1. Research question: how quickly would the MTA need to plan and implement bus routes for them to be relevant and optimal, given the changing nature of traffic needs in NYC?
2. Why this question:
   1. 4-year planning process in [this doc](https://new.mta.info/document/101521).
   2. NYC is expected to add about 22 miles of bus routes this year and last – even though Eric Adams promised 150 miles: <https://www.nytimes.com/2023/08/17/nyregion/eric-adams-buses-nyc.html>
      1. Slowest busses in the nation
      2. New lanes tied up in politics
   3. At the same time, some progressive pundits have launched a broader critique of the slow pace of planning and implementation in liberal jurisdictions:
      1. NYC columnist Ezra Klein has written about, “Everything bagel liberalism” where “[government] tries to accomplish so much within a single project or policy that it ends up failing to accomplish anything at all…one problem liberals are facing at every level where they govern is that they often add too much. They do so with good intentions and then lament their poor results.”
      2. Matt Yglesias on what he calls ‘meetingism’: “We have to trust ourselves to be able to win enough of the argument enough of the time to make the changes we want to see in the country, even knowing that sometimes democratic politics will make the wrong choice. The alternative is endless rounds of stagnation and litigation, arbitrarily advantaging the status quo and whoever has the best lawyers.”
      3. These two writers have slightly different culprits but are both basically interested in progressive governments that build more quickly and dynamically, especially around green energy, infrastructure and housing.
      4. In addition to being a frustrated NYC resident/bus rider, I was also interested in the deleterious effects of a long bus route-planning process, which included 18 community meetings.
3. To tackle this problem, I used NYC’s public data on for-hire vehicles aka rideshares.
   1. 43.9% of Midtown Manhattan traffic is rideshare traffic. So this data can be used to gauge traffic and demand for travel from one area of the city to the next.
4. To process the data:
   1. Each month contains tens of millions of rows, each representing a ride from one taxi zone (about the size of a small NYC neighborhood) to another.
   2. However, I grouped the data by PULocationID, DOLocationID, and hour – reducing the rows and by 90% or more, then concatenated the data into groups of one or two years.
5. I then began EDA and learned some key insights that would guide my model:
   1. The largest proportion of trips are short – about a mile – and decline exponentially from there.
   2. Rideshare usage declined rapidly during the pandemic then slowly climbed back up to almost pre-pandemic levels.
   3. The pandemic changed who commuted into Manhattan during morning commuter hours. Manhattanites and members of adjacent/gentrified Brooklyn/Queens commuted less and outer boroughs commuted more. This aligned fairly precisely with voting patterns in the NYC mayoral election – demonstrating how traffic patterns reflect deeper divisions and changes in how New Yorkers live.
   4. That said, I had some sense that – despite changes – there was also quite a bit of stability in the underlying data. E.g., These are the most common commuter pickup zones for \_\_\_ Wiliamsburg – as you can see, it’s fairly steady. Though there are some key changes (Manhattan neighborhoods often declined) and it changed a lot in the pandemic before reverting.
      1. This was a pattern that would come up later.
6. Next, I built my model to evaluate our changing traffic grid.
   1. First, I used the Python library Geopandas to make a dataframe of taxi zones and adjacent taxi zones.
      1. I used the public spatial file and Tableau’s mapping capabilities to check that it worked.
      2. I also manually added bridges and tunnels.
   2. Second, I used the Python library networkx to create a network represented in this image.
   3. Third, I used the shortest path algorithm to find the best path from place to place in NYC while traveling through the least amount of adjacent traffic.
      1. This was not the model I needed to build for my project, but it was a cool application, aligned surprisingly well with Google maps, and demonstrated that this data and a well-designed graph algorithm reflected NYC’s traffic patterns.
   4. Fourth, I created a custom algorithm to find the path from point A to point B that met the following requirements: a. start point b. end point c. no more than ‘x’ number of zone traversed d. could theoretically replace the maximum amount of rideshare rides
      1. To do this, I used the ‘all simple paths’ algorithm, which tells you every path from point a to point b without going through the same node twice in your graph, given a cutoff for number of nodes.
      2. I created a custom weight by converting each route into a dictionary of tuples that represented potential commutes and the number of rideshare rides for those commutes.
         1. So a route from [A,B,C,D] would be converted to [(A,B),(A,C),(A,D),(B,C),(B,D),(C,D)] as a bus could potentially serve as the transit vehicle for all six of those commutes.
         2. The dictionary would look like this: {(A,B): [number of trips in a given time period],(A,C): [number of trips in a given time period],(A,D): [number of trips in a given time period],(B,C): [number of trips in a given time period],(B,D): [number of trips in a given time period],(C,D): [number of trips in a given time period]}
         3. I created a similar dictionary for total fare paid – which would account for the economic cost, rather than just traffic.
      3. I looked at 15 routes that the MTA has planned and ran the algorithm for each quarter in which we have data: 2017 Q3 – 2023 Q1 for trips 2019 Q1 for economic cost.
      4. For the maximum number of zones traversed: I used the shortest route \* 1.5. There’s no science here, but it gave the algorithm slack to change with traffic patterns.
      5. First I evaluated the results by visualizing them in Tableau and noting how routes changed over time.
      6. Then, I created a function to quantify this change by looking at percentage of ‘commutes’ (our tuples) covered in each quarter compared to the first quarter.
7. Conclusions:
   1. On the model:
      1. NYC’s traffic patterns are largely stable:
         1. While some routes shifted back and forth between two possible ‘best routes,’ 12 of 15 (80%) routes were >90% the same in at least one quarter of 2022 Q2 – 2023 Q1 as in 2017 Q3. EVEN THOUGH THERE WAS A PANDEMIC THAT SUPPOSEDLY CHANGED EVERYTHING.
         2. 6 routes where exactly the same in the final quarter as in the 2017 Q3.
         3. 2 routes changed significantly and stabilized: Bed\_Stuy\_to\_Sheepshead\_Bay and Flatbush\_to\_Downtown\_BK.
         4. So while shifting demand happens, the general rule of NYC’s traffic patterns in stability. And keep in mind: just because there’s a new ‘best route’ doesn’t mean the old route has fallen sharply – it just means another one is now better.
         5. This pattern was even more pronounced when weighting by economic cost (rideshare cost + tip)
      2. Often, the ‘best route’ changed dramatically during the pandemic then reverted to its pre-pandemic state.
   2. Takeaways for the MTA:
      1. While I’m still not a fan of long planning processes, NYC’s traffic patterns are clearly stable enough that we can take time to ensure a route is good and assume it will be relevant for a long time.
      2. By focusing on frequent buses that run through high-traffic corridors, you can reduce the maximum amount of traffic. Most people just want to travel a short distance in a busy area!
      3. Similarly, a huge amount of demand simply lies in circulating through relatively busy areas.
      4. Reducing trips and reducing economic waste on rideshare rides are not at odds: both algorithms created similar routes.