**Project Idea Brainstorm:**

**MTA Service Demand:** Build a model that predicts the demand for public transportation services (subway, buses) on different routes and times of day, helping authorities allocate resources more effectively.

**Scaling it down:**

* For the borough of Manhattan
* Timeframe: June 2022 - December 2022

**Hypothesis:**

* How does better total estimated travel times correlate to demand?
* Is there a correlation between road traffic and MTA bus service and if this correlates train demand (more train transfers, etc.)
* How does construction affect MTA service demand?
  + More construction = More delays/Less demand
  + If construction present on roads, bus is unreliable
  + If construction present in trains, a different service line may see an influx in demand

**Audience of your pitch:**

* Who are they?
  + Users of public transportation
  + Google Maps, Waze, Transit App, Subway Time (GPS app/companies)
  + MTA Apps
  + MTA HQ
* Why should they care?
  + Optimize traffic and travel times for users
  + Allocate resources more effectively
  + Provide more accurate statistics
    - Transit App – Only works if you have service
    - SubwayTime – Convenience, NearBy Lines

**What data are you going to be mining to prove your pitch?**

* Find a dataset with completed route times
* OMNY Data – When people pay for service
* Before construction, when traffic was normal – How many people were taking buses, what was the predicted total travel time?
* After construction, when traffic increased – was there a spike in data?
* Find a dataset where we have the time it takes for buses to complete their route

**Datasets**

Ridership Data - Only contains total riders for MTA Service

<https://data.ny.gov/Transportation/MTA-Daily-Ridership-Data-Beginning-2020/vxuj-8kew>

MTA Developer Data:

<http://web.mta.info/developers/developer-data-terms.html#data>

Unofficial MTA Transit Data:

<https://data.beta.nyc/dataset/unofficial-mta-transit-data-archive>

MTA Bus Customer Journey: Time data

<https://data.ny.gov/Transportation/MTA-Bus-Customer-Journey-Focused-Metrics-2017-2019/8mkn-d32t>

MTA Subway Wait Assessment

<https://catalog.data.gov/dataset/mta-subway-wait-assessment-beginning-2015>

MTA Bus Wait Assessment  
<https://data.ny.gov/Transportation/MTA-Bus-Wait-Assessment-Beginning-2020/swky-c3v4/data>

Street Construction Permits

<https://data.cityofnewyork.us/Transportation/Street-Construction-Permits-2022-Present-/tqtj-sjs8>

Summary Link:

<https://new.mta.info/agency/new-york-city-transit/subway-bus-ridership-2021>

**Notes from 08/14**

**MTA Bus Data**

* + Occupancy/Ridership Data
  + Routes
  + Time of Day
  + Aggregate demand

**Models**

* KNN
* Regression
* Clustering/Association

**Main Goal:** Understanding where we see the most ridership/demand for buses throughout NYC

Compare the routes and identify where the demands come from

“There are deadzone within the 5 boroughs, create an algorithm that helps identify which dead zones should be prioritized for mass transit. This usually gets funded by the govt, research more about ILX”

* Explore the deadzones
* Business understanding: Private buses may find this beneficial so that it may pick up the slack where the MTA lacks

KNN algorithm:

<https://www.geeksforgeeks.org/k-nearest-neighbor-algorithm-in-python/#>

<https://realpython.com/knn-python/>

Regression algorithm:

<https://realpython.com/linear-regression-in-python/>

<https://medium.com/@polanitzer/building-a-linear-regression-in-python-step-by-step-94caeb0a4f18>

Mta metrics

<https://metrics.mta.info/>