# King County Housing Sales Project

Lorin Helfenstein, Elliot Iturbe, Jacob Hoogstra

Data Cleaning Steps

- Dropped the id column
- Replaced nulls with zeros in the columns "waterfront, view, and yr_renovated"
- Converted waterfront, view, and yr_renovated to floats because they were objects
- Replaced the question mark in sqft_basement column with a zero
- Changed date from an object to a datetime

Three most Correlated Features to Price
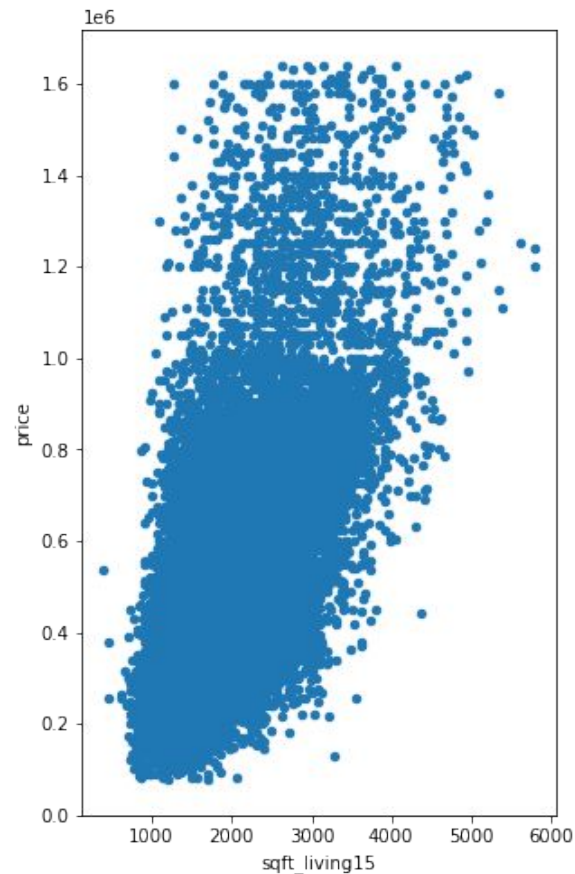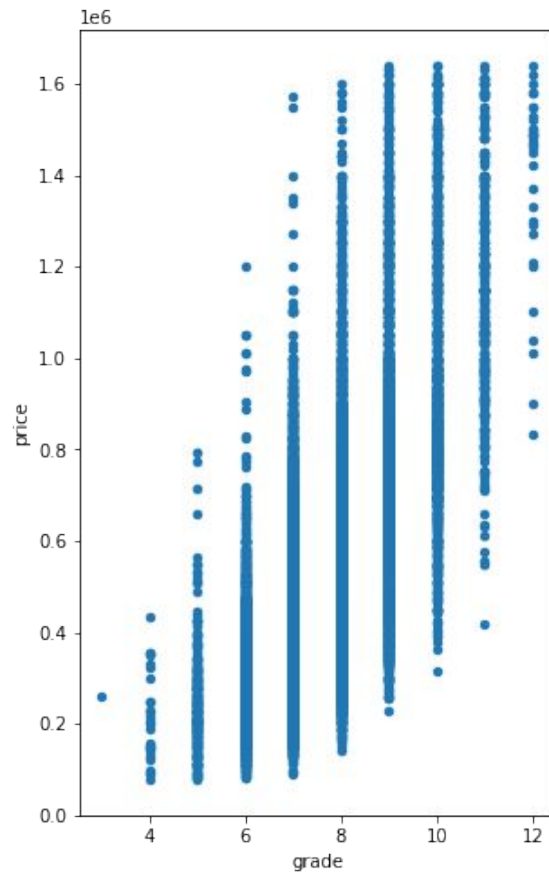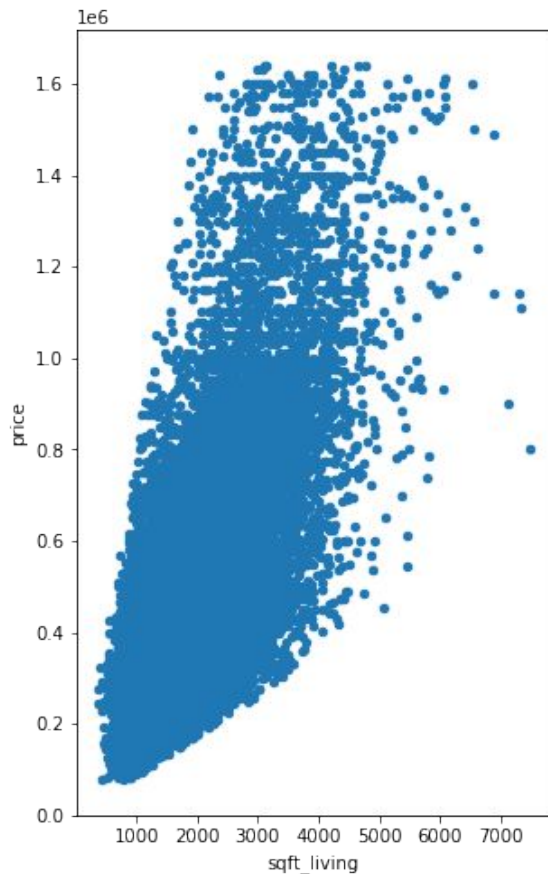
Baseline Model

- Features: sqft_living, sqft_living15, bathrooms, grade
- Target: Price
- Used train test split
- R-squared: .49
- RMSE: 157,108.56
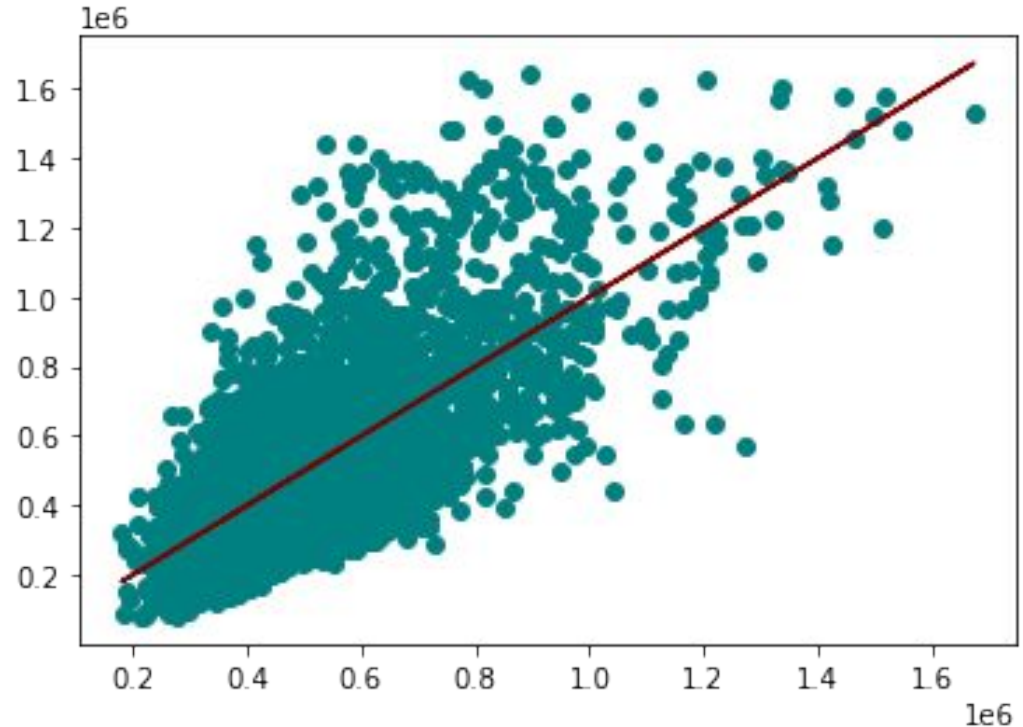
Preprocessing Steps

- Feature engineered a year sold column from date
- Feature engineered a month sold column from date
- Removed major outliers that were more than 3 standard deviations away from the mean of price
- Used StandardScaler
- Log transformed price

Final Model

- Features: sqft_living, sqft_living15, bathrooms, grade, yr_built, lat, condition, long, view, month_sold, zipcode, waterfront, years_sold
- Target: Price
- Then applied polynomial features
- R-squared = .796
- RMSE = .218
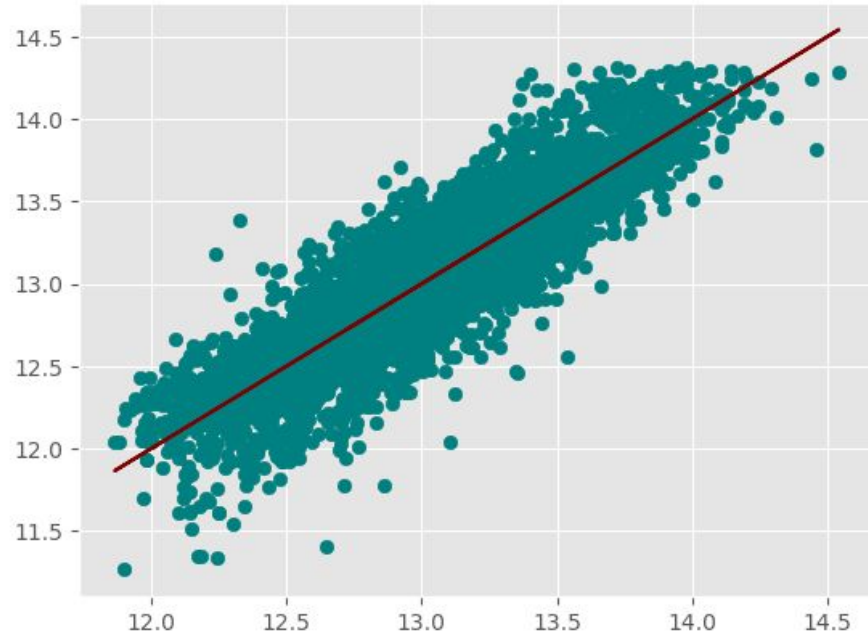
Thank You

Elliott Iturbe
Email: eaiturbe@bsc.edu
GitHub: @eaiturbe

Lorin Helfenstein
Email: lorinh22@gmail.com
GitHub: @lorinhelfenstein

Jacob Hoogstra
Email: jnhoogstra@crimson.ua.edu
GitHub: @jnhoogstra

Questions?