**Linear Regression project to predict college GPA using SAT scores**

Prepared By: **Harshit Kulkarni**

# Introduction

This project is all about predicting college GPA using SAT scores of students. The algorithm used is Linear Regression. Linear regression is used for finding linear relationship between target and one or more predictors. There are two types of linear regression- Simple and Multiple. Simple linear regression is useful for finding relationship between two continuous variables. One is predictor(GPA) or independent(SAT) variable and other is response or dependent variable. It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other.

Y(pred) = AX+B

Where Y = Dependent Variable i.e in our case GPA

A = Gradient of the line

X = Independent variable i.e in our case SAT score

B = constant

The SAT as a predictor of high academic performance at the postsecondary level is a controversial subject. The College Board claims that SAT scores and high school grade point average are more closely linked to later performance in college than high school grades by themselves.

We have a dataset which contains information about relationship between ‘college GPA’ and ‘SAT scores’. Using the training data, a regression line is obtained which will give minimum error. This linear equation is then used for any new data. The values A and B must be chosen so that they minimize the error. If sum of squared error is taken as a metric to evaluate the model, then goal to obtain a line that best reduces the error.



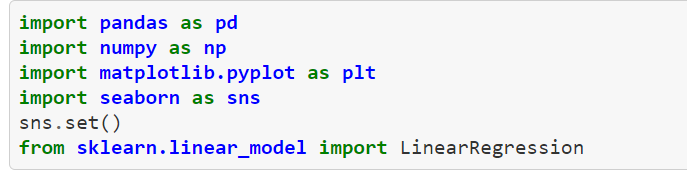
## Technical Solution

Technical solution to the problem is to create a linear regression model (Statistical method) in python using different libraries such as,

* Pandas
* Numpy
* Matplotlib
* Seaborn
* Sci-kit learn

The use of above libraries were made to build the train the

Model.



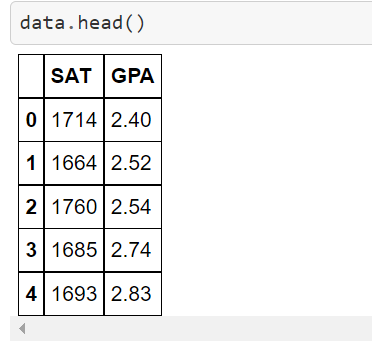
## Implementation

Linear regression assumes a linear or straight line relationship between the input variables (X) and the single output variable (y)

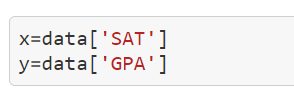
* Importing the libraries
* Importing the dataset (CSV file)



* Reading the head of the data



* Assigning x and y values as SAT and GPA



* Reshaping x and y values
* Fitting the model
* Calculating score , intercept and coefficient
* Making predictions

The predictions made are given below

