/mbtaPredict

Design Review Preparation and Framing

Background

Our project is a predictive modeling tool that will allow a user to input a station and a date/time range and tells the user how busy the station will be in that date/time range. We plan to take data from 2014 (such as mbta ridership, weather, etc.) and use machine learning to correlate this data into a predictive software.

Key Questions

1. What sort of things should be correlated to ridership data? Right now we are considering temperature, weather status, weekday/weekend, and holiday/not holiday.
2. How important are exact numbers to the user? We may be able to get great qualitative data, but not so great quantitative data.
3. What would be a good contingency plan if the MBTA is unwilling to provide us the data we need?

Agenda

I. Project Introduction (2 min)
   A. Design Review Objectives (1 min)
   B. Brief summary of project (1 min)

II. Discussion of Design Decisions (8 min)
   A. Discussion of Predictive Modeling part of project (5 min)
   B. Discussion of User Interface part of project (3 min)

III. Request for Feedback (12 min)
   A. Clarification Questions from audience (5 min)
   B. Questions from us for audience (7 min)

Glossary

Ridership: numerical characteristics defining how busy the T is at any given time.

Regression: determining the relationship between two variables. We would use this technique to determine how MBTA ridership changes with respect to the weather, or other parameters.

Cluster analysis: Grouping sets of objects into groups based on similarity. We want to look for specific trends in ridership, such as when a specific event (e.g. a major snowstorm or a Red Sox game). We might see distinctive behavior associated with these events that would need to be taken into account using a method other than a normal regression tool.