MAKERERE UNIVERSITY

BSE2301 SOFTWARE ENGINEERING MINI PROJECT 2

Final Report

for

Recess Project

Authors:

MBOIZI ERIC	18/U/21103/PS
MAHMOOD ALI	18/U/21125/PS
TAMALE TALEMWA COLLINS	18/U/21135/PS
DOKI GOLDER SYLIVIA	18/U/21090/PS

Table of Contents

List of Figures

1.0 Introduction to the Dataset

Annually, senior four candidates in secondary schools throughout Uganda sit for the Uganda Certificate of Education (UCE) examinations during their fourth year in the lower secondary. The examinations are done over a period of at least two months and are attempted by a diversity of students. Its aim is to test students' understanding of what was taught in the lower secondary and also double as their ticket into the next stage of education; A Level or upper Secondary.

A summary of statistics of students' performance provided by the examination board of the UCE exams in the years 2011-2016, can provide a multitude of interesting insights into this yearly exam.

The statistics provided in this dataset for use in our analysis contain:

- Year in which the exam was done
- Name of the school
- District to which the school belongs
- Total Candidates who sat for exam in that school
- The number of candidates in the school that scored results that fall in each of the different divisions obtainable in the exam
- Percentage of candidates in the school who obtained results that fall in each of those divisions
- The female candidates from the school who sat for the exam
- The number of female candidates in the school that scored results that fall in each of the different divisions obtainable in the exam
- Percentage of female candidates in the school who obtained results that fall in each of those divisions
- The male candidates from the school who sat for the exam
- The number of male candidates in the school that scored results that fall in each of the different divisions obtainable in the exam
- Percentage of male candidates in the school who obtained results that fall in each of those divisions

2.0 The key objectives in this data analysis

The key objectives for analysing the data in this dataset are to;

1. Find out the number of schools in every district.
2. Determine the best performing schools each year.
3. Find out the number of students sitting in every year.
4. Get the overall best performing schools over all the given years(2011-2016).
5. Compare the performance between girls and boys annually.
6. Identify the best performing districts over all the given years(2011-2016).
7. Study the general performance of students as years progress.
8. Identify the regions in which the best performing schools are located.
9. Find out the schools that have managed to maintain a good consistent performance.
10. Look at the general performance of students within Kampala and Wakiso which usually have the majority of best performing schools.
11. Compare the performance between girls and boys in Kampala and Wakiso annually.
12. Study the performance of students outside of Kampala and Wakiso through the years

3.0 Features/columns analyzed and why they were chosen

- *'Year'* column: To get analysis results based on the years in the dataset.
- *'School'* column: To get analysis results based on schools in the dataset.
- Percentage columns of the different divisions: To be able to look at the performance of schools and students as required by the objectives.
- *'District'* column: To get analysis results based on districts in the dataset.
- Total columns of the different divisions: To be able to analyse the performance in the various divisions.

4.0 Processes and techniques used and why those were chosen

4.1 Data organizing/processing

The data from the csv file was read into a structure known as a data frame. A data frame stores the data in rows and columns for easier and further analysis.

4.2 Data cleaning

After organising and processing the data, it was then cleaned. This was due to some errors within the data acquired from the csv file. Removal of the errors resulted in getting more accurate outcomes smoothening our analysis.

Cleaning was mainly done to the column names, NaN values and misplaced values for the various schools.

```
df.columns
```

```
Index(['YEAR', 'DISTRICT ', 'SCHOOL', 'TOTAL CANDIDATES', 'TOTAL DIV 1',
       '% DIV 1', 'TOTAL DIV 2', '% DIV 2', 'TOTAL DIV 3', '% DIV 3',
       'TOTAL DIV 4', '% DIV 4', 'TOTAL DIV 7', '% DIV 7', 'TOTAL DIV 9',
       '% DIV 9', 'TOTAL X', '% X', 'FEMALE CANDIDATES', 'FEMALE TOTAL DIV1 ',
       'FEMALE % DIV1 ', 'FEMALE TOTAL DIV2 ', 'FEMALE % DIV2',
       'FEMALE TOTAL DIV3', 'FEMALE % DIV3', 'FEMALE TOTAL DIV4',
       'FEMALE % DIV4', 'FEMALE TOTAL DIV7', 'FEMALE % DIV7',
       'FEMALE TOTAL DIV9', 'FEMALE % DIV9', 'FEMALE TOTAL X', 'FEMALE % X ',
       'MALE CANDIDATES', 'MALE TOTAL DIV1 ', 'MALE % DIV1 ',
       'MALE TOTAL DIV2 ', 'MALE % DIV2', 'MALE TOTAL DIV3', 'MALE % DIV3',
       'MALE TOTAL DIV4', 'MALE % DIV4 ', 'MALE TOTAL DIV7 ', 'MALE % DIV7',
       'MALE TOTAL DIV9 ', 'MALE % DIV9', 'MALE TOTAL X ', 'MALE % X'],
      dtype='object')
```

Figure 1.0.1: Column headings in capital letters and with some white spaces before cleaning.

```
df.columns
```

```
Index(['year', 'district', 'school', 'total candidates', 'total div 1',
       '% div 1', 'total div 2', '% div 2', 'total div 3', '% div 3',
       'total div 4', '% div 4', 'total div 7', '% div 7', 'total div 9',
       '% div 9', 'total x', '% x', 'female candidates', 'female total div1',
       'female % div1', 'female total div2', 'female % div2',
       'female total div3', 'female % div3', 'female total div4',
       'female % div4', 'female total div7', 'female % div7',
       'female total div9', 'female % div9', 'female total x', 'female % x',
       'male candidates', 'male total div1', 'male % div1', 'male total div2',
       'male % div2', 'male total div3', 'male % div3', 'male total div4',
       'male % div4', 'male total div7', 'male % div7', 'male total div9',
       'male % div9', 'male total x', 'male % x'],
      dtype='object')
```

Figure 1.0.2: Column headings after cleaning with lowercase letters and white spaces removed.

FEMALE % DIV9	FEMALE TOTAL X	FEMALE % X	MALE CANDIDATES	MALE TOTAL DIV1	MALE % DIV1	MALE TOTAL DIV2	MALE % DIV2	MALE TOTAL DIV3	MALE % DIV3	MALE TOTAL DIV4	MALE % DIV4	MALE TOTAL DIV7	MALE % DIV7	MALE TOTAL DIV9	MALE % DIV9	MALE TOTAL X	MALE % X
0.0	NaN	0.0	0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
0.0	NaN	0.0	151	NaN	99.3	150.0	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.7
0.0	NaN	0.0	0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
0.0	NaN	0.0	109	NaN	100.0	109.0	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
0.0	NaN	0.0	73	NaN	97.3	71.0	2.7	2.0	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 2.0.1: Male's percentage results of the different divisions in wrong columns before cleaning.

female total div7	female % div7	female total div9	female % div9	female total x	female % x	male candidates	male total div1	male % div1	male total div2	male % div2	male total div3	male % div3	male total div4	male % div4	male total div7	male % div7	male total div9	male % div9	male total x	male % x
NaN	0.0	NaN	0.0	NaN	0.0	0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
NaN	0.0	NaN	0.0	NaN	0.0	151	NaN	99.3	150.0	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.7
NaN	0.0	NaN	0.0	NaN	0.0	0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
NaN	0.0	NaN	0.0	NaN	0.0	109	NaN	100.0	109.0	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
NaN	0.0	NaN	0.0	NaN	0.0	73	NaN	97.3	71.0	2.7	2.0	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 2.0.2: Male's percentage results of the different divisions in correct columns after cleaning.

female total div7	female % div7	female total div9	female % div9	female total x	female % x	male candidates	male total div1	male % div1	male total div2	male % div2	male total div3	male % div3	male total div4	male % div4	male total div7	male % div7	male total div9	male % div9	male total x	male % x
0.0	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	151	150.0	99.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.7
0.0	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	109	109.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	73	71.0	97.3	2.0	2.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 3.0: Removal of the NaN values.

	year	district	school	total candidates	total div 1	% div 1	total div 2	% div 2	total div 3	% div 3	total div 4	% div 4	total div 7	% div 7	total div 9	% div 9	total x	% x	canc
14383	2016	WAKISO	THE ACADEMY ST.LAWRENCE,BUDO P.O.BOX 29142 ...	20.0	20.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
14384	2016	WAKISO	ST.MARY'S SEC. SCHOOL,KITENDE P.O.BOX 30583...	433.0	430.0	99.3	2.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.2	
14385	2016	WAKISO	UGANDA MARTYRS SS,NAMUGONGO P.O.BOX 5369 KA...	266.0	260.0	97.7	6.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
14386	2016	MUKONO	MT.ST.MARY'S,NAMAGUNGA P.O.BOX 18 LUGAZI	160.0	154.0	96.3	6.0	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
14387	2016	WAKISO	GAYAZA HIGH SCHOOL P.O.BOX 7029 KAMPALA	195.0	187.0	95.9	8.0	4.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Figure 4.0.1: In the year 2016, St. Mary's Secondary School,Kitende is referred to as St. Mary's Sec School, Kitende (second row).

```
df.loc[14384, 'school'] = "ST.MARY'S SECONDARY SCHOOL,KITENDE"
df.loc[14384, 'school']
```

```
"ST.MARY'S SECONDARY SCHOOL,KITENDE"
```

Figure 4.0.2: The name of the school is changed to St.Mary's Secondary School,Kitende to avoid it being recognised as a separate school.

14	ST.HENRY'S COLLEGE,KITOVU	MASAKA MAIN	96.5
15	BP.CIPRIANO KIHANGIRE SS P.O.BOX 27477 KAMPALA	KAMPALA	96.4
16	NAMIREMBE HILLSIDE H/S P.O.BOX 22915 KAMPALA	WAKISO	96.3
17	ST.HENRY'S COLLEGE,KITOVU P.O.BOX 64 MASAKA	MASAKA	96.2

Figure 5.0.1: We have a duplicate of line 14 and 17, this seems unique though.

14	ST.HENRY'S COLLEGE,KITOVU	MASAKA MAIN	96.5
15	BP.CIPRIANO KIHANGIRE SS	KAMPALA	96.4
16	NAMIREMBE HILLSIDE H/S	WAKISO	96.3
19	ST.JOSEPH'S GIRLS,NSAMBYA	KAMPALA	96.2

Figure 5.0.2: After the above duplicate has been removed.

4.3 Data Visualization

With this technique, we displayed the results of our analysis in a graphical manner particularly bar charts and line graphs. We use this technique so we can gain even further insights into the data we are analysing.

5.0 Screenshots of the visualizations generated.

Comparison between girls' and boys' performance in the period of 2011-2016.

A line chart is used because we have two different categories .i.e. boys(blue) and girls(orange), for each of the years in the dataset. So the line chart easily helps us look at the variation between their performance for all the years.

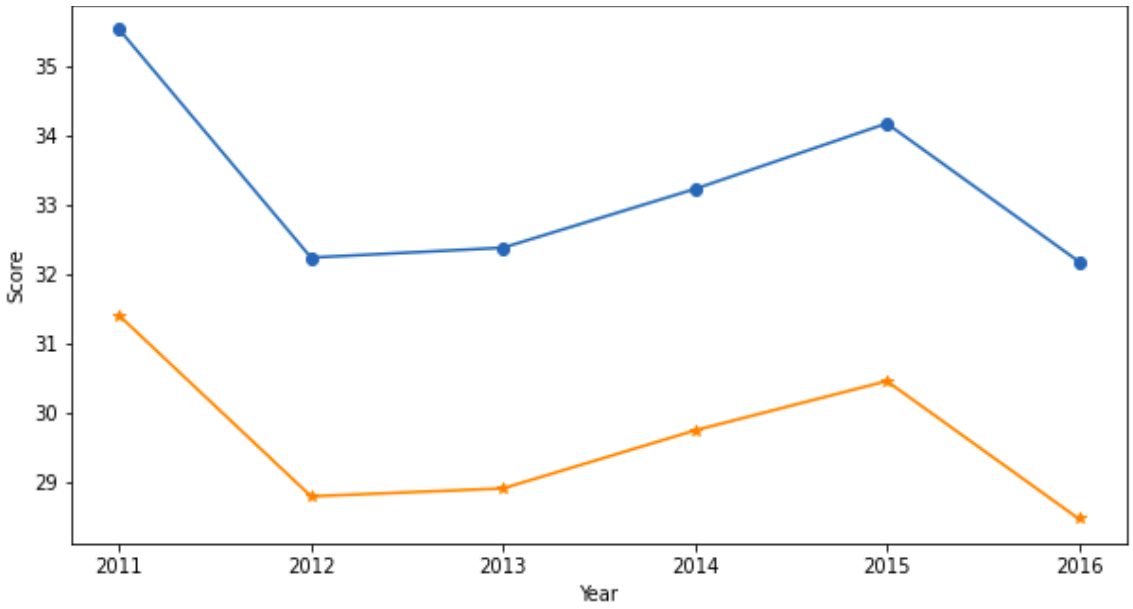

Figure 6.0: Comparison between girls' and boys' performance in the period of 2011-2016.

General performance of students throughout the years.

A bar chart is used because we have more than two categories for each of the years. The bar chart represents the intended data easier and clearer for analysis. It makes variation of the data easy.

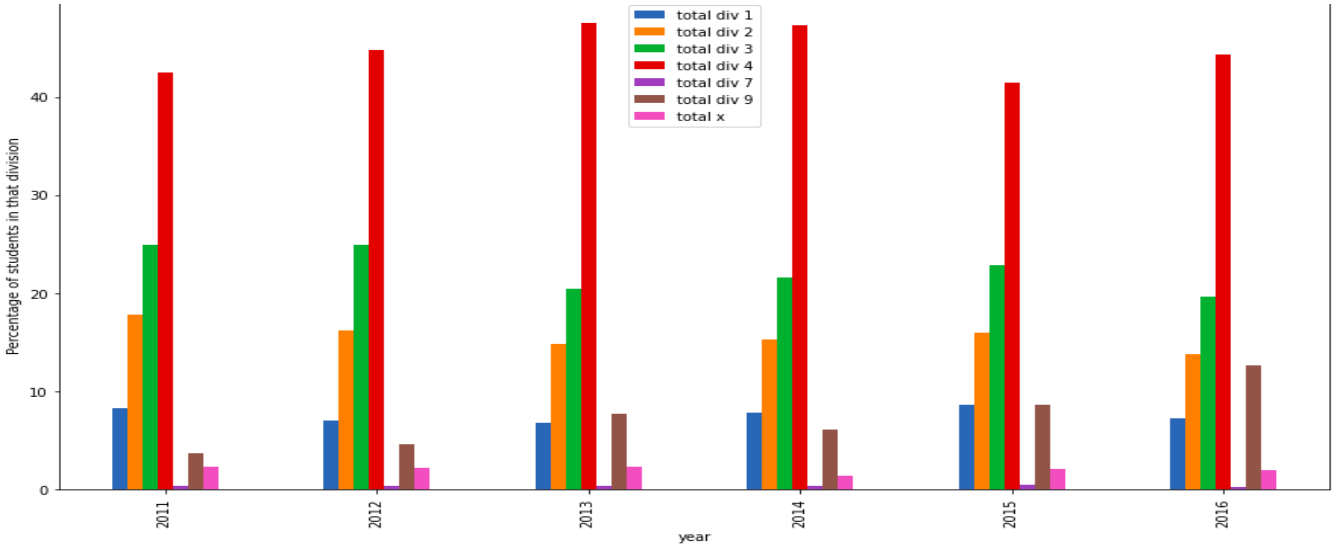

Figure 7.0: General performance of the students over the years.

Students who obtained Division One results throughout the years.

A bar chart is used because it creates a vivid variation of the years and their respective attached values.

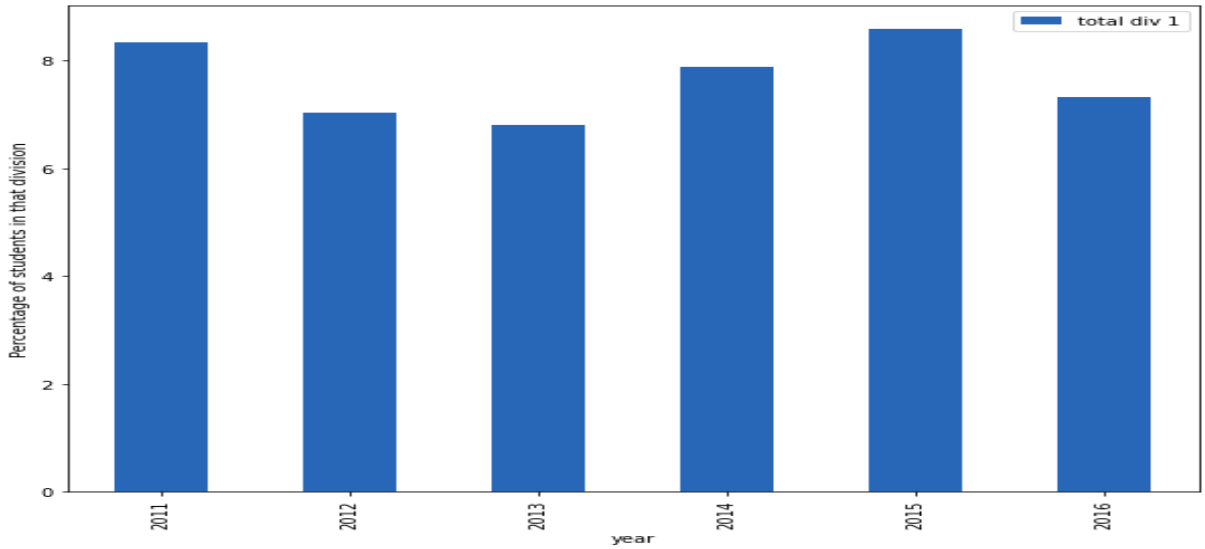

Figure 8.0: Students with Division One throughout the years.

Analysis of students performance in Kampala and Wakiso.

A bar chart is used because we have more than two categories for each of the years. The bar chart represents the intended data easier and clearer for analysis.

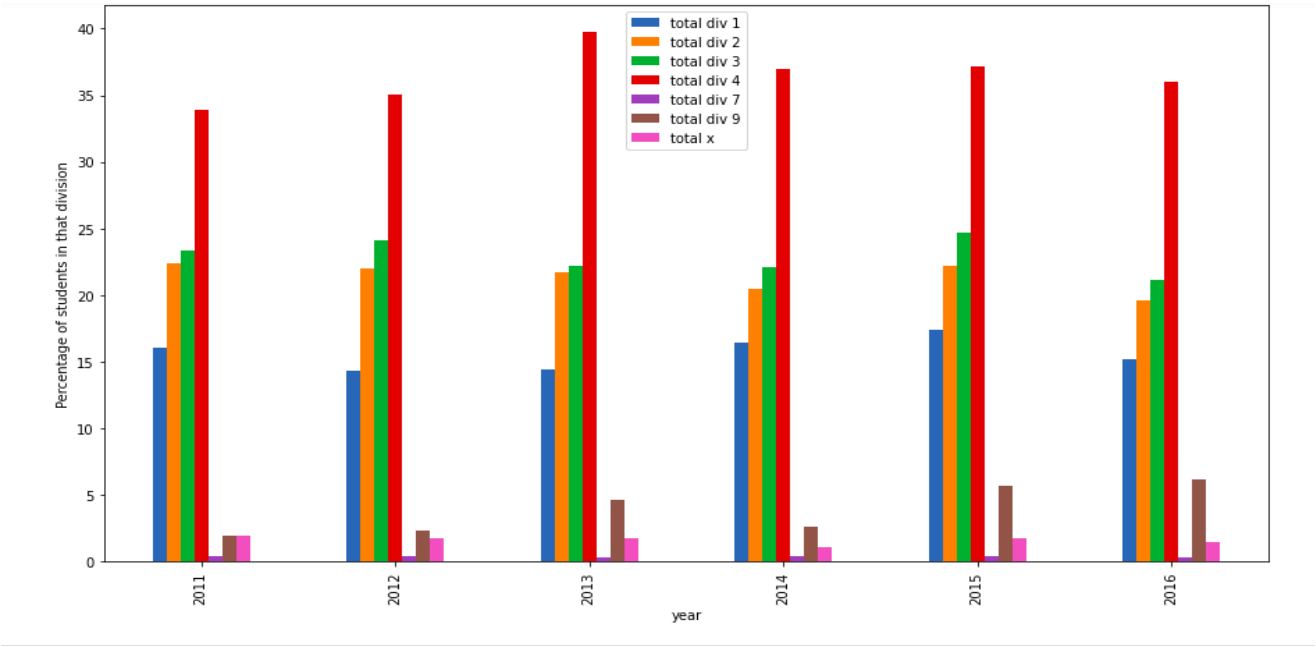

Figure 9.0: General performance of students in Kampala and Wakiso.

Analysis of general performance of students outside Kampala and Wakiso.

A bar chart is used because we have more than two categories for each of the years. The bar chart represents the intended data easier and clearer for analysis.It makes variation of the data easy.

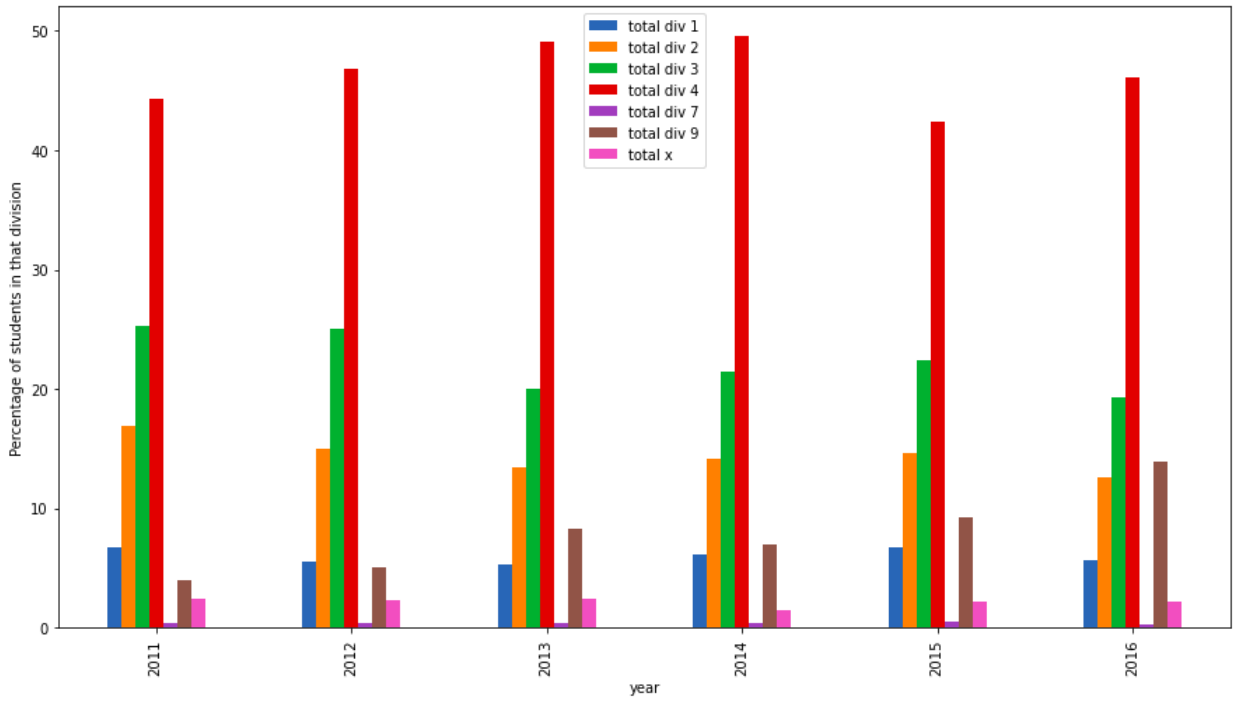

Figure 10.0: General performance of students outside Kampala and Wakiso.

Comparison of girls' and boys' performance in Kampala and Wakiso

A line chart is used because we have two different categories .i.e. boys(blue) and girls(orange), for each of the years in the dataset. The line chart helps clarify any slight variations whether drastic or small.

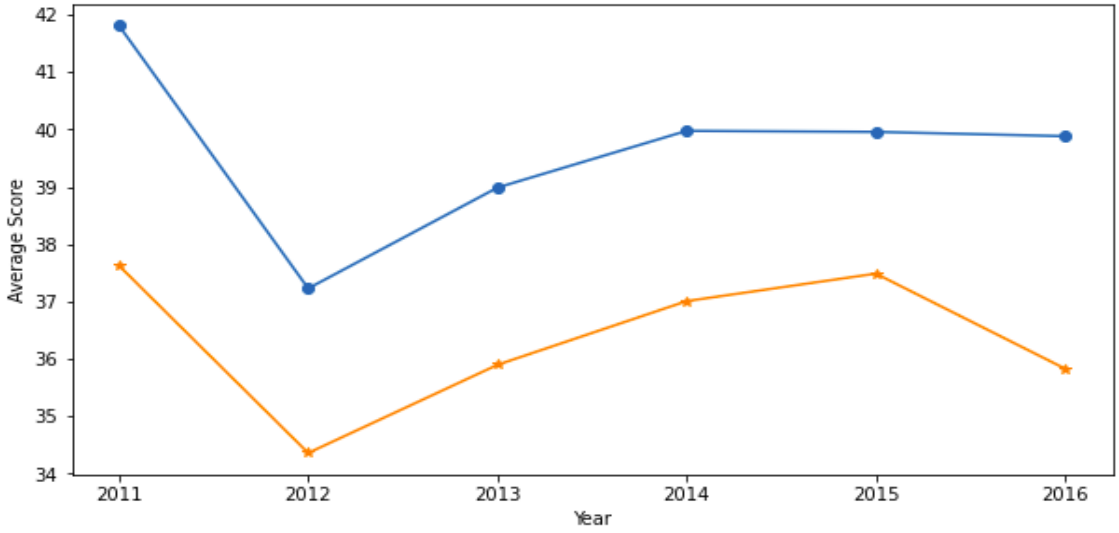

Figure 11.0: Performance comparison between girls and boys from Kampala and Wakiso.

Locations of overall best performing schools in Uganda.

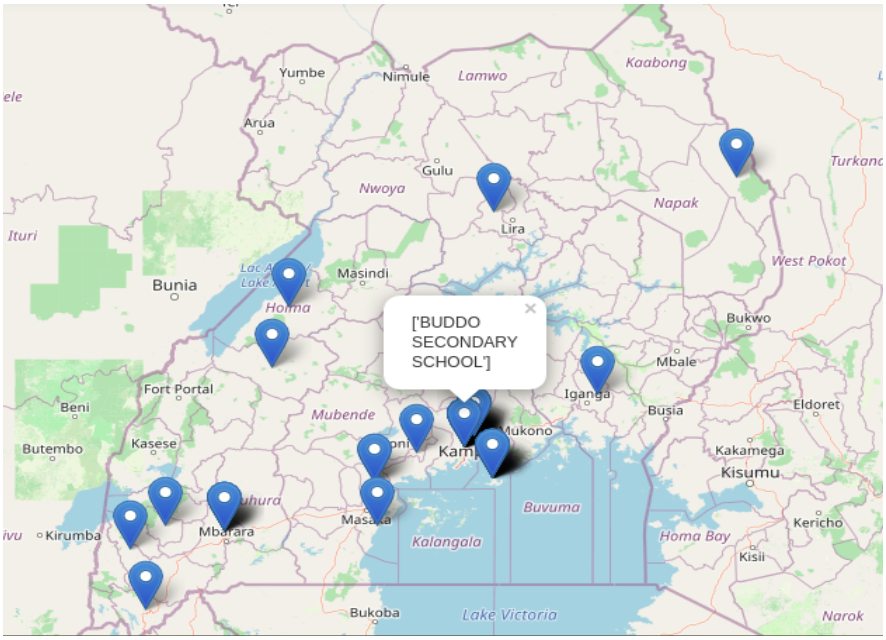

Figure 12.0: Location of overall best performing schools in Uganda.

6.0 Deductions/conclusions for our data.

From our dataset and its subsequent analysis, we were able the draw the following conclusions:

1. The total number of schools registered in Uganda for the UCE examinations as of 2016 was 3256 with Wakiso district having the highest number of 352, followed by Kampala district with 201 and then Mukono district with 126.

2. The total number of candidates that sat for the UCE exams varied with each year. No gradual increase or decrease was noted. The highest number was registered in 2016 with 321,276 students. Following 2016 was 2013 that had 293,435 students. The least number of students was 267,817 students in 2012.

3. We were able to analyse the schools that had the best overall performance over the period of 2011-2016. The top 5 schools in this category were;
    - *The Academy St Lawrence, Budo* with an average score of 100 %
    - *Mt.St.Mary's, Namagunga* with an average score of 99.4 %
    - *Uganda Martyrs SS, Namugongo* with an average score of 98.9%
    - *St.Mary's Secondary School, Kitende* with an average score of 98.9%
    - *Gayaza High School* with an average score of 98 %

4. Comparing the performance between girls and boys from 2011-2016:
    - Boys consistently performed better than the girls.
    - 2011 registered the highest performance by both girls and boys.
    - 2016 registered the lowest performance by both girls and boys.

5. From 2011 to 2016, the 10 overall best performing districts were; Moroto, Bushenyi Main, Kole, Mukono, Kalungu, Wakiso, Otuke, Kampala, Mpigi and Rubirizi respectively.

6. In each of the years, at least 40% of students sitting for the exam scored results that fell in Division Four. Division Three and Division Two followed as the ones that most students lay in. The least division obtained each year was Division Seven.

7. We were able to see that only 38 schools maintained a consistent performance and appear in top 100 schools every year. These schools are, in no particular order, namely;

    St.Charles Lwanga Seminary, Trinity College Nabbingo, St.Joseph Of Nazareth H/S, St.Paul's Seminary Kabale, Iganga Secondary School, Mt.St.Mary's Namagunga, Ntare School, Kitabi Seminary, Tororo Girls' School, Brilliant High School Kawempe, Gayaza High School, St.Mary's College Kisubi, Notre Dame Academy Buseesa, Maryhill High School, London College Of St.Lawrence, Gombe Secondary School, Seeta High School Mukono, St.Joseph's Girls Nsambya, Makerere College School, Turkish Light Academy, Seeta High School, St.Henry's College Kitovu, Bukalasa Minor Seminary, Uganda

Martyrs Ss Namugongo, Kisubi Seminary, Ndejje Secondary School, Mengo Secondary School, Merryland High School Entebbe, Nadiket Seminary Moroto, Nabisunsa Girls' School, Bweranyangi Girls' School, Mbarara High School, Immaculate Heart Girls School, King's College Budo, Jinja College, Namilyango College, Kiira College Butiki and Seeta High School.

8. We were able to deduce that of the overall best performing schools, the majority of them are found in the Central region, followed by the Western region, then Northern and lastly the Eastern region.

9. At least 33% of students within Kampala and Wakiso lie in Division 4, followed by Division 3, then Division 2 and Division 1 respectively in all the years.

10. Comparing the performance between boys and girls within Kampala and Wakiso from 2011 to 2016:
    - Boys performed better than girls each year.
    - 2011 registered the highest performance by both girls and boys.
    - 2012 registered the lowest performance by both girls and boys.

11. Outside of Kampala and Wakiso, at least 42% of students lie mostly in Division 4, followed by Division 3, then Division 2 and Division 1 respectively in each of the years.

12. Regarding the general performance of students through 2011-2016:
    - 2015 registered the highest percentage of Division Ones at 8.3% of the student population, followed by 2011 with 7% . 2013 had the least with 6.8%.
    - The percentage of Division nines is on a gradual increase. 2011 registering the least at 3.6% and 2016 the highest at 12.6%.
    - 2011 registered best overall performance.
    - 2016 registered worst overall performance.