**Objective**

Identify potential patterns in open missing people cases and their geographic and demographic attributes

**Approach**

* Data Collection - Identify public data sources for:
  + Missing persons open cases
  + Demographic information
* Exploratory Data Analysis
  + Study the data acquired for quality
    - Accuracy
    - Completeness
* Data Profiling
  + Ensure all required pieces of data are available for the data set such as
    - Missing persons Name, Missing Date, Location
    - Bonus – possible type (abducted, fled, Other?*)*
  + Ensure data fill rate
    - At least 90% of the data attributes need to be populated
* Questions to answer
  + Is there a relationship between the open missing cases in terms of
    - Ethnicity?
    - Gender
    - Income level?
    - How they went missing?

**Overview**

1. Compare missing people by state
   1. Us map with states and bubble sizes representing the missing people
2. Drill through to county level by demographics
   1. Representative states – *CA, IL, NJ, TX, NY, VA?*
   2. Representative Demographics – Gender, Income Level, Ethnicity
3. Strategy:
   1. Missing people Vs State **Population** Sample T Test
      1. Top 5 states T test
      2. Bottom 5 states T test
   2. The most missing cases by
      1. Gender
      2. Case Type
      3. Education Level
      4. Working Vs. Non Working Women
      5. Combinations
         1. Gender + Case Type
         2. Gender + Education Level
         3. *Other?*