

CHALLENGE

CHAID vs Racial Bias

The dataset contains all cars stopped by the police in 2017 in Minneapolis. The goal is to understand the drivers that lead to people being stopped by the police with a focus on understanding if the race is a key driver or not.

CHAID is a fantastic technique for driver analysis and segmentation.

01

DATA PREPARATION

The most important part of driver analysis is the variable selection. Choose these 5: problem, vehicleSearch, race, gender and policePrecinct.



02

VARIABLE TRANSFORMATION

CHAID requires that all independent variables be categorical. That means all string and continuous variables must be transformed for us to have inputs prepared.



03

NAMES

The last step of the preparation requires us to isolate the names of the dependent and independent variables to use later on.



04

CHAID

I know the preparation takes time but I think it is worth it. Create the model and plot the outcome. Set the max_depth to 2. The last step is to interpret the results. Is there racial bias in the Minneapolis police?



Diogo Resende

