Observations:

* max\_samples = 0.1
  + - increases with n\_estimators, but is logarithmic (slowing down but not leveling off)
    - min\_samples\_leaf = 2,3 are best, higher for 2
  + Standard deviation (STD)
    - min\_samples\_leaf = 2,3 are best, higher for 2
    - increases with n\_estimators, but reaches a SLIGHT peak at n\_estimator = 130 (min\_samples\_leaf = 3)
    - increases with n\_estimators, but reaches a SLIGHT peak at n\_estimator = 125 (min\_samples\_leaf = 2)
    - on average LOWER for min\_samples\_leaf = 3 by
  + verdicts
    - min\_samples\_leaf = 2 on average performs better but has more variance than min\_samples\_leaf = 3

max\_samples change from 0.1 to 0.2:

* Differences:
  + increases , slightly increases standard deviation
  + min\_samples\_leaf seems to have relationship with max\_samples
  + **preferred min\_samples\_leaf changed from 2 to 3**
* Similarities:
  + increases with n\_estimators, but is logarithmic (slowing down but not leveling off)
  + STD almost constant at

min\_samples\_split

* min\_samples\_split = 2,3,4 produces the same , MAE, MSE, and their STD values
  + 4 has shorter fit time than 2,3 → **use 4**
* D

min\_impurity\_decrease

GEOID performs better than STATEFP, but both is better thanjust GEOID.

Feature importances + :

* Both GEOID and STATEFP with everthing else: 0.70231060846335
  + GEOID: 0.216626
  + STATEFP: 0.090070
* Only GEOID with everything else: 0.702206936031992
  + GEOID: 0.305346
* Only STATEFP with everything else: 0.701828360584809
  + STATEFP: 0.276674

Only using STATEFP rather than GEOID → no substantial difference to model performance

Temporal feature importances:

* Month > months\_from\_start > year
* 0.191141 > 0.109396 > 0.031172
* scores:
  + Best with year, month, months\_from\_start
  + But VERY CLOSE with just month, months\_from\_start

suggests there are strong seasonal cycles in deathrate from air quality