

Project Proposal:

Background:

The link between the Subway and Citibike plays an important role in building a more resilient transportation system. Bike sharing has been proven to be a reactionary mode of transport when looking at subway and bike ridership during and after Covid-19, subway delays, disaster relief and crime to name a few.

Leveraging this relationship, Citibike wants to understand what are the most active transportation hubs in NY. Based on that, they want to know if there are underserved hubs where adding more bike racks would increase the user base.

Process:

The question will be answered in four steps:

1. What are the busiest subway stations?
2. For each of those areas how many Citibike stations are there in a defined radius.
3. What is the ratio of subway traffic: number of bikes in a defined radius
4. What are the subway stations with the lowest ratio? How can this inform bike racks addition?

Data Description:

[Remote Unit/Control Area/Station Name Key](#) and MTA Turnstile data sets will be merged to add station name to the MTA Turnstile data. The resulting data set will then be combined with [NYC Transit Subway Entrance and Exit](#)

Data using the Station Name column to get location coordinates and traffic in one dataset

- [MTA Turnstile Data](#)
- [Remote Unit/Control Area/Station Name Key](#)
- [NYC Transit Subway Entrance and Exit Data](#)

The following two datasets will be used to combine longitude and latitude in Citi Bike Trip History with Bike numbers per station in Citibike Station Status. Common column is: Station ID.

- [Citibike Trip History](#)
- [Citibike Stations Status](#)

Tools:

- SQLite
- Numpy and Pandas
- Matplotlib, Seaborn for visualization