

Rosalie Day - CAPSTONE1

Prediction of new housing by existing housing stock by US regions

March 11, 2019

Reports of a housing crisis commonly only contain two variables, the price of real estate and, mostly anecdotally, commuter status. Multiple factors make up price; are closely associated with individual willingness to pay; and hold little opportunity for impact for all but the most draconian measures. This project is going to examine and display housing stock characteristics, grouped regionally and whether the location is in a metropolitan area. Then, the association is going to be explored with household income and commute types, culminating in predictions of housing characteristics that will be demanded.

The use for this information is for:

- developing public policy incentives to alleviate the housing shortage;
- adding to factors to be considered for public transportation planning;
- matching demanded housing characteristics for residential land development; and
- lay persons forming evidenced based housing expectations for cross-regional relocation.

The American Housing Survey data is collected by the Census Bureau and is sponsored jointly with the Department of Housing and Urban Development. All of the variables this project requires are publicly available and provided in csv file format.

The first step of the analysis is going to classify housing on a regional level by type, structures, and year built, testing for expected regional and metropolitan area differences in the existing housing stock. The status of the residential unit is current and are associated with structures built before 1919, which is the starting level, and apply respectively to all the incremental categories of year built up to 2017. The hypothesis tested is that the basic features will be relatively the same for houses built before and after 2010, and the assumption is that combination of feature reflects householders' preferences.

The second step, will use these preferred features, including square footage, detached units or not, and various room configurations to test associations. On the regional/metropolitan levels, regression models will include other variables, such as household income, own or rent, education and commute variables, or potentially a commute index composed of type, distance and commuter days.

Reported trends that Americans are demanding more square footage per person will be examined for income, region, metropolitan area, and commute distance.

The third step will determine what characteristics are preferred in particular regions and metropolitan or not, given income and education, and whether metropolitan house characteristics are more like their respective regions or are more a similar among themselves. The hypothesis is dense large cities are more like each other. Other metropolitan areas are more a reflection of their region. Regression and unsupervised modeling will be utilized, latter for metropolitan areas.

Deliverables will include code, plots and a markdown document for each of 3 steps. The first step, housing characteristics, will have a data exploratory section and descriptive statistics. The second and third steps will employ more relational statistics, modeling and plotting of the same, while the former contains data exploration of the added demographic and commute variables. The final deliverable will include an under 5-page written results summary and a slide deck that captures interesting and significant results for the entire project.

### *More research*

Another data set contains US new construction characteristics annually for single family houses completed since 2000. It is collected for industry insights but share some common variable measurements with the American Housing Survey. This data also can be grouped into geographic regions and whether it is in a metropolitan area and could be used for testing data from the model developed above. This more detailed data set of housing characteristics would allow for trends analysis and prediction of more consumer preferences for housing characteristics. This annual data will provide trend lines for housing features before and after the financial crisis of 2008. <https://www.census.gov/construction/chars/completed.html>