Rail~volution 2017: Improving Transit Improvements Through Big Data

Catherine Manzo
Senior Director, StreetLight Data

catherine.manzo@streetlightdata.com
www.streetlightdata.com
Big Data is Just A Buzzword...

Will Big Data Cure Cancer?
*Fortune* - Nov 2, 2016

While there's a lot of breathless talk about the potential of big data to just ... The Commons now holds the raw genomic data, from the Cancer

Re-thinking Analytics: How big data analytics can revolutionise your ...

Re-thinking Analytics: How big data analytics can revolutionise your ... data sets, and help make sense of the new buzz word “BIG DATA”.

Big Data: Key to Customer Understanding and Service Delivery
MassTransitMag.com (press release) (registration) (blog) - Nov 6, 2016

Transit agencies invest millions to transport people in the safe and reliable manner. Whether public or private, mass

This Startup Is Using Big Data And Smart Tech To Save The Oil ...
*Forbes* - Oct 29, 2016

While experiencing an influx of funding and increased cash flow from ... The failure to adopt a smarter business model driven by big data and ...

Toyota, Intel Leading New Big Data Automotive Consortium
*Windows IT Pro* - Aug 29, 2017

The group, which also includes Denso, Ericsson and Nippon Telephone, aims to get ready for huge connected car data streams ...

How the 'Facebook of music' is using big data to find the next pop star
*Business Insider* - Aug 30, 2017

No longer simply a platform for uploading and discovering new music, the site has employed big-data analysis and large-scale human curation ...
Actual Definition

The Information Explosion

- Data Volume
- Data Variety
- Data Velocity

Analysis Gap

Ability to Analyze

Source: Kay, David and Mark van Harmelen, Delivering benefits from the data deluge, Dec 2012, jisc.ac.uk
What types of big data will you come across?

- On-Board Transit Data
- Sensor Data
- Payment Data
- Cellular Tower
- Navigation-GPS
- Location-Based Services
- Massive Mobile Data or Geolocation Data or Location Data
What types of questions should you ask?

<table>
<thead>
<tr>
<th>Availability</th>
<th>Real time feeds or historic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographic Coverage</td>
<td>Only specific locations? Region-wide? Statewide? Nationwide?</td>
</tr>
<tr>
<td>Behavior Measured</td>
<td>Only one place? Connecting activity between places?</td>
</tr>
<tr>
<td>Sample</td>
<td>Size? Representative?</td>
</tr>
<tr>
<td>Privacy</td>
<td>Personal Identify Information (PII)? Opt-in vs Opt-out?</td>
</tr>
<tr>
<td>Access</td>
<td>Direct feeds? Cloud tools? Need for internal database infrastructure?</td>
</tr>
<tr>
<td>Independence</td>
<td>Useful insights gained from single data stream?</td>
</tr>
</tbody>
</table>
The basics of Massive Mobile Data

- Cellular Tower
- Navigation-GPS
- Location-Based Services
# How different Massive Mobile Data sets stack-up

<table>
<thead>
<tr>
<th></th>
<th>Cellular</th>
<th>Navigation-Based GPS</th>
<th>Location-Based Services</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology Spatial Precision</strong></td>
<td>200-1000 meters</td>
<td>5 meters</td>
<td>5 meters – 50 meters</td>
</tr>
<tr>
<td><strong>Frequency of Data Pings</strong></td>
<td>Irregularly; every 15 min – hours</td>
<td>Regularly; every 1 sec – 1 min</td>
<td>Variable; sometimes triggered by location change</td>
</tr>
<tr>
<td><strong>Type of Trip</strong></td>
<td>Blends personal and commercial trips</td>
<td>Differentiates personal and commercial trips</td>
<td>Blends personal and commercial trips (for now)</td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td>Varies by telco: ~10% of population for small telcos and ~25-30% for large telcos</td>
<td>Varies by region; ~1-4% personal trips; ~10-12% of commercial trips*</td>
<td>Medium – 30M+ US devices in our database (&gt;10% of US adult population)</td>
</tr>
</tbody>
</table>

* Depends on provider
Spatial precision is key for transportation engineering.
This is what you get with 1000m spatial precision
Use Massive Mobile Data to Reveal Opportunities for Converting Vehicle Trips to Other Modes

Scan Large Areas for Transit Opportunities

Drill Down on Refined Areas to Understand Transit Gaps
Use Case 1: Who Lacks Transit Options in Napa Valley?

Regional Study Showed High Commuter Trip Volume → New Bus Routes; Rail Exploration

Internal Trip Starts

- Internal Trips
- Exported Trips
- Imported Trips
- One-way Intra-county Trips
- Pass-through Trips

- 55%
- 12%
- 11%
- 3%
- 18%
Use Case 2: Where Are The Best Places to Expand Transit in NoVa?

Scanning Hundreds of TAZs and Congested Links for Short Trips and Top O-D Pairs

One O-D Pair Represents 30% of AM Trips Between Ralston and Ballston on N. Glebe Rd

Large Volume of Short Vehicle Trips WB on Route 7 during PM Peak

29% of traffic begins at Tysons
22% of trips are < 5 miles
16% of trips are < 2 miles
7% of trips are < 1 mile

Tyson’s Corner

Origins
Destinations

Source of Images: State Smart Transportation Initiative
Use Case 3: Who’s Driving Instead of Using Transit in Denver?

Identify Top O-D Pairs for Vehicle Trips

Origins of Drivers Served by Transit

ZIPs Near Transit Center

Denver Tech Center

Trips that Use I-25 (train route)