**Project 5: Early Childhood Longitudinal Study (ECLS)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Abstract:**

Families who live in poverty are exposed to greater environmental stressors and risk factors than those who do not live in poverty. Children of poverty have greater challenges and these stressors can negatively impact their psychological and social development, which in turn, negatively effects their academic achievement. These environmental influences substantially increase the negative effect on the cognitive processes of a developing child. Children who experience environmental stressors often display signs of behavior problems, depression, anxiety disorders, and/or impaired academic achievement. Bio-ecological factors such as familial relationships, peer relationships, community experiences, and school environments contribute an integral role in the development of a child. When these contributors are presented in the context of poverty-stricken circumstances, consideration of these unique indicators should be closely examined as a possible root-cause for adverse performances in academic environments. In this study, an analysis of data collected from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) will be conducted to evaluate the short-term and long-term impact of environmental stressors on cognitive development and self-efficacy respectively for students who live in poverty.

**Project Statement:**

Develop a comparative study that tells which environmental stressor(s) has the most significant short-term and long-term impact on cognitive development, as defined by reading, math, and science IRT scores, for children in poverty and for children not in poverty

**Expectations:**

A data science-centric analysis of significant attributes that impact the reading, math, & science scores across the waves for children in poverty and children not in poverty that includes

* Data preprocessing
* Model building
* Visualization
* Analysis & Insights
* A Document containing R code, and entire analysis

**Data:**

1. ECLS data for 7 waves
2. Variable description

**Preprocessing Steps Required:**

* Attributes that have values “SUPPRESSED” can be omitted from analysis
* Attributes that have type as “Weight” in attributes Description pdf can be omitted (Eg: BYCW0)
* Attributes that have more than 30% of missing values or NAs can be omitted
* The second position of the attribute name corresponds to the wave number. (Eg: F5Specs belongs to wave5) Based on this data for each wave can be subset
* However, attributes with prefix W (SESL, POVRTY, PARED, INCCAT have W's in the prefix.) have a different interpretation.

The attributes that start with

* + - WK belongs to - Waves 1&2
    - W1 : Waves 3&4
    - W3 : Wave 5
    - W5: Wave 6
    - W7 : Wave7