Unlocking the Value in Vehicular Data Using Analytics

Amir Sayegh, Ph.D.
Data and Analytics Department
12th of September 2019
Executive Summary

- Who is Geotab?
- Big Data at Geotab
- Big Data & Safety
- Current Challenges
  - Data Accessibility Index Project
- Call for Action
A World Leader in GPS Fleet Management

- Engineering company building analytics, telematics & IoT solutions
- #1 Telematics company in World
- Started in North America in 2000, Geotab now has:
  - More than 400 partners and a flourishing ecosystem
  - Over 1000 Employees and offices in Toronto, Las Vegas, Mexico City, London, Madrid, Munich, Aachen, Hong Kong & Adelaide
The Geotab technology stack empowers fleet management.
Open Platform

Enabling 3P data collection and development of data-driven apps
Turning Raw Data into ‘Sensed Data’

- Windshield wiper activations
- Temperature
- Harsh braking
- Speed accelerometer

Road conditions

Traffic

Weather conditions

Leveraging power of big data
Vehicles Become Sensors on Wheels

Road conditions  Weather  Traffic
Big Data at Geotab

>1.8 M
connected vehicles, globally

>40 B
data records processed daily

Richest telematics dataset in the world:
GPS, traffic, accelerometer, engine data, weather, driver behaviour

https://cloud.google.com/customers/geotab/
Innovating with Big Data

- Allows us to aggregate data and provide global insights for both fleets and smart cities
- Sharing aggregated insights via data.geotab.com
  - 12 public datasets
  - Weather, safety, points of interest
Telematics Data can Transform the Community

- Traffic + Safety
  - Intersection Insights
  - Spot speed/virtual pneumatic tubes
  - ...

- More efficient communities
  - Maintenance
  - Ride-sharing
  - ...

- Environmental impact
  - Air Quality assessment
  - Emissions modeling
  - EV Suitability + EVSE Location
  - ...

...
...while Respecting Privacy and Ethics

At Geotab, we make sure our Big Data innovations always RESPECT:

- Delivery of value to our Customers;
- Use is compliant with our Customers’ instructions;
- Data is protected from unauthorized access, use and reconstitution;
- Ethically sourced;
- Doesn’t compromise end users, natural persons or society; and
- Respects data subjects autonomy and their right to make their own decisions.
Fleet Insights Through MVP Program

Predicting battery failures

Fuel economy benchmarking
Smart City Insights Through MVP Program

Empowering smart traffic decisions

Air quality sensor optimization
Safety Using Big Data
Data Enables Safety Understanding and Informs Policy
Helps Fleets, Drivers, Cities Become Safer

Internal Data
- Vehicle
- Telematics
- Mobile app
- Maintenance
- Routing
- Driver Demographics

External Data
- Open data
- Research reports
- Industry trends
- Speed limits

Contextual Data
- Weather
- Dangerous neighborhoods
- Intersection insights
- Vocation

Data from different sources is multiplicative in value
Vehicle & Telematics Data

Real-time and historical data can help predict and prevent crashes

Vehicle Data
- Seatbelt
- Speed
- Odometer
- Turn signals
- Engine faults
- Headlights
- ............

Telematics Data
- Accident reconstruction
- Trip data, speeding profile
- Speeding
- Seatbelt compliance
- Harsh activity (accel., braking, cornering)
- Crashes
- .............

Vehicle and telematics data are vital to understand the “why”
Example: Fleet Safety Policy Design

Insights
- Driver behavior profiling: Speeding, harsh braking, harsh acceleration, etc.
- Impacts of demographics on safety/performance: Ex: tenure
- Most efficient scheduling practices
- Trip optimization – real time routing based on time/distance/outside factors
- Accident predictions: costs, type, fault type
- Collision: both predictive as well as collision reconstruction
- Maintenance and roadside assistance analysis

Fleet Policy Design
- Routing
- Coaching
- Exception rules
- Driver Risk scoring
- Proactive maintenance
- Shift schedule re-design

Using big data, we can benchmark fleets, drivers, and detect issues in real-time
Sensor Data + Big Data + AI = Safer Cities

Machine Learning - Predict Dangerous Driving

Predict Hyper Local Weather Down to 150m

Predict Vehicle Movement and Traffic Patterns
Supporting NYC Vision Zero

- Measure spot speeds and travel times in NYC
- Use the tool to make data-driven safety decisions
  - Road redesigns
  - Traffic calming
  - Lane narrowing
  - Siting speed cameras and humps
Public Safety- Hurricane Dorian Impact

Time-lapse ground-truth probability of precipitation over the last 24 hours

Time-lapse commercial traffic over the last 24 hours
Current Data Challenges
Current Data Challenges

Key Concern: Ongoing real-time access, expressed by major private/govt fleets, leasing companies, etc

Data Silos
- Data is being collected by different stakeholders
- No method for sharing
- No ‘will’ to share

Data Access
- Data access restrictions
- Not all data collected
- Not all data stored
- Different sampling methods

Data Formats
- Different data formats
- Differ by manufacturer

Only ‘hack’ today is reverse engineering; concerns loom over reduced access
Current Work by Geotab: Data Accessibility Index Project

- Current concerns from customers about data access going forward
- The Data Accessibility Index Project is designed to bring reliable data, analytic rigor, and transparency to the vehicular data access debate.
- Empower car owners, car buyers, commercial & public service fleets, mobility providers & innovators, public authorities, service providers

“You cannot manage what you cannot measure.”
The 5 Step Methodology

1. **Identify manufacturer**
   - Make, Manufacturer, and group

2. **Vehicle stats**
   - Region of domicile
   - Distance and time driven
   - Odometer, fuel and seat belt support

3. **Apply Exclusions**
   - Vehicle age
   - Usage requirement, etc

4. **Calculate metrics**

5. **Privacy filter**
Current Output (Sep 2019)

Of the 1M+ vehicles considered:

- 94% have access to 3 basic parameters
  - odometer, fuel, and seatbelt
- average consistency score = 0.6
Magic Quadrant: Sample

OEM Data Accessibility Magic Quadrant

**5 Least Accessible Models**
1. Chevrolet (GM) - Spark M300
2. Chevrolet (GM) - Box_8B
3. Isuzu - NPR / NPR HD
4. Toyota - Corolla
5. Subaru - Forester

**5 Most Accessible Models**
1. Volkswagen - Caravelle Transporter
2. Nissan - Versa Note
3. Freightliner (Dafter) - Sprinter Van
4. Volkswagen - GTI / Saveiro
5. Ford - Fiesta

*Full Data Accessibility = Support for V/S, C/O, Fuel, and Speed*
Data is necessary but not sufficient; simply amassing it does not generate value.
Call For Action!

- Drive a culture shift around openness of data
  - Recall: Data has no ‘value’
  - Recall: Multiplicative value of data
- Create a data trust
  - Privacy and ethics first
- Standardize data collection and storage
  - Use VSS standard as basis
  - Standardize cloud to cloud sharing
- Work with fleets, governments, citizens
- Build research testbeds
- Open sourcing of safety algorithms
How is this made possible?

1.8M + Vehicles  →  Gateway 1  →  Production cluster

Gateway N  →  Big data streaming

Google Cloud Platform

Google BigQuery

40B+ Daily

Near real-time big data cloud ingestion
Professional Services Offering

- Rapid prototyping service
  - Minimum Viable Product (MVP)
- Leverage our Data Science, Data Engineering, and Data Visualization teams to develop high impact solutions that marries Geotab data together with line-of-business data and other 3P sources
- Project based:
  - High Impact
  - Engaged Customer
  - Scalable

Data innovation through ‘learning by doing’
The Importance of Contextual Data...What is it?

<table>
<thead>
<tr>
<th><strong>Examples</strong></th>
<th><strong>Weather</strong></th>
<th><strong>Hazardous Areas</strong></th>
<th><strong>Zoning</strong></th>
<th><strong>Roadway Info</strong></th>
<th><strong>Time of Day</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temperature</strong></td>
<td>Temperature</td>
<td>Historical Incidents</td>
<td>Residential</td>
<td>Highway</td>
<td>Day</td>
</tr>
<tr>
<td><strong>Precipitation</strong></td>
<td>Precipitation</td>
<td>Abnormal Incidents</td>
<td>Industrial</td>
<td>Intersection</td>
<td>Night</td>
</tr>
<tr>
<td><strong>Visibility</strong></td>
<td>Visibility</td>
<td>Intersections</td>
<td>Rural</td>
<td>On/Off Ramp</td>
<td>Weekend</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Construction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Maria and David drive similarly

Maria

- Miles Driven: 5.4 km
- Hard Braking: 2
- Hard Acceleration: 1

David

- Miles Driven: 5.3 km
- Hard Braking: 2
- Hard Acceleration: 1
But they have very different routes

Maria

- Miles Driven: 5.4 km
- Hard Braking: 2
- Hard Acceleration: 1
- Hazard Area Events: 3

David

- Miles Driven: 5.3 km
- Hard Braking: 2
- Hard Acceleration: 1
- Hazard Area Events: 0
Different feedback is required... smarter policy

Maria drives well in hazardous areas compared to peers

David drives poorly when driving on the highway
Let's Stay Connected:

@GEOTAB