Background

Washington D.C. has one of the most sophisticated bike sharing programs in the nation. It also fortunate enough to have an extremely valuable data tracking tools available to help prospective bikers get a heads up on where exactly they can expect to find bikes available (and where they can dock them when done with their use). This data resource is known as the Capital Bikeshare Tracker (CaBi Tracker for short). For each of the many bike share stations located across the D.C. metro area, CaBi Tracker allow a quick view of the status of docks at that station – This way someone interested in grabbing a bike can see if the station they want to use is about to run out, or if their destination is close to a station that can accommodate dropping off a bike they’ve shared.

But this kind of data isn’t just useful to bikers... It’s also useful to the managers of the bike share program, charged with the responsibility of best serving the D.C. public. Being able to anticipate scenarios in which stations are empty (bikers can’t get a ride) or full (bikers can’t drop off) is critical to the effective management of the program. If these “outages” at various stations can be better predicted, intelligent steps can be taken to adjust resources in to reduce the risk of service outages in the future.

That being said, neither biker shortages nor complete occupancy scenarios are easily predicted... but that doesn’t mean we can’t try.... Fortunately we have access to their data, both for the bike share system as a whole (http://www.cabitracker.com/status_history.php) and for individual stations (e.g. http://www.cabitracker.com/station_outage.php?id=98)

On the system-wide bike usage and outage graph on the left, trends in use and station outage are compared between two specific data (red line vs. orange area) across the span of a typical day. The report table below shows an example record of empty and full events (along with their duration) for a particular station of interest over a single day.

Your Task: Use the data available and over the next week develop 3 separate models for predicting the “status” (empty, full or no-outage) of the 3 stations listed below at 8am, 9am, 6pm and 7pm EST on a given weekday (predicted statuses will likely vary by time and day for a given station). Your models must make these predictions at least one full day in advance (i.e. your model(s) must forecast the 9amEST Thursday status of the first station by 9amEST the day before (Wed)).

Stations of Interest: 15th & P St NW (31201); 18th St & Pennsylvania Ave NW (31242); Adams Mill & Columbia Rd NW (31104)

Submit your models by March 24th in spreadsheet form to Dr. Elliot Bendoly (elliot_bendoly@bus.emory.edu). Either individuals or teams may register on submission; team members will split any associated prize.

Starting on March 25th and through March 30th, your models will compete against others submitted for “best predictive capability” (number of “correct” statuses accounted for across the 3 stations x 4 time periods, over the 5 day period of examination = 60 predictions total). If you have additional sources of data necessary for your models to base predictions on (e.g. from other sites), you must submit that information on a daily basis over this evaluation period. The entrant with the best performing set of models will be awarded $500. Second place will received $250. Third place will receive $100. Both the top 3 MBA and top 3 BBA performers will also be separately recognized.