

Fuqing Wang
CS591
Final Project Report1

The main dataset I am going to use can be seen by using the link below:

<https://data.cityofboston.gov/dataset/Parking-Tickets/qbxx-ev3s>

This is a dataset about the issued parking tickets in Boston from 02/23/2015 to 03/01/2015.

Additional dataset I am going to use:

<https://data.cityofboston.gov/City-Services/Street-Sweeping-Schedules/755x-x44q>

This is a dataset of street sweeping schedule of Boston.

Some basic analysis of the main dataset is below:

NOTE since some fields are not helpful for my project at present(maybe will be useful further on), I am going to skip the further analysis of those fields for the moment.

Useful fields:

Field name	Meaning	Example	Value type	Density	Missing value?
Ticket_Loc	Ticket Location	70 FRANKLIN ST	Number + Text	High	No
Issue_Time	Issue Time	9:47:00 AM	Number + Text	Low	No
Violation	Number of rule of violation	38	Number	Low	No
Violation1	Rule of violation	LOADING ZONE	Text	Low	No
Street_Num	Street Number	70	Number	Low	No
Street_Name	Street Name	FRANKLIN ST	Text	High	No
Lat	Latitude	42.3557	Number	Low	Yes
Long	Longitude	-71.0582	Number	Low	Yes

Specific analysis of useful fields:

Ticket_Loc: The issuing of location of the ticket. Number and text(Street number and street name). Density is high since address can be various, but it is able to handle because Boston is not such a big area, and I do believe the police have regular patrol hour and route, so the ticketing place would not be too many. In further analysis I will plot all the locations on the map, so that we are able to see which areas have high risk of being ticketed while violating the parking rules and which areas are not.

Issue_Time: The issuing time of the ticket. Number and text(Issuing time and am/pm.) I am going to use this part to find the ticket issuing distribution in a day. So we can see in which period of a day there is a high or low risk of being ticketed while violating the parking rules.

Violation_: Number of rule of violation. Number. I will use this field to record the frequency of violated parking rules.

Violation1: Rule of violation. Text. To translate the violation number into violation rule in text.

Lat: Latitude. Number. Use to plot location on the map. Missing value will not be a problem since we can find the latitude and longitude via location.

Long: Longitude. Number. Same as above

Potential useful fields:

Issue_Date: The issuing date of the ticket.

Street_Num: Street Number. Number.

Street_Name: Street Name. Text.

Route: Route number. (I'm not sure about this field).

Badge_Numb: Badge number of the agent.

Issue_Agen: Working number of the agent.

Fine_Amt: Fine amount needs to be paid.

Plate_Type: Plate type.(passenger, commercial, etc)

License_St: License state.

Vehicle_Ma: Vehicle make.

Vehicle_Co: Vehicle code.(I'm not sure about this field).

Vehicle_St: Vehicle state.

Comment: Comment.

My hypothesis: Boston police has a regular patrol hour and route that we can predict.

First of all, I will check if any tickets are issued during street sweeping hour and on the sweeping route, then eliminate these data since these cases should be considered as "exceptions".

Second, according to the ticket issued locations, find the density of different area(the frequency of being issued tickets in a area), and plot the result we get on the map according to the density.

Third, according to the time values we have, find the density of different period of the day(the frequency of being issued tickets during some period of the day), and draw a graph based on the results.

Eventually, based on my results from second and third step, create a dictionary in the form of (location, time), plot such points on the map according to the location, link the points according to the time, and eventually we will get a patrol route and hour.