Visualizing Bicycle Safety Patterns for the NYC Citi Bike Rideshare Program

Annie Bui, Brannon Carter, Brian Jigarjian, Edward Owen, Mary Bryan Owen
CSE 6242: Fall 2022 – Data and Visual Analytics, OMSA: Georgia Tech

Introduction

New York City has made considerable improvements to bike safety and infrastructure since the mid 2000s by decreasing bike deaths in the city by 75% *(Bicycle Friendly Report Card*, 2022). Still, there have been more than 100 bike deaths since 2014. NYC continues to stress the need for improvements to achieve its Vision Zero strategy (*Green Wave: A Plan for Cycling in New York City*, 2019).

The innovation for this project focused on answering two questions: ***"What does bike safety look like in New York City?"*** and ***"Are there dangerous areas in the growing Citi Bike network that require attention?"***. We combined New York City Vehicle Collision data, Citi Bike Trip data, and Geospatial data to build an interactive, web-based tool to help users answer these questions. The tool displays all Citi Bike docking stations on a map of New York City and assigns each a "Severity Index" to measure safety. Users can understand which stations or cluster of stations are in relatively dangerous areas for bikers and can dive deeper into understanding why those stations are dangerous. The aim of this tool is to detect risky areas for City planners and Citi Bike to understand where infrastructure investment may be needed and for bikers to understand where they are at risk of a severe accident.

INITIAL DATA

1,644
Docking Stations
Oct. 2022

1.945mil
NYC Collisions
7/2012 - 10/2022

FINAL DATA: *after processing/cleaning*

Did not involve a Bicycle, but will contribute to biker safety: 949,724

Involved a Bicycle: 47,301 | 4.7%

NYC BOROUGH	NEIGHBORHOOD COUNT	BIKE-RELATED COLLISIONS
MANHATTAN	33	20,134
BROOKLYN	30	17,189
BRONX	20	5,053
QUEENS	7	4,925

DATA SOURCES

NYC Collisions Dataset
3.25GB
1,945,052 rows
29% Missing data as NaN
CSV from NYC Open Data
Dates: *7/2012 - 10/2022*

Citi Bike Tripdata
4.88GB
3,021,670 rows
CSV from Citi Bike System Data
Date: *October 2022*
GroupBy: Start Station Name
Unique Docking Stations List

Geospatial GeoJson Files
5 NYC Boroughs
266 NYC Neighborhoods
Latitude and Longitude shape area files with polygon boundaries for creating an interactive map of NYC

METHODOLOGY

Mapping Collisions to Nearest Station

We used python uszipcode package to recover **35,367** collisions with missing Lat/Long coordinates using provided Zipcode. The remaining **1.69** million collisions with Lat/Long coordinates were mapped to the nearest Citi Bike Docking Station.

K-Nearest-Neighbors *(metric = haversine distance)*
nearest station *= knn.BallTree.query(k=1, radius=1mi)*

Ball Tree is a recursive tree-based k-neighbor search algorithm. Trained using station data with the haversine distance metric for calculating the distance between two points on a sphere, in our case the Earth. This reduced the computational cost of a brute force query lookup for each collision's closest station:

from **O[N^2]** *to* **O[log(N)]**

Measuring Collision Severity

Collision Severity Index (SI): using collision features counts

Collision Parameter	B=Bike-Related Weight	N=Non-Bike Weight
V = Vehicle Count	0.2	0.1
I = Number **Injured**	0.6	0.4
K = Number **Killed**	2.0	1.0

$$SI = 0.1 + [\sum_{i}^{n} 0.2V_{Bi} + 0.6I_{Bi} + 2K_{Bi}] + [\sum_{i}^{n} 0.1V_{Ni} + 0.4I_{Ni} + K_{Ni}]$$

Collision distance from its closest docking station provides a weighted dampening factor(α) for farther away collisions.

$$\alpha_i = -0.901(distance_i) + 1$$

$$SI_i = \alpha_i(SI_i)$$

The Severity Index was then normalized for understandability.

A Historical Safety Snapshot

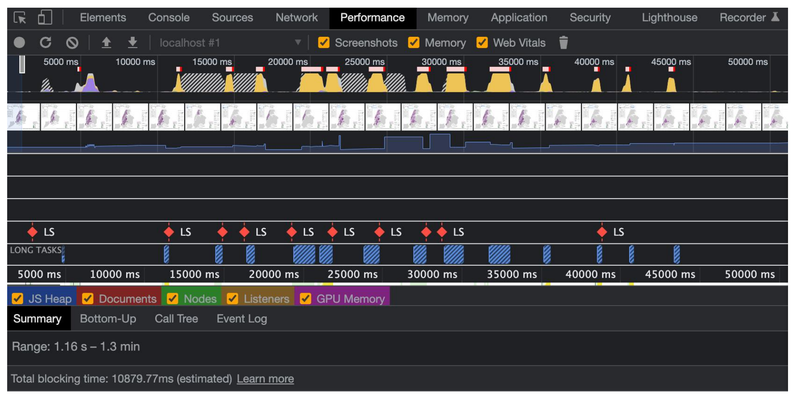
Docking Station 5-Star Safety Rating (SR)
Star Rating from station(d) SI percentile distribution

$$SI_d = \sum_{i}^{n} SI_{di}$$

A D3 Choropleth Map for Visualizing Biker Safety
Our method used historical collision data from 2012 to October 2022 to create an interactive geospatial map of NYC. Users can explore patterns of collision severity as it relates to biker safety around the Citi Bike docking station locations.

EXPERIMENTATION & RESULTS

Performance and UI Testing

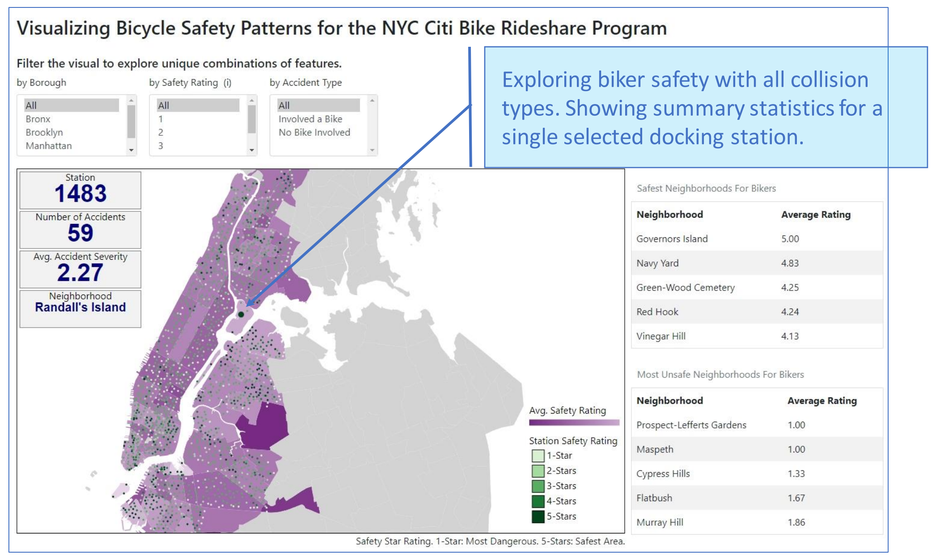


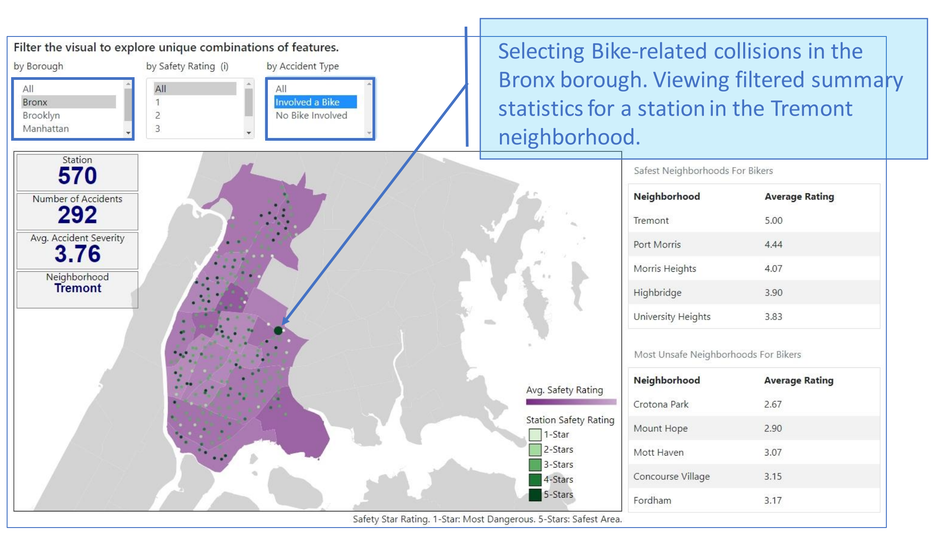
A load-time performance analysis was conducted using the Google Chrome Performance recording tool.

UI interactivity was developed from initial user stories to test and incorporate key safety indicators, color palettes, and filtering options into the final interactivity of the visual.

Validation testing of the generated 5-Star Safety Rating was conducted using a random sampling method on both raw collision data and after aggregating by docking stations.

Final Interactive Visualization





Discussion

This dashboard can be used by several different groups such as bicyclists, urban planners, and policy makers among others. Identifying high safety areas such as the islands (except for Manhattan) can help users find better areas for biking. Outside of these high safety areas, safe neighborhoods such as Navy Yard in Brooklyn and Concourse in the Bronx can be analyzed to help guide safety improvement in other neighborhoods. Lowly rated neighborhoods would be ones first targeted for potential improvement.

Borough	Safest Neighborhood	Most Unsafe Neighborhood	Neighborhood with Most Accidents
Brooklyn	Navy Yard	Mount Hope	Bedford-Stuyvesant
Bronx	Concourse	Prospect-Lefferts Gardens	Kingsbridge
Manhattan	Governor's Island	Murray Hill	Midtown
Queens	Ditmars Steinway	Maspeth	Woodside

There are much fewer 5-star stations surrounding Central Park (Upper West Side, Upper East Side). Conversely, there were more 5-star stations closer to the edges of the boroughs, potentially due to less traffic further from the busiest parts of the city, thus lowering the chance of accidents.