# Can we predict house prices using known features of each house and a supervised learning approach?

Florence Galliers

18/10/2020

## Background:

House prices are an important part of the economy...

### Objectives:

- Understand which independent variables in the data set can be used to predict house price (the dependent variable)
- Minimize the differences between predicted and actual house prices by using model selection to choose the most accurate model.

## Data:

The data used for this analysis contains house prices and information regarding the house features. It has 4600 entries. The original data set downloaded from Kaggle had 17 independent variables, however I felt that 7 of these were not relevant to this analysis and so they have been removed. The dependent variable in this analysis will be the house price (in US dollars).

### Data Preparation:

- changed 2 variables to binary, talk about this here
- dummy variables for condition variable??
- prices are large numbers, change to show in thousands rather than pounds?

```
##     price bed bath sqft_living sqft_total floors condition sqft_basement
## 1  599999   9 4.50        3830       6988    2.5         3             1
## 2  340000   8 2.75        2790       6695    1.0         3             1
## 3 1970000   8 3.50        4440       6480    2.0         5             1
## 4 2280000   7 8.00       13540     307752    3.0         3             1
## 5  840000   7 4.50        4290      37607    1.5         5             0
## 6  999000   7 4.00        3150      34830    1.0         3             0
##   yr_built renovated      city
## 1     1938         1   Seattle
## 2     1977         1 Shoreline
## 3     1959         0   Seattle
## 4     1999         0   Redmond
## 5     1982         0  Issaquah
## 6     1957         1  Bellevue
```

## Methods:

As house price is a continuous variable I have taken a supervised learning approach and will be using regression to look at the relationship between house price and features of each house.

## Results:

## Conclusions:

**Literature Cited:**