	MAPC: Broadband Digital Equity in MA
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 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.
Student Team	Due to high demand for this project, the student team working on this has been split into two teams. We are part of **Team #1**, consisting of the following students: ● Richard Andreas (ra7296@bu.edu) ● Benjamin Sui (bsui@bu.edu) ● Gonzalo Rosales (grosales@bu.edu) ● Jake Lee (thejake@bu.edu) ● Jeffrey Lee (jeffrey1@bu.edu) **PM:** Kamran Arif (kamran55@bu.edu)

Meeting Information	**Weekly Check-in with PM:** Thursday, 1 - 1:30 PM **Bi-Weekly Check-In with Client:** Friday, 1:30 - 2:30 PM
Data Sets	Since this project was split between two teams, our team has been assigned with exploring the following dataset from the **Digital Inclusion Alliance**, in partnership with the FCC. ● https://www.fcc.gov/general/explanation-broadband-deployment-data ● https://www.fcc.gov/form-477-county-data-internet-access-services ● https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477 ● https://www.digitalinclusion.org/data-research/
Suggested Steps	**Step one:** Clean and preprocess the Digital Inclusion Alliance data, 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:** Overlaying the data with Census (county subdivision) data. The initial dataset 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:** Summarize findings using data visualizations for the 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?
Possible Limitations	● FCC data updated as of 2019, so may need to find up-to-date data elsewhere for 2020 ● Distinguishing between cities that own their own broadband infrastructure ● Classifying between different forms of internet access, i.e. 5G vs fiber vs traditional broadband. ● Overlaying broadband data with location data ● Representation of time series visualizations, i.e. representing broadband speed over time for individual municipalities vs. larger regions.
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. Use SQL Lite Processing can be used for query processing. https://realpython.com/python-sql-libraries/#mysql *Machine Learning:* scikit-learn, pytorch for machine learning and regression tools *Data Visualization:* Matplotlib, Seaborn, Tableau for all kinds of interactive visualizations **What is considered good broadband is 25Down 5Up** **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 **Useful Airtables from MAPC** ● https://airtable.com/shrv7Uv7LMWkKDW1b/tblzkWLOP8k1syWEP ● https://airtable.com/shrZkjM3DUASjEVmk/tbl0LhDwl98S1epqT

	Datacommon Links ● https://datacommon.mapc.org/calendar/2020/december ● https://datacommon.mapc.org/calendar/2020/april ● https://www.measurementlab.net/data/ ● https://www.speedtest.net/insights/blog/announcing-ookla-open-datasets/ ● https://www.fcc.gov/document/fcc-annual-broadband-report-shows-digital-divide-rapidly-closing ● https://www.verizon.com/coverage-map/