

## CS 506 Final Project Proposal

CS 506  
Spring 2018  
Computational Tools for Data Science  
Professor Adam Smith  
TF: Sofia Nikolakaki  
2/20/2018

### Group:

- 1) Jason Lu ([jasonlu6@bu.edu](mailto:jasonlu6@bu.edu))
- 2) Zixin "Cindy" Ding ([cindydzx@bu.edu](mailto:cindydzx@bu.edu))
- 3) Fang Qu ([hughqu@bu.edu](mailto:hughqu@bu.edu))

Collaboration with BU Spark!

### Coordinators:

Ziba Cranmer ([zcranmer@bu.edu](mailto:zcranmer@bu.edu))

Professor Dora Erdos ([edori@bu.edu](mailto:edori@bu.edu))

Professor Adam Smith ([ad22smith@bu.edu](mailto:ad22smith@bu.edu))

Contact: Michael Jay Walsh, Ph. D.

Topic: Using Permit, Parcel, Census, data to populate building energy systems inventory

### Description of General Topic:

Our project is about using the databases provided from Carbon Free Boston in order to make an analysis of energy system changes over time for different types of Boston housing. The project will consist primarily of 3 main databases: Boston Assessor's database of tax parcels (heating and cooling systems), Boston's permit approval database, and a master database. In addition, we will employ the techniques taught from lecture of text mining, making a predictive model for the heating system, and use web scraping for the permit approval database. Finally, our project will answer the additional question, via measures of similarity, how well correlated is our data in comparison of the old energy system as opposed to the new energy system.

### Questions:

### From BU Spark:

- 1) What communities more like to have heat pumps (demographics, race, income level, etc. and geography)?
- 2) What communities are not benefitting from energy efficient systems?
- 3) How has this changed over time?

**From Our Group:**

- 1) Does there exist a correlation between the installation of new energy system as opposed to the old energy system for Boston housing permits?
- 2) Will the number of people (occupants) living in Boston housing affect the heating and cooling system cost per family?

**For Michael:**

- 1) How to access Zillow API?
- 2) How to access/understand the CSV entries?
- 3) What range of fiscal year should we use for property assessments?
- 4) What type of housing is inside the parcel data?

**Ideas for Specific Data Sets:**

**Data sources:**

1. Census American Housing Survey
2. Boston Building permits (text analysis for renovations involving installation of new energy systems) (358,472 entries, free to access)
3. Boston Property Assessments (170,910 entries, free to access)
4. Zillow API (description of energy systems)