

Appendix A - Statement of Work

Background

The opioid crisis has become a significant public health issue with devastating consequences for those afflicted. Overprescribed medical opioids such as fentanyl, oxycodone, and codeine are a few of the common drugs that are causing this problem, often leading to addiction, misuse and abuse, with many users turning to cheaper alternatives such as heroin. Although Medically-Assisted Treatment (MAT) for patients afflicted with opioid addictions has shown to be effective in relieving their need for these drugs, the number of incidences related to opioid misuse continues to rise steadily.

Tackling the growing problem has become a priority for many organizations around the nation. Armed with our team of industry experts, years of problem-solving experience, and wealth and breadth of knowledge, Deloitte is well-positioned to positively influence interventions and outcomes. In the past, we have published several examples of thought leadership, including a paper that analyzes the opioid epidemic using an ecosystem framework, and another that explores how data analytics can help healthcare institutions better tackle the problem. We recognize that different lenses and perspectives are essential to fully understanding and tackling such a complex issue. We would like to invite the students at the Northwestern University MS AI program to join us in using skills in machine learning, predictive analytics and AI to more comprehensively understand the nature of this multi-factorial problem and determine potential responses to the crisis.

Through analyzing various sources of publicly available data, the students are expected to determine the key drivers of the growing epidemic and recommend potential actions to prevent and reduce opioid related overdoses and deaths while improving patients' access to necessary treatments.

Description and Objectives

Part I (Required)

The project team will first perform data wrangling and conduct exploratory data analysis (EDA) to determine the key characteristics of the data. Then the team will apply machine learning algorithms (k-means, SVD, k-nearest neighbors, self-organizing maps etc.) to identify unique clusters of U.S. counties or CBSAs that share similar attributes.

After the model is selected and validated, the team will then develop descriptive and predictive profiles of market segments that tell the story of what each segment looks like in terms of market environment and pain management dynamics, as well as risk factors.

Finally, the team will also provide recommendations from market research and insight as to what suggested actions could be taken based on the outcome of your analysis.

Throughout this process, keep in mind the following guiding questions, and make sure to touch upon them in your analysis and when you present your results:

- a. Which socio-economic, demographic, educational, environmental, political factors define a cluster and its risk profile?
- b. Which risk factors are driving opioid usage?
- c. What factors are mitigating factors (e.g. is access to treatment clinics significant)?
- d. Are there anomalies in micro markets and what are possible driving factors for the outliers?
- e. What are the effects of confounding factors?
- f. How do abuse, misuse, and diversion likelihoods differ for sub-populations?
- g. How can health services and intervention programs be optimally located?
- h. Do state policies have an effect on outcomes (e.g. Medicaid expansion)?

Part II (Optional)

From the analysis done in Part 1, the students will have a chance to understand the opioid crisis at a national, state, and county/CBSA level. An optional, open-ended Part II is to encourage them to dig deeper with questions in mind into 1-2 geographies to help address the opioid crisis in these locations.

Various state agencies publish socioeconomic information and more granular data about opioid misuse and abuse that could be analyzed for further insights and suggestions. For instance, since the Illinois Department of Human Services, Division of Substance Use Prevention and Recovery is actively working to curb the opioid crisis, students should look for data sources provided by this organization if they were to pick Illinois as a geographical location on which to focus.

There is also a strong interest in identifying non-traditional (non-governmental) public data sources using advanced AI technologies to supplement state and county level information. For example, using satellite imagery and/or unstructured text to identify and label unique features at the zip code or neighborhood level could help create more granular recommendations and interventions within counties rather than across counties.

The recommendations and interventions can range from changing opioid prescription limits, to increasing social service support for people who lose their homes and jobs due to opioid addiction, to increasing prescriber education or helping patients find a physician more easily. Note that the responsibility falls upon the project team to prospect and obtain permission for any additional source of data used that is not provided by the practicum sponsor.

Deliverables

This project is distinctive in its requirement of a full range of multidisciplinary skills and seeks to replicate the experience of managing a real-life data science project. The goal is for students to not only practice their data science skills, but also to experience the full extent of a project life cycle outside of the classroom. The team is expected to employ creativity, ingenuity and critical thinking throughout every step of the project process to create the following deliverables:

1. A file containing well-documented source code of the modelling process (including data wrangling/cleaning, EDA, selection, fitting, and validation) written in the language of the project team's choice.
2. A written report explaining model findings and results, analyzing the micro markets and sharing other insights. The report should expound upon the existence (or lack of) biases in the underlying data/model, interpret relevant factors and results, and provide a set of recommendations on how to best communicate model findings and act upon its results and indications. It could also include:
 - a. Descriptive and predictive sub-national (zip3 and zip5) insights on various dimensions of the problem
 - b. Quantitative and qualitative profile information about different local markets
 - c. Clusters and cluster predictions for local markets
 - d. Supporting tables and graphics
3. An interactive dashboard to help visualize results, built using a tool like Tableau, Shiny, Bokeh, etc. Here are some suggestions of what could be included:
 - a. Executive views of sub-national maps and drill-downs on individual micro markets
 - b. Heat maps showing relevant metrics by geography; multiple data layers able to be toggled
 - c. A dossier that highlights key findings specific to a micro market as shown in the tool

Data Description

To support analysis and model building, Deloitte will provide the project team with up to 20 data sets from various open source data providers, such as: Substance Abuse and Mental Health Services Administration (SAMHSA), Centers for Disease Control (CDC), and Centers for Medicare & Medicaid Services (CMS). A detailed catalogue of the collected data sets as well as a data dictionary will be provided to the project team.

Most of the data sets are collected at the state/county level; however, some data sources contain data at zip code and address level. Deciding level of granularity for analysis will be left to the discretion of the project team and the team should leverage their skills to 'wrangle' and 'shape' the data into the desired format and granularity.

While up to 20 data sets will be provided to the project team, it is encouraged that the team leverage additional data sources which they believe will improve the granularity of the analysis and/or model accuracy (within the bounds of legal & ethical constraints). The responsibility falls upon the project team to obtain necessary permission(s) for the use of additional data sources outside of the data sets provided by Deloitte.