

# Databnb

mdick, akitatak, jmarkey, ykim117

## Hypothesis

We hypothesize that when the real estate market is successful in a particular neighborhood, the Airbnb market will also be successful in that neighborhood.

## Data

We collected our data from Inside Airbnb and Zillow for ten cities from January 2015 to January 2020. Inside Airbnb periodically scrapes Airbnb and posts the data online for public use. Zillow is the leading online real estate marketplace and uses their own real estate listing information to generate estimates of average housing and rental prices. Both of these data sets need to be cleaned to remove Nan values, uniformize the city names, and parse data values to the proper data type for analyses.

## Findings

**Claim #1:** Success in the real estate market does not have a significant effect on success for Airbnbs in a particular neighborhood.

**Support for Claim #1:** Given only information about housing and rental prices, our multiple linear regression was unable to account for significant variance in the number of Airbnbs across different zip codes (R-squared value of 0.129). We have attached the results of our scatter plot as well. This demonstrated a weak positive correlation that lead us to run the regression.

Year	Correlation	P-Value
2015	0.4392	3.999e-22
2016	0.4979	1.64e-25
2017	0.5126	1.03e-31
2018	0.4375	7.28e-24
2019	0.4267	7.76e-23
2020	0.4046	2.26e-20



**Claim #2:** Rather than one influencing the other, both real estate success and Airbnb success are determined by neighborhood desirability, an attribute not fully captured by our data.

**Support for Claim #2:** When we added a dummy variable encoding zip code, our model accounted for more variation in the number of Airbnbs (R-squared value of 0.927). This suggests that information about real estate is not enough on its own to predict Airbnb success, and instead there is a localized feature to each zip code that determines Airbnb and real estate success.