Proposal of new hub locations

for

Datathon2020

by:

Shamini Puthooppallil Baby
Rishabh Singh Kochhar
Clarine Anslum

ABSTRACT

The aim of the project was to find the new hub locations in order to satisfy the demand of the future need of hub beneficiaries.

Based in primary schools (state, catholic and independent), community hubs serve as gateways to connect families with each other, with their school and with existing services. Hubs build communities through engagement, early childhood services, English conversation support, and pathways to volunteering, training and employment. There are currently 74 HUBS in Australia where 3,986 families engaged with and 176 individuals found jobs as a result of connecting with a hub.

The proposal of new hubs for schools/LGAs was done by exploratory data analysis, visualization, and modelling specifically applied to school profile in Australia using features of existing school hub and geo spatial data of LGAs. The data were provided by the Community Hub Australia.

The results indicated there were seven different groups of schools and the majority of hubs were in three clusters. The efficiency of each cluster was positive. The three clusters that had larger number of hubs were selected and schools were filtered using the specific attributes they used in the past to implement school hubs. And also, some external data related to socio-economic factors including migrant family background in the LGAs were used in the final selection of LGAs.

We concluded that Casey and Wyndham councils are the potential LGAs to choose schools for new hubs as their population related to non-English speaking is high with high developmentally vulnerable kids' percentage.

The Audience of the project is Community Hub Australia and so, the study can be continued based on this analysis for future implementation and enhancements of hub quality. It is advisable to include further attributes such as success rates and efficiency of hubs to the hub profiles so that statistically significant schools could be avoided.

Table of Contents

List of Figures

1. INTRODUCTION

The National Community Hubs Program supports migrant families, while the Refuge of Hope provides assistance to Latin American students living in Melbourne which increases connection between individuals and communities.

Community Hubs serve as gateways that connect families with each other, with their school, and with existing services. Dozens of community hubs operate under the National Community Hubs Program – recognised as a leading model to engage and support migrant women with young children.

Many migrant women can spend most of their time at home, isolated from the rest of the community. Some migrant women cannot drive, some have limited public transport access, many have young children and cannot access child-minding. Women who miss out on available services can then go without opportunities to meet other people, to learn, or to receive health support, and their children can arrive at school without having English language or social skills.

Many adults first come to hubs with their children, but eventually find that hubs also address their own needs, from access to exercise classes, cooking and sewing, or just somewhere for a cup of tea, through to all important access to English and skills training.

Clustering analysis is one of the most frequently utilised analytical techniques applicable to audience segmentation. The project is concerned with performing segmentation analysis on data relating to schools all over Australia based on similar attributes of existing hubs.

The goal of the project relates to interpreting the clusters generated as a result of the analysis. Here, the characteristics of hubs common to the clusters were identified and potential Schools were found using attributes such as habitants of non-English speaking migrants, single mum percentage and social index.

2. DATA

The external data were provided by the Community Hub Australia and there were seven different files to be Analysed. School profiles, Hub locations, Population Australia, Proficiency of English, Socio-Economic Indexes for Areas, Public Health Information Development Unit and Australian Early Development Census by LGA (AEDC).

Apart from these files, the internal data related to current hub profiles were analysed in order to find the insights that could be used for future hub installation.

The AEDC data were combined as to show the health domains under one sheet and Hub locations data were joined with School profiles by deriving GPS location using google map API. School profile data from 2016 were analysed in order to derive the scenarios they used for hub opening.

Table 1: Table of School profile

Calendar Year	AGE ID	School Name	Suburb	State	Postcode	School Sector	School Type	Campus Type	Rolled Reporting Description	...	Equivalent Non-Teaching Staff	Total Enrolments	Girls Enrolments	Boys Enrolments	Full T Equiva Enrolme
2008	3.0	Corpus Christi Catholic School	Bellerive	TAS	7018	Catholic	Primary	School Single Entity	Individual Reporting	...	6.9	410.0	185.0	225.0	4
2008	4.0	Fahan School	Sandy Bay	TAS	7005	Independent	Combined	School Single	Individual Reporting	...	14.6	347.0	339.0	8.0	3·

Table 2: Table of AEDC by health domain with subdomain

	Code	Location	2009	2012	2015	2018	State	Domain	Subdomain
0	10050	Albury (C)	587	624	591	626	New South Wales	Health	children with valid scores
1	10130	Armidale Regional (A)	383	367	319	354	New South Wales	Health	children with valid scores
2	10250	Ballina (A)	481	337	434	471	New South Wales	Health	children with valid scores
3	10300	Balranald (A)	31	31	33	29	New South Wales	Health	children with valid scores
4	10470	Bathurst Regional (A)	482	513	514	510	New South Wales	Health	children with valid scores

Table 3: Table of Existing Hub details

HubRando	Hub_Activi	ActivityDate	ChildPartic	AdultPartic	DSS_Client	ReferralPa	EngagedCc	ExternalVc	HubVolunt	SchoolVol	ShortNam	ServiceTyp	CategoryName
15994	259	4/03/2019	11	7	18	0	0	0	0	0	Playgroups	4	Programmed Activities
15994	260	25/02/2019	12	8	20	0	0	0	0	0	Playgroups	4	Programmed Activities
15994	261	11/03/2019	0	0	0	0	0	0	0	0	Playgroups	4	Programmed Activities
15994	262	18/03/2019	12	8	20	0	0	0	0	0	Playgroups	4	Programmed Activities
15994	263	4/02/2019	10	6	16	0	0	0	0	0	Playgroups	4	Programmed Activities
15994	264	18/02/2019	11	7	18	0	0	0	0	0	Playgroups	4	Programmed Activities

3. METHODOLOGY

The statistical tools and machine learning algorithms were used to analyse data and predict model respectively using python. The columns were selected for model building using the evidence of current hub attributes and meaningless features were dropped.

3.1 DATA PREPARATION

The useful columns (1. ICSEA score, 2. school type, 3. school sector, 4. indigenous enrolment, 5. language other than English, 6. geo location, 7. teaching staff) were selected from school profiles as they directly impact the existence of Hubs at schools. A few columns were filtered for modelling purposes as per the following reasons. The total school profile is 9535.

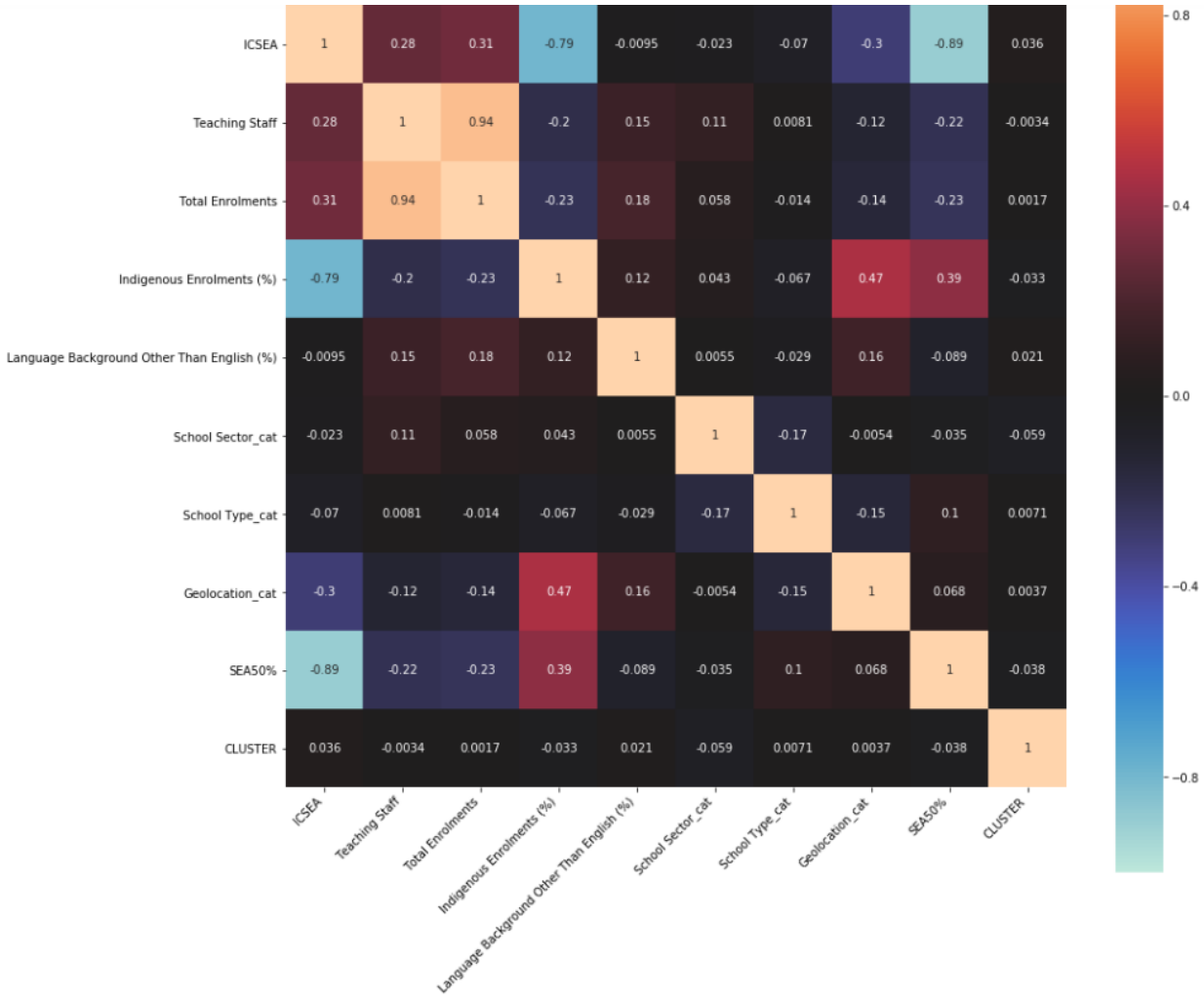


Figure 1: Heat map Schools' features (all over Australia)

1. ICSEA score is the indication of disadvantaged schools where the score is less than 800. And also, all of the scores were lying between 800 and 1200. So, this shows the need of new hub. As per heat map (Figure1), the ICSEA Score is negatively co-related with SEA and indigenous enrolment. Hence, it was decided to drop both and to include ICSEA.

2. Most of the schools are primary schools and some are combined. One outlier was p-12 school. Hence, this feature has been included for modelling purpose.

3. and 4.

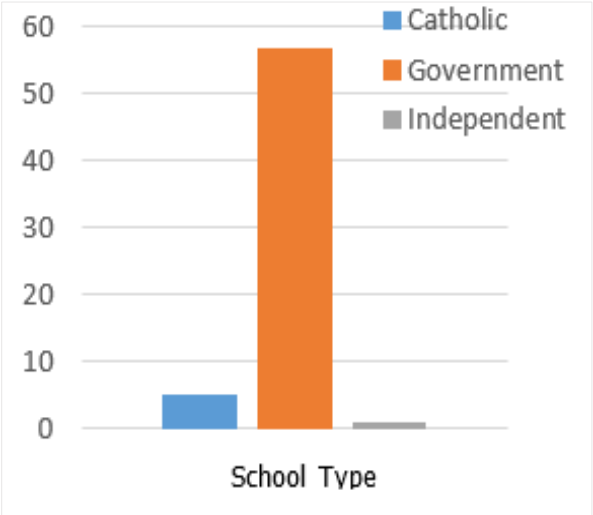


Figure 2: School distribution by type

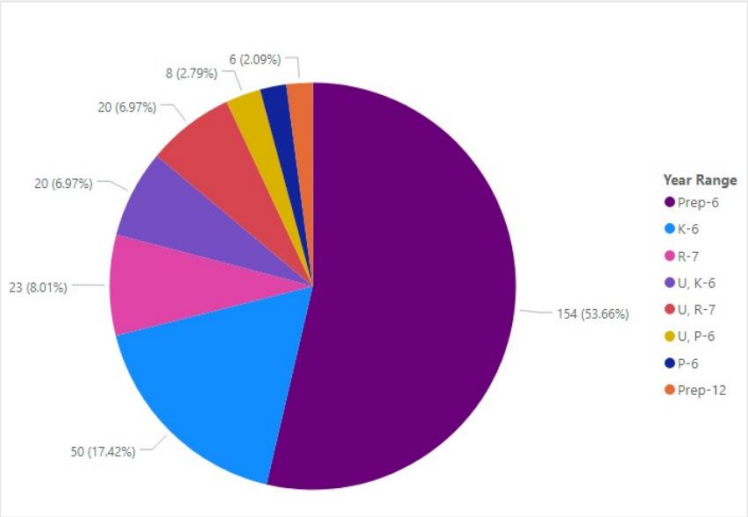


Figure 3: Count of indigenous

Figure 2 and 3 clearly shows the majority of schools with hubs are government schools and indigenous enrolments are high in the same.

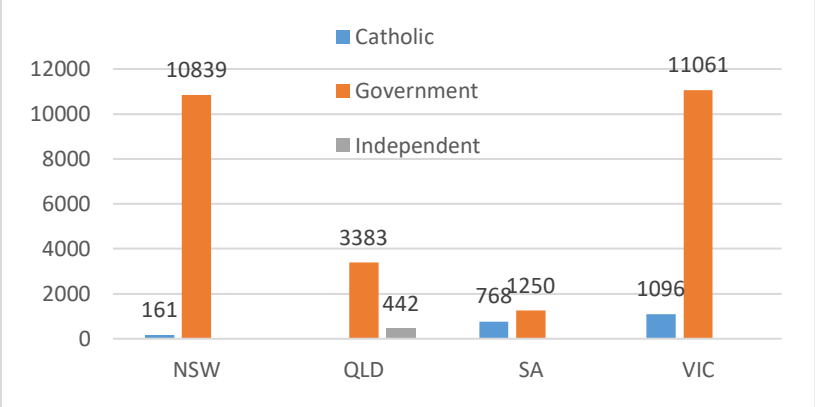


Figure 4: Total student enrolments by state (all schools)

And also figure 4 shows, the majority schools are Government all over Australia. So, these features were included.

5.. Language other than English could be a result of non-English speaking community, so this feature has been taken into account.

6. Among 62 hubs at schools, majority (59) of the schools are located in major cities and 3 others are in inner regional. As this feature had been taken into consideration in the past, this feature was included.

7. As per Figure 1, Total enrolments and teaching staff are highly and positively corelated. Teaching staff were included as they would support community hubs activities.

The remoteness was calculated for later usage of profiling/indexing once the grouping is done. The formula that was used to calculate is:

ICSEA = SEA (up to 75%) + Remoteness + Percent Indigenous student enrolment

Overall, the data related to the project were cleansed including removal of null rows. Total number of total rows 9316 school profiles were used for application of model.

3.2 DATA EXPLORATION

Australia is a fast growing in population and it is significalty facing the need of social networking due to the increase of migrants population.

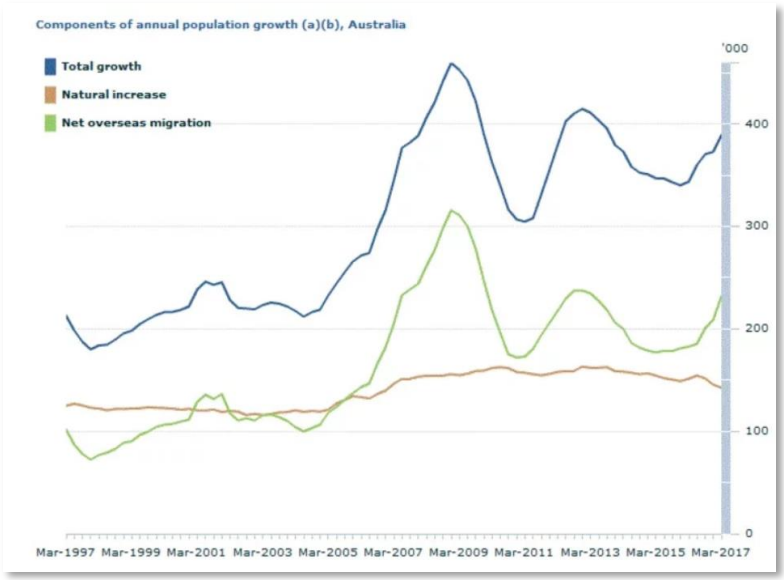


Figure 5: Trend of Australian Population by state (source: Business insider)

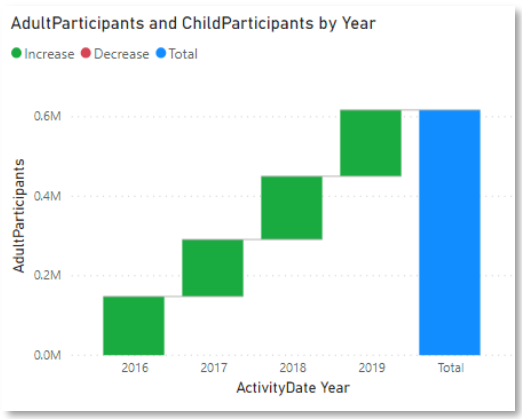


Figure 6:Increment of participants throughout years

The above figures show the growth of population (figure 5) with migrants and the increase in participants in hub activies through out the last four years (Figure 6). This is a clear indication that the new hubs should be implemented.

Since 2002, the Australian Government continues to work with its partners, and with state and territory governments to implement the Early Development Census program nationwide. The following figure 7 illustrates the worth of this program and the health of children at risk still exists.

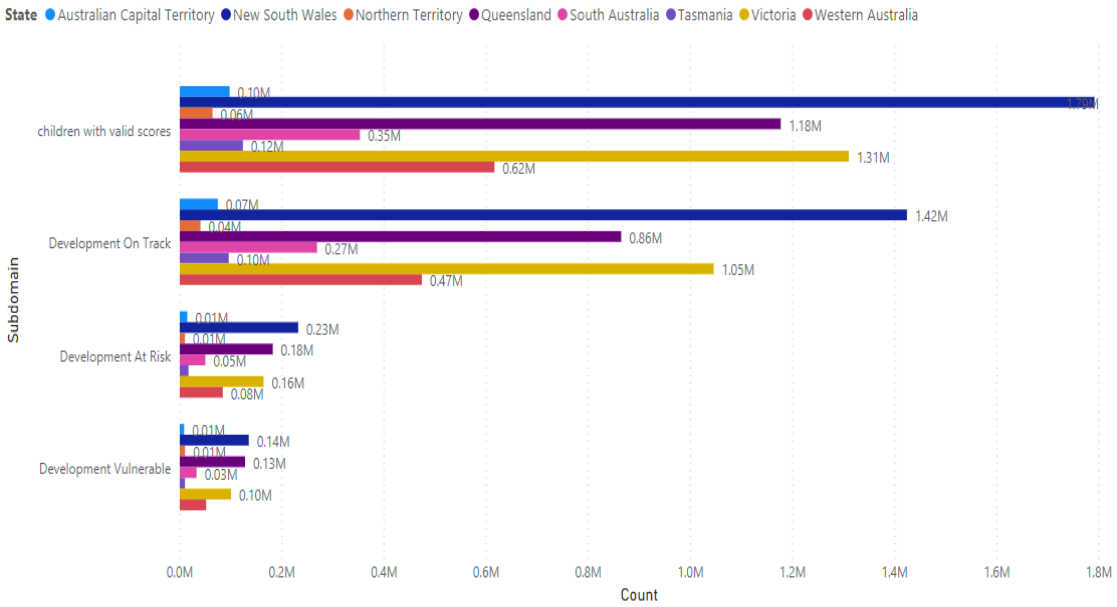


Figure 7:Australian Early Development Census by state and Domain

Community Hubs Australia also providing some activities related to health and well-being and this has been shown in the following pie chart (Figure 8).

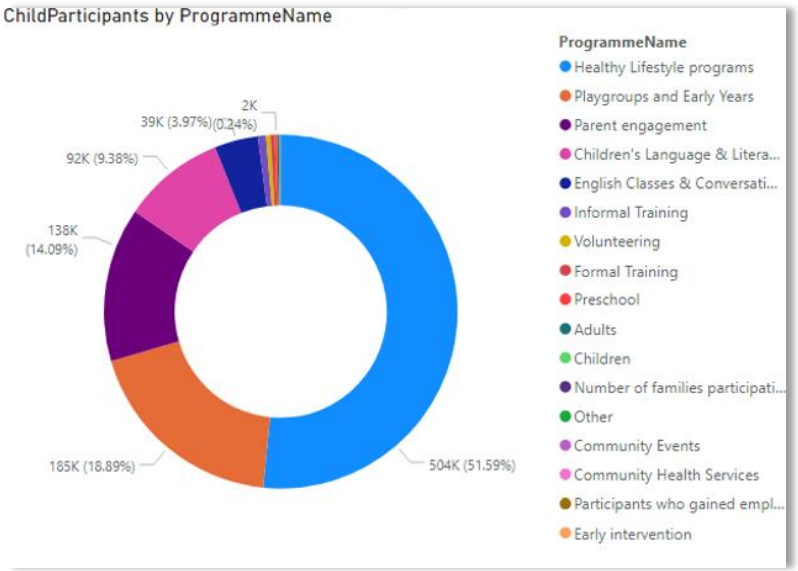


Figure 8: Pie chart of Child participant by programme

As this programme is mostly run at schools, the schools of Australia and its locations have been analysed.

The map of School locations (Figure) is shown in below plot.

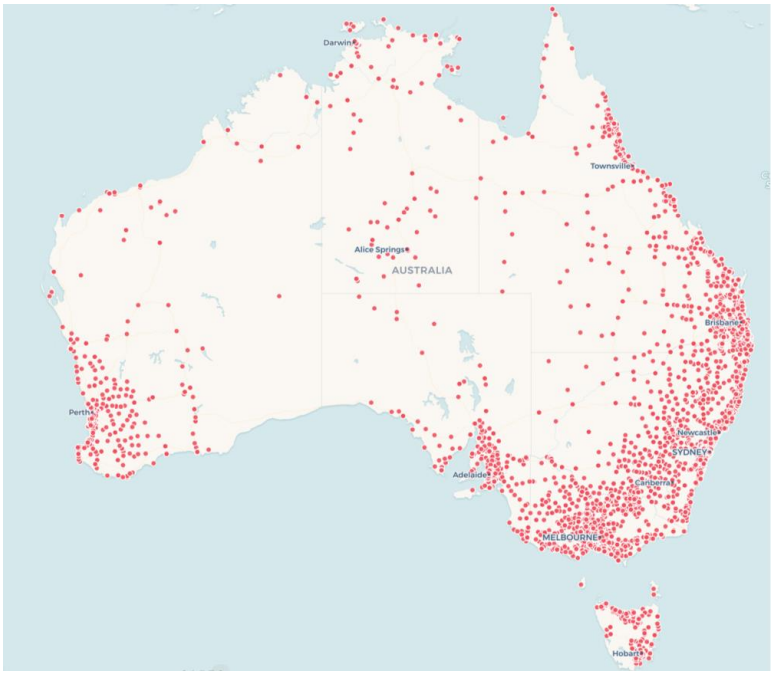
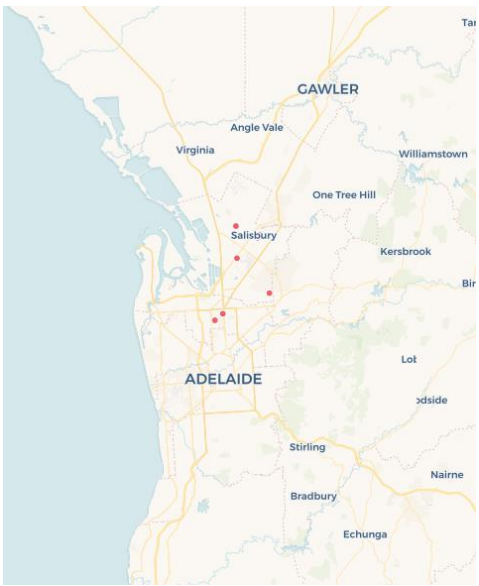
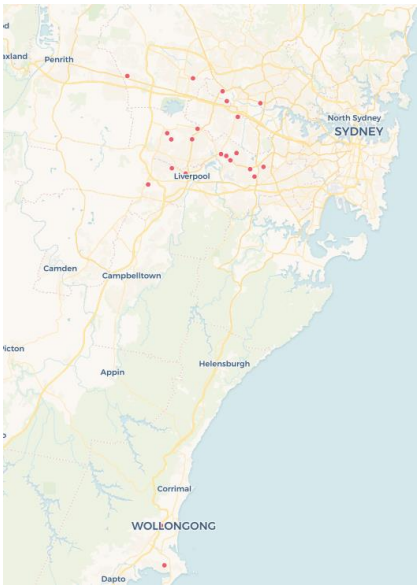


Figure 9:Map of schools in Australia

The existing hub locations by states (NSW, VIC, SA and QLD) shown in the below plots.





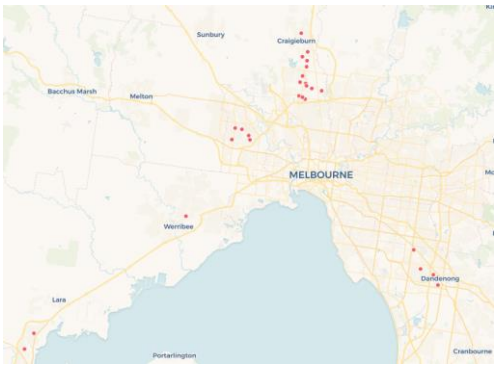


Figure 10:Map of Hub locations in Australia

3.3 FEATURE ENGINEERING

Here, we investigated a light version of Hill climbing about how to select the features for School profile data.

We loaded the shuffle package to randomize the order of features, then we start trying from the 1st feature. After that, we will keep this feature as "selected" if this feature + existing already-selected features can lead to higher accuracy score than that when only using existing already-selected features:

3.4 DATA MODELLING

The normalisation was applied as the scales of features were different. All of them were explored and visualised in order to see the insights. The following section illustrates this further in detail. The unsupervised machine learning methods Kmeans and Hierarchical (Agglomerative) were applied to see the clusters of schools with common attributes and the optimum was 7 clusters from both methods.

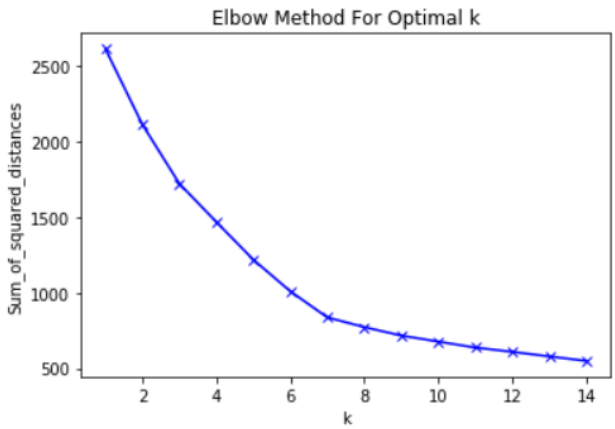


Figure 11:Best k of Kmeans for School profile Data

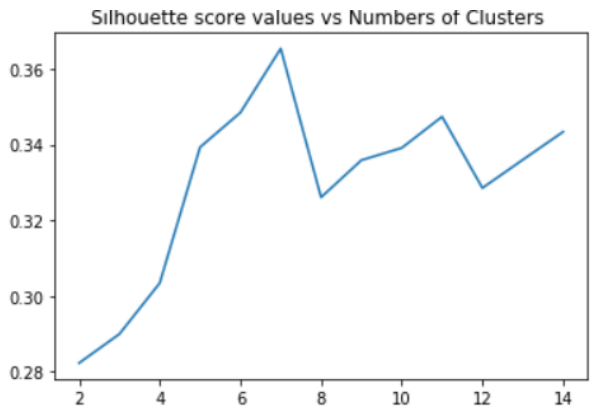


Figure 12:Plot of silhouette score vs number of clusters (kmean)

Hierarchical clustering is a super useful way of segmenting observations, the clusters were selected from the same.

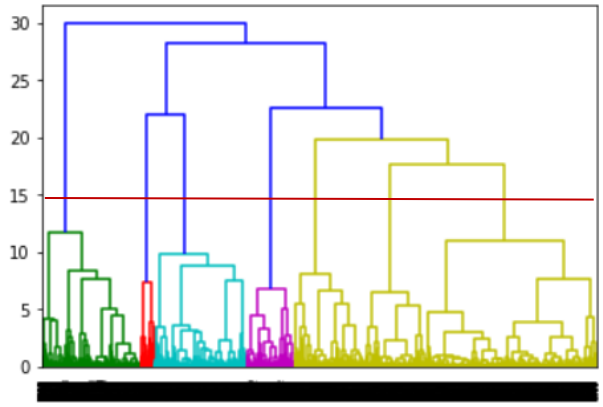


Figure 13: Dendrogram of Hierarchical clustering

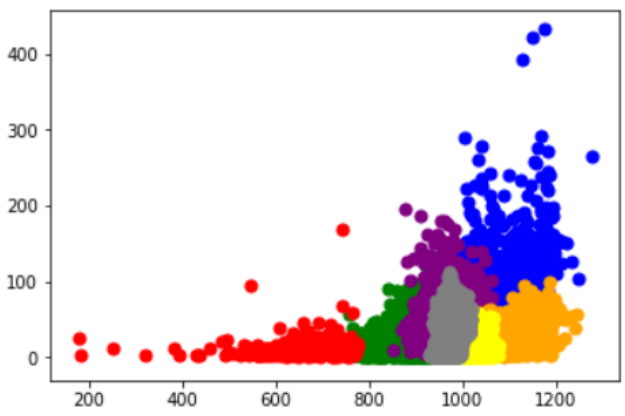


Figure 14: Visual of Clusters by agglomerative model

4. RESULT

We used Silhouette coefficient as it is more efficient for categorical data. Silhouette coefficient is a method to interpret and validate consistency within the clusters of data.

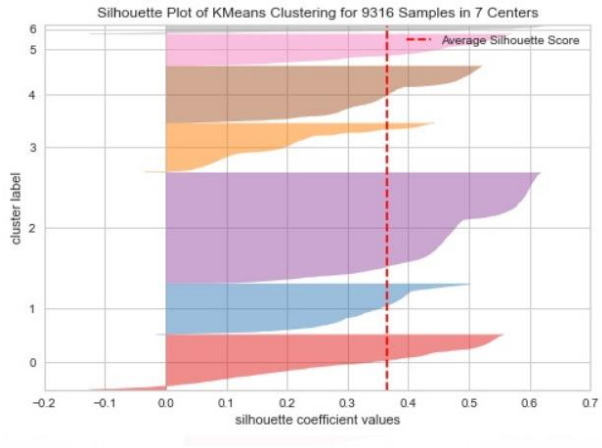


Figure 15: Silhouette coefficients for kmeans clusters

The scores are positive and the largest cluster has 0.6 score where the second largest has 0.55. The first three largest clusters with more hubs would be our potential group to look into as this would have same schools as existing schools with hubs. The insights of these clusters have been discussed in the next section.

5. DISCUSSION

The following chart illustrates the percentage of hub existence in each cluster.

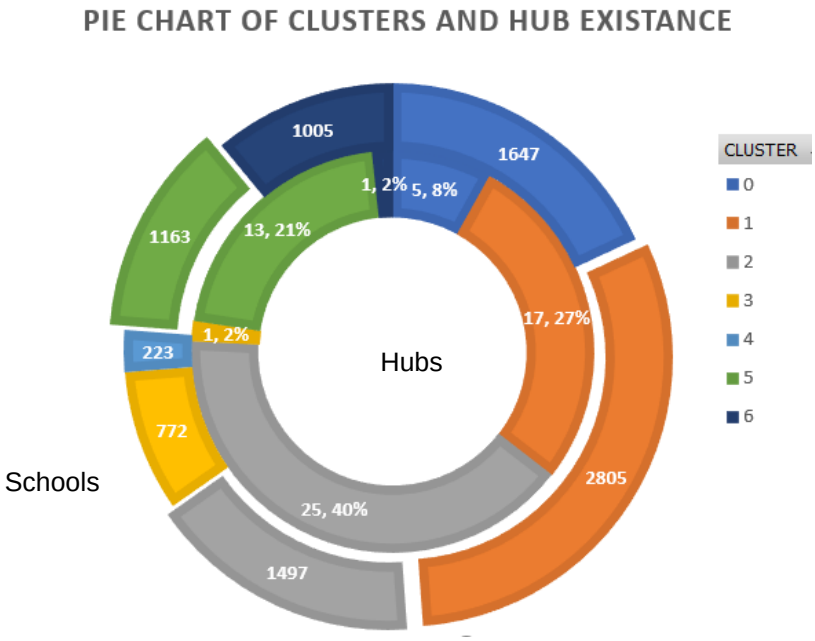


Figure 16: Pie chart of Counts with Hub Existence

After a careful analysis of school profile attributes such as ISCEA greater than 800, indigenous enrolment, SEA(75%) greater than 80%, Geographical location(Major cities), school type(Primary and combined), school category(Government and catholic), Language other than English (greater than 30%) as per current hubs, the schools were short listed. And also, the remoteness score (>750) was used to filter some schools as it is a clear indication of school community with isolated families in the existing hubs.

LGA profile such as AEDC information, non- English-speaking population and Population of children aged between 0 and 14 have been taken into account in choosing LGA for new hub location.

Blacktown	4354
Canterbury-Bankstown	10249
Cumberland	7733
Fairfield	10168
Georges River	4513
Liverpool	4592
Parramatta	5790
Brimbank	5371
Casey	4286
Greater Dandenong	5946
Hume	4230
Monash	3467
Whitehorse	3069
Whittlesea	3135
Wyndham	3877

As per this list of non-English speaking population by LGA's in 2016, the top 15 are listed here and in nine of them Community Hub Australia (**CHA**) is active (between 2015 to 2018). Among the rest of them (highlighted ones), Georges River is a Regional area but school Hubs run only in Major cities or inner regional. Casey, Wyndham and Monash have higher numbers compared to Whitehorse and Whittlesea.

In the past, **CHA** selected LGAs with developmentally vulnerable kids more than 20%. Casey (20.4%) and Wyndham (22.8%) falls in this range whereas Monash has 15.9% according to 2018 statistics. Casey children (age between 0-14) is 22.9% and Wyndham is 25.1 with 2,739 belongs to single income category as per 2017 statistics. These figures also fall within the condition Hubs applied in the past.

6. CONCLUSION

Community hubs focus on reaching women who have pre-school children and leverage existing facilities in primary schools and community centres.

We concluded that Casey and Wyndham Council areas in need of community hubs. Among the following Schools under these LGAs could be selected for new hub implementation based on other criteria such as isolated women population/socio-index rank. And also, readiness of school check list such as principal/school community support, space availability, funds to pay for CHA coordinators and etc could be considered.

Casey

School Name	Suburb
Berwick Chase Primary School	Berwick
Cranbourne Primary School	Cranbourne
Cranbourne West Primary School	Cranbourne
Rangebank Primary School	Cranbourne
Cranbourne Carlisle Primary School	Cranbourne
St Agatha's School	Cranbourne
St Therese's School	Cranbourne
Cranbourne East Primary School	Cranbourne East
Courtenay Gardens Primary School	Cranbourne North
Holy Family School	Doveton
Doveton College	Doveton
James Cook Primary School	Endeavour Hills
Mossgiel Park Primary School	Endeavour Hills
Chalcot Lodge Primary School	Endeavour Hills
Southern Cross Primary School	Endeavour Hills
Hallam Primary School	Hallam
Hampton Park Primary School	Hampton Park
River Gum Primary School	Hampton Park
Coral Park Primary School	Hampton Park
Kilberry Valley Primary School	Hampton Park
St Kevin's School	Hampton Park
Lynbrook Primary School	Lynbrook
Lyndhurst Primary School	Lyndhurst
Fleetwood Primary School	Narre Warren
Fountain Gate Primary School	Narre Warren
Maramba Primary School	Narre Warren
Mary MacKillop School	Narre Warren North
Strathaird Primary School	Narre Warren South
Hillsmeade Primary School	Narre Warren South
Narre Warren South P-12 College	Narre Warren South

Wyndham

Shool Name	Suburb
Mossfiel Primary School	Hoppers Crossing
Woodville Primary School	Hoppers Crossing
Bellbridge Primary School	Hoppers Crossing
Baden Powell P-9 College	Hoppers Crossing
The Grange P-12 College	Hoppers Crossing
St Peter Apostle School	Hoppers Crossing
St James the Apostle School	Hoppers Crossing
Carranballac P-9 College	Point Cook
Werribee Primary School	Werribee
Manorvale Primary School	Werribee
Westgrove Primary School	Werribee
St Andrew's School	Werribee
Manor Lakes P-12 College	Wyndham Vale
Our Lady of the Southern Cross	Wyndham Vale

The link to the map of these schools: https://shamini.carto.com/builder/257628f3-601b-42ac-a1e9-6bf480c30b34/embed

Further analysis on this study could be done by including features such as migrant population (women and families) increase in the selected suburbs and successful rates of school hubs for a fare selection of schools and also to run the hubs humming.

7. REFERENCES

- https://www.communityhubs.org.au/wp-content/uploads/2019/08/National-Community-Hubs-Program-Mid-Year-Update-2019.pdf
- https://issuu.com/communityhubs/docs/community_hubs_guide_live?e=32945844/61166970
- https://www.communityhubs.org.au/publications/
- http://stat.data.abs.gov.au/Index.aspx?DataSetCode=ABS_REGIONAL_LGA2017#
- https://www.w3schools.com/python/default.asp
- https://www.aedc.gov.au/researchers/faqs-for-researchers
- https://www.businessinsider.com.au/australias-population-growth-just-hit-the-accelerator-2017-9

TEAM MEMBER CONTRIBUTION

Jobs	Shamini	Rishab	Clarine
Planning	Idea and method provided to approach the goal	Idea and method provided to approach the goal	Idea and method provided to approach the goal
Data cleansing and exploration	Combine sheets in AEDC and model ready data set. Visuals of AEDC and schools	Visuals of Hubs/school profiles data. Clusters and profiling	Hub location and School profiles data combined. visuals of Internal data
Modelling	Kmeans cluster for combined data set	Agglomerative cluster for the combined data set	Hub location - Kmeans and DBscan
Discussion	Discussed ideas and heat maps for each cluster	Discussed ideas and heat map of School profiles	Discussed ideas and some plots of cluster profiling
Report	Provided plots and report correction	Provided suggestions and plots	Drafted the report
Presentation	Slides completion and Presented	Corrections made and presented	Slides preparation and presented