I trained a *RandomForestRegressor* model on the dataset complete feature vector and an additional random value feature. By evaluating the *feature_importances_* of the model, I realized that the random feature is placed at the 6th rank of the most important features above populationDensity#1 which seems weird.

So, I decided to find the most important column using *drop column importance* method. To mitigate the model retraining cost and get a basic estimate of the features' importance, I trained the model 378 times excluding one feature in each iteration on a portion of observations in the dataset (30%).

It turns out that populationDensity#1 (F) is the most important feature. I performed drill down on F and create the feature F'. Selected threshold is the mean of F to have two bins [0 and 1].

Then, I trained two models. One time including F' as an additional feature and another time by replacing F with F'. Both models returned a slightly lower score in comparison to base model scores.

Following you can find obtained scores:

```
Model Score - base model: 94.689
Model Score - F' added: 94.661
Model Score - F' replaced: 94.308
```