Project 2 Proposal

Team Members

Adam Sohn
Paul Petit

GitHub Repo

UCB-INFO-PYTHON/Project2PetitSohnREPO

Primary Dataset

Our primary dataset is described in the table below.

RESIDENTIAL REAL ESTATE DATA	
Source	https://www.realtor.com/research/data/.
Description	Data contains residential (single family home and condo/townhome) real estate inventory data from May of 2012 through October 2018 at various geographic grains for the U.S.

	According to the website, “Data in this realtor.com library is based on the most comprehensive and accurate database of MLS-listed for-sale homes in the industry. We aggregate and analyze data from hundreds of sources and produce hundreds of metrics for multiple markets, and curate figures and trends where possible for reliability and comparability.”
Fields of interest	Median Listing Price [Y/Y, M/M], Active Listing Count [Y/Y, M/M], Days on Market [Y/Y, M/M]
Grain	Year || Month || Metropolitan Area, [Zip Code, County, State]
Size	10 MB (39,001 rows, 35 columns) at finest grain
Notes	There are many datasets at the source linked above at different location grains (zip/county/state/etc.) and we may elect to use a coarser grain for some analyses if needed

Supplemental Dataset

We plan to adjoin an additional dataset with eviction data by year and geography. This dataset also conveniently includes demographic and income fields like population, poverty rate, median household income, and median gross rent by year and geography, which will allow us to test more sophisticated hypotheses. Further details in the table below.

EVICTION DATA	
Source	https://data-downloads.evictionlab.org/
Description	Data collected by the Eviction Lab, a research organization dedicated to studying the prevalence, causes, and consequences of eviction, including census data. The data is comprised of formal eviction records from 48 states and the District of Columbia from 2000 through 2016.
Fields of interest	eviction filing rate || eviction rate || median household income || median gross rent
Grain	Year || Metropolitan Area, County, [State]
Size	3.5 MB unfiltered (26,160 row, 27 columns) at finest grain

Notes	Demographic and income data isn't available for all years and all geographies, so we'll have to filter the dataset accordingly.

Variable Exploration and Expected Insights

Below is an initial set of hypotheses. It is expected that these insights will lead to further paths of inquiry, which will be followed. These paths may be corollaries to existing hypotheses or new hypotheses.

Hypothesis 1: The affordability of home ownership is disparate between metropolitan areas. It is expected to see major coastal cities as less affordable than major midwest cities.

Column	Variable Type
Eviction.year	Independent
Eviction.name (City)	Independent
Eviction.median-household-income	Independent
Realtor.Median Listing Price	Dependent
(Created) home_own_afrdblty = Realtor.Median Listing Price / Eviction.median-household-income	Dependent

Hypothesis 2: Areas where home ownership is relatively less affordable will similarly be relatively less affordable for renters. This will manifest in a higher rental rates and rental eviction rates.

Column	Variable Type
Eviction.year	Independent
Eviction.name (City)	Independent
Eviction.median-household-income	Independent
Realtor.Median Listing Price	Dependent
Eviction.median-gross-rent	Dependent

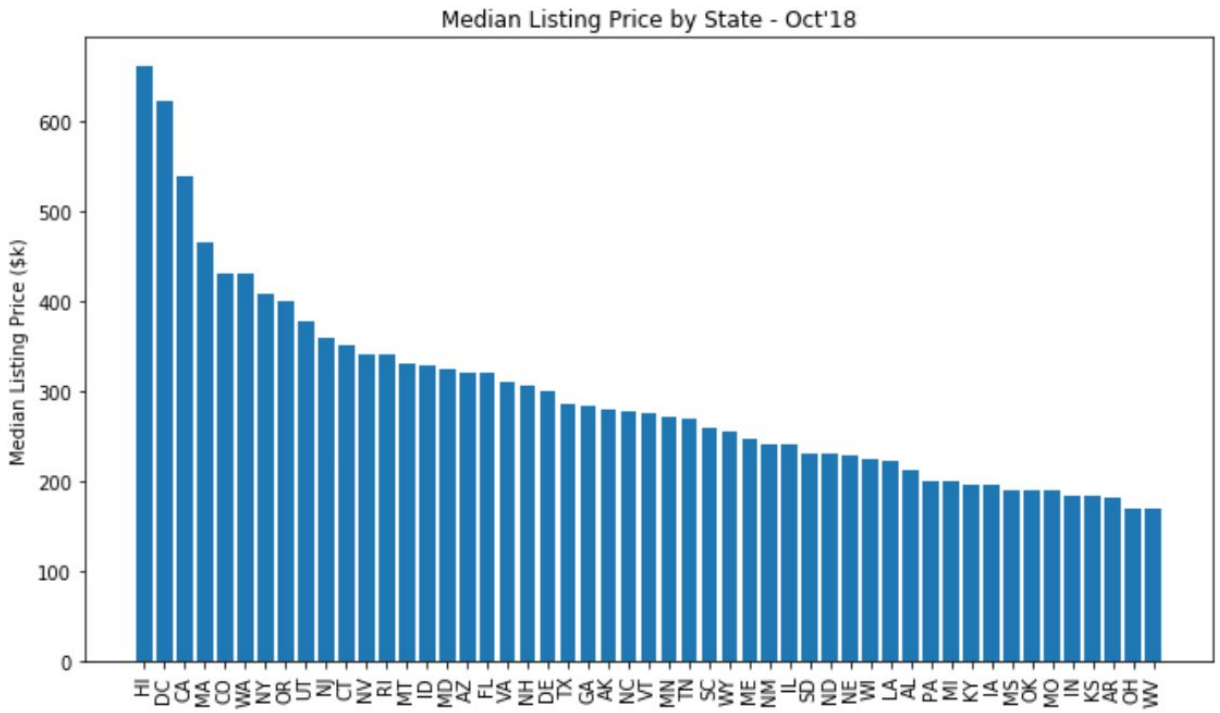
Eviction.eviction-filing-rate	Dependent
(Created) home_own_afrdblty = Realtor.Median Listing Price / Eviction.median-household-income	Dependent
(Created) rent_afrdblty = Eviction.median-gross-rent / Eviction.median-household-income	Dependent

Hypothesis 3: An acceleration of rent cost over time is a larger indicator of an increasing eviction rate than sustained, or linearly-increasing rental affordability issues.

Column	Variable Type
Eviction.year	Independent
Eviction.name (City)	Independent
Eviction.median-household-income	Independent
Eviction.median-gross-rent	Dependent
Eviction.eviction-filing-rate	Dependent
(Created) rent_increase_y_y = Eviction.median-gross-rent[year] - Eviction.median-gross-rent[year-1]	Dependent
(Created) rent_afrdblty = Eviction.median-gross-rent / Eviction.median-household-income	Dependent

Initial Plots, Figures, and Tables

An initial view of the ‘Median Listing Price by State’ (graph below) and ‘Top 10 Metropolitan Areas with Highest Median Listing Prices in U.S.’ (table below) offer a glimpse into the range of residential real estate prices in the U.S.



Top 10 Metropolitan Areas with Highest Median Listing Prices in U.S. in October 2018

	CBSATitle	Median Listing Price
35	San Jose-Sunnyvale-Santa Clara, CA	1099049.50
181	Santa Cruz-Watsonville, CA	900017.00
135	Santa Maria-Santa Barbara, CA	899525.00
10	San Francisco-Oakland-Hayward, CA	899050.00
311	Napa, CA	872050.00
275	Kahului-Wailuku-Lahaina, HI	850050.00
149	Salinas, CA	845275.00
468	Glenwood Springs, CO	799050.00
59	Bridgeport-Stamford-Norwalk, CT	749050.00
1	Los Angeles-Long Beach-Anaheim, CA	727526.25

Bottom 10 Metropolitan Areas with Lowest Median Listing Prices in U.S. in October 2018

	CBSATitle	Median Listing Price
351	Bay City, MI	98950.00
310	Weirton-Steubenville, WV-OH	94900.00
348	Decatur, IL	90550.00
499	Charleston-Mattoon, IL	89750.00
484	Richmond, IN	85049.50
340	Muncie, IN	84950.00
448	Danville, IL	79950.00
289	Johnstown, PA	78075.00
277	Pottsville, PA	75475.00
485	Marion, IN	70024.25

Report Outline

1. A brief introduction to the domain
    a. Include high level data views of average selling price by state and change in average selling price over time.
2. State the questions driving our analysis.
    a. Interested if home ownership affordability is part of a continuum of affordability issues that includes higher rate of rental, rental affordability, and rental evictions.
3. Introduction to datasets used.
4. Discuss data cleansing process and validation.
5. Tell story through:
    a. Hypothesis generation -- relate back to questions.
    b. Analytical investigation of data.
    c. Resulting answer to initial question.
6. Concluding remarks on findings from analysis.