Usage of 311 service across various neighborhoods and socio-demographic groups in New York City. Capstone Project

Diego Garzón*, Juan Pablo Mora†, Yanchao Xu‡, and Nikhil Kishore§
Center for Urban Science and Progress (CUSP)
New York University
*Email: dfg265@nyu.edu † jml1023@nyu.edu ‡ yx1088@nyu.edu § nk1877@nyu.edu

***Abstract*—This paper describes the collaboration with the New York City 311 agency in order to determine the possible relationship between both: socio-economic characteristics of New York City population (both residents and workers), as well as local features of each neighborhood and the number and types of service requests received by the 311 support line. The project scope will include the development of a web-based or mobile visualization tool that can aid the agency to assess the relationships and metrics of their service and the type of requests received within a defined geographical area, and the solution of a two phases machine learning based predictive and classification model. This report will also cover the findings about the relationship between each type of service request and the socio-economic profiles of each area in New York City and the dynamics with neighborhood features.**

I. Introduction

311 is the support service center created by the New York City City government, with the objective of attending all non-emergency inquiries, complaints and service requests from the general public to the local administration. By definition, 311 has a large open database (starting from 2010) that consolidates all of the received citizen requests regarding various topics, ranging from housing, noise and transportation to public health and public services, all of them related to various local government departments. Although some initial approximations have been made to explore potential insights of the data contained in the 311 database, including geographical clustering analysis in order to determine potential hot-spots along the City, theres still various types of analysis that could be performed in order to obtain valuable information. Analysis of 311 database can be useful both for the 311 agency to successfully optimize its operation, as well as to various city agencies that use the information and cases registered by 311 to perform adjustments in order to provide measurements to improve their coverage. A potential analysis that have been targeted as a priority within the 311 agency, is the one aiming to identify the different kind of user profiles segmented by type of service request and geographical area in the City. This kind of analysis would be of great value in order to asses the level of global coverage the service has achieved through the City, and could be useful to identify existing gaps in terms of coverage levels in the services provided, as well as segments of the population that are not being successfully attended. As an answer to that, this project goal is to discover the influence of socio-demographic profile and characteristics of the neighborhood on usage patterns of 311 service requests. In that sense, we think that the ability to link different socioeconomic features such as income, levels of education, housing values, etc, and demographic characteristics such as race, total population, age, etc., of the population living in different areas of the City and the number of service requests 311 received over a specific period of time, grouped by types, would be extremely useful both to 311 as well to other city agencies. Specifically, we aim to determine whether there are different types of users profiles existing per each type of request, and use this analysis as an indicator of the level of universal coverage achieved by the service provided. It is taking into consideration this initial approach that this project has been created. In this research the team will create an analytical model that can generate insights from different data sources containing socioeconomic and demographic features of the population that resides within New York City. We will get this insights by assessing the correlations between these features and the number of service requests received by 311 in a specific period of time, in order to identify certain kinds of user profiles at defined geographical areas. The expectation is that the developed model is not only limited to find these correlations, but can also be a useful tool for generating projections and estimations based on the propensity for 311 service requests. These insights will be based on population features, with the goal of identifying population changes in the City, and also to validate if certain demographic groups are being unattended or have lower probabilities of using the service. Finally, the project aims to provide 311 with a powerful visualization tool that can display intuitive graphical representations about the relationship between population features that are being considered in the analytical model and the number and types of requests by each of the defined areas the team select for the analysis in the City (NTAs or Neighborhood Tabulation Areas). It is also expected that the tool provides a way to easily understand the possible dynamics that currently exist in regards of the volume of 311 service requests and the changes in population in defined areas.

The features that are going to be considered are:

- Income
- Age ranges
- Race
- Proportion of residents vs. households owners
- Single residential units vs. family units
- Car ownership
- Education level
- Housing Price
- Land Use
- Population Density

However, a distintion should be make between resident and incoming population on each area, and demographic features and Local features. One of the distintions that this project want to find is if the service requests are more related to socio demographic characteristics of the population living on a certain area, or to the ones of the population that work in that area. But, since there are some types of requests that can't be explained using just demopgrahic data, a second distintion is adressed for studying the dynamics with local features of each neighborhood.

In the next section, some previous studies and results that are useful to our research are discussed. Then, a complete description of the data, the chosen aggregation and sources is given. After that, the technical methodology is discussed in detail. Sections five and six are focused on the results its implications and limitations. At the end of the paper, the technical and practical conclusions of the study are treated as well as future work and ways to improve the project.

II. Literature review

Up to our knowledge this paper is the first one to estimate the impact of social, demographic and economic characteristics of the population on the 311 service requests. However, a big amount of research has been done so far in order to understand the links between customers behavior and demographic,social, and economic characteristis. The methodology applied in this project have been inspired by some of this studies.

Customer complaint behavior has been one of most important research topics on which both behavioral psychology and marketing have been focusing for the last 40 years. Several academic studies have been deployed to understand some of the features that have a high level of correlation with individual propensities of complaining among different type of individuals, as well as trying to identify the sources of why people decide to complain. However, most of this studies have been focused primarily on relations between private good and service providers and their customers, instead of the relationship between public local government agencies and citizens using their services.

These studies have proven to be useful to the project in terms of providing some insights and possible initial trends in terms of why people decide to complain and what are the most important demographic features that correlate directly with the propensity of service use over time. Particularly in the framework of this project, the team has based part of the methodology proposed and the feature determination selection based on the work of authors such as Andreasen and Best [5], [6] that discuss about the possibility that many of the complaints experienced by customers are not latent and that in fact less than half of them are in fact reported over time.

In that same study, the authors determine that the first identifiable correlation between the propensity to complaint and the individuals comes not from its demographic characteristics, but from the type of product the customer is receiving from the provider (i.e. a product of a service). In that case, the propensity to complain was determined to be higher over time if the good received was a service instead of a product, although a high level of uniqueness is found on a case by case analysis due to the fact that each customer has a different perception and attributes to the product/service she or he is receiving from the provider.

In the same sense, propensity to complain have been studied in terms of the possible correlations with the specific socioeconomic and demographic features of the overall customers that have reported some levels of dissatisfaction with both products and services obtained from the private sector. For those cases, studies performed by Landon [7] and Richins [8] exposed that the correlation between some demographic features such as age, gender, race or income are existent in some cases, and are completely dependent of the type of good and service analyzed. This studies also conclude that there are high variabilities in accordance; however, in general terms, these studies have found that along most categories of goods and services, features like per capita income and education attainments have shown some consistent results over time.

A third component have been identified to be important in terms of propensity to complain is the perceived value of the good and services the individuals are receiving from their providers; its logical to assume that the general propensity of complaints increases as a function of the criticism of the goods and services the individuals receive over time, but authors such as Warland, Hermann and Willitis [9] and Andreasen [4] have shown that this factor can also vary over time depending of the situation and the subjective value that each of the same individual place on each of the goods in a specific period of time. For instance, the importance one individual would place for a specific medical service would not be the same if the request is related to an ambulatory service or if is related to an emergency service.

Our study is deeply inspired by the Urban Lens project developed by the MIT Senseable City Lab (http://senseable.mit.edu/). Even though, it studies the consumer behavior (instead of complaining trends) the techniques used on that project can be extended to our purpose.

The Urban Lens project developed at MIT Senseable City Lab is a novel multi-scale predictive model of Spanish regions, quantifying the distinctive signature of each region based on their spending behavior by identifying indicators regarding the amount of spending, type of spending, demographic groups, and individual mobility.

Urban Lens explores millions of anonymized financial transactions in Spain based on data provided by BBVA (banking transactions). The project is an effort to understand macro trends (consumer habits) from a fine-grained scale of individual economic behavior. That is, the project performs a comparative analysis of city microeconomics, aiming to find how bigger economic patterns could be understood utilizing data of individual economic transactions.

In this large study [10], [11], [12], the authors gathered data from different Spanish cities: demographics, customer activities, average value of transaction, spending diversity, local, and distant mobility. The different geographical areas are clustered according this features, using K-means algorithm, and correlations between them are found. One key point, is the careful treatment that the authors have with multicollinearity issues, overfitting and right scaling of the factors.

Moreover, the authors demonstrated that there are significant deviations from the trends baselines. They attribute this deviations to distinct signatures of cities, forming a solid basis for a scale-free comparison and classification of urban areas.

The authors also believe that although in the paper the approach is applied for the case of Spain, it can be widely applicable to any other country, provided that the appropriate data are available [10] .

III. Data description

In this section, we describe the three main sources of the data and the reasons why we considered those sources.

We use only open data of 311 service requests of the year 2014 posted by the agency in the NYC Open data portal (https://nycopendata.socrata.com/) by mid-June 2016. There are two main reasons for that: first, the final goal of the project is to combine 311 information with socio-demographic information from the U.S census Bureau, and the most updated version of the latter is 2014 estimates. While the reason we are using the mid-June 2016 version of 311 data is that around this time, 311 updated all its historical data to include new types of service requests that weren't being considered before. This version of 311 data includes $1,888,331$ service requests (in contrast, a previous version contained just $1,674,071$ service requests), and 2323 types of service requests (in contrast to 232 types of requests in the previous version). Each service request on the 311 database has its geo-location (latitude and longitude) associated with it.

As stated on the introduction, we want to focus on this aspects of the resident and working population:

- Income
- Age ranges
- Race
- Proportion of residents vs. households owners
- Single residential units vs. family units
- Car ownership
- Education level
- Housing Price

at the Neighborhood Tabulation Area (NTA) level. According to the NYC planning department NTAs have similar boundaries to traditional neighborhoods, however this are not always the same. NTAs were created to project populations at a small area level, from 2000 to 2030 for PlaNYC, the long-term sustainability plan for New York City. Since population size affects the error associated with population projections, these geographic units needed to have a minimum population of $15,000$. This criterion resulted in combinations of neighborhoods that probably would not occur if one were solely designating boundaries of historical neighborhoods. At the end, NTA's are aggregation of census tracts.

We chose NTAs as geographical units for two main reasons: first, the interest of 311 on look at Neighborhood statistics rather than more granular ones. Second, census tracts are too small for an analysis of this type, resulting in higher variablity of the features which reduce significantly the results of regression models in general. There are 195 NTAs in New York City (according to Bytes of the Big Apple: http://www1.nyc.gov/site/planning/data-maps/open-data.page), which is a considerable samplesize.

The team obtained the Social, demographical, and economic characteristics of the population from the American Community Survey (http://www.socialexplorer.com/explore/tables) five years estimates. The American Community Survey (ACS) is an ongoing statistical survey by the U.S. Census Bureau. It regularly gathers information previously contained only in the long form of the decennial census, such as ancestry, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics. These data are used by many public-sector, private-sector, and not-for-profit stakeholders to allocate funding, track shifting demographics, plan for emergencies, and learn about local communities. Sent to approximately $295,000$ addresses monthly (or 3.5 million per year), it is the largest survey after the decennial census that the Census Bureau administers [1].

The five-year estimates version of the ACS uses 60 months of collected data for its statistics. There are several adventages of such version: first, the precision and reliability is higher than the other versions (one-year and three-years estimates), also it contains data for all areas and it is useful for analyzing very small populations [2].

Land use features (such as office, commercial, residential areas) are obtained from NYC department of planning PLUTO data version 16v1 (http://www1.nyc.gov/site/planning/data-maps/open-data/pluto-mappluto-archive.page)

The list of selected features used in the development of the analytical model is divided in two groups: demographic features on each NTA (for resident population and incoming population) and local features (or neighborhood charateristics). Table I shows the demographic characteristics considered in our analysis and their cathegory; the feature cathegory will have an important role in the final model. Table II shows the list of the neighborhood featuers considered in the model.

The third source of data is the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) for 2014 in its more recent version: 7. This dataset comes at the state level and provides, for each

TABLE I
DEMOGRAPHIC VARIABLES USED IN THE ANALYSIS

Population type	Feature group	Feature
Residents	Age	Population under 18
Residents and Workers		Population between 18 and 34
		Population between 35 to 64
		Population 65 and over
	Race	Population white
		Population black
		Population asian
		Population hispanic
		Population other race
	Family or non family	Family households
		Nonfamily households
	Household Education Level	Population education high school or less
		Population education bachelors
		Population education masters
		Population education phd
	Owners vs Renters	Owner occupied units
		Renter occupied units
	Transportation Mean	Car
		Public
		Motorcycle
		Other means
	Household Income	Household income less than 40,000
		Household income from 40,000 to 75,000
		Household income 75,000 and above
	House Value	House value less than 100,000
		House value from 100,000 to 500,000
		House value 500,000 or more
	Rent	Rent less than 1000
		Rent bewteen 1000 and 2000
		Rent 2000 or more

TABLE II
LOCAL VARIABLES USED IN THE ANALYSIS

Local Features
Median House Value
Median Age
Median Rent
Median Income
Cars Per Capita
Building Density
Percentage of commercial area
Percentage of Retail area
Percentage of office area
Percentage of residential area
Inbound Commute Density
Population Density
Percentage of Tenant population

pair of census blocks, the number of people that commute to work from the first census block to the second one according to the census. We used this dataset to produce estimates of the socio demographic features from the previous list, for the working population on each NTA.

IV. METHODOLOGY

A. General Approach

The following list of global topics reflects the basic layout the capstone team defined for the execution of this project:

1) Data exploration and initial predictive model construction: On this initial phase the team estimated a first approximation to the agencys database, using a sub sample of the data (2014 full year), as well as selected external demographic features obtained from the American community survey (ACS) and NYC Pluto database. With this information, the team obtained a first draft version of a predictive model using a machine learning algorithm called constrained linear regression (CLR); this technique was used because it offers the possibility of selecting the best set of features (i.e. those who have the highest explanatory capacity) from the complete list. Once this list of selected features was obtained, the basic levels of correlations between the features, and the number of service requests was defined.
2) Request specific model construction: After the first phase of the project was completed, and the pilot results were shared with the agency, the team proceeded to refine and select the number and types of features that were going to be used in the analytical model. The goal of this step was to move from a single general based model, to various request-specific models that allowed the measurement of the relationships between the selected features and specific types of requests. The team compared the number of service requests per NTA and per type against the selected group of features based on the results obtained in the first steps of the project.
3) Robust model construction: Once the results of the general and specific models were defined, the team applied other types of machine learning techniques (to be fully explained in section 2.3. Technical Methodology) to generate additional insights and improved results regarding classification and prediction of the number and types of service requests using selected demographic attributes.
4) Models analysis and profile creation: After the robust model was proven to be functional and effective, the team proceeded to establish the user profile for each type of service request. That means, given a certain type of service request, a profile of the typical new yorker that requests that type of service is established based on the socio demographic features. These profiles were constructed based on the coefficients obtained from the analytical regressions (previous step).
5) Visualization tool development: The development of the visualization tool started at the same time as the initial data exploration phase, in order to incorporate the milestones and partial results produced in each of the stages of the project. In that sense, the visualization tool construction was a dynamic process that received continuous feedback from the analysis performed and meetings held with the sponsoring agency. The visualization tool aims to display in an intuitive manner a combination of four things:

- First, the summary of the demographic features used

in the construction of the analytical model
- Second, the number of requests per type and per geographical location, as published by 311 open data.
- Third, the number of requests per type and per geographical location, as predicted by our model, based on the chosen demographic features.
- Fourth, the type of service that a certain type of user is more likely to request from the agency. That is, given a certain combination of socio demographic features (or profile) for a hypothetical user, the type of request that this person would probably ask to 311.

Project report and presentation: The final deliverables of this project are going to be: 1. The technical report -a document that explains in great detail the framework, techniques and methodology used in the project, 2. The final presentation -that will serve as a summary of the other tasks performed during its scope, and 3. This Report - that aims to inform to 311, and other city agencies, the principal findings and applications of our research. The team expects that these deliverables serve as a user guide for the agency, an input for future projects to be developed between 311 and CUSP, and a summary of all the task and finding the team achieved during the scope of the project.

B. Data Aggregation

Each service request on the 311 database has its geo-location (latitude and longitude) associated with it. We used this geo-location to aggregate the number of 311 service requests by type at the NTA level, using the NTA location provided in Bytes of the Big Apple. The result of this step is a table where each row represents an NTA and each column represents a type of service request. Then, each cell is the number of complains of the type of request given by the column at the NTA given by the row in 2014.

In the previous subsection, the demographic and economic variables to be analyzed were listed. We can consider each variable as a different group of population (for example: group of people with higher income, group of people on the given age group). Each of these groups is obtained in each area for two different population categories: resident and working population. That is, at the end we have two datasets of variables:

1) Dataset containing the features associated with the resident population in each area, that was extracted directly from the ACS for each NTA within the city.
2) Dataset containing the features associated with the working population in each area, that was extracted from the ACS using the geographical areas defined as origins for working population coming into New York City using the LEHD origin and destination database.

The second dataset was obtained in eight steps:

STEP 1: LEHD origin and destination database contains two parts: a 'main' file(information about workers in New York State with residence in New York State), and an 'aux' file (information about workers in New York State with residence outside the state). Each of this has three columns that we are going to use

- *h_geo*: census block of residence
- *w_geo*: census block of workplace
- *S000*: number of people that lives in *h_geo* and works at *w_geo*

STEP 2: Since the dataset of the previous step contains information about New York State, it is neccesary to filter it just for *w_geo* census blocks inside New York City

STEP 3: Since *h_geo* and *w_geo* are census blocks, they are aggregated by census tracts using Bytes of the Big Apple. For simplify our notation in this paper, we will keep naming *w_geo* and *h_geo* to the resulting census tracts on this step.

STEP 4: Since the *h_geo* from the 'aux' file are outside New York State (and it represents approximately 20% of the total of workers, it was neccesary to download the ACS variables, as detailed in the previous section, associated with each of this census tracts outside the city.

STEP 5: From ACS, we get the collection of $k = 1, ..., l$ the socio-demographic features (population groups) detailed on the previous section for each census tract h_ geo_j. Lets name each of this features as: $Pc_k(\text{ h\_ geo }_j) := Pc_{k,j}$, and $\bar{Pc}_{k,j}$ its normalized version. Then, $\bar{Pc}_{k,j}$ is the probability that a person living in the census tract h_geo_j. belongs to the group k. The result of this step is a probabilities matrix.

STEP 6:

Let w _ geo_i be the geocode of the working place (where $i = 1, ..., m$), and h _ geo_j be the geocode of the residence place (where $j = 1, ..., n$), the LEHD data structure can be seen as follows:

$$\text{w\_ geo}_i \begin{cases} \text{h\_ geo}_1 & n_{i,1} \\ \text{h\_geo}_2 & n_{i,2} \\ \dots & \dots \\ \text{h\_ geo}_m & n_{i,m} \end{cases}$$

where $n_{i,j}$ is the number of workers commuting from h_ geo_j to w _ geo_i

STEP 7: In this step, we merge the results from the previous two steps: we multiply the probabilities in each h_ geo from *STEP 5* by the number of workers in that h_ geo . The result is for each w_ geo the estimated number of workers commin from each h_ geo that belongs to each group.

In a formal way: the goal is to get an approximate of each of this features for each w_ geo_i. Let $F_k(\text{ w\_ geo }_i)$ be the approximate number of workers which workplace is in w_ geo_i, in the k socio-demographic cathegory.

Then,

$$F_k(\text{ w\_ geo }_i) = \sum_j \overline{Pc_{kj}} * n_{i,j}$$

for each w_ geo_i and for each socio-demographic indicator k.

STEP 8: The final step is the aggregation from census tracts, in the previous result, to NTAs. The result of this step is a table where each row represents an NTA and each column represents a feature or population group. Then, each cell is the estimated number of workers that belong to the group given by the column in each NTA given by the row.

C. The mathematical model

After the team obtained the datasets containing the counts of the different types of service requests and the demographic features for each defined geographical area in the city, a matrix that summarizes all of this existing correlations between the number of service requests per capita and demographic features was obtained. This initial analysis was performed as a first step to identify possible existing user profiles.

However, the goal of the study is not simply to understand the correlations of the features and the number of 311 service requests but to understand how this different variables interact together and, with this, identify a profile of the typical 311 user for each type of request. In order to get that result, it is neccesary to apply a multilinear regression.

Specifically, the model works as follows:

For each NTA location A, let the the number of service requests of each type t be $s(A,t)$. Also let its residential population be $P(A)$, while the incoming commuter population be $C(A)$.

Now, for each of the considered socio-economic feature group q (like age, in table I) lets enumerate its possible values by $1, 2, , s_q$ (for example for Owner vs Renters $s_q = 2$ and we can denote owners as 1 and renters as 2).

Let $P_q(A,k)$ be the number of residents for which the value of the quantity q equals to k and $C_q(A,k)$ be the corresponding number of visitors.

Then,

$$P(A) = \sum_{k=1}^{s_q} P_q(A,k)$$

and

$$C(A) = \sum_{k=1}^{s_q} C_q(A,k)$$

for any A and q.

Assume that a number $SR_{q,k,t}$ of service requests of type t made by an arbitrary person in a place of residency with a value k is a normally distributed random variable with a mean $x_{q,k,t}$ and a standard deviation $\sigma^x_{q,k,t}$. Similarly assume that a number $SC_{q,k,t}$ of service requests of type t made by an arbitrary person in a destination of his/her daily commute is a normally distributed random variable with a mean $y_{q,k,t}$ and a standard deviation $\sigma^y_{q,k,t}$. Then for each quantity q, type of service request t and location A one can have the following estimates for $S(A,t)$:

$$S(A,t) = \sum_{k=1}^{s_q} P_q(A,k)x_{q,k,t} + \sum_{k=1}^{s_q} C_q(A,k)y_{q,k,t} \quad (1)$$

For each t and q, we consider the system of equations (1) (for all A treated as different observations) as linear regressions. Then, the problem in the next steps is to provide estimates for $x_{q,k,t}$, $y_{q,k,t}$, as well as the corresponding standard deviations $\sigma^x_{q,k,t}$ and $\sigma^y_{q,k,t}$ (obtained based on the width of the 95% confidence interval which for the normal distribution is approximately 4 standard deviations).

The next subsection will analyze further how we solved the system (1) for every type of request and using the variables included in tables I and II. Solving this kind of problem is not trivial, since incorporating demographic variables and local variables in the same model should be done with caution. The steps that we follow can be summarized as:

1) DEMOGRAPHIC FEATURES ANALYSIS: OLS regressions, with cross validation, for each type of request t and each group of demographic features q in table I. Each of this regressions provides a different R squared value. The mean from this list of R squared values is taken and if it's greater than 0.2, this combination of t and q is taken into account for the next steps.
2) then, a new OLS regression is performed, this time with the whole dataset as training, for combinations of t and q that satisfy an R squared values is taken and if it's greater than 0.2.
   From this step we obtain the coefficients $x_{q,k,t}$, $y_{q,k,t}$, as well as the corresponding standard deviations $\sigma^x_{q,k,t}$ and $\sigma^y_{q,k,t}$ (obtained based on the width of the 95% confidence interval which for the normal distribution is 4 standard deviations).
3) Obtain the estimated number of requests by type as a weighted average of the estimations in the previous regression. This estimation would be noted as $\widehat{S(A,t)}$
4) Estimation of the relative error of $\widehat{S(A,t)}$
5) Study of Spatial Autocorrelation of the error calculated in the previous step using Moran's I (some cases are reported on the Annex 4 at the end of this document)
6) LOCAL FEATURES ANALYSIS: A second regression is performed for taking in to account local variables in table II, following the general form of:
   $$\log(S/\widehat{S(A,t)}) \sim \log(\text{ Spatial Features })$$
   where $\widehat{S(A,t)}$ is the estimation obtained in the regression with the demographic features.
   For this regression, cross validated Lasso regression was used for features selection and then OLS regression was performed for those cases with high R2 with the whole dataset as training

D. System Solution

As stated in the previous subsection, we first analyzed the dynamic between 311 requests and the demographic features in the table I. That means, for each pair t and q we solved the linear regressions (1) obtaining the estimates for $x_{q,k,t}$, $y_{q,k,t}$,$\sigma^x_{q,k,t}$ and $\sigma^y_{q,k,t}$.

Cross-validation (with 10 splits of locations into training and validation sets) was performed and the average preformance for all splits is reported as an average cross-validation $R^2(t,q)$ using OLS regression.

OLS is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the differences between the observed responses in some arbitrary dataset and the responses predicted by the linear approximation of the data (visually this is seen as the sum of the vertical distances between each data point in the set and the corresponding point on the regression line - the smaller the differences, the better the model fits the data). The resulting estimator can be expressed by a simple formula, especially in the case of a single regressor on the right-hand side [13].

For each t we will further consider only those q, having satisfactory $R^2(t,q)$.

Now for each t, once we have the estimates for $x_{q,k,t}$, $y_{q,k,t}$,$\sigma^x_{q,k,t}$ and $\sigma^y_{q,k,t}$ for every q such that $R^2(t,q)$ is satisfactory, formulae (1) allows to produce an estimate for $S_q\widehat{(A,t)}$ (that means, an estimated number of service request based on each group of variables q).

Averaging those estimates over all suitable q we can get a final estimate $S\widehat{(A,t)}$ together with its standard deviation $\sigma_(A,t)$ (that is just the sum of variances $\sigma^x_{q,k,t}$ and $\sigma^y_{q,k,t}$) . For this average, we give more weight to those combinations of q and t with lower standard deviation, so this follows the formulae:

$$S\widehat{(A,t)} = \frac{\sum_q S_q(A,t)\sigma_{q,t}^{-2}}{\sum_q \sigma_{q,t}^{-2}} \tag{2}$$

Once we have averaged estimates $S\widehat{(A,t)}$ we can compare them with the ground-truth and look if the relative error $error(A,t) = \frac{\hat{S}(A,t)-S(A,t)}{\hat{S}(A,T)}$ still has any correlation with spatial context, that is the variables listed in II

In order to answer the question about spatial correlation of error (that means, clusters or groups of NTAs with similar error as calculated in (2)), Moran's I algorithm will be used. Moran's I analyzes the correlation of a variable with itself throughout space. Moran's I is used for finding in which extent the occurrence of an event in one area makes more likely or unlikely the occurrence of an event in a neighboring areal unit. In other words (using this example): if there is a NTA with a high error calculated in (2), how likely, or unlikely, is that the surround NTAs have a similar error?. Specifically, the null hypothesis to be evaluated is that the variable of interest (relative error) is randomly distributed across the study area (New York City) with a significance level of 0.05.

The starting point for Moran's I autocorrelation test is to get an adequate matrix W of spatial weights. In this study the Queen Contiguity Matrix is used. A Queen Weights Matrix defines a location's neighbors as those areas with shared borders or vertices (in contrast to a Rook Weights Matrix, which doesn't include the vertices). For every NTA i, the elements w_{ij} of the matrix W are equal to zero except for those j census tracts that are local neighbors of i according to the Queen Weights Matrix, in that cases $w_{ij} = 1/n$ where n is the number of neighbors for the census tract i.

For instance, in Figure 1 the eight neighbors for the NTA with GeoID equal to 36061009600 are represented, then if $i = 36061009600$, $w_{ij} = 0.125$ for each one of i's neighbors j and $w_{ij} = 0$ for the rest of NTAs.

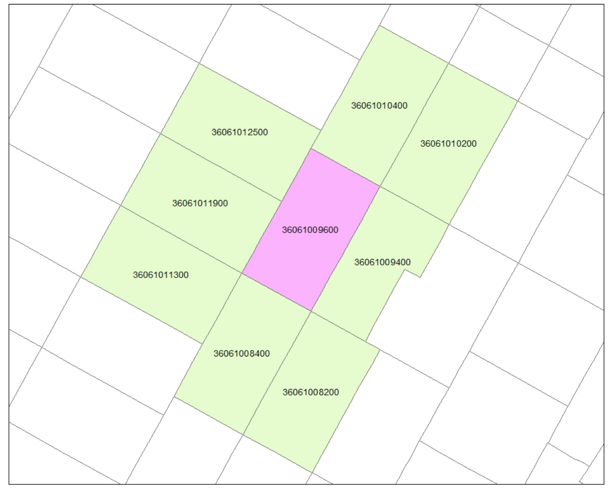


Fig. 1. *Representation of the eight neighbors for the hypothetical NTA with GeoID equal to* 36061009600 *using the Queen Contiguity Criterion*

The Morans I index value follows the formula:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \tag{3}$$

where N is the number of spatial units indexed by i and j (in this case, the total number of NTAs), w_{ij} are the values of the matrix W of spatial weights calculated on the previous step, x is the variable of interest (relative error) and $\bar{x}$ is its mean.

A positive Moran's I index value indicates tendency towards clustering, while a negative Moran's I index value indicates tendency towards dispersion.

In order to verify the statistical significance of the test, a p-value based on the number of permutations is consulted and then compared to the significance level α. If the p-value is less than, or equal to the significance level, then the observed data of the variable is inconsistent with the null hypothesis and this is rejected. However, that does not prove that the tested value hypothesis is true, this test guarantees that a type I error (incorrect rejection of a true null hypothesis) rate is at most α.

Once we have analyzed the spatial autocorrelation of the relative error, we understand if service request behavior can be described just based on the properties of the population, or specific local context still matters on top of that.

In case some of the spatial context features really matter we use an additional regression that follows

$$\log \frac{S(A,t)}{S\widehat{(A,t)}} \sim \sum_i \log(LF_i(A)) \tag{4}$$

where $LF_i(A)$ is the value of the local feature i in the NTA A

For getting strong results of (4), we performed cross validation and feature selection using the LASSO algorithm.

In statistics and machine learning, the Lasso (least absolute shrinkage and selection operator) is a multilinear model that minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant λ. Because of the nature of the constraint, some of the coefficients associated to the variables are exactly 0 which means that those variables have not a meaningful influence over the dependent variable. Hence this model works as a feature selector [3].

After applying the Lasso technique for feature selection, we applied Ordinary leat squares regression on each type of request using the whole dataset as training and just the variables with coefficients different than zero after the feature selection as regressors.

For the Lasso model, a division of the total database was performed in order to define three instances of the data: a training data set, a validation data set and a test data set (70

The model was trained with the data provided by the first instance and then, using the second instance, the parameter λ of the regression was calibrated to obtain the best performance in terms of fit and accuracy to finally perform the regression in the test set; this same procedure was repeated for all of types of service requests.

One of the problems of using the Lasso, is that there is not a simple way to perform a significant test for the resulting coefficients. For our purpose, then, the Lasso regression is just an exploratory tool for understanding the behavior of the variables and its influence over the normalized volume of service requests.

To overcome this issue, we perform Ordinary Least Squares (OLS) regression for each type of service request over each group of variables (after the Lasso feature selection).

Then, OLS regressions are performed for each type of service request. In each case, the log of different features sets act as independent variables while the log of the rate between the volume of requests and the volume of requests as predicted by the demographic variables, is the dependent variable to be fitted. We filter the results of each regression for those with a R-squared value bigger than 0.2.

The result of this experiment is a set of different properties that describes better the profile of the New Yorker that request a given type of service from 311.

E. Web Application

Web application development is carried out in 2 phases. In the first phase, the purpose of the web tool is to visualize the data processed in the previous subsections. This helped in analyzing various features of demographics and 311 service request data. Next phase included implementation of analytical logic explained above for quantifying the impact of demographics and other data on 311 complaints. Implementation of the statistical model in the application is part of phase 2. Currently sufficient work has been done in both the phases. Beta version of our web tool is ready and available online for use. It might be possible to find some bugs in the tool since it is still in its beta phase. Implementation and development of the tool is explained below in 2 parts frontend and backend. Frontend explains the design and visualization methodology while backend explains how data sources are used and connected with front end for the development of the web tool.

Front End Initially layout and functionality of the website were finalized using tools like wix.com, Adobe Photoshop, rough sketches and continuous feedback from 311 department. Once designs and structure for the application was finalized, it was programmed in HTML5, Bootstrap and CSS to build the primary foundation for the web tool. HTML5, Bootstrap and CSS provided the outline structure for the website. It built the space for map and map navigation section(Inspector) comprising of 4 tabs namely options, data, Model and Team. Next Map background layer and NYC map was added using Cartodb and leaflet JavaScript library. In following step each tab of inspector section is assigned functionality using JavaScript and Jquery. Map is connected to the back end and inspector section for providing visualization and navigational functionality. Dropdown menu options like 311 Data Visualization and Demographics Data in option tab were populated with all the 311 complaint and demographics attributes. Next they were connected with map using JavaScript to visualize chosen attributes. Data is calculated as percentages by default for visualization of colors. Although user also has option to visualize map by providing the chosen percentile value in option tab. Feature whose value is less than 10% of the percentile value is given the lowest shade of color. Intensity in shade of color increases from value ranging between 30 - 40%, 40 - 50%, 50 - 70% to 70-90% and values greater than the chosen percentile threshold value. Analytics section also connected to map using JavaScript, D3 JavaScript Library and JQuery. This helps in visualizing top complaint types and demographic attributes like income, households, population etc. Height of bar charts is scaled to percentages for comparing difference between multiple types of complaints. Option tab further provides user flexibility to normalize the complaint or demographic features by population and area. User can also visualize data by their raw values. Other features in option tab includes choice to change background layer from 3 different layer options. This feature is mostly useful if user is sensitive to bright screens. Locations could also be zoomed for better visibility.

Data is the second tab in the inspector section and provides quick insights about the chosen location. If any location is clicked, then data section provides top 20 complaints of that area and some demographic information. Bar chart display complaints and when mouse is hovered over them they show frequency and type of the complaint. Bar charts are also clickable and visualizes the chosen complaint type in the map section. Section also provides quick facts about location including like its per capita income, frequency of total 311 service request, median age and median household income. More demographic information could be scrolled down in the inspector tab. Pie charts provide information on the age classification and number of people who own house vs people who rent. This information helps in understand the situation of any chosen location. It is also a very important part of the project as it suggests demographic profile of people who make certain service requests.

Model tab is the third tab in the inspector section and also one of the most important. It consists of our statistical model which helps in predicting which demographic features are more likely to make certain complaints. It came from the analysis explained above resulting in rich dataset comprising beta coefficient, intercept and predicted calls information. It provides user with a neat form and requests for various demographic information. User is free to provide as much or as less information he wishes. Based on the provided demographic attributes tool would narrow down and visualize a bar chart showing service request which someone with given demographic attributes would most likely make. Tool would visualize 2 bar charts suggesting difference in service requests made by resident and service request made by a someone who is working. Again bar charts are clickable and user can click the service request on bar chart for visualizing it on the map. This provides quick navigational facility for visualizing complaint type.

The Model tab uses the coefficients $x_{q,k,t}$, $y_{q,k,t}$ found as solutions of the system (1) and the known number of residents (and workers) $P_q(A,k)$ $(C_q(A,k))$, in each demographic cathegory

$\widehat{S(A,t)}$ is calculated for all the service requests matching the requirement chosen by the user at the model tab, including the value of the intercept of the regression (1). Next it ranks them in descending order to suggest which request type is most likely to affect the chosen demographic properties. This process is repeated for both residents and workers independently for calculating their respective service request behavior.

Team section is fourth and the last tab on the inspector section. It provides information about the project, project team, project mentor and the sponsor agency.

Similarly, map section of the web application provides information tab at the top left corner of the screen. It provides title of attribute chosen for visualization of the map i.e. attributes of 311 service requests and demographic features. In next line it provides value of chosen attributes followed by Frequency. It also provides information about the location ID, Borough name, and year from which the data is taken. Information tab is activated by simply hovering mouse over the map. Other option include zoom in and zoom out button at the top left corner of the screen. User can also reset position of the map by clicking reset map button. This feature is mostly useful if you navigate somewhere out of the scope where map is not visible. Full view button provides user option to see map in full screen by hiding the inspector tab. User can access back the inspector tab by clicking the Full View button again. [14]

Back End Multiple data files has been used in the backend of the tool. Geojson files of census tract and neighborhood level (NTA) has been used for visualizing the map. Other data files used includes demographics, 311 data files, LEHD data file. All data is processed and cleaned for the purpose of the project. D3.js and JavaScript has been used throughout the application including its implementation in backend and front end. It is attempted to combine as much information on single csv or geojson file so webtool doesnt lag when it is overloaded with lot of data. Various internally developed datasets have been used which comprise of information on model like beta coefficient values, Predicted service request data set etc.

V. Results and implications

Annex number 1 summarizes the results obtained from the first step of the analytical model, where regressions on each type of request were performed using specific groups of demographic features, for both residents and workers independently. These results were filtered based on the capacity of the features to explain the number of request per type (R squared value) . Odd numbered tables in Annex 1 contain information related to the impact that each of the resident features included in the regressions have on the types of request (coefficients). The tables show the confidence interval the model estimated for each of them, and the overall level of R squared for each regression as well; even tables, contain the features from the working population.

A sample of this results were included in form of visualizations in Annex number 3, in order to provide an intuitive form of presenting the information, so the basic concepts can be more easy to understand.

Annex number 2 summarizes the results obtained from the second step of the analytical model, where regressions on each types of request were performed using neighborhood characteristics, and the results from the previous step (demographic characteristics); these results were filtered based on the capacity of the features to explain the number of request per type (R squared)

For those features having positive coefficient, an additional increase of the service requests is likely to be observed, compared to what can be expected based on the population socio-demographic analysis, and in that same sense, those having a negative coefficient - facilitate decrease of the service requests.

In order to provide some guidance on who to use the results presented in Annex 1 and Annex 2, a sample of request types were taken as base examples to understand how they work, and what type of valuable information that can be extracted from them.

Based in the analysis performed, Noise Residential is the type with the higher number of service requests in 2014, followed by Heat/Hot Water and Street Condition So lets see what insights can derived about these types of requests:

The age of the resident population has a capacity to explain the number of received Noise Residential requests complaints through all NTAs, represented on the column R2 os of table 1. This R squared value is the highest in that table (0.544) which means that the selected demographic variable do a better job explaining the number of requests for this type than for any other type. One reason for that, might be that since Noise Residential is the most requested type of service in the dataset, the model had more data to learn trends and draw conclusions.

However, this is not a general rule since none of the working population variables was able to explain the volume of requests of this type.

Surprisingly, the results obtained showed that the population older than 65 is less likely to request this type of service from 311. However, it is imperative to avoid any interpretation without observing first at the dynamics of this population group over other types of requests: for almost every type in table 1, the category of population 65 year old and over has a negative coefficient associated with it. This could mean that, in general, people in areas with a high population of this age group have a low propensity of using 311 services when compared to younger people.

Household Income of Residents show a similar trend in terms of R squared and explanatory capacity (table 7). In this case, request type Noise Residential shows the second higher R squared among other types of service requests (after general Noise request types). From this category it is also noticeable that areas with high concentrations of households with income below $40,000 per year are positively correlated with this type of requests, while wealthier population groups are negatively correlated; in fact, the coefficient associated with low income households is the highest one for that category, closely followed by the Heat/how Water requests.

Another interest relationship can be found in table 9; in this case, the resident population identified as tenants have a positive coefficient, while homeowners have a negative one. This same relationship can be seen on a couple more of requests, such as heat/hot water, heating, flooring stairs, a very intuitive result, as all of this request are directly related to housing conditions.

The results shown in the population race category (table 11) brings also some insightful information: the first noticeable trend is that among all race profiles for residents, the asian population has a negative correlation for almost every type of service request, and even in cases where positive correlation was found the coefficient is close to zero if we consider its confidence interval (For instance the case of Sidewalk Condition request, where the coefficient is equal to 0.00015). For all cases where coefficients with these levels were obtained, it is most likely that all of them are not statistically significant, which means that there is not a certainty of the variable having a positive or a negative impact because of its marginality on the overall explanatory capacity of the model.

If we look for our main example Noise residential request in table 11, we can see that black, hispanic and other races populations are highly positive correlated with this type of request (compared to population white that has a coefficient that is almost a third and a fifth of those obtained for black and hispanic population respectively). Once again, we see that types of requests related to housing conditions, such as Heat/how water, Heating, Electric, Door Window, among others, show similar trends; in the case of the race, this types of requests have a higher relationship with black and hispanic populations.

At this point an important trend can be pointed out: tenants (as opposed to homeowners), living in zones with the lowest household income category (household income less than $40,000 per year), belonging to black and hispanic groups are more likely to request services related to housing conditions.

In that same sense, another example can be seen related to monthly rent levels paid by tenants (table 13), where some request types share similarities with each other; Door/Window, Electric, Heat/Hot Water, Heating, Noise-Residential, Plumbing, Unsanitary Condition, and Water Leak request types present a high positive high coefficient on low rent values (below $1.000 per month), a low positive low coefficient with rents between $1.000 and $2.000 per month, and a negative coefficient with high rents (considered above $2.000 per month).

From this results, it is possible to interpret that residential units with lower rent values are usually accompanied by potentially faulty general conditions, such as broken windows, frequent power or heating failures, noisy neighbors, or relatively higher unsanitary living conditions.

It is important to notice that none of the categories related to housing conditions (with the exceptions of Heat/how water, and heating) are well explained by the local variables (table 19 in annex 2), and even in those cases, the only category that makes a good job at explaining the volume of requests is the population density.

In relation with the Street condition type of request, general analysis had not been performed using the first step of the analytical model, even when its one of the top three types in terms of volume in the city, because the explanatory capacity of the selected demographic variables (tables 1 through 18) are not as high as the one other types of request have. From all demographic categories just four make a good job explaining the volume of requests for this type: House Value for residents (table 5), homeowners v.s. Tenants for resident population (Table 9), residents race (table 11), and the principal transportation mean for residents (table 15)..

Of the four, the last category comes not as a surprise, as places with high volumes of car transit have a higher probability of being positively correlated with this type of service request; moreover, the only transportation category that is negatively correlated is public transportation, reinforcing the general perception that places with high car presences will have more requests associated with Street condition.

Since the demographic variables did not show a high explanatory capacity for the number of Street condition requests, the local variables defined in the second part of the analytical model made a good complementary input. Table 19 in annex 2 shows that all of the selected local variables have an impact on this type of request. A couple of insights that can be derived from this table are that, there is a high volume of requests related to this type in areas with high proportions of retail and office built up areas, while there is a negative correlation with variables such as proportion of residential built up area and population density, and proportion of tenants.

Other types of service request, such as Air Quality, allow for some additional interpretation of the results obtained from the analytical mode.;

Air Quality requests for instance show a very unique behaviour in terms of its relationship with the age ranges of residents, being among the few ones that have a positive correlation with the range age above 65 years old, with a rather moderate result of R squared (0.39) for the whole age category.

Whats is really interesting is to see that the same category for the incoming working population is negatively correlated, perhaps because the size of that particular group is very small when compared to the other two ones.

Additionally, this type of request allow us to see the behaviour of the information related to the education levels of both residents and incoming working population (tables 3 and 4); for the first group it is interesting to note that only the groups with a bachelors and masters degree seem to be positively correlated to the number of request received, with a very high explanatory values of the category (R squared equal to 0.58); this could imply that well educated people are more concern about this type of issues along the city, discounting obviously PhD graduates from the group, as they are a very small portion of the overall population.

This hypothesis can actually be complemented by looking into the information provided by the resident's household income level and levels of monthly rent tenants pay; in the first case, households with a high per month income (above $75.000) have the highest coefficient of the three resulting ones, with the highest explanatory capacity on the category base on the results of the R squared (0.602); for the second case, tenants that pay the highest rent values per month (more than $2.000) have also the highest coefficient with a very high explanatory capacity of the category (R squared of 0.627).

Finally, when analyzing the transportation mean used by the resident group and its relationship with this type of request (shown on table 15), it is noticeable that all groups other than car users (motorcycle, public transportation and other means including biking and walking) have positive correlations, with an high explanatory capacity (R squared of 0.597); this results seems to be logical, as they could be more exposed to exterior conditions that allows them to notice problems in air quality.

All of this results can be actually summarized in a user profile having high income, high levels of education, with ages ranging from 18 to 34 or more than 65 years, that is a tenant paying a high monthly rent price (over $2.000), living in a non-family household, and that uses typically any kind of transportation mean other than car to commute to work, as the most propense to complaint about Air Quality issues.

Finally, we can make some general observations that are applicable through the whole set of results. For instance, features related to the incoming working population have impact on less request types than the resident characteristics. On Annex 1 (Even numbered tables), we can see that in average, the number of request types that meet the criteria of R squared values higher than 0.2 are significantly lower that the one related to residents (Represented on odd numbered tables).

One noticeable exception of the previous point is found in table number 5 (Coefficients and confidence intervals associated with the groups of Resident Homeowner House Values), where it can be seen that this specific group have a significant impact just on a few types of request categories (Damaged Tree, Dead Tree, Overgrown Tree/Branches, and Root/Sewer/Sidewalk Condition,etc.).

The team estimates that this is caused mostly due the nature of this variable; since it is counting just the number of homeowners, seeing a significant correlation between these features and these types of requests (mostly about neighborhoods with green areas in residential zones) is natural. Moreover, in all these cases, it can be seen that the number of homeowners living in houses with higher values are positively correlated with the number of requests for this request types.

In general, the model seems to adjust well depending on the nature of the type of request and its relationship with the defined group of demographic characteristics; for instance in table 16, which contains impact of the working populations transportation means towards 311 request, the type Taxi Complaint that has a R square of 0.704 (high explanatory value), the coefficient obtained by the category commuting by public transit (0.05221) is noticeable higher of those obtained for the other available categories (with the exception of transportation motorcycle, that has a higher coefficient but is not statistically significant according to its confidence interval).

In this case, inference about the propensity of the working population to request a 311 service about this issue can be done on the fact that workers using public transportation, shown by the model, are more likely to request the service, whereas the people commuting by their own vehicles are not, which turns out to be a very intuitive result.

Finally, there are some interesting relationships that can be made from the overall annex; for instance, we can see that asians among the other races are less likely to request 311 services, or that in general, resident population over 65 years old has a low propensity (negative correlation) with the number of request received per type by 311, that highly educated residents, with degree levels between High school and masters, have a higher propensity to request 311 services (excluding PhD graduates), or that low household income levels are in general more correlated to a higher number of requests.

Annex 4 Shows the distribution of the error between the observed number of requests and the number of requests explained by the demographic variables for five different types of requests. It is noticeable that the error in Areas such as the Bronx and some parts of Queens is high for Sidewalk Conditions and Noise. This is caused because the neighborhood dynamics in these areas is having an important effect over the number of 311 requests, and we are not capturing this effects on the first steps of the model.

The same thing could be said about Staten Island for Noise-Residential and Heat/How Water requests. Since Staten Island has certain unique characteristics (high residential population, high road infrastructure, etc.), the first stages of the model are not as accurate at explaining the number of requests for those types based merely on socio-demographic features, so its necessary to look at neighborhood features.

In the web application side, the tool is fully implemented and contains the following featuers:

- Visualizing NYC Map at census tract level
- Visualizing NYC Map at NTA level
- Bar chart for top 20 service request calls of the chosen location.
- Visualizing 68 demographics attributes on map
- Visualizing 180 complaint types on map
- Visualizing clickable bar Charts for complaint type
- Visualizing clickable bar charts for Predicted complaint type
- Visualizing histogram for Total population, Median Age, Income Per Capita, Median Households
- Visualizing Pie chart for visualizing
- Visualizing 4 different background layers
- 2 Normalization option (Population and Area)
- Visualizing map colors based on given percentile value

Figures 2 and 5 provides an overview of the current capabiilities of the web tool, while figure 3 shows how the selection of the demographic characteristics (or user profile) works. After selecting the user profile, the tool displays the types of services that a person of this characteristics request on average as seen on figure 4.

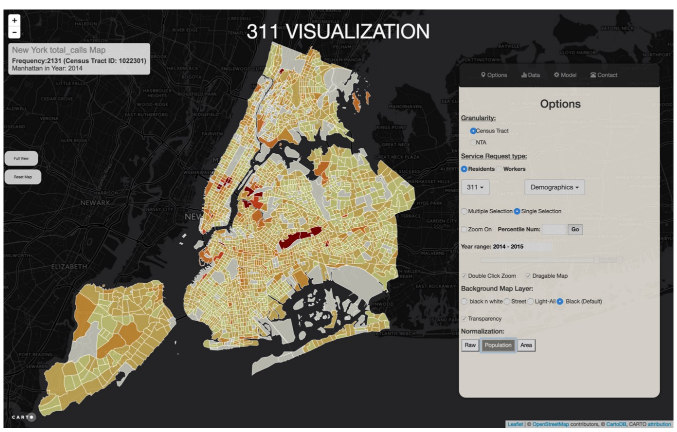


Fig. 2. *Overview of the webtool. On the right side there are some options for visualization: the demographic parameters, the types of service requests, the type of normalization, and the granularity*

VI. Limitations

This section includes some potential limitations and fairness problems that could arise from the use of certain datasets and the implementation of the proposed framework, impacting its results and validity. In this sense, the team has grouped this problematics on to 4 defined topics as follows:

First, 311 service request are not entirely correlated to specific types of problems in certain areas of the city:

The first problem that the team identified in the early phases of the analysis was the fact that the 311 service requests could not be entirely correlated to the real-world problems

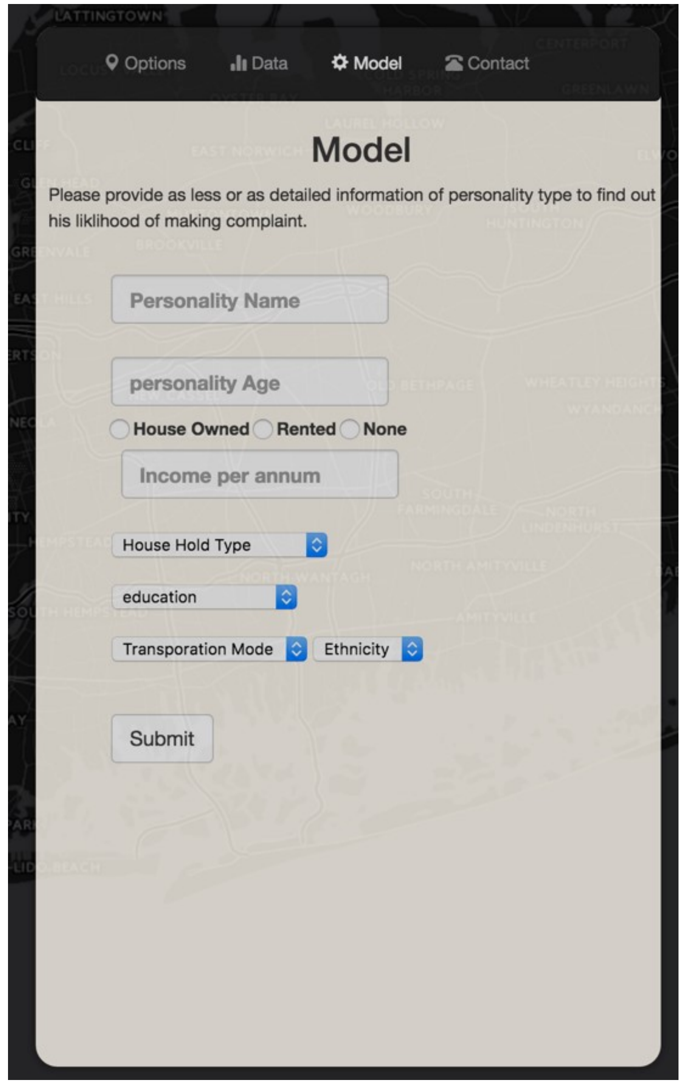


Fig. 3. *Selection of the socio demographic characteristics or user profile*

that are taking place in the City, and that are supposed to be representative of, for instance, some possible mismatching problems have been noticeable in some of the categories since a great portion of the incoming working population is complaining at their working locations, rather than home locations.

Since the list of 311 service request itself is geo-located based on the specific site where the service request was opened, analysis performed on this data would be inconsistent in the sense that it would not be possible to determine the characteristics of the person that actually is making the complaint but only the features of the area where the complaint was made, even if they are not related.

Additionally, making a complaint is related to subjective decision, so it will depend mainly on the complainers tolerance towards a real-world problem. For instance, peoples tolerance towards noise could be greater in Manhattan when compared to Staten Island, because individuals assume that the levels of noise are generally higher in average in Manhattan than

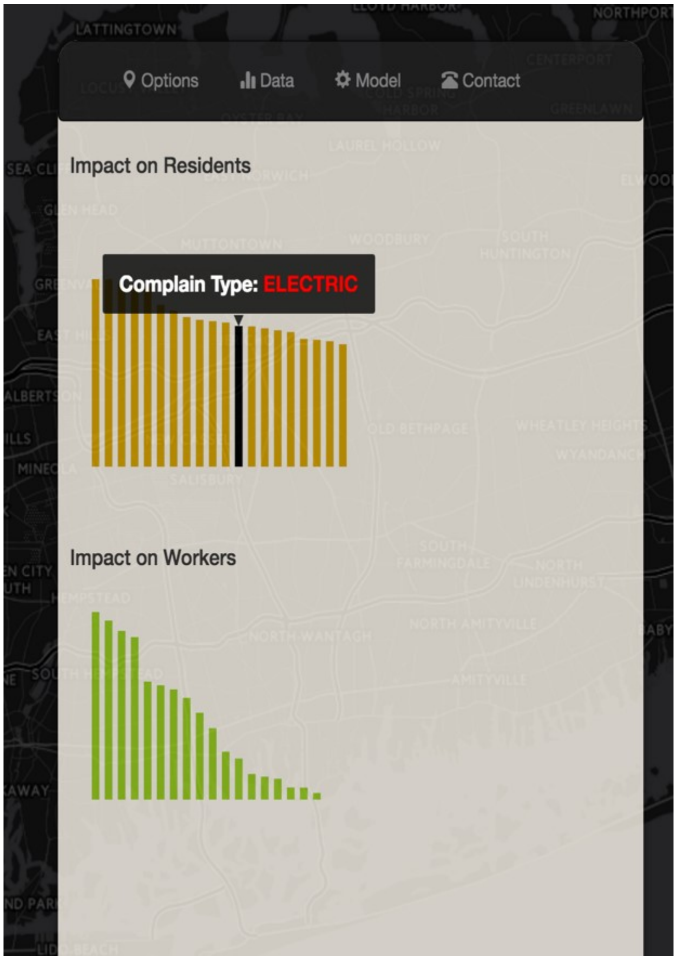


Fig. 4. *After selecting the user profile, the tool displays the types of services that a person of this characteristics request on average*

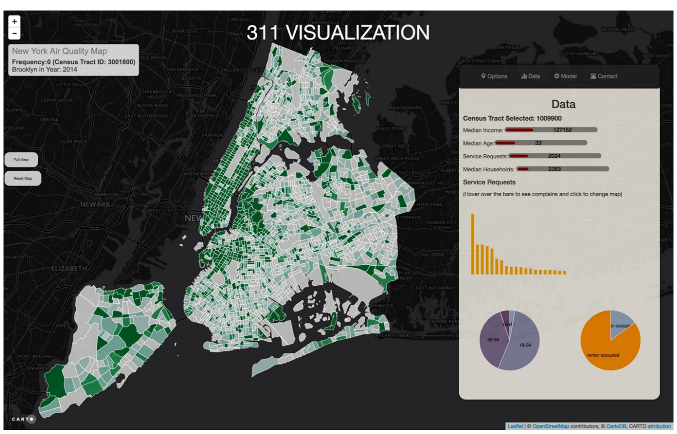


Fig. 5. *The tool offers different types of normalization for visualize the data and the results in a dynamic way*

in Staten Island, although it could be the case that any given day, the numbers of noise complaints from Manhattan and Staten Island are the same; however, it is hardly a persuasive argument that the noise levels of Manhattan and Staten Island are the same over time.

Furthermore, service requests are only one of the factors reflecting real-world problems in the City. When looking into 311 complaints data, it could be a common situation that these requests are not strongly correlated to specific types of problems, since real-world problems are usually complicated and hard to explain and fit into a single category. Hence, 311 data could have limited power to reflect certain problems, and it may underestimate the analysis and results to be produced by this project -311 database could be a good indicator to explain how things go in the city to some extent, but it is certain that it is far from a perfect indicator.

Second, 311 service request data potentially includes levels of skewness in relation with users and geographical areas in the city:

One of the most important objectives of the project, defined in conjunction with the sponsoring department, is the identification of different User profiles defined as a group of demographic features that together can summarize the propensities certain groups of New York City inhabitants have to use 311 to file different types of complaints.

In that sense, one of the expected outcomes of this project by the agency is to validate the level of universal coverage of the service they are providing through the entire City, as their stated mission is to provide the public with quick, easy access to all New York City government services and information while maintaining the highest possible level of customer service; for this purpose the idea of linking a selected group of predefined demographic and economic features obtained from the US Census Bureau American Community Survey (ACS), and the number of service requests grouped by type of call and geographical area seems like a potential and simple way to identify these profiles with an acceptable level of accuracy.

One of the main concerns of the local administration is that the service provided by the 311 agency is limited or contained to some specific areas of the city (especially Manhattan), and its reach is more effective in areas where residents have higher levels of income, higher levels of education and are predominantly white; this could imply that the service itself is uneven along all residents of the city, and that the levels of universal reach and equity wanted by the agency re not being achieved.

Taking this into consideration, one of the problems faced by this project is the use of the 311 service request data base as one of the inputs to construct an analytical model that could result in a tool aimed to identify and define the desired demographic profiles of the users; the problem itself lies in the possibility that if the unevenness in the coverage provided by 311 is in fact existent over time, then potentially all of the analysis performed with the methodology defined by the team is going to just reproduce and represent this inequity in the access when the demographic profiles are modeled based on the skewness of the data to some specific geographical areas of the city or certain types of individuals residing in those areas, creating a limitation in the fairness the model aims to provide.

If the model (in relationship with its design and methodology) itself is not able to guarantee to the agency a certain level of equity and fairness in the results, or at least identify and indicate the levels of underlying bias or skewness in the data as a results of the existing limitations in the current levels of coverage, then the usefulness of it will be seriously compromised and the results and insights drawn from it could not be used as a tool to design policies aimed to improve the conditions of the services of some of the agencies that work with 311 in the city using their database as an input for their own analytical models.

Third, some assumptions in the definition of the demographic characteristics of the commuting workers could result in under/over estimations of the analytical model:

One of the most ambitious parts of this project is related to the estimation of the potential propensity of incoming commuting workers that are not living in the New York City metropolitan area to file certain types of 311 complaints based on their demographic characteristics, and to construct a demographic profile for this population in a similar manner that the project aims to do with the resident population.

For this objective, the use of the Longitudinal Employer-Household Dynamics (LEHD) database has been defined, in order to establish the existing pairs of origin and destination locations for workers that reside outside the 5 boroughs of the city (i.e. New Jersey, Connecticut, Long Island, etc.) but that commute every day to work in them; with the information obtained from this database, the team has aimed to determine the geographical zones on which different incoming workers reside, to use them in conjunction with ACS data to extract their socio economic and demographic features and their destinations in the city to link them with the types of service request received by 311 in each of the Citys areas.

Even though this process seems to be rather straightforward, given the fact that the team has access to all required databases to perform it successfully, through the construction of the projects methodology several complications have arisen given the apparent difficulty to correctly estimate the actual level of demographic features directly related to the incoming working population, that represents only a fraction of the total population that resides in the areas that are defined as origins by LEHD.

To facilitate this measurement, the team has defined a simple process that uses a ponderation of the social economic and demographic features displayed by the origin geographical areas of the incoming workers, based on the proportion of the incoming working population over the total population for each area; the problem with this approach however, is that although allows to have a series of estimated values that can be used by the analytical model to estimate the propensity of complaint of this incoming population and some levels of demographic profiles, the results are not completely accurate given the fact that there is not a complete certainty that the ponderation is actually overestimating or underestimating the demographic characteristics of the working population from the overall population in their origin areas.

This issue represents a great challenge for the proposed objectives that the team has defined for the project, in the sense that the accuracy of the model for the incoming working population component is not entirely assured, and that given the assumptions that the team was required to take in this matter could result in some bias in the model that could result in overrepresentation or underrepresentation certain types of profiles (by income, race, education levels) affecting the level of equality displayed in the final results and deliverables.

Fourth, 311 service request data could include bias related to complaining patterns because of their human nature:

Finally, the team has encountered a final risk in terms of the limitations the data and the analytical model derived from it could have; overall, the definition of contacting 311 for service requests is inherently a personal decision related directly to an individual behaviour, so 311 data may be inevitably biased.

Since the 311 service request data is based on complaints coming from several platforms such as NYC 311 mobile application, NYC 311 call center and online complaint tool, every individual who wants to complain about something could potentially make a complaint to NYC 311, and it will be recorded without any kind of limitation. In that sense, the data and the complaining patterns behind it, which we are trying to figure out, could be biased due to its underlying human nature, and factor that are beyond reasonable understanding or analytical modeling.

Complaint is directly related in one sense or another to reporting issues that surpasses the individuals levels of tolerance; however, not everyone have the same levels of tolerance towards the same occurrence. It could be a common case that people are used to a blocked driveway in Manhattan, but at the same time other people may probably call to 311 to complain a blocked driveway in Staten Island. Also, tolerance levels could be largely affected by the human moods. For example, a white-collar worker would have higher possibility to make a noise complaint about his neighbors party after a busy day, because he is exhausted and wants to sleep early.

It is hard to measure human nature and peoples mood. Thus, our data would probably cause unfairness and may be lack of key information for our analysis.

Facing the human nature of the 311 complaining patterns, the team plans to analyse the socio-economic characteristics of the complainers, which is a main objective of the project. We believe the socio-economic traits of a certain geo-location could be helpful in explaining the complaining patterns in this area. Although we could not keep out the impact from peoples mood, we are measuring the complaining pattern of a person based on income, age, educational level and other demographic features. We are taking a up-down approach to analysing the complaining pattern of an area as a whole based on its demographic characteristics and then try to predict and infer the complaining behaviour of an individual.

On the web application side, there are two potential issues: First, adding lot of information and visualization may slow down the web tool. This may cause higher lags and loading time. Second, web tool may not provide information on census tracts which are parks, airports, or other private or government owned space where data is not sufficient or available.

VII. Conclusion and Future Work

After the analysis of the results and implications on the previous sections, the team have found the following conclusions:

- In term of the geographical grouping units, NTAs provided the most accurate level of aggregation through the development of the project, guaranteeing the best possible results in terms of the analytical model due to their size and homogeneity in terms of demographic and spatial features.
- The final analytical model developed by the capstone team showed statistically significant results related to the explanatory capacity and goodness of a fit of the features (both demographic and spatial) included in it for a high range of request types (more than 110 for the first step and more than 20 for the second step).
- In terms of the population groups used in the model, the resident group showed better results in terms of the number of types of request that successfully explained with a high level of accuracy when compared to the incoming working group, possibly due to the difference in proportions between them.
- Demographic features of the resident population such as age range, household income, owners vs tenants,race and transportation mode use to commute had the highest average capacity to explain a larger set of service request types; in contrast, features such as Education levels, home values and type of household showed the lowest average explanatory capacity.
- The methodology designed to measure the impact of the demographic features in a first phase of the model, to then assess the impact of the spatial features proved to be a highly successful way to increase the overall accuracy of the model in terms of goodness of a fit, allowing for complaints that were not accurately explained by the demographic feature to be assessed in a better way by the use of spatial ones.
- Spatial autocorrelation analysis, which can be seen for some types of requests in Annex 4, was a way to see that in some cases, the selected demographic variables werent able to explain the volume of requests by themselves. Then, another set of regressions was conducted in order to find new ways of explaining the volume of requests using local features this time.
- Even though more than twenty types of requests were successfully explained by the local features, more neighborhood variables will be included in the future, as well as other methods of feature selection in this phase.
- The process of grouping the different types of demographic features into sub categories was definitive and highly helpful in order to successfully generate particular types of User profiles for each type of request, as seen on section 3.1 of this report.
- Due to the high volume of results generated as a final output, as well as the potential to apply this methodology and the analytical model to other 311 historical data (previous to 2014), it is highly recommended that the work initiated with this project is continued in the near future jointly by CUSP and 311.

VIII. Authors Contributions

- Diego Garzón contributed with data analytics, algorithm calibration, algorithm implementation, and report writting (data and methodologies, results and conclusions)
- Juan Pablo Mora contributed with data analytics, algorithm calibration, report writting (project framework, limitations, results and conclusions), and visualizations.
- Yanchao Xu contributed with data analytics, visualizations and report writing (limitations)
- Nikhil Kishore contributed in the design, development and implementation of the web application, and writting about the tool on the report.

IX. Acknowledgments

The team would like to thank to Prof. Stanislav Sobolevsky at NYU- CUSP, for being the mentor of this project and for the guidance that he provided during its process. The team would also like to thank NYC 311 for being the sponsor of this project. In particular, we want to thank Chenda Fruchter, Senior Director at 311, and her team. Without their support and feedback the conclusion of this project wouldnt be possible.

References

[1] US Census Bureau. ”ACS Information Guide”. www.census.gov. p. 8. Retrieved 2015-07-31.

[2] U.S. Census Bureau. A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know. Accessed August 31, 2011 from http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf.

[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc B., Vol. 58, No. 1, pages 267-288).

[4] Andreasen, A.R. (1988) ’Consumer complaints and redress: what we know and what we don’t know’, in The frontier of research in the consumer interest. Proceedings of the International Conference on research in the consumer interest (Racine, Wisconsin, August 16-19, 1986). American Council on Consumer Interests, p. 675-722.

[5] Andreasen, A. R. and Best A. (1977) ’Consumers complain: does business respond?’, Harvard business review, 55 (July-August), pp. 93-101.

[6] Best, A. and Andreasen, A.R (1977) ’Consumer response to unsatisfactory purchases: a survey of perceiving defects, voicing complaints, and obtaining redress’, Law & society review, 11 (Spring), pp. 701-742.

[7] Landon, E.L. (1980) ’The direction of consumer complaint research’, Advances in consumer research, 1980, 7, p. 335-338.

[8] Richins, M.L. (1985a) ’The role of product importance in complaint initiation’ in Hunt, H.K. and Day, R.L. (eds) Consumer satisfaction, dissatisfaction and complaining behavior. Combined Proceedings papers from the eighth Conference: Louisiana, State University, Baton Rouge, Louisiana, March 8-10, 1984 and the ninth Conference: Phoenix, Arizona, February 20-21 1985. Bloomington: Foundation for the School of business, pp. 50-54.

[9] Warland, R. H., Hermann R.O. and Willits, J. (1975) ’Dissatisfied customers: who gets upset and who takes action, Journal of consumer affairs, 9 pp. 148-163.

[10] Sobolevsky, S., Sitko, I., Combes, R. T. D., Hawelka, B., Arias, J. M., & Ratti, C. (2015). Cities through the Prism of People’s Spending Behavior. PLOS ONE, 11(2): e0146291

[11] Sobolevsky, S., Sitko, I., Grauwin, S., Combes, R. T. D., Hawelka, B., Arias, J. M., & Ratti, C. (2014). Mining urban performance: Scale-independent classification of cities based on individual economic transactions. arXiv preprint arXiv:1405.4301. . Fifth ASE International Conference on Data Science in Stanford, CA, May, 2014

[12] Sobolevsky, S., Sitko, I., Tachet des Combes, R., Hawelka, B., Murillo Arias, J., & Ratti, C. (2014, June). Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. the case of residents and foreign visitors in spain. In Big Data (BigData Congress), 2014 IEEE International Congress on (pp. 136-143). IEEE.
[13] Ltkepohl, H. (2001). Journal of Economic Literature, 39(3), 905-907. Retrieved from http://www.jstor.org/stable/2698320
[14] Data Visualizations for the web by Scott Murray, OReilly (2013)

ANNEX 1

Coefficients and standard deviations associated with the demographic variables for types of requests with R2 greater than 0.2

Table 1. Coefficients and standard deviations associated with the groups of RESIDENTS AGE for types of requests with R2 greater than 0.2

Type of Request	Population under 18	population between 18 and 34	population between 35 to 64	population 65 and over	R2 os
APPLIANCE	0.00538 +/- 0.00099	0.00358 +/- 0.00110	-0.00044 +/- 0.00149	-0.00496 +/- 0.00242	0.408
Air Quality	-0.00485 +/- 0.00067	0.00610 +/- 0.00075	-0.00101 +/- 0.00101	0.00470 +/- 0.00165	0.39
Animal Abuse	0.00127 +/- 0.00054	0.00168 +/- 0.00060	0.00092 +/- 0.00081	-0.00176 +/- 0.00131	0.389
Bike/Roller/Skate Chronic	-0.00043 +/- 0.00007	0.00059 +/- 0.00007	-0.00028 +/- 0.00010	0.00065 +/- 0.00016	0.362
Boilers	0.00015 +/- 0.00016	0.00087 +/- 0.00018	-0.00013 +/- 0.00024	0.00010 +/- 0.00040	0.33
Consumer Complaint	-0.00445 +/- 0.00162	0.00670 +/- 0.00180	0.00337 +/- 0.00243	-0.00470 +/- 0.00396	0.359
DOOR/WINDOW	0.01396 +/- 0.00283	0.01132 +/- 0.00315	-0.00096 +/- 0.00425	-0.01491 +/- 0.00692	0.409
Dirty Conditions	0.00193 +/- 0.00167	0.00066 +/- 0.00185	0.00945 +/- 0.00250	-0.00949 +/- 0.00408	0.442
Drinking	-0.00027 +/- 0.00010	0.00046 +/- 0.00011	0.00041 +/- 0.00015	-0.00082 +/- 0.00024	0.411
ELECTRIC	0.01509 +/- 0.00272	0.00812 +/- 0.00303	0.00137 +/- 0.00409	-0.01562 +/- 0.00666	0.445
Electrical	-0.00017 +/- 0.00055	0.00124 +/- 0.00062	0.00103 +/- 0.00083	-0.00200 +/- 0.00136	0.32
FLOORING/STAIRS	0.01029 +/- 0.00232	0.00940 +/- 0.00258	-0.00091 +/- 0.00348	-0.01194 +/- 0.00567	0.365
Food Establishment	-0.00494 +/- 0.00068	0.00527 +/- 0.00076	0.00048 +/- 0.00102	0.00202 +/- 0.00166	0.404
GENERAL	0.00790 +/- 0.00258	0.01225 +/- 0.00287	0.00008 +/- 0.00388	-0.01092 +/- 0.00632	0.385
GENERAL CONSTRUCTION	0.00950 +/- 0.00174	0.00520 +/- 0.00194	-0.00080 +/- 0.00262	-0.00900 +/- 0.00427	0.318
General Construction/Plumbing	-0.00452 +/- 0.00347	0.00805 +/- 0.00386	0.00827 +/- 0.00522	-0.01184 +/- 0.00850	0.39
HEAT/HOT WATER	0.03906 +/- 0.01311	0.04966 +/- 0.01458	0.00323 +/- 0.01971	-0.04367 +/- 0.03207	0.313
Indoor Air Quality	-0.00248 +/- 0.00075	0.00429 +/- 0.00084	-0.00113 +/- 0.00113	0.00281 +/- 0.00185	0.332
Indoor Sewage	0.00050 +/- 0.00012	0.00025 +/- 0.00013	0.00015 +/- 0.00017	-0.00038 +/- 0.00028	0.373
Lead	-0.00087 +/- 0.00020	0.00071 +/- 0.00023	0.00059 +/- 0.00031	0.00059 +/- 0.00050	0.338
Litter Basket / Request	-0.00076 +/- 0.00022	0.00114 +/- 0.00024	0.00064 +/- 0.00032	-0.00033 +/- 0.00053	0.34
NONCONST	0.00446 +/- 0.00083	0.00333 +/- 0.00092	-0.00034 +/- 0.00124	-0.00357 +/- 0.00202	0.452

Type of Request	Population under 18	population between 18 and 34	population between 35 to 64	population 65 and over	R2 os
Noise	-0.04855 +/- 0.00536	0.05827 +/- 0.00597	-0.01030 +/- 0.00806	0.03335 +/- 0.01313	0.448
Noise - Commercial	-0.02564 +/- 0.00412	0.04855 +/- 0.00459	-0.00838 +/- 0.00620	-0.00064 +/- 0.01009	0.385
Noise - Residential	0.03233 +/- 0.01245	0.07796 +/- 0.01385	0.02492 +/- 0.01871	-0.10231 +/- 0.03046	0.544
Noise - Street/Sidewalk	-0.00588 +/- 0.00461	0.03500 +/- 0.00512	-0.00596 +/- 0.00692	-0.00921 +/- 0.01127	0.32
Noise - Vehicle	-0.00272 +/- 0.00145	0.01090 +/- 0.00161	-0.00035 +/- 0.00218	-0.00379 +/- 0.00355	0.432
OUTSIDE BUILDING	0.00032 +/- 0.00013	0.00042 +/- 0.00014	0.00017 +/- 0.00019	-0.00046 +/- 0.00031	0.378
Other Enforcement	-0.00119 +/- 0.00037	0.00150 +/- 0.00041	0.00109 +/- 0.00055	0.00038 +/- 0.00090	0.39
PAINT - PLASTER	0.00671 +/- 0.00132	0.00442 +/- 0.00147	-0.00044 +/- 0.00199	-0.00744 +/- 0.00324	0.36
PAINT/PLASTER	0.02754 +/- 0.00669	0.02178 +/- 0.00744	0.00283 +/- 0.01006	-0.03277 +/- 0.01637	0.337
PLUMBING	0.02980 +/- 0.00601	0.02096 +/- 0.00669	0.00121 +/- 0.00904	-0.03122 +/- 0.01472	0.393
Rodent	0.00155 +/- 0.00150	0.01206 +/- 0.00167	-0.00124 +/- 0.00226	-0.00257 +/- 0.00367	0.482
SAFETY	0.00322 +/- 0.00062	0.00300 +/- 0.00069	-0.00017 +/- 0.00094	-0.00412 +/- 0.00152	0.453
Sanitation Condition	0.00698 +/- 0.00145	-0.00214 +/- 0.00161	0.00711 +/- 0.00218	-0.00913 +/- 0.00355	0.309
UNSANITARY CONDITION	0.02716 +/- 0.00531	0.01941 +/- 0.00591	0.00106 +/- 0.00799	-0.02719 +/- 0.01300	0.407
Urinating in Public	-0.00013 +/- 0.00004	0.00024 +/- 0.00004	0.00002 +/- 0.00006	0.00002 +/- 0.00009	0.329
WATER LEAK	0.01380 +/- 0.00286	0.00961 +/- 0.00318	0.00024 +/- 0.00430	-0.01406 +/- 0.00700	0.377
Water Quality	-0.00013 +/- 0.00008	0.00026 +/- 0.00009	-0.00011 +/- 0.00012	0.00089 +/- 0.00019	0.332
Water System	-0.00278 +/- 0.00237	0.00045 +/- 0.00263	0.01617 +/- 0.00356	-0.01887 +/- 0.00579	0.359

Table 2. Coefficients and standard deviations associated with the groups of WORKERS AGE for types of requests with R2 greater than 0.2

Type of Request	population between 18 and 34_n	population between 35 to 64_n	population 65 and over_n	R2 os
Air Quality	0.00937 +/- 0.00282	0.00175 +/- 0.00459	-0.02355 +/- 0.01046	0.303
Broken Muni Meter	0.05466 +/- 0.02052	-0.04805 +/- 0.03345	0.04235 +/- 0.07614	0.378
Consumer Complaint	0.00464 +/- 0.00471	0.01312 +/- 0.00767	-0.04577 +/- 0.01746	0.355
Fire Alarm - Reinspection	0.00053 +/- 0.00015	0.00063 +/- 0.00024	-0.00286 +/- 0.00055	0.426
Fire Safety Director - F58	0.01812 +/- 0.00242	-0.02014 +/- 0.00394	0.02974 +/- 0.00897	0.446
Food Establishment	0.00805 +/- 0.00224	0.00375 +/- 0.00365	-0.02648 +/- 0.00830	0.424
Food Poisoning	0.00492 +/- 0.00081	0.00032 +/- 0.00132	-0.01041 +/- 0.00301	0.443
For Hire Vehicle Complaint	0.00229 +/- 0.00065	-0.00028 +/- 0.00106	-0.00332 +/- 0.00241	0.431
Found Property	0.00167 +/- 0.00020	-0.00111 +/- 0.00033	0.00013 +/- 0.00076	0.582
Homeless Encampment	0.01087 +/- 0.00180	-0.00150 +/- 0.00293	-0.01742 +/- 0.00666	0.334

Type of Request	population between 18 and 34_n	population between 35 to 64_n	population 65 and over_n	R2 os
Homeless Person Assistance	0.00653 +/- 0.00083	-0.00306 +/- 0.00136	-0.00365 +/- 0.00309	0.442
Non-Residential Heat	0.00129 +/- 0.00055	0.00090 +/- 0.00090	-0.00476 +/- 0.00206	0.365
Taxi Complaint	0.06131 +/- 0.00607	-0.03676 +/- 0.00990	-0.00746 +/- 0.02254	0.653
Taxi Report	0.00203 +/- 0.00018	-0.00211 +/- 0.00029	0.00251 +/- 0.00066	0.585

Table 3. Coefficients and standard deviations associated with the groups of RESIDENTS EDUCATION LEVEL for types of requests with R2 greater than 0.2

Type of Request	population education high school or less	population education bachelors	population education masters	population education phd	R2 os
Air Quality	-0.00170 +/- 0.00052	0.00393 +/- 0.00133	0.00664 +/- 0.00278	-0.00641 +/- 0.00755	0.58
Animal Abuse	0.00420 +/- 0.00051	0.00176 +/- 0.00129	-0.00148 +/- 0.00271	0.00180 +/- 0.00734	0.38
Bike/Roller/Skate Chronic	-0.00032 +/- 0.00005	0.00063 +/- 0.00013	-0.00004 +/- 0.00028	-0.00011 +/- 0.00077	0.464
Blocked Driveway	0.05942 +/- 0.00482	0.02734 +/- 0.01229	-0.01583 +/- 0.02571	-0.14484 +/- 0.06979	0.495
Bus Stop Shelter Placement	0.00009 +/- 0.00002	-0.00011 +/- 0.00006	0.00059 +/- 0.00013	-0.00088 +/- 0.00034	0.339
Derelict Vehicles	0.01062 +/- 0.00104	-0.00277 +/- 0.00265	0.00646 +/- 0.00554	-0.02315 +/- 0.01503	0.349
Dirty Conditions	0.01471 +/- 0.00148	0.00515 +/- 0.00376	0.00478 +/- 0.00788	-0.02049 +/- 0.02138	0.485
Food Establishment	-0.00106 +/- 0.00062	0.00852 +/- 0.00159	0.00012 +/- 0.00332	-0.02152 +/- 0.00901	0.449
General Construction/Plumbing	0.00543 +/- 0.00323	0.01827 +/- 0.00824	-0.00275 +/- 0.01725	-0.03617 +/- 0.04681	0.359
Hazardous Materials	0.00011 +/- 0.00020	0.00052 +/- 0.00051	0.00399 +/- 0.00106	-0.01072 +/- 0.00289	0.411
Illegal Parking	0.02311 +/- 0.00343	0.00716 +/- 0.00874	0.05178 +/- 0.01828	-0.16923 +/- 0.04962	0.403
Illegal Tree Damage	0.00008 +/- 0.00014	0.00166 +/- 0.00035	0.00067 +/- 0.00074	-0.00703 +/- 0.00200	0.454
Lead	0.00044 +/- 0.00015	-0.00069 +/- 0.00038	0.00460 +/- 0.00079	-0.00203 +/- 0.00214	0.58
Litter Basket / Request	0.00062 +/- 0.00019	0.00030 +/- 0.00049	0.00315 +/- 0.00103	-0.00373 +/- 0.00279	0.382
Noise	-0.02835 +/- 0.00411	0.07528 +/- 0.01048	0.01078 +/- 0.02193	-0.16906 +/- 0.05954	0.579
Noise - Helicopter	0.00007 +/- 0.00020	-0.00169 +/- 0.00051	0.00523 +/- 0.00106	0.00601 +/- 0.00288	0.3
Other Enforcement	0.00118 +/- 0.00031	0.00333 +/- 0.00080	0.00143 +/- 0.00167	-0.01091 +/- 0.00453	0.524
Recycling Enforcement	0.00031 +/- 0.00007	0.00039 +/- 0.00018	0.00022 +/- 0.00038	-0.00110 +/- 0.00102	0.318
Sanitation Condition	0.01440 +/- 0.00120	0.00166 +/- 0.00305	0.01206 +/- 0.00639	-0.05400 +/- 0.01735	0.411
Sidewalk Condition	-0.00063 +/- 0.00030	0.00272 +/- 0.00075	0.00453 +/- 0.00158	-0.01596 +/- 0.00428	0.536
Water Quality	0.00023 +/- 0.00007	0.00017 +/- 0.00018	0.00055 +/- 0.00037	0.00045 +/- 0.00100	0.367

Table 4. Coefficients and standard deviations associated with the groups of WORKERS EDUCATION LEVEL for types of requests with R2 greater than 0.2

Type of Request	population education high school and less_n	population education bachelors_n	population education masters_n	population education phd_n	R2 os
Air Quality	0.00198 +/- 0.00158	0.05178 +/- 0.01150	-0.14104 +/- 0.02528	0.27909 +/- 0.04717	0.322
Broken Muni Meter	0.01602 +/- 0.01104	0.30194 +/- 0.08023	-0.87860 +/- 0.17639	1.97821 +/- 0.32916	0.48
Consumer Complaint	-0.00381 +/- 0.00257	0.10949 +/- 0.01869	-0.24214 +/- 0.04109	0.30080 +/- 0.07667	0.394
EAP Inspection - F59	-0.00256 +/- 0.00024	0.00524 +/- 0.00174	-0.00074 +/- 0.00382	-0.01621 +/- 0.00712	0.63
Fire Safety Director - F58	-0.00836 +/- 0.00103	0.01885 +/- 0.00746	-0.00451 +/- 0.01639	-0.05821 +/- 0.03059	0.676
Food Establishment	-0.00405 +/- 0.00114	0.08331 +/- 0.00828	-0.18340 +/- 0.01821	0.22526 +/- 0.03398	0.421
Food Poisoning	-0.00192 +/- 0.00045	0.03111 +/- 0.00330	-0.06733 +/- 0.00725	0.08315 +/- 0.01352	0.433
For Hire Vehicle Complaint	-0.00044 +/- 0.00037	0.01119 +/- 0.00270	-0.02559 +/- 0.00594	0.03991 +/- 0.01109	0.353
Found Property	-0.00029 +/- 0.00011	0.00619 +/- 0.00081	-0.01482 +/- 0.00177	0.02607 +/- 0.00331	0.609
Homeless Person Assistance	-0.00011 +/- 0.00049	0.02026 +/- 0.00355	-0.05377 +/- 0.00780	0.11160 +/- 0.01456	0.392
Other Enforcement	-0.00020 +/- 0.00088	0.03806 +/- 0.00640	-0.08829 +/- 0.01407	0.11224 +/- 0.02626	0.343
Sidewalk Condition	0.00059 +/- 0.00078	0.03073 +/- 0.00568	-0.07592 +/- 0.01248	0.11814 +/- 0.02330	0.315
Taxi Complaint	-0.01164 +/- 0.00347	0.20071 +/- 0.02521	-0.46961 +/- 0.05543	0.80846 +/- 0.10343	0.654
Taxi Report	-0.00045 +/- 0.00009	0.00497 +/- 0.00064	-0.01138 +/- 0.00142	0.02121 +/- 0.00264	0.592

Table 5. Coefficients and standard deviations associated with the groups of RESIDENT HOMEOWNERS HOUSE VALUES for types of requests with R2 greater than 0.2

Type of Request	house value less than 100	house value for 100 to 500	house value 500 or more	R2 os
Damaged Tree	-0.02050 +/- 0.01336	0.02610 +/- 0.00230	0.01149 +/- 0.00178	0.437
Dead Tree	-0.00486 +/- 0.00934	0.01244 +/- 0.00161	0.00859 +/- 0.00125	0.385
Illegal Tree Damage	-0.00134 +/- 0.00167	0.00037 +/- 0.00029	0.00227 +/- 0.00022	0.372
Other Enforcement	-0.00135 +/- 0.00398	0.00075 +/- 0.00069	0.00541 +/- 0.00053	0.349
Overgrown Tree/Branches	-0.01804 +/- 0.01125	0.02406 +/- 0.00194	0.00756 +/- 0.00150	0.433
Root/Sewer/Sidewalk Condition	-0.01594 +/- 0.00702	0.01378 +/- 0.00121	0.00388 +/- 0.00094	0.399
Sewer	-0.02535 +/- 0.02174	0.04118 +/- 0.00375	0.00813 +/- 0.00290	0.394
Sidewalk Condition	0.00035 +/- 0.00381	-0.00226 +/- 0.00066	0.00611 +/- 0.00051	0.444
Street Condition	-0.05631 +/- 0.05056	0.06262 +/- 0.00872	0.06461 +/- 0.00675	0.431

Table 6. Coefficients and standard deviations associated with the groups of WORKERS HOMEOWNERS HOUSE VALUES for types of requests with R2 greater than 0.2

Type of Request	house value less than 100_n	house value for 100 to 500_n	house value 500 or more_n	R2 os
Broken Muni Meter	0.22149 +/- 0.05094	-0.03159 +/- 0.00703	0.00571 +/- 0.00351	0.36
Consumer Complaint	0.02422 +/- 0.01238	-0.00259 +/- 0.00171	0.00084 +/- 0.00085	0.313
Fire Alarm - Reinspection	0.00268 +/- 0.00040	-0.00033 +/- 0.00005	0.00003 +/- 0.00003	0.458
Fire Safety Director - F58	0.00271 +/- 0.00600	-0.00273 +/- 0.00083	0.00443 +/- 0.00041	0.391
Food Establishment	0.01778 +/- 0.00606	-0.00303 +/- 0.00084	0.00132 +/- 0.00042	0.416
Food Poisoning	0.01121 +/- 0.00226	-0.00177 +/- 0.00031	0.00059 +/- 0.00016	0.474
For Hire Vehicle Complaint	0.00478 +/- 0.00170	-0.00077 +/- 0.00023	0.00036 +/- 0.00012	0.394
Found Property	0.00221 +/- 0.00056	-0.00042 +/- 0.00008	0.00021 +/- 0.00004	0.583
Homeless Encampment	0.02155 +/- 0.00499	-0.00366 +/- 0.00069	0.00109 +/- 0.00034	0.377
Homeless Person Assistance	0.01133 +/- 0.00231	-0.00195 +/- 0.00032	0.00070 +/- 0.00016	0.421
Taxi Complaint	0.09270 +/- 0.01699	-0.01734 +/- 0.00234	0.00809 +/- 0.00117	0.64
Taxi Report	0.00182 +/- 0.00049	-0.00046 +/- 0.00007	0.00032 +/- 0.00003	0.569

Table 7. Coefficients and standard deviations associated with the groups of RESIDENTS HOUSEHOLD INCOME for types of requests with R2 greater than 0.2

Type of Request	household income less than 40	household income from 40 to 75	household income 75 and above	R2 os
APPLIANCE	0.01700 +/- 0.00226	-0.00230 +/- 0.00384	-0.00270 +/- 0.00086	0.522
Air Quality	0.00863 +/- 0.00146	-0.01509 +/- 0.00248	0.00903 +/- 0.00055	0.602
Animal Abuse	0.00406 +/- 0.00139	0.00507 +/- 0.00236	-0.00051 +/- 0.00053	0.398
Asbestos	0.00276 +/- 0.00048	-0.00327 +/- 0.00082	0.00213 +/- 0.00018	0.378
Bike/Roller/Skate Chronic	0.00079 +/- 0.00015	-0.00170 +/- 0.00025	0.00084 +/- 0.00006	0.52
Blocked Driveway	-0.08467 +/- 0.01425	0.27883 +/- 0.02417	-0.04893 +/- 0.00540	0.333
Boilers	0.00238 +/- 0.00040	-0.00078 +/- 0.00068	0.00033 +/- 0.00015	0.367
DOOR/WINDOW	0.04552 +/- 0.00652	-0.00128 +/- 0.01106	-0.00786 +/- 0.00247	0.502
Dirty Conditions	-0.00508 +/- 0.00442	0.04165 +/- 0.00751	-0.00213 +/- 0.00168	0.409
Drinking	0.00074 +/- 0.00028	0.00048 +/- 0.00047	0.00001 +/- 0.00010	0.307
ELECTRIC	0.03310 +/- 0.00649	0.01935 +/- 0.01101	-0.01032 +/- 0.00246	0.486
Elevator	0.01354 +/- 0.00238	-0.00400 +/- 0.00403	0.00118 +/- 0.00090	0.332
FLOORING/STAIRS	0.03699 +/- 0.00545	-0.00515 +/- 0.00925	-0.00543 +/- 0.00207	0.462
Food Establishment	0.00764 +/- 0.00178	-0.00949 +/- 0.00302	0.00665 +/- 0.00068	0.448
For Hire Vehicle Complaint	0.00285 +/- 0.00064	-0.00328 +/- 0.00108	0.00158 +/- 0.00024	0.307
Found Property	0.00085 +/- 0.00021	-0.00206 +/- 0.00036	0.00099 +/- 0.00008	0.387
GENERAL	0.04055 +/- 0.00609	-0.00257 +/- 0.01032	-0.00375 +/- 0.00231	0.493
GENERAL CONSTRUCTION	0.02517 +/- 0.00419	-0.00129 +/- 0.00710	-0.00489 +/- 0.00159	0.354
General Construction/Plumbing	0.00785 +/- 0.00913	0.00572 +/- 0.01549	0.00953 +/- 0.00346	0.424
HEAT/HOT WATER	0.19120 +/- 0.03002	-0.00219 +/- 0.05092	-0.02022 +/- 0.01138	0.468
HEATING	0.13272 +/- 0.02752	0.03134 +/- 0.04667	-0.02764 +/- 0.01043	0.361
Hazardous Materials	0.00091 +/- 0.00058	-0.00060 +/- 0.00099	0.00158 +/- 0.00022	0.457

Type of Request	household income less than 40	household income from 40 to 75	household income 75 and above	R2 os
Illegal Tree Damage	-0.00080 +/- 0.00041	0.00221 +/- 0.00070	0.00075 +/- 0.00016	0.387
Indoor Air Quality	0.00416 +/- 0.00202	-0.00295 +/- 0.00342	0.00320 +/- 0.00076	0.39
Indoor Sewage	0.00076 +/- 0.00030	0.00136 +/- 0.00051	-0.00023 +/- 0.00011	0.334
Lead	0.00032 +/- 0.00047	0.00065 +/- 0.00080	0.00155 +/- 0.00018	0.482
Litter Basket / Request	0.00190 +/- 0.00054	-0.00090 +/- 0.00091	0.00161 +/- 0.00020	0.377
NONCONST	0.01249 +/- 0.00197	0.00192 +/- 0.00335	-0.00206 +/- 0.00075	0.509
Noise	0.07007 +/- 0.01180	-0.13703 +/- 0.02001	0.07959 +/- 0.00447	0.654
Noise - Residential	0.21241 +/- 0.03093	0.05446 +/- 0.05246	-0.02546 +/- 0.01173	0.61
Noise - Vehicle	0.01631 +/- 0.00412	-0.00655 +/- 0.00698	0.00432 +/- 0.00156	0.385
OUTSIDE BUILDING	0.00113 +/- 0.00032	0.00101 +/- 0.00054	-0.00017 +/- 0.00012	0.357
Other Enforcement	0.00110 +/- 0.00093	0.00181 +/- 0.00157	0.00230 +/- 0.00035	0.499
PAINT - PLASTER	0.02089 +/- 0.00316	-0.00397 +/- 0.00536	-0.00313 +/- 0.00120	0.389
PAINT/PLASTER	0.09274 +/- 0.01534	0.01370 +/- 0.02603	-0.01897 +/- 0.00582	0.45
PLUMBING	0.09590 +/- 0.01385	0.00115 +/- 0.02349	-0.01548 +/- 0.00525	0.508
Recycling Enforcement	0.00037 +/- 0.00020	0.00033 +/- 0.00034	0.00026 +/- 0.00008	0.324
Rodent	0.03021 +/- 0.00390	-0.01371 +/- 0.00661	0.00496 +/- 0.00148	0.505
SAFETY	0.00898 +/- 0.00157	0.00198 +/- 0.00266	-0.00196 +/- 0.00059	0.493
Sidewalk Condition	0.00338 +/- 0.00084	-0.00550 +/- 0.00142	0.00418 +/- 0.00032	0.574
Taxi Complaint	0.03381 +/- 0.00737	-0.07730 +/- 0.01251	0.03632 +/- 0.00280	0.442
UNSANITARY CONDITION	0.07755 +/- 0.01217	0.02117 +/- 0.02064	-0.01742 +/- 0.00461	0.531
Urinating in Public	0.00041 +/- 0.00010	-0.00032 +/- 0.00017	0.00023 +/- 0.00004	0.333
WATER LEAK	0.03832 +/- 0.00656	0.01167 +/- 0.01112	-0.00976 +/- 0.00249	0.487
Water Quality	0.00003 +/- 0.00019	0.00057 +/- 0.00033	0.00041 +/- 0.00007	0.375

Table 8. Coefficients and standard deviations associated with the groups of WORKERS HOUSEHOLD INCOME for types of requests with R2 greater than 0.2

Type of Request	household income less than 10_n	household income form 10 to 40_n	household income form 40 to 75_n	household income 75 and above_n	R2 os
EAP Inspection - F59	-0.00212 +/- 0.00110	0.00331 +/- 0.00068	-0.00521 +/- 0.00067	0.00133 +/- 0.00008	0.679
Fire Alarm - Reinspection	0.00175 +/- 0.00045	-0.00057 +/- 0.00028	0.00017 +/- 0.00028	-0.00006 +/- 0.00003	0.407
Fire Safety Director - F58	-0.01029 +/- 0.00481	0.01547 +/- 0.00295	-0.02170 +/- 0.00294	0.00465 +/- 0.00034	0.585
Food Establishment	0.01220 +/- 0.00678	-0.00544 +/- 0.00416	0.00403 +/- 0.00415	-0.00076 +/- 0.00049	0.394
Food Poisoning	0.00685 +/- 0.00263	-0.00222 +/- 0.00161	0.00041 +/- 0.00161	-0.00006 +/- 0.00019	0.459
For Hire Vehicle Complaint	0.00029 +/- 0.00187	0.00212 +/- 0.00115	-0.00309 +/- 0.00115	0.00034 +/- 0.00013	0.37
Found Property	0.00179 +/- 0.00065	-0.00076 +/- 0.00040	0.00021 +/- 0.00040	0.00004 +/- 0.00005	0.55
Homeless Encampment	0.02000 +/- 0.00569	-0.00736 +/- 0.00349	0.00143 +/- 0.00348	-0.00016 +/- 0.00041	0.322

Type of Request	household income less than 10_n	household income form 10 to 40_n	household income form 40 to 75_n	household income 75 and above_n	R2 os
Homeless Person Assistance	0.01284 +/- 0.00262	-0.00579 +/- 0.00161	0.00246 +/- 0.00161	-0.00020 +/- 0.00019	0.324
Taxi Complaint	0.07528 +/- 0.02010	-0.03067 +/- 0.01233	0.00670 +/- 0.01230	0.00155 +/- 0.00144	0.618
Taxi Report	0.00109 +/- 0.00054	-0.00033 +/- 0.00033	-0.00028 +/- 0.00033	0.00017 +/- 0.00004	0.61

Table 9. Coefficients and standard deviations associated with the groups of RESIDENT HOME OWNERS vs RENTERS for types of requests with R2 greater than 0.2

Type of Request	owner occupied units	renter occupied units	R2 os
APPLIANCE	-0.00609 +/- 0.00091	0.00522 +/- 0.00043	0.391
Air Quality	0.00284 +/- 0.00071	0.00336 +/- 0.00034	0.344
Asbestos	0.00067 +/- 0.00020	0.00111 +/- 0.00009	0.363
Boilers	-0.00047 +/- 0.00014	0.00092 +/- 0.00006	0.486
Consumer Complaint	0.00117 +/- 0.00160	0.00556 +/- 0.00076	0.331
DOOR/WINDOW	-0.01758 +/- 0.00255	0.01526 +/- 0.00121	0.442
Damaged Tree	0.01789 +/- 0.00126	-0.00350 +/- 0.00060	0.492
Dead Tree	0.01050 +/- 0.00087	-0.00173 +/- 0.00041	0.403
Dirty Conditions	0.00943 +/- 0.00178	0.00567 +/- 0.00084	0.301
Drinking	-0.00015 +/- 0.00010	0.00042 +/- 0.00005	0.339
ELECTRIC	-0.01210 +/- 0.00274	0.01377 +/- 0.00130	0.334
Electrical	0.00069 +/- 0.00055	0.00117 +/- 0.00026	0.341
Elevator	-0.00279 +/- 0.00085	0.00493 +/- 0.00040	0.362
FLOORING/STAIRS	-0.01426 +/- 0.00209	0.01182 +/- 0.00099	0.409
Food Establishment	0.00239 +/- 0.00072	0.00325 +/- 0.00034	0.455
Food Poisoning	0.00093 +/- 0.00034	0.00126 +/- 0.00016	0.389
For Hire Vehicle Complaint	0.00013 +/- 0.00024	0.00096 +/- 0.00011	0.417
GENERAL	-0.01429 +/- 0.00221	0.01477 +/- 0.00104	0.508
General Construction/Plumbing	0.00894 +/- 0.00342	0.00734 +/- 0.00162	0.442
HEAT/HOT WATER	-0.06898 +/- 0.01099	0.07104 +/- 0.00519	0.488
HEATING	-0.05748 +/- 0.01039	0.05284 +/- 0.00491	0.361
Hazardous Materials	0.00118 +/- 0.00023	0.00075 +/- 0.00011	0.38
Illegal Tree Damage	0.00125 +/- 0.00016	0.00042 +/- 0.00008	0.366
Indoor Air Quality	-0.00002 +/- 0.00074	0.00254 +/- 0.00035	0.462
Lead	0.00128 +/- 0.00019	0.00085 +/- 0.00009	0.451
Litter Basket / Request	0.00084 +/- 0.00020	0.00111 +/- 0.00010	0.404
Missed Collection (All Materials)	0.01387 +/- 0.00146	-0.00209 +/- 0.00069	0.324
NONCONST	-0.00391 +/- 0.00079	0.00462 +/- 0.00038	0.361
Noise	0.02365 +/- 0.00618	0.02683 +/- 0.00292	0.321
Noise - Residential	-0.06152 +/- 0.01203	0.08870 +/- 0.00569	0.509

Type of Request	owner occupied units	renter occupied units	R2 os
Noise - Street/Sidewalk	-0.02013 +/- 0.00426	0.02266 +/- 0.00201	0.434
Noise - Vehicle	-0.00358 +/- 0.00142	0.00728 +/- 0.00067	0.452
OUTSIDE BUILDING	-0.00022 +/- 0.00012	0.00062 +/- 0.00006	0.31
Other Enforcement	0.00281 +/- 0.00035	0.00139 +/- 0.00017	0.498
Overgrown Tree/Branches	0.01456 +/- 0.00112	-0.00277 +/- 0.00053	0.442
PAINT - PLASTER	-0.00759 +/- 0.00125	0.00620 +/- 0.00059	0.334
PAINT/PLASTER	-0.03785 +/- 0.00597	0.03424 +/- 0.00282	0.411
PLUMBING	-0.03354 +/- 0.00551	0.03286 +/- 0.00260	0.429
Plumbing	0.00116 +/- 0.00050	0.00071 +/- 0.00024	0.304
Recycling Enforcement	0.00029 +/- 0.00007	0.00028 +/- 0.00004	0.331
Rodent	-0.00377 +/- 0.00135	0.01049 +/- 0.00064	0.542
Root/Sewer/Sidewalk Condition	0.00820 +/- 0.00067	-0.00220 +/- 0.00031	0.476
SAFETY	-0.00335 +/- 0.00063	0.00335 +/- 0.00030	0.388
Sewer	0.02128 +/- 0.00228	-0.00242 +/- 0.00108	0.303
Sidewalk Condition	0.00201 +/- 0.00038	0.00154 +/- 0.00018	0.422
Smoking	-0.00013 +/- 0.00025	0.00082 +/- 0.00012	0.326
Street Condition	0.05969 +/- 0.00490	0.00233 +/- 0.00232	0.446
UNSANITARY CONDITION	-0.02904 +/- 0.00496	0.02977 +/- 0.00234	0.433
Urinating in Public	0.00004 +/- 0.00004	0.00017 +/- 0.00002	0.356
WATER LEAK	-0.01555 +/- 0.00266	0.01467 +/- 0.00126	0.382
Water Quality	0.00056 +/- 0.00007	0.00025 +/- 0.00003	0.391

Table 10. Coefficients and standard deviations associated with the groups of WORKER HOME OWNERS vs RENTERS for types of requests with R2 greater than 0.2

Type of Request	owner occupied units_n	renter occupied units_n	R2 os
Air Quality	-0.00519 +/- 0.00076	0.00387 +/- 0.00049	0.375
Broken Muni Meter	-0.03140 +/- 0.00547	0.02374 +/- 0.00354	0.315
Consumer Complaint	-0.00511 +/- 0.00132	0.00457 +/- 0.00085	0.429
Fire Alarm - Reinspection	-0.00031 +/- 0.00004	0.00025 +/- 0.00003	0.497
Fire Safety Director - F58	-0.00140 +/- 0.00076	0.00288 +/- 0.00049	0.443
Food Establishment	-0.00394 +/- 0.00063	0.00317 +/- 0.00041	0.479
Food Poisoning	-0.00195 +/- 0.00023	0.00157 +/- 0.00015	0.555
For Hire Vehicle Complaint	-0.00101 +/- 0.00018	0.00087 +/- 0.00011	0.484
Found Property	-0.00045 +/- 0.00006	0.00038 +/- 0.00004	0.636
Hazardous Materials	-0.00101 +/- 0.00028	0.00077 +/- 0.00018	0.33
Homeless Encampment	-0.00401 +/- 0.00052	0.00292 +/- 0.00033	0.414
Homeless Person Assistance	-0.00197 +/- 0.00024	0.00154 +/- 0.00016	0.48
Non-Residential Heat	-0.00056 +/- 0.00016	0.00058 +/- 0.00010	0.363
Sidewalk Condition	-0.00174 +/- 0.00041	0.00141 +/- 0.00026	0.349

Type of Request	owner occupied units_n	renter occupied units_n	R2 os
Taxi Complaint	-0.01710 +/- 0.00176	0.01416 +/- 0.00114	0.706
Taxi Report	-0.00040 +/- 0.00006	0.00037 +/- 0.00004	0.6

Table 11. Coefficients and standard deviations associated with the groups of RESIDENTS RACE for types of requests with R2 greater than 0.2

Type of Request	Population white	population black	Population asian	population hispanic	population other race	R2 os
APPLIANCE	0.00050 +/- 0.00019	0.00257 +/- 0.00021	-0.00060 +/- 0.00040	0.00275 +/- 0.00023	-0.00163 +/- 0.00159	0.605
Animal Abuse	0.00065 +/- 0.00013	0.00115 +/- 0.00014	0.00005 +/- 0.00026	0.00121 +/- 0.00015	0.00221 +/- 0.00104	0.457
Boilers	0.00024 +/- 0.00004	0.00025 +/- 0.00004	-0.00011 +/- 0.00008	0.00042 +/- 0.00005	0.00020 +/- 0.00031	0.336
Bus Stop Shelter Placement	0.00005 +/- 0.00001	0.00004 +/- 0.00001	0.00001 +/- 0.00001	0.00002 +/- 0.00001	-0.00000 +/- 0.00005	0.305
DOOR/WINDOW	0.00175 +/- 0.00056	0.00692 +/- 0.00062	-0.00208 +/- 0.00116	0.00819 +/- 0.00067	-0.00476 +/- 0.00467	0.573
Dirty Conditions	0.00271 +/- 0.00041	0.00412 +/- 0.00045	0.00255 +/- 0.00084	0.00221 +/- 0.00049	0.00926 +/- 0.00338	0.4
Drinking	0.00005 +/- 0.00002	0.00000 +/- 0.00003	0.00013 +/- 0.00005	0.00024 +/- 0.00003	0.00105 +/- 0.00019	0.368
ELECTRIC	0.00177 +/- 0.00053	0.00794 +/- 0.00059	-0.00111 +/- 0.00110	0.00698 +/- 0.00064	-0.00239 +/- 0.00442	0.62
FLOORING/STAIRS	0.00132 +/- 0.00047	0.00498 +/- 0.00052	-0.00202 +/- 0.00097	0.00674 +/- 0.00056	-0.00335 +/- 0.00389	0.537
GENERAL	0.00212 +/- 0.00053	0.00589 +/- 0.00058	-0.00158 +/- 0.00109	0.00765 +/- 0.00063	-0.00227 +/- 0.00437	0.56
GENERAL CONSTRUCTION	0.00086 +/- 0.00037	0.00436 +/- 0.00040	-0.00084 +/- 0.00075	0.00400 +/- 0.00044	-0.00609 +/- 0.00303	0.404
HEAT/HOT WATER	0.01085 +/- 0.00274	0.02466 +/- 0.00301	-0.00676 +/- 0.00563	0.03909 +/- 0.00327	-0.02121 +/- 0.02268	0.467
HEATING	0.00664 +/- 0.00257	0.02008 +/- 0.00281	-0.00479 +/- 0.00528	0.02944 +/- 0.00306	-0.02417 +/- 0.02123	0.325
Illegal Parking	0.00986 +/- 0.00087	0.00306 +/- 0.00096	0.00802 +/- 0.00179	0.00326 +/- 0.00104	0.00736 +/- 0.00721	0.362
Illegal Tree Damage	0.00041 +/- 0.00004	0.00002 +/- 0.00004	0.00018 +/- 0.00008	0.00009 +/- 0.00004	0.00093 +/- 0.00031	0.32
Indoor Sewage	0.00012 +/- 0.00002	0.00042 +/- 0.00002	0.00003 +/- 0.00004	0.00015 +/- 0.00002	0.00015 +/- 0.00017	0.612
Lead	0.00062 +/- 0.00005	0.00021 +/- 0.00005	0.00005 +/- 0.00010	0.00018 +/- 0.00006	0.00045 +/- 0.00040	0.331
Litter Basket / Request	0.00061 +/- 0.00005	0.00027 +/- 0.00006	-0.00001 +/- 0.00011	0.00037 +/- 0.00006	0.00046 +/- 0.00045	0.349
NONCONST	0.00072 +/- 0.00017	0.00236 +/- 0.00019	-0.00034 +/- 0.00035	0.00221 +/- 0.00020	-0.00074 +/- 0.00141	0.608
Noise - Residential	0.01219 +/- 0.00245	0.03205 +/- 0.00268	-0.00566 +/- 0.00503	0.04855 +/- 0.00292	0.07302 +/- 0.02026	0.696
Noise - Vehicle	0.00201 +/- 0.00036	0.00110 +/- 0.00040	-0.00198 +/- 0.00075	0.00399 +/- 0.00043	0.00736 +/- 0.00301	0.362
Non-Emergency Police Matter	0.00058 +/- 0.00010	0.00024 +/- 0.00011	-0.00028 +/- 0.00021	0.00027 +/- 0.00012	0.00958 +/- 0.00083	0.4
OUTSIDE BUILDING	0.00015 +/- 0.00003	0.00031 +/- 0.00003	-0.00005 +/- 0.00006	0.00032 +/- 0.00003	-0.00002 +/- 0.00022	0.48
Other Enforcement	0.00094 +/- 0.00008	0.00036 +/- 0.00009	0.00075 +/- 0.00017	0.00017 +/- 0.00010	0.00355 +/- 0.00067	0.575

Type of Request	Population white	population black	Population asian	population hispanic	population other race	R2 os
PAINT - PLASTER	0.00066 +/- 0.00028	0.00316 +/- 0.00030	-0.00090 +/- 0.00057	0.00321 +/- 0.00033	-0.00342 +/- 0.00230	0.459
PAINT/PLASTER	0.00429 +/- 0.00136	0.01468 +/- 0.00149	-0.00358 +/- 0.00279	0.01928 +/- 0.00162	-0.01478 +/- 0.01121	0.51
PLUMBING	0.00408 +/- 0.00117	0.01632 +/- 0.00129	-0.00317 +/- 0.00241	0.01651 +/- 0.00140	-0.00967 +/- 0.00971	0.606
Recycling Enforcement	0.00019 +/- 0.00002	0.00012 +/- 0.00002	0.00005 +/- 0.00004	0.00005 +/- 0.00002	0.00017 +/- 0.00015	0.37
Rodent	0.00284 +/- 0.00036	0.00376 +/- 0.00040	-0.00103 +/- 0.00074	0.00453 +/- 0.00043	0.00266 +/- 0.00298	0.518
SAFETY	0.00043 +/- 0.00014	0.00141 +/- 0.00015	-0.00013 +/- 0.00028	0.00206 +/- 0.00016	-0.00133 +/- 0.00113	0.547
Sanitation Condition	0.00237 +/- 0.00034	0.00419 +/- 0.00038	0.00294 +/- 0.00071	0.00033 +/- 0.00041	0.00779 +/- 0.00285	0.357
Sidewalk Condition	0.00109 +/- 0.00009	0.00021 +/- 0.00010	0.00015 +/- 0.00019	0.00009 +/- 0.00011	0.00157 +/- 0.00076	0.442
Street Condition	0.01169 +/- 0.00132	0.00401 +/- 0.00145	0.00705 +/- 0.00271	-0.00184 +/- 0.00157	0.03681 +/- 0.01092	0.391
UNSANITARY CONDITION	0.00388 +/- 0.00105	0.01463 +/- 0.00115	-0.00167 +/- 0.00216	0.01537 +/- 0.00125	-0.00955 +/- 0.00868	0.599
WATER LEAK	0.00180 +/- 0.00058	0.00757 +/- 0.00063	-0.00078 +/- 0.00118	0.00746 +/- 0.00069	-0.00818 +/- 0.00476	0.555
Water Quality	0.00022 +/- 0.00002	0.00009 +/- 0.00002	0.00003 +/- 0.00004	0.00004 +/- 0.00002	0.00014 +/- 0.00015	0.373

Table 12. Coefficients and standard deviations associated with the groups of WORKERS RACE for types of requests with R2 greater than 0.2

Type Of Request	Population white_n	population black_n	Population asian_n	population hispanic_n	population other _n	R2 os
Broken Muni Meter	-0.01462 +/- 0.00473	-0.01918 +/- 0.00665	0.05452 +/- 0.01799	0.04164 +/- 0.00963	-0.17429 +/- 0.14242	0.355
Construction	-0.00974 +/- 0.00079	-0.00688 +/- 0.00111	0.01692 +/- 0.00299	-0.00559 +/- 0.00160	0.18487 +/- 0.02369	0.433
Consumer Complaint	-0.00631 +/- 0.00105	-0.00510 +/- 0.00148	0.01580 +/- 0.00399	0.00449 +/- 0.00214	0.05627 +/- 0.03160	0.332
EAP Inspection - F59	0.00053 +/- 0.00011	-0.00186 +/- 0.00016	-0.00068 +/- 0.00043	-0.00009 +/- 0.00023	0.01885 +/- 0.00344	0.458
Fire Alarm - Reinspection	-0.00015 +/- 0.00004	-0.00011 +/- 0.00005	0.00042 +/- 0.00015	0.00032 +/- 0.00008	-0.00043 +/- 0.00116	0.336
Fire Safety Director - F58	0.00070 +/- 0.00045	-0.00718 +/- 0.00063	0.00027 +/- 0.00170	0.00181 +/- 0.00091	0.05845 +/- 0.01349	0.562
Food Establishment	-0.00232 +/- 0.00056	-0.00168 +/- 0.00078	0.00799 +/- 0.00212	0.00203 +/- 0.00113	0.00927 +/- 0.01677	0.443
Food Poisoning	-0.00082 +/- 0.00022	-0.00116 +/- 0.00031	0.00123 +/- 0.00084	0.00125 +/- 0.00045	0.01139 +/- 0.00661	0.478
For Hire Vehicle Complaint	-0.00024 +/- 0.00016	-0.00054 +/- 0.00023	0.00022 +/- 0.00062	0.00075 +/- 0.00033	0.00429 +/- 0.00492	0.345
Found Property	-0.00005 +/- 0.00005	-0.00040 +/- 0.00008	-0.00021 +/- 0.00020	0.00049 +/- 0.00011	0.00182 +/- 0.00162	0.559
Homeless Person Assistance	-0.00009 +/- 0.00024	-0.00058 +/- 0.00034	-0.00171 +/- 0.00091	0.00079 +/- 0.00049	0.01166 +/- 0.00719	0.327
Taxi Complaint	-0.00084 +/- 0.00173	-0.01279 +/- 0.00243	-0.01352 +/- 0.00657	0.01383 +/- 0.00352	0.10112 +/- 0.05201	0.614
Taxi Report	0.00006 +/- 0.00005	-0.00051 +/- 0.00006	-0.00050 +/- 0.00017	0.00031 +/- 0.00009	0.00412 +/- 0.00138	0.641

Table 13. Coefficients and standard deviations associated with the groups of RESIDENTS MONTHLY RENT for types of requests with R2 greater than 0.2

Type of Request	rent less than 1000	rent bewteen 1000 and 2000	rent 2000 or more	R2 os
APPLIANCE	0.00999 +/- 0.00100	0.00576 +/- 0.00082	-0.00476 +/- 0.00093	0.572
Air Quality	0.00175 +/- 0.00064	0.00052 +/- 0.00053	0.01196 +/- 0.00060	0.627
Animal Abuse	0.00366 +/- 0.00067	0.00289 +/- 0.00055	-0.00038 +/- 0.00063	0.369
Asbestos	0.00081 +/- 0.00023	0.00062 +/- 0.00018	0.00271 +/- 0.00021	0.359
Bike/Roller/Skate Chronic	0.00014 +/- 0.00006	-0.00010 +/- 0.00005	0.00113 +/- 0.00006	0.632
Blocked Driveway	-0.03965 +/- 0.00704	0.07770 +/- 0.00577	-0.03831 +/- 0.00654	0.342
Boilers	0.00132 +/- 0.00017	0.00082 +/- 0.00014	0.00037 +/- 0.00016	0.463
DOOR/WINDOW	0.02565 +/- 0.00286	0.01899 +/- 0.00234	-0.01294 +/- 0.00266	0.565
Drinking	0.00023 +/- 0.00013	0.00066 +/- 0.00011	0.00018 +/- 0.00012	0.33
ELECTRIC	0.02275 +/- 0.00296	0.01916 +/- 0.00242	-0.01339 +/- 0.00275	0.529
Electrical	0.00097 +/- 0.00071	0.00150 +/- 0.00058	0.00111 +/- 0.00066	0.318
Elevator	0.00592 +/- 0.00108	0.00572 +/- 0.00088	0.00098 +/- 0.00100	0.368
FLOORING/STAIRS	0.01990 +/- 0.00243	0.01436 +/- 0.00199	-0.00968 +/- 0.00226	0.511
Food Establishment	0.00166 +/- 0.00080	0.00158 +/- 0.00066	0.00947 +/- 0.00075	0.471
For Hire Vehicle Complaint	0.00068 +/- 0.00029	0.00050 +/- 0.00024	0.00227 +/- 0.00027	0.394
Found Property	0.00001 +/- 0.00009	-0.00011 +/- 0.00008	0.00136 +/- 0.00009	0.517
GENERAL	0.02350 +/- 0.00261	0.01674 +/- 0.00214	-0.00656 +/- 0.00242	0.585
GENERAL CONSTRUCTION	0.01262 +/- 0.00192	0.01105 +/- 0.00158	-0.00821 +/- 0.00179	0.391
General Construction/Plumbing	0.00324 +/- 0.00442	0.00785 +/- 0.00363	0.01558 +/- 0.00411	0.418
HEAT/HOT WATER	0.09587 +/- 0.01287	0.09630 +/- 0.01056	-0.03756 +/- 0.01196	0.575
HEATING	0.07425 +/- 0.01232	0.07463 +/- 0.01010	-0.04002 +/- 0.01145	0.432
Hazardous Materials	0.00036 +/- 0.00029	0.00048 +/- 0.00024	0.00227 +/- 0.00027	0.381
Homeless Encampment	0.00174 +/- 0.00054	-0.00064 +/- 0.00044	0.00732 +/- 0.00050	0.347
Indoor Air Quality	0.00176 +/- 0.00093	0.00191 +/- 0.00077	0.00478 +/- 0.00087	0.472
Indoor Sewage	0.00089 +/- 0.00014	0.00058 +/- 0.00012	-0.00024 +/- 0.00013	0.353
Lead	-0.00020 +/- 0.00024	0.00121 +/- 0.00020	0.00214 +/- 0.00023	0.388
Litter Basket / Request	0.00064 +/- 0.00026	0.00108 +/- 0.00021	0.00214 +/- 0.00024	0.321
NONCONST	0.00741 +/- 0.00088	0.00592 +/- 0.00072	-0.00317 +/- 0.00081	0.549
Noise	0.00960 +/- 0.00448	-0.00068 +/- 0.00367	0.11092 +/- 0.00416	0.764
Noise - Residential	0.15388 +/- 0.01290	0.09206 +/- 0.01058	-0.03128 +/- 0.01199	0.681
Noise - Street/Sidewalk	0.03518 +/- 0.00565	0.01505 +/- 0.00463	0.01157 +/- 0.00525	0.355
Noise - Vehicle	0.00815 +/- 0.00185	0.00633 +/- 0.00152	0.00638 +/- 0.00172	0.443
OUTSIDE BUILDING	0.00083 +/- 0.00015	0.00085 +/- 0.00012	-0.00017 +/- 0.00013	0.386
Other Enforcement	-0.00046 +/- 0.00048	0.00223 +/- 0.00039	0.00350 +/- 0.00044	0.391
PAINT - PLASTER	0.01031 +/- 0.00145	0.00808 +/- 0.00119	-0.00595 +/- 0.00134	0.441
PAINT/PLASTER	0.04961 +/- 0.00670	0.04928 +/- 0.00550	-0.02995 +/- 0.00623	0.529
PLUMBING	0.05494 +/- 0.00607	0.04098 +/- 0.00498	-0.02589 +/- 0.00564	0.573
Recycling Enforcement	0.00005 +/- 0.00010	0.00043 +/- 0.00008	0.00043 +/- 0.00009	0.303

Type of Request	rent less than 1000	rent bewteen 1000 and 2000	rent 2000 or more	R2 os
Rodent	0.01746 +/- 0.00165	0.00741 +/- 0.00136	0.00519 +/- 0.00154	0.585
SAFETY	0.00510 +/- 0.00071	0.00461 +/- 0.00058	-0.00275 +/- 0.00066	0.541
Sidewalk Condition	-0.00006 +/- 0.00039	0.00091 +/- 0.00032	0.00570 +/- 0.00036	0.588
Taxi Complaint	0.00388 +/- 0.00315	-0.00556 +/- 0.00258	0.04932 +/- 0.00293	0.544
UNSANITARY CONDITION	0.04571 +/- 0.00532	0.04128 +/- 0.00436	-0.02506 +/- 0.00494	0.603
Urinating in Public	0.00015 +/- 0.00005	0.00012 +/- 0.00004	0.00033 +/- 0.00004	0.38
WATER LEAK	0.02288 +/- 0.00290	0.02073 +/- 0.00238	-0.01408 +/- 0.00269	0.546

Table 14. Coefficients and standard deviations associated with the groups of WORKERS MONTHLY RENT for types of requests with R2 greater than 0.2

Type of Request	rent less than 1000_n	rent bewteen 1000 and 2000_n	rent 2000 or more_n	R2 os
EAP Inspection - F59	0.00006 +/- 0.00032	-0.00042 +/- 0.00018	0.00230 +/- 0.00009	0.677
Fire Alarm - Reinspection	0.00074 +/- 0.00012	-0.00035 +/- 0.00007	0.00006 +/- 0.00004	0.383
Fire Safety Director - F58	0.00095 +/- 0.00140	-0.00159 +/- 0.00079	0.00740 +/- 0.00042	0.681
Food Establishment	0.00382 +/- 0.00189	-0.00150 +/- 0.00107	0.00046 +/- 0.00057	0.361
Food Poisoning	0.00240 +/- 0.00073	-0.00120 +/- 0.00041	0.00066 +/- 0.00022	0.406
For Hire Vehicle Complaint	0.00039 +/- 0.00053	-0.00015 +/- 0.00030	0.00041 +/- 0.00016	0.322
Found Property	0.00046 +/- 0.00018	-0.00028 +/- 0.00010	0.00030 +/- 0.00005	0.555
Homeless Encampment	0.00680 +/- 0.00154	-0.00378 +/- 0.00087	0.00143 +/- 0.00046	0.347
Homeless Person Assistance	0.00309 +/- 0.00074	-0.00169 +/- 0.00042	0.00086 +/- 0.00022	0.325
Taxi Complaint	0.02257 +/- 0.00546	-0.01346 +/- 0.00308	0.01179 +/- 0.00163	0.623
Taxi Report	0.00021 +/- 0.00015	-0.00020 +/- 0.00008	0.00048 +/- 0.00004	0.643

Table 15. Coefficients and standard deviations associated with the groups of RESIDENTS TRANSPORTATION MEAN for types of requests with R2 greater than 0.2

Type of Request	transportation car	transportation public	tranportation motorcycle	Transportation Other means	R2 os
APPLIANCE	-0.00472 +/- 0.00112	0.00632 +/- 0.00063	-0.67422 +/- 0.18128	-0.00794 +/- 0.00171	0.337
Air Quality	-0.00127 +/- 0.00063	0.00084 +/- 0.00035	0.33691 +/- 0.10249	0.01003 +/- 0.00097	0.597
Animal Abuse	0.00138 +/- 0.00062	0.00248 +/- 0.00035	-0.03283 +/- 0.10012	-0.00161 +/- 0.00095	0.301
Asbestos	-0.00049 +/- 0.00021	0.00072 +/- 0.00012	0.09072 +/- 0.03410	0.00126 +/- 0.00032	0.357
Bike/Roller/Skate Chronic	-0.00016 +/- 0.00006	-0.00002 +/- 0.00003	0.01763 +/- 0.00972	0.00112 +/- 0.00009	0.501
Blocked Driveway	0.04713 +/- 0.00634	0.02771 +/- 0.00355	-0.63137 +/- 1.02890	-0.03875 +/- 0.00972	0.37
Boilers	-0.00066 +/- 0.00017	0.00091 +/- 0.00009	-0.07885 +/- 0.02707	-0.00025 +/- 0.00026	0.383
Consumer Complaint	0.00035 +/- 0.00171	0.00296 +/- 0.00096	-0.33262 +/- 0.27713	0.01339 +/- 0.00262	0.43

Type of Request	transportation car	transportation public	tranportation motorcycle	Transportation Other means	R2 os
DOOR/WINDOW	-0.01471 +/- 0.00311	0.01856 +/- 0.00174	-1.84389 +/- 0.50561	-0.02289 +/- 0.00478	0.33
Damaged Tree	0.02245 +/- 0.00114	-0.00079 +/- 0.00064	0.23124 +/- 0.18518	0.00072 +/- 0.00175	0.659
Dead Tree	0.01203 +/- 0.00095	-0.00103 +/- 0.00053	0.23898 +/- 0.15416	0.00250 +/- 0.00146	0.406
Derelict Vehicle	0.01608 +/- 0.00125	-0.00157 +/- 0.00070	0.36123 +/- 0.20269	-0.00160 +/- 0.00192	0.416
Dirty Conditions	0.01030 +/- 0.00180	0.00733 +/- 0.00101	0.25905 +/- 0.29163	-0.00255 +/- 0.00276	0.424
Drinking	-0.00001 +/- 0.00011	0.00035 +/- 0.00006	-0.00168 +/- 0.01861	0.00022 +/- 0.00018	0.317
ELECTRIC	-0.00914 +/- 0.00311	0.01813 +/- 0.00174	-1.73971 +/- 0.50491	-0.02353 +/- 0.00477	0.321
Elevator	-0.00347 +/- 0.00101	0.00531 +/- 0.00056	-0.69840 +/- 0.16345	-0.00183 +/- 0.00154	0.309
Food Establishment	0.00030 +/- 0.00068	0.00062 +/- 0.00038	0.15210 +/- 0.11021	0.01081 +/- 0.00104	0.499
Food Poisoning	-0.00005 +/- 0.00033	0.00011 +/- 0.00018	0.01713 +/- 0.05304	0.00498 +/- 0.00050	0.397
For Hire Vehicle Complaint	-0.00034 +/- 0.00025	0.00037 +/- 0.00014	-0.03463 +/- 0.04054	0.00253 +/- 0.00038	0.472
Found Property	-0.00014 +/- 0.00008	-0.00012 +/- 0.00005	0.02156 +/- 0.01362	0.00153 +/- 0.00013	0.46
GENERAL	-0.01290 +/- 0.00275	0.01673 +/- 0.00154	-1.41700 +/- 0.44711	-0.01645 +/- 0.00422	0.403
General Construction/Plumbing	0.00625 +/- 0.00375	0.00391 +/- 0.00210	0.95111 +/- 0.60933	0.01512 +/- 0.00576	0.449
HEAT/HOT WATER	-0.06045 +/- 0.01377	0.07943 +/- 0.00771	-7.39603 +/- 2.23611	-0.07540 +/- 0.02113	0.347
Hazardous Materials	0.00036 +/- 0.00025	0.00036 +/- 0.00014	0.08777 +/- 0.04086	0.00173 +/- 0.00039	0.387
Homeless Encampment	-0.00076 +/- 0.00048	-0.00015 +/- 0.00027	0.12137 +/- 0.07845	0.00798 +/- 0.00074	0.378
Homeless Person Assistance	-0.00073 +/- 0.00027	-0.00003 +/- 0.00015	0.02131 +/- 0.04433	0.00497 +/- 0.00042	0.413
Illegal Parking	0.03036 +/- 0.00415	0.00620 +/- 0.00233	0.69926 +/- 0.67439	0.02595 +/- 0.00637	0.316
Illegal Tree Damage	0.00092 +/- 0.00017	0.00010 +/- 0.00009	0.13024 +/- 0.02731	0.00138 +/- 0.00026	0.389
Indoor Air Quality	-0.00208 +/- 0.00081	0.00172 +/- 0.00045	-0.08008 +/- 0.13140	0.00350 +/- 0.00124	0.429
Lead	-0.00002 +/- 0.00021	0.00086 +/- 0.00012	0.07350 +/- 0.03409	0.00049 +/- 0.00032	0.412
Litter Basket / Request	-0.00018 +/- 0.00021	0.00085 +/- 0.00012	0.11413 +/- 0.03465	0.00092 +/- 0.00033	0.416
Missed Collection (All Materials)	0.01805 +/- 0.00150	-0.00310 +/- 0.00084	0.60806 +/- 0.24305	0.00875 +/- 0.00230	0.406
NONCONST	-0.00348 +/- 0.00093	0.00576 +/- 0.00052	-0.58773 +/- 0.15057	-0.00633 +/- 0.00142	0.36
Noise	-0.00668 +/- 0.00461	0.00017 +/- 0.00258	2.32366 +/- 0.74782	0.11055 +/- 0.00707	0.662
Noise - Commercial	-0.01494 +/- 0.00481	0.01003 +/- 0.00269	1.68352 +/- 0.78110	0.03624 +/- 0.00738	0.37
Noise - Residential	-0.04973 +/- 0.01438	0.10172 +/- 0.00805	-5.11293 +/- 2.33437	-0.09661 +/- 0.02206	0.466
Noise - Vehicle	-0.00529 +/- 0.00160	0.00618 +/- 0.00090	-0.02611 +/- 0.26011	0.00179 +/- 0.00246	0.401

Type of Request	transportation car	transportation public	tranportation motorcycle	Transportation Other means	R2 os
OUTSIDE BUILDING	-0.00030 +/- 0.00013	0.00081 +/- 0.00008	-0.07851 +/- 0.02179	-0.00067 +/- 0.00021	0.346
Other Enforcement	0.00179 +/- 0.00037	0.00080 +/- 0.00021	0.16210 +/- 0.06043	0.00351 +/- 0.00057	0.493
Overgrown Tree/Branches	0.01851 +/- 0.00099	-0.00006 +/- 0.00056	0.04891 +/- 0.16129	-0.00065 +/- 0.00152	0.64
PAINT/PLASTER	-0.03080 +/- 0.00713	0.04275 +/- 0.00399	-4.50888 +/- 1.15767	-0.05115 +/- 0.01094	0.305
PLUMBING	-0.02770 +/- 0.00666	0.03985 +/- 0.00373	-3.98594 +/- 1.08158	-0.04676 +/- 0.01022	0.348
Recycling Enforcement	0.00020 +/- 0.00008	0.00016 +/- 0.00005	0.03837 +/- 0.01314	0.00042 +/- 0.00012	0.344
Rodent	-0.00611 +/- 0.00167	0.00912 +/- 0.00093	0.24412 +/- 0.27067	-0.00310 +/- 0.00256	0.462
Root/Sewer/Sidewalk Condition	0.01082 +/- 0.00061	-0.00090 +/- 0.00034	0.12783 +/- 0.09937	-0.00015 +/- 0.00094	0.641
SAFETY	-0.00280 +/- 0.00073	0.00411 +/- 0.00041	-0.28510 +/- 0.11913	-0.00496 +/- 0.00113	0.31
Sanitation Condition	0.01322 +/- 0.00163	0.00431 +/- 0.00091	0.05930 +/- 0.26428	0.00016 +/- 0.00250	0.344
Sewer	0.02851 +/- 0.00223	-0.00027 +/- 0.00125	-0.11511 +/- 0.36285	0.00609 +/- 0.00343	0.431
Sidewalk Condition	0.00015 +/- 0.00036	0.00035 +/- 0.00020	0.20569 +/- 0.05894	0.00499 +/- 0.00056	0.54
Snow	0.00872 +/- 0.00085	0.00102 +/- 0.00048	0.22672 +/- 0.13802	-0.00061 +/- 0.00130	0.416
Street Condition	0.06179 +/- 0.00538	-0.00370 +/- 0.00301	1.31439 +/- 0.87347	0.05701 +/- 0.00825	0.457
Taxi Complaint	-0.00607 +/- 0.00291	-0.00315 +/- 0.00163	0.65852 +/- 0.47178	0.05382 +/- 0.00446	0.437
UNSANITARY CONDITION	-0.02305 +/- 0.00593	0.03656 +/- 0.00332	-3.24687 +/- 0.96323	-0.04472 +/- 0.00910	0.335
Urinating in Public	-0.00008 +/- 0.00004	0.00011 +/- 0.00002	0.00306 +/- 0.00632	0.00025 +/- 0.00006	0.397
Water Quality	0.00043 +/- 0.00009	0.00016 +/- 0.00005	0.02383 +/- 0.01400	0.00040 +/- 0.00013	0.303
Water System	0.01502 +/- 0.00261	0.00593 +/- 0.00146	-0.43630 +/- 0.42318	0.01199 +/- 0.00400	0.309

Table 16. Coefficients and standard deviations associated with the groups of WORKERS TRANSPORTATION MEAN for types of requests with R2 greater than 0.2

Type of Request	transportation car_n	transportation public_n	tranportation motorcycle_n	Transportation Other means_n	R2 os
Air Quality	-0.01531 +/- 0.00194	0.02791 +/- 0.00390	-0.01990 +/- 0.10139	-0.01601 +/- 0.00292	0.424
Consumer Complaint	-0.01904 +/- 0.00367	0.03911 +/- 0.00741	-0.74134 +/- 0.19242	0.00190 +/- 0.00555	0.317
EAP Inspection - F59	0.00128 +/- 0.00034	-0.00404 +/- 0.00069	0.00345 +/- 0.01800	0.00906 +/- 0.00052	0.666
Fire Alarm - Reinspection	-0.00061 +/- 0.00011	0.00107 +/- 0.00023	0.01264 +/- 0.00601	-0.00088 +/- 0.00017	0.45
Fire Safety Director - F58	0.00160 +/- 0.00148	-0.00741 +/- 0.00299	-0.10834 +/- 0.07765	0.02898 +/- 0.00224	0.622
Food Establishment	-0.00968 +/- 0.00164	0.01646 +/- 0.00330	0.11975 +/- 0.08586	-0.01073 +/- 0.00248	0.52

Type of Request	transportation car_n	transportation public_n	tranportation motorcycle_n	Transportation Other means_n	R2 os
Food Poisoning	-0.00402 +/- 0.00060	0.00624 +/- 0.00122	0.05359 +/- 0.03167	-0.00293 +/- 0.00091	0.534
For Hire Vehicle Complaint	-0.00317 +/- 0.00049	0.00587 +/- 0.00099	-0.11261 +/- 0.02562	0.00138 +/- 0.00074	0.45
Found Property	-0.00096 +/- 0.00016	0.00138 +/- 0.00032	0.00007 +/- 0.00826	0.00013 +/- 0.00024	0.61
General Construction/Plumbing	-0.02245 +/- 0.01140	0.04109 +/- 0.02300	0.51563 +/- 0.59744	-0.04500 +/- 0.01724	0.375
Hazardous Materials	-0.00332 +/- 0.00076	0.00635 +/- 0.00152	-0.02393 +/- 0.03957	-0.00337 +/- 0.00114	0.338
Homeless Person Assistance	-0.00418 +/- 0.00063	0.00619 +/- 0.00126	0.05103 +/- 0.03277	-0.00250 +/- 0.00095	0.356
Indoor Air Quality	-0.01021 +/- 0.00252	0.01838 +/- 0.00509	-0.17341 +/- 0.13217	-0.00463 +/- 0.00381	0.311
Litter Basket / Request	-0.00480 +/- 0.00071	0.00827 +/- 0.00144	0.03605 +/- 0.03739	-0.00605 +/- 0.00108	0.318
Other Enforcement	-0.00535 +/- 0.00125	0.01076 +/- 0.00253	0.03622 +/- 0.06571	-0.00953 +/- 0.00190	0.303
Sidewalk Condition	-0.00568 +/- 0.00104	0.01095 +/- 0.00210	0.05043 +/- 0.05444	-0.00905 +/- 0.00157	0.44
Taxi Complaint	-0.03643 +/- 0.00444	0.05221 +/- 0.00895	0.08974 +/- 0.23244	0.00131 +/- 0.00671	0.704
Taxi Report	-0.00086 +/- 0.00013	0.00109 +/- 0.00026	-0.01775 +/- 0.00688	0.00136 +/- 0.00020	0.677

Table 17. Coefficients and standard deviations associated with the groups of RESIDENTS LIVING IN FAMILY HOUSEHOLDS VS NON-FAMILY HOUSEHOLDS for types of requests with R2 greater than 0.2

Type of Request	family households	nonfamily households	R2 os
Air Quality	-0.00184 +/- 0.00048	0.00721 +/- 0.00041	0.619
Animal Abuse	0.00363 +/- 0.00048	0.00046 +/- 0.00040	0.351
Asbestos	-0.00006 +/- 0.00015	0.00184 +/- 0.00013	0.495
Bike/Roller/Skate Chronic	-0.00033 +/- 0.00005	0.00068 +/- 0.00004	0.512
Blocked Driveway	0.06697 +/- 0.00481	-0.02912 +/- 0.00407	0.405
Dirty Conditions	0.01377 +/- 0.00146	0.00048 +/- 0.00123	0.395
Electrical	0.00148 +/- 0.00048	0.00069 +/- 0.00041	0.301
Food Establishment	-0.00097 +/- 0.00055	0.00618 +/- 0.00046	0.517
Food Poisoning	-0.00077 +/- 0.00026	0.00272 +/- 0.00022	0.403
For Hire Vehicle Complaint	-0.00001 +/- 0.00021	0.00138 +/- 0.00018	0.373
Found Property	-0.00047 +/- 0.00007	0.00081 +/- 0.00006	0.513
General Construction/Plumbing	0.00704 +/- 0.00300	0.00796 +/- 0.00254	0.443
Hazardous Materials	0.00016 +/- 0.00019	0.00137 +/- 0.00016	0.466
Homeless Encampment	-0.00209 +/- 0.00037	0.00473 +/- 0.00031	0.425
Homeless Person Assistance	-0.00137 +/- 0.00021	0.00292 +/- 0.00018	0.393
Indoor Air Quality	-0.00081 +/- 0.00063	0.00414 +/- 0.00053	0.527
Lead	0.00048 +/- 0.00016	0.00130 +/- 0.00014	0.502
Litter Basket / Request	0.00050 +/- 0.00018	0.00145 +/- 0.00015	0.387

Type of Request	family households	nonfamily households	R2 os
Noise	-0.02435 +/- 0.00368	0.06577 +/- 0.00311	0.679
Noise - Residential	0.09340 +/- 0.01285	0.02204 +/- 0.01087	0.321
Noise - Vehicle	0.00277 +/- 0.00136	0.00632 +/- 0.00115	0.314
Other Enforcement	0.00154 +/- 0.00032	0.00181 +/- 0.00027	0.462
Recycling Enforcement	0.00029 +/- 0.00007	0.00027 +/- 0.00006	0.332
Rodent	0.00711 +/- 0.00141	0.00718 +/- 0.00119	0.352
Sidewalk Condition	-0.00026 +/- 0.00030	0.00313 +/- 0.00025	0.516
Taxi Complaint	-0.01620 +/- 0.00231	0.02948 +/- 0.00195	0.579
Taxi Report	-0.00050 +/- 0.00009	0.00085 +/- 0.00008	0.376
Urinating in Public	0.00002 +/- 0.00003	0.00024 +/- 0.00003	0.385
Water Quality	0.00024 +/- 0.00007	0.00038 +/- 0.00006	0.35

Table 18. Coefficients and standard deviations associated with the groups of WORKERS LIVING IN FAMILY HOUSEHOLDS VS NON-FAMILY HOUSEHOLDS for types of requests with R2 greater than 0.2

Type of Request	family households_n	nonfamily households_n	R2 os
Broken Muni Meter	-0.00306 +/- 0.00326	0.00899 +/- 0.00470	0.333
EAP Inspection - F59	-0.00096 +/- 0.00008	0.00215 +/- 0.00011	0.49
Fire Alarm - Reinspection	-0.00001 +/- 0.00003	0.00009 +/- 0.00004	0.404
Fire Safety Director - F58	-0.00307 +/- 0.00030	0.00705 +/- 0.00044	0.655
Food Establishment	-0.00041 +/- 0.00038	0.00143 +/- 0.00055	0.375
Food Poisoning	-0.00045 +/- 0.00014	0.00106 +/- 0.00021	0.426
For Hire Vehicle Complaint	-0.00018 +/- 0.00011	0.00054 +/- 0.00015	0.376
Found Property	-0.00019 +/- 0.00003	0.00038 +/- 0.00005	0.598
Homeless Encampment	-0.00113 +/- 0.00031	0.00207 +/- 0.00045	0.375
Homeless Person Assistance	-0.00062 +/- 0.00015	0.00124 +/- 0.00021	0.437
Panhandling	-0.00004 +/- 0.00001	0.00009 +/- 0.00002	0.316
Taxi Complaint	-0.00729 +/- 0.00103	0.01464 +/- 0.00149	0.667
Taxi Report	-0.00028 +/- 0.00003	0.00055 +/- 0.00004	0.706

ANNEX 2

Coefficients and standard deviations associated with the neighborhood characteristics for types of requests with R2 greater than 0.2

Table 19. Coefficients and standard deviations associated with the neighborhood characteristic for types of requests with R2 greater than 0.2

Type of Request	Median House Value	Median Age	Median Rent	Median Income	Cars per capita	Building Density	Proportion of Commercial area	Proportion of retail area	Proportion of Office area	Proportion of Residential area	Inbound Commute Density	Population Density	Proportion of Tenants	const	R2 os
Lead	8.5e-07+/-3.6e-07	-0.013 +/- 0.011	-0.000 +/- 0.000	5.08e-06+/-4.17e-06	-0.123 +/- 0.404	1.784 +/- 0.631	0.883 +/- 1.114	-3.255 +/- 1.777	-2.939 +/- 0.919	0.550 +/- 1.045	4.6e-07+/-6.4e-07	-0.000 +/- 0.000	0.578 +/- 0.732	-1.212 +/- 1.577	0.144
New Tree Request	3.49e-06+/-4.6e-07	-0.006 +/- 0.014	0.000 +/- 0.000	-0.000 +/- 0.000	-0.511 +/- 0.457	0.921 +/- 0.765	3.183 +/- 1.445	0.646 +/- 2.252	-1.417 +/- 1.169	3.301 +/- 1.367	5.1e-07+/-8.2e-07	-0.000 +/- 0.000	-2.563 +/- 0.865	-2.444 +/- 2.022	0.282
Other Enforcement	1.8e-06+/-3.3e-07	-0.042 +/- 0.010	-0.000 +/- 0.000	6.3e-07+/-3.65e-06	-0.603 +/- 0.320	0.117 +/- 0.537	-3.219 +/- 1.013	6.561 +/- 1.586	0.028 +/- 0.820	-2.428 +/- 0.959	-3e-08+/-5.8e-07	-7.75e-06+/-2.66e-06	-1.590 +/- 0.612	4.771 +/- 1.421	0.167
General Construction/Plumbing	1.65e-06+/-3.4e-07	-0.014 +/- 0.010	-0.000 +/- 0.000	-4.27e-06+/-3.78e-06	0.273 +/- 0.323	1.318 +/- 0.526	-0.557 +/- 1.050	3.938 +/- 1.649	-0.245 +/- 0.812	-0.368 +/- 0.998	-2e-07+/-6e-07	-4.68e-06+/-2.74e-06	-0.165 +/- 0.629	-0.502 +/- 1.440	0.117
Graffiti	1.48e-06+/-5e-07	-0.045 +/- 0.015	0.001 +/- 0.000	-0.000 +/- 0.000	-0.290 +/- 0.478	1.603 +/- 0.779	0.954 +/- 1.553	9.799 +/- 2.438	-1.810 +/- 1.201	0.621 +/- 1.476	-5.9e-07+/-8.8e-07	-0.000 +/- 0.000	-1.070 +/- 0.930	0.119 +/- 2.130	0.137
Taxi Complaint	1.3e-07+/-4.9e-07	0.000 +/- 0.014	0.000 +/- 0.000	0.000 +/- 0.000	-0.491 +/- 0.474	0.587 +/- 0.773	0.486 +/- 1.510	3.806 +/- 2.454	-3.327 +/- 1.170	-1.338 +/- 1.428	-9e-07+/-8.5e-07	4.35e-06+/-3.95e-06	0.484 +/- 0.912	-1.738 +/- 2.075	0.234

Type of Request	Median House Value	Median Age	Median Rent	Median Income	Cars per capita	Building Density	Proportion of Commercial area	Proportion of retail area	Proportion of Office area	Proportion of Residential area	Inbound Commute Density	Population Density	Proportion of Tenants	const	R2 os
HEAT/HOT WATER	0	0	0	0	0	0	0	0	0	0	0	0.000 +/- 0.000	0	-1.075 +/- 0.091	0.152
Noise	1.03e-06+/-2.4e-07	0	0	6.8e-07+/-1.68e-06	0	0	0	0	0	0	6.7e-07+/-3.2e-07	0	0	-0.904 +/- 0.112	0.149
Construction	0	0	0	0	1.576 +/- 0.303	0	0	0	-3.225 +/- 3.348	0	-2.2e-06+/-3.12e-06	0	0	-1.884 +/- 0.299	0.189
HEATING	0	0	0	0	0	0	0	0	0	0	0	0.000 +/- 0.000	0	-0.950 +/- 0.089	0.147
Sidewalk Condition	1.64e-06+/-3.6e-07	-0.016 +/- 0.011	-0.000 +/- 0.000	7.1e-06+/-4.03e-06	-0.508 +/- 0.343	1.007 +/- 0.558	-0.787 +/- 1.113	2.581 +/- 1.754	-0.466 +/- 0.862	-1.073 +/- 1.059	-1.28e-06+/-6.3e-07	-5.04e-06+/-2.94e-06	-1.062 +/- 0.673	1.353 +/- 1.530	0.133
Dead Tree	2.42e-06+/-4e-07	-0.008 +/- 0.012	0.000 +/- 0.000	-7.42e-06+/-4.48e-06	0.780 +/- 0.383	-0.594 +/- 0.624	2.123 +/- 1.244	-2.442 +/- 1.953	-1.502 +/- 0.962	1.971 +/- 1.182	1.4e-07+/-7.1e-07	2.34e-06+/-3.24e-06	0.600 +/- 0.745	-3.586 +/- 1.706	0.118
Street Condition	9.6e-07+/-2.5e-07	-0.021 +/- 0.007	-0.000 +/- 0.000	3.44e-06+/-2.82e-06	-0.066 +/- 0.241	-0.367 +/- 0.393	-0.767 +/- 0.783	2.157 +/- 1.230	0.429 +/- 0.606	-1.477 +/- 0.744	-1.5e-07+/-4.4e-07	-3.48e-06+/-2.04e-06	-0.748 +/- 0.469	2.500 +/- 1.074	0.236
Snow	9e-07+/-3.1e-07	-0.029 +/- 0.009	-0.000 +/- 0.000	1.54e-06+/-3.5e-06	0.417 +/- 0.299	0.337 +/- 0.487	-0.646 +/- 0.970	1.311 +/- 1.524	-1.142 +/- 0.751	-0.297 +/- 0.923	4.5e-07+/-5.5e-07	-9.06e-06+/-2.53e-06	0.372 +/- 0.581	0.568 +/- 1.331	0.151

Type of Request	Median House Value	Median Age	Median Rent	Median Income	Cars per capita	Building Density	Proportion of Commercial area	Proportion of retail area	Proportion of Office area	Proportion of Residential area	Inbound Commute Density	Population Density	Proportion of Tenants	const	R2 os
Indoor Air Quality	0	0	0	0	0	0	0	0	0	0	9.4e-07+/-4.1e-07	0.000 +/- 0.000	0	-0.960 +/- 0.088	0.199
Food Establishment	1.25e-06+/-4e-07	0.026 +/- 0.012	0.000 +/- 0.000	5e-08+/-4.44e-06	0.271 +/- 0.380	0.476 +/- 0.618	-1.782 +/- 1.233	8.681 +/- 1.936	0.381 +/- 0.954	-2.228 +/- 1.172	-1.62e-06+/-7e-07	2.06e-06+/-3.21e-06	0.915 +/- 0.738	-1.347 +/- 1.690	0.284
Root/Sewer/Sidewalk Condition	2.15e-06+/-4.5e-07	-0.010 +/- 0.014	0.001 +/- 0.000	-0.000 +/- 0.000	1.359 +/- 0.432	2.583 +/- 0.740	2.802 +/- 1.357	4.507 +/- 2.154	-3.595 +/- 1.119	2.755 +/- 1.282	-2.8e-07+/-7.8e-07	-0.000 +/- 0.000	0.175 +/- 0.819	-5.201 +/- 1.934	0.264
Air Quality	1.37e-06+/-4.2e-07	-0.010 +/- 0.013	0.000 +/- 0.000	3.92e-06+/-4.74e-06	-0.743 +/- 0.416	-0.753 +/- 0.697	-1.334 +/- 1.315	3.629 +/- 2.059	-1.135 +/- 1.065	-1.920 +/- 1.245	-6e-07+/-7.5e-07	3.06e-06+/-3.45e-06	-0.573 +/- 0.794	1.433 +/- 1.845	0.133
Overgrown Tree/Branches	1.74e-06+/-3.7e-07	-0.051 +/- 0.011	0.001 +/- 0.000	-0.000 +/- 0.000	0.318 +/- 0.353	0.733 +/- 0.575	1.325 +/- 1.146	3.617 +/- 1.807	-2.263 +/- 0.888	1.974 +/- 1.091	-5.2e-07+/-6.5e-07	-0.000 +/- 0.000	-0.666 +/- 0.693	-1.064 +/- 1.576	0.247
Blocked Driveway	5.4e-07+/-4.1e-07	-0.025 +/- 0.012	0.002 +/- 0.000	-0.000 +/- 0.000	-0.384 +/- 0.396	1.656 +/- 0.638	-0.353 +/- 1.271	3.671 +/- 2.004	0.010 +/- 0.984	0.761 +/- 1.210	-5.1e-07+/-8.2e-07	-0.000 +/- 0.000	-1.066 +/- 0.773	-0.503 +/- 1.754	0.193
Damaged Tree	2.01e-06+/-3.3e-07	-0.036 +/- 0.010	0.001 +/- 0.000	-0.000 +/- 0.000	-0.001 +/- 0.309	-0.234 +/- 0.504	1.457 +/- 1.004	1.473 +/- 1.583	-2.002 +/- 0.778	1.843 +/- 0.955	3.2e-07+/-5.7e-07	-8.93e-06+/-2.65e-06	-1.157 +/- 0.607	-0.557 +/- 1.380	0.25
Illegal Tree Damage	1.58e-06+/-3.9e-07	-0.011 +/- 0.011	0.001 +/- 0.000	-7.79e-06+/-4.2e-06	0.303 +/- 0.355	-0.233 +/- 0.581	-0.165 +/- 1.142	7.565 +/- 2.041	-1.339 +/- 0.888	-0.031 +/- 1.083	6.5e-07+/-6.5e-07	-1.08e-06+/-3.05e-06	-0.213 +/- 0.700	-1.009 +/- 1.572	0.121

ANNEX 3

Visual representation of results obtained in step number 1 for selected types of requests

Figure 1. Coefficients for selected request types by resident age category

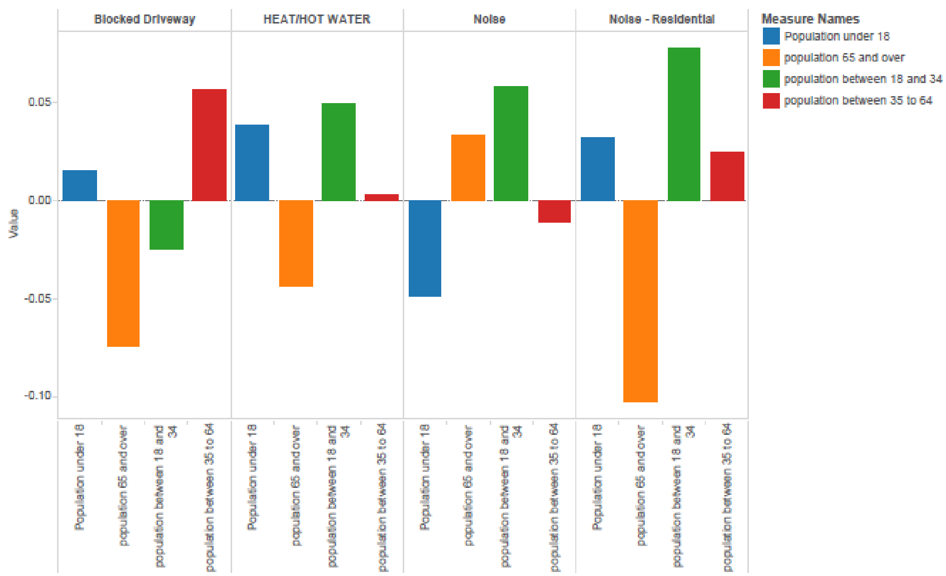


Figure 2. Coefficients for selected request types by workers age category

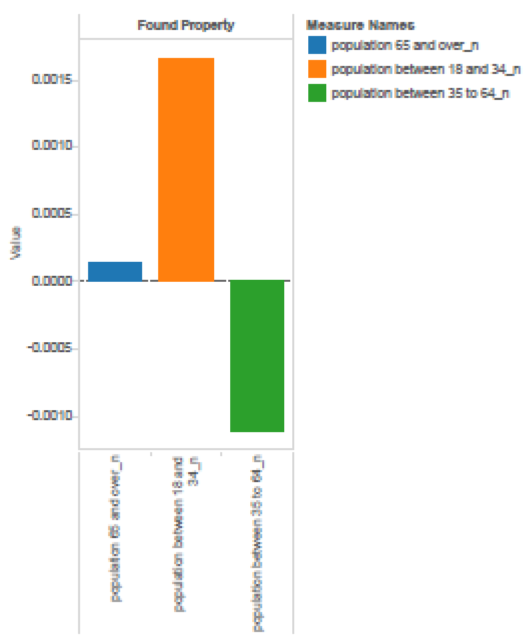


Figure 3. Coefficients for selected request types by resident Race type

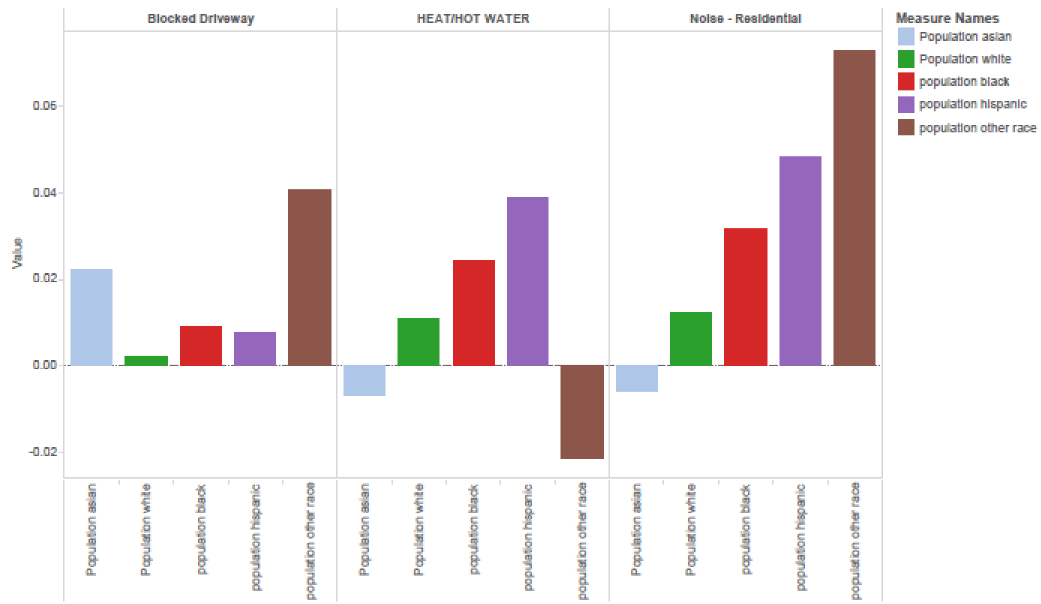


Figure 4. Coefficients for selected request types by resident education level

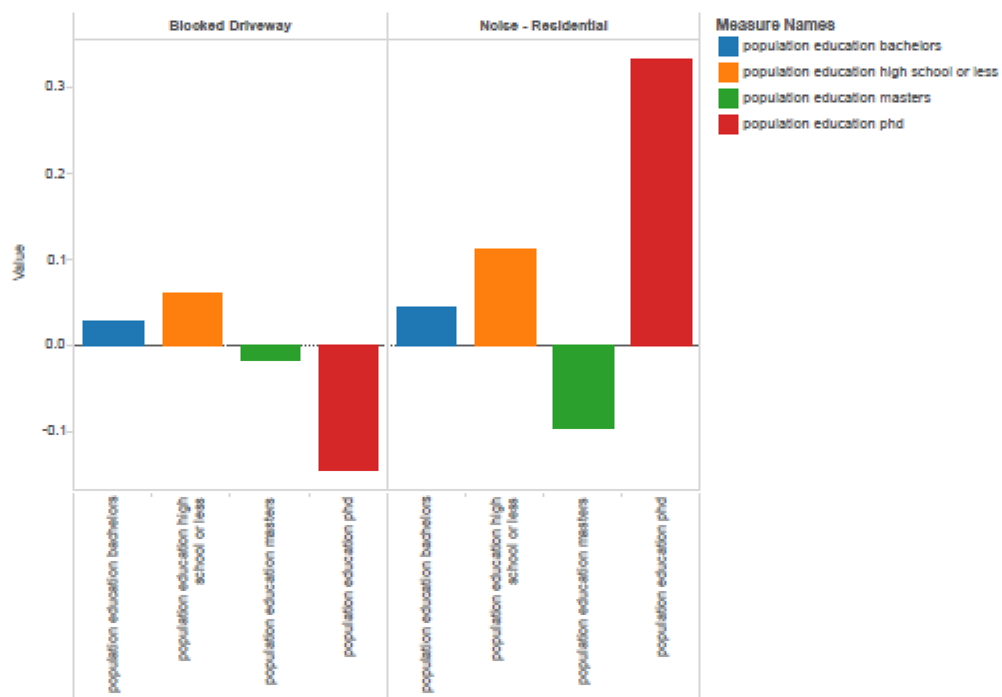


Figure 5. Coefficients for selected request types by worker education level

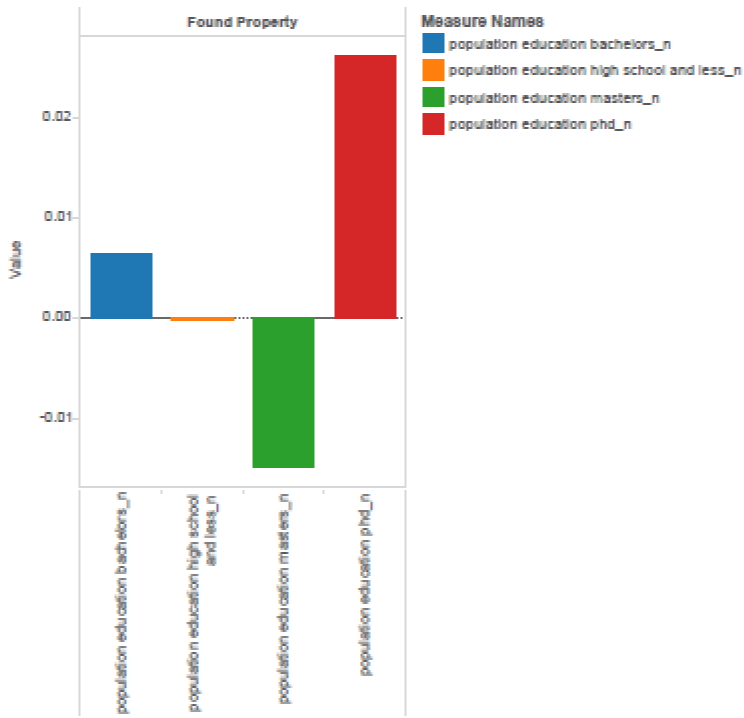


Figure 6. Coefficients for selected request types by resident household income level

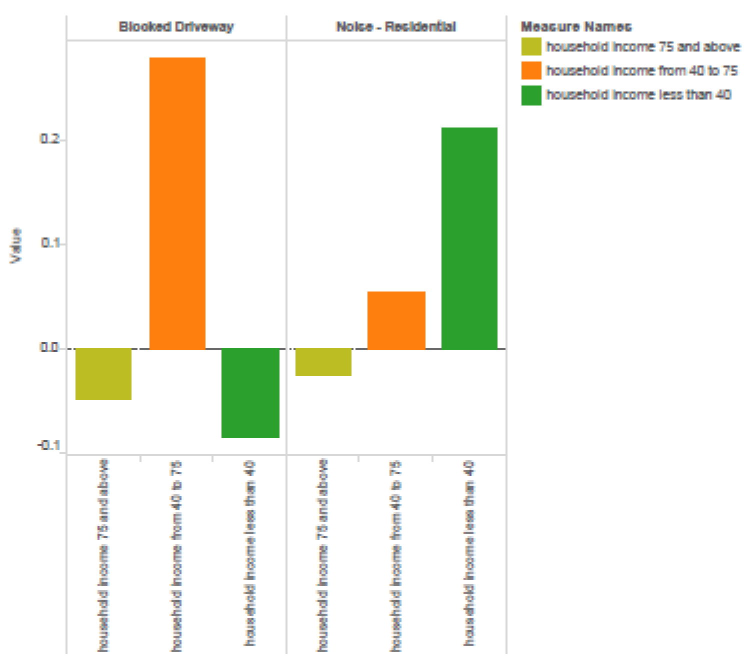


Figure 7. Coefficients for selected request types by worker household income level

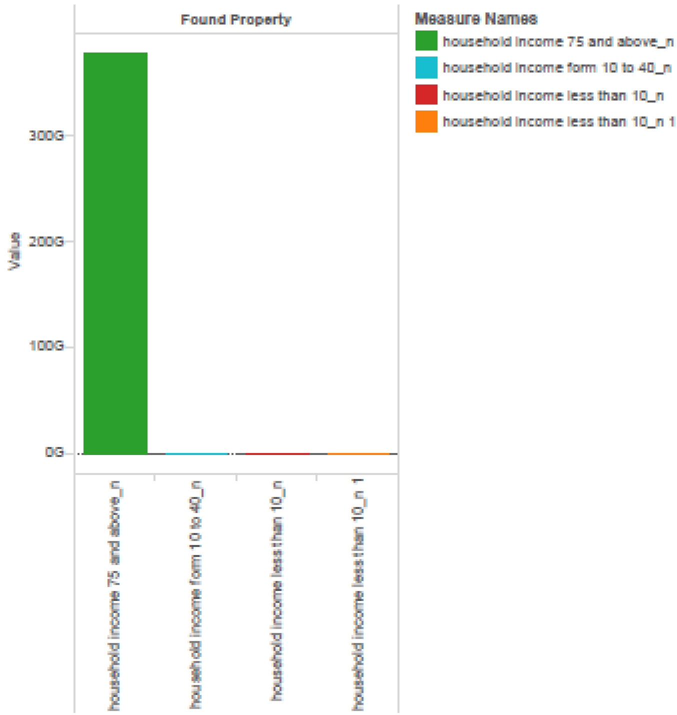


Figure 8. Coefficients for selected request types by resident Owner or renter

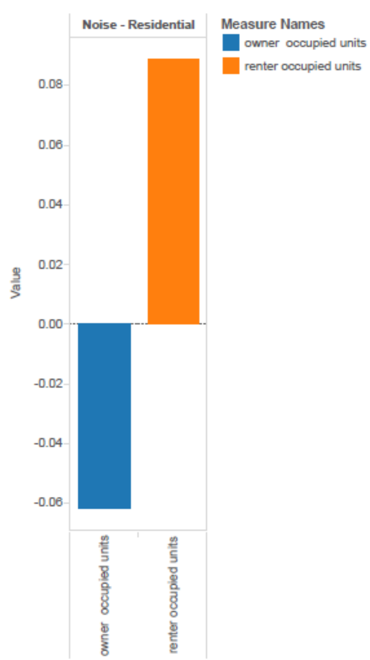


Figure 9. Coefficients for selected request types by worker Owner or renter

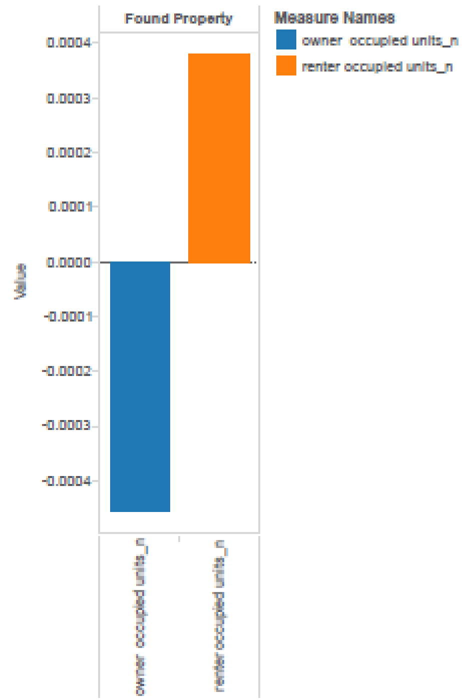


Figure 10. Coefficients for selected request types by resident rent by price

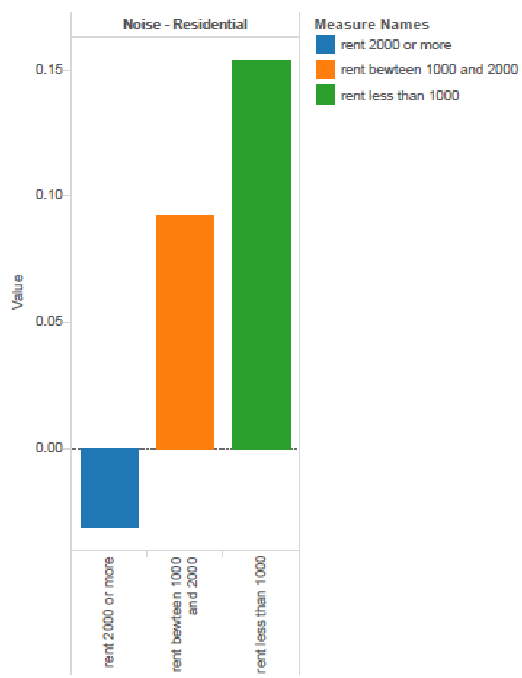


Figure 11. Coefficients for selected request types by worker rent by price

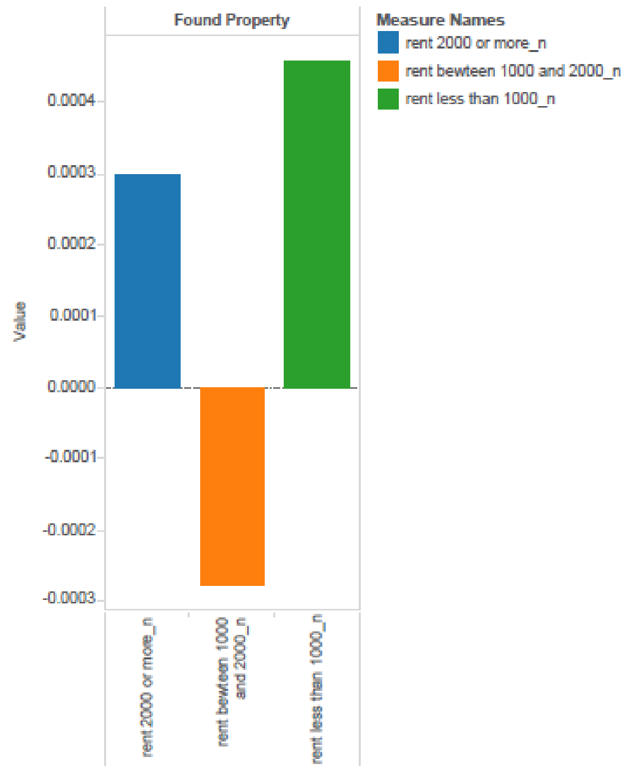


Figure 12. Coefficients for selected request types by resident transportation method

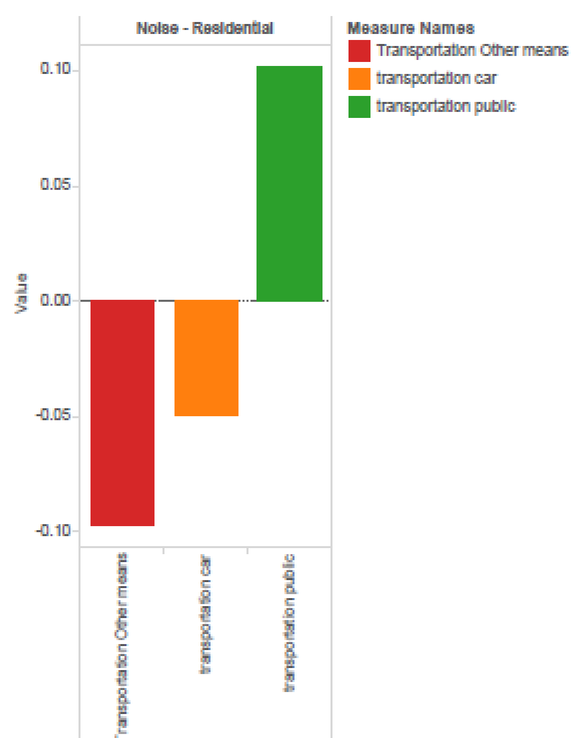


Figure 13. Coefficients for selected request types by worker transportation method

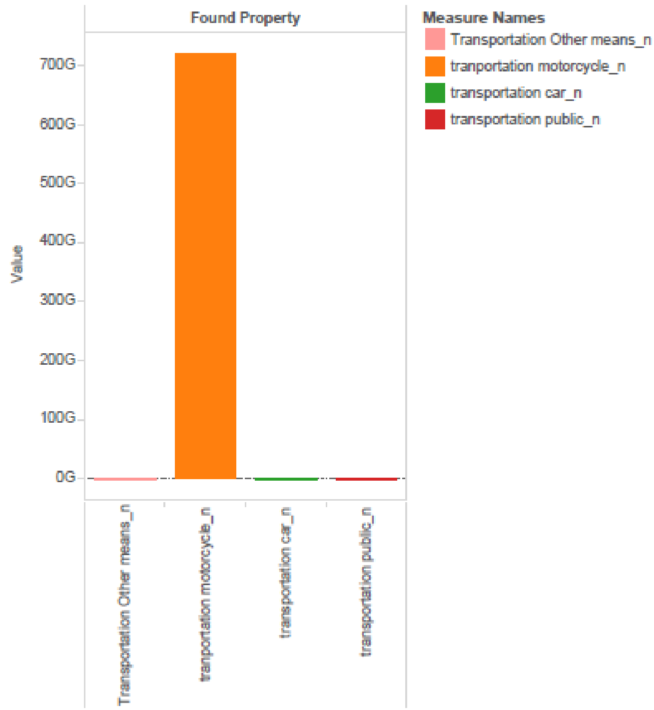


Figure 14. Coefficients for selected request types by resident type of household

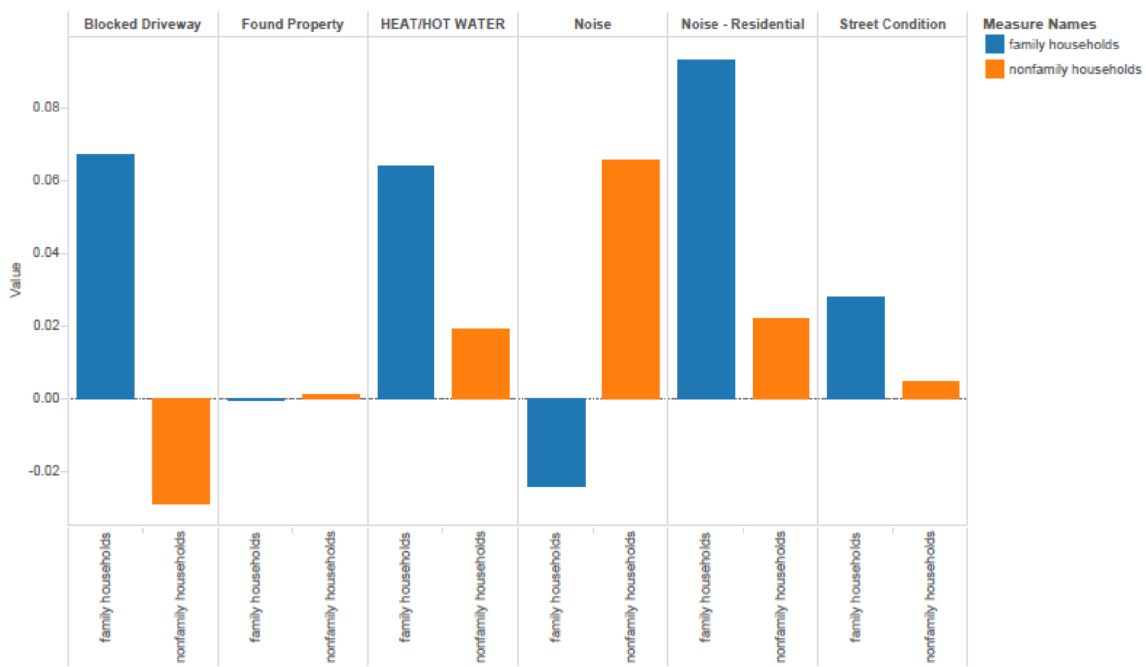
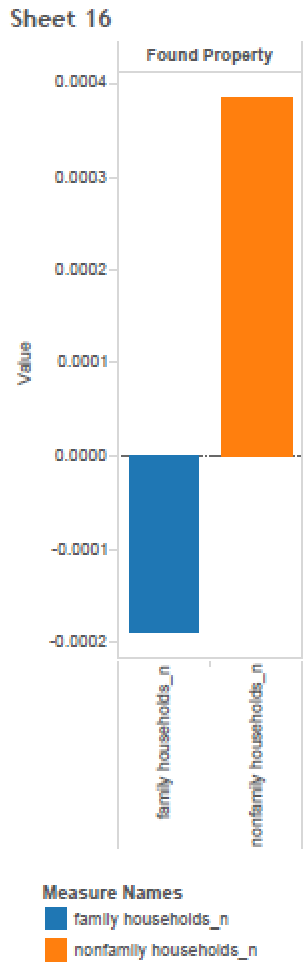


Figure 15. Coefficients for selected request types by worker type of household



Annex 4
Distribution of the error between the observed number of requests and the number of requests explained by the demographic variables