# KING COUNTY HOUSE SALES

## BY
### NORTHWEST REAL ESTATE AGENCY

# WHO ARE WE?

Northwest real estate agency

The bridge between you and your perfect home!!

## Data Description & Problem Statement

An analysis was carried out for King county data for the last two years.

The problem statement that was being tackled in this research was;

**To develop a predictive model that accurately estimates the sale price of houses in King County, Washington.
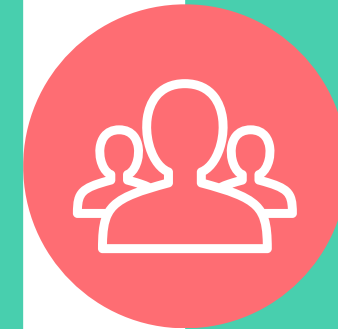
**Between the years 2014 and 2015 21,597 houses were sold to owners from different walks of life.

** House prices ranged from a maximum of 7.7b to a minimum 78k

** Houses were priced at an average of 540k.

# Data & Stakeholders

In the King county data the following variables were used:-

Target variable –
- Price.

Predictor variable –
- No. of Bedrooms
- No. of Bathrooms
- Square feet of living area
- Square feet of lot
- Waterfront
- Renovated
- Grade no
- House Age.
- No. of floors

The stakeholders in this research are the **potential homeowners.**
The researchers interest is to match the homeowners to their ideal homes

**Target variable is the item of interest.

** Predictor variables are the items that the model demonstrates that determine a change in the target variable.

# CORRELATION:- PREDICTOR TO TARGET

- **A test of correlation determines that the variables that most affect the house prices are the ones with a darker shade of red.**
  - **Strongest correlation to price was noted to be square feet of living area.**
  - **Weakest correlation to price was noted to be if the house was renovated.**



Correlation Heatmap

# Data Visualization
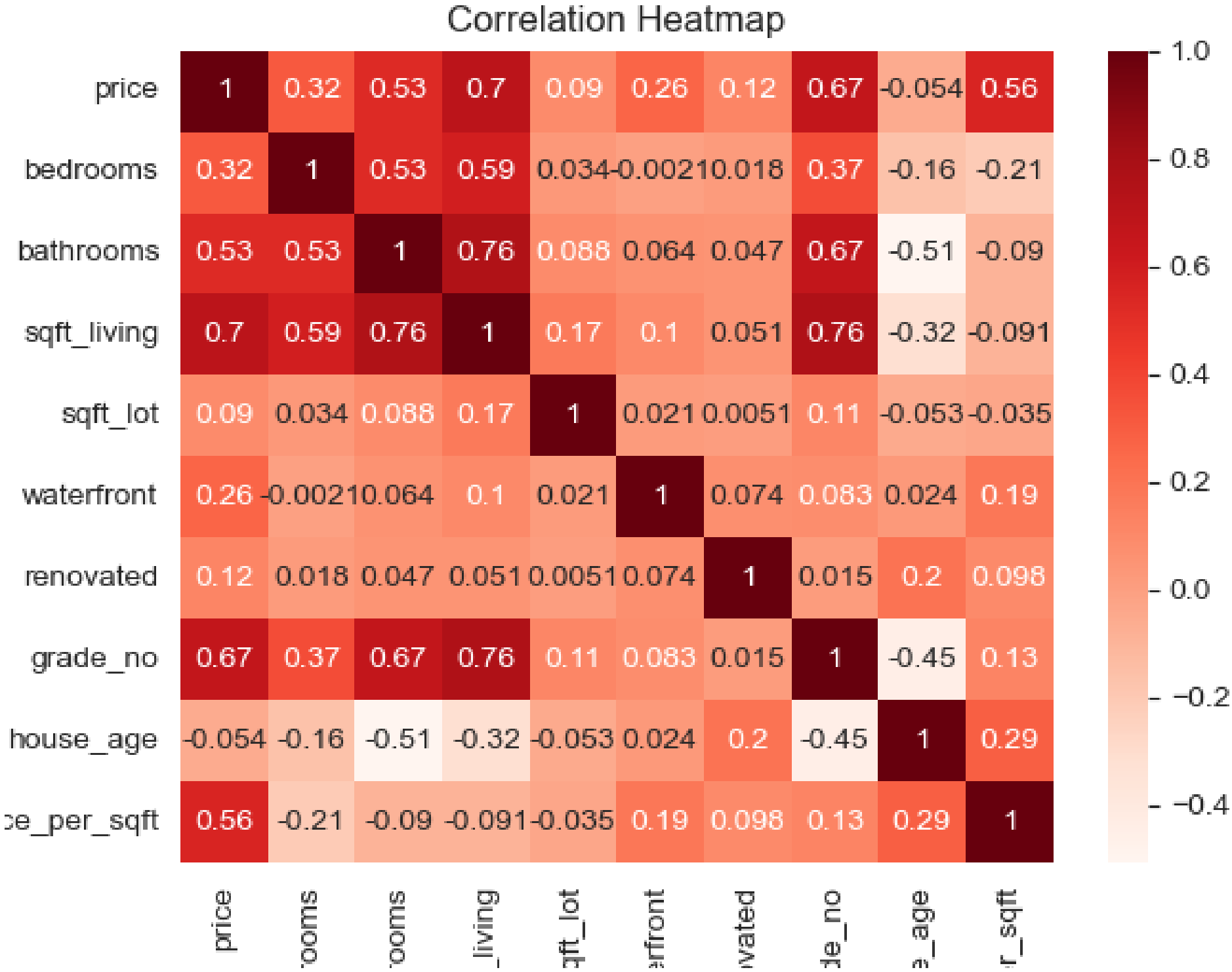

House prices by grade
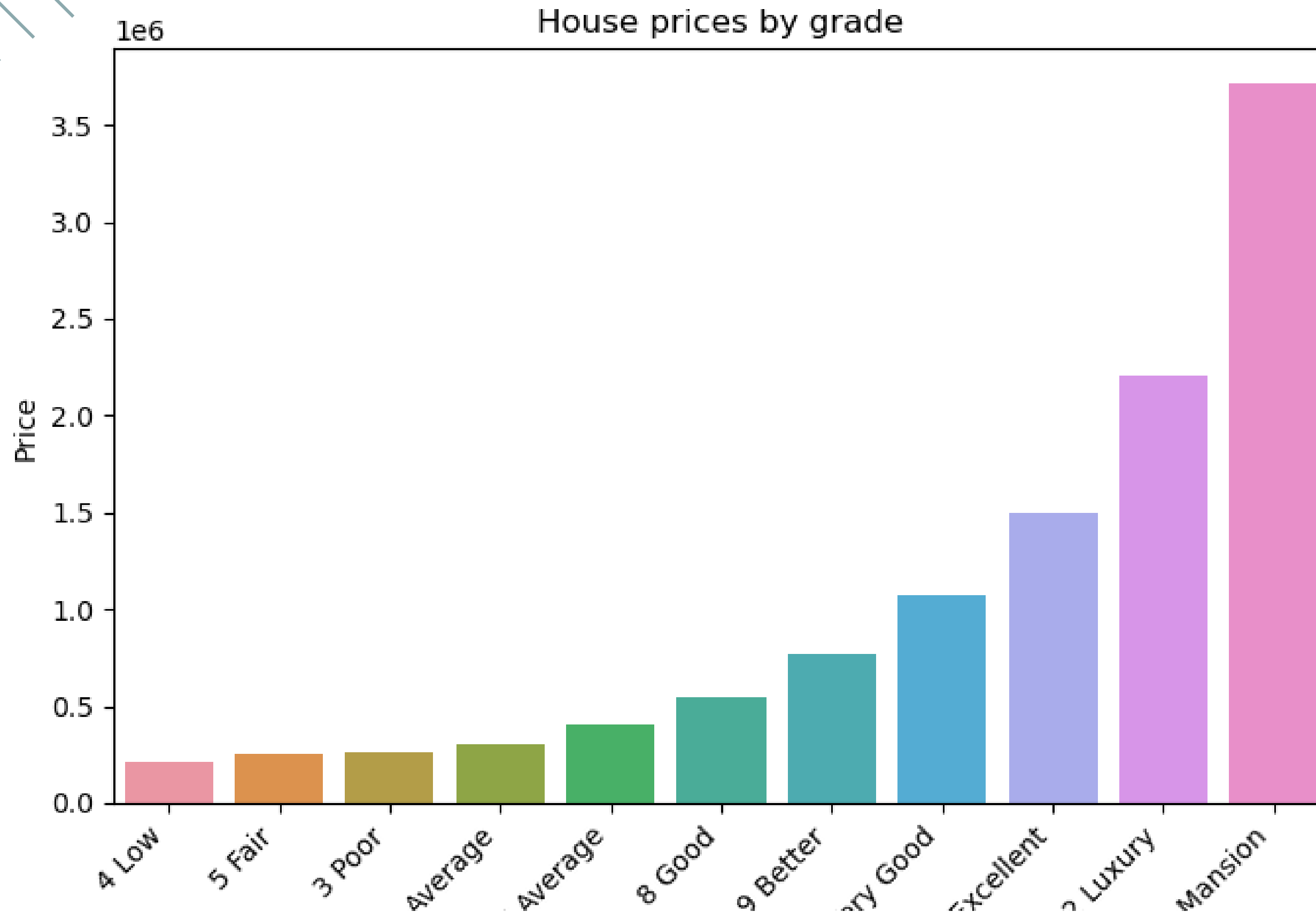
- **The data interprets the average price of houses by grade as:-**
  - **Mansions – $ 3.7m**
  - **Better – $ 700k**
  - **Low – $ 250k**

# Data Visualization cont..

## Number of houses sold per season
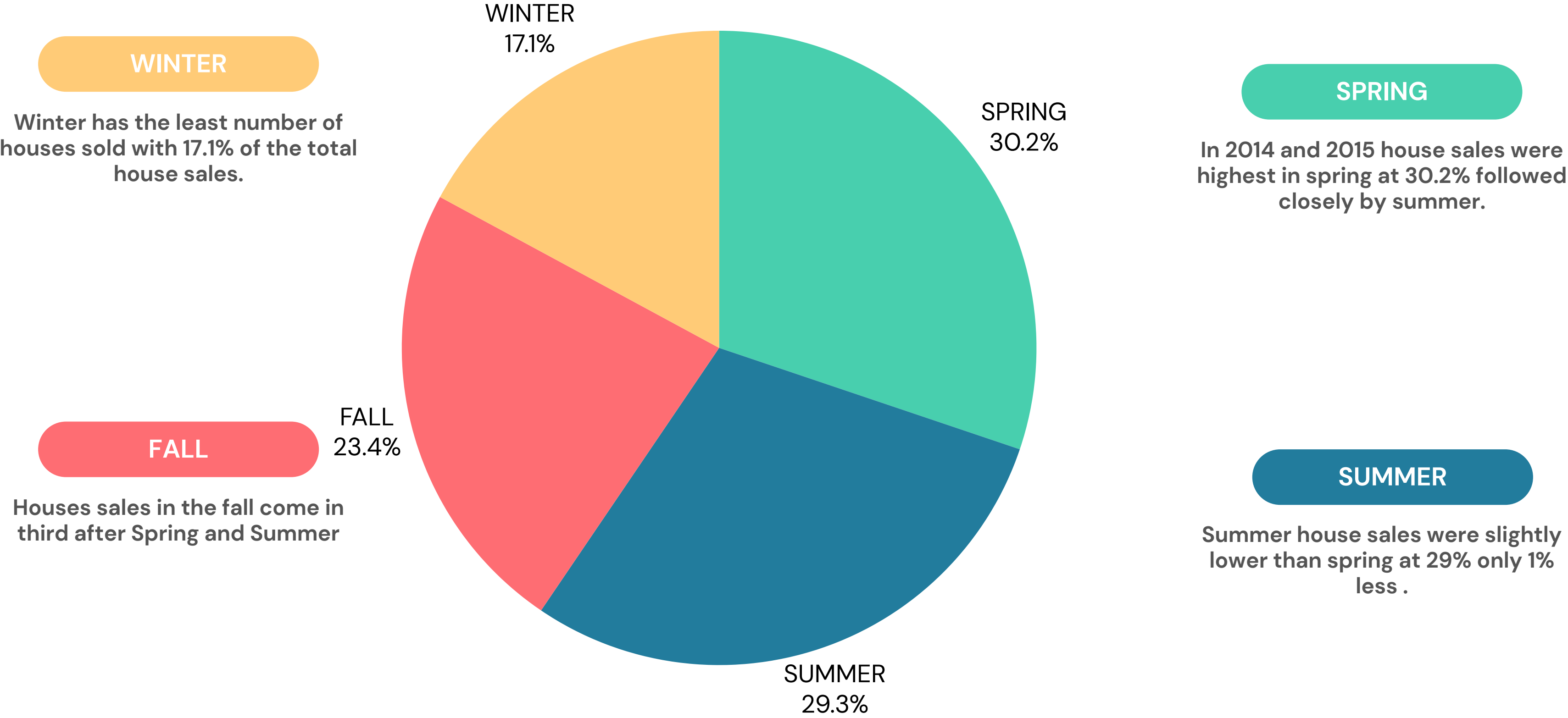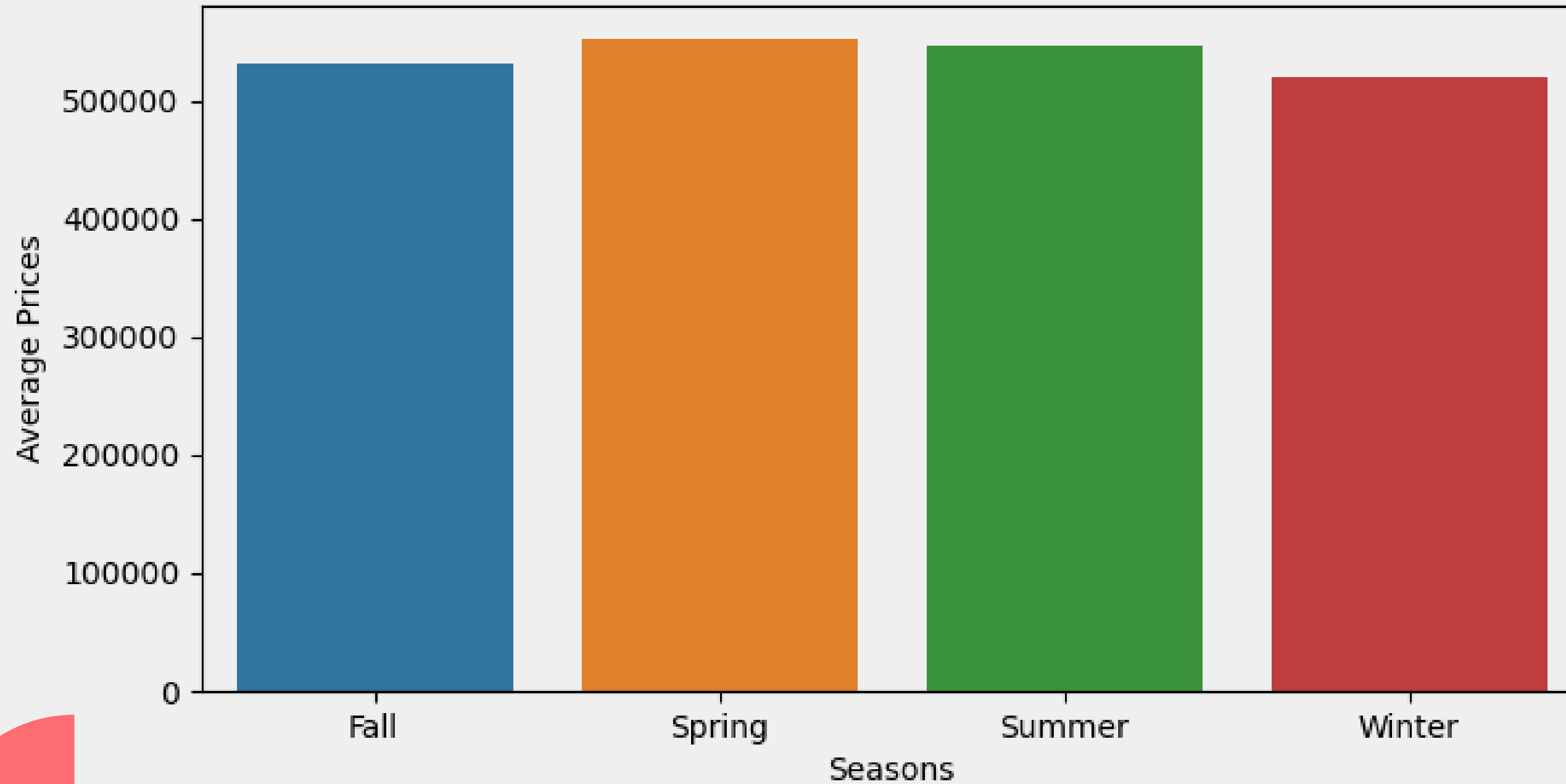


**WINTER**

Winter has the least number of houses sold with 17.1% of the total house sales.

**FALL**

Houses sales in the fall come in third after Spring and Summer

**SPRING**

In 2014 and 2015 house sales were highest in spring at 30.2% followed closely by summer.

**SUMMER**

Summer house sales were slightly lower than spring at 29% only 1% less .

WINTER
17.1%

SPRING
30.2%

FALL
23.4%

SUMMER
29.3%

# Data Visualization contd..

The houses are priced at $2m and above where the age is below 50 years. On the contrary majority of the houses over 50 years are priced below $ 2m.



House prices by house age
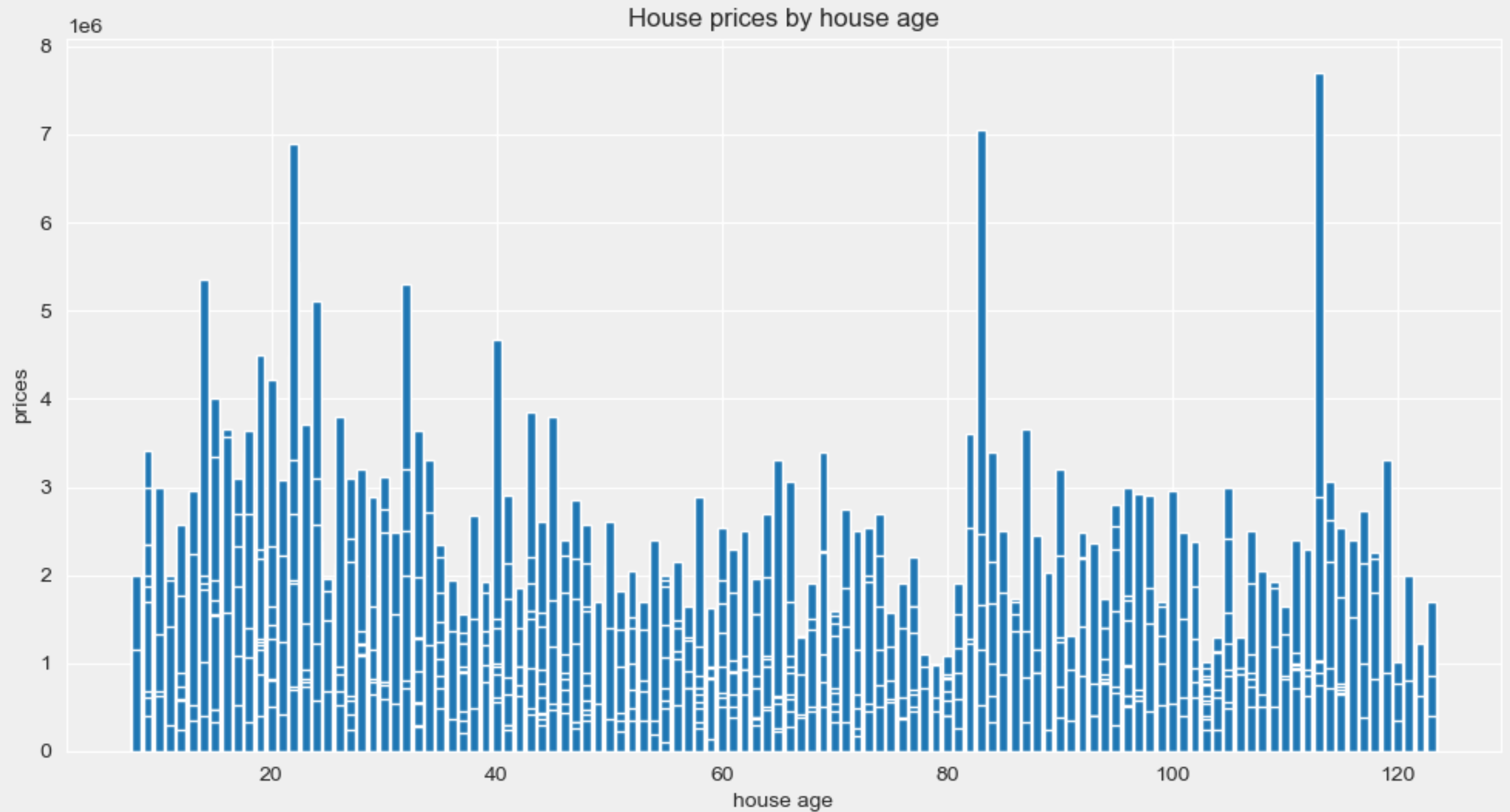
# *Modelling*

```
                          OLS Regression Results
========================================================================
Dep. Variable:              price      R-squared:                 0.493
Model:                        OLS      Adj. R-squared:            0.493
Method:             Least Squares      F-statistic:            2.097e+04
Date:            Mon, 01 Jan 2024      Prob (F-statistic):         0.00
Time:                    05:46:06      Log-Likelihood:        -3.0005e+05
No. Observations:           21596      AIC:                    6.001e+05
Df Residuals:               21594      BIC:                    6.001e+05
Df Model:                       1
Covariance Type:         nonrobust
========================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------
const        -4.401e+04   4410.123     -9.980      0.000   -5.27e+04   -3.54e+04
sqft_living    280.8688      1.939    144.820      0.000     277.067     284.670
========================================================================
Omnibus:                14801.492      Durbin-Watson:             1.982
Prob(Omnibus):              0.000      Jarque-Bera (JB):      542642.481
Skew:                       2.820      Prob(JB):                   0.00
Kurtosis:                  26.901      Cond. No.               5.63e+03
========================================================================
```

## Simple linear regression model

The model shows the following

- R Squared 0.49 – Square foot living explains upto 49% change in the price of houses
- P Value 0.00 – This is well below 0.05%, representing a significant relationship between the square feet living and the price.
- Coefficient 280 – A change of 1 square foot in the living space of a house causes a change in price of $280.

# *Modelling contd..*

```
                         OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.607
Model:                            OLS   Adj. R-squared:                  0.607
Method:                 Least Squares   F-statistic:                     6672.
Date:                Sat, 30 Dec 2023   Prob (F-statistic):               0.00
Time:                        22:56:41   Log-Likelihood:             -2.9729e+05
No. Observations:               21596   AIC:                         5.946e+05
Df Residuals:                   21590   BIC:                         5.946e+05
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -1.162e+06   1.55e+04    -75.098      0.000   -1.19e+06   -1.13e+06
grade_no      1.417e+05   2200.325     64.409      0.000    1.37e+05    1.46e+05
sqft_living    155.1451      3.176     48.843      0.000     148.919     161.371
bathrooms     4.247e+04   3484.316     12.188      0.000    3.56e+04    4.93e+04
renovated     4.715e+04   8921.922      5.285      0.000    2.97e+04    6.46e+04
house_age     3907.1281     67.185     58.155      0.000    3775.440    4038.816
==============================================================================
Omnibus:                    17739.221   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1339226.121
Skew:                           3.466   Prob(JB):                         0.00
Kurtosis:                      40.951   Cond. No.                     2.28e+04
==============================================================================
```

## Multiple linear regression – 2nd model

The model shows the following
- R Squared 0.60 – The predictors; grade no, Square foot living, bathrooms, if renovated and house age explains upto 61% change in the price of houses
- P Value 0.00 – This are well below 0.05%, this representing a significant relationship between the variables.
- Coefficients– The various coefficients in the model represent the respective change in price.

# *Modelling contd..*

```
                          OLS Regression Results
====================================================================
Dep. Variable:               price   R-squared:               0.645
Model:                         OLS   Adj. R-squared:          0.645
Method:              Least Squares   F-statistic:             4906.
Date:             Mon, 01 Jan 2024   Prob (F-statistic):       0.00
Time:                     06:08:03   Log-Likelihood:      -2.9619e+05
No. Observations:            21596   AIC:                  5.924e+05
Df Residuals:                21587   BIC:                  5.925e+05
Df Model:                        8
Covariance Type:         nonrobust
====================================================================
                 coef     std err        t     P>|t|    [0.025     0.975]
--------------------------------------------------------------------
const        -1.001e+06  1.59e+04   -63.100   0.000  -1.03e+06   -9.7e+05
bedrooms     -4.615e+04  2134.313   -21.622   0.000  -5.03e+04   -4.2e+04
bathrooms     5.671e+04  3388.073    16.739   0.000   5.01e+04   6.34e+04
sqft_living    178.5247     3.315    53.856   0.000    172.027    185.022
sqft_lot        -0.2609     0.037    -7.101   0.000     -0.333     -0.189
waterfront    7.51e+05   1.84e+04    40.806   0.000   7.15e+05   7.87e+05
renovated     2.004e+04  8499.545     2.358   0.018   3381.111   3.67e+04
grade_no      1.309e+05  2127.552    61.546   0.000   1.27e+05   1.35e+05
house_age     3870.9806    64.121    60.370   0.000   3745.299   3996.662
====================================================================
Omnibus:                 15563.658   Durbin-Watson:            1.974
Prob(Omnibus):               0.000   Jarque-Bera (JB):    949113.383
Skew:                        2.866   Prob(JB):                  0.00
Kurtosis:                   34.967   Cond. No.             5.45e+05
====================================================================
```

## Multiple linear regression – Final model

This is the best–fit model showing the following
- R Squared 0.65 – The predictors; grade no, Square foot living & lot, bathrooms, bedrooms, if renovated, house age, and waterfront explain up to 65% change in the price of houses.
- P Value 0.00 – These are well below 0.05%, representing a significant relationship between the variables except for if the house was renovated.
- Coefficients– The various coefficients in the model represent the respective change in price.

# RECOMMENDATION

- Our potential homeowners are advised to focus on the variables in the final model when looking for competitively priced homes.

- Where one is looking to buy a home at favorable prices a potential homeowner will need to compromise on one or two items. e.g. waterfront homes or living spaces.

- Our stakeholders are advised to purchase homes in the spring or summer in order to get a good variety of homes to pick from.

- House prices do not vary much from season to season. The average price per house stands at $ 500k.

Thank you