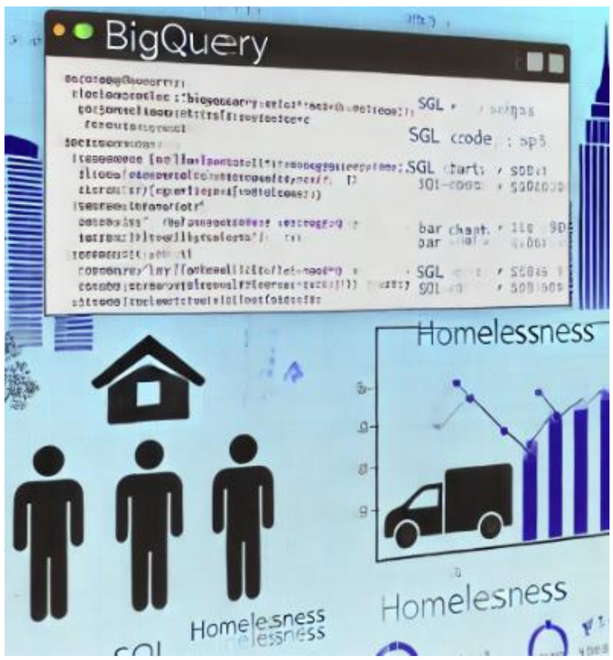
Homelessness in the U.S.

Project Overview

This project focuses on analyzing the **Point-in-Time Homelessness Count** dataset from HUD, using SQL within Google BigQuery to explore homelessness trends across different U.S. states and Continuum of Care (CoC) areas. The goal was to gain insights into the distribution of homelessness, evaluate the effectiveness of shelter programs, and identify locations with specific homelessness challenges.



Tools & Technologies

- SQL
- Google BigQuery
- Exploratory Data Analysis (EDA)

Objectives

- Explore and understand the dataset schema and key metrics around homelessness.

- Answer questions related to homelessness trends using SQL queries.
- Identify regions with specific needs (e.g., unaccompanied homeless youth, unsheltered homelessness).
- Compare homelessness rates with state population data to evaluate representation disparities.

Dataset Details

- **Dataset**: Point-in-Time Homelessness Count
- **Table**: hud_pit_by_coc
- **Source**: HUD (U.S. Department of Housing and Urban Development)

Point-in-Time Homelessness Count

US Dept of Housing and Urban Development

Annual Homeless Assessment Report to Congress

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
CoC_Number	STRING	NULLABLE	-	-	-	-	Number of the Continuum of C...
CoC_Name	STRING	NULLABLE	-	-	-	-	Name of the Continuum of Car...
CoC_Category	STRING	NULLABLE	-	-	-	-	Only for 2018
Overall_Homeless	INTEGER	NULLABLE	-	-	-	-	Total Homeless Population
Sheltered_ES_Homeless	INTEGER	NULLABLE	-	-	-	-	Homeless in Emergency Shelters
Sheltered_TH_Homeless	INTEGER	NULLABLE	-	-	-	-	Homeless in Transitional Housing
Sheltered_SH_Homeless	INTEGER	NULLABLE	-	-	-	-	Homeless in Safe Haven Housing
Sheltered_Total_Homeless	INTEGER	NULLABLE	-	-	-	-	Total Sheltered Homeless
Unsheltered_Homeless	INTEGER	NULLABLE	-	-	-	-	-
Homeless_Individuals	INTEGER	NULLABLE	-	-	-	-	Individuals not with Families
Sheltered_ES_Homeless_Individuals	INTEGER	NULLABLE	-	-	-	-	Emergency Shelter
Sheltered_TH_Homeless_Individuals	INTEGER	NULLABLE	-	-	-	-	Transitional Housing
Sheltered_SH_Homeless_Individuals	INTEGER	NULLABLE	-	-	-	-	Safe Haven Housing
Sheltered_Total_Homeless_Individuals	INTEGER	NULLABLE	-	-	-	-	-

Schema view of the Point-in-Time Homelessness Count dataset showing the field names, types, and descriptions for each column. This helps define the various homelessness categories and the structure of the data used in analysis.

Row	CoC_Number	CoC_Name	CoC_Category	Overall_Homele	Sheltered_ES_H	Sheltered_TH_H	Sheltered_SH_H	Sheltered_Total_	Unsheltered_Ho
1	AK-500	Anchorage CoC	null	1023	665	305	0	970	53
2	AK-500	Anchorage CoC	null	1122	676	394	0	1070	52
3	AK-500	Anchorage CoC	null	1147	704	393	0	1097	50
4	AK-501	Alaska Balance of State CoC	null	761	474	195	0	669	92
5	AK-501	Alaska Balance of State CoC	null	824	461	210	0	671	153
6	AK-501	Alaska Balance of State CoC	null	766	435	184	0	619	147
7	AL-500	Birmingham/Jefferson, St. Clair...	null	1329	387	497	31	915	414
8	AL-500	Birmingham/Jefferson, St. Clair...	null	1469	373	556	31	960	509
9	AL-500	Birmingham/Jefferson, St. Clair...	null	1707	347	630	32	1009	698
10	AL-501	Mobile City & County/Baldwin ...	null	598	287	174	0	461	137
11	AL-501	Mobile City & County/Baldwin ...	null	493	153	107	0	260	233
12	AL-501	Mobile City & County/Baldwin ...	null	634	212	112	0	324	310
13	AL-502	Florence/Northwest Alabama ...	null	209	38	151	0	189	20
14	AL-502	Florence/Northwest Alabama ...	null	223	22	175	0	197	26
15	AL-502	Florence/Northwest Alabama ...	null	192	18	170	0	188	4
16	AL-503	Huntsville/North Alabama CoC	null	536	413	86	0	499	37
17	AL-503	Huntsville/North Alabama CoC	null	586	320	89	0	409	177
18	AL-503	Huntsville/North Alabama CoC	null	607	336	96	0	432	175
19	AL-504	Montgomery City & County CoC	null	490	120	289	0	409	81
20	AL-504	Montgomery City & County CoC	null	515	159	240	0	399	116

A preview of the dataset used for SQL exploration, showing homelessness data by Continuum of Care (CoC) areas, including total homeless populations and breakdowns of sheltered and unsheltered homeless individuals.

Key SQL Queries and Insights

1. Top 3 Areas for Unaccompanied Homeless Youth Under 18 (2018)
2. Homelessness Trends in Delaware
3. Safe Haven Program Analysis (2018)
4. Top 7 States by Homeless Population (2018)
5. Shelter Effectiveness: Locations with Low Unsheltered Homelessness (2018)
6. Overrepresentation and Underrepresentation of Homelessness in Certain States

```
SELECT
  hl.CoC_Name,hl.Homeless_Unaccompanied_Youth_Under_18
FROM
  `Exploration_Project.homelessness` AS hl
WHERE
  hl.Count_Year = 2018
ORDER BY
  hl.Homeless_Unaccompanied_Youth_Under_18 DESC
```

```sql
LIMIT 3;
```

SQL query used to identify the top 3 Continuum of Care (CoC) areas with the highest number of unaccompanied homeless youth under 18 in 2018. This information helps in determining the locations where new programs for unaccompanied homeless children should be developed.

```sql
SELECT hl.Count_Year,hl.Unsheltered_Homeless

FROM

 `Exploration_Project.homelessness` AS hl

WHERE

 hl.State = 'DE'

ORDER BY

 hl.Count_Year;
```

SQL query used to verify whether the number of unsheltered homeless people in Delaware has increased over the past 7 years.

```sql
SELECT hl.CoC_Name,hl.Sheltered_SH_Homeless

FROM

 `Exploration_Project.homelessness` AS hl

WHERE

 hl.Count_Year = 2018

AND

 hl.Sheltered_SH_Homeless > 0;
```

SQL query used to identify how many different Continuum of Care (CoC) locations had at least one person sheltered in Safe Haven housing in 2018. This helps determine the extent to which Safe Haven programs were active in various regions despite funding cuts.

```sql
SELECT

 hl.state,SUM(hl.Overall_Homeless) as total_homeless_population

FROM

 `Exploration_Project.homelessness` AS hl
```

```
WHERE
  hl.Count_Year = 2018
GROUP BY hl.state
ORDER BY total_homeless_population DESC
LIMIT 7;
```

SQL query used to determine the top 7 states with the highest overall homeless population in 2018. By grouping the data by state and summing the total homeless population, this query identifies the regions most affected by homelessness in that year.

```
SELECT
  hl.CoC_Name,hl.Overall_Homeless,hl.Unsheltered_Homeless, Round(hl.Unsheltered_Homeless/hl.Overall_Homeless*100,2) as unsheltered_percentage
FROM
  `Exploration_Project.homelessness` AS hl
WHERE
  hl.Count_Year = 2018
  AND hl.Overall_Homeless > 1000
  AND hl.Unsheltered_Homeless <100
  AND Round(hl.Unsheltered_Homeless/hl.Overall_Homeless*100,2) < 2;
```

SQL queries used to identify locations in 2018 with more than 1000 overall homeless individuals and fewer than 100 unsheltered homeless. Additionally, the query calculates the percentage of unsheltered homeless in these locations, identifying places where less than 2% of the homeless population is unsheltered, highlighting regions that provide effective sheltering.

By comparing the **top 7 states** in terms of **overall homeless population** with their rankings by **total state population**, an interesting trend emerges:

- **Overrepresented States**: States like **New York (NY)**, **Washington (WA)**, **Massachusetts (MA)**, and **Oregon (OR)** have a higher ranking in terms of homelessness compared to their population ranking. This indicates that homelessness is disproportionately higher in these states relative to their total population.

- **Underrepresented States**: In contrast, states like **Texas (TX)**, **Pennsylvania (PA)**, **Illinois (IL)**, and **Ohio (OH)** are underrepresented in terms of homelessness, meaning they have a larger total population but rank lower in homeless population. This suggests that the homelessness issue is less prevalent in these areas relative to their overall population size.

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

US State and Territory* Population Ranking for 2018

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

United States and Territories 2-letter Abbreviations

Challenges and Learnings

- **Handling Missing Data**: Some entries in the dataset had missing or incomplete information. Various techniques like filtering and handling NULL values were applied.
- **State Identification**: The dataset did not directly include state names, so I had to extract state identifiers from the CoC_Number column using the SQL LEFT() function.
- **Insights for Policy-Making**: This project revealed insights that could help governments and non-profits allocate resources effectively to regions most in need of support.

Future Enhancements

- **Data Visualization**: Creating dashboards to visualize key trends using Tableau or Power BI.
- **Further Analysis**: Conducting deeper time-series analysis to predict future homelessness trends based on historical data.

Conclusion

- This project provided valuable insights into homelessness trends across the U.S., highlighting regions that require more support and identifying effective shelter programs. By leveraging SQL in BigQuery, I was able to extract meaningful patterns from large datasets, which can guide policy-making and resource allocation.