

BURP (Balancing Urban Residential Plannings): Visualising the Public Schools Accessibility to Residential Housings

Brendo Austin

School of Information Systems
Singapore Management University
Singapore
brendoa.2015@sis.smu.edu.sg

Goh Li Na Rebecca

School of Information Systems
Singapore Management University
Singapore
rebecca.goh.2016@sis.smu.edu.sg

Tan Peng Chong

School of Information Systems
Singapore Management University
Singapore
pctan.2016@sis.smu.edu.sg

ABSTRACT

Singapore has been going through a lot of development in the education sector. The government has initiated many mergers of schools due to low enrolment. Schools with low enrolment find it challenging to offer a wider range of subject combinations, educational programmes and co-curricular activities (CCAs). The low enrolment level will mean that students will not be fully given opportunities to blossom and pursue subjects of their interests and this might be a hindrance to their future career options.

Declining birth rate contributes to the decline in the number of students entering post-secondary institutions and is a contributing factor to the mergers of post-secondary institutions. To address this issue, this project aims to visualize the impact of accessibility that residential housings have to the different school locations in Singapore to see which residential housings will be most affected if a school location is moved.

Measurement of accessibility includes the usage of Hansen Accessibility Index and time taken to travel from a residential zone to a school using public transportation. The visualization of the impact is shown on an isochrone map with the different intensity of time travel and the number of residential flats within each time travel.

The development of this platform shows the potential of R Shiny as a framework to link large datasets of residential housings, road maps, and school addresses with the planning and analytical models useful for government planners to derive insights on which location of schools to merge into to improve the accessibility.

KEYWORDS

Accessibility, Geospatial Analytics, R Shiny, School mergers, Education

Introduction

Singapore is known to have a world-class education system that has been ranked the best several times. It was rated the best in the world by the Organization for Economic Co-operation and Development.

For more than a decade now Singapore along with Japan, Finland, Taiwan, South Korea, Hong Kong and Shanghai have been near on top of international league table that assesses the children's ability in reading science and maths.[1] The Singapore government

has contributed tremendously to the excellent performance by equipping the education industry modern facilities and qualified workforce to teach in schools. School mergers are necessary to improve the quality of education provided by the Ministry of Education of Singapore.

Due to the low birth rate that Singapore has been experiencing, schools have been facing low enrolment. The low enrolment restricts the variety of subject combinations, educational programmes and CCAs that are offered to the students. The mergers will enable the schools with healthy enrolment needed to continue to provide a rich variety of learning programmes and experiences for students.[2]

The Ministry of Education's priorities in planning schools are to provide quality education, meet the demand for school places at national and local levels and also ensuring accessibility based on proximity to housing developments and public transport.

The schools that are identified for mergers are based on the sufficiency of the enrolment to ensure there is enough provision of educational programmes and CCAs, the geographical proximity of affected schools and the infrastructure capacity to support the merged school. In particular, the JC merge partners were selected based on geography to maintain good spread of JCs across the country. The first assessment is the overall demand and supply of JC places and then finding suitable merge pairs.

Currently, based on our knowledge, there has not been a platform that assesses the accessibility from the residential zones that have been affected by the school merger. For example, when school A moved to school B's location, no analysis has been done to assess how long will it take for people that live around school A to travel to school B.

Therefore, this project aims to visualize the time taken by using public transport to travel to a particular school, the number of residential flats that are within each time travel and also the Hansen Accessibility Index to better understand and then helps urban planner which preferred school location to be the final merged destination.

The paper details the research and development efforts to design and implement a web-based geospatial analytics tool that visualizes the accessibility of residential housings to schools. The paper will include five sections.

The first section discusses how data is obtained, prepared and extracted based on the data that the project requires.

The second section discusses the packages, methodology, development tools and system architecture to develop the geospatial analytics platform.

The third section provides a short description of the functions in the geospatial analytics platform.

The fourth section provides the feedback of the audience during the presentation at the townhall and future works for the application.

1 Data Collection & Preparation

Before performing the accessibility assessment and the time taken to travel from a residential housing zone to a school location, we have used 5 sets of data:

1. Data of school information (School Directory and Information)
2. Data on Housing Development Board (HDB) property information
3. Data on the postal codes
4. Data on General Transit Feed Specification (GTFS)
5. Data on Singapore's OpenStreetMap Protocol Binary Format (PBF)

The first data contains the school information (School Directory and Information) is collected from data.gov.sg. The data set contains the school name, website address, address, postal code, telephone number, second telephone number, fax number, second fax number, email address, nearest Mass Rapid Transit (MRT) station, list of buses that stops near the school, principal name, vice principal name, second vice principal name, third vice principal name, fourth principal name, fifth principal name, vision statement, mission statement, philosophy, culture and ethos, dgp_code, zone code, cluster code, type code, nature code, session code, mainlevel_code, sap_ind, autonomous ind, gifted ind, ip ind, mother tongue code, second mother tongue code, third mother tongue code and the special courses offered. Data wrangling was done to merge the postal codes together with the original data set in order to get the longitude and latitude.

The second data contains the HDB property information is collected from data.gov.sg. The data set contains the address, block number, building name, latitude, longitude, postal code, road name, search value and the X and Y coordinate. Data wrangling was done to filter out residential from non-residential HDB information.

The third data contains the postal codes from OneMap. The data contains the address, block number, building name, latitude, longitude, postal code, road name, search value and the X and Y coordinate. Data wrangling has been done for this section. In the initial data set, the address is in an extended format. There are short forms of the addresses. However, the addresses that are found in data.gov to identify the HDB block addresses are in short form. The conversion was done to expand the short format of the addresses to the long format.

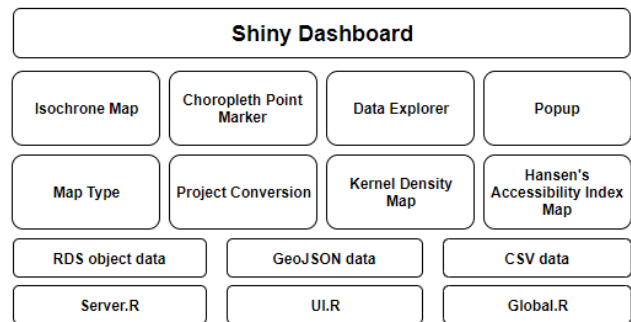
The fourth data contains the GTFS collected from transit.land. The dataset contains the agency file, calendar file, frequencies file,

routes file, stop_times file, stops file and trips file. This data set is collected to get the timings of the different public transports available in Singapore

The fifth data contains Singapore's OpenStreetMap PBF file format collected from Interline. This OpenStreetMap PBF contains geographical details about Singapore, i.e. road lines and the locations of different of public transports in Singapore. The GTFS together with PBF data will be consumed by the backend local server (OpenTripPlanner) which we will explain later in the research paper.

2 Methods

2.1. Application Architecture



Different data types (i.e. RDS object, GeoJSON, and CSV) were used to generate different components of the web application. The UI.R is used to display the output of the functionalities of the application. The Server.R is used to handle the backend and rendering of different plots and map. Lastly, Global.R is used to load the respective libraries as well as the general data like residential data, school data and etcetera that are used throughout the application as it gets run first, even before UI.R and Server.R.

With the usage of R Shiny dashboard, the team developed our application with functionalities for users. These functionalities include:

1. Isochrone map
2. Hansen Accessibility Index
3. Kernel density
4. Distribution of HDB points

With the isochrone map, users are able to see the overall time travel using public transport to the selected school. They could also check the distribution chart to see the number of HDB block within the respective travelling time (i.e. 15 minutes, 30 minutes, 45 minutes, 60 minutes, and 90 minutes). The bar graph in the distribution charts represents the number of HDB within the timeframe while the line graph shows the cumulative of the number of HDB within the timeframe.

For the Hansen Accessibility Index, users can select to show the duration or distance accessibility of Singapore in overall as well as each specific HDB to the selected school.

Using Two-Dimensional Kernel Density Estimation (2D KDE), users can see the overall distribution and density of the residential HDB in a contour plot.

Lastly, with the HDB points, users are able to check it using any of the functionalities mentioned above and they could hover over any specific HDB to obtain their address.

The functionalities are shown as a pop up in the web application, it is movable and minimizable to provide flexibility to the users. However, the legend distribution chart is a fixed pop up that can only be minimized through the checkbox. Users are also able to change the base of the map from a list of map types (e.g. Stamen, Esri, cartoDB, and more) based on their preferences (See Figure 1 and Figure 2).

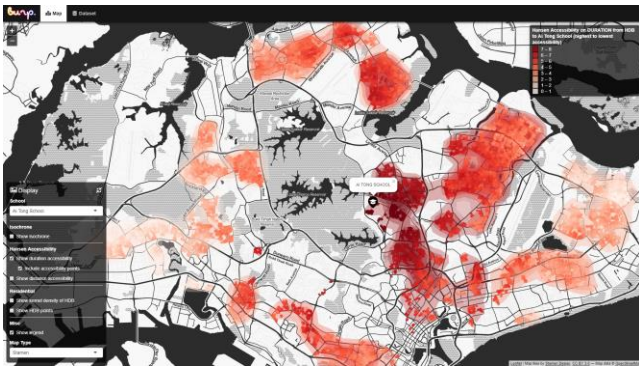


Figure 1: Stamen Map Type.

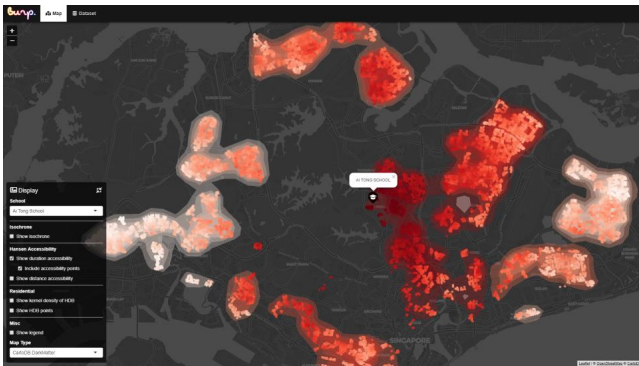


Figure 2: CartoDB.DarkMatter Map Type.

2.2 R Packages

To develop the web application, we used the R Shiny application from RStudio. It allows us to build an interactive web application straight from the R programming language which is commonly used for statistical computing and graphics. As R has a comprehensive library with a large community, there are many packages from third-party that are readily available for us to use.

The packages used during the development process includes the following:

1. OpenTripPlanner (OTP) API is used as the backend local server to retrieve the duration and distance from each HDB flats to each school (See Figure 3).
2. Rgeos package is used to get HDB points within each polygon (See Figure 4).
3. SpatialAcc package is used for its ac function to compute Hansen accessibility index of distance and duration respectively (See Figure 5).
4. GeoJSONIO package is used to read and write GeoJSON file for the isochrone data generated (See Figure 6).
5. Dplyr package is used for data manipulation functions such as select, mutate, etc.
6. KernSmooth package is used for 2D Binned Kernel Density Estimate to compute data for contour line (See Figure 7).
7. DT is used for displaying the data table on the web application (See Figure 8).
8. Sp package is used to convert data into spatial objects.
9. RColorBrewer package is used for the different colour palette of the display and polygon generated
10. Sf package is used to convert data into simple feature
11. Leaflet package is used to display the different maps.
12. RStudio Tool is used as an integrated development environment (IDE) for R.
13. Bootstrap is used for UI design which goes hand in hand with R Shiny.

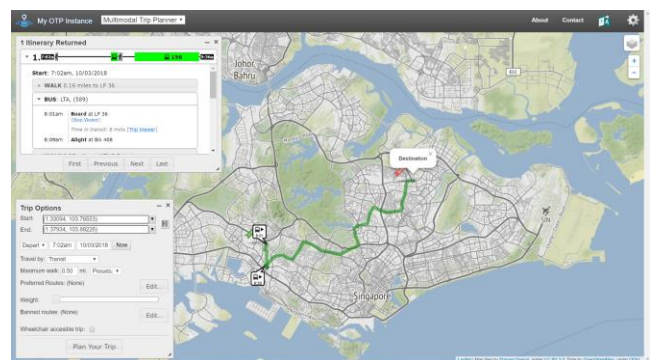


Figure 3: OpenTripPlanner API

```
for(i in 1:nrow(polys)){
  ff = gcontains(polys[i,], points.copied, byid = TRUE)
  points.copied$checker = ff
  sub = points.copied %>% filter(checker == TRUE)
  total = sum(sub$durationHansen, na.rm = TRUE)
  avg = total/length(sub)
  if (is.na(avg)){
    avg = 0
  }
  polys@data$duration_hansen[i] = avg
  points.copied = points.copied %>% filter(checker == FALSE)
}
```

Figure 4: gContains Function of Rgeos

```

180 apply_hansen = function(res){
181   result = res %>% mutate(demand =100)
182   durationMatrix = result %>% dplyr::select('duration')
183   acc_hansen <- data.frame(ac(result$demand,
184     c(10), durationMatrix,
185     power = 0.01, family = "Hansen"))
186   result_duration = result
187   result_duration$durationHansen = acc_hansen[,1]
188   result_duration$durationHansen[is.na(result_duration$durationHansen)] = 0
189
190
191   distanceMatrix = result %>% dplyr::select('distance')
192   acc_hansen <- data.frame(ac(result$demand,
193     c(10), distanceMatrix,
194     power = 0.01, family = "Hansen"))
195   result_final = result_duration
196   result_final$distanceHansen = acc_hansen[,1]
197   result_final$distanceHansen[is.na(result_final$distanceHansen)] = 0
198   # colnames(result_final)[8:9] = c('durationHansen', 'distanceHansen')
199   return (result_final)
200 }

```

Figure 5: ac Function of SpatialAcc

```

139   current <- content(current, as = "text", encoding = "UTF-8")
140   write(current, file = paste0(isopath, '/', filename, ".geojson"))
141
142   if(rb){
143     sp = geosonio::geoson_read(paste0(isopath, '/', filename, ".geojson"), what = "sp")
144     file.remove(paste0(isopath, '/', filename, ".geojson"))
145     return(sp)
146   }

```

Figure 6: read and write Function of GeoJSONIO

```

294 d2d = bkde2D(cbind(residential.infos@coords[,1],residential.infos@coords[,2]),bandwidth=c(0.0025,0.0025))
295 contour(d2d$x1,d2d$x2,d2d$zhat)
296 lines = contourLines(x=d2d$x1,y=d2d$x2,z = d2d$zhat,nlevels = 8)
297 dd1 = sapply(1:length(lines),function(i) Polygon(as.matrix(cbind(lines[[i]]$x,lines[[i]]$y))))
298 dd2 = sapply(1:length(lines),function(i) Polygons(list(dd1[[i]]),i))
299 poly_data = data.frame(value = sapply(1:length(lines),function(i) lines[[i]]$level))
300 dd3 = SpatialPolygonsDataFrame(SpatialPolygons(dd2),data = poly_data)
301 values = unique(sapply(1:length(dd3@data[["value"]]),function(i) dd3@data[["value"]]))[,1]
302 dd3$area_sqkm <- area(dd3) / 1
303 proj4string(dd3) = proj4string(residential.infos)
304 saveRDS(dd3, file='data/objdata/kernel.residential.rds')

```

Figure 7: contourLines Function of KernSmooth

```

output$table <- renderDataTable(datatable({
  if(input$selectTable == 1){
    data = residentialdata@data
    colnames(data) = c('Address', 'Postal Code', 'X Coordinate', 'Y Coordinate', 'Road Name')
    data = as.data.frame(data)
    data %>% mutate(Address = gsub("( ?SINGAPORE.*)", "", Address)) %>% dplyr::select('Address', 'Postal Co

```

Figure 8: datatable Function of DT

2.3 Algorithms

In this project, the team has used various types of algorithms to produce the accessibility results. Some of the algorithms are self-derived by the team. However, the measurement for accessibility uses the Hansen's Accessibility Index.

The first algorithm that we used is to calculate the distance travelled from the HDBs to the desired school location. The calculation steps are as follows:

1. From the HDBs to the nearest bus stop, walking is assumed. The maximum walking distance that was set is at 1.6 kilometers.
2. From the bus stop to MRT stations or other forms of public transportations. There are 3 different ways of taking the public transportations.
3. We generated RDS file to store the data of the distance from 357 schools to each point of HDBs in Singapore.
4. The number of calculations is approximately 357 schools X 9900++ number of HDBs X 3. 9900 refers to the number of HDBs available in Singapore. 3 refers to the number of routes using public transportations. Note that the calculations are approximate as for different schools, the total number of HDB is 9966 but it may vary as the calculations produced by Hansen's Accessibility Index generated are invalid/ NA.
5. Once there is a public transport near the school within the walking distance, the assumption is that people will walk to the desired school.
6. If there is no connected route or walking path, Euclidean distance is used.
7. If there is a walking path, the path line leading to the desired school from the public transport is used.

The second algorithm is on how we used the servers to run the application. The steps are as follows:

1. First, get the GTFS (General Transit Feed Specification) from transit.land for the different timings of the public transportations.
2. Second, go to Interline to get the Singapore's OpenStreetMap PBF (Protocol Binary Format)
3. Third, use OpenTripPlanner to plan and generate different routes of the public transport on the map. This is to get the duration and distance through running a local instance of the OpenTripPlanner.

The third algorithm is on the usage of kernel density of the HDBs. The steps are as follows:

1. The kernel density measures just the density of the HDBs
2. The kernel density estimate uses 2D KDE.
3. The kernel density for our application we used a bandwidth of 250m x 250m.
4. The contour line with the intensity of the HDBs within an area forms the kernel density

5. The polygon of the contour line is used to display the general duration/ distance accessibility for the Hansen's Accessibility Index calculation later on.

The Hansen's Accessibility Index is as shown below.

$$Accessibility_i = \sum_j Opportunities_j \times f(C_{ij})$$

where:

i = index of origin zones (HDB flats)

j = index of destination zone (Schools)

$f(C_{ij})$ = function of generalised time taken and distance of travel

Figure 9: Formula for Hansen Accessibility Index.

For this project, i refers to the index of origin zones which is the HDB flats and j refers to the index of destination zone which is the school's location. The higher Hansen Accessibility Index will mean that the HDB block is more accessible to a particular school location.

3 Results

After the user selects the specific school, the web application will find the corresponding data to display. The data of the map has already been rendered before the user select any schools as this will reduce the time taken to show the corresponding map to the user. Therefore, making it possible to display large data output in a short time. All the distance and duration computation take into consideration the usual school time of 7 am, which is the peak hour.

The isochrone analysis is visualized through the different colour intensity in the different regions. The darkest purple represents the area with the highest accessibility (i.e. travel time within 15 minutes) to the selected school while the lightest purple represents the areas with the lowest accessibility (i.e. travel time within 90 minutes) to the selected school. Using the distribution chart, users could compare between or amongst schools in terms of the number of HDB flats accessibility within 15 minutes. Users could take into consideration when thinking about which specific school should be removed or kept based on their accessibility to residents (See Figure 10).

The Hansen Accessibility is visualized using the colour intensity and individual HDB flat's accessibility points. In this case, the accessibility could be duration or the distance from the respective HDB points to the school. The darkest red represents areas with the highest accessibility while the lightest red represents areas with lowest accessibility. With this accessibility point of each HDB, users are also able to see whether there are any specific HDB flats with lower accessibility despite being within a high accessibility area (See Figure 11).

The kernel density estimate is visualized through the colour intensity as well. Based on the degree of density, the area is being shown in dark green to light green. Those with the darkest green are areas that are more clustered with HDB while those with the lightest green are areas that have HDB that are more spread out (See Figure 12).



Figure 10: Isochrone of BURP application

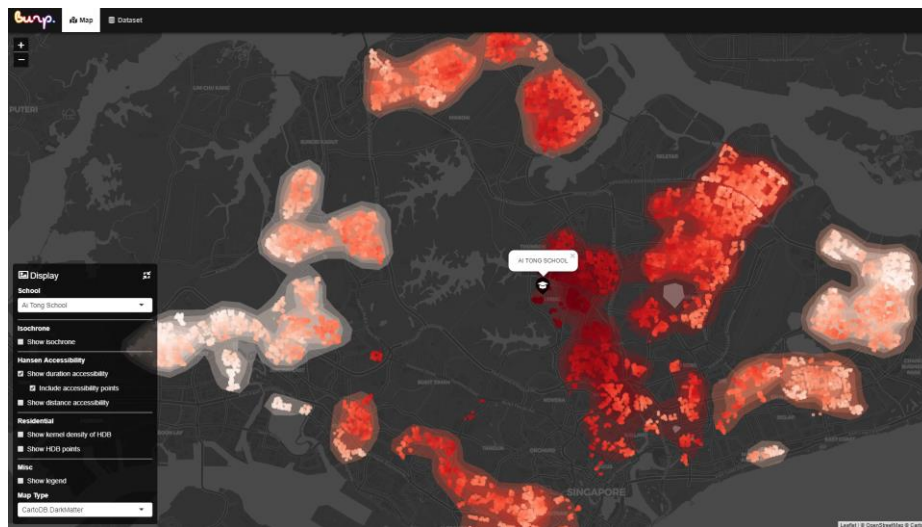


Figure 11: Hansen Accessibility of BURP Application

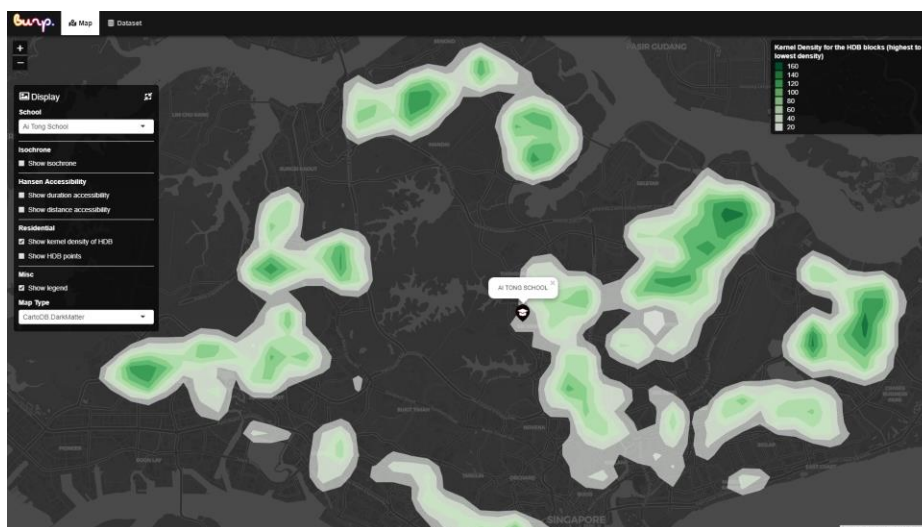


Figure 12: Kernel Density of BURP Application

Other than map display, users can also select data table to obtain more information on the location and accessibility. The location information let users choose between the residential HDB information and school information, and generate information such as the address, postal code, and X and Y coordinates based on the chosen category. Similarly, for the accessibility information, users can choose a specific school out of the 357 schools and the table would generate all the HDB information such as the address, the duration and the distance from the HDB to the selected school (See Figure 13 and 14).

For easier search, both tables have the results sorted according to alphabetical order and users can use the search engine to find specific information. We provide both distance and duration as having the shortest distance. This does not mean having the shortest duration due to the available public transport. An example would be the accessibility to Assumption Pathway School.

As seen in Figure 15, despite Block 135 Cashew Road having a shorter distance, the duration is longer compared to Block 185 Jele Road and this could be due to the lower accessibility to the nearest public transport. Hence, this enables users to get a better insight of the HDB's accessibility to the specific school.

Location Information

Select Table
Residential HDB information

Show 10 entries

Search:

	Address	Postal Code	X Coordinate	Y Coordinate
1	1 BEACH ROAD BEACH ROAD GARDENS	190001	31467.83274	31778.894
2	1 BEDOK SOUTH AVENUE 1 MY FIRST SKOOL	460001	39173.8129	33678.85598
3	1 CHAI CHEE ROAD PING YI GARDENS	461001	37949.02647	34465.73857
4	1 CHANGI VILLAGE ROAD OCBC Changi Village Road - 7 Eleven	500001	45224.4586	41171.52219
5	1 DELTA AVENUE HDB-BUKIT MERAH	160001	27473.0908	30496.63612
6	1 DOVER ROAD DOVER COURT	130001	22430.39934	31652.25293
7	1 EUNOS CRESCENT	400001	35695.14219	33657.23665
8	1 EVERTON PARK EVERTON PARK	81001	28743.72096	28876.85814
9	1 GHIM MOH ROAD GHIM MOH GREEN	270001	22829.04815	32797.13881
10	1 HAIG ROAD MY FIRST SKOOL	430001	35165.4266	32621.49718

Showing 1 to 10 of 9,965 entries

Previous 1 2 3 4 5 ... 997 Next

Figure 13: BURP Data Table on Location Information

Accessibility Information

Accessibility to:
Admiralty Primary School

Show 10 entries

Search:

	Address	Duration (minutes)	Distance (km)
1	1 BEACH ROAD BEACH ROAD GARDENS	100.30	28.02
2	1 BEDOK SOUTH AVENUE 1 MY FIRST SKOOL	112.35	29.30
3	1 CHAI CHEE ROAD PING YI GARDENS	105.60	29.29
4	1 CHANGI VILLAGE ROAD OCBC Changi Village Road - 7 Eleven	107.45	28.53
5	1 DELTA AVENUE HDB-BUKIT MERAH	97.95	23.13
6	1 DOVER ROAD DOVER COURT	76.18	24.04
7	1 EUNOS CRESCENT	102.53	23.69
8	1 EVERTON PARK EVERTON PARK	100.08	26.18
9	1 GHIM MOH ROAD GHIM MOH GREEN	86.47	25.76
10	1 HAIG ROAD MY FIRST SKOOL	110.97	32.52

Showing 1 to 10 of 9,792 entries

Previous 1 2 3 4 5 ... 980 Next

Figure 14: BURP Data Table on Accessibility Information

Accessibility to:
Assumption Pathway School

Show 10 entries Search:

	Address	Duration (minutes)	Distance (km)
975	130 CASHEW ROAD HDB-BUKIT PANJANG	16.00	1.28
1018	132 CASHEW ROAD HDB-BUKIT PANJANG	16.43	1.31
997	131 CASHEW ROAD HDB-BUKIT PANJANG	16.82	1.34
1049	133A CASHEW ROAD HDB-BUKIT PANJANG	17.17	1.37
1036	133 CASHEW ROAD HDB-BUKIT PANJANG	17.60	1.40
1071	135 CASHEW ROAD HDB-BUKIT PANJANG	19.00	1.51
1899	185 JELEBU ROAD HDB-BUKIT PANJANG	17.00	1.57

Figure 15: BURP Data Table Example

Based on our results, we also found some limitations. Firstly, we are unable to factor some conditions such as the number of students in each school, the number of residents in each HDB flats, as well as the traffic conditions of buses. Also, we are unable to take into consideration some factors that are affecting the mergers such as school reputation, the size of the school, the environment of the school, etc. In addition, due to time constraint and lack of data, we did not consider condominiums and private housings as well as the road network for private cars and school buses.

4 Feedback and Future Works

After the completion of the analysis and web application, the team thought of a few potential ways that our application could improve on or work on to provide more insight.

The team also obtained constructive feedback during our Town Hall poster presentation.

One improvement that could be made would be getting the demand data of the school and tying accessibility with it to get more variable to quantify the importance of the school (e.g. reputation, academic offerings, CCA offerings, school fees, and more). Other areas that we could work on would be the accessibility of creating new public amenities and shopping malls. This could provide insights to government or organisations on the Ideal location to have public amenities and shopping malls. The feedback we obtain from our poster presentation was to include a functionality that allows users to search a specific HDB block and show the nearest top 10 schools, including primary school, secondary schools and junior college. This enables residents to plan for the ideal location to stay if they have a specific school in mind. On top of the functionality, we were also given a suggestion to add a supporting functionality that allows users to add a school and see how it changes the accessibility of the specific HDB to the nearest schools.

Lastly, one of the feedback includes finding the ideal HDB for elderly and children by factoring in demographics of the residents. The ideal HDB would be based on the nearby facilities available. These could be potential areas for our application to work on in the future, however there is currently lack of accessibility to some of the data necessary to compute (e.g. the number of residents in each

HDB, the detail information of residents staying in each HDB, the capacity of the facilities, etc.) and also a large amount of data need to be processed.

ACKNOWLEDGEMENTS

The team wants to thank Professor Kam Tin Seong from Singapore Management University for helping the team in providing different resources for us to refer. The team would also like to acknowledge and thank Sean Koh from Centroid-Amenities for giving us tips and ideas on how to develop this application.

REFERENCES

- [1] Joshi, M (2018). Why is Singapore's school system so successful, which factors have contributed to its success? Retrieved from <https://www.msn.com/en-sg/news/singapore/why-is-singapore%E2%80%99s-school-system-so-successful-which-factors-have-contributed-to-its-success/ar-BBR916L>.
- [2] Why is the Ministry of Education merging some schools? (2017) Retrieved from <https://www.gov.sg/factually/content/why-is-the-ministry-of-education-merging-some-schools>
- [3] OpenTripPlanner Basic Tutorial (2019). Retrieved April 13, 2019 from <http://docs.opentripplanner.org/en/latest/Basic-Tutorial/>