Quantitative Population Prediction by Place (USA)

Winston Robson | Galvanize Data Science Immersive g88 (San Francisco, Spring 2019) | https://bit.ly/2Z5IONb

Purpose

This project aimed to utilize machine learning on combined Census and American Community Survey datasets to predict the future population of any place in the United States

First Look Tech

Facebook Prophet is a procedure for forecasting time series data based on an additive model fitting non-linear trends.

Performance Metrics

Mean Absolute Percent Error

*mean(abs(predicted-actual)/actual)*100*

- Model
  - 2016 / 2017: 1.826% / 1.993%
- Baseline
  - 2016 / 2017: 2.052% / 3.855%
- Prophet Cross Validation [1] [2]
  - 2015: 2.29%

Root Mean Squared Error

sqrt(mean(square(actual - predicted)))

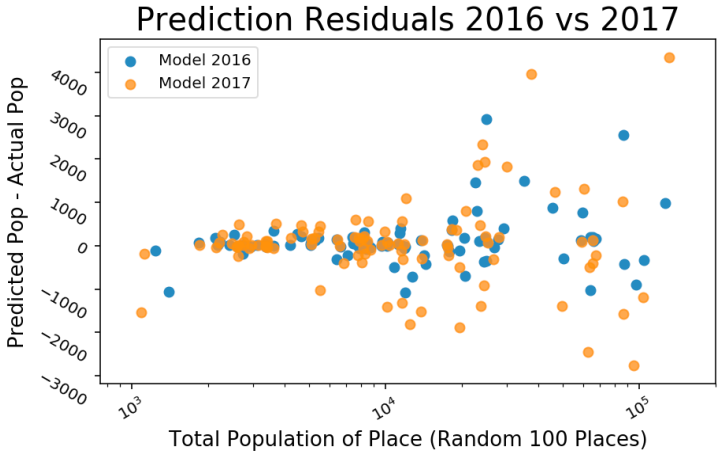
- Model
  - 2016 / 2017: 309.39 / 838.41
- Baseline
  - 2016 / 2017: 433.73 / 868.39

Log Sum Accuracy

log(sum(predicted)/sum(actual))

- Model
  - 2016 / 2017: 0.00318 / -0.00097
- Baseline
  - 2016 / 2017: 0.01245 / 0.02369

Scaled Results

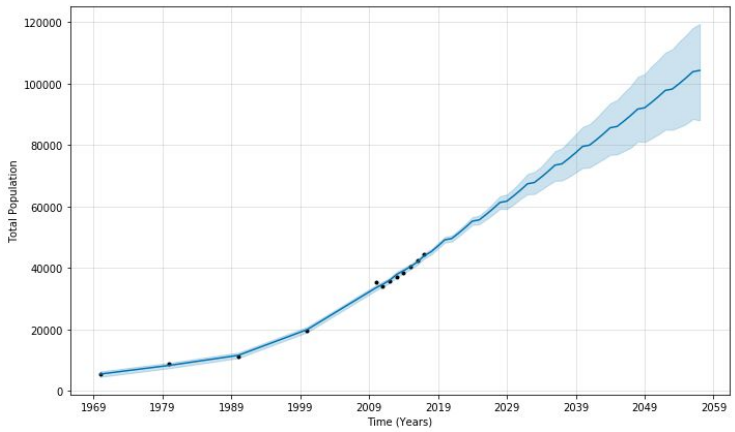


Process

1. Exploratory Data Analysis
   a. Examined geo. filters for Total Population
      i. E.g. Place, 5-digit Zip, County, State
   b. Determined Place to be most usable
2. Combined
   a. Census 1970-2010 dataset
   b. Place ACS 5-year Estimate datasets (‘11-’15)
3. Defined Baseline
   a. Assume max change from year-to-year Place total population, most recent 5 years
4. Engineered Prophet to forecast Total Population
   a. Tuned Model using random samples of 100 to 5,000 Places with various sensitivities
   b. Compared Prediction to Baseline vs Actual Total Population for 2016 & 2017
5. Outcome seemed dim; but eventually broke through with a Model that could consistently outperform Baseline (*MAPE, RMSE, LSA*) across samples

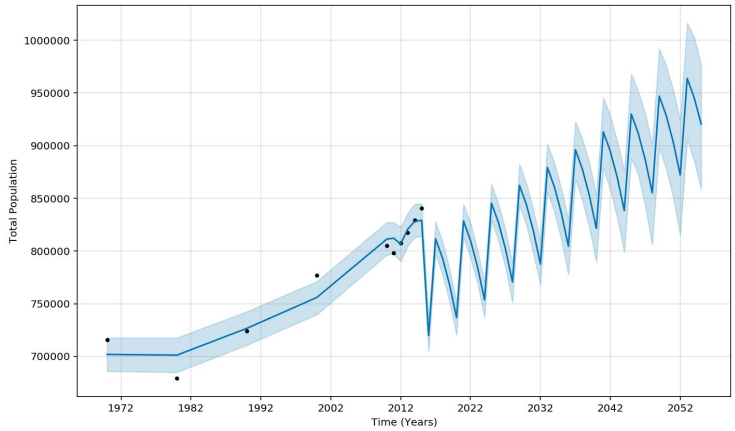
Sample Forecasts

Bentonville, Arkansas (MAPE = 3.7940%)



2000 actual: 19,730 *2016 actual: 42,499*
2020 prediction: 48,152 *2040 prediction: 75,119*

San Francisco, California (MAPE = 0.2748%)



2000 actual: 776,733 *2016 actual: 850,282*
2020 prediction: 897,756 *2040 prediction: 1,109,831*

[1] Model only; preferred method, source: Prophet Paper (Sept. 2017) Avg(5 samples @ 100 Places)

[2] Prophet cross validation building up to prediction of 365 days, every 180 days, starting after 14,235 days (39 years) -- *13 total, 2009-2015*

Project on Github Winston Robson LinkedIn 479.268.0047 / 925.400.8172 warobson@gmail.com