Team The Conquistadors

Kaijie Zhou, Janice He, Tommy Lam, Athina Said, Murtaza Moiyadi, Manuja DeSilva

State Surplus Project Deliverable 0

**DESCRIPTION**: The State Surplus Project aims to efficiently manage surplus lands in Massachusetts in conjunction with the land equity bill. We want to maximize the benefit of building both affordable housing and houses for sale. Over the course of this project, we hope to find the regions of surplus land that can make the most profit to sell. We believe that the best locations for maximizing profit are in regions with the highest income, close proximity to wealthy neighborhoods, and those that possess functional home utility services such as water and electricity. Our anticipated results are a list of rankings with scores(0-10) for each available location in order to estimate how well the land can be sold. With the scores and ranking in mind, the state representative can make an informed decision regarding which locations should be used for affordable housing or to sell. Additionally, we will create a heat map to help translate this information into a visualization of the most valued houses for sale across Boston, and this can be shown using different color gradients. Currently we are still not sure of the exact approach we want to use to create the ranking and heat map representation, but we will keep searching for other APIs that provide such features. We have access to Zillow APIs and owned land data to assist us in estimating land pricing. Our group may also take into consideration some factors that we believe to be more important than others, such as household income, in which we hope to conduct a deeper analysis of which of these factors should be prioritized. Thus we will demonstrate how such attributes can affect the surplus land decision making process.

**DATA SOURCE**:
- [https://docs.google.com/document/d/1w80vbJccsA079McT6Yc1zqVPvTUc_AjDi0ptKldgNEc/edit](https://docs.google.com/document/d/1w80vbJccsA079McT6Yc1zqVPvTUc_AjDi0ptKldgNEc/edit)
- [https://datacommon.mapc.org/browser/datasets/360](https://datacommon.mapc.org/browser/datasets/360)
- [https://github.com/mapc/landparcels](https://github.com/mapc/landparcels)
- [https://dlsgateway.dor.state.ma.us/reports/rdPage.aspx?rdReport=BLA.StateOwnedLand](https://dlsgateway.dor.state.ma.us/reports/rdPage.aspx?rdReport=BLA.StateOwnedLand)

**TOOLS**:
- String matching algorithms with FuzzyWuzzy Python library.
- Scitkit Learn and spaCy for basic machine learning and regression tools.
- ArcGIS for mapping the results. (we will provide a license to this software and will arrange a training in early March)

**APPROACH**:
1. Review existing data
2. Clean existing data
3. Overlay census data
4. Look for proximity to public transport (0.5 miles to public transportation) via Google Maps' API
5. Look for value of homes using various real estate APIs for surrounding geographies (search within 0.5 miles of the address of the land). Look at the real estate value trajectory over time - last 5 years.
6. Identify land with existing buildings (assume water/electricity hookups) - using parcel database
7. Analyze cost of maintenance of land
8. Do analysis of subsets of DOT and geographic concentrations, particularly around major towns, cities, and regionally

**QUESTIONS TO BE ANSWERED**:
- What state land is most attractive for sale to generate funds for affordable housing development as defined by: highest home values in surrounding areas, building viability (access to water/electricity based on presence of existing buildings)?
- What are the geographic concentrations of this land? (map visualization)
- What is the "opportunity cost" of not selling the most valuable land, i.e. how much is the land costing to maintain? How much revenue is the state potentially sacrificing based on the market value of nearby land?
- What additional data sets would help make this a robust analysis, if possible?
- What other recommendations would you make to the state to make tracking and analyzing state land?