Analytics in Business - Problem Set 2

Due: Thursday, 15 Feb 3 PM

This group problem set ask you to apply what we have been learning in Analytics in Business analyses to questions related to predictive policing. Specifically, you will receive a dataset from the Metropolitan Police of London data on (a) stop-and-search events and (b) crimes.

A. Describe (25%)

Using the Met Data for July 2016, describe the relationship you can see, if any, between stop-and-search events and the following variables. On a single page by itself, place four charts clearly labelled charts, with each chart corresponding to one of the four items below, and write a paragraph below the charts of not more than 100 words) to summarise what you see.

1. **Time**. For age, use or create categories for 18-24 25-34, over 34, and under 10.
2. **Age.** Use or create categories for 18-24 25-34, over 34, and under 10.
3. **Gender.** Use categories for Male and Female.
4. **Race.** Use or create categories for White, Black, and Asian. Are things any different if we look at self- versus officer-defined race category?

Tip: State which of the patterns you see are confirming what you'd expect, and which (if any) could possibly be seen as challenging a belief that at least some people hold.

B. Evaluate (20%)

5. **Harassment, or Efficiency?** As developed commercially to date, the idea predictive policing is save resources by looking for crime in places where it is likely to occur. Based on hearings with council members and focus groups with the community, one of the worries people have about predictive policing is a question of reverse causality: does looking for crime in a certain place mean finding more of it? If so, predictive policing might encourage harassment rather than enabling efficiency. (This is an important but hard question that we will not fully answer in this problem set.) Against these questions from, your boss wants to know, how well were the Met doing in July 2016 at looking for crime where it occurs? On a single page by itself and in 250 words or less, tell us what your data set offers that is useful for addressing this question, what other data you might need to add to give a good answer to this question, and what specific question and analysis you would tackle to address what your boss is asking about.

   Tip: Answering this question requires a ratio of events that are like looking for and finding something. For this, create a new variable for stop-and-search outcome (SSO) and code events as "nothing" for "Nothing found – no further action" and "something" for everything else. (Yes, real crimes will be a subset of "something" in this new variable.)

C. Explain (25%)

6. **Racial Fairness?** Your boss is also very keen to know how the Met is doing when it comes to racial fairness. Using your SSO variable, what can you say about how racially fair the Met when it comes to the racial backgrounds of Londoners who get stopped from something versus nothing? Controlling for age, gender, and time of day (0500-1000, 1001-1700, 1701-2300, 2301-0500), do regressions to test for potential effects of race on the dichotomous variable of being for something versus nothing. On a single page by itself, provide a clearly labelled table with your regression results and a summary of no more than 200 words of what you found.

D. Predict (20%)

7. **Does PredPol really work?** The file "Q7 Exhibit" is a PDF of a slide from an analysis of how well the PredPol idea might work in London. In particular, the slide shows the pattern of variation in neighbourhood-level ratios for police events coded something versus nothing (as in SSO, above) as a function of levels of total stops (y axis) and something found (X axis). On a single page by itself, step into the shoes of a manager or supervisor to assess the analysis given and make constructive suggestions. Specifically, please address the following:
   a. If increased police presence (as measured as best we can with data by total stops) does indeed lead to lower crimes, what pattern would we expect to see in the chart, and do we see it?
   b. To do a fair and rigorous analysis of whether increased police presence does lead to fewer crimes by neighbourhood, what kind of analysis would we have to do? Briefly describe what is missing from this chart that you would need to include in your analysis, and say why you need it.
   c. Briefly describe the key variables (columns) you would need for each observation in your data (row), and describe the family of regressions you would need to use to do the analysis.

E. Automate (10%)

8. **Startup pitch.** On a single page by itself and in no more than 200 words, describe the data, analysis and concept you would use to support a socially acceptable approach to guiding police effort to optimise reassuring presence, community trust, the ratio of stop-and-search events that yield nothing versus something to investigate, and — of course — cost. Be creative in the data you would employ, but bear in mind societal concerns about privacy.

   Note: Be prepared to deliver this in class to our guests, and to receive feedback. The pitch judged best will receive a cool prize—not funding, but nice recognition.