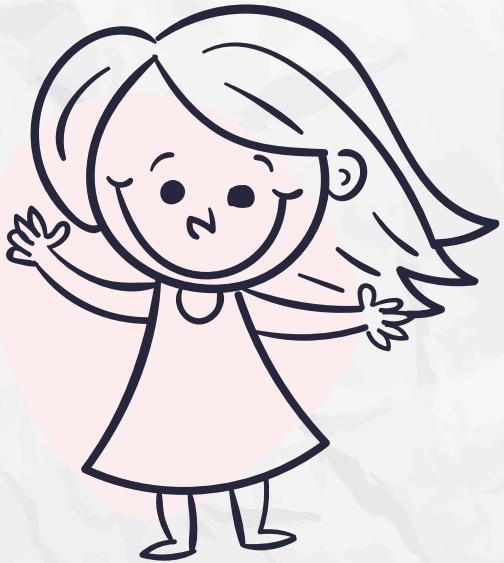


Predicting Primary Schools Application Rate

Table of contents

01

Background

Background

Many Singaporean parents go to great lengths to help secure a place in the primary school of their choice as they want to give their children the best start to their formal schooling years.

In order to have a higher chance of getting into their coveted school, many of them move closer to the school of their choice so that they can be within the school's 1-2km radius.

However, there is an increasing trend of balloting for children staying within 1km of the school at Phase 2C. Phase 2C is the phase where priority is distance-based, given to Singaporeans living within 1-2 km of the school.

today

Singapore

MOE reviewing Pri 1 registration framework to look at increasing spots for pupils living near school

Similarly, Mr Huang, MP for Jurong Group Representation Constituency, said his residents have found it challenging to obtain a spot in the primary school of their choice as the schools in their neighbourhood are usually very popular.

These residents also have to compete with parents who are former students at the school and therefore want their children to attend the school as well.

In his speech, Mr Wong acknowledged that the competition for spaces in more popular schools has intensified in recent years, and so some children do not get to attend a school near their home even with the 20 spots set aside under Phase 2C.

Source: Oh, Tessa, "MOE reviewing Pri 1 registration framework to look at increasing spots for pupils living near school", *Today Online*, March 3, 2021

Problem Statement

The annual Pri 1 registration exercise has morphed into a complicated scheme comprising several phases involving different priority rules.

We want to help parents make better decisions by using machine learning to predict the application rates for each school at each phase. This will help parents decide whether it is worth it to move closer to the school of their choice.

The goal will be to create a machine learning model that predicts the application rate of schools for next 2 years and achieve R-Squared > 0.70 for majority of the predictions.

Data Collection

The data was scraped using BeautifulSoup from Sgschooling.com, a website that collates every primary school's balloting history.

The scraped dataset shows the number of applicants and number of vacancies for each school, split into 7 different application phases. We ran a loop to obtain the data from the years 2009 to 2022.

Data Cleaning and Preprocessing

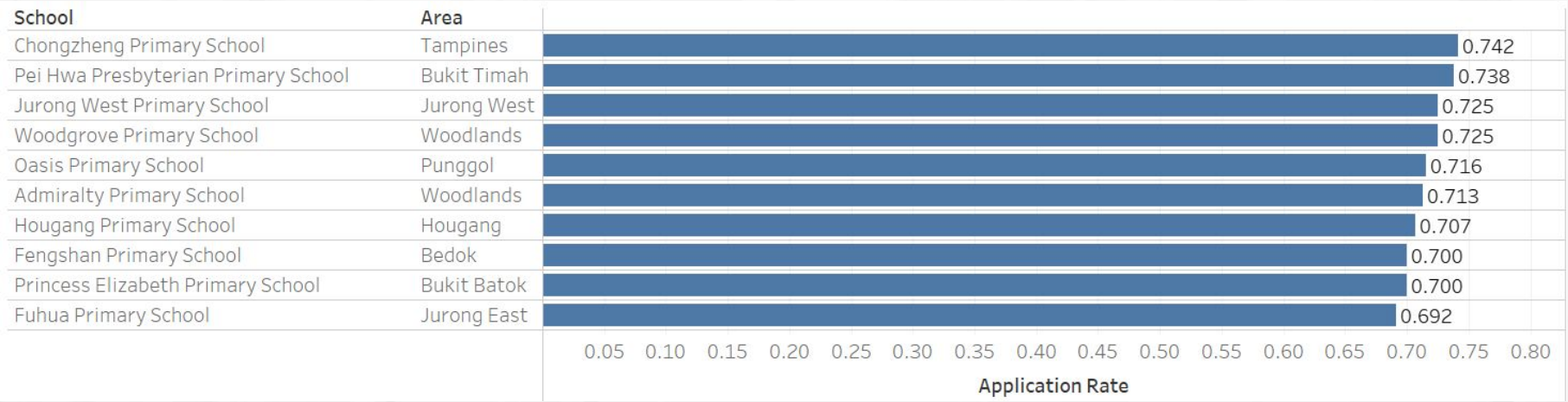
School	Phase 1	2A(1)	2A(2)	2B	2C	2C(S)	3
Admiralty							
↳ Vacancy (300)	-	-	-	63	75	0	0
↳ Applied	-	-	-	51	139	0	-
↳ Taken	168	0	7	51	75	0	-

School	Year	Phase	Vacancy	Applied	Taken	Balloting	Application Rate
Admiralty Primary School	2009	1	300	168	168		0.56
Ahmad Ibrahim Primary School	2009	1	180	57	57		0.316666667
Ai Tong School	2009	1	330	144	144		0.436363636
Anchor Green Primary School	2009	1	240	50	50		0.208333333
Anderson Primary School	2009	1	240	106	106		0.441666667
Ang Mo Kio Primary School	2009	1	210	67	67		0.319047619
Anglo-Chinese School (Junior)	2009	1	270	83	83		0.307407407
Anglo-Chinese School (Primary)	2009	1	240	60	60		0.25
Angsana Primary School	2009	1	180	26	26		0.144444444
Beacon Primary School	2009	1	240	64	64		0.266666667
Bedok Green Primary School	2009	1	210	71	71		0.338095238
Bendemeer Primary School	2009	1	150	44	44		0.293333333
Blangah Rise Primary School	2009	1	210	30	30		0.142857143

- Input missing values using Regex and subtraction
- Combined Phase 2A(1) and Phase 2A(2) into Phase 2A to be aligned with latest 2022 changes
- Rearranged the columns and rows for ease of running model
- Input full name of school
- Created Application Rate column using Applied/Vacancy

02
EDA

Top 10 Primary Schools in Phase 1 in 2022

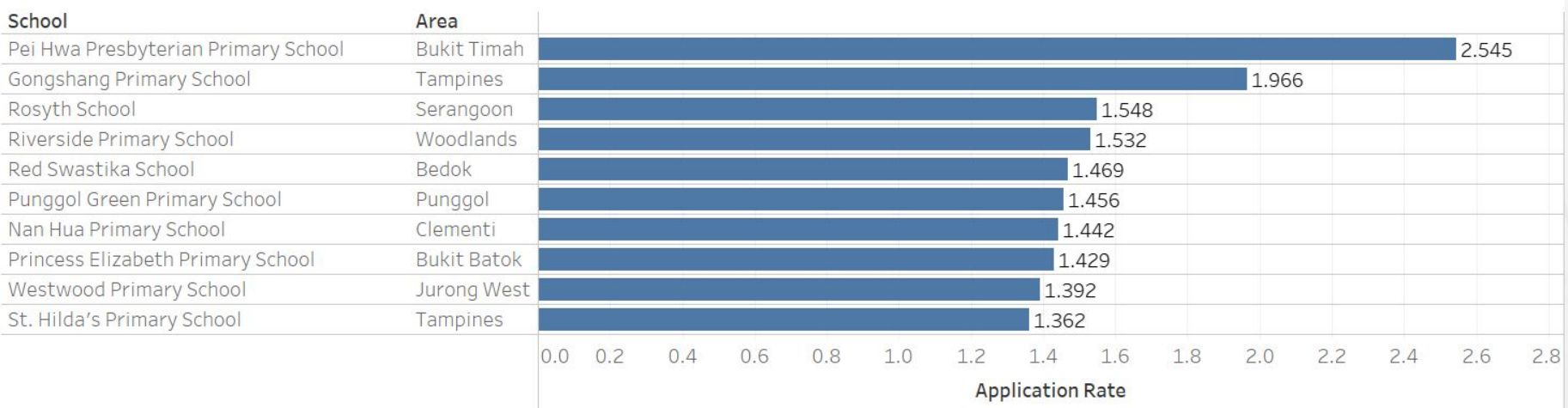


Phase 1 is for a child who has a sibling studying in the primary school.

Popular schools in Phase 1 are all mixed genders schools and all have more than half of vacancies at Phase 1 filled.

No balloting happens at Phase 1. All children with siblings in the same school will get a place.

Top 10 Primary Schools in Phase 2A in 2022



Popular schools in Phase 2A are schools with traditionally strong alumni ties like Pei Hwa Presbyterian, Rosyth, Red Swastika, Nan Hua, St Hilda's etc. These schools tend to be either GEP or SAP schools and parents probably feel that such schools have a good reputation.

Interestingly, neighbourhood schools like Gongshang, Riverside, Punggol Green and Westwood have also became very popular in the recent years at Phase 2A. It could be due to the presence of MOE kindergarten within the school.

All these schools have application rate > 1 so balloting will be required.

Top 10 Primary Schools in Phase 2B in 2022

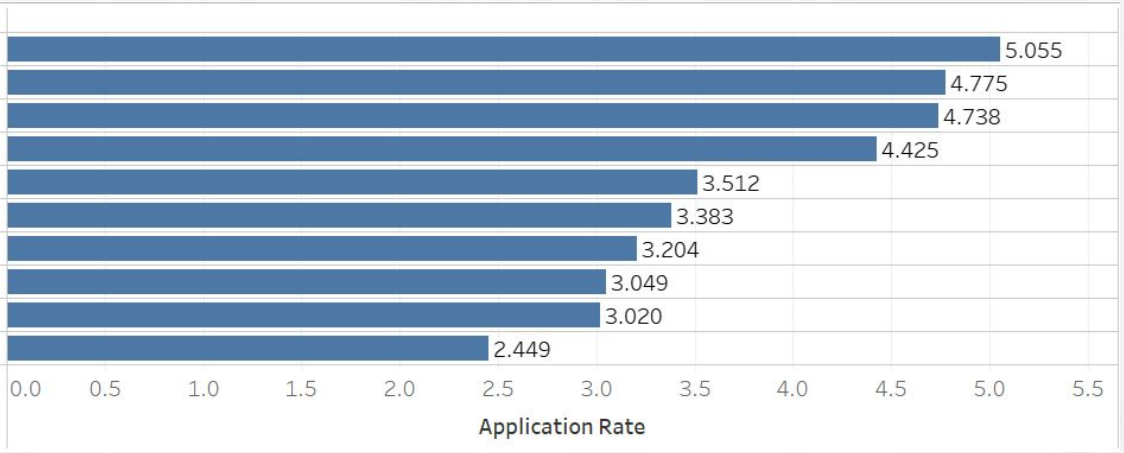
School	Area	Application Rate
Methodist Girls' School (Primary)	Bukit Timah	3.950
Tao Nan School	Marine Parade	2.652
Ai Tong School	Bishan	2.400
Holy Innocents' Primary School	Hougang	2.381
CHIJ St. Nicholas Girls' School	Ang Mo Kio	2.300
Admiralty Primary School	Woodlands	2.227
St. Joseph's Institution Junior	Novena	2.211
Catholic High School (Primary)	Bishan	2.150
Anglo-Chinese School (Junior)	Central	2.111
Chongfu School	Yishun	2.040

We see that the schools which are popular in Phase 2B are affiliated to churches like Methodist Girls', CHIJ St. Nicholas Girls', Catholic High, SJI Junior, ACS Junior etc or schools affiliated to clans like Tao Nan, Ai Tong, Chongfu etc

Balloting is very intense under Phase 2B as it does not confirm a slot even if you live within 1km of the school of your choice.

Top 10 Primary Schools in Phase 2C in 2022

School	Area	Application Rate
Princess Elizabeth Primary School	Bukit Batok	5.055
Nan Hua Primary School	Clementi	4.775
Northland Primary School	Yishun	4.738
South View Primary School	Chua Chu Kang	4.425
St. Hilda's Primary School	Tampines	3.512
Nan Chiau Primary School	Sengkang	3.383
Chongfu School	Yishun	3.204
Rosyth School	Serangoon	3.049
Temasek Primary School	Bedok	3.020
Sengkang Green Primary School	Sengkang	2.449



Phase 2C is distance based. Balloting is even more intense in Phase 2C with as many as 5 people fighting over 1 vacancy. It is certain that balloting happens for these schools within 1km so it is very risky to bet on these schools by moving within 1km of the school.

Schools like Nan Hua, Northland, South View have been consistently on the top 10 list at Phase 2C for the past 4 years.

THE STRAITS TIMES SINGAPORE LOG IN ST

New P1 registration rules from 2022: More places to be set aside under Phase 2C to ensure more kids attend school near home

Source: *The Straits Times*, September 10, 2021

Changes to P1 registration rules:

1. Phase 2C reserved 40 places, instead of 20
2. Merge Phase 2A(1) and 2A(2), so no preference of paying alumni over non-paying
3. Update Home-School Distance, so the 1km/2km radius get bigger

Impact of new PI Registration Rules in 2022

Phase	No. of Oversubscribed Schools		Change (%)
	2021	2022	
2A	18	33	+83.3%
2B	28	33	+17.9%
2C	90	88	-2.2%
Total	136	154	+13.2%

Phase 2A: Increase in number of oversubscribed schools as 20 places from this phase were moved to Phase 2C. Now non paying alumni can also apply so it leads to more alumni applying.

Phase 2B: Those who failed in Phase 2A may try at Phase 2B so it has a rollover effect

Phase 2C: Slight fall in oversubscribed schools due to additional 20 places from earlier phase.

03
Modelling

Model Used

- Dropped Phase 2C(S) and Phase 3 as these phases are not as competitive as previous phases. We will run predictions only for the competitive phases.
- Clean dataset used for modeling consists of 10044 rows
- We used Facebook's Prophet to predict the application rate at each phase for each school for the next 2 years.

Modelling process and challenges

Large number of models to be run (186 schools x 4 phases) and it is not possible to get high accuracy for all using the same parameters. Used moving average instead of raw data for higher accuracy

→ Ran the first model using 3-day moving average and filtered out rows with R-Squared > 0.7

→ Ran the second model with remaining rows using 3-day moving average with parameter changepoint_prior_scale=0.3 and filtered out rows with R-Squared > 0.7

→ Ran the third model with remaining rows using 2-day moving average

Model Evaluation

- After 3 rounds of modeling, 86.7% of the predictions have an R-Squared of more than 0.7. A total of 724 predictions were made.

R-Squared	Number of Predictions	% of Total Predictions
> 0.9	342	47.2%
> 0.8	492	68.0%
> 0.7	628	86.7%

Limitations

- R-Squared can be further improved for remaining predictions with R-Squared < 0.7. We can further tune the hyperparameters and run the model a few more times to obtain higher accuracy

- There was a major policy change in 2022 where Phases 2A(1) and 2A(2) were combined and there were higher number of vacancies allocated for Phase 2C. The current predictions may not be as accurate as there is only 1 year of data with this new policy change effected. As we collect more data over the next few years, we foresee that the accuracy of predictions will improve.

04

Conclusion

Conclusions and Recommendations

- We managed to meet our objective of using machine learning to predict future application rates with R-Squared > 0.7 for majority of the predictions

- Parents should look at the balloting history and application rate predictions to see if there is a high chance of balloting within 1km. If the competition for space at the school is too stiff, it will not be recommended for parents to specially move to a house closer to the school of their choice.

- There has been slight improvement to Phase 2C's application rate after the policy change has been implemented. It has however led to more schools balloting at the earlier phases 2A and 2B and more alumni and parent volunteers may choose to move to within 1km of the school.

Thanks!