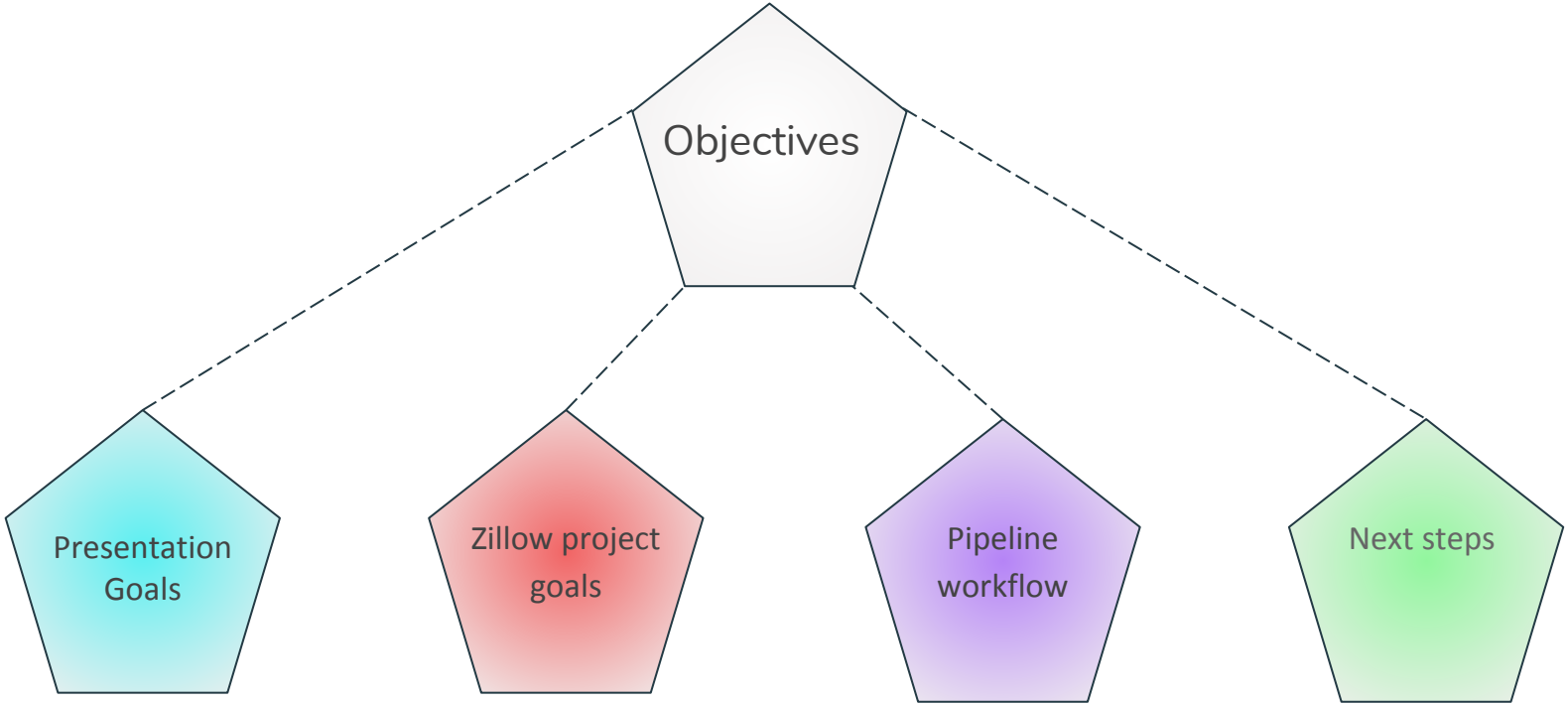
Zestimate

Tim Sotirhos


Objectives
Presentation Goals
Zillow project goals
Pipeline workflow
Next steps

Zillow Project Goals

Create an estimator tool to accurately predict home values using property data from California during select months in 2017

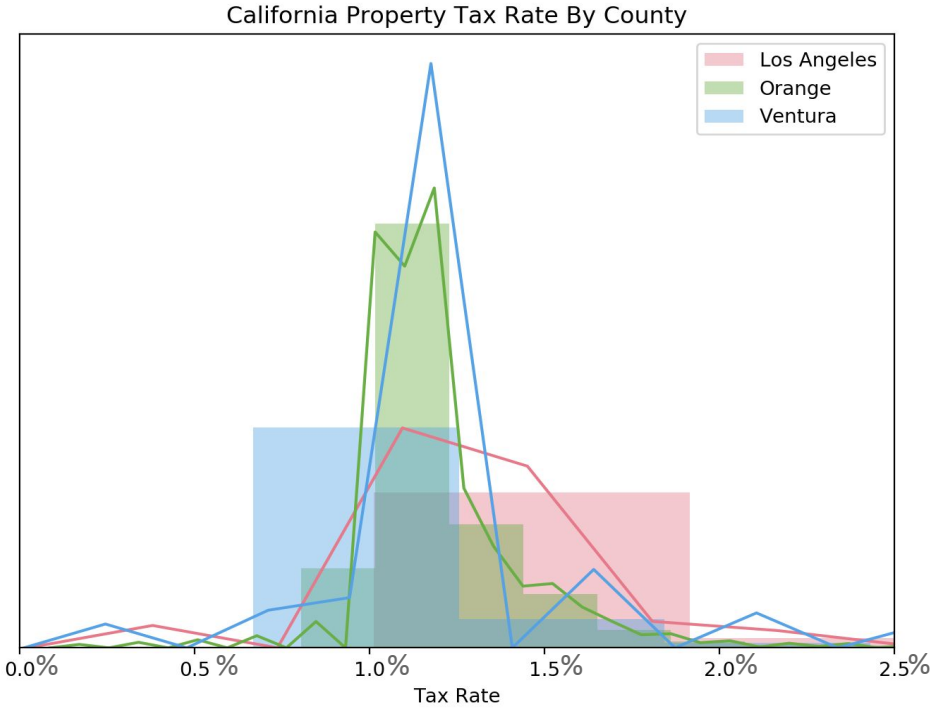
Analyze the distribution of tax rates by county in order to understand how much they vary within the county and the rates the bulk of the properties possess

Build a predictive model which uses number of bedrooms, number of bathrooms, and square footage to estimate the property value

Distribution of tax rates by county

1. Identify counties in the data set

2. Key takeaways
   a. Los Angeles highest values (skew left)
   b. Orange most centered
   c. Ventura most values (skew right)



Pipeline workflow

Acquire - obtain data from SQL zillow database

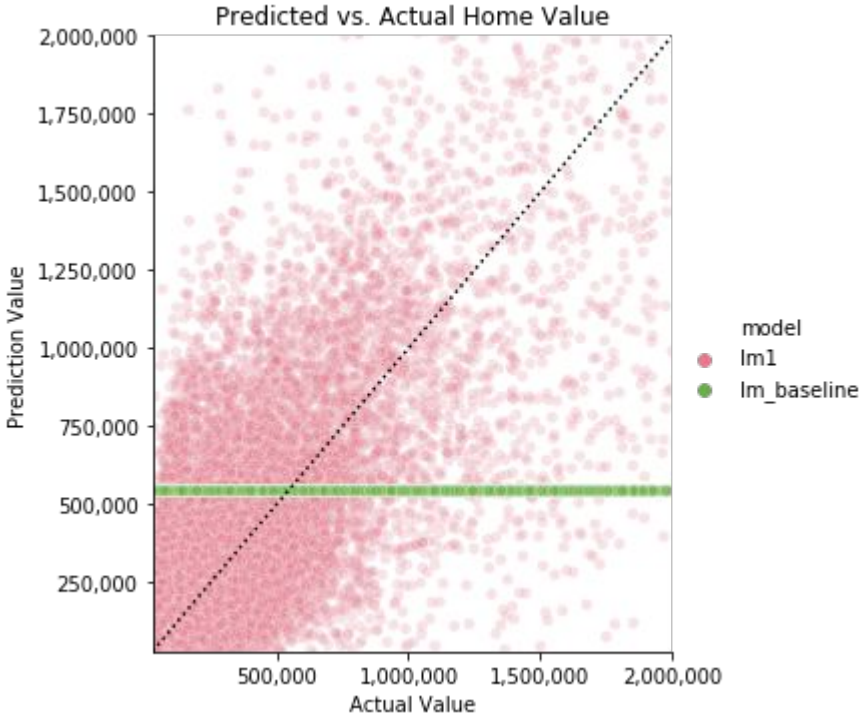
Prep - include only “single family residences” with a “bedroom count” and “bathroom count” greater than zero

Split & Scale - split data into a train size of .80, test size of .20 and scale all selected variables using a Standard Scaler

Data Exploration - visualize correlation and analysis results

Modeling & Evaluation - create baseline using target’s mean and one linear regression model to compare evaluation metrics

	Prediction Value	Property Value
0	598488.582089	534030.0
2	212487.769513	185287.0
4	331318.736508	288579.0



Conclusion

- Best fitting line - baseline model vs. linear regression model
- Use the end of this project as preliminary work for future studies for a more comprehensive predicting system
- Additional feature engineering, additional testing to identify better performance, and more advance modeling