**PROJECT FOR MODULE 4 (DATA MINING AND DATA ETHICS)**

**[75 Points in Total]**

**TERRITORY OPTIMIZATION**

You are a regional sales director for Argentis, a growing financial services firm that specializes in providing investment services (annuities, managing stock portfolios, trust services) for higher net worth clients. Until recently, Argentis has only offered its services online, and you have been asked to help plan its office locations in central Ohio. Your regional president has asked you to put together a plan that (1) places eight agent offices throughout central Ohio and (2) minimizes the amount of time your Top 100 highest net worth clients in the area will have to travel to get to their local agent’s office. You have been given a file that contains the latitude and longitude of these clients’ addresses.[[1]](#footnote-1)

1. **(12 Points)** What analytical process will you use to identify the new agent’s locations? Assuming that your regional president and the agents you will be hiring are unfamiliar with this process, how would you describe it to them?
2. **(14 Points)** Read your geographic data into SAS Studio and run the appropriate algorithm on your data. Paste in the output of your results below. Be sure to include the latitude and longitude locations of the eight new office locations and how many of the Top 100 clients will be served by each location.
3. **(14 Points)** Create a map of central Ohio that shows the location of the eight new agency offices. (You can use whatever software you like. Many free, professional tools are available online: Google Maps, Tableau, Mapline.com, etc.) Write a short paragraph that you might include in a presentation to your colleagues explaining how these locations meet your regional president’s requirements.
4. **(14 Points)** Create appropriate data visualizations to show the distribution of clients’ marital status and investment potential scores for the clients assigned to each agency location.[[2]](#footnote-2)

Do any agency locations seem to have an unfair advantage or disadvantage?

1. **(21 Points)** Several months after your regional president has approved your plan, she pulls you aside and tells you that unfortunately the company has decided to only open seven new central Ohio locations instead of eight.
2. Run your analysis again but now with only seven locations.
3. Provide a map of the revised agency locations.
4. Write a brief paragraph summarizing how the locations have changed. (You do not need to repeat the data visualizations from question 4.) By reducing the number of agency locations by one, how much has the average client-to-agency distance increased?

1. This data file has been randomly generated, so any specific locations contained in the data are purely coincidental. [↑](#footnote-ref-1)
2. Marital status codes: S = single, never married; M = married; D = divorced; W = widowed. Investment potential: a model for Top100 clients (based on demographics) that assigns a 1-5 score based on their potential for having a large amount of investment capital. 1= lowest potential and 5 = highest potential. [↑](#footnote-ref-2)