Handling Bicycle Rack Installations in NYC

Background:

The Department of Transportation (DOT) in New York City installs and maintains numerous sidewalk bicycle-parking racks, like the ones shown to the right. These racks are placed throughout New York City to provide infrastructure support for cycling, alleviating traffic-related problems across the city. Racks are generally placed on City-owned property with wide concrete sidewalks and away from the natural flow of pedestrians. Additional placement considerations include distance from crosswalks and street signs, mailboxes, benches, etc. (see: http://www.nyc.gov/html/dot/downloads/pdf/cityracks-clearance-guidelines.pdf for detailed information).

Your task is to develop a data driven strategy for identifying and prioritizing locations for new rack installations. Beyond the technical specifications above, many other factors can play a role in this decision. For example, demand for racks will depend on the location of major bike routes, cross points of bike routes with other modes of transportations, and proximity to office buildings, shopping points, residential areas, and other points of interest. Further, there can be different types of racks, including multiple and single short-term racks, sheltered long term racks, indoor racks, etc. This link: http://www.nyc.gov/html/dot/html/bicyclists/bicycleparking.shtml provides information on different rack types, existing locations, and some other placement considerations. As you work on this case you may wish to consider yet additional factors that may impact rack placement, we leave this up to you.

Deliverables and schedule:

You are tasked with creating a 10 minutes presentation that outlines your proposed approach. Your presentation should include numerical support from the data available to you. Predictive models, strong descriptions, visualization and other analysis that supports your argument should be included in the presentation. You will deliver your presentation to a panel of judges tomorrow (April 3) starting at noon. All presentations should be emailed to nevod@rpi.edu by 11:30am tomorrow. **Bring a laptop (preferably with HDMI cable connection) and a backup copy of the presentation with you.**

Your presentation will be evaluated on your ability to properly explain and motivate the problem, your displayed understanding of the data, the logic and rigor of your analysis, the level of creativity and sophistication of your proposed solution, and your presentation skills.

Detailed Schedule:

- Saturday/Sunday 8am – 6pm: The Pittsburgh building will be open for team work
- Saturday 9am: mandatory meeting for all teams, Pitt4206.
- Saturday 2-4pm: TA will be available at Pitt4206 to answer any questions.

- Sunday 11:30am: deadline for sending presentations to nevod@rpi.edu
- Sunday 11:45am: deadline for signing up for presentations (Pitt4206)
- Sunday 12pm – 1:30pm: presentations
- Sunday 1:30pm – 2:30pm: Lunch, Pittsburgh building 4th floor
- Sunday 2:30pm – 4pm: Final round presentations and announcement of winners (Pitt4206)

Available data:

One very useful source of information comes from the bike-sharing network that operates in NYC. CitiBike (citibikenyc.com) is a bike sharing system consisting of a fleet of bikes in a network of docking stations around the city. The bikes can be rented and returned to any station in the system. The service is available for use 24 hours a day, year round. While it is unlikely that CitiBike riders will use the DOT racks (since they use the docking stations), you can use CitiBike data to understand bike usage and patterns across NYC, to inform your models. Naturally the data are biased towards existing stations location so keep this in mind.

The following three datasets are available (you can use the stationID field to join the different files):

- *Status feed on Citibike stations*. Two files are available in a dropbox.
    - station-status.csv contains the following fields:
        - id: unique id for a station
        - lastCommunicationTime: unix time[1] of station report
        - availableDocks: number of docks available at station
        - totalDocks: total number of docks at station
        - availableBikes: number of bikes available at station
        - statusKey: 1=>In Service, 3=>Not In Service

      (When a bike is out, the number of available docks goes up by one and the number of available bikes drops by one.)
    - station-data.csv contains the following fields:
        - id: unique id for a station (same as above)
        - stAddress1: street address
        - stAddress2
        - city: city
        - postalCode: zip
        - longitude: longitude of station
        - latitude: latitude of station
        - altitude
        - testStation: false=>station in use, true=>station being tested
        - stationName: mostly same as stAddress1
        - landMark
- *Citibike Trip Histories:* https://www.citibikenyc.com/system-data (files can be downloaded from AWS)
    - A header/row from one of the files contains the following information:
        - *"tripduration","starttime","stoptime","start station id","start station name","start station latitude","start station longitude","end station id","end station name","end station latitude","end station longitude","bikeid","usertype","birth year","gender"*
        - *"471","2014-01-01 00:00:06","2014-01-01 00:07:57","2009","Catherine St & Monroe St","40.71117444","-73.99682619","263","Elizabeth St & Hester St","40.71729","-73.996375","16379","Subscriber","1986","1"*

[1] https://en.wikipedia.org/wiki/Unix_time

Although the data above is about CitiBike, here are some examples of factors you may wish to consider about stations and then use them to infer about possible locations for racks:

1. How would you define a metric to capture usage activity at a station?
2. Do you see stations with low levels of relative activity?
3. How do station locations relate to other public transit options in the area?
4. How does station usage relate to other public transit options in the area?
5. How do bicycling distances impact station locations?
6. Do the demographics of station usage correlate with neighborhood demographics?
7. Are there indicators of commercial or landmark activity amenable to station usage?
8. How does usage by time of day and day of week relate to station locations?
9. Can you aggregate station usage to characterize bike demand by geographic area?

Additional data sources:

- A 2009 Bike Share feasibility study provides good background and insights into the topic: http://www.nyc.gov/html/dcp/pdf/transportation/bike_share_complete.pdf
- DOT data feeds can be found here: http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#Bikes The category "bicycles" provides relevant bike data
- *Venue data from 4SQ*[2] can be used to characterize the surrounding area for a requested location. In particular, the search endpoint can be used to produce a list of venues near a station using its latitude and longitude. You can access documentation on the search endpoint here: https://developer.foursquare.com/docs/venues/search. Please refer to the appendix for some quick instructions on how to connect to this API.
- *Google Maps* can be used to identify bike routes for a trip. Google has a Directions API that you can use to characterize the distance between two stations or a station and another location. The API supports multiple travel modes including a specific bicycling mode. Documentation on the Directions API can be found here: https://developers.google.com/maps/documentation/directions/. Please refer to the appendix for some quick instructions on how to connect to this API.
- You can use the *NYC open data* to find demographic and area data around a location: https://nycopendata.socrata.com/
- https://www.strava.com and http://www.mapmyride.com provide good information on bike routes and usage
- Any other data source your deem relevant

Summary:

Where should racks be placed? This is the challenge you are asked to address in this case. Use all of the above data and other data sources you deem relevant to develop your rack placement strategy. Make sure your proposed strategy is well supported by data and intuitive given the parameters of the case. Remember – this is a rich problem with many factors that come into play. We highlighted some of them in the above descriptions but there are certainly other factors to consider. Take the time to understand the domain and plan your approach before you dive into the data. Keep an open mind.

If you have any questions during the day please email Dorit Nevo: nevod@rpi.edu

Good Luck

[2] https://developer.foursquare.com/overview/venues.html

Appendix

To connect to Google and 4SQ APIs, you need to register and grab free API keys. Instructions on getting an API key:

- Google Maps: https://developers.google.com/maps/documentation/directions/get-api-key#key. Follow the instructions for the Standard API to get a Server key
- Foursquare: https://developer.foursquare.com/overview/auth#userless. Follow the instructions for "userless" access (since you don't need to connect with Foursquare users). To get a client ID and client secret, register an app here: browse to https://foursquare.com/developers/apps and then hit 'Create a New App'. You only need to provide an App Name and a Welcome page URL for the app (you can leave the rest of the form blank as is). For the Welcome page URL, you can use www.rpi.edu (or any other URL). When you hit 'Save Changes', the app will be created and the response page will show you your client ID and client secret.

The Google Maps Directions API has a guide available here:
https://developers.google.com/maps/documentation/directions/intro#Introduction
Once you have your API key, you can make requests using http command line tool like curl or wget (or right in your web browser) to access the API. For e.g.: the following request will give you bicycling directions between stations ids 79 and 82

https://maps.googleapis.com/maps/api/directions/json?origin=40.71911552,-74.00666661&destination=40.71117416,-74.00016545&mode=bicycling&key=**YOUR_API_KEY**

Similarly, the Foursquare Venue/Search API endpoint can be accessed as follows:

```
https://api.foursquare.com/v2/venues/search
  ?client_id=YOUR_CLIENT_ID
  &client_secret=YOUR_CLIENT_SECRET
&ll=40.71911552,-74.00666661&v=20160315
```

You can use 20160315 as the API version date in all your requests to the 4SQ API.

Both APIs will return you JSON in their response. JSON can be easily parsed and read. Use 'python –mjson.tool <filename>' from the command line to produce a pretty print for JSON content in the file <filename>.

You can also script using Python to read the JSON as a Python dictionary and walk the dictionary to extract the key/value pairs you care about. Refer to this StackOverflow thread for some example code.

You can also script access to both APIs. Refer to these publicly available Python client tutorials online:

- Google Maps Directions API: https://github.com/googlemaps/google-maps-services-python
- Foursquare Venue/Search API: https://github.com/mLewisLogic/foursquare