Zillow Housing Data Analysis

BY: RACHEL SPIRO

OVERVIEW

- **The Situation:** A real estate investment firm would like us to run an analysis on a Zillow housing dataset to determine where they should invest. This data includes information about the average home price in each county in the U.S. over a 22 year time period (1996-2018).

- **The Task:** Determine what are the 5 best zip codes to invest in and what are the forecasted 10 year average home values for each of these zip codes.

- **The Solution:** Perform a time series ARIMA analysis that models and forecasts future average home values per county.

METHODOLOGY

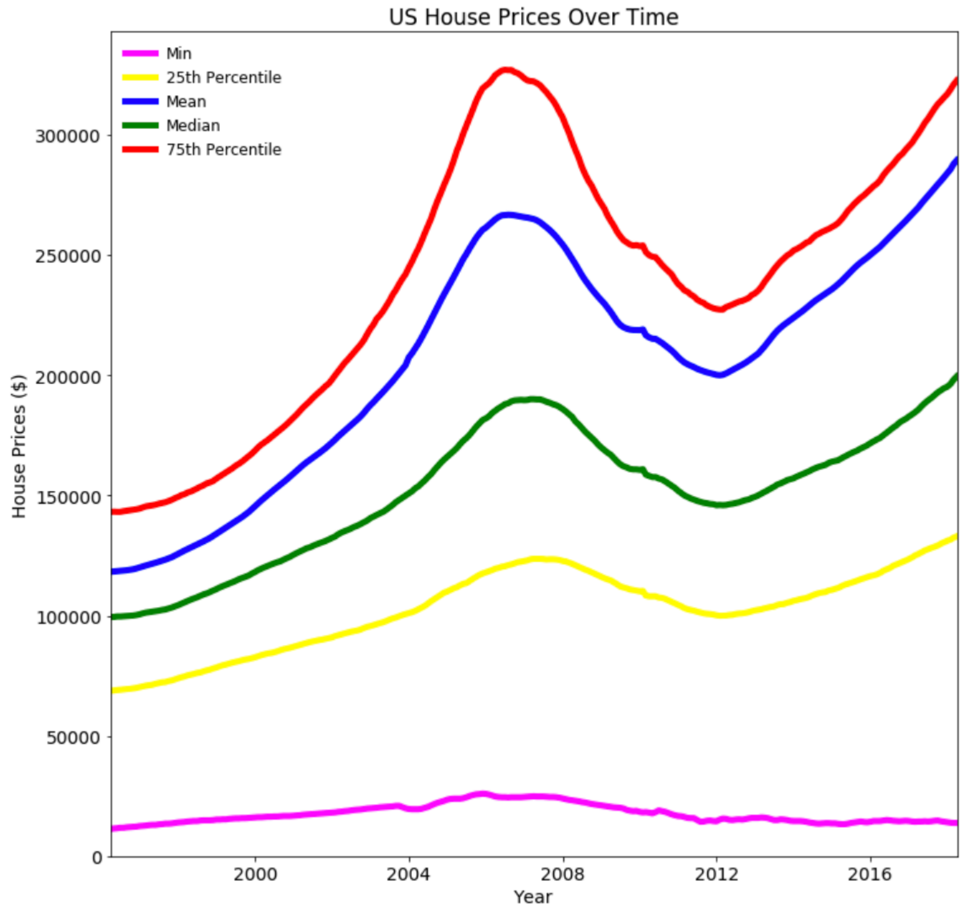
- **Time Series Data** – dataset where the data is captured over the progress of time.
  - Our dataset includes the value of homes ($) in each month from 1996-2018.

- **ARIMA Model** – combines an auto regressive and a moving average models to explain a given time series based on past values or predicts future values. Also accounts for any trends present.
  - Fit model and forecast future values for each region
  - Models based only on previous average home values, no other factors
  - Does not account for crazy exogenous event (i.e. hurricanes, financial crisis)

MARKET AT A GLANCE

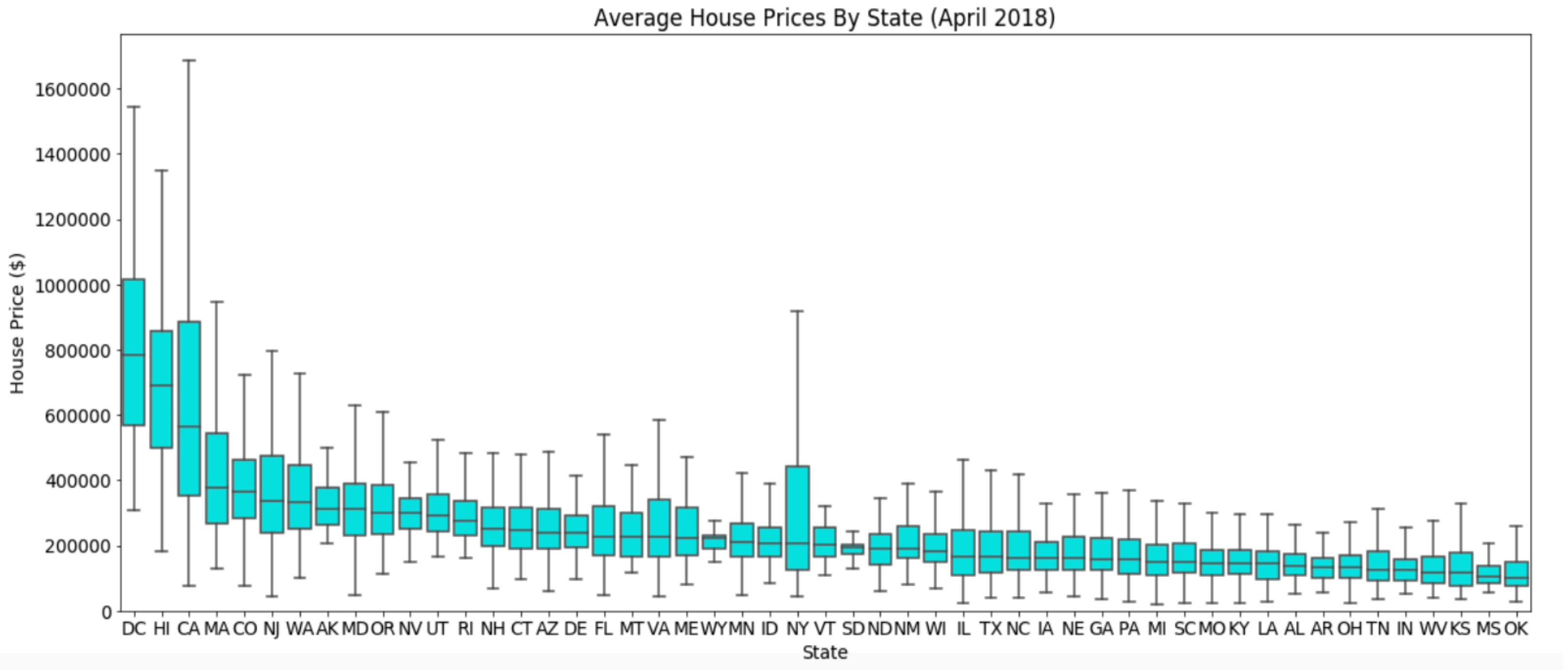
- House prices have varied over the years, mainly as a result of the housing crisis that hit in 2007. Since 2012 though, house prices have been on a steady incline.

- Average home prices in the 75th percentile in recent years are over 20 times higher than the minimum home values.



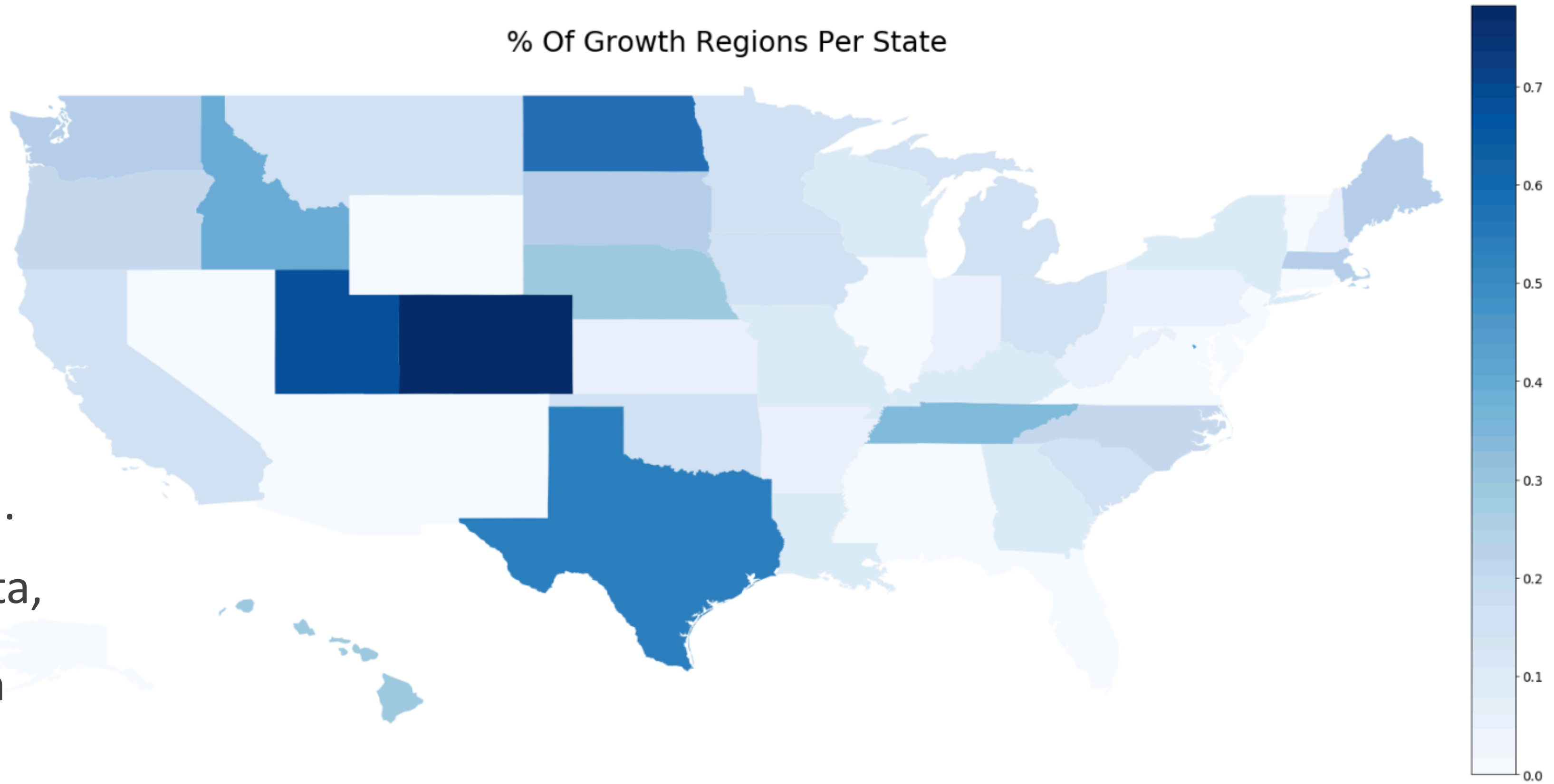
MARKET AT A GLANCE

- In April 2018, DC, Hawaii, California, Massachusetts, and Colorado had the highest median house prices.
- We also see some outliers present in these states along with New York.



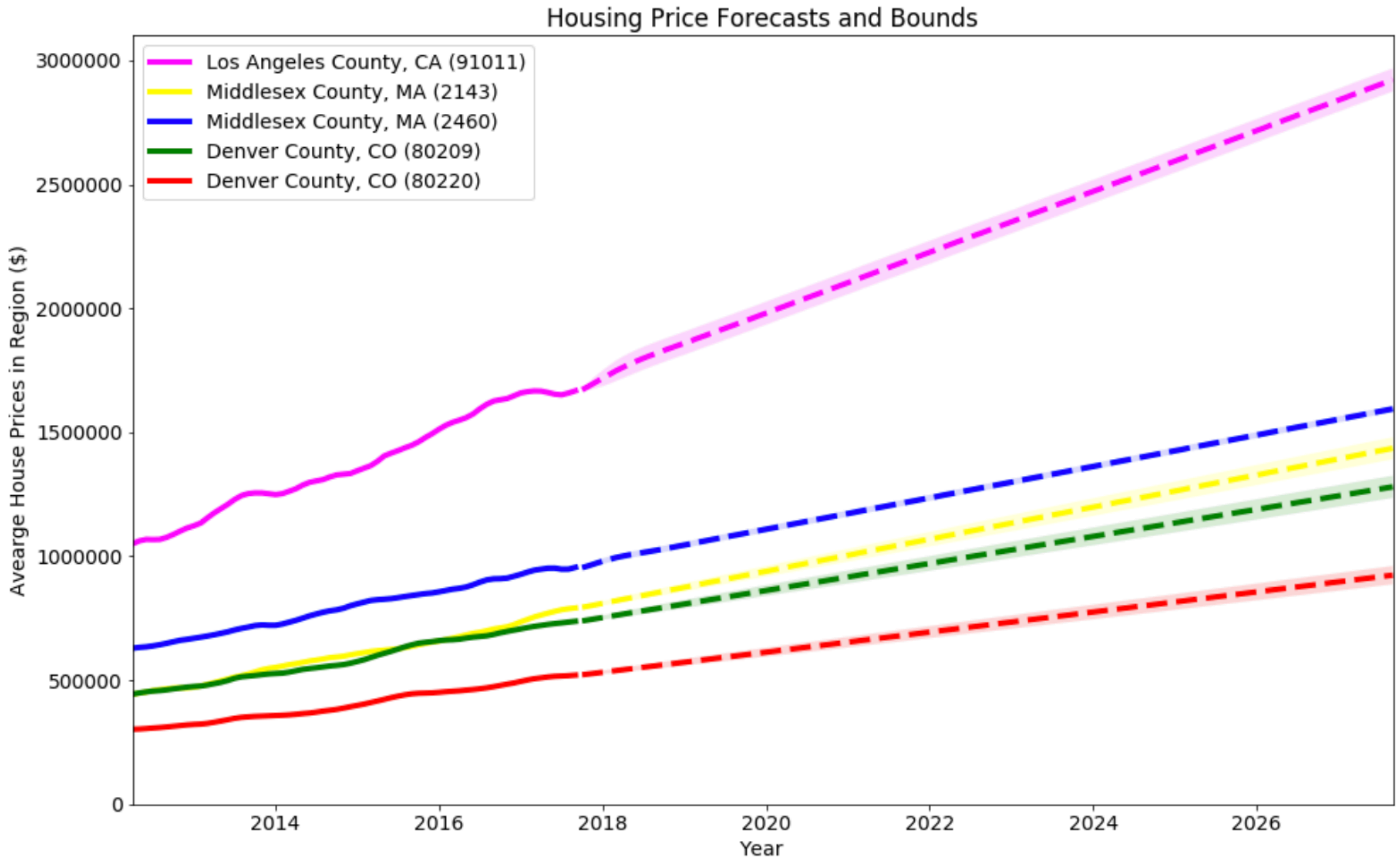
SELECTING REGIONS

- 15% of the 14,500 regions were classified as 'strong growth regions.'
- Strong growth regions – defined as having above average annual growth rate both during the housing recession (2007-2012) and since the recovery of the housing market (2012-2018).
- Colorado, Utah, North Dakota, and Texas have the largest percentage of strong growth regions.



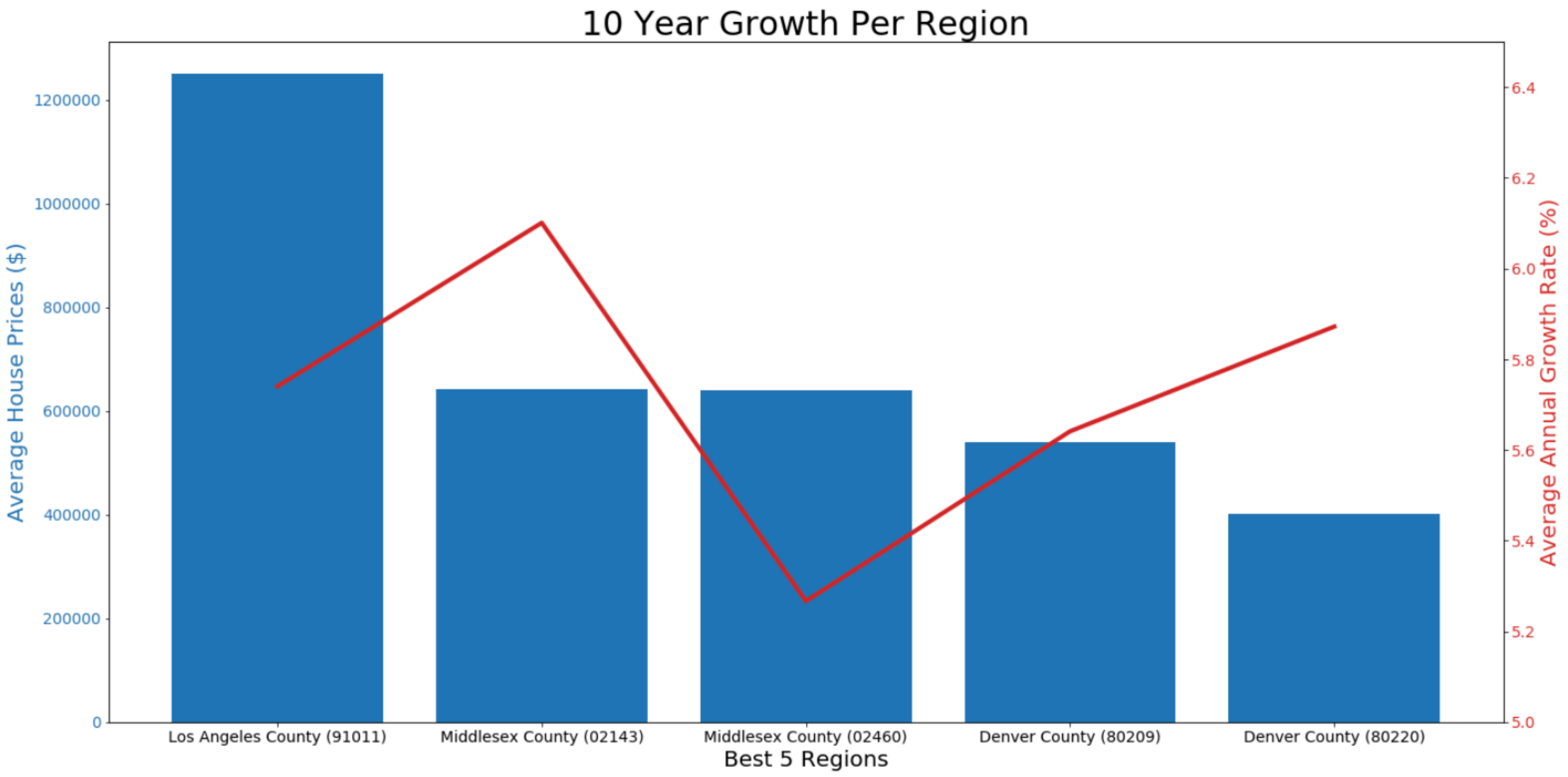
TOP 5 REGIONS

- 5 regions were identified as the best regions to invest in.
- To qualify as a ‘ best region,’ the region must:
  - Qualify as a ‘strong growth region’
  - Be in top 25% of both 5 year and 10 year forecasted average annual growth rate
  - Be in the smallest 50% of prediction interval widths
  - Have a maximum p-value less than alpha=.05



POTENTIAL RETURN

- Los Angeles (91011) has the greatest forecasted 10 year value of $1,249,665.
- The other 4 top regions have the potential to accumulate ~$500,000-$640,000.
- All regions have strong average annual growth rates over a 10 year period.



NEXT STEPS

This analysis was done just looking at the average home values of each county over the 22 year period. There are many other factors that could also play into the potential for investing in these regions. Therefore, my next steps would be to look into some of these other factors, such as:

-Potential for severe weather (i.e. hurricanes, tornadoes)

-Strength of school district

-Crime rate

-Age of homes