

# Assignment

## 1. Introduction

The problem involves analyzing housing prices that were sold between May 2014 and May 2015.

Following are the two tasks:

- 1) Exploratory Analysis: Let us know something that you find interesting
- 2) Machine Learning: Predict the sales price of houses

The dataset contains 21,614 samples and 21 features

## 2. Data Features

- |                            |  |
|----------------------------|--|
| 1) ID                      | — Notation for a house   |
| 2) Date                    | — Date house was sold  |
| 3) Price                   | — House price (target variable)  |
| 4) Bedrooms                | — Number of bedrooms   |
| 5) Bathrooms               | — Number of bathrooms  |
| 6) Sqft_Living             | — Square footage of the home   |
| 7) Sqft_Lot                | — Square footage of the lot  |
| 8) Floors                  | — Total floors (levels) in house   |
| 9) Waterfront              | — House which has a view to a waterfront   |
| 10) View                   | — Has been viewed  |
| 11) Condition              | — How good the condition is overall  |
| 12) Grade                  | — Overall grade given to the housing unit  |
| 13) Sqft_Above<br>basement | — Square footage of the house apart from<br>basement   |
| 14) Sqft_Basement          | — Square footage of the basement   |
| 15) Yr_Built               | — Built Year   |
| 16) Yr_Renovated           | — Year when house was renovated  |
| 17) Zipcode                | — zip  |
| 18) Lat                    | — Latitude coordinate  |
| 19) Long                   | — Longitude coordinate   |
| 20) Sqft_Living15          | — Living room area in 2015 (implies some<br>renovations). This might or might have affected the lotsize area |
| 21) Sqft_lot15             | — Lot size area in 2015 (implies some  |

renovations)

#### **4. Evaluation Criteria**

- We will assess how you approach in this open ended problem, your initial data analysis, why you think a certain approach is suited to solve this problem. We would want to know why you chose a particular evaluation metric, loss function etc, what is your inference from the results.

#### **3. Submission**

- You may use a programming language/framework of your preference
- Please email your results and code