



Predicting House Prices with the King's County Housing Data Set

Module 1: Exploratory Data Analysis
-Sarabi Eventide

Which house costs more?





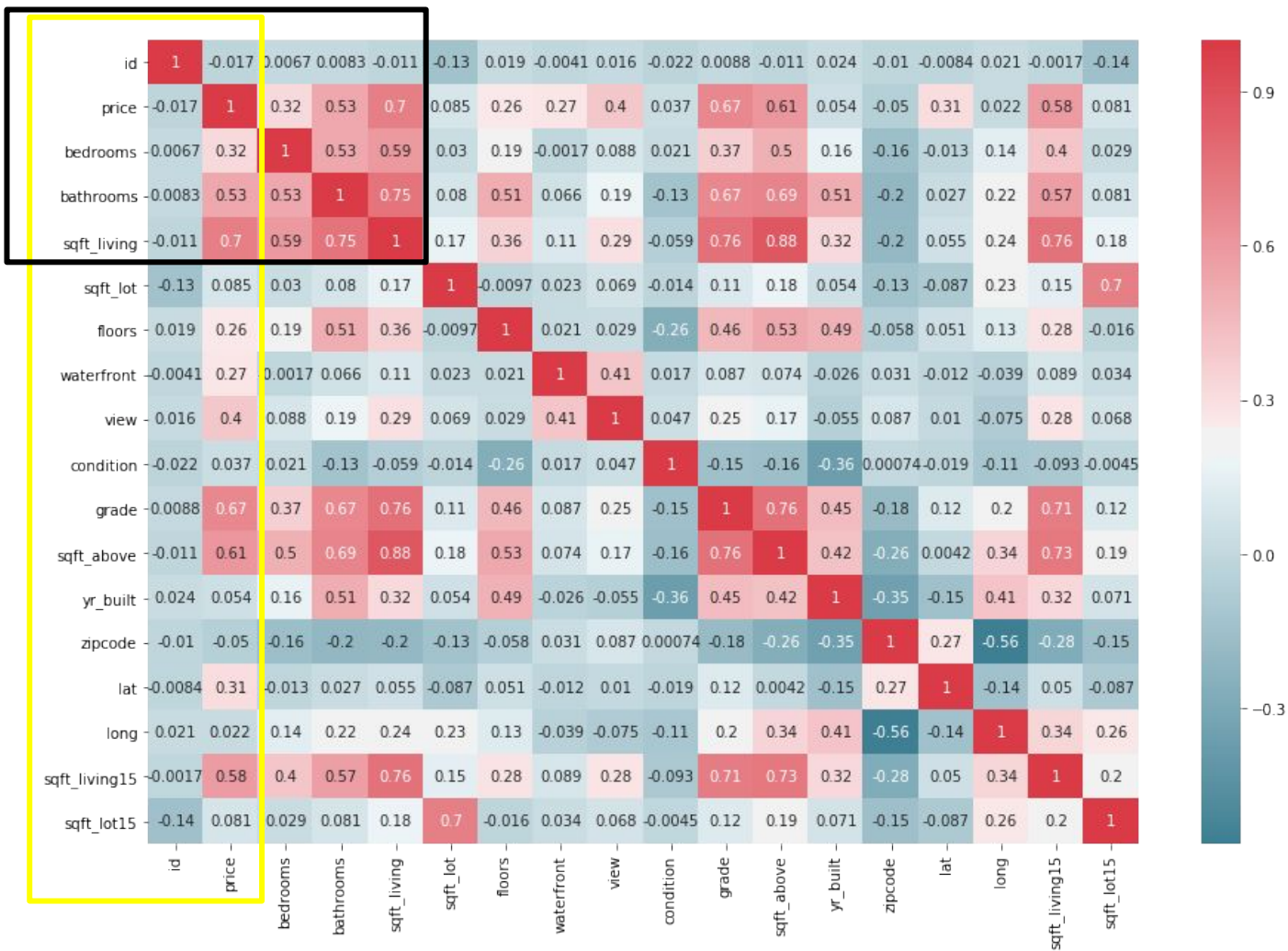
Purpose

Maximize profits



Methodology Overview

1. Clean the Data
2. Explore the Relationships
3. Model the Data
4. Make Predictions





id	1	-0.017	0.0067	0.0083	-0.011
price	-0.017	1	0.32	0.53	0.7
bedrooms	-0.0067	0.32	1	0.53	0.59
bathrooms	-0.0083	0.53	0.53	1	0.75
sqft_living	-0.011	0.7	0.59	0.75	1
sqft_lot	-0.13	0.085	0.03	0.08	0.17
floors	-0.019	0.26	0.19	0.51	0.36
waterfront	-0.0041	0.27	0.0017	0.066	0.11
view	-0.016	0.4	0.088	0.19	0.29
condition	-0.022	0.037	0.021	-0.13	-0.059
grade	-0.0088	0.67	0.37	0.67	0.76
sqft_above	-0.011	0.61	0.5	0.69	0.88
yr_built	-0.024	0.054	0.16	0.51	0.32
zipcode	-0.01	-0.05	-0.16	-0.2	-0.2
lat	-0.0084	0.31	-0.013	0.027	0.055
long	-0.021	0.022	0.14	0.22	0.24
sqft_living15	-0.0017	0.58	0.4	0.57	0.76
sqft_lot15	-0.14	0.081	0.029	0.081	0.18
	id	price	bedrooms	bathrooms	sqft_living



Findings

price = grade + sqft_living15 +
bathrooms + above + condition) + *
bedrooms



The numbers

$$\begin{aligned} \text{price} = & -1080000 + (130200 * \text{grade}) + (77.4393 * \\ & \text{sqft_living15}) + (41590 * \text{bathrooms}) + (70.6386 \\ & * \text{sqft_above}) + (83260 * \text{condition}) + (-8442.6463 * \\ & \text{bedrooms}) \end{aligned}$$

Which house costs more?





Next Steps

- Location, Location, Location
- Effects of renovation
- Age



Questions, comments, concerns?