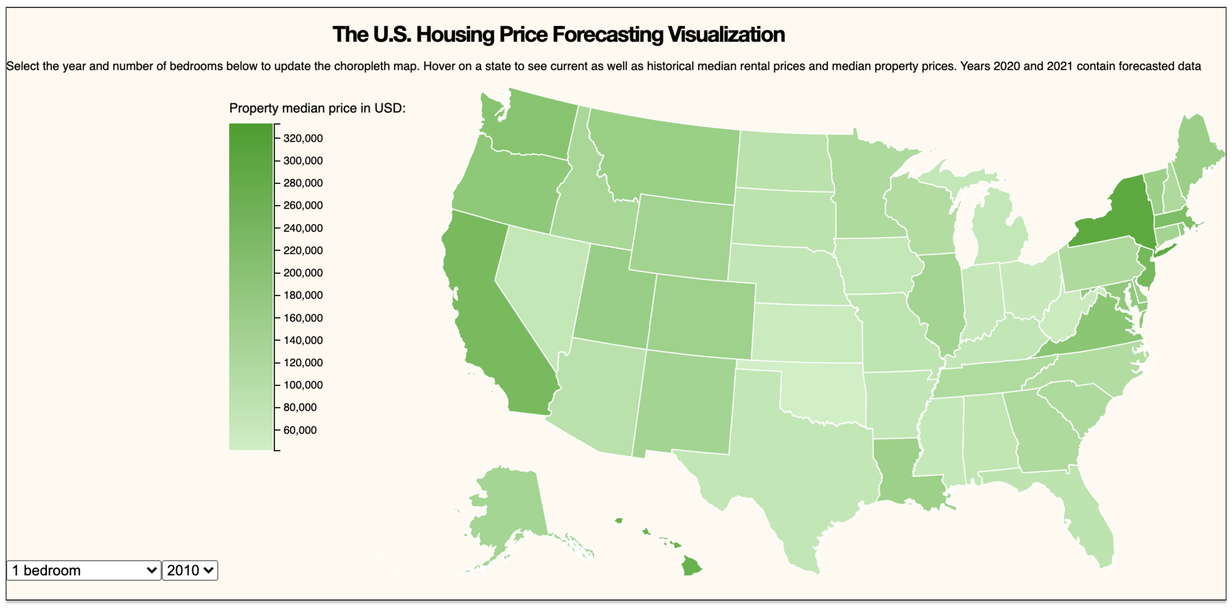
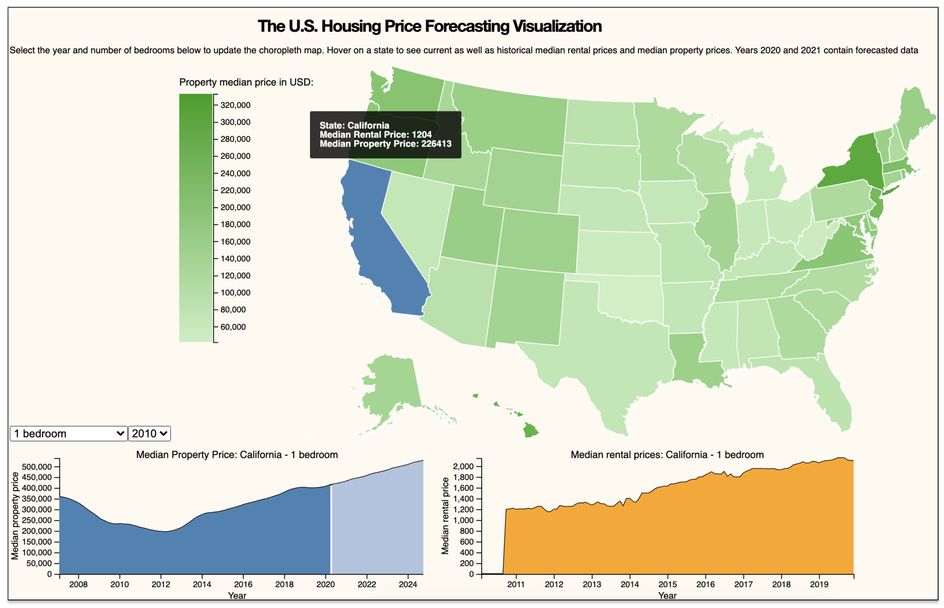
# United States Real Estate Visualization

Team#4 The Underdogs: William Lusty, Leonardo Calizaya, Matthew Pleskow, Haoxuan Huang, Michael Cho, Jun Hyuk Jeon




## Introduction

Housing is a huge driving factor of the economy. In 2011, the total value of residential properties in the US was $16 trillion. In 2020, it was worth $36.2 trillion, meaning that it more than doubled over the past 9 years. Any American who owns a home would be interested in how the market will change in their area. Furthermore, investors looking for the hottest place to buy real estate would want to know where they should grow their business and where they should avoid.

Our team sought to create a web-based real estate visualization tool that allows users to explore historical data and forecasted future data, increasing their market knowledge and savvy. This type of tool that combines machine learning prediction and cutting-edge web based visualization has never been created before for real estate data. Other products on the market visualize real estate data on the local level (Zillow) but do not attempt to forecast changes to the market in the future. No other popular visualizations exist that visualize real estate trends on the national level and leverage machine learning to forecast future trends.

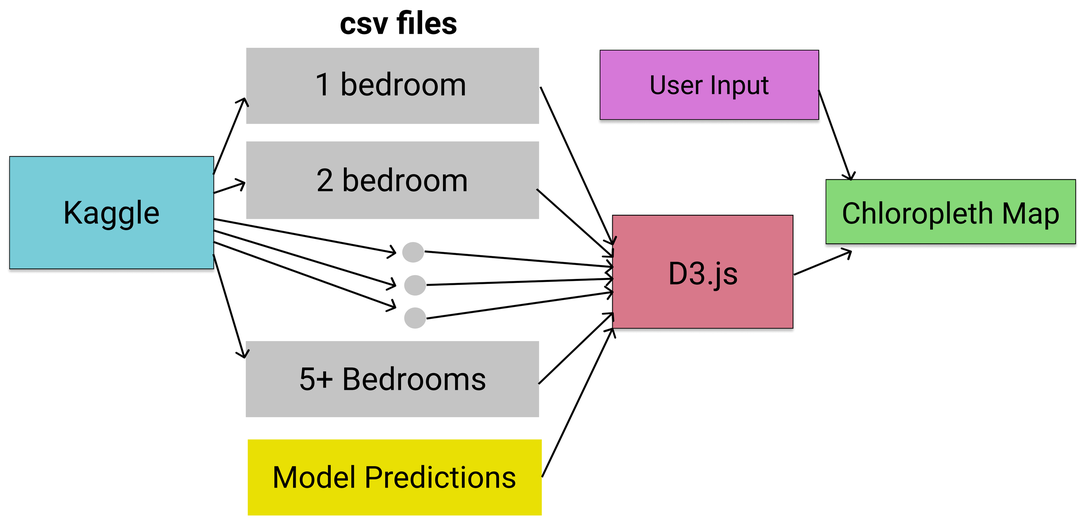
## Approach

- Choropleth map displaying median rental prices
- Colored according to linear scale compared to other states
- User chooses year and number of bedrooms
- User may investigate historical data or forecasted data
- ARIMA model uses machine learning time-series analysis to predict future prices
- Tooltip displays details for the state and a line chart of prices over time
- Implemented in Javascript and HTML using D3

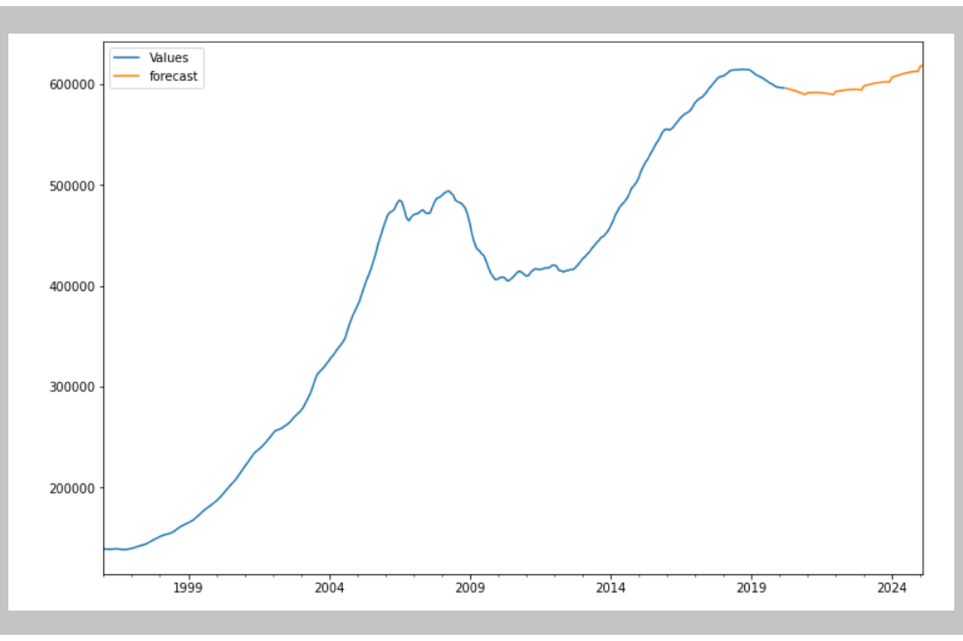
## Data

- Downloaded csv files from Kaggle
- Time series ZHVI(Zillow Home Value Index) from 1996-2020
- Time series median rental price from 2010-2020
- Size: 199.1 MB
- Number of Observations: Approximately 105,000

## Application Schematic

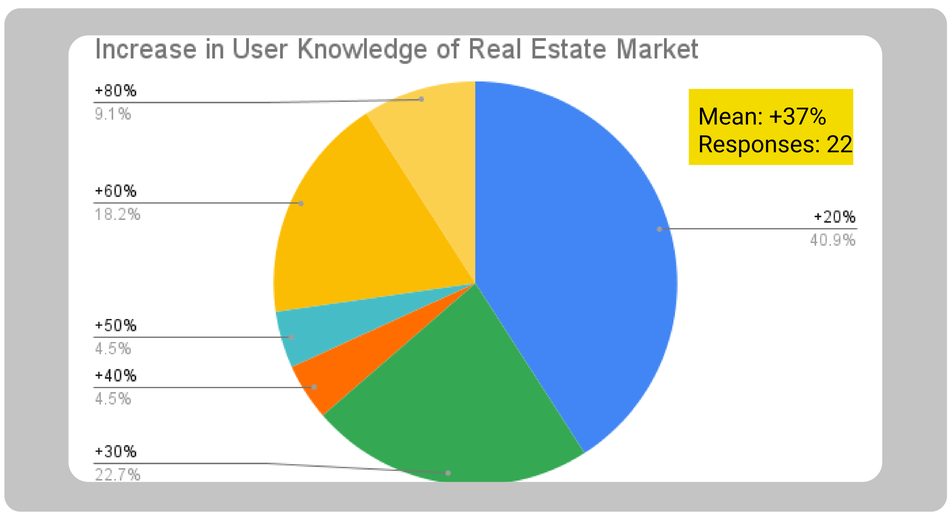



## Price forecast - ARIMA




Example of ARIMA model forecast on a sample (New York, NY)
Blue line - Real price history, Orange line - Forecast

## User Impact Survey

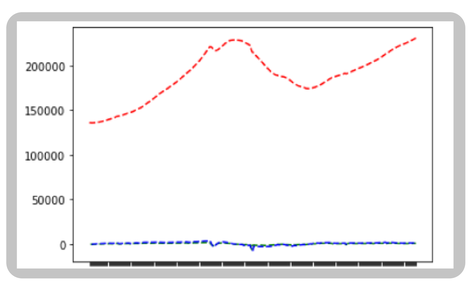



## Process of Data

Our group decided to make the choropleth map that shows color gradient for each state to represent the history and forecasted property price. We believed that users can benefit more if they can look through each state instead of each city to find useful information.

So, we grouped each data observation by its state name and the room size, and then took the average to create our forecast and visualization.

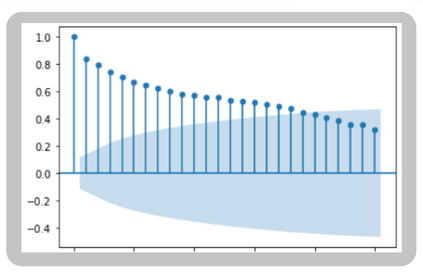
## Parameter Selection



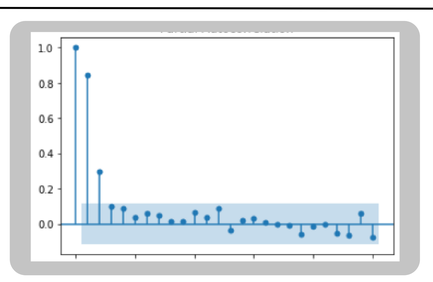

Red line - first order differential, blue line - second order differential

To utilize the ARIMA model for our forecasting, we needed to select appropriate parameters. For ARIMA, there are three parameters existing, which are (p, d, q).

Firstly, we looked at d, or degree of differencing. We observed a significant improvment from the original data to the first order differential, while we observed nearly no difference between the first and second order differential. Therefore, we select d = 1.




ACF Plot on data




PACF Plot on data

Secondly, we generated ACF and PACF plots in order to decide p(lag order) and q(order of the moving average). Since ACF shows a steady decrease of value while the PACF shows rapid decrease after lag 2, we concluded that our time series data resembles the feature of AR model and thus selected p = 0 and q = 1.

Also, we figured that the seasonality exist in the data after some testing, so we decided to add Seasonal parameter to create forecast. We set our seasonal parameter to be (1, 1, 0, 12) since the ACF is positive at the first significant lag, series has stable seasonal pattern over time, and observations were made monthly.

In final, our arima paramater for forecast is (0, 1, 1) (1, 1, 0, 12), and we were able to create a convincing forecast from 2020 to 2025 based on the data from 1996 to 2019.