**Question/need:**

* What is the framing question of your analysis, or the purpose of the model/system you plan to build?

In this project, I wanted to help the city better advertise its electronic drop-off locations, particularly to the large crowds of subway traffic and the metro stations that are closest to the locations.

* Who benefits from exploring this question or building this model/system?

I am advising the city of New York as it is currently against the law to dispose of electronics in the trash, and this will also help the city be more environmentally friendly.

**Data Description:**

* What dataset(s) do you plan to use, and how will you obtain the data?

Along with the give MTA turnstile data, I will use data from the city of New York’s open database that provides the addresses, latitudes/longitudes, store names, and specific identification numbers for every electronics bin in the city. There is also a given map where each bin location is labelled and you can mouse over it for more information.

https://data.cityofnewyork.us/Environment/Electronics-Drop-Off-Locations-in-NYC-Map-/9wzr-uhik

* What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

An individual unit of analysis in this project will be the total volume of people entering or leaving at a particular metro station at a particular 4-hour timeslot, as these values will be the ones I am directly cleaning/analyzing. I will likely aggregate these values by the station and the day of the week. I will be comparing the addresses of the metro stations to those of the bins, but I will not be analyzing the location data, per se.

* If modeling, what will you predict as your target?

I don’t plan to model anything, I’m just trying to determine relationships between the busiest metro stations/those closest to the bins and the bin locations.

**Tools:**

* How do you intend to meet the tools requirement of the project?

Using SQL, I can either do a data join for the two tables to correlate the bin location(s) with the metro stations, or I can just correlate the metro station locations visually with the map (if that is too difficult). I also imagine that I will use at least four of the commands: group by, and / or / not, where, having, case, or create table. The data ingestion guide will be followed, and I imagine I will better understand how exactly to incorporate matplotlib/seaborn into this project once we learn them later this week. Data analysis in Pandas will also be done according to the skills we learned in our pandas lecture today; I will clean the data in pandas.

* Are you planning in advance to need or use additional tools beyond those required?

No.

**MVP Goal:**

* What would a [minimum viable product (MVP)](https://github.com/thisismetis/Metis_Fundamentals/blob/main/project_deliverable_templates/mvp.md) look like for this project?

An MVP for this project would be the analysis of one bin location in relation to one metro station, and being able to explain how much traffic that location could get/if that station would be the best station for the city to advertise at.