Let’s Help Millennials Buy Homes!

RE Investment advice from data, not gurus!

The Problem….

35%
of millennials own a home.

75%
of millennials want to own.

Source: https://www.apmresearchlab.org/housingcost#h1.the_rise_of_renters

The Problem….

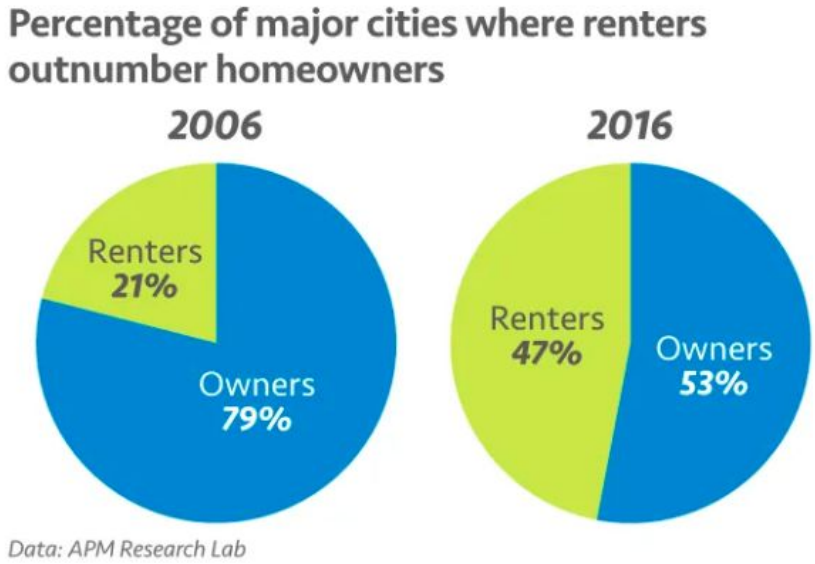
Homeownership among Baby Boomers, Gen Xers, and Millennials in 2015

Generation	Years born	Age	Population	Current homeownership (%)	Homeownership at age 25–34 (%)
Millennials	1981–97	18–34	75,170,263	32.2%	37.0%
Gen Xers	1965–80	35–50	66,441,487	60.4%	45.4%
Baby boomers	1946–64	51–69	74,649,971	75.0%	45.0%

Sources: 1990 and 2000 Decennial Censuses and the 2015 American Community Survey.

- Many well documented benefits of home ownership vs. renting for long term financial well being
  - https://www.investopedia.com/articles/mortgages-real-estate/08/home-ownership.asp

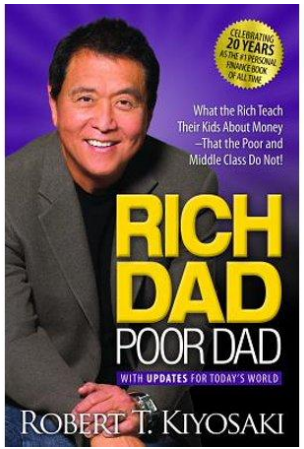
The Problem….

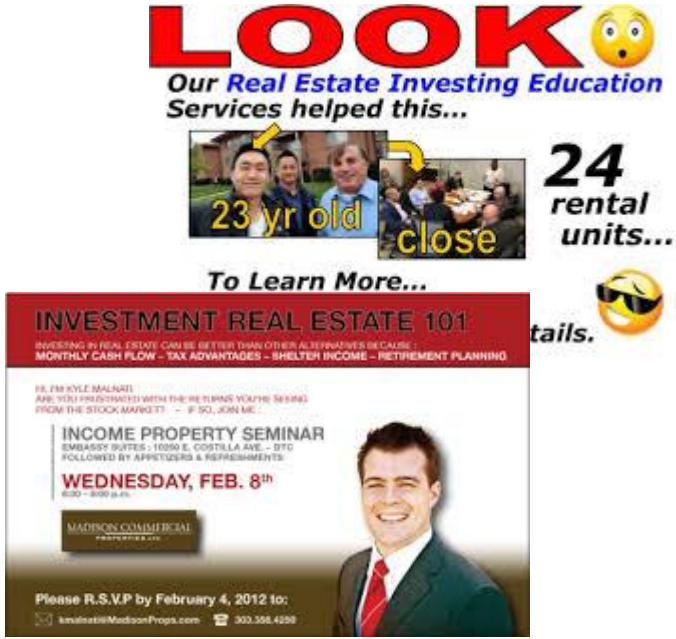


Source: https://www.apmresearchlab.org/housingcost#h1.the_rise_of_renters

Where do millenials go to for advice on RE?

- Thousands of sources of “expert advice”
  - Just google “Real Estate Investing”





Common Claims:

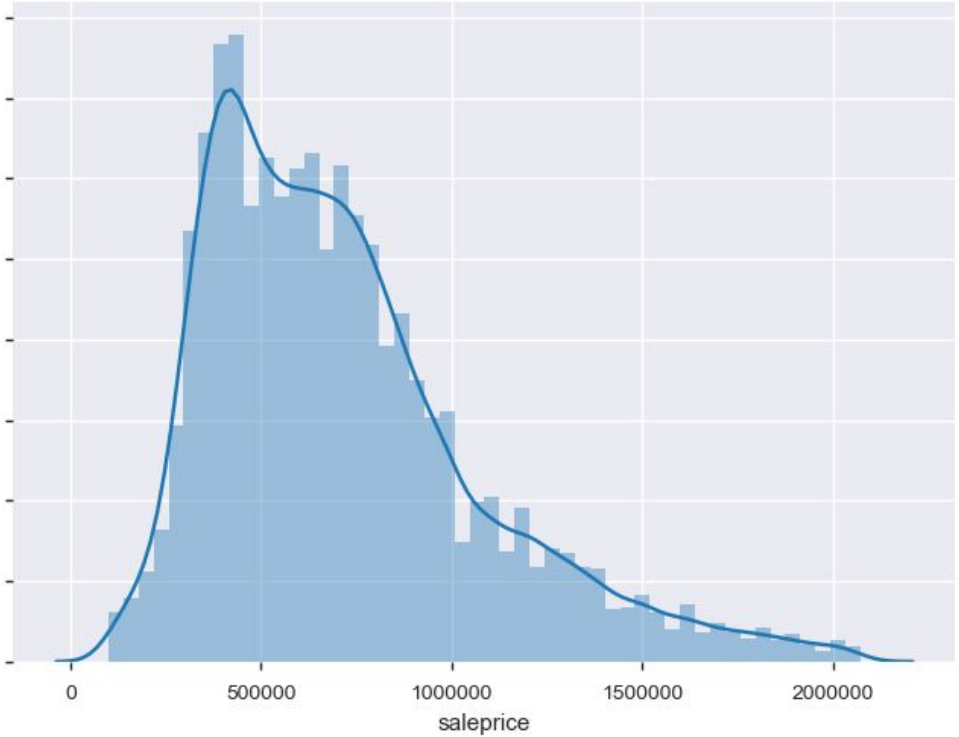
1. Higher square footage increases home sale price
2. The house filling a higher proportion of the overall lot decreases home sale price
3. The cost per square foot is lower in duplexes than in single-family homes

Is this advice out-of-date, irrelevant to King County, or otherwise inaccurate?

Bring in the data!

1. Utilizing data from 2018 sales…
2. Looking at houses that sold from $100,000 to $2.3 million
3. In King County
4. 27,813 sales
5. Up to 155 variables recorded for each sale

Bring in the data!



Bring in the baseline model!

1. Multiple Regression Model
2. Our model considered the following predictive features, to address the experts' claims :
    a. Total square feet of living space
    b. The size of the lot
    c. How much of the lot the house took up
    d. If it was a duplex, or not

Bring in the baseline model!

1. Multiple Regression Model
2. Our model considered the following predictive features, to address the experts' claims :
    a. Total square feet of living space
    b. The size of the lot
    c. How much of the lot the house took up
    d. If it was a duplex, or not

R^2 Value of .317…. Seems tough to address the claims with this model fit

Bring in the advanced model!

1. Multiple Regression Model
2. Many iterations considering things like bed and bath count, porches and decks, waterfront location, and many others…
3. Finally included:
    a. Total square foot living space
    b. The footprint ratio
    c. If it was a duplex or not
    d. Different Views
    e. And… What zip code it was in!

Bring in the advanced model!

R^2 Value of .720, so it seems like we can test some of those expert opinions….

1. Higher square footage increases home sale price
2. The house filling a higher proportion of the overall lot decreases home sale price
3. The cost per square foot is lower in duplexes than in single-family homes

1. Higher square footage increases home sale price

The total square foot of living space has a coefficient of 196.9, standard error of 1.7, and a P score ~0.

Therefore we can reject the null hypothesis and we support this claim as being valid advice!

2. The house filling a higher proportion of the overall lot decreases home sale price

The Footprint Ratio predictor has a coefficient of -129100, standard error of 13200, and a P score ~0.

Therefore we can reject the null hypothesis and we support this claim as being valid advice!

3. The cost per square foot is lower in duplexes than in single-family homes

The Footprint Ratio predictor has a coefficient of -10050, standard error of 10050, and a P score ~.340.

Therefore we fail to reject the null hypothesis and it appears that this may be inaccurate advice!

Predictions

$515,000 predicted vs.
$519,000 actual

$1,255,000 predicted vs
1,395,000 actual

Next Steps

Make a predictor function that takes in all of our final model features

Make a model for each zip code, since zip code is such a strong predictor. How do the other features predict within a given zip code?

Normalize the features to zip code

Use address to create a finer resolution geographic location class than zipcode

Zillow is wrong and they should hire us