Bias in 311 Response

ECO395M: Data Mining

Scott Cohn

Last compiled on 11 May, 2021

Abstract

This paper explores the interaction between citizen satisfaction in local government with 311 service usage, and related bias in who may be using these services more. I combine open-access data on 311 service usage and property violations from KCMO OpenData with Census data, and a proprietary citizen satisfaction survey conducted by the Kansas City Office of Performance Management. I find that certain demographics like gender, race, and income affect likelihood of contacting city services, but the strongest predictors of 311 usage are pre-existing and regular interactions with city services — like recycling, visiting parks, and visiting the city website. **Disclaimer:** Burnout this semester has killed my ability to produce work that I deem high quality. Regardless, I hope you find this satisfactory.

Keywords: 311 Complaint, City governance, Classification, Bias

Introduction

Many cities use 311 systems to capture non-emergency complaints of their citizens in real time ("About 311, our mission." 2021). These cities include Austin, New York, Kansas City, Washington, DC, and more. In parallel, access to these data are often granted to the public via open data portals. These portals allow access to many types of city metrics, including: traffic, 311, city governance, and transit. This data availability has led to many governance decisions. For example, the New York City Fire Department (FDNY) uses a tool called FireCast 3.0 to analyze three years' worth of data for every building in the city and assign a fire risk score.[1] Data availability combined with smart governance allows a city department to be more effective in their response.

However, some instances of city department response may be prone to bias. Bias can also exist in the ways that residents make use of city services, in addition to bias of response to them. Here, I ask who is asking for help from their city government? I make use of 311 data, publicly available from Kansas City's (KC) OpenData portal, in conjunction with results from a citizen satisfaction survey administered by the Kansas

---

[1] https://www.nfpa.org/-/media/Files/News-and-Research/Resources/Research-Foundation/Current-projects/Smart-FF/SmartFFNFPAJournalDataDriven.ashx

City Office of Performance Management (OPM) to explore potential bias in 311 reporting ("311 call center service requests" 2021; "Citizen and business satisfaction survey results." 2021).

The use of the citizen satisfaction survey, administered quarterly, provides depth to the 311 data, allowing me to estimate likelihood of 311 contact from demographic and socioeconomic characteristics as well as political engagement and overall citizen satisfaction. The survey is a stratified sample across city council districts based on a balanced population against census demographics.

Following Kontokosta and Hong (2018), I hypothesize that increased citizen satisfaction and political participation is associated with increased likelihood for 311 usage. Largely, I replicate their research process. Although, it should be noted that do not publish their code or data (or respond to emails). All data-pulls, cleaning, and model creation is mine (errors included). This paper is organized as follows: section two describes the data sources and estimation methodology, section three presents the results, and section four concludes.

Data and Methods

I integrate numerous datasets, described in Table 1, to pursue the research question of whether there is bias in 311 usage. OPM graciously provided me with the data for the citizen satisfaction survey. This survey samples "are stratified across the city council districts based on a statistically significant random sample of the balanced population against census demographics, and responses are collected via mail, phone, and webpage" ("Citizen and business satisfaction survey results." 2021). The 311 reports and property violations data capture local disparities in 311 calls and outcomes. Figure 3 illustrates the 311 service request process.

Dataset	Records	Time range	Temporal Resolution	Spatial Resolution	Source and Description
Citizen satisfaction survey	12,644	2014-2017	Quarterly	Individual (household)	Citizen survey on level of satisfaction with city services, use of city services, and household characteristics each fiscal year provided by Kansas City (Missouri)
311 service requests	380,687	2013-2017	Daily	GPS (X, Y)	All citizen service requests and non-emergency complaints, provided by KC 311 through OpenData KC
Property violations	410,460	2015-2018	Daily	GPS (X, Y)	Neighborhood code enforcement violations issued by the KC Neighborhood and Housing Service (NHS), provided by KC NHS through OpenData KC
Census population	240	2010	Annual	Neighborhood	KC Census population in 2010, U.S. Decennial Census and American Community Survey

Table 1: Description of dataset used

Figure 1 illustrates the spatial patterns of 311 call and response. According to the survey, 47% have reported using 311 at least once. In 2016, 103,955 complaints were reported. Of the total reported, 54% (55,743) were related to code enforcement, which is handled by the Neighborhoods and Housing Services Department (NHS). The NHS provides inspections and will close the case unless and inspection is found. Of the cases given the NHS, 22% (12,411) ended up with served violations. A single 311 call that is routed to the NHS can yield many violations. These 12,411 calls that yielded violations, resulted in 71,356 separate violations. The location of these vary wildly over the city.

(a) 311 reports at neighborhood level

(b) 311 reports per capita at neighborhood level

(c) Property violations at neighborhood level

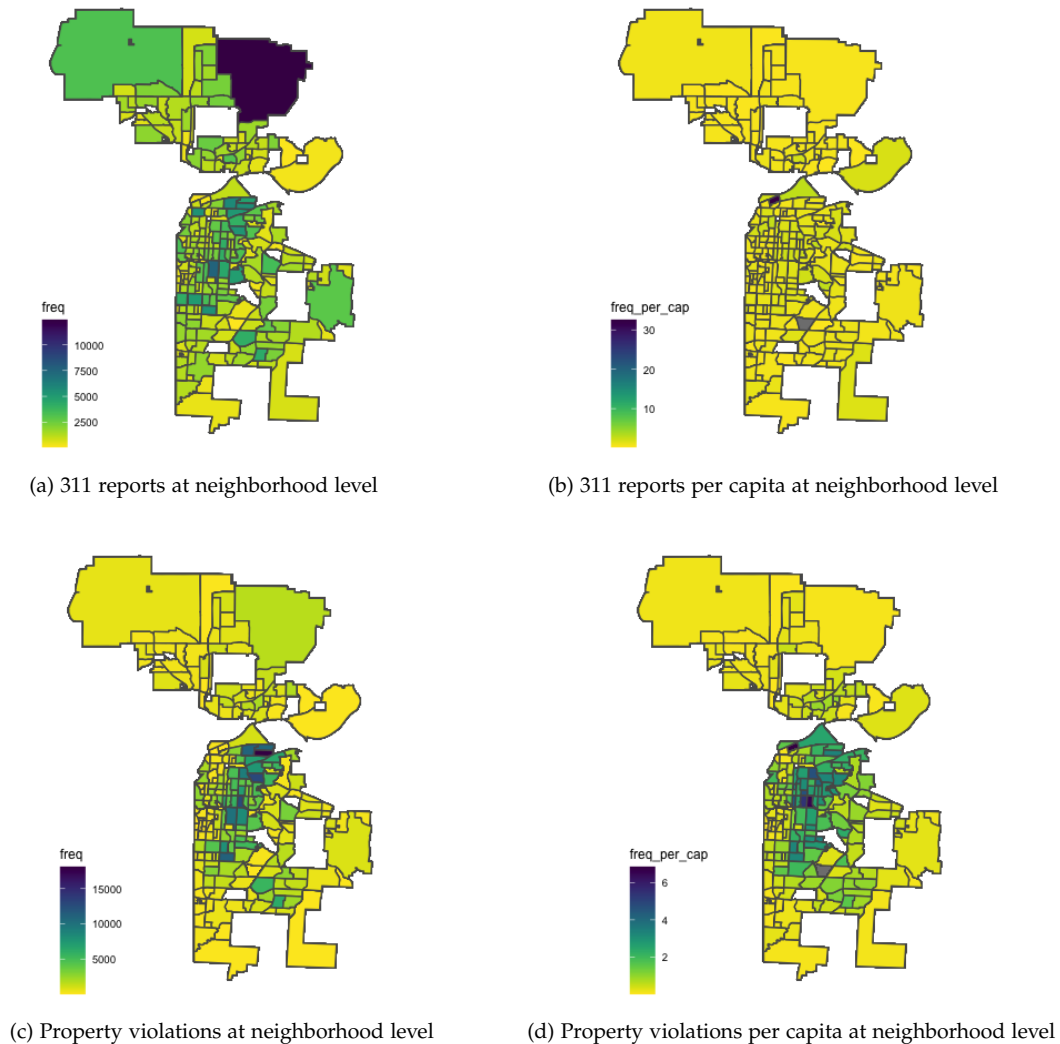(d) Property violations per capita at neighborhood level

Figure 1: Spatial patterns of 311 reports and property violations

The survey data provides insight on the demographics of households that have and have *not* contacted 311. Since the survey data is randomly sampled across neighborhoods, there should be no disparity due to individual level factors. The survey is split on a single question:

"Have you or anyone in your household contacted City's 311 Call Center in the past year?"

I then use relevant characteristics from the survey to predict the likelihood of 311 usage. To predict the likelihood of 311 usage, I develop two models: a penalized logistic regression and a random forest model. The dependent variable is the binary response to above question:

- **0** (311 User) : Respondent has used 311
- **1** (311 Non-User): Respondent has not used 311

The coefficients used, and their categories/type are in Table 2.

The logistic regression uses a tuned penalty hyperparameter to give less weight to less relevant predictors in the modeling process. This penalty is the total amount of regularization in the model and is tuned over an expanded grid with the alpha value. I have set $\alpha = 1$ in this specification, leading to a pure LASSO.

The random forest model uses two tuned hyperparameters with 1000 trees: `mtry` and `min_n`. These are the number of variables to possibly split at each node and the minimal node size, respectively. Again, these are tuned over an expanded grid. This process is sped up using parallelization.

Results

Table 3 and Figure 2 show the results of the logistic model. Figure 2 also shows the results of the random forest model. Both results perform similarly (AUC of 0.67 and 0.655). These are comparable with the results of Kontokosta and Hong (2018) who achieve an AUC of 0.69 using a logistic regression.[2]

The results suggest that older women with lower incomes, are non-white, and are property owners are more likely to make use of 311 services. The intuition behind these demographics make sense. Households with lower income may be more reliant on these services to take care of non-emergencies and households with lower incomes may be more likely to face these issues in the first place. Since the survey is anonymous and unable to be linked directly to census data and voting records, these kind of extrapolations are made with caution.

Further, as hypothesized, civic engagement seems to be more strongly associated with usage of 311 services. Those who visit one or more of the city website, the community website, and park, as well as those who recycle, are associated most strongly with 311 service requests. This might suggest that interaction with city governance more regularly might make a household more aware that 311 is available, or increase the service's salience in their life. Figure 4 in Appendix A illustrates measures of variable importance for both

---

[2]No details are given in their paper about how they clean or adjust covariates.

| Category	Variable	Questionnaire
Demographic/socioeconomic characteristics	Gender	Gender (Categorical)
	Age	What is your age (Categorical)
	Income	Total annual household income (Categorical)
	Race/Ethnicity 1	Describe your race ethnicity (Categorical)
	Race/Ethnicity 2	Are you Hispanic Latino Spanish (Categorical)
	Home ownership	Own or rent current residence (Categorical)
	Internet access	Have access to Internet at home (Categorical)
	Recycling	How often use curbside recycling (Categorical)
Civic engagement	City website	Visited city's website (Categorical)
	Community center	Visited KCMO community center (Categorical)
	Park	Visitied any parks in KCMO (Categorical)
Citizen satisfaction	Neighborhood service	Quality of neighborhood service (Ordinal)
	Quality of life	Quality of life in the city (Ordinal)
	Safety	Feeling of safety in the city (Ordinal)

Table 2: Input variables
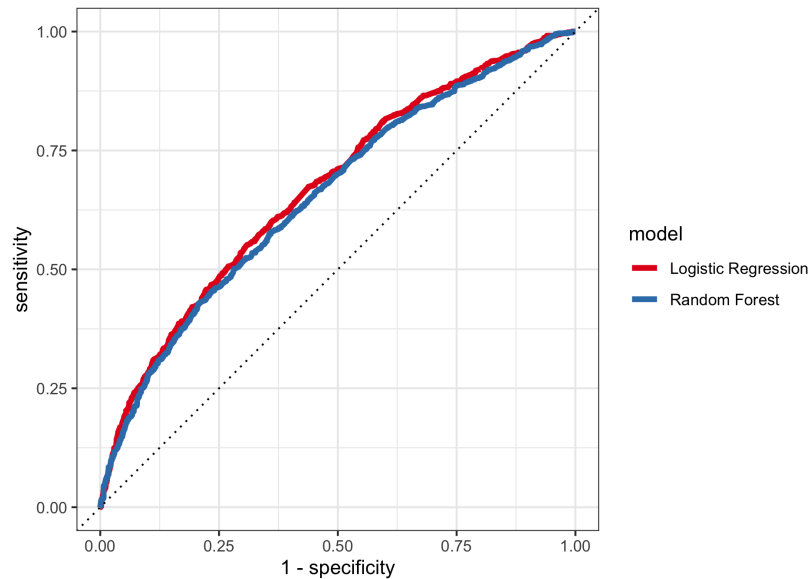
Figure 2: ROC (Receiver Operating Characteristics) Curve of the logistic regression and random forest model

Table 3: Results of the logistic regression model for 311 use

Term	Estimate
(Intercept)	-1.376
female	0.091
age_18_24	-0.786
age_25_34	-0.509
age_35_44	0.049
age_45_54	0.000
age_55_65	0.091
age_over65	-0.070
income_under30	0.096
income_30_59	-0.126
income_60_99	-0.260
income_over100	-0.587
asian_pac_isl	-0.497
white	-0.058
amer_ind_esk	-0.080
black	0.352
race_other	0.198
hispanic	-0.016
own_prop	0.365
rent_prop	-0.014
internet_access	-0.054
recycle_freq_weekly	0.820
recycle_freq_biweekly	0.655
recycle_freq_monthly	0.592
recycle_freq_never	0.430
visit_city_website	0.776
visit_comm_center	0.237
visit_park	0.296

models.

Overall, these results suggest non-trivial associations with between socioeconomic characteristics, civic engagement, and citizen satisfaction with 311 reporting behavior.

Conclusion

Previous studies on 311 reporting has to extrapolate the characteristics of the household making the report from unlinked census data about the area. Information is thus limited to expected income for the building, and other associated characteristics with the neighborhood or block-level characteristics. This was an obstacle in analyzing 311 data from New York City's OpenData portal. The citizen satisfaction survey in Kansas City is unique in that it allows for a link between 311 usage and characteristics of interest that allow the city to understand how their services are being used and to what extent they are valued.

Acknowledgements

References

"311 call center service requests: 2007 - march 2021," (http://www.kcmo.org/CKCMO/Depts/CityManagersOffice/The3-1-1ActionCenter/index.htm, 2021).

"About 311, our mission." (http://www1.nyc.gov/311/our-mission.page, 2021).

"Citizen and business satisfaction survey results." (https://dashboards.mysidewalk.com/kansas-city-mo-citizen-survey, 2021).

Kontokosta, Constantine E, and Hong, Boyeong, "Who calls for help?" (2018) (Bloomberg Data for Good Exchange Conference).
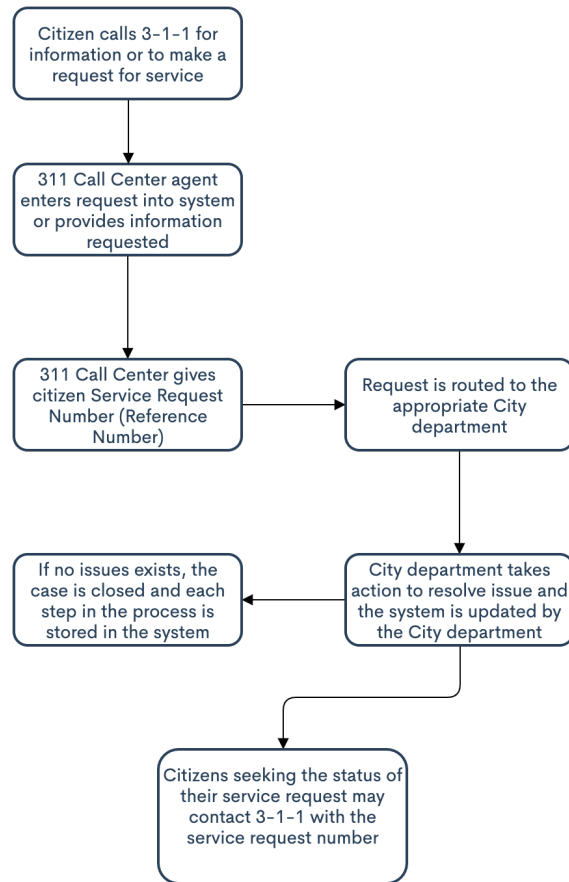
Appendices

Appendix A – Figures

Figure 3: 311 Service Request Process

Figure 4a — Logistic regression: recycle_freq_weekly, age_18_24, visit_city_website, recycle_freq_biweekly, recycle_freq_monthly, income_over100, asian_pac_isl, age_25_34, recycle_freq_never, income_60_99, visit_park, visit_comm_center, own_prop, black, white, income_30_59, amer_ind_esk, female, rent_prop, age_over65 (x-axis: Importance).

Figure 4b — Random Forest: visit_city_website, recycle_freq_weekly, rent_prop, own_prop, age_25_34, visit_comm_center, black, visit_park, white, age_18_24, income_over100, recycle_freq_never, age_35_44, asian_pac_isl, female, age_55_65, age_45_54, internet_access, income_30_59, recycle_freq_biweekly (x-axis: Importance).

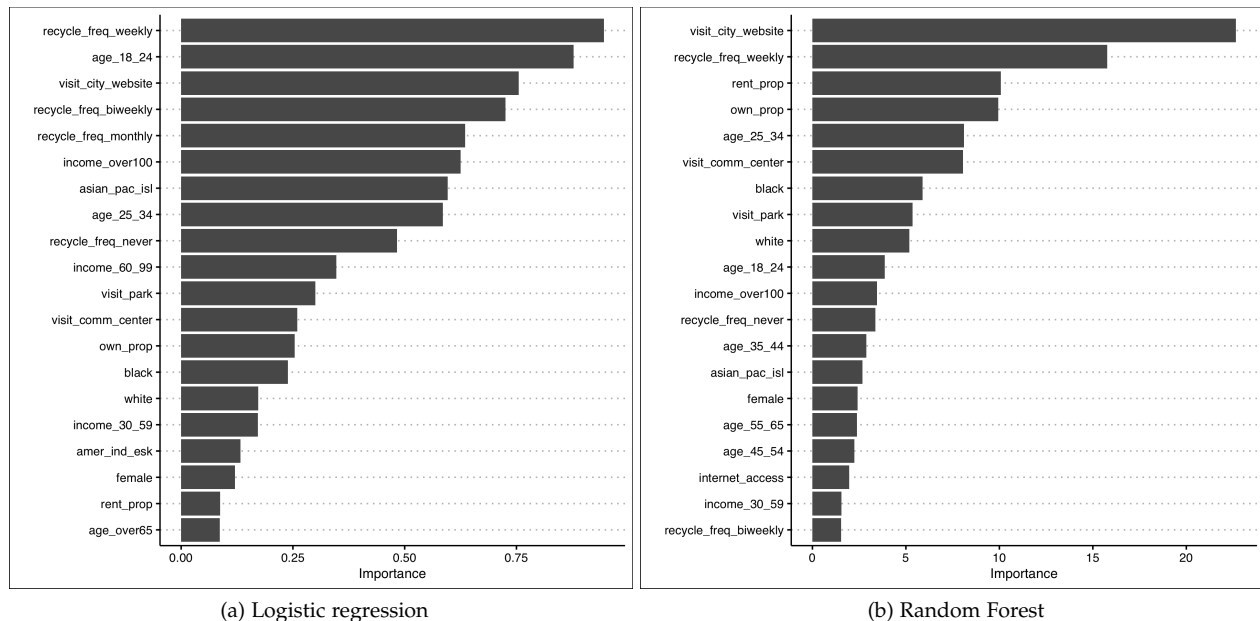(a) Logistic regression   (b) Random Forest

Figure 4: Variable importance

Appendix B – Session Information

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] ggthemes_4.2.0    kableExtra_1.2.1  forcats_0.5.0     stringr_1.4.0
##  [5] purrr_0.3.4       readr_1.3.1       tidyr_1.1.2       tibble_3.0.5
##  [9] tidyverse_1.3.0   mosaic_1.8.2      ggridges_0.5.2    mosaicData_0.18.0
## [13] ggformula_0.9.4   ggstance_0.3.4    dplyr_1.0.2       Matrix_1.2-18
## [17] ggplot2_3.3.3     lattice_0.20-41
##
## loaded via a namespace (and not attached):
##  [1] ggdendro_0.1-20   httr_1.4.2         viridisLite_0.3.0 jsonlite_1.7.1
##  [5] splines_4.0.3     modelr_0.1.7      assertthat_0.2.1  blob_1.2.1
##  [9] cellranger_1.1.0  yaml_2.2.1        ggrepel_0.9.1     pillar_1.4.4
## [13] backports_1.1.6   glue_1.4.2        statquotes_0.2.2  digest_0.6.27
## [17] RColorBrewer_1.1-2 polyclip_1.10-0  rvest_0.3.6       colorspace_1.4-1
## [21] htmltools_0.4.0   plyr_1.8.6        pkgconfig_2.0.3   broom_0.7.3
## [25] haven_2.2.0       webshot_0.5.2     scales_1.1.1      tweenr_1.0.1
```

```
## [29] ggforce_0.3.1     generics_0.1.0   farver_2.0.3       ellipsis_0.3.1
## [33] withr_2.2.0       cli_2.2.0        magrittr_2.0.1     crayon_1.3.4
## [37] readxl_1.3.1      evaluate_0.14    tokenizers_0.2.1   janeaustenr_0.1.5
## [41] fansi_0.4.1       fs_1.5.0         SnowballC_0.7.0    MASS_7.3-53
## [45] xml2_1.3.2        tools_4.0.3      hms_0.5.3          lifecycle_0.2.0
## [49] reprex_0.3.0      munsell_0.5.0    compiler_4.0.3     rlang_0.4.10
## [53] grid_4.0.3        rstudioapi_0.13  htmlwidgets_1.5.1  crosstalk_1.1.0.1
## [57] mosaicCore_0.8.0  rmarkdown_2.5    gtable_0.3.0       DBI_1.1.0
## [61] R6_2.4.1          gridExtra_2.3    lubridate_1.7.8    knitr_1.30
## [65] tidytext_0.2.5    stringi_1.5.3    Rcpp_1.0.4.6       vctrs_0.3.4
## [69] wordcloud_2.6     leaflet_2.0.3    dbplyr_1.4.4       tidyselect_1.1.0
## [73] xfun_0.21
```