Data Science Assessment Test

You are one of the Data Scientists of the Marketing Science team of a major media company. One of your clients is a medium-sized aged care provider that operates nationally in Australia. The client has been in the business for more than 10 years and has a pretty good reputation, however they are looking at increasing their market share in Australia over the next few years. They get in touch with a number of investors, who could potentially fund the creation of 200 new aged care facilities across the country. The client needs to prepare a detailed project plan to secure the investment, and they get in touch with you to get some help on this. In particular, they ask you to run an analysis to determine the best possible geographical location and capacity of their 200 new aged care centres. You gladly accept the task and immediately ask if they can share with you their first party data, but unfortunately they are unable to access their own data in a short time because they currently have no resource looking after their internal database. Their request is pretty urgent though, so you will need to rely only on publicly available datasets.

Luckily, the Australian Government maintains a list of aged care providers which is available at this web address:

https://www.gen-agedcaredata.gov.au/Resources/Access-data/2017/October/2017\_Aged\_Care\_Services\_List

You also remember that the ABS (Australian Bureau of Statistics) has recently published the aggregated results of the 2016 census. Information about number of residents by postal code or by age group might be useful for your project. The data are available here:

<http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20TableBuilder>

Last, but not least, you can search the internet for any other data sources that might help you come up with a very original and robust analysis, and you can definitely consult with your colleagues.

Considering all of the above, your task for this exercise consists of the following 5 points:

1. Formulate one or more hypotheses to identify the best geographical location and capacity (max number of people that the facility can serve at the same time) for 200 new aged care centres in Australia.
2. Gather the data (publicly available on the web) necessary for your analysis. You might also need to do some data cleaning and merging.
3. Implement an analysis to test your hypothesis(es). You are free to use any programming language or data analysis tool from the open source stack (preferably R or Python).
4. Visualize results, find insights, and draw conclusions from your analysis. If your response is in static form, please provide it in PDF or Microsoft Office compatible format. Dynamic/interactive visualisations should be web based or in Microsoft Office compatible format. OMG uses both Internet Explorer (v11) and Chrome and Microsoft Office 2016. Please ensure no installation is required, as executables cannot easily be installed on OMG computers.
5. Include your results into a presentation. We will ask you to present your findings in 15 minutes over the job interview, the same way you would present them to your client.
6. The first slide of the presentation should immediately and clearly state your findings.
7. You may explain your methodology in a *non-technical* way in the body of the presentation.
8. Please, explain your methodology in a *technical* way into an appendix of the presentation, as you would present to your boss or data scienstist colleagues. We will ask you to run us through your technical methodology over the job interview.