1. What is the problem you want to solve?

Develop an approach to calculate and report the levels of “underutilization” of female and minority groups at New York City Transit (NYCT).

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

The client is the Equal Employment Office (EEO) of New York City Transit (NYCT). When a subgroup is deemed to be underutilized, appropriate goals should be established to increase the number of females and/or minorities.

3. What data are you going to use for this? How will you acquire this data?

American Community Survey data provided by the United States Census Bureau, more specifically the population of the available workforce by Ethnicity, Gender and education attainment for individuals in New York City, as well as, EEO occupational group information for all NYCT employees by Ethnicity and Gender.

4. In brief, outline your approach to solving this problem (knowing that this might change later).

First, obtain ACS data via R studio and then import data on NYC’s available workforce by Ethnicity, Gender, and Education attainment. Next, import NYCT employee data by EEO category, Ethnicity, and Gender. Merge the two sets of data and calculate underutilization using the 80% Rule. Under the 80% rule, underutilization exists if the percentage of females and/or minorities in a particular job group is less than 80% of the final percentage.

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

A code that can be used to extract the data used to calculate and report the potential underutilization of females and minorities within NYCT. Moreover, a report which will display the subgroups that is underutilized.