Selecting A School in Singapore:

Beyond Grade-based Entry Requirements and Popular "Top" Schools

By DC LIM

Introduction

Singapore is a small nation-state in South-east Asia, with a population of about 5 million people. Its education system ranks among the best in the world, where its students consistently come in among the top few worldwide in PISA scores. A good education is highly prized for it brings promise of a stable, high-paying job in the future. Hence, a place in the popular "top" schools is highly sought after, right down to primary (elementary) school. Good foundations pave the way for good test scores for entry into the best schools in the next level of the education system, so goes the general thinking.

Practically every Singaporean parent with school-going children knows the game. The popular "top" schools are household names too. Some schools are also known for being the best at certain sports or competitive games. Based on performance in preliminary exams sat for in the final year of elementary and high school, each student roughly knows what "tier" of schools he/she can enter in the next level of the system. Grades are make or break, literally.

However, basing a major decision on just 1 factor is a little shoddy and boring, isn't it? What about life outside school? Where can a student hang out with friends or go on a date after school? What are the lunch options available? For the tiger moms/dads out there - are there entertainment options aplenty nearby that might distract their children from going straight home for yet more study and revision?

For this project, I will put aside test scores and "top schools", and explore the neighbourhoods of schools in Singapore to identify various groups of schools that might cater to different profiles of students and their parents.

Data

Data on Singapore schools can be obtained from data.gov.sg, a publicly-accessible database maintained by the Singapore government that contains datasets on different aspects of the country.

The dataset relevant to this project can be downloaded as a csv file from the School Directory and Information page on the website, link below. The dataset contains the names of schools in Singapore and basic information for each such as address (including postal code), phone number, level (e.g. primary/secondary school) and so on.

Postal codes are essential for the purposes of this project, since they will be geocoded to obtain the latitude and longitude of each school. This will allow us to tap on Foursquare's database to obtain the nearby venues for analysis and subsequent clustering using the K-means clustering machine learning technique.

We import the relevant Python libraries needed for this project as below, and ingest the dataset for exploration and analysis.

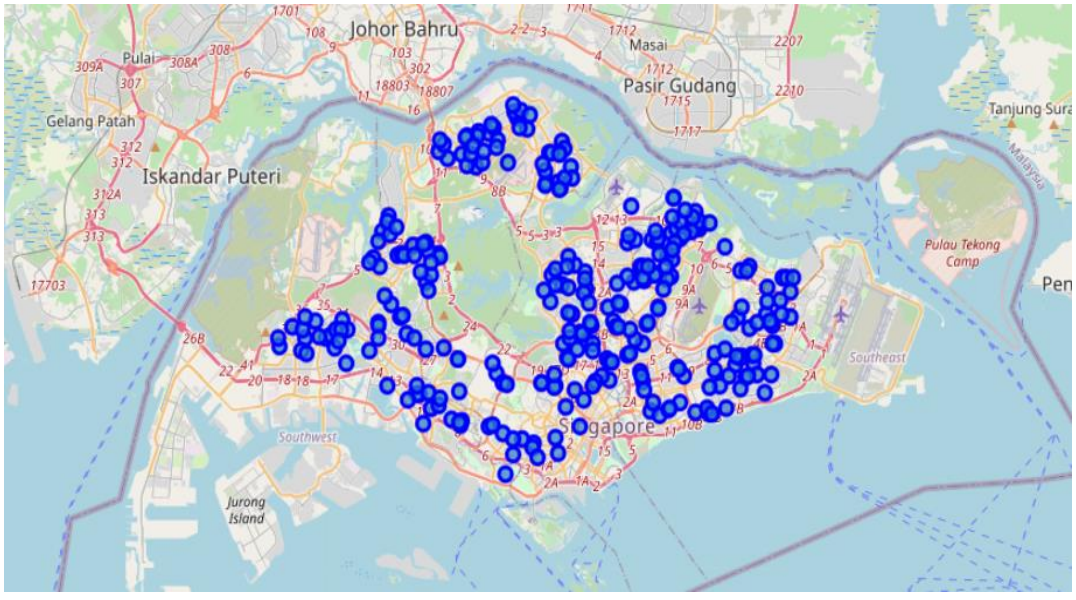
https://data.gov.sg/dataset/school-directory-and-information

An extract of the dataset is shown below:

	school_name	url_address	address	postal_code	telephone_no	telephone_no_2	fax_no	fax_no_2	email_ad
0	ADMIRALTY PRIMARY SCHOOL	http://www.admiraltypri.moe.edu.sg/	11 WOODLANDS CIRCLE	738907	63620598	na	63627512	na	ADMIRALTY_PS@MOE.ED
1	ADMIRALTY SECONDARY SCHOOL	http://www.admiraltysec.moe.edu.sg	31 WOODLANDS CRESCENT	737916	63651733	63654596	63652774	na	Admiralty_SS@moe.e
2	AHMAD IBRAHIM PRIMARY SCHOOL	http://www.ahmadibrahimpri.moe.edu.sg	10 YISHUN STREET 11	768643	67592906	na	67592927	na	aips@moe.e
3	AHMAD IBRAHIM SECONDARY SCHOOL	http://www.ahmadibrahimsec.moe.edu.sg	751 YISHUN AVENUE 7	768928	67585384	na	67557778	na	aiss@moe.e
4	AI TONG SCHOOL	http://www.aitong.moe.edu.sg	100 Bright Hill Drive	579646	64547672	na	64532726	na	AITONG_SCH@MOE.ED

Methodology

We use the Python Nominatim library to centre the map on Singapore and the geopy library to geocode the postal codes of the Singapore schools. We then add the latitudes and longitudes of the schools to the original dataset, and use it to create a map of Singapore with the schools indicated by markers, as shown:



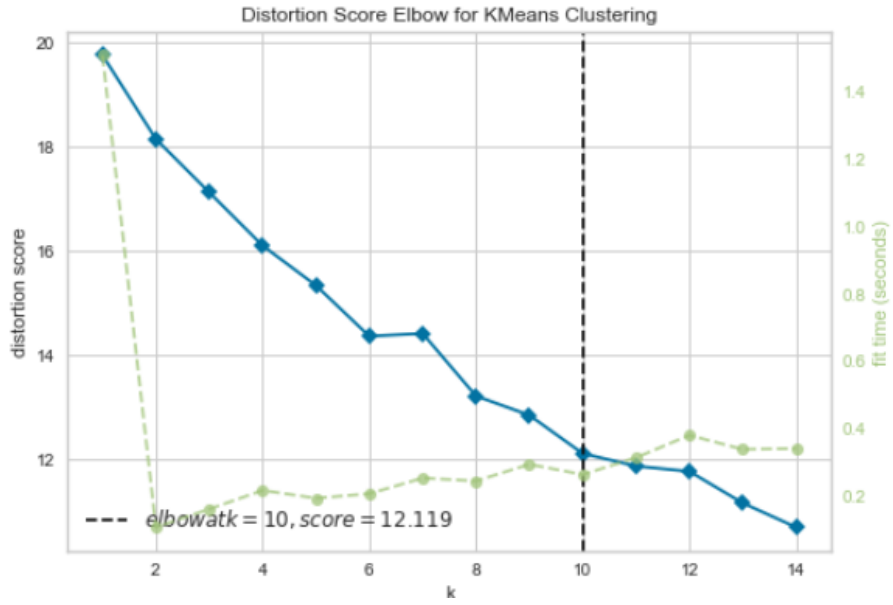
As seen from the above, schools are pretty much present in all parts of Singapore except the water bodies (reservoirs), since land is limited on the island.

Next, we explore the neighbourhoods within 1 km radius around the schools by using Foursquare to obtain data of nearby venues, as shown:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ADMIRALTY PRIMARY SCHOOL	1.443015	103.800282	Kampung Admiralty Hawker Centre	1.439939	103.800774	Food Court
1	ADMIRALTY PRIMARY SCHOOL	1.443015	103.800282	McDonald's	1.445931	103.798101	Fast Food Restaurant
2	ADMIRALTY PRIMARY SCHOOL	1.443015	103.800282	Starbucks	1.439761	103.800659	Coffee Shop
3	ADMIRALTY PRIMARY SCHOOL	1.443015	103.800282	Superpets Warehouse / HQ	1.436848	103.806059	Pet Store
4	ADMIRALTY PRIMARY SCHOOL	1.443015	103.800282	Asia Ghani	1.437431	103.795388	Malay Restaurant

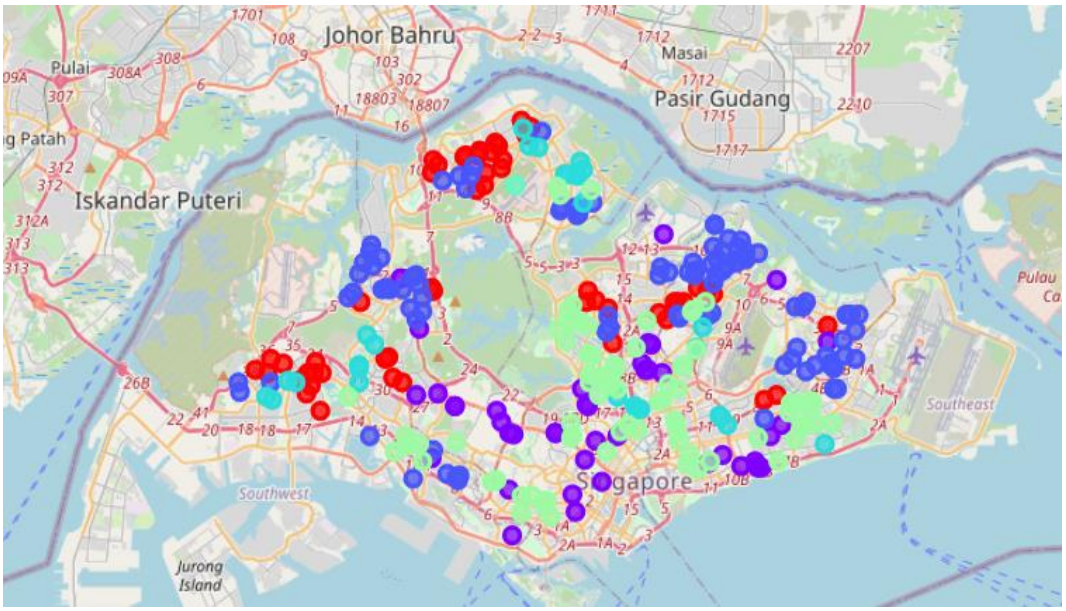
We identify categories that may be similar variations of the same venue type and clean up such data to reduce noise in the subsequent analysis. For example, both "Hainan Restaurant" and "Hakka Restaurant" are both Chinese restaurants, which for the purposes of our analysis can be treated the same (since this isn't a study targeted at, say, food gourmands).

We prepare the dataset for K-means analysis to group the various school neighbourhoods into clusters by using one-hot encoding. Next, we initialise K-means modelling and find the optimal K value using the elbow method – optimal K is 10.



Results

We create a map to visualise the resulting clusters:



By doing a count of the number of members in each cluster, we see that the most significant clusters are clusters 2, 6, 0, 1 and 4 in descending order of size. Clusters 8, 7, 5 and 3 are much smaller in size. For the purposes of our analysis, let us focus on the first group of clusters since it will offer more choices for our potential users within each cluster.

```
schools_merged["Cluster_Labels"].value_counts()
```

```
2    97
6    78
0    53
1    50
4    29
8     1
7     1
5     1
3     1
Name: Cluster_Labels, dtype: int64
```

By subsetting the dataset according to the each of the significant clusters, we examine each cluster in turn to understand the profile of the neighbourhood around the schools. For example:

```
schools_merged[schools_merged.Cluster_Labels == 4]
```

	school_name	latitude	longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
31	BOON LAY GARDEN PRIMARY SCHOOL	1.342899	103.712983	4	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Shopping Mall	Food Court	Indian Restaurant	Snack Place	Farmers Market
64	CHONGFU SCHOOL	1.438767	103.839296	4	Coffee Shop	Food Court	Fast Food Restaurant	Convenience Store	Thai Restaurant	Multiplex	Indian Restaurant	College Bookstore
335	YUHUA PRIMARY SCHOOL	1.342698	103.740936	4	Chinese Restaurant	Coffee Shop	Food Court	Indian Restaurant	Halal Restaurant	Pool	Diner	Ramen Restaurant
95	EUNOS PRIMARY SCHOOL	1.324709	103.904580	4	Food Court	Chinese Restaurant	Coffee Shop	Asian Restaurant	Noodle House	Bubble Tea Shop	Restaurant	Furniture / Home Store
266	ST. ANDREW'S SECONDARY SCHOOL	1.330980	103.866011	4	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Food Court	Convenience Store	Market	Ice Cream Shop	Photography Studio
249	SEMBAWANG SECONDARY SCHOOL	1.445661	103.816660	4	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Fast Food Restaurant	Food Court	Café	Steakhouse	Betting Shop

Discussion

Observations based on the clusters generated from Kmeans modelling and visualisation of the cluster locations

Cluster 2 - Majority of the school neighbourhoods have Coffee Shop as the first most common venue. Fast Food Restaurant occurs frequently in the top or second top venue. Japanese Restaurant, Shopping Mall, Sandwich Place and Supermarket appear more frequently in this cluster compared to other clusters. Schools in this cluster are predominantly located in the northern half of Singapore.

Cluster 6 - This cluster is predominantly comprised of F&B businesses in the top 10 venues. Majority of the school neighbourhoods have Chinese Restaurant as the first most common venue. Food court appears frequently in the top or second top venue. Cafe, Fried Chicken Joint, Frozen Yogurt Shop and Vegan/vegetarian Restaurant appear more frequently in this cluster compared to other clusters. Schools in this cluster are predominantly located in the southern half of Singapore.

Cluster 0 - =Majority of the school neighbourhoods have Food Court as the first most common venue. Park, Bar and Bus Station appear more frequently in this cluster compared to other clusters. Sports/exercise venues such as Pool, Gym, Trail and Soccer Stadium appear with the highest frequency in this cluster compared to other clusters. Schools in this cluster are predominantly located in north and west of Singapore.

Cluster 1 - Wide mix of venue categories come under first most common venue in this cluster - no obvious trend. Hotel, Fish & Chips Shop and Filipino Restaurant appear more frequently in this cluster compared to other clusters. Overall, there is a wide mix of F&B and

recreational/entertainment venues; the widest amongst the 5 clusters. Schools in this cluster are predominantly located in south and east of Singapore.

Cluster 4 - Like Cluster 6, this cluster is predominantly comprised of F&B businesses in the top 10 venues, though to a lesser extent since it has more recreational venues like Bowling Alley, Multiplex and Bookstore. Majority of the school neighbourhoods have either Chinese Restaurant or Coffee shop as the first most common venue. Fast Food Restaurant, Food Court and Chinese Restaurant also occur frequently in the top 4 venues. This is the smallest cluster, and the schools are scattered throughout Singapore, except for the south.

What type of "users" would prefer each cluster profile?

Parents may prefer to send their children to a school in either Cluster 6 or 4 if they want their children to have a quick meal after school before heading home. There are not many recreational options that may pose distractions. These may be for families where both parents have full-time day jobs, and are unable to pick up their kids immediately after school.

Parents may prefer to send their children to a school in Cluster 0 if they are keen to have their children adopt a healthy lifestyle that can make use of the amenities that tend to be close by, such as Pool, Soccer Stadium etc.

Students may choose to enroll in a school in either Cluster 2, 6 or 1 if they want access to a wider variety of recreational and entertainment options after school where they can hang out with friends.

Conclusion

In conclusion, we can see that the neighbourhoods around schools in Singapore can be clusterd into 5 big and distinct groups, each with their own profile based on the most common venue types found in them. The study can provide further information for both students and parents alike in selecting schools based on the type of neighbourhoods they are in, in addition to the usual grade-based entry requirements, popular brand names, sporting excellence and so on.