

## Project Deliverable 2

More data should have been collected to perform a more thorough analysis of the data and attempt to answer one additional question relevant to your project proposal which you will submit as a pull request.

### Checklist

1. Collect and pre-process a secondary batch of data
  2. Refine the preliminary analysis of the data performed in PD1
  3. Answer another key question
  4. Refine project scope and list of limitations with data and potential risks of achieving project goal
  5. Submit a PR with the above report and modifications to original proposal
- 

The four steps to finish the target is as follows:

1. 15-Minute City Parcel Identification Where are the parcels in Boston which have access to all essential amenities (healthcare, food access, recreation, education, commercial, and social amenities) within a 15-minute walk, bike, or transit ride? Where are the parcels that have limited access (ie 3 or fewer of the categories covered)? Where are the parcels with no access?
2. Neighborhood-scale trends Which zip codes are predominantly made up of parcels with 15-minute access to all amenities? Which zip codes are predominantly parcels with limited or no 15-minute access?
3. Gap Analysis For each zip code that is missing some or all amenities within 15-minute access, which amenities are missing for each zip code?
4. Demographic Analysis How many residents live in 15-minute city parcels vs non 15-minute city parcels? What are the demographics of residents living within 15-minute parcels? What are the demographics of residents not living within 15-minute parcels? (Specifically: Income, Race, English-speaking vs Non-English Speaking )

So far we have almost finished data gathering and preprocessing. For essential services, we have found 5554 services (grocery:1876 |healthcare:2493 |supermarket:68 |hospital:156 |park:373 |open space:588). For parcels, we originally have 142924 parcels (these are parcels defined in zoning view).

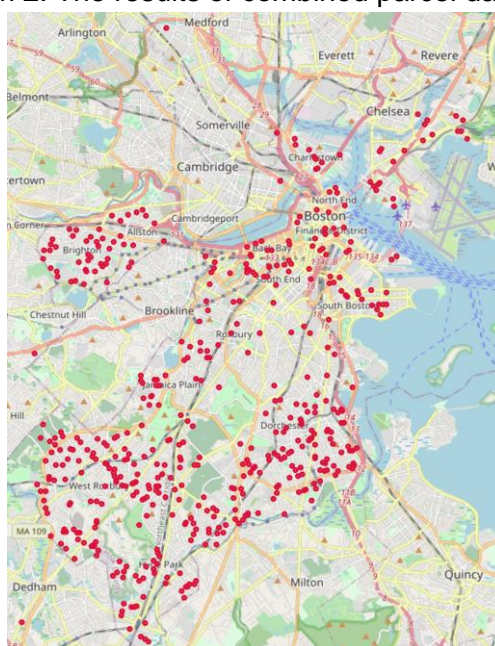
As shown in Diagram1, there are parcels located on the same street. To reduce the necessary calculation, we clustered those parcels that are close to each other into a bigger parcel using latitude and longitude and updated the lat/long range of the bigger parcel according to the newly added parcel. An example of the updated bigger parcels is Diagram 2:

'27 -37 CHESTNUT ST #103, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #107, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #109, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #102, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #104, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #106, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #108, CHARLESTOWN,02129',  
 '27 -37 CHESTNUT ST #110, CHARLESTOWN,02129',

<Diagram 1. Examples of adjacent parcels that are combined into one bigger parcel>

	lat	lon	addresses
0	42.000528	-71.326871	[8 WASHINGTON ST, CHARLESTOWN,02129, 449 -463 WASHINGTON ST #8A, BOSTON,02111, 449 -463 WASHINGTON ST #8B, BOSTON,02111, 449 -463 WASHINGTON ST #8C, BOSTON,02111, 449 -463 WASHINGTON ST #8D, BOSTON,02111, 449 -463 WASHINGTON ST #8E, BOSTON,02111, 10 WASHINGTON ST, CHARLESTOWN,02129, 10 WASHINGTON ST #1, CHARLESTOWN,02129, 10 WASHINGTON ST #2, CHARLESTOWN,02129, 10 WASHINGTON ST #3, CHARLESTOWN,02129, 449 -463 WASHINGTON ST #10A, BOSTON,02111, 449 -463 WASHINGTON ST #10B, BOSTON,02111, 449 -463 WASHINGTON ST #10E, BOSTON,02111]
1	42.000829	-71.326667	[12 WASHINGTON ST, CHARLESTOWN,02129, 12 WASHINGTON ST #1, CHARLESTOWN,02129, 12 WASHINGTON ST #2, CHARLESTOWN,02129]
2	42.002777	-71.325398	[36 30 WASHINGTON ST, CHARLESTOWN,02129, 30 WASHINGTON ST #30-2A, CHARLESTOWN,02129, 32 WASHINGTON ST #32-B, CHARLESTOWN,02129, 32 WASHINGTON ST #32-1, CHARLESTOWN,02129, 32 WASHINGTON ST #32-2, CHARLESTOWN,02129, 32 WASHINGTON ST #32-3, CHARLESTOWN,02129]
3	42.004037	-71.325050	[40 WASHINGTON ST, CHARLESTOWN,02129]
4	42.010260	-71.322008	[72 WASHINGTON ST, CHARLESTOWN,02129, 72 WASHINGTON ST, HYDE PARK,02136, 82 WASHINGTON ST, CHARLESTOWN,02129]
5	42.013311	-71.320596	[86 WASHINGTON ST, CHARLESTOWN,02129]

<Diagram 2. The results of combined parcel data points>



<Diagram 3. Visualization of the 500 sampled large parcels>

After merging, we have 12265 parcels and we are ready to move on to the other key question -- “What percentage of residents are 15 minutes within essential amenities in a parcel of land?”. Firstly, for each convinced parcel, we find six corresponding essential services by estimating the distances between services and parcels using Manhattan distance. It is reasonable to believe Manhattan distance is a good approximation of real distance in modern city.

After that, we randomly sampled 500 combined parcels (geo locations are shown in Diagram 3). We used Matrix Distance API to generate the distance, along with the cycling time, between each parcel and the closest amenities. Diagram 4 shows the visualization of the 500 sample parcels.

	Parcel lat	Parcel lon	supermarket dist	supermarket cycling	parks dist	parks cycling time	grocery dist	grocery cycling time	healthcares dist	healthcares cycling	hospitals dist	hospitals cycling time	openspaces dist	openspaces cycling
0	42.37625199	-71.05970413	7476	1397	565	127	239	35	106	16	531	104	318	172
1	42.2338921	-71.1328608	3115	702	248	45	534	94	1488	388	2735	702	249	46
2	42.28253673	-71.0673458	404	107	776	208	396	122	279	67	651	237	576	142
3	42.2904597	-71.0881017	1760	409	2171	484	74	67	451	72	730	203	820	137
4	42.28672913	-71.17110727	1579	355	718	188	1480	329	1104	256	2183	453	1219	295
5	42.3373555	-71.0333902	2214	453	200	58	287	108	61	7	3613	817	293	111
6	42.29479684	-71.06274506	1439	334	391	86	162	59	125	51	1925	414	503	132
7	42.34519799	-71.14812767	913	251	340	87	74	33	469	66	504	132	376	76
8	42.2936349	-71.0635427	1324	268	685	152	411	92	50	8	899	209	338	54
9	42.33795117	-71.0396833	1670	340	407	93	273	111	228	63	3277	788	829	211
10	42.27244455	-71.10344482	2407	544	1796	489	545	162	858	229	3158	791	815	191
11	42.3250093	-71.0642507	431	97	899	281	201	37	150	27	1778	367	1205	335
12	42.33771665	-71.0450829	1507	310	1138	262	442	127	201	104	3113	758	937	193
13	42.2690757	-71.0944341	3236	800	1227	320	19	3	722	179	2872	690	667	173
14	42.29394595	-71.14246109	1261	255	872	158	781	168	1509	397	1140	276	1744	370
15	42.26766188	-71.11488972	1806	401	2050	443	628	111	24	7	2866	692	521	90

<Diagram 4. Distance and cycling time between parcels and amenities>

To represent the result of analysis on “15-minute” parcels, we use 1 and 0 to mark whether or not a parcel has 15-minute access to each kind of amenities.

	A	B	C	D	E	F	G	H
1	Parcel lat	Parcel lon	supermarket	parks	grocery	healthcares	hospitals	openspaces
2	42.37625199	-71.05970413	0	1	1	1	1	1
3	42.2338921	-71.1328608	1	1	1	1	1	1
4	42.28253673	-71.0673458	1	1	1	1	1	1
5	42.2904597	-71.0881017	1	1	1	1	1	1
6	42.28672913	-71.17110727	1	1	1	1	1	1
7	42.3373555	-71.0333902	1	1	1	1	1	1
8	42.29479684	-71.06274506	1	1	1	1	1	1
9	42.34519799	-71.14812767	1	1	1	1	1	1
10	42.2936349	-71.0635427	1	1	1	1	1	1
11	42.33795117	-71.0396833	1	1	1	1	1	1
12	42.27244455	-71.10344482	1	1	1	1	1	1
13	42.3250093	-71.0642507	1	1	1	1	1	1
14	42.33771665	-71.0450829	1	1	1	1	1	1
15	42.2690757	-71.0944341	1	1	1	1	1	1
16	42.29394595	-71.14246109	1	1	1	1	1	1

<Diagram 5. Bit representation of “15-minute” parcels>

Assuming that we defined the parcels to be underserved when their access to essential amenities is less than 4, the result shows that 92.4% of parcels have access to all 6 types of amenities within 15 minute of cycling; 1% of parcels are underserved in terms of essential amenities and 0.8% of parcels have no 15-minute access to any amenities:

6 amenities:  $462/500 = 92.4\%$

5 amenities:  $31/500 = 6.2\%$

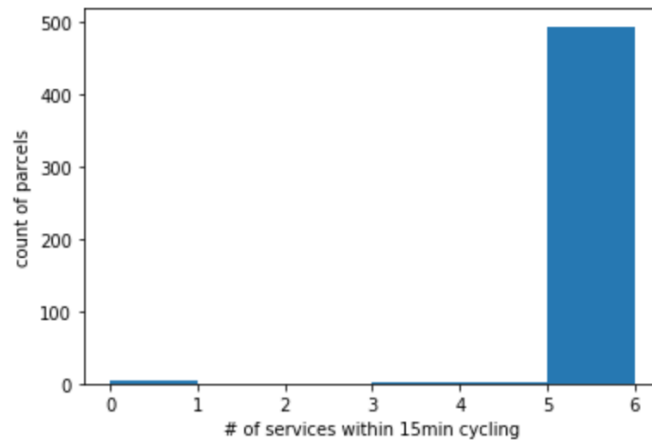
4 amenities:  $2/500 = 0.4\%$

3 amenities:  $1/500 = 0.2\%$

2 amenities:  $0/500 = 0\%$

1 amenities:  $0/500 = 0\%$

**0 amenities:  $4/500 = 0.8\%$**

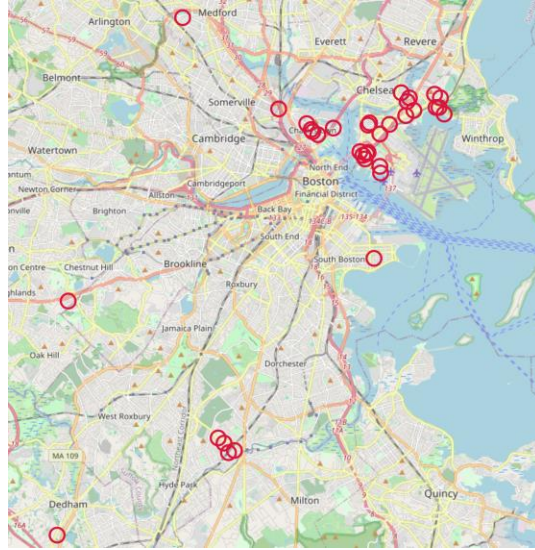


<Diagram 6. Number of essential amenities deployed in sample parcels >

Among the 500 samples, the coverage rate of supermarket, parks, grocery, healthcares, hospitals, openspaces is [0.972, 0.988, 0.99, 0.992, 0.942, 0.992] respectively. The overall coverage is pretty good. But we can still discuss the comparative shortage. Considering the fact that open spaces are parks are similar. The most restricted amenities are supermarkets and hospitals. The shortage of hospitals is foreseeable because medical resources are also very limited. But the reason for the shortage of supermarkets is not clear.

The visualization of sample parcels which does not have all 6 amenities are shown below. Their total number is 38. Their zip code and count are shown below:

'02114': 1,  
'02124': 1,  
'02126': 3,  
'02127': 1,  
'02128': 22,  
'02129': 6,  
'02132': 1,  
'02135': 2,  
'02136': 1.



<Diagram 7. Visualization of the sampled parcels which doesn't have all six amenities>

For the next step, we will repeat the same step on the whole dataset, then conduct demographic analysis on the parcels of land, by comparing the number of residents in 15-minute parcel land to non 15-minute parcel land. Then we will analyze the demographic features – race, income, non-English speakers, education, and housing density – of residents living within the 15 minute parcel lands and those not living in them. As for the final step, we will perform more visualizations with the data collected and start writing a draft of the report.

\* list of limitations with data and potential risks of achieving project goal do not change; please refer to previous deliverables.