Replica

How it Works

Raw Data Layer

We leverage a diverse set of third-party source data to create our models.

This composite approach is both a risk-mitigation strategy and aligned with our objective to show a holistic view of the built environment.

Location Data

Consumer & Resident Data

Built Environment

Economic Activity

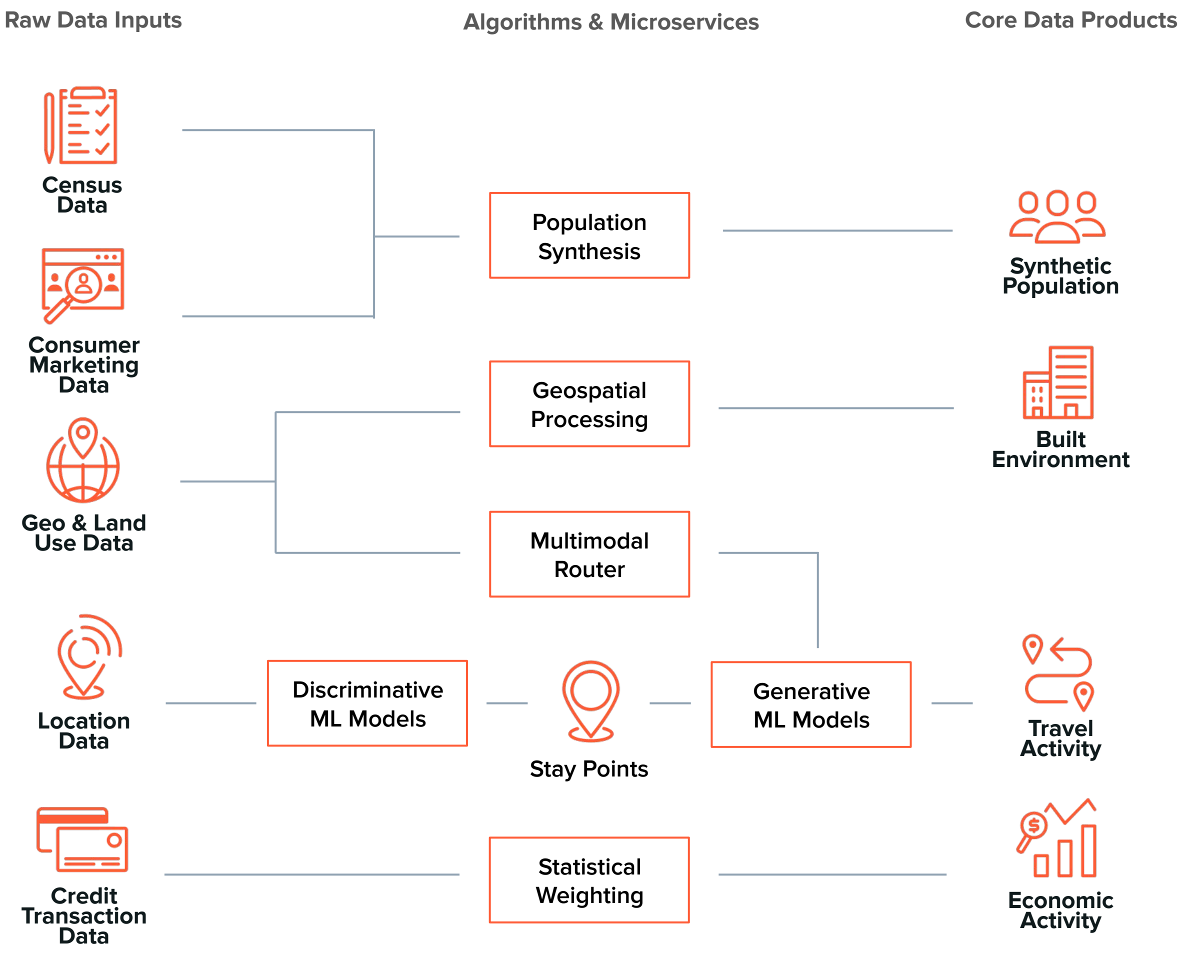
Ground Truth Data

How it Works

Algorithms Layer

Replica generates its data by running large scale, computationally-intensive simulations — a “replica” of transportation and economic patterns.

As the quantity and variety of available raw data continues to grow, we introduced a privacy-preserving algorithms layer that produces composite synthetic core data sets in a unified schema.



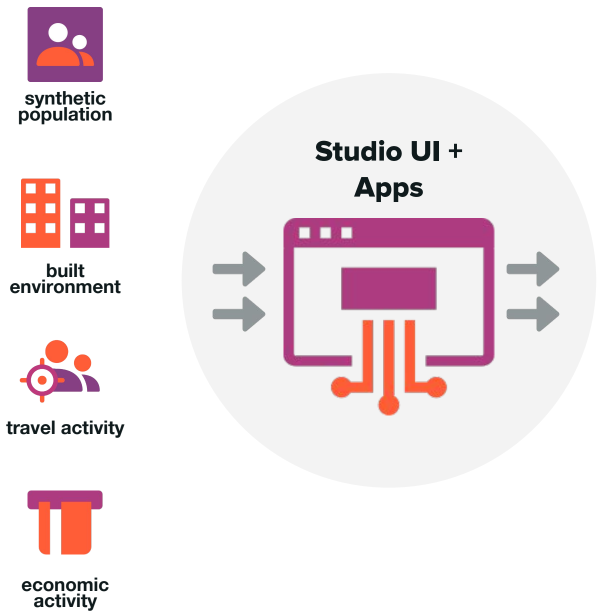
How it Works

Core Data Access APIs

The Core Data Products produced by the Algorithms Layer both feed our platform applications and can be delivered directly to customers.

3rd party applications are built via core data access API.

Core Data Products



Platform Applications

Trends
Weekly updates, less geospatial fidelity, essential feature set

Places
Seasonal updates, more geospatial fidelity, exhaustive feature set

Direct Access + Apps
Customer ingestion of CDPs for proprietary or custom use and/or applications targeting high value use cases

REPLICA

Population

Census Households Surveys + Consumer Marketing Data

Public Use Microdata Files

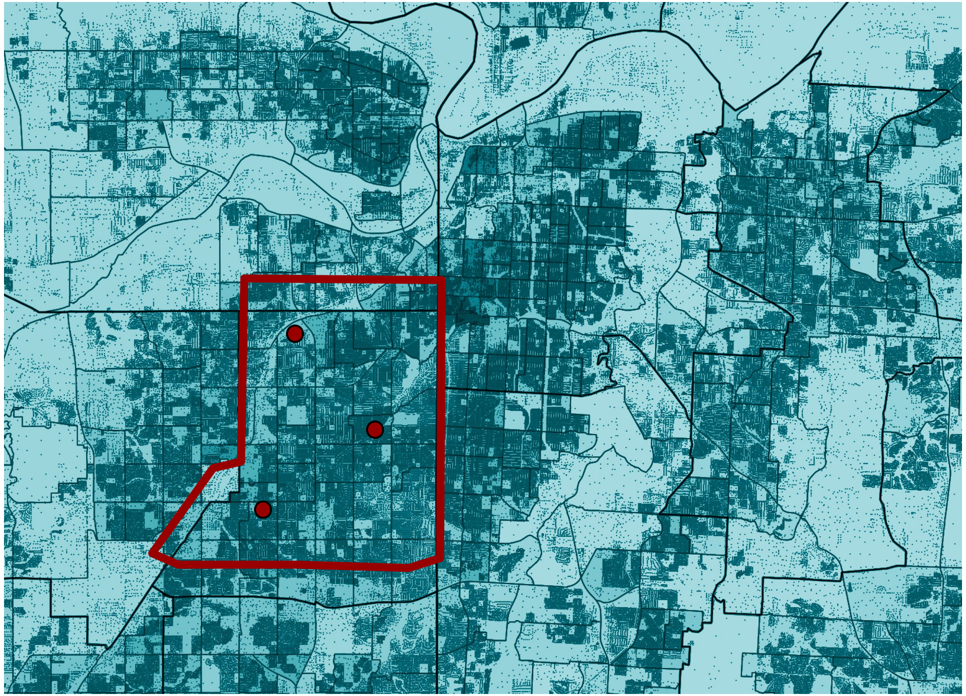
- Records of individual households and household members, 100s of attributes
- 1% of Public Use Micro Area population, 100K+ people

American Community Survey

- aggregates at Census Tracts level, censored at low values
- 10 to 50 Tracts per PUMA, 100s of attributes, 25% or 100% sample

Consumer Marketing Data

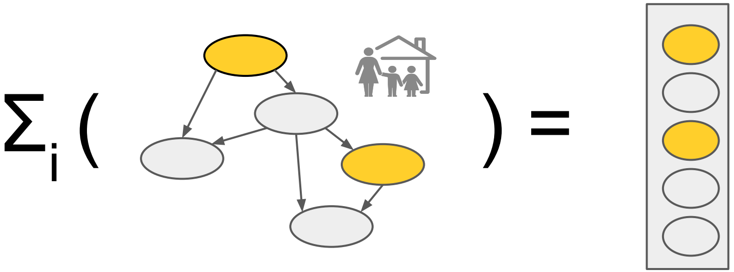
- EV ownership



Synthetic Population - Methods

Allocation with Convex Optimization

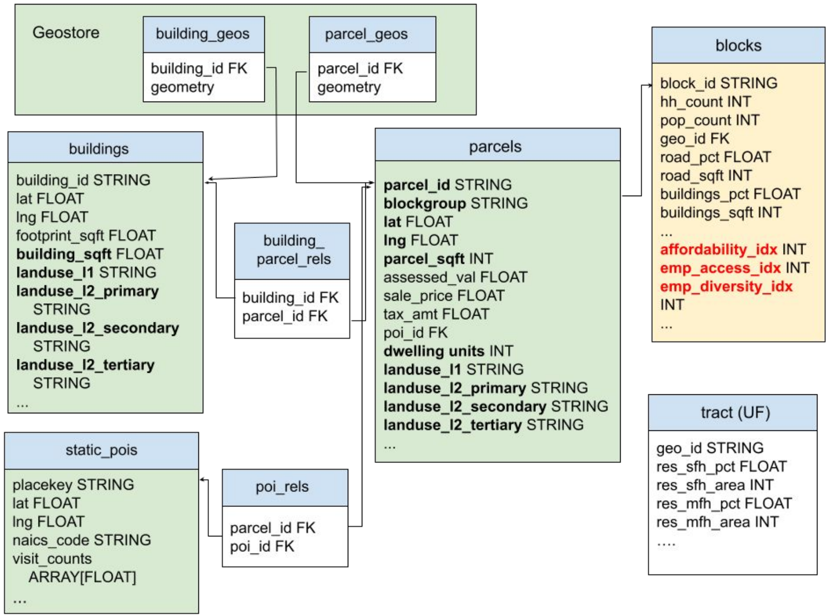
- Allocate N households composed of M people into T tracts so that the attributes match Census totals
- Census expansion weights vs fit to marginals trade-off formulated as a constrained convex optimization problem

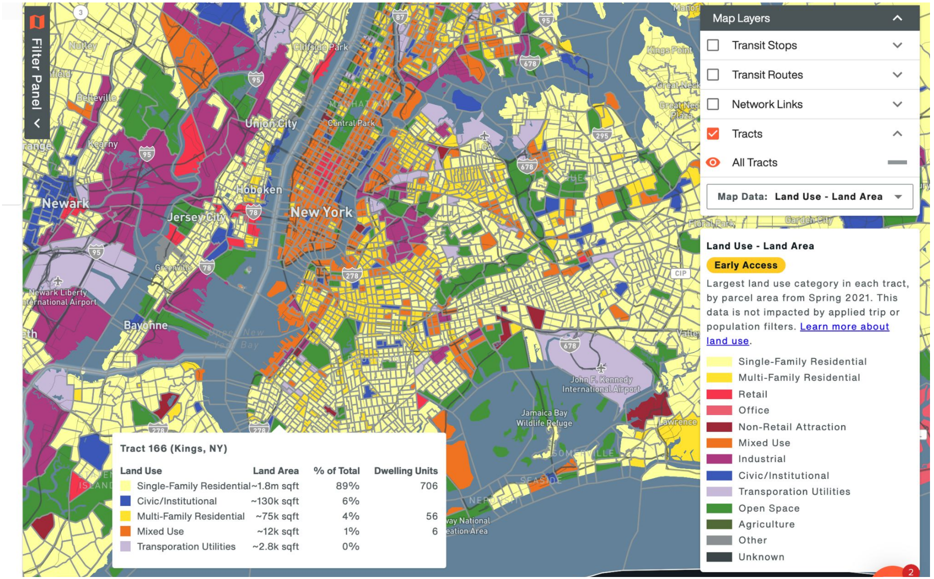


- Within each tract, allocate households to housing units matching block-level household totals
- Assign work and school locations, using iterative proportional fitting from a combination of CTPP, LEHD, and LBS data sources

Land use

Nationwide parcel-level land use with buildings footprints and POIs

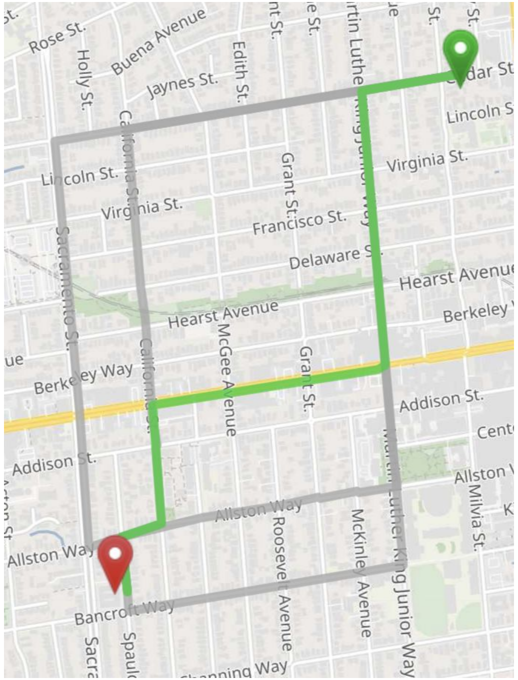


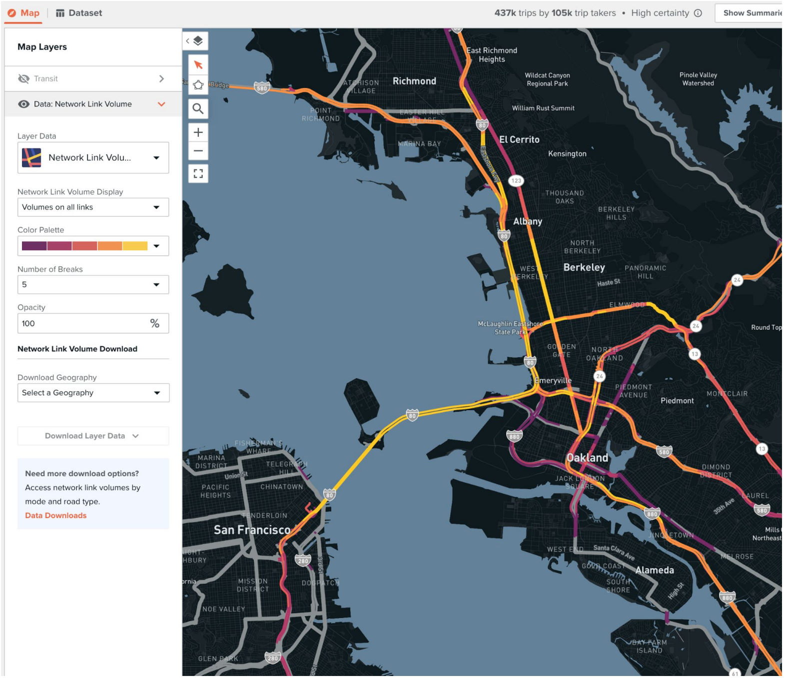


Transportation networks

Nationwide, OSM + GTFS

- **Multimodal nationwide** router (auto, freight, transit, walking, biking)
- **Time-dependent** routing to account for congestion; multiple profiles for local / long distance trips
- Places build takes (330M ppl) x (4 modes) x (4 trips a day) = **5B requests per release**
- Nationwide transit routing, GTFS data repo





Persona Trainer

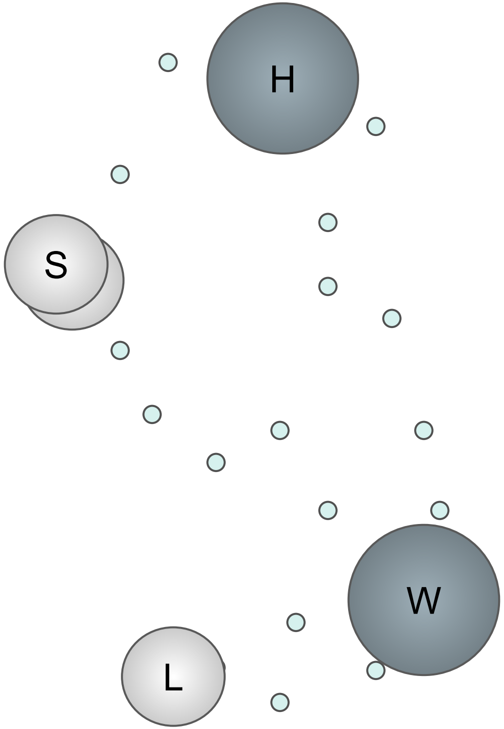
LBS Data Processing and Modeling

- Locational data analysis
  - Typically, 3 weeks up to 3 month of data per device
  - Mobility traces segmentation into staypoints and trips
  - Attributes of visited places and travel choices context
- Modeling: interpretable and policy responsive models
  - Day structure and activity sequence model (ASM)
  - Location choice model (LCM)
  - Mode choice model (MCM)

Persona Trainer

LBS Data Processing and Modeling

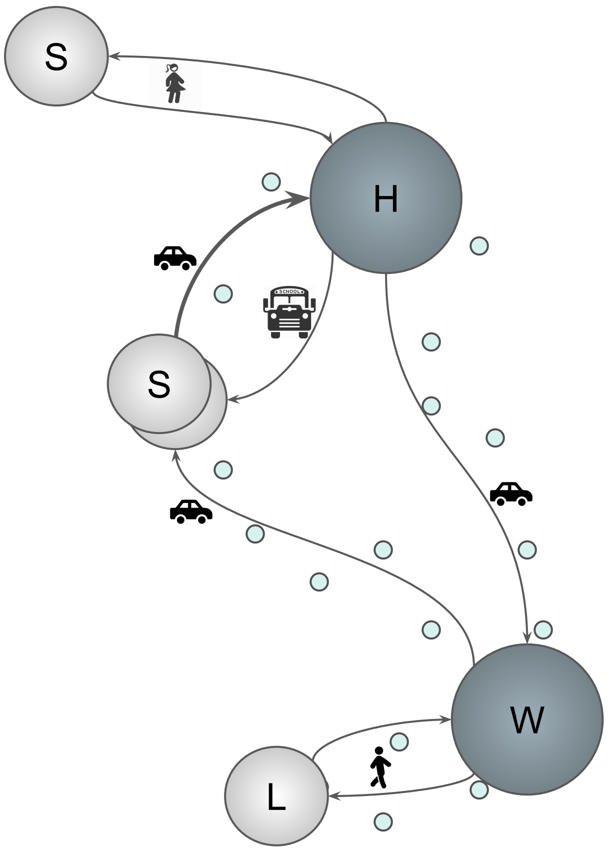
- Locational data analysis
  - Typically, 3 weeks up to 3 month of data per device
  - Mobility traces segmentation into staypoints and trips
  - Attributes of visited places and travel choices context
- Modeling: interpretable and policy responsive models
  - Day structure and activity sequence model (ASM)
  - Location choice model (LCM)
  - Mode choice model (MCM)



Persona Trainer

LBS Data Processing and Modeling

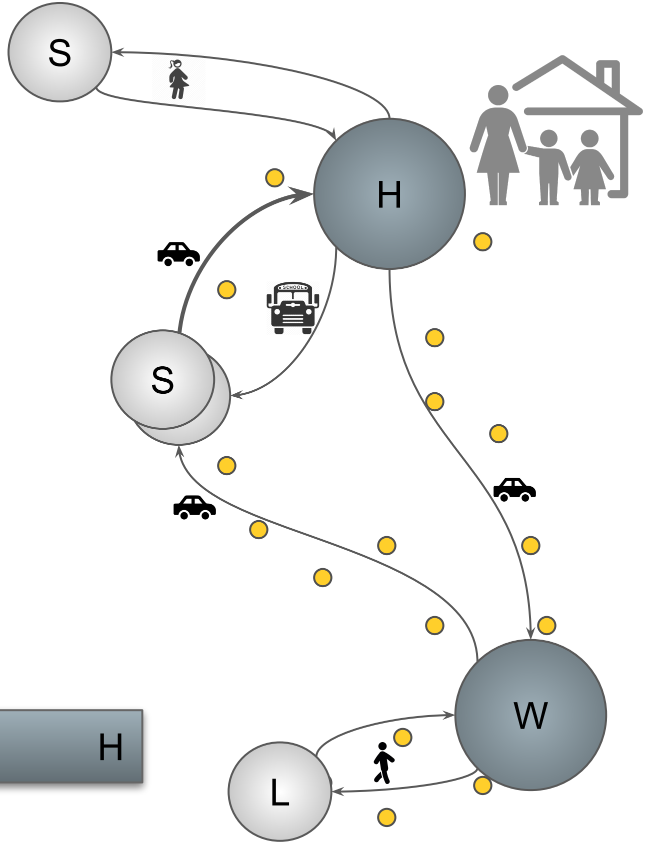
- Locational data analysis
    - Typically, 3 weeks up to 3 month of data per device
    - Mobility traces segmentation into staypoints and trips
    - Attributes of visited places and travel choices context
- Modeling: interpretable and policy responsive models
    - Day structure and activity sequence model (ASM)
    - Location choice model (LCM)
    - Mode choice model (MCM)

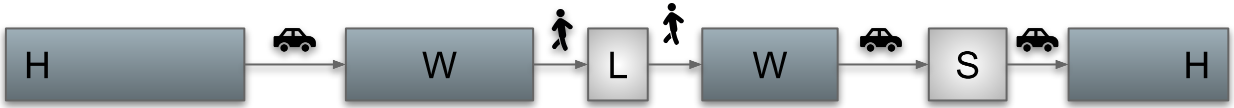


Persona Trainer

LBS Data Processing and Modeling

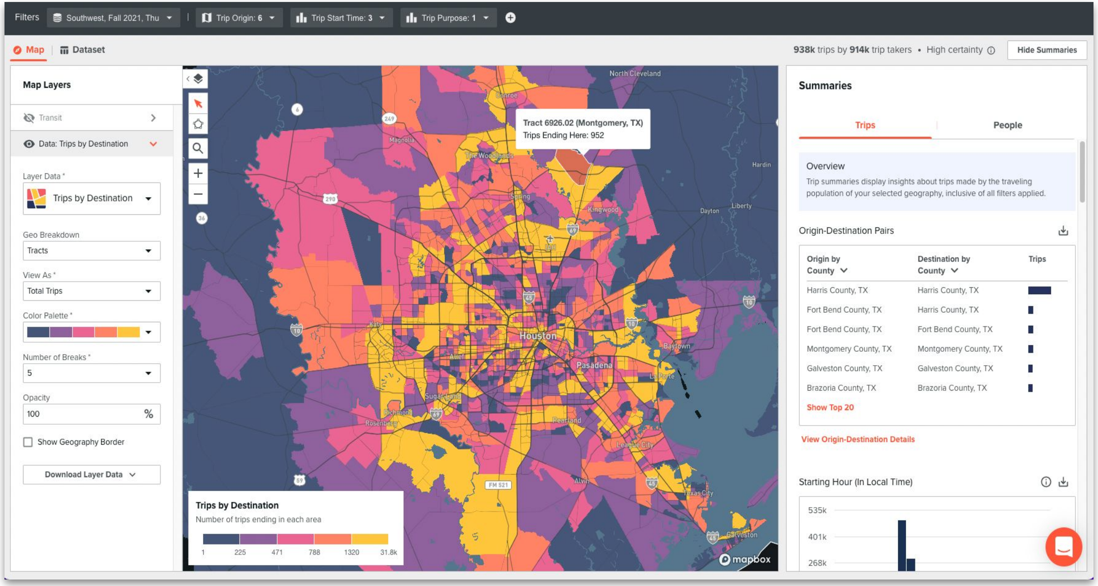
- Locational data analysis
  - Typically, 3 weeks up to 3 month of data per device
  - Mobility traces segmentation into staypoints and trips
  - Attributes of visited places and travel choices context
- Modeling: interpretable and policy responsive models
  - Day structure and activity sequence model (ASM)
  - Location choice model (LCM)
  - Mode choice model (MCM)





Places

High-fidelity **activity-based travel models**, representing specific regions during specific seasons, with data outputs down to the network link level.



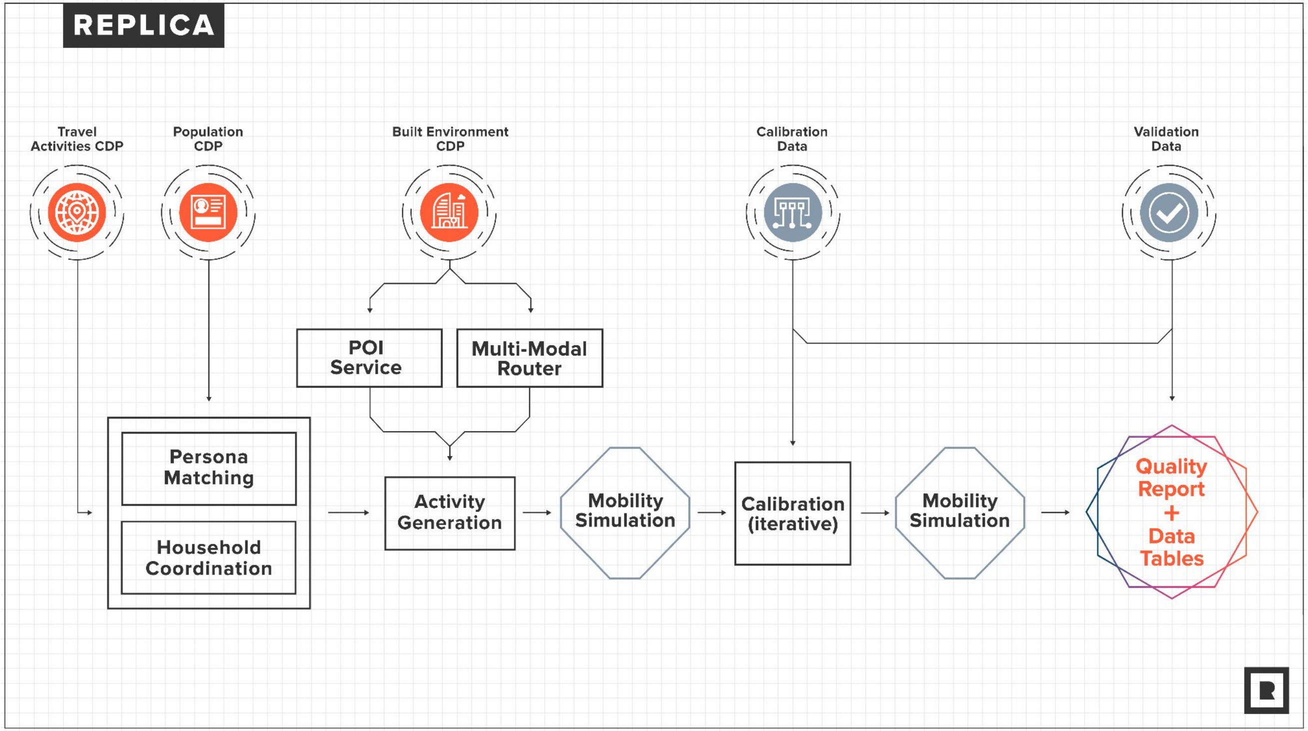
Showing first 100 of all 937,826 rows

Activity ID	Trip Origin Block Group	Trip Origin Tract	Trip Origin County	Trip Origin State	Trip Destination Block Group	Trip Destination Tract	Trip Destination County	Trip Destination State	Primary Mode	Trip Purpose
8673634482875025000	2 (Tract 3402.01, Harris, TX)	3402.01 (Harris, TX)	Harris County, TX	Texas	2 (Tract 3332.02, Harris, TX)	3332.02 (Harris, TX)	Harris County, TX	Texas	private_auto	work
12320064160658230000	3 (Tract 3422, Harris, TX)	3422 (Harris, TX)	Harris County, TX	Texas	1 (Tract 3506.01, Harris, TX)	3506.01 (Harris, TX)	Harris County, TX	Texas	private_auto	work
14828863641052529000	1 (Tract 3403.02, Harris, TX)	3403.02 (Harris, TX)	Harris County, TX	Texas	1 (Tract 3506.01, Harris, TX)	3506.01 (Harris, TX)	Harris County, TX	Texas	private_auto	work
3990376502905317000	2 (Tract 3421, Harris, TX)	3421 (Harris, TX)	Harris County, TX	Texas	4 (Tract 3502, Harris, TX)	3502 (Harris, TX)	Harris County, TX	Texas	private_auto	work
13724276556411600000	1 (Tract 3430, Harris, TX)	3430 (Harris, TX)	Harris County, TX	Texas	1 (Tract 3340.01, Harris, TX)	3340.01 (Harris, TX)	Harris County, TX	Texas	carpool	work
1032375517898096000	1 (Tract 3402.02, Harris, TX)	3402.02 (Harris, TX)	Harris County, TX	Texas	1 (Tract 3333, Harris, TX)	3333 (Harris, TX)	Harris County, TX	Texas	private_auto	work

How it Works

Places
Typical Day

- Generative models of travel demand
  - Activity-based, individual level
  - Responsive to changes (travel delays, road closures, accessibility)
- Model of network dynamics
  - Traffic flow and congestion
  - Transit route choice
  - TNC supply
  - Walkability

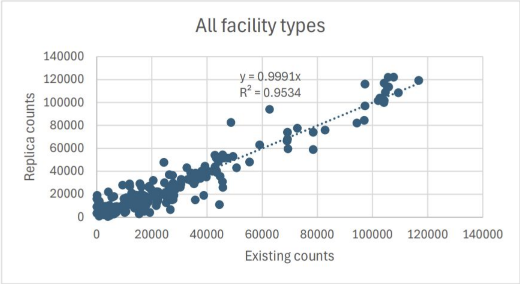


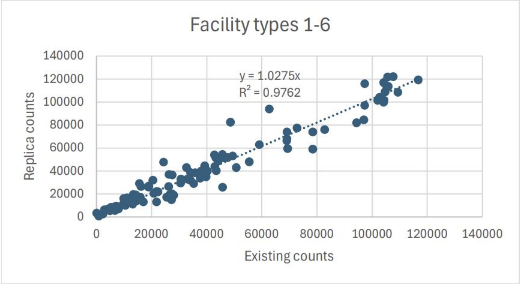
How it Works

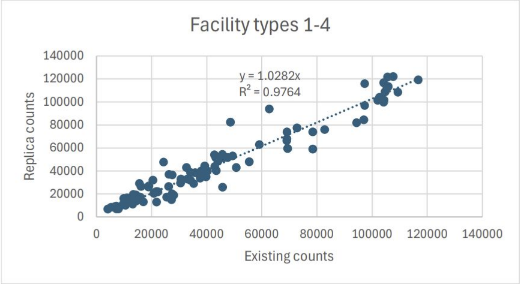
Places Typical Day

- Generative models of travel demand
  - Activity-based, individual level
  - Responsive to changes (travel delays, road closures, accessibility)

- Model of network dynamics
  - Traffic flow and congestion
  - Transit route choice
  - TNC supply
  - Walkability



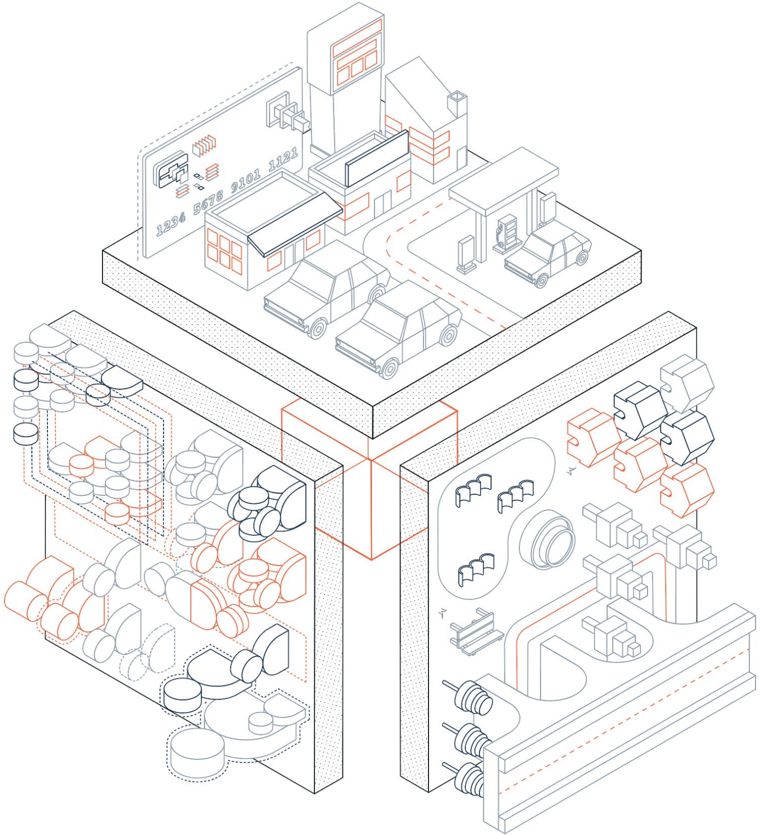




Scenario

Uses our data and activity-based model to forecast future conditions based on potential changes to the population, land use, and transportation infrastructure.

- The first release of Scenario can evaluate **Growth Scenarios** – changes to population and employment and the associated impact on travel and infra demand, with no infra changes
- **Custom Scenarios (available now)**
  A managed service where we re-run our ABM with custom inputs, assumptions, and geographies

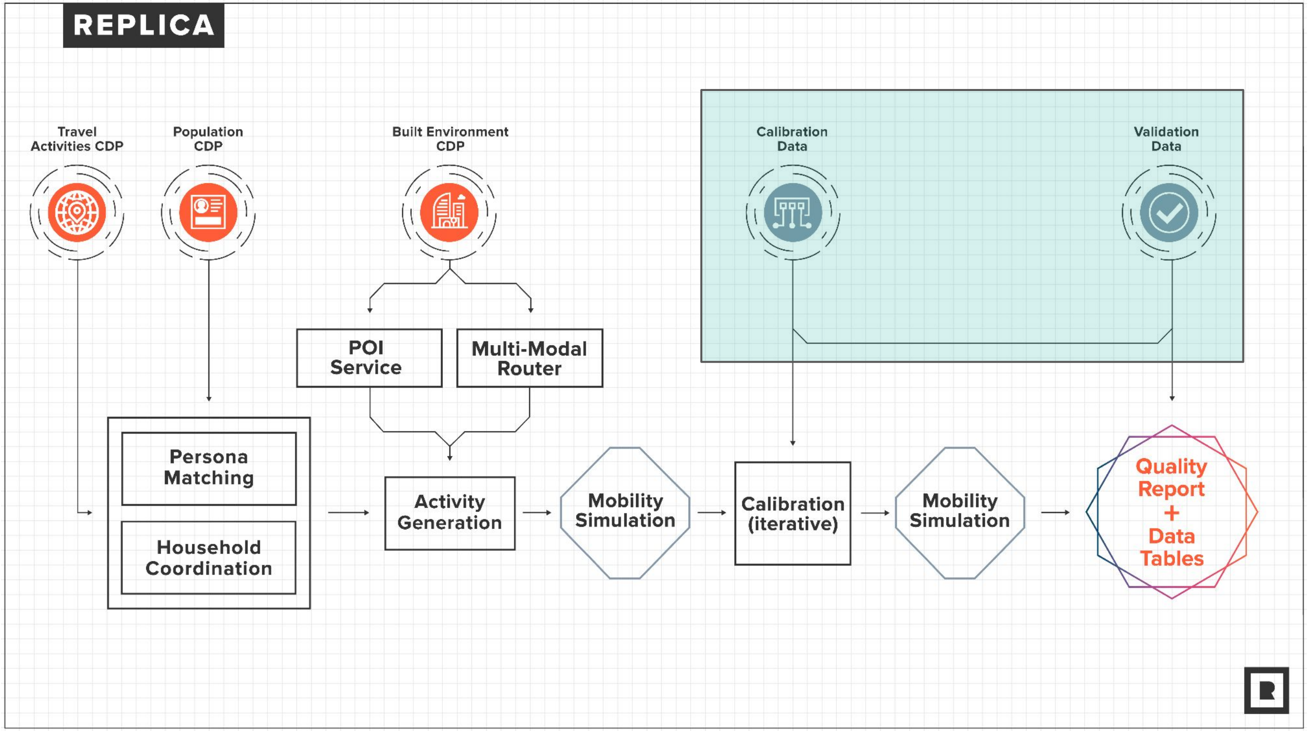


How it Works

Scenario

- Generative models of travel demand
  - Activity-based, individual level
  - Responsive to changes (travel delays, road closures, accessibility)

- Model of network dynamics
  - Traffic flow and congestion
  - Transit route choice
  - TNC supply
  - Walkability



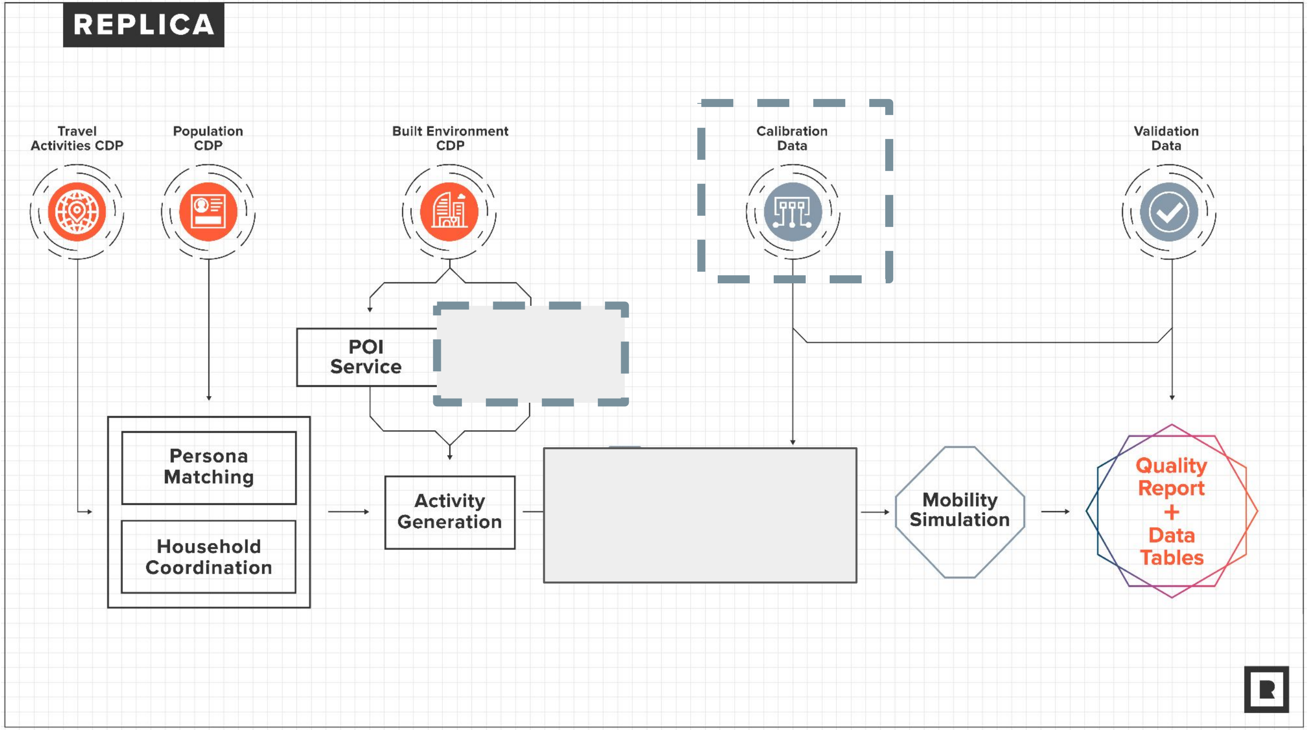
Trends

Nationwide activity-based model, with near-real time data at the census-tract level covering mobility, consumer spend, and land use.

How it Works

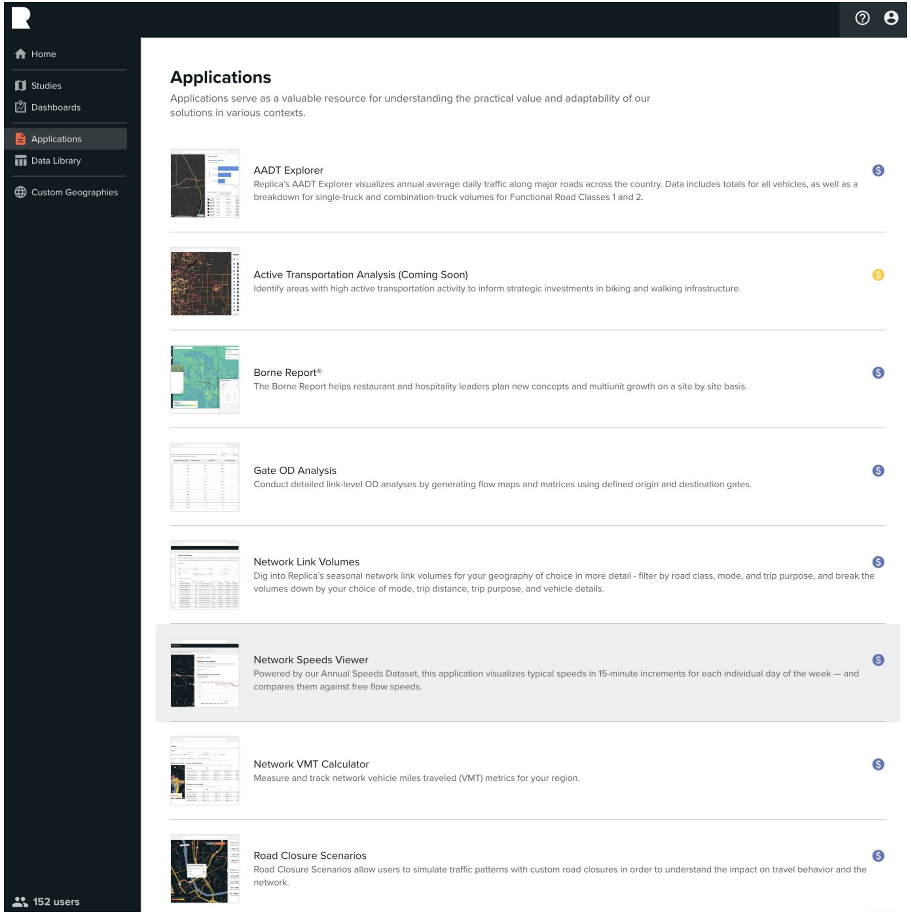
Trends

- Generative models of travel demand
    - Activity-based, individual level
    - Responsive to changes (travel delays, road closures, accessibility)
- Travel time / accessibility
    - Proxy traffic flow and congestion
    - Proxy transit accessibility
    - Walkability

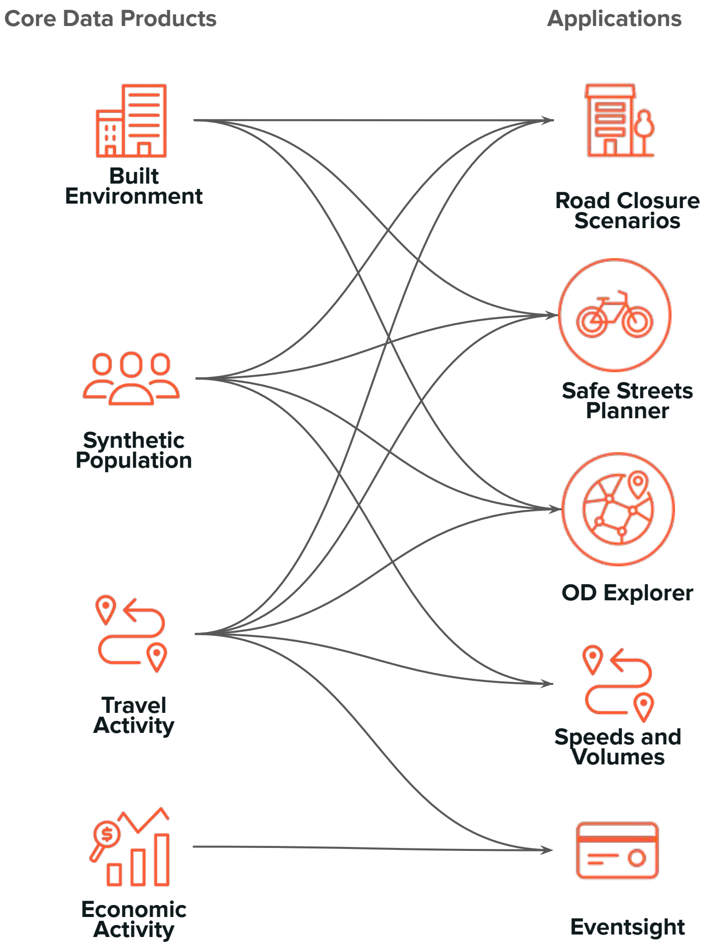


Applications

API access to all Core Data Products and derivative data (Places and Scenario) allows building targeted applications, both by Replica and 3rd parties



Data Sources + Data Flow



APPLICATIONS

Patterns

Measure the relative uptick (or drop-off) in popularity at specific locations on specific days and times.

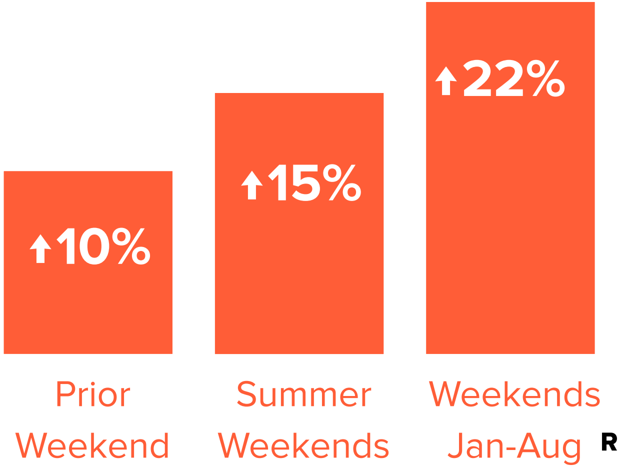
Create a custom geography, select the date and time of interest, and choose the comparison window.

The tool quantifies the relative change in activity in that location between the two periods and unlocks dozens of analysis:

- ROI on Major Investments
- Economic Impact + Spillover,
- Impacts of Work from Home
- Weather Impacts on Economic Activity
- YoY and MoM Comparisons
- Traffic, Tourism, Public Safety Management

Highlighted area is the downtown zone for analysis.

Uptick in Visits As Compared To:



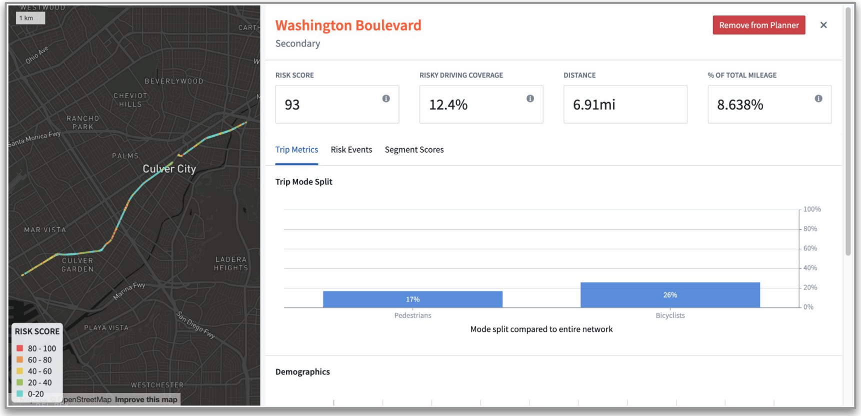
SAFE STREETS PLANNER

Culver City

Replica's detailed multimodal data with driving event data, enables agencies to map existing conditions, analyze specific corridors, generate their own High Conflict Corridor and Safety Action Plans and Reports.

Rather than investing in a one-off analysis or third-party report that will quickly become out of date as conditions on the ground change, Replica's Safe Streets Planner updates each season with the most recent data tied to specific road segments, including:

- Rapid acceleration events
- Hard braking / near miss events
- Speeding events
- Collisions
- Phone handling



Analyze Individual Corridors

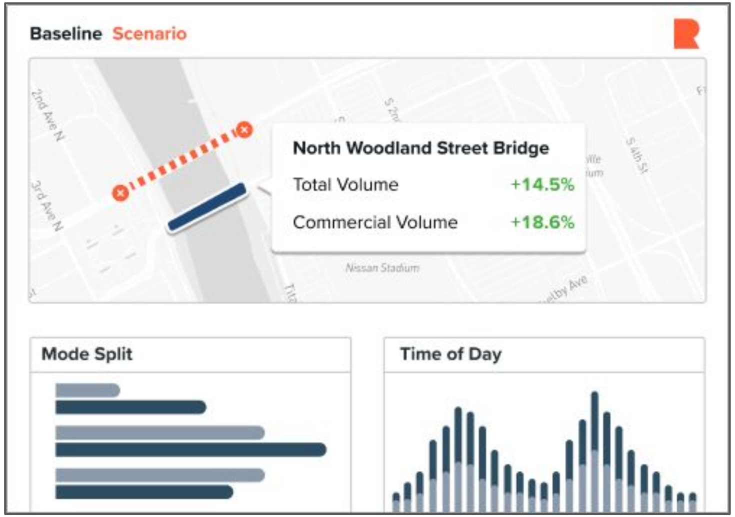
High Conflict Corridor Map

ROAD CLOSURE SCENARIOS

Large-Scale and Small-Scale Event Planning	Major Infrastructure Updates	Safety and Construction Planning	Emergency Planning and Preparedness

What this Unlocks:

- Identify Traffic Diversion Patterns: Find alternative routes with ease and understand congestion impacts.
- Estimate Impacts on Daily Travel: Quantify changes in travel volumes, trip distances, mode split, and more.
- Understand Socio-Economic Impacts: Identify who is most impacted by these closures and how.
- Assess Environmental Impact: Understand changes in emissions and vehicle miles traveled (VMT).



Key Features:

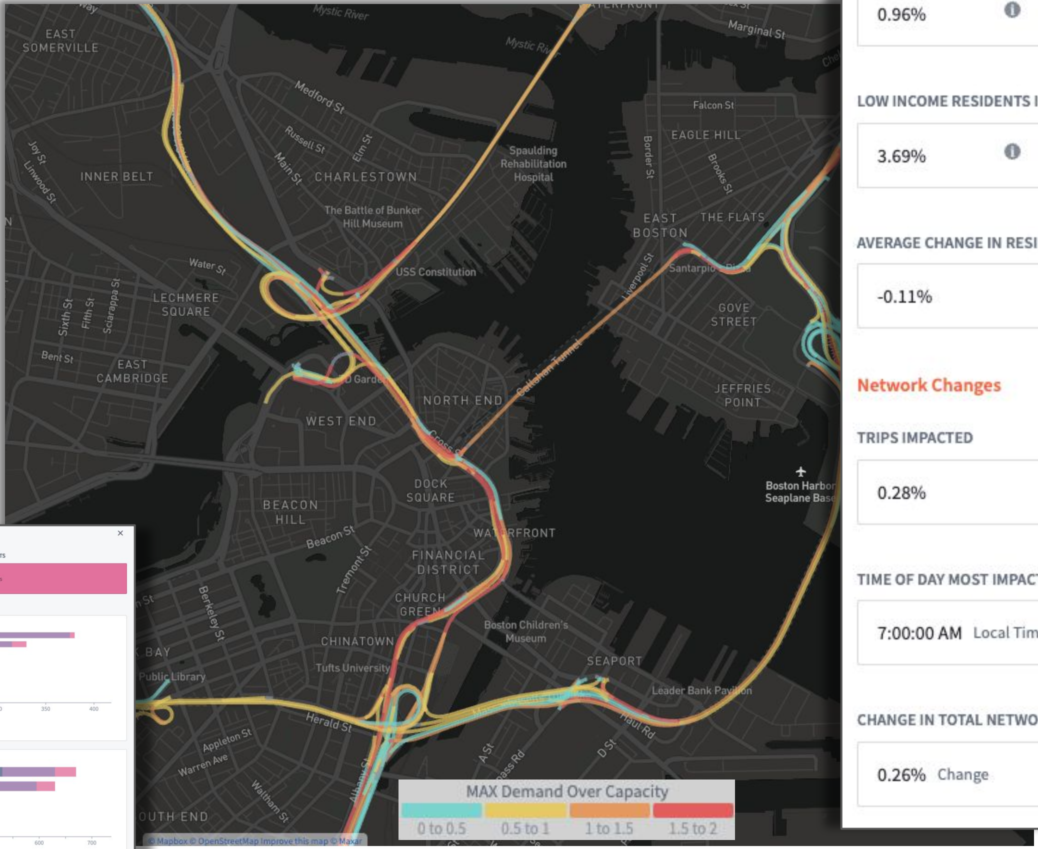
- Interactive Map-Based Tooling: Easily designate roads for closure, and visualize traffic diversion patterns and changes in travel time for affected routes.
- Comprehensive Analysis: Dynamic charts and visualizations showing forecasted impacts.
- Collaborate and Share: Share insights with your colleagues for review and collaboration.

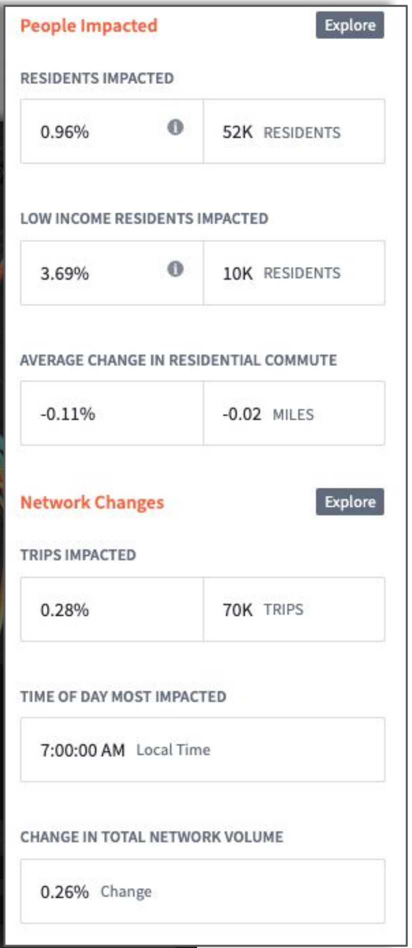
ROAD CLOSURE SCENARIO: MASSACHUSETTS

Sumner Tunnel Restoration

The Sumner Tunnel is undergoing a restoration that began in the spring of 2022. This work requires the tunnel to be closed to traffic periodically.

MassDOT performed a Road Closure Scenario to quantify traffic diversions and to understand people impacted.





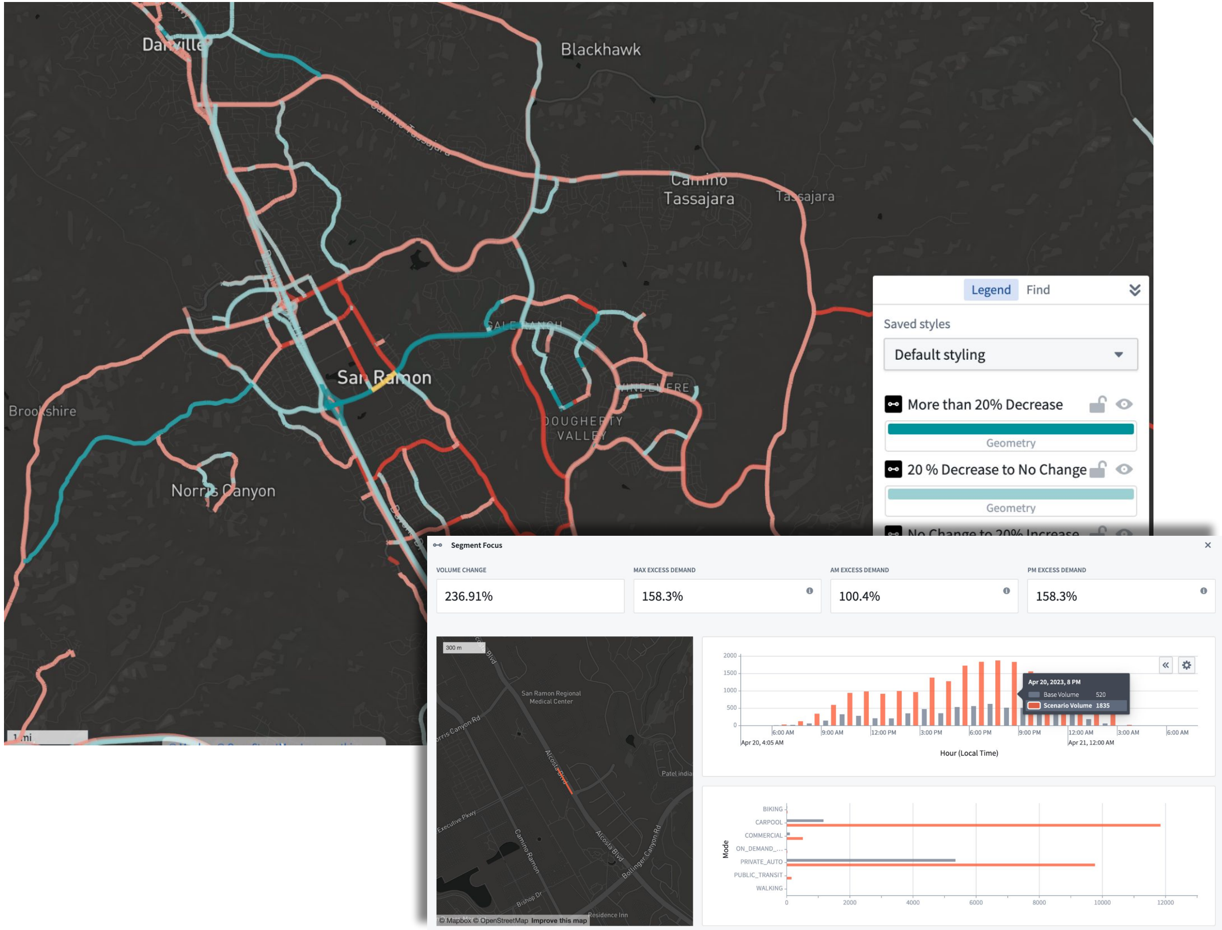
massDOT
Massachusetts Department of Transportation
REPLICA

ROAD CLOSURE SCENARIO

San Ramon Bollinger Canyon

In order to install a bike and pedestrian overcrossing, the City of San Ramon needed to stop traffic on Bollinger Canyon Road for nine days, forcing roughly 40,000 trips to be rerouted each day.

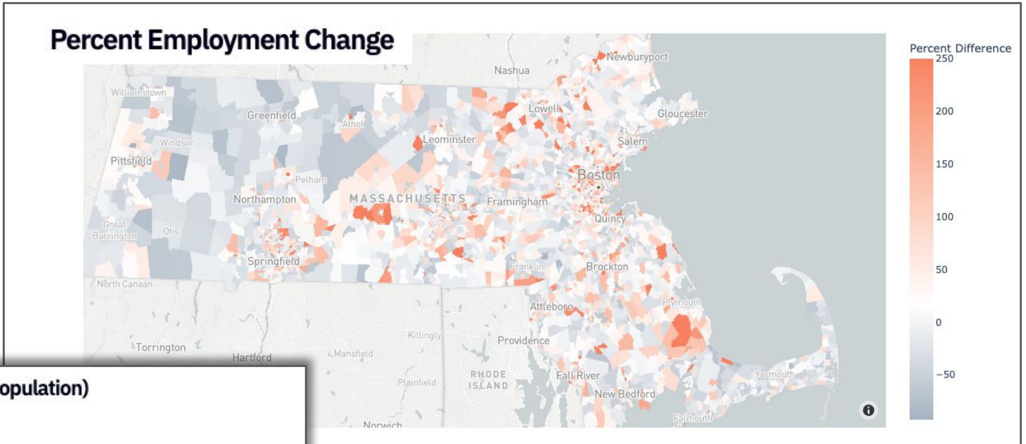
The findings of this Scenario run helped San Ramon create **a data-driven outreach plan** that centered around the most congested pain points and **informed a diversion plan** that includes consideration of a shortcut through a parking lot that Replica's router identified.



CUSTOM SCENARIOS

MassDOT

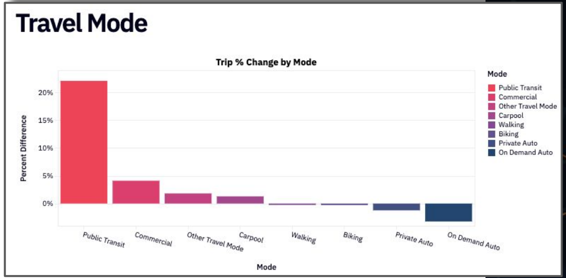
MassDOT and Boston Region MPO (CTPS) used Replica's Scenario product to model mobility and employment changes by their proprietary TAZs, to understand the future demand for transit, stressors to road network, and changes to VMT based on their population growth estimates.

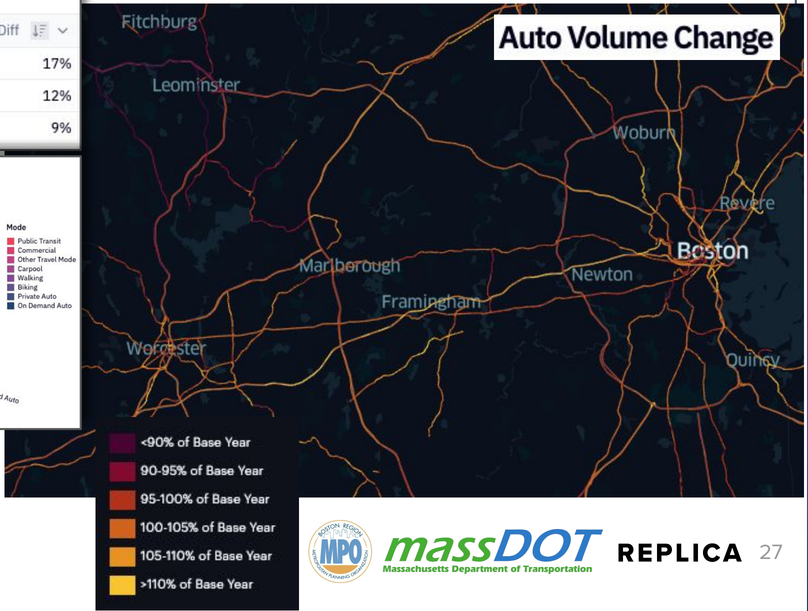


VMT Metrics by City (25k+ Population)

Total VMT

City	Base	Build	Diff	% Diff
Woburn, MA	745.4K	895.3K	149.9K	17%
Methuen Town, MA	1.3M	1.4M	177.2K	12%
Marlborough, MA	930.6K	1.0M	94.5K	9%

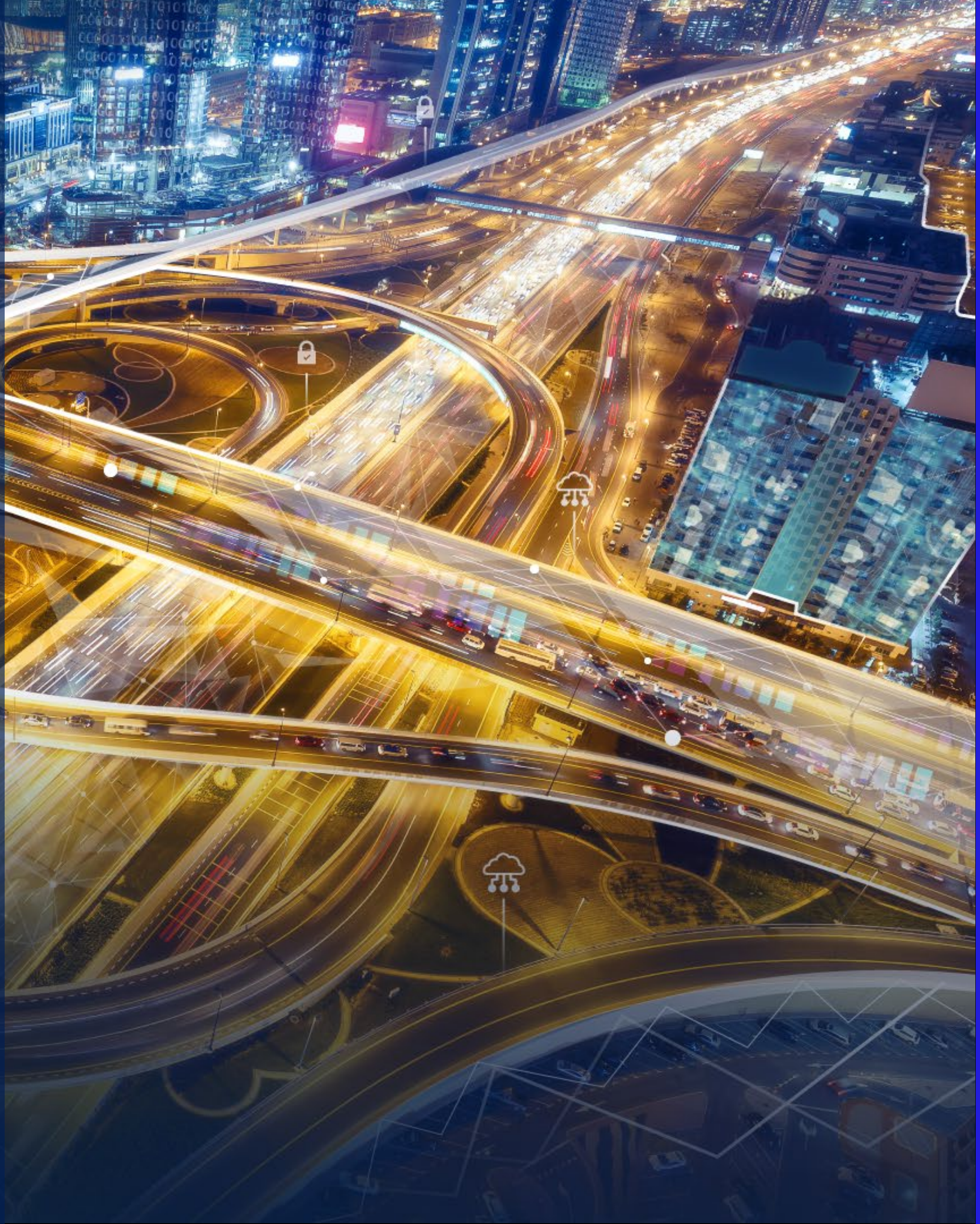




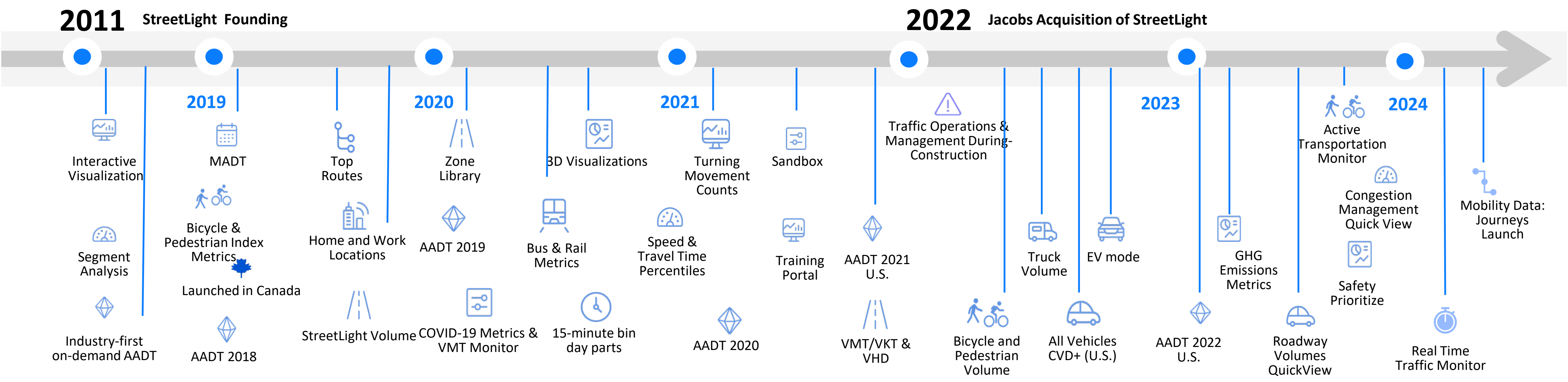
MPO massDOT Massachusetts Department of Transportation REPLICA

STREETLIGHT Mobility Data Solutions

Connected Vehicle Data



StreetLight Data – Big Data to Make Transportation Better



1M+
PROJECTS AND COUNTING

7+
BILLION MONTHLY TRIPS PROCESSED

100s
OF DATA SOURCES INCORPORATED

>5PB
IN OUR HISTORY WE'VE PROCESSED OVER 5PB POINTS OF DATA

20,000+
POWERING ANALYSES PER MONTH

StreetLight's product suite tackles a wide array of transportation challenges with the most complete portfolio in the industry

PLANNING	OPERATIONS	CLIMATE	MOBILITY DATA	COMMERCIAL
Traffic & Intersections	Traffic Monitor	VMT Analytics	Connected Vehicle Journeys	Advanced Traffic Counts
Transportation Modeling	Real-Time Traffic Monitor	GHG Analytics	Connected Vehicle Hard Braking	Advanced Truck Counts
Transportation Planning	Operational Performance Analytics	Vehicle Electrification Analytics		Transportation Planning
Truck Traffic		Evacuation Analytics		Trucking & Logistics
Truck Modeling				Active Transportation
Freight Planning				Visitation & Site Selection Analytics
Active Transportation				Multi-Trip Analytics
Safety Prioritize				
Multi-Trip Analytics				

Mobility Data IS Big Data!

Snapshot of data points
averaged over a week

- **14.9 billion** points
- **10.2 million** vehicles *(expect ~10+ million when we look across the whole month)*
- **39 million** journeys
- Ping count every **3 seconds**

STREETLIGHT

14.9B

POINTS PER DAY

0.72B

VISA TRANSACTIONS PER DAY

VISA

8.5B

GOOGLE SEARCHES PER DAY

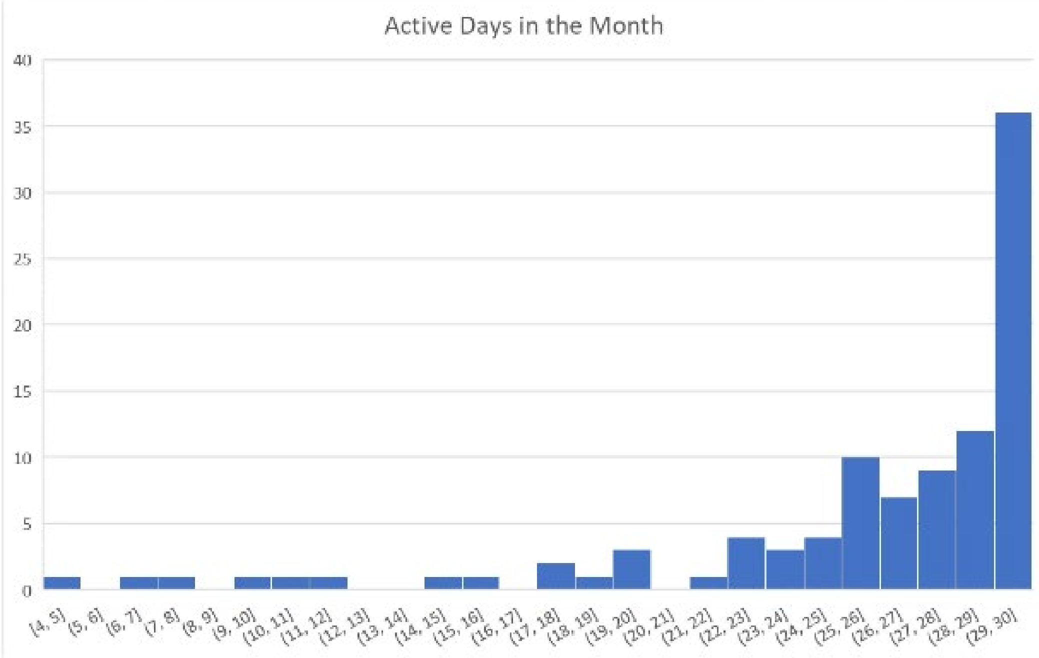
Google

Source: Google and Visa data points from 2023 - Worldwide

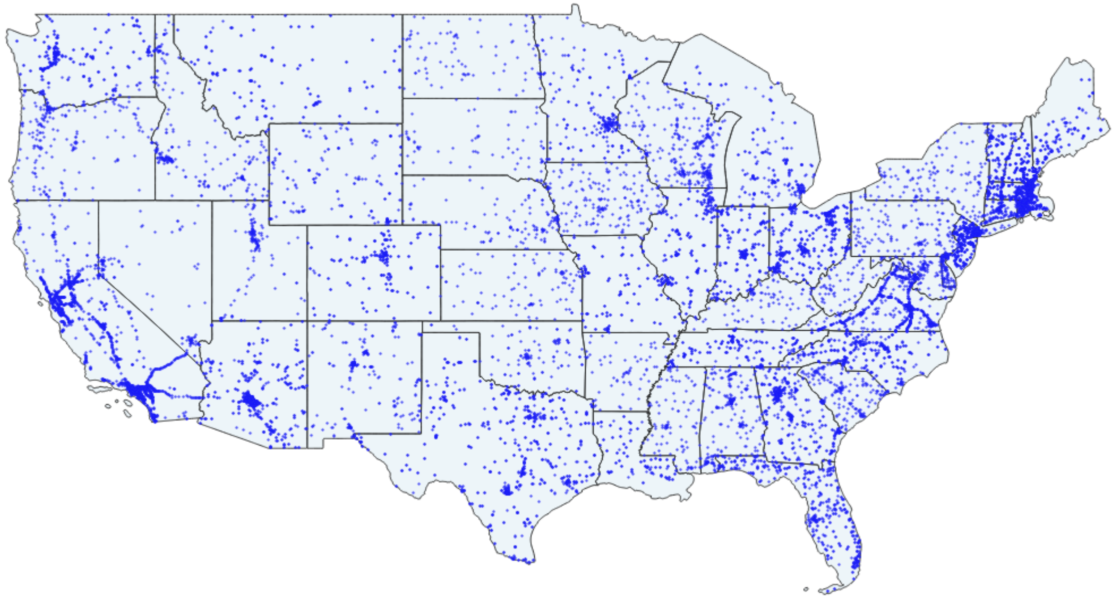
Fleet Statistics

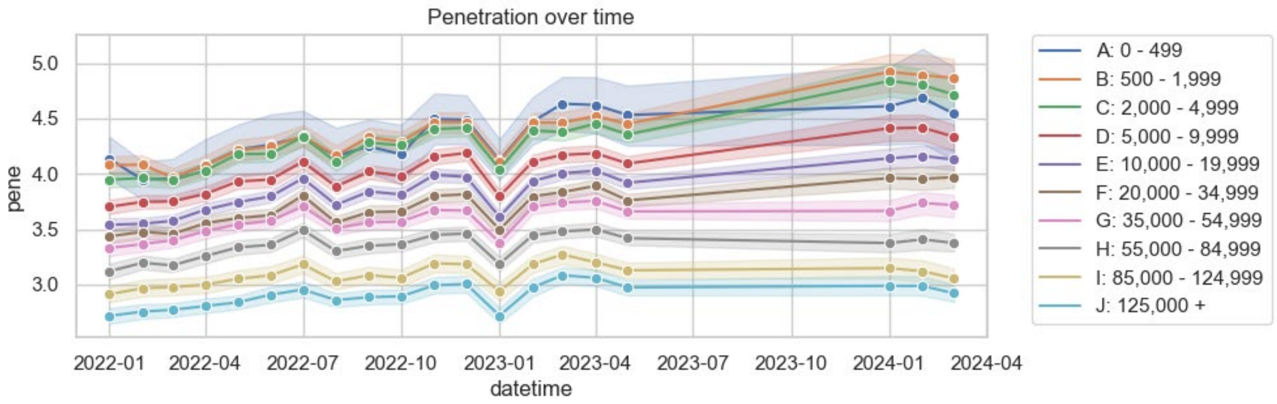
Nationwide:

- 10M+ unique vehicles sampled every 3secs
- Model year 2015 and later
- 55% SUV, 30% Pickup Truck, 15% Sedan/Saloon
- Most drive 25+ days/month

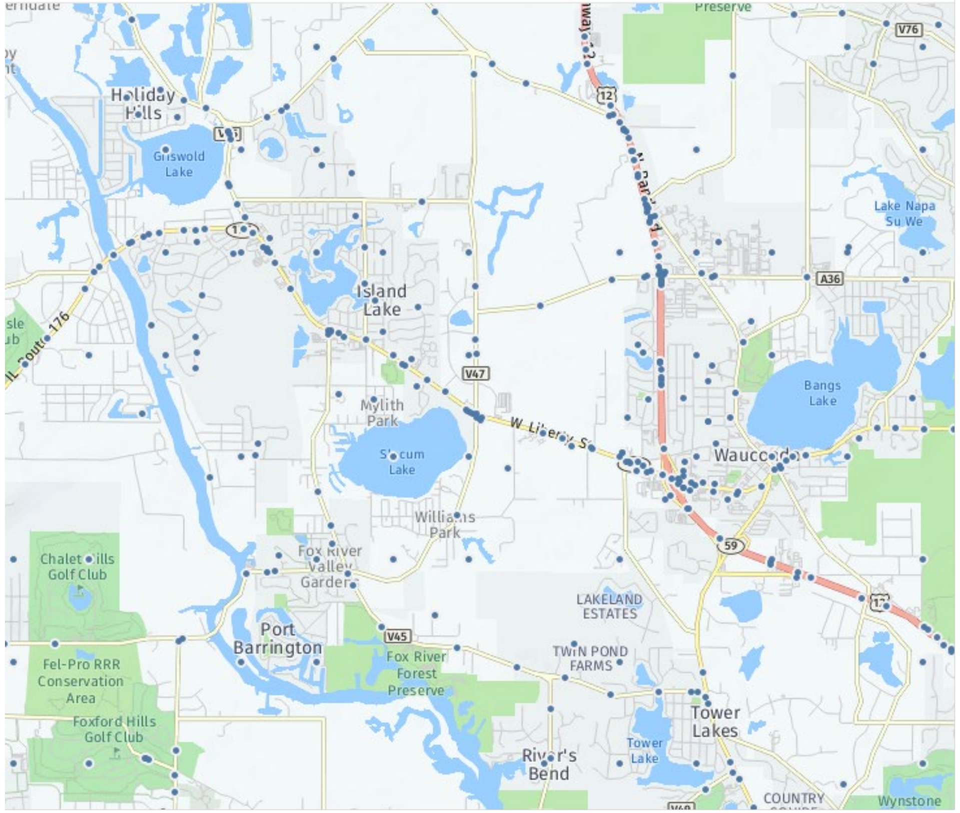


StreetLight's Counter Network (14,000)





Data privacy rules in place to protect StreetLight CVD



Location Blurring

Lat/Long coordinates will be truncated to two decimal points to blur trip points that fall within approximately 0.5 miles of a frequently visited location. Frequently visited locations are defined as the top 5 locations a vehicle visited in the past 30 days. This definition may be changed by a Third-Party Licensor at any time, and we will notify customers with reasonable notice, if possible.

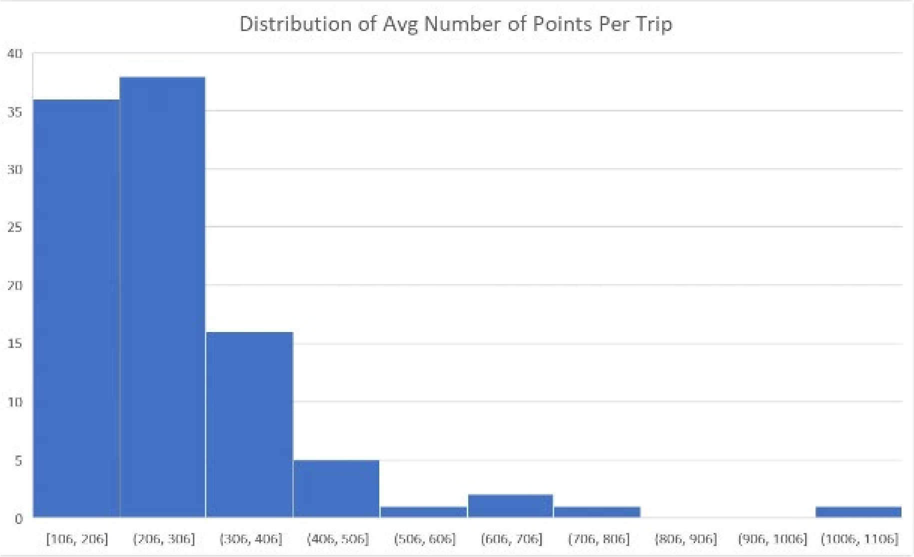
Low Density Filtering

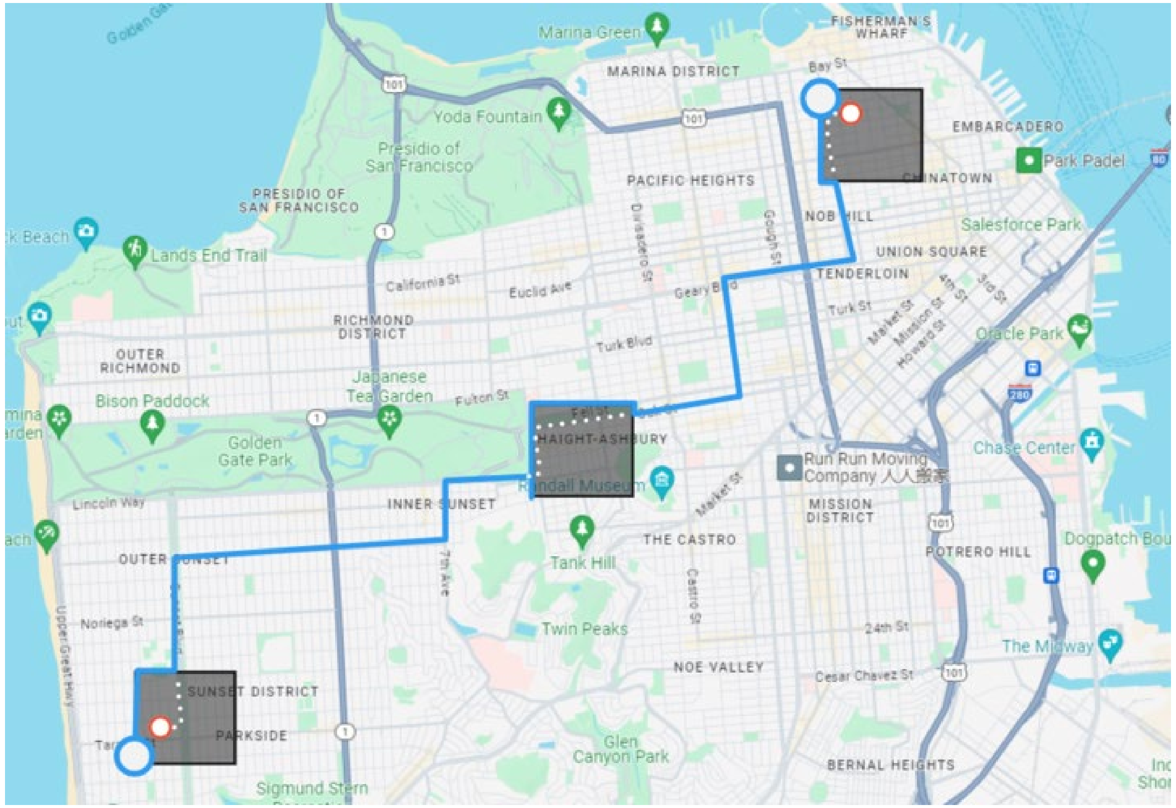
Data points representing locations in any individual square half mile area where the Data contains only one or two vehicles within that area during the most recent 3-month period.

Privacy Blurring - Statistics

Nationwide: ~15B points,~40M Journeys daily

- ~20% of all points are blurred
- ~23% of KeyON points blurred
- ~23% of KeyOFF points blurred
- ~14% of Journeys have both KeyON and KeyOFF blurred
- ~9% of journeys are completely blurred

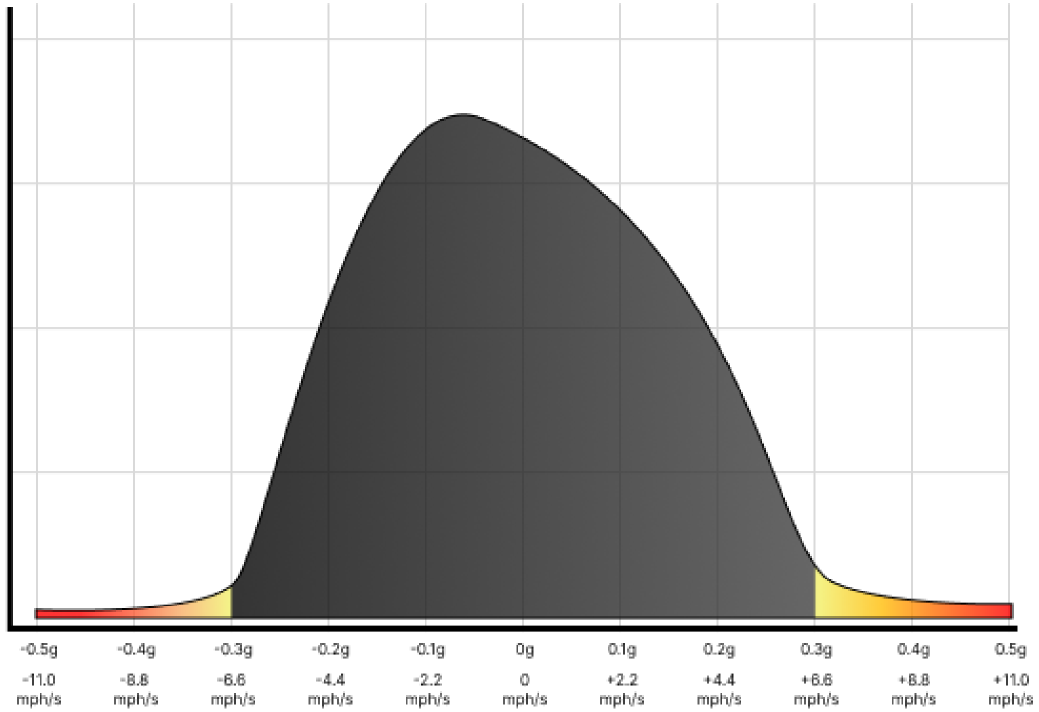




Most trips 100-300 points, assuming 3s interval, 5-15 mins

Hard Braking/Acceleration Data

- 10M+ unique vehicles, telemetry data every 3s
- Acceleration derived using vehicle-reported speeds
- 99.95% of pings are removed, leaving only hard braking + acceleration events
- gForce is kept as an attribute, allowing for exploration of different severity of acceleration behaviors



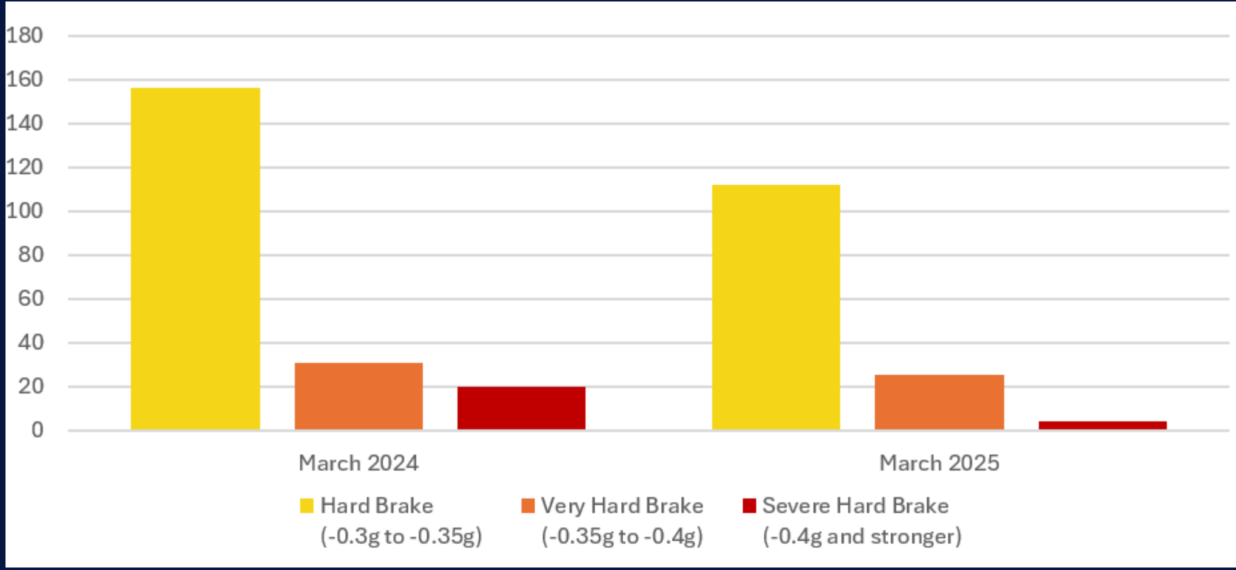
Established Thresholds:
Purdue (2024): -0.27g
Purdue (2023): -0.4g to -0.5g
Jun et al. (2007): -0.4g

Geotab (2023): -0.27g
Wejo (2022): -0.32g
(Mousavi, 2015): -.37g
(Kamla et al., 2019): -0.2g to –0.8g

EXAMPLE BEFORE & AFTER STUDY

STREETLIGHT

Safer Turns in the Roundabout Capital of America: Measuring Impact with Hard Braking Data



CHALLENGE

Carmel, Indiana aimed to improve safety at the busy three-way intersection of 106th Street and Hazel Dell Parkway.

With high traffic volumes on both roads, the city needed a way to reduce risky driving behavior.

DATA-BASED SOLUTION

StreetLight analyzed hard braking events before and after the intersection's conversion to a multi-lane roundabout.

Using a high magnitude of precise locations for hard braking & acceleration events data, the study compared March 2024 to March 2025 to evaluate changes in driver behavior and quantify the safety impact of the new design.

AGENCY & PUBLIC BENEFIT

The analysis revealed a 32% drop in total hard braking events, with the sharpest decline in severe incidents at 80%.

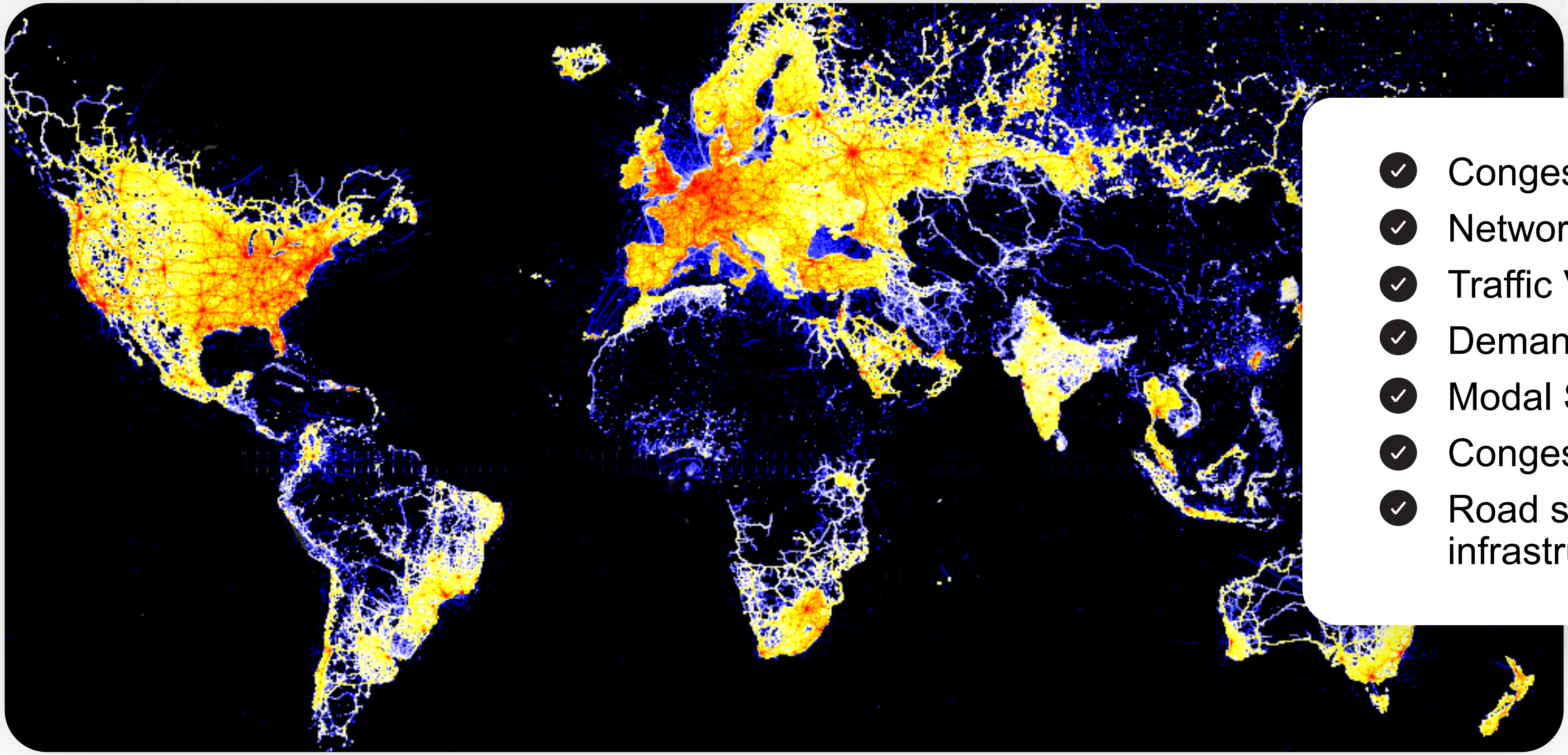
This immediate, measurable insight can help validate the roundabout’s effectiveness, support future design decisions, and demonstrate the power of using braking data as a rapid safety indicator.

Congestion and traffic demand Analysis using Floating-car Data

Ralf-Peter Schaefer / VP Product Management

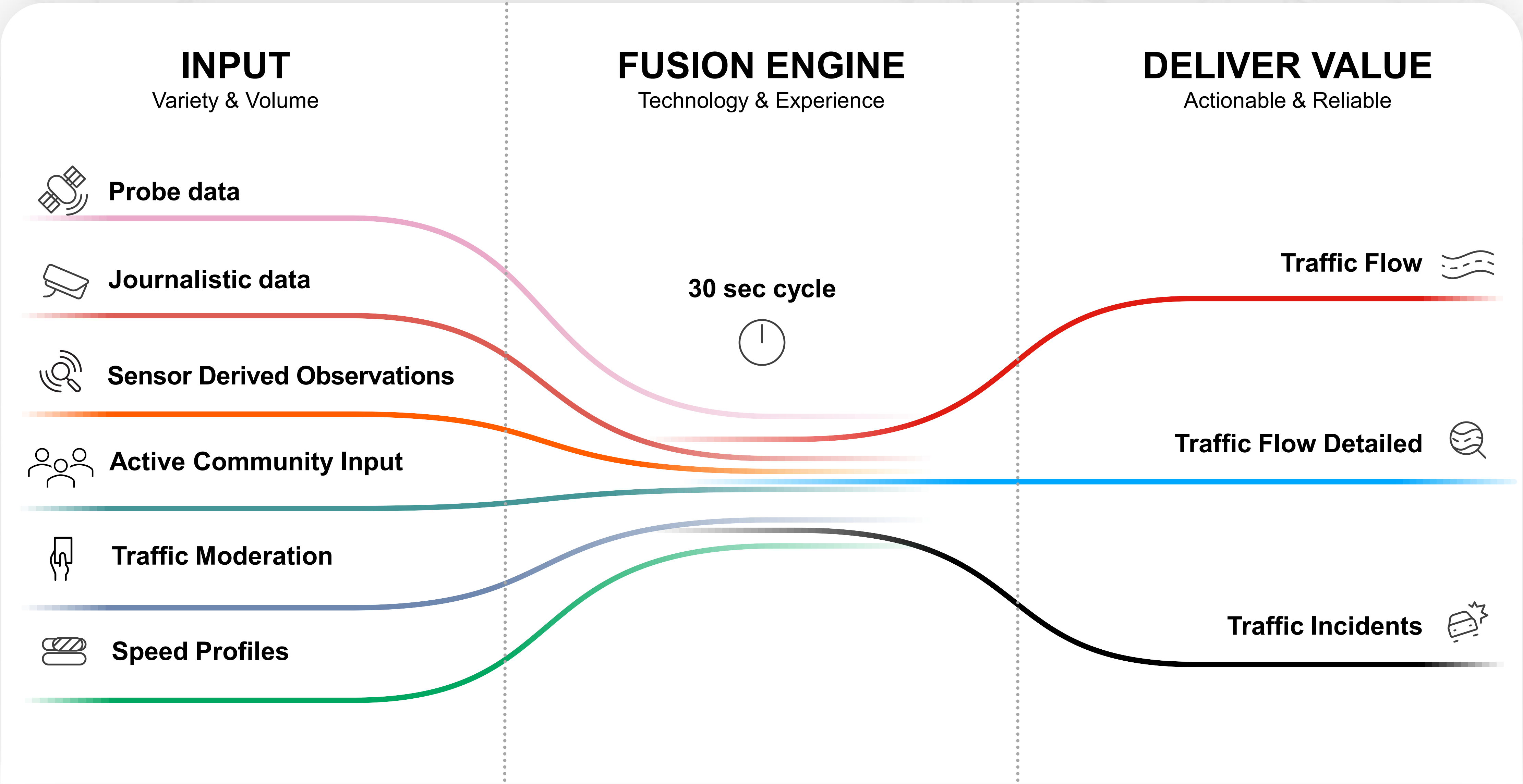
City and DOT Challenges

Understanding the Challenge using Big Data



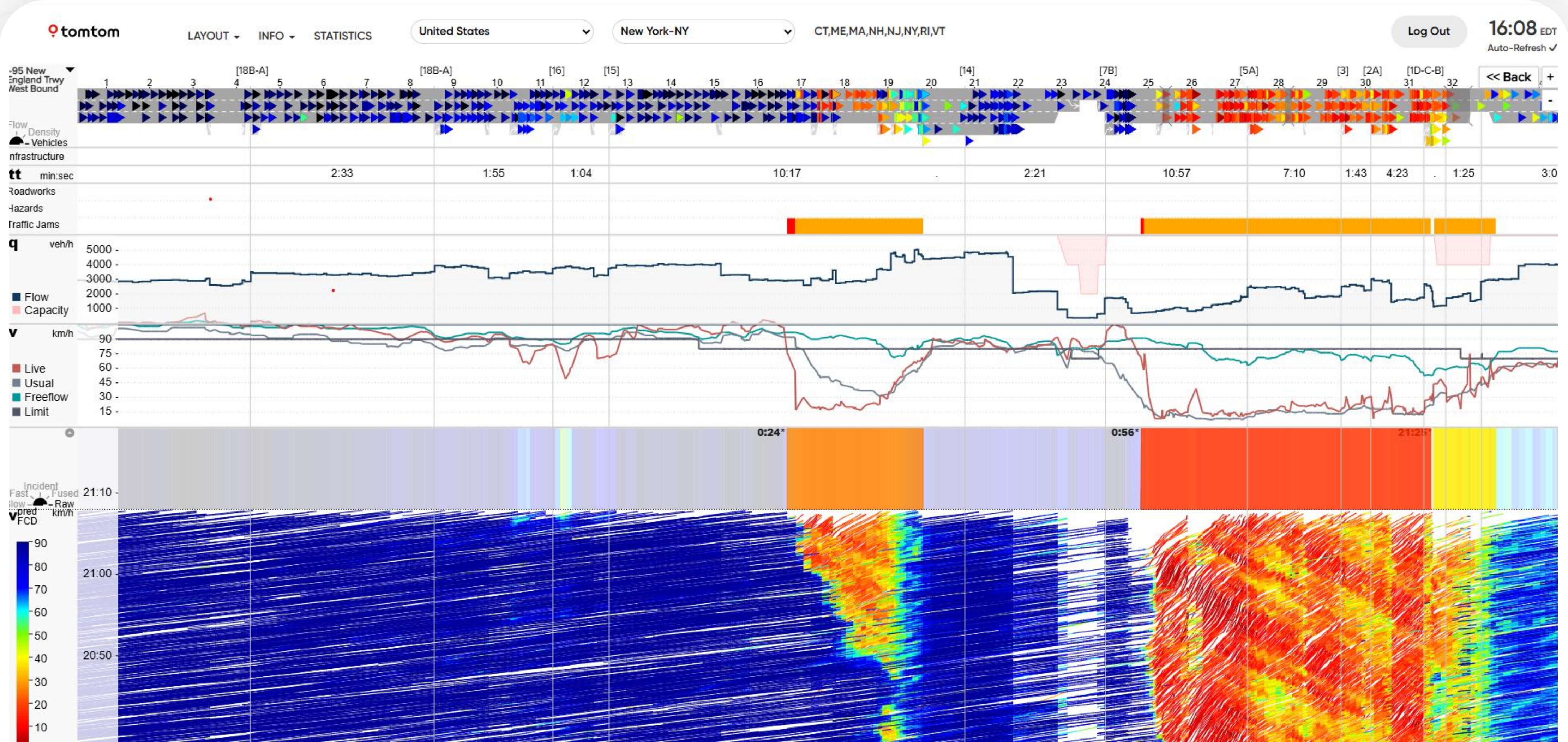
- Congestion
- Network Capacity
- Traffic Volume
- Demand
- Modal Split
- Congestion root cause
- Road safety and infrastructure risks

How we build live traffic



Live Traffic Manhattan

I-95 WB 19 March 2025 4:08pm



TomTom MOVE Data Products

A product suite to build global mobility metrics

Traffic Stats
Historical Traffic

- Speeds per segment
- Average speeds
- Average travel-time
- GPS probe counts
- Per segment

O/D Analysis
Historical Traffic

- Trip based analysis
- Origins, destinations, via-points
- Measure movement between areas, or through selected road connections

Route Monitoring
Real-Time Traffic

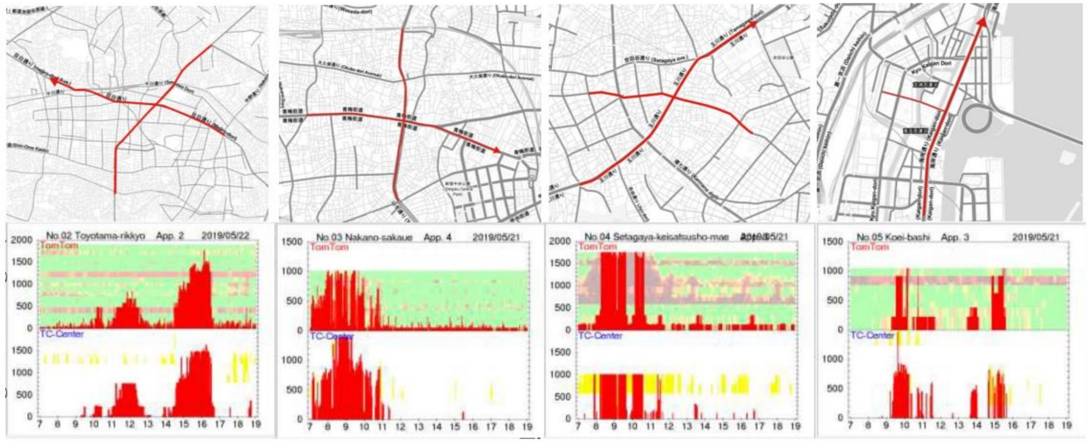
- 24/7 measurements of pre-defined routes
- Real-time speeds and travel-time
- Per route
- Per segment

Junction Analytics
Real-Time Traffic

- Analyze queue length, delay, travel time
- Optimize traffic lights
- Optimize intersections in real-time

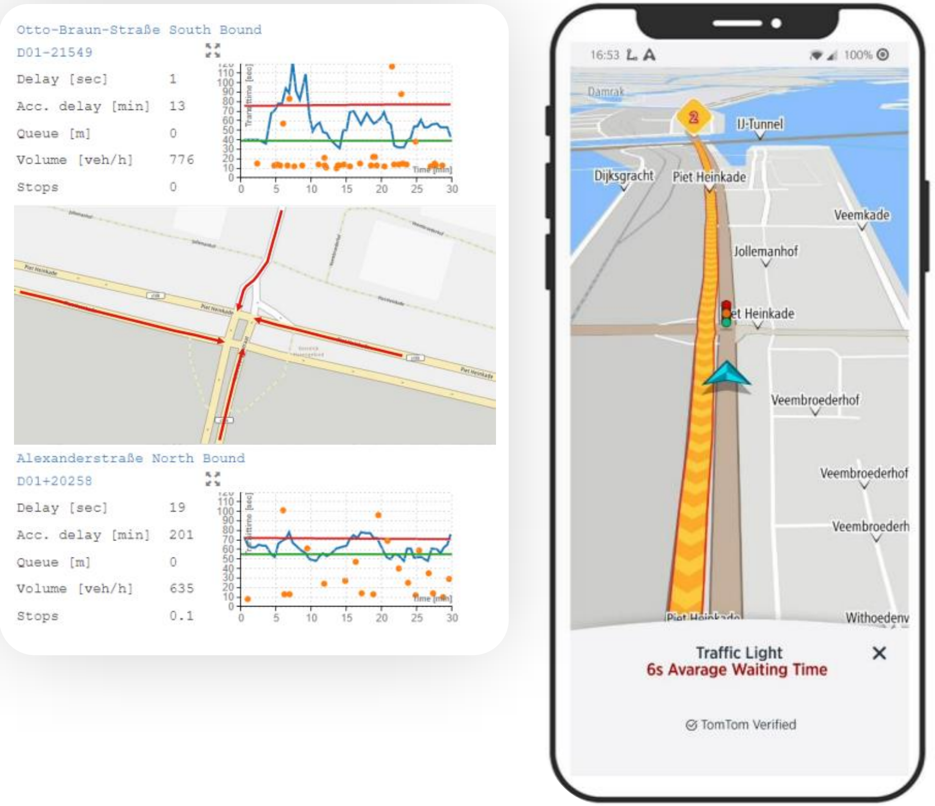
Closing the Loop

Signal Time Optimization
and Time2Green Service



- Comparison of count measurements vs. TomTom Junction Analytics probe measurements
- Ground truth from 4 Junctions in Tokyo, Japan

Speed Advice Service using Time2Green Data



TomTom Traffic Index Results 2024

Annual congestion level comparison of 500 global cities

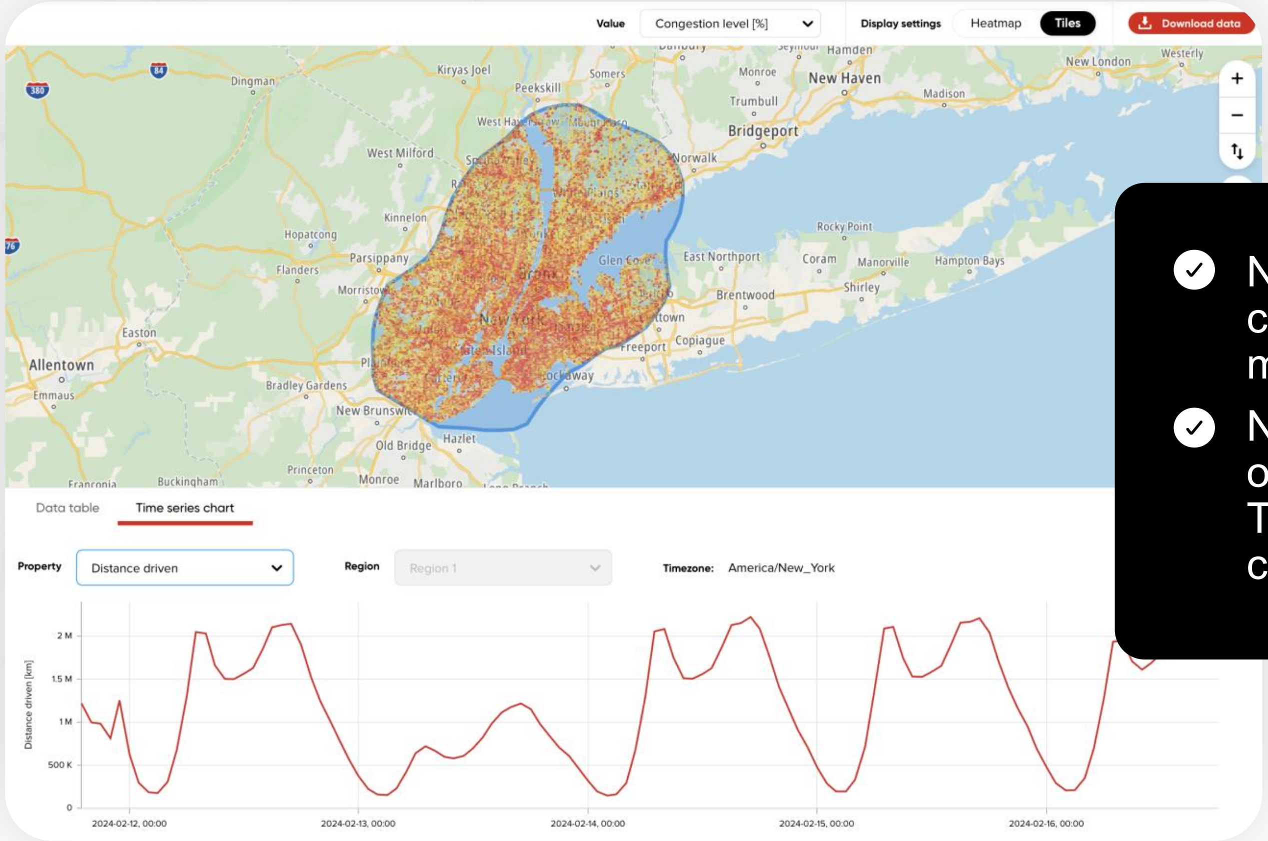
North America Ranking

Rank by filter	World rank	City	Average travel time per 10 km	Change from 2023	Congestion level %	Time lost per year at rush hours	Congestion world rank
1	17	Mexico City Mexico	31 min 53 s	+ 1 min	52%	152 hours	1
2	25	New York, NY United States of America	31 min 6 s	+ 40 s	30%	98 hours	222
3	36	Puebla Mexico	29 min 57 s	+ 1 min	30%	73 hours	218
4	65	Guadalajara Mexico	27 min 8 s	+ 40 s	42%	103 hours	28
5	69	Vancouver Canada	27 min 3 s	+ 50 s	35%	86 hours	96
6	75	San Francisco, CA United States of America	26 min 32 s	+ 30 s	32%	84 hours	160
7	79	Leon Mexico	26 min 27 s	no change	28%	71 hours	260
8	95	Toronto Canada	25 min 13 s	+ 1 min 20 s	31%	77 hours	180
9	117	Halifax Canada	23 min 31 s	+ 50 s	30%	83 hours	217
10	129	Winnipeg Canada	23 min 1 s	+ 20 s	26%	74 hours	308

Global Ranking

Rank by filter	World rank	City	Average travel time per 10 km	Change from 2023	Congestion level %	Time lost per year at rush hours	Congestion world rank
1	1	Barranquilla Colombia	36 min 6 s	- 20 s	45%	130 hours	16
2	2	Kolkata India	34 min 33 s	+ 10 s	32%	110 hours	169
3	3	Bengaluru India	34 min 10 s	+ 50 s	38%	117 hours	64
4	4	Pune India	33 min 22 s	- 1 min	34%	108 hours	128
5	5	London United Kingdom	33 min 17 s	+ 40 s	32%	113 hours	150
6	6	Kyoto Japan	33 min 16 s	+ 20 s	39%	95 hours	60
7	7	Lima Peru	33 min 12 s	+ 1 min 30 s	47%	155 hours	9
8	8	Davao City Philippines	32 min 59 s	- 30 s	49%	136 hours	3
9	9	Trujillo Peru	32 min 56 s	+ 30 s	34%	102 hours	119
10	10	Dublin Ireland	32 min 45 s	+ 40 s	47%	155 hours	10

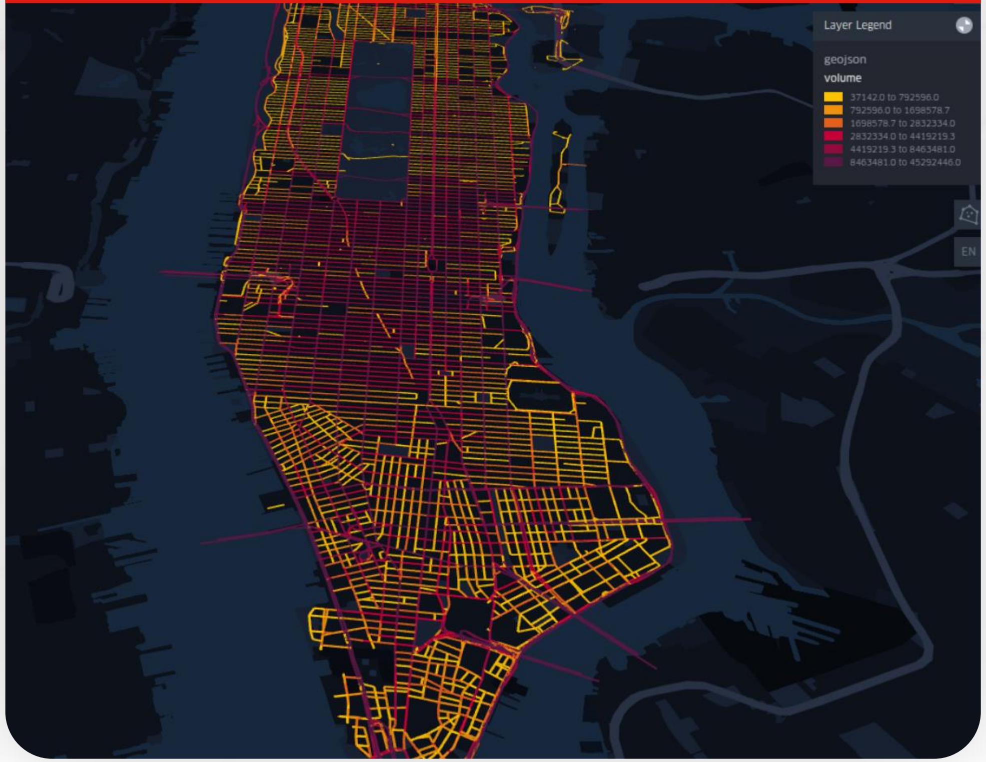
Measuring Congestion Impact
Congestion Index API



- New API to custom congestion level monitoring
- New API based on annual TomTom Traffic Index campaign
- Filtering in time horizon and geography
- Based on million of probe vehicle trips

Analyze Emission Impact with Traffic Volumes and Traffic flow data

Manhattan Area



- In 2023, vehicles in Manhattan drove ~ **4.09 billion miles.**
- Each mile driven emits about 400 grams of CO_2. (*)
- This results in a total of approximately **1.8 million tons** of **CO_2 emissions** from vehicle traffic in Manhattan for the year.

(*) https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle, Neglecting city traffic/congestion/vehicle fleet composition

2025-03-13 19-15-44

Analyze Road Safety Risk and Impact