A look into Homelessness Data in America

Janine Par

2022-05-15

Introduction

Homelessness is one of the social problems our nation is facing now. The pandemic, inflation, and other economic events have been contributing factors to the number of homeless Americans, which continues to increase significantly in an alarming rate. McKinney–Vento Homeless Assistance Act of 1987 defines homelessness as people lacking a fixed, regular, and adequate nighttime residence. Adults, children, families of different age and race struggle with homelessness. According to the U.S. Department of Housing and Urban Development (AHAR) about 18 of every 10,000 people in United State experience homelessness in United State. California, New York, and Florida are the top states with homeless populations however the state with the lowest homeless population is seeing an increase as well. Even before the Pandemic, Chicago sees a two percent increase over 2019. State and Federal government is allocating and spending billions of dollars to fund initiatives that aim to end and reduce homelessness. Data science and technology can be used to guide the government in making informed decisions that they can use to fight the homelessness crisis.

For this project, my goal is to understand and correlate multiple variables that contribute to the increasing rate of homelessness in America, focusing on the state with the top homeless population. I would like to do more analysis on the factors and variables and use them to identify Americans at risk or on the verge of homelessness and those that may be repeated homelessness (Chronic homeless).

Research questions

1. Identify the current statistics on homelessness in America and focus on the Top state with homeless population.
2. Compare these variables between states with highest and lowest homeless population
3. Identify factors and variables that influence the decline in homelessness population
4. What are the main factors and variables that influence the increase of homelessness?

Approach and How your approach addresses (fully or partially) the problem.

My approach to understanding homelessness crisis is to explore and analyze data on homelessness population and look at factors including but not limited to:

- Demographic Factors
- Economic Factors
- Environmental Factors
- Health and Behavior Factors

This includes running multiple transformations, filtering, slicing and selecting data to make an informed decision on this chosen topic. I plan to focus on the state with the highest rate of homelessness and compare that information with other states that show the rise of homelessness and validate if the variables (predictors) show the same trend. I also plan to look at these variables and identify the ways they are related and if the relationship can be measurable. This analysis also attempts to use these variables as predictors and show how those increase/decrease the homelessness rate as the outcome variable.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

1. 2021 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S. 2007 - 2021 Point-in-Time Estimates by State

Point-In-Time (PIT) count and Housing Inventory Count (HIC) conducted in January 2021. Specifically, this report provides 2021 national, state, and CoC-level PIT and HIC estimates of homelessness, as well as estimates of chronically homeless persons, homeless veterans, and homeless children and youth.

Data is aggregated by state with 26 continuous variables of count for each population including Age, race, family and veteran information.

2. Homeless Population USA FACTS(2005-2020) This statistic is collected by the Department of Housing and Urban Development and reflects a snapshot of the homeless population as of a single night at the end of January each year from 2005-2020. This includes homeless population by living situation, age, ethnicity, gender, race ethnicity and state

3. CoC Housing Inventory Count Reports The reports tally the number of beds and units available on the night designated for the count by program type, and include beds dedicated to serve persons who are homeless as well as persons in Permanent Supportive Housing. This data is currently in PDF format, I'm still researching if raw data is available to be downloaded. This is helpful reference to check if these available homeless resources per state is effective and if data can show decline in homelessness.

Required Packages

Reading, selecting, Transformation, Slicing

plyr dplyr readxl magrittr tibble

Plotting and RMD

ggplot2

Analysis

lm.beta car

Plots and Table Needs

- Frequency Distribution to validate normality in distribution of gathered data
- Correlation and Covariance using the different factors and variables with Homelessness rate

Questions for future steps

1. Do I have enough data to proceed with the analysis? With multiple definitions of homelessness, do we have good quality data representing America's homeless population?

2. Do I clearly define our project goals and outcome for this analysis?

3. I need to define assumptions on the data used for this project clearly.

4. What are other variables that I can gather that might have a correlation and causation with homelessness? (i.e Domestic Violence rate, Job availability, Drug abuse rate)

References

2021 ahar: Part 1 - pit estimates of homelessness in the U.S. 2021 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S. | HUD USER. (n.d.). Retrieved May 15, 2022, from https://www.huduser.gov/portal/datasets/ahar/2021-ahar-part-1-pit-estimates-of-homelessness-in-the-us.html

USAFacts. (n.d.). Homeless population. USAFacts. Retrieved May 15, 2022, from https://usafacts.org/data/topics/people-society/poverty/public-housing/homeless-population/?utm_source=bing&utm_medium=cpc&utm_campaign=ND-StatsData&msclkid=a229785546631d9d16451ad4d1a50df6

COC housing inventory count reports. HUD Exchange. (n.d.). Retrieved May 15, 2022, from https://www.hudexchange.info/programs/coc/coc-housing-inventory-count-reports/

Publisher data.cityofnewyork.us. (2022, May 12). DHS Daily Report. DHS Daily Report - CKAN. Retrieved May 15, 2022, from https://catalog.data.gov/dataset/dhs-daily-report

System performance measures. HUD Exchange. (n.d.). Retrieved May 15, 2022, from https://www.hudexchange.info/programs/coc/system-performance-measures/#data