# KING COUNTY HOUSING

## Predictive Analysis

# OUR MISSION

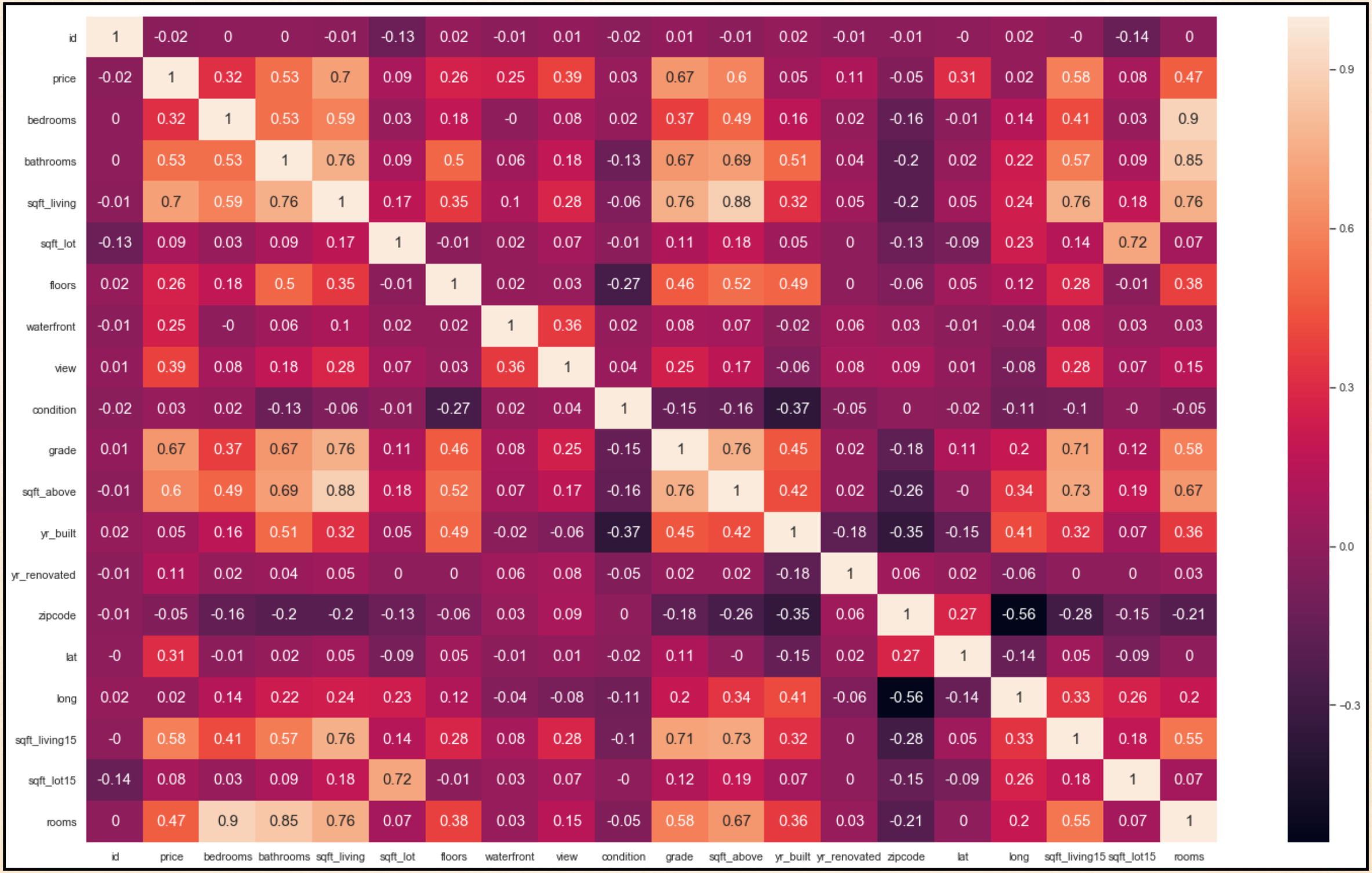
- Construct a regression model with particular variables that accurately predict house prices in King County

- With our predicted prices we can advise our clients on what features they can add or sacrifice to maximise the possible value of their budget

# OUR APPROACH

- Initial data exploration with some cleaning
- Formulate economic theories relating to the areas of analysis
- Regression analysis
- Conclusions drawn from regression output

# CORRELATION MAP



# MAIN REGRESSION STATUS

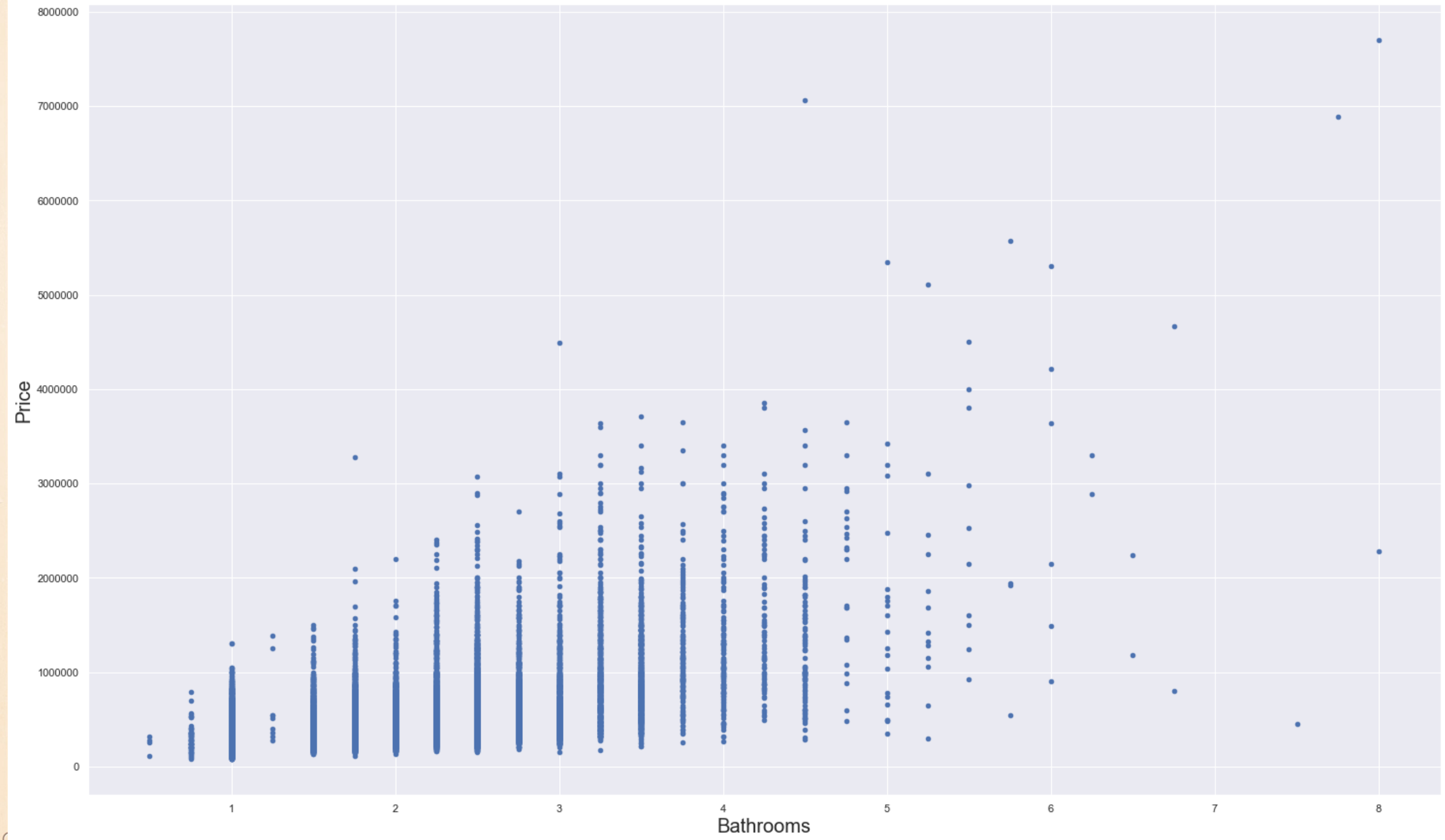
**R-squared:** 0.637

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 6.072e+06 | 1.31e+05 | 46.388 | 0.000 | 5.82e+06 | 6.33e+06 |
| **yr_built** | -3535.7599 | 69.160 | -51.125 | 0.000 | -3671.321 | -3400.199 |
| **sqft_living** | 178.7667 | 3.066 | 58.310 | 0.000 | 172.757 | 184.776 |
| **waterfront** | 7.906e+05 | 2.11e+04 | 37.416 | 0.000 | 7.49e+05 | 8.32e+05 |
| **grade** | 1.383e+05 | 2536.383 | 54.531 | 0.000 | 1.33e+05 | 1.43e+05 |

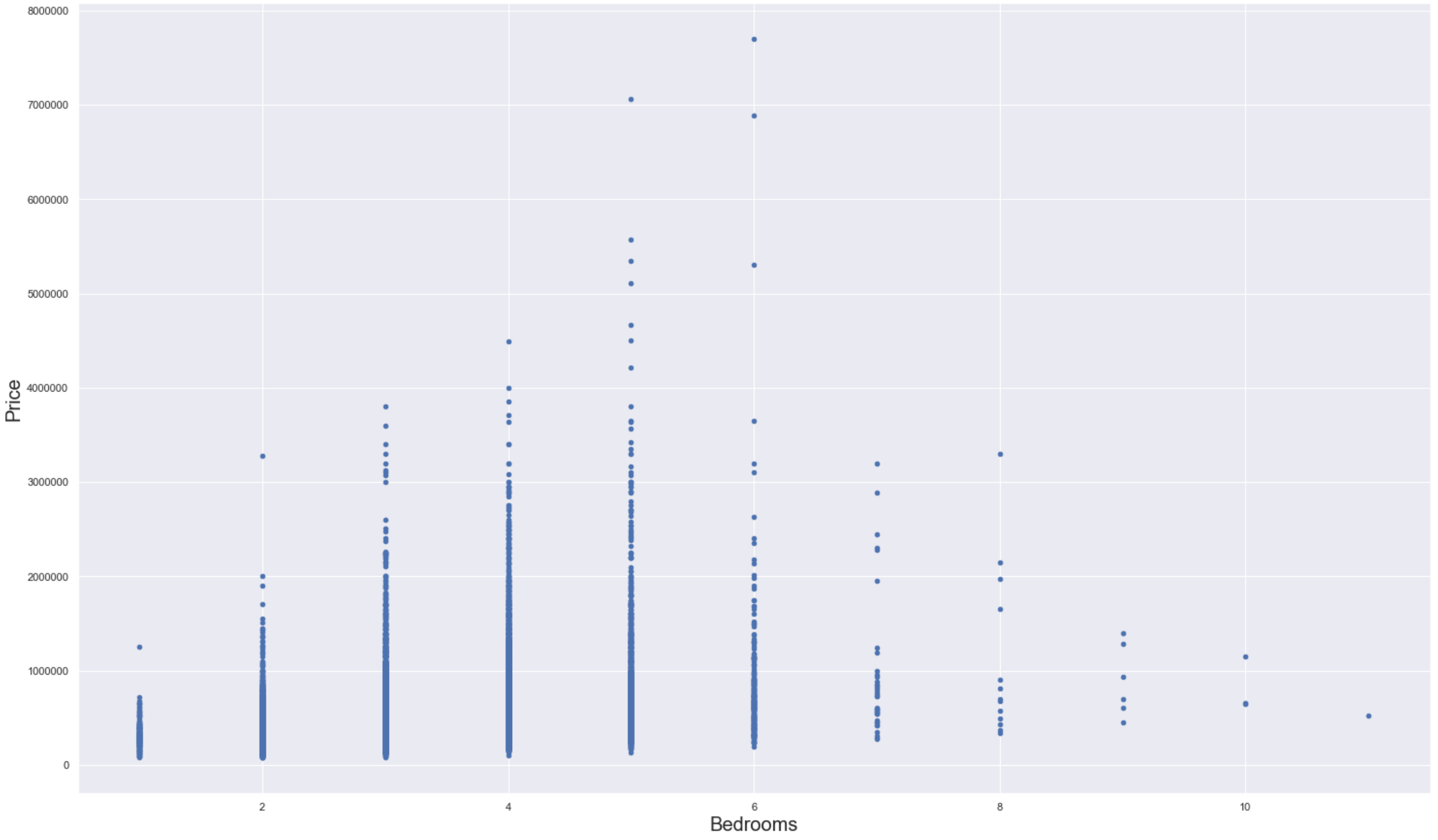
# QUESTION 1

Analysing  bed and bath rooms effect on price

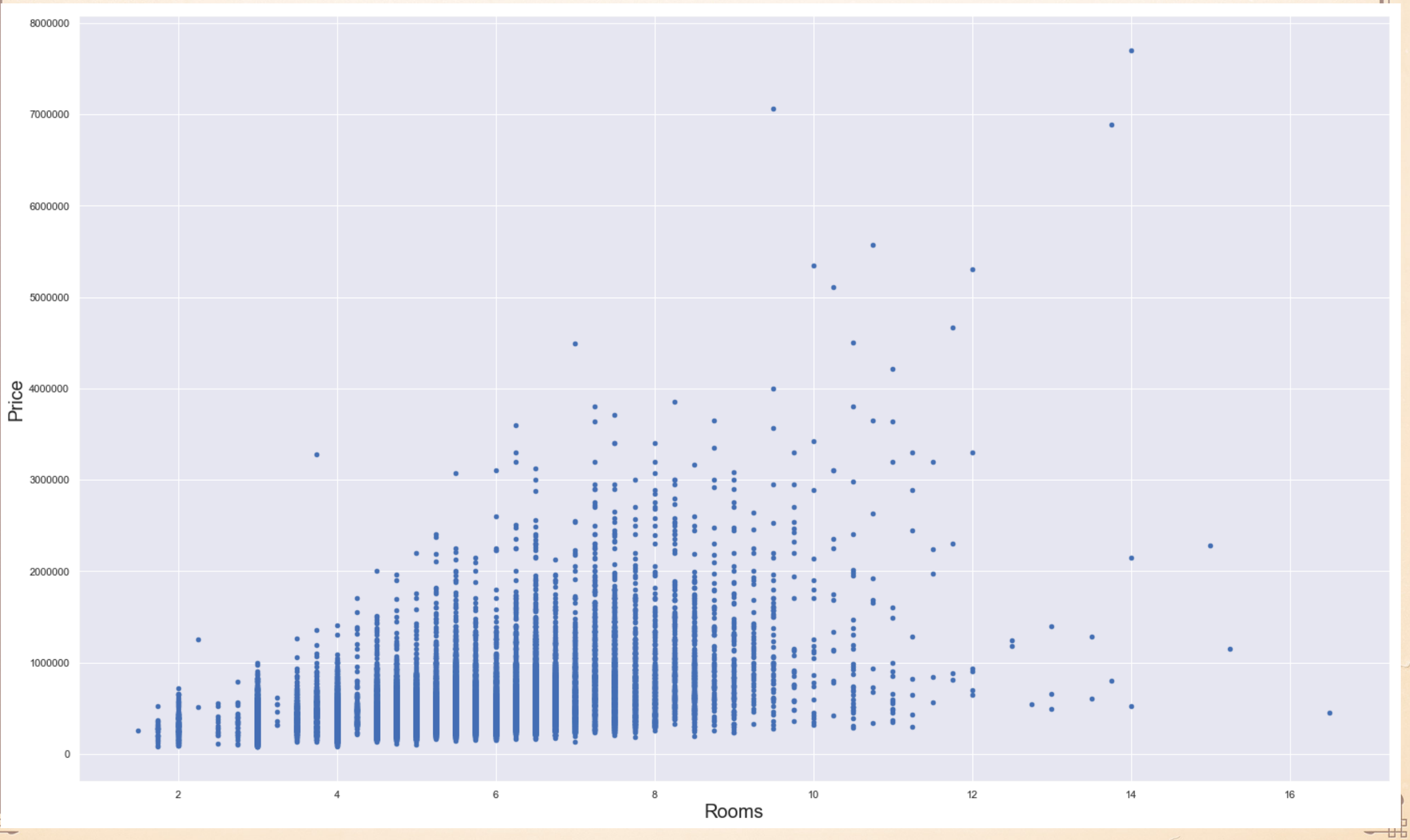
# BATH & PRICE DISTRIBUTION



# BED & PRICE DISTRIBUTION



# COMBINED & PRICE DISTRIBUTION



# COMBINED & PRICE DISTRIBUTION

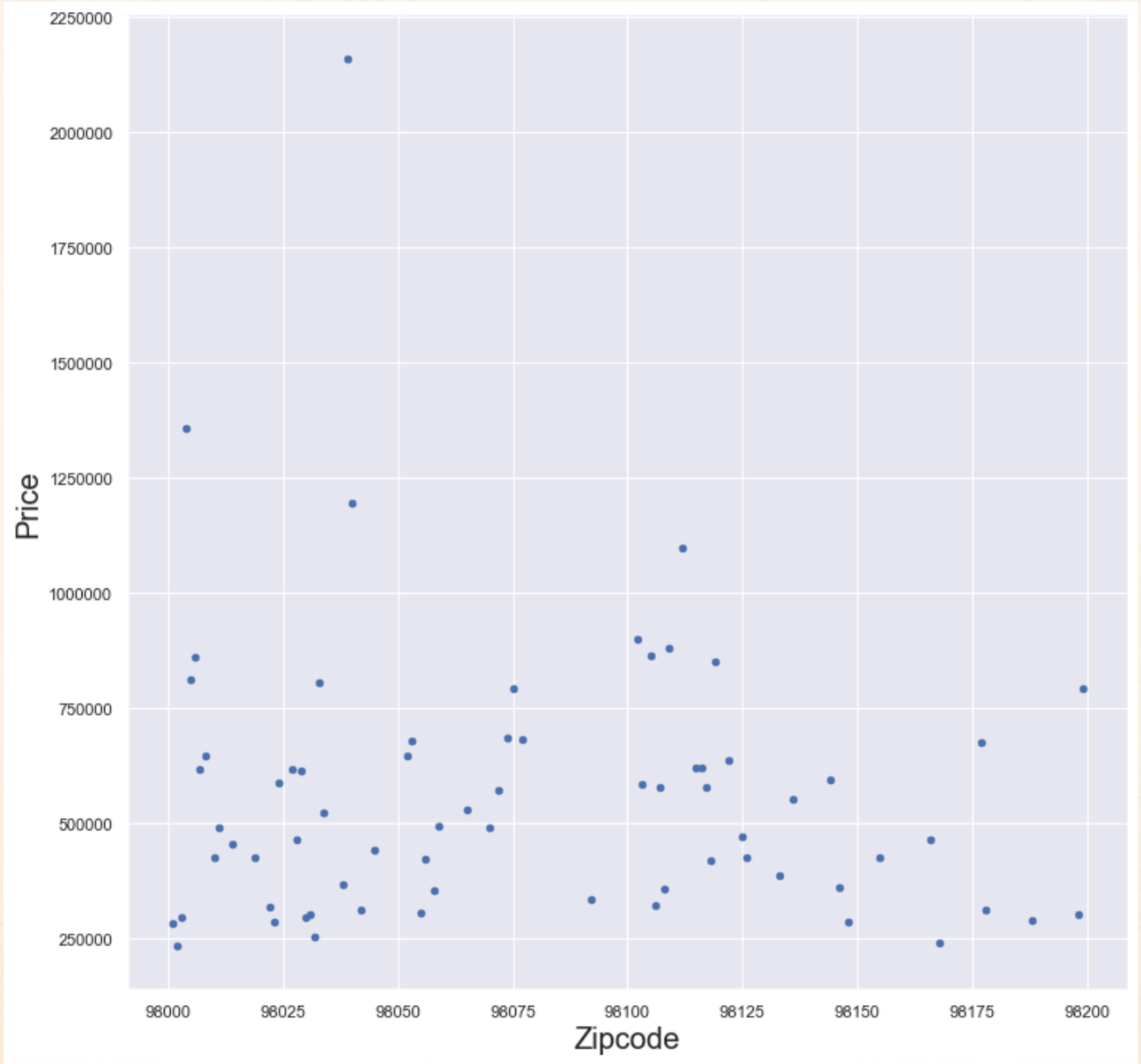
| | |
|---|---|
| **R-squared:** | 0.532 |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | -1.011e+06 | 1.37e+04 | -73.999 | 0.000 | -1.04e+06 | -9.84e+05 |
| **rooms** | 1.795e+04 | 1805.350 | 9.942 | 0.000 | 1.44e+04 | 2.15e+04 |
| **waterfront** | 9.016e+05 | 2.4e+04 | 37.602 | 0.000 | 8.55e+05 | 9.49e+05 |
| **grade** | 1.828e+05 | 2160.426 | 84.596 | 0.000 | 1.79e+05 | 1.87e+05 |
| **sqft_basement** | 152.8349 | 4.950 | 30.878 | 0.000 | 143.133 | 162.537 |
| **sqft_lot** | 0.0838 | 0.049 | 1.702 | 0.089 | -0.013 | 0.180 |

# QUESTION 2

Analysing the differences between Top and Bottom quartiles of average price per zipcode

# ZIP VS MEAN PRICE



# BOTTOM END ZIPS

Theory

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 4.109e+05 | 1.12e+05 | 3.685 | 0.000 | 1.92e+05 | 6.3e+05 |
| **rooms** | -3157.5844 | 1261.068 | -2.504 | 0.012 | -5630.038 | -685.131 |
| **sqft_basement** | -13.7096 | 3.208 | -4.273 | 0.000 | -20.000 | -7.419 |
| **sqft_lot** | 0.4833 | 0.022 | 22.353 | 0.000 | 0.441 | 0.526 |
| **waterfront** | 4.589e+05 | 1.61e+04 | 28.437 | 0.000 | 4.27e+05 | 4.91e+05 |
| **grade** | 2.837e+04 | 1791.567 | 15.835 | 0.000 | 2.49e+04 | 3.19e+04 |
| **sqft_living** | 93.7506 | 2.879 | 32.568 | 0.000 | 88.107 | 99.394 |
| **yr_built** | -243.7253 | 59.395 | -4.103 | 0.000 | -360.175 | -127.275 |

R squared
0.677

Stepwise

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | -6.563e+04 | 9160.977 | -7.164 | 0.000 | -8.36e+04 | -4.77e+04 |
| **sqft_living** | 80.5128 | 2.125 | 37.891 | 0.000 | 76.347 | 84.679 |
| **waterfront** | 3.592e+05 | 1.76e+04 | 20.414 | 0.000 | 3.25e+05 | 3.94e+05 |
| **sqft_lot** | 0.4429 | 0.021 | 20.622 | 0.000 | 0.401 | 0.485 |
| **grade** | 2.791e+04 | 1578.283 | 17.683 | 0.000 | 2.48e+04 | 3.1e+04 |
| **view** | 2.478e+04 | 2002.949 | 12.370 | 0.000 | 2.09e+04 | 2.87e+04 |
| **yr_renovated** | 20.4302 | 3.465 | 5.895 | 0.000 | 13.636 | 27.225 |

R squared
0.689

# TOP END ZIPS

Theory

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 7.629e+06 | 3.74e+05 | 20.378 | 0.000 | 6.89e+06 | 8.36e+06 |
| **rooms** | -1.126e+04 | 5290.076 | -2.128 | 0.033 | -2.16e+04 | -883.399 |
| **sqft_lot** | -1.6232 | 0.200 | -8.131 | 0.000 | -2.015 | -1.232 |
| **waterfront** | 1.153e+06 | 5.13e+04 | 22.481 | 0.000 | 1.05e+06 | 1.25e+06 |
| **grade** | 1.022e+05 | 6579.933 | 15.525 | 0.000 | 8.93e+04 | 1.15e+05 |
| **sqft_living** | 298.8767 | 9.636 | 31.016 | 0.000 | 279.983 | 317.770 |
| **yr_built** | -4217.4814 | 197.277 | -21.378 | 0.000 | -4604.270 | -3830.692 |

R squared
0.675

Stepwise

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 7.113e+06 | 3.75e+05 | 18.952 | 0.000 | 6.38e+06 | 7.85e+06 |
| **grade** | 9.943e+04 | 6518.502 | 15.254 | 0.000 | 8.67e+04 | 1.12e+05 |
| **sqft_living** | 275.1018 | 7.518 | 36.592 | 0.000 | 260.361 | 289.842 |
| **waterfront** | 1.019e+06 | 5.39e+04 | 18.905 | 0.000 | 9.13e+05 | 1.12e+06 |
| **yr_built** | -3954.1561 | 197.537 | -20.017 | 0.000 | -4341.455 | -3566.857 |
| **view** | 5.003e+04 | 6134.179 | 8.155 | 0.000 | 3.8e+04 | 6.21e+04 |
| **sqft_lot15** | -2.1755 | 0.268 | -8.123 | 0.000 | -2.701 | -1.650 |

R squared
0.682