To Buy or to Rent a House?
Predicting future housing price in MA

Project Background:

When a new real estate property comes out, often people wonder if they should buy it or rent it. Having the ability to predict future housing price is the key to answer this question. In this project, we want to build two regression models to predict houses' sale prices and rent prices. Based on our prediction, we can give advice to people on whether they should buy or rent the property.

Project Milestones:

1. Project Selection:

Form a team of 2-3 people and select a project from the project list.

2. Data Collection/Cleaning:

There are a few ways you can collect data. We recommend using housing websites' API (NOTE that these APIs often have limits of calls per day, so it's recommended to start getting the data early). If you are interested in data other than MA, it's ok to switch to another state. If the dataset is too large, it's also ok to only look at a specific city's data.

- Zillow API: http://www.zillow.com/howto/api/APIOverview.htm
- Trulia API: http://developer.trulia.com/

Besides the housing data, there are other factors that can influence housing price, such as interest rate, unemployment rate, population etc. You are welcome to use data sets that contain this information.

3. Method Proposal:

In this proposal, you will write a detailed plan of solving this problem. Your proposal will have to address the following aspects, including (but not limited to):

- ***Exploratory Statistics***: how do you plan to explore the data? What plots/values are you interested in looking at?
- ***Performance Metrics***: what metrics are you going to measure your models' performance, and why choosing those metrics?
- ***Feature Selection***: You can get many features associated with a house, but how are you going to choose the most significant ones to train your model?
- ***Baseline Model***: What model do you plan to compare your model against? A couple idea could be the housing price from last year/month, an average of prices from past few years, or a simple linear regression model.
- ***Regression Model***: What regression models do you plan to use?