

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:

- San Jose/Santa Clara City & County CoC - 506
- Oregon Balance of State CoC - 243
- Las Vegas/Clark County CoC- 214

SQL Code Used:

```
SELECT
  CoC_Name,
  Homeless_Unaccompanied_Youth_Under_18
FROM `capstone-practice-434814.Exploration_Dataset.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:

- This statement is true due to the fact that when we queried the number of unsheltered homeless people in Delaware, we saw that, outside of 2013, the numbers were steadily increasing over time.

SQL Code Used:

```
SELECT
  Count_Year,
  Unsheltered_Homeless
FROM
  `capstone-practice-434814.Exploration_Dataset.homelessness`
WHERE
  State = 'DE'
ORDER BY
  Count_Year
```

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:

90

SQL Code Used:

```
WITH sheltered_homeless_2018 AS (  
  SELECT  
    CoC_Name,  
    Sheltered_SH_Homeless  
  FROM `capstone-practice-434814.Exploration_Dataset.homelessness`  
  WHERE  
    Sheltered_SH_Homeless >= 1  
    AND Count_Year = 2018  
)
```

number of locations with sheltered homeless

```
SELECT  
  COUNT(CoC_Name) AS number_of_Shelters  
FROM  
  sheltered_homeless_2018
```

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

Answer:

- Philadelphia CoC
- Reno, Sparks/Washoe County CoC
- Indianapolis CoC

SQL Code Used:

```
WITH sheltered_homeless_2018 AS (  
  SELECT  
    CoC_Name,  
    Sheltered_SH_Homeless  
  FROM `capstone-practice-434814.Exploration_Dataset.homelessness`  
  WHERE  
    Sheltered_SH_Homeless >= 1  
    AND Count_Year = 2018  
)
```

top 3 locations with sheltered homeless

```

SELECT
    CoC_Name
FROM
    sheltered_homeless_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:

- CA: 128877
- NY: 90884
- FL: 31030
- TX: 25310
- WA: 22304
- MA: 20068
- OR: 14476

SQL Code Used:

```

WITH overall_homeless_2018 AS (
    SELECT
        State,
        Overall_Homeless
    FROM
        `capstone-practice-434814.Exploration_Dataset.homelessness`
    WHERE
        Count_Year = 2018
    GROUP BY
        State,
        Overall_Homeless
    ORDER BY
        Overall_Homeless DESC
)

```

count of overall homeless by state

```

SELECT
    State,
    SUM(overall_homeless_2018.Overall_Homeless) AS homeless_population
FROM
    overall_homeless_2018
GROUP BY
    State
ORDER BY
    homeless_population DESC

```

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?
 - i. This is not the case. While CA and FL do seem to have a correlation, the other 5 states have none. Texas is shown to have the second highest overall population and New York having the 4th highest but Texas only ranks 4th as far as Overall Homeless population and New York ranking 2nd. The next three highest in population don't appear on the next 3 spots in the Overall Homeless Population; WA, MA and OR have the next highest Overall Homeless population and they rank towards the middle of the population rankings
- 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?
 - i. Washington
 - ii. Maryland
 - iii. Oregon
 - iv. New York

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:

- Delaware Statewide CoC: 8.6%
- Anchorage CoC: 8.59%
- Rhode Island Statewide CoC: 4.63%
- Omaha, Council Bluffs CoC: 4.54%
- Cleveland/Cuyahoga County CoC: 4.31%
- Cincinnati/Hamilton County CoC: 4.22%
- Lynn CoC: 3.96%
- Maine Statewide CoC: 3.9%
- Yonkers, Mount Vernon/Westchester County CoC: 2.41%
- Quincy, Brockton, Weymouth, Plymouth City and County CoC: 2.02%
- Nassau, Suffolk Counties CoC: 1.34%
- Springfield/Hampden County CoC: 1.31%

SQL Code Used:

```
WITH overall_homeless_2018 AS (
SELECT
  CoC_Name,
```

```

Overall_Homeless,
Unsheltered_Homeless
FROM
`capstone-practice-434814.Exploration_Dataset.homelessness`
WHERE
Count_Year = 2018
AND Overall_Homeless > 1000
AND Unsheltered_Homeless < 100
ORDER BY
Overall_Homeless DESC
)

```

locations with low unsheltered population in relation to overall homeless shown in percentage

```

SELECT
CoC_Name,
ROUND((Unsheltered_Homeless / Overall_Homeless)*100, 2) AS unsheltered_percentage
FROM overall_homeless_2018
ORDER BY
unsheltered_percentage DESC

```

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

Answer:

- Nassau, Suffolk Counties CoC: 1.34%
- Springfield/Hampden County CoC: 1.31%

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