STUDENT ATTENDANCE DATA ANALYSIS

# 1. Business Understanding

## 1.1 Business Overview

Increasing the percentage of school attendance is a critical factor in determining a population’s ability to pull themselves out of poverty. Low school attendance can make it difficult for countries to spread important health information, teach real-world skills, and ultimately grow their economies.

High unemployment rates have been one of the major crises affecting the population in Kenya overtime and according to sources from various case studies.This can be attributed to low literacy levels.

In accordance with a case study done by the Organisation for European Economic Co-operation its been proved that in general, people with higher levels of education have higher employment prospects. A major difference is particularly marked between those who have attained upper secondary education and those who have not.

## 1.2 Business Objective

### Main Objective

Our main objective is to determine the province that is most affected with low school attendance.

### Other Objectives

1. Determine low attendance per district and the reason behind it
2. Identify the most affected county by low school attendance
3. Identify the most affected gender with low school attendance, the reason behind that and how to resolve that.
4. Determine the transition rate between different levels of education per province.

## 1.3 Business Success Criteria

To come up with a list of the provinces with the lowest school attendance and identify the reasons behind that which will enable us to find ways of resolving the situation.

## 1.4 Assessing the Situation

### Resource Inventory:

* Dataset:

School Attendance Data[[Link](https://www.opendata.go.ke/datasets/bbcd64cbce5347189ead08c23d6a0d38)]

* Softwares:

Trello [[Link](https://trello.com/aristocrats2/home)]

Google Collaboratory [[Link](https://colab.research.google.com/drive/1dTrd088avd6yYyjI6gkVcddPTTd60opk#scrollTo=y2BE4OP4j_MY)]

### Assumptions:

* Pre-primary students are from age; 3 - 6 years
* Primary school students age; 6 - 14 years
* Secondary school students; 14 - 18 years
* University/Tertiary; 18 years and above.
* Data that is over 100% could be**:**

1. Students unaccounted for during census
2. Students in the school are above the required maximum age
3. Students from other districts

### Constraints:

Our data was last updated 4 years ago hence it might have changed over time.

## 1.5 Data Mining Goals

Our data mining goals are as follows:

* Formulate the province with the lowest and highest school attendance
* Find out the district with the lowest school attendance
* Find what gender has the highest number of school attendance and group by provinces
* Identify the highest school dropout level

## 1.6 Data Mining Success Criteria

Our success will be measured by identifying the province with the lowest school attendance

# 

# 

# 

# 

# 2. Data Understanding

### 2.1 Data Understanding Overview

For this project we will use a dataset that we were able to discover through our own research.

Student Attendance Data by District dataset -This dataset provides oversight to the total student population at different school levels comparing them to the total number of students counted during the census.

### 2.2 Data Description

* Student Attendance Data by District dataset - This dataset contains 31 columns consisting of the name of the district, county and province, the percentage population of students in school, the population by age, the number of students who have over three years in the different school levels i.e. Pre-Primary, Primary, Secondary, University, Tertiary and Madrassa and the total population of students in the school levels

### 2.3 Verifying Data Quality

The dataset had a few missing values but we were able to drop them

# 

# 

# 3.Data Preparation

The following steps were followed in preparing of our data:

## 3.1 Loading Data

Loaded the dataset from the CSV file and loaded it to a Python Collab Notebook.

## 3.2 Data Cleaning

### Data Validity

Dropped non-relevant columns and converted some data types.

### Data Completeness

We determined the records with Null values and found out that they had no real impact to our dataset. We then dropped all null values.

### Data Accuracy

We calculated and confirmed that the total school going population was accurate.

We calculated and confirmed that the total student population conducted during the census was accurate.

### Data Uniformity

Repositioning and renaming some specific columns to aid in readability of the dataset.

## 3.3 Investigating Outliers

Lastly, we were able to identify outliers within our dataset.

We took the following steps to deal with them:

* We calculated the total number of students attending school
* We calculated the ratio of the total number of students attending school to that of total student population
* We presented our findings on a boxplot graph where we discovered Rift Valley has the highest number of outliers. This could be attributed to high number of students not attending school.

# 4. Data Analysis.

Our data analysis resulted to the following findings:

1. The county with the lowest overall school attendance was Lamu County.

This can be attributed to the rise in Insecurity due to frequent Al-Shabaab attacks.

It can also be attributed to the low population hence, likely to have fewer students when compared to other counties

1. The county with the least female attendance was Kitui County

This can be attributed to prevalence in early marriages and pregnancies in the region

1. The district with the lowest school attendance was Chalbi District in Marsabit County

This can be attributed to constant migration as most of the communities are nomadic. It may also be due to harsh environmental conditions which could make learning difficult.

1. The highest number of students attending school in all provinces are male.

This can be attributed to cultural practices and beliefs whereby some communities believe that the girl child should not go to school, instead they should stay home and learn how to raise a family.

1. The North Eastern province had the lowest number of students attending school.

(A Bar graph and pie chart visualisation showed this to be true)

On further analysis, Wajir, Garissa and Mandera County have mainly majored in religious education i.e. Madrassathus, low school attendance outcomes.

1. Majority of school going students are from the Primary school level. This can be attributed to the prevalence of free primary education
2. Generally, in all provinces, the majority of the students are not able to attend school past the Secondary school level. This can be attributed to lack of enough funding because university education is costly.

Our conclusions to our analysis were obtained from case studies which can be accessed from the following links:

[Lamu Case Study](https://www.saferworld.org.uk/long-reads/inside-kenyaas-war-on-terror-the-case-of-lamu)

[Lack of Education](https://helpfulprofessor.com/lack-of-education/)

The above analysis was done using a Python Notebook. The complete analysis can be found in the following link: [Python\_Notebook\_Analysis](https://colab.research.google.com/drive/183Zct28gHDJPZe0_imyGEFCo-jjQ-vWv?usp=sharing)

# 5. Recommendations.

# Based on our analysis we recommend the following:

* Offering of scholarships to students so as to cater for them in their university education
* Educating the society on importance of educating the girl child especially in North Eastern province
* Engage communities to develop and deliver positive and culturally relevant messages about the importance of education
* On the issue of insecurity, the government should come up with strategies to manage rising border conflicts.

# 

# 

# 

# 

# 

# 

# 

# 

## 