Problem 1:

1. We downloaded the csv files for the request: . Then we generated summary statistcs for them.
2. Please find various summary statistics for the 311 requests:

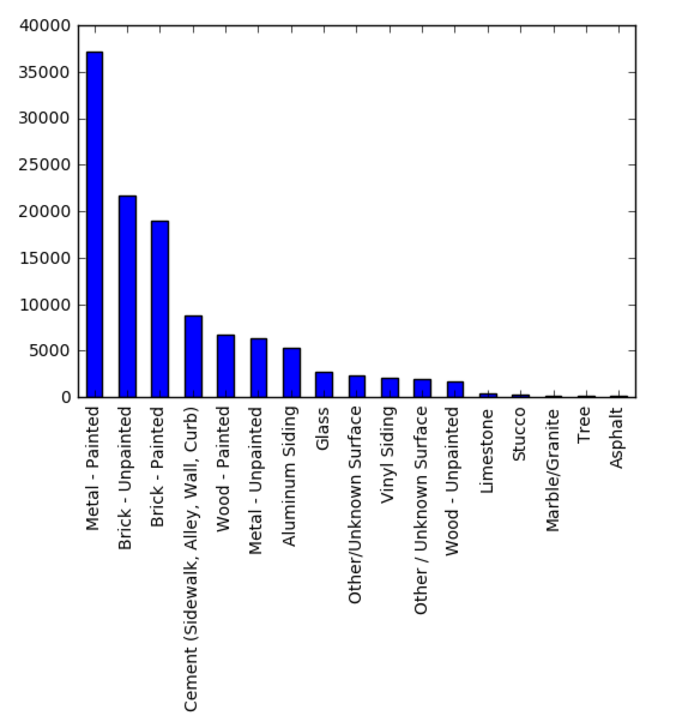
| **Type of Service Request** | **Count** |
| --- | --- |
| **Graffiti Removal** | 117327 |
| **Pothole in Street** | 56161 |
| **Sanitation Code Violation** | 21518 |
| **Vacant/Abandoned Building** | 4750 |

* 1. As we can see maximum number was reported for Graffiti removal

(variable: count\_311\_combined in our code)

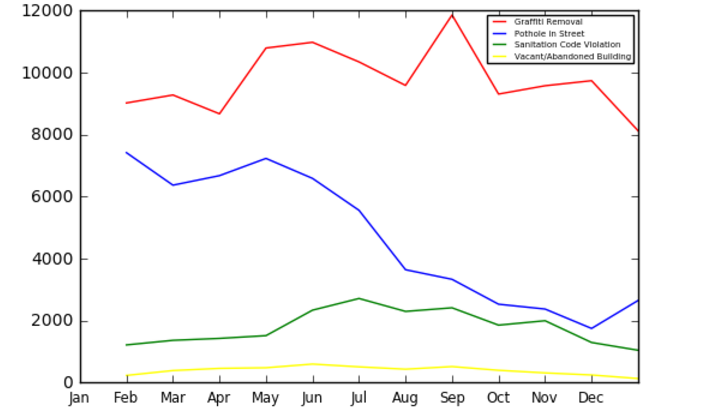
* 1. Metal and walls were the ones with maximum graffiti

(variable: count\_graffitti in our code)



| **Type of Service Request** | **Graffiti Removal** |
| --- | --- |
| **What Type of Surface is the Graffiti on?** | **Count** |
| **Aluminum Siding** | 5257 |
| **Asphalt** | 105 |
| **Brick - Painted** | 19041 |
| **Brick - Unpainted** | 21731 |
| **Cement (Sidewalk, Alley, Wall, Curb)** | 8794 |
| **Glass** | 2689 |
| **Limestone** | 467 |
| **Marble/Granite** | 152 |
| **Metal - Painted** | 37208 |
| **Metal - Unpainted** | 6392 |
| **Other / Unknown Surface** | 1970 |
| **Other/Unknown Surface** | 2290 |
| **Stucco** | 317 |
| **Tree** | 148 |
| **Vinyl Siding** | 2055 |
| **Wood - Painted** | 6685 |
| **Wood - Unpainted** | 1696 |

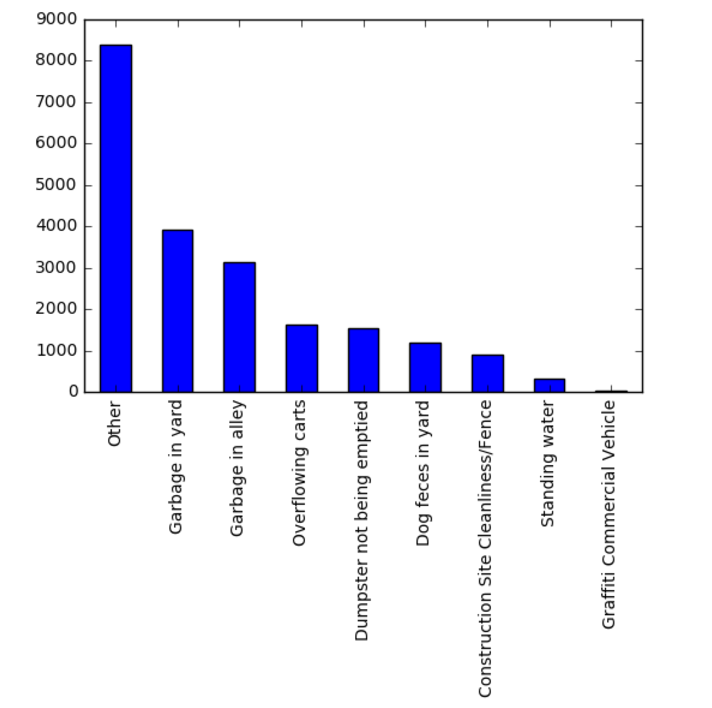
c. Details for number of complaints per month



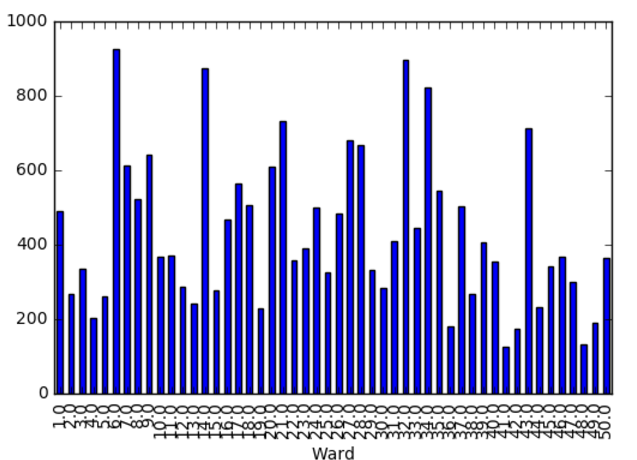
* 1. The service response time for most 311 requests is less than 10 days

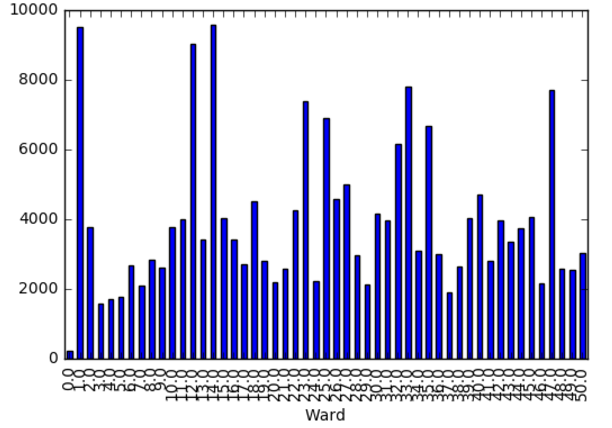
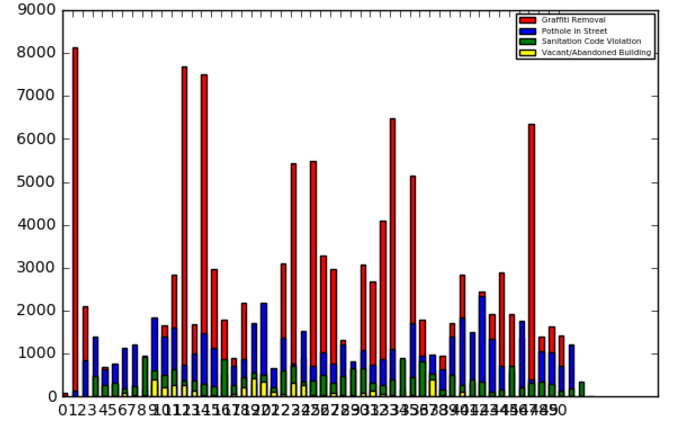
| **Type of Service Request** | **Graffiti Removal** | **Pothole in Street** |
| --- | --- | --- |
| **Response Time** |  |  |
| **0 days** | 67947.0 | 9349.0 |
| **1 days** | 33213.0 | 3227.0 |
| **2 days** | 6786.0 | 2710.0 |
| **3 days** | 6286.0 | 2139.0 |
| **4 days** | 2179.0 | 2035.0 |
| **5 days** | 510.0 | 1997.0 |
| **6 days** | 122.0 | 2015.0 |
| **7 days** | 38.0 | 1829.0 |
| **8 days** | 13.0 | 1502.0 |
| **9 days** | 4.0 | 1114.0 |
| **10 days** | 7.0 | 908.0 |

* 1. Sanitation:



* 1. Sanitation per ward.



* 1. Ward wise data for all requests (combined). We can see a similar pattern.
  2. Ward wise data (in a single graph segregated for individual request types.

1. Based on the summary statistics, we can observe the following:
   1. There are certain wards which have the highest number of complaints. However we do not know how these wards are categorized. Do they have some similar relationships? Like having similar dempgraphics, etc.
   2. Sanitation and graffiti follow a similar complaint and response pattern.
   3. There is a low number of potholes around end of year and an increased number of potholes 311 complaints during the first quarter of the year.
   4. The complaints for vacant building are consistent and low. However there are certain wards with higher complaints, which was surprising.
   5. Certain areas have a lot of graffiti issues and very less of other issues. I would have imagined some areas are worse in all verifiable aspects, rather than them being good at certain things and worse at others.

Question 2:

For this problem, I have taken the following requests for each block:

code\_represents = {"B01002\_001E": "age", #median age

"B23025\_002E": "employment\_labor\_force", #Number of persons, age 16 or older, in the labor force.,"B23025\_005E": "employment\_unemployed", #Number of unemployed, age 16 or older, in the civilian labor force. "B19013\_001E": "income", # Median household income in the past 12 months (in 2013 inflation-adjusted dollars)."B19301\_001E": "income\_per\_capita", #Per capita income in the past 12 months (in 2013 inflation-adjusted dollars)."B01002\_003E": "median\_female\_age", #Median age by sex (female)."B01002\_002E": "median\_male\_age", #Median age by sex (male)."B01003\_001E": "population", #Total population"B17001\_002E": "poverty", #Number of persons whose income in the past 12 months is below the poverty level}

Interestingly, there were many blocks for which poverty of the data was missing. However, the same data was available for other blocks.

Due to lack of time, I performed analysis of top 10 blocks from where there is maximum and minimum complaints.

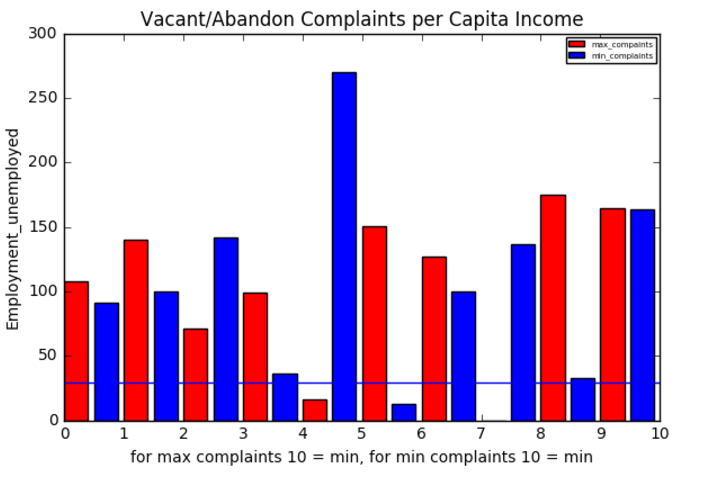
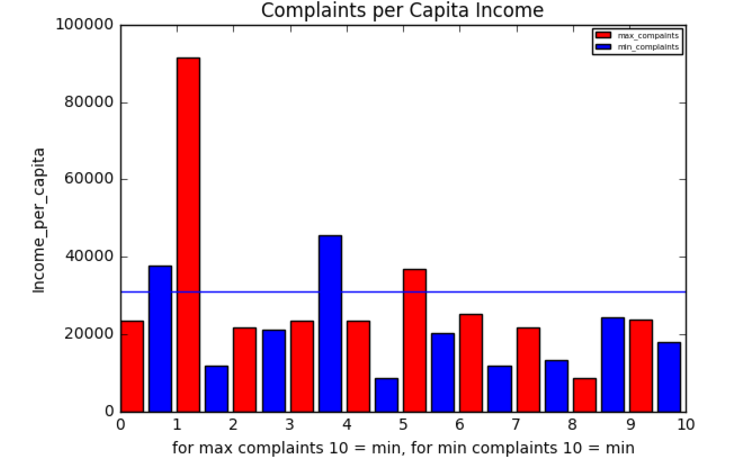
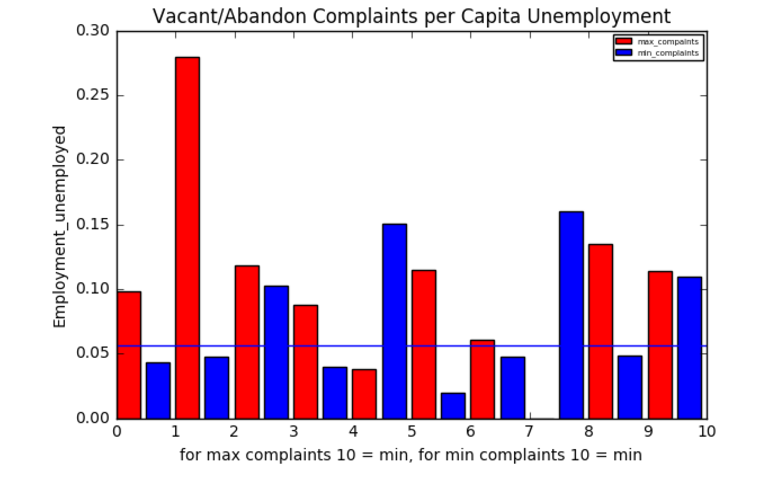
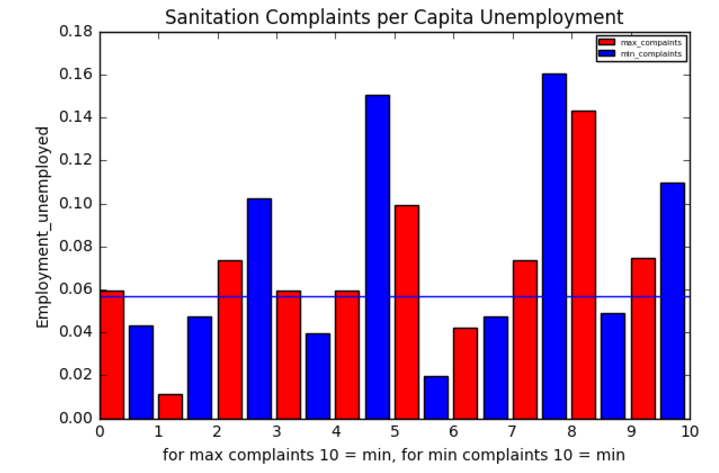
The line represents the average cook county data to make a sense check of the values. As we can see, blocks with higher average salaries complained more about sanitation and blocks with low unemployment rate complained more.

This seems to be consistent with the data as people with higher salaries and hence lower unemployment rates tends to be more educated about issues and hence make more 311 requests.

What is surprising to note that blocks with higher unemployment rates made more complaints. Though this too can be explained as blocks with higher unemployment tend to be higher in poor areas where people tend to abandon their house due to various reasons. But this fact does not confirm with per capita income data as median salary is higher for people who make the highest complaints.

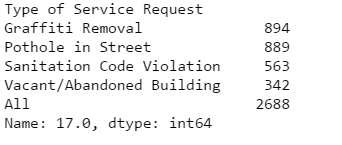
This thing is consistent across time, though there is a total increase in number of complaints during few months

Another interesting thing to observe is that sanitation complaints and abandoned buildings have opposite moving data. Perhaps areas with high vacant buildings have greater sanitation issues, so people tend to make one complaint only.

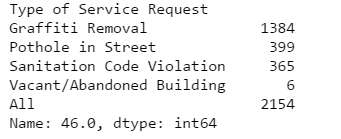
Question 3

1. Our address lies in the ward 17. From what we can see, for ward 17 there is an equally high likelihood of having a Grafitti Issue or Sanitation issue.

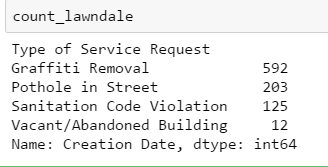


1. We find out the number of request of all types in each locality.

For Uptown: graffiti is 60% of the phone calls. Uptown area comprises of ward 46 of Chicago.



For Lawndale: graffiti is 63% of the calls



Hence if the call is for graffiti, its more likely that it is from Lawndale

1. Probability of Call is for Graffiti = P(G)

= 260/1000 = 0.26

Probability of Call from Engelwood = P(E) = 0.6

Probability of Call from Englwood for Graffiti = P(G|E) = 0.16

Probability of Call from Uptown = P(U) = 0.4

Probability of Call from Uptown for Graffiti = P(U|E) = 0.4

Find:  
 P(E|G) and P(E|G)

Using Bayes Theorem:

P(E|G) = P(G|E)\*P(E)/P(G)

= 0.16\*0.6/0.26 = 0.369231

P(U|G) = P(G|U)\*P(U)/P(G)

= 0.4\*0.4/0.26 = 0.615385

As we can see that the probability of call from uptown is higher, hence its way more likely (25%) more likely.