MBTA Rider Segmentation
The Who, What, and When of Public Transit

Problem Statement

The Massachusetts Bay Transportation Authority (MBTA) is the largest public transportation agency in New England, providing over 1.3 million trips on an average weekday and supporting the dynamic economy of the Greater Boston Area. The system is complex, as it delivers subway, bus, commuter rail, light rail, and ferry services to riders.

A persistent limitation for the MBTA as it considers changes to its fare structure and service planning is the organization’s lack of knowledge around rider time and price flexibility, habits of use, and the interaction of these characteristics. The MBTA runs periodic surveys to obtain federally mandated rider data, but this data does not allow the agency to group riders by nuanced patterns, such as schedule variability for populations with lower abilities to pay, for example.

The goal of this project is to develop a rider segmentation model that can group individuals according to time flexibility and pattern-of-use dimensions. Key variables in this analysis could include likely place of residence (and inferred demographic data), time of day usage, trip patterns and segments (including origin/destination), and other ridership habits. Ideally, this model would be easily re-run by the MBTA to periodically assess how rider pattern-of-use profiles might change over time.

The analysis will focus on the “core system” (encompassing subway and local bus) of the MBTA, as trip data is more plentiful and reliable in this service area than in commuter rail or ferry. Inside this service area, students will be able to analyze movements of unique CharlieCards across space and time. The MBTA will readily provide support with subject matter expertise, past survey data, and the use of the ODX Model (which predicts likely trip origin and destination routes for cards) designed by the MIT Transit Lab.

To provide context for this project, the MBTA recently began a multiyear process to replace its fare collection system across all modes. This new technology will allow the agency to consider future fare structures and data analyses that are often not possible with the current technology. These new fare structures can include policies such as subway peak-time pricing or distance-based fares. Implementing such policies would require substantial research regarding
possible ridership, revenue, and equity implications. A flexible, repeatable rider segmentation model would make these and other analyses much more robust.

**Data Resources** Students will need to use a variety of MBTA and publically available datasets.

- MBTA Automated Fare Collection Data (AFC) going back several years:
  - Card validations, purchases, and location of use.
  - Card status (Senior, Student, etc.) and Sales Channel (Corporate, University).
- MIT/MBTA ODX Data
- US Census or other public data
  - Demographics from the American Community Survey
  - Workforce data from the Department of Labor
- MBTA rider survey data
  - Rider demographic survey
  - Panel survey data
  - Ad-hoc MBTA analyses

**Project Goals**

- Develop a segmentation model that clusters riders into categories based on time of trips, pattern of use, and time flexibility on the subway and bus system. The analysis will need to use MBTA fare collection data, origin/destination data, and may also require US census or MBTA survey datasets. This model will need to take a form that the T may re-use over time, such as a code package with proper documentation. To the extent that different rider groups can be characterized by different price sensitivities, (as observed by past fare changes), the MBTA may also like this feature to be included in the model to help inform future fare policymaking.

- Deliver report on outcomes of rider segmentation analysis using the model. In this report, students should suggest which rider features appear most analytically meaningful for the agency’s future exploration of alternative fare policies, such as peak time increases or night and weekend decreases.