Case study - Kenya Schools

The dataset contains data about 111 schools in 7 provinces across 31 regions in Kenya. The data is for approximately 13,000 pupils from grades 1-5 from the end of an undisclosed school term in the past five years. Each school year has three terms, and they consist of ~3-3.5 months each.

The data files

- *"Lesson completion"*: file provided at the teacher level. This means there is a unique row for each teacher. The file contains the grade that each teacher teaches and the average lesson completion rate over the term of interest.
- *"Pupil attendance":* file provided at the pupil level (that means there is a unique row for each pupil). This file includes the unique school ID, the unique pupil ID, the pupil's grade, the attendance records, and the present records.
    - The attendance records mean the total number of times that a pupil's teacher took attendance.
    - The present records mean the total number of times a pupil was present out of the attendance records.
- *"Pupil scores":* file provided at the pupil subject level (that means that there is more than one row per pupil). This file includes the unique school ID, unique pupil ID, the pupil's grade, the subject for this assessment, and the score obtained.
- *"School information"*: file provided at the school level. It includes the region and province of each school, the unique school ID, and the "treatment status" (yes/no) for a given tutoring program.

To tackle this dataset, I took these steps:

- Cleaning the data
- Calculating KPIs
- Making descriptives – graphs and charts
- Writing my suggestions in a memo-like manner.

Calculating KPIs

One of the identified KPIs was "Percent Pupils Present" which means "The percentage of pupils who were present, out of all pupils - across all days in the term to date ". In other words, the percentage of pupils who were present (for each pupil in the "Pupil attendance" file, this is displayed in the "present_records" variable), out of pupils who had attendance records (the "attendance_records" variable in the same file). I calculated this KPI in two different ways. First, calculate this KPI for all pupils at once. Then, the percentage for each school, and an average at the school level. I also evaluated when one approach would be more appropriate than the other.

This was my answer to that - The interpretation changes because the second focuses more on a niche than the general Kenya regions. The first approach would be more appropriate at a government level to determine if pupils of a particular age attend the schools while the second would be more appropriate for the school stakeholders to determine how many students attend on average.

Descriptives

I created charts to answer the following questions-

- The average fluency scores for each of the five grades.
- The regions which have the lowest and highest average fluency scores across all grades.
- A bar chart with the share of pupils scoring under 10 for reading and pupils above this threshold for each grade.
- The

- school with the highest share of pupils scoring under the above threshold in grade 3.

The Memo

Dear School(s) Board of directors,

MEMO: PUPIL SCORES IN KENYA

Following our curated Key Performance Indices, the data analyst team found some interesting insights into the data we collated.

- Pupil attendance on a general level was on average, 75.93%. However, digging into each school, we found the attendance rate to be as low as 58.37%. It is recommended that we look for factors affecting attendance and create action on how to solve the problems.
- The average reading fluency scores grade-wise increased in each higher grade as expected.
- The highest threshold of low reading scores was highest in Grade 1.
- Unfortunately, Bomet has the lowest reading fluency score across all grades.
- School ID 223941 has the highest number of pupils with low reading fluency scores in Grade 3.

I believe with these insights, the planning team would create steps to increase reading fluency across all the schools covered, focusing on the schools and regions with the lowest scores.

Thank you.