nyconnect

The competitiveness of the telecommunications industry in New York City

---

Final Report

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. Understanding this industry is important for the renewal of franchise agreements between NYC and internet providers in July 2020, which will have a lasting impact on the long-term economic benefits related to broadband internet access. 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. Tweets involving telecom were mined to analyze New Yorkers' sentiments towards ISPs and the broader market, thereby reflecting the status quo and guiding the relevant policy interventions to address the raised concerns. 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. Furthermore, NYC's government should enhance connectivity amongst less well-connected demographic groups through means-tested subsidies.

Introduction

Problem Statement and Research Outcomes

The team's goals are multifold. The first is to explain the structure of urban broadband markets, and the supply and demand factors influencing availability of broadband internet and subscription rates. The second is 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 final goal is to produce a policy white paper for DoITT replete with policy suggestions to increase broadband subscription rates.

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, about 18% of residents do not have any means to connect to the internet (NYC CTO, 2020). Even residents with wired home broadband service perceive the service as extremely costly (Talbot *et al.,* 2018). Unevenness in 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 due to the structure of urban telecommunication markets, greater competition does not necessarily result in lower prices and maximum consumer benefit. Telecommunications infrastructure incurs high fixed capital costs subject to rapid technological change. Only big firms can achieve economies of scale. Meanwhile, smaller companies can neither surmount the barriers to entry nor enjoy the economies of scale. Hence, measuring industry concentration and competition using the Herfindahl-Hirschman Index (HHI) may not capture the market's latent efficiencies 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 help New Yorkers become potential paying customers 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 towards a broadband subscription would help connect those who cited price as an impediment to any internet service (Horrigan, 2009). This would help areas where internet infrastructure may be underutilized. However, this subsidy should not be limited to wired connections. Lee (2017) found that people of lower incomes preferred and prioritized a wireless internet service over wired broadband service.

As a case study, Singapore's Infocomm Media Development Authority (IMDA) has implemented the Home Access scheme since 2014. 24,000 low-income households in the bottom decile receive subsidies of up to 80% and pay between 0.3% and 1.3% of their monthly income for wired broadband service (IMDA, 2020). Consumer choice and market competition is maintained as consumers get to choose their preferred ISPs.

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

Fundamentally, the methods employed in this project sought to identify spatial, demographic, and economic structures and variables that could approximate supply and demand within the spatial boundaries and context of NYC. Geoprocessing in Python and QGIS matched the datasets to spatial boundaries of NYC in terms of census blocks, tracts, and NTAs. Shapefiles were created and reprojected. This enabled visualization of the data as requested by DoITT.

KMeans clustering, which identifies groups of observations with similar values, was performed to identify any directly unobservable but 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.

A Bayesian Network was constructed to identify demographic factors that impact broadband subscription rates at the Census block level. Bayesian networks model causation between data features by representing conditional dependence with edges in a directed graph. Data used included demographic data, broadband subscription rates, and the number of broadband ISPs.

Brake and Atkinson's (2019) argument of spatial segmentation can only be tested by calculating the degree to which contiguous spatial units reflected similar values, and whether said observed patterns were randomized outcomes. This is known as local and global spatial autocorrelation. A queen contiguity-based spatial weights was obtained from the geospatial datasets. 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. 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. The two-stage least squares regression used introduced a spatially-lagged term and allowed the coefficients to vary by boroughs (Drukker, Egger & Prucha, 2013; 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 their accuracy.

Text mining of tweets provided insights into how the public's perception of current service changes over time. They serve as testing metrics for the desired states, from the consumers' perspective. Twitter analysis involves tweets retrieval, sentiment analysis and word cloud visualization. Tweets mentioning a provider and services between 2010 and 2020 were scraped from webpages. Each tweet's sentiment was evaluated on a scale of 0 (negative) to 1 (positive) using a Word2vec-LSTM model, and then was aggregated by year to display the time series trend. A word cloud plot showcased keywords 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. Clusters of spatially contiguous tracts that share a similar level of infrastructural provision in terms of price are found across the whole of Staten Island, Manhattan south of Central Park, Washington Heights, downtown Brooklyn, Williamsburg, and most of Queens. In terms of ISPs, the patterns are similar, with the addition of the Bronx and neighborhoods up to the northern end of Central Park. At a confidence interval of 95%, these patterns seen in Figure 15 are not randomized outcomes.

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, as shown in Figure 8, found that broadband subscription rates are impacted more by the percentage of residents in a Census block that make less than $50,000 per year than the other factors. Additionally, the network found that Census blocks with more white residents tend to have access to more ISPs and more people who make more than $150,000 per year. These results demonstrate that adoption of broadband internet is impeded more by socioeconomic disparities than access to better internet service providers.

This observation is further supported by regressing socioeconomic and infrastructural variables against broadband subscription rates. The regression models also found that income levels were significant variables. 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 1. The less elastic demand curve suggests people are less responsive to an increase of internet providers. Hence, a change in the number of ISPs may not lead to a significant alteration in the consumer's expectation of broadband price. The intersection of these two curves is considered to be the equilibrium of the broadband market at a price of $70/month. There are small movements between 2017 and 2018, but the price decrease is insignificant. The optimal number of providers city-wide is 5 which was met in 28.22% of all census blocks. 58.58% of blocks have 4. However, given differing conditions in each borough, the optimal number differs by borough as shown in Figure 13.

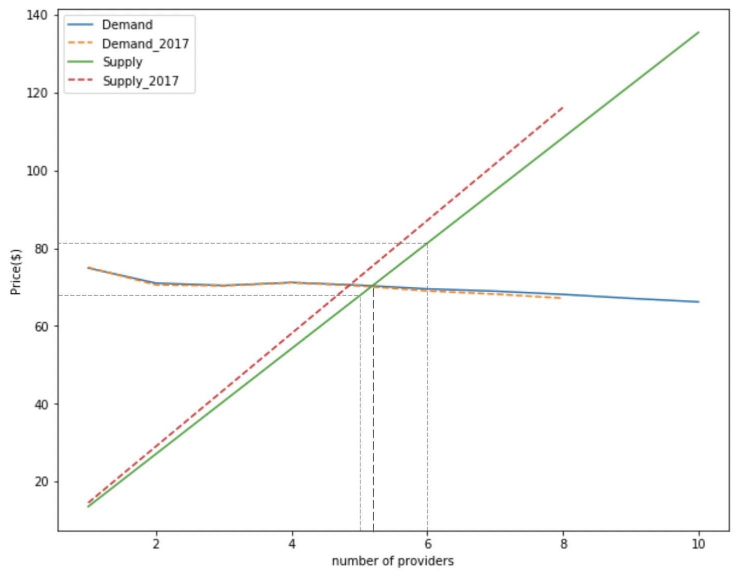

Figure 1: Demand curve and supply curve of broadband services

Figure 9 in the Appendix shows over half of the mined tweets were complaints. This might indicate that in most cases, users only tweet when there is a service problem. As shown in Figure 10, sentiment gradually improved between 2013 and 2018 while there were fewer tweets (Figure 11) between 2015 and 2018. This could mean that between 2015 and 2018, consumers were more satisfied, and fewer complaints were tweeted. However, the public sentiment trended downward in recent two years, suggesting service can be improved. Verizon, AT&T, T-Mobile, Sprint accounted 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. Crucially, providers of subsidized service under the federal Lifeline program were not mentioned. Considering the tweets with extremely negative sentiment as shown in Table 3, mobile service was most commonly referenced. Other significant keywords show that most complaints were network performance-related such as ease of connectivity, accessibility, and speed. Thus, regulators should not only focus on the quantity of providers in NYC but also service quality.

Implications: Policy Recommendations

At the federal level, the FCC's Lifeline program provides up to $9.25 a month to households who earn up to 135% of the federal poverty line. As qualifying providers in NYC are barely tweeted about, it is unclear if households are aware of such a scheme. Besides, a $9.25 monthly subsidy is insufficient for broadband plans that cost as much as $70 a month. Of the six authorized Lifeline providers in NYC, Verizon is the only provider with wired broadband services.

NYC can provide a supplementary top-up to low-income households modeled on Singapore's Home Access program as explained in our literature review while including measures of service quality in their 2020 franchise agreements. The 2018 ACS survey revealed that 46% of NYC households without an internet subscription, or approximately 231,000 households, earn less than US$20,000 annually. The NYC government can negotiate with ISPs to provide broadband plans at wholesale prices below market rates. The government can then administer the subsidies for households earning less than US$20,000 annually and sign up. Prices can be means-tested and capped at 0.5% to 2% of monthly income. The total cost is dependent on rates negotiated by the government and the total number of households that sign up.

Limitations

A key limitation faced was the lack of updated open information. For example, the latest American Community Survey and FCC data available were 2018 and June 2019 respectively.

Second, other data such as the number of subscribers per ISP, and updated pricing information were proprietary. This hampered the team's ability to provide current analyses as economic and demographic conditions have changed significantly since then, and proxies had to be estimated.

Third, the FCC 477 form, which was a cornerstone of our analysis, relies on a flawed data collection methodology. It is self-reported by ISPs. A census block is "served" even if just one building in a block has access. An overestimation of infrastructural provision is possible.

Conclusions

This project is among the first to incorporate spatial regression and autocorrelation into an analysis of how to narrow the digital divide within urban telecoms markets. Based on current demographic and socio-economic conditions, current infrastructural provision in terms of the number of ISPs, is already close to the optimal number of 5. As household incomes are the most important factors determining household internet connectivity, a subsidy program targeting low-income households will be more suited to narrowing the digital divide in NYC.

These findings thus serve as a primer for future research into how broadband can be made more accessible for all. Future projects can look into other major urban markets such as Los Angeles or the greater New York City metropolitan area including Long Island and New Jersey.

Team Collaboration Statement and Research Plan

Wesley took the lead in geospatial visualizations and analyses, as well as KMeans clustering. Yichen created the tweet-based consumer sentiment analysis model. Ziyu was in charge of market and economic analyses. Erik built the website showcasing our final results, and the

Bayes Network analysis. The team worked together on policy suggestions, and a white paper that has been sent to DoITT. The team was able to adhere to all deadlines and deliverables agreed upon at the start of this project.

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.

Deng, Y. (2018). Estimation for the spatial autoregressive threshold model. *Economics letters*, *171*, 172-175.

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}

Drukker, D. M., Egger, P., & Prucha, I. R. (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews*, *32*(5-6), 686-733.

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/.

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.

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.

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

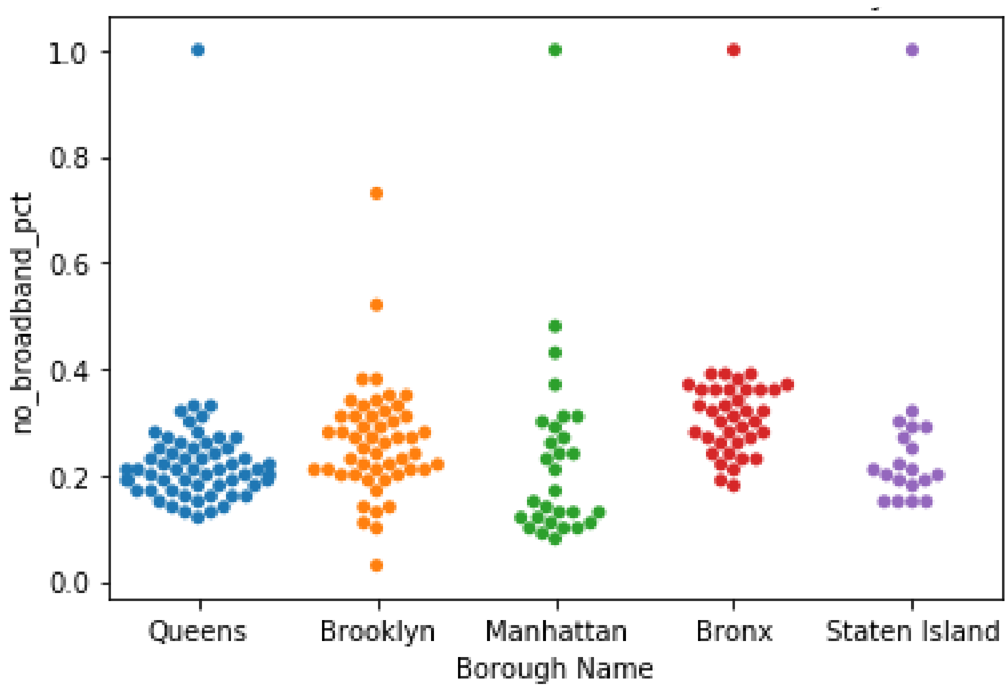

Figure 3: Households (%) with no form of broadband connection in each NTA by borough

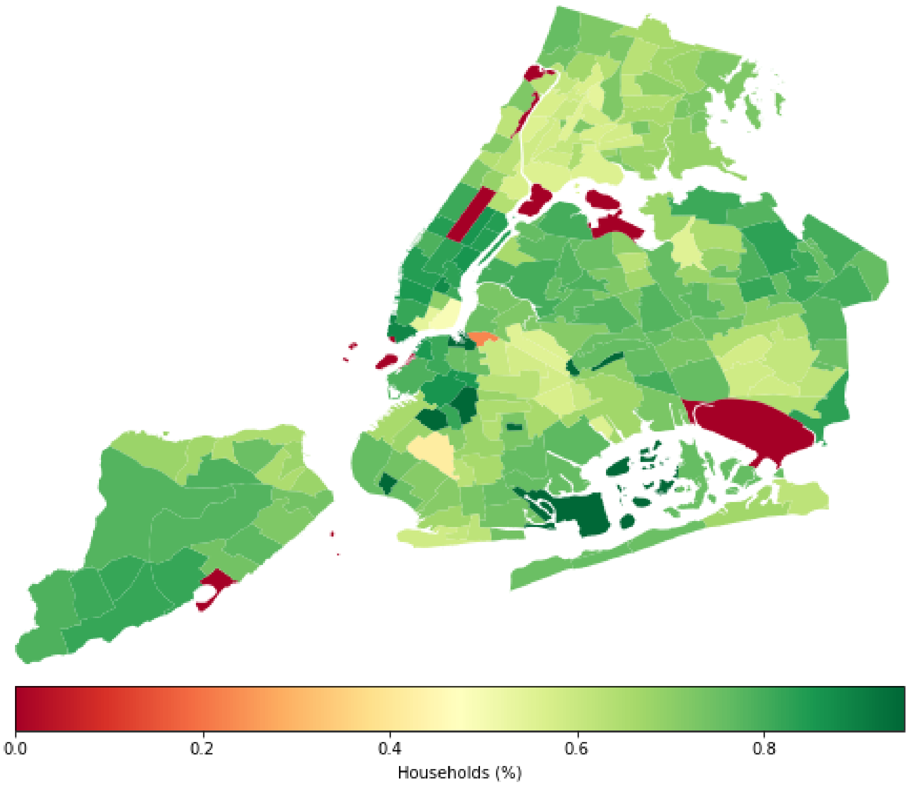

Figure 4: Geographical distribution of households in each NTA without a broadband subscription

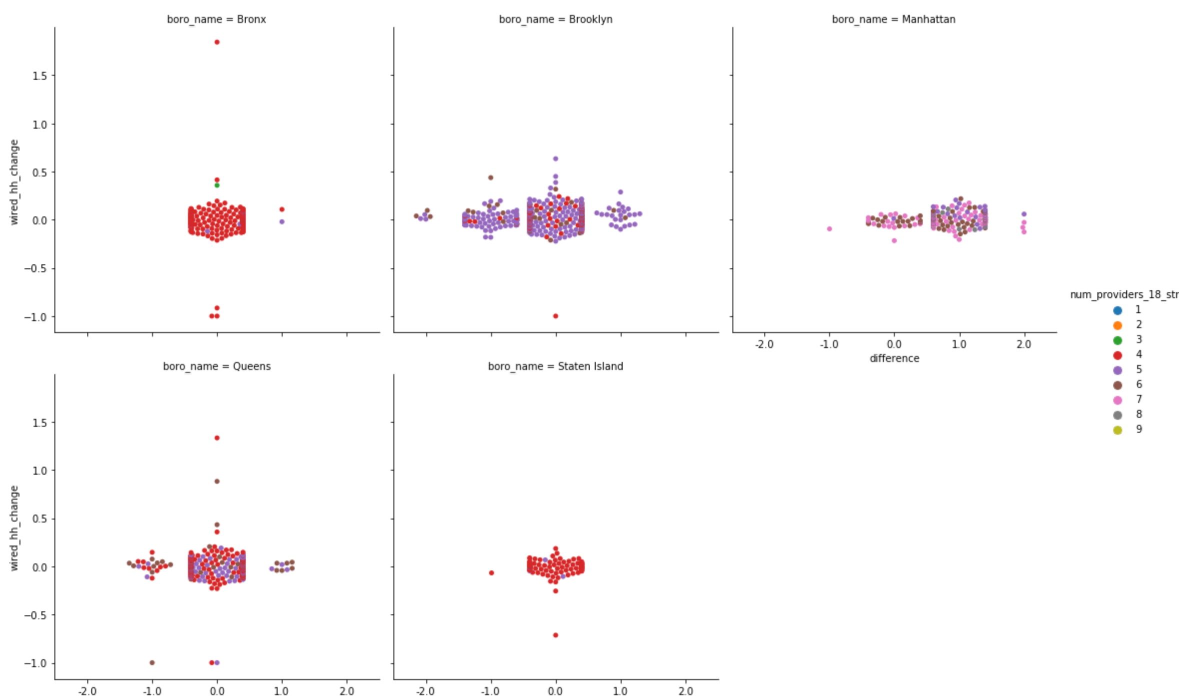

Figure 5: Swarm plot of changes in ISPs and household wired broadband subscription rate by borough

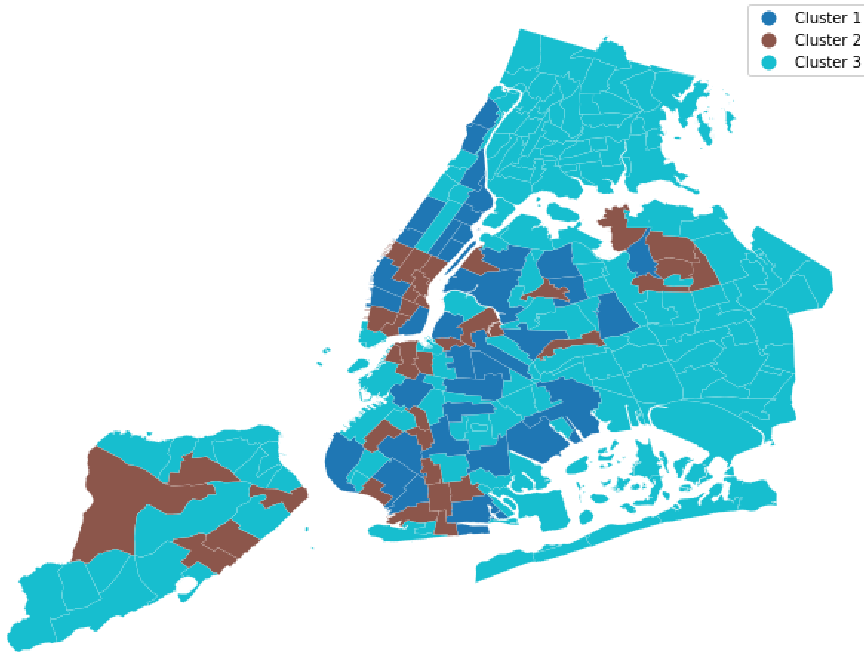

Figure 6: KMeans clustering of census tracts based on business patterns

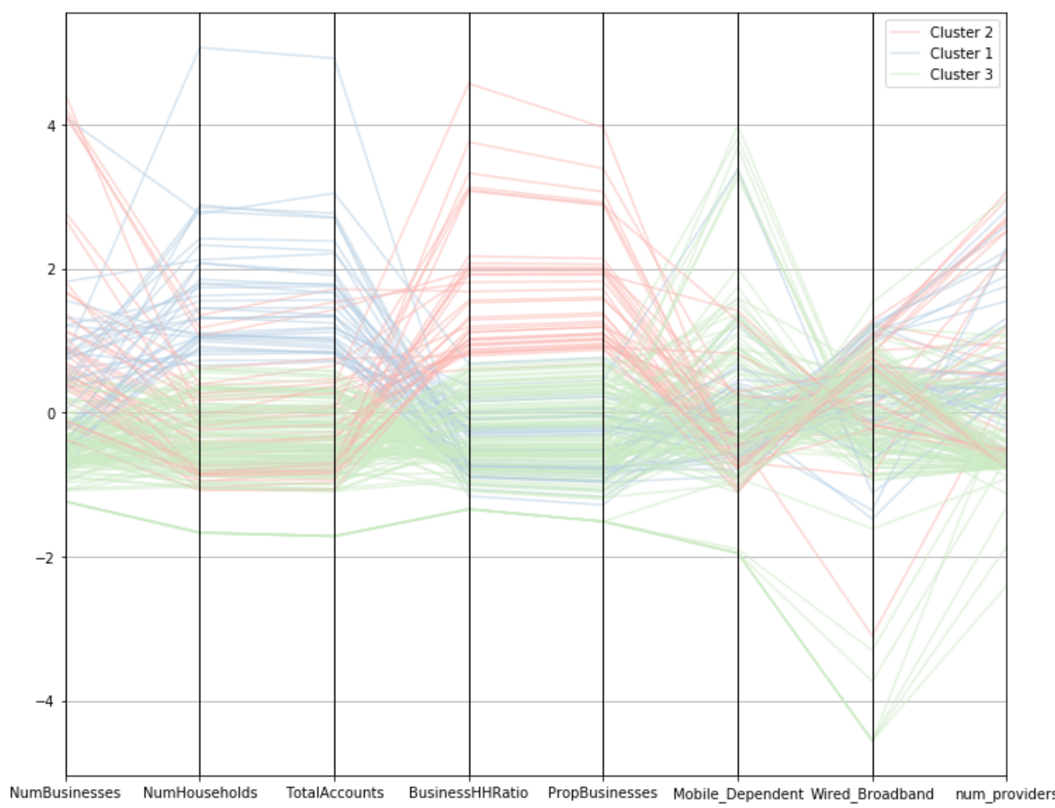

Figure 7: Parallel coordinates plot of KMeans clustering based on business patterns (Figure 6)

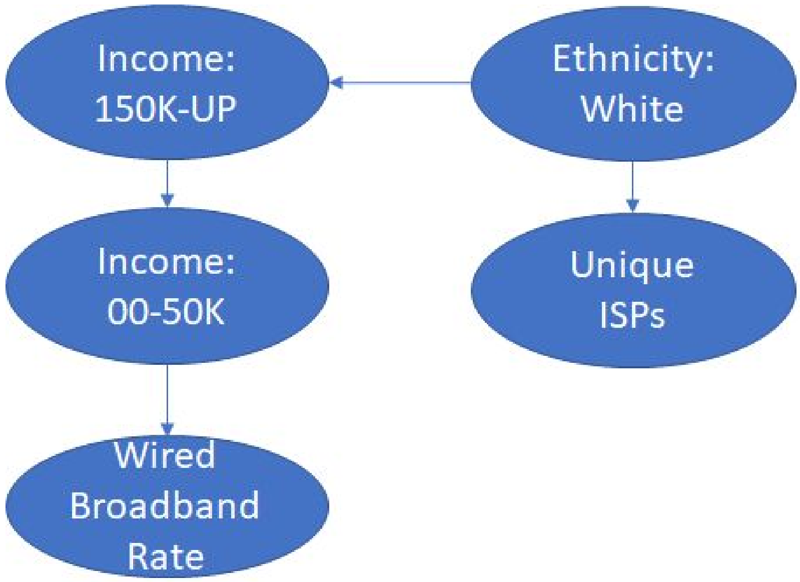

Figure 8: Bayesian Network

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

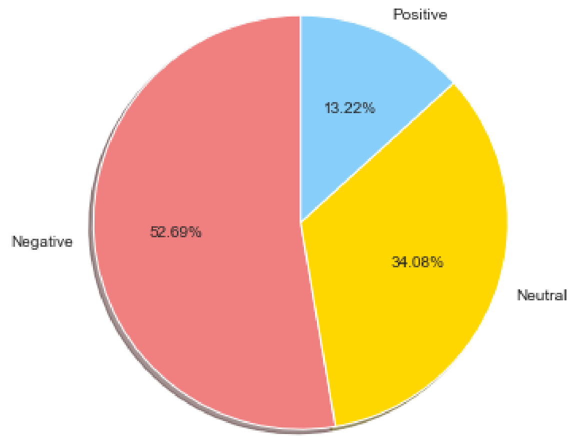

Figure 9: The proportion of the three sentiments

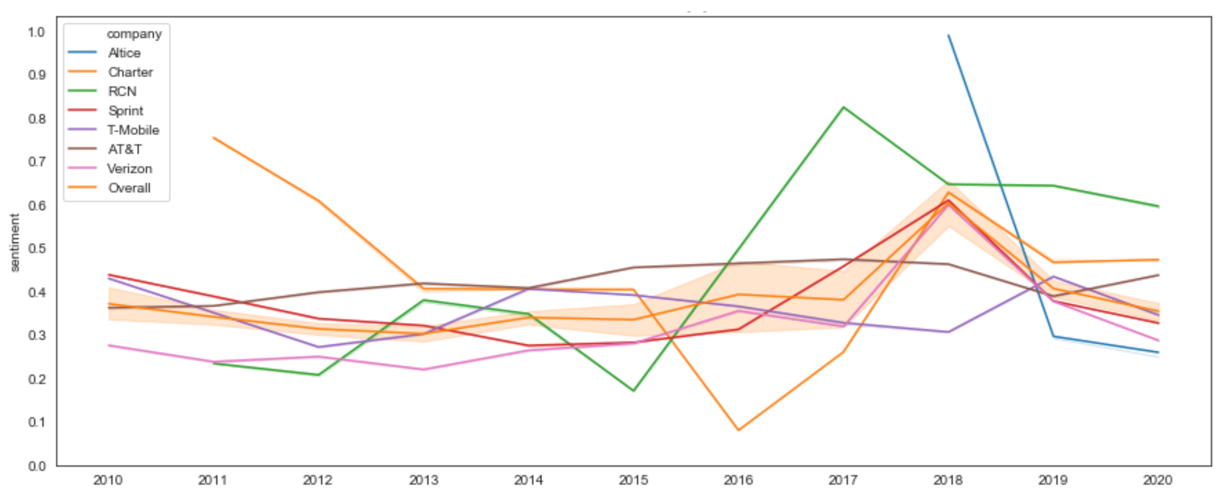

Figure 10: tweet sentiment by year and provider

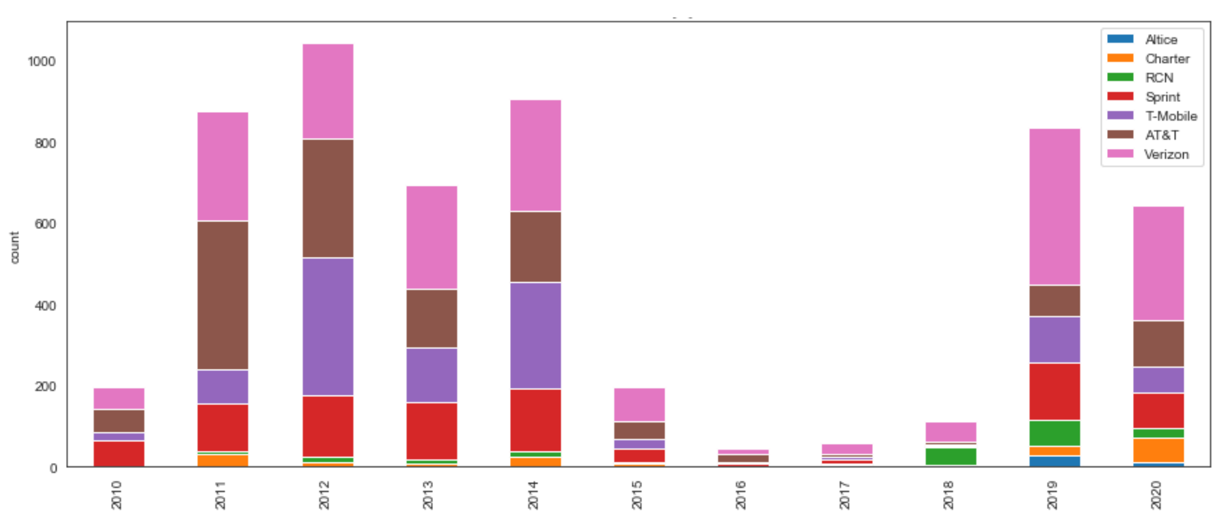

Figure 11: tweet count by year and provider

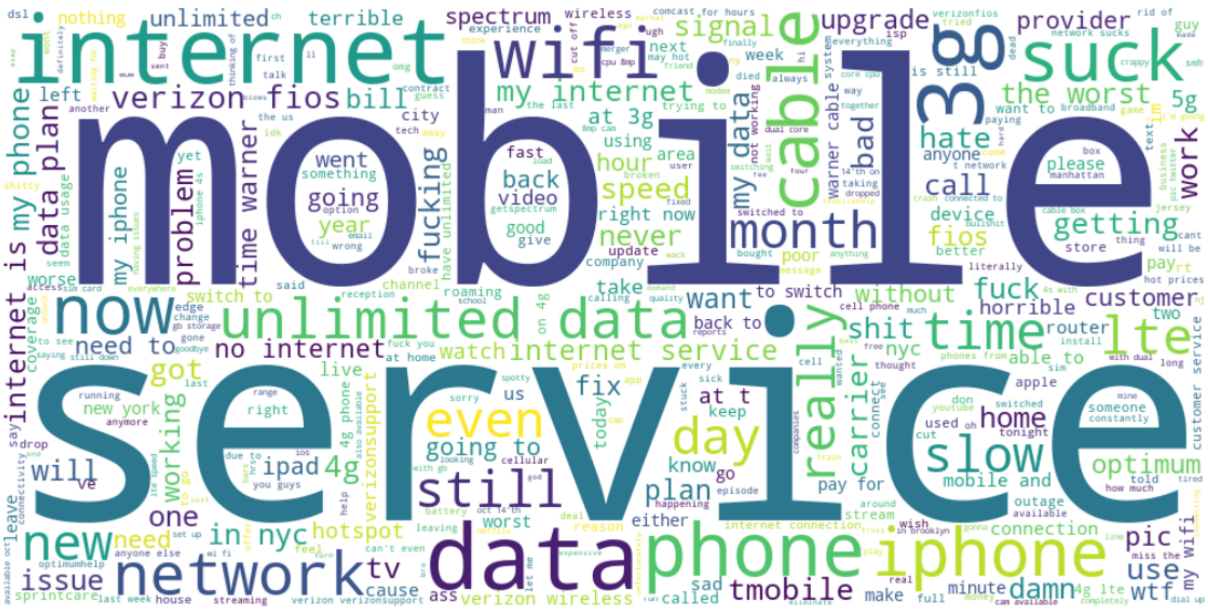

Figure 12: Word cloud for tweets with very negative sentiment

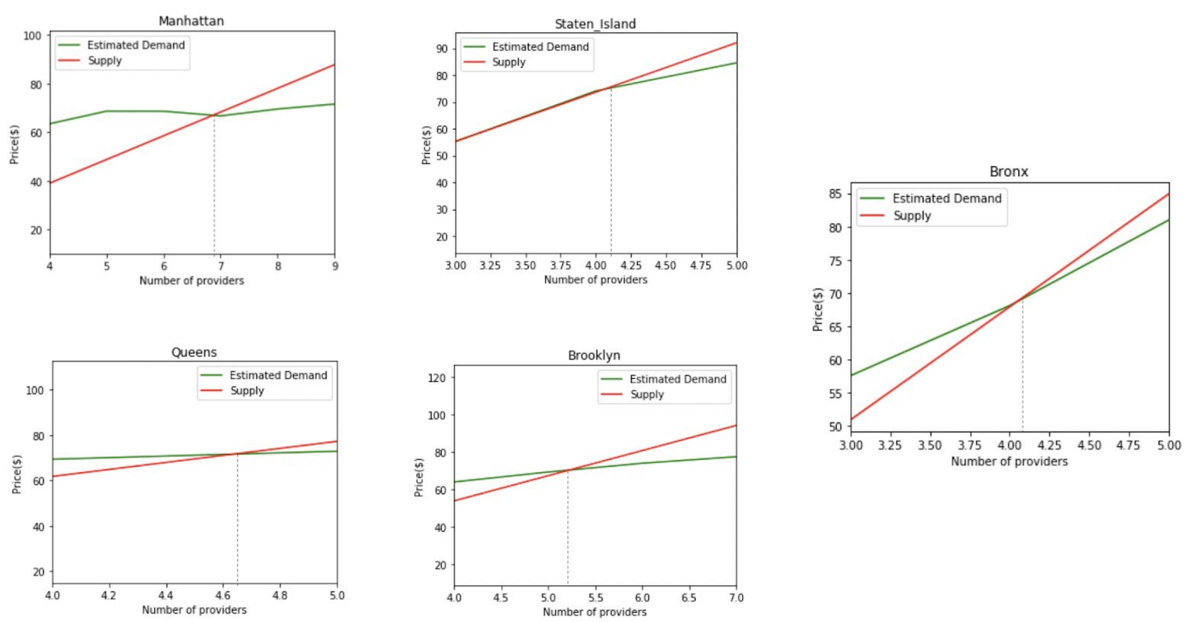

Figure 13: The optimal number of providers in each borough

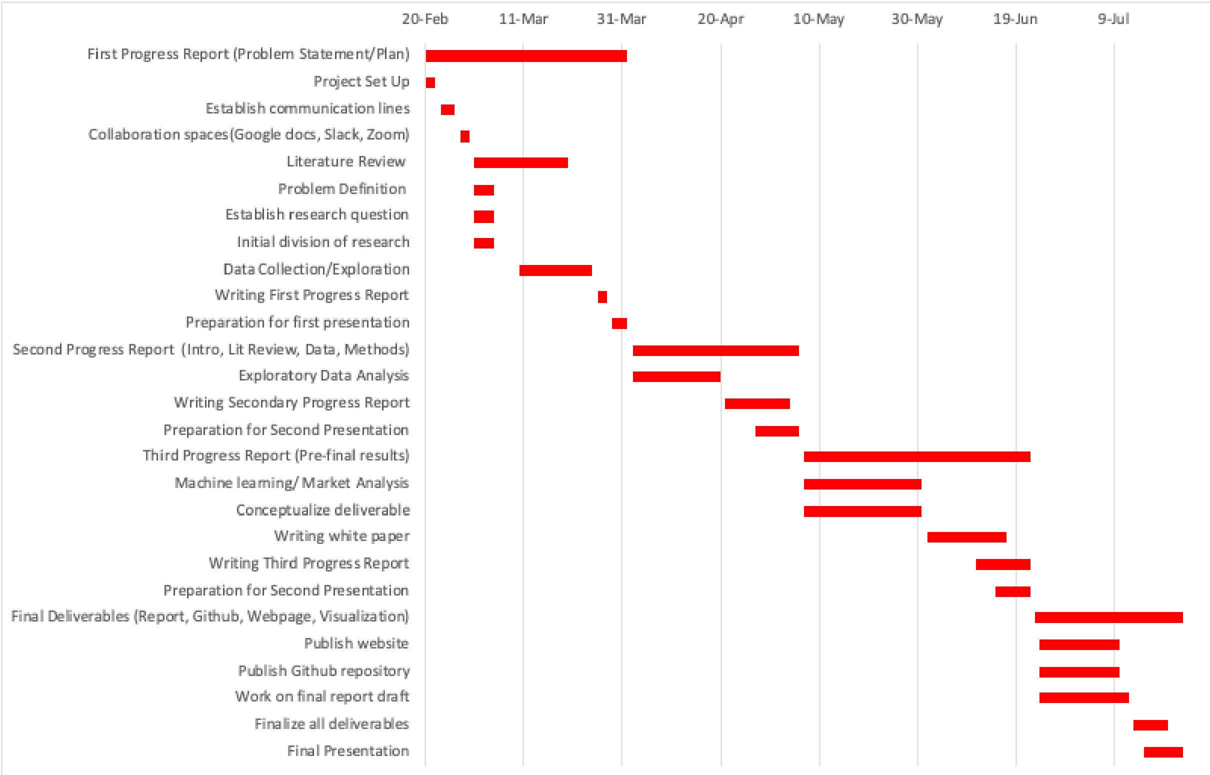

Figure 14: Gantt Chart for the whole project

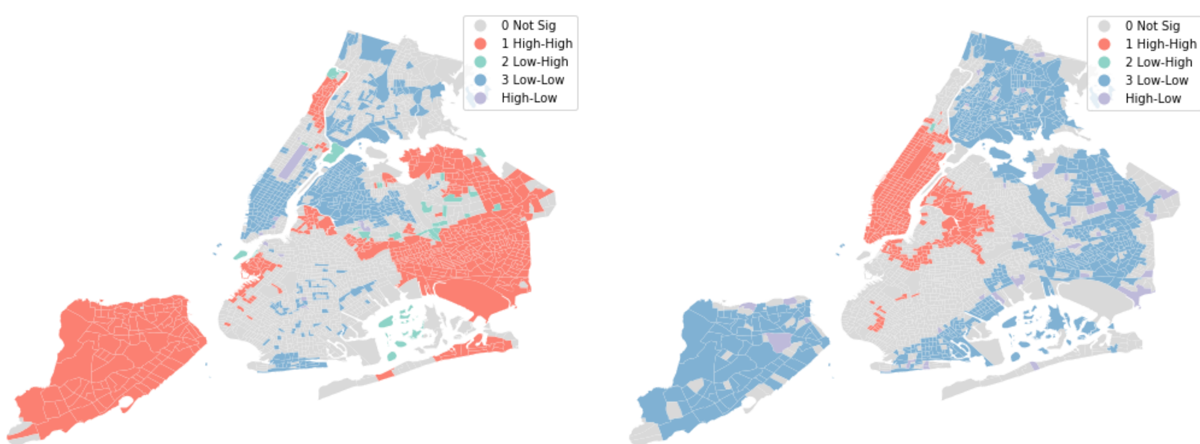

Figure 15: Spatial autocorrelation maps for subscription prices (left) and number of ISPs (right)