Quality of NYC Schools - Survey Analysis

Sandro Mikautadze

Last compiled on 11/07/2022

Contents

1 Introduction

This is a TLDR. Enjoy!

1.1 What Is the Project About?

- Do student, teacher and parent perceptions of NYC school quality appear to be related to demographic and academic success metrics?
- Do students, teachers, and parents have similar perceptions of NYC school quality?

2 Data

2.1 Raw Data Analysis

2.1.1 Initial remarks

In `data\raw-data` 5 files are available: **combined.csv**, **masterfile11_gened_final.txt**, **masterfile11_gened_final.xlsx**, **masterfile11_d75_final.txt** and **masterfile11_d75_final.xlsx**.

Without importing the files yet, from the *Survey-Data-Dictionary* file in `data\metadata` we can notice that **masterfile11_gened_final** and **masterfile11_d75_final** differ by a small thing: **gened** contains information on all community schools, while **d75** from all District 75 schools. As the Dictionary states, "these files display one line of information for each school, by DBN, that includes the response rate for each school, the number of surveys submitted, the size of the eligible survey population at each school, question scores, the percentage of responses selected, and the count of responses selected".

Both files come with two different formats: *.txt* and *.xlsx*. I decide to work working with *.txt*, because the Excel version requires paid software to be visualized (i.e. Microsoft Excel).

NOTE: having a look at the *.txt* datasets, we can notice that they are actually saved as tsv (tab separated value) files.

Instead, the **combined** dataset has been pre-cleaned as an exercise and contains combined information on different NYC schools based on SAT, AP scores and geographical data.

2.1.2 Loading the dataset

Using the `readr` under `tidyverse`, I will import the datasets as `combined`, `general` and `district`.

```
dim(combined)
```

```
## [1] 479  30
```

```
dim(general)
```

```
## [1] 1646 1942
```

```
dim(district)
```

```
## [1]   56 1773
```

The first thing we notice is that

2.2 Processed Data