

Thuy Pham
Taylor Hazlett
Menting Song
Xiaoyi Zhang

Team Name: TMTX

Motivation/Background:

Last semester, another CS506 student team worked with state legislator Elugardo to determine the value of state-owned surplus lands for her equity bill. The primary motivation being that she felt there is an opportunity to address the issue of lack of affordable housing across the state. With the team's assistance, she was able to determine that there was ~\$1.2 billion of state-owned surplus land that could potentially be utilized for this purpose. The data provided by the team assisted with the passage of legislator Elugardo's bill through its first committee.

Now, legislator Elugardo is once again seeking help from CS506 students to push her equity bill through its second committee. This primarily requires categorizing the available state-owned surplus land into two sections: 1) land that should be prioritized for affordable housing development and 2) land that should be prioritized for selling off.

Goals: Based on what has been done last semester, we are trying to come up with a scheme to distinguish surplus lands that are more suitable for affordable housing development and lands that are profitable for sale and hopefully could generate revenue to fund affordable housing.

Non-Goals: Though we will need to use real estate APIs to estimate the values of homes in areas of interest, we are not interested in coming up with a model to predict future housing market trends. Additionally, we are not going to consider lands that might satisfy the criteria for surplus land but are privately owned.

Hypothesis: Our hypothesis is that by providing recommendations for standardization within the dataset (specifically entries under the owner_name tuple) and utilizing data points from ally districts (i.e: New Bedford), we will be able to conduct data analysis will make the bill more influential in front of the committee, as well as make the data more insightful and useful to legislators moving forward with the bill.

Outcome(s): By the end of the project, we want to offer standardization recommendations for the dataset, as well as an analysis of which areas are best to sell for profit and which are best to use for affordable housing. This analysis will be presented as a comprehensive report that can be presented to the committee. We also want to offer a trained model to the committee that can be used to determine which parcels of land are better suited for selling or building upon moving forward.

End Result/Definition of Done:

We will deliver an analysis report including explanation in the following areas:

- Quantitative list of the most attractive state land for the development of affordable housing and the influential factors such as proximity to public transport, demographics: low income/minorities, rapidly rising housing costs in surrounding areas, and building viability (access to water/electricity based on presence of existing buildings), etc.
- Quantitative analysis of the attractiveness of state land owned by specific agencies for the development of affordable housing, e.g. the MBTA, DOT, and DOE
- A list of state-owned surplus land that should be prioritized for affordable housing development based on data analysis
- A trained model that given data about a piece of state-owned surplus land can produce a (quantitative) evaluation of whether it should be prioritized for affordable housing development or for selling to earn profit
- Answers to strategic questions, including map visualization of the geographic concentrations of the land types, the “opportunity cost” of not selling the most valuable land, etc.
- Recommendations for additional data sets to be tracked which help analyzing state-owned surplus lands in the future

Open Questions:

- What are other datasets/factors (outside of housing prices and proximity to public transportation) that might affect whether or not land should be sold or used?
- Are there other tuples in the dataset lacking standardization, and could this affect the outcome of our project?
- How much revenue does the state need to generate from selling unused land?
- What other districts are potential ally districts?
- Are there specific regions that should be focused on because they are stronger allies?
- What data standardization practices does the state/do the municipalities already adhere to, if any?