Parking in Chicago

Understanding How to Avoid Illegal Parking Tickets in Chicago

Audience

The intended audience for this deep dive into Chicago's parking data is primarily individuals who park in the Chicagoland region but is also relevant to individuals who work for the city issuing tickets. The portions of data involving the ticket issuer are very high-level recommendations that may assist new ticket issuers in performing their duties with more efficiency and awareness.

Motivation

After receiving a parking violation from the City of Chicago, I was distraught because I felt that the ticket was not fair. It was this situation that sparked a desire for me to learn more about the ticketing habits of Chicago. The initial goal was to figure out the probability of getting a parking ticket based upon where a vehicle was parked. One way to achieve this goal was to focus on where the tickets are more commonly issued. Although a good starting point, I needed to factor in the plethora of other variables that increase the chances of getting tickets in some areas over others.

Some of the variables that motivated my desire to understand how to avoid getting illegal parking tickets included: officer on duty, weather, events, time of day, and weekday vs. weekend. With these variables in mind, I realized the goal required a more holistic approach towards synthesizing the data. Eventually, I was able to arrive at a conclusion through a naive based outlook.

Methodology

The goal of the project was to build a Folium time series heatmap using naive based methodology. This method required the use of data on the time the ticket was issued, the

location of the ticket and weekday vs. weekend. Further analysis of this data painted a more cohesive image that captured the relationships between these variables.

Data Set Description

The project used the data set provided by the City of Chicago. The set includes roughly 18,000,000 separate parking tickets issued over an 11-year span. It was necessary to clean some portions of the extensive data sets in order to produce organized and accurate results.

The data synthesized included some of the aforementioned variables of who issued the tickets, location of the ticket, time of ticket, type of ticket, how the ticket was addressed as well as a list of other miscellaneous facts.

Data Wrangling

Like many other large public data sets, the Chicago parking data had a number of missing sections. To ensure the data was relevant to the project early on, it was necessary to make sure that there was an actionable time series. In order to properly use and analyze the data, we created a time series made up of day and hour columns. Beyond the creation of the time series, we also had to use some other libraries to help fill in with additional information.

The Chicago parking data set used is comprised of all the parking tickets issued in Chicago from 2007 through spring of 2018. There are 18 tickets about 15 different sections of info collected on the tickets. We were provided with the ticket number, issue date, violation location, license plate state, plate type, violation code, violation description, unit description, vehicle make, fine level amount, the amount due, ticket queue date, notice level, hearing disposition, notice number, officer and address. Using python, we added columns containing latitude, longitude, hour and date time of the ticket.

At first glance, it was evident that there was an a lot of attention to detail and accuracy of submitting data to the set over the years. The initial data set had tickets from the year of 2007 to the spring of 2018. Out of 18 million rows, 90% of the columns were not missing any information. Most of the missing values were in columns that I later removed or that were not

relevant for the task at hand. Some of the removed columns included the license plate number and the make of the vehicles.

One import column that needed to be fixed was the zip code column. Initially, there were 1-1.5 million incorrect zip codes. Upon early inspection, we found that the zip codes were wrong or alphanumeric based. We fixed the erroneous zip codes by running a geolocator program to output the proper zip code for the location of the ticket. The correct zip code was not found for every line in the data frame. I chose to not collect data on the years and days not relevant to the days I focused my research on.

The other challenge that arose in the data wrangling process was producing the latitude and longitude of the address where each ticket was issued. This issue was remedied through the use of 25 lines of python that required research and practice on my part before I felt confident to tackle the problem. The data cleaned up and worked quite well. I am very pleased with the outcome that we produced.

Exploration of the Data

The data narrative for the parking project began early on in the data harvesting process. My goal in completing this assignment was to cover the programs I used and explain how the python libraries were utilized to find the necessary information. Upon completion of tasks for the project thus far, I feel that I entered the course with an incorrect or shallow understanding of what data science involves. In hindsight, I am pleased to have entered into the program with an incomplete understanding because my newfound skills and more holistic view of data science has been grown through hard work and knowledge acquisition efforts.

After getting a ticket for being parked on a street that needed street cleaning, I decided to learn more about the parking data in the Chicagoland region. My desire to learn more about the world around me using data was bolstered by my intent to figure out the probability of getting a ticket if I parked my car illegally for a set number of hours. I wanted to base the conclusions from this study on the historical parking ticket data in the area. Upon starting this data science course, I quickly realized that I need a firmer foundation in data analytics before I

can dive deeper into my capstone project and into understanding how to use data to solve problems I encounter daily.

The first hurdle that I had to overcome was getting the data set to upload to Github. After many days with little success, my mentor and I decided to break down the data set into smaller quantities. After breaking it down, we were able to produce a sample of the data set that was smaller than 2GB. At this point, the dataset was loaded to Github using Gitlfs.

While looking at the data early on, we figured out that the zip code of the data was not showing up in the proper format. To fix the zip code issues, we decided to use Geocoder. While using Geocoder, we set up a function that would take the ticket issue address then add Chicago, IL to the end of the address geocoder which would append the zip code to a new list. It took a while to get the data in the correct format. Some of the initial problems we faced included time-out errors and struggling with the function crashing randomly. After a few weeks of fidgeting with the data and reducing the data down to a single day, we were able to produce the correct zip codes.

The only problem with using the zip codes in the data was that we were planning on using Folium to create a heat map of the ticket issue locations. This presented an issue because Folium works with latitudes and longitudes rather than zip codes. We then had to re-run the code with Geolocator to produce the latitude and longitude coordinates. Finally, using Folium we were able to produce a heat map of the parking tickets in the Chicago area.

After seeing the completed map, the visual prompted a new question: are there more tickets issued near the police stations in comparison to areas without increased police presence? To answer this, I found a public data set of police station locations in Chicago given in latitude and longitude coordinates. I imported the dataset, then used the children method of Folium to place the fifteen or so police stations on the map. The process of importing the second data set was a lot quicker for me because I was able to practice the same technique as before. I found that there was a loose correlation between more tickets being given around police stations. I found that the driving factor behind there being more tickets in some areas over other areas had to do with the type of parking available. The areas with fewer tickets issued, were typically safer areas with lower crime rates and areas with free street parking.

I did not expect to find the number one ticket area to be at O'Hare Airport. I originally thought that the area with the most tickets would be downtown. Through my research, I did find that there are special tickets that can be given downtown for parking during rush hour. Finally, I broke down the number of tickets per hour. After studying the trend, it appears that most ticket issuers start their shift at 6 AM and finish at 3 PM.

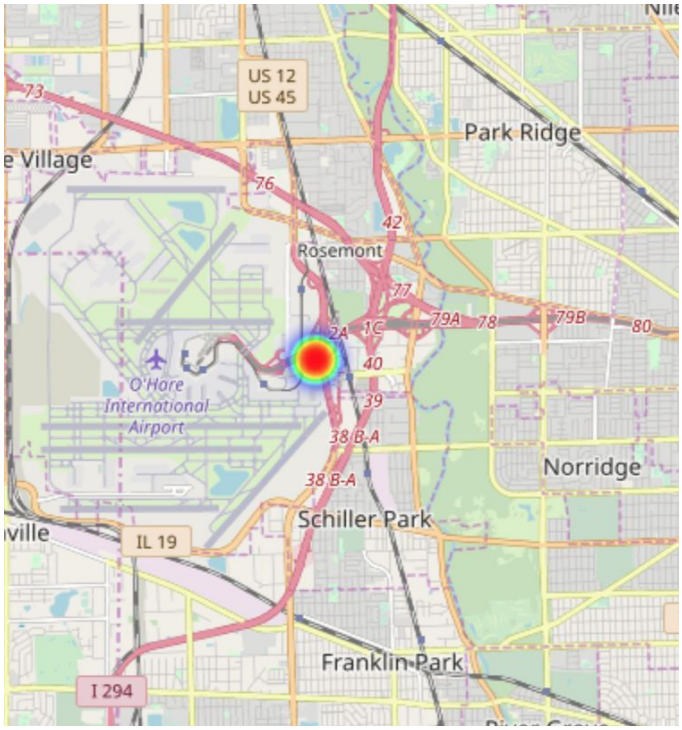
While my mentor and I were brainstorming how to create a machine learning aspect to the data, we decided to run some basic naive based statistics on the provided info. I wanted to learn more about my daily illegal parking habits and the probability of acquiring a ticket. To figure out the probability, I recreated June 9, 2017, and then broke down tickets by zip code. Since I typically park in 60657 zip code and I cannot be parked on that street from 6pm to 6am without risking getting a parking ticket, I chose to investigate this further.

My first finding from this data showed that 8,052 tickets were issued that day where 449 of that total were issued within the 60657 zip code. At this point, I was becoming a little concerned because this averaged out to over 15 tickets per block in the neighborhood. While this prompted concern due to the high ticket total per block radius, I was reminded that tickets could only be issued from 6 pm to 6 am. I broke this data set down further to see how many of the 449 tickets were issued between 6 pm and 12 am of that day. I found that 21 tickets were issued within that six hour time frame. Overall, 4.6% of the tickets issued that day were given the time that I spent illegally parked. After analyzing this data, I feel like I can confidently take on those odds.

Models Summarized

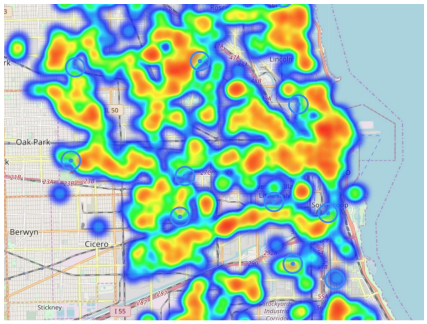
Through the modeling section, I will explain my finding as well as provide images that will help further analyze the data for my audience. There are 6 images that will help explain what the summarized data shows about the ticketing and parking habits of the residents of Chicago.

Where Does the Top Ticket Issuer Operate?



As I canvassed the data set, I noticed officer #716 continually appeared as a major ticket issuer. This individual has issued around 190,000 tickets. Officer 716 has issued more tickets than anyone else in Chicago. I wanted to figure out the region this individual oversaw and after a little time with Folium, I found that officer 716 is located at O'Hare airport. When being confronted with the idea of an individual issuing 190k in tickets, I hope it makes you think twice about leaving your car in areas out of the proper zone, with "no parking" signs or street cleaning warnings.

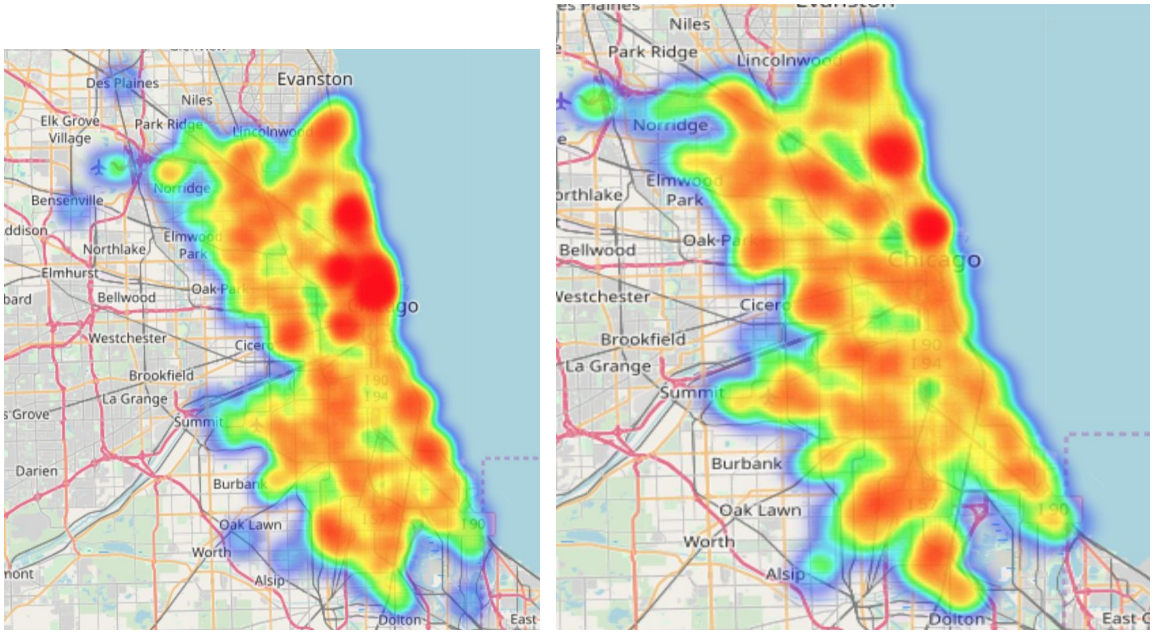
What Does a Slow Day of Ticketing Look like in Chicago?



Do you ever have a rainy day and think individuals issuing tickets may be taking it easy that day? Even though the number of tickets issued is down 60% on raining days, are you still safe?

Technically, you are not safe. The illustration above shows that on a day with only 3900 tickets there are few safe places to hide. The areas with the most tickets are issued in the most traveled areas of Chicago. The more red an area appears on the map, the higher the number of tickets issued. I included the location of the police stations in Chicago to show that there are more tickets issued near the police stations.

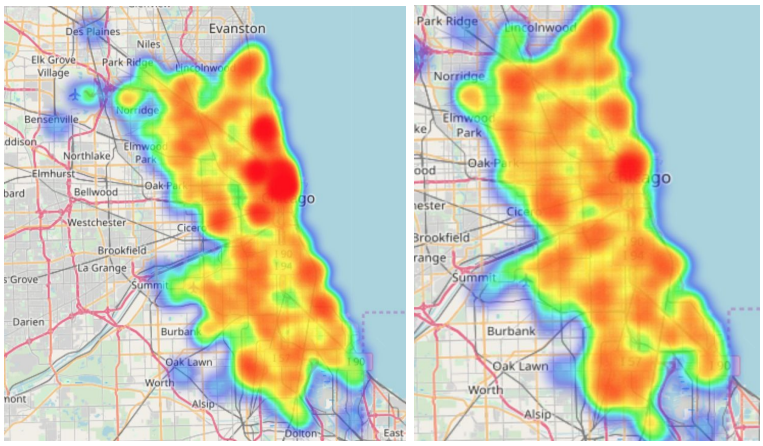
Weekday Vs. Weekend



The image on the left shows a summer weekday and the image on the right shows a summer weekend. When examining the image on the left, it is possible to see the large red dot hanging over the center of Chicago. The large red dot is a product of the expired meter central business district ticket. The EMCBDT tickets can be only issued in downtown Chicago and this kind of ticket makes up close to 10% of the issued tickets.

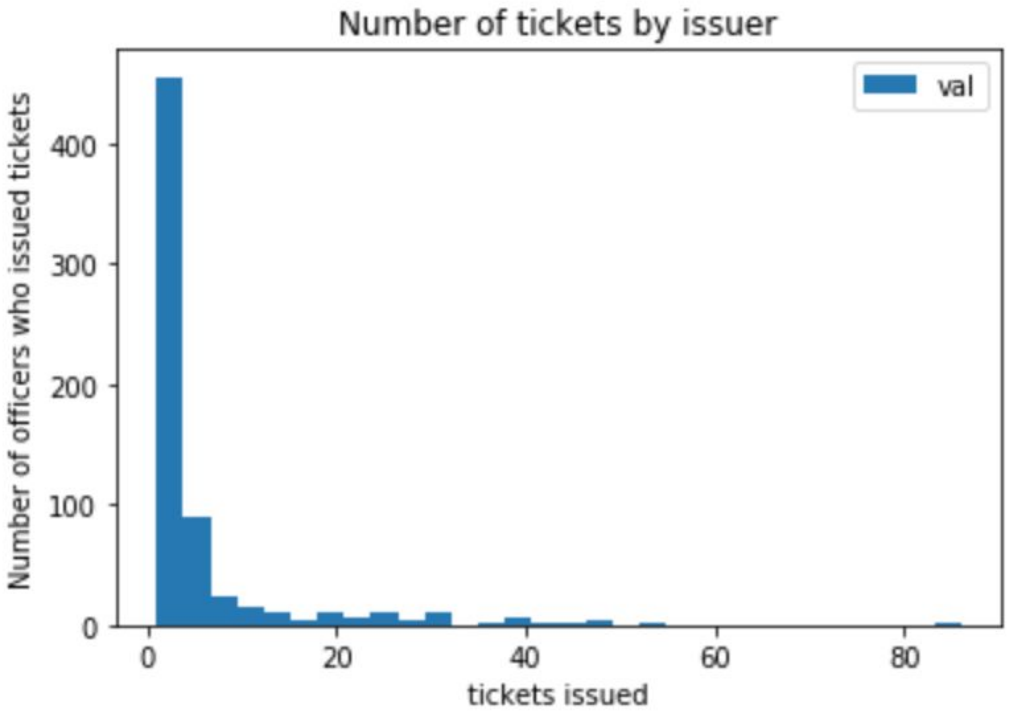
In the image on the right, you can see that the large red dot over downtown disappears. Typically downtown Chicago receives much less traffic on the weekend when the individuals making a 9am-5pm commute are back in the suburbs and surrounding areas for the weekend. Looking at this, it is evident that the number of tickets at Montrose Beach area grows tremendously. If you are at a picnic on the weekend make sure you are parked legally.

Winter Vs. Summer



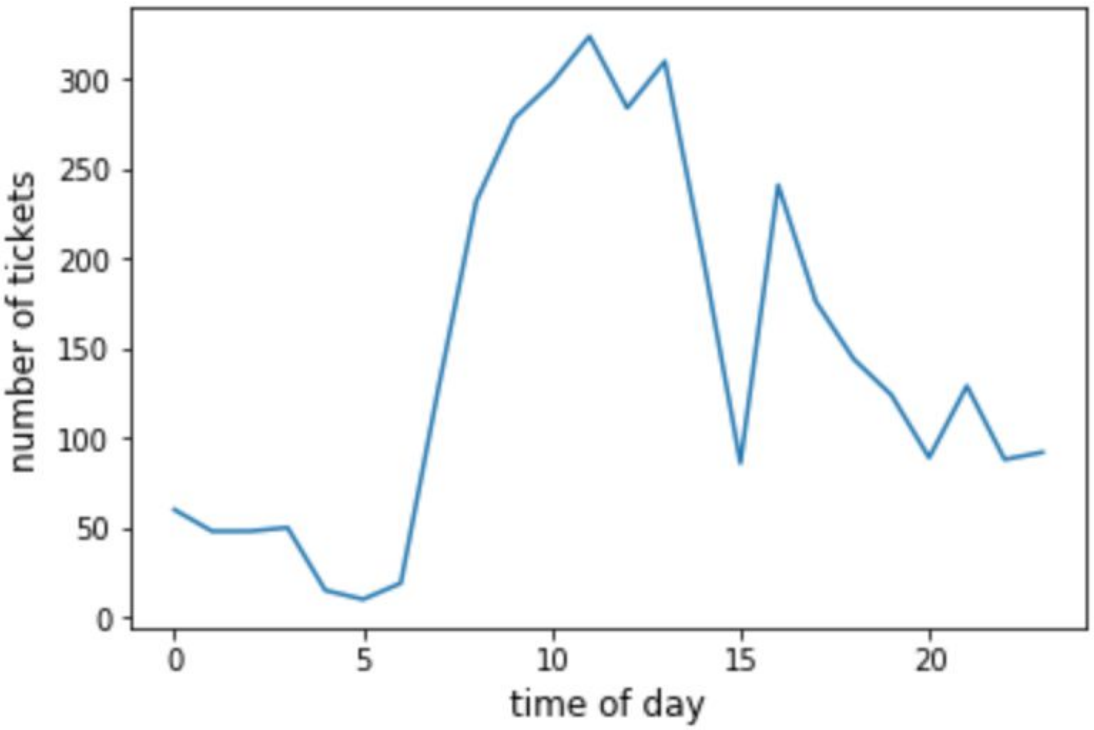
The image on the left is a distribution of tickets on a winter day, where the image on the right is a ticket distribution on a summer day. In the winter of 2017, around 479,000 tickets were issued where the summer of 2017 saw 576,583 tickets issued. Chicago saw 90,000 fewer tickets in the winter than the summer. The other trend between the two images shows a huge decrease in tickets on the waterfront during the winter time.

Histogram of ticketing activity



I wanted to see a distribution of how many tickets were issued by each officer on a specific day. Upon investigation of December 26, 2017 I found that the large majority of officer issued fewer than 10 tickets. Around 450 officers issued a meager 1 ticket that day where the most aggressive officer issued 86 tickets. I looked further into who issued 86 tickets that day and it was officer 716 located at O'Hare airport.

When are the tickets issued



This graph shows the breakdown by the hour when tickets are issued. The most tickets are issued between 6 am and 3 pm. The decrease in tickets at 3 pm is due to shift changes. After the shift changes happen, the number of tickets issued increases again.

Navien Based Stats

449 tickets in 60657	6 pm to 12 pm 21 tickets	4.6% of tickets issued within 6 hours	Low probability of getting a ticket issued in that time

This image takes a Friday night in the 60657 zip code of Chicago and breaks down the probability of getting a parking ticket. On this day in the 60657 zip code, 449 tickets were issued. In zone 383 of 60657, you cannot be parked from 6 pm to 12 am without a parking tag. If your car is left in the neighborhood from 6 pm to 12 am in that time, it is good to be reminded

that an average of 21 tickets was issued in that same time frame. These 21 tickets are 4.6% of the daily tickets issued within 60657 zip code. There is a low probability of getting a ticket while your car is unattended.

Walk Away

As a person who has definitely parking illegally in Chicago, I have surmised that it comes down to a game of calculated risk. No matter how good you are at parking illegally, you will be caught eventually. As the data shows, play it smart when parking on the lake during the summer and be cautious when parking downtown during the week or any time while at the airport. In other words, when you see signs warning you to not leave your car unattended at O'Hare, be reminded of notorious officer 716 who is out to ruin your day.