**Overview of the Dataset**

The data set originally consisted of 6 main variables and 13,944 cases that pertain to school attendance data for 3 schools namely the Ann Street School, the Louis A Spencer School and Chancellor Avenue School that are located in the New Ark School District. The students are identified by unique ID numbers and are enrolled from grades PK 3 to grade 8. Furthermore, each school is identified by a distinct SCHOOLID.

About 2 new variables were added to the dataset which pertained to approximating the percentage attendance, and the percentage absence of each student over the term. Furthermore, the students were categorized by their attendance into four main categories.

The attached excel and R code contains attendance data for students from 3 schools along with the newly created variables. Below is a data dictionary that defines the variables in the file.

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| STUDENT\_NUMBER | Student number (unique identifier) |
| SCHOOLID | Student’s school code for the 2015-2016 year |
| SCHOOL\_NAME | Student’s school name for the 2015-2016 year |
| GRADE\_LEVEL | Student’s grade level for the 2015-2016 year |
| Memsum | Student’s days enrolled (in NPS) in 2015-2016 year |
| Attsum | Student’s days attended (in NPS) in 2015-2016 year |
| Percentage\_Attendance | Percentage\_Attendance is the percent of days the student attended the school (in NPS) in 2015-2016 year |
| Percentage\_Absence | Percentage\_Absence is the percent of days the student was absent from the school (in NPS) in 2015-2016 year |
| Student\_Absenteeism | Categorizes the students by their Percentage Absence into 4 main classes based on set thresholds: a) Good Attendance, b) Approaching Chronic Absence , c) Chronically Absent, d) Severely Chronically Absent |

**Exploratory Data Analysis**

The average number of days a student was enrolled in each of the three schools approximates to around 176 days . The minimum number of days any student was enrolled was 7 and the maximum number of enrolled days were calculated to 191. Average attendance for all the students in the 3 schools amounts to 167 days. Similarly, 7 is the minimum number of days that a student attended class and 190 is the highest number of recorded attendances by a student. Plotting showed concentration of most student enrollment and attendance numbers to be around the mean with few outliers, with the standard deviation for the two metrics being 24.5 and 27 respectively.