Big Data for Transportation Models

Neal Bowman
Sales Engineer
neal.bowman@streetlightdata.com
Agenda for Today’s Presentation

I. What is Big Data?
II. Why Am I Talking About It?
III. Different Data for Different Models
IV. Real-World Examples
V. Q&A
What is Big Data?
Big Data is A Meaningless Buzzword...

Will Big Data Cure Cancer?
*Fortune* - Nov 2, 2016
While there's a lot of breathless talk about the potential of big data, it's tempting to think that the world is a better place now that the raw genomic data, from the Cancer Genome Atlas, is being released to the public.

Big Data: Key to Customer Understanding and Service Delivery
Transit agencies invest millions to transport people in the most seamless, efficient, safe and reliable manner. Whether public or private, mass transit organizations have the opportunity to be smart, proactive and dynamic in their service delivery.

Why Cultural Change Is Necessary For Big Data Adoption
*Forbes* - Nov 8, 2016
Love it or hate it, big data is here to stay. As data volumes and sources of data proliferate at ever increasing rates, leading companies will be...

Re-thinking Analytics: How big data analytics can revolutionise your ... *BusinessTech* - Oct 25, 2016
Re-thinking Analytics: How big data analytics can revolutionise your ... data sets, and help make sense of the new buzz word “BIG DATA.”

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 ...

Big Data Unleashes Business Opportunity
*Forbes* - Nov 1, 2016
Businesses have gone data crazy. You can’t blame them. Streaming, real-time data analysis promises to bring the type of predictability that cuts ...

Mitigating the Security Risks of Big Data
*CIO* - Nov 1, 2016
...But Geospatial Records Created by Mobile Devices Can Enhance and Improve Data-Driven Modeling

LBS Data Location
Circle radii vary: they accurately reflect the spatial precision of each unique data point

Navigational GPS Location
Circle enlarged for visibility

Note: This image shows a filtered subset of data to improve visibility
Today, We’ll Discuss Three Types of Big Data

- Cellular Tower
- GPS
- Location-Based Services
The Impact of Spatial Precision is Significant

---Proprietary and Confidential---
This Is What You Get With 1000m Spatial Precision
<table>
<thead>
<tr>
<th></th>
<th>Cellular</th>
<th>Navigation-Based GPS</th>
<th>Location-Based Services</th>
<th>Key Takeaway</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial Precision</strong></td>
<td>200-1000 meters</td>
<td>5 meters</td>
<td>5 meters – 50 meters</td>
<td>Cellular data doesn’t show exact routes, origins, or destinations because it is spatially imprecise.</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>
<td>Infrequent pings mean cellular data is messy; a LOT additional processing is required to clean incomplete trips.</td>
</tr>
<tr>
<td><strong>Accuracy of Output Metrics</strong></td>
<td>Medium-Low for small geographies</td>
<td>High, assuming good normalization and contextualization algorithms</td>
<td>Medium-high, assuming good normalization &amp; contextualization algorithms.</td>
<td>Lack of spatial precision and infrequent pinging of cellular data leads to inaccurate Metrics.</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>
<td>Only GPS data segments commercial traffic from personal traffic. Commercial fleet management systems are tagged.</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 (for INRIX)</td>
<td>Medium – 30M+ US devices in our database (&gt;10% of US adult population)</td>
<td>Cellular provides largest sample, but GPS and LBS sample sizes are statistically valid for any transportation project.</td>
</tr>
</tbody>
</table>
But Why Should You Listen to Me?
Who Is StreetLight Data?

StreetLight Data is a mobility analytics provider that makes it easy, affordable and intuitive for the transportation industry to use Big Data.

With our easy-to-use web app, you can accurately measure transportation behavior for almost any project in just a few minutes.
We are the First Company to Provide Transportation Metrics Derived from Big Data via a Web App

Input: Big Data
Processing: RouteScience® Technology
Web App Output: StreetLight InSight Metrics

- Anonymous and accurate Location Data
- Road network, land use, parcel, census and more Contextual Data
- Core Metrics: Origin-Destination, Select Link, Zone Activity
- Premium Add-On Metrics: Trip Attributes, Traveler Attributes, Commercial Tours
- Customization: Day Parts, Day Types, Data Period, Trip Type
- Speed: Most Metrics are Ready in Minutes
# Drilling Down on GPS Data from Our Partner INRIX

<table>
<thead>
<tr>
<th>Analyze</th>
<th>To Understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precise Origins &amp; Demographics</td>
<td>- Trip generation</td>
</tr>
<tr>
<td></td>
<td>- Transportation equity</td>
</tr>
<tr>
<td></td>
<td>- Transit and Facility Accessibility</td>
</tr>
<tr>
<td>Precise Destinations &amp; Day/Time</td>
<td>- Commuting &amp; work locations</td>
</tr>
<tr>
<td></td>
<td>- Transportation centers</td>
</tr>
<tr>
<td></td>
<td>- “Last mile” and parking</td>
</tr>
<tr>
<td></td>
<td>- TDM Solutions</td>
</tr>
<tr>
<td>Actual Routes and Trip Duration</td>
<td>- Actual travel times, VMT</td>
</tr>
<tr>
<td></td>
<td>- Variations, reliability, circuitry</td>
</tr>
<tr>
<td></td>
<td>- Route choice, diversion, arterials</td>
</tr>
<tr>
<td>Vehicle Type and Class</td>
<td>- Personal vs. commercial trips</td>
</tr>
<tr>
<td></td>
<td>- Criteria emissions</td>
</tr>
<tr>
<td></td>
<td>- Commercial weight classes</td>
</tr>
<tr>
<td>Custom Date Ranges</td>
<td>- “Before and after” studies</td>
</tr>
<tr>
<td></td>
<td>- Seasonal variance</td>
</tr>
<tr>
<td></td>
<td>- Trends over time</td>
</tr>
</tbody>
</table>

--Proprietary and Confidential--
Drilling Down on LBS Data from Our Partner Cuebiq

Our New Data Partner: Cuebiq
- Location Based-Services Data (SDK provider for mobile apps)
- Archival and anonymous
- Spatial precision 50% better than 25m, 75% better than 50m
- Typically, devices collect 30 distinct activity locations per day
- 30M device sample size → 10% of adult US population
- “Privacy first” policy

New and Enhanced StreetLight InSight Metrics
- True Trip Purpose (Work, Home, Other)
- Additional Trip Purposes
- Activity-Based Modeling, including personal travel days
- Improved Normalization
- Enhanced Demographics

--Proprietary and Confidential--
**StreetLight InSight Metrics**

<table>
<thead>
<tr>
<th>Core Metrics: .shp, .csv, visualizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin/Destination Matrix:</td>
</tr>
<tr>
<td>Relative volume and average travel time of trips between Origin and Destination Zones</td>
</tr>
<tr>
<td>Origin/Destination + Middle Filter (Select Link)</td>
</tr>
<tr>
<td>Relative volume and average travel time of trips that pass through Middle Filter Zones, or links, when traveling between Origin and Destination Zones</td>
</tr>
<tr>
<td>Zone Activity Analysis</td>
</tr>
<tr>
<td>Relative volume, average travel time, and average length of trips that originate in, have destinations in, or pass through each Zone analyzed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Premium Bundles: .shp, .csv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premium A: Trip Attributes</td>
</tr>
<tr>
<td>Trip time, trip length, trip speed, and trip circuitry</td>
</tr>
<tr>
<td>Premium B: Traveler Attributes</td>
</tr>
<tr>
<td>Simple trip purpose, household income, race, education level of head of household, and family status</td>
</tr>
<tr>
<td>Premium C: Commercial Tours (2 Month Lead Time; Not Yet in App)</td>
</tr>
<tr>
<td>The relative volume of tours (within 24 hours) and dwell times at each intermediary stop for tours that begin or end in specified Zones</td>
</tr>
</tbody>
</table>

---

**Project Options**

*Available for All Metrics*

- **Custom Day Types**
  Define Weekends and Weekdays
- **Custom Day Parts**
  Six day parts; customize with bins as small as one hour
- **Data Period**
  Choose the months to analyze from January 2014 through 1-2 months ago
- **Trip Type**
  Commercial vs. Personal
- **Truck Type**
  Heavy Duty vs. Medium Duty
Different Models, Different Data
No One Type of Data is Ideal for Every Model: Choose The Right Combination for Your Application

Micro Simulation

GPS
No One Type of Data is Ideal for Every Model: Choose The Right Combination for Your Application
No One Type of Data is Ideal for Every Model: Choose The Right Combination for Your Application

- Macro Simulation
- GPS
- LBS
- Cell.
Real-World Examples
Calibrating a Travel Demand Model in North Carolina

Transport. Demand Modeling

Origin-Destination for North Carolina MPO

Need: Accurate O/D for calibration or transportation demand model without expensive/time consuming survey for personal and medium/heavy duty commercial trips.

Zones: 155 census tracts

Metric: O/D with demographics

Months: Spring 2016 (3 months)

Time to Run: 9 minutes
A Corridor Impact Study for a New Route

Corridor Impact in Canada

Need: Route choice between 2 highways to measure the impact of offering a new option.

Zones: 6 routes, 3 origin zones, 6 destinations zones.

Metric: O/D with Middle Filter

Months: May, June 2016

Time to Run: 2 min for each month.
Questions?