Analysis of Khan Academy Data Set

John Chao

Table of Content

01 Introduction

Introduction

- Data is a Khan Academy real data.
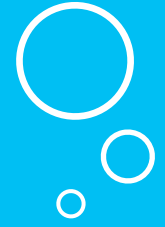
- Two kinds of information in the dataset –

  Num of Students and Learning Time.

02 Introduction to Data

Introduction to Data

District Name	School Name	Total Number of Students	Numbers of active students (per month)											
			M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
District 1	ZPP 1	133	2	0	7	44	37	6	7	0	4	2	1	1
District 1	ZPP 2	135	5	2	12	59	23	8	13	11	6	12	5	0
District 1	ZPP 3	69	0	0	4	31	32	18	4	6	3	1	0	0

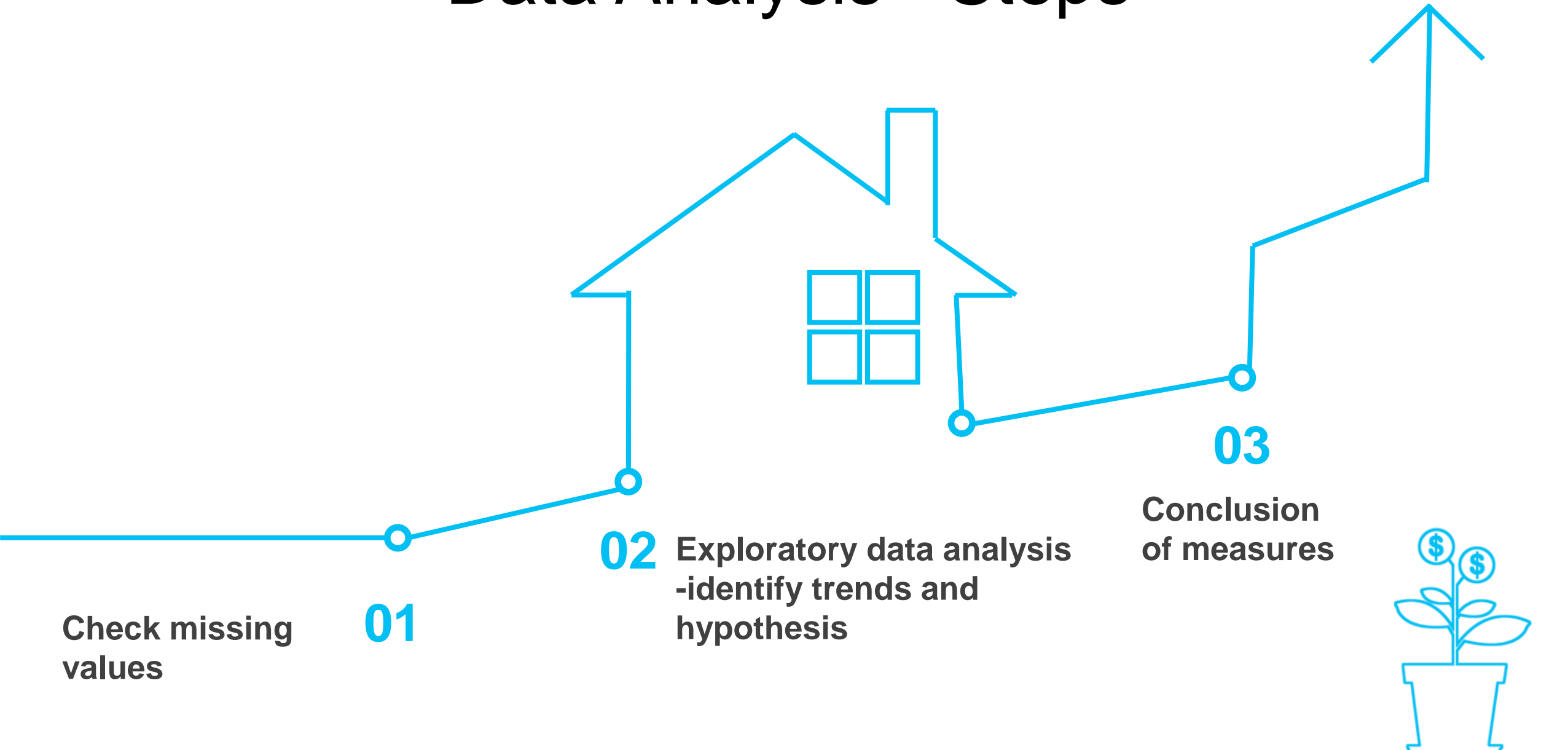| District Name	School Name	Total Number of Students	Average learning time on KA per active students (in minutes per month)											
			M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
District 1	ZPP 1	133	0	0	47.2	57.7	19.1	31.4	0	0	20.8	12.9	15	0
District 1	ZPP 2	135	3	24.8	30	24.2	18	23.4	34.3	34.5	29.5	23.8	0	0
District 1	ZPP 3	69	0	0	42.8	73.9	76.5	45.1	29.8	25.1	11	0	0	0

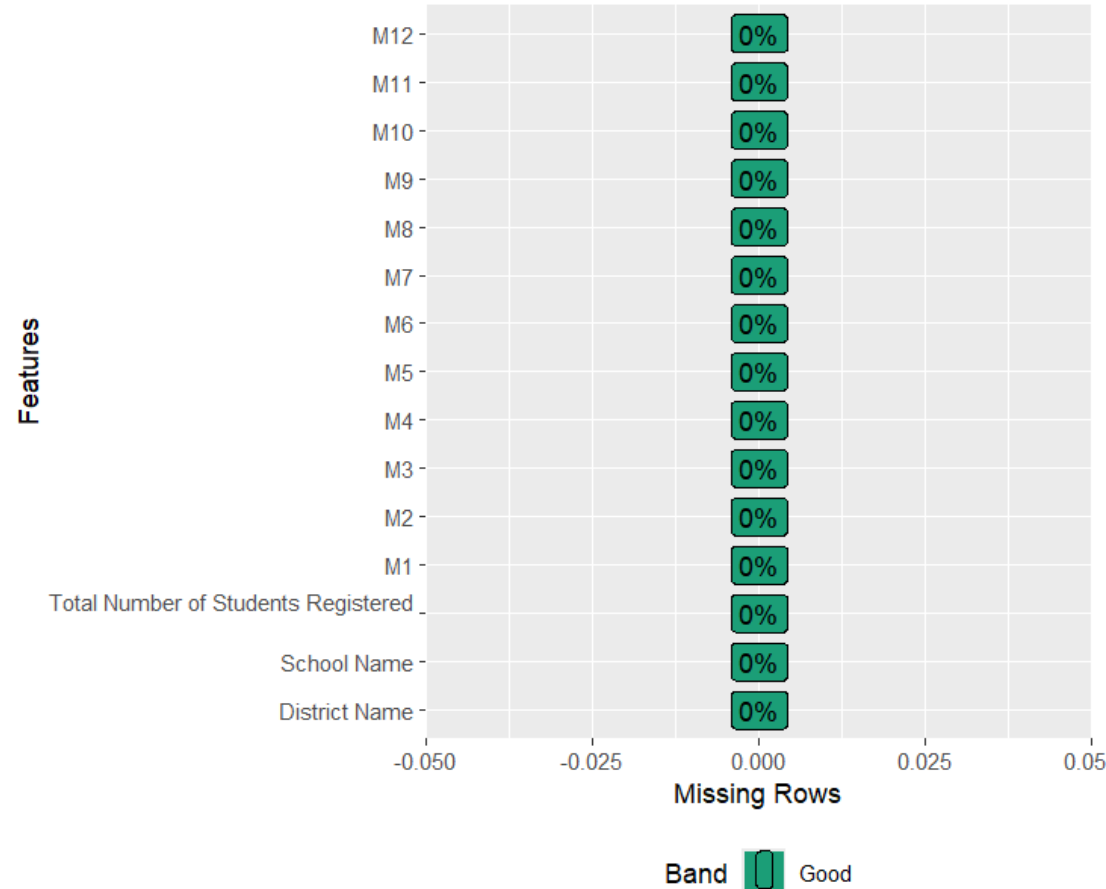| Variable	Definition	Type
District Name	Name of the district	Nominal
School Name	Name of the school	Nominal
M1 – M12	Average value or counts in January - December	Numeric

03 Data Analysis – Trends and Hypothesis

Data Analysis - Steps

01

Check missing values

02 **Exploratory data analysis -identify trends and hypothesis**

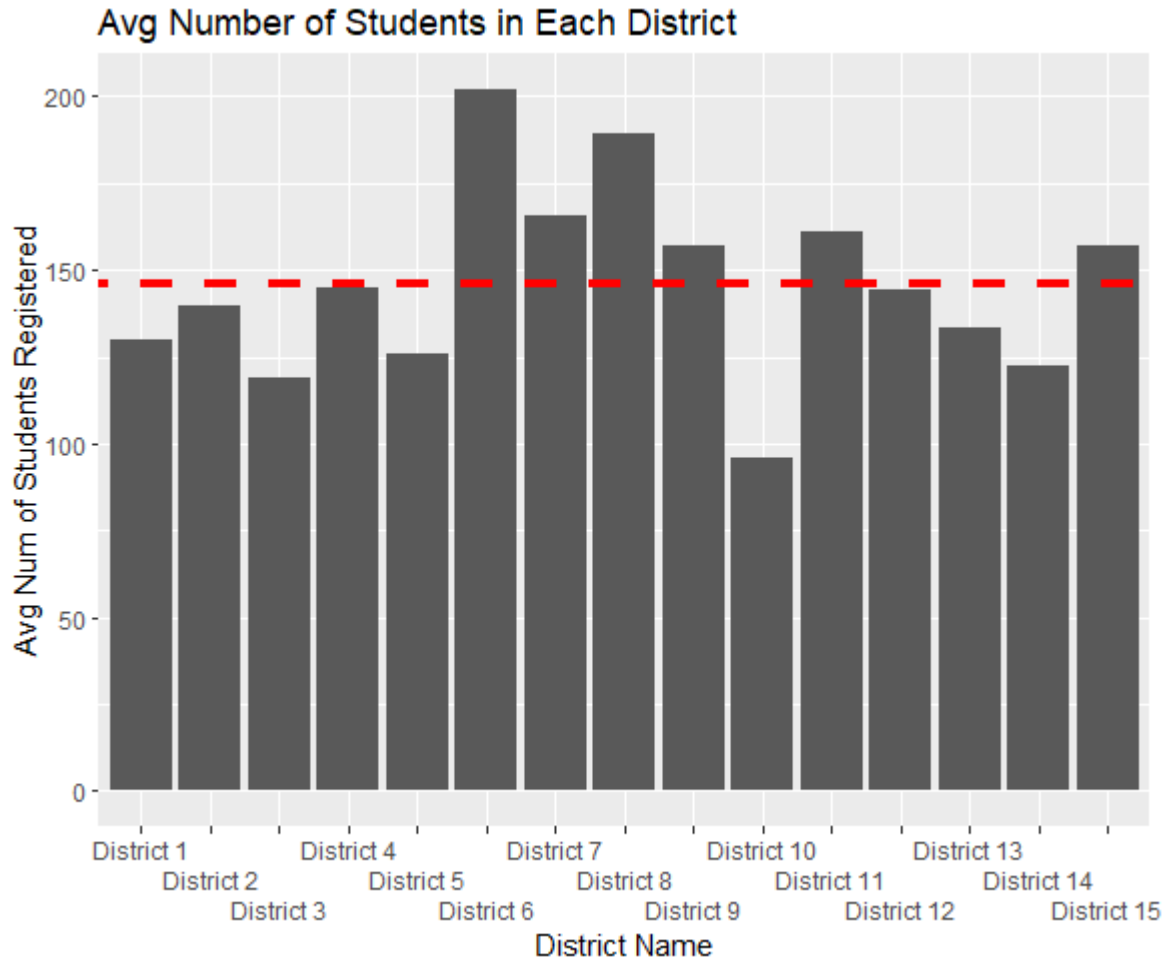**03**

Conclusion of measures

Data Analysis – Missing Values Detection

- R package - DataExplorer
- No missing values in both datasets

Data Analysis – Exploratory Analysis of Trends

Avg Number of Students in Each District

- Trend - The average number of students in each district is different. Districts 6-8 have more enrolled students compared to other districts.

- Hypothesis – The area of Districts 6-8 may be larger than that of the other districts, or more schools, so there are more students.
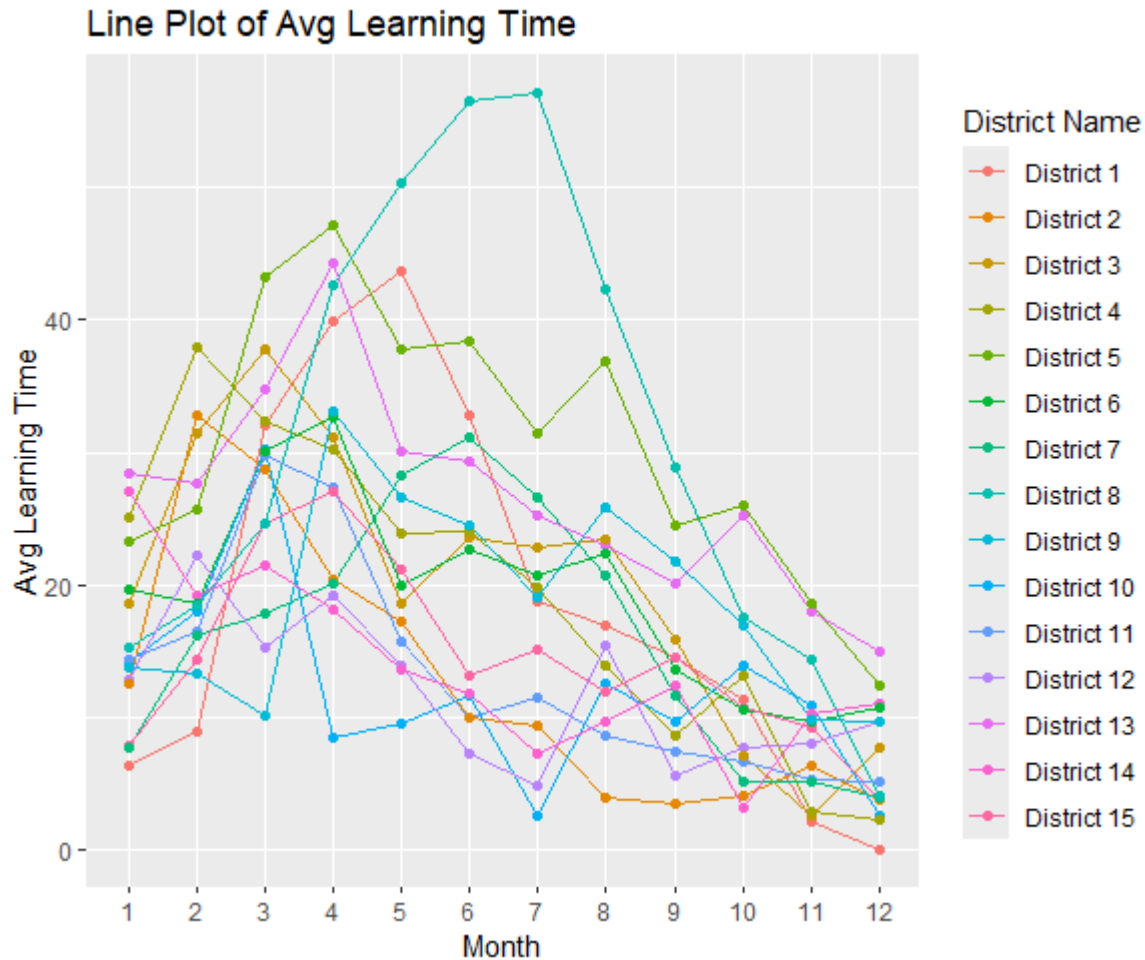
Data Analysis – Exploratory Analysis of Trends

Line Plot of Avg Num of Active Students
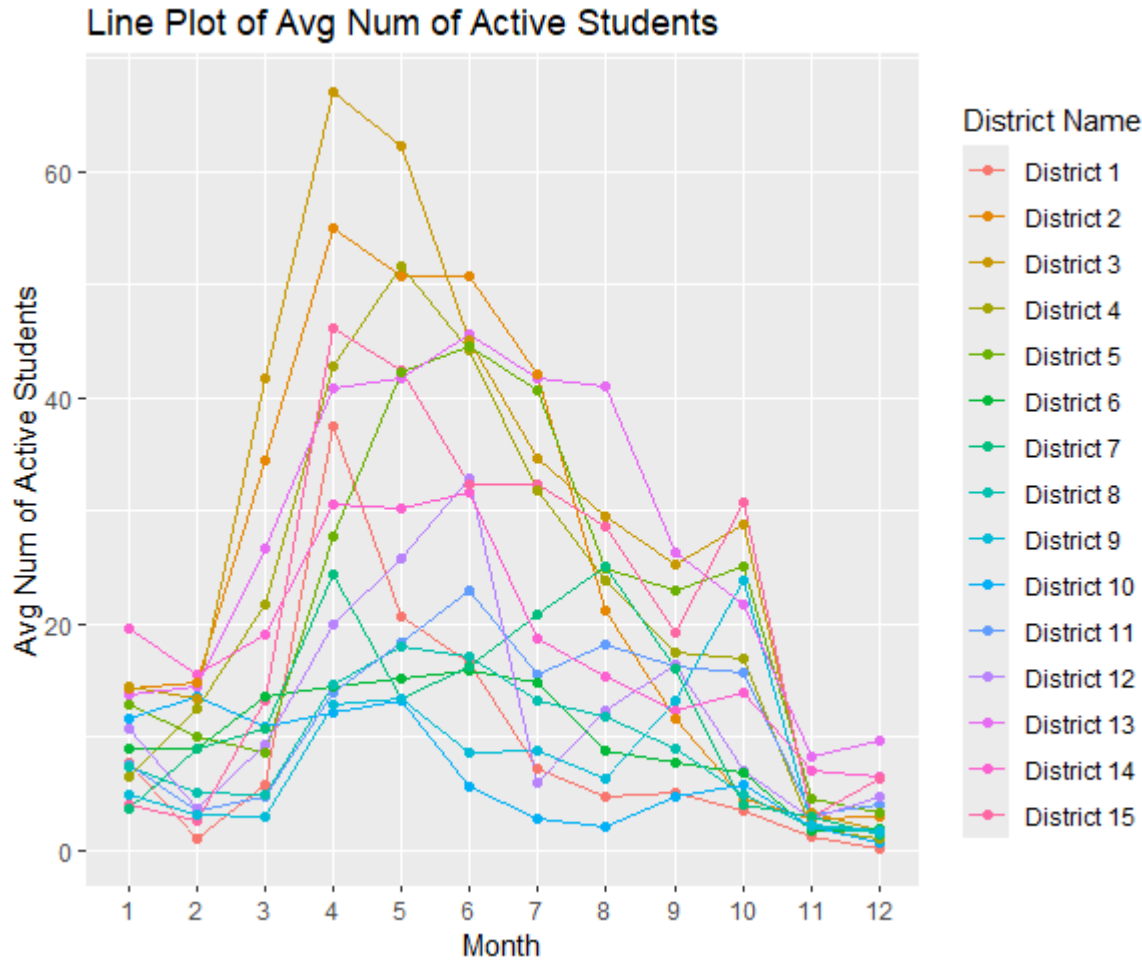
- Trend - There's a trend in the average number of students in a year. Months 4-6 apparently have more students with new supplements. However, after that, there is a decreasing trend in the average number. The average number of active students is even lower than before the supplements were used.

- Hypothesis – Students found out there's no improvement in their academic performance, or most of the questions in the test come from Khan Academy, so there's no need for so many students to do the practice. They can just share the answers to each other.

Data Analysis – Exploratory Analysis of Trends
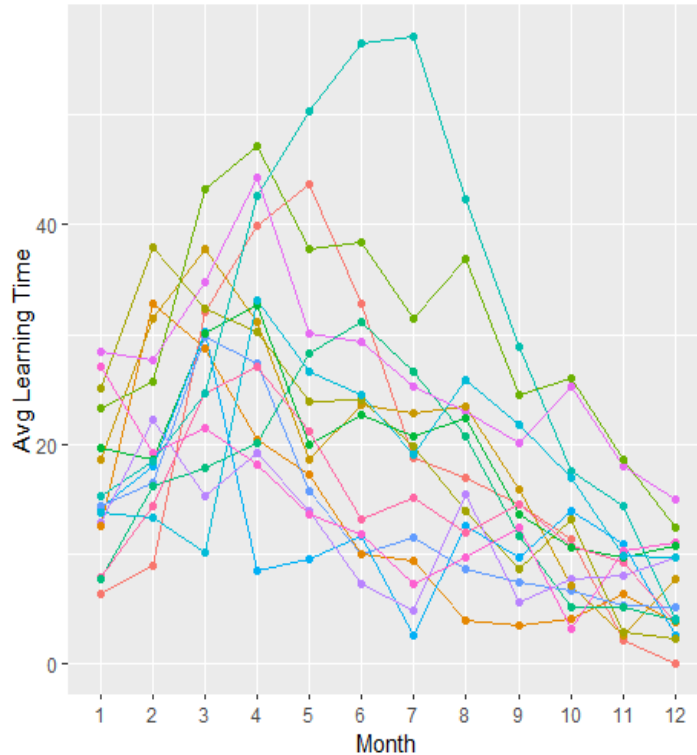

Line Plot of Avg Learning Time

- Trend - In general, the average learning time shows roughly the same trend.

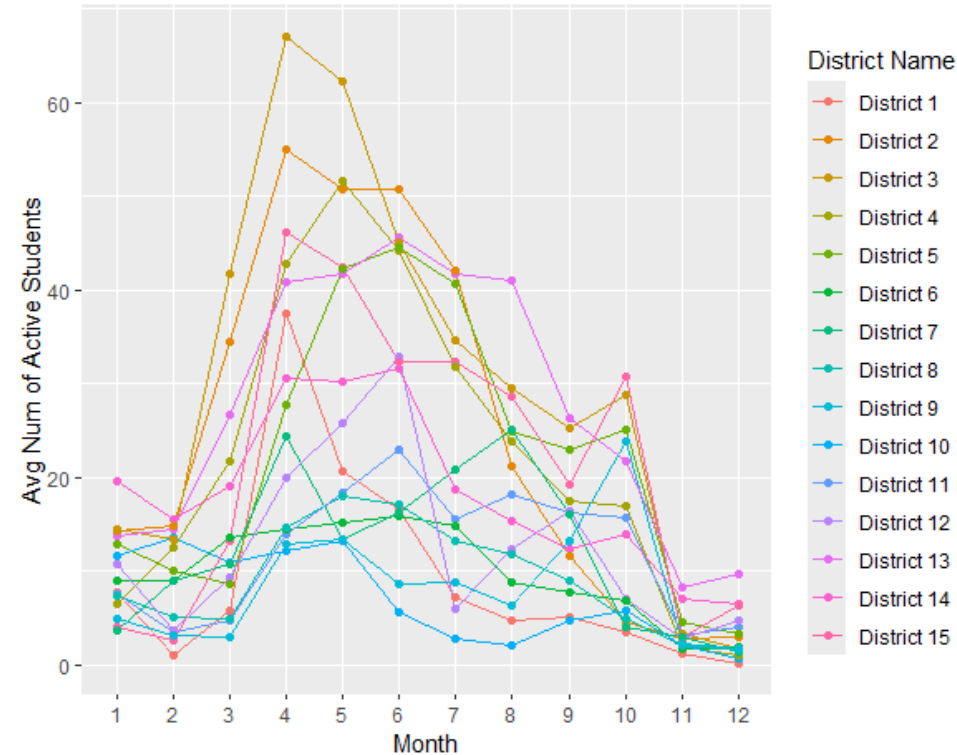- Hypothesis – There are fewer students, so it makes sense to have fewer average learning time.

Data Analysis – Exploratory Analysis of Trends

Line Plot of Avg Learning Time

Line Plot of Avg Num of Active Students

District Name
- District 1
- District 2
- District 3
- District 4
- District 5
- District 6
- District 7
- District 8
- District 9
- District 10
- District 11
- District 12
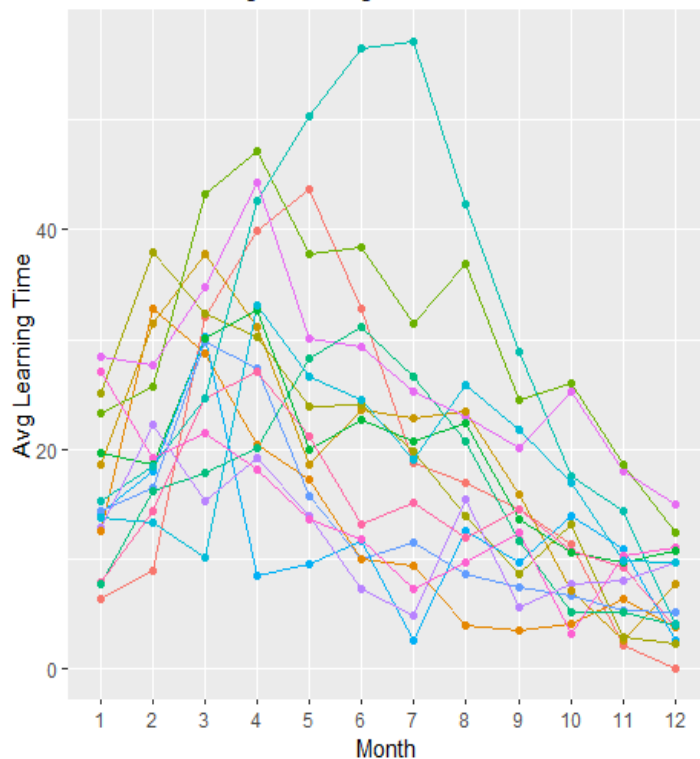- District 13
- District 14
- District 15

Trend - The value of the average learning time of District 8 is obviously more than the other districts from May to August. However, during that time, it is district 3 had the most average number of active students.
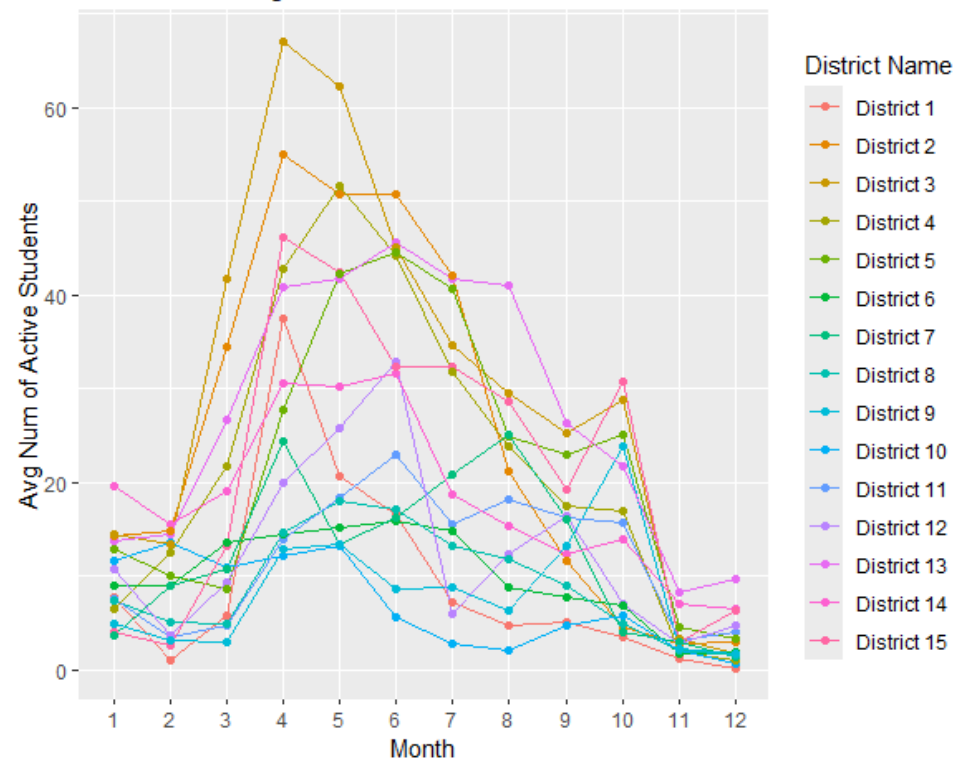
Data Analysis – Exploratory Analysis of Trends

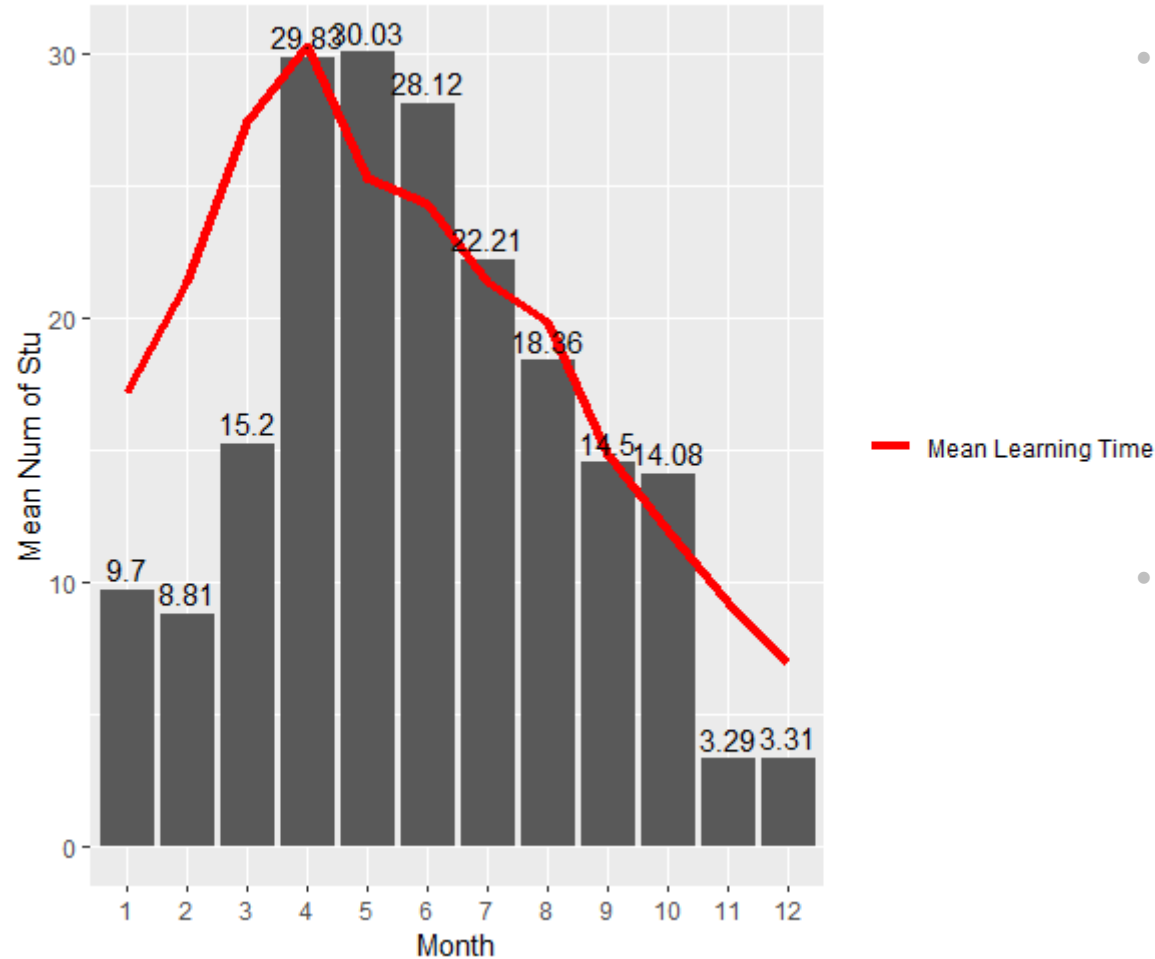
Line Plot of Avg Learning Time

Line Plot of Avg Num of Active Students

Hypothesis – Maybe the schools in District 8 are elite, so they don't recruit too many students.

Data Analysis – Exploratory Analysis of Trends

- Trend - The average learning time shows roughly the same trend as the average number of students. One thing I noticed is that I thought the red line would stay at the peak in May and June, but it went down instead.

- Hypothesis – The teachers still preferred using their own materials, but due to the policy, they still have to ask students to sigh up Khan Academy.
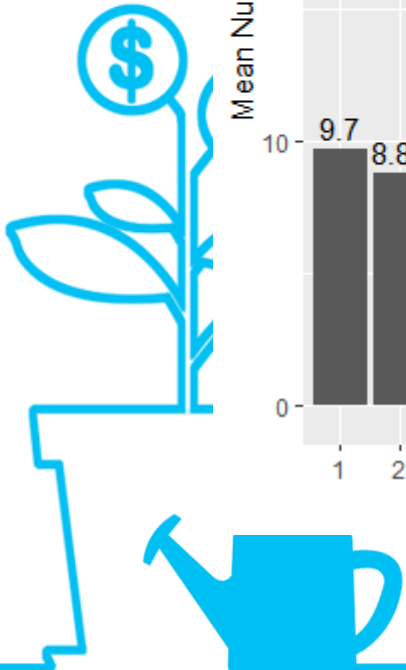
04 Answers of Measures

Answers of Measures

- Myself – Discuss with other educators and other officials using Khan Academy whether they are facing the same situation and how they solve the problem. Keep doing data analysis to see whether the new policy brings any improvement.

- State officials – Provide the benefits of Khan Academy to people who use their materials, like funding. Let students know the advantages of using their online supplements.

- Teachers – Encourage the teachers to use Khan Academy more by providing them with the guidelines. Ask them to collect feedback on using Khan Academy and teachers' material from students.

Thank you