Part 1: Mapping questions to data

Dataset name	Description	Metrics
Vacant-Foreclosed-Abandoned Properties	List of vacant, abandoned and foreclosed properties	Property type (vacant/foreclosed/abandoned), address, violation date (2014 to 2020)
QoL_calls	Summary of Quality of Life calls received by Public Safety for each month of 2020.	Call type (barking animal, blocked driveway, noise complaint, etc), number of calls, month.
Annual report Board of Adjustments	List of residential and commercial units that come up in front of Paterson Board of Approvals	Date of the meeting (2017-2019), applicant name, address, phone number, block number, lot number, number of units, commercial type
Incidents Non-Fatal shootings	List of non-fatal shootings during 2020	Date, address, hour
911 Calls for Service	911 log of calls	Date, call hour, call type, status, closed date, last activity date, outcome, latitude, longitude, zip code, address.

For each dataset, what metrics would be useful for our problem? Map each question to the metric that might be used to answer it. Describe in detail what the analysis should be, as well as any assumptions you made or further questions you have about the data.

Are application permits being processed efficiently?

I would start by looking at the **Annual Report for the Board of Adjustments** dataset. I'd be curious to group by **lot numbers** and filter for those which appear in consecutive years (**date of meeting > year**). This may serve as a surrogate for "inefficient processing." Assuming this filter produces results, I'd then look to see if these repeating lots have similar or different applicants (**applicant name**); the former may indicate that permits are not going through, the latter may indicate lots with particularly high turnover. I would then be interested in looking closer at the data to ask the following questions: Are there zip codes (**address**) which are enriched for high turnover? Are there trends in the type of application that are or are not being processed? Perhaps the commercial applications *are* being processed but residential applications are not, or vice versa. Are there trends in application processing for certain size residential properties (**number of units**) or certain types of commercial applications (**commercial type**).

What is the trend in opening/closures of businesses?

In addition to the relevant insights gained after asking the questions listed above (some of which also pertain to this question), I would again look at the **Annual Report for the Board of Adjustments** dataset. I'd group first by **block number** and filter by **date of meeting** to see if there are any coarse trends in the number of applications for a given block, which would point to turnover frequency by location. For **lot numbers** associated with blocks that passed this filter, I'd inspect **applicant names** to see if the applications were being sent by the same person or different people, which might help us understand if a developer is trying many different businesses in the same location or if many people are trying the same location. I would also group by **date of meeting** and look at the frequency of individual **commercial types** to see if certain types of businesses are coming up for application more frequently. I'd also want to look at the **Vacant-Foreclosed-Abandoned Properties** dataset. I'd begin by asking if there are particular zip codes (**address**) enriched for certain **property types** in this dataset, and could do this test across time (**violation date**) to observe the trajectory of a particular area. This might help make choices about which locations are worth focusing on immediately and which locations might serve as either positive or negative controls for what is/isn't working.

Are existing businesses faring well?

With the given data sets, we can ask about some facets of what that question might be getting at. "Fairing well" is a fairly wide question and I think I'd like to get more specific about what this means. The data provided can address, in particular, whether or not **QoL** or **911 Calls for Service** are occuring around businesses, whether there are high concentrations of **Vacant-Foreclosed-Abandoned Properties** around businesses, and whether there have been frequent reports of **Incidents of Non-fatal shootings** around businesses. For each of these data sets, the **addresses** may be used to identify certain businesses that are in areas that have high rates of QoL issues, crime, or unused properties. This doesn't necessarily address how the business is "fairing" but again, I think I'd like to understand what's behind the question to address it more completely.

Do residents feel that the City is safe and liveable?

This question is most directly answered by the **QoL calls** dataset, as it provides a direct connection to the direct emotional expressions of individuals. The question that comes to mind when I consider looking at this data set is "are QoL reports called in by certain demographics more often than others?" I mention below how demographic data could help point to representation differences within these data, but I think it would be very important to know who's calling in reports so that any choices made in response to those reports could be weighted by any representation differences present in the data.

Are Public Safety response times decreasing?

I would look at the 911 **Calls for Service** dataset. Assuming the **last activity date** includes the time, I would extract the hour, calculate the difference in last activity hour and **call hour** for each call, average these differences across weeks and plot across all weeks for which data is present. One could also consider all call times and group by **zip codes** to see if average call response times vary by location. A fancy infographic might combine both and allow one to scrub through time on a map to see how response times vary by location and time. I am picturing a

dynamic irregular heatmap where a map of zip code regions have response times encoded to a color scale. As you move a slider through time, each zip code region changes color based on response times.

Are shootings (fatal & non-fatal) decreasing?

For non-fatal shootings, I would look first to the **Incidents Non-Fatal shootings** and calculate and plot the **number of incidents** across **dates** for the 2020 year. Since that's only going to be a small window — and during a global pandemic — I would also look at the **911 Calls for Service** dataset and attempt to filter **call type** to be shooting reports, further filtering that by **outcome** in an attempt to glean fatal and non-fatal outcomes. Again, I would also want to see these data plotted by **zip code** to understand how different areas are affected. The **QoL_calls** data set may also have gunshot reports that could be cross-referenced against the two aforementioned data sets.

What is the trend for other crime types?

As described similarly above, the **911 Calls for Service** data set would be my first start. I would filter out **call types** by shootings and group by each type, plotting the number of calls over time first, then breaking up into different **zip codes** to see location-dependent trends.

What are the locations?

I think I anticipated and answered this in the questions above.

Are vacant, abandoned and foreclosed properties dealt with in a timely manner?

I would begin by merging the **Vacant-Foreclosed-Abandoned Properties** and **Annual report Board of Adjustments** datasets by address. I would filter **addresses** from the Vacant-Foreclosed-Abandoned Properties set which appear in the adjustment meetings set, then further filter these addresses to comprise only those which have adjustment meeting dates after violation dates. For each of these addresses, I would calculate in a new column the difference in time between the **violation date** and the adjustment **meeting date.** Though this would be limited to the 2017-2019 time range in the Board of Adjustments data set it would still help give a sense of change over time.

What other metrics would ideally help the City of Paterson track the evolution of their efforts to revitalize the downtown area? Create a new list of metrics and datasets we can request the city to collect and/or share.

I'd like to have access to demographic data associated with many of these datasets. Understanding the representation of meeting attendees, permit applicants, and business owners could help address potential

inequities in the processes. Is it the case that certain demographics are applying for permits more/less often? Is it the case that permits are being granted more/less often to certain groups?

I think getting a better understanding about what, exactly, they mean by “revitalize” would be very helpful. Who, specifically, do they want to bring downtown? What metrics will *they* be using to assess success? Do they want mixed-income residences or “luxury” apartments? What types of businesses do they want to attract? Who are, historically, the primary patrons of these businesses? Do the incomes of individuals who live near the “revitalization” align with the cost of goods sold by the businesses they want to attract? Do they want chains run by out-of-state developers or small businesses run by local residents? A mix? What is the timeline? Who will evaluate, when and what will shift in the approach if success isn’t being met within that timeline? I think more conversations are needed to understand their needs and wants before specific metrics requests can be made of them.

I would also like for them to provide some examples of cities that have done work in this area. Which cities embody success as they define it? Who do they look up to? What worked for them? What didn’t work for them? In what ways are they similar or dissimilar and what obvious pitfalls can be avoided by looking across multiple cities that have done this work? Are there obvious solutions that don’t require extensive data modeling to implement? What are examples of both short-term and long-term projects that have seen success? Or failure? Are there projects that consistently “seem” like they work—or are attractive to the media or trends—but don’t actually move the needle on any important metrics? Where did other cities get their investments? What type of developers or investors might be more or less-likely to align with their long-term values? If success is met, what happens next? How do you continue to support the businesses/residents you brought in?