MAPC Digital Equity

Team 1 - Final Report Draft V1.0

—

Gonzalo Rosales

Benjamin Sui

Richard Andreas

Jeffrey Li

Jake Lee

Overview

The Metropolitan Area Planning Council (MAPC) is the regional planning agency serving the people who work and live in the 101 cities and towns of the metropolitan Boston area.

The MAPC would like to allocate funds from the Coronavirus Aid, Relief, and Economic Security (CARES) Act towards increasing broadband access across the state of Massachusetts, deciding how best to allocate this money based on their dataset of historical broadband speeds with hundred of features such as income, ethnicity, and speed data.

We, Team 1 of this project, will be specifically focused on mapping out the digital divide across municipalities in Massachusetts to analyze discrepancies between the resident's needs and existing service from current providers. We will accomplish this by analyzing FCC data, in addition to the MAPC's existing features.

Dataset

Description

Through its mandatory Form 477 reporting process, the FCC collects a variety of data from Internet Service Providers (ISPs) every six months. This data comes in two forms: one revolves around internet access and available broadband subscriptions, while the other focuses on maximum advertised upload and download speeds reported by ISPs. Both datasets are classified by census tract. For this project, focus is on the latter set with max advertised speed data.

Limitations

- FCC data is updated and published about a year or more after each 6-month reporting period. The current data is from 2019, which means there is unfortunately no data available yet for 2020 and beyond.
- Since ISPs self-report data, and they're only required to report maximum advertised upload and download speeds, the FCC dataset is very flat. For example, in most municipalities, Comcast reports 850 mbps for download data, giving the impression that all localities receive this speed. As the dataset exclusively represents advertised speeds, there is little bearing on real-world speeds.
- Difficult to distinguish between cities that own their own broadband infrastructure.
- Census tracts to zip codes is not a one to one mapping.

Useful Links

- https://www.digitalinclusion.org/data-research/
- https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477
- https://www.fcc.gov/form-477-county-data-internet-access-services
- https://www.fcc.gov/general/explanation-broadband-deployment-data
- https://datacommon.mapc.org/calendar/2020/december

Leading Strategic Questions

While we are aware that finding answers to all the following questions may not be possible through the provided FCC dataset, they drive our approach to the problem and are ones we have kept in mind throughout the process:

- What are discrepancies in coverage and speeds among MA municipalities?
- Is there any evidence of digital redlining? For example, do municipalities with a high percentage of underrepresented groups receive poor coverage relative to the rest of the population?
- How do broadband providers vary in the three gateway cities, Revere, Everett, and Quincy, that the MAPC has asked us to focus on?
- What are the leading predictors for higher broadband speeds in MA?
- What is the case for extending FCC requirements beyond just advertised speeds?
- Current standards set by the FCC classify good broadband as 25 mbps upload and 5 mbps download. Are these speeds still sufficient?

Insights and Progress

General Insights

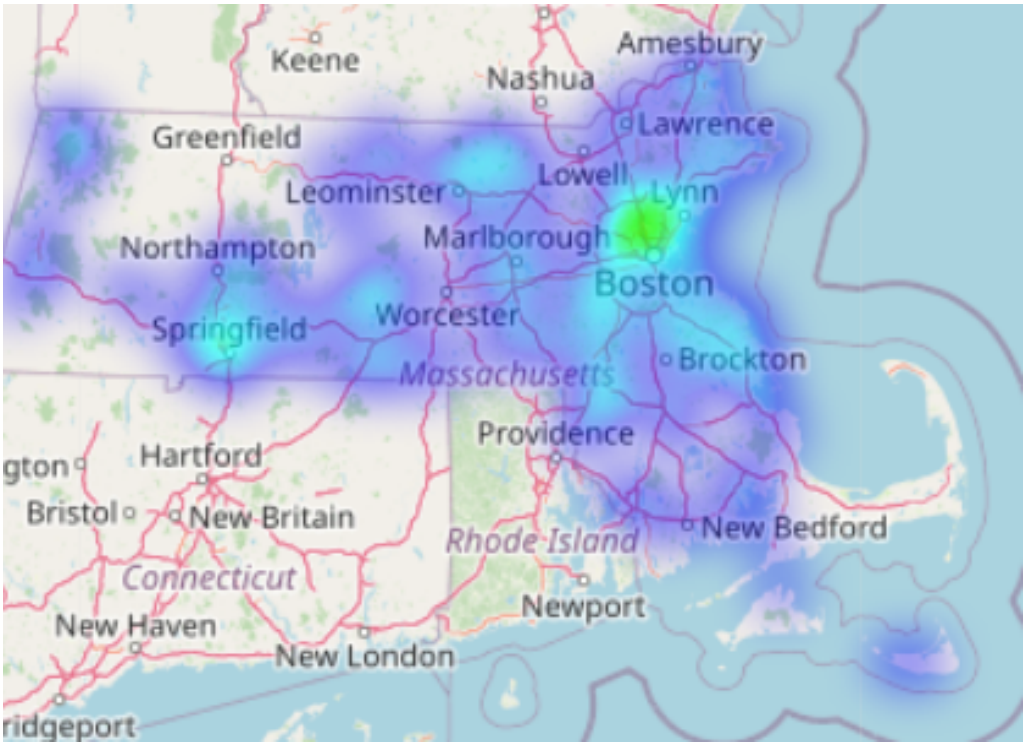
- Given the aforementioned flatness of the FCC data, much of our findings further justified this sentiment. By integrating the work of our partner team for this project into our own, we expect to gain better insight on the true speeds observed per area and take note of any possible discrepancies between what is advertised and what is actually provided to customers.
- Combining our reported data with our partner team's real-world data could potentially show which tracts and/or providers have the worst juxtaposition.

Tools and Methods

- Jupyter Notebooks were our main development tool, as they provided the impactful ability to run demanding code blocks separate from the rest of the program, as well as providing in-line visualizations.
- Pandas and NumPy acted as our main libraries for preprocessing and displaying data as DataFrame data structures.
- SQLite provided query processing and local database access.
- Matplotlib and Folium allowed for data visualization.

Mapping from Census Tracts to Zip Codes

As part of our process to import the FCC data, we created a folder that maps municipalities to their respective census tracts. However, every municipality maps to multiple census tracts. When visualizing the data in a heatmap created using Folium, this overlap causes the number of providers per area to appear denser than in reality.

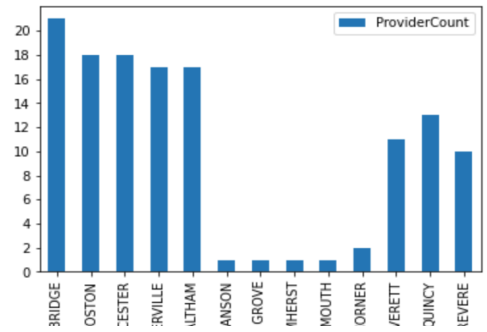


Comparison Between Providers

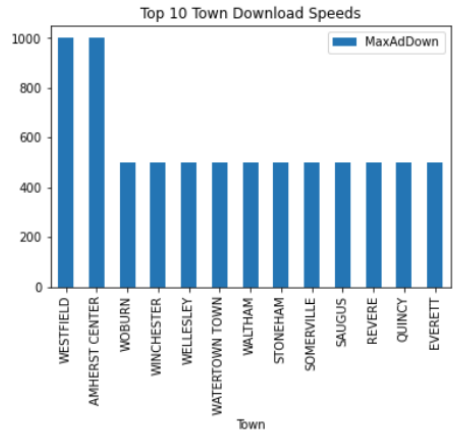
We were able to filter the number of providers per municipality. We then focused on the top five municipalities with the most providers, the bottom five municipalities with the least number of providers, and the three gateway cities: Revere, Quincy, and Everett.

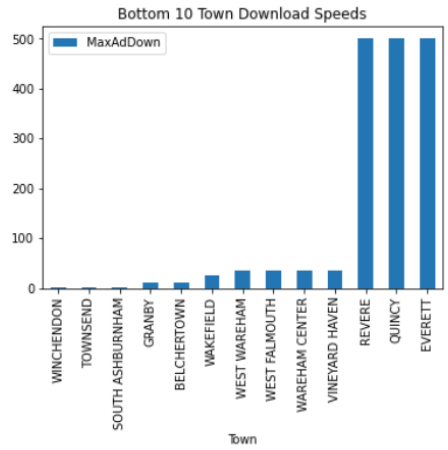
Cambridge was the most popular city with the highest count of providers, while Hanson, Ocean Grove, South Amherst, and West Falmouth all stood at only one provider. The towns of Everrett, Quincy, and Revere all had an average number of providers.

	Town	ProviderCount
0	CAMBRIDGE	21
1	BOSTON	18
2	WORCESTER	18
3	SOMERVILLE	17
4	WALTHAM	17
5	HANSON	1
6	OCEAN GROVE	1
7	SOUTH AMHERST	1
8	WEST FALMOUTH	1
9	BLISS CORNER	2
10	EVERETT	11
11	QUINCY	13
12	REVERE	10

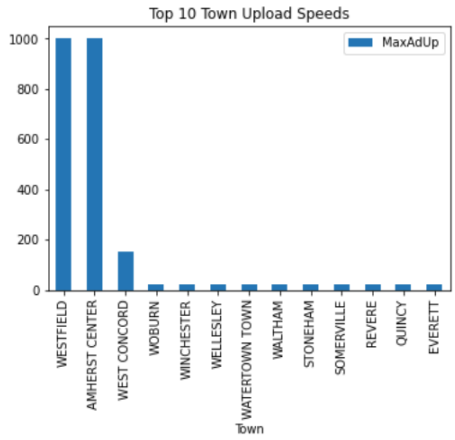


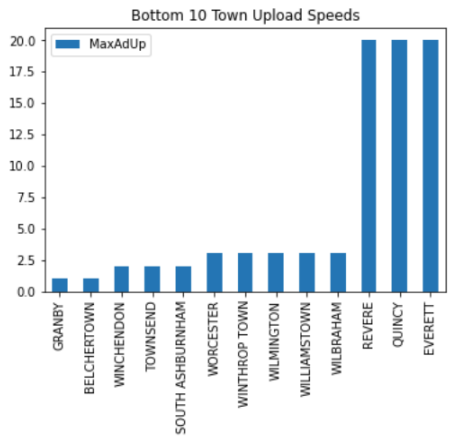
	MaxAdDown	Town
0	1000.0	WESTFIELD
1	1000.0	AMHERST CENTER
2	500.0	WOBURN
3	500.0	WINCHESTER
4	500.0	WELLESLEY
5	500.0	WATERTOWN TOWN
6	500.0	WALTHAM
7	500.0	STONEHAM
8	500.0	SOMERVILLE
9	500.0	SAUGUS





	MaxAdUp	Town
0	1000.0	WESTFIELD
1	1000.0	AMHERST CENTER
2	150.0	WEST CONCORD
3	20.0	WOBURN
4	20.0	WINCHESTER
5	20.0	WELLESLEY
6	20.0	WATERTOWN TOWN
7	20.0	WALTHAM
8	20.0	STONEHAM
9	20.0	SOMERVILLE

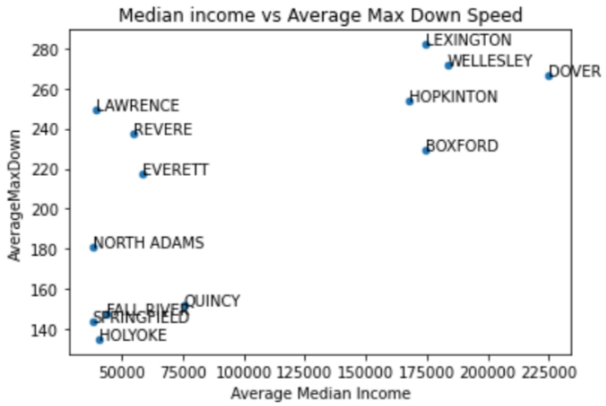




When mapping out the towns with the highest and lowest speeds, Westfield and Amherst Center both come out on top with maximum advertised download and upload speeds of 1 Gig. Interestingly, the difference among the highest download speeds is not as significant as the difference among the highest upload speeds. After Westfield and Amherst Center, we found that there is a large contrast with the next highest upload speed. With an advertised upload speed of 150mpbs, West Concord is the municipality with the third highest *upload* speed. One potential reason is that Westfield and Amherst Center are serviced by ISPs providing Gig speed internet. The presence of gigabit internet providing ISPs can be a reason for the large variance amongst municipalities as fiber optic internet is still widely unavailable in Massachusetts.

Note that the earlier point about FCC data being very flat is relevant here. We can see that many towns share the same speeds, even though this may not be indicative of actual conditions

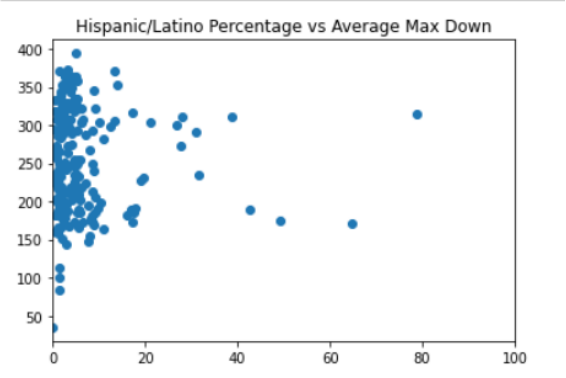
Demographics Data and Speeds



When analyzing the relation between a census tract's median income against the average max download speed provided in that region, we note an upwards trend. Disregarding the inclusion of the three gateway cities, we can observe clear clustering - the five districts with the highest median income against the five districts with the lowest median income. However, this observation is much like a truism; wealthier districts having higher average internet speeds is to be expected.

Currently, the observed visualization of the data does not have any clear indications of digital redlining. However, we must once again consider the fact that such measurements are based on advertised speeds provided by their respective ISPs. There are obvious PR and financial incentives for ISPs to advertise appealing speeds. Thus, we can not definitively dismiss the occurrence of digital redlining until we are able to observe municipality broadband services beyond just advertised speeds.

	Town	AverageMaxDown	AverageMaxUp	Average Median Income
0	DOVER	266.60	116.52	224784.000000
1	WELLESLEY	271.43	114.13	183744.285714
2	BOXFORD	229.41	98.39	174340.000000
3	LEXINGTON	281.84	92.27	174233.166667
4	HOPKINTON	253.57	112.75	167733.500000
5	SPRINGFIELD	143.97	5.65	38017.162162
6	NORTH ADAMS	181.18	8.01	38142.500000
7	LAWRENCE	249.54	110.73	39653.166667
8	HOLYOKE	135.08	5.35	40761.636364
9	FALL RIVER	147.38	5.87	43967.920000
10	REVERE	237.55	19.68	55013.500000
11	QUINCY	152.04	11.08	75583.850000
12	EVERETT	217.56	17.50	58378.500000



Challenges

- As previously stated, the FCC data is very flat due to the metrics being advertised speeds. There is not much insight into real world conditions allowed. We make the case that these plots prove that there is a need for more transparent reporting from ISPs, as they are very potentially misleading. We suspect the digital divide is greater.
- Mapping data from census tracts to zip codes and vice versa is a tedious and imperfect process due to inconvenient naming conventions within the FCC data.
- Lots of variables contribute indirectly to broadband access and speed, so lots of trial and error was necessary when determining next steps. This obtuseness was compounded by the flatness of the provided FCC dataset. We are curious to see what Team 2, which used real time data from Ookla, found.