Estimating the Probability of Receiving a Meter Violation in San Francisco

Tessa Weiss

Introduction

The objective of this project was to investigate crime in San Francisco. In particular, we decided to take an in-depth look into parking citations given throughout all of San Francisco obtained from unpaid meters.

The main party who would benefit from this analysis is a common citizen of San Francisco. Looking at things from their perspective, the main question one would want to be able to answer is "if I do not pay the meter on a specific street S during a specific time interval T, how likely am I to get a parking ticket?" The intention is to avoid locations and times where getting a ticket is most probable, and this project aims to help the user make informed decisions about whether or not it is worth it to pay for the meter, thereby mitigating the financial consequences of failing to pay for a meter. In order to do this, the question was transformed into a probability problem.

Additionally, information about locations where meter violations are most common would be beneficial to the city, as this information would enable them to assess if tickets are being distributed fairly across neighborhoods, after normalizing for factors such as neighborhood population, number of meters in a neighborhood, etc. Also, the city could use this information to devise more effective traffic regulations in regions where traffic violations are most common.

Data Sources and Quality Control

The main data source used for analysis was the SFMTA Parking Citations Dataset, which is a tabular dataset consisting of about 19 million rows and 10 columns providing citation information from 2008 to present[1]. Some of the columns in this dataset were paramount in constructing the probability of citations, such as `Citation Location`, `Citation Issued DateTime` (can obtain day of the week from this variable), and `geom` (geometric endpoints of a street segment). For future reference, a street section will be defined as a corridor – the name of the street the section belongs to – between two limits – the two streets at either end of the section (e.g. Mission St. between Spear St. and Main St.). This dataset was filtered to only provide information about meter violations in 2022-2023. Three other datasets were also used in the analysis: the Street Sweeping Schedule dataset[2], the meter transaction dataset[3], and the meter location dataset[4].

The street sweeping dataset was used to obtain geometric endpoints for each street segment and was joined to both the parking citations dataset as well as the meter locations dataset. In order to correct for some of the incorrect mappings, we used the US Census Bureau Geocoder to run the address strings for each citation, giving us the correct geometry for each street section[5].The meter transactions and locations datasets contained information about which meters were active, which streets they were on, and intervals of when each meter was paid throughout the day. With this information, we gain knowledge of all instances when someone was parked illegally on a street due to not paying a meter, with the assumption that all metered spots are always taken. A detailed visualization of the data pipeline is shown in Figure 1 below:

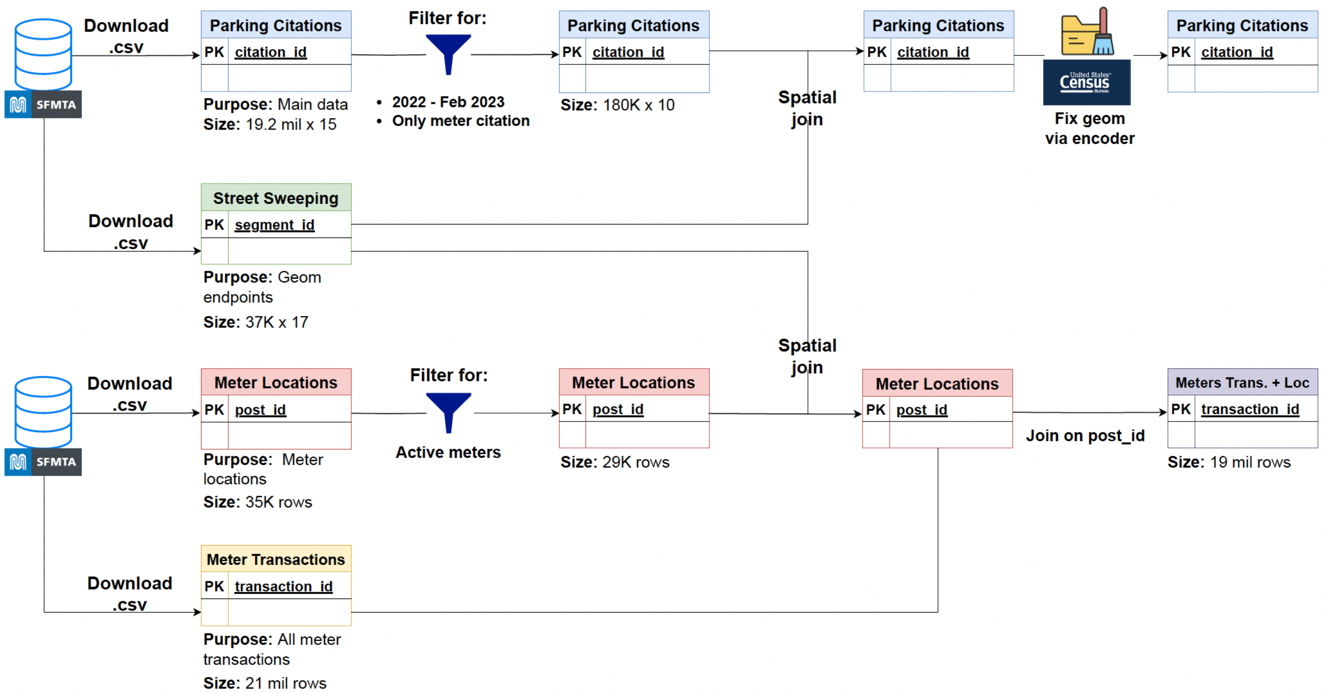

Figure 1: Data pipeline.

EDA

Referencing the main question of determining which streets parking enforcement will be on during which times, it became clear that we should look for potential trends corresponding to days of the week and possibly hours of the day. Figure 2 displays two barplots: the left barplot shows the number of meter citations aggregated over each day of the week from January 2022 - February 2023, and the right barplot shows the total number of meter citations aggregated over each hour during the same time period. Note that these hours range from 9:00am to 6:00pm, which are the hours that meters in San Francisco are active.

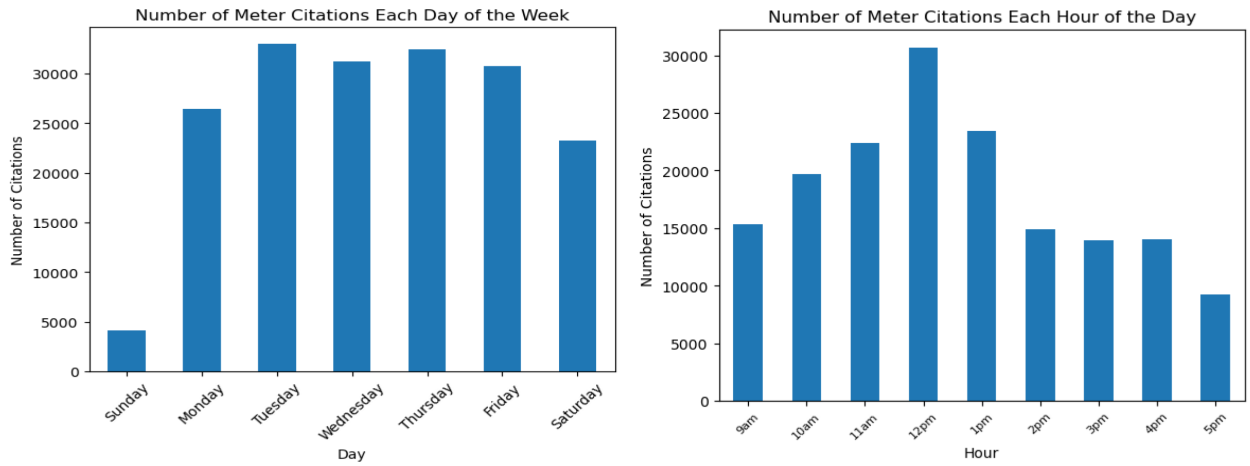

Figure 2: Citations aggregated over each day of the week (left) and over each hour of the day (right).

From these two plots, some immediate trends can be noted. First, the left plot in Figure 2 indicates that there is a major difference in the number of tickets for weekdays compared to weekends (particularly Sunday). This makes sense, as parking is typically free on Sundays. On the right plot in Figure 2, there is a large spike 12:00pm, a typical time when people would park for lunch. The trends shown on both of these plots indicate that not only weekday can be used as a feature in our analysis, but also hour of the day (or maybe even a more granular unit of time such as five minute bins) can be used as features.

One last important aspect to consider is the matter of having street segments of equal lengths. For example, a longer street segment has more parking spots than a smaller street segment, and therefore will naturally have a higher number of citations. Thus, it should be checked that the length of each street is roughly the same. Figure 3 shows the distribution of the lengths of each street segment, calculating the length of a segment by finding the Euclidean distance between the two endpoints (longitudinal and latitudinal coordinates) of that segment. The plot is presented on the log scale in order to better view outliers and get a clearer sense of the distribution.

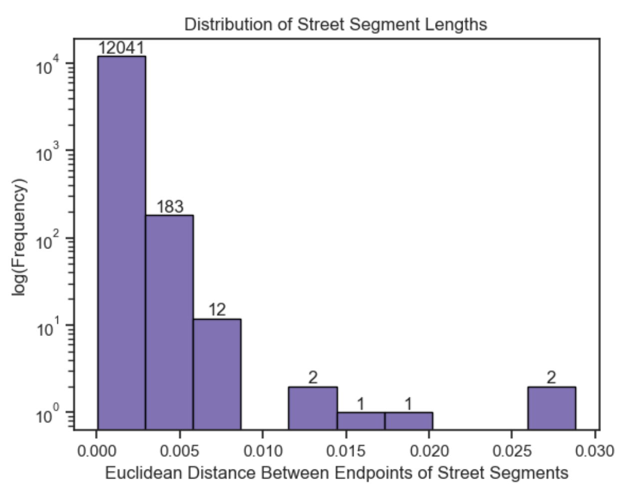

Figure 3: Distribution of segment lengths (Euclidean distance between geometric endpoints) presented on log scale.

From this output, it appears as though a majority of the streets throughout San Francisco are roughly the same length, and there are just a few outliers that have longer street lengths.

Analysis

With reference to the original goal, we want to calculate the probability of getting a parking ticket given a street section and day of the week/time of day and given someone has not paid the meter. At a high level, the idea behind calculating probabilities of receiving a parking ticket is to divide each 24 hour day into fifteen minute time bins, conditioning on this as well as a street segment and weekday (e.g. probability of getting a citation from 9:00-9:15am on a typical Friday on street segment A), thereby answering the main question from the citizen posed in the introduction. In terms of notation, define

- Weekday: $W \in \{\text{Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday}\}$
- Street Segment: $S \in \{1, 2, \ldots, N\}$ where N is the number of unique streets
- Start of Time Interval (15 minute increments): $T \in \{9{:}00\text{am}, 9{:}15\text{am}, 9{:}30\text{am}, \ldots, 5{:}45\text{pm}\}$
- $E =$ event that enforcement happens
- $I =$ event that illegal parking happens

Then, the probability of getting a ticket given a time bin T, a street segment S, and a day of week W is denoted as

$$\begin{aligned} p_{t,s,w} &= P(E \mid I, T = t, S = s, W = w) \\ &= \frac{P(E \cap I \mid T = t, S = s, W = w)}{P(I \mid T = t, S = s, W = w)} \end{aligned}$$

where we used the fact that probability of getting a ticket is equivalent to probability of enforcement being on the same street in the same time interval. Then, the following formulas were used to empirically calculate estimates of the numerator and denominator:

$$\widehat{P(E \cap I \mid T = t, S = s, W = w)} = \frac{\sum\limits_{\text{weekday(date)}=w} \mathbb{1}[\#\text{tickets}(s, t, \text{date}) > 0]}{\#\text{weekday(date)} = w}$$

$$\widehat{P(I \mid T = t, S = s, W = w)} = \frac{\sum\limits_{\text{weekday(date)}=w} \mathbb{1}[\#\text{unpaid meters}(s, t, \text{date}) > 0]}{\#\text{weekday(date)} = w}$$

where #tickets() is a function that takes in a street segment, time bin, and date, and outputs the number of tickets given, while weekday() is a function that takes in a date and outputs what day of the week it is on. To give an interpretation of these formulas, first consider the numerator. We have that each citation that occurred on day w in street segment s was assigned to a corresponding 15 minute bin. If there were multiple tickets on the same day in the same time bin, we recorded only one ticket, thus giving the indicator of there being at least one instance of enforcement and illegal parking. Lastly, the number of tickets in each time bin t was summed and divided by the total number of weekdays corresponding to day w in the dataset, thus giving an estimate of the numerator. Figure 4 shows a visual representation of how this was calculated:

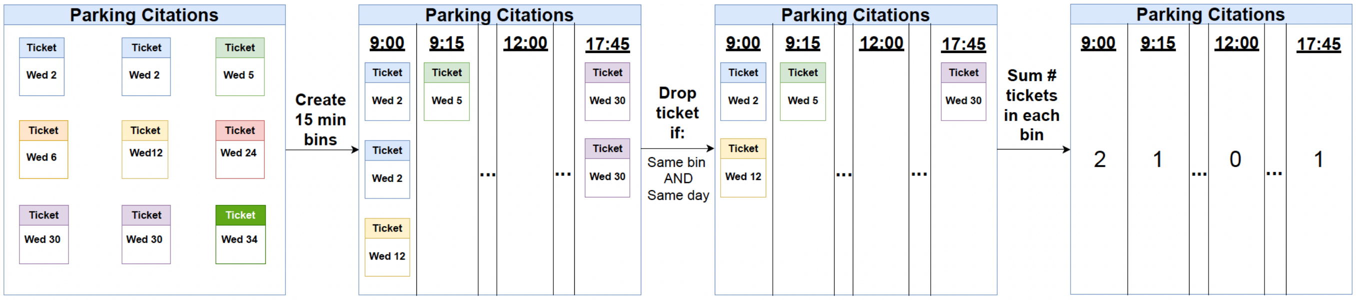

Figure 4: A simple example estimating the numerator where W = Wednesday on one particular street segment.

Now, considering the denominator, we have that for a street segment s and weekday w, the 15 minute time bins that each transaction covers for each meter on street s was recorded. If there were no transactions that covered a 15 minute time bin, or if there were gaps in transactions that exceeded a 3 minute grace period (time allowed for one car to leave and another car to park and pay), we considered the meter to have been unpaid during that time bin. Next, we aggregated all the time intervals for when there was at least one unpaid meter and used this to generate an indicator of an unpaid meter for the corresponding time bin on street s. Lastly, these indicators were summed up over all weekdays corresponding to day w and divided by the total number of weekdays corresponding to day w in the dataset, thus giving an estimate of the denominator. Figure 5 shows a visual representation of how this was calculated over one day:

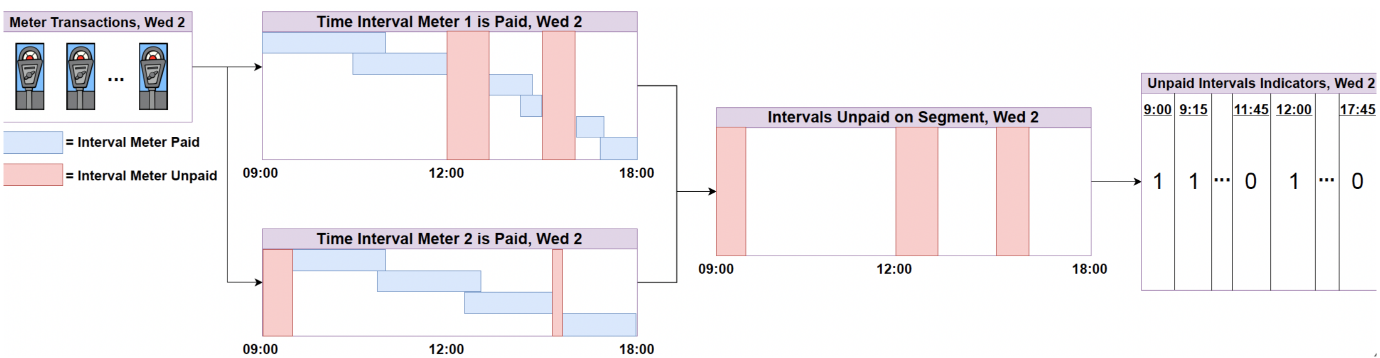

Figure 5: A simple example estimating the denominator where W = Wednesday on one particular street segment.

With these estimates of the numerator and denominator, we obtain our final estimate of the probability of getting a parking ticket due to not paying the meter given a 15 minute time bin defined by T, a street segment S, and day of the week W as notated by

$$\widehat{p}_{t,s,w} = \frac{\widehat{P(E \cap I \mid T = t, S = s, W = w)}}{\widehat{P(I \mid T = t, S = s, W = w)}}$$

In devising this model, the following assumptions were made. The first is that at any given time, every metered spot available will be filled. In this sense, the assumption is made that if there are any meters not being paid for on a particular street segment during a particular time interval, then we conclude that there is someone parking illegally (this is not always true and thus results in many underestimated probabilities). The second assumption is that parking enforcement will spend at most fifteen minutes on a street segment. The final assumption is that if parking enforcement is on a street segment, they will ticket all cars violating parking regulations with 100% certainty.

Final Product

Using the estimates for the probabilities of obtaining a parking ticket above, the final product involved creating a web application built on deck.gl and React that anyone can use to view these probabilities. The app allows the user to enter a start time to park, a day of the week, and the duration of parking, ranging from 15 minutes to 2 hours. Given these inputs, a colored map of San Francisco is displayed, showing the probabilities of obtaining a parking ticket with colors ranging on a spectrum from white to dark red, where the color white means no probability or no meters, and the color dark red means high probability. Hovering over a specific street on the map will output text containing the street name as well as the exact predicted probability of getting a meter ticket during the selected time interval. Additionally, the user may click on a street to output a graph of the expected cost of paying or not paying for a meter on that street starting at the selected time, and lasting for a duration from 15 minutes to 2 hours. In calculating the probability of getting a parking ticket across multiple 15 minute intervals (for example, 30 minutes is two intervals), the probability was determined by

$$\widehat{p}_{d,s,w} = 1 - \prod_{i=1}^{k}(1 - \widehat{p}_{t_i,s,w})$$

where d denotes the duration of parking with hourly units and k denotes the number of 15 minute intervals. Assumptions were that the probability of getting a meter ticket for each time bin on a given street and weekday are independent of each other. To calculate the expected cost of paying for a meter, we assume that the cost of a meter per hour is \$3.50 (which is the average price of all meters in San Francisco), in which the expected cost is calculated as $\$3.5d$. To calculate the expected cost of not paying for a meter, we further assume that the cost of a meter ticket is \$93.50 (which is the average price of all meter tickets in San Francisco). Then, the expected cost is calculated as $\$93.50 \times \widehat{p}_{d,s,w,}$. An example image of the final product is shown in the appendix. Most probabilities seem to suggest it is not worth it to illegally park, as the estimated expected costs are typically a lot higher on most streets when not paying the meter, especially when parking for longer than 15 minutes.

Next Steps

The methods in this paper discuss how data was used to obtain estimates of the probability of getting a parking ticket given a street segment S, a time interval T, and a weekday W, given one has not paid the meter. Next steps of this project would include fixing the issue of probabilities being underestimated due to areas that violate the modeling assumption of each metered spot being taken 100% of the time, of which there are many. Unfortunately, the dataset used for this analysis did not contain this information. In order to do this, one would ideally conduct a survey to gather data of true illegal parking rates, or the city could install sensors for a sample of meters throughout San Francisco in order to capture a more accurate estimate of the presence of vehicles. Additionally, expected costs could be further corrected by separating downtown streets and non-downtown streets, since the price of meters and meter tickets downtown are higher than non-downtown areas. Lastly, the city could use the probabilities calculated in this analysis to conduct an ANOVA across neighborhoods in order to see if certain areas are being ticketed more often than others. This could help identify a possible equity issue, for example if wealthier neighborhoods are found to be ticketed less than poorer neighborhoods.

Appendix

Probabilities of getting a parking ticket due to unpaid meters:

Figure 6: Final product: Probabilities of getting a parking ticket due to unpaid meters.

The above screenshot displays the estimated probabilities of getting a ticket due an unpaid meter given a start time, weekday, and duration ranging from 15 minutes to 2 hours, as well as expected costs for paying vs. not paying for the meter. White indicates no meters or no probability of getting a citation, while dark red indicates high probability of getting a citation.

References

[1] SFMTA - Parking Citations
`https://data.sfgov.org/Transportation/SFMTA-Parking-Citations/ab4h-6ztd`

[2] Street Sweeping Schedule
`https://data.sfgov.org/City-Infrastructure/Street-Sweeping-Schedule/yhqp-riqs`

[3] SFMTA Parking Meter Detailed Revenue Transactions
`https://data.sfgov.org/Transportation/SFMTA-Parking-Meter-Detailed-Revenue-Transactions/imvp-dq3v/data`

[4] Map of Parking Meters
`https://data.sfgov.org/Transportation/Map-of-Parking-Meters/fqfu-vcqd`

[5] US Census Geocoder
`https://census-geocoder.readthedocs.io/en/latest/`

[6] Ning Jia (2022): What are the odds of getting a parking ticket in Toronto?
https://towardsdatascience.com/what-are-the-odds-of-getting-a-parking-ticket-in-toronto-1f090d d0c608

[7] Song Gao, Mingxiao Li, Yunlei Liang, Joseph Marks, Yuhao Kang & Moying Li (2019): Predicting the spatiotemporal legality of on-street parking using open data and machine learning
https://www.tandfonline.com/doi/full/10.1080/19475683.2019.1679882