

# Kings County Investment Opportunities

# Contents



Data



Model



Results &  
Discussion



# Data



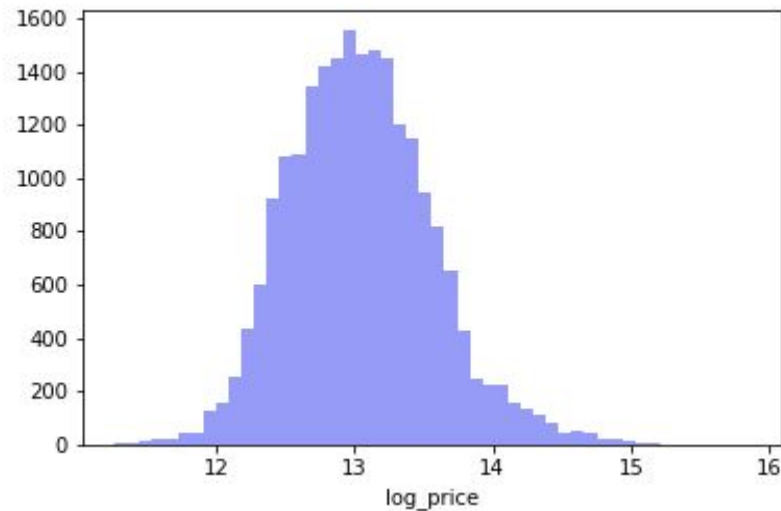
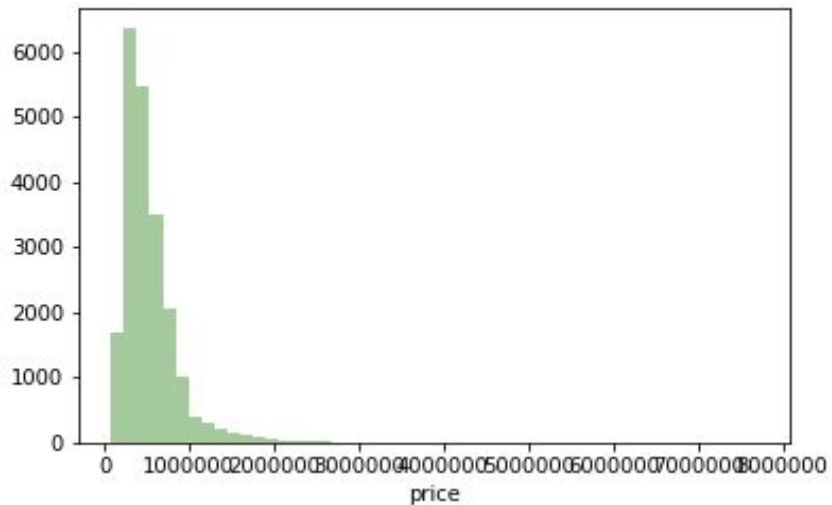
# King county house prices dataset

- Number of bed/bathrooms
- Size of the property (including floors)
- Size of surrounding properties
- View quality & whether facing waterfront
- Data on the condition of the house
- Time of building/ renovation
- Location
- Date the house was sold
- Price

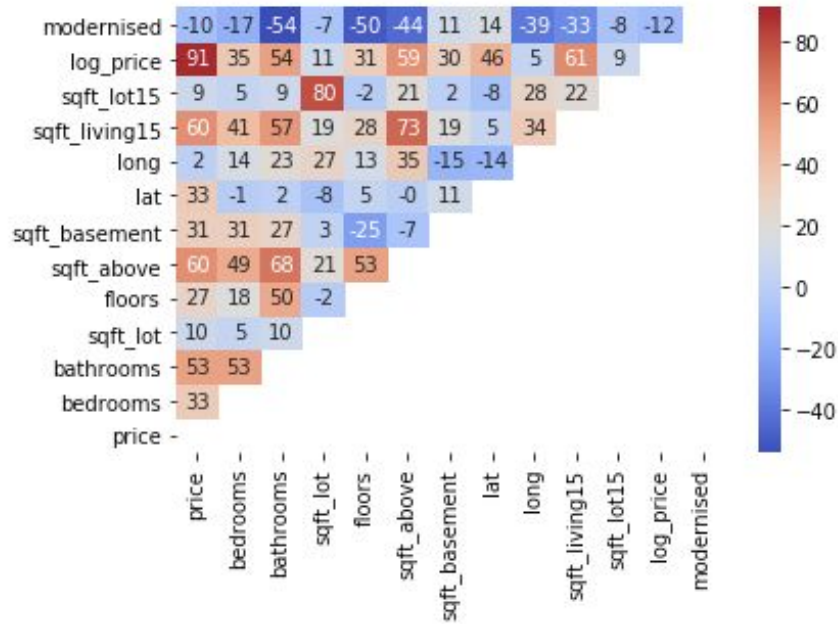
# Dataset description

- 21,597 observations
- 21 variables
- Missing Data for three variables
  - Waterfront - 11%
  - View - 0.3%
  - yr\_renovated 17.8%
- 65 outliers

## ► Price & Log Price



# Correlations



# Model





# Selected Features

18 significant and relevant features (incl. dummies):

- Number of bathrooms
- Size of the living space (excl. basement)
- Size of the basement
- Size of the lot
- Floors
- Location (longitude/ latitude)
- View quality
- Years since last modernisation
- Condition
- Size of the living space & lot of the nearest 15 neighbours

# Model Results

## OLS Regression Results

<b>Dep. Variable:</b>	log_price	<b>R-squared:</b>	0.713
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.713
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	2949.
<b>Date:</b>	Wed, 10 Jun 2020	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	10:56:56	<b>Log-Likelihood:</b>	-2845.4
<b>No. Observations:</b>	21391	<b>AIC:</b>	5729.
<b>Df Residuals:</b>	21372	<b>BIC:</b>	5880.
<b>Df Model:</b>	18		
<b>Covariance Type:</b>	nonrobust		

# Results & Discussion



# House Price Prediction

	Total no of properties	Percent of dataset
Underpriced (>5%)	8,934	41.4%
Overpriced (>5%)	9,124	42.2%
Fairly Priced	3,539	16.4%

# Investment Opportunities

	1049010620	7999600180	3530210260
Living space (sqft)	790	900	3,030
Lot (sqft)	2,640	8,580	45,004
Year built	1973	1918	1987
Price (USD)	90,000	84,000	274,975
Valued at	334,369	293,608	955,510
Profit	244,369	210,608	680,535
In %	272%	254%	247%

## ► Further research needed...



- Model has mediocre predictive powers (R-Squared)
- Model predicts merely 16.4% of the prices



- Test other models and further features
- Mitigate multicollinearity



Thank you for your  
attention!!