# **Multiple Linear Regression Modelling:**

**Objective**: To build a model which predicts the Average Performance Index (API) of Elementary Schools for the year 2000 based on various contributing factors.

**Software to Use**: SAS

**Data Information**: We have data of various attributes like class size, enrollment, poverty, parent education, student performance, teachers’ credentials from 400 elementary schools from the California Department of Education's API 2000 dataset.

|  |  |
| --- | --- |
| VARIABLE NAME | Label |
| acs\_46 | class size 4-6 |
| acs\_k3 | class size k-3 |
| api00 | Academic Performance Indicator 2000 |
| api99 | Academic Performance Indicator 1999 |
| avg\_ed | Average number of Years of High school parent education |
| col\_grad | % of Parents who are college grad |
| dnum | district number |
| ell | % of Students who are English language learners |
| emer | % of part Time Teachers |
| enroll | Number of students |
| full | % of full Time Teachers |
| grad\_sch | % of Parents who are attended grad school |
| growth | growth 1999 to 2000 |
| hsg | % of Parents who are high school graduate |
| mealcat | free meals in 3 categories |
| Smeals | % of Students who opt for free meals |
| mobility | Dropout Rate |
| not\_hsg | % of Parents who are not high school graduate |
| snum | school number |
| some\_col | % of Parents who are attended some college |
| yr\_rnd | year round school – School open through the year or not |

Source: <https://www.data.gov>

**Challenges**:

* Numerous missing values, invalid observations and outliers.
* Highly correlated independent variables.

Refer to ‘**Codes\_Model1\_7.doc’** for the approach.

Refer to ‘**Regression Equation.txt’** for the equation.

Refer to PPT ‘**MLRM Case Study – Vishesh Goel’** for key outline and summary of the model and business insights.