DS 2 - Project Proposal

Competition: Child Mind Institute - Problematic Internet Use (https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use)
Data taken from: The Healthy Brain Network Dataset

Dataset Description

The Healthy Brain Network (HBN) dataset is a clinical dataset containing records from approximately 5000 people ranging from 5 to 22 years old. The broader goal of this study and dataset is to identify biological markers that predict the mental health of children. The goal of this particular Kaggle competition is to predict the impact of children's problematic (excessive) internet use using demographic, physical, and questionnaire data (from both parents and clinicians) along with activity data provided by a fitness watch. This impact is measured using the Severity Impairment Index (SII), which ranges from 0 to 3.

Column Category Overview

A paraphrase/reproduction of https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use/data

- Demographics (age and sex)
- Internet Use (in hours)
- Children's Global Assessment Scale
- Physiological / Physical Measures
- The FitnessGram Assessment
- Bio-electric Impedance Analysis
- Physical Activity Questionnaire
- Sleep Disturbance Scale
- Data from a "bio-tracker" / fitness-watch-type device

Questions

1. Does the CGAS depend on the season it was administered?
2. Do activity scores tend to diverge before/after the PCIAT is administered?
3. Does PCIAT correlate with age?
4. Can we predict SII without using a complex ML / NN model?

How Dataset Will be Used to Answer Questions

1. Analyze and build regression model to establish if there is a relationship between CGAS scores and CGAS participation season
2. Perform statistical analysis or regression to see if PCIAT questionnaires differ significantly before, around, and after PCIAT is administered
3. Build and analyze a correlation model / regression to determine if the age of participants has a direct correlation with the PCIAT score