

Minneapolis 311 Calls: Exploratory Data Analysis

Langhan Dee

Significant variables in determining how long a 311 case will remain open:

Case Types:

In 2016 and 2017, the average number of days a case stayed open was 3 days. However, this average differed quite a bit across case types. Even looking at the highest level of case types, Subject (also known as Department), we saw a variation in the number of days until a case was closed.

- **Animal-Related** cases had the fastest response time. More than 75% of cases were closed within 1 day of being opened. As these cases included reports of abandoned, stray, or mistreated animals, they were the only call types affecting the immediate safety of living creatures.
- **Public Safety** had the highest median of 11 days open. This subject included one of the top ten volume drivers, Graffiti complaints.
- **Sidewalks** had the second highest median days open at 8 days, however, 75% of these complaints took 38 days to close. Sidewalk Snow and Ice complaints built up each winter and then were closed en masse when the weather warmed up. They were especially bad in 2017. These cases were so problematic that in 2018 the City plans to enforce inspections and fines to encourage people to keep their sidewalks clear.

Correlations between pairs of independent variables or between an independent and a dependent variable?

- The case types (Subject, Reason, and Title) are three levels of description detail. Because of this dependence, only one level will be used when doing statistical inference.
- The chi-squared test was used to examine the relationship between the Case Type variables and Neighborhoods. The test produced a p-value near 0.0, indicating a strong correlation.

What are the most appropriate tests to use to analyze these relationships? All of the factors in this data set are categorical variables (case types, neighborhood, date/month opened or closed) or discrete variable (temperature, inches of precipitation, snow depth, days open). This means we would have to get very creative to use linear regression. To determine a correlation between these types of variables, the chi-squared test was used.