KING COUNTY HOUSE SALES

MODULE 1 FINAL PROJECT | CYNTHIA PEDRASA

TABLE OF CONTENTS

Business Problem
Data Science Framework
Data Set Features/Predictors
Exploratory Data Analysis
Multiple Regression Assumptions
Interpretation

BUSINESS QUESTION

Problem Statement

Which features have the biggest impact on home sale prices?
- What is the Median Home Sale Price by Zip Code?
- Where are the Top 20 Most Expensive Homes Sold located?
- Do more bedrooms add value?
- Do more bathrooms add value?
- How does the condition affect the price of a home?
- How does the grade (quality of construction and design) affect home pricing?
- Which number of floors add the most value to the sale of a home?
- What is the home construction history like at King County
- What is the best month and day to sell a home?
- What is the distribution of homes sold with views, waterfronts, or have undergone renovation?
- Which features add the most value to home sales?

DATA SCIENCE PROCESS

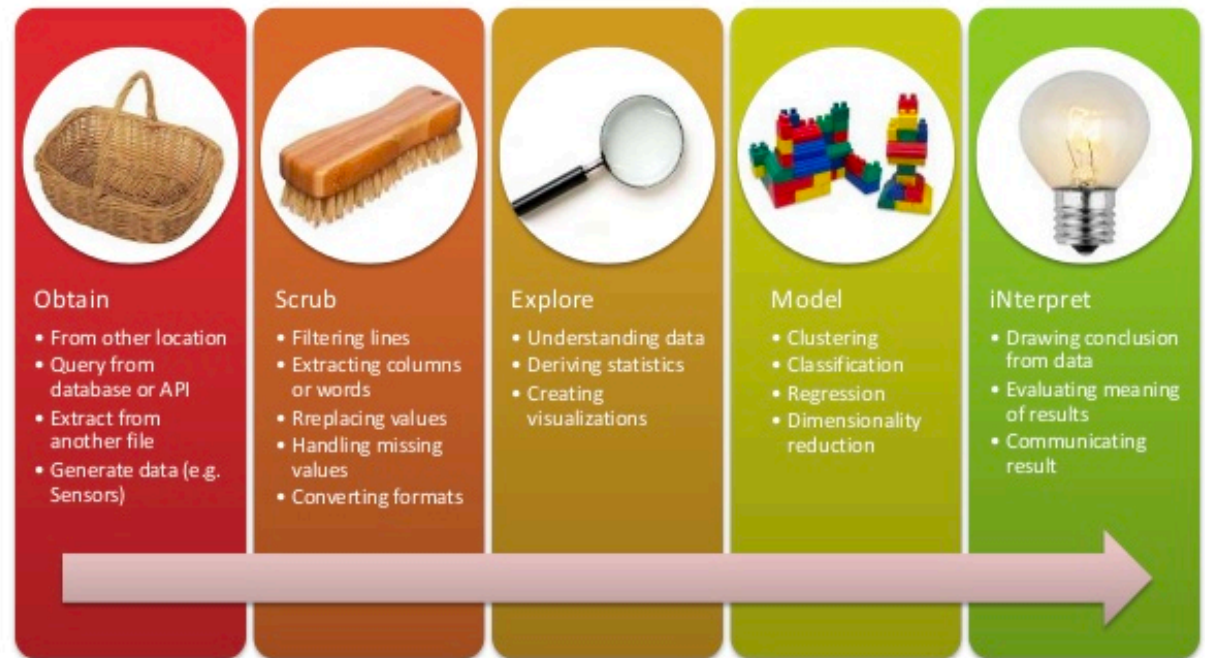Obtain-Requirements and information gathering on the problem.

Scrub- Pre-processing our data (removing nulls, outliers, normalization, feature engineering, and one-hot encoding)

Explore-Check the data to see if it meets the assumptions necessary for the model (regression)

Model-Build and tune the model

Interpret - and communicate results to stakeholders.

OSEMN model

Obtain
- From other location
- Query from database or API
- Extract from another file
- Generate data (e.g. Sensors)

Scrub
- Filtering lines
- Extracting columns or words
- Rreplacing values
- Handling missing values
- Converting formats

Explore
- Understanding data
- Deriving statistics
- Creating visualizations

Model
- Clustering
- Classification
- Regression
- Dimensionality reduction

iNterpret
- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result

Source: A Taxonomy of Data Science

DATASET COLUMNS

id unique id for each home sold

date Date of the home sale

price Price of each home sold - This is the prediction target

bedrooms Number of Bedrooms

bathrooms Number of Bathrooms

sqft_living square footage of the home's interior living space

sqft_lot square footage of the land space

floors number of floors (levels) in house

waterfront House which has a view to a waterfront

view an index of 0-4 of how good the view of the property was

condition an index from 1-5 on the condition of the home

grade an index to 1-13, where 1-3 falls short of building construction and design, 7 has an average level, and 11-13 have a high quality level of construction and design

sqft_above square footage of the interior housing space that is above ground level

sqft_basement square footage of the interior housing space that is below ground level

yr_built the year the house was originally built

yr_renovated the year of the house's last renovation

zipcode zipcode area the house is in

lat Latitude coordinate; long Longitude coordinate

sqft_living15 square footage of interior housing living space for the nearest 15 neighbors

sqft_lot15 square footage of the land lots of the nearest 15 neighbors
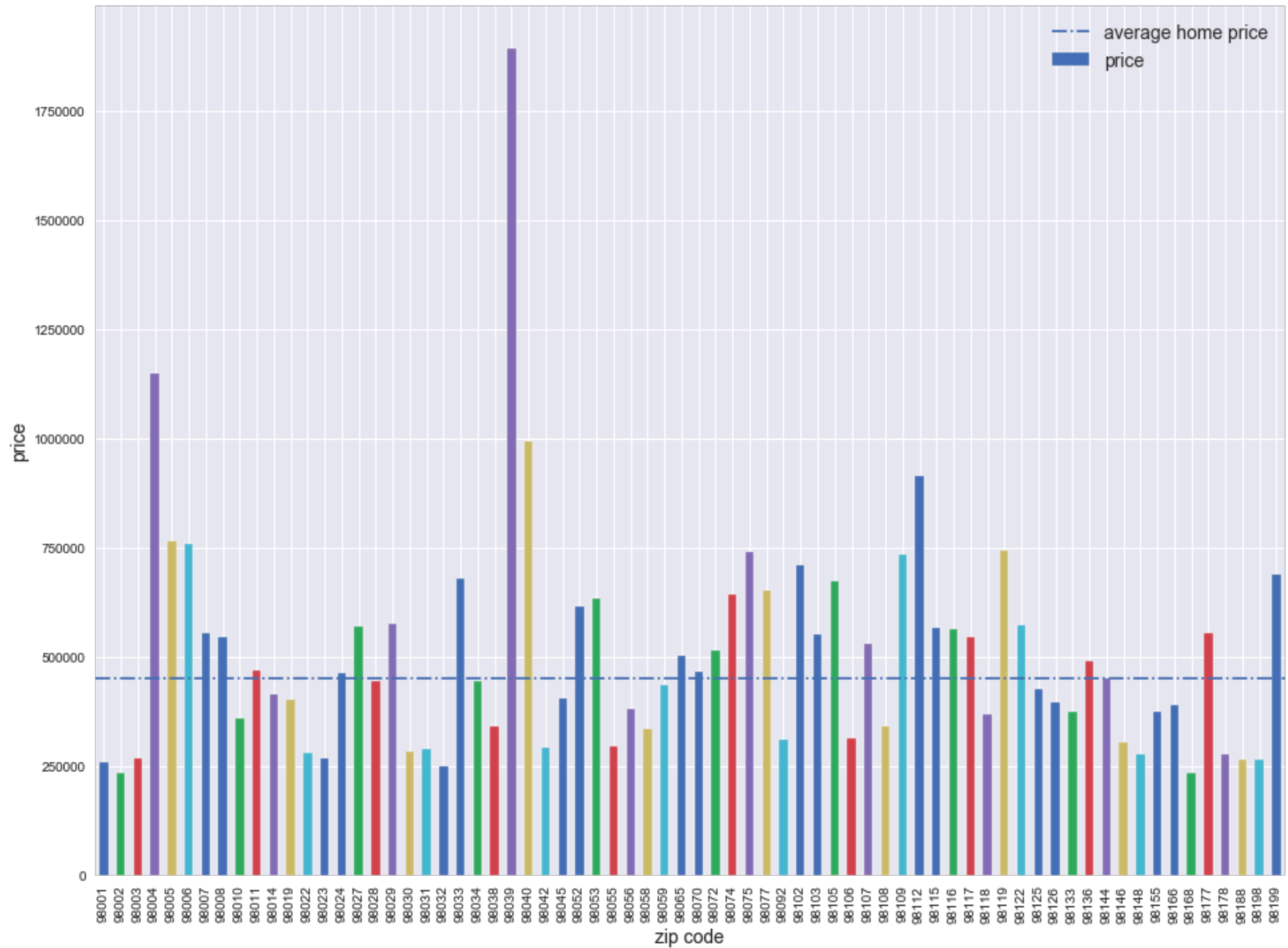
WHAT IS THE MEDIAN HOME SALE PRICE BY ZIP CODE?

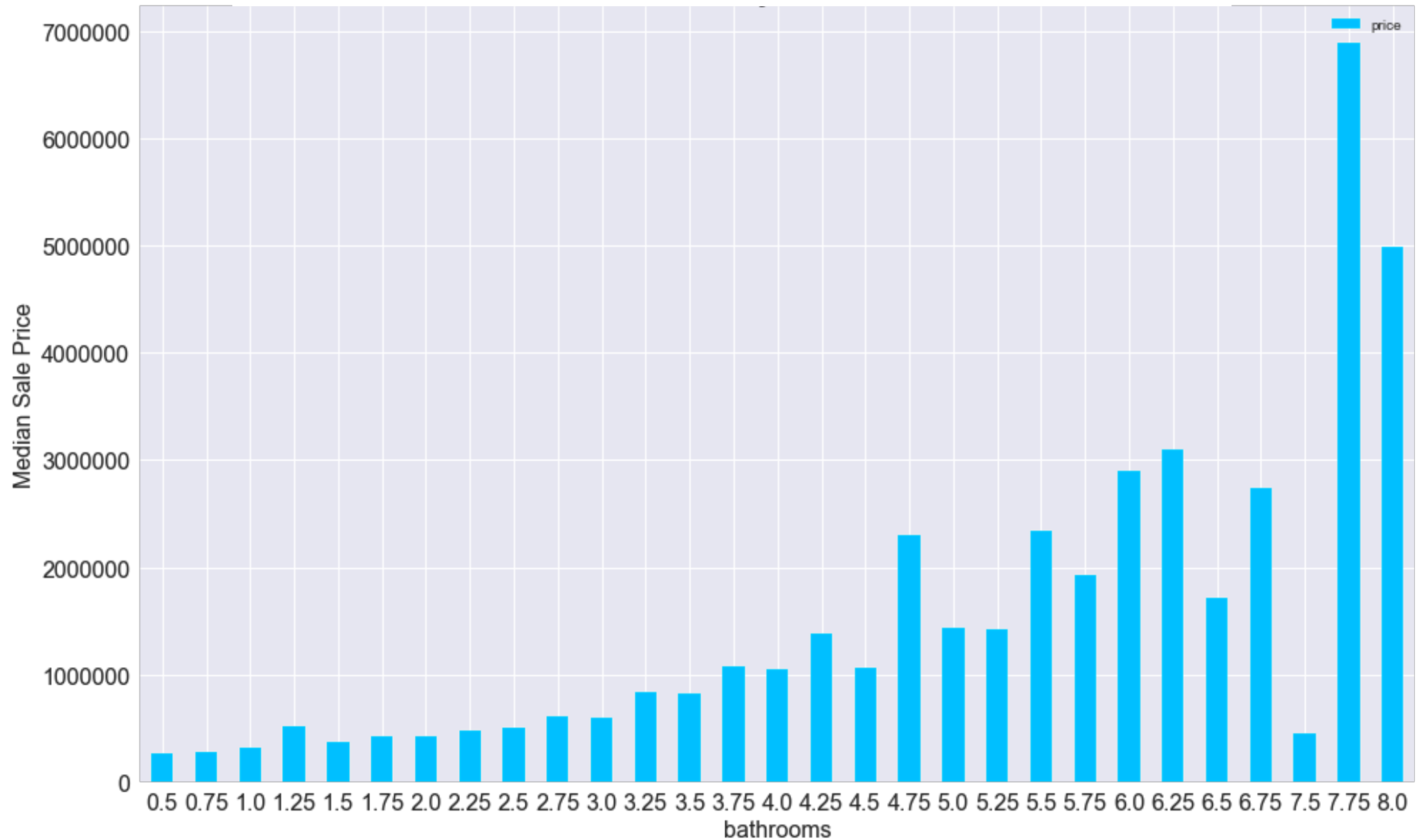# WHERE ARE THE TOP 20 MOST EXPENSIVE HOMES SOLD LOCATED?

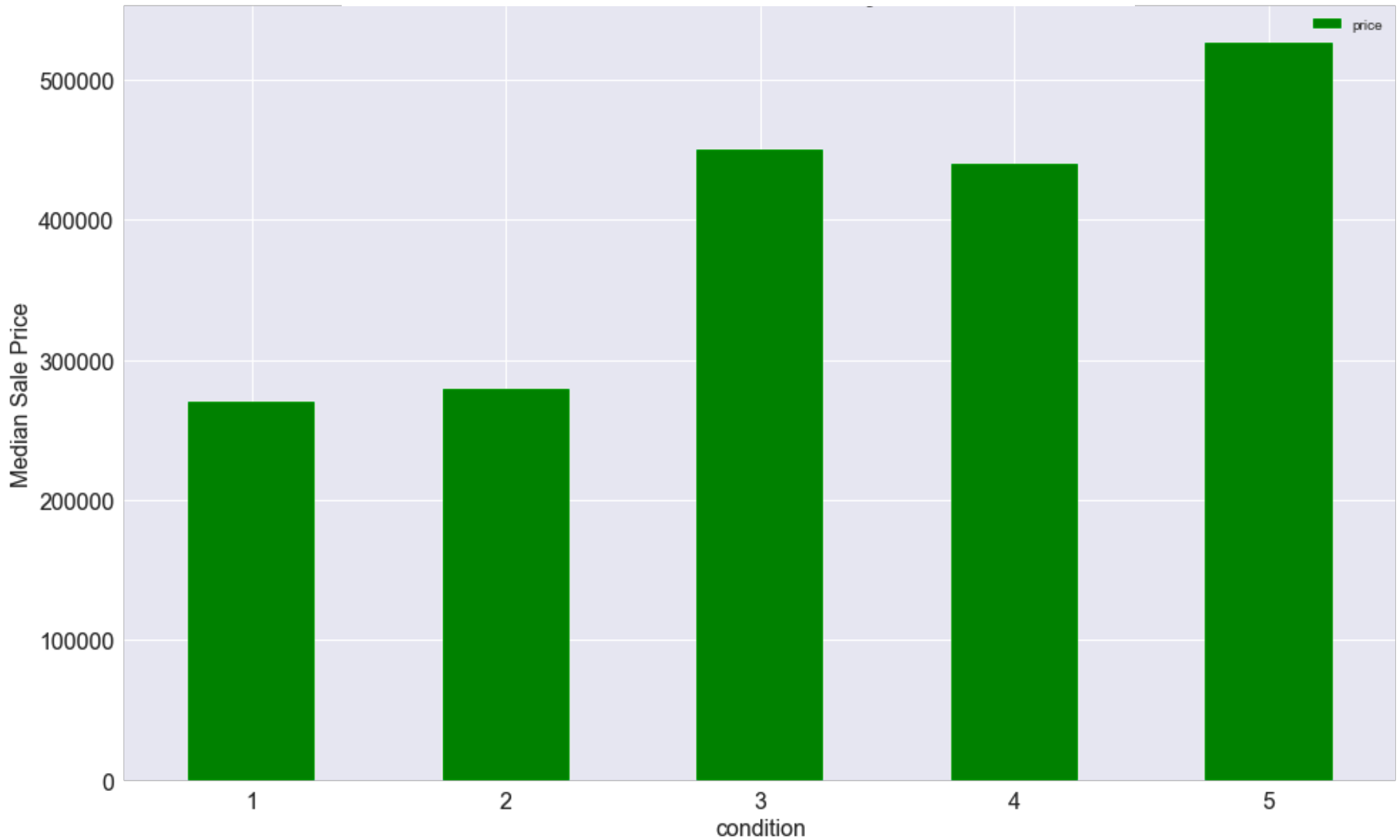# DO MORE BEDROOMS ADD VALUE?

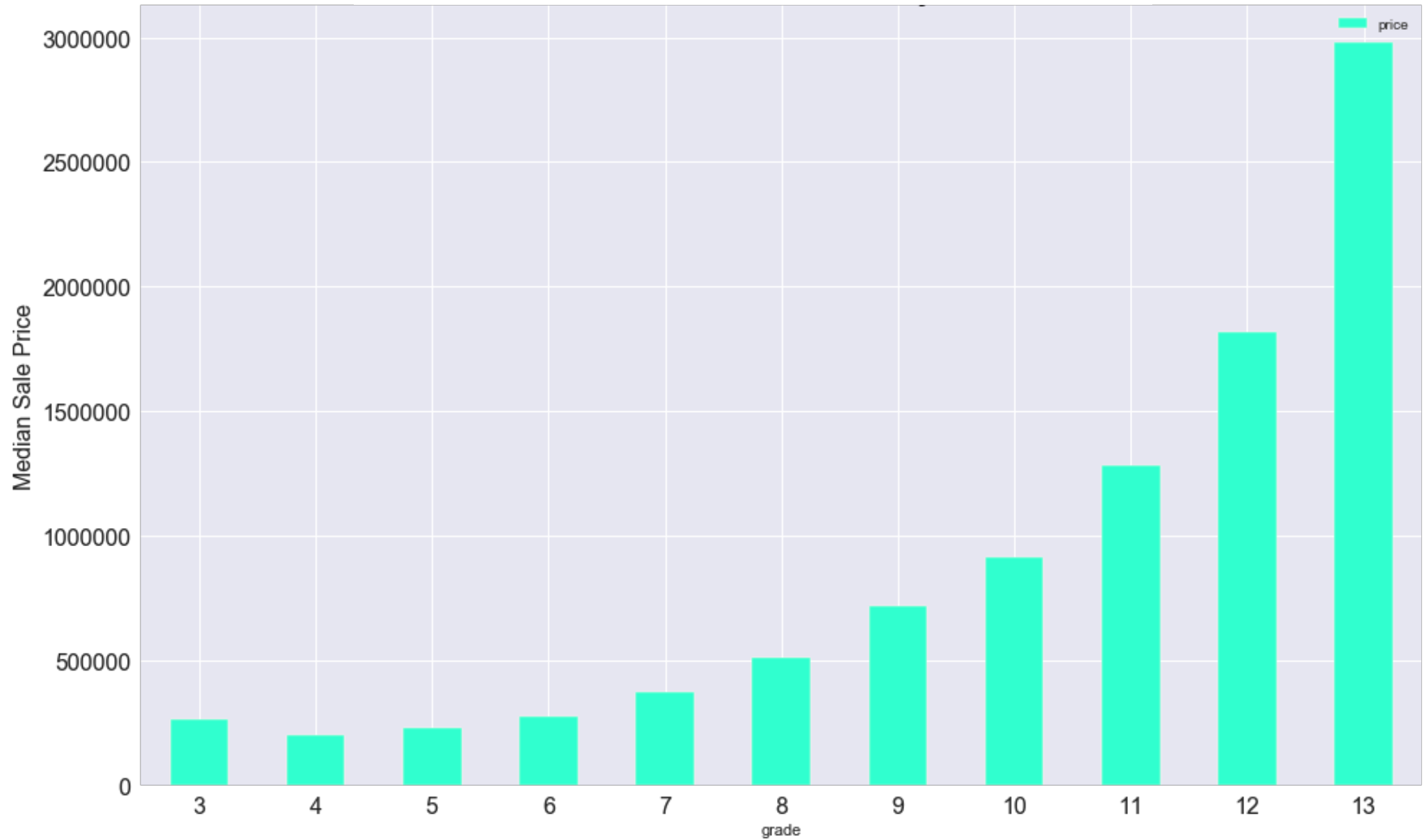# DO MORE BATHROOMS ADD VALUE?

HOW DOES THE CONDITION AFFECT THE PRICE OF A HOME?

Bar chart showing Median Sale Price versus condition. Legend: price (green).

condition	Median Sale Price (approx.)
1	270000
2	280000
3	450000
4	440000
5	530000
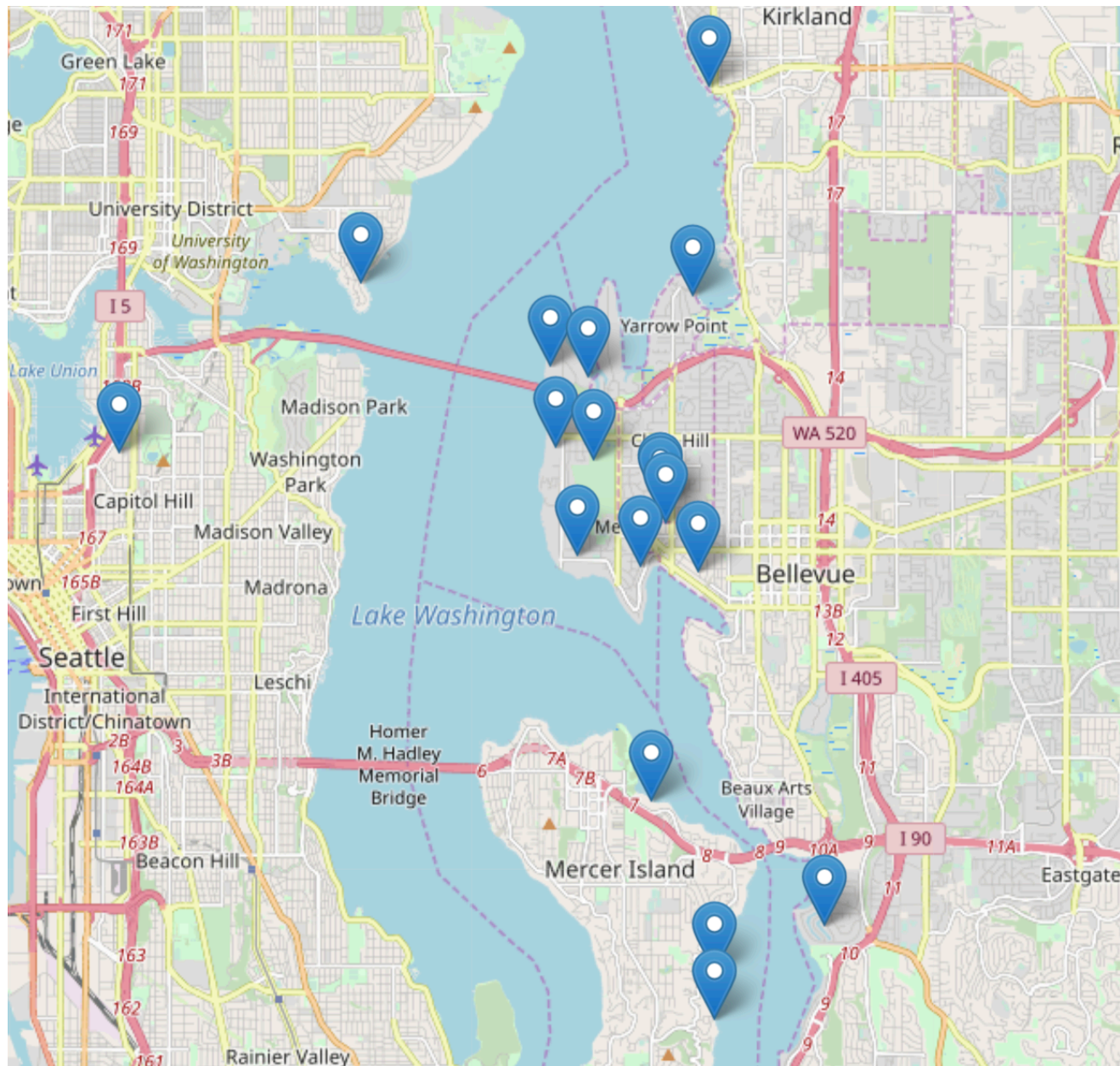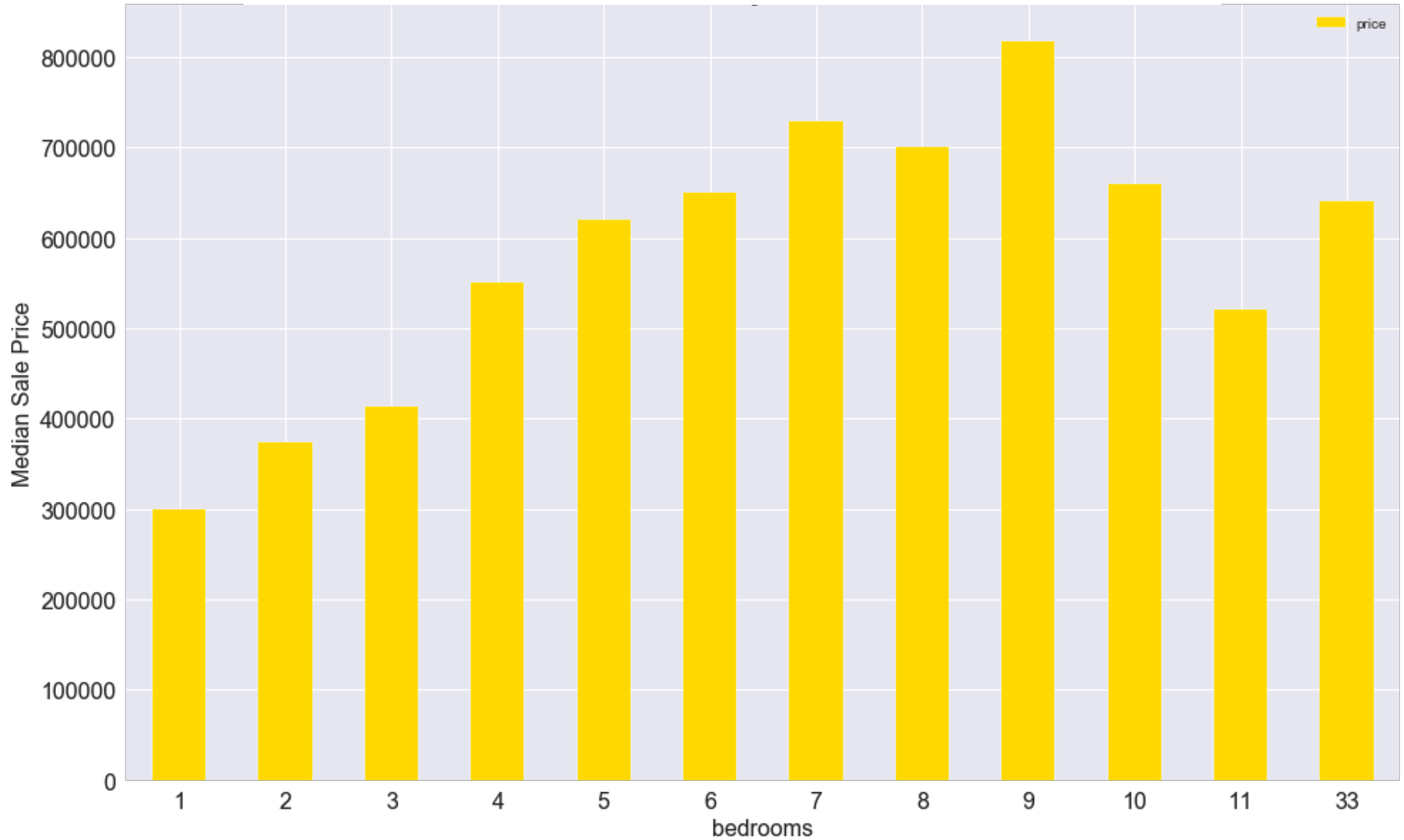
HOW DOES THE GRADE (CONSTRUCTION AND DESIGN) AFFECT HOME PRICING?

WHAT NUMBER OF FLOORS ADD THE MOST VALUE TO THE SALE OF A HOME?
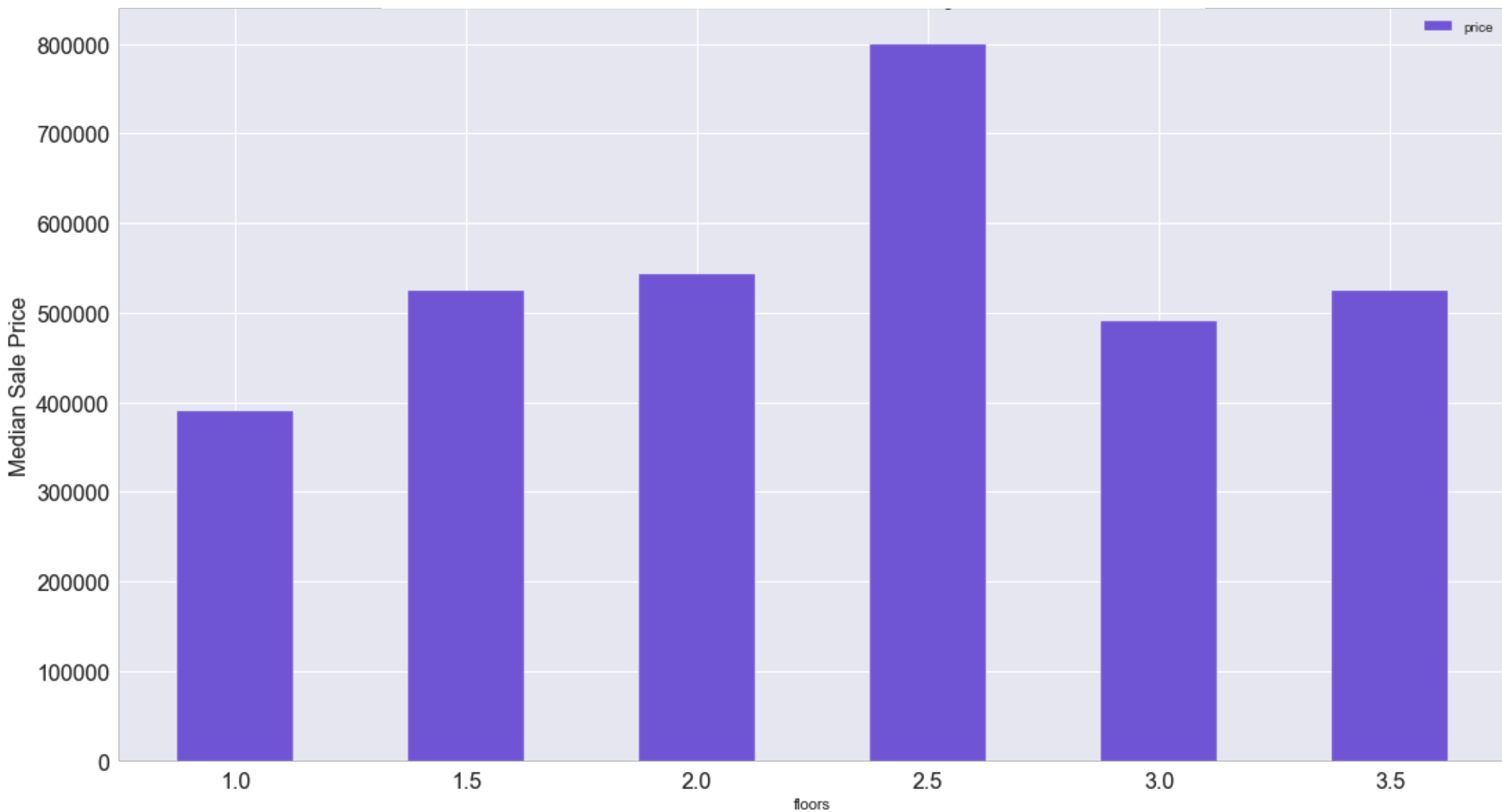
WHAT IS THE DISTRIBUTION OF HOMES SOLD WITH WATERFRONTS, VIEWS, OR RENOVATION?

DISTRIBUTION OF VIEW
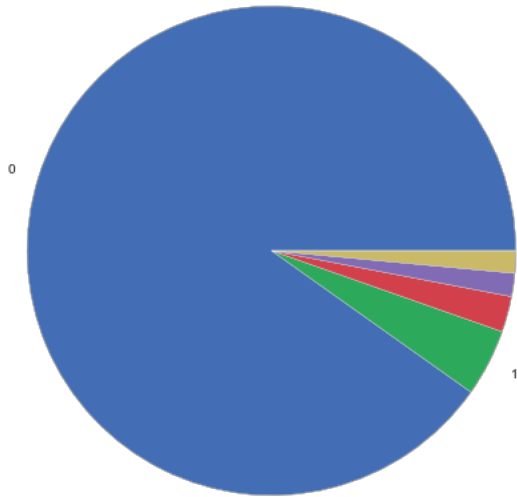
DISTRIBUTION OF WATERFRONT
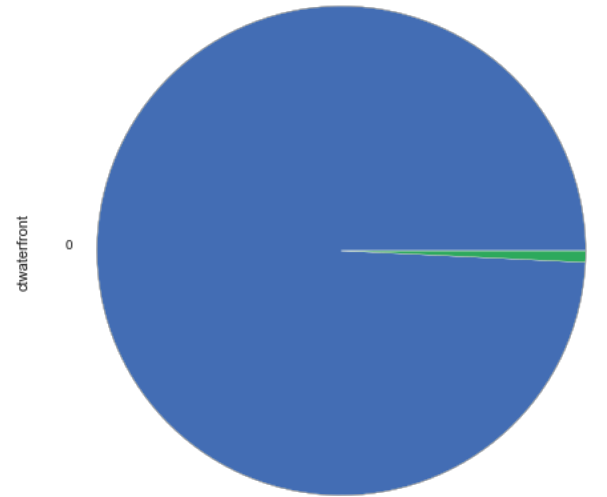
DISTRIBUTION OF YEAR RENOVATED
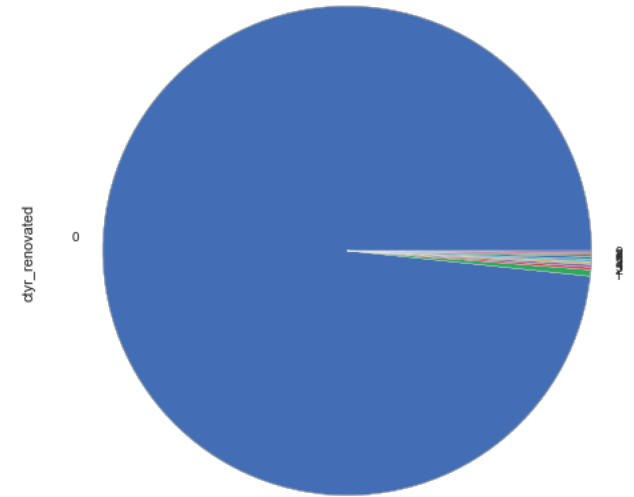
Median Home Price with & without waterfront
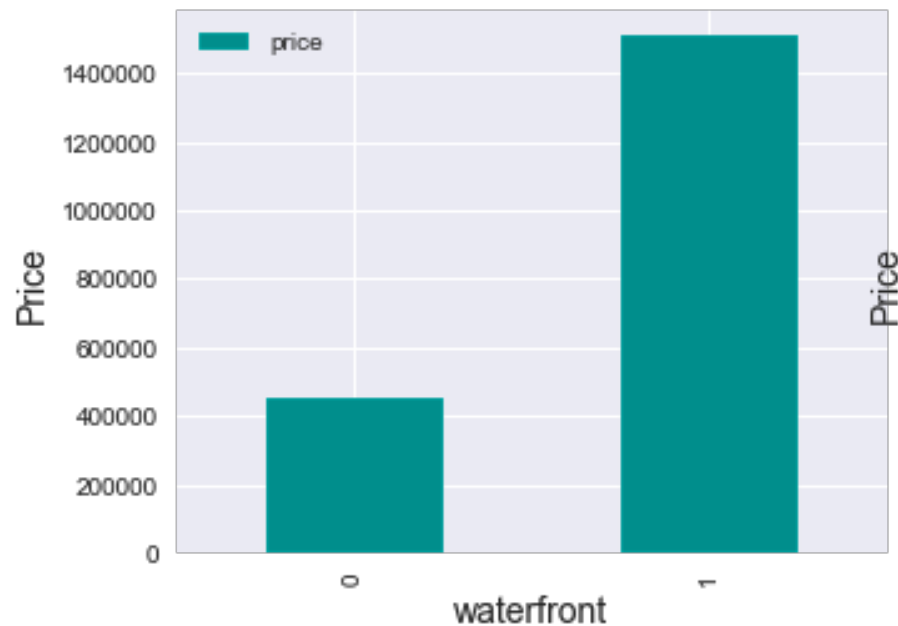
Median Home Price by Views
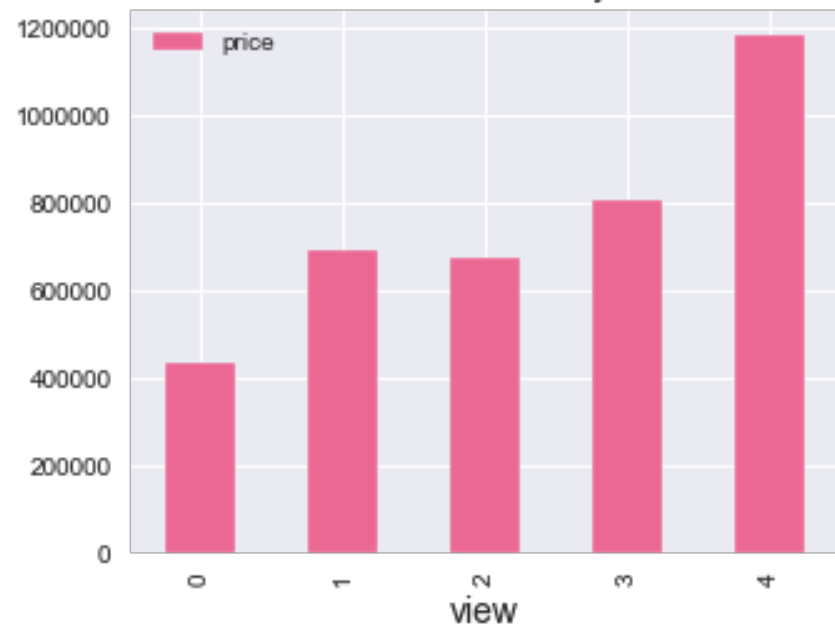
HOW MANY HOMES WERE BUILT IN KINGS COUNTY FROM 1900-2014?
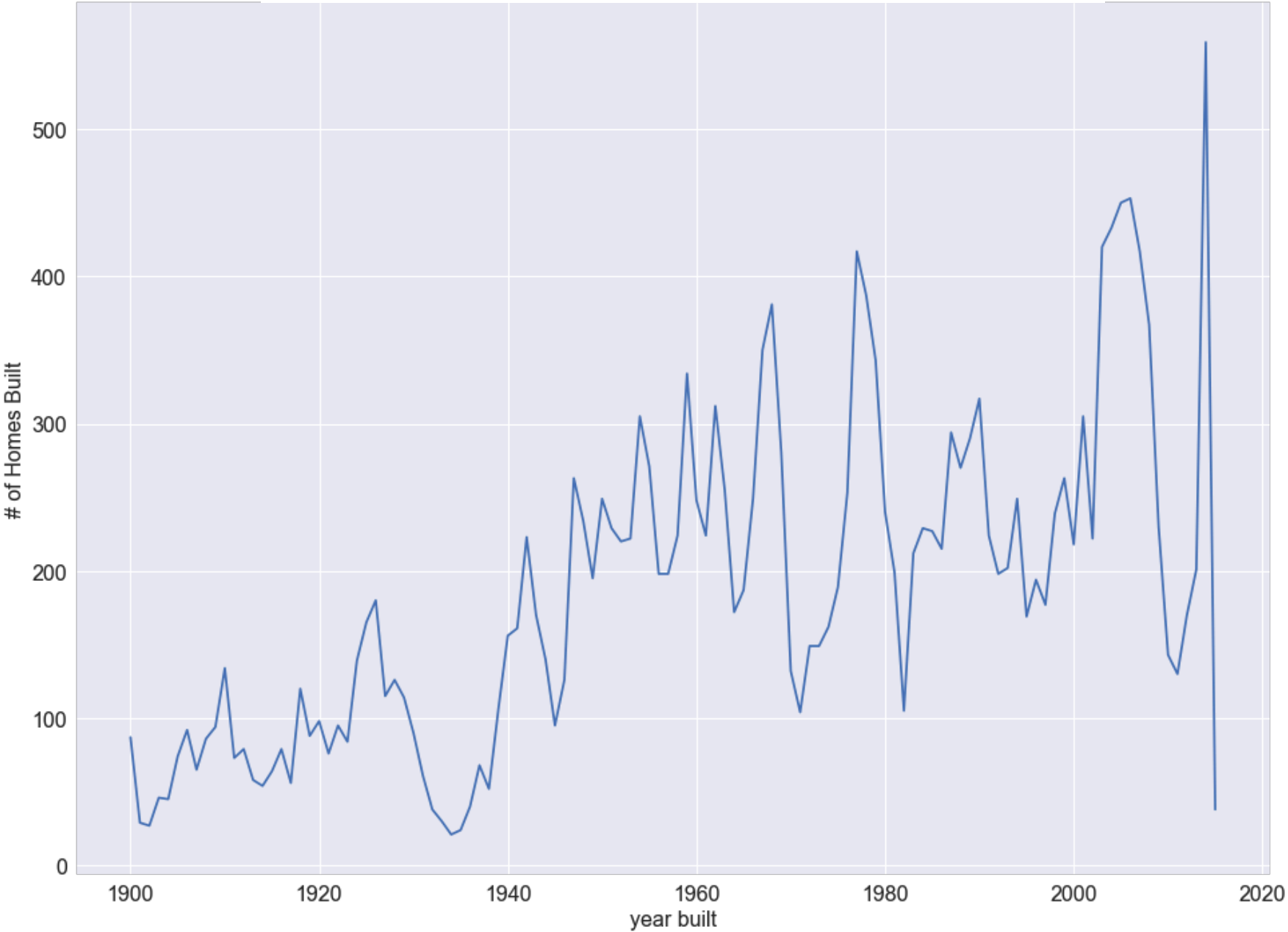
WHICH YEARS WERE HOME CONSTRUCTIONS THE HIGHEST? LOWEST?

Number of Houses Built in King County
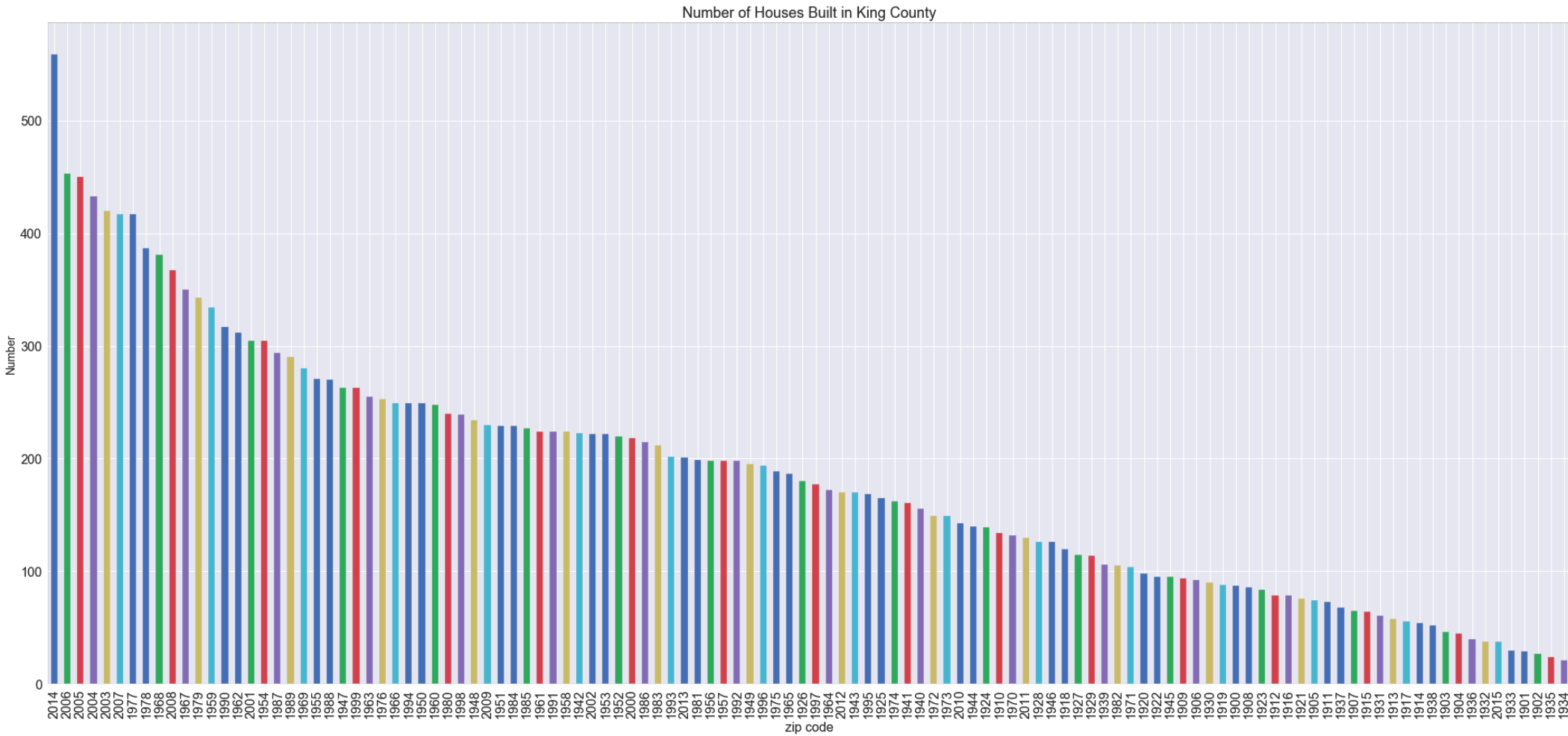
The top 10 years were most homes were built are in the years 2014, 2006, 2005, 2004, 2003, 2007, 1977, 1978, 1968 and 2008. The years where constructions were the lowest are 1934, 1935, 1902,1901,1933, 2015,1932,1936,1904, and 1903

WHAT IS THE BEST MONTH TO SELL A HOME?
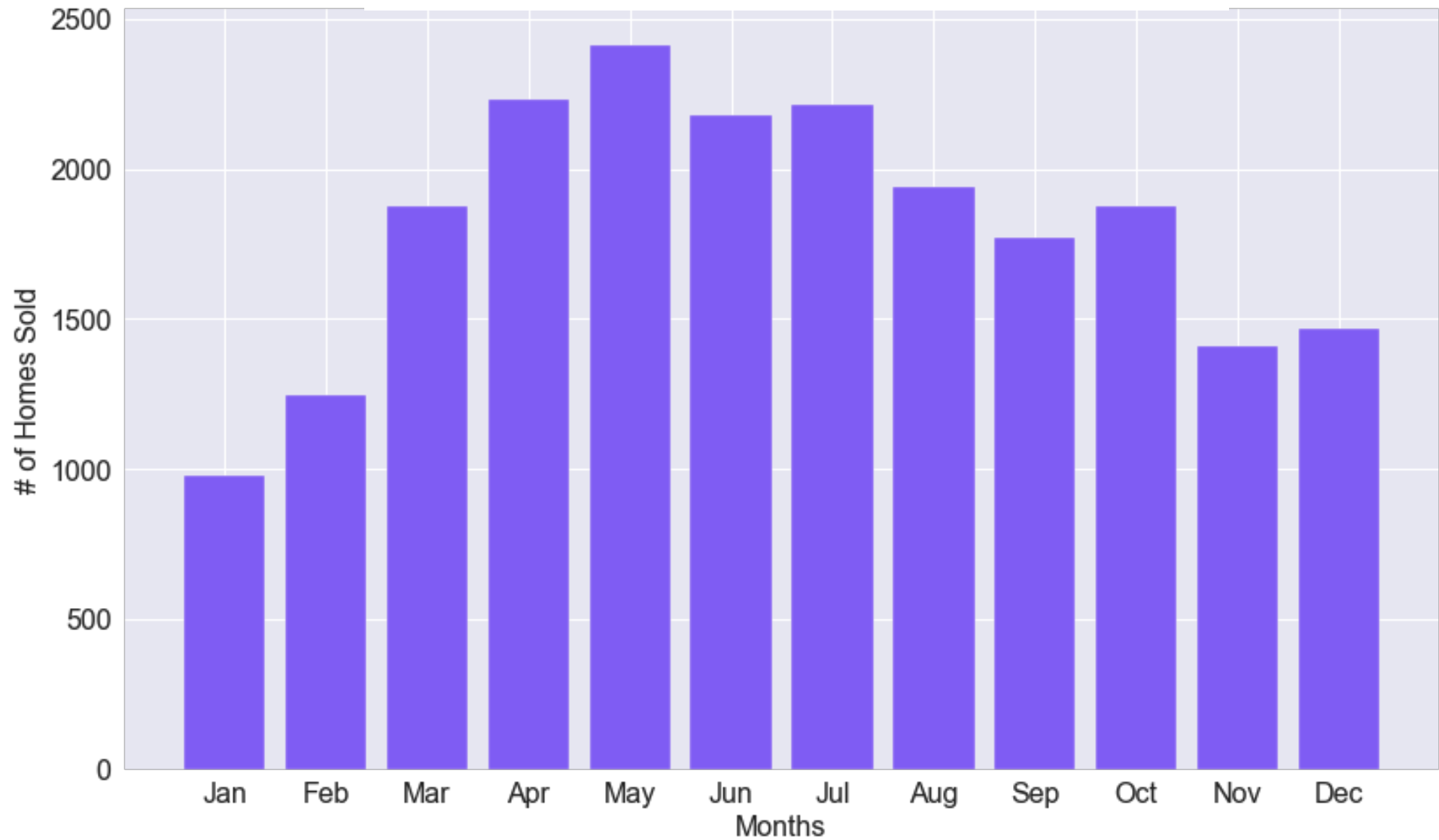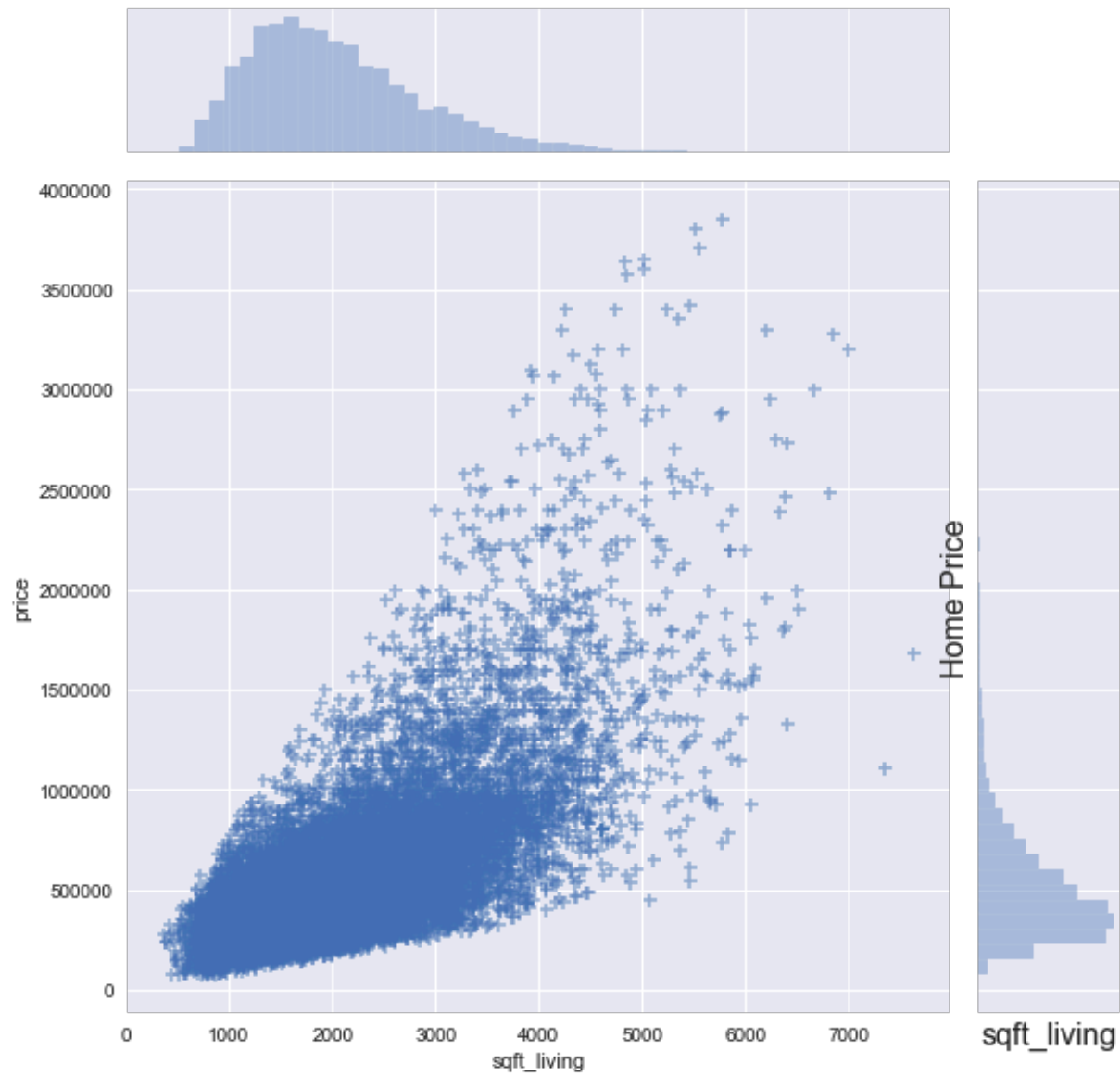
WHAT IS THE BEST DAY OF THE WEEK TO SELL A HOME?
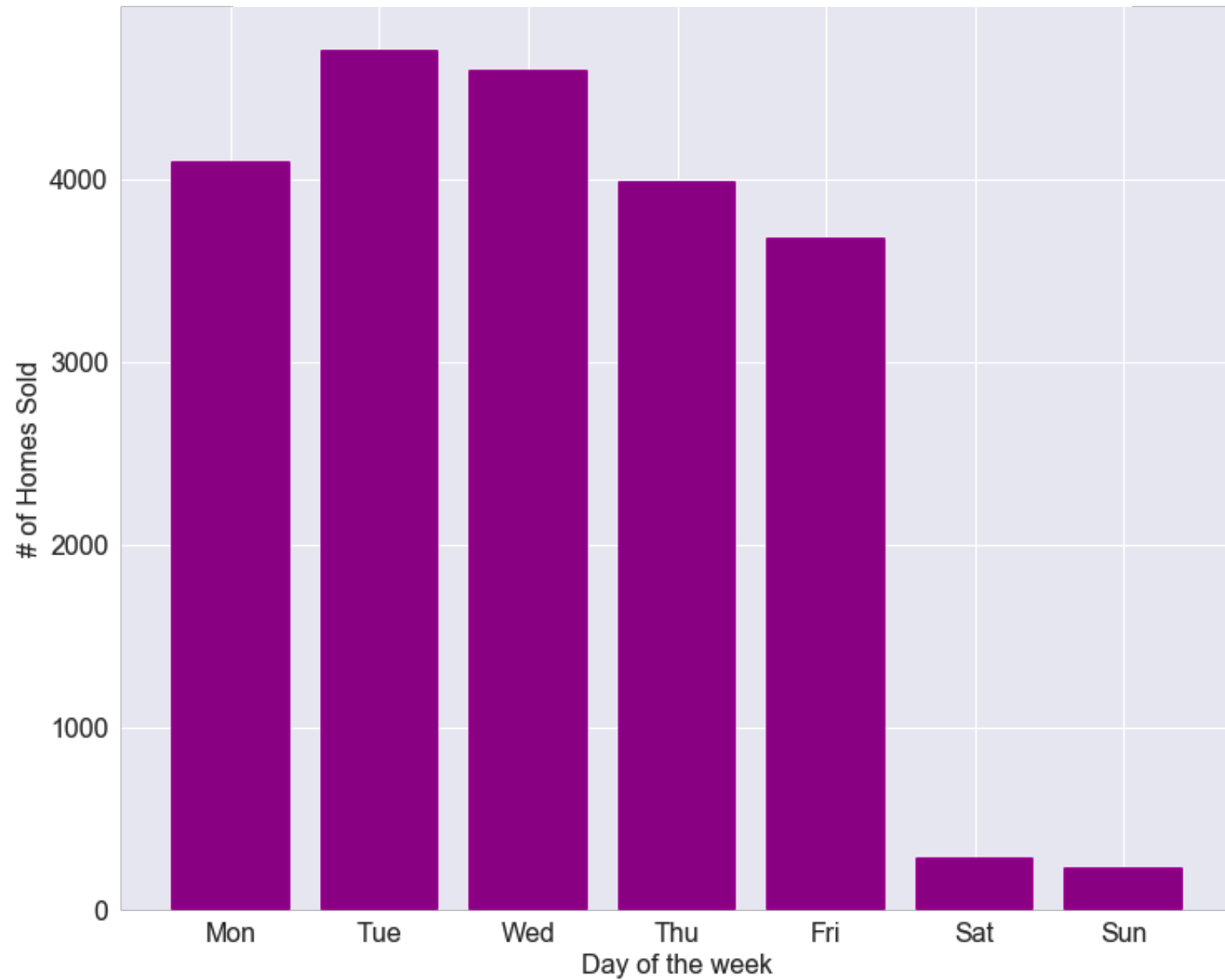
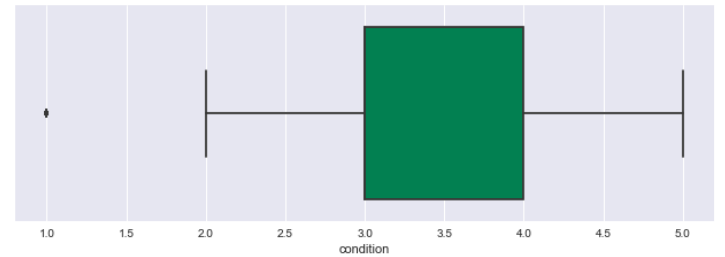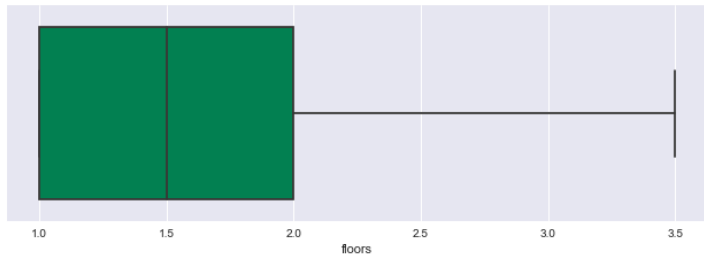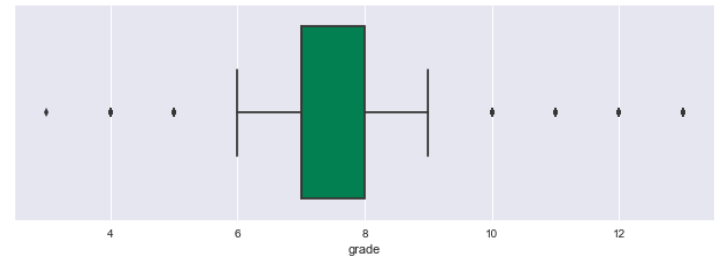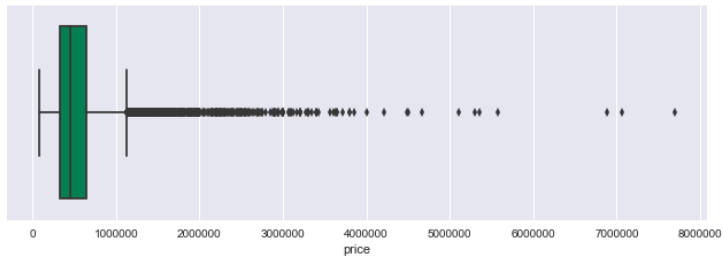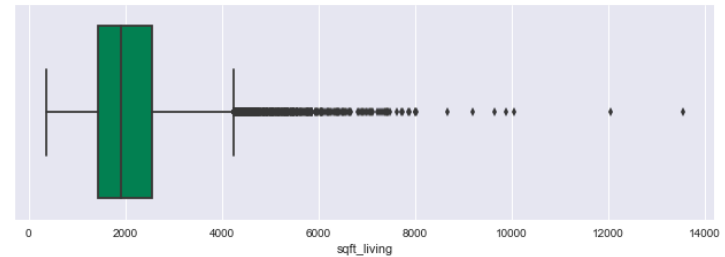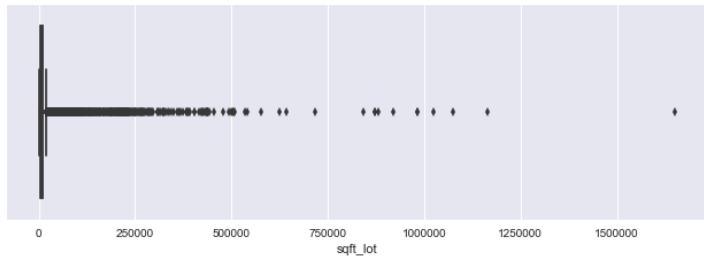# WHAT ARE THE TOP 10 DATES OF THE MOST NUMBER OF HOMES SOLD?

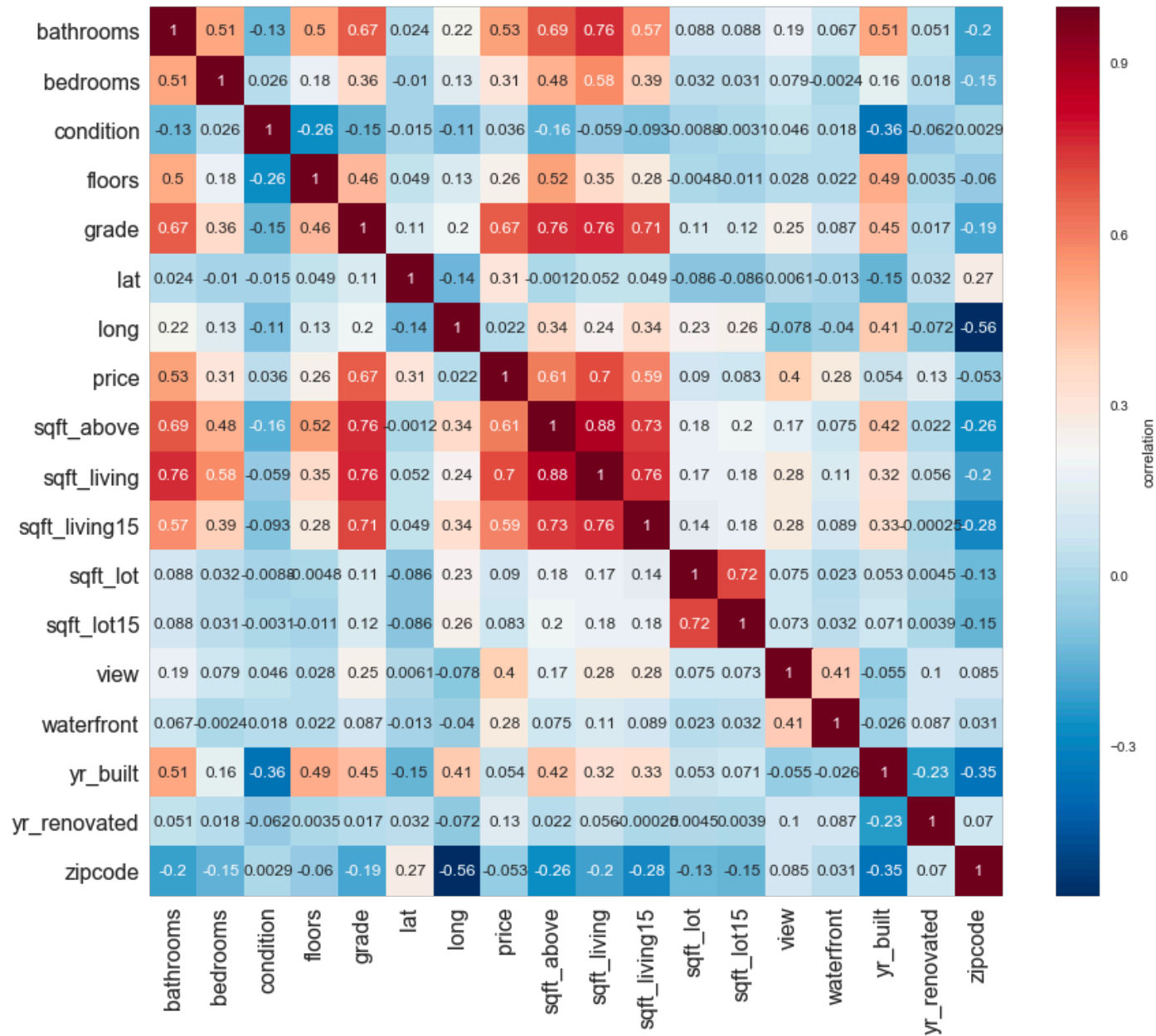
Top 10 Dates of Most Number of Homes Sold
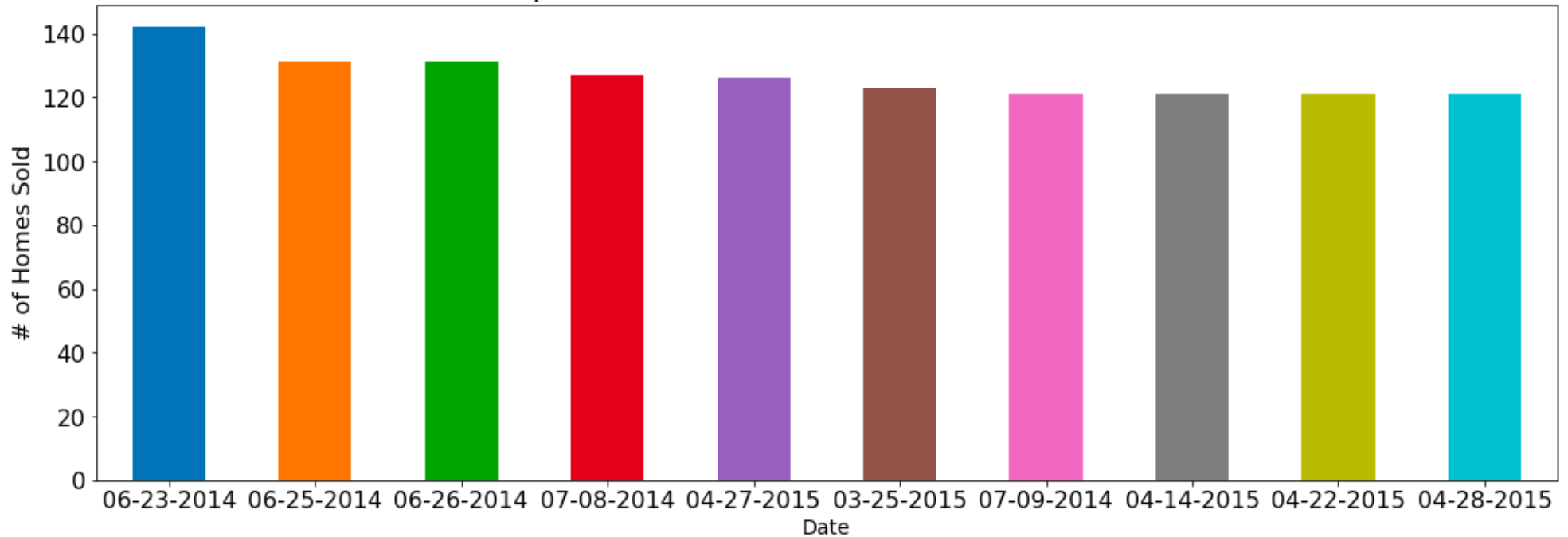
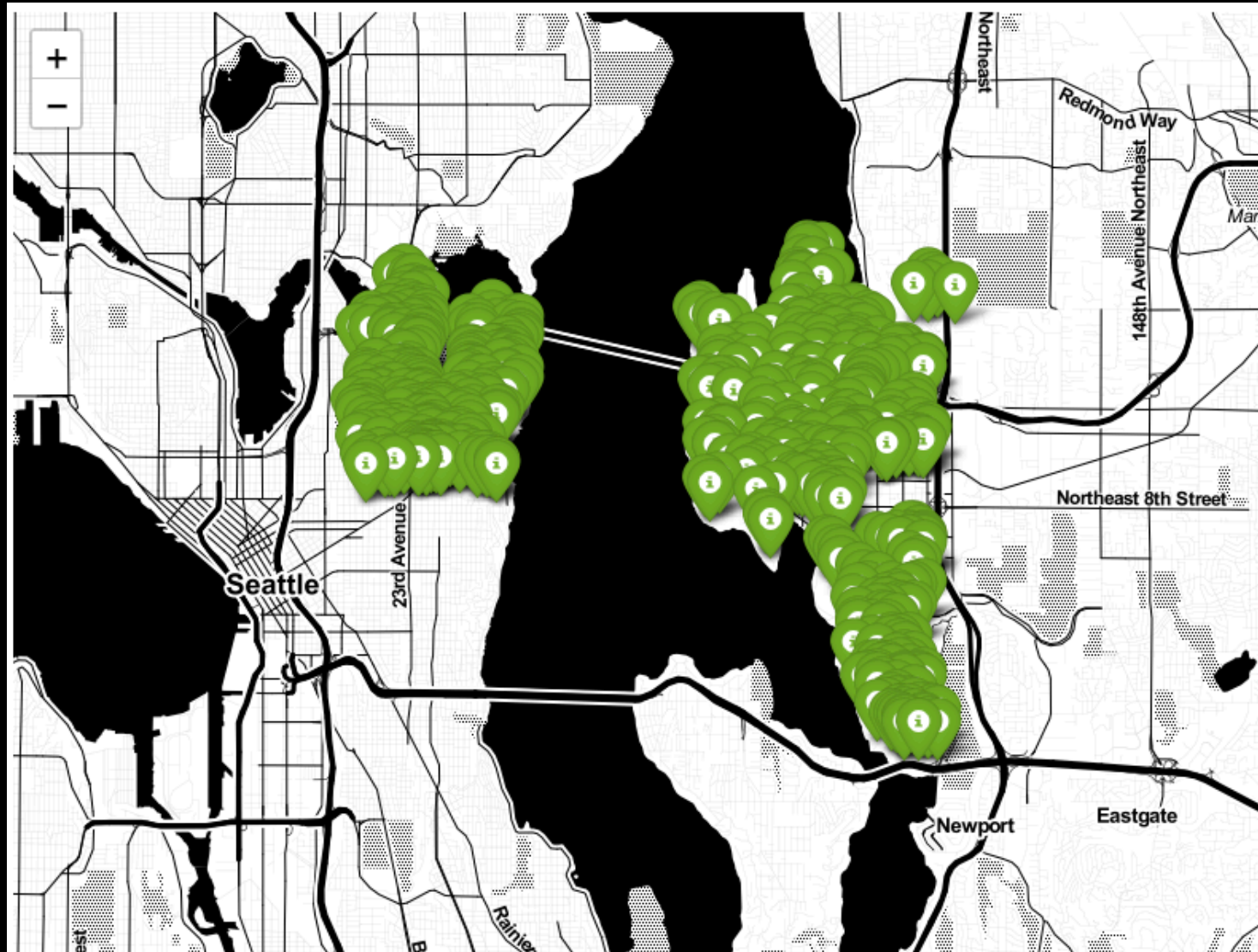# IS SQUARE FOOTAGE OF LIVING SPACE MOST IMPACTFUL ON HOME PRICING?

ARE THERE ANY OUTLIERS IN OUR DATA SET THAT WILL AFFECT PRICE ESTIMATION?
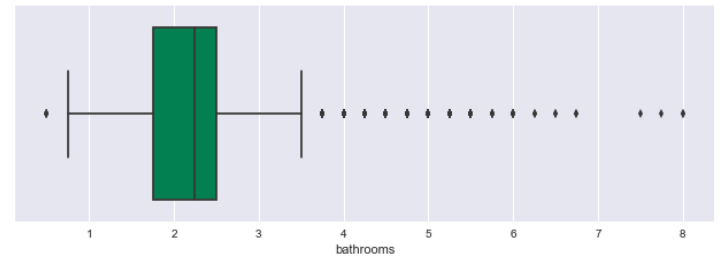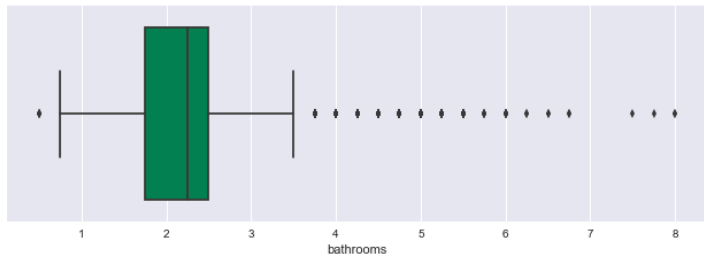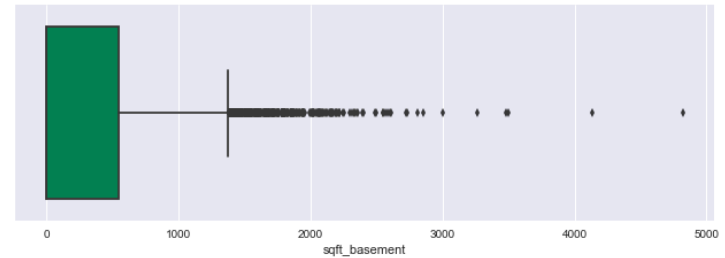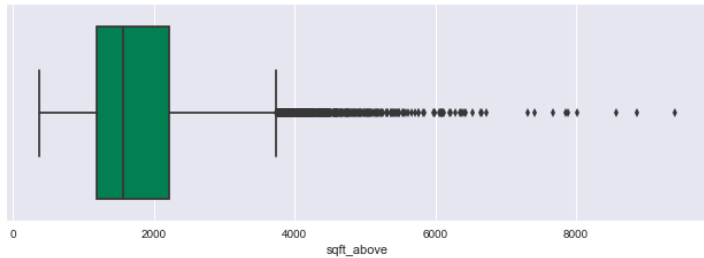
CORRELATION BETWEEN HOME PRICE FEATURES

TOP ZIPCODES

INTERPRETATION

In this project, we were able to determine the relative influence of one or more predictor values to analyze the value of houses. The model predicted the top features that impact home sale price as follows:
The square footage of living area, location of the homes (zip codes 98004, 98039 and 98112) have a strong correlation to the price of a home.

How can we translate the data into higher home sale price? Investigating how adding more bathrooms or bedrooms or renovations could potentially increase the home value or investing in specific locations. Future enhancement of the model includes review of the completeness of data and larger sample to ensure accurate appraisal. Including other predictors like proximity to transportation, school rating, parks, buyer's age, or type of home community will improve the House Price Model.