

Bike Share data

Data Source

New York's CitiBike system is one of the major public bikeshare systems around the world and the largest in the United States. The CitiBike system was launched in May 2013 with 330 stations and 6,000 bicycles in the lower half of Manhattan and some part of northwest Brooklyn. In 2017, the system size expanded to 750 stations with 12,000 bicycles. According to CitiBike report, the number of annual subscribers were nearly 130,000 on July 2017. The trip itinerary dataset (from January 2017 to June 2017) of the CitiBike system is the primary data source employed (<https://www.citibikenyc.com/system-data>). The ridership dataset provides information on start and end time of trips, their origin and destination, geographic coordinates of stations (latitude and longitude), travel time or trip duration, user types, and age and gender for members' trips. The trip data was augmented with other sources including: (1) built environment attributes such as number of restaurants and park area derived from New York City open data (<https://nycopendata.socrata.com>); (2) socio-demographic characteristics at the census tract/zip code level gathered from US 2010 census; (3) the weather information corresponding to the Central Park station retrieved from the National Climatic Data Center (<http://www.ncdc.noaa.gov/data-access>).

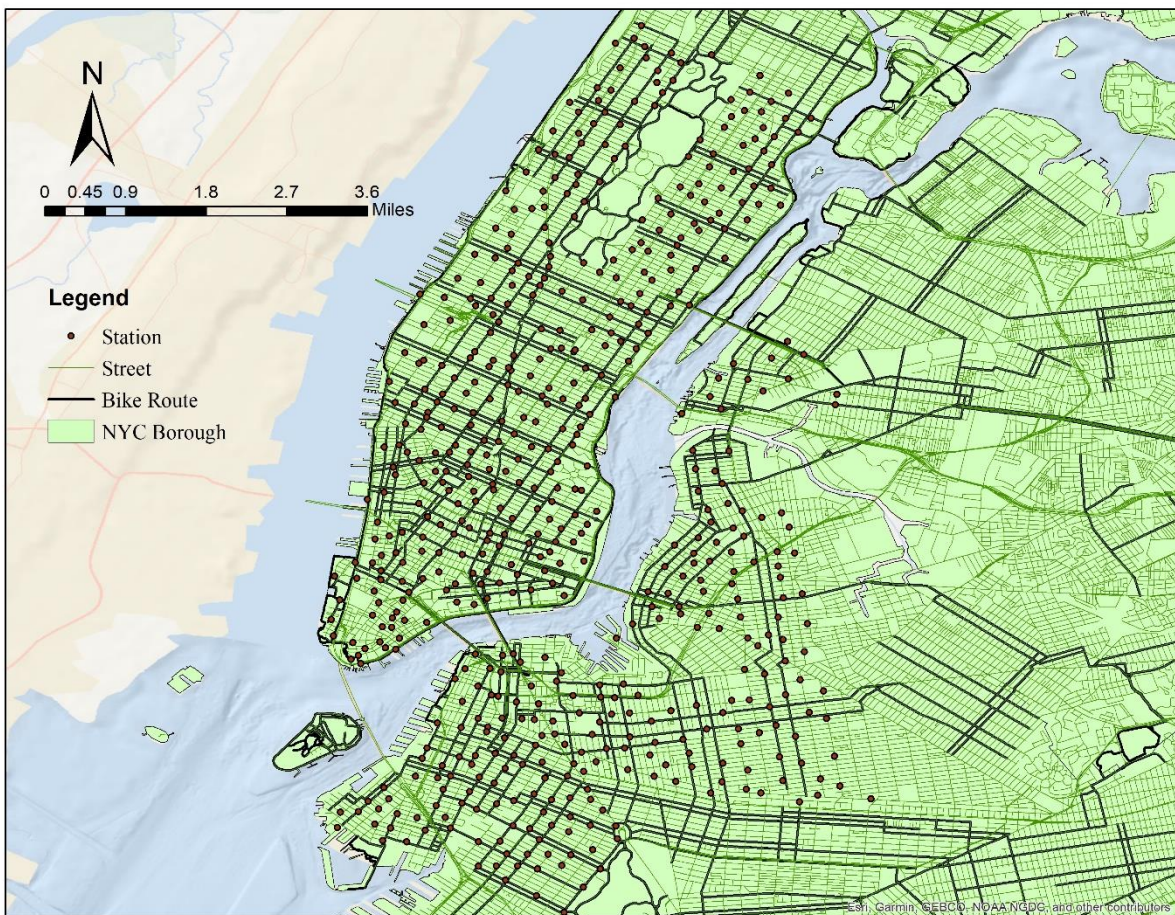


FIGURE 1 NYC's bicycle-sharing system (CitiBike).

Independent Variable Generation

Several independent variables were generated in our study. These can be grouped into six categories: 1) Trip attributes, 2) Socio-demographic variables, 3) Bicycle and transportation infrastructure variables, 4) Weather attributes, 5) Temporal attributes and 6) Land use and built environment variables. Trip attributes includes the network distance between each origin-destination station pair estimated using the shortest path algorithm. While the actual trip might involve a different route, the shortest network distance would be an appropriate indicator of the distance traveled. The socio-demographic attributes considered are population density, job density and establishment density. Population information was collected from US census 2010 and projected for 2017 at the census tract level. Job density data was estimated at the census tract level while establishment density was calculated at the zip code level for 2016.

Bicycle and transportation infrastructure variables include CitiBike station attributes, bike route length, and public transit stations. For these attributes a 250-meter buffer around each station was created. The 250-meter buffer seems a reasonable walking distance based on the distances between CitiBike stations and the dense urban form of New York City (50). The variables created at the buffer level include length of bike routes, length of roads (minor and major roads). The number of CitiBike stations and total dock's capacity within 250 meter buffer (excluding the station considered and its capacity) were estimated to capture the impact of neighboring stations on cycling trips. Number of subway stations and bus stops in the 250 meter buffer were generated to examine the influence of public transit on cyclist's preference of destination station. Weather variables include average temperature, relative humidity and precipitation over the week. Several interaction variables were also created. Seasonality is the only temporal variable considered. We consider winter (January-March) and Spring (April-June) as dummy variables.

Finally, several land use and built environment variables were considered including the number of facilities (schools, colleges, hospitals), the number of point of interests (museums, shopping malls), and the number of restaurants (including coffee shops and bars), total area of parks and commercial space (office, industry, retail) within 250 meter buffer, station elevation, and distance of destination from Times Square. Non-motorized vehicle score (average of walk score and bike score) and transit score associated with each CitiBike station was considered at the census tract level.