

## BigQuery Data Exploration Part 1

Step 1 - While exploring the dataset, look at the Schema tab and read through the descriptions provided for each column, to get a sense of what data is in this table. This will help you answer the questions below.

Step 2 - Answer the following questions

1. There are several acronyms used in this dataset. For each acronym below, write out what it stands for:
  - a. CoC = Continuum of Care Area
  - b. Sheltered\_ES = Homeless in Emergency Shelters
  - c. Sheltered\_TH = Homeless in Transitional Housing
  - d. Sheltered\_SH = Homeless in Safe Haven Housing
2. What are the only 3 columns that are NOT an Integer type?
  - CoC\_Number
  - CoC\_Name
  - CoC\_Category
3. Is there any way to determine which state each row of data is located in?  
The CoC\_Number contains the state code as a part of its identification using the WHERE clause and LIKE to extract the state code from the number.
4. How many total rows of data are there? 2678
5. What might be some reasons that someone would use this dataset (no wrong answers here, just trying to think about how this data could be used)?
  - Out of the total homeless in a specific state, what is the percentage of them in the different types of shelters?
  - Which shelter has the highest number of occupants per state?
  - How many homeless are chronically homeless?
  - How many of the homeless are families per state?
  - Is the number of homeless on the rise?
  - What five states have the highest number of homeless and how many are chronically homeless versus transitional? Which five have the lowest?

Step 3 - While you have the dataset selected, click on the Query button, and choose "In New Tab". In the new query window, you can delete all of the SQL code there, and copy and paste the following code, then run it to create a new table:

```
CREATE TABLE Exploration_Project.homelessness AS
SELECT CoC_Number, LEFT(CoC_Number, 2) AS State, CoC_Name, Overall_Homeless,
Sheltered_ES_Homeless, Sheltered_TH_Homeless, Sheltered_SH_Homeless, Sheltered_Total_Homeless,
Unsheltered_Homeless, Homeless_Individuals, Homeless_People_in_Families, Chronically_Homeless,
Homeless_Veterans, Homeless_Unaccompanied_Youth_Under_18, Count_Year
FROM `bigquery-public-data.sdo_hud_pit_homelessness.hud_pit_by_coc`
```

6. Do a quick Google search to figure out what the SQL function LEFT( ) does. In the query we just ran, what does the line of code, LEFT(CoC\_Number, 2) AS State, do?

Pulls the 2 leftmost characters in the string and place them in a column labeled State

Step 4 - Open the new table you just created and use the Preview tab to look at the data and make sure all of the columns are appearing properly.