City of Boston 15-minute city (Spring 2021)

Team members: Mingcheng Xu, Liangzhu Zhang, Jieying Liu, Soowhan Park, Xingyu Wang

This is a SPARK project.

Client:

Nayeli Rodriguez

Technologist, Mayor's Office of New Urban Mechanics

City of Boston

nayeli.rodriguez@boston.gov

Organization:

City of Boston

Organization Description:

The municipal government for the city of Boston overseeing all public works and governance for Boston.

Project Type:

Data Science

Project Description

We would like to research the features that can make Boston a "15-minute city" where essential services and resources are all 15 minutes from each other via train, bus, or walk. Essential services within the city include: hospitals, healthcare providers, grocers, supermarkets, parks/green spaces, etc. We will target dense urban areas for the 15-minute city concept, this will be done by identifying parcels of land in the city with the highest concentration of essential services and then finding the shortest paths between them. Many cities such as Melbourne that have implemented the plan seem to heavily emphasize bike lanes and access to parks in the scheme, so a worthwhile addition to the project might be identifying locations that can easily be integrated into wide bike lanes and potential public lands to be marked out for green spaces/parks. Further, the C40 mayor's report emphasizes the use of bike lanes as a post-COVID economic recovery idea to build more closely knit cities and improve accessibility and reduce the reliance on public transit and cars that posed a COVID risk for many.

Data Sets:

Boston Master businesses and establishments dataset

Grocers dataset from MAPC

Environmental justice demographics

Open space dataset

Suggested Steps:

Step One: Combine the three datasets and process them into clear categories of essential amenities - grocers, green spaces/parks, hospitals/healthcare services, supermarkets, gyms/recreational centers. Use the business_type fields within the Boston Master businesses and establishments dataset. Base geographic unit will be land parcels.

Step Two: Identify the 15-minute city parcels - Where are these parcels of land located in Boston with the essential amenities within a 15-minute commute of one another? Employ the CLARANS algorithm to cluster parcels of dense essential amenities.

Step Three: Identify zip codes with higher concentrations of these “15-minute parcels” and the zip codes with limited “15-minute parcels”

Step Four: Analyze the missing essential amenities within zip codes that are starved of “15-minute parcels”.

Step Five: Conduct demographic analysis on the parcels of land, compare the number of residents in 15-minute parcel land to non 15-minute parcel land. What are the demographic features of residents living within the 15-minute parcel lands and those not living in them? This would include: race, income, non-English speakers, housing density, etc.

Questions to be answered:

1. What percentage of residents are 15 minutes within essential amenities in a parcel of land?

   the percentages of parcels that lack each kind of essential amenity (supermarkets, groceries, hospitals, health cares and green spaces) are 0.468, 0.0298, 0.472, 0.0316, 0.039.

2. Which areas of the city are underserved in terms of a lack of essential amenities?

   In aspect of groceries and health care (hospital not included), there should not be a city area that is underserved. However, areas including Northeast Corndor and West

Roxbury seem to have a low density of parks and some parcels in those areas may not be within a 15-min range. When it comes to hospital and supermarket, more areas (e.g., parts of Roxbury and Dorchester) in the south are clearly not covered in a 15-min range (walking or biking).

After a more careful analysis, the most supermarket-underserved areas are 2132,2126 and 2129; the most hospital-underserved areas are 2136, 2128, 2127 and 2210. Please refer to deliverable 3 for more details.

Tools and Methods:

Data processing: Pandas and NumPy will be used to clean and process the datasets into one master dataset.

Clustering & Graph search: CLARANS algorithm - Works much better for high-dimension spatial clustering than DBScan that becomes inefficient and inaccurate when multiple attributes are added. We have multiple essential service types so it would be best to use CLARANS.

Kruskal's algorithm implementation in Python - To find the shortest distance between multiple sets of points, this will be done within our high-density parcels.

Data Visualization: Matplotlib, Seaborn, Tableau for visualizing the geospatial data as needed.

Additional step after step two if there is additional time:

Within each parcel of land find the shortest distance between the essential amenities via Kruskal's algorithm - that finds the shortest distance between a set of edges and nodes. We can further customize this by excluding paths (edges) that are not viable for bike lane expansion.

Once shortest pathways are established between essential services within each dense parcel of land, find junctions of roads and paths that can easily be connected to increase mobility between the city. In the City of Melbourne report, this is emphasized as an important step.

Explore whether the multiple parcels of lands can be interconnected and find routes that are shortest to each parcel of land from one another. Can repeat Kruskal's algorithm or try Djikstra's algorithm for this step.

Further background information:

One of the metrics used to measure the value of dense cities is called WalkUP (Walkable Urban Places) premiums as defined by SmartGrowth Traffic. SmartGrowth takes the share of total retail, office, and multifamily housing space located in WalkUPs and then ranks the metros. The WalkUP premium shows what people are willing to pay more rent for dense areas based on their desirability.

List of limitations:

- Many datasets are old.
- The open space dataset contains some entries that have incomplete information so it is not possible to locate its exact location on map.
- We use Google map api, which has a usage cap. This may limit the amount of analysis we can do.
- This analysis is done using parcels as the basic unit, but we don't have information about population in each parcel. Incorporating population data with parcel data can better reflect the real situation.

Potential risks of achieving project goal:

- Because many data are old, our work may not be able to reveal the most up-to-date result. For example, some shops may be closed due to this pandemic.