

Selling Sunset...Or Just the Waterfront

Customer Analysis & Insights (June 2020)

Scott Louis Andrew Johnson

Problem Statement:

As a junior data scientist at real estate company PropertiesInc., I have been tasked with **investigating house sales (EDA)** in the King County area and building a **model to predict sale price**. Key executives are keen to launch an advertising campaign directed towards home owners in that area who might consider selling their house, **focusing on higher-end residential properties (high revenue)**.

Business Value

- **Targeted** marketing campaign with right **timing and locations (zipcode/region)** —> high revenue
- Predictive modeling will allow for feature valuation based on historical data and market value validation of “fleeting desires” of potential buyer (aka “a nice view”)

Methodology

OSEMiN

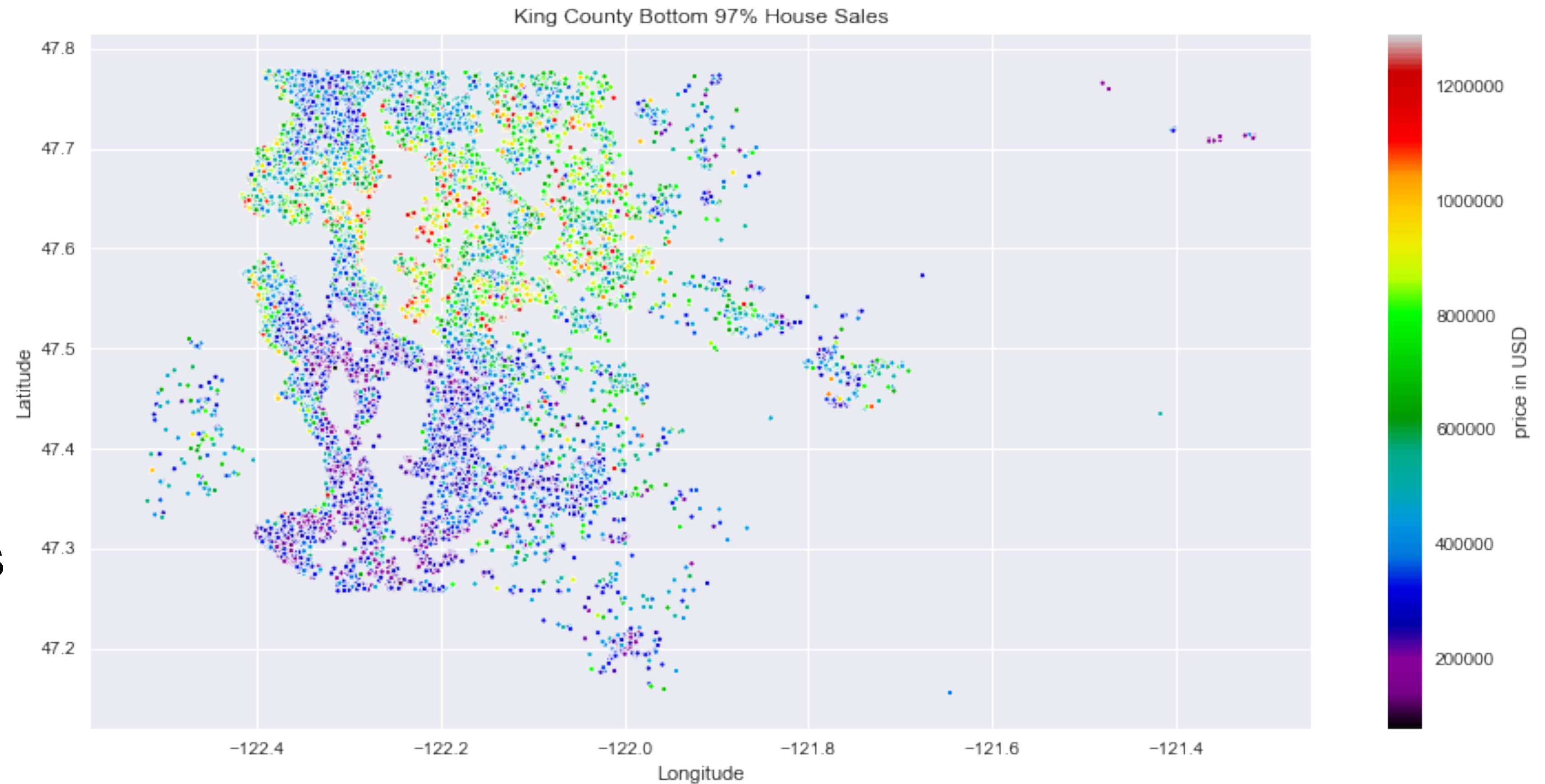
- Obtain
- Scrub
- EDA
- Model
- Interpret



Location

Highest Value Homes

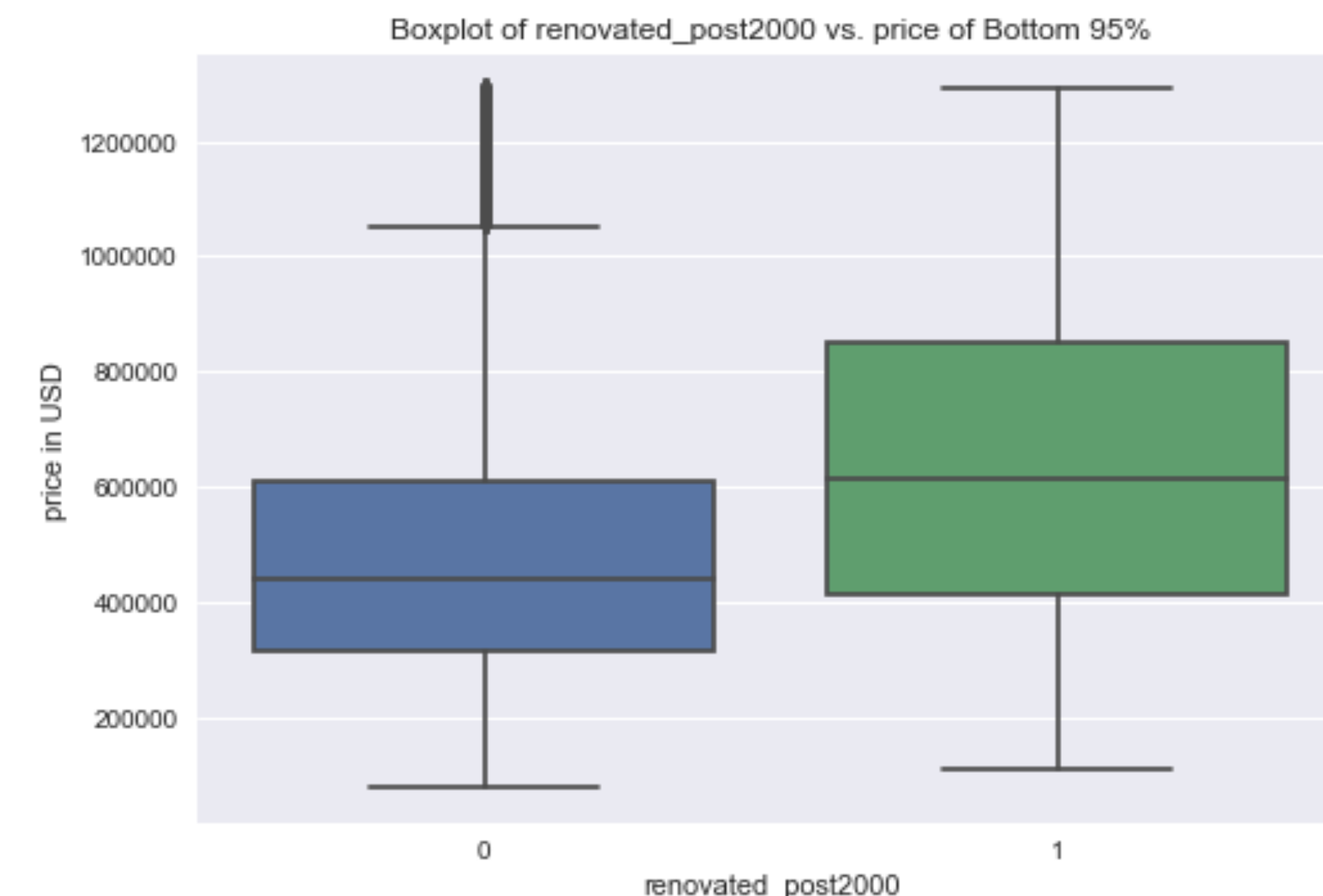
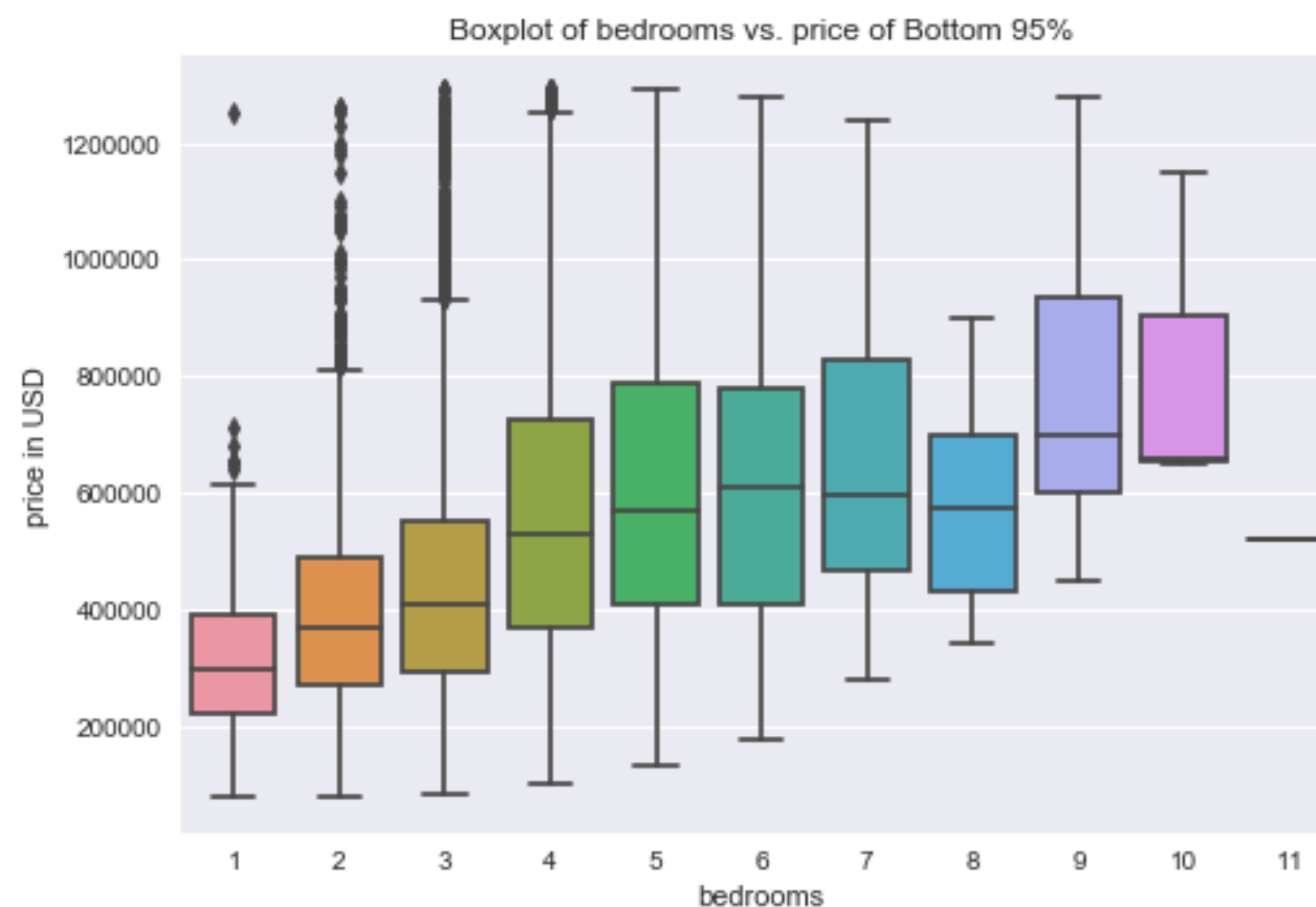
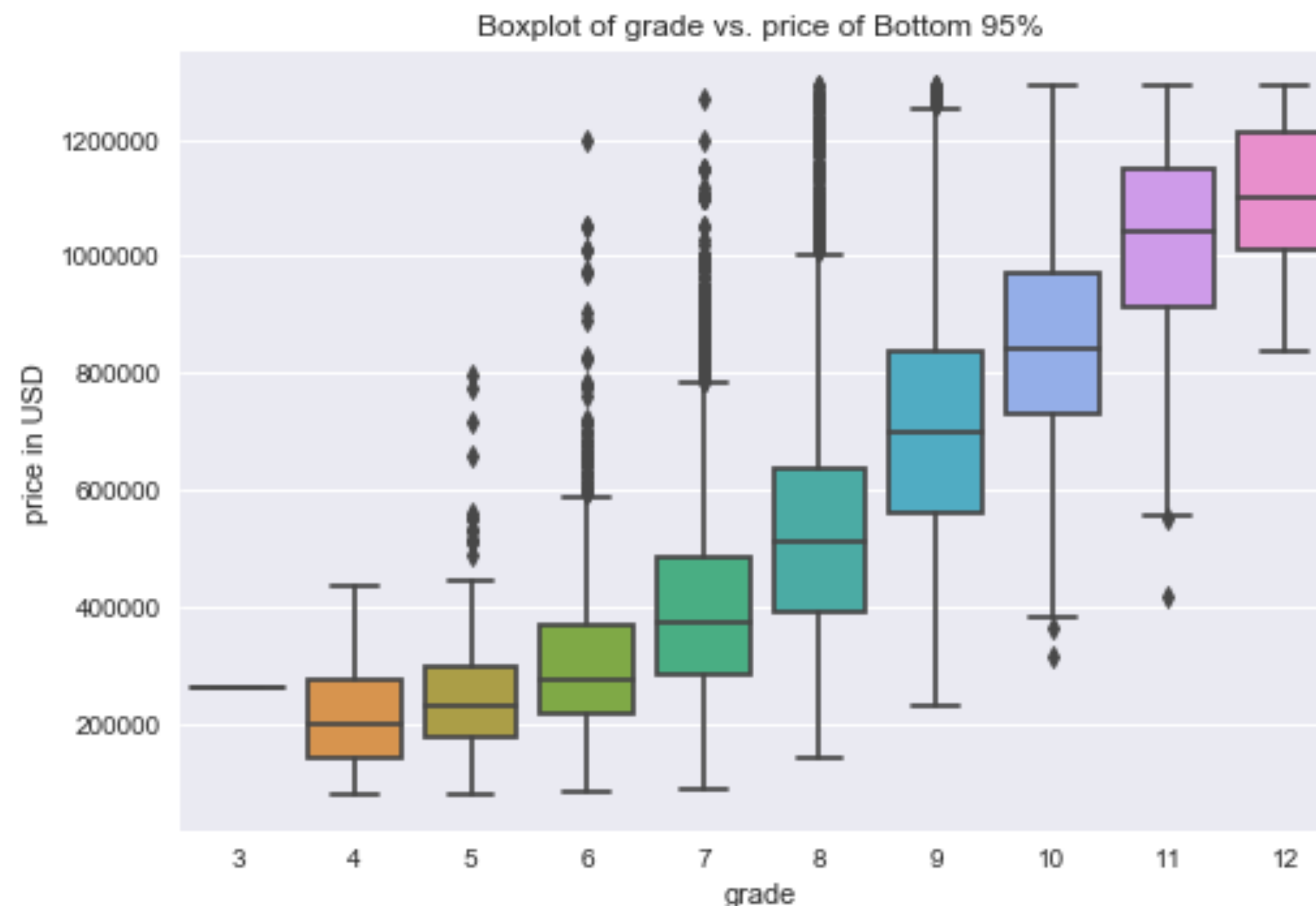
- Near the water front
- Hot spot near lat,long (47.6, -122.2)
- In Tier 1 / 2 Zip codes
 - 14 highest median home value zip codes



Feature RICH

Positive price levers

- Grade (quality)
- Renovations post 2000
- Bedrooms



Feature RICH

Positive Price Levers

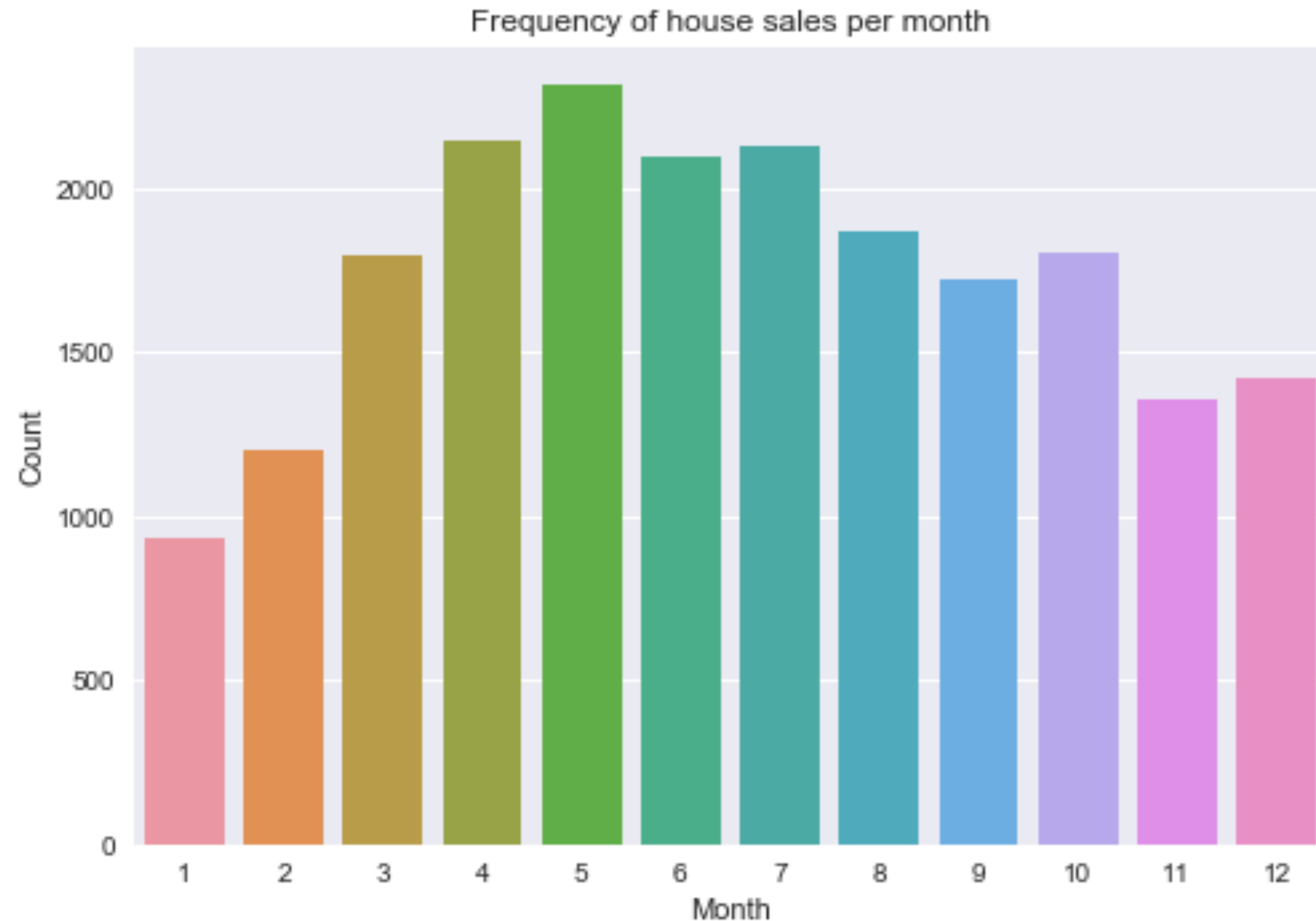
- Living area square footage
- Grade also correlated with living area square footage



Timing is Everything

When to market/sell

- Highest number of houses sold during the beginning of the summer
- Marketing prep Feb/March
- Sell in Q2 (April-July)



Modeling

R-squared represents accuracy

Model for Bottom 95%	Description	Num Features	R-squared
	Model 2 With interactions	39.0	0.789
	Model 1 With zipcode tiers	27.0	0.774
	Base Model One-hot encoding	18.0	0.500
Simple Linear Regression	Sqft_living	1.0	0.414

- Most accurate (Model #2) has an accuracy of 79% and accounts for confounding interactions between parameters
- Second most accurate (Model 1) incorporates tiered zip codes
—> uses less # of features to predict price

Model # 2

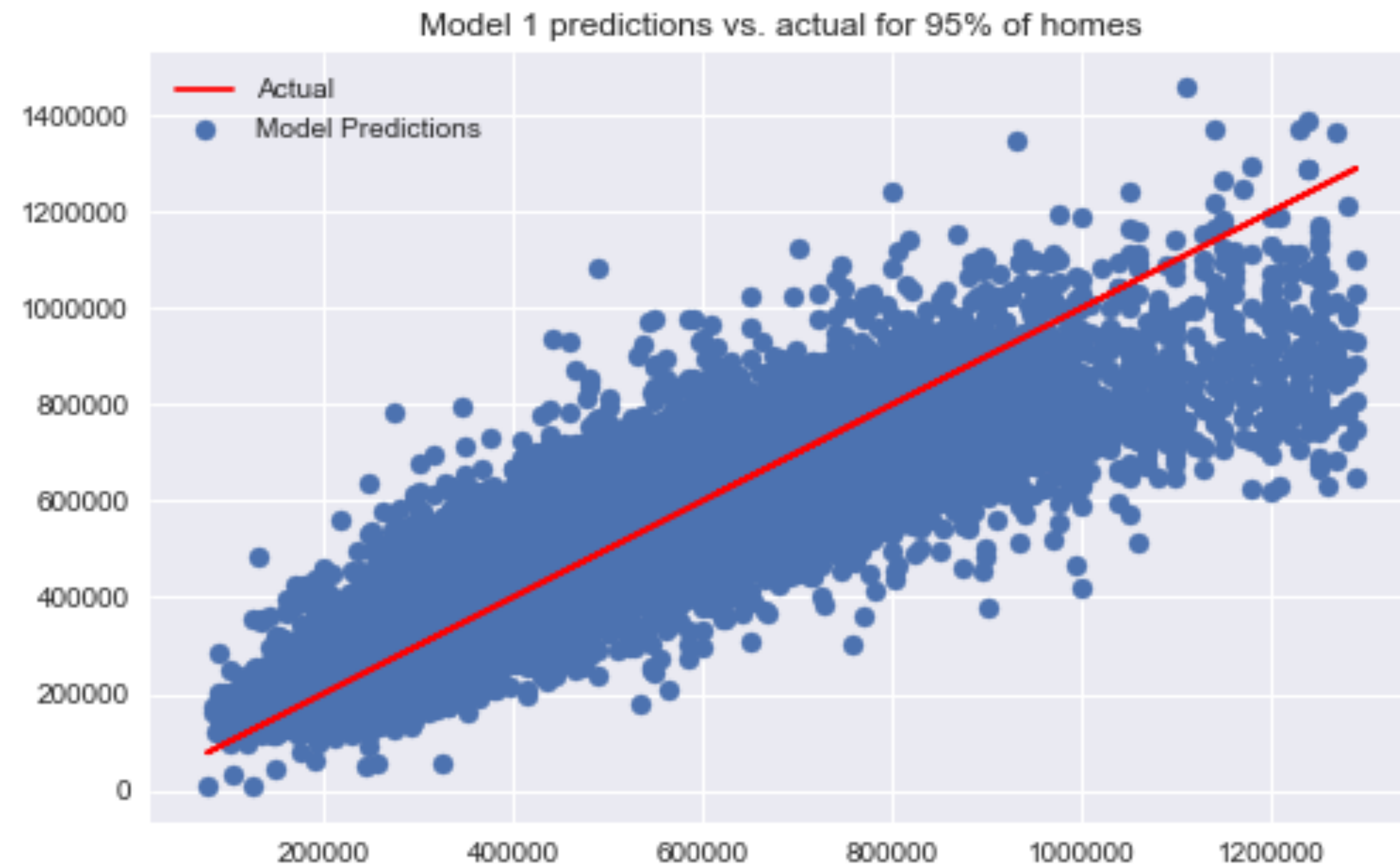
Accounts for Confounding Interaction

Accuracy of 79%

Underfits more expensive house, fits house <900K well

OTHER INSIGHTS

- waterfront valued @ 234K
- 2000 sqft living space is worth 153K more than 1000K
- tier 1 zip codes worth 70k more than tier 2 zipcodes



Conclusions/Actionable Insights

EDA

- **Location:** near the waterfront, in tier 1/2 zipcodes
- **House features:** living area square footage, grade of materials, # of bedrooms, renovations since 2000
- **When:** prep to sell during April-July, starting marketing in earlier Feb/March

Modeling

- What's the waterfront worth ?
~ **240K**
- Model 1 & 2 Accuracy >78%
- living area square footage worth \$153 - \$200 per sqft

Questions

Thank You