Hyperiondev

TASK

Exploratory Data Analysis on the 2016 Matric Reports DataSet

Visit our website

Introduction

This dataset describes matric student performances in schools from 2014 to 2016, separated by quintile, which is a school ranking system (1 - 5) used by the education department to classify a school in terms of school facilities (science lab, library, computer lab, pools etc), where it is based (rural, informal, CBD), teacher qualifications (B.Ed., Masters etc.) and other characteristics not mentioned.

Table 1: Summary of Data

	Quintile	Wrote 2014	Passed 2014	Pass (%) 2014	Wrote 2015	Passed 2015	Pass (%) 2015	Wrote 2016	Passed 2016	Pass (%) 2016
count	886.00	886.00	886.00	886.00	886.00	886.00	886.00	886.00	886.00	886.00
mean	2.57	79.35	54.52	65.76	100.14	61.38	59.69	94.38	60.11	62.87
std	1.17	55.52	46.62	23.07	69.93	53.17	24.59	71.63	52.47	24.85
min	1.00	5.00	0.00	0.00	5.00	1.00	2.20	4.00	0.00	0.00
25%	2.00	40.00	21.00	49.70	50.00	23.00	39.40	43.00	21.00	43.65
50%	3.00	66.00	41.00	67.70	85.00	46.00	58.40	77.00	46.50	63.50
75%	3.00	103.75	74.00	85.00	132.00	84.00	80.00	126.00	81.00	85.25
max	5.00	416.00	343.00	100.00	648.00	428.00	100.00	707.00	378.00	100.00

Dataset = 1000 rows x 15 columns

DATA CLEANING

The following columns were dropped by passing a list of the column names in a for loop to a method that called the drop() method:
emis, **centre_no**, **progressed 2015 and 2016**

Duplicated() method was employed to find duplicates and drop_duplicates() to drop the rows. Initially I could not delete the duplicated rows as they seem to have the same school's name but different centre number. Further investigations based on the centre numbers provided the following example:

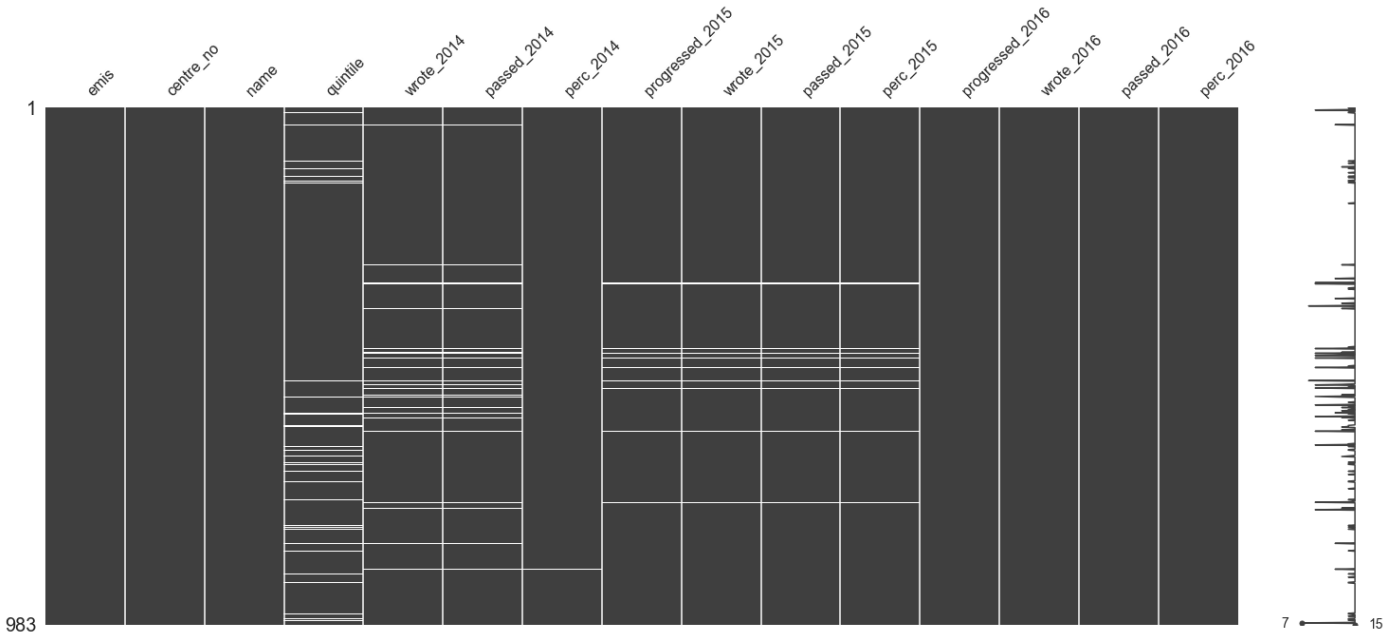
4161035 Zwelandile Senior SecondaN
4141025 Zwelandile Senior SecondaN.

It seems that even though the centre number is different, the school is the same. With this information, I have decided to drop the 17 duplicated rows in the 1000 row database

0 Duplicate columns were found and dropped
17 Duplicate rows were found and dropped

Quintile column had values of 99.0, which is inconsistent. Quintiles are from 1 to 5. Replaced all 99.0 with NaN

MISSING DATA



Approximately 1.8% of the data was missing by either NaN or irregularities in the data which were then converted to NaN. The small amount of missing data removed wouldn't affect the data significantly and still be able to run analysis on the data available. These schools are independent and since there's no regional category, all these schools are taken as a whole. So the data can then be assessed on what is available, and will not bias the data significantly.

DATA STORIES AND VISUALIZATIONS

The dataset provides data on schools in South Africa with matriculant pass rates. No geolocation data was present. I believe that some of the main questions here are, what schools had the highest pass rates? What's the top 10 schools in South Africa? what quintile are those schools? and do students perform better at higher quintile schools than lower ones?

Top 10 schools in South Africa with the highest pass rates over the years, based on the number of enrolled students are presented in the following table:

Table 2. Top 10 Schools in South Africa

	2014	Q_2014	2015	Q_2015	2016	Q_2016
1	HUDSON PARK HIGH SCHOOL	5	CAMBRIDGE HIGH SCHOOL	5	PEARSON HIGH SCHOOL	5
2	STIRLING HIGH SCHOOL	5	FICHARDTPARK SS	5	FICHARDTPARK SS	5
3	FICHARDTPARK SS	5	GREY BOYS SCHOOL	5	VICTORIA PARK HIGH SCHOOL	5
4	EUNICE SS	5	WESTERING HIGH SCHOOL	5	CAMBRIDGE HIGH SCHOOL	5
5	PEARSON HIGH SCHOOL	5	PEARSON HIGH SCHOOL	5	GOUDVELD SS	5
6	WESTERING HIGH SCHOOL	5	EUNICE SS	5	WESTERING HIGH SCHOOL	5
7	GREY BOYS SCHOOL	5	C EN N H MEISIESKOOL ORANJE	5	EUNICE SS	5
8	ST PATRICKS SENIOR SECONDARY SCHOOL	3	Hoerskool NICO MALAN	5	SASOLBURG SS	5
9	CLARENDON GIRLS HIGH SCHOOL	5	SELBORNE COLLEGE BOYS HIGH	2	C EN N H MEISIESKOOL ORANJE	5
10	C EN N H MEISIESKOOL ORANJE	5	CLARENDON GIRLS HIGH SCHOOL	5	CLARENDON GIRLS HIGH SCHOOL	5

It is evident from table 2, that the majority of the schools that fall in the top 10 are those of quintile 5. Many of these schools fluctuate within the top 10, and mainly maintain their position.

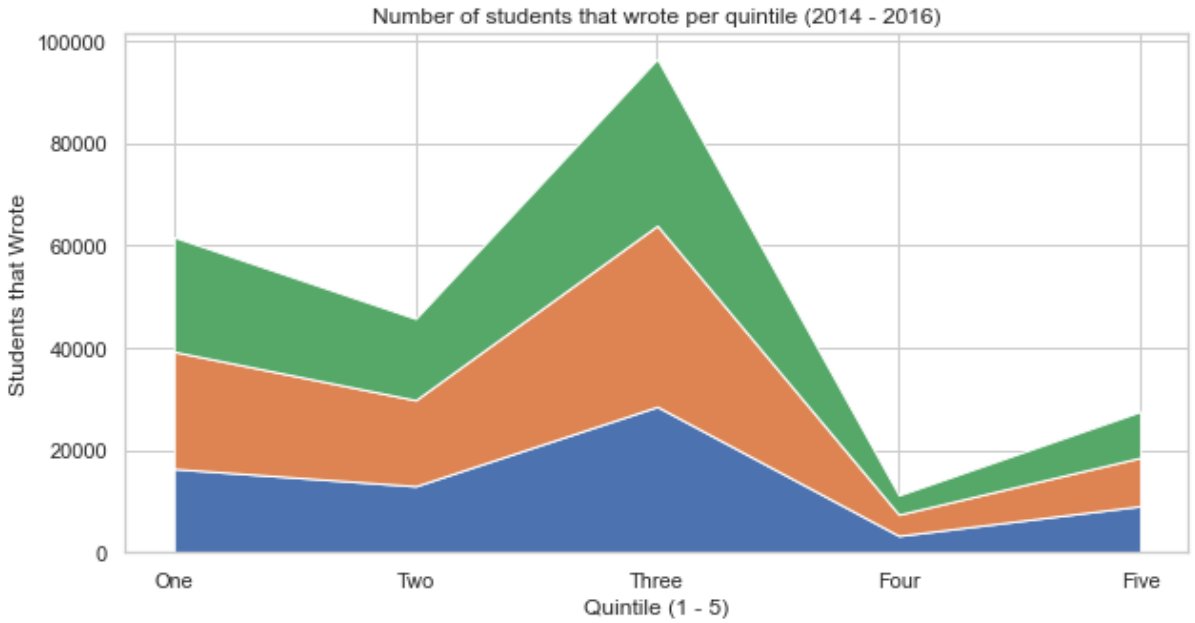


Figure 1. Number of students per quintile over the past three years

The number of students that wrote the matric examinations is presented graphically by figure 1. It appears evident that the majority of students are from quintile 1 to 3, with quintile 3 schools having significantly larger numbers of students.

It is important to note that Quintiles 1 – 3 schools are "no fee" schools and the state compensates the schools for a lack of income, while quintiles 4 and 5 are "fee paying" schools. This makes sense, given the poverty in South Africa, why quintile 1 to 3 schools would have a much higher student count than fee paying schools of quintile 4 and 5.

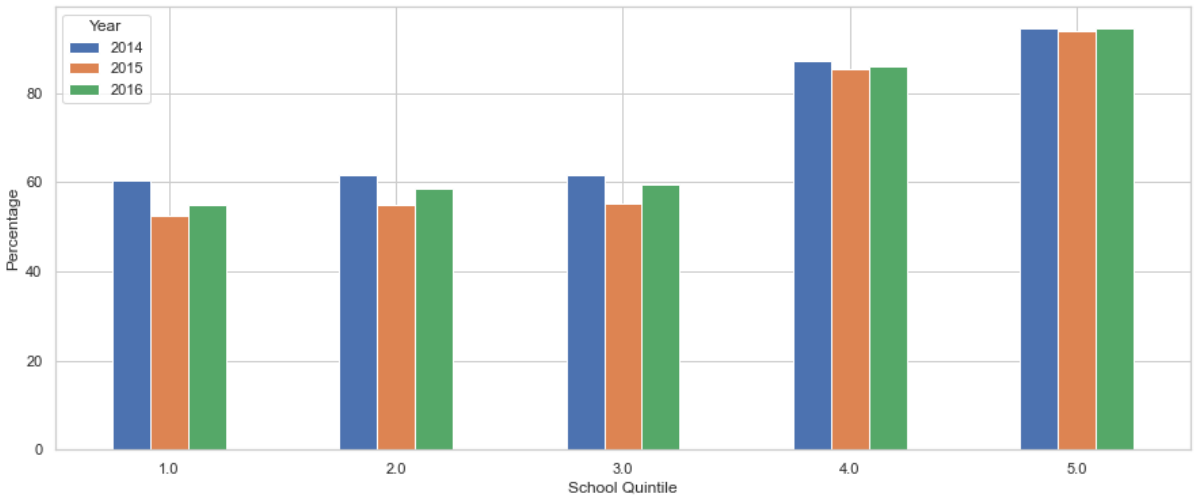


Figure 2. Pass percentage per quintile, per year

The illustration in figure 2, indicates that over the years schools that are quintiles 1 – 3 have more or less a similar student pass rate pattern, while schools characterized as quintile 4 and 5 have much larger percentage of pass rates. So, the question posed is, do students perform better in higher quintile schools? And if so, what are the underlying contributing factors that promote this? These are some questions that cannot be answered by the lack of data in this dataset leading to those type of investigations.

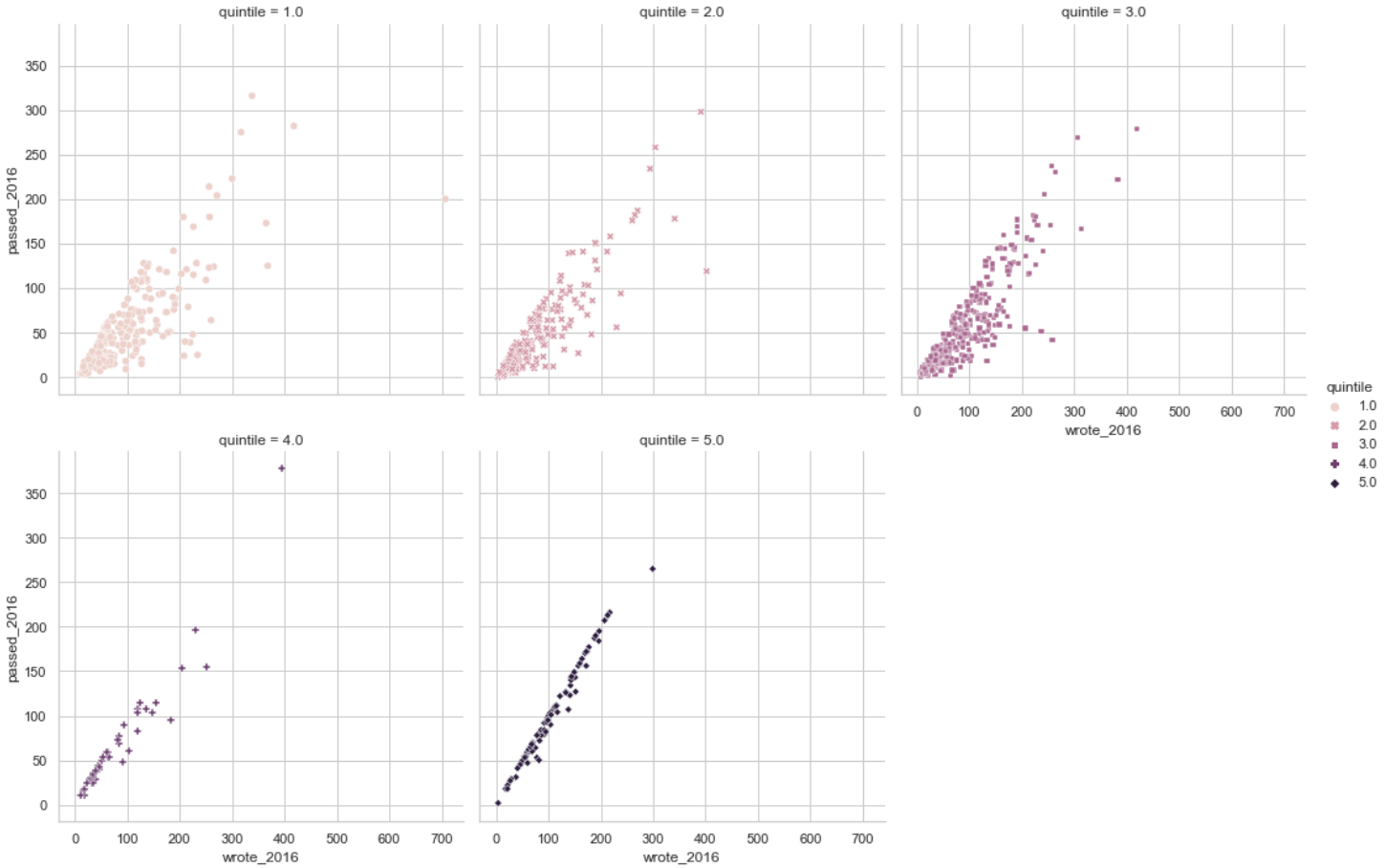


Figure 3. Student pass rates vs those that wrote per quintile in 2016

The illustration in fig 3 helps to answer this question. Looking at the quintile 1 to 3 schools, there are large frequencies of data at the lower end, and very few schools with high student pass rates in the upper quartiles, and these schools seem to follow a very similar pattern.

How do quintile 4 and 5 school pass rates compare from 2014 to 2016 with other quintile schools, is another question to pose.

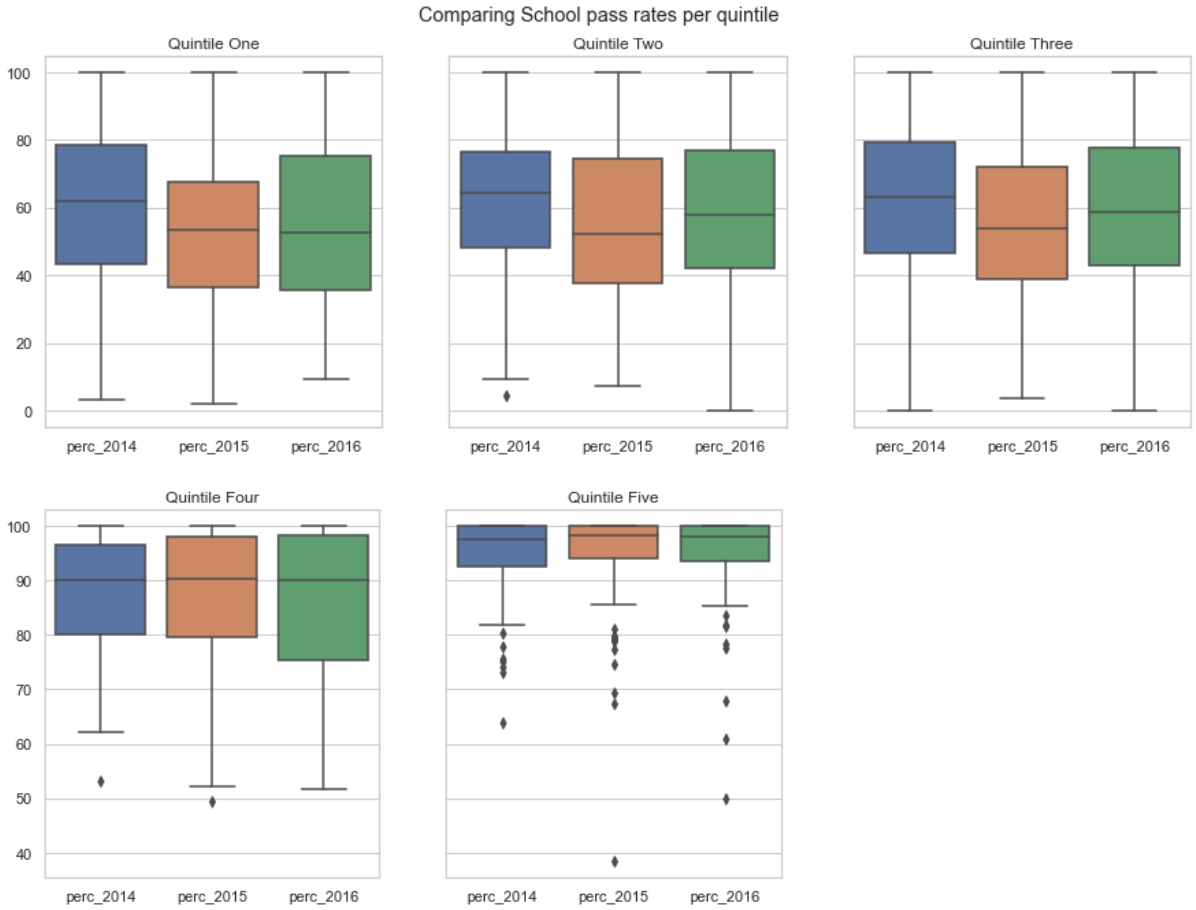


Figure 4. School pass rate per quintile in the past three years

The bar plot presented by figure 4, validates the information presented in figure 3. It can be seen that the trend in figure 3 is similar for the past three years (2014 - 2016). However there many outliers displayed in quintile 5 schools, this indicates that even though those schools are of quintile 5, providing the best education facilities, still there are many schools which have lower pass rates than the expected standard that these schools have set.

The total number of students that passed matric over the 3 years, even though display a high throughput (fig 5), still indicates that many students do not.

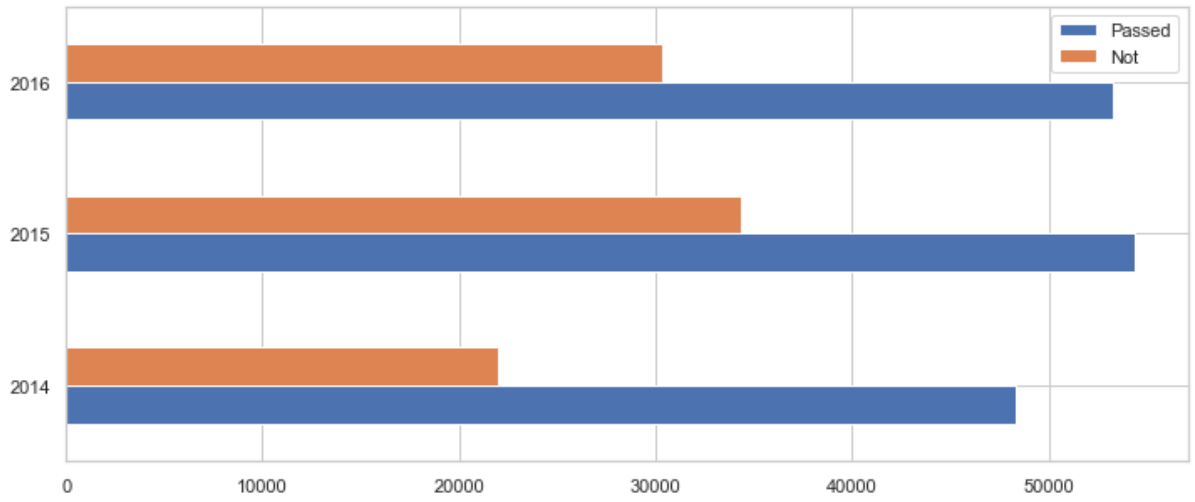


Figure 5. Total number of students that passed and that did not over 3 years

As discussed previously, many contributing factors can be discussed on this topic had there been more data on geolocation and the types facilities and funding these schools have.

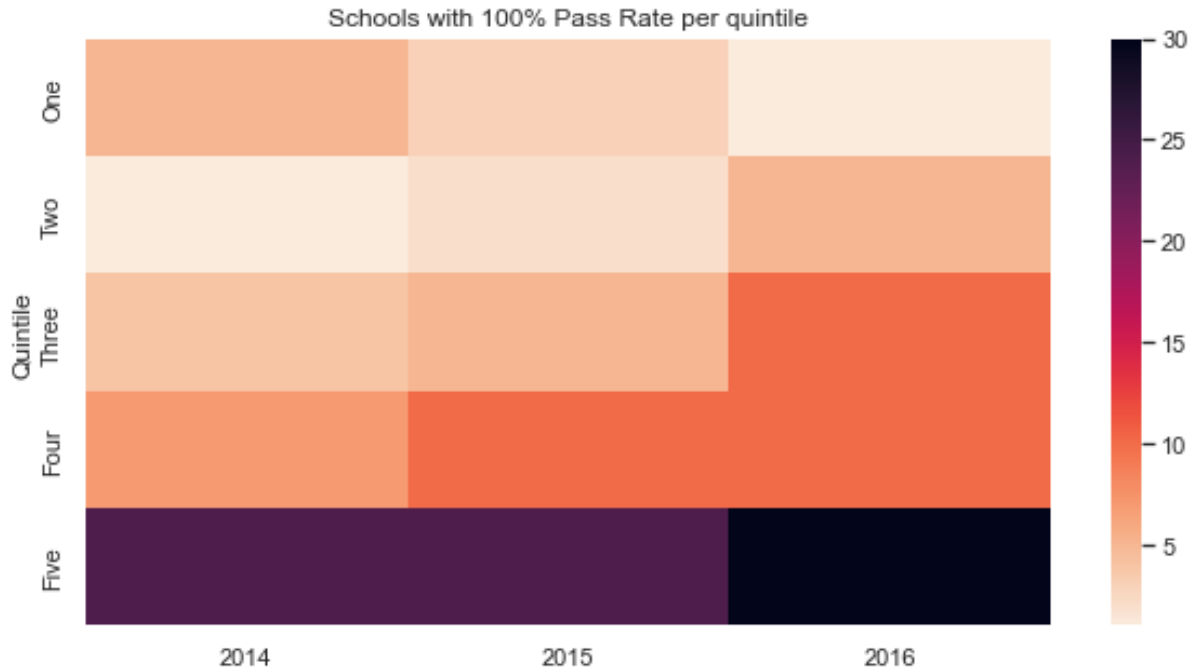


Figure 6. Heatmap displaying number of schools per quintile with 100% pass rate

The heatmap (fig 6) indicates most schools with 100% pass rate were of Quintile 5, and as of 2016, that number had increased in schools of quintiles 2, 3 and 4.

There is a slight decline over the years in Quintile 1 schools, and those reasons cannot be ascertained in this dataset. However, it may be attributed to the types of students that attend, and the facilities that are provided.

Table 3. Schools at 50th Percentile

Quintile	2014	2015	2016
One	61.70	53.30	52.70
Two	64.45	52.25	57.85
Three	63.20	54.00	58.65
Four	89.90	90.20	90.10
Five	97.45	98.10	97.9

Table 3 displays the 50th percentile for each quantile over the past 3 years. It is apparent that quintiles 4 and 5 schools are definitely high achievers.

The dataset discussed herein this document, that increasing quintile schools doesn't necessarily increase a schools pass rate. The distinction here is that quintile 1 – 3 are "no fee" paying schools which are compensated by the state, and quintile 4 and 5 are "fee paying" schools, furthermore, free schools, though have a higher student count, have a lower pass rate in comparison to fee paying schools, and this may be attributed to the lack of facilities provided by those schools, however more data needs to be provided in order to properly assess this avenue. It can be concluded, that in essence, Quintile 4 and 5 schools do have a higher pass rate than the lower quintile 1 to 3 schools.

THIS REPORT WAS WRITTEN BY : Raman Sewjugath