

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?
 - a. CoC_Number
 - b. CoC_Name
 - c. CoC_Category
3. Is there any way to determine which state each row of data is located in?
 - a. The state abbreviation in CoC_Number
4. How many total rows of data are there?
 - a. 2768
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)?
 - a. to use data as a tool to inform more effective, equitable, and sustainable solutions to homelessness.

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?
 - a. The state abbreviation in CoC_Number is moved to a new column with header 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.

BigQuery Data Exploration Part 2

Step 1 - Open the dataset you created last time, which should be a table named Homelessness. Once you have opened it, click on the Query button and open a new query in a new tab.

Step 2 - Using this dataset, answer the following questions. These first few questions can be answered by using ORDER BY and WHERE clauses. For reference, a list of all of the [state abbreviations is listed HERE](#).

1. We want to develop some new programs to help unaccompanied homeless children under 18 years old, and need some locations to start some programs. What are the top 3 Continuum of Care areas (CoC_Name in the table) with the highest number of unaccompanied homeless youth under 18 in the year 2018.

Answer:

Row	CoC_Name	State	Homeless_Unaccompan...
1	San Jose/Santa Clara City & County CoC	CA	506
2	Oregon Balance of State CoC	OR	243
3	Las Vegas/Clark County CoC	NV	214

SQL Code Used:

```
SELECT
  CoC_Name,
  State,
  Homeless_Unaccompanied_Youth_Under_18
FROM
  `merit-america-data-project-ew.Exploration_Project.homelessness`
WHERE
  Count_Year = 2018
ORDER BY
```

Homeless_Unaccompanied_Youth_Under_18 DESC
LIMIT 3;

2. We suspect that in Delaware (state abbreviation is “DE”), the number of unsheltered homeless people has been increasing over the past 7 years. Is this statement true? How do you know?

Answer: I ran the query and looked at the total_unsheltered numbers for each year in Delaware. When I reviewed the data year over year, I noticed a general upward trend—from earlier years to the most recent. This confirms that the suspicion is true: the number of unsheltered homeless people in Delaware has indeed been increasing over the past 7 years.

Row	Count_Year	total_unsheltered
1	2012	22
2	2013	10
3	2014	37
4	2015	37
5	2016	51
6	2017	58
7	2018	93

SQL Code Used:

```
SELECT
  Count_Year,
  SUM(Unsheltered_Homeless) AS total_unsheltered
FROM
  `merit-america-data-project-ew.Exploration_Project.homelessness`
WHERE
  State = 'DE'
GROUP BY
  Count_Year
ORDER BY
  Count_Year ASC;
```

3. The Safe Haven program was created in 1992 to provide shelter for people who are homeless and have a serious mental illness. However, funding for Safe Havens was cut in

2009, so no new Safe Havens could be created. Looking at data from only 2018, answer the following questions:

- a. In 2018, how many different locations had at least 1 person as a Sheltered_SH?

Answer:

Row	locations_with_s...
1	90

SQL Code Used:

```
SELECT
COUNT(DISTINCT CoC_Name) AS locations_with_safe_haven
FROM
`merit-america-data-project-ew.Exploration_Project.homelessness`
WHERE
Count_Year = 2018
AND Sheltered_SH_Homeless > 0;
```

- b. In 2018, what was the CoC_Name of the top 3 locations in terms of number of Sheltered_SH?

Answer:

Row	CoC_Name	State	Sheltered_SH_Ho...
1	Philadelphia CoC	PA	235
2	Reno, Sparks/Washoe County C...	NV	185
3	Indianapolis CoC	IN	68

SQL Code Used:

```
SELECT
CoC_Name,
State,
Sheltered_SH_Homeless
FROM
`merit-america-data-project-ew.Exploration_Project.homelessness`
WHERE
```

Count_Year = 2018

ORDER BY

Sheltered_SH_Homeless DESC

LIMIT 3;

4. What are the top 7 states in terms of Overall Homeless population in the year 2018? (Hint: This question will require the use of a GROUP BY clause in your query)

Answer:

Row	State	total_overall_ho...
1	CA	129972
2	NY	91897
3	FL	31030
4	TX	25310
5	WA	22304
6	MA	20068
7	OR	14476

SQL Code Used:

SELECT

State,

SUM(Overall_Homeless) AS total_overall_homeless

FROM

`merit-america-data-project-ew.Exploration_Project.homelessness`

WHERE

Count_Year = 2018

GROUP BY

State

ORDER BY

total_overall_homeless DESC

LIMIT 7;

5. [Below you will see a table](#) of all of the states and their populations in 2018. If the Overall Homeless population were only correlated with the total state population, then the top 7 states for total population and the top 7 states for Overall Homeless population would be the same.

- a. Is this the case? If not, which states are not lining up the same?

No, the top 7 states by overall homeless population do **not** match the top 7 states by total population in 2018. This suggests that factors beyond just population—such as housing costs, mental health resources, or climate—may contribute to higher homelessness rates.

- b. We would say that homelessness is overrepresented in a state if that state listed higher in the Homeless ranking than it did in total population ranking, and it would be underrepresented if the homeless ranking was lower than the population ranking. With that in mind, which states in the top 7 homelessness list would be overrepresented for homelessness?

The states that appear overrepresented in the top 7 for homelessness in 2018 are Washington (WA), Massachusetts (MA), and Oregon (OR). These states rank much lower in total population but are disproportionately high in their homeless populations. This suggests that factors like housing affordability, cost of living, or social services availability may play a significant role in these areas.

6. In order to create better policies to support homeless individuals, we want to study locations that are doing a good job providing shelter. To do so, we need to identify places that have a relatively large number of homeless, but a relatively small number of unsheltered homeless.

- a. Generate a list of all locations, in 2018, that have more than 1000 Overall Homeless, but less than 100 unsheltered homeless.

Answer:

Row	CoC_Name	State	Overall_Homeless	Unsheltered_Ho...
1	Nassau, Suffolk Counties CoC	NY	3868	52
2	Springfield/Hampden County C...	MA	3368	44
3	Maine Statewide CoC	ME	2516	98
4	Yonkers, Mount Vernon/Westch...	NY	1827	44
5	Cleveland/Cuyahoga County CoC	OH	1808	78
6	Omaha, Council Bluffs CoC	NE	1411	64
7	Cincinnati/Hamilton County CoC	OH	1114	47
8	Rhode Island Statewide CoC	RI	1101	51
9	Anchorage CoC	AK	1094	94
10	Quincy, Brockton, Weymouth, Pl...	MA	1090	22
11	Delaware Statewide CoC	DE	1082	93
12	Lynn CoC	MA	1061	42

SQL Code Used:

```
SELECT
  CoC_Name,
  State,
  Overall_Homeless,
  Unsheltered_Homeless
FROM
  `merit-america-data-project-ew.Exploration_Project.homelessness`
WHERE
  Count_Year = 2018
  AND Overall_Homeless > 1000
  AND Unsheltered_Homeless < 100
ORDER BY
  Overall_Homeless DESC;
```

- b. From that same list, in which locations do Unsheltered Homeless account for less than 2% of the Overall Homeless population?

Answer:

Row	CoC_Name	State	Overall_Homeless	Unsheltered_Ho...	unsheltered_perc...
1	Springfield/Hampden County C...	MA	3368	44	0.0131
2	Nassau, Suffolk Counties CoC	NY	3868	52	0.0134

SQL Code Used:

```

SELECT
  CoC_Name,
  State,
  Overall_Homeless,
  Unsheltered_Homeless,
  ROUND(Unsheltered_Homeless / Overall_Homeless, 4) AS unsheltered_percentage
FROM
  `merit-america-data-project-ew.Exploration_Project.homelessness`
WHERE
  Count_Year = 2018
  AND Overall_Homeless > 1000
  AND Unsheltered_Homeless < 100
  AND (Unsheltered_Homeless / Overall_Homeless) < 0.02
ORDER BY
  unsheltered_percentage ASC;

```

US State and Territory* Population Ranking for 2018

Rank	State	Population	Rank	State	Population
1	California (CA)	39776830	29	Connecticut (CT)	3588683
2	Texas (TX)	28704330	30	Puerto Rico (PR)	3193354
3	Florida (FL)	21312211	31	Iowa (IA)	3160553
4	New York (NY)	19862512	32	Utah (UT)	3159345
5	Pennsylvania (PA)	12823989	33	Nevada (NV)	3056824
6	Illinois (IL)	12768320	34	Arkansas (AR)	3020327
7	Ohio (OH)	11694664	35	Mississippi (MS)	2982785
8	Georgia (GA)	10545138	36	Kansas (KS)	2918515
9	North Carolina (NC)	10390149	37	New Mexico (NM)	2090708
10	Michigan (MI)	9991177	38	Nebraska (NE)	1932549

11	New Jersey (NJ)	9032872	39	West Virginia (WV)	1803077
12	Virginia (VA)	8525660	40	Idaho (ID)	1753860
13	Washington (WA)	7530552	41	Guam (GU)	168678
14	Arizona (AZ)	7123898	42	Hawaii (HI)	1426393
15	Massachusetts (MA)	6895917	43	New Hampshire (NH)	1350575
16	Tennessee (TN)	6782564	44	Maine (ME)	1341582
17	Indiana (IN)	6699629	45	Montana (MT)	1062330
18	Missouri (MO)	6135888	46	Rhode Island (RI)	1061712
19	Maryland (MD)	6079602	47	Delaware (DE)	971180
20	Wisconsin (WI)	5818049	48	South Dakota (SD)	877790
21	Colorado (CO)	5684203	49	North Dakota (ND)	755238
22	Minnesota (MN)	5628162	50	Alaska (AK)	738068
23	South Carolina (SC)	5088916	51	Washington DC (DC)	703608
24	Alabama (AL)	4888949	52	Vermont (VT)	623960
25	Louisiana (LA)	4682509	53	Wyoming (WY)	573720
26	Kentucky (KY)	4472265	54	Virgin Islands (VI)	101365
27	Oregon (OR)	4199563	55	Northern Mariana Islands (MP)	50304
28	Oklahoma (OK)	3940521			

United States and Territories 2-letter Abbreviations

STATE(TERRITORY)		STATE(TERRITORY)		STATE(TERRITORY)	
Alabama	AL	Kentucky	KY	Ohio	OH
Alaska	AK	Louisiana	LA	Oklahoma	OK
Arizona	AZ	Maine	ME	Oregon	OR
Arkansas	AR	Maryland	MD	Pennsylvania	PA
American Samoa	AS	Massachusetts	MA	Puerto Rico	PR
California	CA	Michigan	MI	Rhode Island	RI
Colorado	CO	Minnesota	MN	South Carolina	SC
Connecticut	CT	Mississippi	MS	South Dakota	SD
Delaware	DE	Missouri	MO	Tennessee	TN
District of Columbia	DC	Montana	MT	Texas	TX
Florida	FL	Nebraska	NE	Trust Territories	TT
Georgia	GA	Nevada	NV	Utah	UT
Guam	GU	New Hampshire	NH	Vermont	VT
Hawaii	HI	New Jersey	NJ	Virginia	VA
Idaho	ID	New Mexico	NM	Virgin Islands	VI
Illinois	IL	New York	NY	Washington	WA
Indiana	IN	North Carolina	NC	West Virginia	WV
Iowa	IA	North Dakota	ND	Wisconsin	WI
Kansas	KS	Northern Mariana Islands	MP	Wyoming	WY