NYU | Center for Urban Science + Progress

Urban Science Intensive
Final Technical Report

Hearing Noise Complaints

Data-Driven Optimization of Rapid Respond to Urban Noise Complaints

Qinyu Goh (qg412)
Zoe Martiniak (zem232)
Sam Ovenshine (sgo230)
Siddhanth Shetty (sds695)
Sung Hoon Yang (shy256)

NYC 311

Abstract

In New York City, the Department of Environmental Protection (DEP) handles outdoor noise complaints from sources ranging from construction activity to the jingle of ice cream trucks. Since 2010, the growing volume of noise complaints has increased the agency's response times and hindered enforcement of the city's noise code. This capstone project provides a data-driven approach to optimize the DEP's processes to better address noise complaints. To accomplish this, we deployed two machine learning models: an LSTM neural network to predict spatial and temporal complaint volume, and a random forest classifier model to predict complaint enforceability. Model features were 311 complaints, socioeconomics, PLUTO land use, weather conditions, and construction variances. Based on sponsor feedback, we opted for the random forest classifier model and to optimize its performance for recall. In evaluation, the model precision was 9.3% and recall was 91.1%. The model's enforceability predictions were incorporated into an interactive data visualization for DEP inspectors to identify clusters of unresolved noise complaints with a high likelihood of enforceability. The implications of our work are to improve noise code enforcement and reduce the DEP backlog.

Acknowledgements

The team would like to express their sincerest gratitude to the mentors, Charlie Mydlarz, Mark Cartwright, and Vincent Lostanlen for their guidance and patience throughout this entire project.

Thank you for inculcating in us the spirit to never be settled with our analysis and for always challenging us to be better versions of ourselves!

Many of the ideas and steps adopted in this project would not have been possible without discussions with fellow SONYC researchers (Carlos Bautista, Fabio Miranda and Joao Rulff). We are inspired by their ongoing work.

The team would also like to give a special shout-out to our sponsors, the New York City Department of Environmental Protection, for their unwavering support and candid feedback. We look back to each sponsor meeting fondly as they were always filled with memorable inspection and enforcement stories. Thank you, Oscar Gonzalez, for tirelessly working to get us relevant internal information and gathering relevant personnel, as well as for being our main point of contact!

Lastly, we would like to thank our family, loved ones, and friends for their love and support throughout our lives and especially during this tough but rewarding master candidacy.

Table of Contents

1. Introduction

Noise has become synonymous with New York City, long maintaining the status of the most common 311 complaint type (Wellington, 2015; Cook, 2017). But while the volume of noise complaints has been increasing, the response rate to such complaints has not kept pace (DiNapoli, 2018). A 2017 report by the state comptroller found that the noise management system in New York City leaves much to be desired. There is no immediate relief from noise from lodging a complaint with 311, as the two enforcement agencies NYPD and DEP both reported that less than 30% of noise complaints could be verified and enforced. For DEP in particular, of the 230,000 investigations made from 2010 to 2015 in response to complaints, the agency was able to confirm noise for only 3%, and most noise complaints were closed without a violation (Office of the New York State Comptroller, 2017).

With limited manpower to combat the increasing number of complaints, the DEP is faced with mounting pressure to improve its effectiveness at identifying and mitigating noise complaints. The primary motivation of this project is to provide a technological approach to improve the current state of affairs. More specifically, the main goal is to showcase—visually, on a map in real time—areas where there are higher probabilities of verifying noise complaints so resources can be directed to such areas at the discretion of the inspector coordinator.

We formulated our problem statement based on the two goals stated by our sponsor: to reduce the time to resolution for noise complaints and improve the enforcement rate of the noise code. We believe that achieving the first goal will help achieve the second, since responders will be more likely to arrive while noise is still occurring.

1.1 Literature Review

There are two main streams of literature related to the detection and modeling of urban noise. The first stream focuses on the utilization of sensors and the creation of noise models to represent the spatial extent of noise over time. One such study from 2017 used 35 sensors to measure noise in selected locations, focusing on how these data can aid in New York City DEP's noise enforcement efforts (Mydlarz, Shamoon and Bello, 2017). The research found that while the sensors were helpful in acting as evidence of noise violation in real time, their effectiveness was limited by false positives and the inability to accurately identify the source of the noise.

The second stream of literature instead focuses on the utilization of an array of datasets such as 311 complaints and road network data to gauge noise over space and time. A study conducted by Zheng et al. in 2014 made use of 311 complaints, social media posts, road network data, and other points of interest to model noise in New York City. The conclusion was that noise pollution varies depending on time and days of the week as citizens' tolerance of noise changes.

Additionally, research has helped identify features of interest that are correlated with noise complaints. For instance, the frequency of 311 noise complaints has been found to be positively correlated with Lower Manhattan's high income and education levels (Wang, Qian, Kats, Kontokosta & Sobolevsky, 2017). Noise complaints have also been found to be associated with major construction and after-hours work (Hong, Kim & Widener, 2019). Land use has long been established to affect noise tolerance (Galloway & Bishop, 1970). As a result, demographics, construction related data and land use-related data will be considered in this project.

2. Data

Our primary data source is the 311 complaint dataset available in the NYC Open Data portal. Each complaint record is tied to two natural dimensions: space and time. The complaint counts over time are localized and many complaints take place in temporal clusters, with counts decaying from the onset. Complaint intake is highly seasonal, peaking in late spring and mid-fall. The average time to resolution for a complaint is 100 hours (4 days), which exhibits little seasonal variation. At any given point, the DEP has between 300 and 900 open complaints (**Figure A**). Complaints peak on Saturday and at the start of workdays (**Figure B**).

Other datasets used by our models include demographic features (e.g., population and racial makeup) from Census TIGER shapefiles, land use features (e.g., zoning district and land value assessment) from NYC MapPLUTO, and weather features (e.g., average temperature and precipitation) from NOAA.

DEP stated that one of the most important features for deciding whether to prioritize a complaint was the presence of an after-hours variance (AHV) near the complaint site. AHV permits are granted by the Department of Buildings (DOB) and allow construction work outside the standard hours of the New York City Noise Code. Because no citywide report on AHVs is available, we wrote a web scraper to obtain information about all recently issued AHVs (see **Supplementary Materials: Web Scrapers**).

The combination of all data sources produced 660 total features, which were reduced as discussed below.

3. Methodology

There were two principal machine learning models undertaken for this project: an LSTM neural network to predict aggregated noise complaints and a random forest classifier to predict the enforceability for each complaint. Each is described below, along with the decision to opt for the latter as the sponsor deliverable.

A variety of neural networks were built with the PyTorch library and were trained for evaluation. The result is returned with an under-complete autoencoder architecture that employs LSTM encoder with two subsequent layers of DenseNet to reduce dimensionality and then to decode to the output dimensionality (Li, J., Luong, M. T., & Jurafsky, D., 2015). With the aim of predicting the daily volume of complaints for 29 neighborhood tabulation areas in Manhattan, the model was provided with 49 spatiotemporal features, including weather, after-hour variances, complaint time to resolution, and 28 autoregressive features. We used the mean squared error (MSE) loss function to compute the stepwise loss. The model was trained with data from 2011 to mid-2016 and evaluated on data from mid-2016 to 2017 (75:25). Because the MSE loss function was used, a random, concentrated onset of noise complaints severely penalizes the performance of the models. In order to mitigate a severe penalization on a single random onset, as well as to reformulate the question such that the prediction can serve as an actionable insight to DEP, the Softmax function was added to the top of the network to convert predictions to pseudo probability.

After developing the neural network, we presented two modeling options to DEP: using the neural network to predict the volume of future complaints by neighborhood or using a classification model to assess the enforceability of each new complaint. The sponsor weighed both options and opted for the classification model due to its ability to be leveraged in real time, its potential to reduce response time, and its greater usefulness to everyday inspector operations.

A random forest classifier was built to classify new complaints as enforceable or unenforceable. The model was trained on closed complaints with actual enforcement outcomes. The random forest was an appropriate choice given its ease with the number of features in our data (660) and for being able to distill features down to the most important. It also provides an efficient way of including additional features with minimal preprocessing, as input data do not need to be rescaled or transformed and can be in binary, categorical or numerical form.

We performed training on 2016 noise complaints since historical AHV data was available only through then. Some feature engineering was conducted to transform data inputs into the appropriate format for model consumption. For example, categorical resolution descriptions were mapped to numeric enforceability scores (see Figure C for full list). In order to link AHV permits with complaints, an assumption was made that if an open AHV existed at the time of the complaint and within 200 meters of the complaint (an approximation of how far construction noise can travel), the two were related.

One difficult aspect of noise complaint data is that DEP enforces minor violations like dog barking, along with more serious violations like illegal after-hours construction. Two different data subsets were trained on, one with all complaint types and one with construction complaints. Additionally, in order to combat the risk of overfitting, model were created with all 660 features and with the 10 most important features. This resulted in four variants of the model (Models A, B, C, and D in Figure H).

For each model variant, two training methods were used in order to handle the skew of enforceability in the dataset (out of 46,322 complaints, only 1,470 were actually enforced, or about 3.2%). The first training method used a standard train-test split of 0.77 to 0.33 on the entire dataset, while the second training method trained only on an equal subset of enforceable and less enforceable complaints with a train-test split of 0.77 to 0.33 and tested the trained model on a random subset of 1,000 rows.

4. Results

4.1 Neural Network

To assess model accuracy, we compared the neural network's predicted top 15 neighborhoods by daily complaints against the actual top 15 neighborhoods by daily complaints. We recorded an accuracy of 74%, correctly guessing an average of 11 neighborhoods from the daily top 15 (**Figure D)**. The model performed consistently better than the static ranking of neighborhoods by all-time complaint volume, which scored about 56% accuracy. Overall, the neural network provided a good estimate of the true standardized mean of complaint volume but failed to capture complaint variance, which is high due to the fast onset and decay of related noise issues (**Figure E**). These findings suggest that a neural network model can learn the baseline volume and periodicity of complaints within targeted spatial regions but struggles to predict specific complaint clusters and complaint peaks.

4.2 Random Forest Classifier

Random forest classifier performance for each variant and training method is summarized in **Figure H**. It is evident that the precision is consistently higher than recall in the training method 1, whereas the recall is high and the precision low when we test with the second training method on 1,000 records. This is understandable, as training method two has a greater proportion of enforceable complaints and therefore classifies more complaints as enforceable.

The models with the highest F score were initially selected as the final product for the sponsor, which were Models C and D using training method 1. However, in a later model review session, the sponsor requested an option with higher recall to optimize for detecting more enforceable complaints at the cost of pursuing more false positives.

Model B using training method 2 satisfied the sponsor's criteria with its high recall and F score and thus was selected for the final product. Model B.2 recall is 94.1%, with a precision of 9.9% and an F score of 0.179, which is five times better than chance in selecting complaints that are enforceable. That figure is derived from the division of the model's F-score (0.179) over the percentage of complaints that was actually enforced (3%).

To build an interactive tool for our inspectors to see the model enforceability probabilities, we leveraged Tableau Public, a free dashboarding tool that supports web embedding. The dashboard colors complaints based on enforceability and has filters chosen by DEP inspectors (see screenshots, **Figures F-G**).

5. Conclusion

Our project aimed to provide a data-driven improvement to DEP's current process for identifying and responding to noise complaints. We developed two machine learning models to do so: a neural network to forecast aggregated complaint volumes by neighborhood and a random forest classifier to predict the enforceability of a given noise complaint. DEP opted for the latter model, and it was the basis of the project deliverables.

Some limitations of our approach should be noted. First, due to DEP privacy regulations, we were unable to secure access to the real-time 311 complaint stream and had to rely instead on redacted NYC Open Data on a 1-day lag. Thus, we could not create a real-time dashboard or use the features of 311 complaints that are redacted prior to public access. In addition, the random forest classifier is a supervised learning model which learns from biases existing in the training data. In rough terms, this suggests that if DEP's strategy of enforcement for certain complaint types has changed over time, the model will extend whatever strategy was in place in 2016.

The final results of the random forest classifier model and the dashboard were presented to DEP operations manager Oscar Gonzalez and multiple inspectors in a feedback session in July 2019. The sponsors welcomed the deliverables and agreed they could be used to spot high-enforceability complaints, see clusters of related complaints, and ensure no high-priority complaints evade inspector notice. “I really love the tool you created, and with a little tweaking I believe it will be a powerful tool to help solve complaints across the city,” Gonzalez said. After the session, the team compiled sponsor recommendations and implemented many of them. Updates included adding a check for construction complaints with other nearby complaints and retooling the dashboard to surface AHV status. To our knowledge, our approach is the first machine learning model built for DEP inspectors, the first tool that predicts complaint enforceability, and the first model to incorporate inter-agency data in the form of AHV permits.

6. References

Cook, L. (2017). Newsday | Long Island's & NYC's News Source | Newsday. Retrieved from https://www.amny.com/news/noise-tops-list-of-complaints-to-nyc-s-311-last-year-report-says- 1.12972804.

DiNapoli, T. (2018). Noise in New York City Neighborhoods, Assessing Risk in Urban Noise Management. Retrieved from https://www.osc.state.ny.us/reports/health/noise-in-nyc.pdf

Galloway, W., & Bishop, D. (1970). NOISE EXPOSURE FORECASTS: EVOLUTION~ EVALUATION, EXTENSIONS, AND LAND USE INTERPRETATIONS. Retrieved 24 July 2019, from https://apps.dtic.mil/dtic/tr/fulltext/u2/711131.pdf.

Hong, A., Kim, B., & Widener, M. (2019). Noise and the city: Leveraging crowdsourced big data to examine the spatio-temporal relationship between urban development and noise annoyance. Environment And Planning B: Urban Analytics And City Science, 239980831882111. doi: 10.1177/2399808318821112

Mydlarz, C., Shamoon, C., & Bello, J. (2017). Noise monitoring and enforcement in New York City using a remote acoustic sensor network. Internoise. Hong Kong.

Office of the New York State Comptroller. (2017). New York City Department of Buildings - Responsiveness to Noise Complaints Related to Construction Projects [Issued 08/31/17] [NY]. Retrieved from https://osc.state.ny.us/audits/allaudits/093017/16n3_1.htm.

Wang, L., Qian, C., Kats, P., Kontokosta, C., & Sobolevsky, S. (2017). Structure of 311 service requests as a signature of urban location. PLOS ONE, 12(10), e0186314. doi: 10.1371/journal.pone.0186314

Wellington, B. (2015). Mapping New York's Noisiest Neighborhoods. Retrieved from https://www.newyorker.com/tech/annals-of-technology/mapping-new-york-noise-complaints.

Li, J., Luong, M. T., & Jurafsky, D. (2015). A hierarchical neural autoencoder for paragraphs and documents. *arXiv preprint arXiv:1506.01057*.

Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., & Chang, E. (2014). Diagnosing New York City's Noises with Ubiquitous Data. UBICOMP. Seattle.

7. Figures

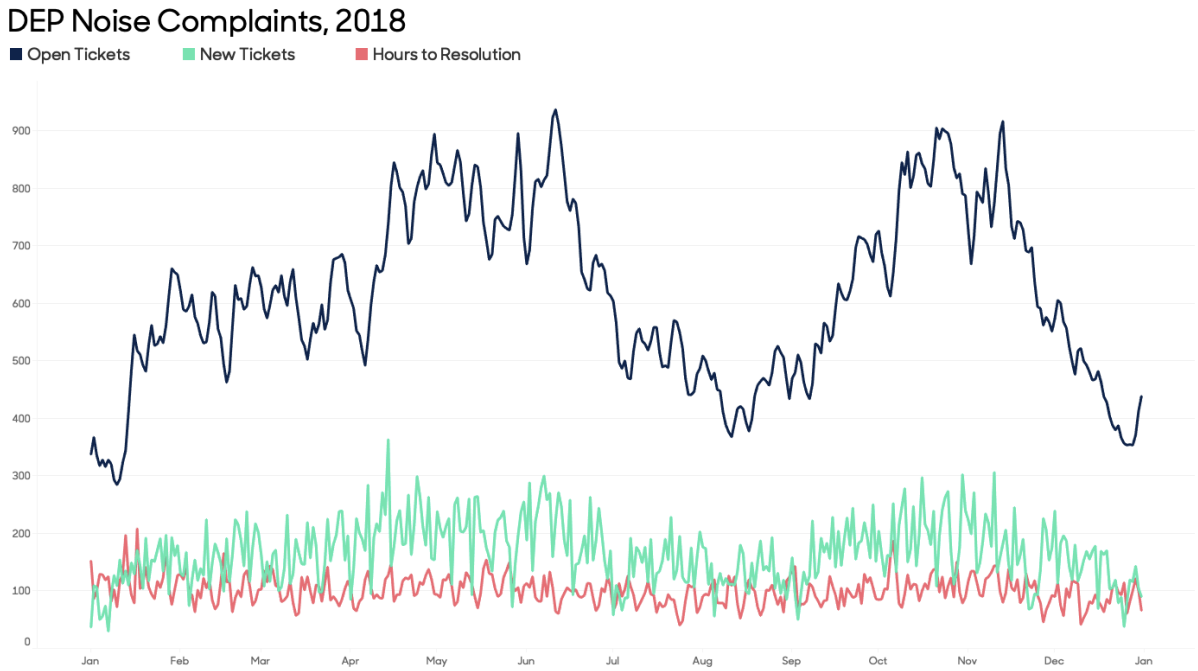

Figure A: Noise complaint volume in 2018 with seasonality visible. The black line represents the number of open tickets as of the beginning of the calendar day. The green line represents the number of new tickets received daily. The red line represents the average time to resolution (TTR) in hours for complaints received that day. New complaint volume and open tickets are highly seasonal. TTR is steady and weakly seasonal.

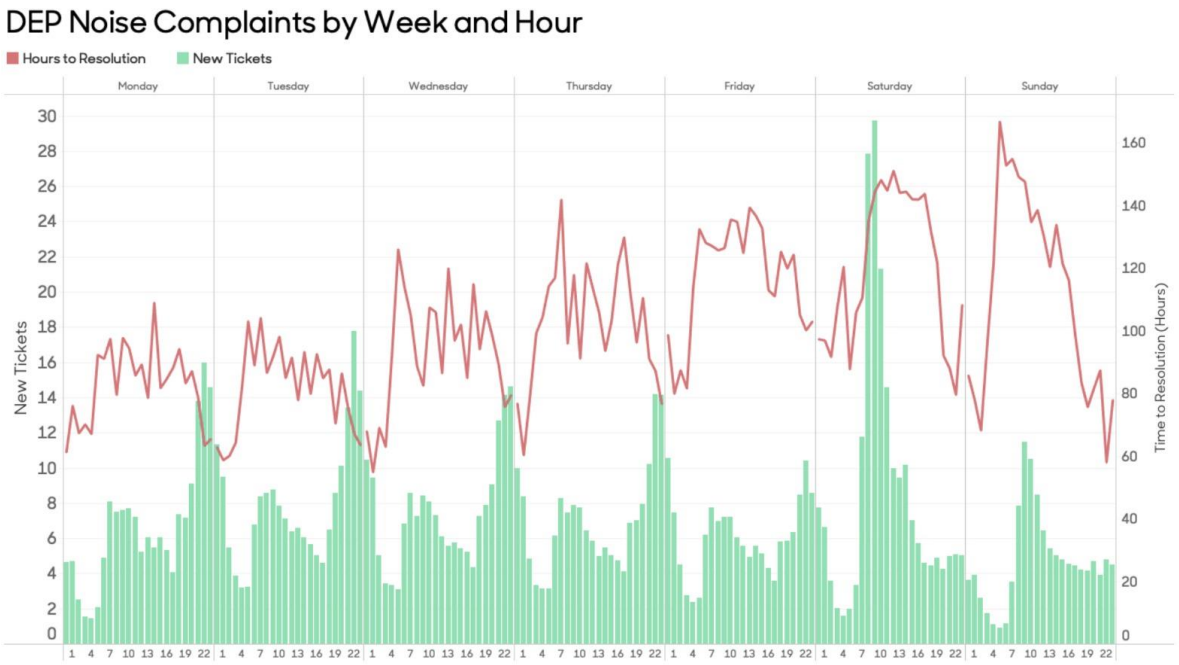

Figure B: Complaint volume by weekday and hour. Complaints are highest Saturday morning, with smaller peaks early in the workday and after work hours.

Resolution description	**Enforceability Score**
The Department of Environmental Protection investigated this complaint and shut the running hydrant.	1
The Department of Environmental Protection conducted an inspection and determined that "No Horn Honking" or "No Idling" signs are warranted. A request will be sent to the Department of Transportation to have the sign(s) installed.	1
The Department of Environmental Protection investigated this complaint and made a repair.	1
The Department of Environmental Protection observed a violation of the New York City Air/Noise Code at the time of inspection and issued a notice of violation.	1
The Department of Environmental Protection attempted to inspect this complaint but could not gain access to the location. If the problem still exists, please call 311 and file a new complaint with additional access information. If you are outside of New York City, please call (212) NEW-YORK (212-639-9675).	2
The Department of Environmental Protection researched this complaint and determined that it could be closed.	2
The Department of Environmental Protection attempted to contact the complainant by phone, but the phone number provided was incorrect. If the problem still exists, please call 311 and file a new complaint with the correct phone number. If you are outside of New York City, please call (212) NEW-YORK (212-639-9675).	2

Resolution description (cont'd)	Enforceability Score
The Department of Environmental Protection has inspected your complaint and determined that further investigation is required. More information will be available once the condition is resolved. Please visit nyc.gov/311 or call 311 at a later time to check the status of your complaint.	2
The Department of Environmental Protection conducted an inspection and determined that "No Horn Honking" or "No Idling" signs are not warranted. If the problem still exists, please call 311 and file a new complaint. If you are outside of New York City, please call (212) NEW-YORK (212-639-9675).	2
The Department of Environmental Protection did not observe a violation of the New York City Air/Noise Code at the time of inspection and could not issue a notice of violation. If the problem still exists, please call 311 and file a new complaint. If you are outside of New York City, please call (212) NEW-YORK (212-639-9675).	2
The Department of Environmental Protection sent a letter to the complainant and/or respondent and the letter was returned as undeliverable.	2
The Department of Environmental Protection closed or canceled this complaint at the complainant's request.	2
The Department of Environmental Protection requires an appointment to inspect this complaint type. Complainant information was not provided and an appointment could not be scheduled. The complaint has been closed. If the problem still exists, please call 311 and file a new complaint including your contact information.	2
The Department of Environmental Protection received a letter or phone call from the alleged dog owner in response to a letter or inspection.	2
The Department of Environmental Protection requires contact with the complainant to investigate this complaint. The complaint was closed because attempts made to contact the complainant by phone and/or letter received no response.	2
The Department of Environmental Protection resolved this complaint by speaking to the complainant on the phone.	2
The Department of Environmental Protection scheduled an inspection.	2
The Department of Environmental Protection determined that an inspection is not warranted to investigate this complaint at this time and sent a letter to the complainant and/or respondent.	2
The Department of Environmental Protection received a response from the complainant with additional information that was requested to assist in investigation of the complaint (see resolution code A22).	2
The Department of Environmental Protection determined that an inspection is warranted to investigate this complaint.	2
The Department of Environmental Protection determined that this complaint is not under its jurisdiction and referred it to the New York Police Department for further action.	2

Resolution description (cont'd)	Enforceability Score
The Department of Environmental Protection attempted to investigate this complaint but the address was incorrect. If the problem still exists, please call 311 and file a new complaint with the correct address. If you are outside of New York City, please call (212) NEW-YORK (212-639-9675).	2
The Department of Environmental Protection requires contact with the complainant to investigate this complaint. A message was left for the complainant at the phone number provided.	2

Figure C: Full list of 311 complaint resolution descriptions with assigned enforceability mapping. Enforceability score of 1 refers to higher enforceability, while 2 is indicative of lower enforceability.

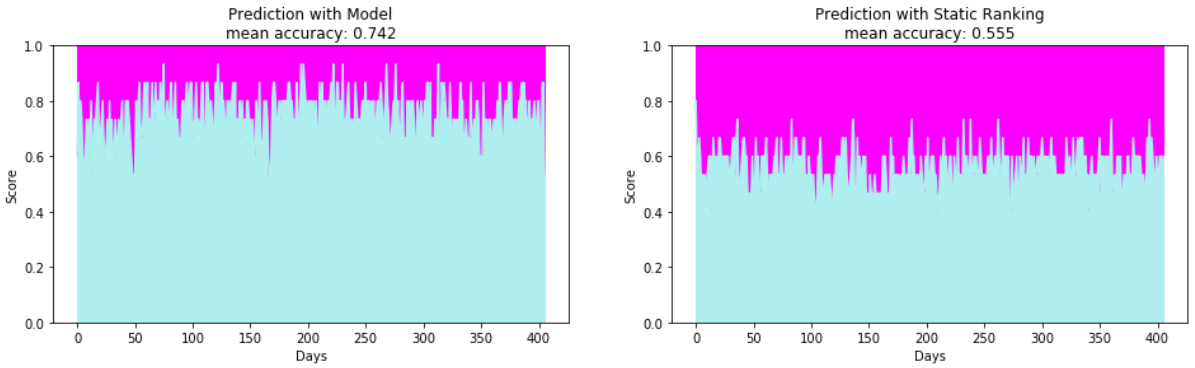

Figure D: Dynamic top 15 results in precision-at-15 accuracy of 74% (left), whereas the static ranking results in 56% (right).

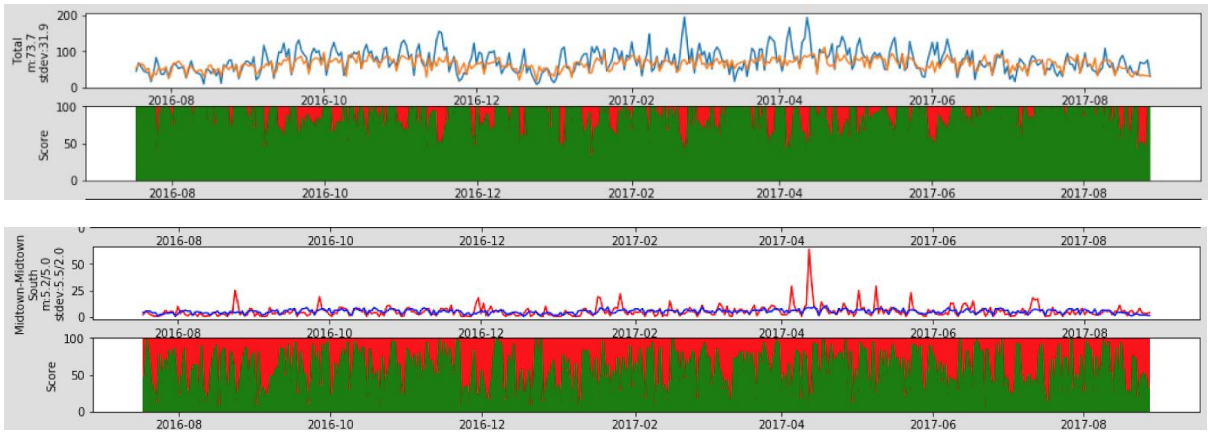

Figure E: Predicted noise complaints (green) compared actual noise complaints (red) for Manhattan and Midtown South, 2016-2017. Neural network models succeed in learning the periodicity of data, but fail to predict high-variational volume.

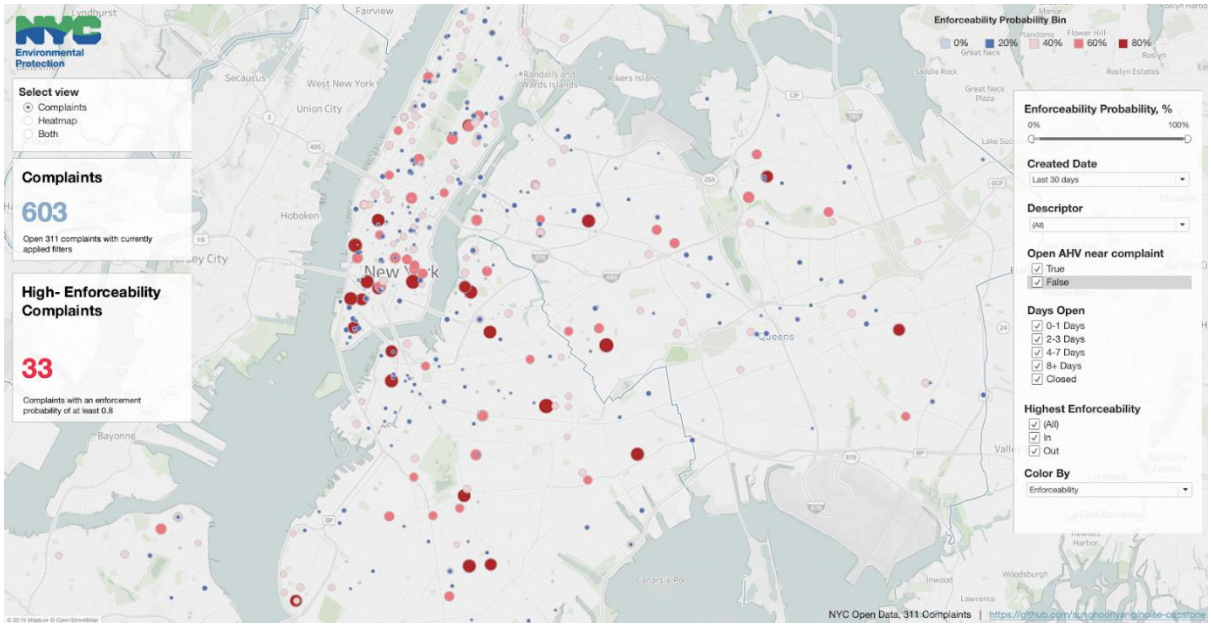

Figure F: A screenshot of the default view for the interactive dashboard visualization. Points represent complaints from one week in July 2019. Complaints are filterable by descriptor, AHV status, and model enforceability score.

Figure G: A screenshot of an alternative view from the dashboard showing heatmap and individual complaint points. Multiple complaints in a closed area are automatically aggregated into a heat cluster.

Model	Training Method 1			Training Method 2		
	Model performance, no subset (0.77 train, 0.33 test)			Model performance on random subset of 1,000 rows of data (0.77 train, 0.33 test)		
	F	P	R	F	P	R
A	0.165	0.304	0.113	0.138	0.074	0.911
B	0.144	0.636	0.081	**0.179**	**0.099**	**0.941**
C	0.197	0.347	0.137	0.128	0.068	0.941
D	0.197	0.347	0.137	0.146	0.080	0.882

Figure H: The performance of each of the models compared using F-score, precision and accuracy. Model A was based on all complaint types and all 660 features. Model B was based on all complaint types and the 10 features with the highest computed feature importance. Model C was based on construction complaints only and all 660 features. Model D was based n construction complaints only and the 10 features with the highest computed feature importance. Model B.2 was selected for its high recall and F score in accordance with sponsor specifications.

8. Supplementary Materials

8.1 Neural Network

We selected using Autoencoder approach to encode meaningful patterns in sequence through compression of information, hence encoding. The data that has daily batch shaped as 29 neighborhoods by the number of features is collapsed as a one-dimensional array and fed to some type of RNN for pattern recognition.

An LSTM model was chosen as the best performer out of RNN and LSTM and was employed as the first layer of an under-complete autoencoder/decoder pair, with two densenets to reduce dimensionality by 50% at each step, before recovering it again from 25% to 100% through the decoder that is of the same architecture as the encoder, but with increasing dimensionality to the correct output dimensionality for the final projection to the target space. The model employs normalization of data and the introduction of healthy noise using batch normalization and dropout procedures. Batch normalization is used for all linear layers, while dropout was used for LSTM layers. Softmax layer is used to output pseudo-probability density of complaint volume of each spatial bin on each day, which allowed us to rank the neighborhoods by probability and compute the Precision at 15 accuracy metric.

For the complete setup of the model, please visit the GitHub analysis/neural network subdirectory.

8.2 Random Forest Classifier

The random classifier was built using the features from American Community Survey (ACS), Primary Land Use Tax Lot Output (PLUTO), National Oceanic and Atmospheric Administration (NOAA) climate data and After Hour Variance Permits.

The after-hour variance permits are what allows a construction to be done after hours. The DEP indicated that this could be a useful tool to identify enforceable complaints. However, since this data is not publicly available, we had to make use of data scraped by NYU researcher Fabio Miranda (https://engineering.nyu.edu/student/fabio-miranda). This scraper had complaints only until mid-2017 and we had to scrape the permits for the 2019 data. Each of these files had differences in format and were tackled in separate notebooks.

The steps taken to model were to spatial join the complaints to all the shapefiles. Most of these features were already numerical but the non-numerical features had to be encoded. With these features we built the models. Multiple iterations of models were tested. The settings we thought would affect the model were:

- Construction/All type of complaints
- Proportion of Enforceable complaints used for training
- Number of features used

We tested each of these permutations and finally chose the model trained on top 10 features, all types of complaints and equal proportion of enforceable and unenforceable complaints. We used the F-score and recall to pick this model. The models were trained using complete 2016 complaints and then run on a week of 2019 complaints to finally produce a dashboard that the DEP could test and provide feedback. Ideally, we would have trained on complaints closer to 2019 but the limitation in our work was the availability of historical AHV permits. The complete notebooks on modelling and assigning AHV permits are present here.

8.3 Web Scrapers

Two web scrapers were written to extract data from other New York City public agencies that is not otherwise available in aggregated form, namely an after-hours variances (AHV) scraper and a DOT permit scraper. Currently, information about AHVs is available only through the Department of Building's website and only by searching for one building at a time and navigating a series of forms in the agency's archaic web portal. DEP repeatedly requested AHV data be available, as they do not currently have a source for AHVs and have to manually look up each complaint site. Full instructions on how to use the scraper are available in Github under the ahv scraper subdirectory. Several researchers or public advocates previously scraped AHV data (SONYC, BetaNYC) through 2016 and that static data was fed for training into the random forest classifier. The previous scrapers, however, performed poorly by looking up and looping through each city building identification number (BIN). The scraper written for DEP was intended to be lightweight and retrieve results more quickly. However, it should be noted that there is a ceiling in script performance due to server-side throttling or blocking from frequent client requests. To retrieve as many AHVs using as few web requests as possible, the scraper exploits the vulnerability that AHVs IDs are numeric and incremented. By identifying a starting point, AHVs can be returned by incrementing the AHV ID and feeding the new ID into the webpage's allkey query parameter. AHV form data is scraped using Beautiful Soup and exported to CSV format. The script automatically detects when it has reached the latest available AHV and on future runs will resume requests from that AHV.

In testing, the AHV scraper returns about 4,000 AHVs per hour and was used to evaluate testing data. In discussion with DEP, we provided guidance on how to productionize the scraper and recommended that it be run hourly to minimize the consecutive web requests and to have data in near real-time availability.

In the final sponsor meeting in late June 2019, DEP inspectors pointed out that NYC Department of Transportation Permits are also valid variances and should be taken into account in the model. Due to the belated request, we could not write the scraper and incorporate it into the model, but we did develop and circulate a methodology to retrieve this information in CSV form by manipulating the request parameters of the API endpoint used in the DOT website. See the dob_scraper subdirectory in GitHub for full documentation.

8.4 Dashboard Visualization

Output from the random forest classifier model is exported in Pandas to two CSV files, open complaints and closed complaints. We selected Tableau Public to create the interactive dashboard visualization. The data source was the union of the two CSV files, with a small transformation to cast closed_date as datetime. A Tableau extract (proprietary in-memory cache) was taken to improve performance. The model's enforceability probability was binned by quintile. DEP picked out filters and visualization components. A full list of filters and options is available in the dashboard subdirectory or by interacting with the dashboard on our website.

9. Team Collaboration

Project work was evenly undertaken by the five group members.

Qinyu led administration and logistics, which included responsibilities for coordinating meetings with sponsors and clients, enforcing team deadlines, and taking meeting notes. She also reviewed NYC Department of Buildings (DOB) construction permits to identify seasonal patterns and worked closely with Siddhanth to set up the random forest classifier and interpret its findings.

Zoe shared additional responsibility for contacting stakeholders and planning group work sessions. She edited written reports for cohesiveness and transformed PLUTO zoning data to facilitate spatial joins for modeling. Zoe also built the final project report website.

Sam investigated temporal patterns in 311 complaint data and the average complaint time to resolution. He wrote the AHV and DOT web scrapers and created the interactive Tableau dashboard. He also reviewed reports for grammar and voice.

Siddhanth located demographic and census data that will be used for additional features in our developing model and served as lead for the random forest classifier model, including presenting results to sponsors and honing the model based on sponsor feedback.

Duke leveraged his prior experience as a data engineer to set up a shared team PostgreSQL database on ADRF for ad hoc analysis. He also analyzed how 311 complaints are dispersed temporally and spatially and built the LSTM neural network model. Finally, he administered the team GitHub repository.