nyconnect

The competitiveness of the telecommunications industry in New York City

---

Progress Report Three

Erik Lopez, Wesley Chioh, Yichen Liu, Ziyu Yan
CUSP Mentor: Junaid Khan
External Sponsor: Zachary Gold, New York City DOITT

Abstract

The New York City Department of Information Technology and Telecommunications (DoITT) wants to understand if the current telecommunication industry regulations are effective at generating a competitive industry environment that can provide cheap and reliable internet access to all New Yorkers. This is due to the long-term economic importance of universal broadband internet access and the renewal of franchise agreements between NYC and the providers in July 2020. This project tested for the presence of an optimal number of Internet Service Providers (ISPs) in New York City that can improve Internet accessibility. Current internet infrastructure coverage data was overlaid with pricing and socio-economic data to determine the relationships that can lead to the desired state. In a departure from just infrastructural, economic and demographic data, tweets were also mined for analysis of New Yorkers sentiments towards ISPs and the broader internet market. It is estimated that there is an optimal number of ISPs for the NYC urban internet market but is yet to be achieved due to infrastructural and socio-economic unevenness. NYC's government may have to directly intervene and enhance connectivity in less well-connected neighborhoods.

Problem Statement and Research Outcomes

The team wants to discover if there is an optimal number of ISPs for the New York City market, thereby improving Internet accessibility in terms of coverage area, household adoption rates, and affordability for all New Yorkers.The aim is to produce a policy white paper that explains NYC's broadband market structure, the factors influencing availability of broadband internet and subscription rates, and how DoITT can optimize the regulatory framework.

Literature Review

The Federal Communications Commission (FCC) defines a wired home broadband service as an “Internet access connection providing speeds of at least 25Mbps download and 3Mbps upload.” (Talbot, Hessekiel, & Kehl, 2018, p.4) A reliable broadband internet connection is increasingly crucial to NYC’s long-term economic vitality. Universal broadband access is estimated to increase economic output in the city by US$142 billion, and create up to 165,000 jobs (The New York City Mayor’s Office of the Chief Technology Officer [NYC CTO], 2020).

However, in New York City, more than 1.5 million residents do not have any means to connect to the internet (NYC CTO, 2020). Even among residents with wired home broadband service perceive it as extremely costly (Talbot *et al.,* 2018). Unevenness in broadband affordability and availability is a microcosm of social inequality and ineffectiveness of competition and regulation of the industry. The shortcomings can be attributed to the lack of competition between ISPs because the industry is effectively a spatially segmented oligopoly (Brake and Atkinson, 2019).

Brake and Atkinson (2019) argue that the structure of urban telecommunications markets means that greater competition does not necessarily result in lower prices and maximum consumer benefit. Broadband infrastructure incurs high fixed capital costs that are subject to rapid technological change. As a result, only big firms can achieve economies of scale and thrive. Meanwhile, smaller companies can neither surmount the barriers to entry nor enjoy the economies of scale, necessitating subsidies. In this case, the Herfindahl-Hirschman Index (HHI) that measures industry concentration, and competition may not truly capture the latent efficiencies within the market that are a function of the present regulatory framework.

Presently, telecommunication companies own the infrastructure. NYC's Department of Buildings (DOB), the FCC, and DoITT regulate the installation of this infrastructure. Internet service provision is licensed by the city through cable franchises. An alternative is for cities to construct their own fiber broadband networks. (Talbot *et al.*, 2018) Examples of cities that have done this successfully include Chattanooga, TN and Santa Monica, CA (Lampland and Mitchell, 2014). Talbot *et al.* (2018) argue that a majority of these cities have been able to offer faster broadband connectivity at lower prices, forcing incumbents to improve their services in order to effectively compete. However, the sample size of this study was small, given the limited number of community-owned fiber networks. Plus, these networks serve communities with less than a million residents; it is unclear whether NYC will be able to replicate and scale up effectively.

Public intervention can focus on helping all New Yorkers become potential customers who are able to pay for such services. Siangjaeo (2016) found that a price subsidy for lower-income residents in Minnesota was most effective at improving broadband adoption rates. A subsidy to help pay for a broadband subscription would help connect those folks who cited price as an impediment to get a better or any internet service (Horrigan, 2009), this would specifically help areas where internet infrastructure may be underutilized. However, this subsidy should not be limited to wired connections, as Lee (2017) found that people of lower incomes preferred and prioritized a wireless internet service over wired broadband service.

Data

A more detailed description of the datasets used can be found in Table 2 in the appendix.

Dataset	Description
Internet Master Plan: Adoption and Infrastructure Data by Neighborhood	Produced by the NYC Mayor's Office of the CTO as a checkpoint for NYC's progress in meeting its goals

Internet Master Plan: Broadband Choice and Speed by Census Block	Produced by the NYC Mayor's Office of the CTO to reflect FCC 477 Form data specific to NYC
Mobile Telecommunication Franchise Poletop Installation Locations	Locations approved by DOITT for the installation of telecommunications infrastructure
FCC Form 477 (2017-2019)	Telecommunication firms' declaration of service provision to FCC
American Community Survey 2017-2018	US Census Bureau survey of economic and demographic conditions, sourced for NYC Census block age, race, income information, and broadband subscription rates.
TIGER Census Blocks shapefile	Produced by the United States Census Bureau based on the 2010 Decennial Census to denote population groups by area
NYC NTA shapefile	NYC Department of City Planning grouped neighborhoods
2010 Census Tract to NTA Equivalency Table	Produced by NYC Department of City Planning to assign tracts to neighborhoods
Broadband Provider-Compare_v.2	Provided confidentially by DoITT, listing the price and service content for main ISPs in NYC

Table 1: Data Sources

Methods

Geoprocessing in Python and QGIS matched the datasets to spatial boundaries of NYC in terms of census blocks, tracts, and NTAs. New shapefiles were created, and existing ones were reprojected. This enabled visualization of the datasets, as requested by DoITT.

To test Brake and Atkinson's (2019) argument of spatial segmentation, local and global spatial autocorrelation was calculated. A queen contiguity-based spatial weights was obtained from the geospatial datasets. The degree to which contiguous spatial units reflected similar values, and whether said observed patterns were randomized outcomes were then calculated. This was performed on the ISPs, pricing, and broadband subscription datasets.

Spatial autocorrelation was positive. Thus, the cumulative effect of the spatial configuration of each census tract's independent variables on the dependent variable should be considered in any regression. Furthermore, since spatial autocorrelation differs by boroughs, it cannot be assumed that the dependent variable behaves uniformly across space. Spatial heterogeneity must be considered when measuring the impact of demographic, racial, and infrastructural variables. This was done by introducing a spatially-lagged term in a two-stage least squares regression model while allowing the coefficients to vary by boroughs (Drukker, Egger & Prucha, 2010; Deng, 2018). Cross-section regression of pricing data was also performed by identifying potential demand side and supply side factors affecting broadband prices. However, the lack of updated pricing data may affect the accuracy of the regression models..

KMeans clustering was also performed to identify any latent structures within the demographic and business activity data that could have resulted in the observed broadband subscription patterns. However, it was largely inconclusive where demographic data were concerned. It was more informative with the business activity dataset.

A Bayesian Network was also constructed using demographic data such as race, income, age, and broadband subscription rates, plus the number of unique broadband ISPs at the Census block level. The network showed which factors had an impact on broadband subscription rates.

Text mining based on Twitter data was also performed to provide more insights into the public's perception of current telecom service. Twitter analysis includes tweet retrieval, text preprocessing, exploratory data analysis, sentiment analysis and word cloud visualization. Tweets mentioning a provider and services between 2010 and 2020 were scraped. A

Word2vec-LSTM model evaluated each tweet's sentiment on a scale of 0 (negative) to 1 (positive). A word cloud plot showcased words that appeared frequently and their frequency. However, tweets are relatively short and may not convey a lot of meaningful information.

Results

Exploratory data analysis revealed that on average 68.7% of households in each NYC Neighborhood Tabulation Area (NTA) have at least a broadband connection (Figure 4). Nearly all of NYC census blocks have at least one provider of cable or fiber broadband-speed Internet access. Only 19% of blocks are limited to just one ISP. Among the main ISPs, Altice covers the Bronx and East Brooklyn, Charter serves the rest of NYC, and Verizon is available across NYC.

Internet service infrastructure displays strong signs of spatial autocorrelation at the scale of census tracts. There are clusters of spatially contiguous census tracts that share similar levels of infrastructural provision in terms of price and the number of ISPs. At a confidence interval of 95%, the spatial patterns seen in Figure 1 are not randomized outcomes.

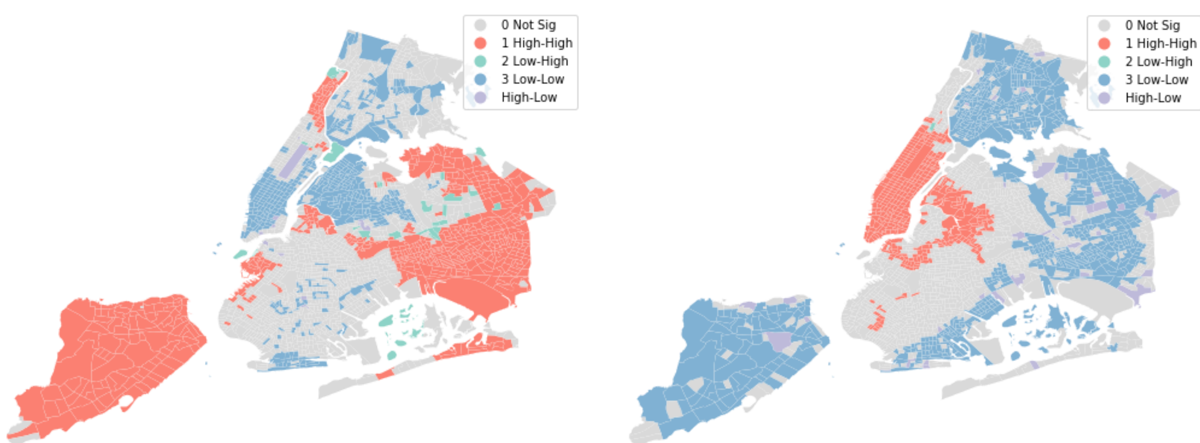

Figure 1: Spatial autocorrelation maps for subscription prices (left) and number of ISPs (right)

Between 2017 and 2018, the only two years in which data were available, the number of ISPs in Manhattan increased, but decreased in Brooklyn. However, changes in the number of ISPs in each census tract appear to have very little impact on the rate of household broadband subscription. Subscription rates remained fairly stable, as shown in Figure 5 of the appendix.

KMeans clustering was performed on the estimated number of business and household accounts in each NTA as shown in Figure 6. The results suggest that clusters with a higher business to residential accounts ratio are associated with above-average ISP counts (Figure 7). The converse is true for NTAs that are more residential in nature. This was not entirely unexpected as there are ISPs that specifically cater to businesses and do not serve residential accounts. Unsurprisingly, these clusters tend to be found in Manhattan.

The Bayesian network showed that broadband subscription rates are impacted by the percentage of residents in a Census block that make less than $50,000 per year. Additionally, more white affluent areas tend to have access to more ISPs. These results demonstrate that access to broadband internet is more a function of income than the number of service providers.

Income levels were significant variables in the wired and mobile broadband regression models. Census tracts with wealthier residents are associated with higher wired broadband subscription rates. More interestingly, the number of ISPs was negatively correlated with subscription rates. This implies that in the long-run, an increase in the number of ISPs can result in market fragmentation that reduces subscription rates.

A proxy for the connectivity of New York City is the demand and supply of ISPs in NYC as represented by Figure 2. The less elastic demand curve suggests people may be less responsive to the increase of internet providers unless they can afford one. The demand curve is relatively more price-elastic than the supply curve. Hence, a change in the number of providers may not lead to a significant decrease in the consumer's expectation of broadband price. The intersection of these two curves is considered as the equilibrium of the broadband market. The optimal number of providers is around 2, which is already met in 80% of all NYC census blocks. The equilibrium monthly price ranges from $69.10 to $89.00 per subscription.

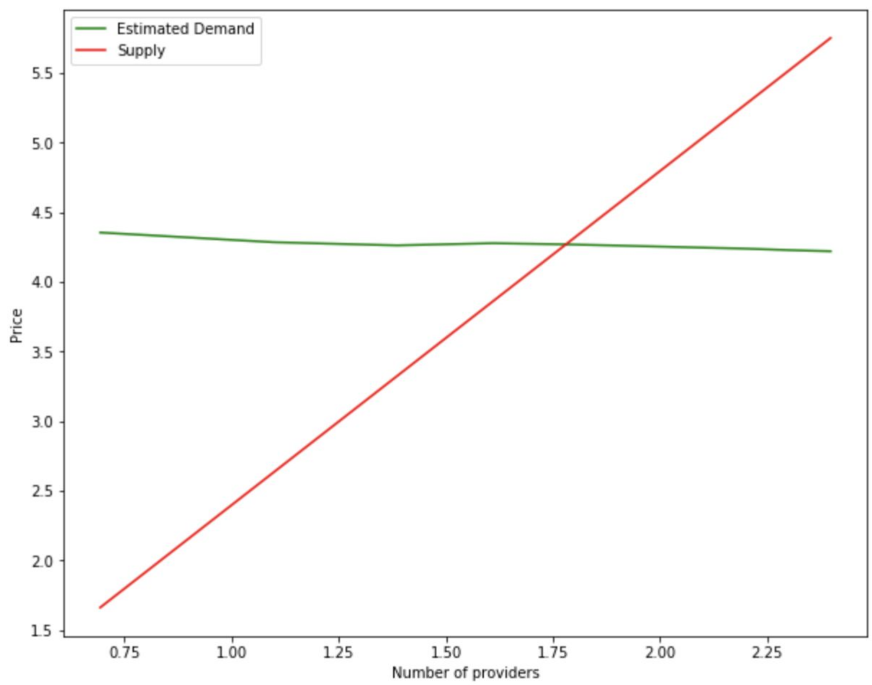

Figure 2: Demand curve and supply curve of broadband services

According to Figure 8, more than half of tweets are complaints. This might indicate that in most cases, users only tweet when there is a problem with the service. As shown in Figure 9, the tweet sentiment gradually increased between 2013 and 2018, and then quickly decreased between 2018 and 2020. There were fewer tweets (Figure 10) between 2015 and 2018 as compared to that of other years. This could mean that between 2015 and 2018, consumers were basically more satisfied, and fewer less complaints were tweeted. Verizon, AT&T, T-Mobile, Sprint account for a vast majority of the tweets. Among them, Verizon covers mobile service, cable and fiber while the other three mainly focus on mobile broadband in NYC. Thus in most cases people tweet about their mobile provider and mobile service.

Considering the tweets which display extremely negative sentiment shown in Table 3, the word cloud (Figure 11) shows what words are commonly used in these tweets. As guessed above, mobile service is most commonly mentioned in this kind of tweets. Other significant keywords (e.g. internet, data, wifi, cable, network, speed, 3g, signal, LTE, slow) show that most people complained about the performance aspect (connectivity, accessibility, speed). In an oligopoly market, monopoly companies do not have enough incentive to maintain high service quality. This result indicates that related government agencies and public policy should not only focus on the quantity of providers in NYC but also pay attention to the service quality of providers.

Research Plan

The team has largely been able to adhere to the timeline set out in the previous progress report. Most analyses are complete at this stage. The only outstanding items are an estimation of the number of beneficiaries under the Lifeline program, and the website on which work has already begun. A Gantt chart can be found in the appendix as part of Figure 12.

Challenges

A key challenge faced was the lack of updated information. For example, the latest American Community Survey data available is 2018, and the FCC only had estimates valid as of June

2019. This hampered the team’s ability to provide current analyses as economic and demographic conditions have changed significantly since then.

A corollary was the lack of open data. Important information such as the number of subscribers per ISP, and updated pricing information were proprietary. Proxies to estimate the number of users had to be created, as a result.

Third, the FCC 477 form data, which is a cornerstone of our analysis relies on a flawed data collection methodology. It is self-reported by ISPs which opens up the possibility of error. Further, a census block is “served” even if just one building in a block has access to a particular ISP. This opens up the possibility of an overestimation of infrastructural provision.

Team Collaboration Statement

Wesley took the lead in geospatial visualizations and analyses, as well as KMeans clustering. Yichen worked on creating a tweet-based consumer sentiment analysis model. Ziyu was in charge of market and economic analyses. Erik is currently building the website where our final results will be published; he also took care of the Bayes Network analysis. The team worked together on policy suggestions, and a white paper that has been sent to DoITT for review.

Sources

Brake, D., & Atkinson, R. D. (2019). *A Policymaker's Guide to Broadband Competition*. Information Technology and Innovation Foundation.

Chen, Yidan, and Hafshah Ashrawi. "Equity of Access to Public Wifi in NYC." *RPubs*, 2019, rpubs.com/yidanchen95/561896.

Department of Public Service Office of Telecommunications. In the Matter of a Study on the State of Telecommunications in New York State, Staff Assessment of Telecommunications Services (2015). Retrieved from http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId={1CBAE64C-C4B9-483A-8317-C76CB914344D}

Executive Office of the President. (2015). *Community-Based Broadband Solutions*.

Federal Communications Commission (2010). *Connecting America: The national broadband plan.* ERIC Clearinghouse

Guowang Miao; Jens Zander; Ki Won Sung; Ben Slimane (2016). Fundamentals of Mobile Data Networks. Cambridge University Press. ISBN 978-1107143210.

Horrigan, J.B. (2009) Obama's online opportunities II: If you build it, will they log on? Washington,DC: Pew Research Center. Available: https://tech-insider.org/internet/research/acrobat/0901.pdf

Hung-Yu Wei, Jarogniew Rykowski, Sudhir Dixit. WiFi, WiMAX and LTE Multi-Hop Mesh Networks: Basic Communication Protocols and Application Areas. Wiley, (2013).

Lampland, E., & Mitchell, C. (2014). *Santa Monica City Net: An Incremental Approach to Building a Fiber Optic Network*. Institute for Local Self-Reliance. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.7810&rep=rep1&type=pdf

Lee, H. J., & Whitacre, B. (2017). Estimating willingness-to-pay for broadband attributes among low-income consumers: Results from two FCC lifeline pilot projects. Telecommunications Policy, 41(9), 769–780. https://doi.org/10.1016/j.telpol.2017.04.001

McKetta, Isla. "LinkNYC Proves Public Wi-Fi Can Be Free, Fast and Far-Reaching." *Speedtest Stories & Analysis: Data-Driven Articles on Internet Speeds*, Speedtest Stories & Analysis: Data-Driven Articles on Internet Speeds, 31 Oct. 2017, www.speedtest.net/insights/blog/linknyc-2017/.

Palmer, K. (2015). *Broadband internet*. New York: Novinka.

Siangjaeo, Sorawit (2015). Minnesota Broadband Development Policy: An Analysis of Fixed Cost and Direct Usage Subsidies. Humphrey School of Public Affairs. https://conservancy.umn.edu/bitstream/handle/11299/172663/Siangjaeo,_Broadband%20subsidies%20analysis.PDF?sequence=1

Talbot, D., Hessekiel, K., & Kehl, D. (2018). Community-owned fiber networks: Value leaders in america. *Berkman Klein Center Research Publication*, (2018-1).

Telecommunications- NAICS 517, Statista Industry Report – USA, January 2020

The New York City Internet Master Plan, The New York City Mayor's Office of the Chief Technology Officer (2020). Retrieved from https://tech.cityofnewyork.us/wp-content/uploads/2020/01/NYC_IMP_1.7.20_FINAL-2.pdf

U.S. Congressional Research Service. Broadband Data and Mapping: Background and Issues for the 116th Congress. (R45962; Oct. 16, 2019), by Colby Leigh Rachfal. Text from: Congressional Research Digital Collection; Accessed: March 27, 2020

Ripamonti, P. (2019). *Twitter Sentiment Analysis*, https://www.kaggle.com/paoloripamonti/twitter-sentiment-analysis

Ranjan, Sandeep & Sood, Sumesh & Verma, Vikas. (2018). Twitter Sentiment Analysis of Real-Time Customer Experience Feedback for Predicting Growth of Indian Telecom Companies. 166-174. 10.1109/ICCS.2018.00035.

Nur Azizah Vidya, Mohamad Ivan Fanany, Indra Budi, Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers, Procedia Computer Science, Volume 72, 2015, Pages 519-526, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.12.159.

Appendix

Dataset	Structure	Main characteristics	Limitations
Internet Master Plan: Adoption and Infrastructure Data by Neighborhood	CSV file with 52 attributes at the scale of NTA	Basic population count of the NTA with the availability and utilization of different means of internet access and street infrastructure survey	Amalgamation of data, some of which are estimated or outdated
Internet Master Plan: Broadband Choice and Speed by Census Block	CSV file 14 attributes at the scale of each census block	Count of the number of commercial and residential ISP providers within each census block	FCC's reporting may not truly reflect availability for consumer
Mobile Telecommunication Franchise Poletop Installation Locations	CSV file with 20 attributes at the scale of each location	Description of installation at said location with longitude and latitude, and street address provided.	Dataset lists the firm installing the infrastructure but not the specific company leasing it
FCC 477 (2017-2019)	CSV file with 21 columns where each row is a provider and internet service details at the Census block.	Contains an internet provider's offered speeds (download/upload), infrastructure used for delivery, location of service (Census block)	June 2019 is the latest available
TIGER Census Blocks shapefile	ESRI shapefile with geometry	Spatial representation of census blocks in the US	From the 2010 Decennial Census
NYC NTA shapefile	ESRI shapefile with geometry	Spatial representation of NTAs in NYC, identified by NTA code	Incompatible with datasets on the scale of census tracts and blocks without further transformation
2010 Census Tract to NTA Equivalency Table	CSV file with 5 attributes at the scale of census tracts	Provides information on which NTA each tract corresponds to	Inconsistent naming of boroughs and tract IDs
Broadband Provider-Compare_v.2	Excel spreadsheet consisting of 6	Provides the current and historical price of mobile broadband, cable and fixed	Historical data is available only for cable service, but

dataset	sheets	broadband for ISPs in NYC as well as according detailed service content and coverage	not for mobile broadband and fixed broadband. Might be outdated.

Table 2: Detailed description of data sources

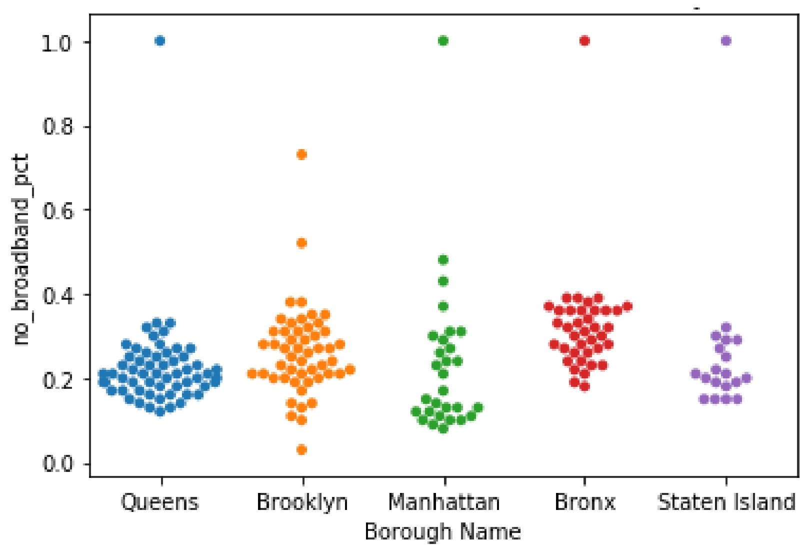

Figure 3: Households (%) with no form of broadband connection in each NTA by borough

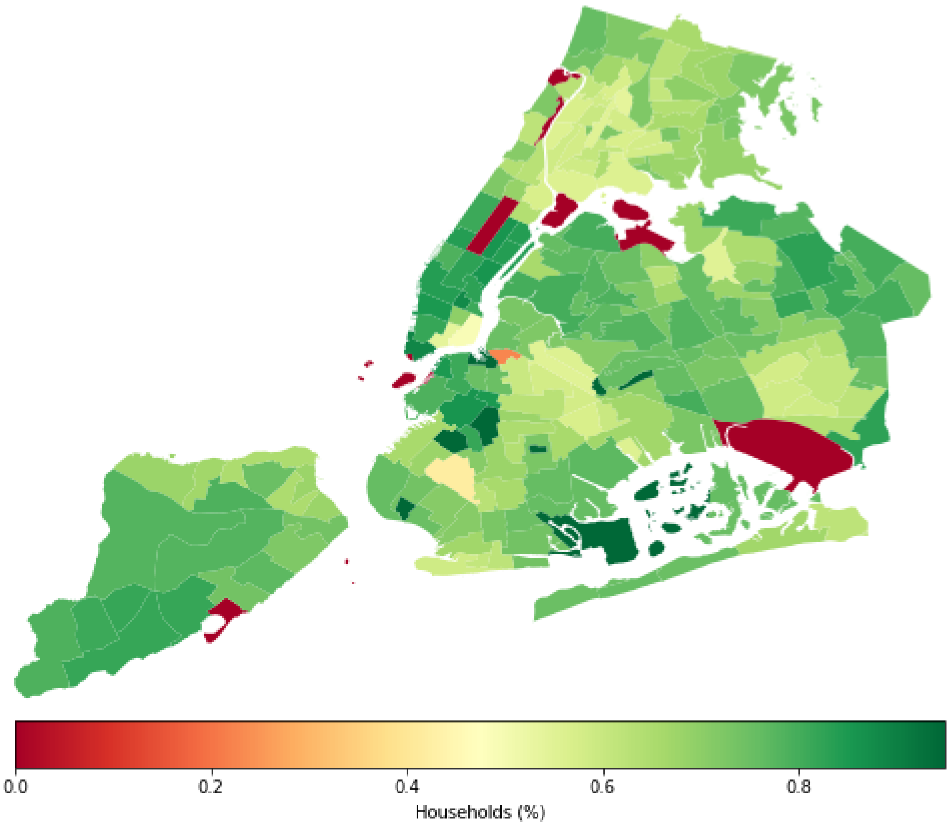

Figure 4: Geographical distribution of households in each NTA without a broadband subscription

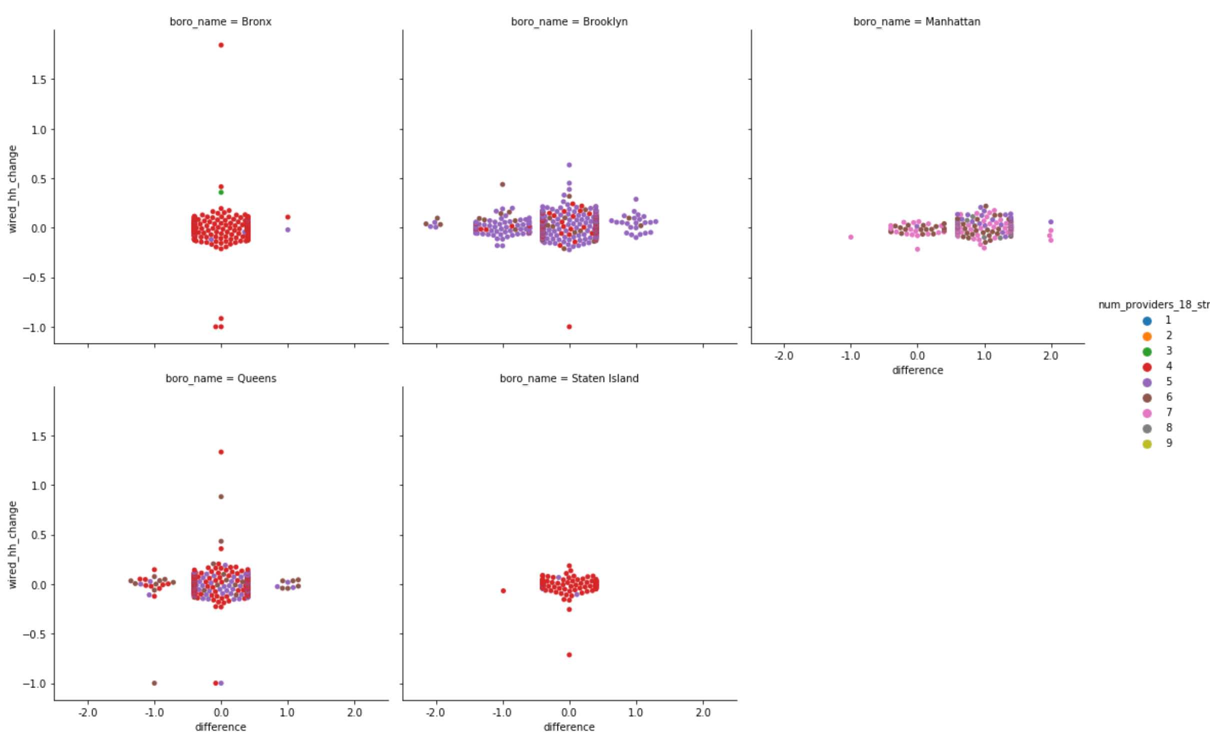

Figure 5: Swarm plot of changes in ISPs and household wired broadband subscription rate by borough

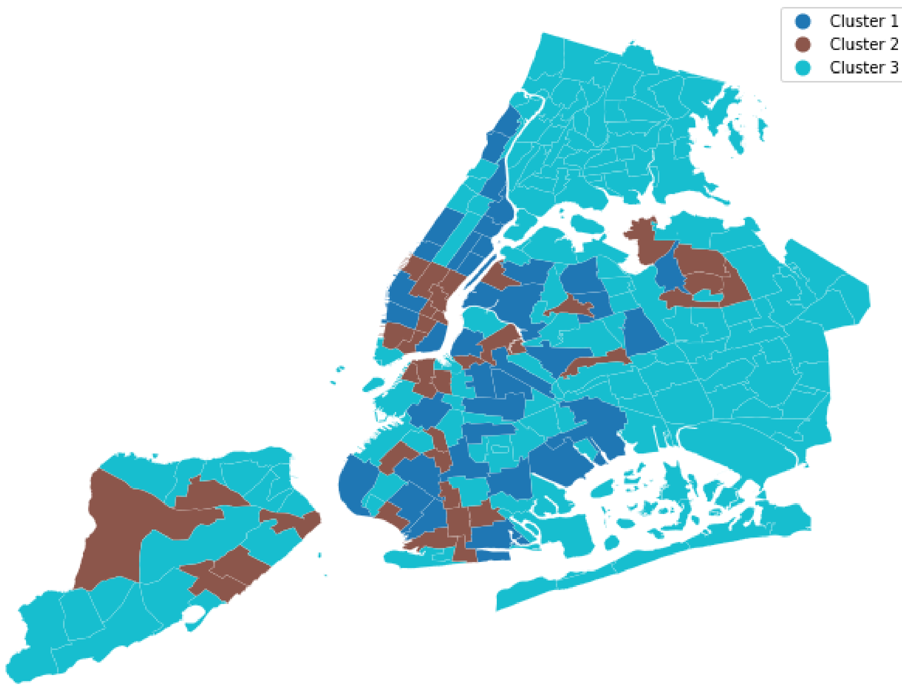

Figure 6: KMeans clustering of census tracts based on business patterns

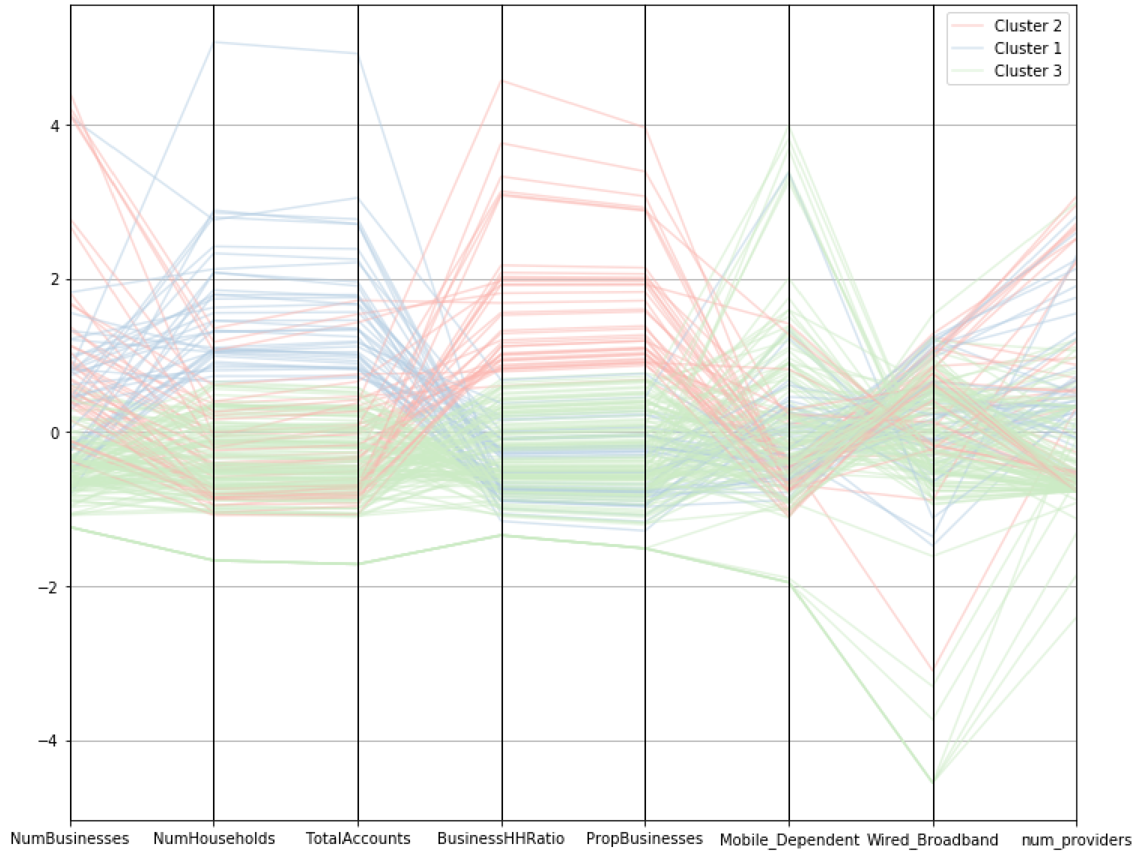

Figure 7: Parallel coordinates plot of KMeans clustering based on business patterns (Figure 6)

Tweet content	score	Company
"@sprint I've been havin too mny problems wit my !!! Sad real sad!!! ??!!"	0.00319	Sprint
'Verizon wireless wifi fucking sucks'	0.0041384	Verizon
'@Verizon wifi sucks in my house just sucks'	0.004889	Verizon
'Verizon Internet sucks!!! '	0.0050412	Verizon
'@verizon internet sucks!'	0.0050412	Verizon
'I really need to switch data plans... Verizon really sucks.'	0.0051683	Verizon
'AT&T sucks. The 3G signal is in the stone age... 3 minutes for uploading a picture on Facebook. That really sucks.'	0.005189	AT&T
'"@TheRealRick_: Fucking Verizon Wifi Box Sucks 😒"Wrrdd'	0.0059957	Verizon
"My cable box really broke Verizon Really ain't shit only a little drop of water got in it & now it's completely broke"	0.0072183	Verizon
"@Pmazz21 wish I was but unfortunately sucks and I don't have cable or Internet"	0.0073278	Verizon
'I hate verizon fios I have no tv no Internet no phone called they said the main box broke my f day off with nothing to do its raining out'	0.0074932	Verizon
'i hate verizon now i gotta fucking call these skumbags to fix my damn wifi house phone and computer!! ugh'	0.007557	Verizon
"AT&T 3G is still down, so I'm walking around with a Verizon LTE MiFi in my jacket pocket. Sad."	0.0075979	Verizon
"AT&T 3G is still down, so I'm walking around with a Verizon LTE MiFi in my jacket pocket. Sad."	0.0075979	AT&T
"Fckin Verizon's internet sucks! Can't work."	0.0078632	Verizon
"My Internet, WiFi and television is down. Pissed to no end!! Service won't be repaired till tomorrow bet 8&5pm. Verizon sucks"	0.0078916	Verizon
"And this is @oskrNYC's sad AT&T LTE \xa0"	0.008118	AT&T
'no access to internet fios sucks'	0.0085835	Verizon
'I hate 3G on sprint. Shit is slow'	0.0088	Sprint
'@bonjourev ugh they suck. My Verizon Internet is super slow'	0.0094012	Verizon

Table 3: Top 20 most negative tweets with score and company

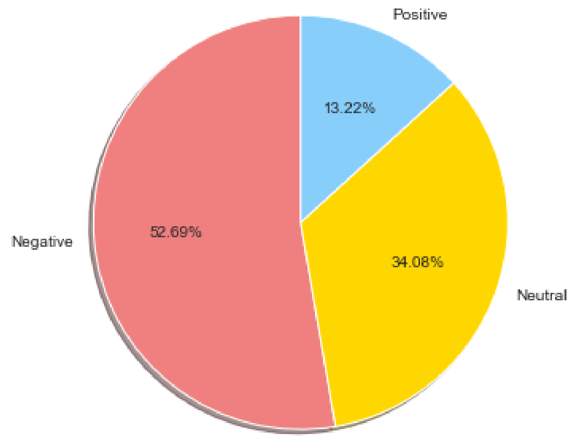

Figure 8: The proportion of the three sentiments

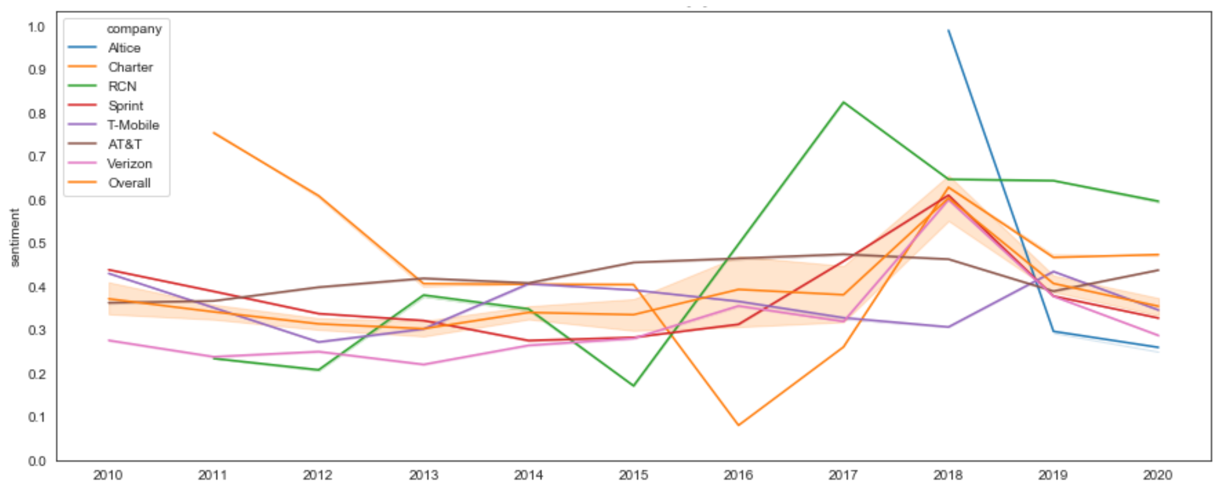

Figure 9: tweet sentiment by year and provider

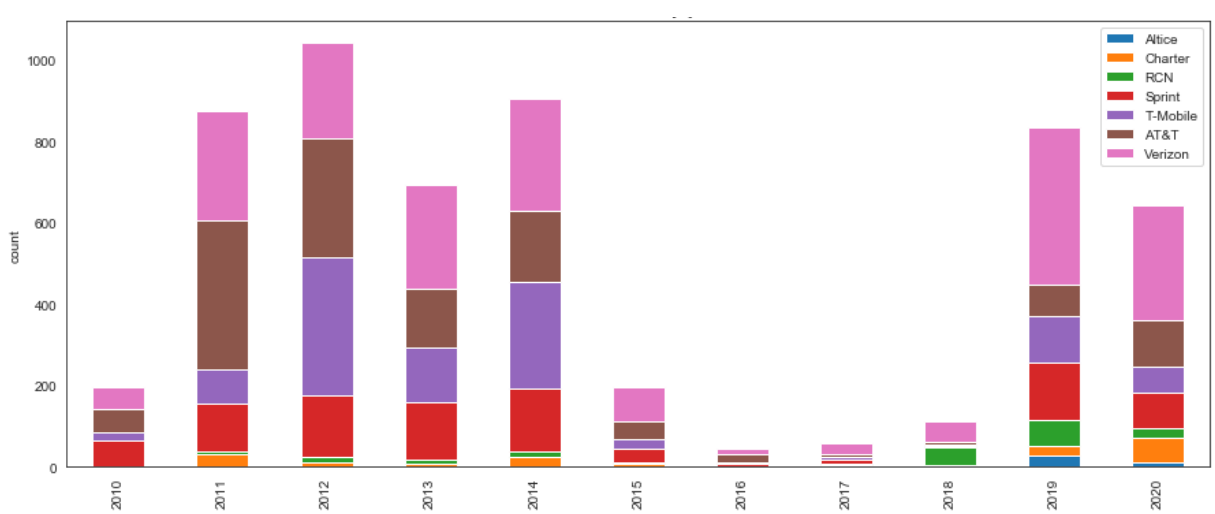

Figure 10: tweet count by year and provider

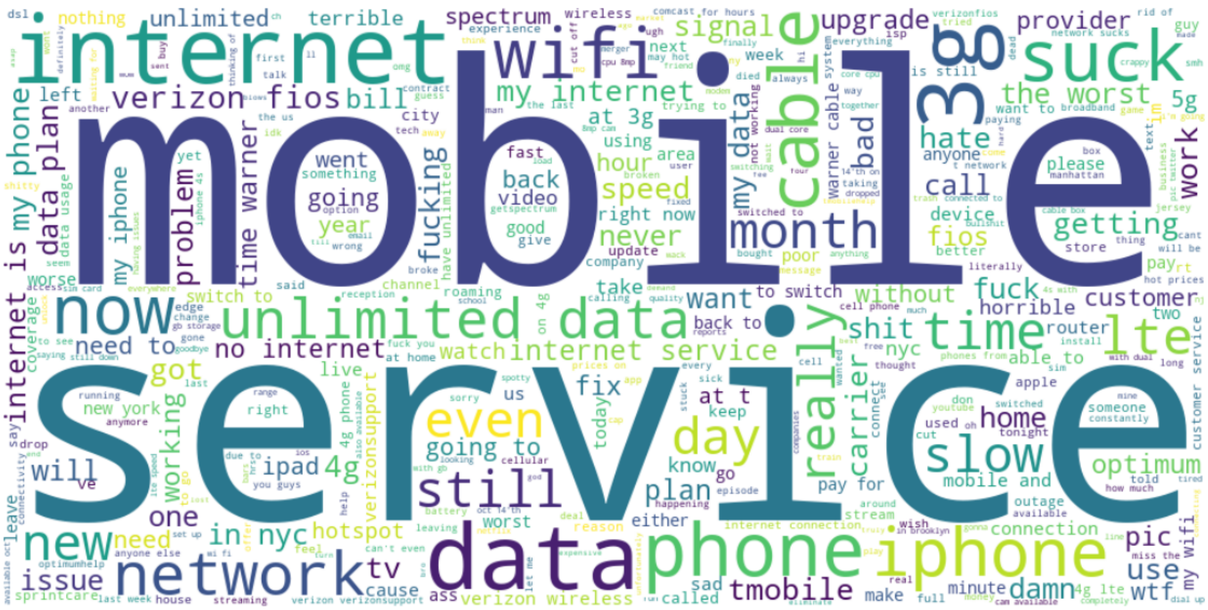

Figure 11: Word cloud for tweets with very negative sentiment

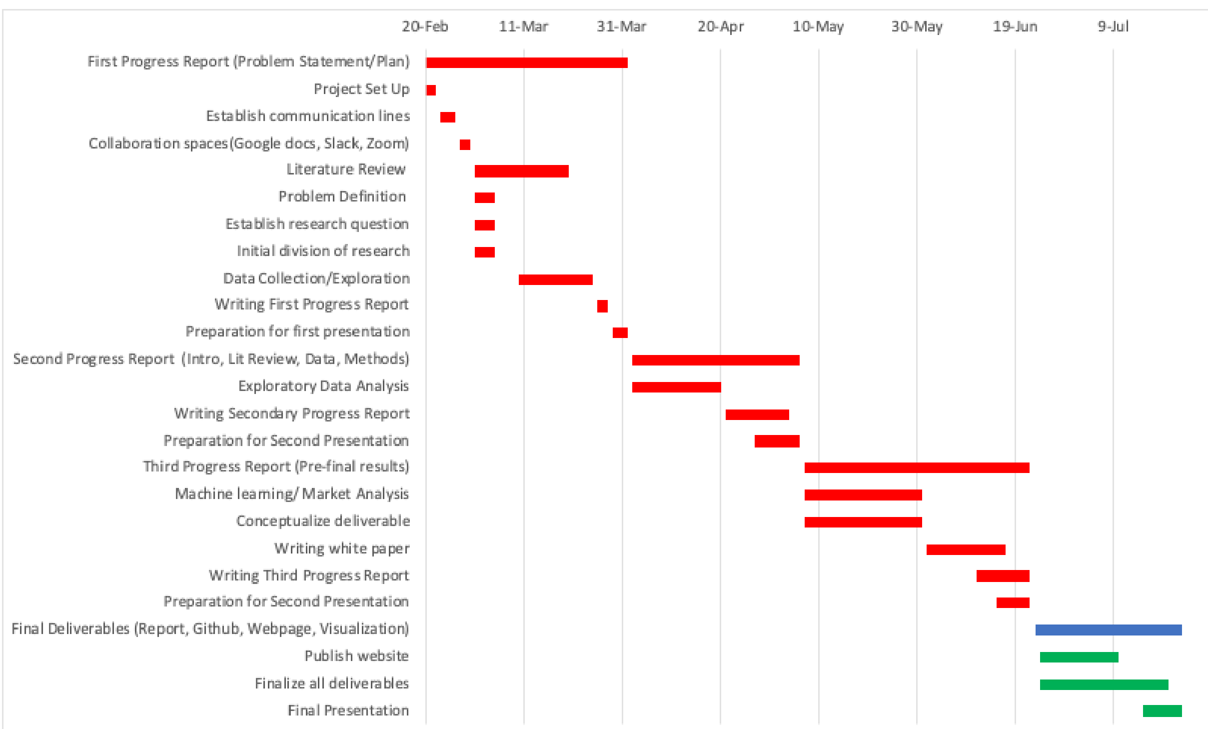

Figure 12: Updated Gantt Chart for Progress Report 3