proposal

Team

11/12/2021

The tentative project title

The group members (names and UNIs)

Motivation Have you ever been stuck in the NYC subway with no signal on? In a car that does not move at all? Or in a bus that is never on time? According to the INRIX scorecard, NYC was ranked No. 1 for the worst traffic in the United States, and the NYC subway on time performance has always been an issue as well. In this project, we were interested in whether bicycles can outcompete other transportations in terms of travel time in Manhattan. We focused on the time in a day, weekday versus weekend and seasonal variation to investigate the best time-saving way to commute, if not bicycle.

The intended final products

To the end, we are expected to create several interactive dashboards of the relationship between the travel time differences and comparisons of the public traffic (citibike/bus/subway/taxi) within daytime across the Manhattan area in New York. Analyzing the factors that might influence the actual travel time could reflect the road congestion situation in New York. Furthermore, it could be used as a marketing tool to help people in need of public transport decide the most cost-effective way to travel. Using visualization, we are expected to obtain an effective mapping to show the public what vehicle to choose in a particular area and specific period. Meanwhile, it can be a practical analysis to explore how public transit changes from 2020 to 2021 after people have options to receive the COVID-19 vaccine.

The anticipated data sources

- Main dataset - NYC Taxi and Limousine Commission Trip Record Data: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
- Dataset on Citi Bike Trip Histories: https://ride.citibikenyc.com/system-data
- Neighborhood Tabulation Areas: https://data.cityofnewyork.us/City-Government/2010-Neighborhood-Tabulation-Areas-NTAs-/cpf4-rkhqs

The planned analyses / visualizations / coding challenges

Identify the original destination pairs around the city across different time periods. We can also divide four time periods for our study, and we can visualize transportation preferences on a map using a specific zip code. Random sampling can be applied since a large sample might be computationally time-consuming. Exogenous variables should be adjusted. We can set up exogenous variables accordingly. We can also add MTA attributes and bus station attributes. Panel mixed multinomial logit model is widely used for travel time. R package "mlogit" https://cran.r-project.org/web/packages/mlogit/vignettes/c5.mxl.html

Visualization: Barplots compare the travel time of trips by four means of transportation during different time periods of a day, weekend vs. weekday, seasonal variations, etc. Map of available routes and average travel time for each mean of transportation from one neighborhood to another in nyc

Code challenges: Need to properly divide data into subgroups based on different times of the day, day types, etc. Panel mixed multinomial logit model may be difficult to deploy.

The planned timeline

November 13: Create shared Github repository, submit project proposal November 13-20: Narrow dataset, decide data visualizations, data cleaning (Import and tidy NY citibike/subway/bus/taxi data, merge datasets, generate descriptives) November 21-27: Create Visualizations (interactive descriptives, interactive plots) November 28-December 4: Final integrate code, proofread and test for glitches, two-minute screencast December 5-December 10: Finishing report for detailed information of the projects. December 11: Final project due