**Baltimore**

**Parking Violations**



**Tamara Monge**

September 26, 2017

Springboard Data Science Career Track

Capstone #1 Proposal

­

1. **What is the problem you want to solve?**

Parking frustrations abound in nearly every major metropolitan area. Baltimore, MD is no exception. One measure of the difficulty in finding adequate legal parking is the number of citations issued in a metro area. In this project I will analyze patterns in parking citations issued by the City of Baltimore over the past two years.

1. **Who is your client and why do they care about this problem?**

Three clients will be interested in this study. The first stakeholder will be the residents of Baltimore. Even the most law-abiding citizens among us are sometimes tempted to bend the rules to make life easier. How many times have you found yourself circling the block hoping for a parking space to open up, while simultaneously performing a cost-benefit analysis in your head: “Is it better to wait for a legal space to become available and be late to my [work, meeting, appointment, sporting event, date, concert, etc.]? Or, is punctuality more important?” When the latter wins, and I have a choice of parking illegally in two different areas, the question then becomes: “Does one location offer a reasonable expectation of a lower fine, should I get cited?”

The second stakeholder in this study is the City of Baltimore. Paying a fine is a matter of civic responsibility and is a source of significant revenue for the city. Accordingly, the City of Baltimore will be interested to know what percentage of fines are paid on time, what percentage are still outstanding, and whether fines are high enough to deter repeat offenses. In addition, city planners could use the results of this study to identify which areas could benefit from more parking solutions (for example, additional parking structures, or perhaps an app-based option for remotely refilling parking meters.).

The final stakeholder in this study is the local police force. The Baltimore Police department will be interested to learn whether parking violations in their city are more common in particular areas and/or at particular times of day. This will inform their work distribution (scheduling and routing of officers throughout the shift), thereby reducing hours and saving money. A further point of interest for the police department could be whether or not there exist any patterns in their ticketing (for instance, more citations given to certain car makes, older cars, or out-of-state license plates). Armed with this information, the police department could the investigate whether such patterns are due to discriminatory ticketing, or whether they reveal true patterns in which particular car characteristics can be seen as a type of proxy for a driver who is more likely to park illegally.

1. **What data are you going to use for this?**

The dataset for this project will come from the City of Baltimore’s website: <https://data.baltimorecity.gov/Transportation/Parking-Citations/n4ma-fj3m/data>. As this data is freely available on the web, I will acquire it via simple download. The dataset comprises approximately 1.3 million rows of data (citations) and 20 columns per citation in CSV format. The interesting columns/features of each citation are as follows: date, citation number, license plate number, expiration date, state, make of vehicle, address, violation code, description, fine, penalty, police district, and council district.

1. **In brief, outline your approach to solving this problem.**

My approach will consist of six steps: (1) Data acquisition, (2) Data cleaning, (3) Exploratory Data Analysis, (4) Supervised learning algorithm, (5) Visualization and (6) Story compilation. Steps 1 and 2 are self-explanatory. In steps 3, I will explore relationships between violation code, violation description, fine, location, time-of-day, delinquency in paying fine, state of license plate, and make of vehicle. To accomplish this, I will use regression, correlation, histograms, heat maps, and geographic mapping. Step 4 will involve trying to predict the amount of the fine one would receive for a parking citation in a given region, at a given time (month, day of week, hour of day). In step 5, I will visualize the results of the machine learning algorithm. Finally, in step 6, I will pull together the results of the study into a comprehensive story.

1. **What are your deliverables?**

My deliverables for this project will be: a summary report, a PowerPoint slide deck, and the code created to perform the analysis.