**Citi Bike Data** **Description[[1]](#footnote-1)**

Citi Bike is a private bicycle sharing service in New York City. This is a subsample of the Citi Bike System Data (<https://www.citibikenyc.com/system-data>) and contains information about demand (the number of trips) between different stations. The file contains a random sample of taken from June 1, 2017 to May 31, 2018 (31,452 out of 15.7 million rides). The information has been supplemented with demographic, economic, and weather information pulled from a variety of sources.

We extracted the demographic and economic data from the American Community Survey completed by the US Census Bureau (<https://www.census.gov/programs-surveys/acs>). Summary data was extracted at the ZCTA (Zip Code Tabulation Area) level and matched by zip code to each Citi Bike station address.

We extracted the weather data from the NOAA National Centers for Environmental Information historical weather archive (<https://www.ncdc.noaa.gov/>). Data was collected at the Central Park, NYC weather station (WBAN Number: 94728) and matched by date and time to Citi Bike trips. The records indicate the general weather for the NYC area during the ride and are not specific to the weather at any Citi Bike station.

Citi Bike has two basic user types. A ***customer*** buys a temporary pass (a 24-hour pass or 3-day pass). A ***subscriber*** purchases access to bicycles for an entire year. Because subscribers register with Citi Bike, there is more information about them in the database.

The file contains these following fields:

**Customer\_prop:** The proportion of Customers

**DemandDate:** the date the bicycle was rented, and the trip started

**DemandTime:** the time of day (daytime vs evening) the bicycle was rented, and the trip started

**DayOfWeek:** the day of week the bicycle was rented, and the trip started

**Month:** the month the bicycle was rented, and the trip started

**Demand**: the number of trips between the start station and the end station

**StartStationId:** A unique ID for the station from which the bicycle was rented (the ‘start’ of the trip)

**StartStationName:** Name (often the street location) of the start station

**StartStationLatitude, StartStationLongitude:** Latitude and Longitude of the start station

**StartNeighborhood:** General neighborhood of the start station (e.g., East Harlem, Soho, Park Slope)

Similar fields are defined for the End station (EndStationId, EndStationName, EndStationLatitude, EndStationLongitude, EndNeighborhood).

For each row (which corresponds to a specific combination of start and end stations), we define

**TripDurationMinutes:** average duration of trips starting at the start station and ending at the end station, in minutes

**DistanceMiles:** Approximate distance between start and end station, for trips starting at the start station and ending at the end station, in miles. For example, in row two, **DistanceMiles= 0.8489,** which means that theapproximate distance for trips that started in StartStationId 83 and ended in EndStationId 3368 is 0.8489 miles.

**MilesPerHour:** Approximate speed of the trip, based on the trip duration and the distance between stations (we assume – possibly incorrectly – that the trip was direct travel between stations)

**StartCommuteTime, EndCommuteTime:** The mean travel time in minutes spent commuting to work in the neighborhoods where the trip started and ended.

**Temperature:** The temperature (in Fahrenheit) at the Central Park weather observation station at the start of the trip.

**DailySnow:** The total snow falls in inches that day at the Central Park weather station.

**DailyPrecipitation:** Total precipitation in inches that day at the Central Park weather station.

**StartPerCapitaIncome, EndPerCapitaIncome:** The mean income for each person living in the neighborhoods where the trip started and ended. It is calculated by dividing the aggregate income of the residents by the total number of residents in the neighborhood.

**StartPctHouseholdsNoVehicle,** **EndPctHouseholdsNoVehicle:** The percentage of households in the neighborhoods at the start and end of the trip that report not having a vehicle available.

1. This dataset is adopted from Decision Science course taught by Prof. Robert Shumsky and Prof. Jim Smith at the Tuck School of Business at Dartmouth. [↑](#footnote-ref-1)