Jonathan Dyer
jbdyer

Pittsburgh Parking Authority

Questions: What are the analytics questions?

In the case of setting the 'optimal' rates for new electronic parking meters throughout Pittsburgh, several important issues come immediately to mind regarding business and strategy decisions that could be benefited through data analysis. These directly align with the goal of meeting the PPA's express goals and duties[1]:

- How do we best ensure "available parking to residents and merchants"? In other words, how do we make sure there is enough parking **in the areas** it is demanded *and* **at the time** it is demanded?
- How can we guarantee a "fair and equitable" amount of parking spaces for people? That is, how do we establish **hourly rates** and **time limits** that help alleviate parking problems and allow enough time for the average person to accomplish all that they would like while they are parked?
- What is the best way to "promote the proper flow of traffic" in Pittsburgh, or rather how do we leverage changes in parking rates to help **manage congestion** and increase the use of underutilized space?
- In the interest of "protect[ing] the rights of the parking public", where and at what times should PPA officers focus their attention in order to **enforce parking meter restrictions** during the day?

All of these questions represent the ultimate **business value** inherent in pursuing analytics for parking. Data science techniques will be useful for tackling all of these problems (as we see below), but it is crucial to keep the end in mind as we attempt to do so, or else we risk losing sight of what the project is all about.

Data: What data is relevant to this scenario?

There is one **primary dataset** that should be readily available to the Pittsburgh Parking Authority that will help us answer the above questions through analytics. This would be the Parking Meter Usage Data (I will call it the PMUD), which should include (at least) the following details:

- How many people used this meter to pay for parking?
- At what times throughout the day?
- For how long?
- What was the rate?
- Did they extend their time? How often/how long before it expired?

[1] http://www.pittsburghparking.com/enforcement-meter-policies

This and similar data should be readily available for every electronic *Pay by License Plate* meter in Pittsburgh, for every day that meter has been functional. Such a dataset is a treasure trove of information to help us answer our analytics questions.

Some **secondary datasets** that could be useful to have for getting more accurate or detailed insights are found below:

- Weather data for the same days as the parking data above would allow us to draw correlations between weather patterns and parking patterns. An example of where to find such data would be at climate.gov[2], where past weather is available in tabular form. Such data would include:
    - What was the temperature throughout the day?
    - What was the weather? Was there rain/snow/wind?
    - Was the weather localized/where were different weather patterns found?
- An events/holidays/public meetings calendar would help to identify otherwise anomalous parking patterns at certain times of year or during special events. In this way we could learn how to plan for events where large unusual changes in parking may occur. This data would include:
    - When and where did these events take place?
    - Roughly how many people attended these events?
- The PPA has a Customer Satisfaction Survey[3] that allows for qualitative feedback about particular meters or areas. This data (especially if the survey were widely circulated) could help identify problem areas throughout the city that might otherwise go unnoticed.

Note that these secondary datasets are not essential, but would simply be nice to have for data analytics purposes.

Methods

EDA and Visualization

Although not a flashy answer by any means, the truth is that much of the low-hanging fruit in this situation will be captured through exploratory data analysis (EDA) and data visualization. By exploring the PMUD and searching systematically for anomalies related to our questions, an analyst will find patterns that indicate an area where progress could be made. An example should help illustrate what I mean:

> You want to deal with the issue of setting appropriate time limits in commercial areas throughout the city, as per our second analytics question above. In some feedback from

[2] https://www.climate.gov/maps-data/dataset/past-weather-zip-code-data-table
[3] http://www.pittsburghparking.com/meters

> customers you have noticed that a particular section of South Side seems to be particularly troublesome.
>
> The PMUD has usage data for those meters since they are the new electronic kind. After some investigation, you see that on certain days, many people are paying for a long period of parking and then renewing as much as possible until their time runs out. After overlaying a calendar of events in Pittsburgh with those unusual days, you discern the cause right away: those are game days for the Steelers or the Pirates. This particular section of South Side has numerous sports bars where people go to watch sports games (and Pittsburgh has a notoriously devoted fanbase).
>
> When you check the maximum time allowed at the meters in that section, you find that it is only 2 hours—hardly enough to watch a full game!
>
> You start searching for a solution that allows these fans to enjoy enough time in South Side without blocking any other potential customers from ever finding a spot. By generating heatmaps representing the density of parking utilization throughout the day in that section, you notice that just a block or two off the main street, there are a number of meters that are consistently underutilized, even on game days.
>
> At this point there are options to alleviate the problem, such as lengthening the time allowed on these underutilized meters and improving signage to lead more drivers to those locations. Alternatively, the maximum time at any given meter could fluctuate throughout the day in response to shifts in demand or "unique demand generators"[4] such as the sports bars.

The key here is not the exact steps taken to generate the insight about the sports bars, but that the rule regarding maximum time was no longer seemingly arbitrary but instead became data-driven. Hopefully this fictional example illustrated the type of detective work that can lead to some of the most useful insights. It isn't glamorous, but as long as we keep the end in mind—our questions up above—this kind of activity can be highly fruitful.

Predictive Models

Another option to make use of the PMUD and related data would be to employ a more advanced machine learning technique such as LSTM or other time-predictive models in order to predict demand for parking in the future. If an accurate model of this type were available, then the results could be used to guide answers to our questions, e.g. where enforcement officers should spend the most time, if rates should go up or down in response to changing demand throughout the day, etc.

---

[4] https://www.parknews.biz/article?id=11

Jonathan Dyer
jbdyer

Actionable Next Steps

There are three main steps that can be acted on right away, that will facilitate the goals set out above. I recommend that they are followed in order—they can be done in any order, but following them as outlined will lead to the greatest value being generated as quickly as possible, and will set you up for future success.

1. **Prepare a data pipeline.** The first step is to establish a consistent, repeatable method of collecting and preprocessing the necessary data. This will set the stage for all future analysis, and involves at least the following:
    a. Acquiring/extracting the PMUD dataset mentioned above
    b. Ingesting the data into a usable form that allows for processing by data analysts
    c. Cleaning/preparing the data for modeling and exploration

    Once the data is prepared in this way, you will be able to pursue other goals and techniques much more rapidly, saving time and money. Having a pipeline like this will also allow you to continue inputting future data as it is collected, which will improve your results.
2. **Identify problems and search for anomalies.** This step has two parts:
    a. *Identifying problems*: The type of investigative analytics described in the fictional story above is much easier to perform if there is a specific problem that can be addressed. In the story, the problem stemmed from complaints via a qualitative survey that the PPA issues. Such feedback is an excellent way to identify potential issues to focus on.
    b. *Searching for anomalies*: This is really just another way to identify problems. As data is visualized, less obvious patterns will start to emerge. For instance, searching for hotspots where a very high number of people park for a very short time can help identify areas of focus for enforcement officers, or opportunities to fine tune the hourly rates in response to demand throughout the day.
3. **Develop predictive models that support your goals.** In accordance with your specific needs as identified in the previous step and elsewhere, you should use this data to develop models that allow you to stay ahead of parking trends and proactively solve potential problems. For instance, you may find that certain parking areas are over- or under-utilized when nearby events are taking place, and therefore by predicting demand in those areas for future events you can prepare appropriate signage and hourly rates to help balance out utilization and decrease congestion.