Saptarshi Ghose

Discussed concepts with Natasha and Andrea

Machine Learning Homework 1 Write-Up

Problem 1

**Based on these summary statistics, tell me 5 interesting things you learned (about Chicago and the different neighborhoods) using the 311 data.**

1. Among the 311 requests in Chicago in 2017 for the service types we are working with, there were 112,779 requests for graffiti removal, 27,896 for alley lights out, and just 3,666 for vacant/abandoned buildings. As such, there were more than 4X as many graffiti removal reports than alley lights out reports. And there were more than 30X as many graffiti removal reports than vacant/abandoned building reports.
2. Considering all three types of 311 requests in 2017 we are exploring in this assignment, September, October, and November were the months with the most total requests by creation month. In contrast, June and July had the fewest total requests. In our limited dataset for just 2017, there were far more 311 requests (for graffiti removal, alley lights out, and vacant/abandoned buildings) in colder months than warmer months.
3. Of the 3148 vacant/abandoned buildings 311 requests without relevant missing data, 1267 (40%) had people still using the property (i.e. homeless people, children, gangs).
4. Among all explored 311 requests in 2017, the average response time (completion date – creation date) was 6.4 days.
5. 40,329 (36%) out of the 112,669 requests for graffiti removal in 2017 were for graffiti on metal surfaces. Unpainted brick had the next highest number of graffiti removal requests with 19,482 (17%).

Problem 2

**Based on this augmented data, provide some descriptive statistics to describe:**

**1. What types of blocks get “Vacant and Abandoned Buildings Reported”?**

**2. What types of blocks get “Alley Lights Out”?**

**3. Does that change over time in the data you collected?**

**4. What is the difference in blocks that get “Vacant and Abandoned Buildings Reported” vs “Alley Lights Out”?**

1. In blocks with vacant and abandoned buildings reported, the mean white population (according to the American Community Survey API data) is far lower than the mean white population for blocks with reports of graffiti or alley lights out. More specifically, the average white population for blocks with vacant and abandoned buildings reported is just 11,837 – compared to an average of 27,352 white people in blocks with alley lights out requests and 42,131 white people in blocks with graffiti removal requests.
2. In blocks with alley lights out reported, the average response time in days (from 311 creation date to completion date) is 33.5 days and the median response time in days is 14 days. In contrast, in blocks with graffiti removal requests reported, the average response time in days is just .67 days and the median response time in days is just 1 day. As such, it appears that alley lights out 311 requests in blocks usually takes much longer to complete the response to.
3. The American Community Survey data added and analyzed in parts 1 and 2 are unlikely to change meaningfully over time in 2017 because they refer to population estimates (i.e. white population and female population by block). These are variables that are not very susceptible to meaningful temporal changes within a year.
4. Blocks that get vacant/abandoned building requests typically have a far lower white population than blocks that have alley lights out requests. More specifically, the mean white population is nearly 3X higher in blocks with alley lights out requests than for blocks with vacant/abandoned building requests. The average number of never married males is very close between blocks with vacant/abandoned buildings vs. blocks with alley lights out. Blocks with vacant and abandoned building requests have, on average, have female populations that are ~10% lower than the female populations in blocks with alley lights out requests.

Problem 3

**Assume you are running the 311 call center for Chicago. You get a call from 3600 W Roosevelt Ave.**

**A. Of the three types of requests you have data for, which request type is the most likely given the call came from 3600 W Roosevelt Ave? What are the probabilities for each type of request?**

**B. Let’s now assume that a call comes in about Graffiti Removal. Which is more likely – that the call came from Garfield Park or Uptown? How much more or less likely is it to be from Garfield Park vs Uptown?**

**C. Now assume that you don’t have access to all the raw data and you know the following things:**

**There are a total of 1000 calls, 600 from Garfield Park and 400 from Uptown. Of the 600 calls from Garfield Park, 100 of them are about Graffiti Removal. Of the 400 calls from Uptown, 160 are about Graffiti Removal. If a call comes about Graffiti Removal, how much more/less likely is it that the call came from Garfield Park versus Uptown?**

1. The zip code for the given address is 60624. Using zip code for grouping in this analysis, alley lights out is the request type that is the most likely with 612 requests. In 2017 in our dataset, there were 311 service requests for graffiti removal and 146 for vacant and abandoned buildings. With just the data we have, the probability of an alley lights out request is 57% and 29% for graffiti removal and 14% for vacant and abandoned buildings.
2. The zip codes associated with Garfield Park are 60612 and 60624. The zip codes associated with Uptown are 60613 and 60640. Using zip codes as our grouping variable for this analysis, there were a total of 2261 graffiti removal requests in Garfield Park out of a total of 3412 requests (including alley lights out and vacant and abandoned buildings). 66.12% of requests in Garfield Park are for graffiti removal.

For Uptown, there were a total of 5120 graffiti removal requests out of a total of 5494 requests (including alleys light out and vacant and abandoned buildings). 93.19% of the requests in Uptown were for graffiti removal.

As such, it is more likely that the call came from Uptown. It is 1.41X more likely for a call to come from Uptown vs. Garfield Park.

1. For Garfield Park: (100/600) = .166

For Uptown: (160/400) = .4

.4/.166 = 2.4

Therefore, if a call comes for graffiti removal, it is 2.4X more likely to come Uptown than from Garfield Park.