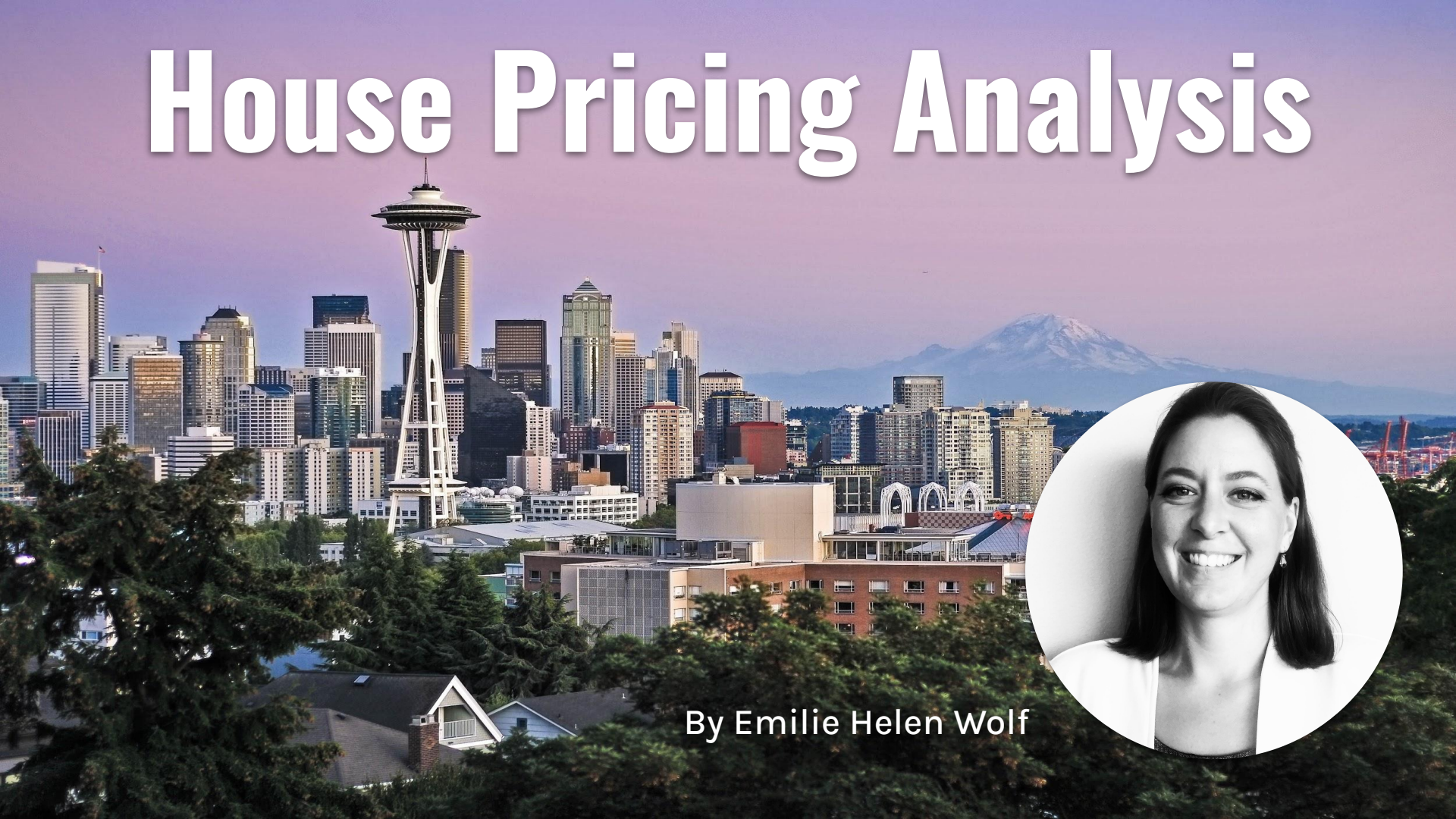
# House Pricing Analysis





By Emilie Helen Wolf


# First Milestone Objective

Analyze and plot past data of house sales in the Seattle Metropolitan Area in preparation for predicting optimal prices for future home sales to ensure a better selling experience

# Overview of the Data Collected

**21,613 Observations**
At least 99% of the data is complete and error-free

**22 Features**
Location, house and lot measurements, condition, time of sale, improvements

**1 Target Variable: Price**
House prices ranging from $75K to $7.7M

A

B

C

D

E

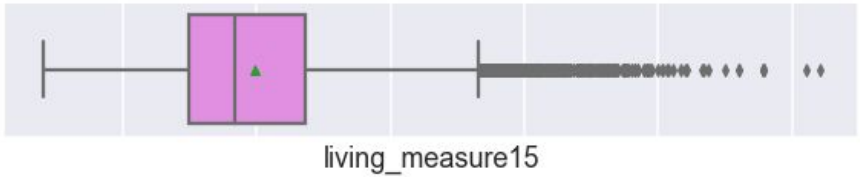
F

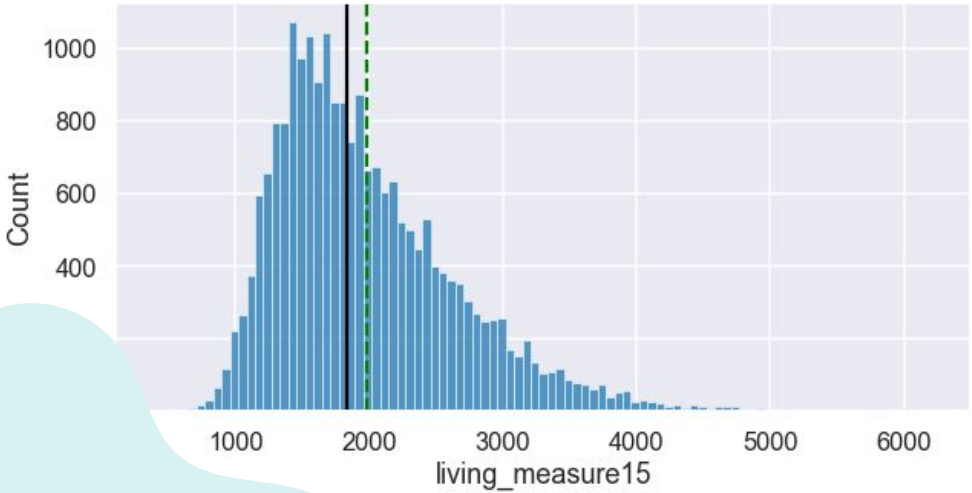
**1 Year of Data**
House sales from May 2014 through May 2015

**70 Zip Codes**
Inner city, waterfront island, suburb, and mountain town neighborhoods

**1900 - 2015**
Years the homes were built

# Univariate Analysis Summary




## Half of the features have a Normal distribution

Some narrow, some wide, all of them right-skewed as seen here

## Outliers detected

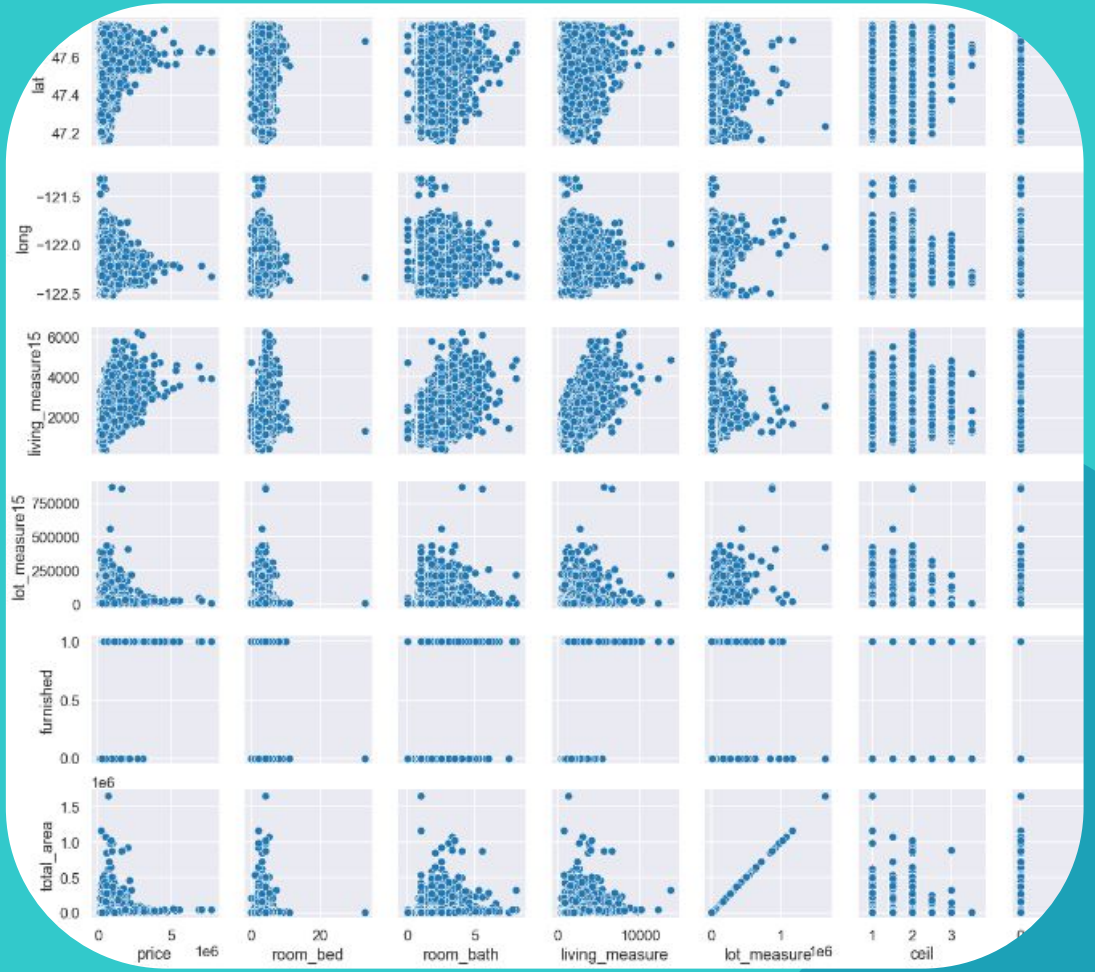
All but 3 of the features showed outliers in the boxplots as seen here

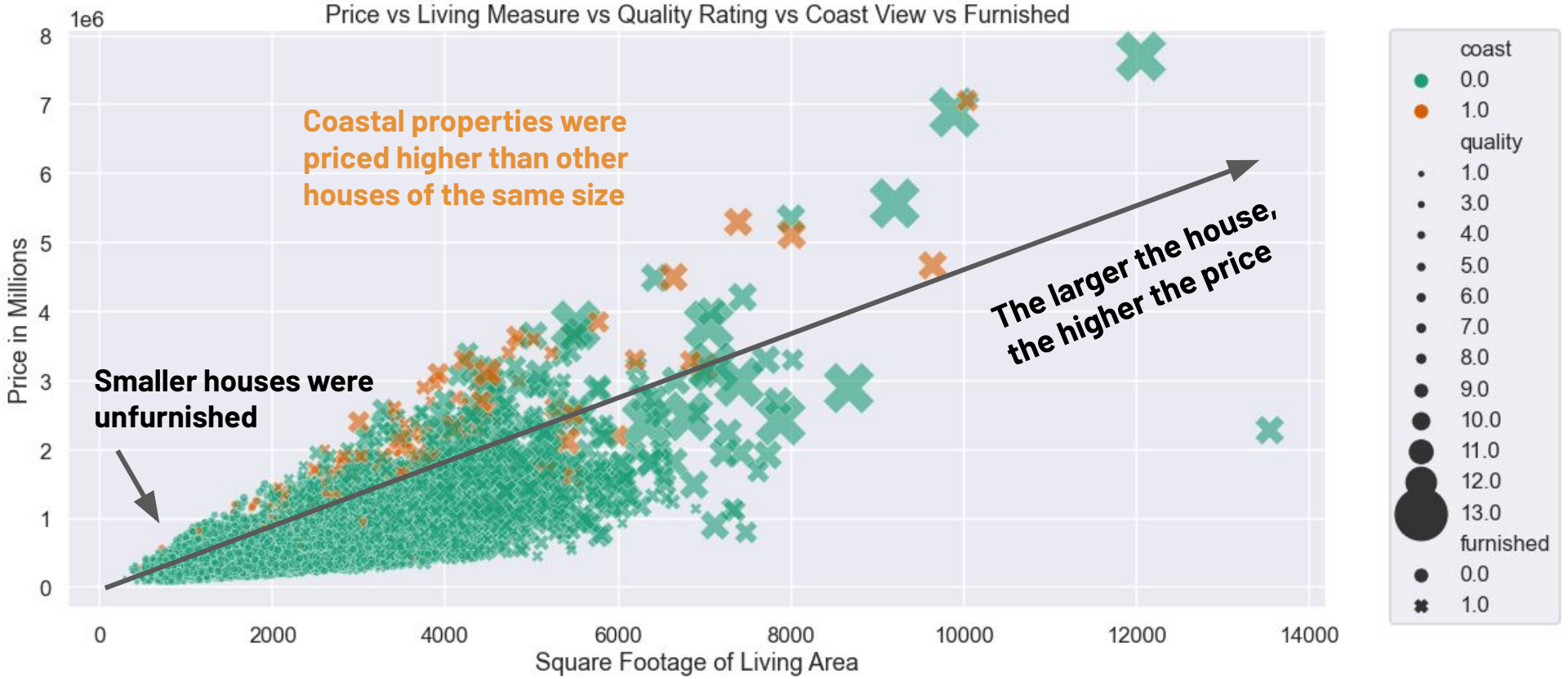
## Feature engineering needed

Some features such as “yr_renovated” have 0’s which will need to be treated

# Bivariate and Multivariate Analysis

There were many interesting correlations among features, especially the positive trend between **price** and **living measurements** (next slide).

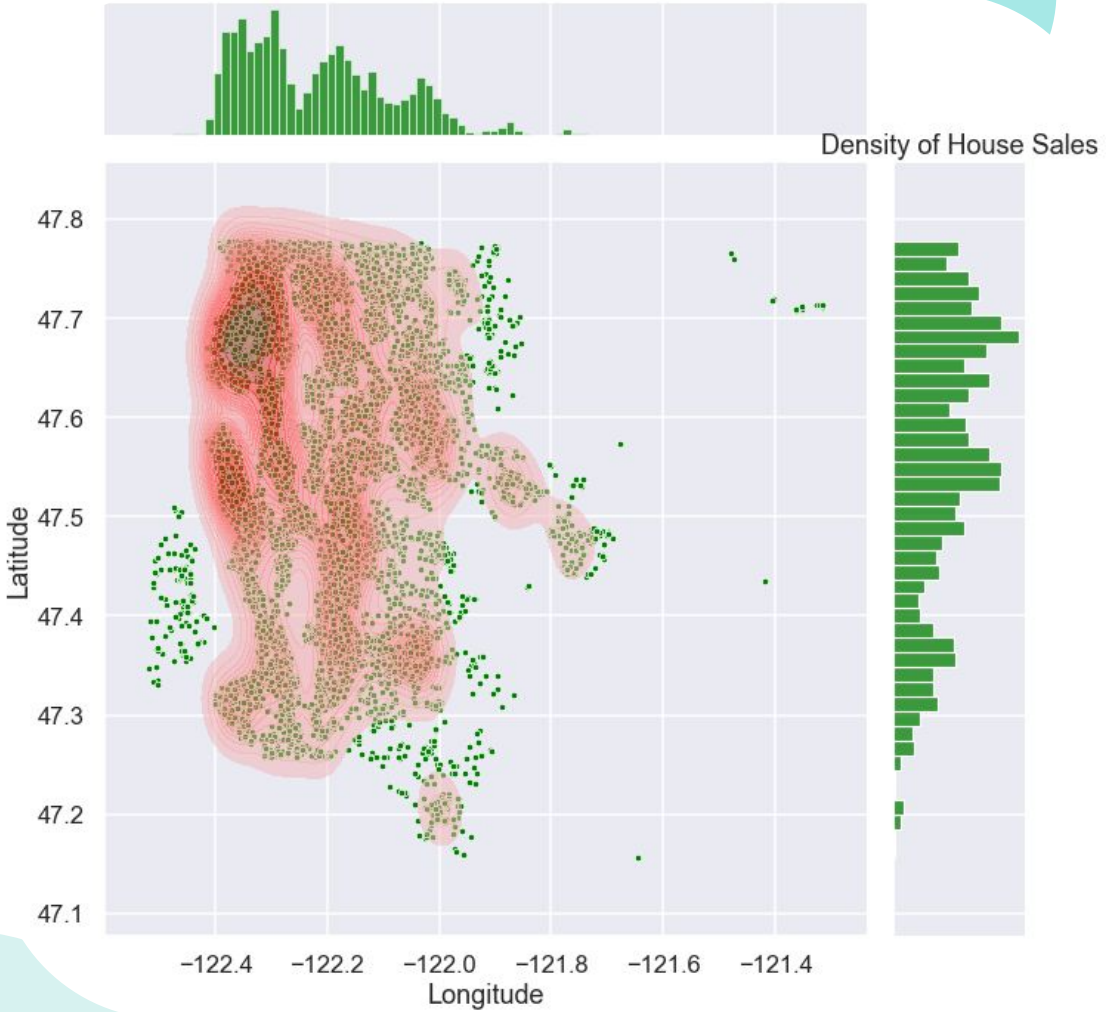



Price vs Living Measure vs Quality Rating vs Coast View vs Furnished
1e6
Price in Millions
Square Footage of Living Area
Coastal properties were priced higher than other houses of the same size
Smaller houses were unfurnished
The larger the house, the higher the price
coast
0.0
1.0
quality
1.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
furnished
0.0
1.0

# Density of House Sales by Latitude and Longitude

Most house sales are concentrated around the city centers of Seattle and Bellevue.

You can see the shape of Vashon Island to the left of the red contour.



## More Plots in Notebook

Click here for the full Jupyter Notebook to see the full analysis in Python and many more insightful graphs.

## What's Next?

In the Second Milestone, we will tidy and engineer the data further to make it ready for linear regression. For questions, please email: emilie.h.wolf@gmail.com

Thank you!