

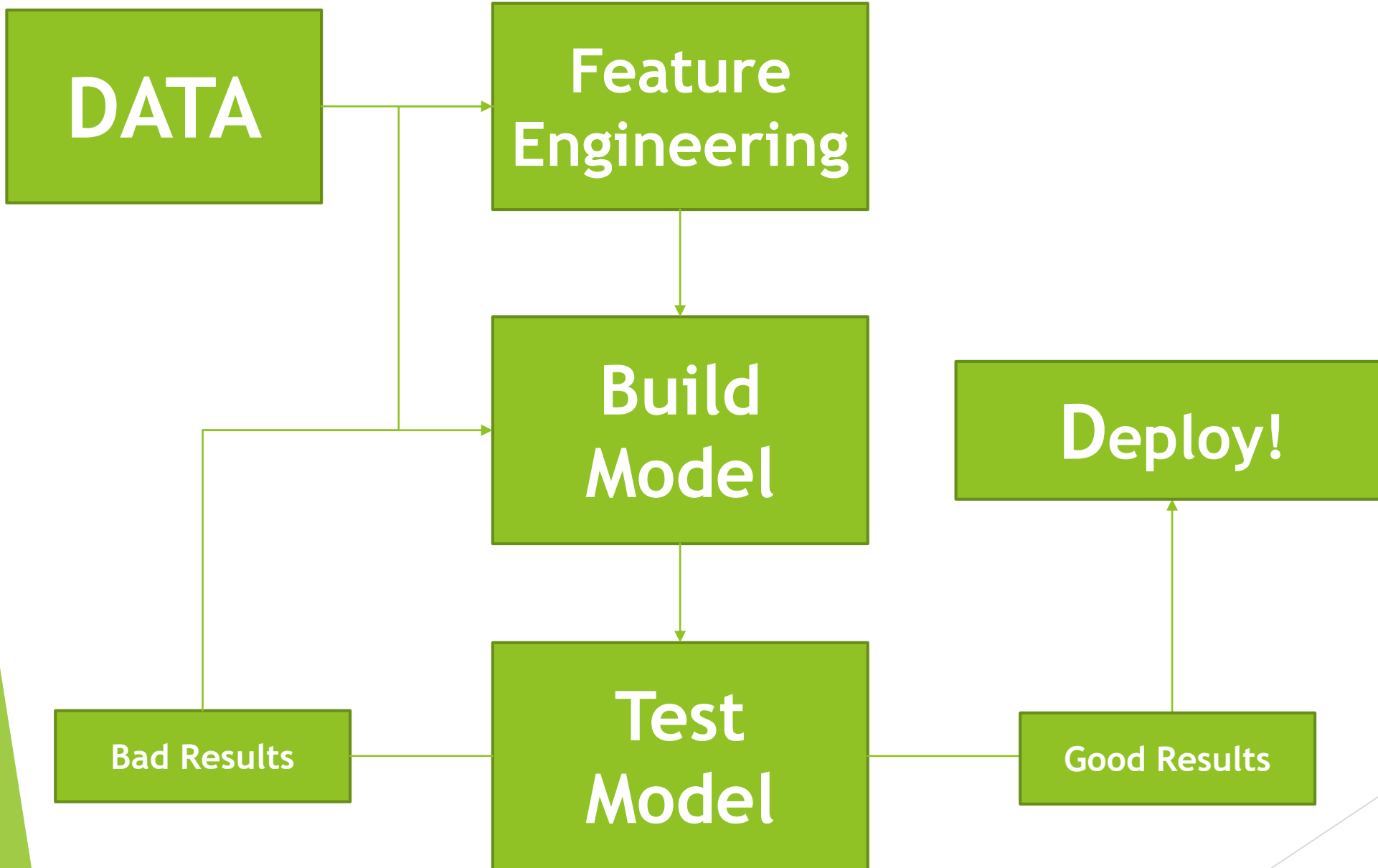
Linear Regression Modeling

King's County Housing



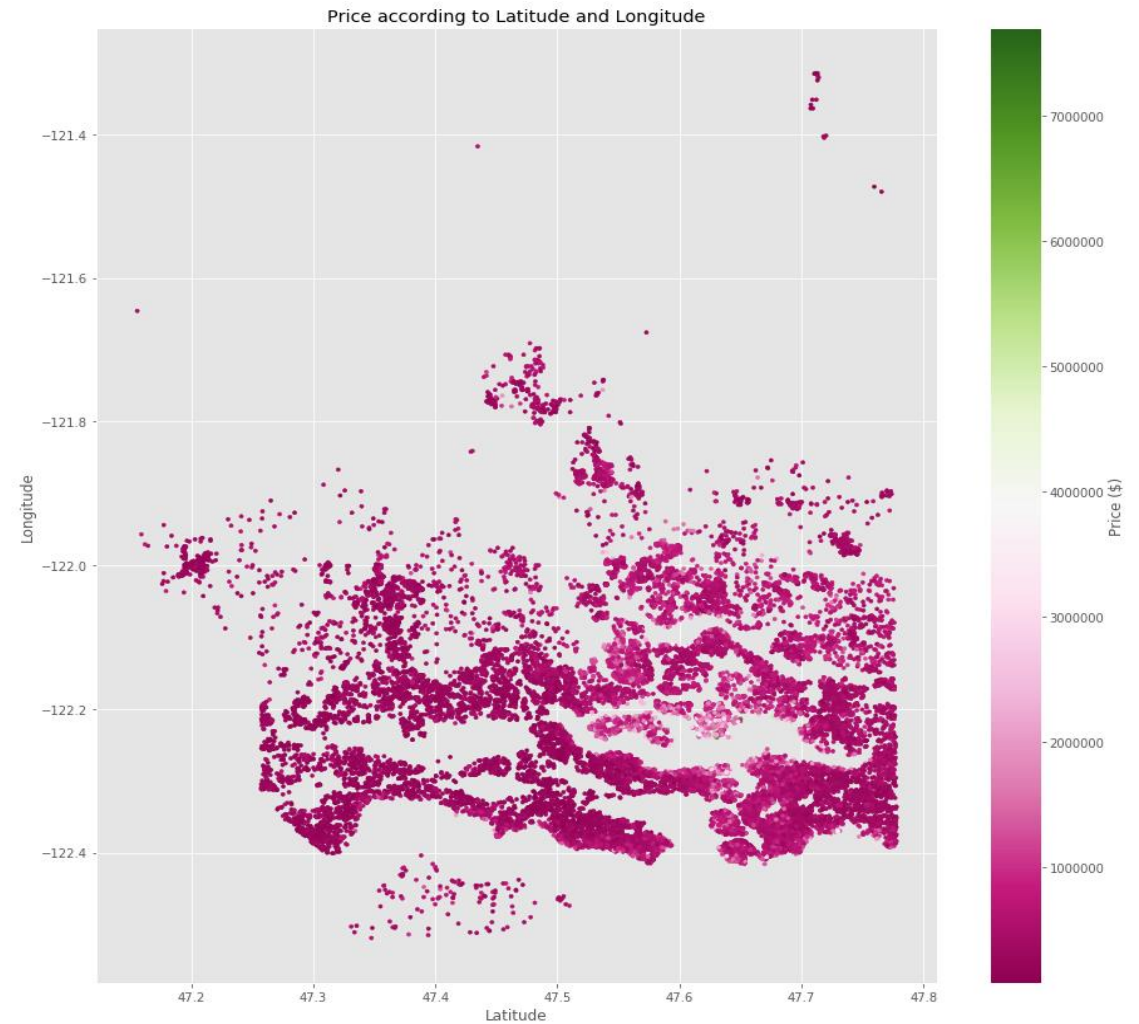
The Data

- ▶ Target:
 - ▶ Goal is to predict: **House Price**
- ▶ Predictor Variables:
 - ▶ Number of Bedrooms, Bathrooms, and Floors
 - ▶ Sq. footage - Internal and Lot
 - ▶ Overall Condition (1-5)
 - ▶ Geographical Location (Based on Clustering)
 - ▶ Waterfront (yes / no)
 - ▶ Year Built
 - ▶ Renovated (yes / no)



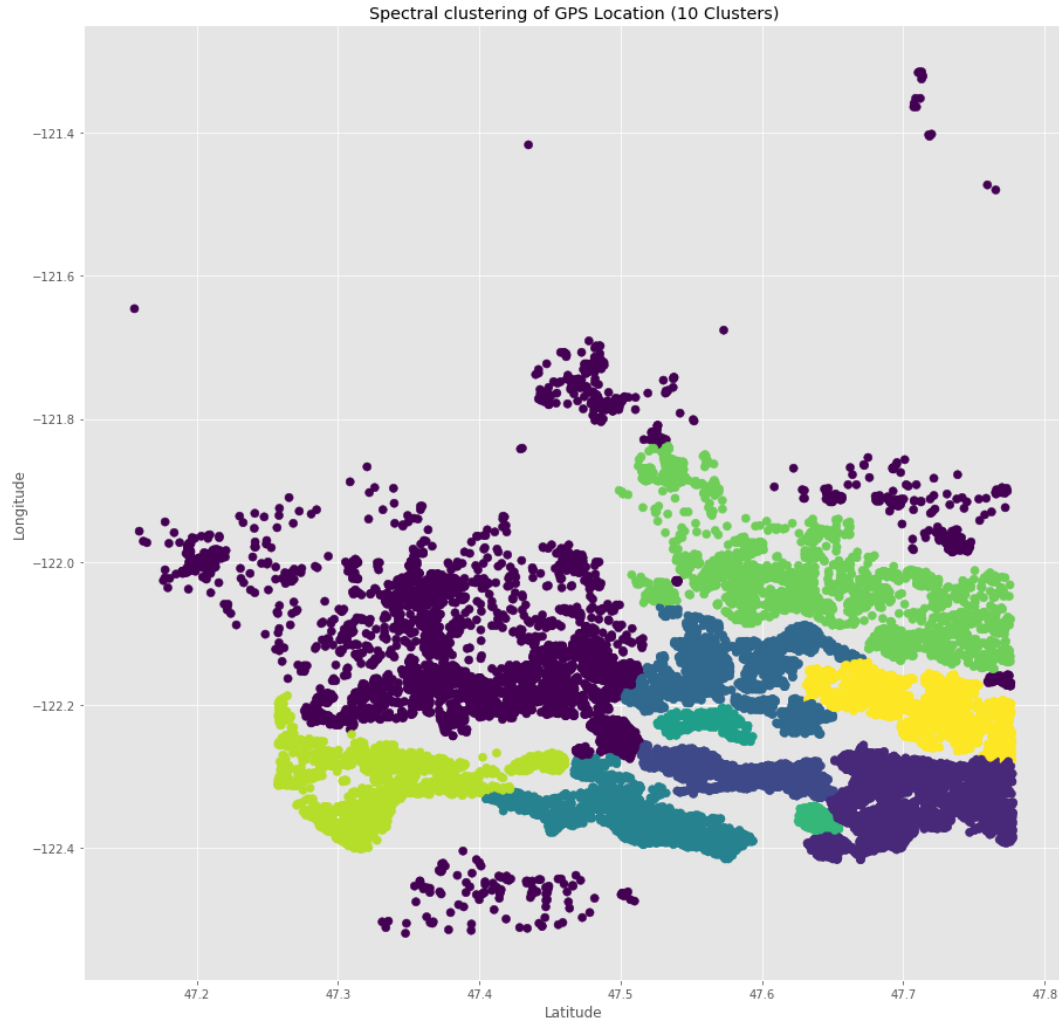
Problem: Dealing with Location Data

- Price depends on exact location
- Difficult to use latitude, longitude, zip code in linear model
- Still want to capture affect of location in model



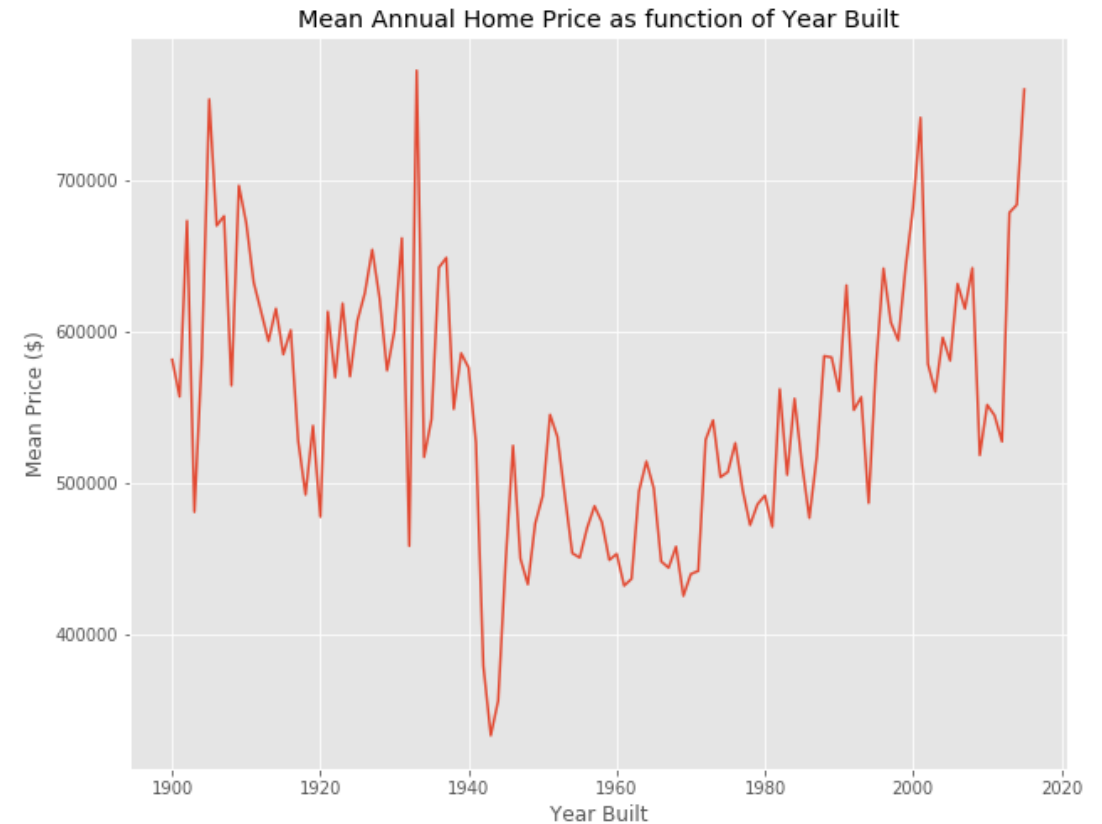
Solution: Spectral Clustering

- ▶ Finds patterns in data and separates into groups
 - ▶ Works well with complicated patterns
 - ▶ Gives good way to distinguish geographical affect of price
- ▶ Works well in linear model
 - ▶ Extracts meaning from location data



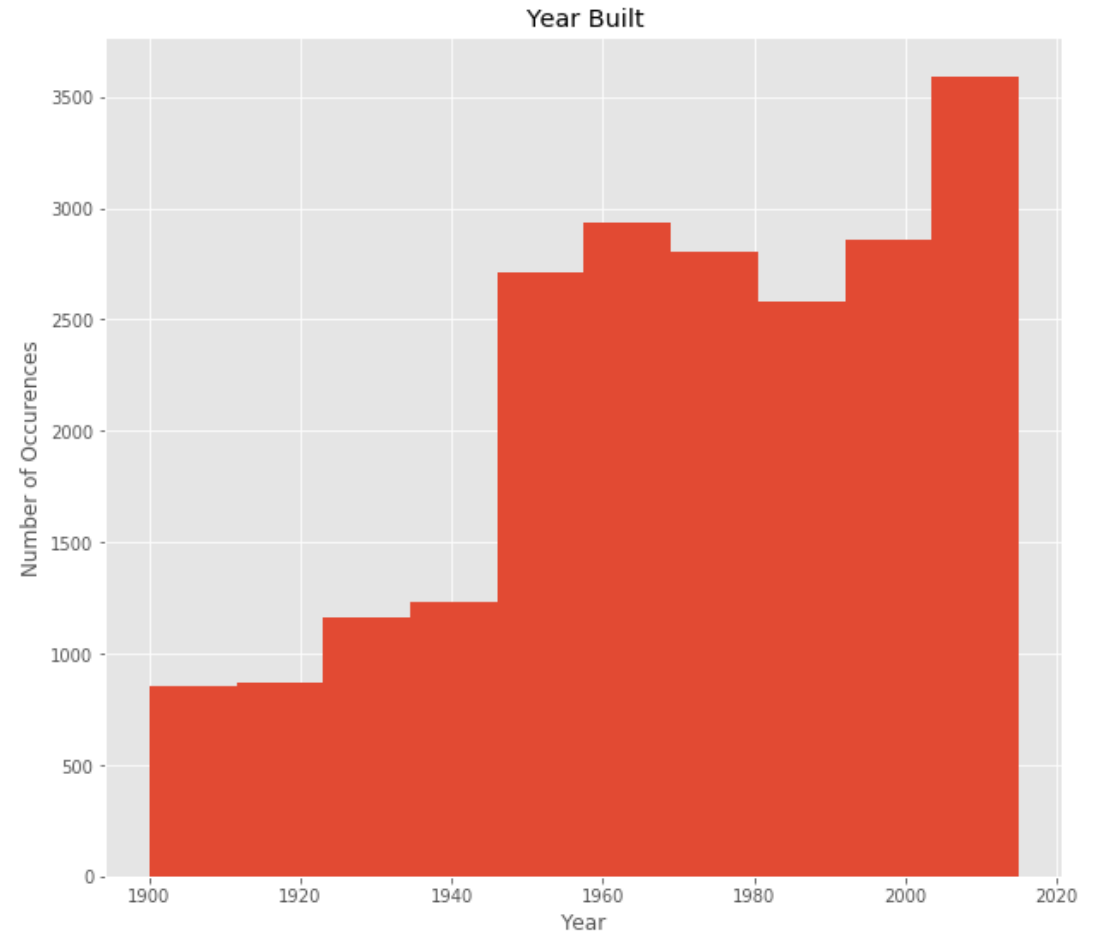
Problem: Dealing with time series data

- ▶ Year built has clear affect on price
- ▶ Won't work well in linear model
- ▶ Still want to capture affects in linear model



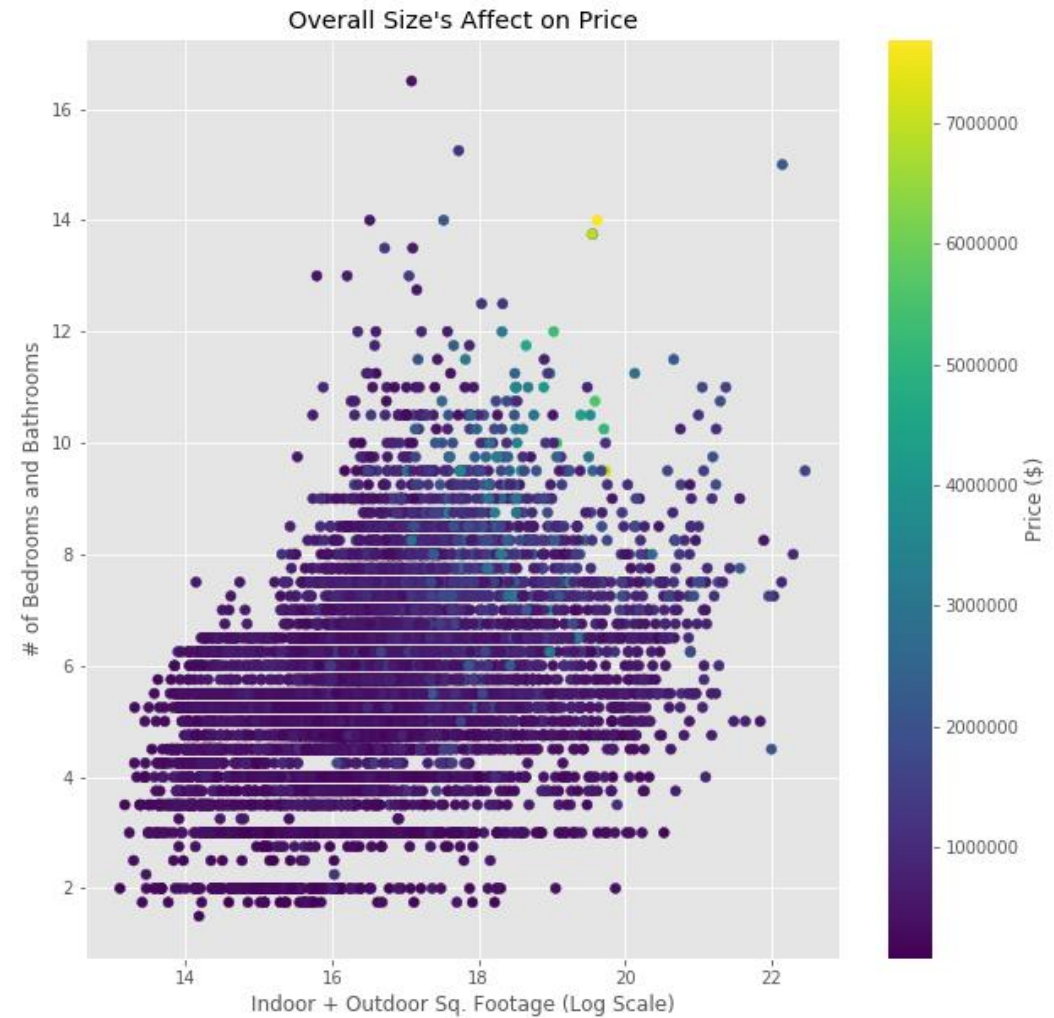
Solution: Binning

- ▶ Reduce complexity: Keep only which decade house belongs to
- ▶ Create new features representing each bin
- ▶ Linear model captures time affect - Performs better



The Model

- ▶ Best Features:
 - ▶ Number of Total Rooms & Floors
 - ▶ Indoor Sq. Footage
 - ▶ Overall Condition
 - ▶ Location
 - ▶ Year Built
- ▶ Achieved R^2 Score of
 - ▶ ~0.76 on Test Set
 - ▶ ~0.77 on Training Set
- ▶ Results:
 - ▶ Housing prices are volatile, so these results are good for a linear model
 - ▶ (See how spread out visual is)



Main Takeaways

- ▶ Feature engineering + Log scaling:
 - ▶ Achieved higher R2: ~17% improvement over raw data model
 - ▶ All features statistically significant - will perform well on unseen data
- ▶ Recommendation:
 - ▶ Keep recording data in same way
 - ▶ Model depends highly on clustering - try many different clusters
 - ▶ Use provided techniques to transform data
 - ▶ Retrain model with more data, specifically with more expensive homes
- ▶ Things to watch for:
 - ▶ As houses get more expensive, price becomes harder to predict