At Urbint, we pride ourselves on our ability to make accurate predictions about the state of urban infrastructure- an high-fidelity model of the future state of infrastructure is the first step in a coherent infrastructure plan. In NYC, perhaps no infrastructure is as familiar as our subway system. In this exercise, you will take advantage of the publicly available NYC Subway Turnstile Data in order to forecast ridership numbers.

Your response should include any code used to answer the following questions and instructions for running that code (reproducibility is important!). Use whatever languages or tools you like, but please make portability a consideration.

Questions:

1. Using January to November as a training set, we wish to step through every day in December, and using all data prior to that date, forecast the total number of entries on that day. Ensure that you do whatever data cleaning and filtering is necessary. What is the most appropriate metric for evaluating success?
2. Similar to 1, forecast turnstile entries for every station individually for every day in December. Is this problem easier or harder than forecasting for the system as a whole?
3. Augment your solution to 2. With whatever external data (one or two sources) you think is most appropriate. Explain why you think this data would be useful in predicting ridership and the improvement in predictive accuracy