

EDA PROJECT

KING COUNTY, SEATTLE (USA)
HOUSING DATA

LUKAS KAUFMANN

22.09.2023

- I. introduction
- II. data overview
- III. hypotheses
- IV. eda steps
- V. analysis & visualisation
- VI. conclusions

I. CLIENT INTRODUCTION

- | William Rodriguez
- | Buyer
- | Two people (want two houses)
- | country house (best timing & non-renovated)
- | & city house (fast & central location)

* added information after talk with clients:

- | budget of 400.000 \$ per house
- | looking for two single apartments (limited size)
- | no multistory houses / apartments (live accessibly in the future)
- | want to renovate their country house themselves
- | city house should be in good condition

II. DATA OVERVIEW

- | House & Sales Details
- | Data collected between 2014/05 and 2015/05
- | >21.000 Rows and >20 Columns

III. HYPOTHESES

A. Questions:

- | When is the best time to buy a (country) house?
- | Are the houses in the country side bigger?
- | Are the houses in the city more expensive?

B. Hypotheses:

- | There is no time of the year where it is “better” to buy a house.
(price drops within a year or month, etc.)
- | House prices in King County generally increase over time.
- | The newer the house, the better building grade it has.
(construction quality / meets code).
- | The better rated a house, the more expensive it is.

IV. STEPS OF EDA

Step 0: Research Questions

- What do I want to find out?
- > Generate Hypotheses

Step 1: Data Understanding / Exploring

- Dataframe shape
- head, tail, columns
- dtypes
- describe
- info, etc.

Step 2: Data Cleaning

- Remove unnecessary columns
- Identifying duplicated rows/columns
- Renaming Columns
- Turn integers into date-time-objects
- Check/handle missing values
- Look for and handle possible errors

Step 3: Data Preparation

- filter houses according to clients needs and wishes
- Make decisions about which houses to recommend to clients

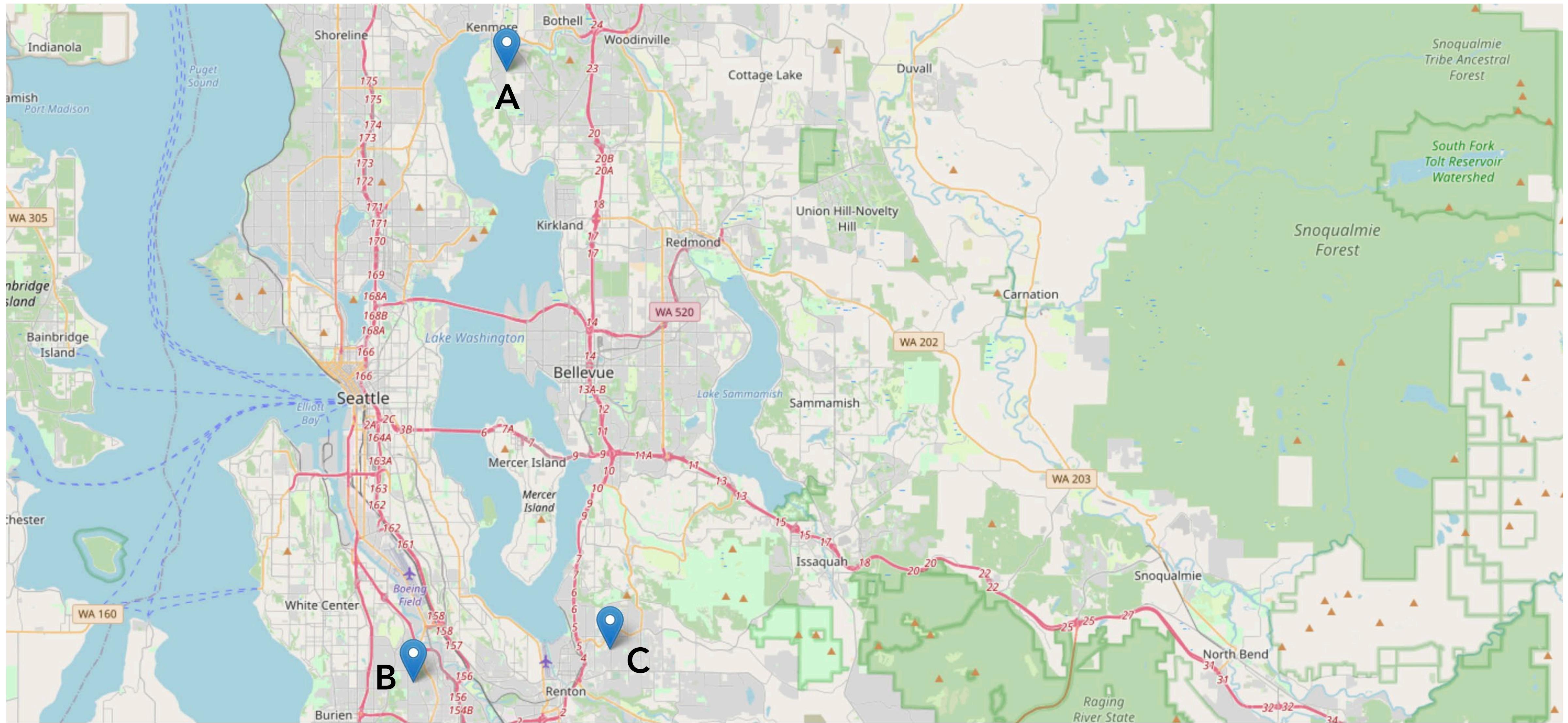
Step 4: Feature Understanding & Relationships - Data Visualisation

- Map locations
- Boxplots
- Scatterplots
- Heatmap
- Pairplots
- (Multi-)Lineplots
- Compare Top Houses

Step 5: Answer questions about the data

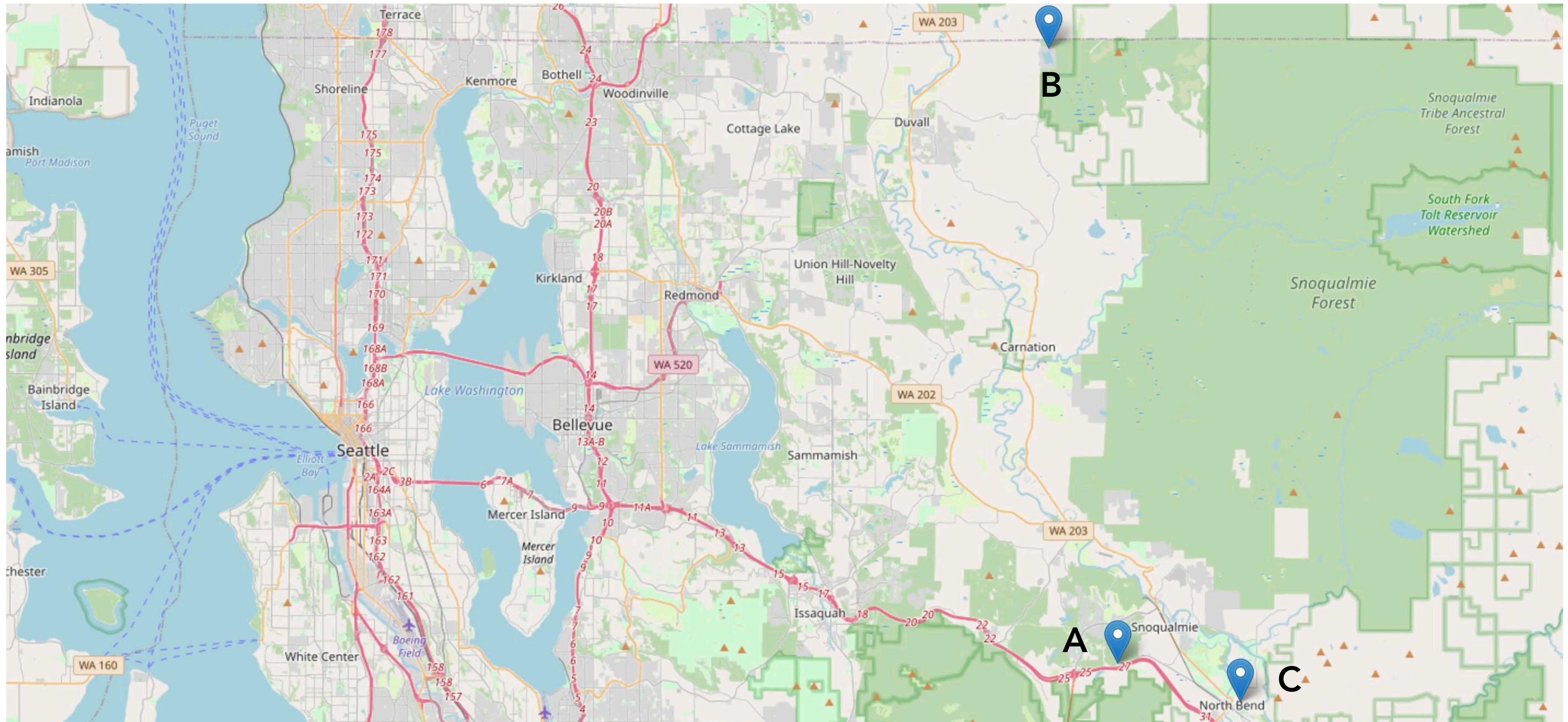
- When is the best time to buy the country house?
- Show top three country and city houses
- Insights and recommendations for client

V. ANALYSIS & VISUALISATION



map 01 - top three city house locations

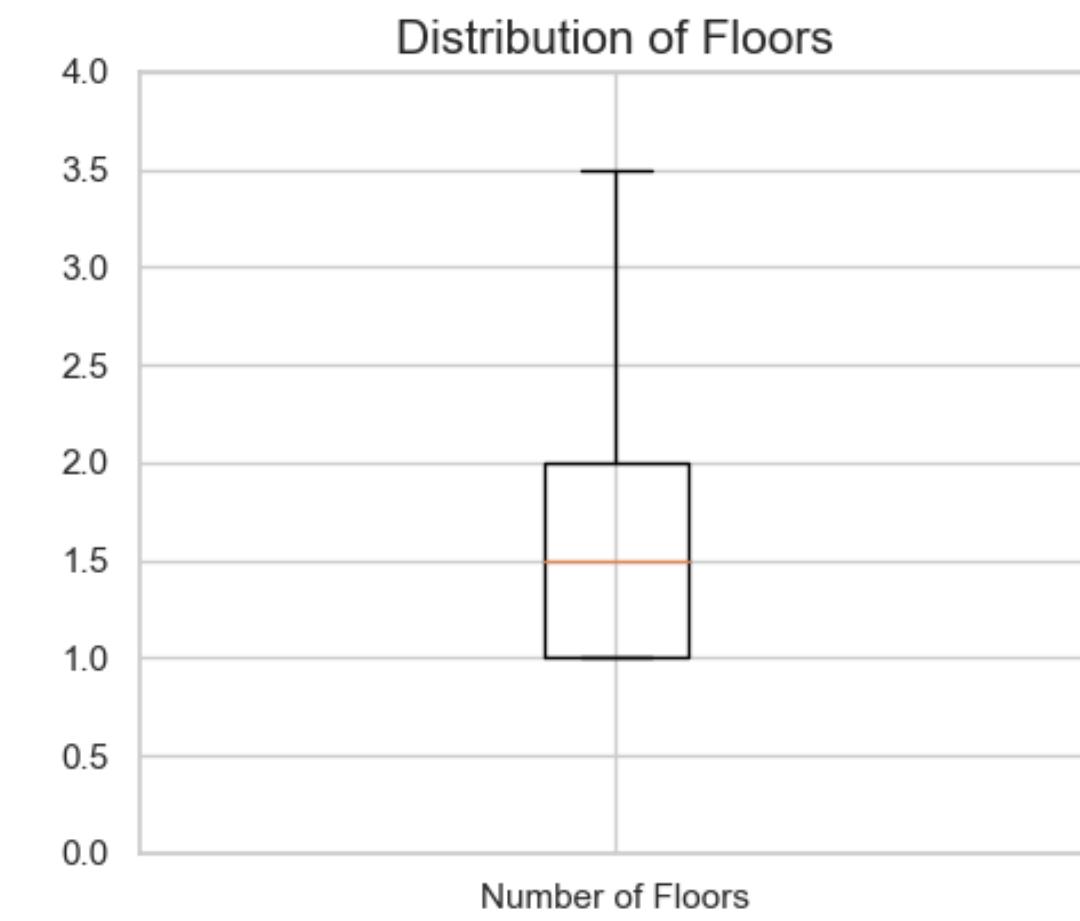
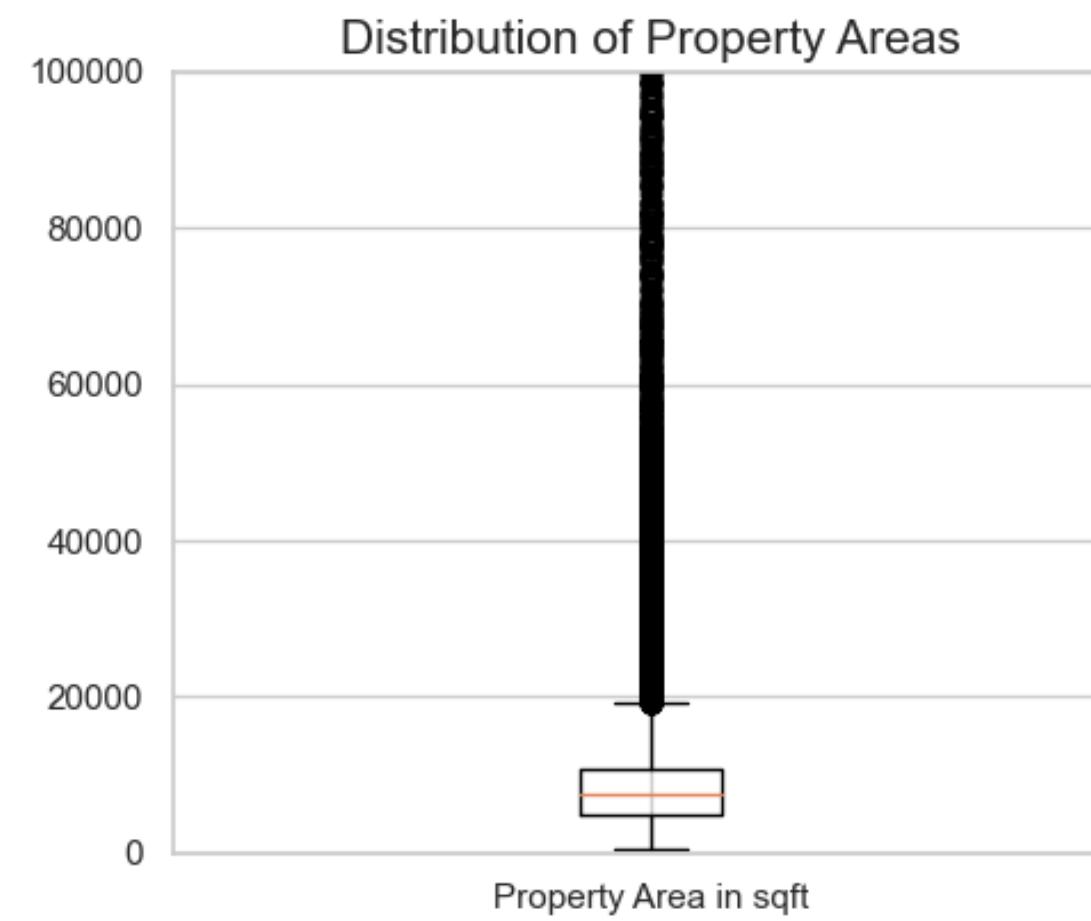
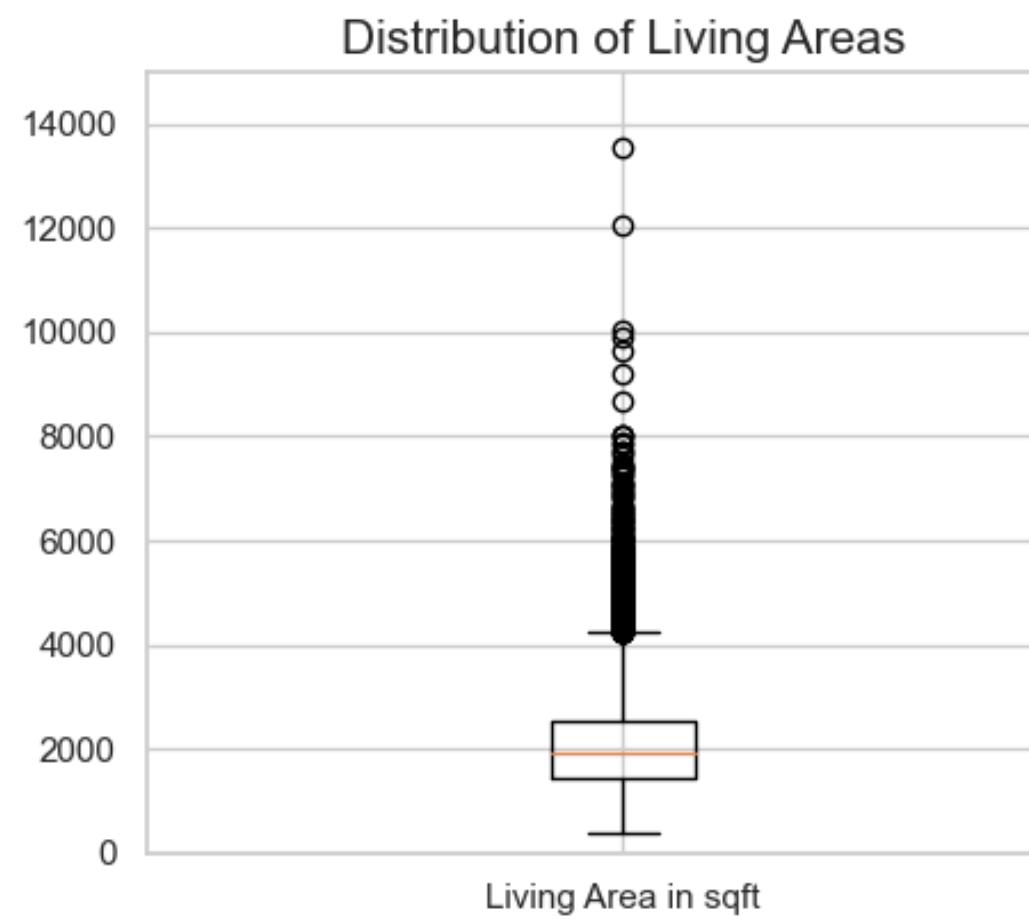
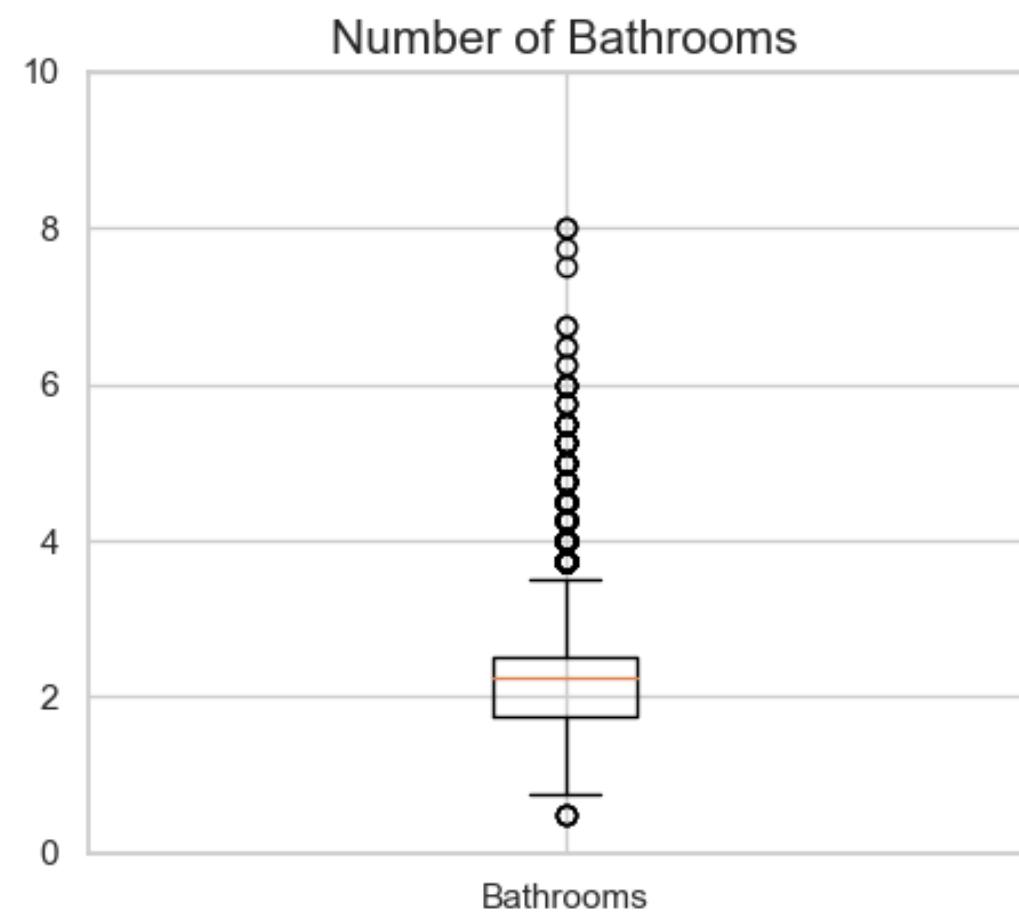
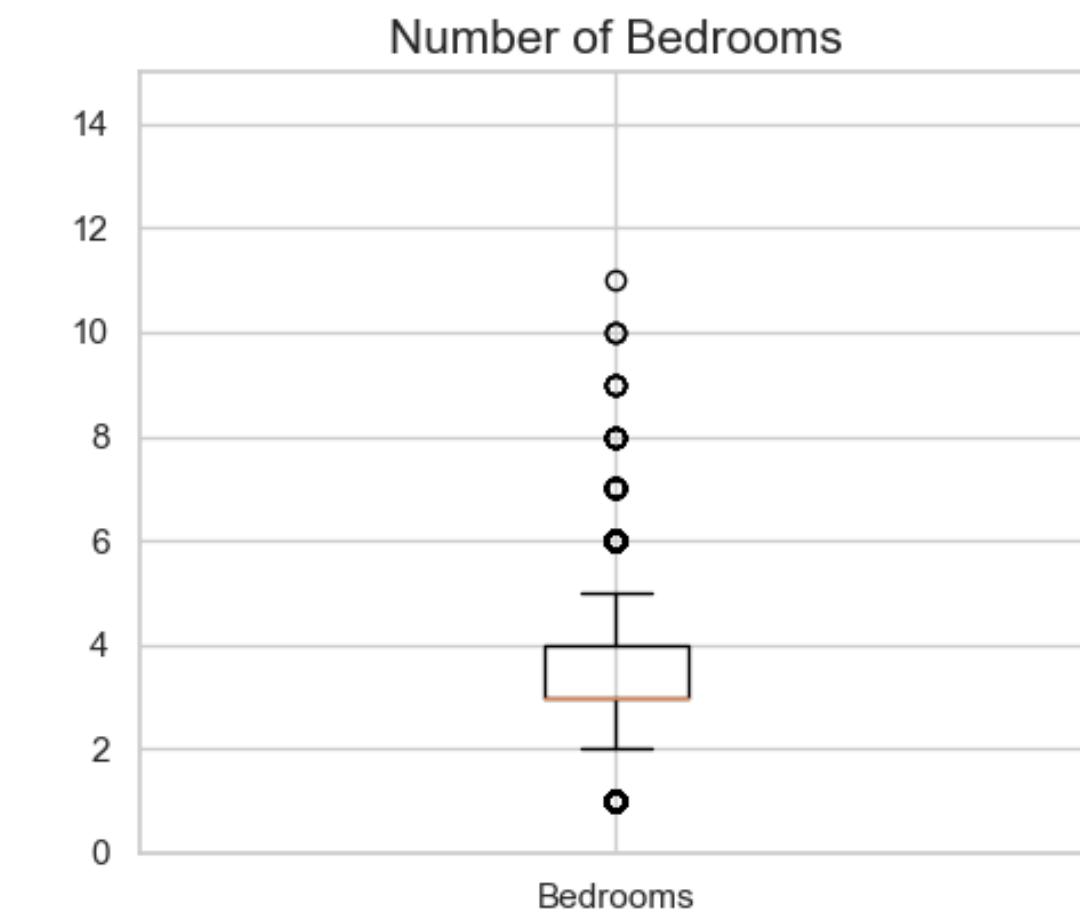
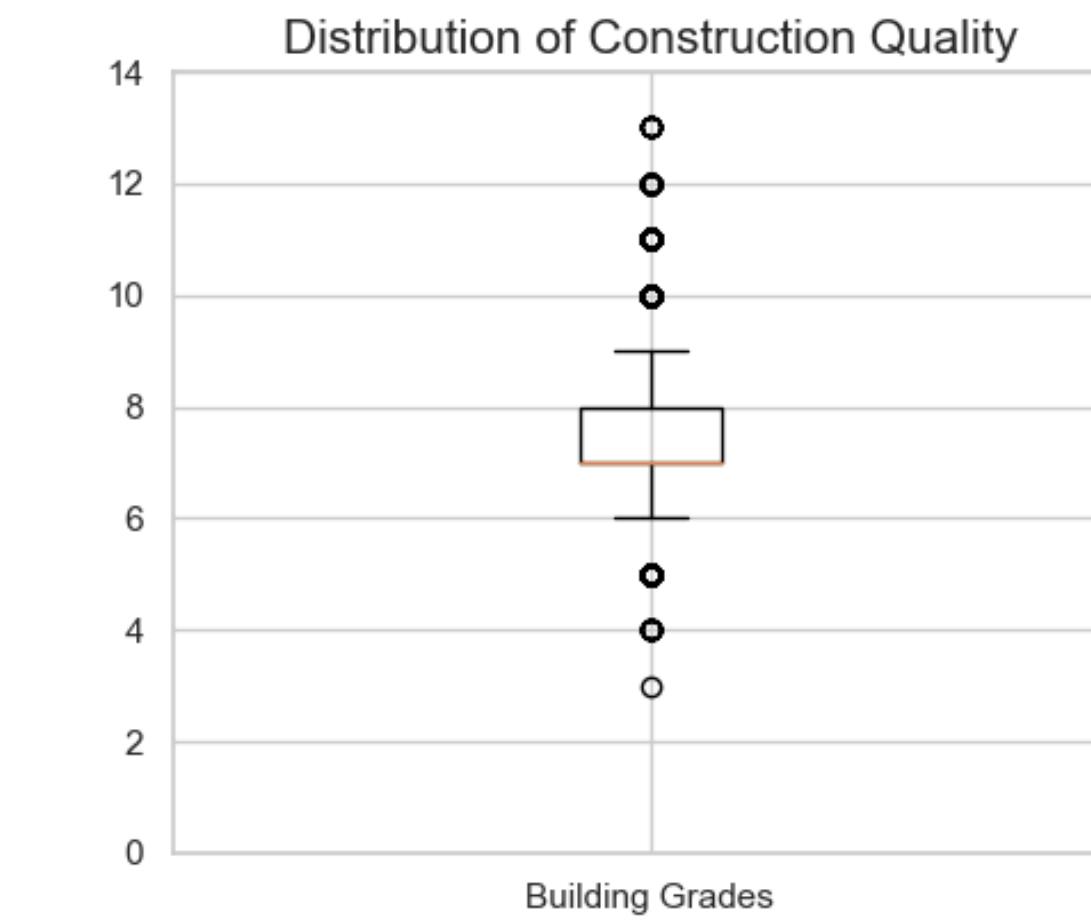
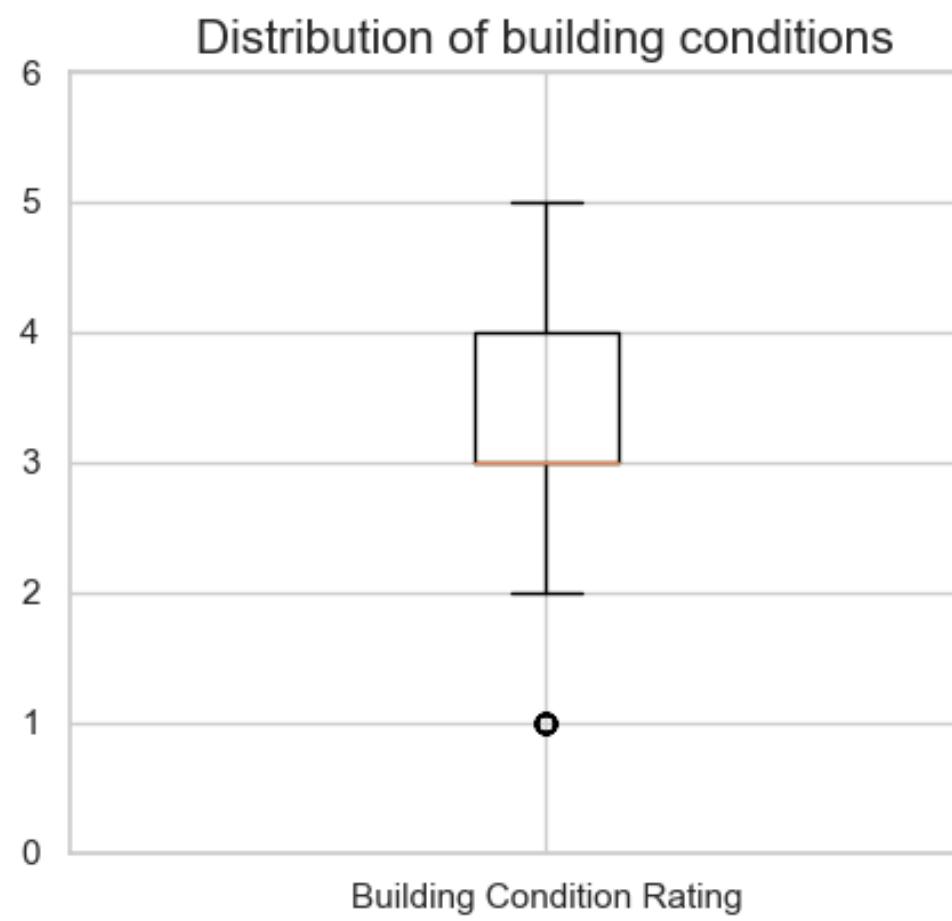
V. ANALYSIS & VISUALISATION



map 01 - top three country house locations

V. ANALYSIS & VISUALISATION

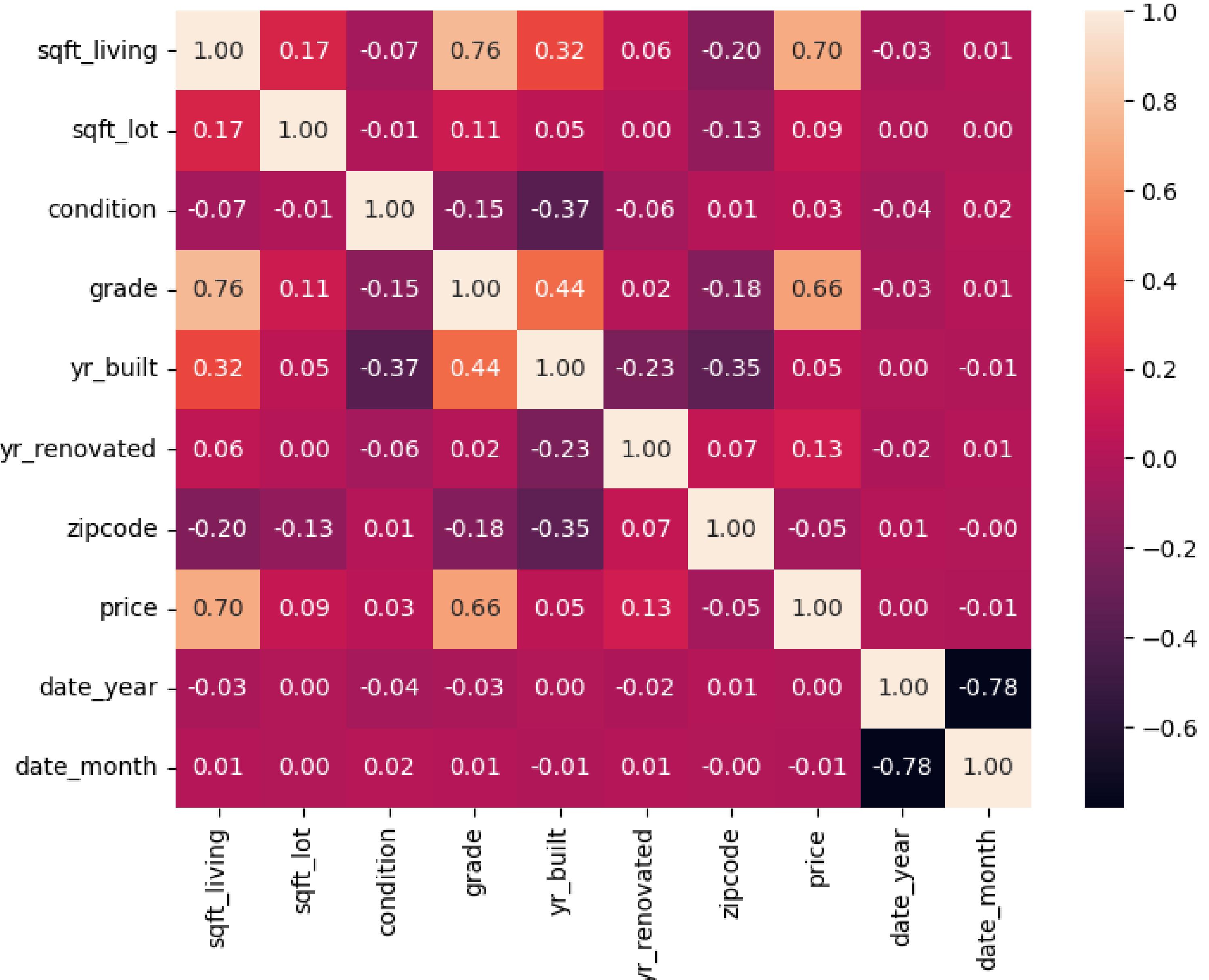
Distribution of Numeric Housing Values in King County, Seattle (USA)



boxplot - average distributions of numeric housing values

V. ANALYSIS & VISUALISATION

- year_built has an effect on the grade and to sqft_living
- grade and sqft_living have a high correlation to the price of the house
- grade and sqft_living are related to each other as well

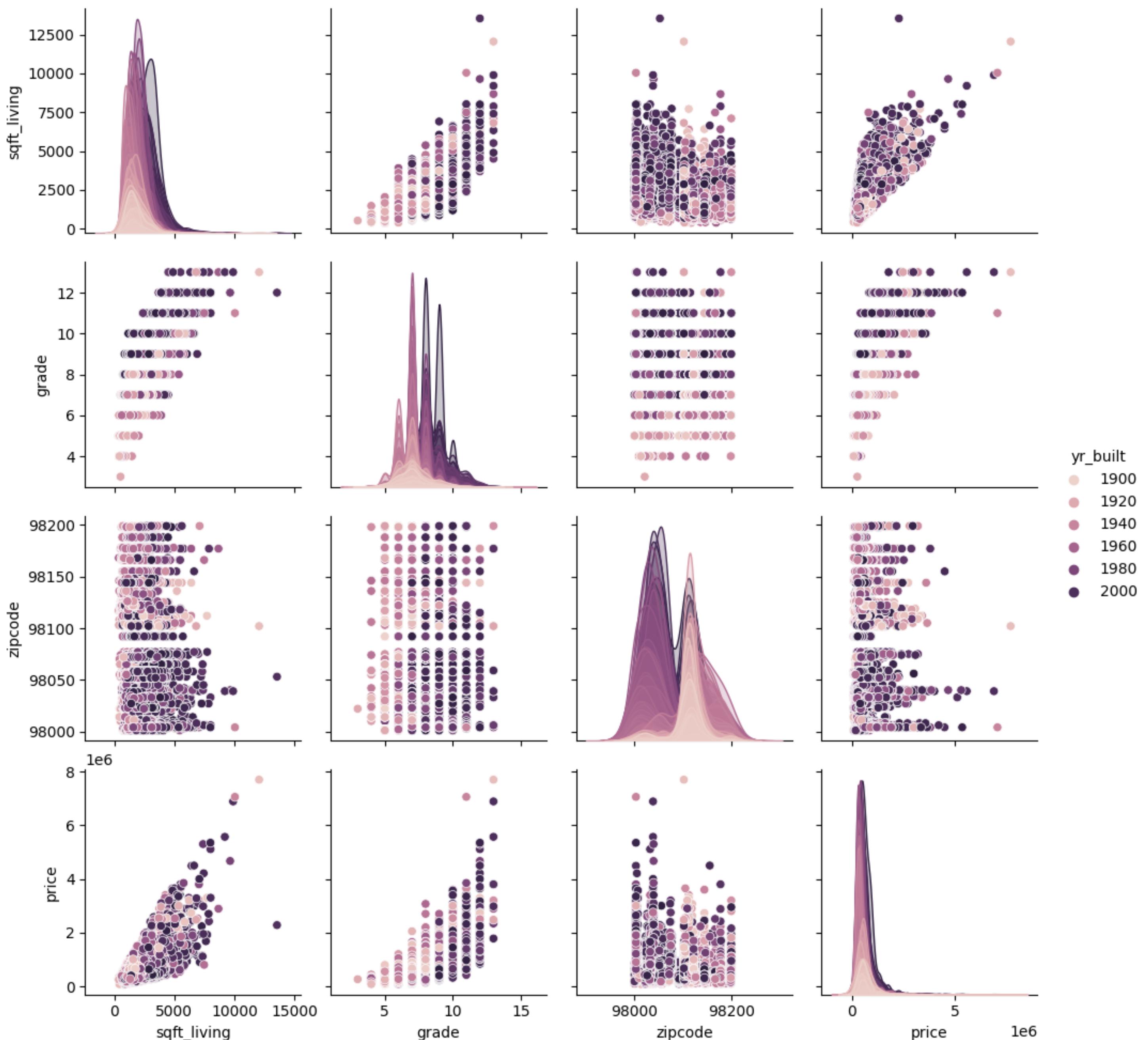


heatmap - feature correlation values

V. ANALYSIS & VISUALISATION

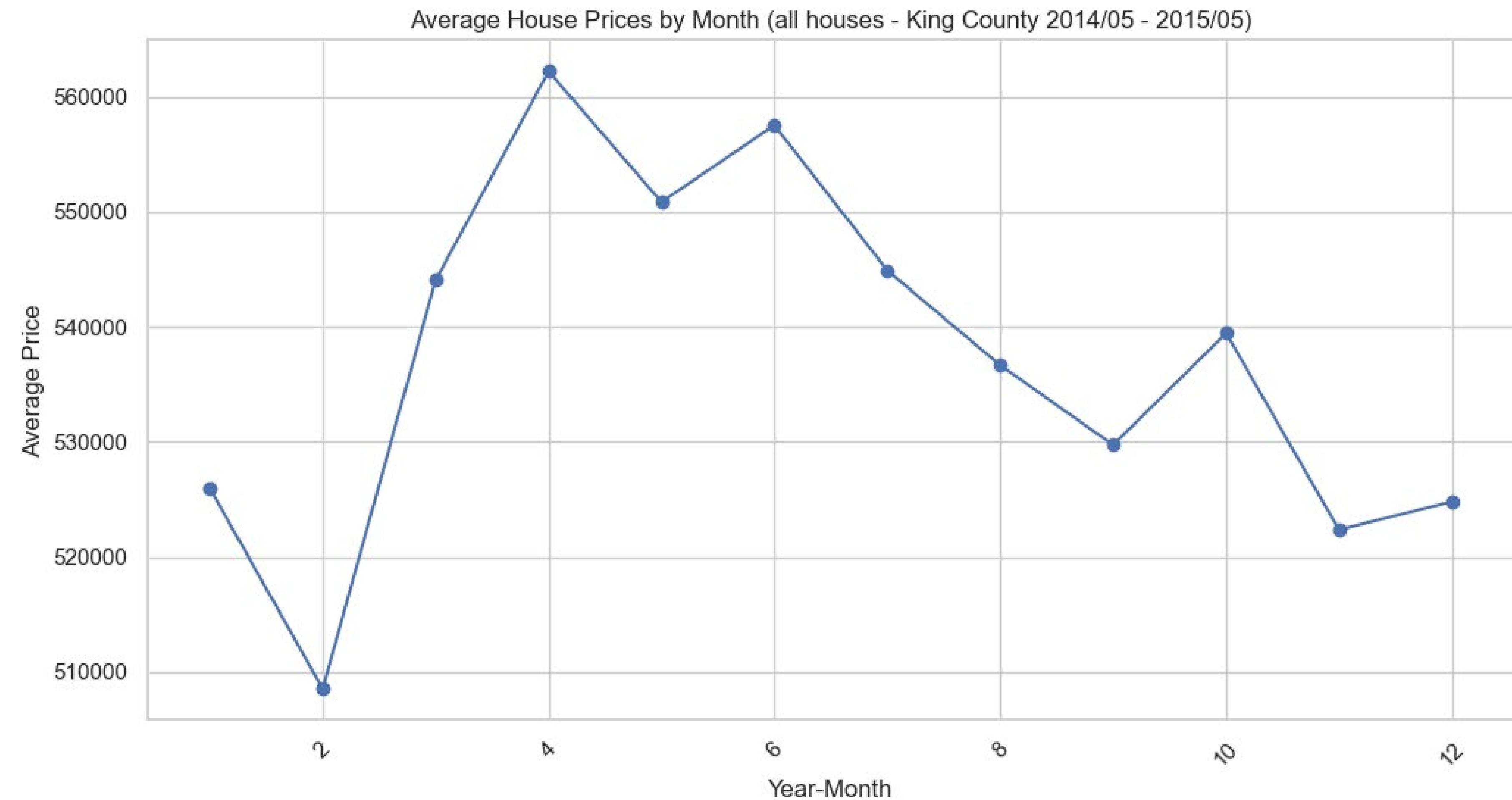
- older houses are smaller in size
- newer houses have better grades
- The better rated and the bigger a house, the more expensive it is
- The price range is rather evenly distributed throughout the zipcodes
- Is the time of the sale related to the price?

pairplot - correlation visualisations



V. ANALYSIS & VISUALISATION

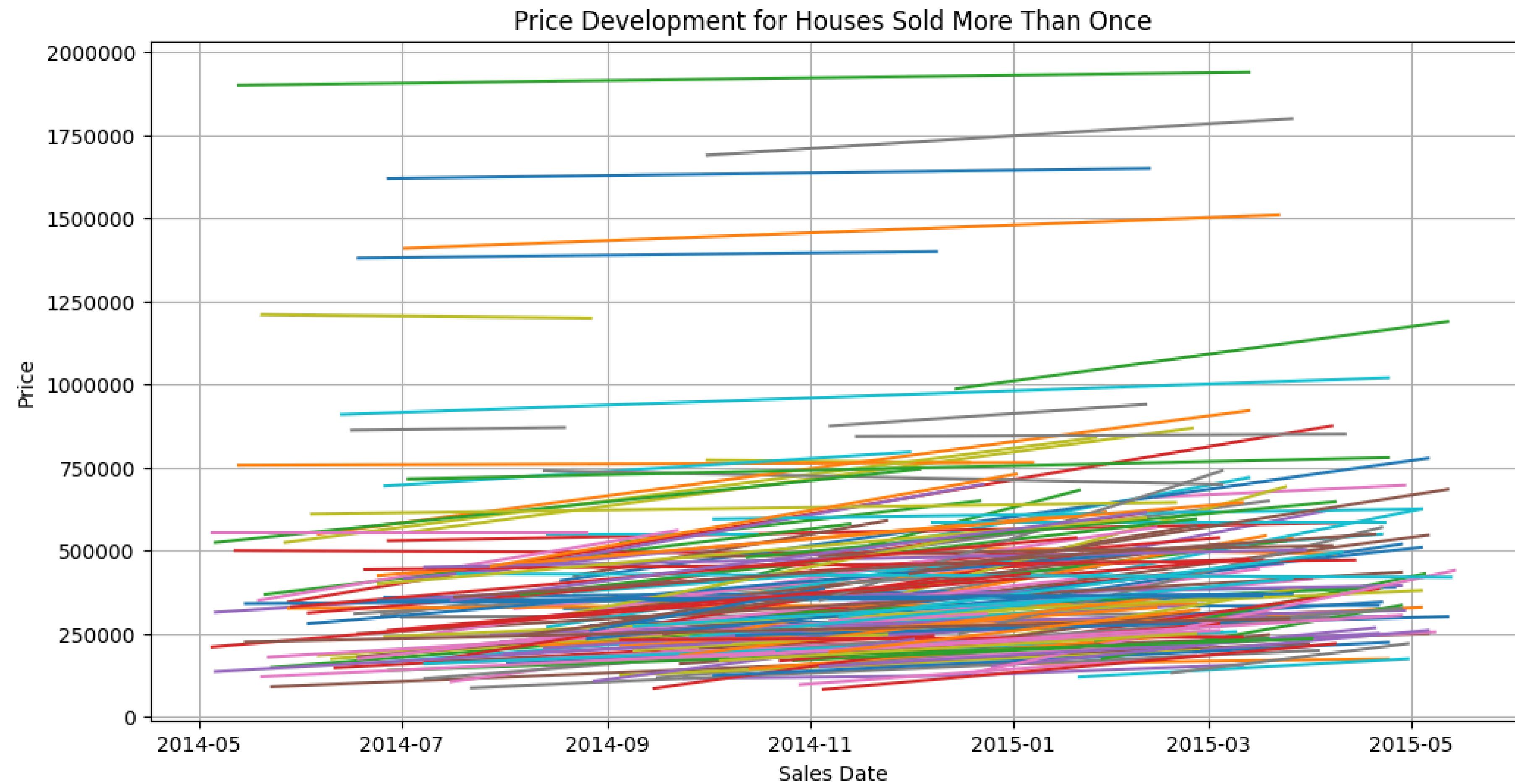
- no long term data
- too many factors
- random distribution



lineplots - average house prices by month of sale

V. ANALYSIS & VISUALISATION

- as expected:
upward trend
- price decrease:
only 3.4%
(6 of 176 houses)
- = outliers/exceptions



lineplots - price development for houses sold more than once

V. ANALYSIS & VISUALISATION

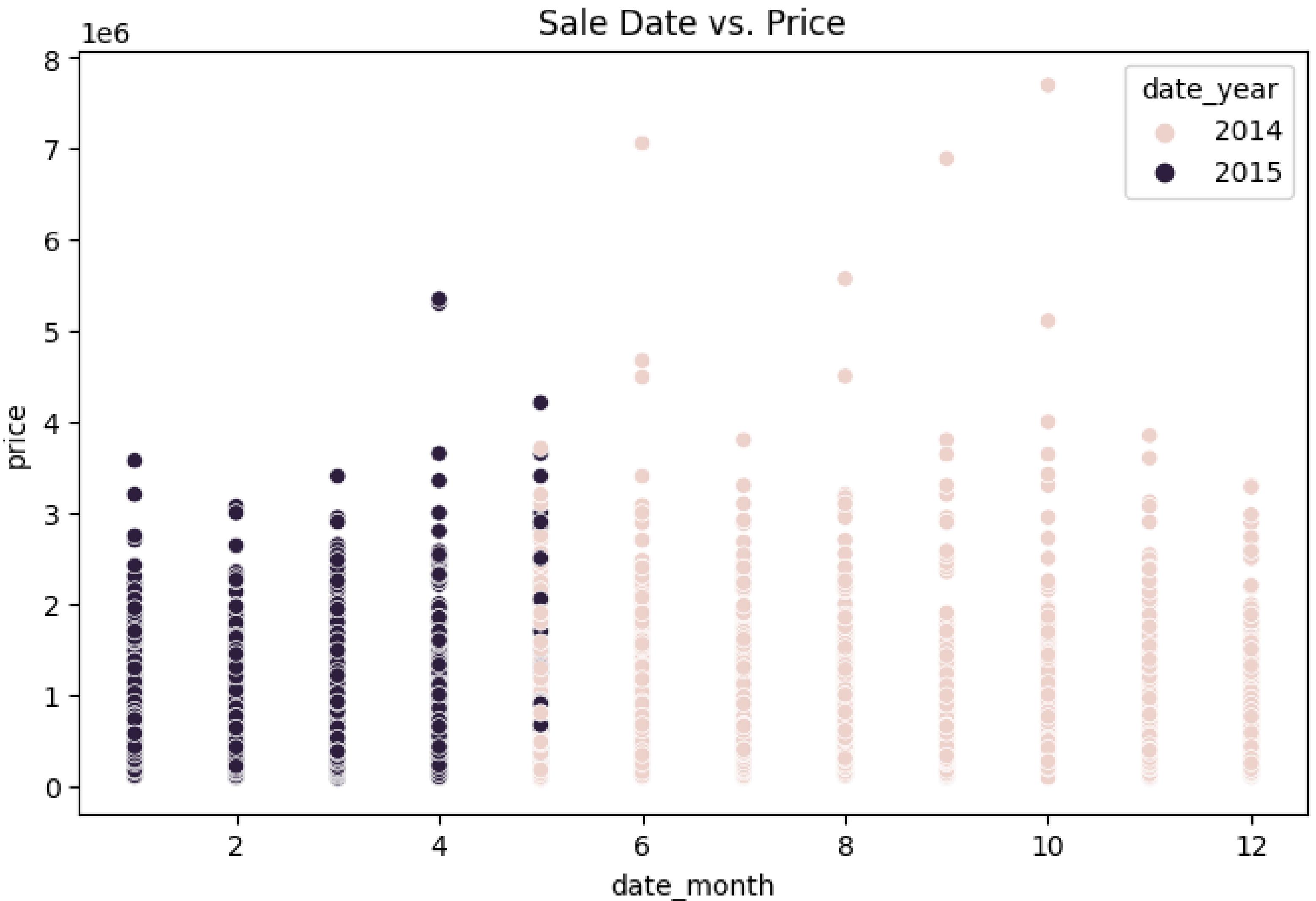
Conclusion:

Analyzing the “best time” to buy a house based solely on the month is challenging for several reasons:

- real estate prices are influenced by a multitude of factors (market conditions, economic factors, location, individual property attributes, etc.)
 - No long-term trends available. (Real estate trends often follow long-term cycles.)

Recommendation:

- > the best time to buy a house depends on individual circumstances and goals.



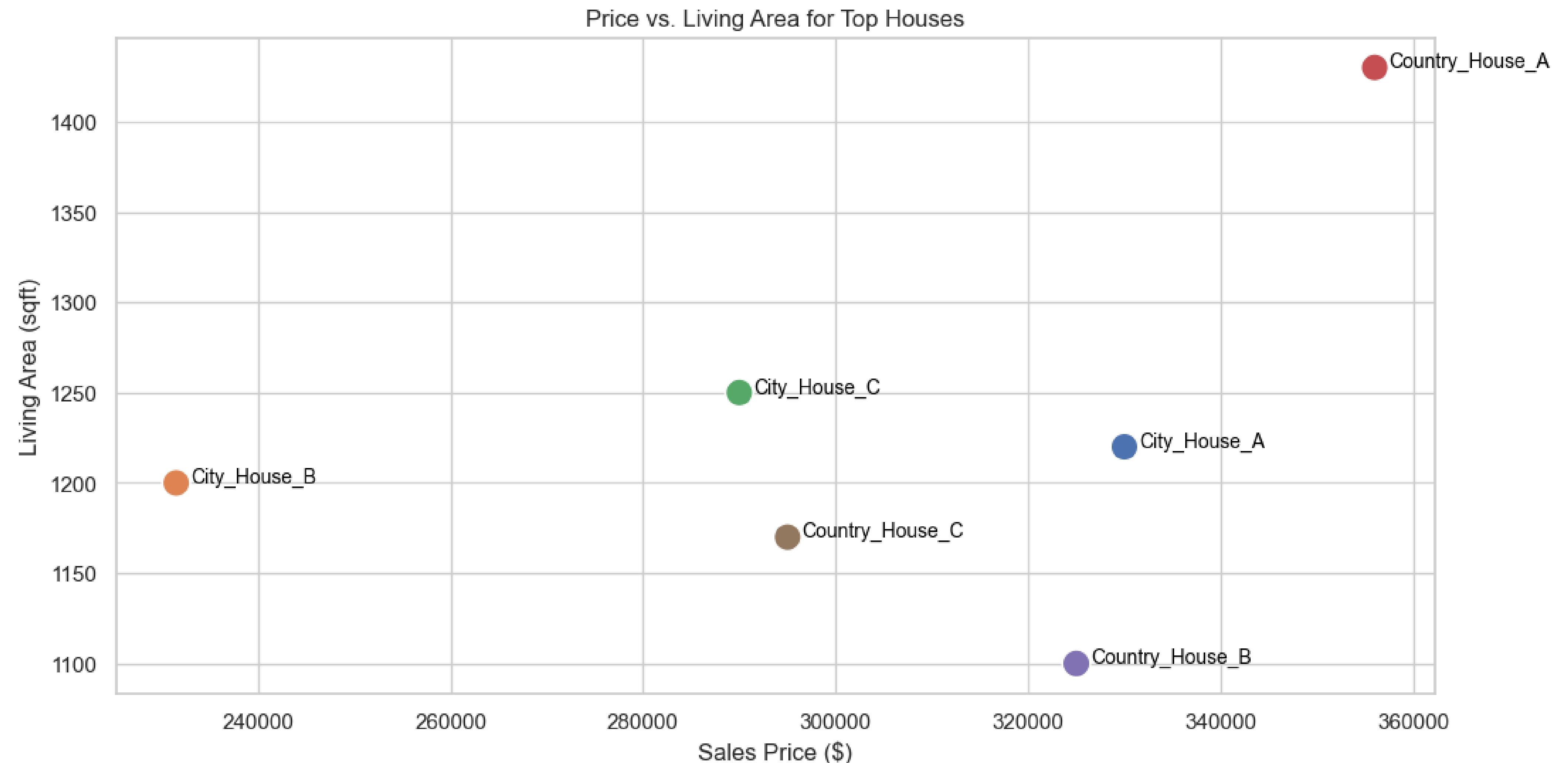
scatterplot - distribution and relation between month of sale and price

V. ANALYSIS & VISUALISATION

	Top_3	price	bedrooms	bathrooms	sqft_living	sqft_lot	condition	grade	yr_built	zipcode
0	City_House_A	330000.000	2.000	1.000	1220.000	10000.000	5	7	1950	98028
1	City_House_B	231500.000	2.000	1.000	1200.000	9488.000	5	7	1941	98168
2	City_House_C	290000.000	2.000	1.000	1250.000	12507.000	5	7	1958	98056
3	Country_House_A	356000.000	2.000	1.000	1430.000	365904.000	3	7	1991	98065
4	Country_House_B	325000.000	2.000	1.000	1100.000	17817.000	3	7	1980	98019
5	Country_House_C	295000.000	2.000	1.000	1170.000	10621.000	3	7	1963	98045

data - top houses

V. ANALYSIS & VISUALISATION



scatterplot - top houses - price vs. linging area

V. ANALYSIS & VISUALISATION



scatterplot - top houses - price vs. property area

V. ANALYSIS & VISUALISATION



house images - country house a

V. ANALYSIS & VISUALISATION



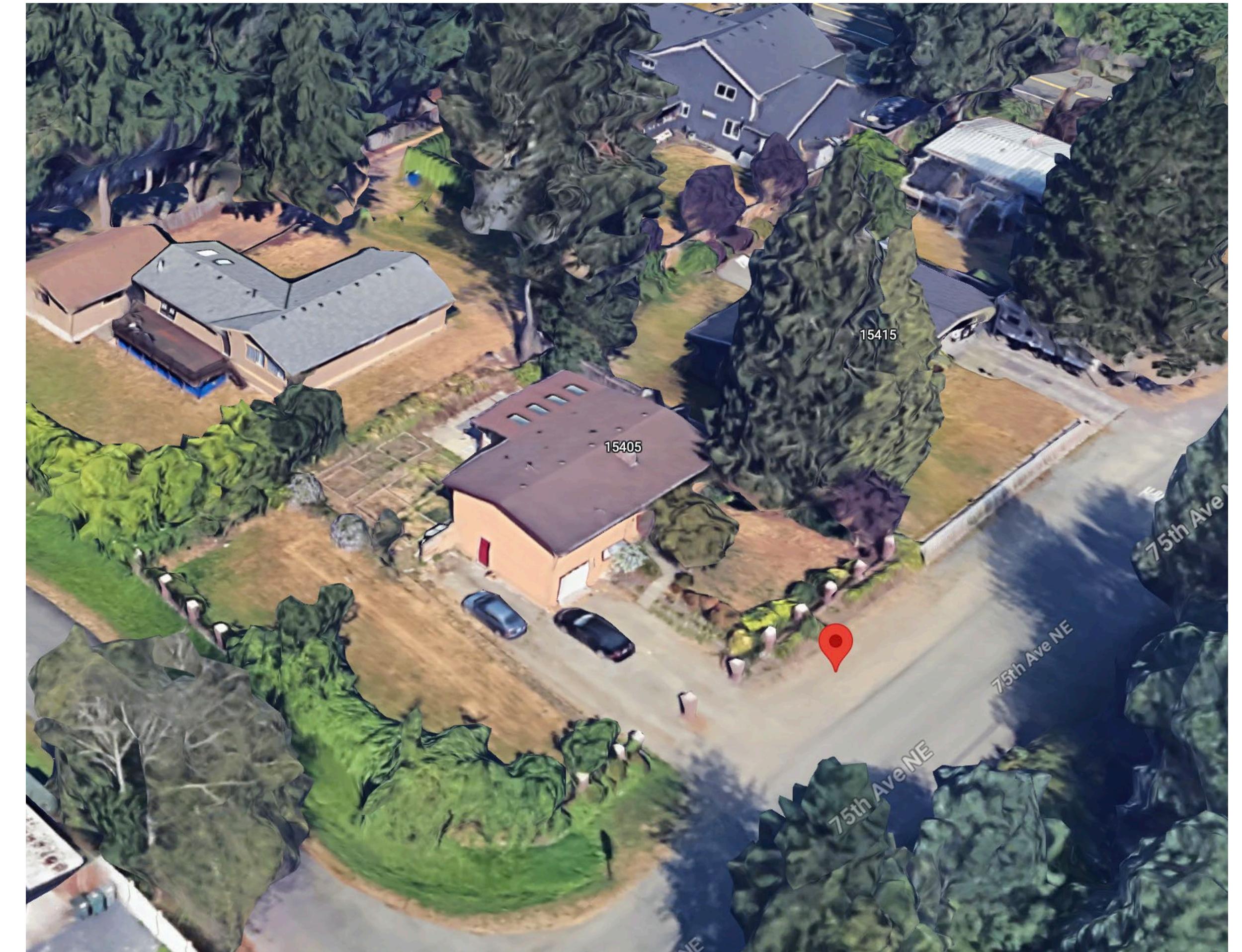
house images - country house b

V. ANALYSIS & VISUALISATION



house images - country house c

V. ANALYSIS & VISUALISATION



house images - city house a

V. ANALYSIS & VISUALISATION



house images - city house b

V. ANALYSIS & VISUALISATION



house images - city house c

VI. CONCLUSIONS

A. Insights:

- | Older houses tend to be smaller. Newer Houses tend to have better grades.
- | The better rated and the bigger a house, the more expensive it is usually.
- | The price range is relatively evenly distributed throughout the zipcodes.

B. Recommendations:

- | Country House A has the biggest plot, but Country House B is on a lake.
Country House B is recommended.
- | City House C has the best size-to-cost ratio and is also the newest of the Top 3 city houses.
- | The best time to buy a house depends on your individual circumstances and goals.
The sooner the better though, since prices tend to increase over time.