Revelio Labs Modeling Assignment

Metropolitan Statistical Areas (MSAs) define a convenient grouping of a large city with its nearby cities and townships, comprising an effective working population. MSAs are a useful tool by which to evaluate aggregate populations of workers, perform site selection analysis, and to evaluate a company's current geographic reach. Just as important is the ability to identify areas that are non-metropolitan; in other words, rural areas that do not associate with an MSA.

Traditionally, MSAs have been defined manually by government entities, but this approach has many limitations. Our main concern is that MSAs are only available within the United States, which limits their usefulness when evaluating international companies and workforces.

Provided to you along with this assignment is a downloadable CSV file (link available here) that lists all US cities, along with their geo-locations and population estimates. In addition, you have been provided with a noisy match to the standard MSA groupings within the US.

For this assignment, you are tasked with the following:

1. Please develop a model / algorithm to cluster cities within the US into MSAs, and provide useful labels. As part of this task, you will likely develop MSA clusters that do not exactly align with the standard MSA groupings. Can you give a reasonable explanation as to why your approach is better and more useful than the standard MSAs?
2. Apply your algorithm to cities outside of the US in order to create international MSAs.
3. Pay close attention to non-metro areas. These should be included as separate entities. For instance, rural upstate New York does not belong in a metropolitan area.

Keep in mind that there is ***no single correct approach***. The quality of your categorization depends on how useful it will be to practitioners.

Please provide us with the following:

1. A jupyter notebook, python file, or other piece of runnable code that takes in the csv file given and outputs a csv file with the following columns:
    a. City
    b. State
    c. Latitude
    d. Longitude
    e. Population
    f. Traditional MSA
    g. Your MSA
2. Your final csv file that was created as the output of your model