**Tips for DAP’s Submission #2 – Using Secondary Data**

If you plan on using publicly available data, you must follow the following criteria to meet the assignment expectations:

**Relevance to Research Question**:

* Ensure that the dataset is appropriate for answering the research problem or question outlined in the project.
* The data must allow for meaningful statistical analysis related to the research question.

**New Analysis**:

* Conduct a new, original analysis using the dataset.
* Avoid merely replicating existing studies or analyses performed with the same data.
* Clearly explain how your analysis differs from previous analyses done with the same data.
* Provide justification for what new insights your analysis will bring compared to existing research using the dataset.

**Data Source**:

* The dataset should come from reliable and credible sources.
* If you are using a publicly available dataset, you must use the raw dataset file to carry out the analysis. Acceptable sources include publicly available databases such as (but not limited to):
  + Pew Research
  + American National Election Studies (ANES)
  + World Bank datasets
  + General Social Survey (GSS)
  + Kaggle
  + Roper Center Public Opinion Archives
  + Gallup Public Dataset Collection
  + European Social Survey (ESS)
  + Data.gov
  + World Values Survey (WVS)
* Always verify the legitimacy and credibility of the data source before using it.

**Dataset Description**:

* Provide a detailed overview of the dataset including:
  + Source of the dataset (e.g., publicly available database, survey)
  + Number of observations (sample size)
  + Time period the data covers
  + Types of variables included (categorical, continuous, etc.)
* Ensure that the dataset contains all the necessary variables to test your hypotheses.