Customer Segmentation:
A Deeper Understanding of Ridership Patterns

Washington Metropolitan Area Transit Authority (WMATA)

Session 1066: Transit Data Challenge: Innovative Analytical Tools to Improve Public Transit

Highlights

Customer segmentation methodology identifies distinct segments of WMATA's ridership using machine learning on passenger movement and farecard data.

Addresses shortcomings in earlier ridership forecasting models that do not address unique responses to fare and service level changes among portions of WMATA's ridership.

Produces enhanced datasets and tools that can be used to understand how ridership by segment has changed over time and how fare and service level changes can affect individual customer segments.

Project Context

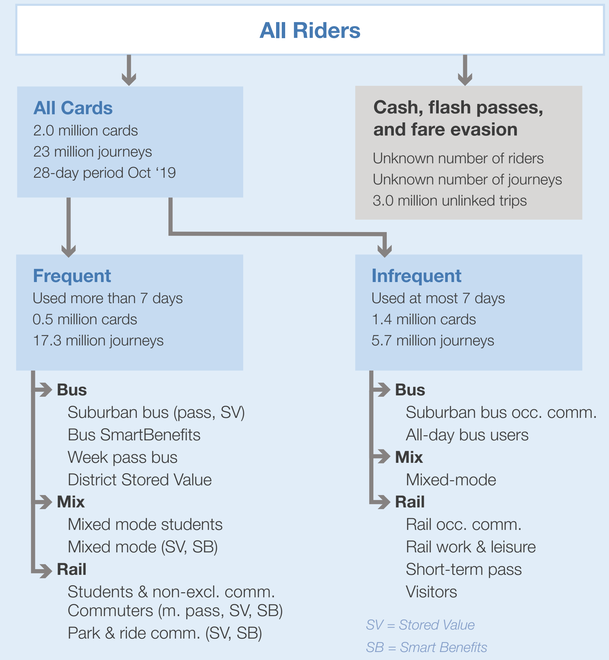
Previous ridership models have struggled to accommodate the diversity of travel behavior on transit, especially as COVID-19 has changed who rides transit and how they ride it.

To address this and other forecasting challenges, WMATA and its consultant team (see Project Team) developed a novel Customer Segmentation model. With this model, WMATA can associate individual SmarTrip cards with a segment based on their cardholder's travel behavior and other characteristics. This allows WMATA to quantify and forecast how different types of customers respond to fare and service changes.

Example Findings

- **Customers using pre-tax transit benefits are much less sensitive to fare changes than Stored Value users**
- **Off-peak fare and service changes are more impactful to rail customer behavior than peak-hour changes.**
- **Tourists have returned to transit more quickly than other infrequent rail customers since COVID-19**

Nested Customer Segments



Methodology

So that the Customer Segmentation model can be continuously applied to new card and passenger movement data, it draws as much as possible from automatically collected data. Major data sources include:

Automatically Collected Data	Demographic Data	Event Data

These datasets were used to define features for each fare card: travel frequency, fare instrument, mode(s) used, time of day of travel, etc. Notably, a "visitor score" for each card was produced using a neural network trained on cards that had been created and loaded at the National Airport Metrorail Station. The visitor model can then be applied to other cards with similar travel behaviors.

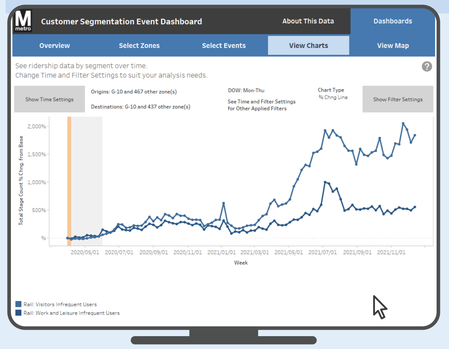
After more than 3,614 different segmentations of cards were generated and tested using machine learning methods, 21 nested customer segments were identified (see figure on the left).

Outcomes

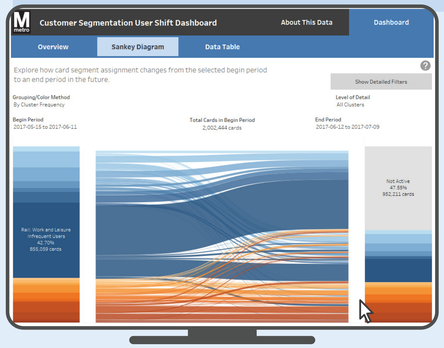
The Customer Segmentation model has broad applicability in planning, marketing, and performance analytics. The project team produced novel tools to make insights from the model widely available within WMATA, including:

- **Interactive dashboards** that support analysis of how ridership activity by segment and segment composition change over time
- **An elasticity modeling spreadsheet tool** that can be used to understand how service level and fare changes affect ridership by segment

Event Dashboard



User Shift Dashboard



Elasticity Model

Cluster ID	Cluster Description	% of Trips	Base ridership	Base Value	New Value	% Increase	Elasticity	Ridership % change	Pct Change - V2*	Post-Change Ridership	Post-Change Ridership - V2*	Net Change
3211	bus - SB suburb + district	0.38%	944	2	2.25	13%	0.031	0.39%	0.00%	948	944	4
3103	rail - short-term pass	0.49%	1,220	2	2.25	13%	0.024	0.30%	0.00%	1,224	1,220	4
BR.3204.3102.3	Bus Riders - SV	0.58%	1,460	2	2.25	13%	-0.110	-1.38%	-1.38%	1,440	1,440	-20
3203	bus - district sv	0.92%	2,293	2	2.25	13%	-0.136	-1.70%	-1.70%	2,254	2,254	-39
3200	students (bus/mix)	0.99%	2,482	2	2.25	13%	-0.013	-0.16%	-0.16%	2,478	2,478	-4
3099	Cash riders	1.05%	2,627	2	2.25	13%	0.033	0.41%	0.00%	2,638	2,627	11
3106	rail - visitors	1.14%	2,851	2	2.25	13%	-0.386	-4.83%	-4.83%	2,713	2,713	-138
3104	rail - occasional commuters	2.77%	6,936	2	2.25	13%	0.054	0.67%	0.00%	6,982	6,936	47
3210	rail - students and non-exclusiv	2.96%	7,403	2	2.25	13%	-0.019	-0.23%	-0.23%	7,385	7,385	-17
3206	park-n-ride - sv	4.97%	12,421	2	2.25	13%	-0.284	-3.56%	-3.56%	11,979	11,979	-442
3209	rail - commuter month pass	5.43%	13,575	2	2.25	13%	-0.023	-0.29%	-0.29%	13,536	13,536	-39
3212	mixed mode sb	6.34%	15,857	2	2.25	13%	0.028	0.36%	0.00%	15,914	15,857	56
3205	park-n-ride - sb	6.64%	16,599	2	2.25	13%	0.037	0.46%	0.00%	16,675	16,599	76
3100	rail - work and leisure	7.94%	19,848	2	2.25	13%	-0.325	-4.07%	-4.07%	19,041	19,041	-807
MM.3208.3105	Mixed Mode	11.33%	28,315	2	2.25	13%	-0.260	-3.25%	-3.25%	27,394	27,394	-921
3202	rail - commuter sb	22.30%	55,759	2	2.25	13%	0.007	0.09%	0.00%	55,810	55,759	51
3201	rail - commuter sv	23.76%	59,410	2	2.25	13%	-0.291	-3.64%	-3.64%	57,248	57,248	-2162
		Total Base	**250,000**					**Scenario Total**	**-1.74%**	**245,659**		**-4341**
								With positive fare elasticities set to zero	**-1.84%**	**245,411**		**-4589**
								With unclustered elasticity estimate	**-4.20%**	**239,500**		**-10500**

TIME PERIOD	TOTAL RIDERSHIP LEVEL:	VARIABLE
AM	250,000	Fares

Project Team

WMATA:

- Catherine Vanderwaart, corresponding author (CEVanderwaart@wmata.com)
- Andrew Gerritsen

IBI Group:

- Nihit Jain
- Rubén Morgan

Korbato:

- Gabriel Sánchez-Martínez
- Gabriel Goulet-Langlois
- Mary Rose Fissinger

EBP:

- Cecilia Viggiano
- Naomi Stein
- Jenna Goldberg

Foursquare ITP:

- Wylie Timmerman
- Adam Recchia
- Rebecca Martin