September 2022

# Take-Home Data Challenge

**Description**

N1 Health's client, a national Medicare Advantage plan, is seeking to understand how best to use its resources to address food access challenges in its membership. N1 Health has been tasked with presenting a short analysis (~5-10 minutes) of publicly available data to the Chief Medical Officer to answer this question. The guiding questions are:

1. Where should we deploy a food access program?
2. Which segment of the population might benefit the most from the program?
3. What do we know about the potential impact of the food program on medical costs or utilization?

**Data**

Please use two of the following three datasets as the primary source for analysis:

- CMS – State/County Medicare Utilization Summary - https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-geographic-comparisons/medicare-geographic-variation-by-national-state-county
- CDC - 500 Cities Project - https://chronicdata.cdc.gov/500-Cities/500-Cities-Census-Tract-level-Data-GIS-Friendly-Fo/k86t-wghb
- FDA – Food Atlas - https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/

**What we're evaluating for:**

- Can you take novel, messy datasets and produce a compelling analytical story from them?
- Can you structure your code in a way that will be extensible and re-usable in future analyses?
- Can you build an appropriate predictive model for the context given?
- Can you speak to how this model/analysis might be deployed in production?

**Some more specific guidelines:**

1. Please build one statistical model as part of this analysis, and then tell us how well it did. You'll be evaluated only on your ability to build and understand the tradeoffs of the model, not on the performance of it, so don't spend much time tuning it.
2. Please submit your code and visualizations in one of these forms:
    1. Jupyter Notebook
    2. RMarkdown file
    3. Github repository link
    4. PDF of results + code used for analysis
3. If an environment is necessary to run your code, please include relevant configuration files (e.g. requirements.txt) with your submission.
4. Please create 2-3 key data visualizations to answer the 3 questions detailed in the description. More visualizations are fine, but think of them as an appendix rather than a part of the deliverable.
5. You'll be given a chance to present the analysis in person, and you will be asked questions about both the results and the code.
6. This project is expected to take between 4-6 hours of your time spread out over a week. The questions are open-ended, so there is room to spend far more than that. We are evaluating for your instincts and thought process in this work, not for the time you spent, so if you find yourself spending more than this, we recommend stopping work on it and simply describing your suggested approach for future work during the presentation.
7. If any questions come up, please feel free to ask them ahead of delivering a result.