FINAL REPORT

Optimizing our university courses…

Team Fantastic 5

Sea Hao Feng, Sam (A0229849J)

Benjamin Ng Jet Hon (A0229888A)

April Chow Siew Peng (A0229846N)

Ng Zhi Wen (A0229906W)

Quek Ka Wei (A0229869E)

Table of Contents

List of Tables

List of Figures

A. Context and Landscape

In the 1960s, Singapore was an undeveloped country. Since then, it has maintained high levels of economic growth by riding on the traction of rapid industrialisation and exports. Unlike its neighbouring countries, Singapore does not have many forms of natural resources and the nation's success is highly dependent on human capital.

To ensure the nation's global competitiveness, Singapore must develop the necessary skillsets for different industries through education. Technological advancements are fast changing the way we live, work and play. The future workforce will need to be ready to adapt to all the changes. This requires the deliberation of what skills the future workforce needs.

Universities curricula need to be reviewed to ensure that the knowledge, and skills taught remain current and relevant. The effectiveness of Singapore's education system will play an important factor in its progression forward.

Against this backdrop, the team carried out extensive desk research and thought about how we might be able to optimize undergraduate courses. The need for higher learning institutions to change strategies to groom IT talent was also highlighted in a recent Channel NewsAsia article.

Introducing Okun's Law - Postulated by Yale professor and economist Arthur Okun in the early 1960s. Okun's Law examines the statistical relationship between a country's unemployment and economic growth rates and states that a **country's gross domestic product (GDP) must grow at about a 4% rate for one year to achieve a 1% reduction in the rate of unemployment.**

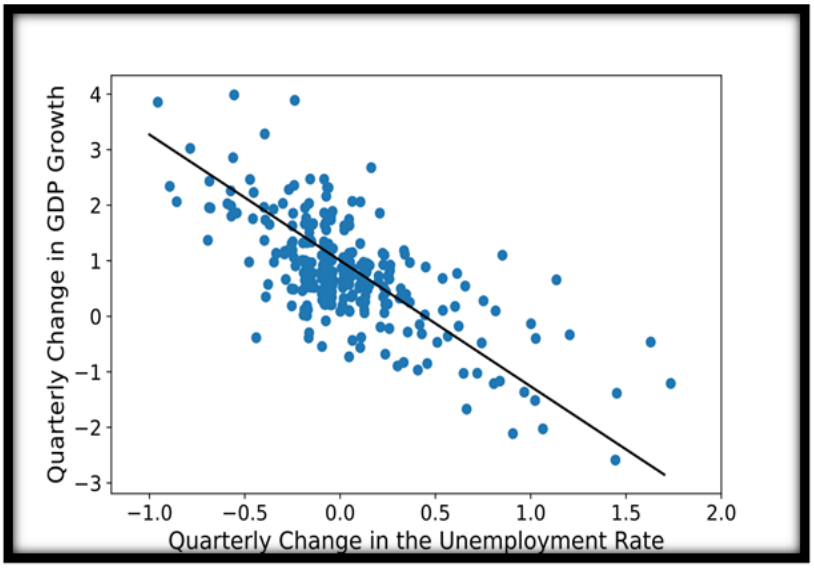


Figure 1: Okun's law

Bringing together all the various sources of information gathered from secondary research, our analytics project sought to establish the link of what we know about the relationship between unemployment rate to GDP – (refer to as Okun's Law) to graduate employment rate and graduate enrolment courses. And then looking at the relationship between these variables how we can optimize the undergraduate courses…

B. Target Audience

For the purpose of our project, we have identified our target audience to be Ministry of Education (MOE) as they are responsible for planning and implementing education policies for the nation.

Figure 2: Ministry of Education

C. Business & Technical Objectives

The key business problem can be broken down into the following business objectives with the corresponding technical objective:

	Business Objectives	Technical Objectives
1	To find out if the current education system meets the demands of the economy	• Ascertain if correlation exists between (1) Economic growth and employment rates and (2) Overall employment and graduate employment • Examine the relationship between course enrolment number and GDP. (e.g. Low course enrolment number and high GDP) • Identify the major contributors (Y-O-Y % growth) to economic growth (through examining the GDP by sectors) • Map education courses to the economic sectors of Singapore.
2	To discover insights on how the education system can meet the demands of the future economy?	• Identify the education courses in the various economic sectors of Singapore (Growth sectors, declining sectors) • Does the increase in student enrolment correlate to the growth or decline? (Trends of student enrolment should match the trends of the sector) • Does the percentage contribution of the sector to GDP correlate to the percentage of enrolment across all courses? (is there a gap?) (Horticulture example) • Examine GDP by sector for foreign countries versus Singapore

Table 1: Business and technical objectives

D. Project Vision

To optimize and improve our local university undergraduate course curriculum to better prepare and equip our workforce with the necessary and relevant skills to fuel Singapore's economic development.

E. Project Timeline

The team put together a project timeline with breakdown of tasks, timeline and personnel involved, in the form of a Gantt Chart.

Fantastic 5
Project Lead

Project Start: 2021-02-27
Display Week: 6

TASK	ASSIGNED TO	PROGRESS	START	END
Formulation of Proposals				
Project Proposal Discussion 1,2,3 (4 days each)	All	100%	2/27/21	3/10/21
Project Proposal Finalisation	All	100%	3/11/21	3/14/21
Project Proposal Submission				
Define Business Problem	All	100%	3/15/21	3/18/21
Define Project Scope and Design, Develop Project Schedule	Benjamin Ng Chow Siew Peng	100%	3/19/21	3/24/21
Source and Prepare Data	Quek Ka Wei Sea Hao Feng	100%	3/25/21	4/1/21
Analyse and Validate Data	Benjamin Ng Chow Siew Peng Ng Zhi Wen	100%	4/2/21	4/11/21
Project Demo Presentation				
Prepare and Fine-tune Data	Benjamin Ng Chow Siew Peng Ng Zhi Wen	100%	4/12/21	4/17/21
Analyse Data	Quek Ka Wei Sea Hao Feng	100%	4/18/21	5/1/21
Visualise Data	All	100%	5/1/21	5/6/21
Develop Findings and Recommendations	All	100%	5/6/21	5/12/21
Collate and Submit	All	100%	5/12/21	5/19/21

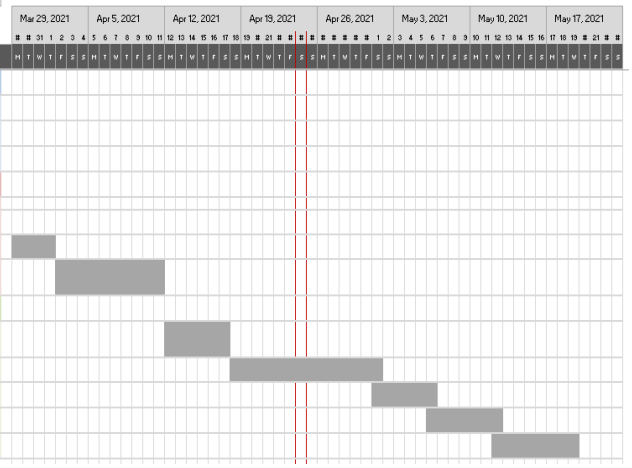


Figure 3: Gantt Chart

F. Project Design

Project Methodology

Disciplined Agile Delivery approach was used as features were developed in iterations while the project has a fixed timeframe.

Building features in iterations allow the team to be goal-oriented, experiment and fail fast if it does not work out.

CRISP-DM framework was deployed under the sandbox (analytics) and Disciplined Agile Delivery was used for the entire project, starting with the project inception phase where we came up with the project's vision along with the business and technical objectives, identification of project risks, setting up the project design and some initial iterations to establish the foundation.

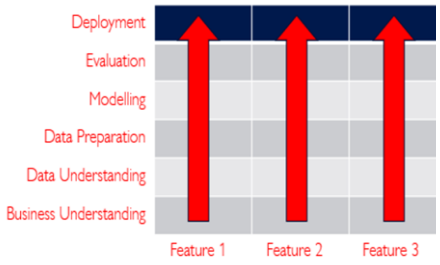


Figure 4: Disciplined Agile Delivery Approach

Data Sourcing

This project will use structured aggregated data from the following sources:

S/N	Title	Data Provider
1	Singapore Gross Domestic Product at Current & Chained Prices, By Industry (SSIC 2015 Version 2018), Annual	Singapore Department of Statistics (DOS)
2	Labour Force Singapore	Singapore Department of Statistics (DOS)
3	Graduate Employment Survey - NTU, NUS, SIT, SMU, SUSS & SUTD	Data.gov.sg
4	2020 GES Employment Rates and Salaries of NUS, NTU, SMU and SUSS Graduates by Course Cluster	Ministry of Education, Singapore
5	NUS: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Ministry of Education, Singapore
6	NTU: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Ministry of Education, Singapore
7	SMU: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Ministry of Education, Singapore
8	SUTD: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Ministry of Education, Singapore
9	SIT: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Ministry of Education, Singapore
10	SUSS: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Ministry of Education, Singapore
11	SUTD: 2018 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Singapore University of Technology and Design
12	Universities - Intake, Enrolment and Graduates by Course	Data.gov.sg
13	Intake, Enrolment and Graduates by Institutions	Data.gov.sg
14	US – GDP per capita (current US$)	World Bank
15	US - GDP growth (annual %)	World Bank
16	US – GDP by industry	Economic Research at the St. Louis Fed
17	US – Bachelor’s degrees awarded, by field	National Center for Science and Engineering Statistics, National Science Foundation

Table 2: Data sources

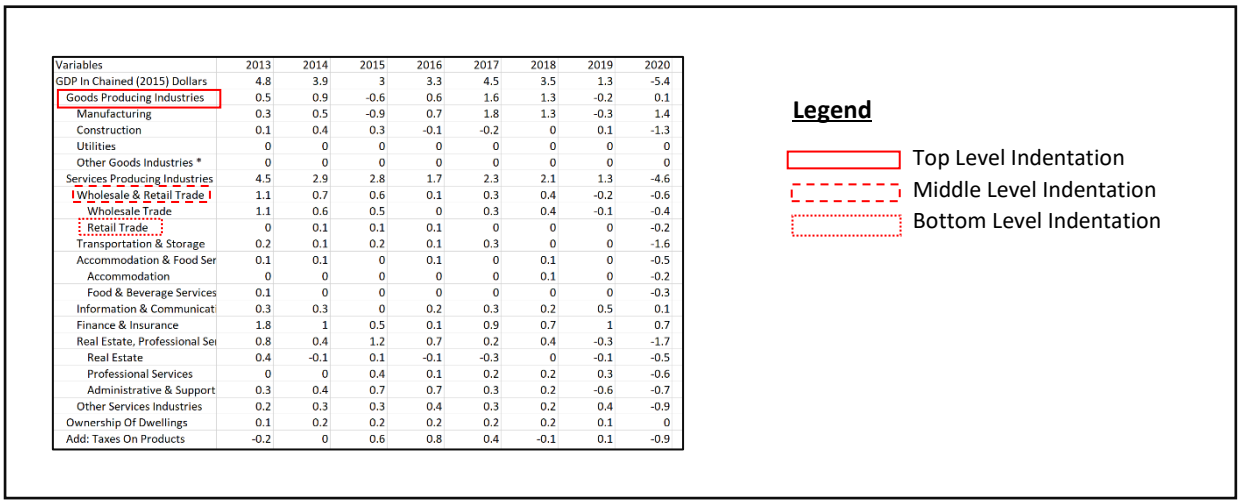
G. Data Pre-Processing

As mentioned by Trifacta (a software company developing data wrangling tools), 80% of the time in an analytics project will be spent on cleaning the dataset before performing any advanced analytics (e.g. visualizations, creating machine learning models). Therefore, it is imperative that the team gets this fundamental step right so that the project will not be affected downstream.

Data Set 1 - SG GDP Growth Rate by Industry ANNUAL

This dataset contains the Gross Domestic Product (GDP) Growth Rate of Singapore and the contributions of individual industries towards it.

Records of the GDP were from 1961 to 2020. Given that the timeframe of this project aligns to the compiled Graduate Employment Survey (Data Set 4), the team trimmed down the window of this dataset to the timeframe of 2013 to 2019. The dataset was also flattened out into 3 levels of category (Overview, Industry and Sector) because sub-categories were placed in the same column as the main domains.



Variables	2013	2014	2015	2016	2017	2018	2019	2020
GDP In Chained (2015) Dollars	4.8	3.9	3	3.3	4.5	3.5	1.3	-5.4
Goods Producing Industries	0.5	0.9	-0.6	0.6	1.6	1.3	-0.2	0.1
Manufacturing	0.3	0.5	-0.9	0.7	1.8	1.3	-0.3	1.4
Construction	0.1	0.4	0.3	-0.1	-0.2	0	0.1	-1.3
Utilities	0	0	0	0	0	0	0	0
Other Goods Industries *	0	0	0	0	0	0	0	0
Services Producing Industries	4.5	2.9	2.8	1.7	2.3	2.1	1.3	-4.6
Wholesale & Retail Trade	1.1	0.7	0.6	0.1	0.3	0.4	-0.2	-0.6
Wholesale Trade	1.1	0.6	0.5	0	0.3	0.4	-0.1	-0.4
Retail Trade	0	0.1	0.1	0.1	0	0	0	-0.2
Transportation & Storage	0.2	0.1	0.2	0.1	0.3	0	0	-1.6
Accommodation & Food Ser	0.1	0.1	0	0.1	0	0.1	0	-0.5
Accommodation	0	0	0	0	0	0.1	0	-0.2
Food & Beverage Services	0.1	0	0	0	0	0	0	-0.3
Information & Communicat	0.3	0.3	0	0.2	0.3	0.2	0.5	0.1
Finance & Insurance	1.8	1	0.5	0.1	0.9	0.7	1	0.7
Real Estate, Professional Se	0.8	0.4	1.2	0.7	0.2	0.4	-0.3	-1.7
Real Estate	0.4	-0.1	0.1	-0.1	-0.3	0	-0.1	-0.5
Professional Services	0	0	0.4	0.1	0.2	0.2	0.3	-0.6
Administrative & Support	0.3	0.4	0.7	0.7	0.3	0.2	-0.6	-0.7
Other Services Industries	0.2	0.3	0.3	0.4	0.3	0.2	0.4	-0.9
Ownership Of Dwellings	0.1	0.2	0.2	0.2	0.2	0.2	0.1	0
Add: Taxes On Products	-0.2	0	0.6	0.8	0.4	-0.1	0.1	-0.9

Figure 5: Trimming down dataset 1

The output of this dataset transformation is shown below:

Overview	Industry	Sector	2013Y	2014Y	2015Y	2016Y	2017Y	2018Y	2019Y	2020Y
Goods Producing Industries	Manufacturing	Manufacturing	0.3	0.5	-0.9	0.7	1.8	1.3	-0.3	1.4
Goods Producing Industries	Construction	Construction	0.1	0.4	0.3	-0.1	-0.2	0	0.1	-1.3
Goods Producing Industries	Utilities	Utilities	0	0	0	0	0	0	0	0
Goods Producing Industries	Other Goods Industries *	AGRICULTURE, FISHING AND QUARRYING	0	0	0	0	0	0	0	0
Services Producing Industries	Wholesale & Retail Trade	Wholesale Trade	1.1	0.6	0.5	0	0.3	0.4	-0.1	-0.4
Services Producing Industries	Wholesale & Retail Trade	Retail Trade	0	0.1	0.1	0.1	0	0	0	-0.2
Services Producing Industries	Transportation & Storage	Transportation & Storage	0.2	0.1	0.2	0.1	0.3	0	0	-1.6
Services Producing Industries	Accommodation & Food Services	Accommodation	0	0	0	0	0	0.1	0	-0.2
Services Producing Industries	Accommodation & Food Services	Food & Beverage Services	0.1	0	0	0	0	0	0	-0.3
Services Producing Industries	Information & Communications	Information & Communications	0.3	0.3	0	0.2	0.3	0.2	0.5	0.1
Services Producing Industries	Finance & Insurance	Finance & Insurance	1.8	1	0.5	0.1	0.9	0.7	1	0.7
Services Producing Industries	Real Estate, Professional Services A	Real Estate	0.4	-0.1	0.1	-0.1	-0.3	0	-0.1	-0.5
Services Producing Industries	Real Estate, Professional Services A	Professional Services	0	0	0.4	0.1	0.2	0.2	0.3	-0.6
Services Producing Industries	Real Estate, Professional Services A	Administrative & Support Services	0.3	0.4	0.7	0.7	0.3	0.2	-0.6	-0.7
Services Producing Industries	Other Services Industries	Other Services Industries	0.2	0.3	0.3	0.4	0.3	0.2	0.4	-0.9

Figure 6: Output of transformed dataset 1

Data Set 2 - SG GDP Total and by Industry ANNUAL

Data Set 2 contains the absolute figures of Singapore's Total GDP and contributions by its respective industries. It faced the same timeframe and data structure issue as Data Set 1.

The team used the pre-processing technique used for Data Set 1 and the output is as follows:

Industry	Sector	2013	2014	2015	2016	2017	2018	2019	In millions (SGD)
Manufacturing	Manufacturing	67,885	71,809.7	76,598.2	77,325.5	88,265	105,762.7	99,360.7	
Construction	Construction	17,127.3	19,183.8	20,433.8	19,673.1	17,864.3	17,827.1	18,486.4	
Utilities	Utilities	5,221.4	5,430.3	5,815.9	5,615.1	5,500.2	5,719.4	5,795.1	
Other Goods Industries *	Other Goods Industries *	132.5	138	138.1	139.2	145.9	153.2	165.4	
Wholesale & Retail Trade	Wholesale Trade	61,886.9	57,829.3	58,306.1	66,397.9	74,324.9	79,669.4	78,511.7	
Wholesale & Retail Trade	Retail Trade	6,608.4	6,895.2	7,278.4	7,604	7,781.3	7,936.5	7,613.8	
Transportation & Storage	Transportation & Storage	24,756.6	27,092.7	30,014.6	27,617.5	31,531.8	30,684.5	31,682.4	
Accommodation & Food Services	Accommodation	3,470.1	3,639.7	3,604.7	3,763.8	3,891.8	4,224.5	4,309.5	
Accommodation & Food Services	Food & Beverage Service	4,738.8	5,025.6	5,158.3	5,470	5,525	5,643.5	5,784.7	
Information & Communications	Information & Communic	14,774.3	15,798.8	15,779.3	17,254.1	18,718.2	19,755.2	22,325.5	
Finance & Insurance	Finance & Insurance	42,892.6	45,968.7	49,874.5	51,636.5	56,853.3	62,535.7	68,168.3	
Real Estate, Professional Services And Administrative & Support Services	Real Estate	19,253.4	18,970.8	18,790.1	17,418	15,703.3	16,421.9	16,592.0	
Real Estate, Professional Services And Administrative & Support Services	Professional Services	21,673.2	22,107.2	24,218	24,895.9	25,745.7	27,039	29,085	
Real Estate, Professional Services And Administrative & Support Services	Administrative & Support	14,734.9	16,856.4	20,114.3	23,766.5	25,428.9	26,837.1	23,967.6	
Other Services Industries	Other Services Industries	40,293.1	42,655.4	44,963	47,623.3	50,525.4	52,584.1	54,461.8	
Ownership Of Dwellings	Ownership Of Dwellings	17,251	17,797.2	18,100.1	17,589.4	17,436.3	17,751	18,453	
Taxes	Taxes	22,170.8	21,749.1	24,256.7	26,582.4	28,873.8	26,579.1	25,975.2	

raw | GDP per sector | GDP Total

Figure 7: Output of transformed dataset 2

Data Set 3 - Labour Force SG

This dataset contains the Labour Force statistics of Singapore from 1970 to 2020. A sample of its raw format is shown below:

Subject: Labour, Employment, Wages and Productivity
Topic : Labour Force
Title : M182331 - Labour Force, Aged 15 Years And Over, (June), Annual

Variables	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	198
Total Labour Force (Thousand)	701.7	-	-	-	-	-	-	-	-	-	1,1
Total Employed Persons (Thousand)	644.2	-	-	-	-	-	-	-	-	-	1,0
Total Unemployed Persons (Thousand)	57.5	-	-	-	-	-	-	-	-	-	
Total Unemployment Rate * (Per Cent)	8.2	-	-	-	-	-	-	-	-	-	
Resident Unemployment Rate * (Per Cent)	na	na	na	na	na	na	na	na	na	na	na

Figure 8: Structure of dataset3 (pre-processed)

The data was trimmed from 2013-2019 in alignment to Data Set 4 (Singapore Graduate Employment Survey 2013-2019) and the last 2 variables (Total Unemployment Rate and Resident Unemployment Rate) were dropped. The remaining rows and columns were then transposed to form the following table:

Year	Labour Force (thousands	Employed (thousands)	Unemployed (thousands)
2013	3443.7	3352.9	90.7
2014	3530.8	3440.2	90.7
2015	3610.6	3516	94.6
2016	3672.8	3570	102.8
2017	3657	3550.1	106.9
2018	3675.6	3575.3	100.2
2019	3742.5	3631.7	110.8

Figure 9: Output of transformed dataset 3

Data Set 4 - Singapore GES 2013-2019

Data Set 4 comprises of raw data from different data sources:

Data Source	Title	Purpose
1	Graduate Employment Survey - NTU, NUS, SIT, SMU, SUSS & SUTD	Bedrock for this Data Set
2	SUTD: 2018 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Compensate missing records in Data Source 1
3	2020 GES Employment Rates and Salaries of NUS, NTU, SMU and SUSS Graduates by Course Cluster	To retrieve the list of education clusters and its corresponding degree programmes as specified by Singapore's Ministry of Education
4	NUS: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	Supplement Data Source 1 with 2019 Graduate Employment Survey results for all autonomous universities in Singapore as Data Source 1 only contains records from 2013 to 2018
5	NTU: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	
6	SMU: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	
7	SUTD: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	
8	SIT: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	
9	SUSS: 2019 GES Employment Rates and Salaries of Graduates by Bachelor Degree	

Table 3: Data Sources of Data Set 4 (SG GES 2013-2019)

The raw format of Data Source 1 contains the following fields:

1. year
2. university
3. school
4. degree
5. employment_rate_overall
6. employment_rate_ft_perm
7. basic_monthly_mean
8. basic_monthly_median gross_monthly_mean
9. gross_monthly_median
10. gross_mthly_25_percentile
11. gross_mthly_75_percentile

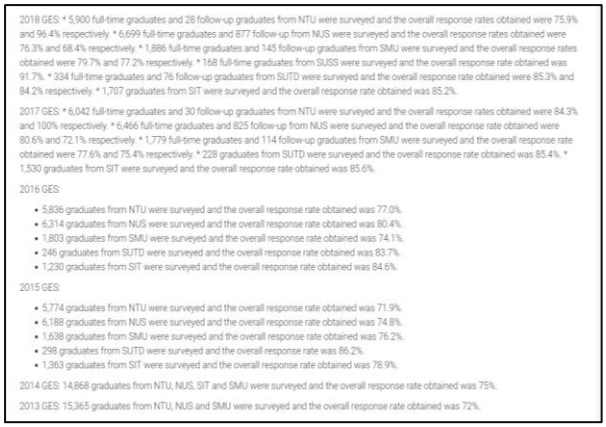
We retained column A to F (figure 10).

	A	B	C	D	E	F
1	year	university	school	degree	employme	employr
2	2016	Singapore Manageme	School of Accountancy (4-year progra	Accountancy	98.5	96
3	2017	Singapore Manageme	School of Accountancy (4-years progr	Accountancy	98	96.6
4	2019	Nanyang Technologic	College of Business (Nanyang Business	Accountancy	97.5	94
5	2015	Singapore Manageme	School of Accountancy (4-year progra	Accountancy	97.3	93
6	2016	Nanyang Technologic	College of Business (Nanyang Business	Accountancy	97.3	95.1
7	2018	Singapore Manageme	School of Accountancy (4-year progra	Accountancy	96.4	92.3
8	2014	Singapore Institute of	DigiPen Institute of Technology	Bachelor of Arts in Game Design	100	86.7
9	2019	Singapore Manageme	School of Accountancy (4-year progra	Accountancy	96.4	93.7
10	2016	Singapore Institute of	The Glasgow School of Art	Bachelor of Arts with Honours in Interior Design	97.4	76.3
11	2017	Singapore Institute of	DigiPen Institute of Technology	Bachelor of Arts in Game Design ^	96.2	92.3
12	2018	Nanyang Technologic	College of Business (Nanyang Business	Accountancy	94.6	92.2
13	2014	Singapore Institute of	The Glasgow School of Art	Bachelor of Arts with Honours in Interior Design	93.9	90.9
14	2015	Nanyang Technologic	College of Business (Nanyang Business	Accountancy (3-yr direct Honours Programme)	97.3	96.5
15	2017	National University of	School of Design and Environment	Bachelor of Arts (Industrial Design)	93.3	60
16	2019	National University of	Faculty of Arts & Social Sciences	Bachelor of Arts (Hons)	92.7	68.2
17	2017	Nanyang Technologic	College of Engineering	Bachelor of Communication Studies (Hons)	92.6	62.2
18	2017	Singapore Institute of	The Glasgow School of Art	Bachelor of Arts with Honours in Interior Design	92.3	87.2
19	2018	Singapore Institute of	The Glasgow School of Art	Bachelor of Arts with Honours in Interior Design	92.3	76.9
20	2018	Singapore Institute of	DigiPen Institute of Technology	Bachelor of Fine Arts in Digital Art and Animation	91.2	52.9
21	2016	Singapore Institute of	The Glasgow School of Art	Bachelor of Arts with Honours in Communication Design	91.1	58.9
22	2016	National University of	Faculty of Arts & Social Sciences	Bachelor of Arts (Hons)	90.7	74.1

raw | course-cluster | **processed** | response-rate | response-rate-stats | overall-grad-employment | overall-grad-employment rate | g ...

Figure 10: Processed sheet from Data Set 4

The team also tabulated the total number of participants polled and the response rate from Data Source 1's description (fig. 11) into a table in the "response-rate" sheet as follows:



	Year	University	Total No. of Grads	Response Rate (%)
2	2013	NUS, NTU, SMU	15,365	72
3	2014	NUS, NTU, SMU, SIT	14,868	75
4	2015	NUS	6,188	74.8
5	2015	NTU	5,774	71.9
6	2015	SMU	1,638	76.2
7	2015	SIT	1363.00	78.9
8	2015	SUTD	298.00	86.2
9	2016	NTU	5,836	77
10	2016	NUS	6,314	80.4
11	2016	SMU	1,803	74.1
12	2016	SUTD	246	83.7
13	2016	SIT	1,230	84.6
14	2017	NTU	6,042	84.3
15	2017	NTU	30	100
16	2017	NUS	6,466	80.6
17	2017	NUS	825	72.1
18	2017	SMU	1,779	77.6
19	2017	SMU	114	75.4
20	2017	SUTD	228	85.4
21	2017	SIT	1,530	85.6
22	2018	NTU	5,900	75.9

raw | course-cluster | processed | **response-rate** | response-rate-stats

Figure 11: Transformation of response rate sheet

Checking through the processed information, the team realised that records of SUTD is missing despite the description of Data Source 1 stating it contained the records of all local autonomous universities from 2013 to 2018. SUTD's 2018 Graduate Employment Survey (GES) data was sourced from its official website and added to existing records:

year	university	school	degree	employme	employr
2018	Singapore University of Technology and Design		Bachelor of Engineering (Engineering Product Developme	94.7	86.3
2018	Singapore University of Technology and Design		Bachelor of Engineering (Information Systems Technolog	97.3	93.3
2018	Singapore University of Technology and Design		Bachelor of Engineering (Engineering Systems and Design)	90	90
2018	Singapore University of Technology and Design		Bachelor of Science (Architecture and Sustainable Design)	95.1	88.5

Figure 12: Snapshot of SUTD 2018 (GES data)

The team realised that the timeframe of the Data Source 1 can be more comprehensive by including the year (2019) before the COVID-19 pandemic changed the fundamentals of the economy in 2020.

Data regarding the GES 2019 were available through the Ministry of Education's (MOE) website, but the records were not compiled together in one data source. The team transcribed the information from each autonomous university's 2019 GES (Data Source 4 to 9) and put them through the same data cleaning process as Data Source 1.

After combing and cleaning Data Source 1, 4, 5, 6, 7, 8 and 9 to establish the foundation for Data Set 4, the team created a reference table for education cluster by transcribing the information from Data Source 3 into a table in the "course-cluster" sheet as shown in Figure 13 below:

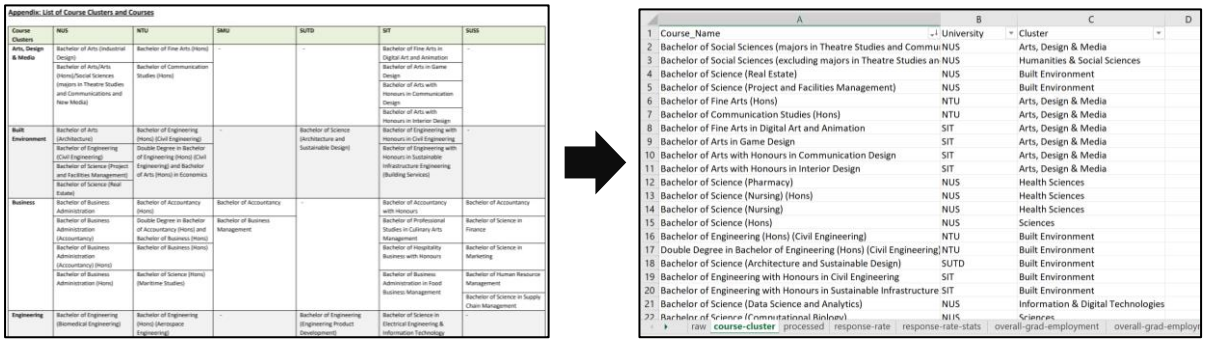


Figure 13: Course cluster sheet transformation

On the "processed" sheet, we the cleaned "degree" field (Column D) using a custom Excel formula to remove all special characters (besides parentheses) as some degree names were appended with special characters (e.g. # ^ % *).

After removing the special characters (besides parentheses), we clustered the "cleansed_degree name" field (Column G in figure 14) and mapped them to "course-cluster" sheet via the VLOOKUP function. It is expected that the first round of clustering will result in some N/A values because some degree names provided in Data Source 1 were not prefixed with its degree level (e.g. bachelor) or type (e.g. science, arts, engineering).

To circumvent this issue, the team filtered the N/A values and sorted its "cleansed_degree name" field in A-Z alphabetical order to ease the manual mapping process of the education cluster. No N/A values were dropped in this data construction process. The following is an output of the entire clustering process:

	D	E	F	G	H	I
1	degree	employme	employ	cleansed_degree name	Clustered1	Clustered2
2	Accountancy (3-yr direct Honours Programme)	97.1	95.7	Accountancy (3-yr direct Honours Programme)	#N/A	Business
3	Accountancy (4-years programme)	94.2	93.2	Accountancy (4-years programme)	#N/A	Business
4	Accountancy (4-years programme) Cum Laude and above	95.2	95.2	Accountancy (4-years programme) Cum Laude and abo	#N/A	Business
5	Accountancy and Business	97.4	96.1	Accountancy and Business	#N/A	Business
6	Aerospace Engineering	95.3	95.3	Aerospace Engineering	#N/A	Engineering
7	Art, Design & Media	81.6	61.4	Art, Design Media	#N/A	Arts, Design & Media
8	Arts (with Education)	100	100	Arts (with Education)	#N/A	Arts, Design & Media
9	Bachelor of Applied Science (Hons)	97.3	97.3	Bachelor of Applied Science (Hons)	#N/A	Sciences
10	Bachelor of Computing (Communications and Media)	88.9	77.8	Bachelor of Computing (Communications and Media)	#N/A	rmation & Digital Technologies
11	Bachelor of Computing (Electronic Commerce)	88.9	83.3	Bachelor of Computing (Electronic Commerce)	#N/A	rmation & Digital Technologies
12	Bachelor of Dental Surgery	100	100	Bachelor of Dental Surgery	#N/A	Health Sciences
13	Bachelor of Engineering (Bioengineering)	74	60	Bachelor of Engineering (Bioengineering)	#N/A	Engineering
14	Bachelor of Laws (LLB) (Hons) #	98.8	98.2	Bachelor of Laws (LLB) (Hons)	#N/A	Humanities & Social Sciences
15	Bachelor of Medicine and Bachelor of Surgery (MBBS) #	100	100	Bachelor of Medicine and Bachelor of Surgery (MBBS)	#N/A	Health Sciences
16	Bachelor of Science (Pharmacy) (Hons) #	96.4	96.4	Bachelor of Science (Pharmacy) (Hons)	#N/A	Health Sciences
17	Bachelor of Social Sciences	88.1	76.8	Bachelor of Social Sciences	#N/A	Humanities & Social Sciences
18	Bioengineering	81.3	68.8	Bioengineering	#N/A	Engineering
19	Biological Sciences	78.6	68.1	Biological Sciences	#N/A	Sciences
20	Biomedical Sciences (Traditional Chinese Medicine) #	90.7	88.4	Biomedical Sciences (Traditional Chinese Medicine)	#N/A	Health Sciences
21	Business (3-yr direct Honours Programme)	90.9	85.7	Business (3-yr direct Honours Programme)	#N/A	Business
22	Business and Computing	87.5	87.5	Business and Computing	#N/A	Business

raw | course-cluster | processed | response-rate | response-rate-stats | overall-grad-employment | overall-grad-employment rate | g ...

Figure 14: Output of cleansing process

Data Set 5 - US GDP By Industry

This dataset contains information from FRED Economic Data: Gross Output by Industry, Billions of Dollars, Seasonally Adjusted at Annual Rates. It spans over a few pages with dynamic hyperlinks and it is not possible to copy and paste into Excel due to the structure of the website. We used Rvest package in R to scrape the data from site.

As we compared Data Set 5 – US Sector GDP vs SG GDP Total and by Industry ANNUAL, there are some data mapping which took place.

For US Sector GDP, the following data were mapped

- Educational services, health care, and social assistance merged into Other services, except government

For SG GDP Total and by Industry ANNUAL, the follow data were mapped:

- Other Goods Industries = US' Agriculture, forestry, fishing, and hunting
- Transportation & Storage = US's Transportation and warehousing
- Accommodation, Food & Beverages Services merged into US' Arts, entertainment, recreation, accommodation, and food services
- Information & Communications renamed as per US' Information
- Finance & Insurance, Real Estate, Ownership Of Dwellings combined and renamed as per US' Finance, insurance, real estate, rental, and leasing
- Professional services and Administrative & Support Services combined and renamed as per US' Professional and business services
- Other Services Industries renamed as per US' Other services, except government

This ensured the final outputs of the 2 data sets are in line and grouped into the following bins:

- Manufacturing
- Construction
- Utilities
- Agriculture, forestry, fishing, and hunting
- Wholesale Trade
- Retail Trade
- Transportation and warehousing
- Arts, entertainment, recreation, accommodation, and food services
- Information
- Finance, insurance, real estate, rental, and leasing
- Professional and business services
- Other services, except government
- Mining

H. Modelling

To make certain analysis possible, the data must be modelled in a way that is helpful to the analytics end user. This section will showcase the various modelling techniques that the team had performed to setup the stage for the analysis that we want to do.

Modelling 1

This modelling process prepares Data Set 1 for a trend analysis on Singapore's Overall GDP Growth via trended line graph. It uses the Excel pivot table function on the "processed" sheet of Data Set 1 with the following configuration:

- Rows: Values
- Values: Years (2013 to 2019)

The output of this data modelling process is shown below:

Values	
2013	4.8
2014	3.6
2015	2.2
2016	2.2
2017	3.9
2018	3.3
2019	1.2

Figure 15: Output of SG overall GDP growth rate by year

Modelling 2

Similar to Modelling 1, this modelling process seeks to display the trend line of various industry's contribution to Singapore's GDP Growth Rate from 2013 to 2019. It also uses a pivot table on the "processed" sheet of Data Set 1 with the following configuration:

- Columns: Industry
- Rows: Values
- Values: Years (2013 to 2019)

The result of this data modelling process is as follows:

Column Labels												
Values	Accommodation & Food Services	Construction	Finance & Insurance	Information & Communications	Manufacturing	Other Goods Industries *	Other Services Industries	Real Estate, Professional Services And Administrative & Support Services	Transportation & Storage	Utilities	Wholesale & Retail Trade	Grand Total
2013	0.1	0.1	1.8	0.3	0.3	0	0.2	0.7	0.2	0	1.1	4.8
2014	0	0.4	1	0.3	0.5	0	0.3	0.3	0.1	0	0.7	3.6
2015	0	0.3	0.5	0	-0.9	0	0.3	1.2	0.2	0	0.6	2.2
2016	0	-0.1	0.1	0.2	0.7	0	0.4	0.7	0.1	0	0.1	2.2
2017	0	-0.2	0.9	0.3	1.8	0	0.3	0.2	0.3	0	0.3	3.9
2018	0.1	0	0.7	0.2	1.3	0	0.2	0.4	0	0	0.4	3.3
2019	0	0.1	1	0.5	-0.3	0	0.4	-0.4	0	0	-0.1	1.2

Figure 16: Output of Industry contribution to SG overall GDP growth rate by year

Modelling 3

This modelling process prepares the data for a trend analysis of Singapore's Graduate Employment Rate from 2013 to 2019. It uses the Excel pivot table function on the "processed" sheet of Data Set 4 with the following configuration:

- Rows: Values
- Values: Average of employment_rate_overall

The output of this data modelling process is shown below:

Row Labels	Average of employment_rate_overall
2013	90.20
2014	90.60
2015	90.89
2016	90.00
2017	90.53
2018	90.93
2019	90.87
Grand Total	**90.60**

Figure 17: Output of average employment rate by year

Modelling 4

Like Modelling 3, this data modelling seeks to prepare the information needed for a trend analysis of Singapore's Graduate Employment Rate (by Education Cluster) from 2013 to 2019. It also uses the Excel pivot table function on the "processed" sheet of Data Set 4 with the following configuration:

- Columns: Clustered2
- Rows: Values
- Values: Average of employment_rate_overall

The result of this data modelling process is shown in the following figure:

Average of employment_rate_overall	Column Labels								
Row Labels	Arts, Design & Media	Built Environment	Business	Engineering	Health Sciences	Humanities & Social Sciences	Information & Digital Technologies	Sciences	Grand Total
2013	87.42	94.03	93.91	88.42	97.43	89.33	91.59	83.91	90.20
2014	88.26	95.98	93.37	88.31	99.57	88.45	93.21	85.58	90.60
2015	84.23	95.58	95.55	88.92	99.46	90.65	92.95	84.88	90.89
2016	88.23	92.73	94.55	86.53	96.21	90.69	92.54	85.62	90.00
2017	90.24	93.98	95.43	87.56	97.61	87.88	95.88	83.45	90.53
2018	88.79	90.83	94.85	90.16	96.42	87.44	95.62	85.59	90.93
2019	87.23	95.18	93.07	88.50	98.53	87.20	95.28	88.94	90.87
Grand Total	**87.87**	**94.23**	**94.38**	**88.33**	**97.92**	**88.75**	**94.02**	**85.51**	**90.60**

Figure 18: Output of SG graduate employment rate by education cluster and year

Modelling 5

This data modelling process put together the required information for a correlation analysis between Singapore's GDP and Overall Employment. This process will result in a table with 3 columns:

Column	Name	Modelling Method
1	GDP At Current Market Prices [In millions (SGD)]	Sums up the total GDP per year from the "GDP per sector" sheet in Data Set 2
2	Employed (thousands)	Retrieved from the "processed" sheet in Data Set 3
3	Overall Employment [Millions]	The column "Employed (thousands)" divided by 1000

Table 4: Data for correlation analysis (SG GDP and overall employment)

The last column was created for standardization and consistency before proceeding on to the correlation analysis because the first two columns were using different units of measure (Column 1 in millions and Column 2 in thousands).

The outcome of this modelling process is shown below:

Year	GDP At Current Market Prices [In millions (SGD)]	Employed (thousands)	Overall Employment [Millions]
2013	384870.3	3352.9	3.35
2014	398947.9	3440.2	3.44
2015	423444.1	3516	3.52
2016	440372.2	3570	3.57
2017	474115.1	3550.1	3.55
2018	507123.9	3575.3	3.58
2019	510737.8	3631.7	3.63

Figure 19: Output of standardised overall employment(thousands)

Modelling 6

This data modelling process seeks to prepare the information required for a correlation analysis between Singapore's Overall Employment and Graduate Employment. Similar to Modelling 5, this process will create a table with three columns:

Column	Name	Modelling Method
1	Employed (thousands)	Retrieved from the "processed" sheet in Data Set 3
2	Employed Graduates	Averages the total Graduate Employment Rate per year from the "processed" sheet in Data Set 4
3	Employed Graduates (thousands)	The column "Employed Graduates" divided by 1000

Table 5: Data for correlation analysis (SG overall employment and graduate employment)

Similar to Modelling 5, the last column was created for standardization and consistency before proceeding on to the correlation analysis as the first two columns were using different units of measure (Column 1 in thousands and Column 2 in ones).

The result of this data modelling process is shown in the following figure:

Year	Employed (thousands)	Employed Graduates	Employed Graduates (thousands)
2013	3352.9	10029	10.03
2014	3440.2	10167	10.17
2015	3516	10337	10.34
2016	3570	10944	10.94
2017	3550.1	12551	12.55
2018	3575.3	12626	12.63
2019	3631.7	12900	12.90

Figure 20: Output of standardised overall employed graduates (thousands)

I. Preliminary Investigations

Before continuing onto the next phase of the analytics project, the team must conduct Exploratory Data Analysis and verify that the fundamental technical objective holds true to ensure that we are on the right track so that we can set the basis for this project.

Building on the modelling processes in the previous section, the team developed data visualisations to help us better understand the trend in each dataset.

Visualization 1

Figure 21 (Built on Modelling 1) shows the Overall GDP Growth Rate of Singapore:

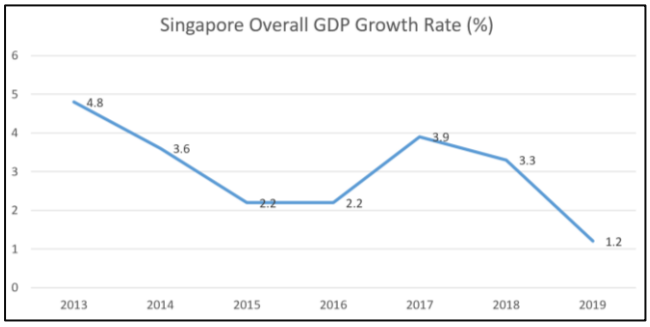


Figure 21: SG GDP growth rate by year

Visualization 2

Figure 22 (Built on Modelling 2) displays the GDP Growth Rate of Singapore by sector:

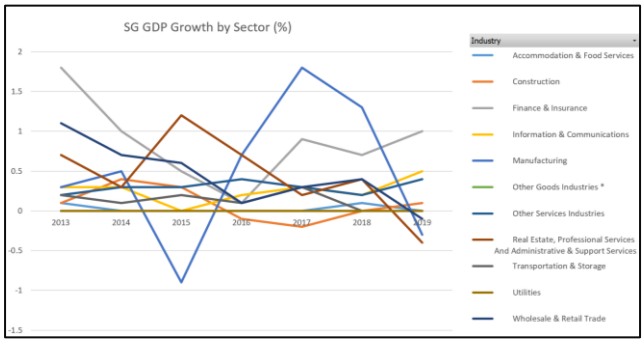


Figure 22: SG GDP growth rate by sector

Visualization 3

Figure 23 (Built on Modelling 3) shows the Overall Graduate Employment Rate of Singapore:

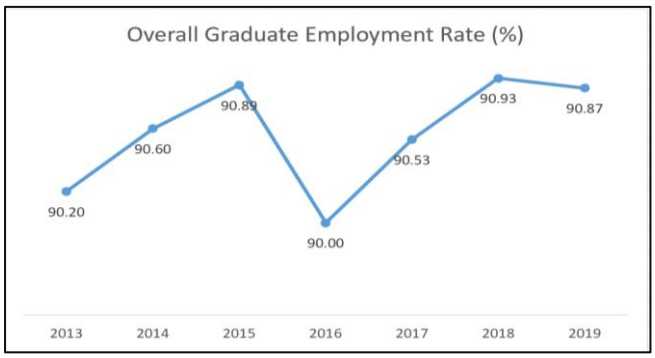


Figure 23: Overall Graduate Employment Rate by year

Visualization 4

Figure 24 (Built on Modelling 4) explores the Graduate Employment Rate of Singapore by Education Cluster:

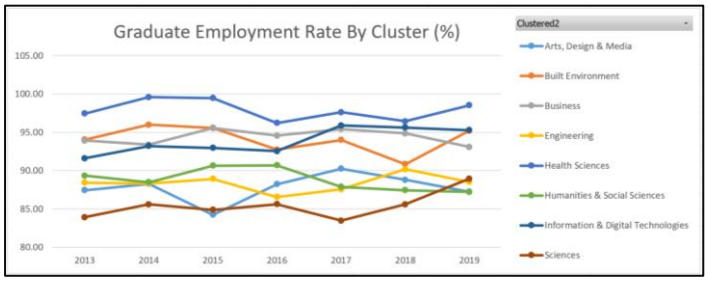


Figure 24: Graduate employment rate by cluster

Visualization 5

Figure 25 (Built on Modelling 5) plots the trendline of Singapore's GDP and Employment (Correlation calculation was performed as well):

Year	GDP At Current Market Prices [In millions (SGD)]	Employed (thousands)	Overall Employment [Millions]
2013	384870.3	3352.9	3.35
2014	398947.9	3440.2	3.44
2015	423444.1	3516	3.52
2016	440372.2	3570	3.57
2017	474115.1	3550.1	3.55
2018	507123.9	3575.3	3.58
2019	510737.8	3631.7	3.63
Correlation between Singapore GDP and Employment	0.888445239	There is a very strong positive correlation between GDP and Overall Employment	

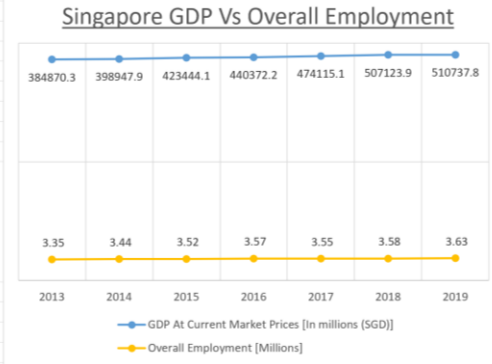


Figure 25: Trendline and Correlation between SG GDP and overall employment

A scatterplot was also used in Tableau to show the correlation between the two variables as shown in Figure 26:

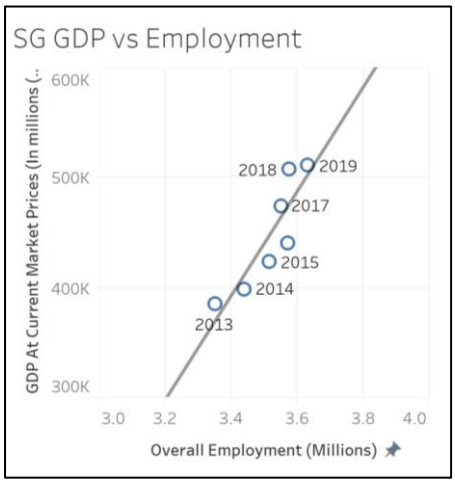
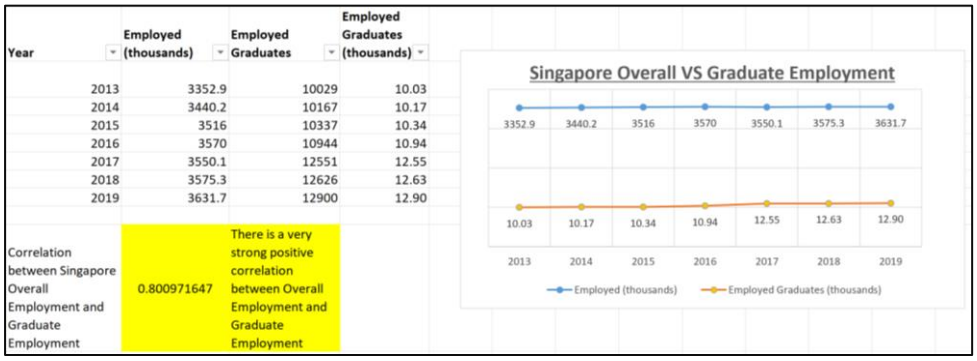


Figure 26: Scatterplot between Singapore GDP and overall employment

Visualization 6

Figure 27 (Built on Modelling 6) plots the trendline of Singapore's Employment and Graduate Employment (Correlation calculation was performed as well):



Year	Employed (thousands)	Employed Graduates	Employed Graduates (thousands)
2013	3352.9	10029	10.03
2014	3440.2	10167	10.17
2015	3516	10337	10.34
2016	3570	10944	10.94
2017	3550.1	12551	12.55
2018	3575.3	12626	12.63
2019	3631.7	12900	12.90

Correlation between Singapore Overall Employment and Graduate Employment	0.800971647	There is a very strong positive correlation between Overall Employment and Graduate Employment

Figure 27: Trendline and Correlation between SG overall and graduate employment

Similarly, a scatterplot was utilised to better visualise the correlation between these two variables as shown in Figure 28:

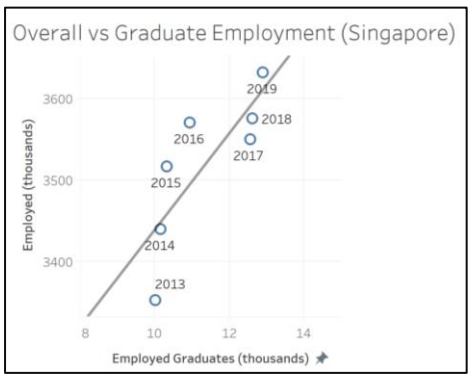


Figure 28: Scatterplot between SG overall and graduate employment

From the outputs of Visualisations 5 and 6, we have the metrics to evaluate correlation between the two key sets of variables for this analytics project:

Set of variables	R-value	P-value
Singapore GDP & Employment	Strong positive correlation of 0.89	0.007
Singapore Employment & Graduate Employment	Strong positive correlation of 0.80	0.03

Table 6: Correlation and P -Value

Section Findings

With both sets of correlation achieving better than recommended values in both evaluation metrics (R-value > 0.7 and P-value < 0.05), we can assume that there exists a linear relationship between Singapore's GDP, Overall Employment and Graduate Employment. However, we must keep in mind that correlation does not necessarily equates to causation; it just means that these 3 variables move in tandem.

Given the sequence of Visualisations 5 and 6, we presume that when GDP goes up, Overall Employment goes up and when Overall Employment increases, so does Graduate Employment.

Hence, working backwards we can see that GDP is the underlying factor for the increase in Graduate Employment. Thus, the team will investigate what are the main contributors to Singapore's GDP and recommend re-calibration of education resources for the respective industries accordingly so that the Ministry of Education can better anticipate the respective sectors growing or decreasing need for skilled workers.

J. Project Findings

Project Findings 1- "Unproductive" courses identified

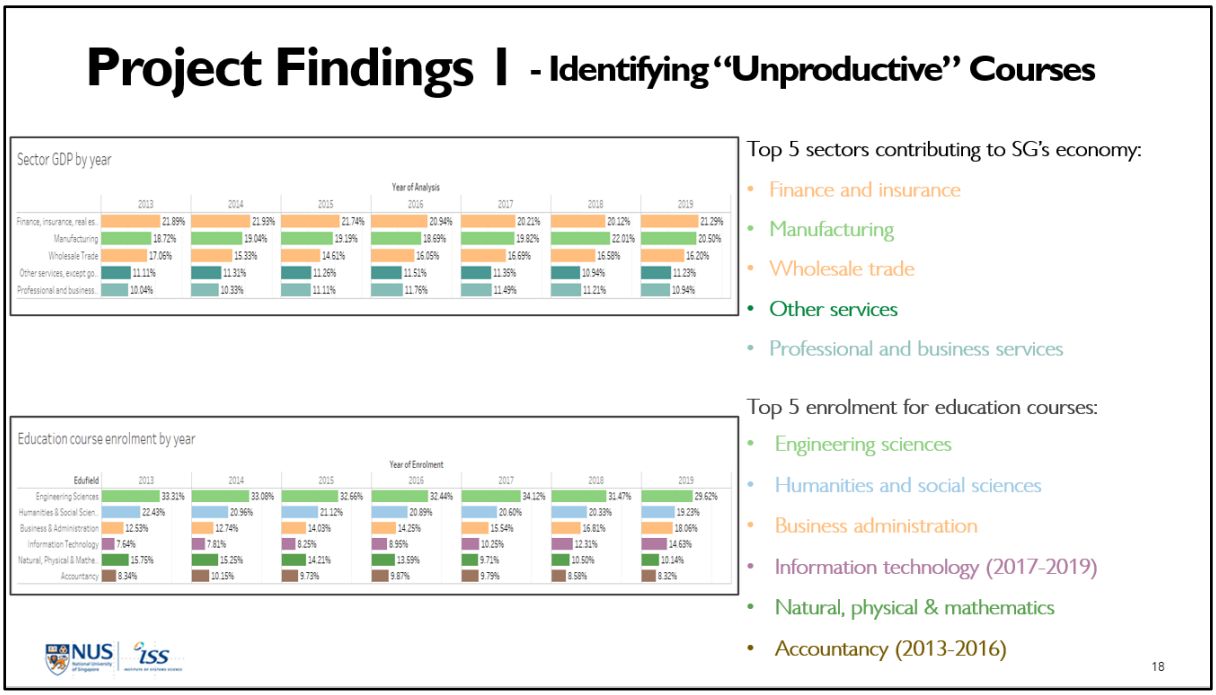


Figure 29: Project Findings 1-Identifying "Unproductive" Courses

For project findings 1, the top 5 industry sectors ("sector") contributing to Singapore's economy, as well as the top 5 education cluster ("cluster") enrolment by year were identified, as shown in the above figure.

For industry sectors: Finance and insurance, Manufacturing, wholesale trade, other services, professional and business services.

For education clusters: Engineering sciences, Humanities and social sciences, Business administration, Information technology (2017-2019), Natural, physical and mathematics, Accountancy (2013-2016)

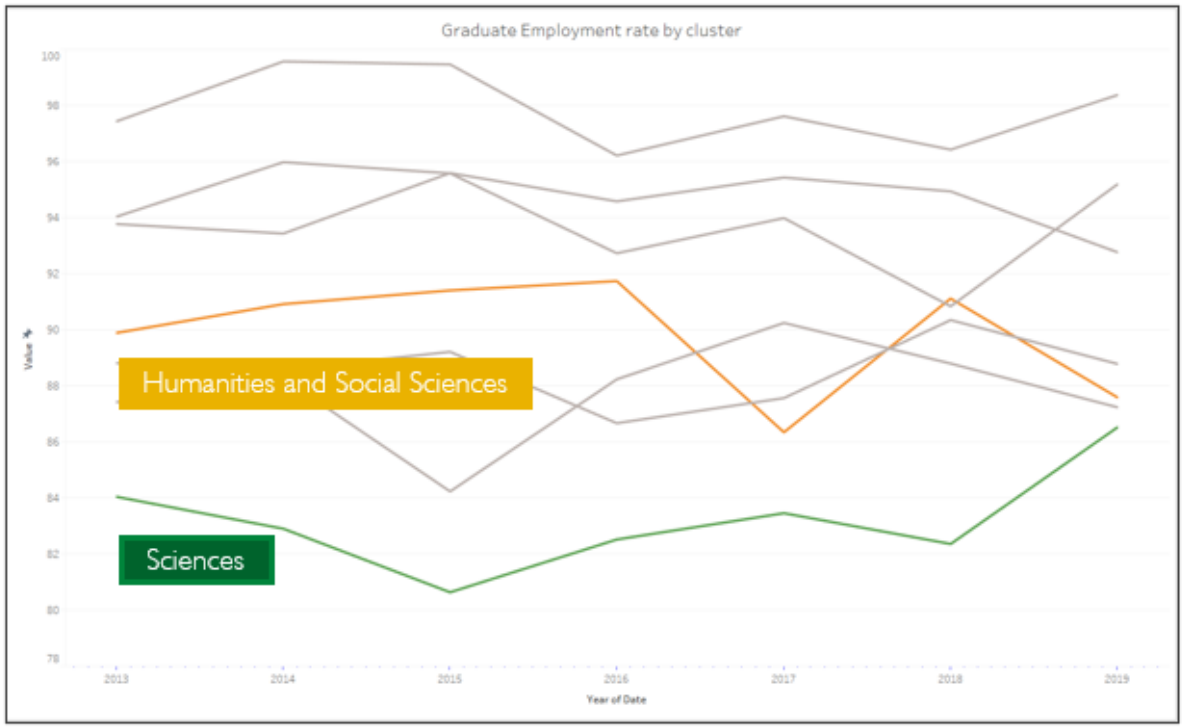


Figure 30: Graduate employment rate by cluster

Graduate employment rate by education cluster was also taken into consideration, as shown in figure 30 above.

Project Recommendation 1- "Unproductive" courses identified

Using data from industry sector GDP contribution and education cluster enrolment, we mapped the industry sectors to the education course clusters based on the core competency required for each industry sector. For example, Manufacturing (industry) was mapped to Engineering Sciences (course cluster) and Wholesale Trade (industry) mapped to Business Administration (course cluster). Based on the mapping in the figure below, most education course clusters could be mapped to a top 5 industry sector in terms of GDP contribution, lest for 2 anomalies. The anomalies were Humanities and Social Sciences as well as Natural, Physical and Mathematics.

Course Rank	Course	Industry Sector	Sector Rank (GDP)
1	Engineering sciences	Manufacturing	2
2	Humanities and social sciences	?	
3	Business administration	Finance and insurance Wholesale trade	1 3
4	Information technology	Information	8
5	Natural, physical & mathematics	?	

Figure 31: Industry sector and education cluster mapping

The observed anomalies indicate that despite having a significant enrolment size, these education clusters do not seem to have a clear contribution to GDP. For Humanities and social sciences, enrolment was 15.9% of the 2019 cohort and for Natural, physical and mathematics, enrolment was 8.43% of the 2019 cohort. This suggests that there might room for optimization or augmentation for these two courses to better align with Singapore economy.

Analysing the graduate employment survey (Figure 30) provided additional support to the above insight, where the two course clusters also held the lowest employment rate numbers across the education clusters, further suggesting that such graduates have the lowest market demand and that there is room for optimization.

Project recommendation 1 suggests that these education clusters should augment their faculties, courses, or curriculum to incorporate skillsets required for high contributing and growing industry sectors. For example, NUS is already leading this shift by merging the Faculty of Arts and Social Sciences (FASS) with Sciences, with a focus to include new competencies and skillsets for a digital world in the curriculum. We concur that there must be a focus on digital competencies and skillsets as no industry today is spared from the digital cataclysm.

We therefore strongly recommend that other universities to investigate these course clusters and start augmenting them for the digital world.

Project Findings 2- Misaligned courses identified

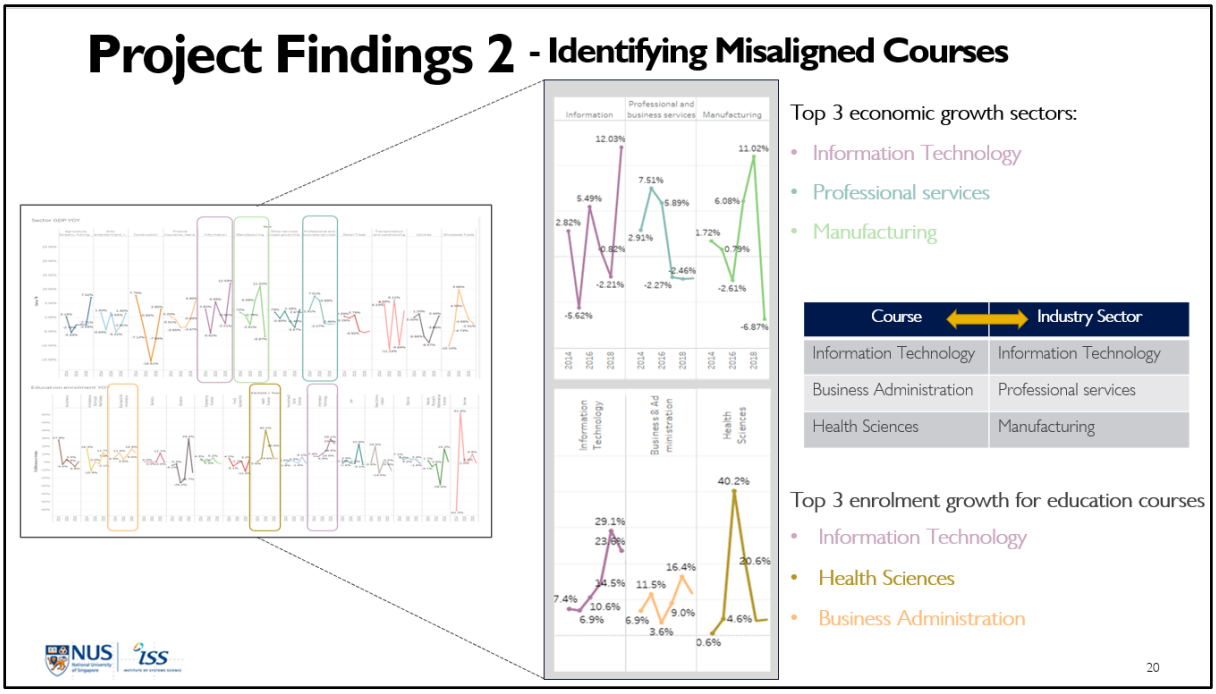


Figure 32: Project Findings 2-Identifying misaligned courses

For project findings 2, the top 3 industry sectors with highest year-on-year GDP growth, as well as the top 3 education cluster with highest year-on-year employment growth were identified, as shown in the above figure.

For industry sectors: Information technology, Professional services, Manufacturing

For education clusters: Information technology, Health sciences, Business Administration

These data points were based on the past 3-year average.

Information technology→ 3.54%

Business Administration→ 3.41%

Wholesale trade→ 0.35%

Project Recommendation 2- Misaligned courses identified

Continuing with the approach in project finding 1, the team mapped the industry sectors to the education course clusters based on the core competency required for the industry sectors, as per the table below.

Course	Industry Sector
Information Technology	Information Technology
Business Administration	Professional services
Health Sciences	Manufacturing

Figure 33: Education cluster to industry sector mapping

Based on the mapping, all education course clusters with the greatest enrolment growth could be mapped to a top 3 industry sector with the greatest GDP growth. Thus, this suggests that our university enrolment growth numbers appear to be aligned with our economic growth sectors. That is, the hot economic sectors are accompanied with the strongest university enrolment growth.

However, when taking a deeper look, the team noticed that for Information Technology in particular, the GDP grew 12.03% whilst the education cluster enrolment grew by 23.6%. Looking at this data point at face value, hints that our education system is producing an over-supply of IT graduates. Upon further analysis, this data does reflect the economy of today, as Information Technology is cross cutting across almost all industry sectors.

In this age of disruptive economy, the rise of new technology companies challenges the notion of traditional companies. For example, GRAB which started out providing ride-hailing service would be considered a transport company traditionally, yet it is driven by technology to co-ordinate demand and supply of transportation needs via a mobile application. Likewise for a manufacturing company that might have automated a lot of the processes via robots, the company would now rely heavily on a different skillset such as robotics as compared to purely manufacturing sciences.

In addition, with the upsurge in upskilling and advocacy of lifelong learning, the workforces' graduate degrees may no longer accurately reflect the nature of work or sector that they may be employed in.

This suggests that as the world and economy changes, the traditional way of categorising and collecting data may need to be updated.

Project recommendation 2 suggests the following for deeper analysis and richer insights:

For industry sectors: Categorisation of industries needs to be updated to reflect the core and supporting competencies required in the company. For example, is GRAB a transport or IT company? The ideal would be to also measure GDP not via industry sectors in silo, but rather by the nature of work. That is GDP to be also measured horizontally, instead of just vertically.

For graduate employment data: To capture current employed sector, nature of work and relevant qualification acquired for the job.

Project Findings 3- Identified sectors to re-allocate resources to

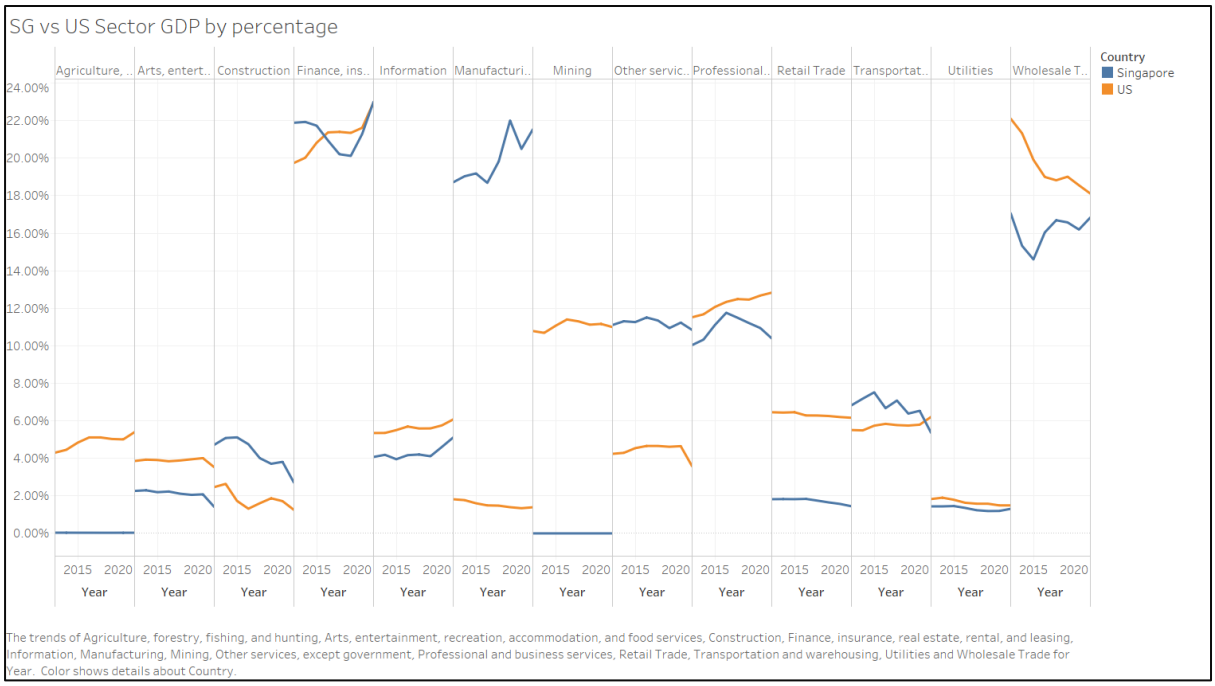


Figure 34: Project Findings 3 -SG vs US Sector GDP by Percentage

Given US is one of the world's first developed country hence, considered more advanced in terms of economic growth and development than Singapore. By looking at the GDP trends between these two countries, we hope to gain some insights as to how the future contribution to GDP by sector might potentially look like for Singapore.

Publicly available was obtained and grouped according to following sectors:

- Agriculture, forestry, fishing, and hunting
- Arts, entertainment, recreation, accommodation, and food services
- Construction
- Finance, insurance, real estate, rental, and leasing
- Information
- Manufacturing
- Mining
- Other services, except government
- Professional and business services
- Retail Trade
- Transportation and warehousing
- Utilities
- Wholesale Trade

Sectors such as 'Agriculture, forestry, fishing, and hunting' and 'mining' has been excluded in our analysis due to the different kind of natural resources available to Singapore.

Arts, entertainment, recreation, accommodation, and food services:

There is a drop of GDP Contribution by both countries, this might be due to pandemic and it should not be taken at face value.

Construction:

Contribution by this sector has been declining. For developed countries, there would not be as much construction as developing countries. We can see US's contribution is much lower, and it can be used as a good predictor for said industry in Singapore.

Construction related education (such as architecture) can be reduced.

Finance, insurance, real estate, rental, and leasing:

Contribution by this sector is on the rise, except for Singapore in 2018. This might be due to rising interest rates around the world and trade wars affecting stock prices. We can see US is not as unaffected. Singapore's financial and real estate sector might be more susceptible than US' to global events. However, in the years where there is a decline in contribution by Singapore, it is stagnant for US counterparty. It has since bounced back and has been on the rise.

Education in this sector can be increased accordingly.

Information:

Information sector is mostly identified with computer programming and system design. There is a steady rise in Singapore and US. It is a larger contributor to GDP for US than Singapore and it can be used as a forecast of what will happen to the same sector in Singapore.

Education in this sector can be increased accordingly.

Manufacturing:

Manufacturing industry contributes to more than 20% of Singapore's GDP and our government do have plans to further grow this industry by going into 'advanced manufacturing'. As this is the direction by the government, we are unable to advise on whether we should increase or reduce intake based on our data.

Other services, except government:

Consistent for both countries, we are unable to determine graduates from which sector contributes to these and we are unable to advise based on our data.

Professional and business services:

This sector has been decreasing in Singapore, however there is a steady rise for US. There might be potential for this sector and the MOE can consider ramping up education in this sector.

Retail Trade:

This sector has seen a slow decline. Due to pandemic, technology is accelerated and there is a decrease in retail activities. This will be the direction moving forward as technology grows into a larger part of our lives. Education in this sector can be decreased accordingly.

Transportation and warehousing:

There is a decline in transportation for Singapore, compared to increase in US. We are however unable to advise as we are different geographically.

Utilities:

This sector is consistent for both countries. It should remain so as it is a staple part of the population's lives.

Wholesale Trade:

Wholesale Trade industry is on the rise for Singapore contrary to US decline. As there are initiatives put in place by government to digitalise this sector (https://www.imda.gov.sg/wholesale-trade-idp), we might see further growth until Singapore has hit US stage of economy. Education in this sector can be increased accordingly.

Project Recommendation 3 - Identified sectors to re-allocate resources to

Arts, entertainment, recreation, accommodation, and food services	Unable to advise
Construction	Decrease resources
Finance, insurance, real estate, rental, and leasing	Increase resources
Information	Increase resources
Manufacturing	Unable to advise
Other services, except government	Unable to advise
Professional and business services	Increase resources
Retail Trade	Decrease resources
Transportation and warehousing	Unable to advise
Utilities	Remain unchanged
Wholesale Trade	Increase resources

Table 7: Summary of project recommendation 3

K. Conclusions

✓ **Re-design / Re-invent "Unproductive" Faculties & Courses**

- Humanities & Social Sciences and Natural, Physical & Mathematics are courses with room for improvement.
- Courses should prepare for the digital world.

✓ **Re-design Graduate Employment Survey & GDP Measurement**

- Capture data points like current employed sector and current nature of work (eg. IT).
- Explore measuring GDP not just by sectors but by nature of work (Measure horizontally instead of just vertically).

✓ **Re-design / Re-invent Faculties & Courses of Sunset Sectors**

- Re-invent construction related courses such as Civil Engineering and Architecture.

L. Project Limitations

- The scope of our study only covers graduates and excludes non-graduates.
- Assumed graduates will select employment in sectors of related to course of study.
- Data is not available for certain courses or years (due to the small number of graduates and/or low response rate)
- This project only analyses the graduates from the 6 local autonomous universities in Singapore (NUS, NTU, SMU, SUTD, SIT, SUSS)
- Figures from the following universities only started in the year:
  - SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN (FROM 2013)
  - SINGAPORE UNIVERSITY OF SOCIAL SCIENCES (FROM 2017)
  - SINGAPORE INSTITUTE OF TECHNOLOGY (FROM 2017)
- Deliberately excluded 2020 data because it is a black swan year

M. Data governance

We assumed ourselves to be members of the Data & Analytics practice in MOE (Ministry of Education) and evaluated data governance considerations from that persona.

There were two key datasets which we used in our analysis **which are owned by MOE**:

1) Graduate Employment Survey (GES)
2) Universities Intake Enrolment and Graduate by Course (UIEG)

While the individual universities are responsible for the collection of the data for both datasets, MOE is responsible for the collation and publishing of the data. The datasets are published on data.gov.sg and in both their metadata files, MOE is listed as both the publisher and the source. The contacts of relevant MOE individuals are also given in the metadata files.

Governing the Key Datasets

We understand that not all data is governed data. Judging from the data quality issues observed (to be explained in detail below), it is highly possible that the two datasets are currently not governed. However, we would like to make the following recommendations to MOE to govern the datasets and make improvements to the data quality and data points which will allow for improved analysis and richer insights on whether our university courses are optimal in supporting our current and future economy. Due to the ever-changing economy landscape, this analysis has to be re-done periodically, such as once a year to ensure that our university courses are optimal.

Data Quality Issues Observed:

- **Completeness** – Based on the dataset's description found in the metadata, the GES dataset should contain SUTD's 2018 data. It is however missing.

- **Timeliness** – It is already 2021. Minimally, the 2019 GES data should be in. The more recent data is actually available (on other MOE links) but has not be consolidated into the dataset. We were able to find GES data up till 2020 but they were scattered across multiple MOE links as each link contained data for a certain year and for a certain university.

- **Consistency** – Between the GES and UIEG datasets, the way the courses are clustered are different. For example, in GES, we saw "Engineering" while in UIEG, we saw "Engineering Sciences" instead. While we could find the underlying degrees that made up the clusters for the GES, we could not find the same for UIEG. This made it hard for us to compare apple against apple in our analysis and some guesswork had to be done.

Recommendations to tackle Data Quality Issues:

- To prevent re-inventing the wheel and to find out if we could leverage on any existing structures or practices, we did some online research on the existing data governance structure and practices of the MOE. However, we were not able to find much about it. The only piece of information found was that MOE does have some existing data governance structure in the form of its Research and Management Information Division.

- **On Structure:** Based on the data quality issues identified, it is likely that these two datasets are currently ungoverned. This would also mean that there are no data stewards assigned to govern the data. From our research, it appears that the subject matter expert of the

datasets will be a senior person from the Higher Education Planning Office and thus, someone of this background should be assigned as the data steward for the datasets.

- **On Trifecta:**

Data Quality Issues	Policies / Procedures	Performance Measures
Completeness	• Include in **data quality policy** that data should be **100% complete** • Conduct data quality audit once a year	• 0 missing data found during data quality audit • Automate checks within 1 year
Timeliness	• Include in data quality policy that data should be **refreshed every 6 months** • Conduct data quality audit once a year	• 0 deviations found during data quality audit • Automate checks within 1 year
Consistency	• Data dictionary of how courses are clustered • Conduct data quality audit once a year	• 0 deviations from data dictionary during data quality audit • Automate checks within 1 year

Table 8: Trifecta

Beyond the policies and procedures suggested above, we also recommend MOE to explore working with CPF to automate some of the data collection for these datasets. For example, through CPF contribution data, we can determine whether the graduate is employed, when he was employed, which industry he is employed in, etc. This automation will:

1) Increase number of collected records - Tackle the issue of no responses (~20%) in the survey.
2) Provide for more data points such as time taken to find employment, industry, etc for richer analysis and insights.

Data Governance Recommendations on our Analysis & Insights (Output):

- **Assigning the "Right" Data Steward**
  - One whose interests is aligned with the value of our output.
  - One that has policy making authority or influence with regards to higher education.
  - One that has a hunch on whether our output "make sense".
  - **From our research, this person should be a senior person in MOE's Higher Education Planning Office**

- **Data Quality**
  - Comply with the data quality policies set by Research and Management Information Division.
  - (If not already done as part of the above) **Peer-reviewed by at least two senior MOE staff, with 1 person from the Higher Education Planning Office (SME in terms of higher education data and policies) and 1 from the Research and Management Information Division (SME in terms of data analytics and data governance in MOE).**

- **Data Security**
  - Output to be properly **classified** according to MOE's data classification bands.
  - Assuming that existing bands are Confidential, Restricted, Internal & Public. **Output is minimally Internal as insights are sensitive due to it having an impact on higher education policy making.**

- **Data Compliance**
  - As only aggregated data is being displayed, there is no personal identifiable information in the output. **Thus, there are no concerns with regards to PDPA.**

N. Risks and Challenges Encountered

Below are the key risks and challenges encountered together with the corresponding mitigation measures.

Legend:

- Risk Scale for Impact and Probability: 1 – Low, 2 – Medium, 3 – High.
- I : Impact | P : Probability | R : Risk Exposure (I X P)

S/N	Risks	I	P	Explanation	R	Mitigation
1	**Missing Data Records** • Missing records in the datasets. • After conducting a quick check on our datasets, we found out that records were missing in one of our key datasets – Graduate Employment Survey. The 2018 SUTD records were missing.	3	3	• The impact is high as missing data will result in us making wrong the analysis and resulting in inaccurate insights. • The probability is high as we detected missing records from a quick check.	9	• Do a thorough check by pivoting the dataset and finding any missing records. • Supplement the missing records by finding other data sources.
2	**Data Unavailability** • We require GDP and graduate's data from other countries such as the USA for our analysis. • Preliminary findings are that such data are hard to find for foreign countries.	2	3	• The impact is medium as this is a supplementary analysis in our project. • The probability is high as no data is easily available from a quick search.	6	• Web scrape of the data using R (Source: FRED economic data) is carried out to collect data which we cannot find.
3	**Insufficient Time** • Project is on top of existing responsibilities.	3	2	• The impact is high as a lack of time will directly impact the effort and quality of our project. • The probability is medium as we have verbally expressed our commitment to the project.	6	• Commitment to block out time to collaborate on the project. • Constant updates & communication. • Employ use of burn down charts to track project progress carefully

4	**Unexplainable Findings** • Findings / Insights uncovered deviates significantly from our hypothesis. • Unable to make sense of findings and provide actionable recommendations.	3	1	• The impact is high since this would mean that our project has failed its goal. • The current assessed probability is low based on our preliminary research. There have been articles published by academics that supports our hypothesis.	3	• Conduct background research to look for evidence and studies which support our hypothesis. • Conduct high level analysis such as simple correlation to quickly validate hypothesis.
5	**Unclear Objectives** • While the project team (us) were aligned on the general theme of the project. • We struggled to define clear business objectives on our first inception iteration.	3	1	• The impact is high as an unclear scope will result in vague outcomes. • The probability is low as the team is committed to come to a consensus on the business objectives.	3	• Conduct 1 more inception iteration to thrash out what the business objectives should be before concluding the inception phase.

Table 9: Risk and challenge table

O. References

Listed below are the data files that were used:

Singapore - Graduates from Higher Degree Courses By Type Of Course, Annual

Source: https://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=15200

Singapore - GDP per capita (current US$)

Source: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=SG

Singapore - GDP growth (annual %)

Source: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=SG

US – GDP per capita (current US$)

Source: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=US

US - GDP growth (annual %)

Source: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=US

US – Bachelor's degrees awarded, by field

Source: https://ncses.nsf.gov/pubs/nsb20197/trends-in-undergraduate-and-graduate-s-e-degree-awards

Graduate Employment Survey - NTU, NUS, SIT, SMU, SUSS & SUTD

Source: https://data.gov.sg/dataset/graduate-employment-survey-ntu-nus-sit-smu-suss-sutd

2020 GES Employment Rates and Salaries of NUS, NTU, SMU and SUSS Graduates by Course Cluster

Source : https://www.moe.gov.sg/-/media/files/post-secondary/ges-2020/joint-web-publication-4-aus-ges-2020.pdf

Universities - Intake, Enrolment and Graduates by Course

Source : https://data.gov.sg/dataset/universities-intake-enrolment-and-graduates-by-course?resource_id=115bf8a7-46df-466c-b7fc-375ef3c1b425

Graduates from Higher Degree Courses By Type Of Course, Annual

Source : https://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=15200

Intake, Enrolment and Graduates by Institutions

Source : https://data.gov.sg/dataset/intake-enrolment-and-graduates-by-institutions?resource_id=2264a6ed-51f5-45d6-accb-1a980e32e632

FRED Economic Data: Gross Output by Industry, Billions of Dollars, Seasonally Adjusted at Annual Rates

Source: https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2013-01-01

Source: Okun's Law: Economic Growth and Unemployment (investopedia.com)

Source: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=SG

Trifacta blog: Cleaning Dirty Data & Messy Data

https://www.trifacta.com/blog/messy-data/

Annex A: R Source Code for Data Set 5

```
install.packages('tidyverse')
install.packages('rvest')
install.packages('writexl')
library('rvest')
library('tidyverse')
library("writexl")

#2013 Q1 to 2

url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2013-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2013_A.xlsx')

#2013 Q3 to 4

url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2013-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2013_B.xlsx')

#2014 Q1 to 2

url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2014-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2014_A.xlsx')

#2014 Q3 to 4

url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2014-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2014_B.xlsx')

#2015 Q1 to 2

url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2015-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2015_A.xlsx')
```

#2015 Q3 to 4

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2015-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2015_B.xlsx')
```

#2016 Q1 to 2

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2016-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2016_A.xlsx')
```

#2016 Q3 to 4

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2016-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2016_B.xlsx')
```

#2017 Q1 to 2

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2017-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2017_A.xlsx')
```

#2017 Q3 to 4

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2017-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2017_B.xlsx')
```

#2018 Q1 to 2

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2018-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2018_A.xlsx')
```

#2018 Q3 to 4

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2018-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2018_B.xlsx')
```

#2019 Q1 to 2

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2019-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2019_A.xlsx')
```

#2019 Q3 to 4

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2019-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2019_B.xlsx')
```

#2020 Q1 to 2

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2020-04-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2020_A.xlsx')
```

#2020 Q3 to 4

```
url<-'https://fred.stlouisfed.org/release/tables?rid=331&eid=1993&od=2020-10-01#'

gdp_table<-read_html(url) %>% html_nodes('table') %>% html_table
write_xlsx(gdp_table,'C:\\Users\\Ka Wei\\Documents\\NUS Y1S1\\Practice
Module\\Webscrape\\gdp_table_2020_B.xlsx')
```

Annex B: Excel Formula for Data Set 4 - Singapore GES 2013-2019

```
Function RemoveSpecial(Str As String) As String
'updatebyExtendoffice 20160303
   Dim xChars As String
   Dim I As Long
   xChars = "#$%^*&"
   For I = 1 To Len(xChars)
      Str = Replace$(Str, Mid$(xChars, I, 1), "")
   Next
   RemoveSpecial = Str
End Function
```