

LEVEL: Intermediate

DURATION: Full day. Split over two afternoons

Pre-requisites

Participants should have some experience with the tidyverse (e.g. the pipe operator, dplyr selectors, and other basic elements). Some modeling experience, on the level of linear or logistic regression, is also needed.

Overview

The workshop will create a regression model for predicting public transportation ridership. Simple models will be used to start and then more complex methods of crafting features will be used to improve performance. The model performance will be characterized using a time-series resampling method.

This workshop will introduce the basics of fitting models with the tidyverse. A basic case study will be used to demonstrate how to create models, engineer model features, and evaluate performance using resampling

Trainer

Max Kuhn is a software engineer at RStudio. He is currently working on improving R's modeling capabilities. He was a Senior Director of Nonclinical Statistics at Pfizer Global R&D in Connecticut. He was applying models in the pharmaceutical and diagnostic industries for over 18 years. Max has a Ph.D. in Biostatistics.

