

New York City Bike Share Data: Methodology

The Open Bus
2015



1 Introduction

New York City's bike share program (Citi Bike) streams live data on bike availability across all stations in the system. This service allows users to look up the number of currently available bikes for a particular station, as well as the number of open docks available to return a bike that has been signed out. The Open Bus (TOB) logs this data to provide retrospective data and summary statistics on the performance of stations. When a dock is completely empty or completely full it becomes unusable to riders who need to sign out or return a bike respectively. TOB will provide accurate statistics on dock performance in order to inform users as to the reliability of particular stations.

2 Data Collection

TOB collects data via Citi Bike's [json live stream](#). Data is captured approximately every 10 minutes, returning the number of functioning docks, the number of bikes at the station, and (consequently) the number of empty docks. Observations are stamped with the time of observation. Data for every station in the system is recorded.

Data collection began on March 1st, 2015. Data collection has been executed continuously since that date. Collection relies on the proper operation of GPS equipment, public servers, and internet connectivity, meaning interruptions in any of this infrastructure will cause periods of incomplete data. These events are rare and can be easily identified in the data as instances of longer than usual time between observations.

3 Web Interface

The web interface provides averaged statistics for one month intervals, with the month indicated on each page. Data displayed on web pages excludes weekends. Docks are allocated to defined neighborhoods.

Stations are fully functional for all riders except in instances wherein the station is either completely empty or full. Therefore, the web interface highlights the percentage of the time the station has either no available bikes, or no empty docks. This information can be compared to the accompanying graph displaying the averaged number of bikes available by hour of day. For example, if a station is often empty, the time of day that this frequently occurs can be inferred by identifying the hour of the day that bikes are scarce.

The capacity of the dock is indicated by the horizontal red line.

4 Raw Data

Raw data will be posted in the first week of each month, providing data collected in the previous month. Raw data is provided in .csv format. Raw data includes observations covering 24 hours per day, 7 days per week within the month indicated. Dock IDs and names are provided and are set according to the Citi Bike json feed. Latitude and longitude coordinates are provided for each observation.

5 Raw Data Codebook

Two data sets are provided for the New York City bike share system. The first is time stamped raw data providing station conditions through time. The second data set provided is collapsed down to a single observation for each station, providing the average number of bikes/docks available over the particular month.

bikeshare_nyc_raw.csv:

dock_id: official dock id provided by Citi Bike
dock_name: official dock name provided by Citi Bike
date: year-month-day format
hour: hour of day (am/pm is determined by ‘pm’ variable)
minute: minute of hour
pm: 0–am 1–pm
avail_bikes: number of available bikes
avail_docks: number of available (empty) docks
tot_docks: number of functioning (online) docks
_lat: degrees latitude of station
_long: degrees longitude of station

empty_full_results.csv:

dock_name: official dock name provided by Citi Bike
dock_id: official dock id provided by Citi Bike
percent_full: percentage of time the station had no empty docks available
percent_empty: percentage of time the station had no bikes available