

# About Dataset

This dataset includes scores from three test scores of students at a (fictional) public school and a variety of personal and socio-economic factors that may have interaction effects upon them.

## Remark/warning/disclaimer:

- This datasets are **fictional** and should be used for **educational purposes only**.
- The original dataset generator creator is Mr. Royce Kimmons
- There are *similar datasets* on kaggle already but this one is **different** and **arguably better** in two ways.
  - > 1) has **more data** (>30k instead of just the 1k the other datasets have),
  - > 2) has extended datasets with **more features** (15 instead of 9) and has **missing values** which makes it ideal for data cleaning and data preprocessing.

## Data Dictionary (column description)

1. **Gender:** Gender of the student (male/female)
2. **EthnicGroup:** Ethnic group of the student (group A to E)
3. **ParentEduc:** Parent(s) education background (from some\_highschool to master's degree)
4. **LunchType:** School lunch type (standard or free/reduced)
5. **TestPrep:** Test preparation course followed (completed or none)
6. **ParentMaritalStatus:** Parent(s) marital status (married/single/widowed/divorced)
7. **PracticeSport:** How often the student practice sport (never/sometimes/regularly))
8. **IsFirstChild:** If the child is first child in the family or not (yes/no)
9. **NrSiblings:** Number of siblings the student has (0 to 7)
10. **TransportMeans:** Means of transport to school (schoolbus/private)
11. **WklyStudyHours:** Weekly self-study hours(less than 5hrs; between 5 and 10hrs; more than 10hrs)
12. **MathScore:** math test score(0-100)
13. **ReadingScore:** reading test score(0-100)
14. **WritingScore:** writing test score(0-100)