DATA SCIENCE CAPSTONE

Linear Regression Model for

Housing Price Prediction

By:

Rajan Gyawali

PRESENTATION OUTLINE

- Introduction/Business Problem
- Data Collection/Preprocessing
- Methodology
- Results
- Discussion
- Conclusion

INTRODUCTION/BUSINESS PROBLEM

- Linear Regression Model for prediction of Housing Price
- Linear Regression Model is used for predicting the continuous target value that is dependent upon number of independent variables.

Business Problem:

1. How much more price one can sell his/her house with additional bedroom/bathroom?
2. How does the price of house differ with increase in square feet?
3. What is the impact of number of years built for the pricing of houses?

DATA COLLECTION/ PREPROCESSING

The dataset consists of attributes like:

1. Price
2. Number of Bedrooms
3. Number of Bathrooms
4. Living Room Area
5. Number of Floors
6. Waterfront
7. View
8. Condition of the House
9. Year built
10. Year renovated, etc

METHODOLOGY

A basic linear regression model can be formulated as:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \ldots$$

The attributes used for linear regression model are:

- price
- Bedrooms
- Bathrooms
- sqft_living
- sqft_lot
- floors
- waterfront
- view
- condition
- grade
- sqft_above
- sqft_basement
- number_of_days_built

METHODOLOGY

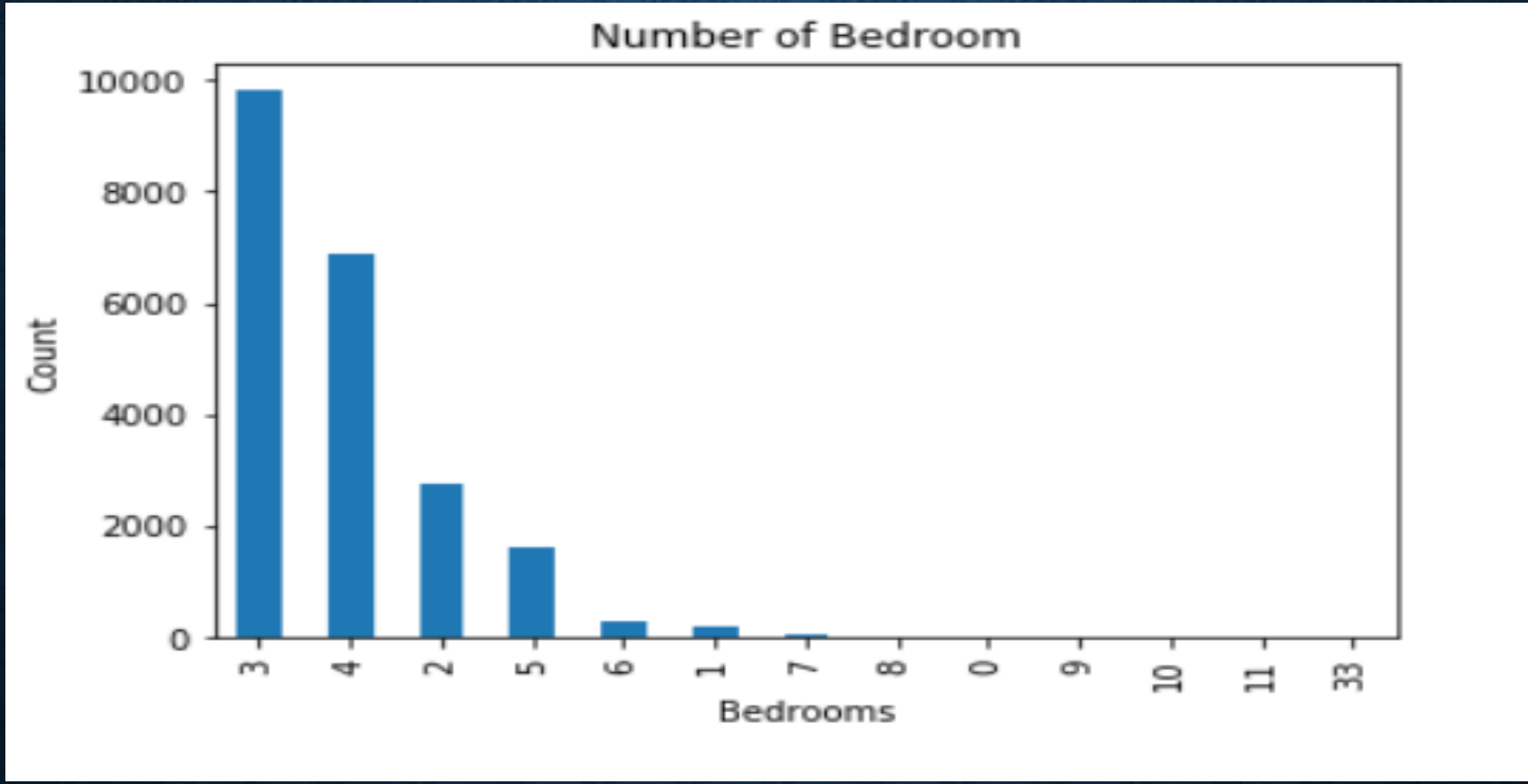

Figure: Plot of Number of Bedrooms Vs Count of Houses

METHODOLOGY

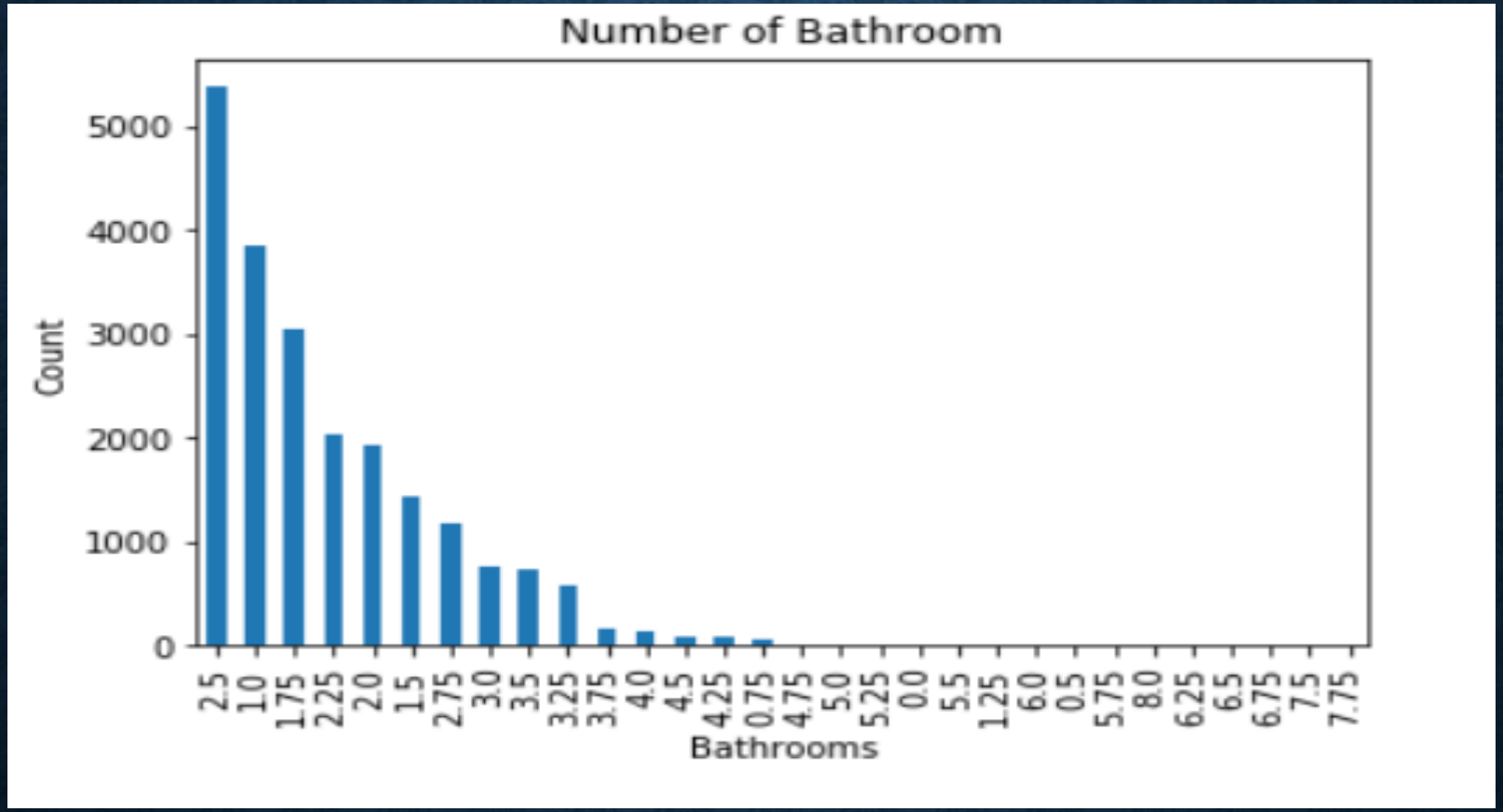

Figure: Plot of Number of Bathrooms Vs Count of Houses

METHODOLOGY

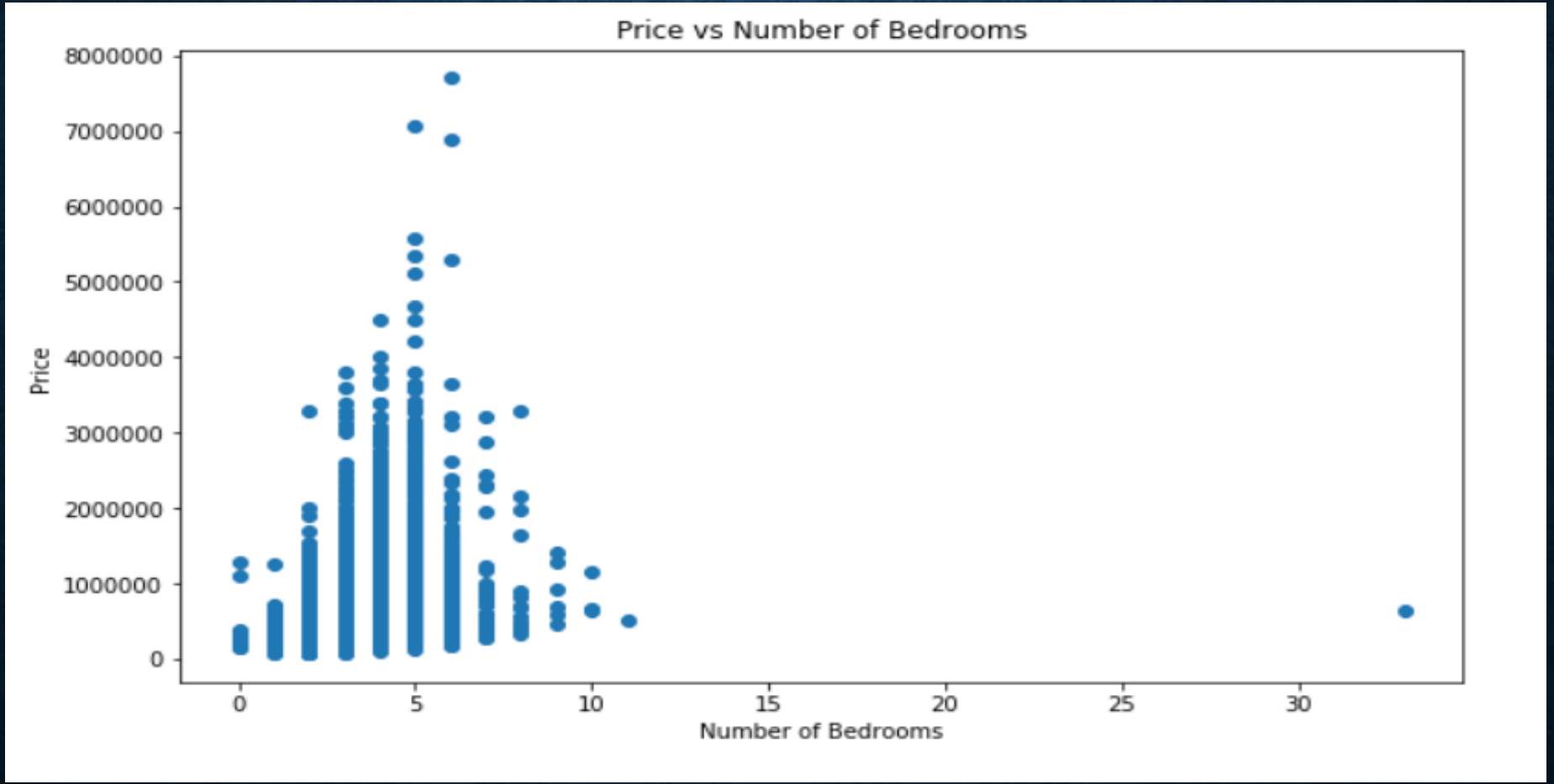

Figure: Scatter Plot to show relationship between price and number of bedrooms

METHODOLOGY

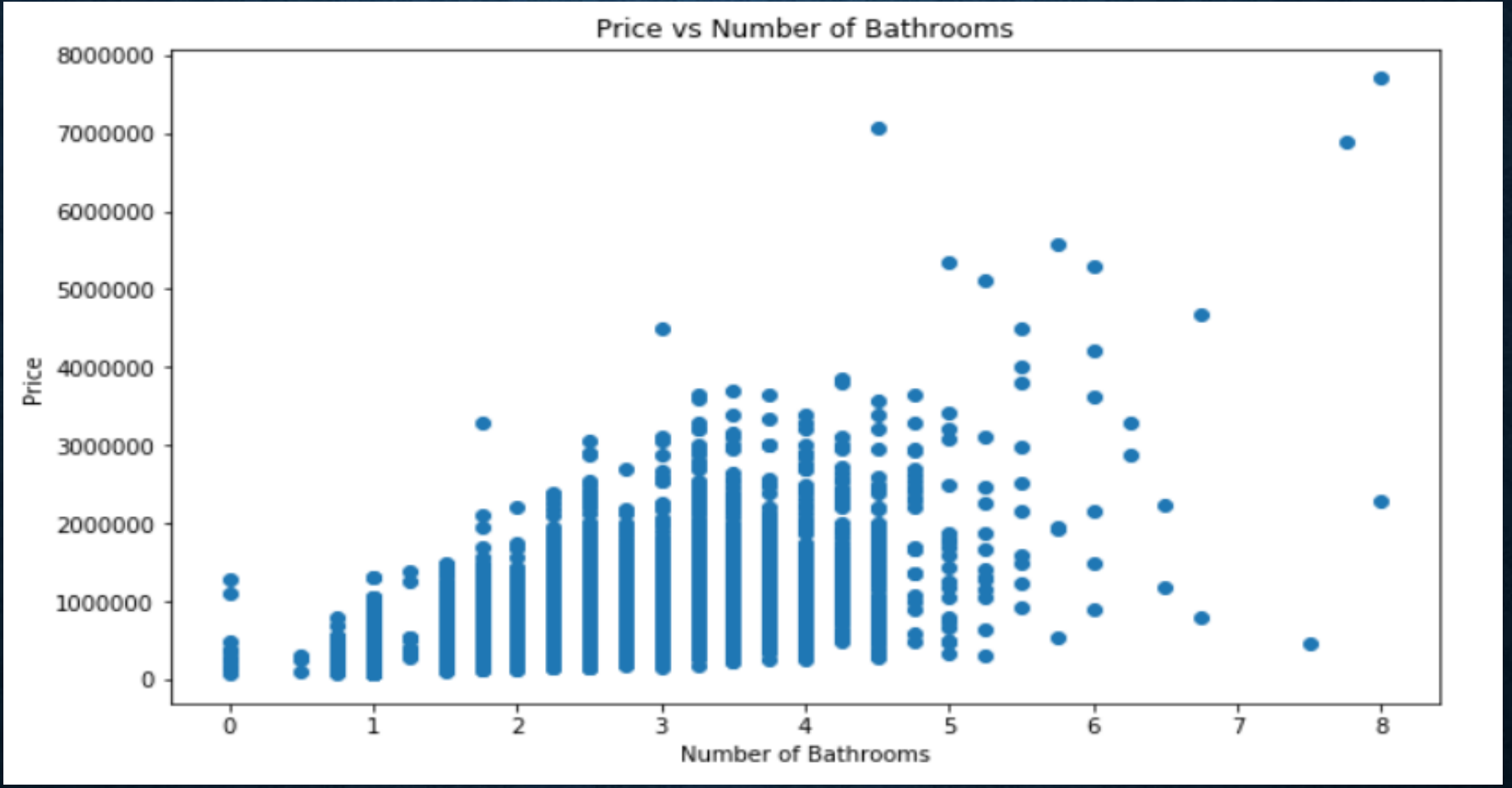

Figure: Scatter Plot to show relationship between price and number of bathrooms

METHODOLOGY

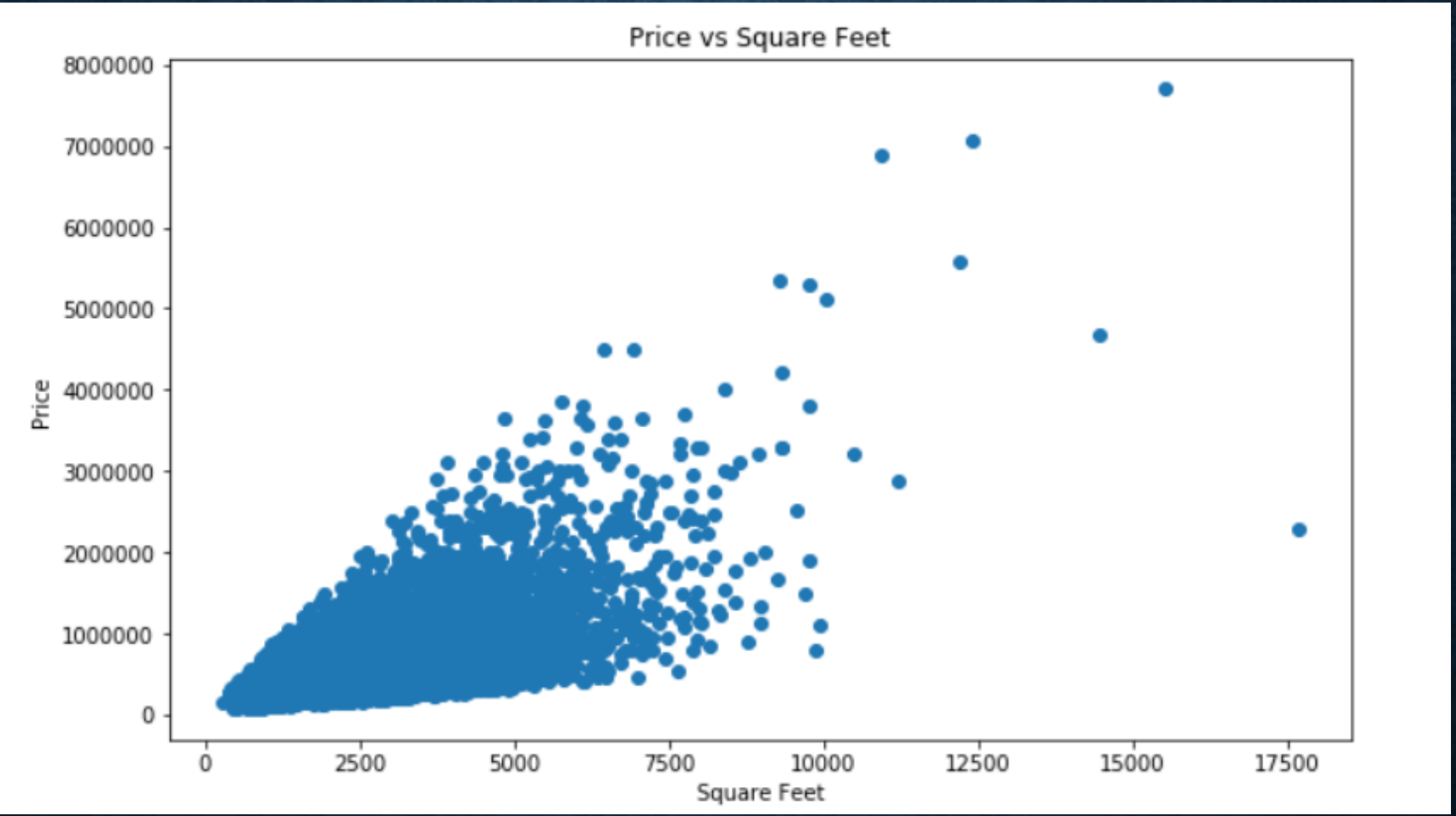

Figure: Scatter Plot to show relationship between price and squarefeet

METHODOLOGY

- From the overall dataset, 80% is converted to training set and rest 20% is converted to test set.
- With these training samples, the machine learning model for linear regression is modeled and fitted with the scikit learn library.
- The fitted model is tested with the test dataset and a regression score is calculated.

RESULTS

Predicted Price	Actual Price
666735	735000
1455081	1150000
337335	350500
1183792	860000
191910	122000
750749	725000
644535	417000
551787	594950
511011	471000
820174	634950
518060	500000
770764	768000
486381	323000
418779	430000
609564	625000
609511	710000
455584	620000
1015401	665000
1609385	1600000
1109880	875000

DISCUSSION & CONCLUSION

- Housing price prediction is done with Linear Regression Model
- Linear Regression Model Can be used for predicting continuous target value.
- With increase in number of training samples, accuracy of the model can be increased.

DATA SCIENCE CAPSTONE

Thank You !!!