NYU

NYC Bikeshare Trips Prediction

NYU Marron's Urban Data Hackathon

Tao Liang

Agenda

- Background
- Data
- Models
- Outcome
- Discussion

Background

NYU

Background

Since 2015, **bike-sharing** has gradually entered people's lives. As a new public and environmentally friendly mode of transportation, it enriches people's travel choices, improves people's travel convenience, and reduces the pressure of the traditional transportation system. However, with the popularity of shared bikes, the lack of **bike lanes** and supporting facilities have become increasingly apparent.

(Source: citibikenyc.com)

Problem Statement

- How do transportation planners decide where to install a new bike lane when only have a limited budget?
- What factors influence bike-sharing trips?
- How will the bike-sharing trips change after the design intervention?

This project will try to answer these questions by developing a bike-sharing trips prediction model.

Project Abstract

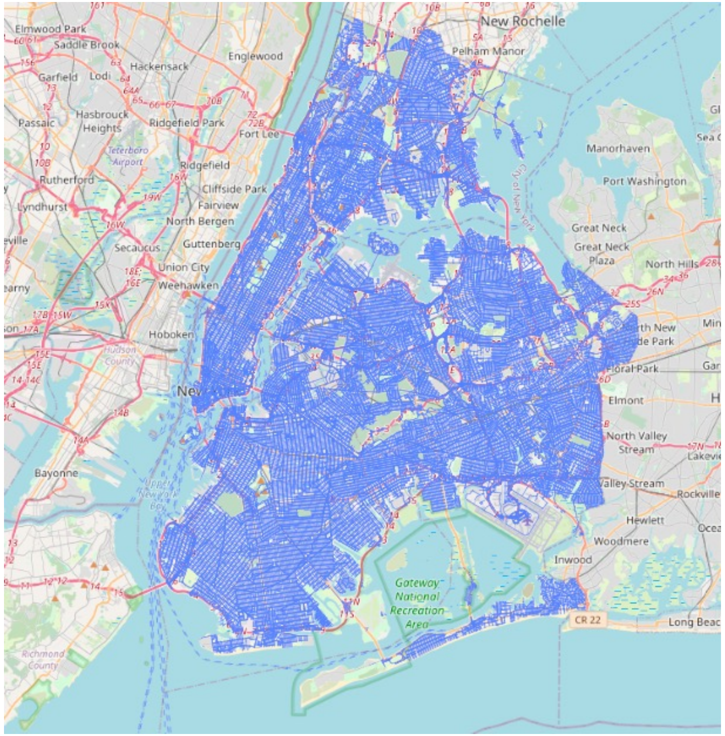
Traditionally, planners may prioritize bike lane investments based on the number of collisions between cars and bicyclists or the value of a particular roadway to the network. However, this project provides a different way to prioritize bike lanes based on **bike-sharing trips** and **simulate the impact of installing a new bike lane on ridership**. In addition, this project will try to reveal the relationship between **social factors and shared bike ridership** and select the **interpretive model**. This could help planners convince the residents who are skeptical of bike lanes.

Data

LION Street Centerlines

Version: LION 21A

Since this project focuses on predicting the link bike flow. We used LION to construct the NYC segment network as the base map, which also contains some segment information, like lane width, has bike lane or not, protected bike lane or unprotected bike lane, etc.



Citi Bike

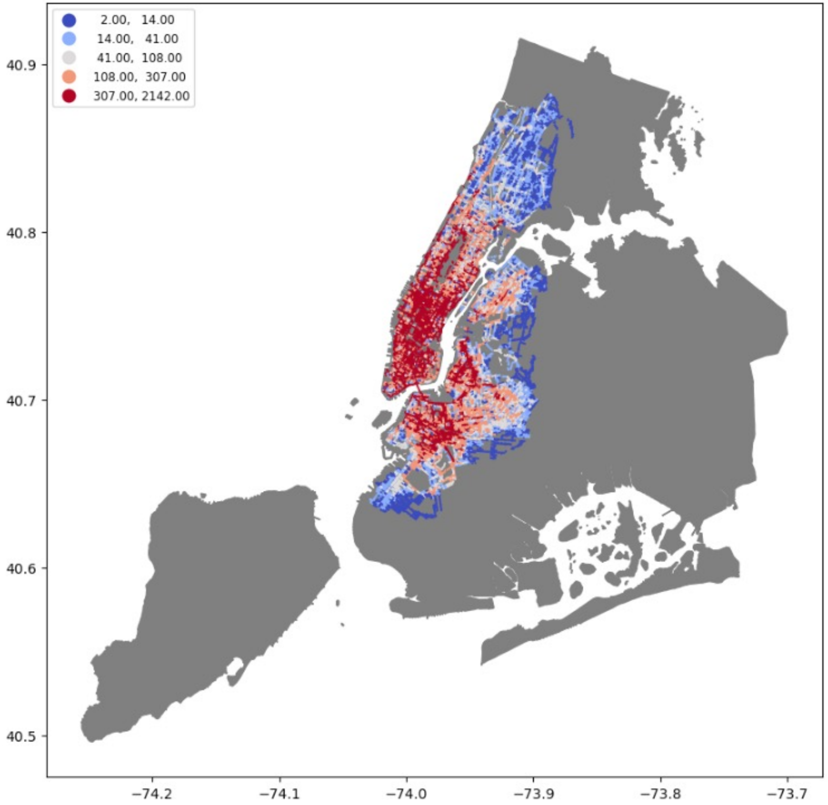
Data Size: 3280560 rows

Date: 9/2021

This project chose Sep 2021 Citibike trips data because it has the highest volume. This data set includes start time, start station, and end station information.

Citi Bike – Link Flow

Since our goal is to find out which street needed to install the bike lane, we need trip data on each segment. However, the Citi-bike data shared with the public is **origin-destination data**, meaning you know the stations where a trip starts and ends but not the route the cyclist took in between. In order to derive the flow on street, we used **OSMnx** to **estimate the trajectories** between start stops and end stops, then assign these trips to the corresponding segment.



Demographics

Source: ACS 5 years estimate, 2020

Date: Unit: Census Tract

Indicators:

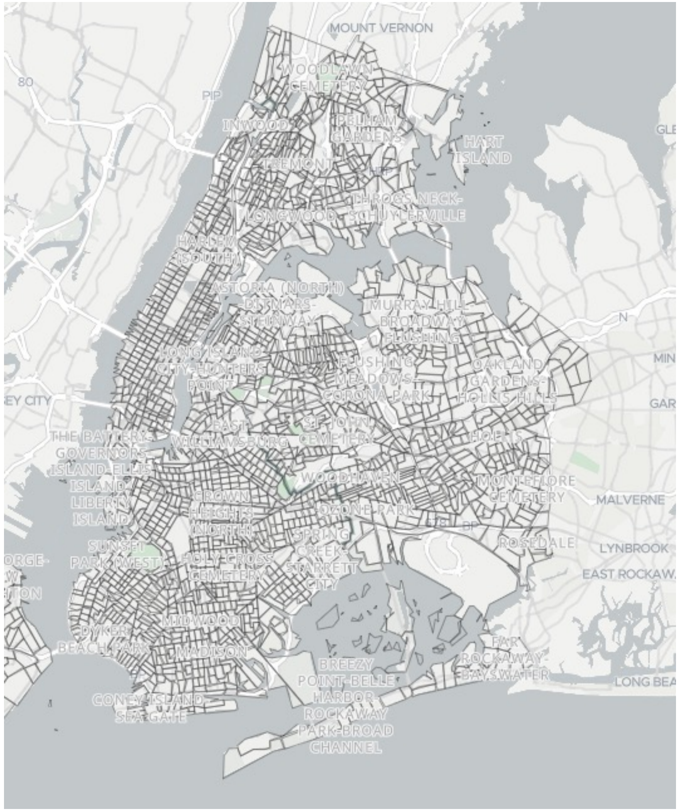
Population

Median Age

Median Income

etc.

We spatial join the segment buffer with the corresponding census tract to get the mean demographics value of each segment.



NYU

Final Features

23 Inpendent Variables

Dependent Variable

- Citi Bike Median Daily Trips

Segment

- Street Width
- Numbers of Travel Lanes
- Numbers of Parking Lanes
- Numbers of Lanes
- With Bike lane or not
- Protected Bike Lane or not
- Unprotected Bike Lane or not
- Truck Route or not
- Borough Dummy Variables
- Distance to Nearest Subway Entry

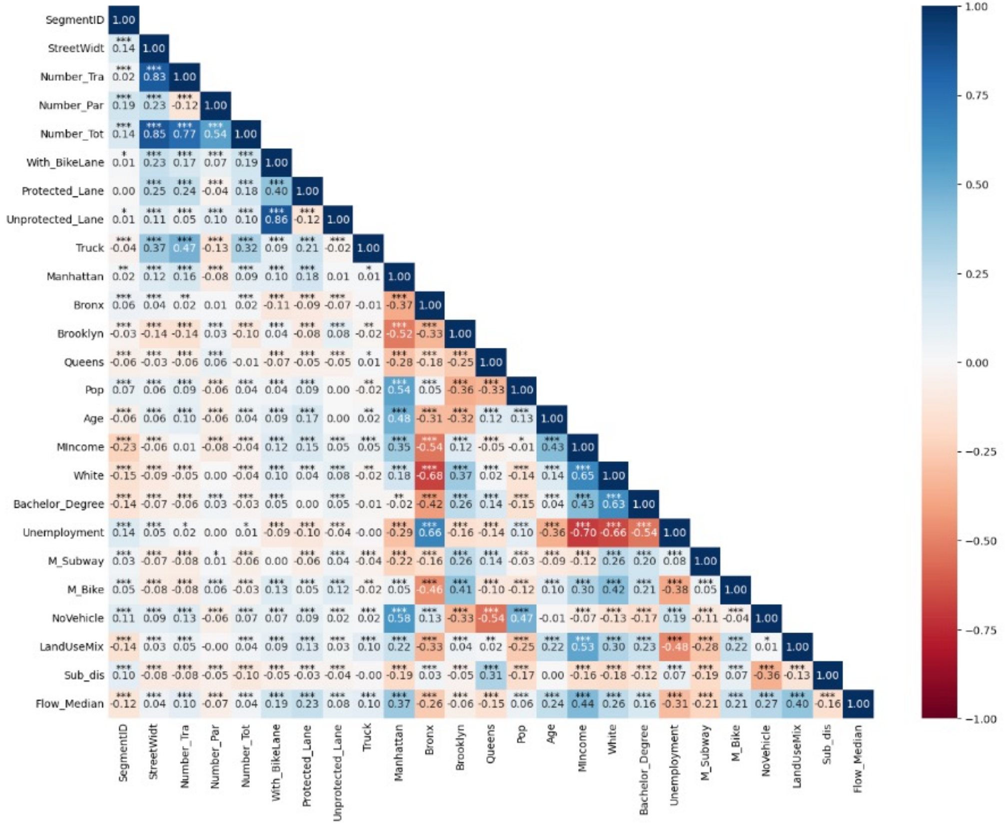
Demographics

- Population
- Median Age
- Median Income
- Percent White Residents
- Percent with Bachelor’s Degree
- Unemployment Rate
- Mode Choice Split: Subway
- Mode Choice Split: Bike
- Percent without Vehicle
- Land Use Mix

Pearson Correlation

P-Value:

- *   < 0.05
- **   < 0.01
- ***   < 0.001



Model

Model

Model	MAE	RMSE
Random Forest	54.5	111.2

Hyperparameters	Final Model
n_estimators	830
max_depth	28
min_samples_split	3
min_samples_leaf	2
max_features	7

Feature Impacts

Model	MAE	RMSE
Random Forest	54.5	111.2

Hyperparameters	Final Model
n_estimators	830
max_depth	28
min_samples_split	3
min_samples_leaf	2
max_features	7

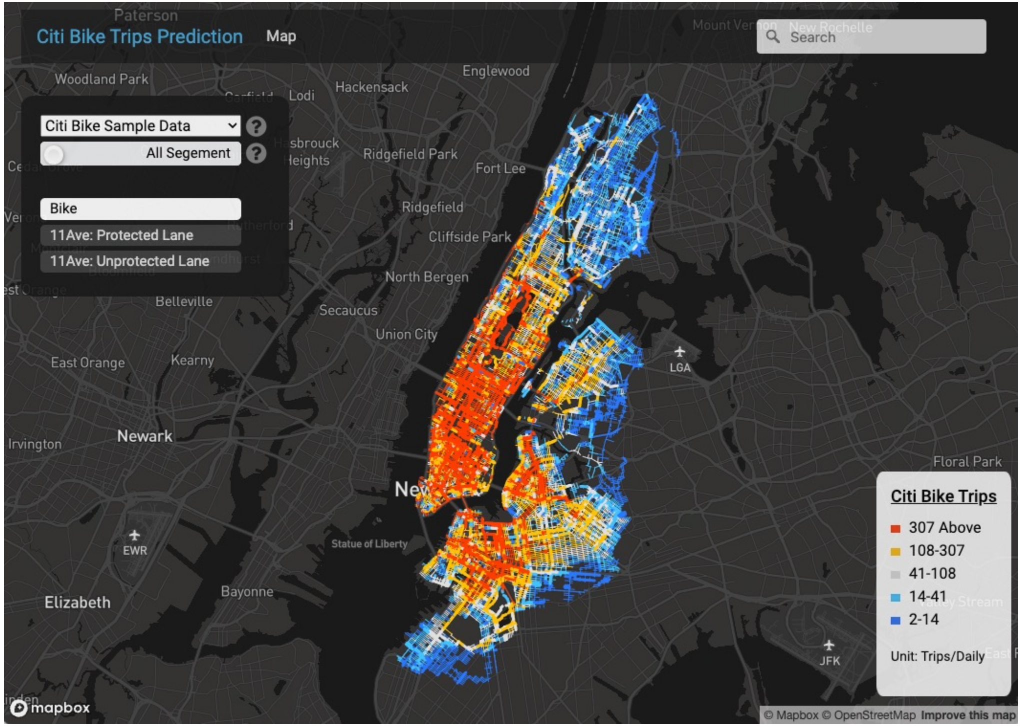
Outcome

NYU

Citi Bike Trips

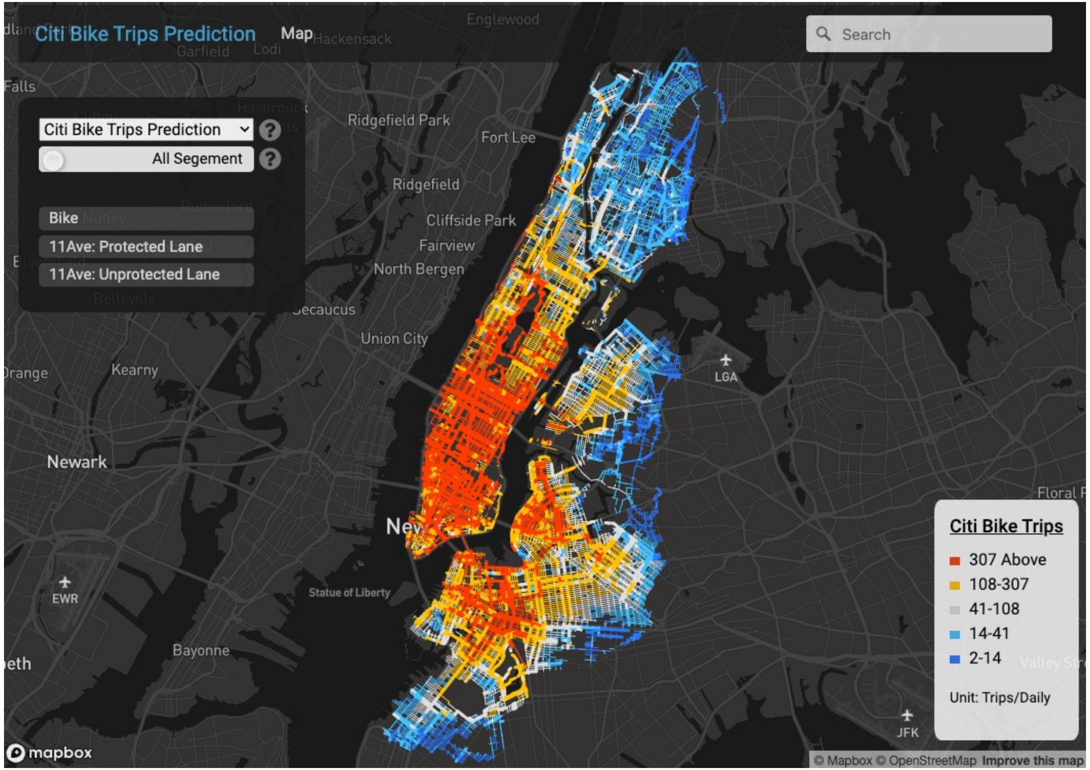
This heat map shows the original Citi Bike Trips in each segment.

[Interactive Map]



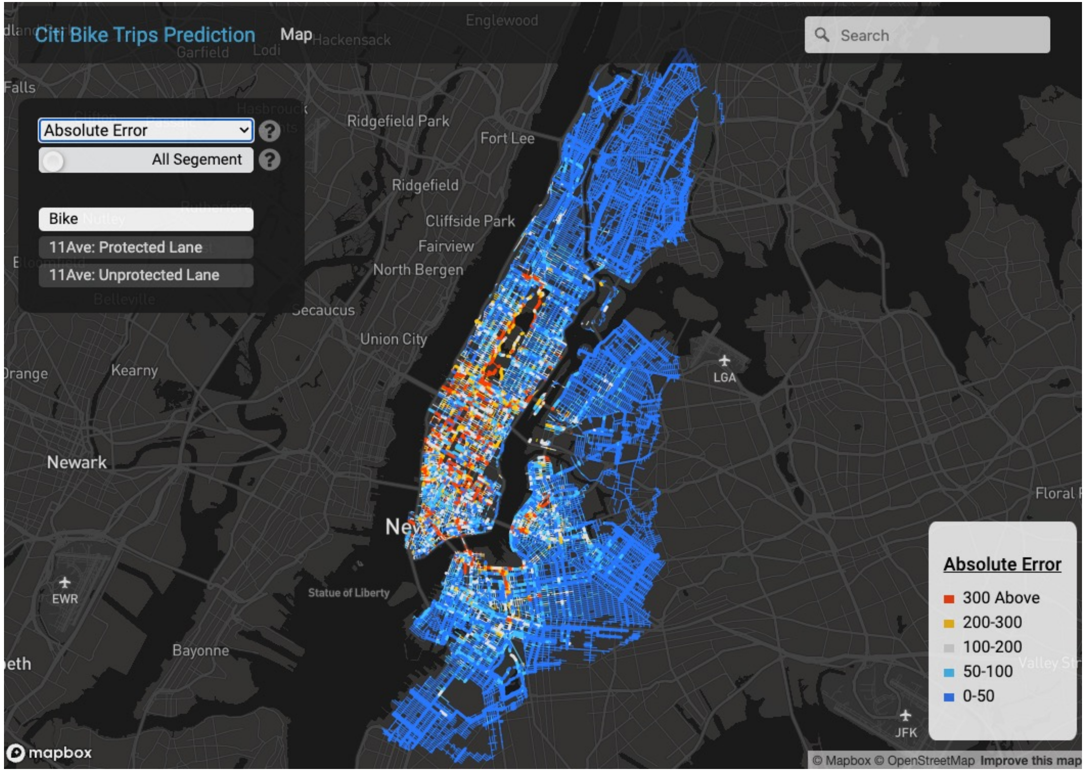
Prediction Trips

From the dropdown, we could select to show prediction trips.



Absolute Error

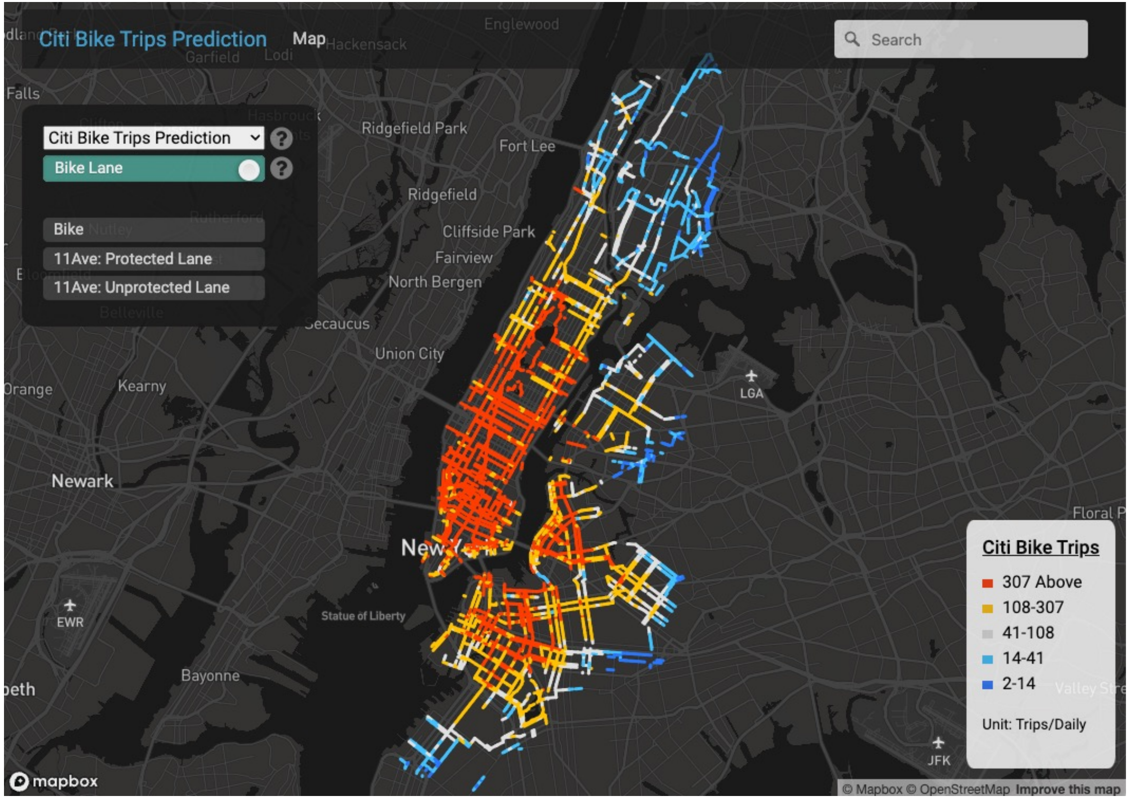
We can see that Manhattan trips are overestimated by the model.



NYU

Bike Lane

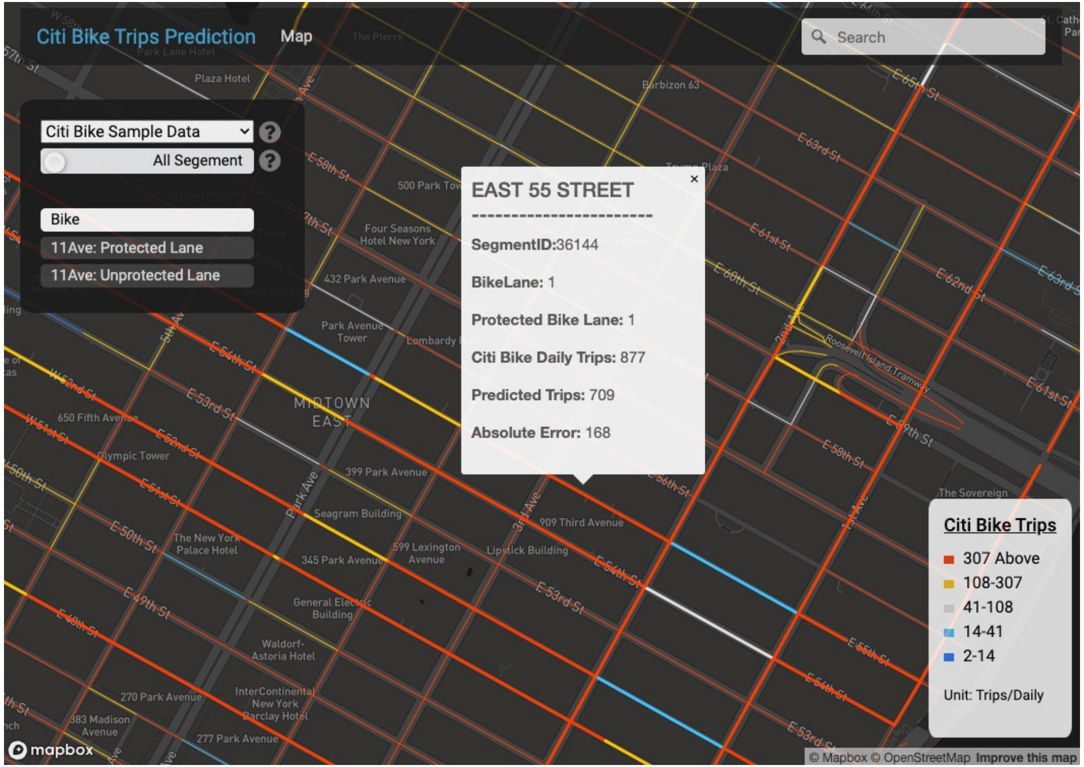
Could only show segments with bike lanes.



NYU

Pop-up

Click the segment to show the detail information.

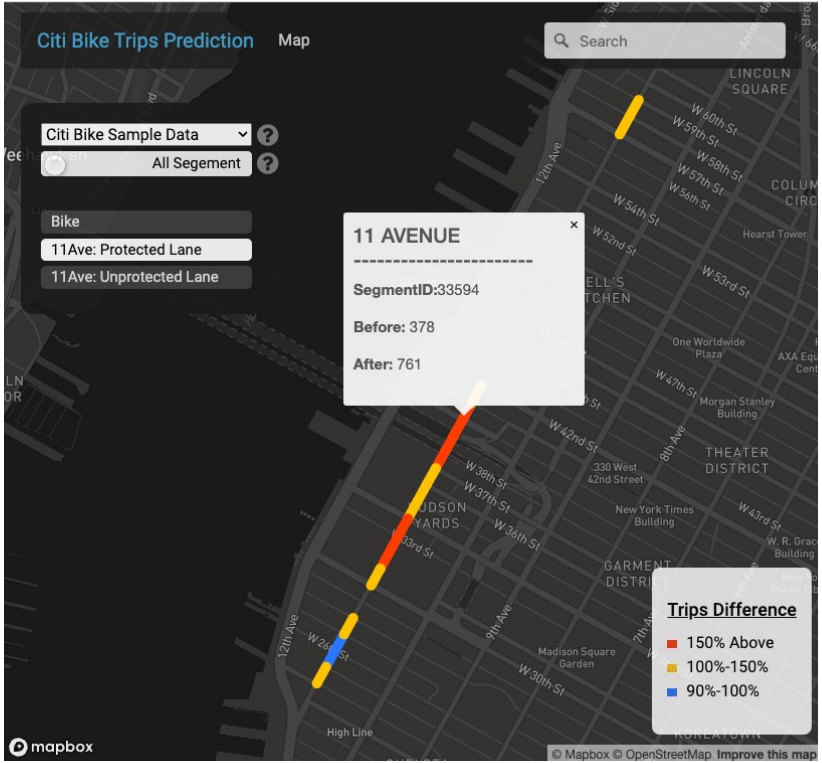


NYU

Case Study 1

We took the segments of **11 Ave** that without bike lane as study area.

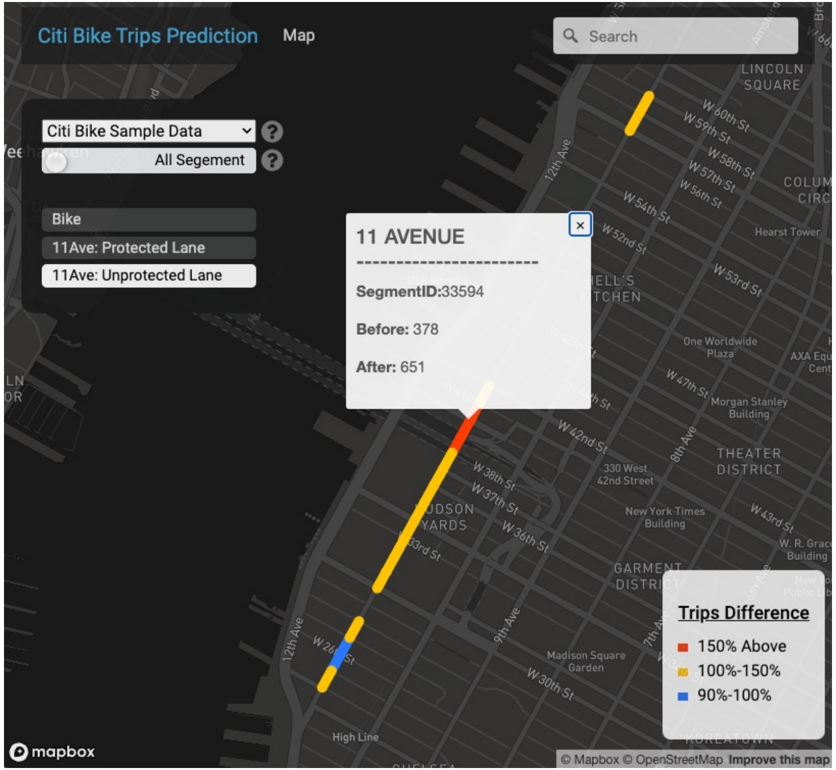
This map shows after installed **protected bike lane** on all of segments, how much bike trips will change comparing with before prediction trips.



NYU

Case Study 2

This time shows after installed **unprotected bike lane** on all of segments, how much bike trips will change comparing with before prediction trips.



NYU

Discussion

NYU

Limitation

- The routes generate using OSMnx may not be the optimal method, Google API and HERE API have better routing service, but the problem is too costly.
- Due to the limited time, some features may increase model performance did not included, like Distance to Attractions, Numbers of Jobs, and How important the segments in the whole network (could use PageRank algorithm to estimate)
- This project did not consider the direction of the link flow. In the future, the direction of link flow can be taken into account and modeled using GNN. This approach will give us a better understanding of which side the bike lane should be installed on.