Project Proposal

Abstract

311 is a phone number used in North American communities that allows its citizens to request non-emergency municipal services (McEntire, 2021). The objective of this project is to perform exploratory analysis of the 311 data of New York City to improve the quality of life of the residents by providing insights to the local government so that it may allocate its time and resources more efficiently and equitably. The percent of cost savings brought on by the implementation of the team's recommendations will also be estimated.

Current Practices and Team Approach

New York City has made the data available to public and it encourages the site visitors to share their work by featuring the projects on their gallery. There is no indication if any insights are derived and implemented from the submissions. This may be attributed to a lack of detailed analysis in the submissions, for example, a similar project was found in the gallery that provided nothing more than a summary of the data.

Research teams all over North America have critiqued the use of 311 data for policymaking. Other teams have used the data to generate insights in respective, albeit limited, contexts. Attempts have been made to establish correlations between one complaint category and other co-occurrences (Maureen H. Murray, 2018). The approach may be appealing, since strong correlations may make some categories redundant, thereby reducing the size of data collected. However, it may not be possible to establish a correlation between many seemingly related issues. Other studies, although narrow in scope, provide a great framework for dealing with data with social implications (Li, 2020) (Tong, 2021). Books have also been written on increasing citizen engagement towards 311 calls to further improve quality of life, however, they provide little insight on actual analysis (Gavin Newsom, 2014).

The team will begin with summarizing the most prevalent issues in the neighbourhoods and then establish a timeline of said issues. Analysis of 911 calls (Cramer, 2011) may provide an overarching approach to analysis, however, the size of dataset for this study is exponentially larger. After cleaning the data and performing initial analysis, the team will attempt to investigate if budget allocated to departments had any impact on closure of issues. Similar studies on customer complaints for financial companies show the importance of techniques like regression to improve service levels, however, they fall short of categorizing their complaints effectively (Ayres, 2013). The call center industry provides another treasure trove of insights that can be used for guidance. Research shows a spike in calls made to call centers during the pandemic and analysing how those centres process increased volumes can have a direct impact on resident engagement (Andersen, 2021). Moreover, the team will compare New York City's performance to that of another city. This analysis will not only attach a financial component to the grievances of the residents but also put the city's performance in perspective when compared against another similar city.

The findings of this analysis will be relevant to New York City government, the residents, as well as budget steering committees. The study will also be relevant to different departments as it will allow them to track their performance over the years. One may also use New York City as a pilot to assess the applicability of 311 data in other regions.

Project Proposal

Impact

The study has the potential to impact how a city allocates its resources. Upon briefly skimming the raw data, it was apparent that not all requests were closed. A study like this can allow the city to focus on the backlog and use it as a key performance indicator for respective departments. Given the recent and overwhelming increase in government spending due to the pandemic, a gradual reduction in funding is expected. In that scenario, it will be crucial for governments and city departments to be extremely judicious while prioritizing areas of concerns. The allocation of resources will influence the quality of life in the city and if handled appropriately, improve the standard of living.

Risks

Like any data analysis project, the accuracy of insights is contingent upon the cleanliness of the given dataset. Since the records are created by humans, and 32 of the 41 fields are plain text fields, it would be unreasonable to assume that the raw data is devoid of any user entry errors (Ma, 2020). Also, the data includes no information on the caller, making it impossible to differentiate between an individual who may file identical complaints repeatedly and another individual who may only file a complaint once (Ariel White, 2016). There is no information on the administrative costs associated with vetting and updating the data, however, there has been research on how 311 operators can create value for predictive analysis. Although insightful, those studies focus more on organizational agility than analytics (Chatfield, 2018).

Since there is no standard template for logging 311 calls, there is an inherent risk of making faulty deductions when comparing cities (Lingjing Wang, 2017). To mitigate this risk, the team will assess each field and compare like-for-like parameters (Nalchiga, 2017).

Most importantly, the data does not have any information on the demographic of the callers. The risk of using this dataset in isolation to dictate resource allocation may lead to improvement of services in certain areas while service levels in other areas dwindle. This has the potential to negatively impact lower income areas, and areas with residents who do not speak English as their native language (Constantine E. Kontokosta, 2021).

Cost Factors

Since the dataset exceeds 10 GB and has over 26 million records, it will be imprudent to perform analysis locally on a machine. Therefore, a cloud service like Amazon Web Services (AWS) may prove to be more effective. AWS has many products that can aid in analysis of this scale, namely EC2, Glacier, and S3 (Taneja, 2020). For implementation, an instance of Amazon EC2 must be initiated. As per AWS' quick estimate guide, one instance that uses Windows Server, with 4 CPUs, a memory of 64 GB and general-purpose storage of 64 GB will incur an annual cost of US$483.09.

Timeline and Checkpoints

Since all team members are part time students, the entire project will take anywhere from 6 to 8 weeks for completion. All team members will make similar amount of effort for project completion.

Item	Checkpoint	Duration
1.	Data cleansing	14 days
2.	Data consolidation, focus on most prevalent grievances	14 days
3.	Focus on insights, proper visualization techniques	14 days-21 days
4.	Final report	7 days

Project Proposal

References

Andersen, D. (2021, September 10). *How to Analyze Call Center Data to Improve Efficiency*. Retrieved from Invoca: https://www.invoca.com/blog/how-to-analyze-call-center-data-to-improve-efficiency

Ariel White, K.-S. T. (2016, July 5). *The Promises and Pitfalls of 311 Data.* Retrieved from Ariel White: https://arwhite.mit.edu/sites/default/files/documents/promises311data_RR2_unblinded_June2016.pdf

Ayres, I. L. (2013). Skeletons in the Database: An Early Analysis of the CFPB's Consumer Complaints. *SSRN Electronic Journal*.

Chatfield, A. T. (2018). Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services. *Government Information Quarterly*, 336-347.

Constantine E. Kontokosta, B. H. (2021). Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. *Elsevier*.

Cramer, D. &. (2011). Predicting 911 Calls Using Spatial Analysis. *Studies in Computational Intelligence*, 15-26.

Gavin Newsom, L. D. (2014). *Citizenville: How to Take the Town Square Digital and Reinvent Government.* Penguin Books.

Li, Y. H. (2020). 311 service requests as indicators of neighborhood distress and opioid use disorder. *Scientific Reports*.

Lingjing Wang, C. Q. (2017, October 17). Structure of 311 service requests as a signature of urban location. *Plos One*. Retrieved from Plos One: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0186314

Ma, M. P. (2020). Data Sets, Modeling, and Decision Making in Smart Cities: A Survey. *ACM Digital Library*, 1-28.

Maureen H. Murray, R. F. (2018). Public Complaints Reflect Rat Relative Abundance Across Diverse Urban Neighborhoods. *Frontiers in Ecology and Evolution*.

McEntire, K. (2021, April 29). *What is 311 and When Should I Use it?* Retrieved from Safewise: https://www.safewise.com/blog/what-is-311/

Nalchiga, S. &. (2017). Achieving Interoperability of Smart City Data: An Analysis of 311 Data. *Journal of Smart Cities*.

NYC Open Data. (2021, October 15). *NYC Open Data*. Retrieved from NYC Open Data: https://opendata.cityofnewyork.us/data/

NYC Open Data. (2021, October 15). *NYC Open Data - Projects*. Retrieved from NYC Open Data: https://opendata.cityofnewyork.us/projects/

Project Proposal

Oleh Dubno, D. A. (2021, October 15). *311 Data & Life in NYC*. Retrieved from NYC Open Data: https://opendata.cityofnewyork.us/projects/311-data-life-in-nyc/

Taneja, M. (2020, September 28). *What is AWS? Why Every Data Science Professional Should Learn Amazon Web Services*. Retrieved from Analytics Vidya: https://www.analyticsvidhya.com/blog/2020/09/what-is-aws-amazon-web-services-data-science/

Tong, H. &. (2021). Relationships between noise complaints and socio-economic factors in England. Sustainable Cities and Society. *Science Direct*, 65.