

# FORECASTING DAILY COVID- 19 SPREAD IN REGIONS AROUND THE WORLD.

CONFIRMED CASES AND  
FATALITIES



# *What does the data tell us?*

- Negative values in the TargetValue field
- Province\_State had data for *US, UK and China*



Days since first infected



*What features matter*

## *Data enrichment*



population density



population median age



urban population  
percentage

# Model Selection and evaluation

MODEL	R2 SCORE ON CONFIRMED CASES	COMMENTS
Ridge Regression	-404026.412	Performs badly
ElasticNet	0.079	Relatively better than Ridge Regression
ExtraTreesClassifier	0.217	Relatively better than Ridge Regression
Linear Regression	0.475	Relatively better than Ridge Regression
SGDRegressor	0.518	Relatively better than Linear Regression
GradientBoostingRegressor	0.7979	One of our top 4 models
HistGradientBoostingRegressor	0.862	One of our top 4 models
RandomForestRegressor	0.892	One of our top 4 models
DecisionTreeRegressor	0.899	One of our top 4 models