

Phase 3 Case Study

Victorian Education

Elements of Data Processing

Name: Akira Wang

Research question

What are the major key factors that determine a school's performance and the outcome of students in Victoria (through the VCAA education system)

Why?

The Victorian education system has had numerous changes in the recent years, as well as the increasing number of demands for higher achieving schools with families. I thought it would be interesting to see if the location of a school as well as the students that attend it would have any impact on a schools overall performances. Especially in Melbourne, we are diverse in high schools with a large bulk of independent and catholic run schools.

Main topics of research

High Performing Schools

What separates them from normal schools?

- Selective entry?
- Number of students?
- Variety of subjects offered?
- Location?

Locality & educational availability

Does the locality influence education?

- What's the correlation?
- Differences between localities
- Do certain suburbs offer a larger variety of school sectors?

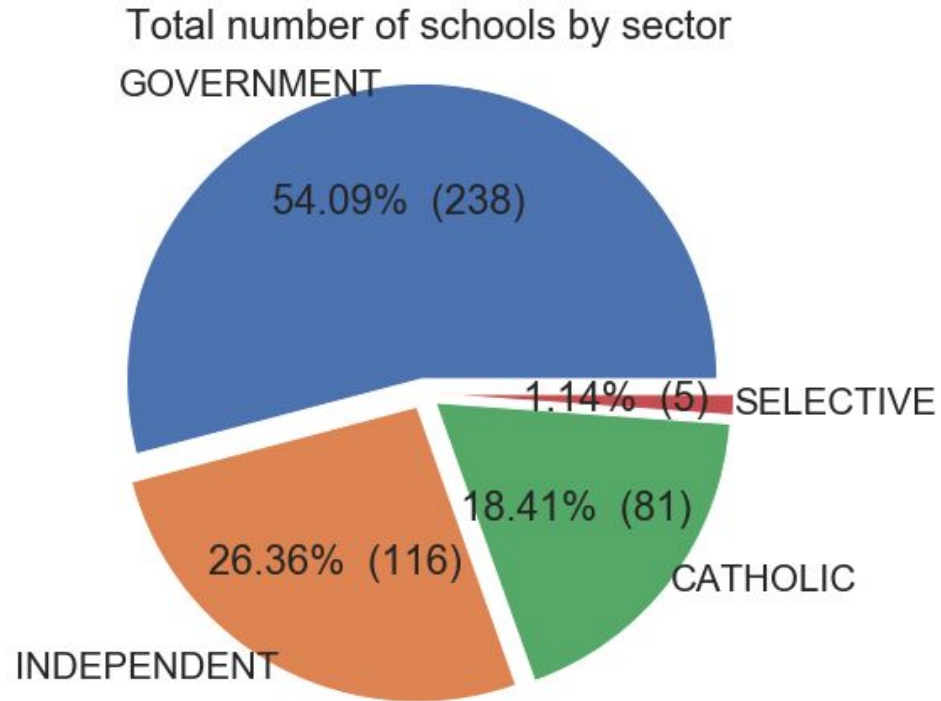
Diversity of Graduate Pathways

What do the statistics of graduate pathways suggest?

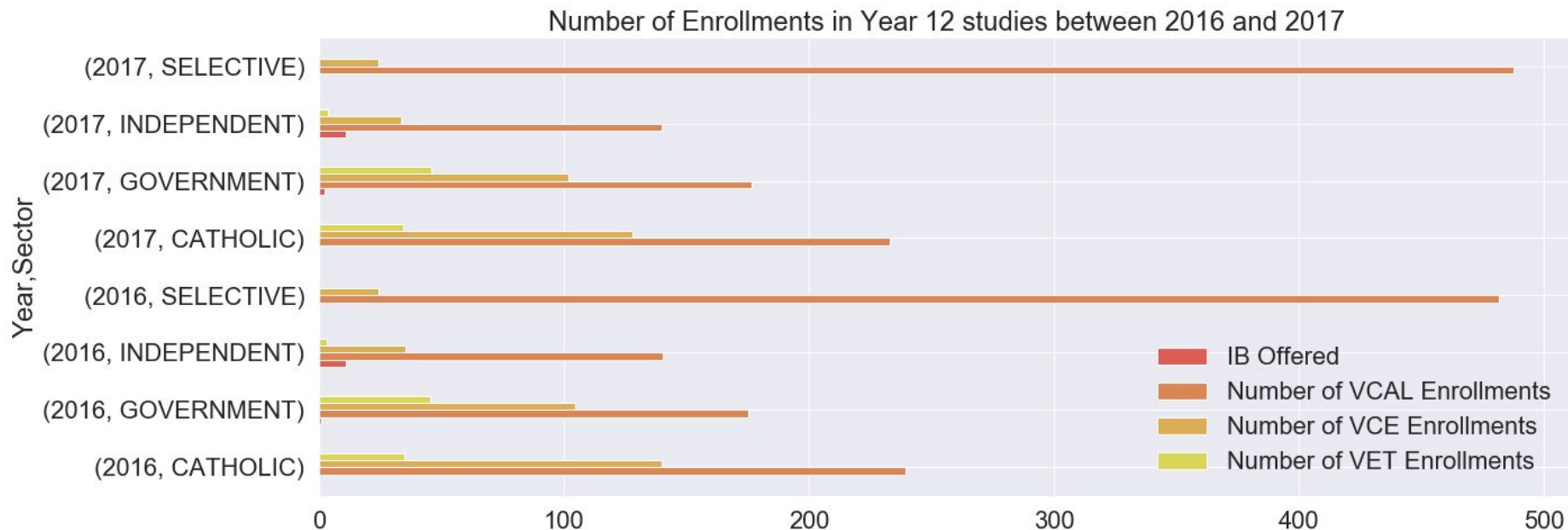
- Which sector is better for pathways
- Where do graduates mainly go

General Distribution of Sectors

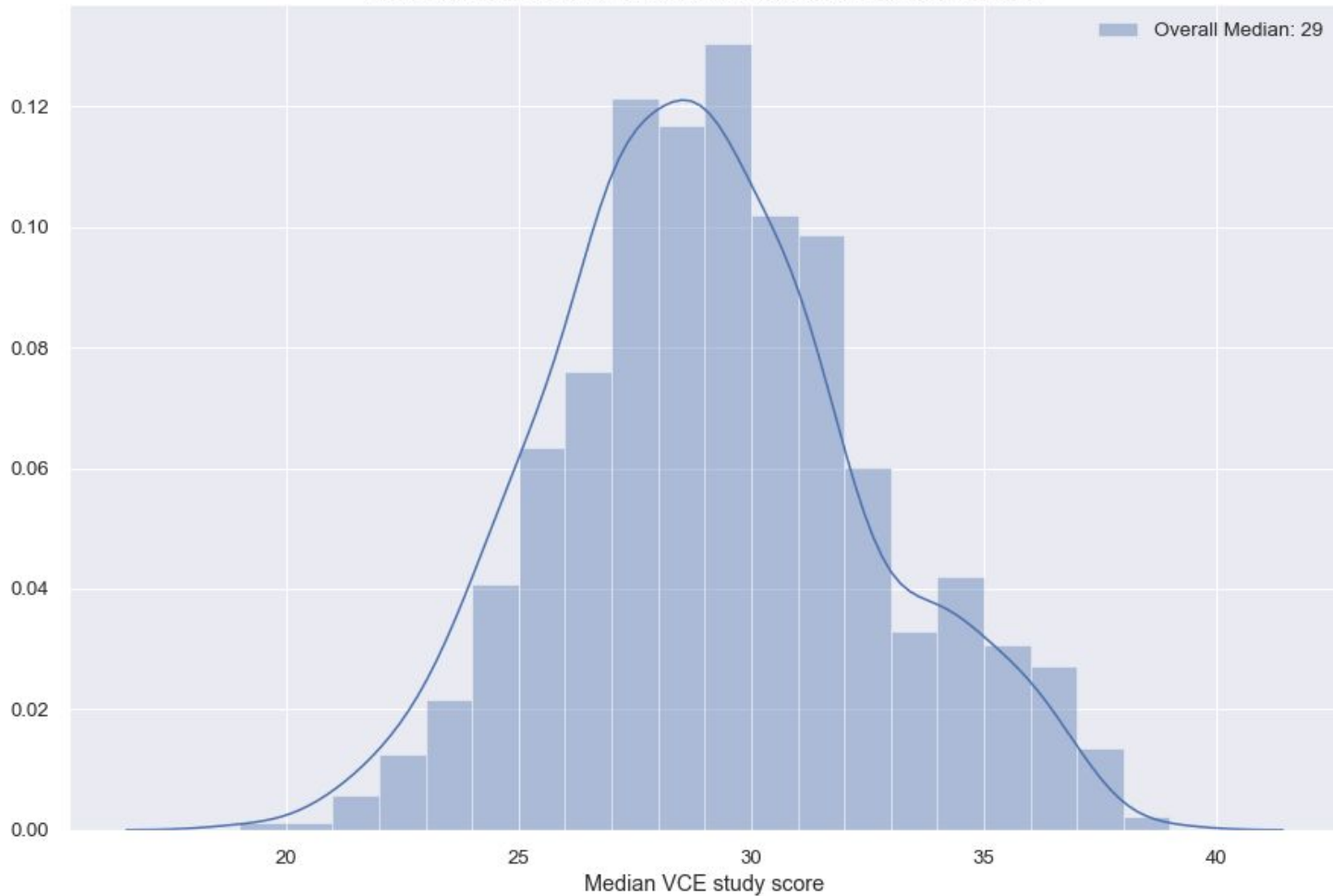
Selective schools are funded by the government



Average number of student enrollments

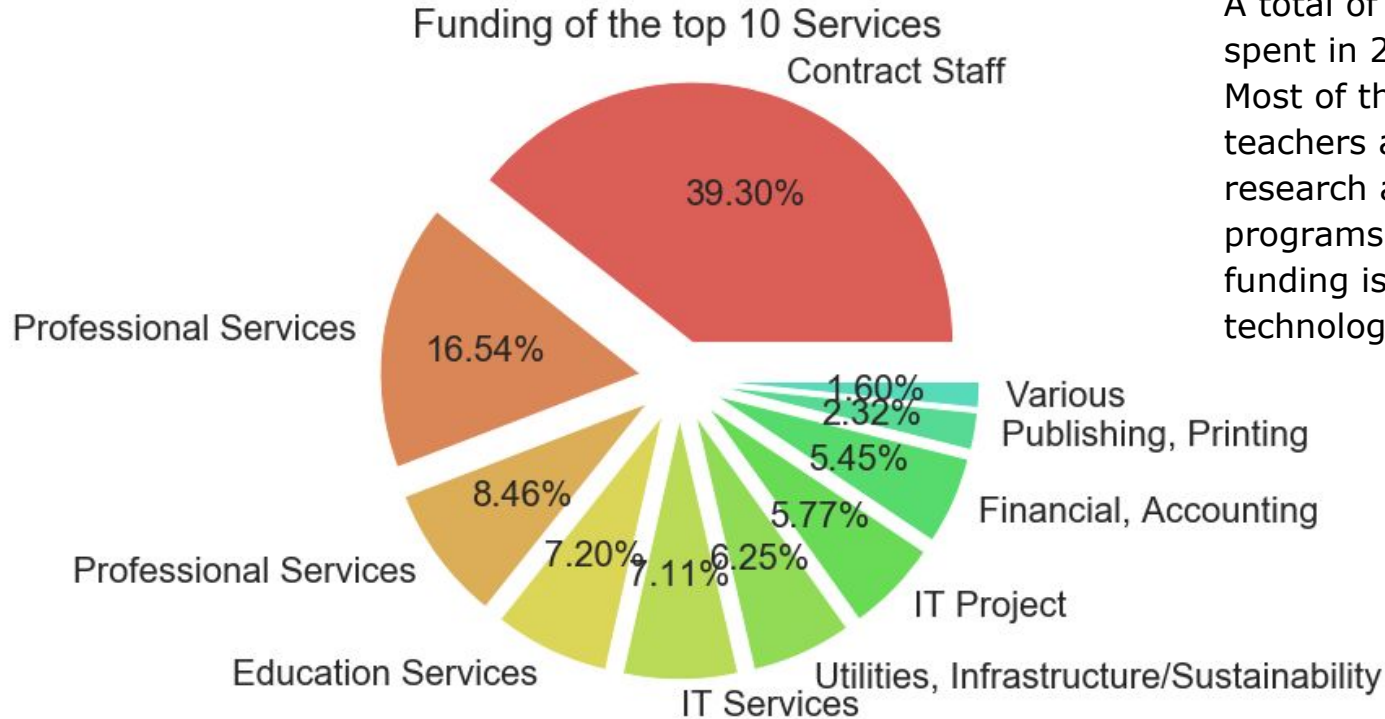


Distribution of VCE study scores not including Adult or Small Schools



The actual median is 30 (calculated using a bell curve), but for the sake of this analysis I have ignored adult institutions and schools with less than 20 students under recommendation from VCAA

Education Funding

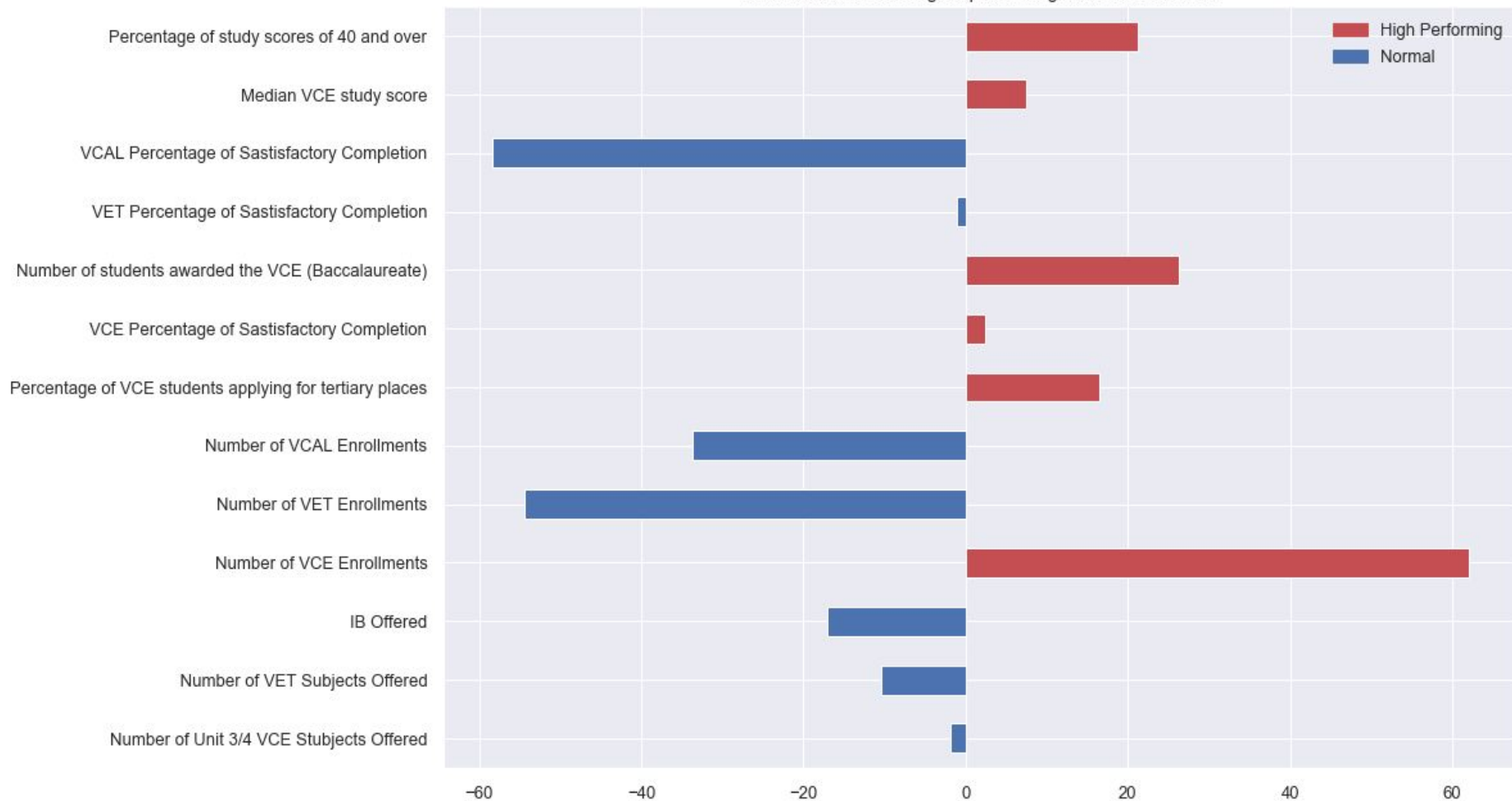


A total of \$133,586,693 was spent in 2016 on contractors. Most of the funding goes to teachers and staff, followed by research and development programs. Quite a fair bit of the funding is also poured into technology

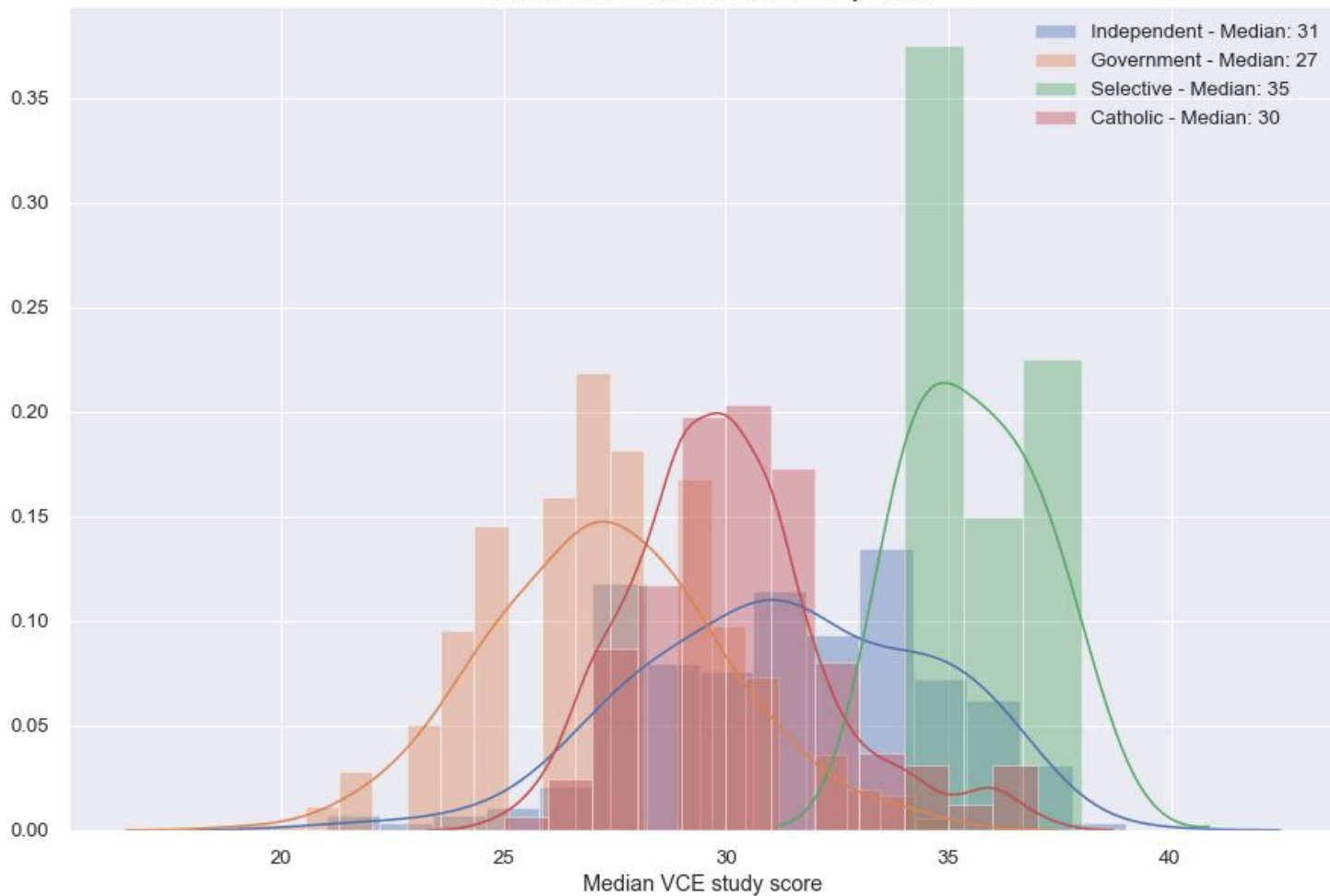
What separates a high achieving school from a normal school?

Schools with a median study score of 35 and above are classified as high achieving / performing

Differences between higher performing and normal schools



Distribution of Sector Median VCE study scores

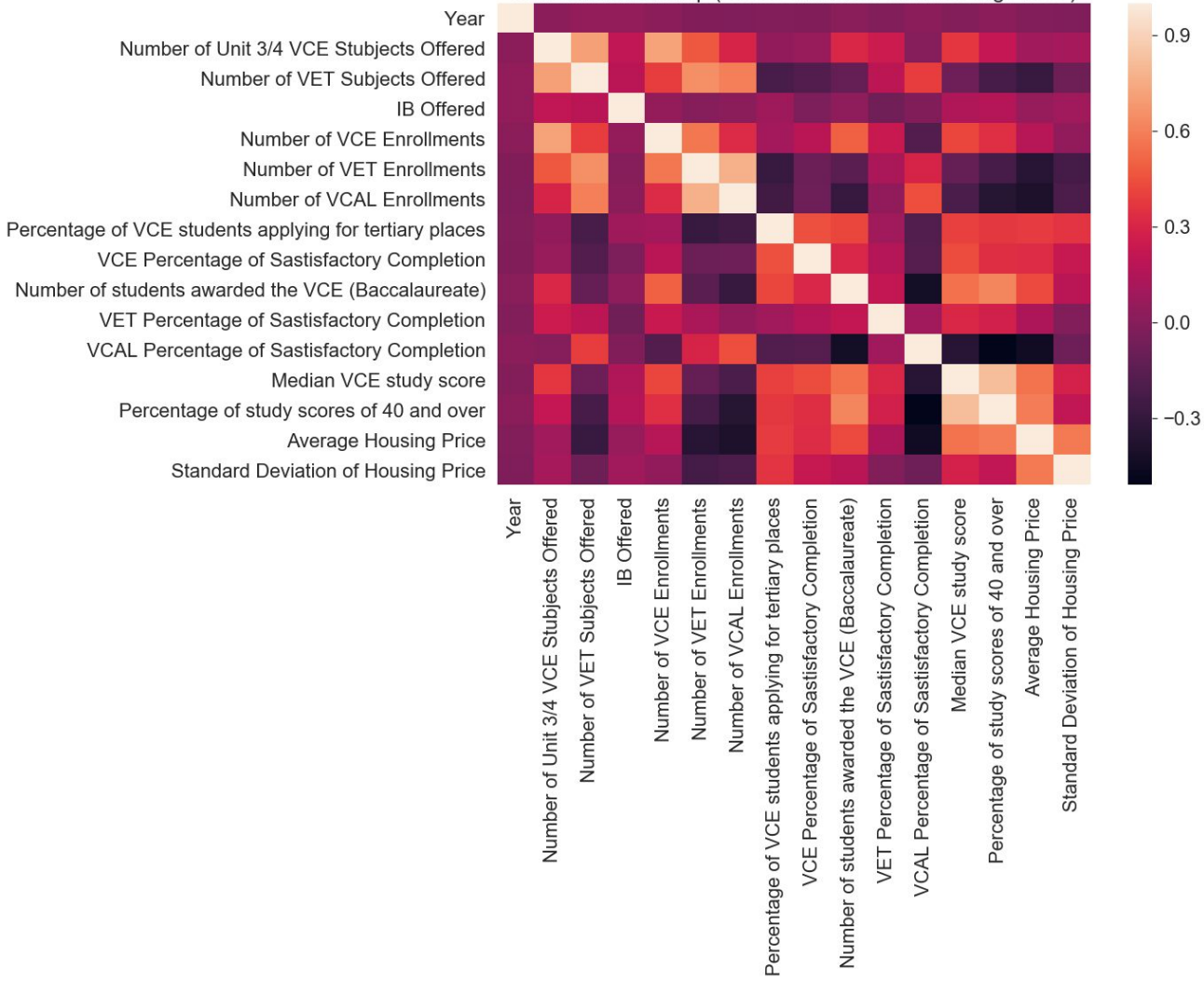


Selective entry schools perform well above any other sector overall



Does the locality influence
education?

Correlation Heatmap (School attributes and Housing Prices)



Correlation:

Median VCE study score: 0.5542

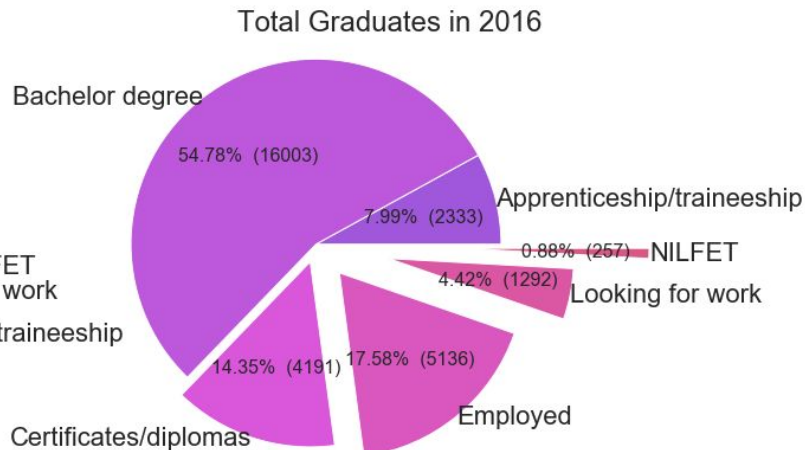
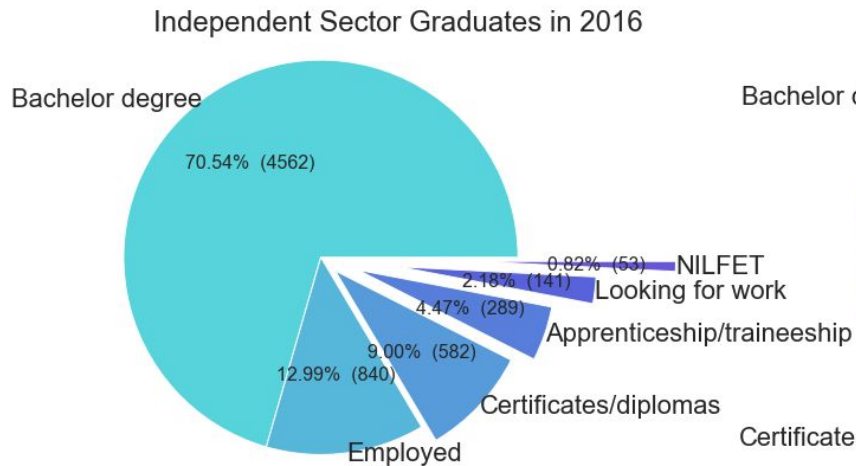
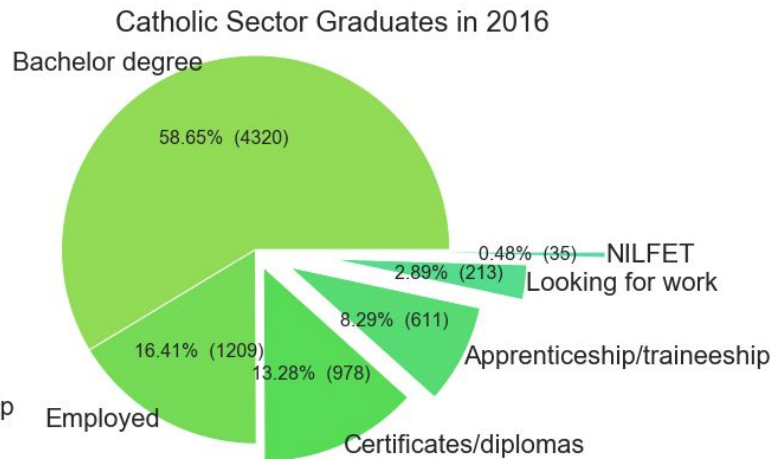
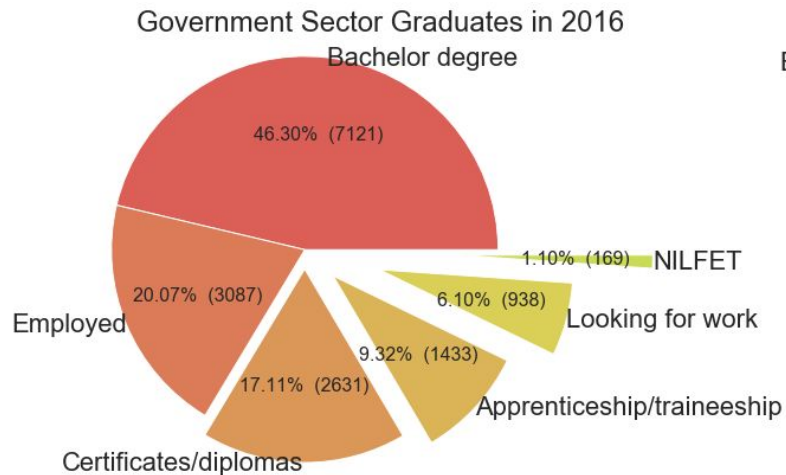
Percentage of study scores of 40 and over: 0.5899

Top 10 suburbs with the best availability

	Average Housing Price	Median VCE study score	Schools	Schools offering IB
Locality				
KEW	481.000	33.000000	10	3
BERWICK	385.000	32.428571	10	0
SHEPPARTON	266.875	27.500000	10	0
WERRIBEE	333.125	30.500000	9	2
BALLARAT	270.625	30.666667	9	0
CRANBOURNE	370.000	28.000000	9	0
ESSENDON	377.500	29.500000	9	0
BROADMEADOWS	339.375	28.800000	8	0
CANTERBURY	462.875	33.500000	8	0
WARRNAMBOOL	290.625	29.000000	8	0



What do the graduate pathways
look like?

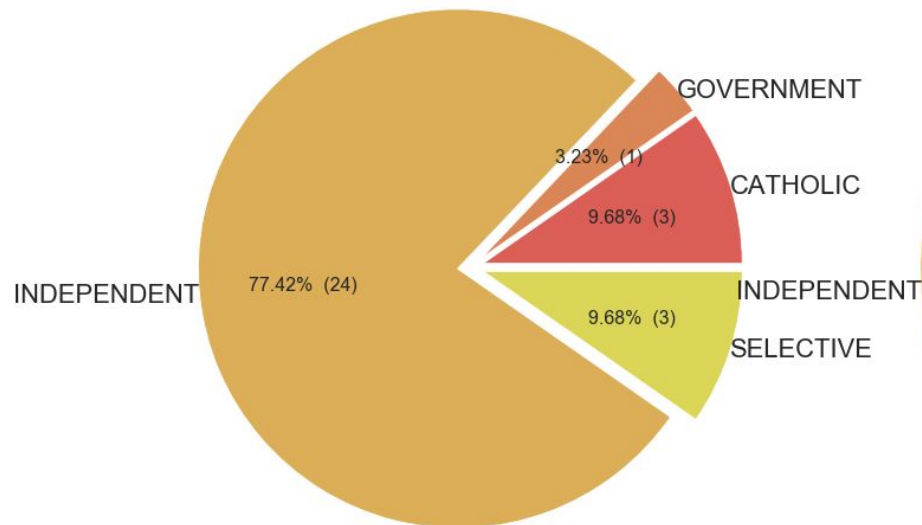


2016 Graduate Pathway Statistics



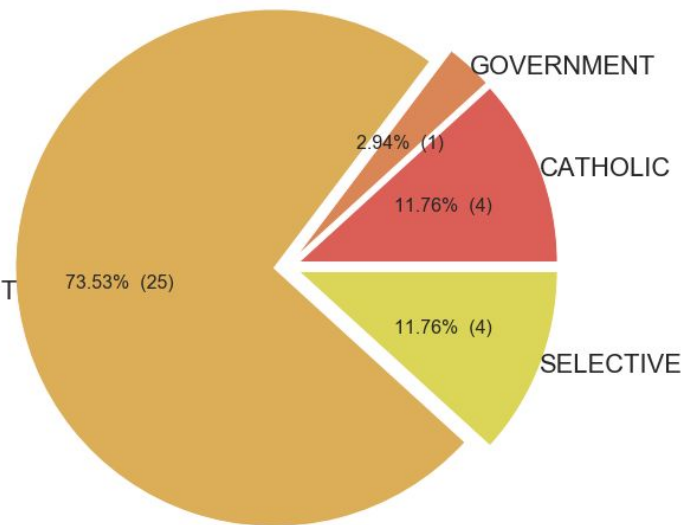
Statistics of the Top Schools

2016 Top Performing Schools by Sector



- SACRE COEUR
- MORTLAKE COLLEGE
- BIALIK COLLEGE
- MACROBERTSON GIRLS HIGH SCHOOL

2017 Top Performing Schools by Sector



- ST KEVIN'S COLLEGE TOORAK
- VICTORIAN COLLEGE OF THE ARTS SECONDARY SCHOOL
- KOROWA ANGLICAN GIRLS' SCHOOL
- MACROBERTSON GIRLS HIGH SCHOOL

	Sector	Locality
School		
BENDIGO SENIOR SECONDARY COLLEGE	GOVERNMENT	BENDIGO
MACROBERTSON GIRLS HIGH SCHOOL	SELECTIVE	MELBOURNE
MELBOURNE HIGH SCHOOL	SELECTIVE	SOUTH YARRA
BIALIK COLLEGE	INDEPENDENT	HAWTHORN EAST
WERRIBEE SECONDARY COLLEGE	GOVERNMENT	WERRIBEE

Top 5 overall schools boasting in:

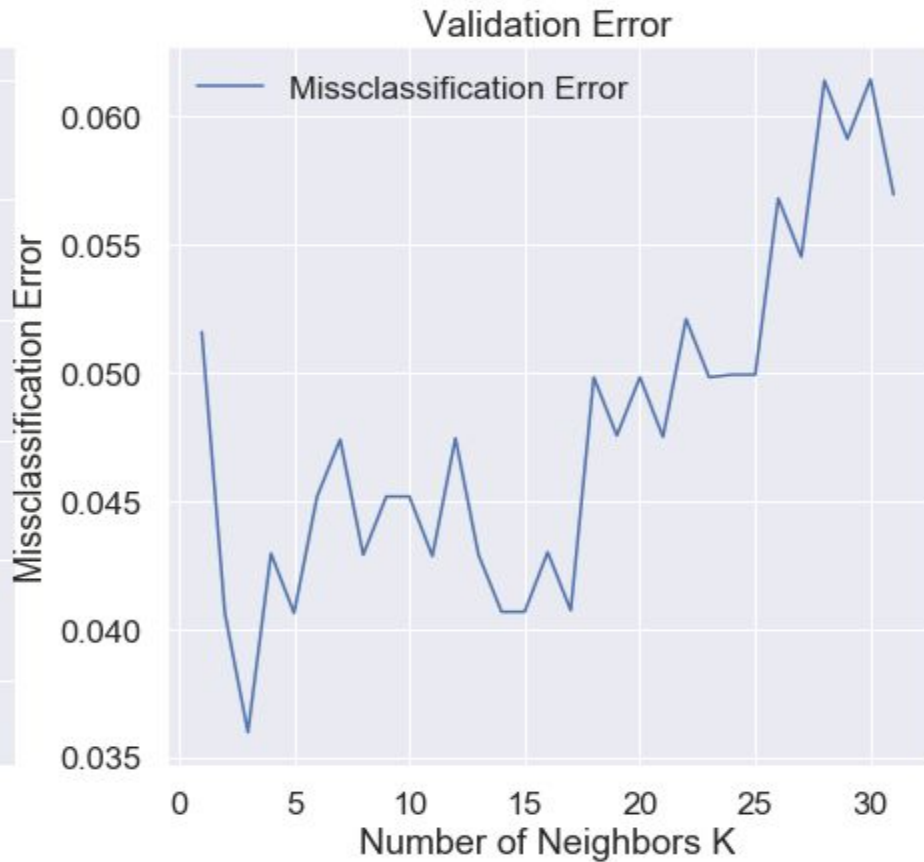
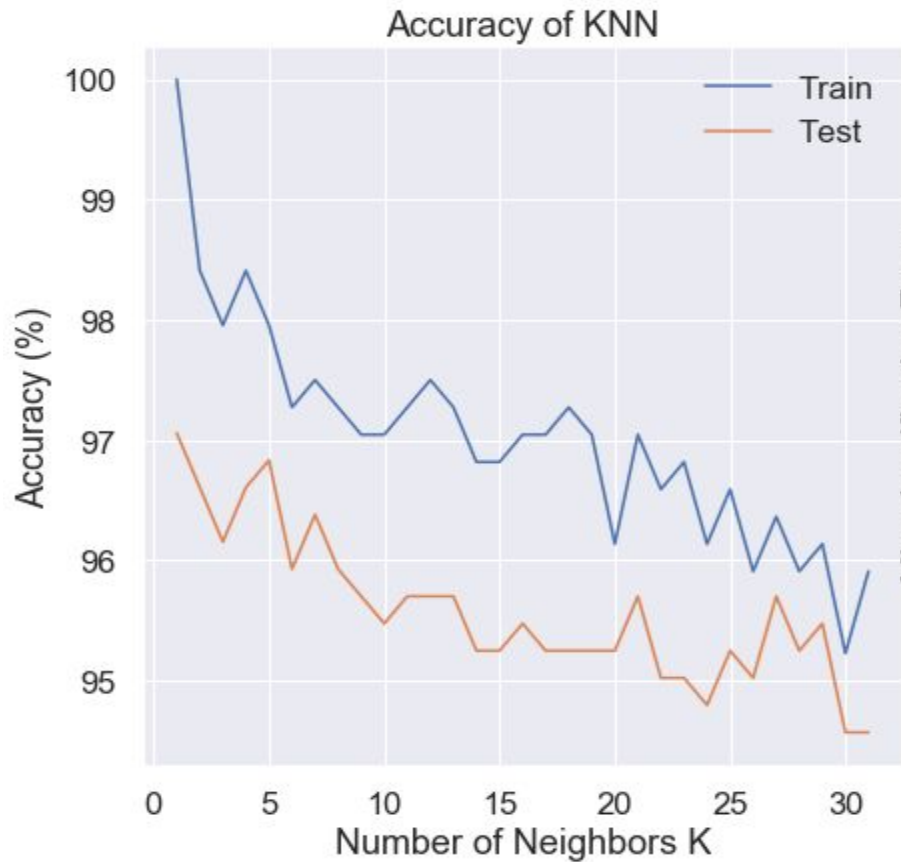
- High number of student enrollments (Average of 500+)
- High number of subject variety (75+)
- High average median study score (34)
- At least 95% complete VCE with a pass and roughly 90% apply for tertiary education

Note that Bendigo SSC and Werribee SC are **NOT** high achieving schools, but are overall top due to the sheer number of variety and student enrollments

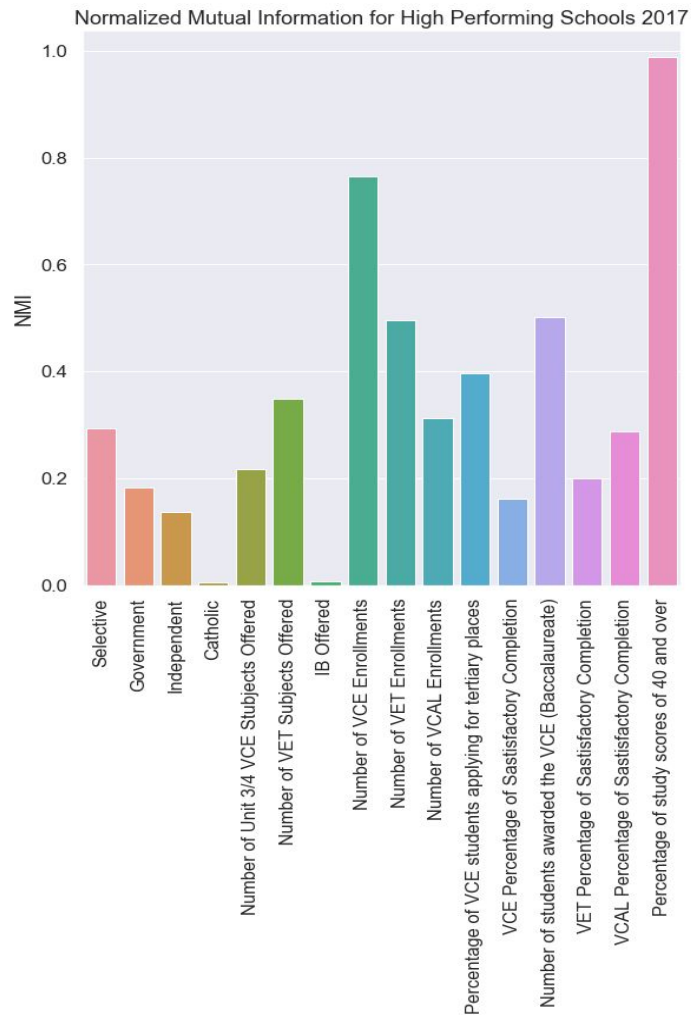
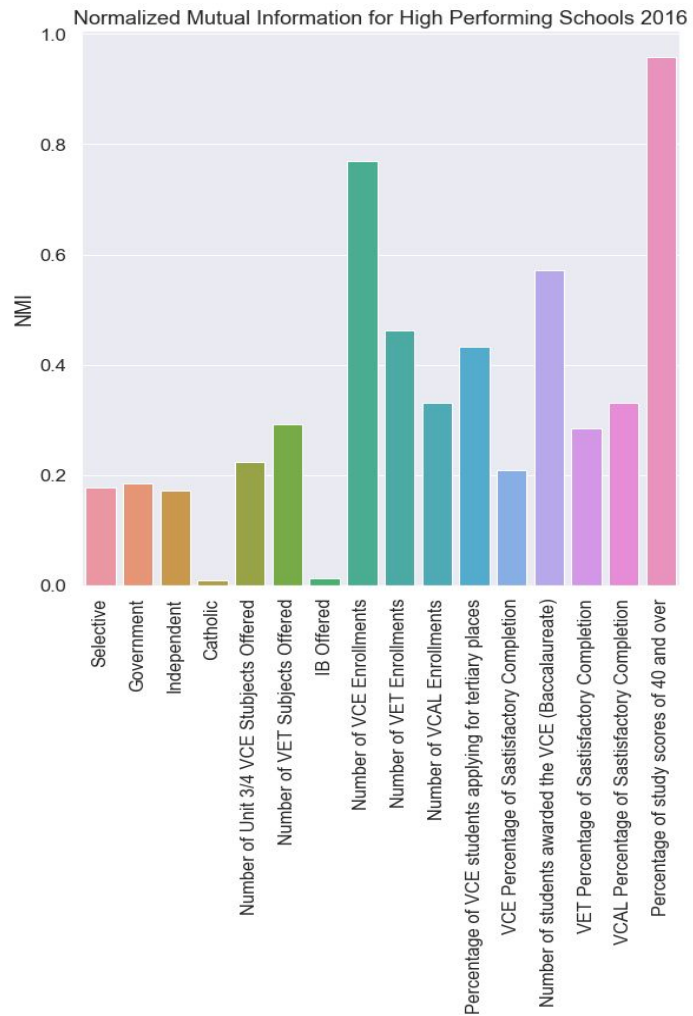


Labelling and Predicting future High Performing Schools

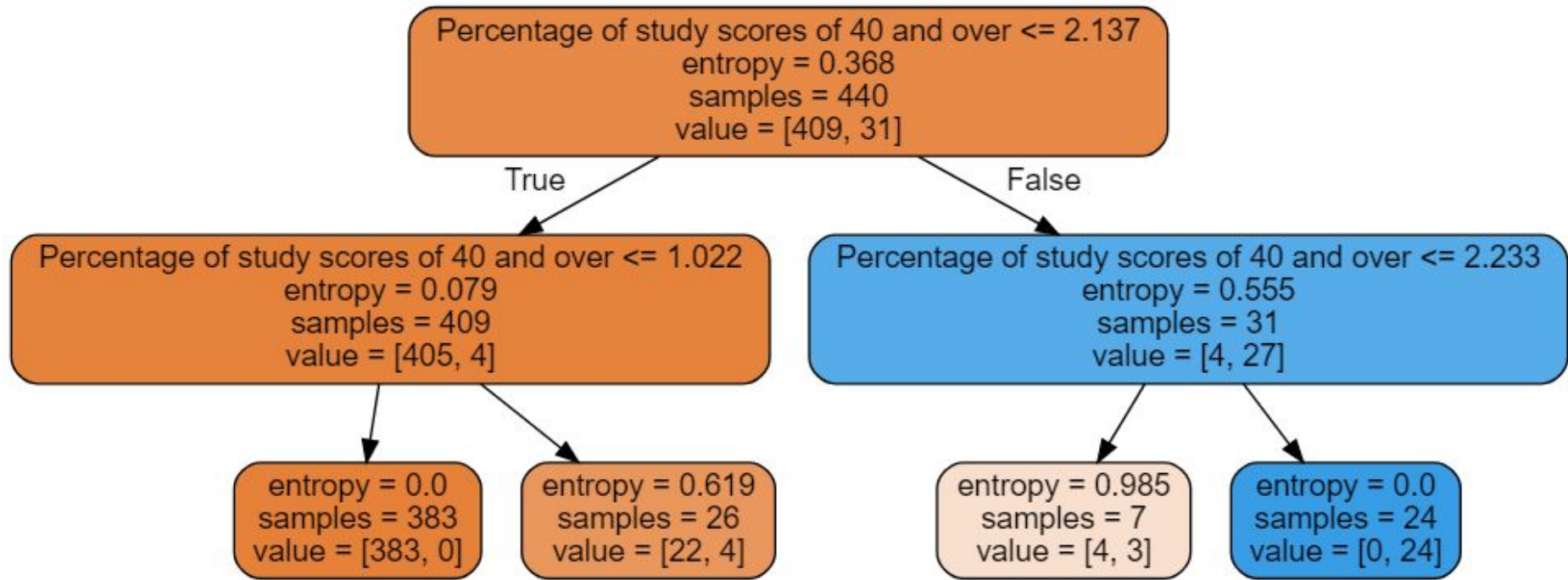
Using the year 2016 as the training data to predict the schools of 2017



With the optimal number of neighbours given as [3] - misclassification error of [0.03596195]:
Train Accuracy: 97.95%
Test Accuracy: 96.15%



The NMI suggests that the most important factor of a school is relative to the number of student enrollments and the percentage of those students that achieve a study score over 40



As expected, using an optimal max_depth of 2 - the results verify that the percentage of study scores over 40 is the biggest factor that determines a schools performance. The overall trend can be fitted to a linear regression model to predict schools which can grow into high achieving schools

97.74% with train accuracy 99.77%

Conclusion

- High achieving schools generally have a higher student enrollment count within a location with a higher rental cost per week.
- For households enrolling their children into a “local” school, their best chance would be to move to the Eastern suburban areas of Victoria which perform better compared to their counterpart schools in other suburbs.
- Students that generally attend selective entry schools will perform significantly better compared to any other sector as expected
- Sending your children to an independent school will yield in a better education compared to a local school - but is it really worth the money? (The cost can be from anywhere between 10k to 50k a year)
- Whilst most independent schools result in graduates pursuing a tertiary education, those in the government sector may choose to enter the workforce straight after graduating. Overall, half the students will end up studying at a tertiary level

Problems when dealing with the datasets and future changes

- Bad formatting (symbols such as *, Y, <, A filled the dataset)
- Excel parsing (spreadsheets imported were terribly formatted and easier to fix in Excel)
- Merging housing prices with the VCAA dataset - same localities but different names (Example ALBERT PARK-MIDDLE PARK-WEST ST KILDA = ALBERT PARK, after fixing it an extra 44 localities were recovered)
- Be able to predict a sector / school performance given less attributes (location, number of students, sector)
- Work with bullying rates within schools to see if there is an effect on school performance (higher bullying rates = lower overall student performance)

Extra information, code and dataset cleaning

<https://github.com/akiratwang/Victorian-Education-Analysis/blob/master/Analysis.ipynb>