage\_of\_O-level\_Cohort |
|  |  | Filter year from 2016 to 2020 | Primary\_Students\_and\_Teachers |
| column1 | Name | Renaming Column Name | Polytechnic\_graduate\_ survey |
|  |  | Remove Null column | Polytechnic\_graduate\_ survey |
| Fresh\_ Graduates | Fresh\_ Graduates\_2019 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| column3 | Fresh\_ Graduates\_2020 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| column4 | Fresh\_ Graduates\_2021 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| Post\_NS\_ PNS\_Graduates | Post\_NS\_ PNS\_ Graduates\_2019 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| column7 | Post\_NS\_ PNS\_ Graduates\_2020 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| column8 | Post\_NS\_ PNS\_ Graduates\_2021 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| Combined\_Fresh\_ and\_Post\_NS\_PNS\_ Graduates | Combined\_Graduates\_2019 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| column11 | Combined\_Graduates\_2020 | Renaming Column Name | Polytechnic\_graduate\_ survey |
| column12 | Combined\_Graduates\_2021 | Renaming Column Name | Polytechnic\_graduate\_ survey |
|  |  | Sort year column by descending | Students\_and\_Teachers\_ under MOE |
|  |  | Sort year column by descending | Pre-university\_students\_and\_ teachers |

#### How is the dataset related to problem / hypothesis?

|  |
| --- |
| **Enrolment\_by\_institutions** contain enrolment information for post-secondary education collected by the Ministry of Education (MOE) which will help us to determine if Singapore Education is **superior** for the student to study based on the number of Enrolment for each institution and it will help to determine if the Singapore Education has **improved** by looking at the enrolment number.**Enrolment\_moe\_kindergartens** contain the number of enrolments for kindergartens school collected by the Ministry of Education (MOE) which will help to determine if MOE is giving kids a **Holistic development** and integrated, active learning and it will help to determine if Singapore Education has **improved** by looking at the enrolment number.**Enrolment\_pre-university** contain enrolment information for pre-university education collected by the Ministry of Education (MOE) which will help us to determine if student prefer to take up **subject-based teaching** instead of **course-based teaching** and it will help to determine if the education **has improved** or if the education is flexible for the students to study by looking at the enrolment number.**Enrolment\_primary** contain enrolment information for primary education collected by the Ministry of Education (MOE) which will help us to determine the number of enrolments for the **last past 5 years** and it will help to determine if the education has **improved** by looking at the enrolment number.**Percentage\_of\_N-level\_Cohort** contain the percentages of student progress to post-secondary education collected by the Ministry of Education which will help to determine MOE school is putting a lot of effort in teaching and helping the students to progress to their **desired pathway** and it will help to determine if the education has **improved** by looking at the enrolment number.**Percentage\_of\_O-level\_Cohort** contain the percentages of student progress to post-secondary education collected by the Ministry of Education which will help to determine if MOE school is putting a lot of effort into teaching and helping the students to progress to their **desired pathway** and it will help to determine if the education has **improved** by looking at the enrolment number.**Primary\_Students\_and\_Teachers** contain the number of students and teachers under MOE which is collected by the Ministry of Education which will help us to determine if MOE is giving more **needs and support** to students and it will make the education **flexible** as new teachers are arriving for primary school to teach.**Secondary\_Students\_and\_Teachers** contain the number of students and teachers under MOE which is collected by the Ministry of Education which will help us to determine if MOE is giving **more needs and support** to students and it will make the education **flexible** as new teachers are arriving for secondary school to teach.**Enrolment\_secondary\_by\_level\_and\_ course** contains the number of students in each stream for secondary school which is collected by the Ministry of Education to find out if there is **any improvement** in the education stream by looking at the enrolment number.**Polytechnic\_graduate\_survey** contain the percentage of employed, full-time, freelancing, part-time, unemployed for fresh graduates, post ns and combined fresh and post ns graduates and the monthly gross salary which is collected by the Ministry of Education to determine if having a diploma certificate give you a **high pay salary** and the **rate of finding a job** and it will help to determine if the education has **improved** in having the certificate by looking at the employment rate.**Graduate\_employment\_survey\_ntu-nus-sit-smu-suss-sutd** contain the basic\_monthly\_mean and median, degree, employment\_rate, gross\_monthly \_25 and 75, gross\_monthly\_mean and median, school, university and year which is collected by the Ministry of Education to find out having a degree certificate what is the **salary** and the **rate of finding a job and** it will help to determine if the education has **improved** in having the certificate by looking at the employment rate.**Polytechnics\_information** contain the number, of course, enrolment, graduates, intake, sex and year which is collected by the Ministry of Education which will help us to determine if student prefer to take up **subject-based teaching** instead of **course-based teaching** and it will help to determine if the education has **improved** by looking at the enrolment number.**Universities-enrolment** contains the number, of course, enrolment, graduates, intake, sex and year which is collected by the Ministry of Education which will help us to find the **number of enrolments** in university in Singapore and it will help to determine if the education has **improved** by looking at the enrolment number. |

#### Data Visualisation

#### Describe all visual charts created

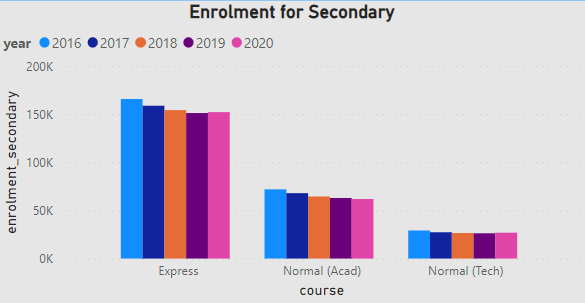
#### 

This bar chart shows the **number of enrolments** for Kindergarten Schools from 2018 to 2022.

Chart, pie chart

Description automatically generated

This pie chart shows the **number of enrolments** for Primary School from 2016 to 2020.

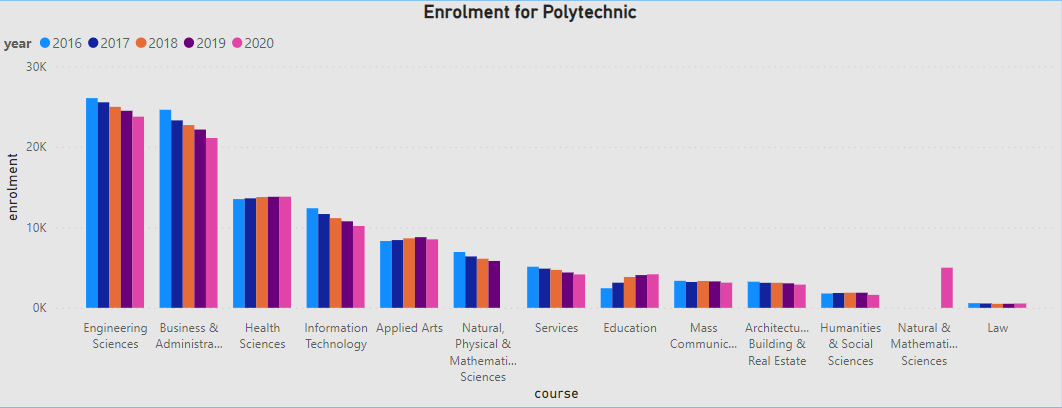


This bar chart shows the **number of enrolments** for each stream in Secondary School from 2016 to 2020.

Chart, pie chart

Description automatically generated

This pie chart shows the **number of enrolments** for ITE from 2016 to 2020.



This bar chart shows the **number of enrolments** for the course in Polytechnic School from 2016 to 2020.

Graphical user interface, text

Description automatically generated with medium confidence

This card shows the **number of graduates** in Polytechnic School from 2016 to 2020.

Graphical user interface

Description automatically generated with low confidence

This card shows the **number of intakes** in Polytechnic School from 2016 to 2020.

Chart

Description automatically generated

This bar chart shows the **number of enrolments** in Pre-University School from 2016 to 2020.

A screenshot of a computer

Description automatically generated with medium confidence

This bar chart shows the **number of enrolments** for the course in University School from 2016 to 2020.

Chart, line chart

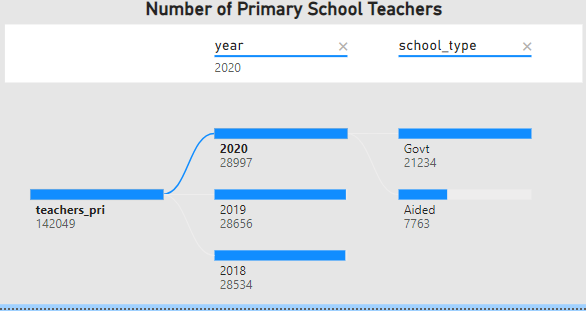
Description automatically generated

This line chart shows the **Percentage of the N-level cohort** from 2015 to 2019.

Chart, line chart

Description automatically generated

This line chart shows the **Percentage of the O-level cohort** from 2016 to 2020.



This decomposition tree shows the **number of primary school teachers** under MOE from 2016 to 2020.

Graphical user interface

Description automatically generated

This decomposition tree shows the **number of secondary school teachers** under MOE from 2016 to 2020

Background pattern

Description automatically generated

This clustered bar chart shows the **polytechnic monthly salary and percentage** from 2019 to 2021

Graphical user interface

Description automatically generated with medium confidence

This decomposition tree shows **University Monthly Salary and Percentage** for degree students from 2015 to 2019**.**

#### Statistical Modelling

#### Describe the statistical modelling undertaken

I plan to use **Key Influencers** **Visualization for all the enrolment from kindergartens to university** to view which year the influencer rating (year) is **low** or **high** for enrolment students and to view the chart given to know the average excluding selected and the **purpose** of using Key Influencers is to view which factors **affects** the metric being analyzed.