Analysis of NYC Noise Complaints

Final Project in Applied Data Science

Pratik P. Watwani (ppw220)

Abstract

With the introduction of the NYC Noise Code Guide, NYC residents recognized that they now have the potential to report complaints of noise to the city agencies. In recent years, the "311" department has recorded more than four hundred thousand complaints of noise alone. This project aims to discover if there exists a relationship between specific demographic characteristics in the population that would explain the tolerance in the population to report noise nuisance to the "311" department. To attend to this, we focus our work on New York City, as it's radically an immigrant city with an assorted gamut of races dwelling in it. Further ahead, we concentrate on the demographic characteristics of Age, Gender, Race as well as Education and attempt to discover patterns and correlations with reporting potential respecting NTA census tracts.

Introduction

Launched in 2003, NYC 311 was originally created to filter non-emergency calls away from the emergency phone line, 911. Since 2010 when 311 complaints dataset has been released to the public, it soon became an information hub which gives a window for the government to know the citizens' requirements and insights about the possible avenues to provide a better service for the public.

Living in cities, people are encountered with different kinds of noises every day. According to the results shown by the World Health Organization's *Guidelines for Community Noise* (1999), the general population are increasingly exposed to community noise. Community noise includes environmental noise, residential noise and domestic noise, which are mainly from traffic, neighborhoods and construction. The indoor sources of noise include machines, house fixing and so on. Near some open spaces, people may complain about barking dogs for unregulated pets' activity. The health department has stated that noise may result in several adverse effects, including hearing loss, stress and anxiety. Furthermore, there is increasing evidence that long term exposure to high levels of noise can cause direct health effects such as heart attacks and other health issues, which creates a significant public health concern. In most countries, regulations of community noise, such as construction noise and traffic noise, have been released by limiting the fixing time

or applying emission standard of industry plants. However, due to the lack of methods to define and measure indoor noise, this part is much more difficult to control.

In this research, we are trying to use 311 complaints data and demographic data from the US census to understand the relationship between complaints and demographic features. After cataloging all noise types, we can figure out what kind of noise is complained most in certain areas.

Literature Review

In New York City Department of Environmental Protection's *A GUIDE TO NEW YORK CITY'S NOISE CODE* (2018) it provides an overview of the noise code and the most common sounds in a city. The code divides all noise complaints into several categories, which are construction noises, animal noises, noises from vehicles, indoor noises and noise from the neighbor. Based on the categories, a cartographic design firm CartoDB[1] mapped publicly available 311 noise complaint data by census tract and created a dashboard that allows users to study those complaints. From the complaints map, Manhattan is by far the loudest borough of NYC (Laura Bliss, 2016). She also found that there might be seasonal difference in noise complaints. Besides that, calls about noise from neighbors are most frequent in zones between racially homogenous neighborhoods (Laura Bliss, 2015). Sule Yilmaz et al (2017) found that females tended to report noise exposure or a decrease in their tolerance to noise more. In their results of questionnaire, females had significantly higher scores than men in terms of both the total and the attentional and emotional dimensions, when assessing their sensitivity to noise.

1 Mapping last year's 311 noise complaints, http://bl.ocks.org/nerik/raw/90c087a3f0fe96f8a2ce/#13/40.6994/-73.9653

Data and Data Preprocessing

Of 22 million observations in 311 complaints data, 1.1 million complaints related to noise are reported within the 10 years. For a further study, census data are collected from the NYC Open Data[2], including demographics, economy, housing and socio data. All census data are at NTA level, which are aggregations of census tracts that are subsets of New York City's 55 Public Use Microdata Areas (PUMAs). In order to visualize them, Boundaries of Neighborhood Tabulation Areas dataset[3] from the NYC Department of City Planning is also collected, using whole census tracts from the 2010 Census as building blocks.

I. Initialization

With such a huge amount of data to be worked on, it was crucial for it to be loaded entirely and more importantly as efficiently as possible. The team chose to work on SODA api, using which we were able to fetch the data in real time. Furthermore, this API call was passed in "modin" package to process and load data in a reduced time frame as compared to pandas standard library. As compared to pandas standard library the data set when loaded via modin, took ~650 seconds to load completely.

II. Data Cleaning

In order to bring to a consistent and usable form, each data set was thoroughly cleaned, but before was made sense by understanding the inter- and intra- relation of the data sets. How each data was distributed and the data contents. The process began with understanding the structure and organization of the data with the help of the metadata provided by the respective agencies. Taking into consideration the metadata, we conducted an exhaustive study of the patterns and the values in the datasets, what each column represented, how it varied, it's distribution, types of values in each attribute, and subsequently what each of these values meant in the given context.

Understanding distribution

Understanding the distribution of the data in each column and what kind of values did each column held in order to understand the range, context and values.

2 Demographics and profiles at the Neighborhood Tabulation Area (NTA) level, https://data.cityofnewyork.us/City-Government/Demographics-and-profiles-at-the-Neighborhood-Tabu/hyuz-tij8

3 Boundaries of Neighborhood Tabulation Areas, https://data.cityofnewyork.us/City-Government/Neighborhood-Tabulation-Areas-NTA-/cpf4-rkhq

Type Conversion

One of the most prominent issues of the data sets were problems with inaccurate data types. Columns that held numeric values were string formatted, thus required a suitable conversion to numeric data type. A particular column, "Created Date" which informed about the time, was essential to get it into a proper DateTime format in order to establish our preliminary analysis.

String formatting, Spell Check, and Removing Special characters

The datasets involved inaccurate string formatting, repeating characters, words with abbreviations, spelling errors, special characters in string values; column "City" had values like "NA," "N/A," or "No Value" were replaced with the help of geospatial coordinates from data.

Range Constraints

A few of the data points fell out of the range of the values that the column held. In the case of two complaints, the created date was 1960. For this, the data points were generalized to the previous data point.

Eliminating irrelevant/redundant attributes

After extraction, certain attributes were irrelevant for further analysis. "Zipcode" used to aggregate the data to the census level NTA tracts, was eliminated. Likewise, agency, agency_name, location_type, complaint_type, close_date, were then decided upon by the team to be eliminated as they did not serve any purpose for the prospective analysis.

III. Spatial Join

New York City is reasonably populated with roughly ~8.3 million people that reside in the city, given this, it would be unintelligible to perform analysis on noise reports on the zip code level consecutively taking into account different demographic characteristics, as NYC is decently made up of a good blend of a variety of races, ages, and income groups.

Thus, to ensure a more accurate analysis, the 311 data that was obtained on the zip code level was then consequently transformed to census level tracts, NTA, called Neighborhood Tabulation Areas. New York City, with it's all 5 boroughs have more than 50+ NTA regions. The aim was to establish some patterns, and correlate demographic data to noise complaints, since on a higher level, using zip code, it would rather be imprecise to make a judgement call on the hypothesis.

The image(Appendix A, Figure 1) furnishes a convenient pictorial representation between an NTA area and a zip code.

Methodology

Our project can be divided into 2 parts, exploratory data analysis and correlation analysis. In EDA, we implement spatial join method to combine 311 noise complaints data and NTA boundaries data. In this way, we can explore its temporal characteristics, which will be discussed in the visualization and temporal analysis part. In the correlation research, we try to use demographic data, economic data, housing data and socio data to find potential factors of noise complaints. In this part, regression method is also implemented. Detailed methods are in the Appendix A, Figure 2.

Visualization and Temporal Analysis Results

I. Overview

In 5 boroughs, Manhattan has the most noise complaints. As for the noise type, there are 31 distinct types of noise complaint categories(Appendix A, Figure 3), “Loud Music/Party” accounts as the most reported type of noise with more than 400k+ reports. “Construction” follows as a major source of noise in the surroundings. Other categories that follow “Banging/Pounding,” “Loud Talking,” “Barking Dog” et cetera. (Appendix A, Figure 3) projects the top 10 most reported noise categories.

II. Time series analysis

To understand how different categories of noise complaints span across the time range, we visualize all complaints by “Months” and “Hours”. Monthly plot helps us to understand the frequency of noise complaints in each month and an insight about seasonal trends. Hourly plot gives a direct view of the trend of noise complaints hourly on a given day. Furthermore, the two histograms are subcategorized into different categories of noises to provide a better understanding of the data.

From the monthly visualization(Appendix A, Figure 4), we can observe that noise complaints are reported more frequently in summer. Allegedly, summer season drives out people for rooftop

parties, gatherings and more social outgoing events, which thusly involves more noise production. Besides that, an analysis indicates that the rise in the complaints of noise overall can also be attributed to the introduction of common complaint portal in the city[4].

Digging into noise categories, there is no significant difference in construction noise. The decrease of winter noise complaints may be attributed to less loud music and party complaints. A possible reason to it is that in winter, people hold less party outside, so noises from neighbors are reduced in the source. Besides that, people generally close windows for better heating performance, which blocks the noise to a certain extent.

There is a visible difference between the number of noise complaints when observed as day vs night(Appendix A, Figure 5). The highest volume of noise complaints is collected at 11:00 pm, near midnight. A decrease is identified during the initial hour of the day, at five o'clock and at noon - which corresponds to the second rest period. Noise complaints show an increasing pattern after people get off work. There is a fair recognition of categories and their distribution in the "Hour" plot when compared to the "Month" plot, as it reports several noise types and their contribution to the whole count, these include construction noise, loud music and jack hammering noise.

A notable difference is originated from the loud music and party. During daytime, a small number of people choose to report complaints of loud noises from neighbors, streetwalkers during night, parties and social gatherings lead to an increase in noise nuisance. The construction noises are complained most in the 9:00 am and near midnight. Possible reason is that construction generally begins in the morning. People are more sensitive to noise in the night when it is comparably quiet in the surroundings. As a result, construction noise and jack hammering noise will be extremely 'noisy' to people even though the sound may not reach that high decibel level. Although the Department of Buildings receives a significant income of funds via work permits, they keep issuing

4 Is New York Getting Too Loud, or Are New Yorkers Just Too Whiny?, Vice Report. A Guide to New York City's Noise Code, nyc.gov

extended permits, which leads to a number of construction activities take place after midnight hours[5,6]

From these figures, we already know that the loud music & party, construction noise and banging and pounding are the most influential types of total noise complaints, which count the biggest ratio of total change of it. In order to understand the change pattern of each category monthly and hourly, we plot them with normalized value. From the monthly figure (Appendix A, Figure 6), we can see that loud music & party and loud talking show the same pattern, which are both higher in the summer. A possible reason might be that they both belongs to the type, noises from neighbors. In this case, we try to explore the change pattern in each main category of noises. We also found evidence in hourly normalized figure (Appendix A, Figure 7). Loud music & party, car music, loud talking and loud television all show patterns with higher value in the night time, while environmental noise, including construction noise and jack hammering, seems to be complained more in the morning.

III. Spatial analysis

For spatial analysis, we first used a spatial joint to map all 311 noise complaints to each NTA and visualized the frequency of noise complaints in the NYC map.

The chart (Appendix A, Figure 8) indicates the distribution of 311 noise complaints. Considering the extremely high level in Manhattan, we also plot the log version (Appendix A, Figure 9).
From the two analysis(Appendix A, Figure 8 and 9), we observe that:

- The highest number of noise complaints are reported in Manhattan- Lower Manhattan, Upper West Side and Washington Height, and around Downtown Brooklyn. This might be because of more active commercial and entertainment entities - noisy events and more population involvement in these regions.
- The least number of noise complaints are reported in JFK Airport in Brooklyn, LaGuardia Airport in Queens and the Great Kills Park in Staten Island. Though airports may have engine

5 Noise Construction Sleep, Jeffrey C. Mays, September 27, 2019, New York Times
6 Noise from Construction, NYC 311 Portal

idling problem, people there are basically travelers and working staff who will probably not complain about it.

- Staten Island turns out to be the borough with the least complaints about noise. Fewer people live there, so the complaints collected there are also less. In this case, the population can be proved to be an important factor for the total number of noise complaints in plain sight.
- Complaints from park areas are significantly fewer than those from areas around the parks. With less traffic and construction works, parks are places with less unpleasant noise, while people in the parks may also be more tolerant to the noise of entertainment and tend to not report about them.

In order to figure out what kind of noise is reported most in each NTA, we use a pivot table to count the value of noise complaints in each type in each NTA. After sorting them, we got the most complained type of noises in each area. Then we plot it on a NYC map. The result is shown below.

Observing (Appendix A, Figure 10) we conclude:

- In most areas, loud music and parties are the main sources of noise. As proven in the overview part, it also counts the biggest part in the total number of noise complaints.
- Construction before/after hours comes to the second place among the reasons for noise complaints affecting widespread areas. NTAs with the construction noise as the most noise complaints distribute mainly in Upper East, Upper West, Midtown, Lower Manhattan, Brooklyn Height, Red Hook, Kensington, Forest Hills, Starrett City where there are more buildings requiring fixing.
- In John F. Kennedy International Airport and LaGuardia International Airport, the engine idling is the biggest problem, which can be understood explicitly.
- In Staten Island, noise from barking dogs can be a big problem. In Great Kills Park, Rikers Island and Soundview, there are rare machines running or party holding. Open spaces create good places for pets to have fun, which becomes a relatively significant source of noise if dogs are not trained well. Surprisingly, in Great Kills Park, loud talking can even be the most annoying noise.
- In Bronx, complaints about the noise from ice cream truck becomes the main component of 311 noise complaints in New York Botanical Garden, Bronx Zoo and Pelham Bay Park. Away from commercial area, there is less construction works in these areas. People live in a more quiet

neighborhood, such as Pelham Bay Park, the largest park with golf course and riding trails. Without other kinds of noises to complain about, never-ending ice cream truck jingles becomes music from hell[7].

Correlation Analysis and Testing

Null Hypothesis: Demographic characteristics are not an influence in determining the tolerance/sensitivity of the population in reporting complaints.

Our analysis began with testing the effect of each variable with the complaints density(Appendix B, Tables 1-4). When considered separately, each variable tends to have had a considerable effect. Say, (Appendix B, Table 1) which is a regression analysis for the data of Job Type , a good amount of attributes fetch high significance value.

However, the case completely alters when all of the variables are considered altogether at the same time. With this, we reach 61 features that hold some weight in the complaints; of these, 22 features talk about jobs, 13 about race, 7 about education, 6 about origin, and 13 about salary, and to continue with regression testing (Appendix B, Table 5), we conduct a PCA dimensionality reduction, and the contribution of each attribute in the PCA, which is furnished in (Appendix A, Figure 10).

To understand which components in the PCA analysis have higher weightage in the data is to look for the eigenvalue of each attribute. (Appendix A, Figure 11) is the heat map that shows the eigenvalue of each attribute, and observationally "Race" seemingly does not have a major effect in the relationship. Moving ahead in (Appendix A, Figure 11), one of the categories that have significantly good results is "Education" where the attribute "Edu 07" which stands for "Graduate/Professional Degree" captures the highest attention with a value of 0.48. It is thus evidently convenient to say that a population that has an educational attainment of an advanced degree has a significant weightage in the relationship. Next follows is the attribute "Edu 06" which stands for "Bachelor's Degree" which produces a score of 0.45 - the next significant attribute.

7 Incessant Ice Cream Truck Jingles Force Neighbors Into Meltdown, 2014, https://www.dnainfo.com/new-york/20141201/new-york-city/incessant-ice-cream-truck-jingles-force-locals-into-meltdown/

Considering another category of the data, “Or”; talks about the origin of the population, say, a native or an immigrant from another continent. “Or5” which stands for “Latin America” has the highest relationship score. Evidently from the analysis none of the other origins have such a strong relationship. Salaries have multiple attributes that contribute in a considerable effect overall, falling in the mid range of the eigenvalue scores. From the analysis, population that have higher salary ranges are more active in reporting complaints than those in lower earning scales, for example, attribute “Sal5” is a representative for a payscale of $200,000+ earnings, with a score of 0.26.

Conclusion

In this paper, temporal analysis is presented. The present study focuses on establishing a relationship between noise sensitivity and different demographic characteristics and how they affect the reporting sensitivity/tolerance. Regression between complaints density and selected features is included. While individually a lot of attributes were strong, collectively, only a small portion of them attended to be holding a strong relationship. The data we considered focused on points beginning 2010 and beyond. Our analysis estimates Education and Origin factors taking the lead in building the relationship, followed by Salary and Race. Further research can be undertaken to incorporate an extensive range of demographic characteristics like Housing, Neighborhood, Area Type et cetera. We also suggest that a more drilled down analysis can be undertaken beyond NTA tracts, more precisely BBLs which can be a more thorough analysis.

References

1. World Health Organization, Geneva. “GUIDELINES FOR COMMUNITY NOISE.” *GUIDELINES FOR COMMUNITY NOISE*, 1999, www.who.int/docstore/peh/noise/Comnoise-1.pdf.
2. Environmental Protection, NYC. “A GUIDE TO NEW YORK CITY’S NOISE CODE.” Noise Code - DEP, Mar. 2018, www.nyc.gov/site/dep/environment/noise-code.page.
3. Mapping last year’s 311 noise complaints, http://bl.ocks.org/nerik/raw/90c087a3f0fe96f8a2ce/#13/40.6994/-73.9653
4. Bliss, Laura. “Mapping Where New Yorkers Can't Stand the Commotion.” CityLab, 25 Jan. 2016, www.citylab.com/design/2016/01/mapping-new-york-city-noise-complaints-311/426606/.
5. [1] Bliss, Laura. “When Racial Boundaries Are Blurry, Neighbors Take Complaints Straight to 311.” CityLab, 25 Aug. 2015, www.citylab.com/equity/2015/08/when-racial-boundaries-are-blurry-neighbors-take-complaints-straight-to-311/402135/.

6. DiNapoli, Thomas P. “Responsiveness to Noise Complaints Related to Construction Projects.” Responsiveness to Noise Complaints Related to Construction Projects, Aug. 2017, www.osc.state.ny.us/audits/allaudits/093017/16n3.pdf.
7. Yilmaz, Sule, et al., “Assessment of Reduced Tolerance to Sound (Hyperacusis) in University Students.” *Noise & Health*, Medknow Publications & Media Pvt Ltd, 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5437755/.

APPENDIX

Appendix A: Figures

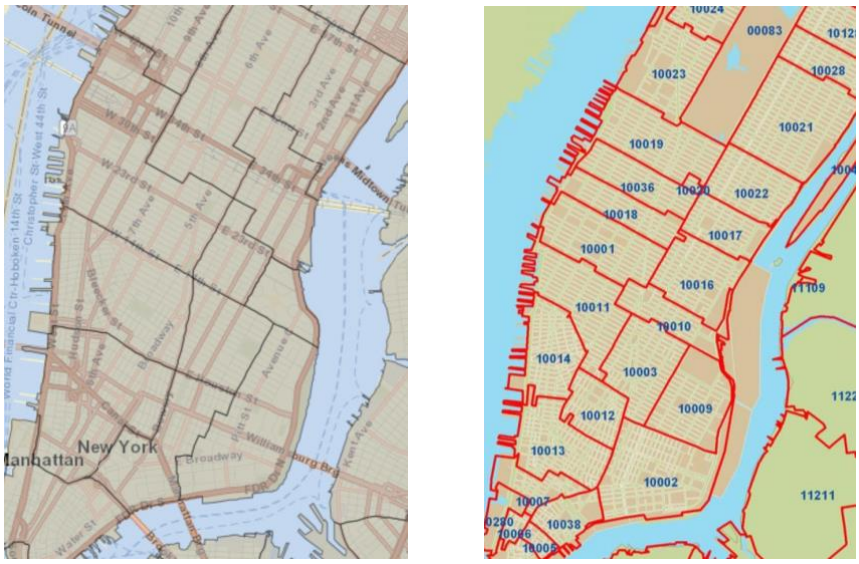

Figure 1: Pictorial representation of NTA areas(left) and zip code areas(right) in New York City

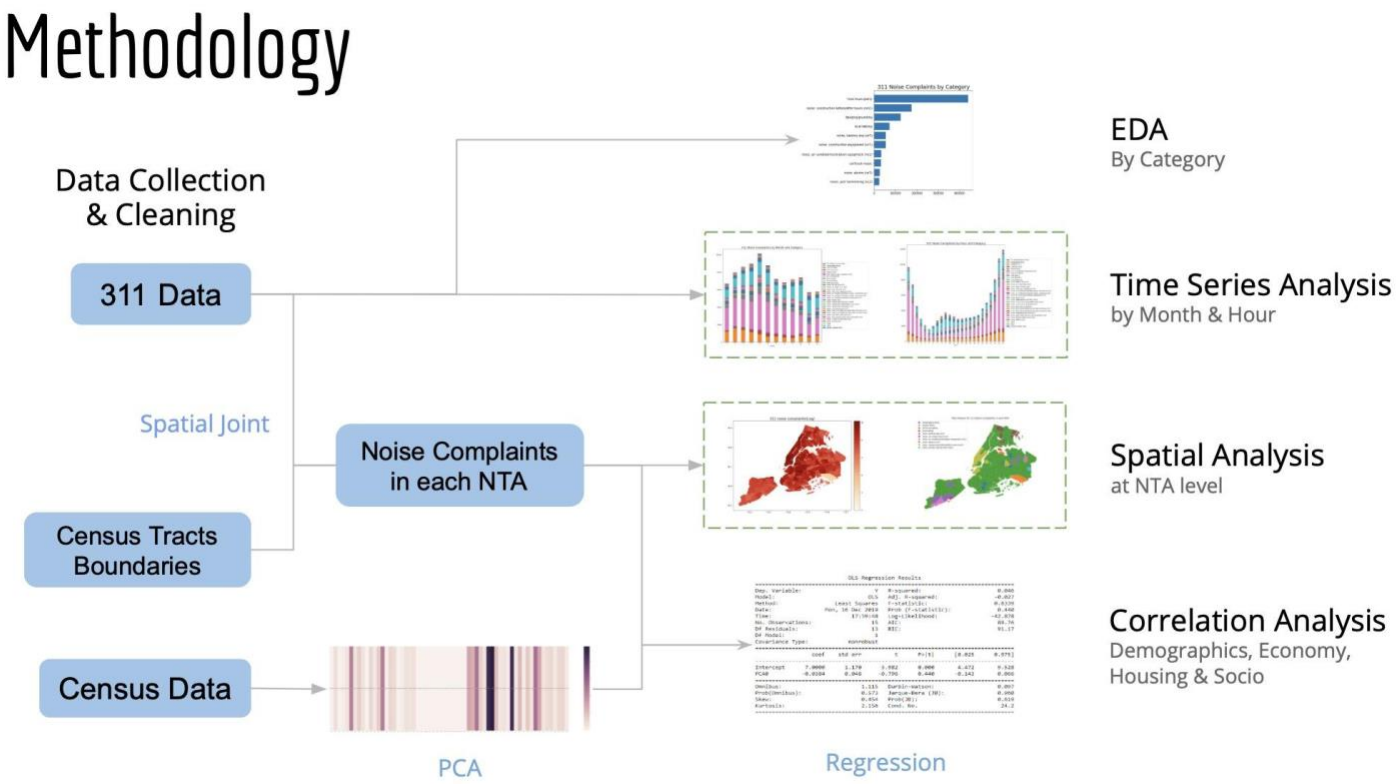

Figure 2: Methodology workflow chart

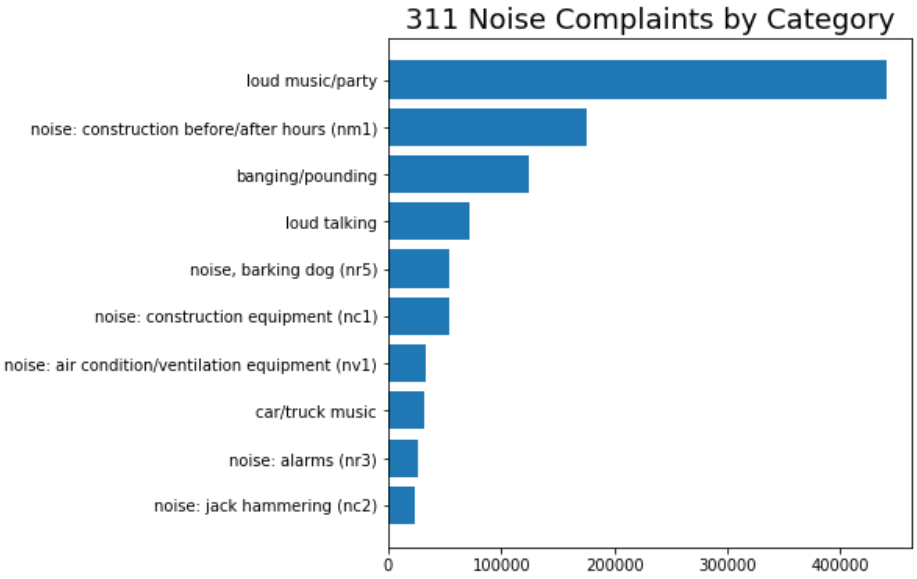

Figure 3: 311 noise complaints by category

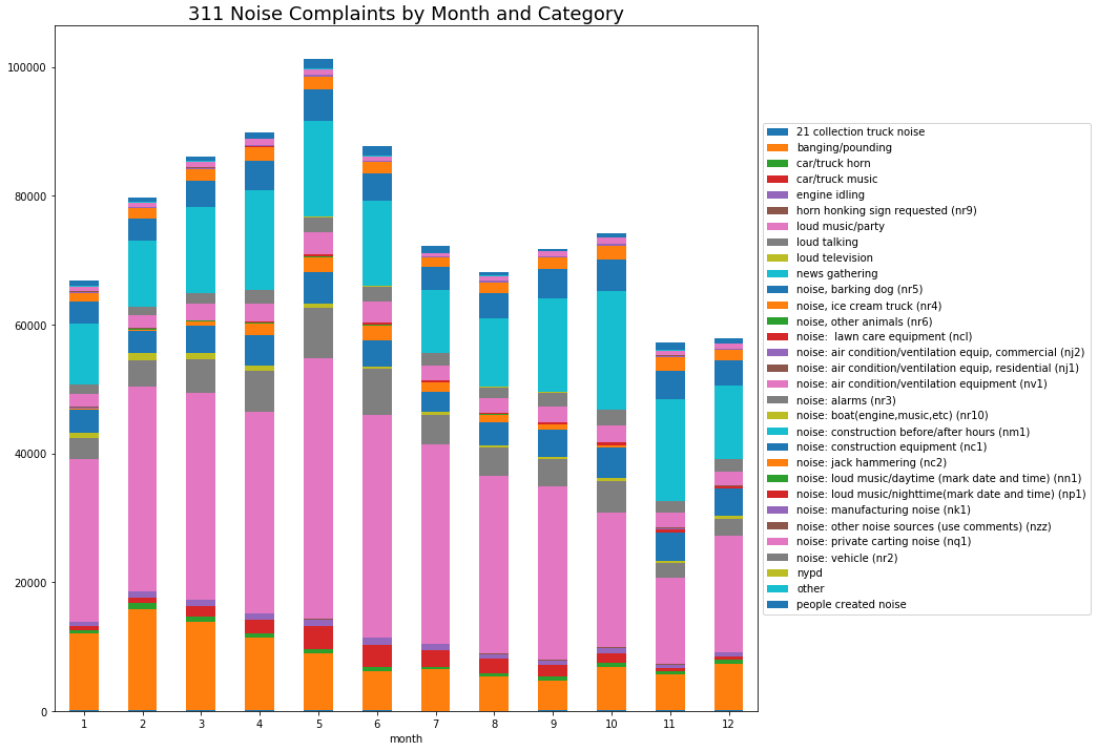

Figure 4: 311 noise complaints by month and category

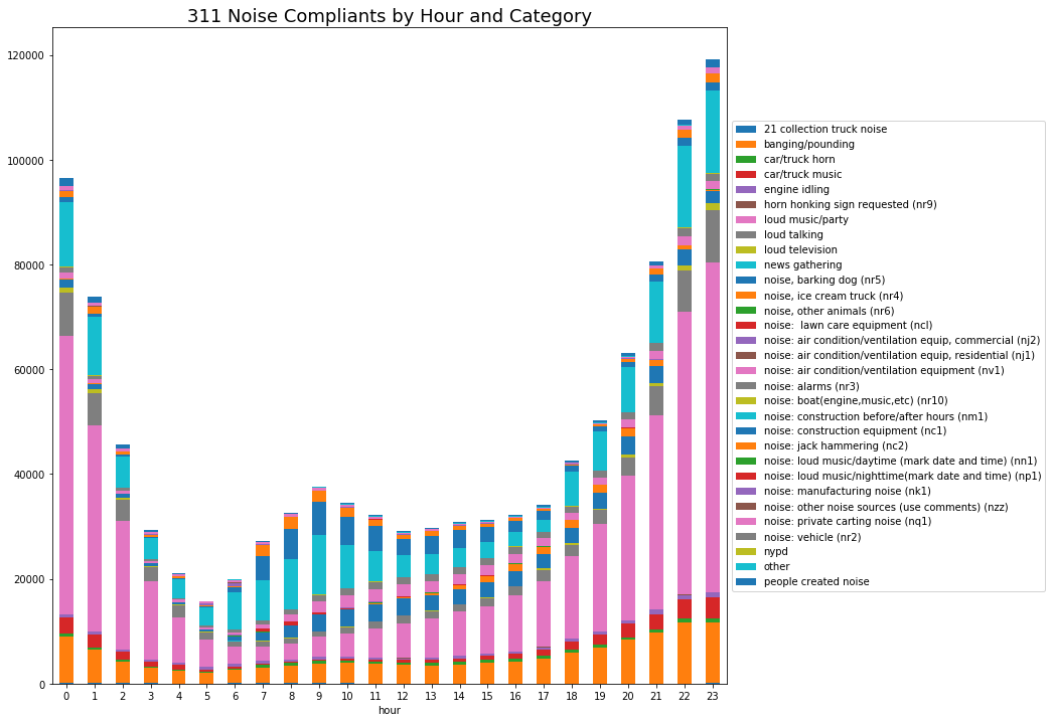

Figure 5: 311 noise complaints by hour and category

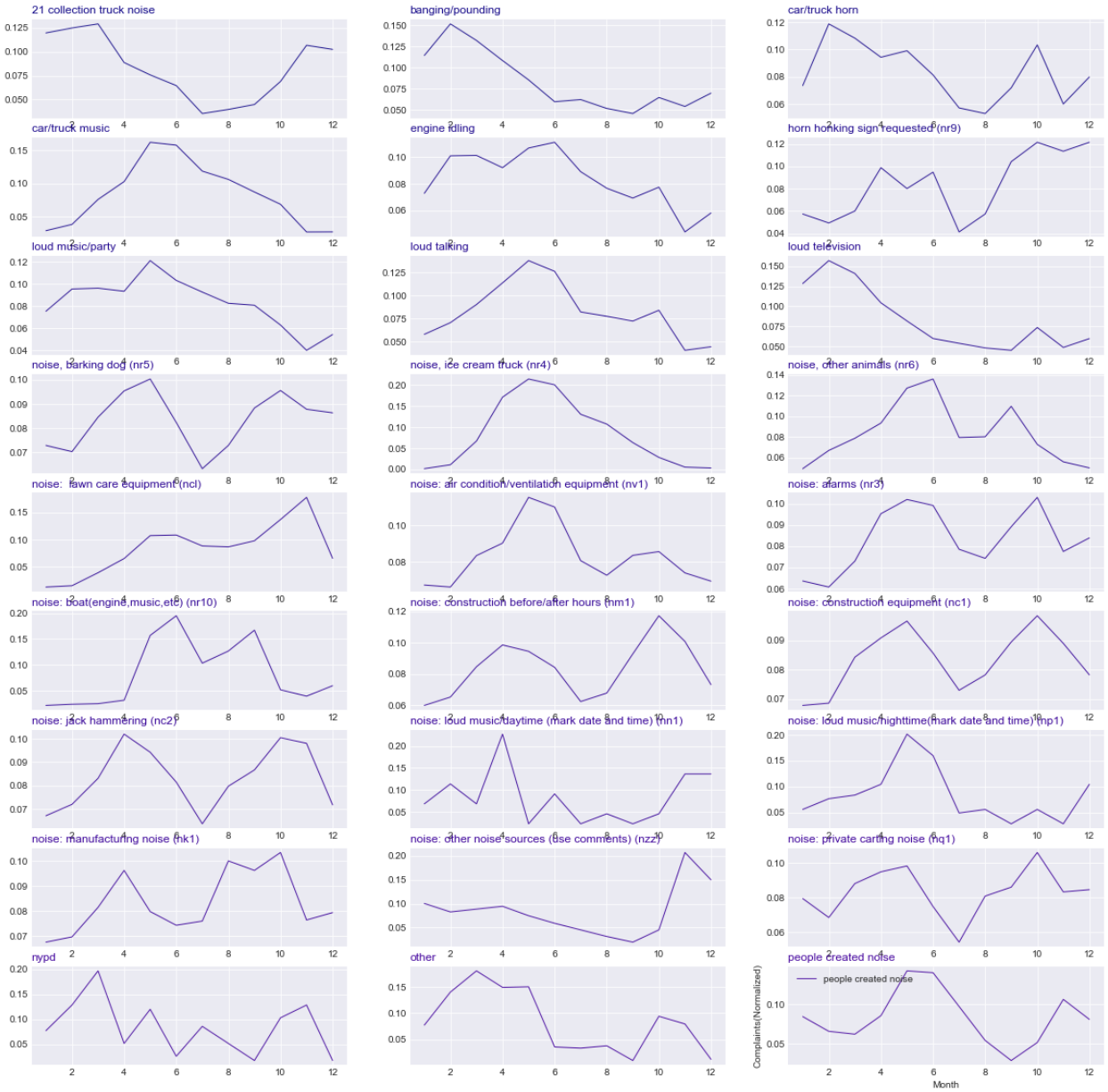

Figure 6: normalized 311 noise complaints by month and category

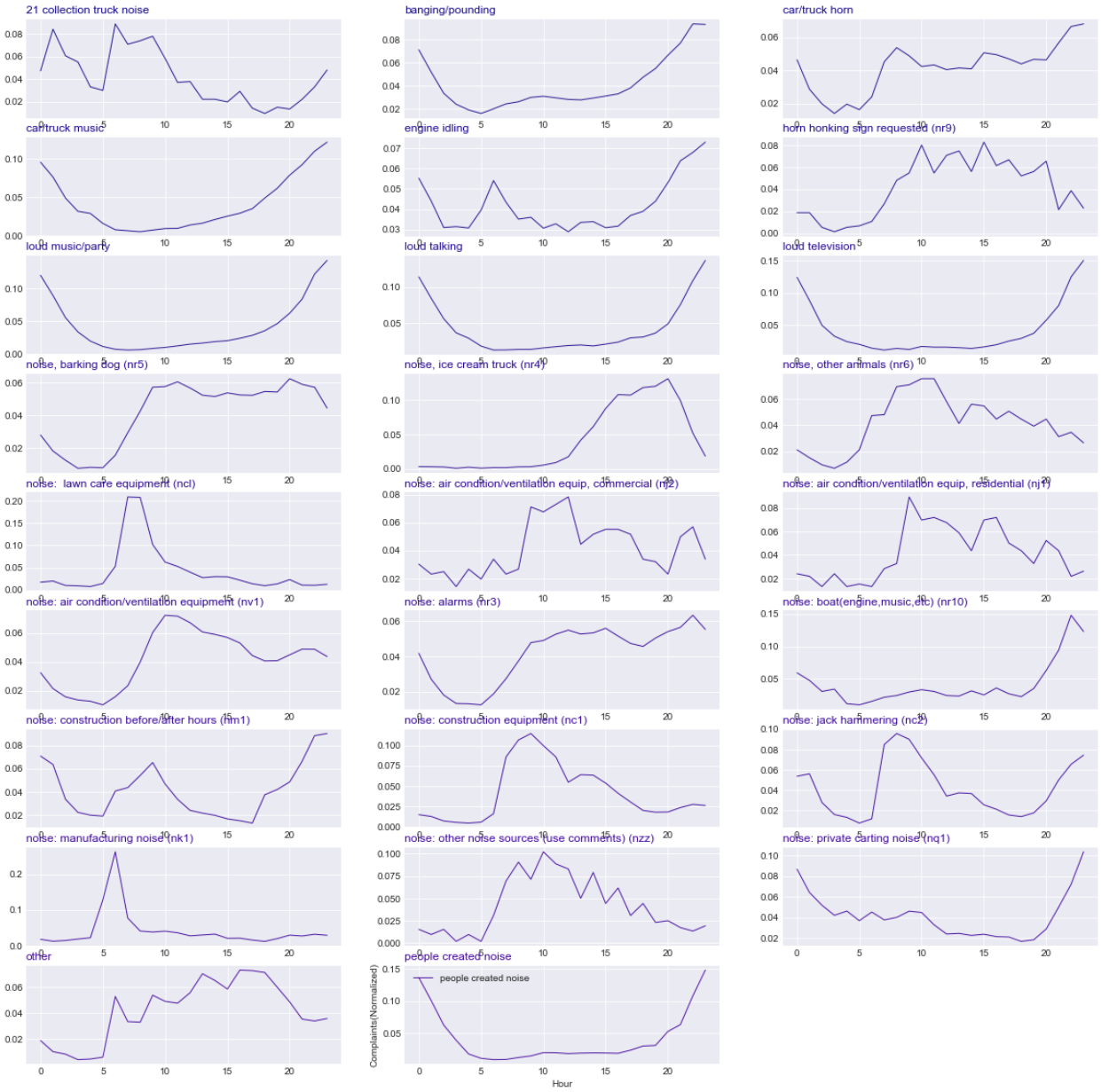

Figure 7: normalized 311 noise complaints by hour and category

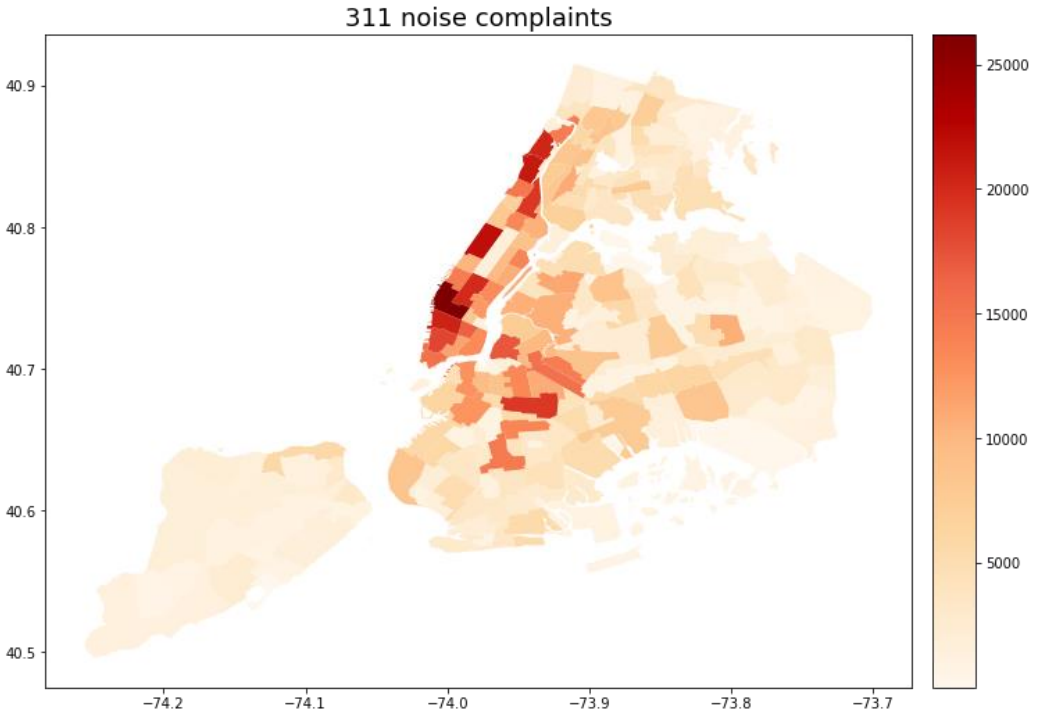

Figure 8: 311 noise complaints in NYC

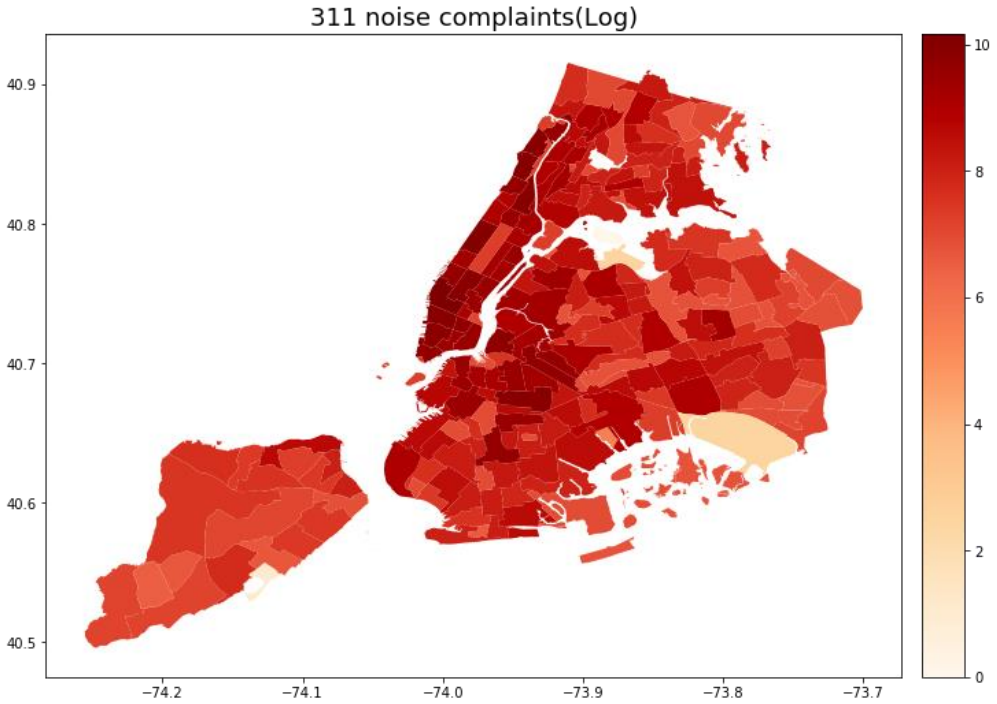

Figure 9: 311 noise complaints (log version) in NYC

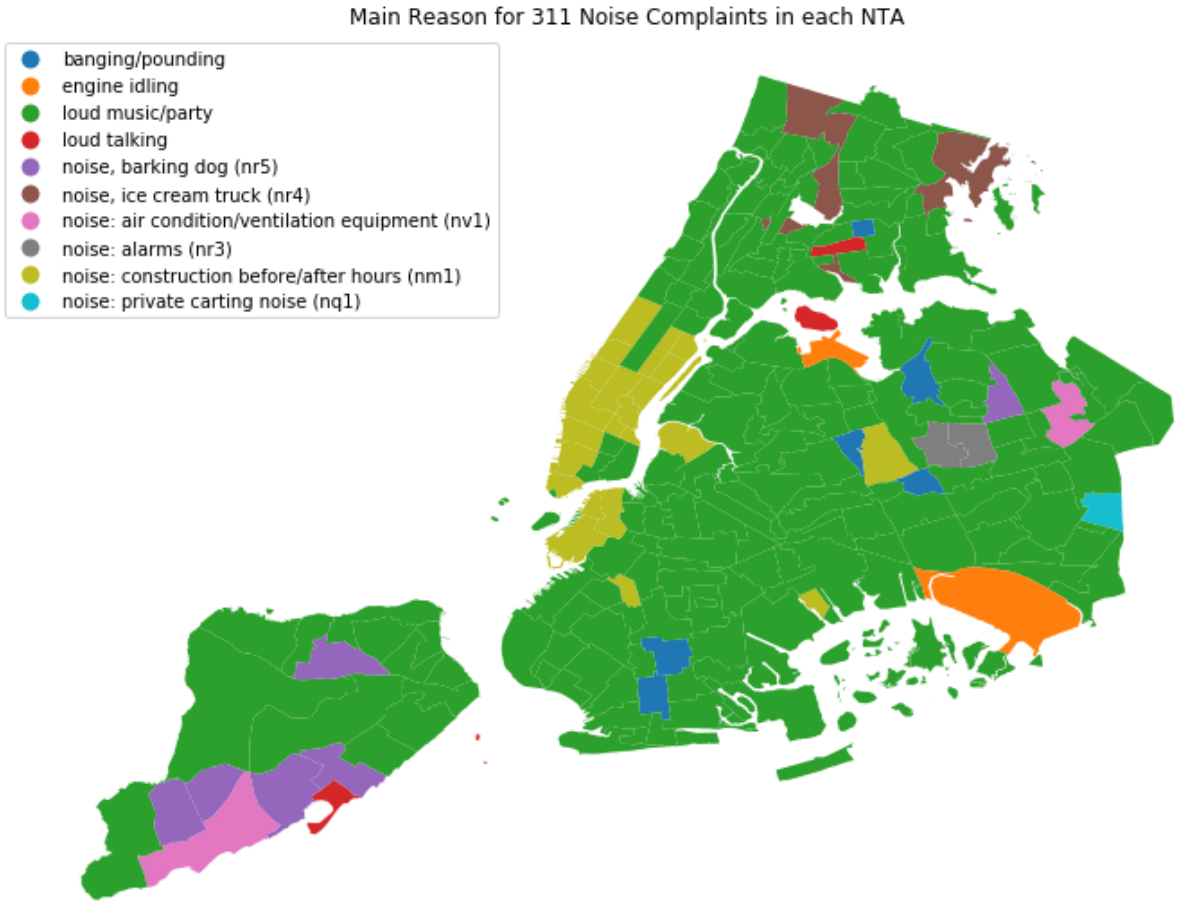

Figure 10: Main type of noise complaints in each NTA

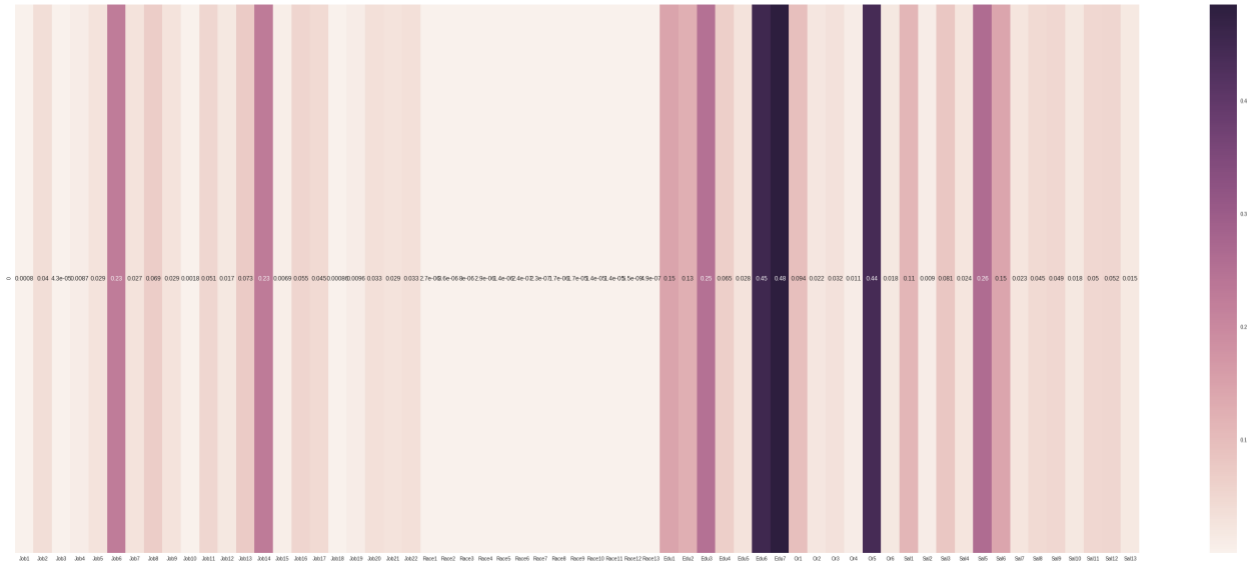

Figure 11: PCA analysis of attributes multiple attributes of Jobs, Education, Salary, and Country of Origin

Appendix B: Tables

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.948
Model:	OLS	Adj. R-squared:	0.930
Method:	Least Squares	F-statistic:	53.10
Date:	Sat, 14 Dec 2019	Prob (F-statistic):	2.51e-34
Time:	20:33:16	Log-Likelihood:	-230.09
No. Observations:	91	AIC:	508.2
Df Residuals:	67	BIC:	568.4
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	57.0529	1.452	39.296	0.000	54.155	59.951
F0	-0.4752	0.018	-26.110	0.000	-0.512	-0.439
F1	-0.0138	0.011	-1.304	0.197	-0.035	0.007
F2	0.0002	0.001	0.210	0.834	-0.001	0.002
F3	-0.0078	0.010	-0.795	0.429	-0.028	0.012
F4	-0.0007	0.001	-1.081	0.284	-0.002	0.001
F5	-0.0041	0.003	-1.525	0.132	-0.009	0.001
F6	-0.0017	0.001	-2.773	0.007	-0.003	-0.000
F7	-0.0038	0.001	-4.029	0.000	-0.006	-0.002
F8	-0.0017	0.002	-1.041	0.302	-0.005	0.002
F9	0.0039	0.002	1.764	0.082	-0.001	0.008
F10	0.0010	0.002	0.466	0.642	-0.003	0.005
F11	0.0028	0.003	1.055	0.295	-0.002	0.008
F12	0.0002	0.001	0.111	0.912	-0.003	0.003
F13	-0.0017	0.002	-1.088	0.280	-0.005	0.001
F14	-0.0002	0.001	-0.218	0.828	-0.002	0.002
F15	0.0050	0.003	1.812	0.074	-0.001	0.011
F16	0.0018	0.001	1.763	0.082	-0.000	0.004
F17	0.0009	0.002	0.573	0.569	-0.002	0.004
F18	-0.0317	0.019	-1.643	0.105	-0.070	0.007
F19	0.0012	0.003	0.448	0.655	-0.004	0.006
F20	-0.0005	0.001	-0.461	0.647	-0.003	0.002
F21	-0.0062	0.002	-2.890	0.005	-0.010	-0.002
F22	0.0015	0.001	2.008	0.049	8.71e-06	0.003

Omnibus:	28.557	Durbin-Watson:	2.178
Prob(Omnibus):	0.000	Jarque-Bera (JB):	54.966
Skew:	1.208	Prob(JB):	1.16e-12
Kurtosis:	5.943	Cond. No.	3.62e+04

Table 1: OLS Regression results for Jobs

[Armed Forces, Unemployed, Agriculture Forestry Fishing Hunting and Mining,Arts Entertainment Recreation Accommodation and Food Services, Construction, Finance and Insurance, Government Workers, Information, Manufacturing, Natural Resources Construction and Maintenance Occupations, Other Services, Production Transportation, Professional Scientific Waste management Services, Public Admin, Self Employed, Warehousing and Transportation, Unpaid Family Workers, Wholesale Traders, With Supplemental Security Income, With Public Assistance Income, With Retirement Income]

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.301
Model:	OLS	Adj. R-squared:	0.181
Method:	Least Squares	F-statistic:	2.512
Date:	Mon, 16 Dec 2019	Prob (F-statistic):	0.00646
Time:	17:25:19	Log-Likelihood:	-404.78
No. Observations:	90	AIC:	837.6
Df Residuals:	76	BIC:	872.5
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	17.8488	27.353	0.653	0.516	-36.629	72.326
F0	-2.1070	6.845	-0.308	0.759	-15.741	11.527
F1	-0.3661	8.621	-0.042	0.966	-17.537	16.805
F2	-14.1316	9.834	-1.437	0.155	-33.717	5.454
F3	20.8268	12.248	1.700	0.093	-3.567	45.221
F4	12.4952	4.943	2.528	0.014	2.651	22.339
F5	-13.7478	6.266	-2.194	0.031	-26.227	-1.268
F6	-8.4375	6.444	-1.309	0.194	-21.271	4.396
F7	13.6909	9.475	1.445	0.153	-5.179	32.561
F8	-10.0785	7.045	-1.431	0.157	-24.109	3.952
F9	5.9055	5.205	1.135	0.260	-4.461	16.272
F10	-0.3268	7.878	-0.041	0.967	-16.018	15.364
F11	-2.3470	6.304	-0.372	0.711	-14.902	10.208
F12	12.5798	8.622	1.459	0.149	-4.592	29.751

Omnibus:	5.137	Durbin-Watson:	0.488
Prob(Omnibus):	0.077	Jarque-Bera (JB):	2.597
Skew:	-0.131	Prob(JB):	0.273
Kurtosis:	2.210	Cond. No.	144.

Table 2: OLS Regression results for Salary

['$100,000 to $149,999', '$100,000 to $149,999.1', '$150,000 to $199,999', '$150,000 to $199,999.1', '$200,000 or more', '$200,000 or more.1', '$25,000 to $34,999', '$25,000 to $34,999.1', '$35,000 to $49,999.1', '$50,000 to $74,999', '$50,000 to $74,999.1', '$75,000 to $99,999', '$75,000 to $99,999.1']

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.683
Model:	OLS	Adj. R-squared:	0.366
Method:	Least Squares	F-statistic:	2.152
Date:	Mon, 16 Dec 2019	Prob (F-statistic):	0.167
Time:	17:53:19	Log-Likelihood:	-34.624
No. Observations:	15	AIC:	85.25
Df Residuals:	7	BIC:	90.91
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-9.5903	18.434	-0.520	0.619	-53.181	34.000
F1	0.0811	0.649	0.125	0.904	-1.454	1.616
F2	1.4282	1.680	0.850	0.423	-2.545	5.401
F3	0.0618	0.698	0.089	0.932	-1.588	1.712
F4	-2.1337	1.078	-1.979	0.088	-4.683	0.416
F5	4.9760	2.060	2.415	0.046	0.104	9.848
F6	0.6146	0.421	1.461	0.187	-0.380	1.609
F7	-0.0650	0.402	-0.162	0.876	-1.015	0.885

Omnibus:	3.779	Durbin-Watson:	1.425
Prob(Omnibus):	0.151	Jarque-Bera (JB):	1.870
Skew:	0.848	Prob(JB):	0.393
Kurtosis:	3.338	Cond. No.	661.

Table 3: OLS Regression results for Education

[Less than 9th grade, 9th to 12th grade, no diploma, High school graduate (includes equivalency)Some college, no degree, Associate's degree, Bachelor's degree, Graduate or professional degree]

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.636
Model:	OLS	Adj. R-squared:	0.362
Method:	Least Squares	F-statistic:	2.325
Date:	Mon, 16 Dec 2019	Prob (F-statistic):	0.134
Time:	17:58:18	Log-Likelihood:	-35.665
No. Observations:	15	AIC:	85.33
Df Residuals:	8	BIC:	90.29
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	11.7688	5.480	2.147	0.064	-0.869	24.407
F1	-0.9681	0.471	-2.056	0.074	-2.054	0.118
F2	0.1364	0.220	0.621	0.552	-0.370	0.643
F3	-1.8451	0.887	-2.081	0.071	-3.890	0.200
F4	9.8895	4.805	2.058	0.074	-1.190	20.970
F5	0.0551	0.127	0.434	0.676	-0.238	0.348
F6	-2.3454	3.527	-0.665	0.525	-10.478	5.787

Omnibus:	0.594	Durbin-Watson:	1.143
Prob(Omnibus):	0.743	Jarque-Bera (JB):	0.330
Skew:	-0.337	Prob(JB):	0.848
Kurtosis:	2.728	Cond. No.	133.

Table 4: OLS Regression results for Population Origin
[Europe, Asia, Africa, Oceania, Latin America, Northern America*]*

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.046
Model:	OLS	Adj. R-squared:	-0.027
Method:	Least Squares	F-statistic:	0.6339
Date:	Mon, 16 Dec 2019	Prob (F-statistic):	0.440
Time:	17:59:48	Log-Likelihood:	-42.878
No. Observations:	15	AIC:	89.76
Df Residuals:	13	BIC:	91.17
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	7.0000	1.170	5.982	0.000	4.472	9.528
PCA0	-0.0384	0.048	-0.796	0.440	-0.143	0.066

Omnibus:	1.115	Durbin-Watson:	0.097
Prob(Omnibus):	0.573	Jarque-Bera (JB):	0.960
Skew:	0.454	Prob(JB):	0.619
Kurtosis:	2.156	Cond. No.	24.2

Table 5: OLS Regression results for multiple attributes of Jobs, Education, Salary, and Country of Origin, collectively as a PCA component