Combining two datasets to get which counties have more overdose and successful reversals by naloxone, then the governor can pay attention to these countries to prevent overdose. But I found it is hard to combine two datasets.

use "Number of successful reversals" and "Incident County " as variables and county as the response. then we can get which county use more Naloxones and when people use more naloxones to rescue patients under overdose. Also, we can use both two datasets to predict the county and risk time.

Also, I have searched dataset about Centers of Excellence and PacMAT in your opinion, but I still can't find the dataset under Centers of Excellence. Furthermore, treatment happens after overdose, not before overdose. In other words, it is a little bit different with our topic Prevention Target. maybe it can be as an additional suggestion in our project.

Methadone is a way to treat opioid addiction. So, maybe we should focus on what is an effective opioid treatment. I notice that in the 'Treatment' link on data.gov.pa, there is a distinction between the types of treatment centers in PA. There are Centers of Excellence and Pennsylvania Coordinated Medication Assisted Treatment (PacMAT). Perhaps we could identify which of these treatment facilities/programs is most effective? This would also fit into the Case Study we chose, 'Targeting Prevention' in terms of guiding the state with where funding dollars should be spent. For example, if we can show, statistically, that the Centers of Excellence are more effective at treating opioid addiction, then we could recommend to open more centers like that.

This is only an idea as I have been digging more into the data. Also, another thing that makes the think the above idea is better than a predictive model, is that every data set I looked at today was a frequency (that is, count) data. For example, the count of some statistic in a particular county. I am not finding any individual data on patients/citizens. In order to do predictive modeling, we need each row to represent a patient/citizen, not a count. So, we may end up doing something like t-tests, ANOVA, etc… Especially if we were comparing certain statistics at treatment centers.