# **Public Policy Question Assignment**

# Introduction

For this assignment, you will write a proposal synthesizing the information that you have learned to design an analysis to support data-driven decision making in public policy. In this assignment, you will not be conducting the analysis, but you will be designing your "ideal" plan. Propose a question that is relevant for local, state, or federal government, identify the information needed (data source or data sources) to conduct an analysis, and propose the analytical strategy to be used. Choose the model covered during the semester that best fits your analysis (a decision tree, hierarchical cluster analysis, or k-means cluster analysis). You must choose a data source *other* than the data that we have used during the semester.

## Directions

Any of the lecture and reading material for the semester may be useful for the assignment, but several readings in particular will be good resources. This includes *Practical Data Science with R*, Chapter 1, "Chapter 3: What is the Question?" *Research Methods for Public Administrators*, "Brilliant Analytics for Smart Cities," *A Practical Guide to Analytics for Governments: Using Big Data for Good*, and *Practical Data Science with R*, Chapter 5, Section 5.1, Chapter 11.

The completed assignment should adhere to the following guidelines:

* 1. Include your answers on the assignment document.
  2. Write your answers using complete sentences with correct punctuation, grammar, and spelling.
  3. Submit your completed assignment through the Blackboard portal in Lesson 3. It may be submitted early, if desired.

Answer the following questions:

1. What is the public policy question? (2 points)

In 2018 the CDC released an observation stating that for the first time since 1990 we as a nation had drop in annual overdose-related deaths. Should I advise my state to reduce its funding to our intervention-programs? If so, to what degree?

<https://www.nytimes.com/interactive/2019/07/17/upshot/drug-overdose-deaths-fall.html>

1. What are the data sources that you would use to conduct your analysis? Identify the source of the data (the organization that produced the data and where the data are found). Describe the characteristics of the data including the unit of analysis, sample, and variables. (2 points)

CDC Overdose Data - <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm>

1. Count of Overall Deaths
2. Demographics – Age, Gender, Race, Socioeconomic Status.
3. Comorbidities

NIDA State Data - <https://www.drugabuse.gov/drugs-abuse/opioids/opioid-summaries-by-state>

1. Count of Overall Deaths
2. Demographics – Age, Gender, Race, Socioeconomic Status.
3. Comorbidities

Proprietary Claims Data – State’s DSHS dept

1. Cost of drug-abuse programs
2. Count of Drug abuse claims/services used
3. Which machine learning model would you use and why? Describe the stages of your analysis from the original assessment of the data through the analysis, and the subsequent model diagnostics. (4 points)

Initially I would confirm CDCs notion that there is a reduction in my state. Depending on the results of that, assuming we follow the national trend, a drill down into the services that make up our drug-intervention program would ensue.

Within that drill down I’d summarize utilization of our services, cost of our services and effectiveness of our services. Further, via unsupervised k-means clustering of claims data, I’d explore/identify novel clusters with the hope that I find location-based (city vs rural) differences in claims.

1. What information would your analysis produce that a policy maker could use to make decisions about the public policy question that you have identified? (2 points)

In an ideal scenario, I’d have basic values of avg cost-per-individual utilizing state/federally funded services, cost per taxpayer, utilization of our current programs, and an impact analysis of reducing services identified in our clusters.

## Scoring

The assignment is worth 10 points. Points will be based on the real-world usefulness of the question, the appropriateness of the analytic strategy for the question, and the thoroughness and attention to detail of the answers.