Questions: Taxis & Crime

Carlos Petricioli

1 Assumptions

Assumption 1.1. Crime occurrences will be divided into two levels: *high crime* and *low crime* with respect to the total number of observed number crimes in a given period of time which will be defined later. So crime level can be defined at every time t an location l as a binary variable for *high crime* and *low crime* when the total number of crime at is respectively higher and lower than the average for that time and location.

Assumption 1.2. To simplify things, any location is considered to have a distance to a subway entrance equal to

- 0 if the location has any number available subway entrances.

- 1 if any of its neighbors location has a distance of 0.

- 2 in every other case.

NOTE: This assumption will be revised later to mark differences when a location has a lot vs just one entrance because it is important while making a very granular analysis.

Assumption 1.3. Any taxi ride will be categorized as a *short ride* if the distance form the pickup location to drop off location is considered as: could be walked in in less than a given time (i.e. 10 minutes) for an average person. And *long ride* any other case.

2 Questions

1. Is the average number of taxi pickups different in areas that have different levels of crime rates?

2. Is the average number of taxi pickups different in areas that have different levels of crime rates grouping by the proximity of an area to a **subway**? Any other public transportation options such as **Citibikes stations**?

3. Is the average number of taxi pickups different in areas that have different levels of crime rates when we compare at times with and without **rain** (or different weather variables)?

4. Does the average number of **short** rides hve a different average number of taxi pickups in areas that have different levels of crime rates compared to the average of **long** rides?

5. Does the answer of the previous questions change for special dates such as **holidays**?

6. When categorized by the **severity of crime**, is the average number of taxi pickups different in areas that have different levels of crime rates for a given level of severity in crimes?

7. **Run away vs feel attracted** to crime. When there are "high profiled" crime incidents in a given location does the average number of drop offs is greater, lower or similar to the average?

   NOTE: This is a similar question to [1].

8. Given crime rates changes across time for a given area, is the average number of taxi pickups changing too? If so, same direction? With a time lag? does it lasts long? Does severity has an impact?

9. How does the error in the prediction of crime rates changes when modeling it with traditional demographics vs the taxi rides vs both as features?

   NOTE: This is the main question of the paper from KDD 2016 [2].

References

[1] J. Bendler, T. Brandt, S. Wagner, and D. Neumann, "Investigating crime-to-twitter relationships in urban environments - facilitating a virtual neighborhood watch.," in *ECIS* (M. Avital, J. M. Leimeister, and U. Schultze, eds.), 2014.

[2] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime rate inference with big data," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 635–644, ACM, 2016.