State Surplus Land Assessment - Affordable Housing - Spring 2021	
Contact	Representative Nika Elugardo Nika.Elugardo@mahouse.gov Frank Mendoza (aide) francis.mendoza-mccarthy@mahouse.gov 203-517-5399 **copy both on emails Office of Representative Nika Elugardo 24 Beacon Street Room B1 Boston, MA, 02133 Phone:617-722-2582
Organization	Office of Representative Nika Elugardo Suffolk 15th District
Organization Description	1. Nika Elugardo is a State Representative who represents the 15th Suffolk District in the Massachusetts House of Representatives. She represents the towns of Boston and Brookline.
Data Sets	**IMP: Processed Data from the Summer and Two Semesters Ago (more on this in the tools and approaches section)** **Municipal Housing Authority Dataset** **State Housing Land** **State Owned Land Data District Summary** **State Dept. of Housing Owned Land District Summary** **Real Estate APIs:** ● Property Web API ● Estated API ● ATTOM API

	• Zillow API request link to determine housing price trends **Available Affordable land Data (and Report) by Boston Federal Reserve** (row R: available affordable units per 100 low income residents) **Affordable housing thresholds in Mass** (town names top left, % of affordable housing bottom right) **Transportation APIs:** • MBTA: https://docs.digital.mass.gov/dataset/massgis-data-mbta-rapid-transit • Bus Routes: https://docs.digital.mass.gov/dataset/massgis-data-mbta-bus-routes-and-stops • Look up access to transportation threshold to evaluate transportation access. Here is an API that might be useful for this exercise: https://www.walkscore.com/professional/walk-score-apis.php
Approach	**Review existing data (THIS IS CRITICAL).** Review work done two semesters ago / summer - Final report and scripts are here and here. Identify universities and hospitals near (define proximity – will vary for urban - adjacent vs. suburban/rural) affordable housing need (based on low income; utilization in community) and near housing authorities Establish the % of affordable housing for each municipality/town by extracting data from this document (Some housing units have expired and we need to get an updated dataset or exclude those units)(pdf scrape needed if original data cannot be obtained) Assign a relative "transit friendly" score for each parcel Look for proximity to public transport (.5 miles to public transportation). There are lots of approaches to this e.g. Transit Score or GoogleMaps API, etc. Some of this analysis has already been completed last semester, you can build on this. - Walk Score API looks promising, but we may have to research other options if the volume of the data is too large Do analysis of subsets of DOT and geographic concentrations, particularly around major towns and cities and regionally? (see map) Generate visualizations: TBD with client using software such as ArcGIS or tableau as a final deliverable along with the list data.

Limitations	● Some of our data is out of date, and might include closed hospitals and universities, or have expired affordable housing units. If we cannot get updated data, this may lead to inaccuracies in our final result ● The land value in our datasets is self-assessed, and may not be very accurate. We have very limited knowledge on computing actual land value based off the information we have, and such calculations may be inaccurate or entirely useless ● We have limited knowledge about creating a transit score and the applicability of public APIs. Arriving at a satisfactory result may be difficult for large datasets
Tools and approaches	Using the processed data: ● Why? - The MAPC Land Parcel Database has three datasets that are virtually impossible to merge by themselves because of inconsistencies in data formatting. There is a statewide geodatabase available that expands to about 30GiB in memory, and therefore, any operations on this object takes a long time. Using geofeather, we optimized the storage mechanism, creating a much smaller initial database, stored as initial_data.feather (ensure to have initial_data.feather.crs to retain the geographic coordinates). ● Preprocessing Done - There are multiple functions defined at the top of the analysis file (Analysis.ipynb) which we used to preprocess data to remove erroneous entries, and only retain data that we need. This data is saved as final_data.csv and final_data.geojson. The geojson preserves geographic data. ● Most important to you are the filter_luc, filter_bldg and filter_out_bldg functions. The filter_luc function takes in the original df and a filter name, and filters out parcels with the given affiliation. Filter_bldg takes in the original df and filters out parcels with buildings. Filter_out_bldg does the opposite. ● A sample use case would be if you wanted to get all the parcels owned by the housing department, you would load up final_data.csv, and run filter_luc(df, 'housing'). Scitkit Learn and spaCy for basic machine learning and regression tools. Tableau and ArcGIS for mapping.
Other Readings	Background Testimony from Nika Elugardo: https://drive.google.com/open?id=1JgOE7YqXsC6MrKkULXJ2tdjt05XBQJ2C