School of Computing and Information Systems

ISSS616 – Applied Statistical Analysis with R
Group Project Final Report

Prepared by G2 Group 7:

TAN Hui Wen

LAI Yoke Chin

TAN Shing Kwan

KWOK Pei Shan

MA Anbo

Table of Contents

1.   Introduction

Every year, Singaporean parents have to actively plan for their children's primary school education. There are five phases in the primary school registration process, as follows:

- **Phase 1**: For a child who has a sibling studying in the school. All children registered under this phase are guaranteed places in the schools.
- **Phase 2A**: For a child whose parent is an alumni, School Advisory or Management Committee member, or staff member of the school. It also covers a child who attended the Ministry of Education (MOE) Kindergarten located within the primary school.
- **Phase 2B**: For a child whose parent has joined the school as a parent volunteer, is endorsed by the church or clan directly connected with the primary school, or endorsed as an active community leader.
- **Phase 2C**: For a child who is not yet registered in a primary school.
- **Phase 2C (Supplementary):** For a child who is not yet registered in a primary school after Phase 2C.

Based on the updated MOE Primary One registration process[1], the determining factors for an application at any phase will be:

(a) distance of primary school from the residential address, and
(b) the number of vacancies at each phase

However, there is currently no one-stop platform use case that parents can use to find all the information necessary to make an informed decision on the school(s) they should apply for. This is because data around school details (e.g., school location, number of vacancies, number of applications, etc.) are dispersed across various websites.

The project team aims to collate and use the data available in these websites to build a one-stop comprehensive data visualisation tool that can help parents know the chances of their child enrolling at the nearby schools in the upcoming ballot. We aim to target parents who are planning for primary school education paths for their children. We believe this tool we build to be a very compelling use case for parents that can also be commercialised, if needed one day. The geospatial representation of data can even help couples select housing areas when they take their family planning into account.

2.   Overall Concept

Project Objectives: Develop an interactive tool through R Shiny app, covering the following:
- Visualisation that provides the geospatial view of the primary schools within set radius distance away from a postal code
- Visualisation that provides a comparison table of vacancies available, number of applicants, and number of spaces taken up at the respective phases for 2020, 2021, and 2022.

---

[1] https://www.moe.gov.sg/primary/p1-registration

Impact: One-stop comprehensive data visualisation tool which focuses on the determining factors of the Primary One registration, thus allowing parents to make informed decisions.

3.   Data
3.1.   Data Sources

We have utilised the following data sources for our project:.

Data Sources:
1) Maply[2]
2) SGSchooling[3]
3) Google Developer Console[4]
   a) Maps Javascript API
   b) Geocoding API

We used the longitudinal and latitudinal data to calculate distance from residential location (based on postal code) to the school that the parents are interested in sending their child to.

Variable	How this variable will be used
Postal code of residential location	Input from users, used to calculate the distance with other primary schools
Longitude of residential location	For Shiny app to map residential location to the longitude based on postal code input
Latitude of residential location	For Shiny app to map residential location to the latitude based on postal code input
Postal code of School location	For users to view postal code of school location in Shiny app
Longitude of School location	For Shiny app to calculate home-school distance based on postal code residential location
Latitude of School location	For Shiny app to calculate home-school distance based on postal code residential location
Number of Applicants in each Phase of Primary One registration for each Primary School	To calculate the success rate* of getting into the desired primary school at each phase based on the past 3 years data (2020, 2021, 2022)
Number of Places taken up in each Phase of Primary One registration	To calculate the success rate* of getting into the desired primary school at each phase based on the past 3 years

[2] School Location: https://geo.sg/gmaps/singapore-schools/show_data?direction=asc&sort=latitude
[3] Number of Vacancies, Applicants, and Places Taken Up: https://sgschooling.com/school/
[4] https://console.cloud.google.com/apis/dashboard

for each Primary School	data (2020, 2021, 2022)

* Success Rate = (number of places taken up / number of applicants) x 100%. We use the Success Rate to measure the probability of getting into the desired school at any given phase.

3.2. Data Preparation

The above data was available across multiple websites, the project team then manually compiled the number of vacancies, applicants, and places taken up across the last 3 years/phases. They were then able to calculate the Success Rate for each year at each phase.

There were however multiple errors in the sheet because of 2 notable exceptions:
1. No Vacancies at a certain phase
2. Vacancies but No Applicants at a certain phase ("#DIV/0!" error in the .xls)

In order to resolve these errors, the group subsequently decided to convert all cells with no vacancies to 0%. All cases with vacancies but no applicants were converted to 100%. These conversions were considered for the prospects of future applicants as these statuses would similarly reflect their equivalent in future years chances, if they were "qualified" in each phase i.e. if no applicants were there, and there were vacancies your chance to get a slot if you qualified would therefore be a 100%.

We were eventually able to come up with a dataset that spanned 181 rows (# of schools) and 23 columns (Data of success rates across phases/years + geographical information of the school), with a sample shown here

Name	Phase1_20	Phase2A_	...	Postal Sec	Postal Dis	Address	Zip/Postal	Info	Lat	Lng
PRINCESS	100%	70%		65	23	30 BUKIT E	659163	PRINCESS	1.34948	103.741
NAN HUA	100%	69%		12	5	30 JALAN	128806	NAN HUA	1.31961	103.7615

Sample Screenshot of our Schools.csv

In order to make sense of the data, we wanted to cluster the schools to see if there were schools that had success rates that were "closer" to each other. The dataset was imported into SAS JMP Pro to be run through Hierarchical Clustering because of the numerical nature of the success rates (ranging in %s) as well as the number of the schools.
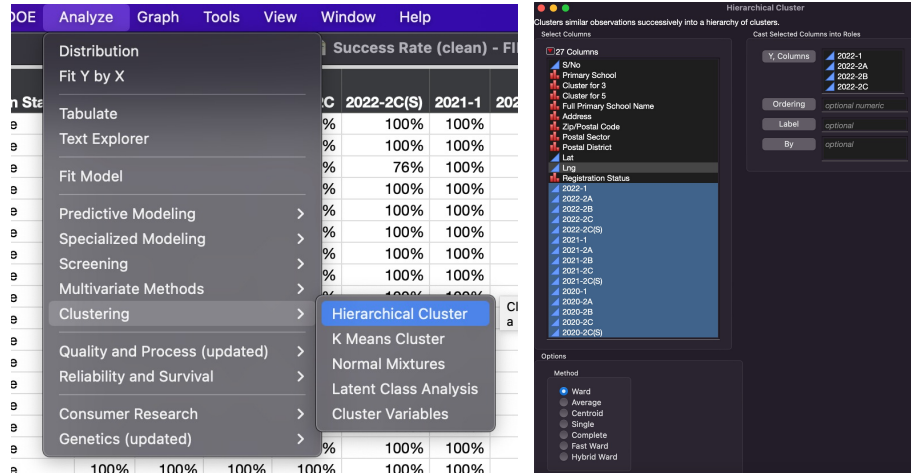
Sample Screenshot of our the modes selected in SAS JMP Pro

For hierarchical clustering in SAS JMP Pro, we selected Ward's Minimum Variance method (see screenshot above) for the clustering options selection. Ward's Method is meant to minimise the "Error Sum of Squares" between the values of each cluster as they take each next clustering step (which the group opined is doable for 181 schools). We also "clustered" using **only** the numeric continuous variables that were success rate values (see screenshot above).

From the distance between the clusters, we then took the decision to select 5 clusters from the dendritic tree:
1. At 5 Clusters, this was where the distance between clusters began to stop falling "drastically", dropping by less than a value of 1 between each cluster;
2. Qualitatively, we believed with more than 5 clusters, the classification of the schools would also lose meaning for our users.

Graph of the Distance between Clusters (Left at 180 Clusters, Right with 1 Cluster)

Clustering History

Number of Clusters	Distance	Leader	Joiner
6	8.42508755	152	159
5	9.13064339	100	104
4	10.84478655	1	34
3	13.95822677	152	161
2	18.62853147	100	152
1	29.99094696	1	100

Clustering Distances from 1-6 Clusters

This allowed us to map 5 clusters onto our dataset.

4. Descriptive Statistics

Once the clustering was completed, we were able to come up with the following dendrogram, which showcased the decay into 180 different clusters in tandem with the distance between clusters:
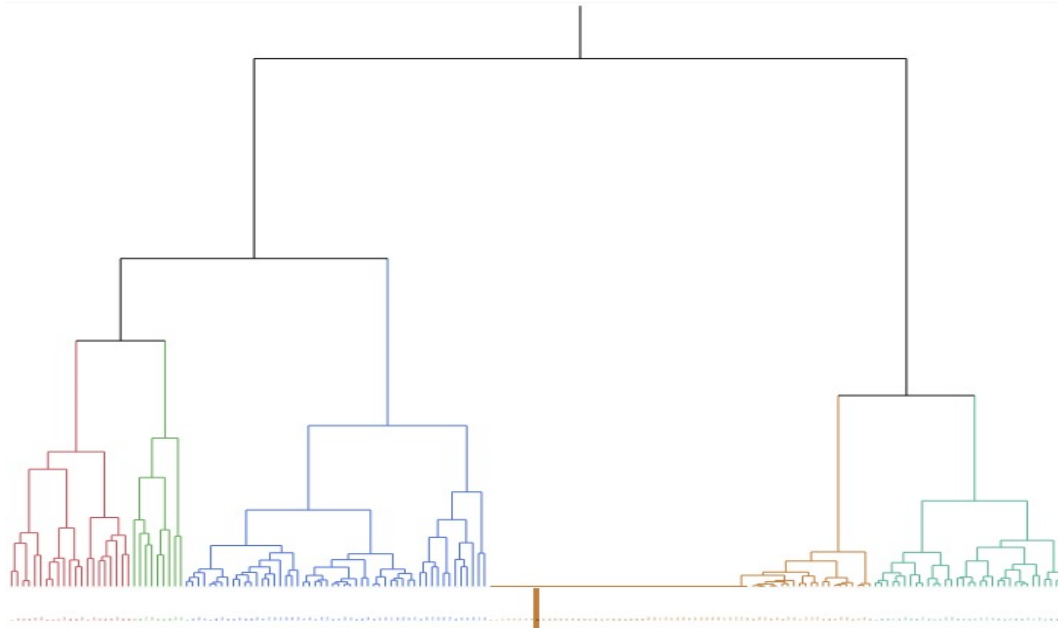
The summary statistics are also covered here as shown, Phase 1 has been excluded for analysis purpose as the success rate across all schools in Phase 1 is 100% for all three years:

Cluster	Phase 2A	Phase 2B	Phase 2C	Phase 2C (S)	Group's Renaming	# of Schools in Cluster
	Average success rates from 2020 - 2022					
1	96%	63%	49%	0%	Class B	21
2	77%	81%	34%	0%	Class A	9
3	98%	99%	65%	0%	Class C	52
4	100%	100%	100%	97%	Class E	66
5	100%	100%	96%	37%	Class D	33

Using the summary statistics, we then named the different Classes and analysed them further below.

5. Inferential Statistics

Overview Analysis for Class A, B, C, D, E Schools

Upon closer analysis, we found that the schools in each of the 5 classes generally have these properties:

Class	Description
A	The most popular schools (the most difficult schools to enrol in, usually oversubscribed in each phase of the Primary One Registration) E.g.: Nan Hua, Pei Hwa Presbyterian, Rosyth School
B	Second most popular group of schools (still quite difficult to get into these schools, especially after Phase 2B) E.g.: Ai Tong, Kuo Chuan Presbyterian, Tao Nan
C	Easier to get in compared to Class A and B schools. Will still be challenging to enrol especially in Phase 2C. E.g.: Henry Park, Princess Elizabeth, Radin Mas
D	These are the schools that are relatively easier to enrol in, but still might not be guaranteed a slot at Phase 2C(S). E.g.: Kheng Cheng, Naval Base, Tampines
E	These are the schools which are the easiest to enrol in (Usually More vacancies than applicants) E.g.: First Toa Payoh, Mayflower, Teck Whye

Class A Schools

Phase 2A:

Row Labels		Average of 2022-2A	Average of 2021-2A	Average of 2020-2A
⊟A		0.698888889	0.722222222	0.888888889
	CHIJ ST. NICHOLAS GIRLS' SCHOOL	0.8	0.64	0.95
	FRONTIER PRIMARY SCHOOL	0.76	0.74	1
	GONGSHANG PRIMARY SCHOOL	0.51	0.63	0.89
	HORIZON PRIMARY SCHOOL	0.86	0.61	1
	NAN HUA PRIMARY SCHOOL	0.69	0.85	0.92
	PEI HWA PRESBYTERIAN PRIMARY SCHOOL	0.39	0.83	0.97
	ROSYTH SCHOOL	0.65	0.79	0.98
	RULANG PRIMARY SCHOOL	0.84	0.73	0.67
	TEMASEK PRIMARY SCHOOL	0.79	0.68	0.62

From the above table, we can see that the average success rate of getting into Class A schools at Phase 2A has decreased significantly, from 89% in 2020 to about 70% in 2022. It is especially difficult to get into Pei Hwa Presbyterian Primary School in Phase 2A (the chance of getting into Pei Hwa Presbyterian at Phase 2A is only 39% in 2022).

Phase 2B:

Row Labels		Average of 2022-2B	Average of 2021-2B	Average of 2020-2B
⊟A		0.768888889	0.812222222	0.851111111
	CHIJ ST. NICHOLAS GIRLS' SCHOOL	0.43	0.5	0.66
	FRONTIER PRIMARY SCHOOL	1	1	1
	GONGSHANG PRIMARY SCHOOL	1	1	1
	HORIZON PRIMARY SCHOOL	1	1	1
	NAN HUA PRIMARY SCHOOL	0.53	0.62	1
	PEI HWA PRESBYTERIAN PRIMARY SCHOOL	0.54	0.68	0.69
	ROSYTH SCHOOL	0.83	0.68	0.64
	RULANG PRIMARY SCHOOL	0.59	0.83	0.67
	TEMASEK PRIMARY SCHOOL	1	1	1

At Phase 2B, the average success rate of getting into the Class A schools has decreased significantly, from about 85% in 2020 to about 77% in 2022.

The Class A schools that are the easiest to enrol at Phase 2B are Frontier Primary School, Gongshang Primary School, Horizon Primary School, and Temasek Primary School. Hence, parents might want to consider putting in effort to qualify for Phase 2B, such as becoming a parent volunteer in the desired school.

It would be difficult to enrol into CHIJ St. Nicholas Girls' School at Phase 2B, with a success rate of only 43% in 2022. The probability of getting into Nan Hua Primary School and Pei Hwa Presbyterian Primary School at Phase 2B is also only about half (from 2022 data).

Phase 2C:

Row Labels	Average of 2022-2C	Average of 2021-2C	Average of 2020-2C
A	0.433333333	0.295555556	0.303333333
CHIJ ST. NICHOLAS GIRLS' SCHOOL	0.63	0.33	0.43
FRONTIER PRIMARY SCHOOL	0.54	0.5	0.32
GONGSHANG PRIMARY SCHOOL	0.42	0.26	0.23
HORIZON PRIMARY SCHOOL	0.55	0.26	0.49
NAN HUA PRIMARY SCHOOL	0.21	0.15	0.19
PEI HWA PRESBYTERIAN PRIMARY SCHOOL	0.47	0.4	0.33
ROSYTH SCHOOL	0.33	0.22	0.21
RULANG PRIMARY SCHOOL	0.42	0.22	0.27
TEMASEK PRIMARY SCHOOL	0.33	0.32	0.26

At Phase 2C, the average success rate of getting into the Class A schools is consistently quite low, although it increased to 43% in 2022.

It is especially difficult to enrol in Nan Hua Primary School at Phase 2C, with a success rate of only 21% in 2022 (and average success rate of 18% if we aggregate over the 3 years). In 2022, the probability of enrolling into Rosyth School and Temasek Primary School at Phase 2C is similarly quite low at 33%, but at Phase 2B the probability is at 83% and 100% respectively. Thus, if parents are aiming to send their child to these schools, they would get a much better chance if they can apply at Phase 2B.

Class B Schools

Phase 2A:

Row Labels	Average of 2022-2A	Average of 2021-2A	Average of 2020-2A
B	0.921428571	0.985238095	0.97952381
ADMIRALTY PRIMARY SCHOOL	1	1	1
AI TONG SCHOOL	0.75	1	0.86
ANGLO-CHINESE SCHOOL (JUNIOR)	1	1	1
ANGLO-CHINESE SCHOOL (PRIMARY)	1	1	1
CATHOLIC HIGH SCHOOL	0.74	0.97	0.96
CHIJ PRIMARY (TOA PAYOH)	1	1	1
CHONGFU SCHOOL	1	1	1
FAIRFIELD METHODIST SCHOOL (PRIMARY)	0.9	1	1
HOLY INNOCENTS' PRIMARY SCHOOL	0.85	0.9	1
KONG HWA SCHOOL	1	1	1
KUO CHUAN PRESBYTERIAN PRIMARY SCHOOL	1	1	1
MARIS STELLA HIGH SCHOOL	1	1	1
METHODIST GIRLS' SCHOOL (PRIMARY)	0.78	1	1
NAN CHIAU PRIMARY SCHOOL	1	1	1
NANYANG PRIMARY SCHOOL	0.84	0.93	0.91
NORTHLAND PRIMARY SCHOOL	0.97	1	1
SINGAPORE CHINESE GIRLS' PRIMARY SCHOOL	0.97	1	1
SOUTH VIEW PRIMARY SCHOOL	0.82	0.95	1
ST. HILDA'S PRIMARY SCHOOL	0.73	0.97	0.84
ST. JOSEPH'S INSTITUTION JUNIOR	1	1	1
TAO NAN SCHOOL	1	0.97	1

From the above table, we can see that the average success rate of getting into the Class B schools at Phase 2A has decreased slightly from 2020 to 2022. However, there is still quite a high chance of getting into these schools at Phase 2A, with an average success rate of 92% in 2022.

Phase 2B:

Row Labels	Average of 2022-2B	Average of 2021-2B	Average of 2020-2B
⊟B	0.523333333	0.670952381	0.699047619
ADMIRALTY PRIMARY SCHOOL	0.45	0.83	1
AI TONG SCHOOL	0.42	0.53	0.74
ANGLO-CHINESE SCHOOL (JUNIOR)	0.47	0.61	0.44
ANGLO-CHINESE SCHOOL (PRIMARY)	0.51	0.59	0.5
CATHOLIC HIGH SCHOOL	0.47	0.57	0.51
CHIJ PRIMARY (TOA PAYOH)	0.68	0.71	0.56
CHONGFU SCHOOL	0.49	0.65	0.89
FAIRFIELD METHODIST SCHOOL (PRIMARY)	0.61	0.56	0.49
HOLY INNOCENTS' PRIMARY SCHOOL	0.42	0.45	0.71
KONG HWA SCHOOL	0.49	0.96	1
KUO CHUAN PRESBYTERIAN PRIMARY SCHOOL	0.57	1	0.67
MARIS STELLA HIGH SCHOOL	0.6	0.82	0.75
METHODIST GIRLS' SCHOOL (PRIMARY)	0.25	0.6	0.69
NAN CHIAU PRIMARY SCHOOL	0.52	0.7	0.89
NANYANG PRIMARY SCHOOL	0.54	0.65	0.44
NORTHLAND PRIMARY SCHOOL	0.91	0.75	1
SINGAPORE CHINESE GIRLS' PRIMARY SCHOOL	0.59	0.84	0.77
SOUTH VIEW PRIMARY SCHOOL	0.67	0.91	1
ST. HILDA'S PRIMARY SCHOOL	0.5	0.41	0.54
ST. JOSEPH'S INSTITUTION JUNIOR	0.45	0.6	0.71
TAO NAN SCHOOL	0.38	0.35	0.38

In Phase 2B, the average success rate of enrolling into the Class B schools has decreased significantly, from about 70% in 2020 to 52% in 2022.

In particular, it is especially difficult to enrol into Methodist Girls' School (Primary) and Tao Nan School at Phase 2B. However, it is relatively much easier to enrol into Northland Primary School at Phase 2B, with a success rate of at least 75% for three years.

It might be worth noting that for some schools, the success rate decreased significantly over the years, for example, Admiralty Primary School, Methodist Girls' School (Primary), and Kong Hwa Primary School. This may indicate that these schools became more popular among parents from 2020 to 2022.

Phase 2C:

Row Labels	Average of 2022-2C	Average of 2021-2C	Average of 2020-2C
⊟B	0.511904762	0.475714286	0.474285714
ADMIRALTY PRIMARY SCHOOL	0.43	0.54	0.63
AI TONG SCHOOL	0.49	0.3	0.31
ANGLO-CHINESE SCHOOL (JUNIOR)	0.77	0.65	0.61
ANGLO-CHINESE SCHOOL (PRIMARY)	0.8	0.63	0.57
CATHOLIC HIGH SCHOOL	0.59	0.44	0.45
CHIJ PRIMARY (TOA PAYOH)	0.79	0.77	0.71
CHONGFU SCHOOL	0.31	0.26	0.31
FAIRFIELD METHODIST SCHOOL (PRIMARY)	0.46	0.33	0.42
HOLY INNOCENTS' PRIMARY SCHOOL	0.42	0.32	0.53
KONG HWA SCHOOL	0.5	0.58	0.48
KUO CHUAN PRESBYTERIAN PRIMARY SCHOOL	0.78	0.73	0.74
MARIS STELLA HIGH SCHOOL	0.6	0.86	0.8
METHODIST GIRLS' SCHOOL (PRIMARY)	0.57	0.53	0.67
NAN CHIAU PRIMARY SCHOOL	0.3	0.31	0.32
NANYANG PRIMARY SCHOOL	0.46	0.33	0.26
NORTHLAND PRIMARY SCHOOL	0.21	0.15	0.23
SINGAPORE CHINESE GIRLS' PRIMARY SCHOOL	0.55	0.77	0.41
SOUTH VIEW PRIMARY SCHOOL	0.23	0.15	0.19
ST. HILDA'S PRIMARY SCHOOL	0.28	0.17	0.28
ST. JOSEPH'S INSTITUTION JUNIOR	0.74	0.69	0.76
TAO NAN SCHOOL	0.47	0.48	0.28

In Phase 2C, the average success rate of enrolling into the Class B schools is higher than that of Class A schools, at about 50% on average.

There are several schools which the success rate decreases significantly at Phase 2C:
- Chongfu School: the average success rate (over 2020, 2021, and 2022) to enrol in this school is 100%, 68%, and 29% at Phases 2A, 2B, and 2C respectively.
- Nan Chiau Primary School: the average success rate (over 2020, 2021, and 2022) to enrol in this school is 100%, 70%, and 31% at Phases 2A, 2B, and 2C respectively.
- Northland Primary School: the average success rate (over 2020, 2021, and 2022) to enrol in this school is 99%, 89%, and 20% at Phases 2A, 2B, and 2C respectively.
- South View Primary School: the average success rate (over 2020, 2021, and 2022) to enrol in this school is 92%, 86%, and 19% at Phases 2A, 2B, and 2C respectively.
- St. Hilda's Primary School: the average success rate (over 2020, 2021, and 2022) to enrol in this school is 85%, 48%, and 24% at Phases 2A, 2B, and 2C respectively.

For parents who would like to enrol their child at any of these schools, it would be best to qualify at Phase 2A or make efforts to qualify at Phase 2B to increase their chances.

Class C Schools

Class C is deemed as the middle class because the primary schools in this category are easier to get in as compared to the more popular schools listed in Classes A and B. However, parents will still face challenges especially in Phase 2C.

The data also highlights that there are at least 3 schools within the postal sector, postal district and area in the following table that are grouped in this Class.

Postal Sector	Postal District	Area
51	18	Pasir Ris
53	19	Hougang
82	19	Punggol
64	22	Jurong West
65	23	Bukit Batok
73	25	Woodlands
75	27	Sembawang

Analysis of the success rates at the respective phases of the Primary One Registration Process for Class C points out that three primary schools, namely Henry Park Primary School, Pei Chun Public School and Red Swastika School, have much lower success rates across all the phases as compared to the other schools. Parents who are keen on these schools may have to have alternative school plans in case they fail to enrol their child into their desired school.

In addition, the impact that MOE Kindergartens have on the success rate of Phase 2A is clearly shown for the primary schools listed in Set A below. These schools generally have lower success rates compared to those listed in Set B below. With MOE Kindergartens, there will be more applicants qualifying for Phase 2A, hence resulting in the chances of getting to the desired school in that phase to be lower. Thus, parents who are interested in the schools in Set A, will have to make efforts to qualify for Phase 2B to stand a better chance. For parents who are interested in Set B as well as those remaining schools in this Class, the parents have more options. They can work on the eligibility criteria to get themselves to qualify for earlier phases (either Phase 2A or Phase 2B) in order to have higher odds in securing their desired schools.

A	B
PUNGGOL PRIMARY SCHOOL	XINMIN PRIMARY SCHOOL
PRINCESS ELIZABETH PRIMARY SCHOOL	ST. ANTHONY'S PRIMARY SCHOOL
WESTWOOD PRIMARY SCHOOL	PAYA LEBAR METHODIST GIRLS' SCHOOL (PRIMARY)
RIVERSIDE PRIMARY SCHOOL	PASIR RIS PRIMARY SCHOOL
OASIS PRIMARY SCHOOL	MAHA BODHI SCHOOL
WEST SPRING PRIMARY SCHOOL	CANBERRA PRIMARY SCHOOL
PUNGGOL VIEW PRIMARY SCHOOL	RADIN MAS PRIMARY SCHOOL
WELLINGTON PRIMARY SCHOOL	YU NENG PRIMARY SCHOOL
	CHONGZHENG PRIMARY SCHOOL

Class D and E Schools

There are a total of 99 schools in Class D (33) and E (66), which makes up around 55% of the total number of primary schools in Singapore. For Class D and E schools, the chances of getting into Phase 2A and 2B are 100%. At Phase 2C, the average success rates are still high at 96% for Class D and 100% for Class E.

Parents who would like their child's guaranteed entry to a primary school near their homes may consider moving to the following postal districts which have the top 4 most number of Class D and E primary schools. Refer to the table and chart below for the distribution of Class D and E primary schools across the postal districts in Singapore.

- Postal district 19 (Serangoon Garden, Hougang, Punggol): 16 schools
- Postal district 23 (Hillview, Dairy Farm, Bukit Panjang, Choa Chu Kang): 15 schools
- Postal district 18 (Tampines, Pasir Ris): 10 schools
- Postal district 27 (Yishun, Sembawang): 10 schools
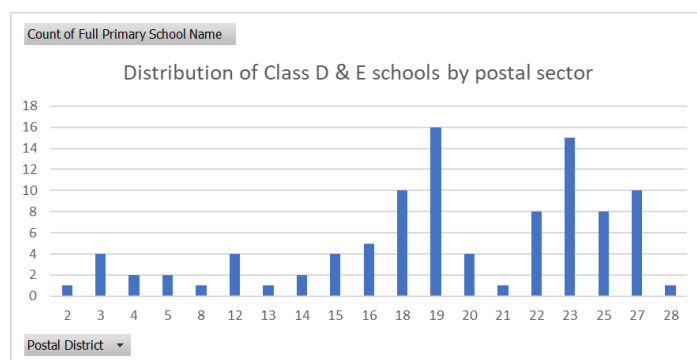
Postal districts	General location	A	B	C	D	E	# of Schools
2	Anson, Tanjong Pagar	0	0	0	0	1	1
3	Queenstown, Tiong Bahru	0	0	1	2	2	5
4	Telok Blangah, Harbourfront	0	0	1	1	1	3
5	Pasir Panjang, Hong Leong Garden, Clementi New Town	1	1	1	1	1	5
8	Little India	0	0	0	0	1	1
9	Orchard, Cairnhill, River Valley	0	1	2	0	0	3
10	Ardmore, Bukit Timah, Holland Road, Tanglin	0	1	1	0	0	2
11	Watten Estate, Novena, Thomson	0	3	1	0	0	4
12	Balestier, Toa Payoh, Serangoon	0	1	2	1	3	7
13	Macpherson, Braddell	0	1	1	0	1	3
14	Geylang, Eunos	0	1	1	2	0	4
15	Katong, Joo Chiat, Amber Road	0	1	1	2	2	6
16	Bedok, Upper East Coast, Eastwood, Kew Drive	1	0	3	1	4	9
18	Tampines, Pasir Ris	1	1	5	3	7	17
19	Serangoon Garden, Hougang, Punggol	2	2	12	6	10	32
20	Bishan, Ang Mo Kio	1	3	1	1	3	9
21	Upper Bukit Timah, Clementi Park, Ulu Pandan	1	1	0	1	0	3
22	Jurong	2	0	4	3	5	14
23	Hillview, Dairy Farm, Bukit Panjang, Choa Chu Kang	0	1	6	4	11	22
25	Kranji, Woodgrove	0	1	3	2	6	12
27	Yishun, Sembawang	0	2	4	3	7	16
28	Seletar	0	0	2	0	1	3
	Total number of schools	9	21	52	33	66	181

At Phase 2C(Supplementary), the success rates between Class D and Class E schools are more starkly different, at an average of 37% for Class D versus 97% for Class E.

Singapore Citizen children are given priority over permanent resident/ foreigner children in the primary school registration exercise. There are also caps on PR/foreigner intake for some schools. Hence, parents of children who are permanent residents or foreigners may consider

going for Class E schools, where the chances of getting into the school is still high in Phase 2C(S).

Hypothesis Testing
The key changes to the MOE Primary 1 registration process and phases which were implemented in 2022, are as follows. These changes were made to increase the chances for a child to get a place in a primary school in Phase 2C.
- Phases 2A1 and 2A2 are combined to be Phase 2A.
- Number of pre-allocated vacancies in Phase 2C is increased from 20 to 40.
- One-third of the remaining vacancies at the end of Phase 2A will go to Phase 2B, while two-thirds of the remaining vacancies at the end of Phase 2A will go to Phase 2C.

Given the above-mentioned key changes, hypothesis testing was performed on the difference in population means in Phases 2A, 2B and 2C respectively for 2021 and 2022 at 5% significance level to understand the impact on success rate at each phase before and after the change.

1. A lower-tailed test is performed for Phase 2A, with the null hypothesis that the average success rate before the change in registration process (μ_1) is higher than or equal to the average success rate after the change (μ_2).
   The null hypothesis is $\mu_1 \geq \mu_2$ and the alternative hypothesis is $\mu_1 < \mu_2$.

```
#Hypothesis Testing for population mean difference for two paired/matched dependent populations
#before: Success rates of all primary schools for combined Phase 2A1 and Phase2A2 in 2021 (previous MOE P1 registration process)
#after: Success rates of all primary schools for Phase 2A in 2022 (current MOE P1 registration process)
#lower-tailed test

data <- read.csv("C:/Users/yzand/OneDrive/Desktop/data1.csv")

#HO: mu1 >= mu2
#H1: mu1 < mu2 ('after' is better than 'before')

#Given the following sample information:
before <- c(data$A21)
before
after <- c(data$A22)
after

#t-test calculation
diff <- before - after
dbar <- mean(diff)
sd <- sd(diff)
n <- length(diff)

t <- dbar/(sd/sqrt(n))
#t = 3.939372

#lower-tailed t-test and significance level = 0.05
qt(0.95,n-1)
#qt(0.95,180) = 1.653363

#Since t = 3.939372 > -1.653363, we do not reject the null hypothesis at alpha=5%.
#Therefore, there is insufficient evidence that the average success rate at Phase 2A has increased
#after the change in the MOE Primary 1 registration.
```

As shown in the screenshot above, at significance level = 0.05, we do not have sufficient evidence to conclude that the average success rate in Phase 2A after the change has increased.

2. Similarly, a lower-tailed test is performed for Phase 2B, with the null hypothesis that the average success rate before the change in registration process is higher than or equal to the average success rate after the change (i.e., $\mu_1 \geq \mu_2$):

```
#Hypothesis Testing for population mean difference for two paired/matched dependent populations
#before: Success rates of all primary schools for Phase 2B in 2021 (previous MOE P1 registration process)
#after: Success rates of all primary schools for Phase 2B in 2022 (current MOE P1 registration process)
#lower-tailed test

data <- read.csv("C:/Users/yzand/OneDrive/Desktop/data2.csv")

#H0: mu1 >= mu2
#H1: mu1 < mu2 ('after' is better than 'before')

#Given the following sample information:
before <- c(data$B21)
before
after <- c(data$B22)
after

#t-test calculation
diff <- before - after
dbar <- mean(diff)
sd <- sd(diff)
n <- length(diff)

t <- dbar/(sd/sqrt(n))
#t = 4.059651

#lower-tailed t-test and significance level = 0.05
qt(0.95,n-1)
#qt(0.95,180) = 1.653363

#Since t = 4.059651 > -1.653363, we do not reject the null hypothesis at alpha=5%.
#Therefore, there is insufficient evidence that the average success rate at Phase 2B has increased
#after the change in the MOE Primary 1 registration.
```

As shown in the screenshot above, at significance level = 0.05, we do not have sufficient evidence to conclude that the average success rate in Phase 2B after the change has increased.

3. Finally, a lower-tailed test is performed for Phase 2C, with the null hypothesis that the average success rate before the change in registration process is higher than or equal to the average success rate after the change (i.e., $\mu_1 \geq \mu_2$):

```
#Hypothesis Testing for population mean difference for two paired/matched dependent populations
#before: Success rates of all primary schools for Phase 2C in 2021 (previous MOE P1 registration process)
#after: Success rates of all primary schools for Phase 2C in 2022 (current MOE P1 registration process)
#lower-tailed test

data <- read.csv("C:/Users/yzand/OneDrive/Desktop/data.csv")

#H0: mu1 >= mu2
#H1: mu1 < mu2 ('after' is better than 'before')

#Given the following sample information:
before <- c(data$C21)
before
after <- c(data$C22)
after

#t-test calculation
diff <- before - after
dbar <- mean(diff)
sd <- sd(diff)
n <- length(diff)

t <- dbar/(sd/sqrt(n))
#t = -2.848318

#lower-tailed t-test and significance level = 0.05
qt(0.95,n-1)
#qt(0.95,180) = 1.653363

#Since t = -2.848318 < -1.653363, we reject the null hypothesis at alpha=5%.
#Therefore, there is sufficient evidence that the average success rate at Phase 2C has increased
#after the change in the MOE Primary 1 registration.
```

As shown in the screenshot above, at significance level = 0.05, we have sufficient evidence to conclude that the average success rate in Phase 2C after the change has increased.

We understand that the intention of MOE increasing the vacancies for Phase 2C in 2022 was to increase the chances of students getting into the respective schools at Phase 2C. Our hypothesis testing shows that the average success rate at Phase 2C has statistically significantly increased in 2022 compared to 2021. This is in line with the outcome that MOE had intended to achieve.

Interactive Map Tool - Shiny App

Based on our preliminary research, there isn't an interactive tool on the internet that provides a comprehensive list of Singapore primary schools with the admission rates across different phases and interactively show their location on the map. As mentioned above in the data part, we have the Latitude and Longitude of 181 Singapore primary schools. From the side Panel numeric input, we would retrieve the postal code entered by the user.

```r
# Input: BOX for entering Postal code ----
numericInput(inputId = "housepc", label = h3("Postal Code:"),
             min = 0, max = 999999, step = 1, value = 348196),
```

The geocode function under the ggmap package would help to generate the Latitude and Longitude of the postal code entered. We are able to compute the distance between the entered postal code address and the schools, with Lat and Lon of two addresses.

```r
# distance calculation function
cul_dist <- function(Lng_1,Lat_1, S_Lng, S_Lat){
  geo_dat <- data.frame(lon = c(Lng_1, S_Lng), lat = c(Lat_1, S_Lat))
  distance <- (distGeo(geo_dat[1, ], geo_dat[2, ]))/1000
  return(distance)
}
```

In the data frame that stores all the school information, each row is one school, and the distances could be calculated and stored in the new column at the end of the rows. Using this new column called distance, we can filter out all the schools within the radius of user-entered distance (in km). The Lon and Lat can be used again to help locate those filtered schools in the map panel.

```r
mydata$filtered <- df %>%
  rowwise() %>%
  mutate(distance = cul_dist(Lng, Lat, h_Lng, h_Lat)) %>%
  filter(distance <= input$Dist_Input)
```


Note: Class A: most challenging, Class E: easiest
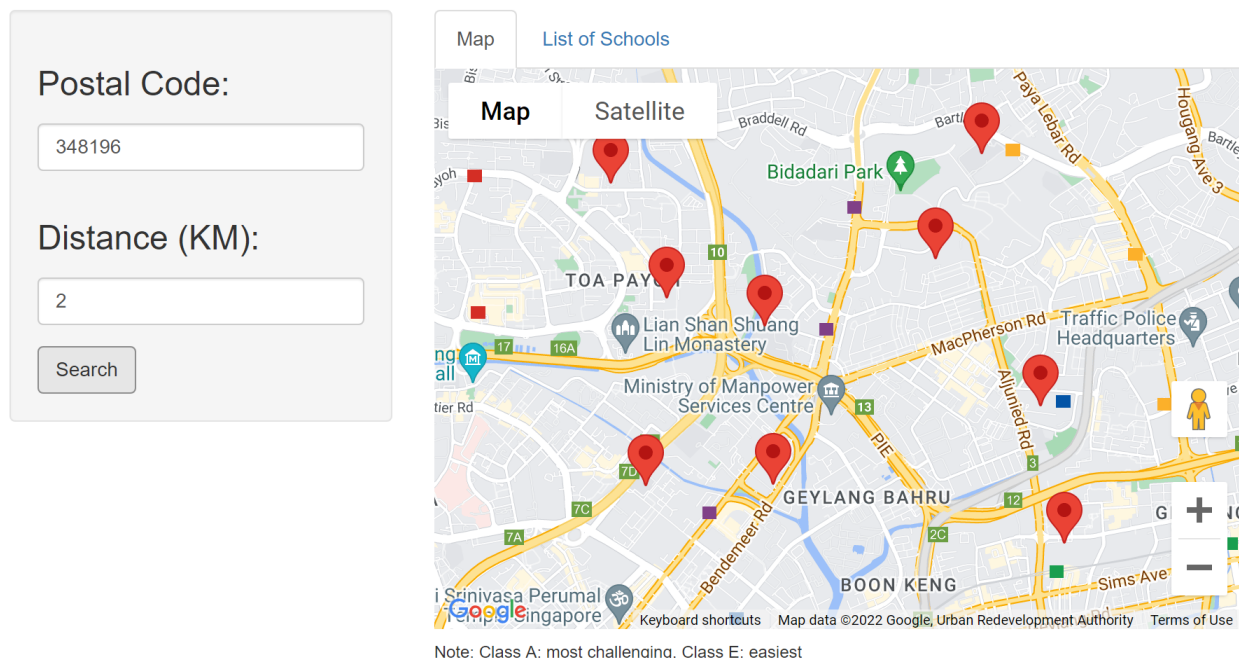
Interactive School Table Tool - Shiny App

In the Map, you can mouse over the pins to see the school name and the school class, that is all. And if you would like to know more about the schools, we have designed a panel to showcase the admission information, including the admission rate of each phase from 2020 to 2022, in the form of a table.

Map	List of Schools						
Class	**Name**	**Phase1_2022**	**Phase2A_2022**	**Phase2B_2022**	**Phase2C_2022**	**Phase2C.S._2022**	**Phase1_2021**	**Phase2A_2021**
C	RADIN MAS PRIMARY SCHOOL	100%	89%	100%	49%	0%	100%	100%
C	ALEXANDRA PRIMARY SCHOOL	100%	100%	100%	77%	0%	100%	100%
E	BLANGAH RISE PRIMARY SCHOOL	100%	100%	100%	100%	100%	100%	100%
E	CANTONMENT PRIMARY SCHOOL	100%	100%	100%	100%	100%	100%	100%
D	CHIJ (KELLOCK)	100%	100%	100%	100%	100%	100%	100%
E	GAN ENG SENG PRIMARY SCHOOL	100%	100%	100%	100%	100%	100%	100%

6. Conclusion
6.1. Limitations

- The MOE P1 registration process will fill up the vacancies with "less than 1km" applicants first before "between 1km to 2km" applicants. However, there is currently no data available on the number of places taken up by "less than 1km" and "between 1km to 2km" respectively in a phase. It will be helpful to have and analyse this piece of information to better understand the relationship between the probability of securing a place with respect to home-school distance.
- Similarly, there is also no data on the number of places taken up by Singapore Citizens and Permanent Residents respectively in a phase. Inclusion of this analysis will add another dimension to the conclusion and will be beneficial for different target groups.

6.2.   Assumptions

We made several assumptions in our project pertinent to the tool:

1. We assumed prospective chances of entry to parents were "equal" regardless of where they stayed - as we did not have further clarity on how spaces were allocated using distance data by MOE's system
2. We also assumed that the success rate is 0% for no vacancies, and success rate is 100% for cases where there were vacancies but no applicants, to reflect the prospective chances of parents should they apply for their children at these phases
3. We assumed that the data that we obtained from SGSchooling is accurate.

7.   Recommendations

Based on our hypothesis testing, we saw that the average success rates at Phase 2C have increased in 2022 compared to 2021. This is in line with the outcome that the Ministry of Education (MOE) had wanted to achieve, which was to increase the chances of success at Phase 2C.

Given that this is only the first year of change, we recommend that MOE continue to assess the relevance of the criteria in Phase 2A and Phase 2B and the number of vacancies allocated, to provide a more level playing field to students/parents who do not qualify for these 2 Phases.

Our App remains relevant and useful to facilitate parent's planning and strategizing of the registration process, as it allows parents to look into the historical success rates for each Phase and compare that across the schools of their choice.

8.   Future Work

To enhance our R Shiny Interactive App, we propose the following future work:

- Include a targeted display of success rates for the registration phases that the user can qualify for based on a Q&A filter
- Include information such as whether the school has a MOE Kindergarten affiliation (which would allow more children to qualify for Phase 2A)
- Include details on parent volunteering programs to qualify for Phase 2B (such as number of parent volunteers and the skill sets required).
- Include other details of the school such as the school website, Co-curricular activities, awards, academic track record, etc.
- Include more details for non-Singaporeans (PRs/etc) within the application
- Include click-through details of other user searches for users to understand how many "views-in-app" a school has received for users to judge school popularity based on views.

- Adjust chances depending on distance to school, instead of showing equal %s based on our assumptions today
- Publish our SHINY app by partnering with relevant commercial sites such as SGSchooling, ELITE or SCHLAH for them to drive user hits and improve UX

9. References

- MOE Primary One Registration Process: https://www.moe.gov.sg/primary/p1-registration
- Maply Primary School Locations: https://geo.sg/gmaps/singapore-schools/show_data?direction=asc&sort=latitude
- Number of Vacancies, Applicants, and Places Taken Up: https://sgschooling.com/school/
- Google Developer Console: https://console.developers.google.com/
- "Changes to P1 registration: Are you likely to face balloting at the school of your choice?" The Straits Times: https://www.straitstimes.com/multimedia/graphics/2022/06/primary-1-registration-school-ballot/index.html?shell

10. Appendix

R Packages Used:
library(shiny) : Build interactive Shiny web apps straight from R
library(dplyr): Calculation and filtering in the data frame
library(googleway): Accesses Google Maps APIs to Retrieve Data and Plot Maps
library(ggmap): Use geocode function to return Lat and Lon based on the given postal code
library(geosphere): Calculate distance based on the Lat and Lon of two addresses