

Use Numbers: Assignment 2 of 2
Determinants of Demand

Due September 23, 2023 (Lecture 5) before class to Google Drive

For this assignment, we look at data that helps us focus on demand. Generally, it is difficult to use data we observe to separate the influence of demand from supply. This is because what we observe is the equilibrium, and demand and supply both determine the equilibrium. To get around this problem, we are going to focus on a market where (over a short period) supply is relatively fixed: the New York City subway.¹

To be clear: if we compare two of the same days of the week (e.g., two Wednesdays) at the same time, differences in transit ridership should be due to demand, rather than supply factors. In other words, we're going to assume that the same number and size of trains are running, that the stations have the same capacity, and that all other relevant supply-side features are constant.

I have prepared transit data that allow you to make such comparisons. There are three separate datasets, as listed below, in csv (good for Excel) or rds (for R) format.

1. ridership for the entire system by day, selected days

- These data contain information for the week ending February 22, 2020, and the week ending March 28, 2020.
- csv format: [ridership_by_day_20230912.csv](#)
- rds format: [ridership_by_day_20230912.rds](#)

2. ridership by station and day, selected days

- csv format: [ridership_by_station_20230912.csv](#)
- rds format: [ridership_by_station_20230912.rds](#)

3. daily ridership for the 238th St. and 86th St. stations on the 1 Line, selected days

- csv format: [ridership_by_station_1lineonly_20230912.csv](#)
- rds format: [ridership_by_station_1lineonly_20230912.rds](#)

These data come from NY's [Metropolitan Transit Authority](#) (MTA) and I have pre-processed them for your ease of use.

You can find a map of the NYC subway system [here](#). This map shows subway stations with dots. Below the station name, the map notes to which line (e.g., 1/2/3) the station belongs. For example,

¹I wanted to do this for the DC Metro, but similar data are not available, to the best of my knowledge.

the 79th Street stop in Manhattan’s Upper West Side is on the 1 line.

Importantly, there are different points on the map that have the same station name. For example, there are multiple 96th Street stations: one on the 1, 2 or 3; one on the B and C lines; one on the 6 train, and one on the Q train. To differentiate between these 96th Street stations, you need to know which line the station is on. Our dataset has one variable for the station and another variable for the line to help you with this issue.

As with the previous assignment, Google and news archives should be sufficient to answer these questions. This is not a major research paper, so please scale your effort accordingly.

Also as in the previous assignment, you are welcome to discuss parts of this assignment with other students. However, any work you turn in must be your own and written in your own words.

To make graphs, you can use Excel, R, or the software of your choice. We can support technical questions in Excel or R.

1 Questions

1. Assume (and this is likely true) that the amount of transit supplied was about the same in both February and March 2020. Graph subway boardings for the days included in the dataset for these two months. (Use the “ridership for the entire system by day” dataset.) What demand feature explains the difference? How do you see this?

2. Even absent large-scale public health disasters, there is also variation in subway usage across stations. Use the “ridership by station and day” data and choose three stations (the data has all stations, but this is too many!). Make a graph of pre-pandemic daily ridership at those three stations that illustrates these other demand differences (pre-pandemic daily ridership has two separate weeks in September 2019 and February 2020; use whichever week you prefer). Explain whether differences in ridership across stations are due to transit demand, transit supply, or both.

3. Use the final dataset (“ridership on the 1 line”) to compare ridership at two stations on the 1 train – 238th St. and 86th Street – for the five weeks in September in the dataset.

(a) Which station has a higher level of ridership? Give two hypotheses for this difference. (Remember that they are on the same line, so supply should not be an explanation.)

(b) Which station has a larger change in ridership after the pandemic? Which features of demand explain this relative change (that is, bigger at one station than the other)?

4. Use the data provided to make a final graph that shows one more dimension of demand (the answers to questions 1, 2 and 3 are also displaying dimensions of demand) using these data. By “dimension of demand,” I mean some feature that drives demand. Return to our list from class about the determinants of demand as needed. Write a brief paragraph that discusses how your graph shows this additional element of demand.

2 How to turn it in

- Name the file “use_numbers_2of2_[lastname].pdf”. For example, my file would be “use_numbers_2of2_brook”
- Turn in via this [google survey](#).
- Make sure your name is at the top of the submission.
- You do not need to type your submission. Any **legible** submission in pdf format is ok. For example, you can write the problem set with hand-drawn graphs, take a picture, make a pdf, and submit the pdf.

3 Data

Daily Ridership

The daily ridership dataset has one row per day, for the weeks ending
Feb. 22, 2020
March 28, 2020
June 27, 2020
September 19, 2020
September 14, 2019
September 25, 2021
September 24, 2022
September 9, 2023

The daily ridership dataset has the following variables (columns in excel-speak).

Variable	Definition and Source
DATE	Date as text in format MM/DD/YYYY
date_format	Date in date format
total_boardings	Total number of entries into subway
year	Year associated with date
month	Month associated with date
day	Numeric day associated with date

Daily Ridership by Station

Both daily ridership by station datasets have the following variables (columns in excel-speak).

Variable	Definition and Source
DATE	Date as text in format MM/DD/YYYY
date_format	Date in date format
total_boardings	Total number of entries into subway
year	Year
month	Month
day	Numeric day associated with date
day_of_week	Day of the week (Monday, Tuesday, . . .)
STATION	Name of station (note that some names repeat across lines; you may also need LINENAME)
LINENAME	Name of subway line (note that some STATION names repeat across lines)

Please be sure to carefully read the information about stations and lines in the introduction to the assignment.