Hyperiondev

TASK

Exploratory Data Analysis on the 2014 Matric results Data Set

Visit our website

Introduction

This dataset is comprised of matric results from 2012,2013 and 2014. It also has data from schools part of the Dinaledi programme aimed at empowering previously disadvantaged schools. Quintile categories are use to classify schools based on resources and how they previously performed. What type of data do we have?

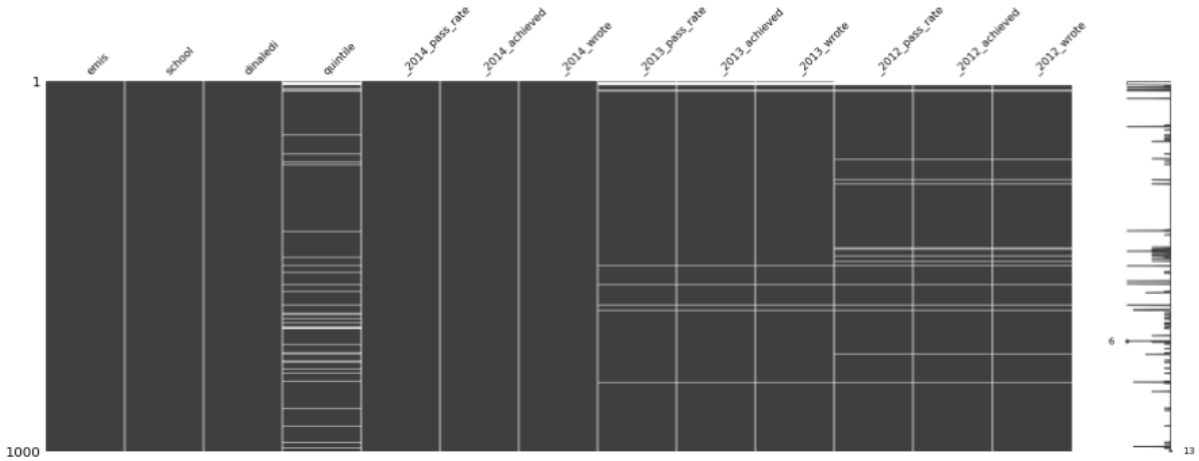
1. emis [nteger] [Categorical] Unique school identifier
2. exam_no [nteger] [Categorical] Unique exam identifier
3. school [string] [Categorical] Name of school
4. dinaledi [string] [Categorical] Is the schoold part of Dinaledi development programme
5. quintile [integer][Categorical] Classification according to school's prehistoric available resources
6. _2014_pass_rate [double] [Continuous] 2014 Pass rate
7. _2014_achieved [integer][Continuous] How many students passed
8. _2014_wrote [integer][Continuous] How many students wrote
9. _2013_pass_rate [double] [Continuous] 2013 Pass rate
10. _2013_achieved [double] [Continuous] How many students passed
11. _2013_wrote [double] [Continuous] How many students wrote
12. _2012_pass_rate [double] [Continuous] 2012 Pass rate
13. _2012_achieved [double] [Continuous] How many students passed
14. _2012_wrote [double] [Continuous] How many students wrote •

DATA CLEANING

- The column 'exam_no' was dropped as we already have a unique identifier column with the ‘emis’ number
- The missingno library was used to identify and visualize missing data on figure 2

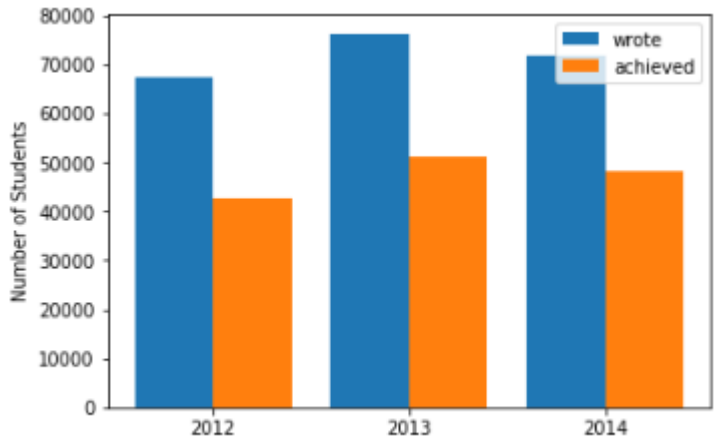
MISSING DATA

- Below we visualise missing data with the missingo library. We note we can simply remove missing records by using ‘_2012_wrote’ column, we will then be left with missing data only in the ‘quintile’ column

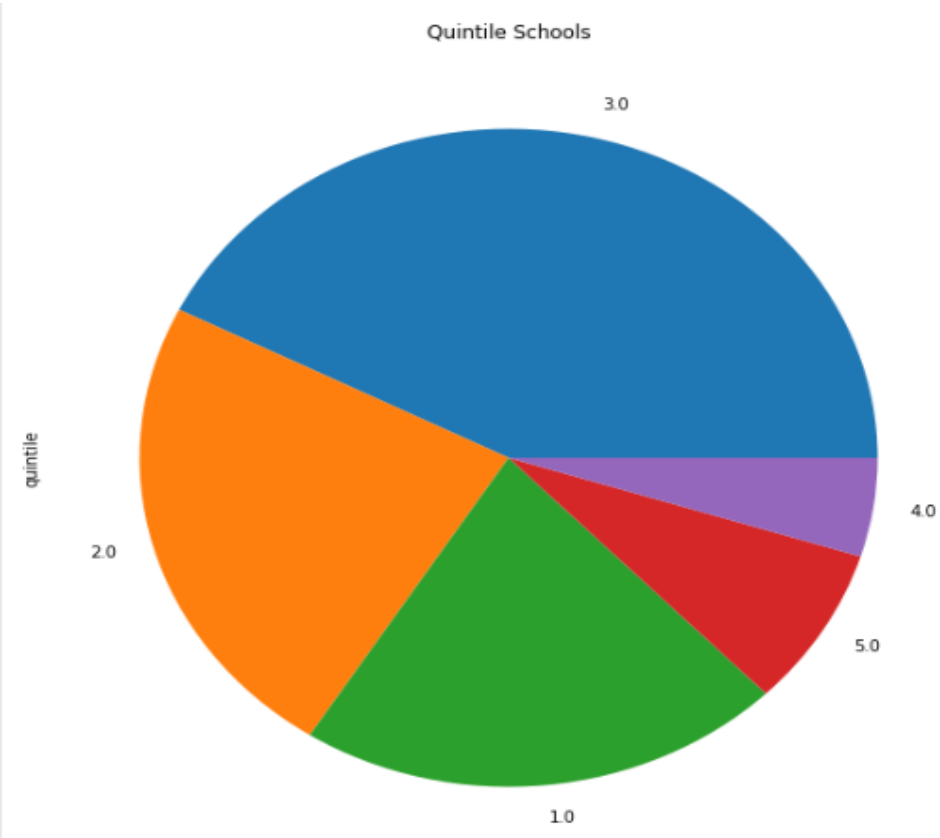


DATA STORIES AND VISUALIZATIONS

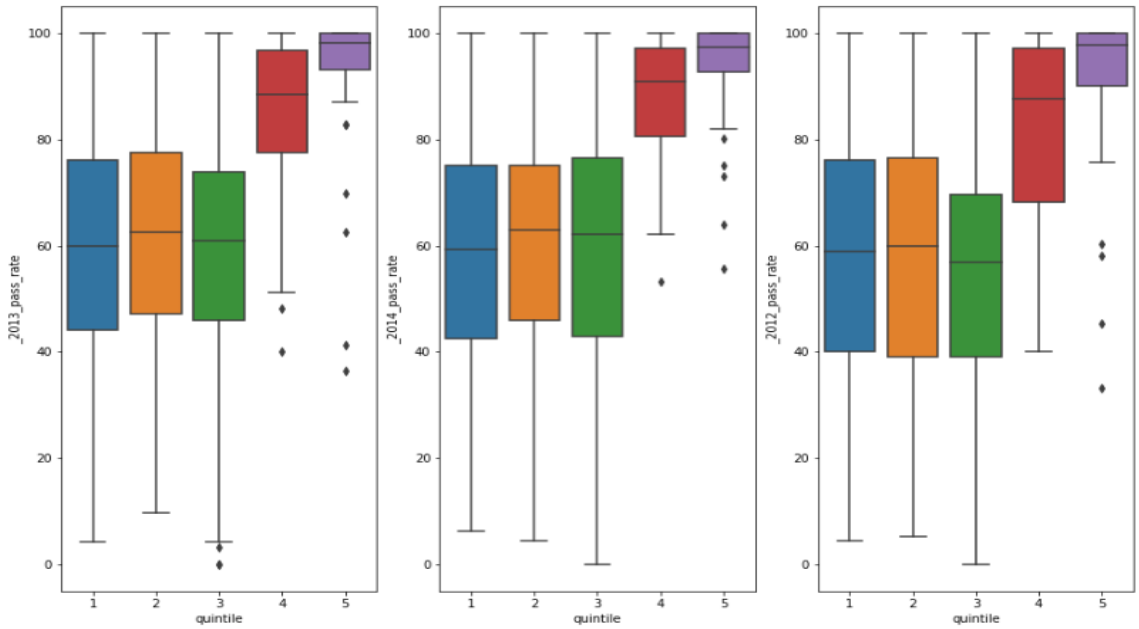
- The number of students who wrote decreased in the year 2014 compared to the previous year. We also note the ratio of those who achieved to those who wrote has been relatively constant which is alarming



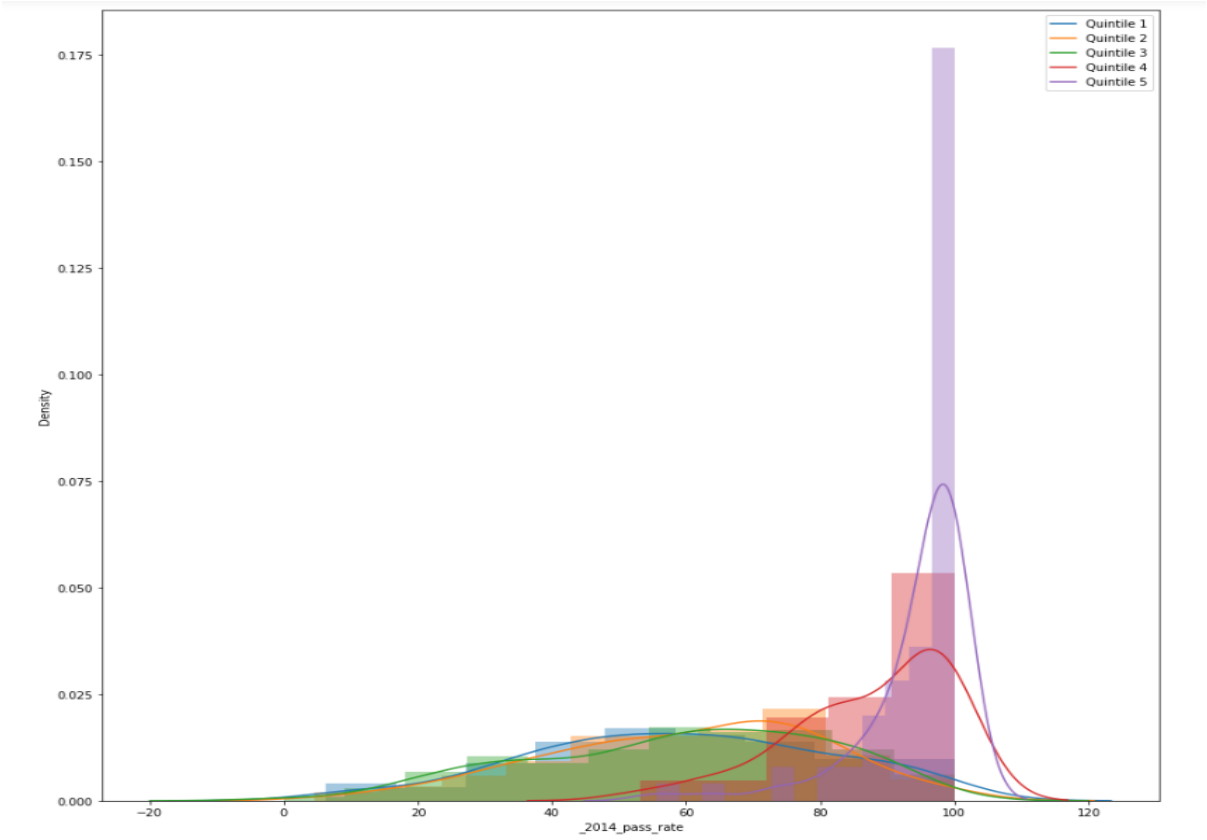
- We note there are more schools from previously disadvantaged back grounds



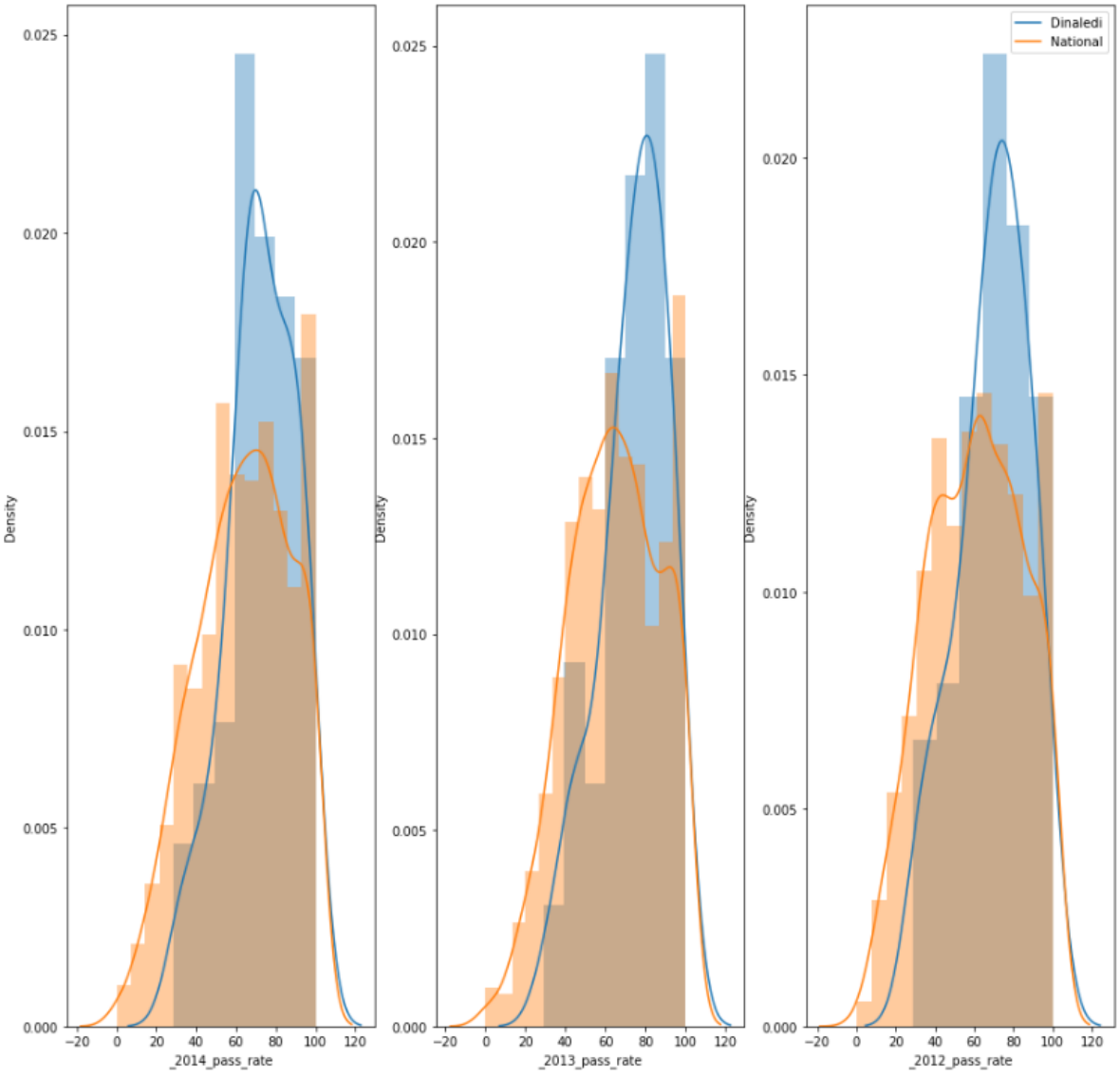
- We note from the above three boxplots for all three years results have not changes by any significant amount for all quintile classifications.



- We can see from the distribution that quintile 4 and 5 schools are concentrated on 90-100% pass rate for the year 2014 while the lower quintile schools have bell distribution peaking at 60% rate



- We see the Dinaledi programme is effective as dinaledi schools are consistently exceeding the national pass rate average for all three years



THIS REPORT WAS WRITTEN BY : SIVIWE TSHANANDA