This is a checkpoint document.

The date is 10.05.2021. The last changes were made 2 weeks ago.

Issues we’ve encountered so far include

1. messy data entered by the CASAS system, to be excluded from final analysis.
   1. For example – data entered on two different scales. Numeric (in the 100s) and grade category.
   2. Measures to be taken include drafting and sending a memo to Ta Tanisha.
2. and early correlations not showing up in the final logistic regression.

Other intermediary measures to take:

1. create some visuals to depict yearly trends in enrollment by race, gender, citizenship etc.

---- Note: From here on out, the protocol for the ML Model project, will be, not to copy over files and folders from previous weeks. But instead to upload the repo altogether and to download the entire repo whenever I’m working on it.

* Note about data updates: trust that the original data has not been modified, and is true to the original pull. Work off of this assumption, until it comes time to create another data pull.