Crimson Global Academy Report

Progress –

So far, I have extracted the percentage change of enrollments within schools over a time period from 2020 to now. I have also calculated the avg change of enrollment percentage revealing the average change in population of a school. Although data is not taken from a longer period of time, leaving the results inaccurate this is the safest method to find a consistent change in enrollment history of schools.

By displaying the Avg_Percentage_Change of each school in descending order we get the schools listed at the top with the highest or most dramatic increase in enrollment size since 2020. This result shows a trend revealing which schools are growing the fastest. This could potentially be an indication of population increase in certain areas.

Complications

When calculating average percentage change from 2020 to present day, there is no previous enrollment history data available for the first year 2020. This means that the averages are incorrectly calculated as they are considering an extra 0 term. We must remove the 2020 total enrollments rows to correctly calculate change in percentage.

The problem was arising because I was setting null values to 0. Specifically for rows where year was 2020, there was no previous enrolment data for 2019 since we are considering data from 2020 onwards. This meant that 0 was being considered as a value in the average when it should be dismissed. When keeping the values set to null, sql ignores this values and uses only 3 from 2021,2022,2023 to calculate average change. Now we have a more accurate avg percentage change in enrollment count of schools.

Hypothesis 1

Areas that may have a high ICSEA may also have a lower percentage change in enrollments over time. This may be because suburbs with high ICSEA may be more stable in terms of population size than suburbs with lower ICSEA. This might mean that enrollments for these areas may not change much over time. Parents tend to be comfortable with their kids' enrollments so advertising here may not be ideal.

avg growth factor of a suburb based on their avg ICSEA percentile
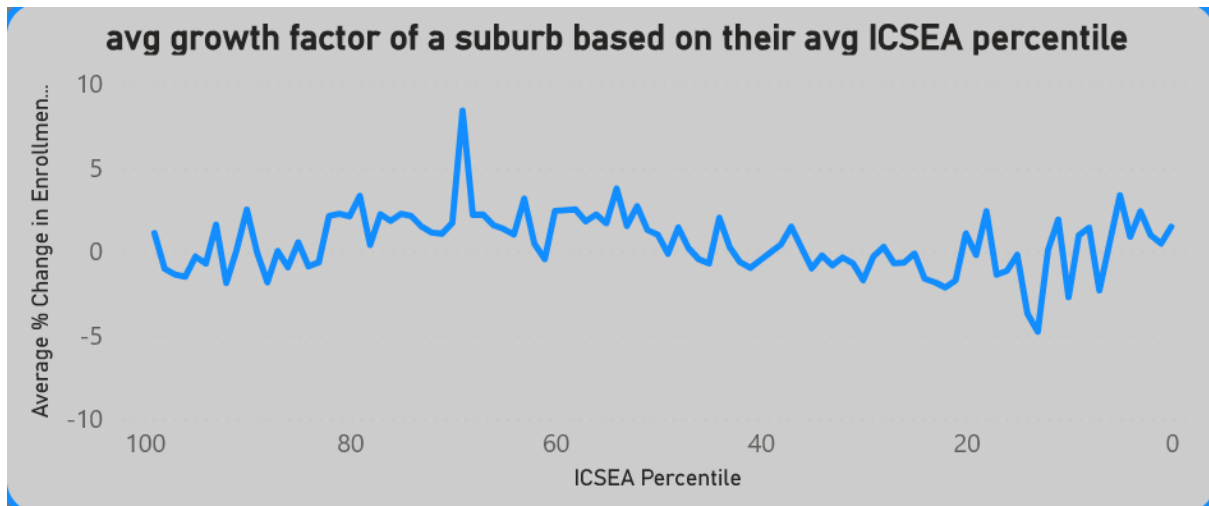
By looking at suburbs by their average ICSEA percentile we can distinguish which suburbs tend to have the schools with the highest ICSEA percentile. This can help us figure out if the ICSEA percentile of a suburb has any affect on the avg percentage change of enrollments within them. From the data above it is clear that regardless of the ICSEA percentile of a suburb, the avg percentage change of population is not affected.

The data above models the avg change in enrollment numbers based off of the average ICSEA percentile of schools in a suburb. This forms an avg ICSEA percentile for each suburb and then models the change in enrollments since 2020. From the data it is clear that ICSEA percentile of a suburb does not have a correlation with its growth in enrolments. The change in enrollments over the years based on ICSEA percentile tend to remain consistent regardless of the ICSEA percentile. From this we can conclude that This data will be irrelevant in figuring out which areas are most ideal to advertise.

Insights

It turns out that **Dayton** being listed as the fastest growing school was because it was a newly developed school in a newly developed area. So realistically this actually has no correlation to our problem and thus is an outlier in the data. Therefore, Dayton can be ignored and thus the suburbs with the highest avg growth in enrollments are **Theebine**, **Deanside** and **Bullarah**.

Methodology

Rural and regional suburbs –

- Take the avg ICSEA scores of rural suburbs and their avg total enrolment count over the last 4 years.
- Create a weighted score based on ICSEA and enrolment count of a suburb.
- Visualise data based on this weighted score

Weighted score formula:

Normalisation process:

$$\text{weighted average score} = T(x, y, z) = W(x', y', z')$$

$$\text{suppose} \quad a = \text{span}\{a_1, a_2, \ldots, a_n\}$$

$$\& \quad x, y, z \in a$$

$$x' = \frac{x}{\max(x)} \quad , \quad y' = \frac{y}{\max(y)} \quad , \quad z' = \frac{z}{\max(z)}$$

then,

$$W(x', y', z') = x' + y' + z'$$

The problem with the formula above is that it does not account for the size difference in values for each attribute. If x is the ICSEA percentile and y is the avg total enrolment count then x has a min of 72 and a max of 81 whilst total enrolment ranges from 118-984.

Therefore, we must add additional parameters to account for this (normalisation).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad , \quad y' = \frac{y - \min(y)}{\max(y) - \min(y)} \quad , \quad z' = \frac{z - \min(z)}{\max(z) - \min(z)}$$

This effectively accounts for the discrepancy between ranges for each attribute. Applying this formula to Suburbs based on 2 attributes : (avg_ICSEA_percentile & avg_total_enrolments) gives us an accurate indication of the best rural suburbs to advertise to

Result:

	Suburb	avg_ICSEA_percentile	avg_total_enrolments	weighted_score
1	Redlynch	64	1804	0.89
2	Cairns North	71	984	0.69
3	Manunda	29	1726	0.65
4	Kalgan	80	634	0.64
5	Berrimah	71	795	0.63
6	Rapid Creek	83	505	0.62
7	Kamerunga	74	645	0.61
8	Edge Hill	56	1045	0.61
9	Freshwater	73	665	0.61
10	Larrakeyah	81	464	0.59
11	Kirwan	33	1394	0.58
12	Pambula Beach	65	731	0.58
13	Belgian Gardens	73	593	0.58
14	Annandale	56	918	0.58

Complications

The formula and the result set above give equal significance to the 2 attributes. For this business model we are more concerned with the ICSEA percentiles of this suburbs rather than their enrolment numbers.

New formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad , \quad y' = \frac{y - \min(y)}{\max(y) - \min(y)} \quad , \quad z' = \frac{z - \min(z)}{\max(z) - \min(z)}$$

$$W = \lambda_1 x' + \lambda_2 y' + \lambda_3 z' \quad , \quad \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$$

For our case our weights are 0.67 for ICSEA and 0.33 for total enrolments.

RESULT:

	Suburb	avg_ICSEA_percentile	avg_total_enrolments	weighted_score
1	Redlynch	64	1804	0.847
2	Kalgan	80	634	0.749
3	Rapid Creek	83	505	0.748
4	Cairns North	71	984	0.744
5	Larrakeyah	81	464	0.724
6	Berrimah	71	795	0.708
7	Kamerunga	74	645	0.703
8	Freshwater	73	665	0.699
9	Belgian Gardens	73	593	0.685
10	Jindabyne	73	555	0.677

The result shows that Redlynch should be the primary focus of our rural advertisements. Although it's average ICSEA percentile is not that high, its sheer enrolment count suggests that it is still more favourable for advertising.

Further Complications

Upon further investigation, it is clear that when we look at the data from the perspective of suburbs and from the perspective of schools for advertisement, we get drastically different results. This is because suburbs tend to have 1 school that will be an outlier in terms of avg ICSEA percentile and its enrolment numbers and some suburbs may have a school with very low values for these attributes. This means that when considering the data from a suburbs perspective, we get a very skewed result set that isn't accurate.

RESULT:

	School_Name	Suburb	avg_ICSEA_percentile	avg_total_enrolments	weighted_score
1	Townsville Grammar School	North Ward	90	1235	0.756
2	Cairns School of Distance Education	Manunda	59	4005	0.735
3	The Essington School	Rapid Creek	90	811	0.724
4	The Cathedral School of St Anne and St James	Mundingburra	85	1065	0.706
5	Trinity Anglican School	White Rock	87	871	0.706
6	St Andrew's Catholic College Redlynch Valley	Redlynch	76	1658	0.684
7	St Joseph's Catholic School, The Strand	North Ward	86	379	0.661
8	Snowy Mountains Grammar School	Jindabyne	84	423	0.65
9	Great Southern Grammar School	Kalgan	80	634	0.636
10	Larrakeyah Primary School	Larrakeyah	81	464	0.63
11	Mother of Good Counsel School	Cairns North	82	361	0.63
12	Geraldton Grammar School	Geraldton	79	516	0.62
13	St Augustine's College	Cairns	75	777	0.609
14	Parklands School	Albany	81	135	0.606
15	Darwin High School	The Gardens	69	1220	0.598
16	Darwin Middle School	The Gardens	73	777	0.595
17	Peace Lutheran College	Kamerunga	74	645	0.592
18	Annandale Christian College	Annandale	74	629	0.591
19	Southern Cross Catholic College	Annandale	66	1389	0.589
20	Milkwood Steiner School	Berrimah	79	119	0.589

The school's data suggests we should advertise to schools that are in suburbs that don't even exist in the top suburb selection. For example:

	School_Name	Suburb	avg_ICSEA_percentile	avg_total_enrolments
1	Townsville Grammar School	North Ward	90	1235

Townsville Grammer School is the most ideal suburb to advertise to just by looking at its attribute values. It is important to note that this school is located in the suburb North Ward which is not included in the result set for top 10 suburbs. For this reason, the perspective of this project must be shifted and we must look for data on specific schools to advertise to.

Outliers

When querying for rural schools best suited for our advertisement need, the data set reveals some obvious outliers that must be examined.

2	Cairns School of Distance Education	Manunda	59	4005	0.735

…

Developing a Map Representation

At this stage we have the recommendations for Suburbs and Schools to advertise to based on:

- Rural Schools
- Rural Suburbs
- Special Needs Schools

Complications

We must now gather the longitude and latitude of these schools and suburbs to get an accurate representation of their location on a visual map. The problem with this is that the school's data set does not contain any data on the locations of the schools in the database. There is however, a separate data file called Locations which has the longitude and latitude values of all schools in the school's data base.

Methodology

Add the Locations data file into the database for schools that we are using. This allows the database to have multiple tables allowing us to connect attributes between them and find the locations of schools that we have suggested for advertisement.