

A large, light gray, stylized house graphic serves as a background. It features a simple roofline with a chimney on the left side and a gabled roof. The house is centered horizontally and vertically.

# King County House Sales

Priscilla Suero

A large, light gray stylized house icon is positioned in the background, spanning across the white and teal sections of the slide. The house has a simple outline with a chimney on the left side.

# Housing Sales

- Predicting house prices from prior house sales
- Provided the King County House Sales Dataset
- Data from 2014-2015
- Over 19,000 house sales

---

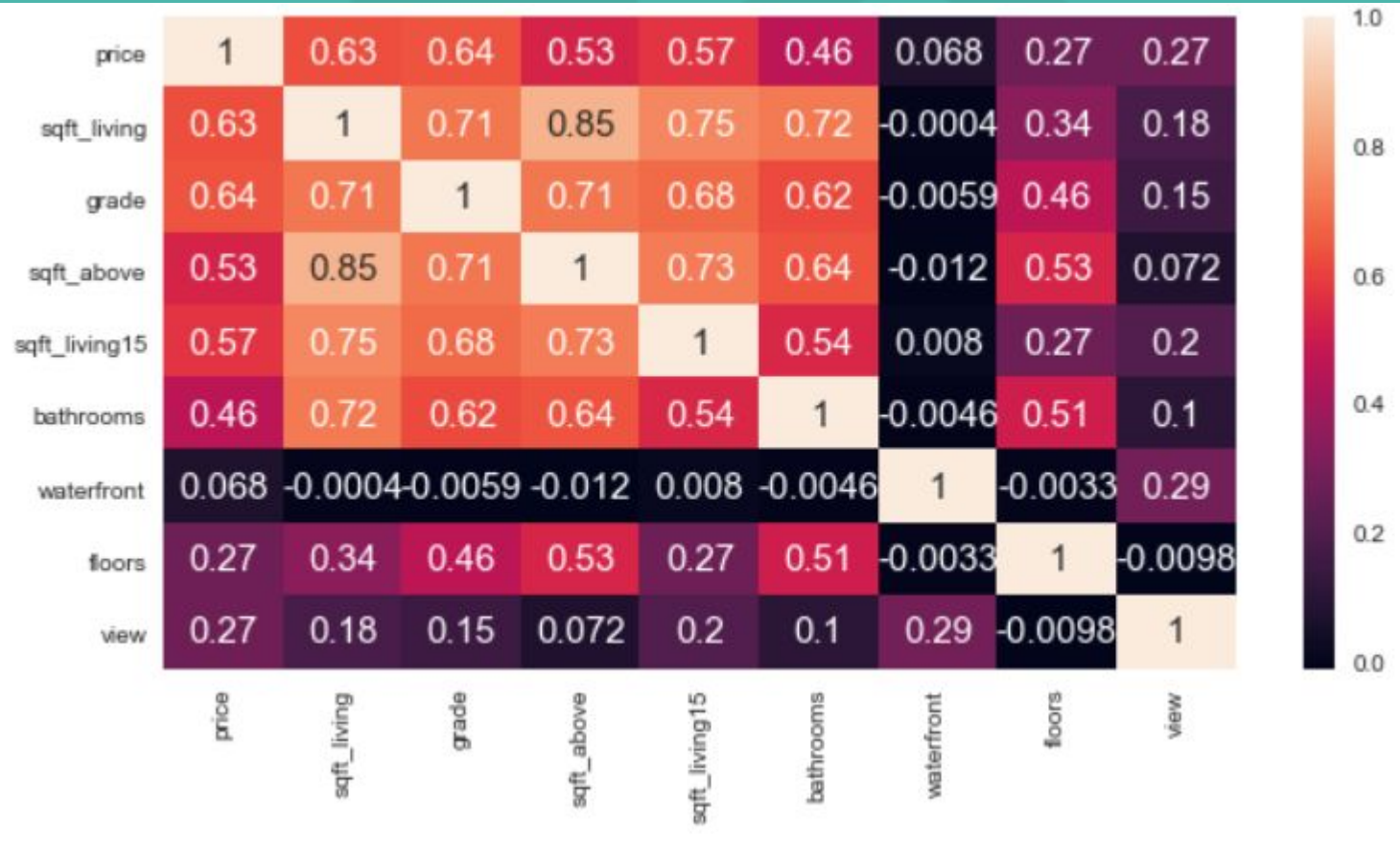
# Data Snapshot

- Bedrooms: 1 - 8
  - Bathroom: .5 - 4.75
  - Square Footage of Lot:
    - 520 sqft - 49,936 sqft
  - Built Year: 1900-2015
-

# Features

- **id** - unique identified for a house
- **dateDate** - house was sold
- **pricePrice** - is prediction target
- **bedroomsNumber** - of Bedrooms/House
- **bathroomsNumber** - of bathrooms/bedrooms
- **sqft\_livingsquare** - footage of the home
- **sqft\_lotsquare** - footage of the lot
- **floorsTotal** - floors (levels) in house
- **waterfront** - House which has a view to a waterfront
- **view** - Has been viewed
- **condition** - How good the condition is ( Overall )
- **grade** - overall grade given to the housing unit, based on King County grading system
- **sqft\_above** - square footage of house apart from basement
- **sqft\_basement** - square footage of the basement
- **yr\_built** - Built Year
- **yr\_renovated** - Year when house was renovated
- **zipcode** - zip
- **lat** - Latitude coordinate
- **long** - Longitude coordinate
- **sqft\_living15** - The square footage of interior housing living space for the nearest 15 neighbors
- **sqft\_lot15** - The square footage of the land lots of the nearest 15 neighbors

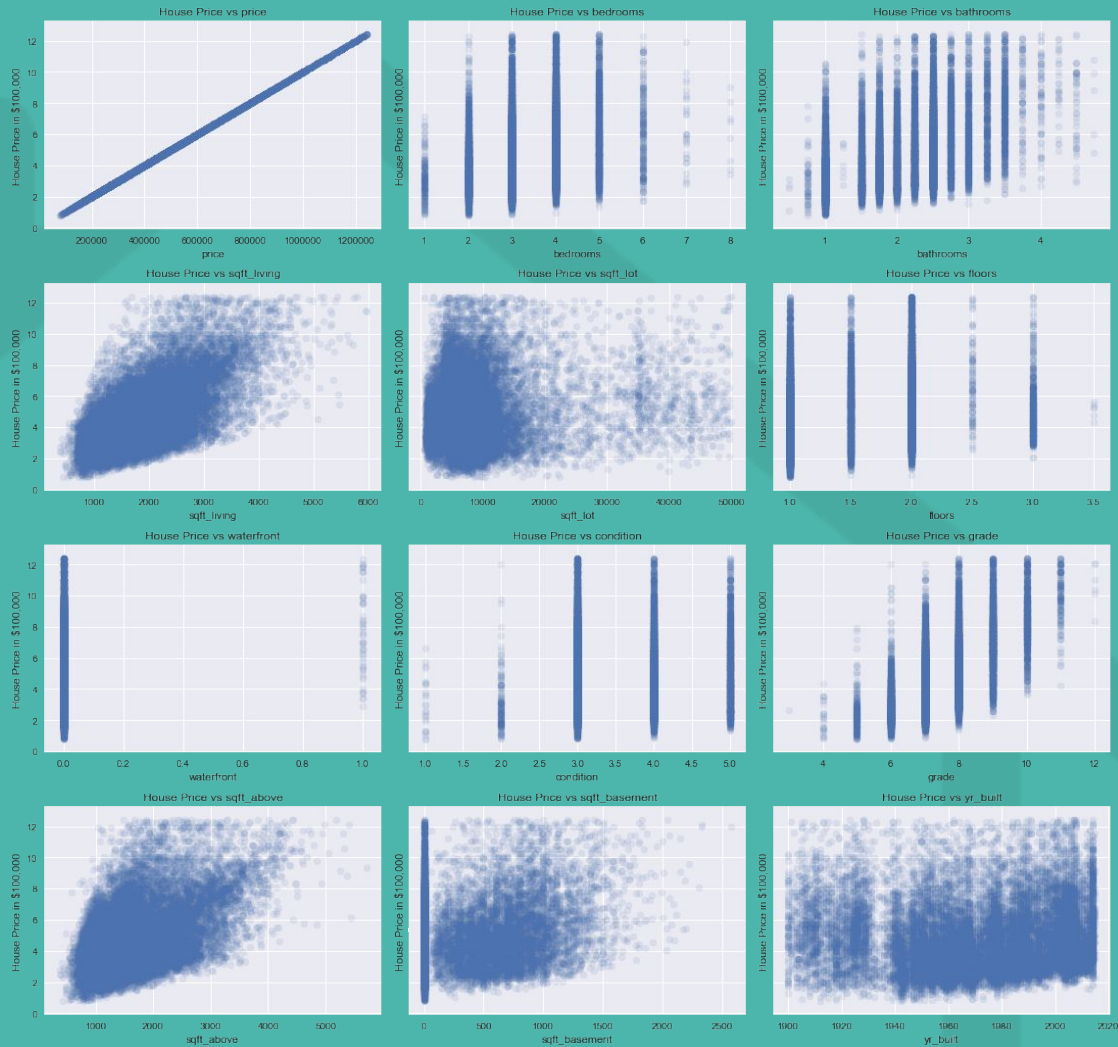
# Feature Correlation



# Continuous vs Categorical

- Continuous:
    - Price, Sqft features
  - Categorical:
    - Bedrooms, bathrooms, floors, conditions, grade, and waterfront
-

# Linearity



# Initial Modeling

Train Mean Squared Error: 18734722588.94  
Test Mean Squared Error: 18450854084.6  
R Squared: 0.612  
Mean Absolute Error: 104553.88  
Root Mean Squared Error: 135833.92  
Average Predicted Price: 480476.9  
Average Actual Price: 480593.35  
Difference: -116.45

Out[5]:

OLS Regression Results

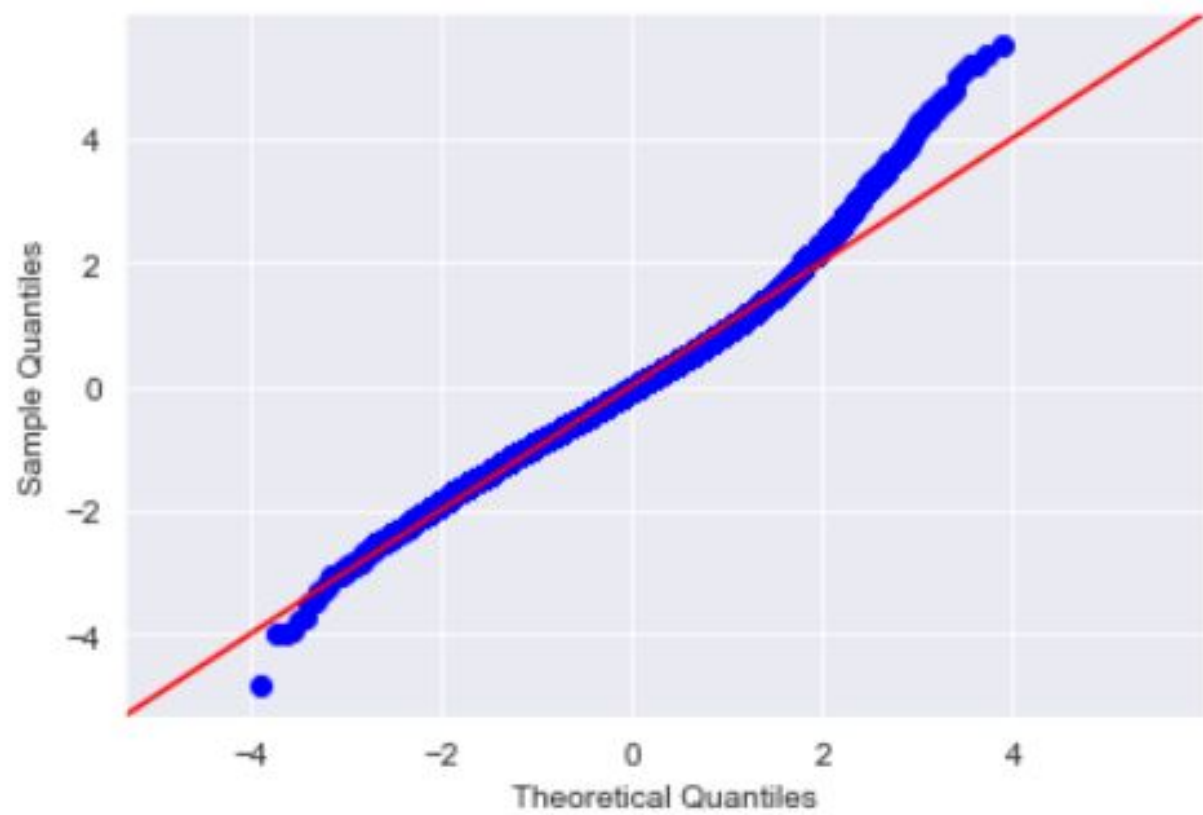
<b>Dep. Variable:</b>	price	<b>R-squared:</b>	0.613
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.612
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	613.7
<b>Date:</b>	Wed, 04 Nov 2020	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	12:58:45	<b>Log-Likelihood:</b>	-2.6278e+05
<b>No. Observations:</b>	19842	<b>AIC:</b>	5.257e+05
<b>Df Residuals:</b>	19790	<b>BIC:</b>	5.261e+05
<b>Df Model:</b>	51		
<b>Covariance Type:</b>	nonrobust		



Train Mean Squared Error: 18734760468.99  
Test Mean Squared Error: 18451757113.48  
R Squared: 0.612  
Mean Absolute Error: 104551.05  
Root Mean Squared Error: 135837.24  
Average Predicted Price: 480481.55  
Average Actual Price: 480593.35  
Difference: -111.8

#### OLS Regression Results

<b>Dep. Variable:</b>	price	<b>R-squared:</b>	0.613
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.612
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	626.0
<b>Date:</b>	Wed, 04 Nov 2020	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	14:03:05	<b>Log-Likelihood:</b>	-2.6278e+05
<b>No. Observations:</b>	19842	<b>AIC:</b>	5.257e+05
<b>Df Residuals:</b>	19791	<b>BIC:</b>	5.261e+05
<b>Df Model:</b>	50		
<b>Covariance Type:</b>	nonrobust		



# Next Steps

- Look deeper into the location
- How location affects price
- Zipcodes
  - But also the longitude and latitude that was provided

---

**Thank you!**



[LinkedIn](#)



[Github Repo](#)



[sueropriscilla@gmail.com](mailto:sueropriscilla@gmail.com)

---