Customizing Forecasting Tools Using Big Data

presented to

NYSAMPO 2023

presented by

Cambridge Systematics, Inc.

Nikhil Puri

**CAMBRIDGE SYSTEMATICS**

Think ➤ *Forward*

May 9th, 2023

Context

» Over 400 Metropolitan Planning Organizations (MPO) across the U.S.

» MPOs oversee the expenditure of federal transportation funds

» Planning process and decisions directly impact communities

» Good portion of the planning process relies on analytics and forecasts
  - Requires forecasting tools (yes, travel demand models!)

» Evolution of forecasting tools is needed

» Customizable data-driven forecasting tools

CAMBRIDGE SYSTEMATICS

- Founded in 1972
- National leader in innovative solutions- travel demand modeling and data
- Authored guidelines on industry best practices
- LOCUS – Big Data platform developed by Cambridge Systematics
- Successfully developed and applied innovative, <u>customized</u> big data-driven models
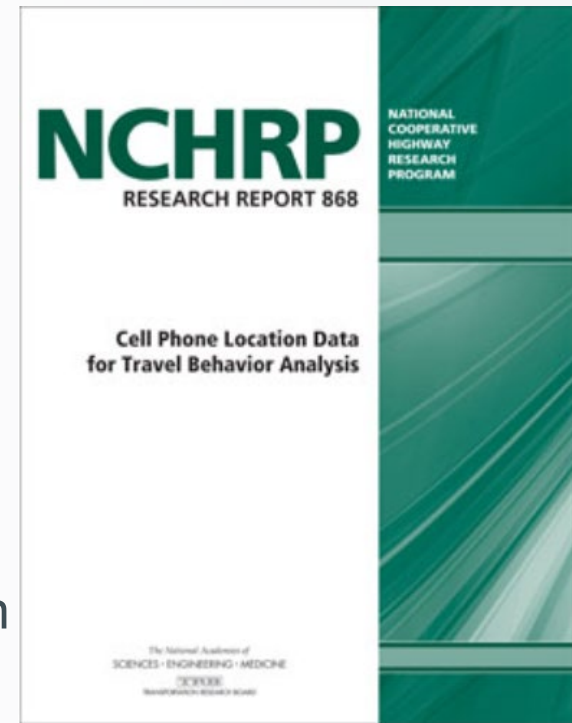
NYMTC Best Practice Model
(Big data used in Validation)

Orange County, NY
(Big data-driven model)

NJ Transit
(Big data-driven model)

San Francisco Bay Area
(Big data-driven model)

CAMBRIDGE SYSTEMATICS

Data….

» Data - Factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation. *Merriam-Webster Online Dictionary*

» Big data- Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions. *Oxford Languages*

  • Can be used to complement traditional data sources and fill in gaps as needed

CAMBRIDGE SYSTEMATICS

Questions That Can be Effectively Answered Using Big Data

» What does the spatial distribution of visitors to a facility look like? In-state/out of state visitors?

» How does seasonality affect travel in a region? Proportion of weekend travel?

» Are most trips work related? Do they vary by time of day?

» What percentage of trips are from disadvantaged communities?

» How has travel changed as a result of COVID?

» How can analytical approaches be scaled?

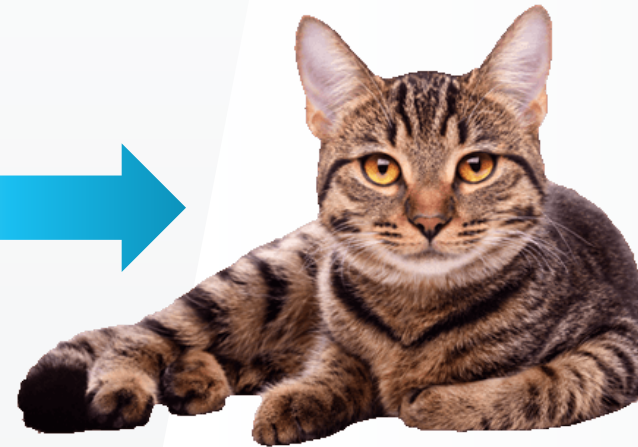» Can we measure the resiliency of communities after a climate event?

CAMBRIDGE SYSTEMATICS

A Model…..

> **"..a simplified representation of a part of the real world-the system of interest - which concentrates on certain elements considered important for its analysis from a particular viewpoint." - Ortuzar & Willumsen.**

Real world   **Modeled World**

CAMBRIDGE SYSTEMATICS

Model Inputs and Outputs

Inputs

- Transportation Networks
- Socioeconomic Data
- External Data
- Special Generators
- Model Parameters

Outputs

- Trips by Mode
- Traffic Volumes
- Congested Speeds
- Transit Volumes
- Summary Information

Customizing Forecasting Tools Using Big Data

» Leverage big data to develop better predictive tools

- Large sample sizes
- Persistent data
- Traditional data gaps

» Additional step of ground-truthing data sources – all data have limitations

» Tools still require validation (and calibration)

» Sensitive to regional policy changes

- Parameters representing local characteristics

» How can we use big data, technology and improved computing power to improve forecasting tools?

CAMBRIDGE SYSTEMATICS

LOCUS – Big Data Platform
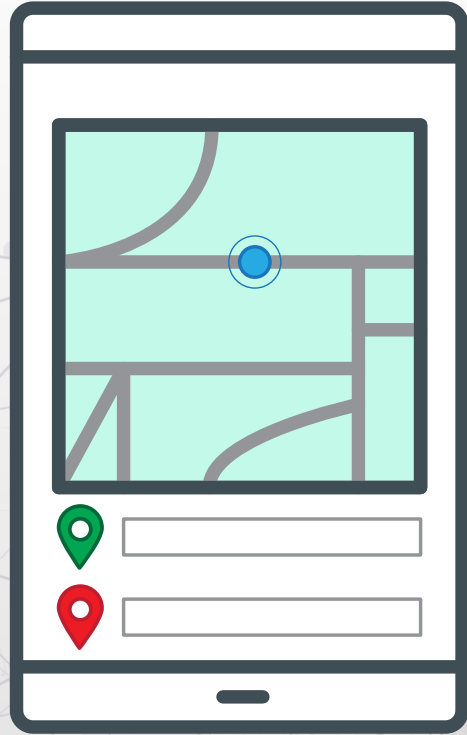
- ❑ **Smartphones**
- ❑ **Location data collection by Apps**
- ❑ **GPS quality**
- ❑ **User permission**

LOCATION-BASED SERVICES DATA

DATA CHARACTERISTICS

- ❑ **Large sample size**
- ❑ **Messy datasets**
- ❑ **Spatial precision**
- ❑ **Persistent ID**
- ❑ **Anonymized**

LOCUS

CAMBRIDGE SYSTEMATICS

Customized Data-driven Models –
SF Bay Area - transit share using a sketch planning tool

» Sketch planning tool

» Autonomous transitway corridor

» Location-based services (LBS) data + ridership data

» Transit share using travel time competitiveness

CAMBRIDGE SYSTEMATICS
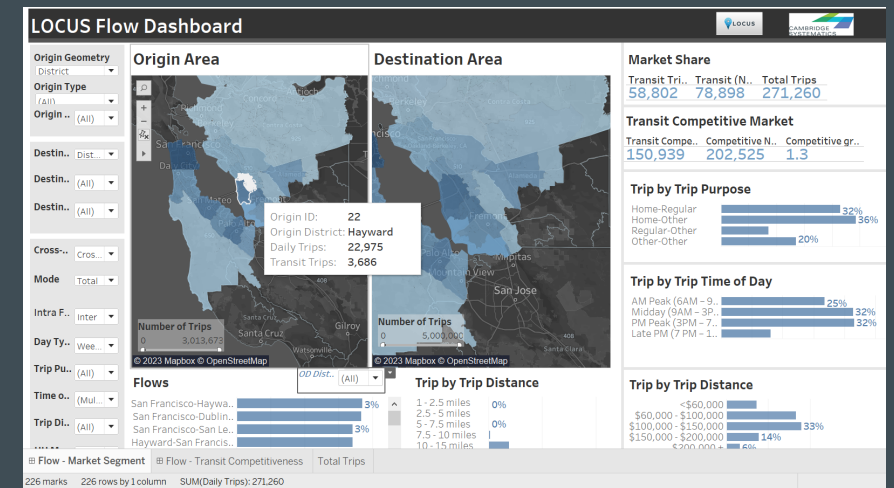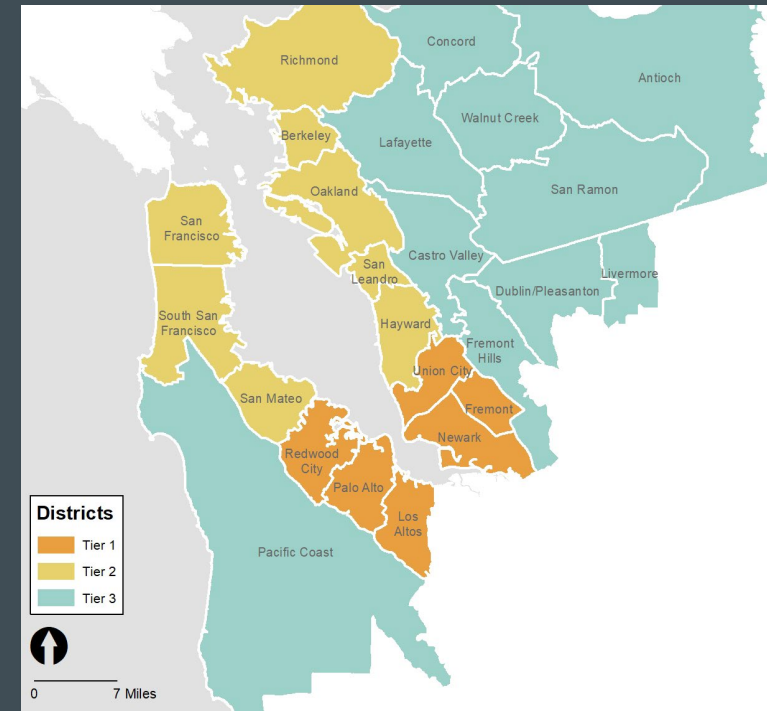
Customized Data-driven Models –
Orange County, NY - model built from scratch

» No Household Travel Survey

» Estimated trip generation models using LBS data (LOCUS platform)

» Combined LBS with traffic counts for external and truck flows

» Assigned trips to existing network

» Validated model

» Approved by Interagency Consulting Group (ICG)

» Easy and quick to run

Internal Trip Generation Models

→ Separate models for trip origins and trip destinations

→ Purposes
  » Home to Work/School
  » Work/School to Home
  » Home to Other
  » Other to Home
  » Non-home based

→ Time of day
  » AM Peak -6AM to 9AM
  » Midday - 9AM to 2PM
  » PM Peak - 2PM to 6PM
  » Off-peak - 6PM to 6AM
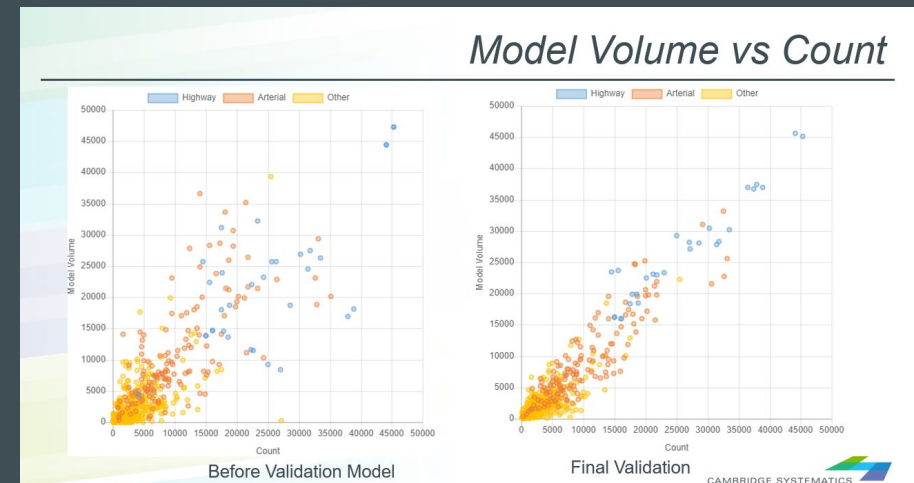
$$Y = B_0 + B_1 X_1 + B_2 X_2 + ... + B_n X_n$$

where:

Y = Dependent variable
B_i = Estimated coefficients
X_i = Independent variables

Model Volume vs Count

Before Validation Model

Final Validation

CAMBRIDGE SYSTEMATICS

Customized Data-driven Models –
New Jersey Transit - special generator + mode choice

» Non-traditional approach

» Transit ridership forecasts – 2026 Soccer World cup

» Key elements

  • Regional flows beyond current models

  • Time period slices

  • Arrival /departure patterns

» Mode sensitivity to cost and travel times

CREATIVE SOLUTION TO A NONTRADITIONAL CHALLENGE

CHALLENGE	APPROACH
Predominantly weekend travel	Leveraged LOCUS – allows scalability add tailored solutions
Flexibility to expand service market	Developed a sketch planning tool using best practices
Ability to consider markets beyond NJTDFM region	Pivoted off NJTDFM and Meadowlands Recreational Transit Model (NJT)
Limited data availability	
Short timeframe	

Hourly Distribution

HOURLY DISTRIBUTION OF DEPARTURES AFTER A NFL GAME

73% DEPARTURES within **2 HOURS** of event ending

Hour before game end	Hour 0-1 after game	Hour 1-2 after game
10%	**39%**	**34%**

RAIL + TRANSITWAY DEPARTURE BY HOUR

Hourly Departure Splits	Hour	Outbound
39%	Hour 0-1	4,075
34%	Hour 1-2	3,543

CAMBRIDGE SYSTEMATICS

In summary….

» Distinguish between data and models

» Question data quality and whether a data source is applicable

» Leverage advantages of big data to complement traditional data

» Obtain an up-to-date picture of travel patters and emerging trends

» Customize forecasting tools to address policy questions

» Create tools of different complexity to address range of policy needs

CAMBRIDGE SYSTEMATICS

Thank you!

Nikhil Puri

Cambridge Systematics

Email : npuri@camsys.com

Cell: 201 926 0425

CAMBRIDGE SYSTEMATICS