Proposal for Midterm Project

Xiaozhou Lu

2020.10.5

**1 Personal Statement**

My career goal is to be a statistical consultant in a consulting firm. Working for a consulting firm means there will be many cases to deal with. And cases could vary in all fields of expertise, such as product strategies, profits, geographic marketing and facility locating. The question I am going to explore is the factors that could have influence on the price of real estate in Boston. Problems like this may always exists if I work as a consultant. So I believe the problem I have chosen could serve as a practice for my future career.

**2 Question**

Questions that I have come up with is to make a prediction for the price of real estate with the variables contained. Besides try to see which factor might influence the price more significantly and if there is a interactive relationship between each variables. More questions might be proposed when the project has begun. It is always interesting and challenging to face the unknown problems that may hidden behind the numbers.

**3 Data Source**

The dataset is called Dragon Real Estate Dataset - Price Predictor downloaded from Kaggle:

<https://www.kaggle.com/arslanali4343/real-estate-dataset>

The data is a matrix with 14 variables and more than 500 observations. The variables contains information about crime rate, location, air condition, age of houses, educational resource and the price of each observations in 1000 USD etc.

**4 Time Line**

Tentative plans to finish the project:

1. EDA: by November 15;
2. Data processing: by November 20;
3. Modeling and Validation: by November 27;
4. Write up: by November 30.