Deliverable 1:

This week the team successfully filtered the MAPC dataset by the following categories:
- Land use codes (all codes under 91*,92*,97*)
- Building data
  - Ziba specified:
    bldg_value - for condos, generally includes land value
    bldg_area - may include garages, stairwells, basements, and other uninhabitable areas.
    bldgv_psf - building value $ per sq foot
  - Additional:
    sqm_bldg - parcel area estimated to be covered by a building (sq meters)
    pct_bldg - % parcel area estimated to be covered by a building
- Poly_typ
  o Fee
  o Tax

In addition to filtering the dataset, we were able to scrape state agency names and addresses from https://www.mass.gov/state-a-to-z. We found a total of 315 agencies on the website. The data has been stored in a csv file.

One important thing to note regarding the web scraping is that we were not able to obtain all agency addresses due to the fact that the website we scraped from did not have a unique webpage from every agency.

This information will be useful going forward. We plan to use it to clean the owner_name category of the MAPC dataset and use it as a baseline for standardizing state agency names.

Our next steps are to clean the owner_name category and begin to look into demographic stats on land parcels/blocks

Filter_bldg:

	bldg_value	bldg_area	bldgv_psf	sqm_bldg	pct_bldg
2	307400.0	4040.0	76.089109	13.312941	0.388600
3	258200.0	3648.0	70.778509	-0.999900	0.000000
4	199300.0	3782.0	52.696986	182.962820	6.404145
5	206500.0	4752.0	43.455387	129.203406	8.479037
6	276200.0	4634.0	59.602935	236.458540	10.219435

Filter_luc/poly_typ:

	luc_1	luc_2	luc_adj_1	luc_adj_2	poly_typ
1385	9280	9280	928.0	928.0	FEE
1427	9280	9280	928.0	928.0	FEE
1456	9280	9280	928.0	928.0	FEE
2137	9700	9700	970.0	970.0	FEE
2199	9700	9700	970.0	970.0	FEE