Homework 1
Shuyu Zhou

GitHub Link:
https://github.com/szhou12/MachineLearning4PublicPolicy/tree/master/pa1

Problem 1 (refer to Q1_DataAcquisition_Analysis.ipynb)

3. Five interesting things:
    1) The number of 311 requests for graffiti removal is far more than the number of requests for vacant/abandoned buildings and the number of requests for alley lights out combined. This could be due to the fact that the functionality of buildings or lights is relatively more stable while graffiti as a culture among young people is more turbulent.
    2) Top community areas (e.g. west Englewood, Englewood) with high volume of 311 requests for vacant/abandoned buildings are also areas with high crime rates. The number of requests for vacant/abandoned buildings may work as an index for local crime rates.
    3) Contrasting with the above, top community areas (e.g. West Town, Logan Square, South Lawndale, Lower West Side) with high volume of 311 requests for graffiti removal are areas with low crime rates.
    4) Response time to graffiti removal is short. Generally speaking, graffiti will be removed within 4 days. However, response time to lights out is more spread out. Within 6 days, the percentage of incomplete request for graffiti removal is more than the percentage of incomplete request for lights out. However, over 6 days, the percentage of incomplete request for graffiti removal will be less than the percentage of incomplete request for lights out.
    5) The peak volume of requests for graffiti removal happens from August to October. This is could be due to the nice weather so that people especially young people are more likely to be involved in outdoor activities. The change of requests for alley lights out over time is similar to that for graffiti removal. This again could be due to the weather.

Problem 2 (refer to Q2_DataAugmentation_APIs.ipynb):

1. Blocks that get “vacant/abandoned buildings reported” has relatively small portion of whites. Based on the statistics in out[223], only 0.7% of total population is white people in top 10 blocks that get “vacant/abandoned buildings reported”. Based on the statistics in out[327], 34.7% of population is white in all blocks that get “vacant/abandoned buildings reported”. About half of population in those areas is male. Based on out[327], 47.7% of total population is male. Based on out[337], average median family income in those blocks is $48611.2 and the average family size is 2.95.
2. About half of population in blocks that get “alley lights out” is white. The statistics in out[327] show that 47.7% of total population is white. Similarly, half of population in those blocks is male, which is about 48.5%. Based on out[337], average median family income in those blocks is $62699.2 and the average family size is 2.81.
3. Based on out[300] and out[293], small number of blocks are observed to be remained in top 10 blocks that get either “vacant/abandoned buildings reported” or “alley lights out” for the past 3 months. So blocks change over time.

4. Out[327] and Out[337] provide comparison between blocks that get "vacant/abandoned buildings reported" and blocks that get "alley lights out". Blocks that get "vacant/abandoned buildings reported" have less white population than blocks that get "alley lights out" do. Male population is about the same in both types of blocks. The average family size in blocks that get "vacant/abandoned buildings reported" is larger than that in blocks that get "alley lights out". However, the average median family income in blocks that get "vacant/abandoned buildings reported" is about $10,000 less than that in blocks that get "alley lights out".

Problem 3 (refer to Q3.ipynb):

1. The data I had have no information about '3600 W Roosevelt Ave' but information about '[number] W Roosevelt Rd'. Also, there is only one request call from '3600 W Roosevelt Rd'. Therefore, in order to have enough large sample, I collected requests call from '2000 W Roosevelt Rd' to '4999 W Roosevelt Rd'. Based on the data I collected (shown in Out[103]), the probability of the request for graffiti removal is 83.87%, the probability of the request for vacant/abandoned buildings is 1.075%, and the probability of the request for allege light out is 15.05%.
2. As shown in Out[107] and Out[108], given a call about graffiti removal, the probability of the call coming from Uptown is 0.016 while the probability of the call coming from Garfield Park is 0.007. So this call is more likely to be from Uptown, about 0.009 more likely.
3. The calculation is shown below:

	Graffiti removal	Others	
Garfield Park	100	500	600
Uptown	160	240	400
	260	740	1000

$$P(Garfield|Graffiti) = \frac{P(Garfield \cap Graffiti)}{P(Graffiti)} = \frac{100}{260} = 0.385$$

$$P(Uptown|Graffiti) = \frac{P(Uptown \cap Graffiti)}{P(Graffiti)} = \frac{160}{260} = 0.615$$

This says if a call comes about graffiti removal, it is about 0.23 more likely that it comes from Uptown than from Garfield Park.