

MAPC: Broadband Digital Equity in MA

This document is a transcription of the information listed for this spark! project, as given before pitch day. It's purpose is to record an original version of the goals, steps, and desired outcomes for this work.

Contact

Ryan Kelly

RKelly@mapc.org

Digital Services lead at the MAPC

Matt Zagaja

mzagaja@mapc.org

Lead civic web developer at the MAPC

Organization

The Metropolitan Area Planning Council - MAPC

Organization Description

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

Project Type

Data Science

Project Description

The MAPC would like to allocate newly released funds from the [CARES act towards increasing broadband access](#) across MA. They will decide how best to allocate this money based on their dataset of historical broadband speeds with hundreds of features such as income, ethnicity, % of uploads/downloads, etc. Time my series analysis on internet use by hour and day will also be done to capture broad trends.

The second part of the project will specifically focus on Municipal Digital Divide Planning efforts on Gateway cities and analyze differences amongst provider speeds so they can choose the best provider to expand broadband access.

Data Sets

The dataset will be provided by the client from MLab and Ookla, it is part of a larger dataset containing historical broadband speeds across the world coming in at Terabytes of data. However, we will only be working with MA historical data.

Ookla dataset (Broadband speed data)

MLab dataset (Broadband speed data)

FCC dataset (Broadband provider coverage data) - Shows how many broadband providers cover each census tract. The census tracts will mostly nest within municipalities.

Census county subdivision data for MA

Suggested Steps

Step one: Clean and preprocess the MLab & Ookla data for MA, this will involve some format of SQL queries and Pandas preprocessing in Python after. Duplicate the client's approach in querying the data, the approach will be provided by the client.

Step two: Overlay the MLab, Ookla, and FCC datasets with the Census municipality (county subdivision) data. The initial three datasets should have geographic units down to the municipality level.

Step three: Analyze discrepancy in broadband coverage and speeds across MA municipalities using demographics information with the merged dataset above. Is there a noticeable presence of digital redlining - Communities of color receiving poorer coverage and speeds? Setup and conduct a regression test to predict broadband speeds using demographics, income levels, and housing density as predictors. We want to best understand which of these variables contribute towards faster broadband speeds. A further step could also be clustering similar broadband speeds and analyzing their similarities.

Step three: Focus on the Gateway cities - Revere, Everett, and Quincy and compare differences in provider speeds here. The outcome should be visualizations showing the difference. This step will help with their Digital access plan.

Step four: Conduct a time series analysis for the state by month, day, and hour to study trends of when internet usage is concentrated and find possible explanations.

Step five: Summarize findings using data visualizations for the time series analysis and provider speed differences in the gateway cities.

Strategic Questions

1. What are the discrepancies in coverage and speeds among MA municipalities? Identify key features
2. Is there presence of digital redlining? Black communities and communities of color receiving poor coverage relative to the rest.
3. How do broadband provider speeds vary in the three gateway cities - Revere, Everett, and Quincy?
4. What are the hours of highest internet activity? Test this as a hypothesis and find possible explanations.
5. What are the leading predictors for higher broadband speeds in MA?

Additional Information

Tools & Methods

Data pre-processing: Pandas, NumPy for processing and cleaning the data. BigQuery to create SQL queries and obtain subsets from the database.

Machine Learning: scikit-learn, pytorch for machine learning and regression tools

Data Visualization: Matplotlib, Seaborn, Tableau for all kinds of interactive visualizations

Related Links & Resources

- [Running List](#) of potential report visualizations and reports
- [Running List](#) of terms and definitions
- NDIA Cleveland AT&T Digital Redlining [Report](#)
- ISLR [Video](#) discussion about NDIA Report
- ISLR blog on the topic of [broadband](#)
- RWJF thread with link to Brookings [Report](#)