The Oxford Summer School in Economic Networks
Group Project

The government of Colombia has just announced funding for a new metro station in Medellin, Colombia. Medellin is the country's second largest city, composed of nearly 4 million people across 10 municipalities and 66 communes (administrative region). Your task is to decide **in which commune the new stop should be built**. Commune leaders are already squabbling: they all have arguments as to why their commune would benefit the most from the new metro-stop. You should give a data-driven assessment of which of the 66 communes should be awarded the metro-stop (and why), and how it should be connected to the existing infrastructure.

Research Question:

The criterion you use to decide on the municipality is somewhat open-ended, but you should make an effective argument about why your stop will better connect a particular municipality to new opportunities.

We have provided an array of data sources that you can use to help make your recommendation. You may also add any additional datasets available online to help inform your decision.

You can assume that building a station anywhere costs the same, but the cost of rail-line is approximately proportional to the length of the line. Your answer should provide basic information regarding how the metro stop is connected to the current network (What line it is on, and which stop or stops it is adjacent to). It is acceptable to suggest building a metro-stop in a commune that already has a stop, if your analysis supports this decision.

Your group will have 10min to present your recommendation to the National Planning Department on Friday.

Data Sources:

- The Medellin metro system is comprised of nearly 100 stops across subway, tram, bus rapid transit (BRT), and cable cars (tramway). Daily ridership averages 1 000 000+. The current network is given in "*medellin_metro.csv*" each row represents a station, complete with its latitude and longitude coordinates, the commune it is in, and what line it is on (as well as the type of line). Transfer stations are indicated. Successive stops along the same line are successive in the file. For instance, Sabaneto and La Estrella are adjacent on Line 1, but La Estrella (Line A) and San Antonio (Line B) are not.
- The file "*comunas_communtes.csv*" contains the relative number of trips between each commune and has been constructed from call data records. For

each row, there are three columns - the first gives the origin (an index), the second gives the destination (another index), and the final column gives the relative number of trips. The file `*comuna.csv*' matches the index with the corresponding communes.

- "*comunas_gmaps.csv*" contains the driving times, transit times and walking times (in seconds) between the different communes which comes from Google Maps. Edge (i,j) is how long it takes to commute from commune i to j (at 9am in the morning during the week).
- "*comunas_pop.csv*" gives the average population in 2020 within each commune.
- "*comunas_firms.csv*" gives the predicted number of commercial firms that are visible in Google Street Imagery (both formal and informal) in each commune.
- "*comunas_stratum.csv*" contains the average stratum of the neighborhoods in each commune. Stratum is socio-economic indicator that is based on the quality of housing and infrastructure. The highest socio-economic scale is 6 (wealthy areas), and the lowest 1 (poor areas).
- Optional: We have also included the file "*comunas_boundaries*" which contain the shapefiles of the various communes. This file labels communes in Macrozona numbers, and it can be linked to the commune indexes of the rest of the data in the "comuna.csv" file.