San Antonio 311 Service Calls Analysis Proposal

By
Yixiao Li
Andrew Parson
Manuel Perez
Ben Polan
Dylan Rowe

The University of Texas at San Antonio
DA 6813

Background

The city of San Antonio, Texas has been making great strides in becoming a "Smart City" and have been making it easier for various departments to be connected. The main way they have been improving this interdepartmental connection for citizens is by improving their "311" line. They have principally improved the "311" line regarding their data collection of the different calls and requests from citizens. The city has been standardizing this data from 2011 to 2018 and will continue to do so to bring the city to a full "Smart City" status. In addition, they have made great strides in improving transparency of the city by making this data open source. Citizens have the ability to enter grievances, requests, and general city questions via web portal, call center, or mobile app. By allowing citizens the ability to contact the "311" line in a variety of ways the city hopes to raise public engagement, improve efficiency, and raise accountability. The 311 app is managed by Cityflag, a local startup company, and the call line is managed by the city of San Antonio. The services the city provides are for Animal Control Services, Graffiti Removal, Pothole Repair, Garbage/Recycle Services, Traffic Sign Maintenance, and many others.

Motivation

San Antonio, with its continued goal of transparency and free access to its data, has given its citizens an incredible opportunity and power. Now, ordinary people have the ability to make a tangible, positive impact in their community with the amazing amount of resources and data at hand. Our analysis aims to be an evaluation of San Antonio's continued efforts to be a "Smart City", specifically, to answer critical questions about its 311 service in hopes to improve foresight, responsiveness, and efficiency.

The analysis will primarily focus on:

- **Which types of request take the longest to fulfill?**
  Some requests require more time and resources. If the city can anticipate these requests, they can vastly improve their response time therefore providing the residents a better service.
- **What causes the delay in these requests?**
  If the factors which are delaying these requests can be identified, then they might shed light and provide insight on why certain requests are more prevalent and what can be done to counter that.
- **Can we identify trends in call frequency and anticipate high volume call times?**
  Identifying the highest and lowest volume call times and dates will help the city in terms of staffing decisions for that particular time and date. If service requests are more frequent in the summer, then more manpower will be required for the summer months. Inversely, periods with low call frequency won't require as much staff.
  We also want to identify when and where to anticipate different types of requests. Instances of vandalism will likely occur more frequently during the summer months

when school is not in session. Less obvious insights should be detectable through further analysis of the data.

- **Does the source of the request (call, online, or mobile) impact response time?** The city needs to know if their call and online systems are working at a comparable rate. If not, it is possible they need more people in the call center, or a more user-friendly online reporting system. The city has a reward system for enticing citizens to use their mobile app, but would citizens use it if the response to their query was too slow?
- **Does the city council district impact response time?** Which areas of the city can use a boost in response time to their requests? Is there a difference among districts and response time? Is there an inherent bias toward one council district over another? Answering these questions can help improve the lives of citizens who aren't getting their problems solved in a timely manner. This can also potentially show a need for a shift in personnel from one district to another.

Description of the Data

The initial dataset is comprised of 554,676 observations with 18 variables describing the data. The data comes from data.sanantonio.gov and covers 311 service calls in the city of San Antonio. 311 is a service offered for non-emergency calls and complaints to report problems like graffiti or road damage. This dataset spans the time frame of 01/2011 to 12/18, an 8-year period. The variables describing the data include service call category, the date the case was opened and closed, the number of days late the case was responded to, address of the caller/complaint, case status, the department assigned the task, and several other descriptive features.

In moving forward with the analysis of 311 service calls we may subset the data to represent a more recent time frame. We are thinking about cutting the start date of our data from 01/2011 to potentially 01/2015, however, this is something we will investigate more to determine if this is a well-represented and efficient point to subset at.

As for the cleanliness of the data, the data is fairly clean and does not have too many missing values. All those missing values, however, are from the same column the Closed Date Time column. The missing values are all from cases of multiple reports from one person. The main potential issue of the data is the sheer large number of observations. Due to the large amount of in-depth information provided, we do not anticipate additional information needing to be augmented with the dataset.

Proposed Analysis

Our analysis of the 311 data will comprise of logistic and linear regressions, random forest models, time series analysis, chi-square tests and visual representations of the data analysis such as histograms to help interpret the information.

Through the use of logistic regression, we will try to explore what variables caused a request to be late. The use of this type of model can be utilized because the response variable, whether a service was late or not, is categorical. The explanatory variables, such as department

of request and source of request, may give insight into what causes these requests to be late. Hopefully, we may see if it's a department issue, some may need more resources, or type of contact issue, the city may need to rework the way they are contacted via certain platforms.

A random forest will help to identify the reasons why a problem was solved late. The decision trees in the forest should accurately predict the late problems found in the logistic regression with flexibility and without concern for types of variables. These insights could help the city be more efficient in its response to these specific problems and provide an amount of foresight in how to prevent problems from even happening if proactive steps are taken.

Based on the timestamps of when the request was first issued we hope to analyze through a time series analysis various quality of type of requests. We hope to see if there are certain seasonal relationships between types of requests, such as graffiti maintenance, and perhaps month or period of year, such as summer. Through this type of analysis, we hope to provide insight to the city so that they may know when to provide more resources to certain departments or when to hire more seasonal workers to mitigate an overflow of requests for certain departments.

We want to explore the variables type and council district to see the main differences and similarities between those two categorical variables. We will also explore a relationship between whether a case was late versus the type of case that was reported. In order to accomplish this analysis, we will run several chi-square tests to break the variables down into their respective category frequencies and cross analyze them by category and interaction with the other respective variable. This can help us determine which type of case is reported the most per council district in San Antonio and which types of cases tend to be completed late. In doing this we can potentially identify departments that may need extra funding or additional resources.

Additionally, we want to explore the differences between cases made using the new 311 app established in August 15th, 2018 versus the old app used that was provided by outside providers, Lagan. In doing this analysis we will determine whether the new app is more efficient in communicating a service call and explore the relationship between which app was used and the promptness with which a case was resolved as well as whether the new app has had an impact in decreasing the usage of other methods of filing a service complaint.

A basic linear regression will be used to see if their relationship between the number of late days, the response variable, and the type of request, an explanatory variable. This type of analysis can be used to show if there are certain requests that take longer for city departments to complete. Consequently, this type of analysis can lead to further analysis regarding what may have a strong correlation to the requests that are considered late. This may lead the city to contact special services for certain requests so that it may mitigate the lateness of service requests so that certain departments resources are not overtly depleted.

References

"All Service Calls" City of San Antonio Open Data, https://data.sanantonio.gov/dataset/service-calls.
"311 Customer Service Office" City of San Antonio, https://www.sanantonio.gov/CustomerService.

Appendix – Variable Descriptions

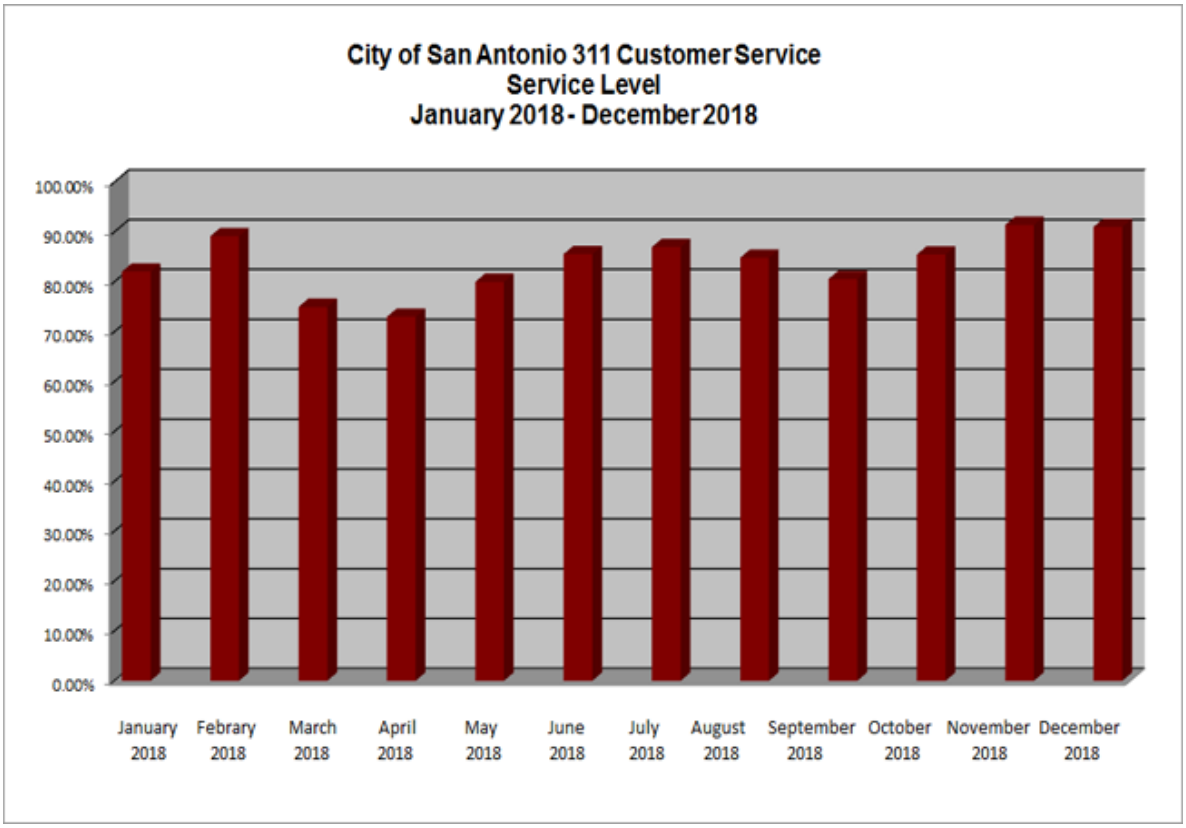
The raw data: https://data.sanantonio.gov/dataset/service-calls

- Includes observations from September 29, 2011 to November 9, 2018

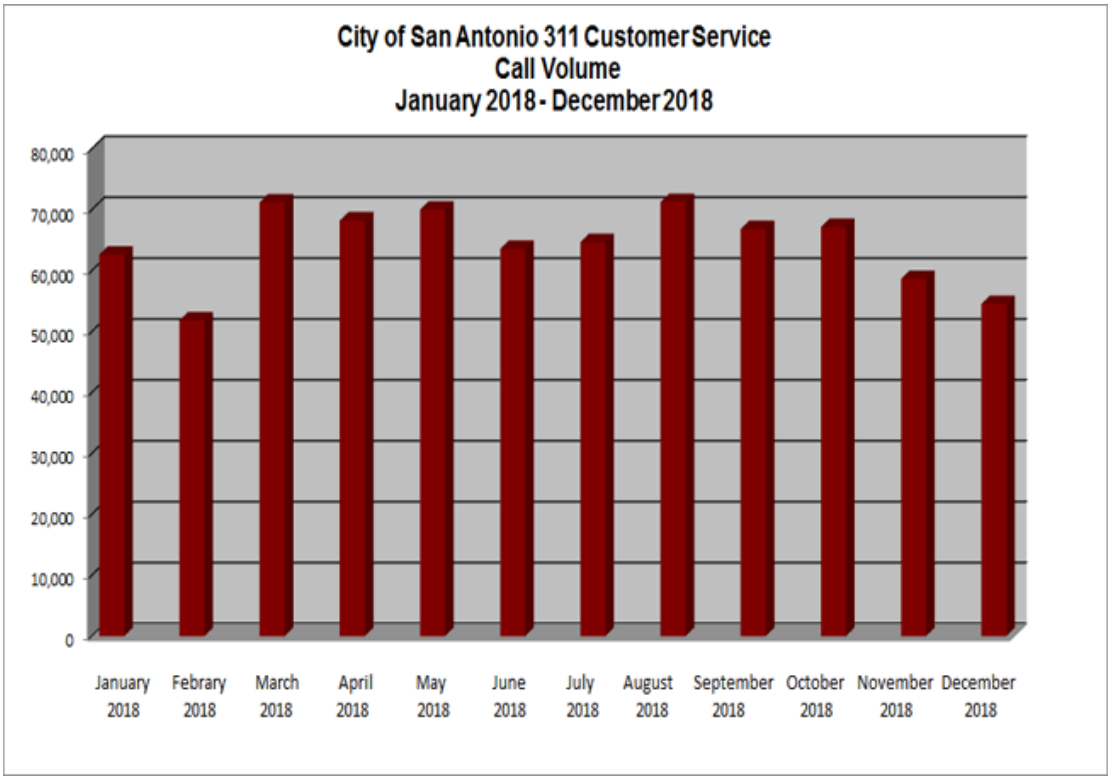
The data came as one CSV Excel file with 554,677 observations with 18 variables.
The variables are described as follows:

- _id: The number for the respective case on the spreadsheet, automatically assigned by the export.
- CATEGORY: This general category was developed to place 311 services in a high level category, different than their respective department.
- CASEID: The unique case reference number is assigned by the 311 Lagan customer relationship management system.
- OPENDATETIME: The date and time that the case was submitted.
- SLA_Date: The SLA Date is the due date and time for the request type based on the service level agreement (SLA).
- CLOSEDDATETIME: The date and time that the case/request was closed. If blank, the request has not been closed as of the Report Ending Date.
- Late (Yes/No): This indicates whether the case has surpassed its Service Level Agreement due date for the specific service request.
- Dept: The City department to whom the case is assigned.
- REASONNAME: The department division within the City department to whom the case is assigned.
- TYPENAME: The service request type name for the issue being reported. Examples include stray animals, potholes, overgrown yards, junk vehicles, traffic signal malfunctions, etc.
- CaseStatus: The status of a case which is either open or closed.
- SOURCEID: The source id is the method of input from which the case was received.
- OBJECTDESC: The address or intersection for the reported case/service requested.
- Council District: The Council District number from where the issue was reported.
- XCOORD: The X coordinate for the case reported.
- YCOORD: The Y coordinate for the case reported.
- Report Starting Date: The start date range for the case open date for this extract file.
- Report Ending Date: The end date range for the case open date for this extract file.

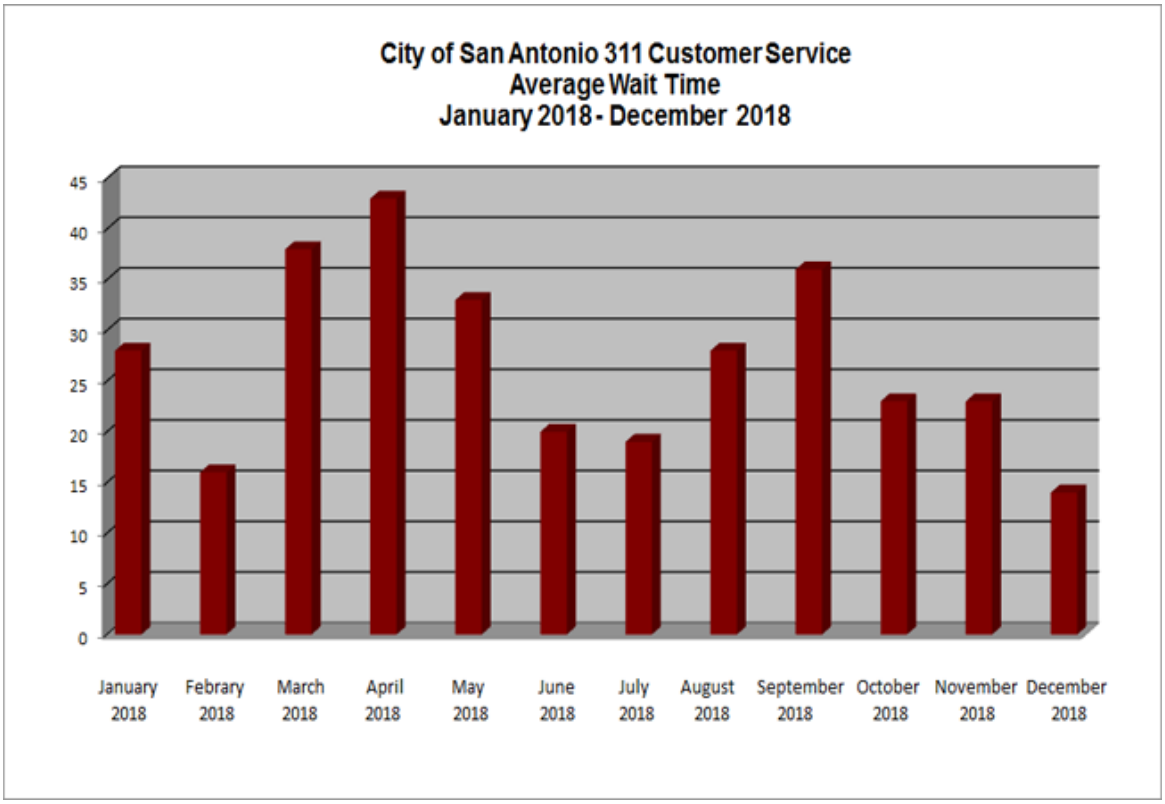
This metric represents the percentage of monthly calls answered within 45 seconds.



This metric represents the number of monthly calls presented during operating hours.



This metric represents the monthly average wait time in seconds for calls answered.


City of San Antonio 311 Customer Service
Average Wait Time
January 2018 - December 2018
45
40
35
30
25
20
15
10
5
0
January 2018
Febrary 2018
March 2018
April 2018
May 2018
June 2018
July 2018
August 2018
September 2018
October 2018
November 2018
December 2018