Brecht Lauwers r0830649

Project description

I’m going to create a model that predicts how much students will score on a math exam in a certain school and certain grade. These are schools situated in NYC and they are all labelled with an ID named DBN (District Borough Number). For example, a school in a certain district and certain grade will get a mean of 83.0% on their math exam. You can infer that this is a regression problem.

The dataset consists of students in a certain school with there math results from grades 3 through 8 (*2de leerjaar tot 2de middelbaar*). The data is not cleaned and contains empty/missing/wrong values with 33.5K rows and 16 columns in total. I haven't found any implementations online of this dataset for a regression problem either so this seems perfect for this assignment.

The only problem is that I don’t have any additional information from the features or database in general so I have to figure this out myself.

Here you can find the database:
https://data.cityofnewyork.us/Education/2006-2012-Math-Test-Results-All-Students