Kings County Crime Rate Report

A County Suffering from Extreme High Population Density

Ziwei Meng, Ao Liu

Abstract

In 1990, Our client Kings county (Brooklyn) suffered from highest crime rate among the US, and asked for our help to determine which factors influenced the number of serious crimes per county. The goal was to implement policies that would lead to the reduction of the number of serious crimes in their county.

This article reports analysis on county demographic information (CDI) data set from Applied Linear Statistical Models, 5th edition, by Kutner, Nachtsheim, Neter, and Li. The results of Poisson Regression Model and Gradient Boosting Model show that poverty rate, region, percentage of youth are key factors associated with high crime rate for most of counties. And for Kings county, extreme population density and per capita income plays an significant role in explaining its abnormal high crime rate.

This study may provide insights into possible policies for Kings county and other counties with an willing to reduce crime rate.

Keywords: Crime Rate, Poisson Regression Model, Gradient Boosting

1. INTRODUCTION

In CDI data set, demographic information of 440 counties in US was collected, each line of the data set has an identification number, a county name, a state abbreviation for a single county together with another 14 variables, including geographic data of area, region; demographic data of total population(population), percentage of population aged 18-24(perc.young), percentage of population with bachelors degree(perc.bs); and economic data of percentage below poverty level(perc.poor), total personal income(tot.income), per capita income(per.income) etc. In order to get most relevant variables with crime rate, we transform the original variables into following new variables:.

- population density (pop.density) = $\frac{population}{area}$
- physicians per 1000 people (physician.per.1000) = $\frac{physician}{(population/1000)}$
- hospital beds per 1000 people (beds.per.1000) = $\frac{hospitalbeds}{(population/1000)}$
- crime rate per 1000 people (crime.rate.per.1000) = $\frac{crimes}{(population/1000)}$

After examining the correlation between each pair of variables (heatmap.png), we select variables which are relatively higher correlated with crime rate and relatively lower correlated with each other.

/pics/.jpg

The remaining variables are: area, percentage of young people, percentage of old people, percentage of population with high school degree, percentage of population with bachelors degree, poverty rate, unemployment, total income, region, population density, physician per 1000 people, beds per 1000 people.

	1	2	3	4	5
1	1,1	0,0	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0	0,0
3	0,0	0,0	1,1	0,0	0,0
4	0,0	0,0	0,0	1,1	0,0
5	0,0	0,0	0,0	0,0	1,1

From [table1] and [table2] we can see that Kings county stands out by showing highest population density and highest crime rate among all counties. In the next section we use Poisson Regression Model and Gradient Boosting Model to verify whether it is a coincident.

	1	2	3	4	5
1	1,1	0,0	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0	0,0
3	0,0	0,0	1,1	0,0	0,0
4	0,0	0,0	0,0	1,1	0,0
5	0,0	0,0	0,0	0,0	1,1

2. METHODS

We split the data set into training set and test set, then use R to carry out two models, compare their results and use their suggested predictors as complementary to each other. In the first model we assume that the number of serious crimes in per 1000 people follows a Poisson distribution with mean parameter as linear combination of selected predictors, and we use quasi-likelihood to deal with over-dispersion. After filtering insignificant predictors, our model is as follows:

$$\begin{aligned} log(crime.rate.per.1000) = {} & \beta_0 + \beta_1 * perc.young + \beta_2 * perc.poor + \\ & \beta_3 * per.income + \beta_{42} * region2 + \beta_{43} * region3 + \\ & \beta_{44} * region4 + \beta_5 * log(pop.density) + \\ & \beta_6 * physician.per.1000 + \beta_7 * beds.per.1000 + \\ & \beta_8 * perc.bs + \beta_9 * unemployment \end{aligned}$$

The result is shown in [table 4]. In the second model, we use the same predictors as in first one, but fit a Gradient Boosting Tree Model to the data. These two models give us consistent results, but slightly different in ranking relative importance of predictors. In the conclusion section we will consider both of them to give our suggestions.

3. RESULTS

From the result of Poisson Regression Model [table 4], we see that percentage of youth, poverty rate, region, population density and beds per 1000 people are most significant predictors to crime rate. This model suggests that, to obtain 1 unit decrease in crime rate per 1000 people, we should consider (separately) decreasing percentage of youth by $\exp(0.003) = 1.003$, or decreasing poverty rate by $\exp(0.00003) = 1.00003$, or decreasing population density by 0.00003. Although region plays a significant role here, since its impossible to change the region of a county, we can only change other variables that are implicitly related to region.

	1	2	3	4	5
1	1,1	0,0	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0	0,0
3	0,0	0,0	1,1	0,0	0,0
4	0,0	0,0	0,0	1,1	0,0
5	0,0	0,0	0,0	0,0	1,1

From the result of Gradient Boosting Model [table 5], we see that most important predictors suggested by this model are: poverty rate, region, physicians per 1000 people, and population density.

	1	2	3	4	5
1	1,1	0,0	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0	0,0
3	0,0	0,0	1,1	0,0	0,0
4	0,0	0,0	0,0	1,1	0,0
5	0,0	0,0	0,0	0,0	1,1

Both models are successful in predicting/explaining crime rate in test data[compare.png], but they fail at one county, our client Kings county.

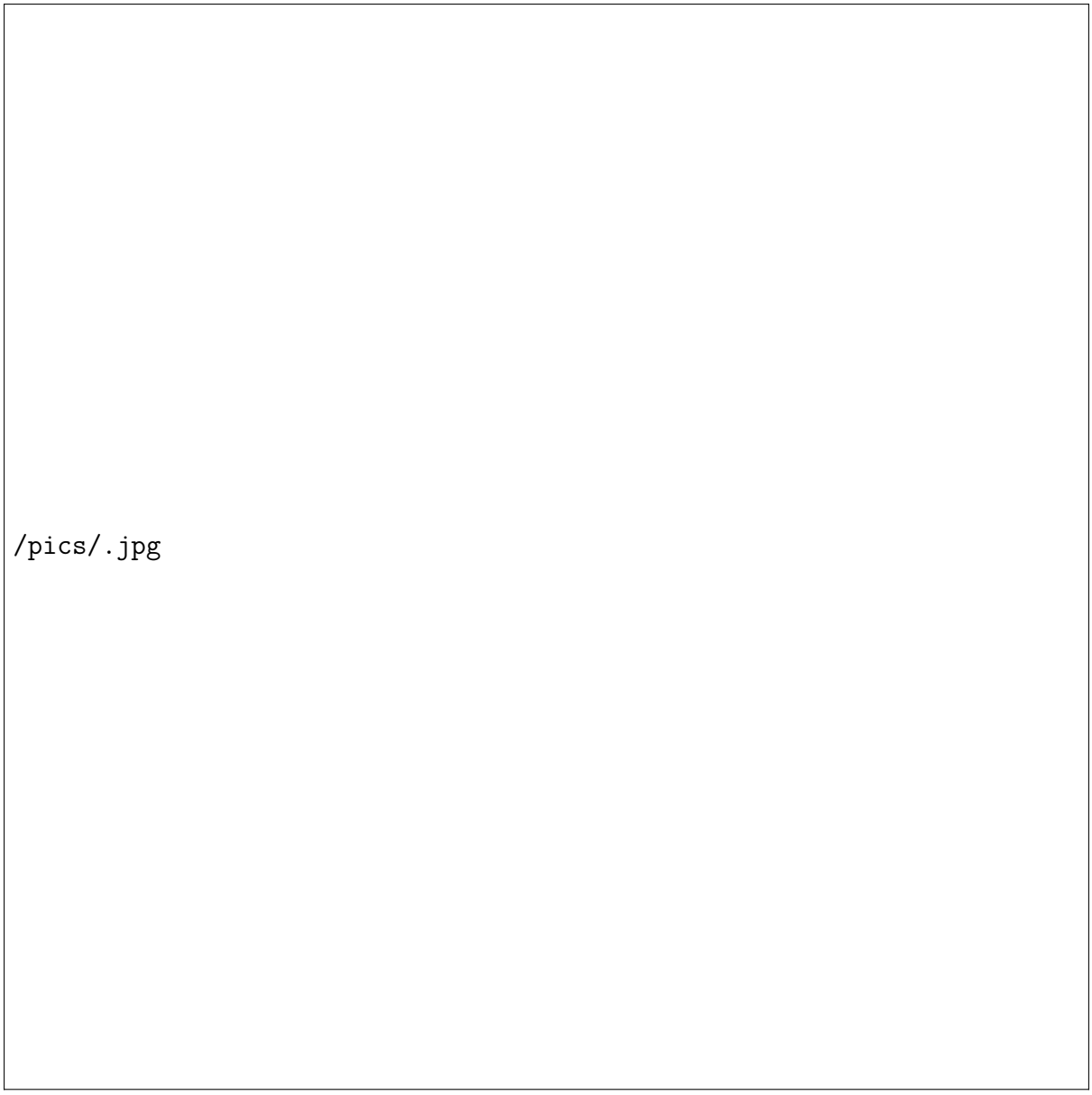
	1	2	3	4	5
1	1,1	0,0	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0	0,0
3	0,0	0,0	1,1	0,0	0,0
4	0,0	0,0	0,0	1,1	0,0
5	0,0	0,0	0,0	0,0	1,1

From [table 6] we see that, in Kings county, poverty rate is of 2.33 standard deviation higher than the average poverty rate in all US counties, and population density is of 19.76 standard deviation higher than the average population density in all US counties.

Since Kings county is located at New York, we also compare Kings county with other counties in New York, and from [table 7] we see that the percentage of population with high school degree in Kings county is of 3.10 standard deviation lower than the average percentage of population with high school degree in all other New York counties, the poverty rate in Kings county is 3.22 standard deviation higher than average other New York counties, and the population density in Kings county is 4.32 standard deviation higher than average other New York counties.

	1	2	3	4	5
1	1,1	0,0	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0	0,0
3	0,0	0,0	1,1	0,0	0,0
4	0,0	0,0	0,0	1,1	0,0
5	0,0	0,0	0,0	0,0	1,1

Now we can give a possible explanation for the failure of our model at Kings county. We suspect that when population density is higher than a threshold, its effect on crime rate would become different from when the population density is lower than that threshold. In other words, as long as population density is not too high, the effect of population density on crime rate is restricted, however, once the population density breaks that red line, the crime rate would increase very quickly. Since Kings county is the only county with such high population density (population.png), the parameters obtained from training a model on other counties are actually invalid to it.



4. CONCLUSION

Given above analysis, we provide slightly different suggestions to Kings county and to potential clients who are suffering from high crime rate. General suggestions:

- Reduce poverty rate by providing food, shelter and career training to low-income people.
- Educate young people about how to resolve disputes and conflicts through law process.
- Equip hospitals with more physicians and beds.

Specific suggestion for Kings county:

- Increase effective land area, provide assistance to people who are willing to relocate.

5. REFERENCE

[What Statistical Consultants Do: Report of a Survey] http://csyue.nccu.edu.tw/ch/What[Interpreti
Crime Data and Statistics] http://www.iaca.net/ExploringCA/2Ed/exploringca$_c$*hapter*8.*pdf*[*AStargaz*
//jakeruss.com/cheatsheets/stargazer.html

Appendix A. Section in Appendix