Quantitative Population Prediction by Place (USA)

Purpose

When listing any house, the numerous measurements any good REALTOR® will consider range from the condition of the property to what's for lunch that day. Most prevalent, though, is their assessment of demand; after all, a million-dollar mansion (or million-dollar shed in San Francisco), is only worth $1,000,000 if someone will purchase it at that price.

As humans take up space, one main driver of demand in any housing market is population. The more people seek to live in any location, the more valuable that location is.

While it is entirely reasonable for you to build a $7.9m single-family resort in Sidney, Nebraska, you may encounter a few difficulties trying to sell it in 2019.

On the other hand, however, if you were the one acquiring the modest 4 bed, 4 bath, 4,437 sqft abode at 24 Clarendon Ave, San Francisco, CA on August 5th, 1996 for $588,000,

you could comfortably list it for a mere $4,400,000 and expect to sell in about 40 days, while moaning how you should have bought the spot down the street, just listed at $5,995,000.

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

Tech Used

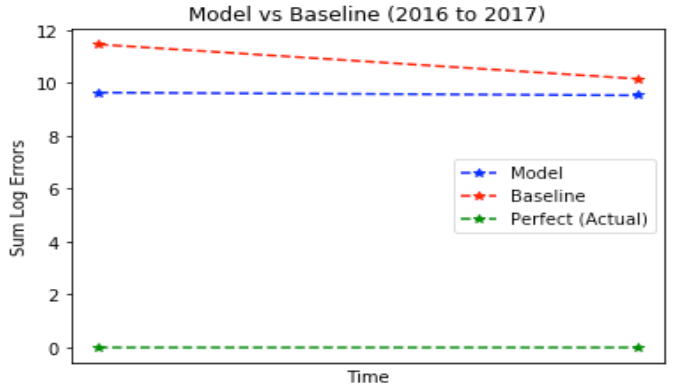
PROPHET

python™ NumPy scikit learn matplotlib pandas

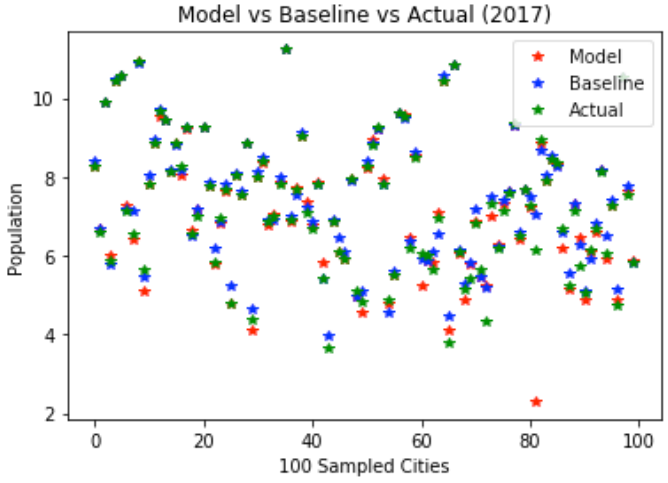
$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

Process

1. Exploratory Data Analysis
    a. Examined large number of Geographic filters on Total Population
        i. E.g. Place, 5-digit Zip, County
    b. Determined Place to be most usable
        i. Counties were too ranged in Number of Measurements, some since 1790 and having 20+ w/ others < 10
        ii. 5-digit Zip measurements were initialized too recently, no multi-decade historical data
2. Combined Place Census 1970-2010 Total Population data with each Place ACS 5-year Estimate 2011-2015 of Total Population (Age/Sex)
3. Defined Baseline as assuming Continued Trend from Place's population change 2014-2015
4. Engineered Generalized Additive Time-Series Model using Facebook's Prophet to forecast Total Population
5. Randomly sampled 100 places Coexisting across ACS years 2011-2017 and 2+ Census 1970-2010
    a. Allows interpretation on per year and decades-long basis
6. Fit Model on each sample place
7. Measured Model Outcomes v. Baseline Assumptions
    a. Compared to Actual 2016 & 2017 Populations



- Model about 15.9% more accurate than Baseline for 2016 predictions of Population for Sampled Places
- Model about 6.2% more accurate than Baseline for 2017 predictions of Population for Sampled Places



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