# 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.

A close-up of a computer screen

Description automatically generated

**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)

A screenshot of a web page

Description automatically generated

A screenshot of a computer

Description automatically generated

*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.*

A screenshot of a computer

Description automatically generated

*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

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.*

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.*

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.*

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.

A table with numbers and names

Description automatically generated

**US State and Territory\* Population Ranking for 2018**

A table with names and states

Description automatically generated

**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.