Data Querying 101

Getting the right information from your data

It's not enough to collect data. You need to collect the right data and know how to find it. Any procured software should help with this task just as much as its main function. Any platform, from scheduling & dispatch to fare payment or CAD/AVL systems, are part of transit systems' technology stacks to help agencies accomplish certain service goals. As a result, an agency should be able to use those same platforms to track if and how they are helping with that effort. To do so requires access to useful datapoints that answer the questions that pop up about certain aspects of operations, and that can be combined with datasets from other sources to address bigger-picture questions and make larger planning decisions. To ensure you get the right data, knowing how to ask for and query it is a critical skill.

Setting Up the Data Environment

Pre-Procurement

While procuring new technology, understand what the platform should accomplish and then how to monitor whether or not it is accomplishing that goal. As a result, contracts with vendors need to specify the data that the platform will collect, and how agency staff will be able to access and use it when and how it works for them. This will take some upfront work that will pay off down the line.

First, answer a few questions that will guide your efforts:

1. What is this platform going to accomplish? How will it affect service or operations?

   _____
   _____
   _____
   _____

2. What are you trying to answer, solve, or track to make sure this is succeeding?

   _____
   _____
   _____
   _____

3. What datapoints can help to accomplish that? How will you measure success (i.e. what are your KPIs)? Make sure to answer why each datapoint will help.

   _____
   _____
   _____
   _____

4. How do you need the data displayed? For example, in a spreadsheet? In a map?

   _____
   _____
   _____
   _____

Also, *check out our Data Ownership Factsheet for more information on how to establish data relationships*

Need some examples of datapoints that transit systems rely on? You'll probably recognize a lot! Below are a sample of the universe of possible transit data points.

Examples of Common Datapoints	Examples of Advanced Datapoints
Overall ridership	Ridership broken down by origin/destination/bus stop/route
Revenue miles/hours	Areas where delays consistently occur
Passengers per revenue mile/hour	Ridership by vehicle or certain timeframe (demand-response)
Cost per passenger trip	On-time performance
Cost per hour	Customer wait time

Interested in learning more about datapoints and how to use them? Take our Data Dashboards 201 course!

Implementation

Once an agreement is in place, here are a few points to check on the data side during the system's setup phase:

- Ensure data is reporting the way you want it to. Is the formatting correct? Is the information populating accurately?
- Ensure all relevant staff know how to access reporting and that it is easy to show new staff how to find it
- Double-check all tech system reports to ensure they are accurate and account-for as requested

Trying to improve the data environment with a current vendor?

- Document what you want and what's lacking in your current process
- Try to work with the vendor to improve data access and quality
- If that doesn't work, use this to inform a new vendor procurement

Ask Good Questions

Now that the data is set up, it's time to understand how to make good use of it. The key is to know how to ask a good question. Otherwise it's easy to find yourself swimming in numbers without a clear direction in which to use them.

There are three primary elements that make a useful data query in the transit space:

1. **Characteristic**: what about your system are you studying? Routes? Stops? Specific vehicles?
2. **Key Performance Indicators (KPIs)**: what do you want to know about the characteristic(s) that you are studying? Ridership? Delays? Maintenance issues?
3. ***Timeframe***: a longer timeframe can help you identify trends, and a shorter timeframe can help you understand how the events of a specific event or incident played out

These three elements can help you craft the right questions to make more-informed decisions.

Understand

First, go back to the beginning: what is the problem or question you would like to address or track? There are a few types of questions you can ask:

- Descriptive: Which **routes** were the least productive ***since our last redesign*** based on the **KPIs** we established with the new network?
- Comparative: What are the **common factors** of our **least-productive routes** compared to our most-productive routes in the ***new network***?
- Analytical: How do we think our proposed service change might affect **ridership** on **these new or adjusted routes** ***over the next five years***?

The elements of a descriptive data query: This basic type of query simply helps describe what has happened to certain elements of an agency. In this case, **routes** are specific, definable parameters to which you can attach performance data; ***since the last redesign*** is a time period that you want to understand; and **KPIs** are the datapoints that describe what occurred to the **parameters (routes)** during ***the time period that matters to you (redesign through now)***. Importantly, the elements of this query are all specific and definable, which means they can be collected and tracked by a machine (i.e. software).

The elements of a comparative data query: This builds on the descriptive query by analyzing the difference among multiple parameters. The **common factors** are the various **KPIs** or other **defining characteristics (e.g. headways, service hours)** collected regarding your routes; the **least-productive routes are the parameters**, which were described in the first query and are now comparing; the ***new network*** is another way of defining ***the time period*** from the network redesign through the present. It might make sense here to compare various descriptive queries side by side to identify differences. That is, comparing descriptive queries helps to understand what makes them different.

The elements of an analytical data query: Now the comparative analysis can lead to a conclusion about how you can affect change. You'll likely have to combine data across multiple platforms and use publicly available information, such as from the Census, to draw these conclusions. In this case, the **new/adjusted routes** are the **parameters** (and the service change is the overarching project that redefines the parameters); **ridership** is the specific **KPI** you're researching in this case; and the ***five-year horizon*** helps to project the ***medium-term*** effects of changes and make decisions based on those predictions of the future.

In all of these cases, the queries use measurable points that have already been defined, are specific and can be used to compare like elements.

A few things to keep in mind when you develop a data-based question:

1. Tie your question to a goal
2. Define your goal with measurable KPIs
3. Build your questions so you not only understand what is happening but why
4. Be able to answer why these questions and answers matter

Forming a query depends on the available software, but it's key to address the three elements discussed above. When using the reporting function of your software, think of a query like this:

During _____*(time period)* what is/are the _____*(KPI(s))* for _____*(characteristic(s))*?

Querying data is just the first step in a process of making that information useful. Below are follow-on steps to keep in mind:

Confirm Data Quality: is it accurate and reliable?

Analyze the Data: what is it telling you?

Communicate Findings: who does this impact, and what should they know about it?

Act on Your Findings: what will you do with this information? Turn these into a nice graphic or pathway. bus stops on a road?