**Capstone Project Proposal**

***Introduction to Data Science by Springboard***

Physical Fitness Test **Analysis**

**By Sumita Palanisamy**

**Introduction**

By law (California Education Code Section 60800), all public local educational agencies in California are required to administer the Physical Fitness Test (PFT) annually to all students in grades five, seven, and nine. The main goal of the test is to help students in starting life-long habits of regular physical activity.

The test has six parts that show a level of fitness that offer a degree of defense against diseases that come from inactivity. The test results can be used by students, teachers, and parents. During the month of February, March, April, or May, the governing board of each LEA maintaining grades five, seven, or nine shall administer to each student in those grades the PFT.

**The Problem**

The goal of this project is to determine if, given certain characteristics of the student (age,race, gender,school type,socio economic information) can the fitness level of the student be predicted. Furthermore, this project aims to explore if there are any trends or patterns in types of schools and school districts and fitness levels and if age plays a part in the fitness level of a student.

**The Clients**

The main clients for this analysis are:

* School officials : Teachers to design the curriculum for physical education programs
* Government officials such as California Department of Education : this report will help them compare how children from  independent charter schools perform in comparison with the children from non-independent charter schools. Help identify school districts that have fitness level lower than others. It might help with budgeting for the future.
* Parents, Guardians, Students : To assess levels of health-related fitness and to plan personal fitness programs

**The Data**

The data for this project is the Physical Fitness Test (PFT) data the state of CA, year 2017. This dataset contains about 1048875 rows and 24 columns. This dataset was accessed from [https://www.cde.ca.gov/ta/tg/pf/pftresearch.asp](https://www.cde.ca.gov/ta/tg/pf/pftresearch.asp%20)

**The Approach**

The following steps highlight the strategy to be adopted for carrying out the analysis for the Capstone project:

1. Data Wrangling and Cleaning
   * Dropping columns that are irrelevant to the analysis
2. Exploratory Data Analysis
   * Perform Regression Analysis to determine factors that most influence the fitness of a student
   * Identify patterns and correlation between the different variables
   * Use data visualization for graphical analysis to answer questions on the dataset
3. Predictive Analysis
   * Apply machine learning techniques to see if the student’s fitness level can be predicted based on various characteristics.

**The Deliverables**

The deliverables for this project will include a report of all the findings, a slide deck and R code, all of which will be published on GitHub.

**Specific questions to be explored**

1. Does the ethnicity of student influence the fitness level?
2. Does the age of the student affect the fitness level? Are there more students passing the test in the 5th grade vs the 7th grade vs the 9th grade?
3. Does the gender of the student affect the fitness level?
4. Does the student’s socio-economic condition affect the fitness level?
5. Does fitness level vary in the multiple school districts?
6. Does fitness level vary in the multiple counties?
7. Do students do better or worse in a certain exercises (abdominal strength, flexibility, trunk extension) as they age?
8. Does socio economic condition play a part in the kind of exercises the students are good/bad at?
9. Are students in independent charter schools more fit than children in non-independent charter schools?