



Homeless Hospital Visits in California

09.15.2022

Theresa Parker

Executive Summary	2
Data Sources	2
Data Collection Method	2
Data Overview	3
Rows and Columns:	3
Columns and Descriptions:	3
Data Profile:	4
Cleaning the Data	4
Column Derivations:	5
Descriptive Statistics / Understanding the Data:	5
Descriptive Statistics	5
Checking the Average and Variance	5
Understanding the Data	6
Limitations & Ethics:	7
Bias	7
Timely	7
Manual Errors:	7
Exploratory Questions:	7
Who:	7
What:	8
Where:	8
When:	8
How:	8
Why:	8
Results and Findings:	8
Recommendations:	10

Executive Summary

Hospital demographics are very important as they can help guide programs, services, and staff members accordingly. By looking at this data to see if there are any trends or correlations, we can find out if there are homeless areas that are more populated than others, or if demographics may be a correlation with homeless encounters in the California Hospitals.. By seeing where there may be trends or correlations, hospitals can help provide:

- Patients with appropriate resources
- Programs (financial, mental health, etc) to help homelessness
- Provide staff with appropriate training on how to work with and care for homeless patients.

Knowing patient demographics can help the hospitals to better serve their community, which in turn will better serve the hospitals business needs.

Data Sources

Hospital Encounters for Homeless Patients is a dataset that is open source from the state of California. Hospitals have provided this information and the government owns it, but have made it open for public records. This is deemed a reliable source as it came from the California Health and Human Services Open Data Portal.

[Homelessness across the USA](#) is an open source dataset from datahub.io. This was a very small part used for the time series analysis of the project; however, this was not used during any final analysis. This data is obtained from HUD exchange. Because it was such a small part of the project and not used in the analysis, the information on it will not be included below.

Data Collection Method

Hospital Encounters for Homeless Patients is the main data frame that we used and the only one that was used for analysis. This dataset is administrative data because in order to meet California's needs in making informed decisions and through partnerships with healthcare facility data providers, The Department of Health Care Access and Information

(HCAI) has been collecting quality, timely, secure, and comprehensive inpatient discharges since 1983 — and emergency department and ambulatory surgery encounters, from both hospital-based ambulatory surgery centers and licensed freestanding ambulatory surgery clinics, since 2005. Data submitted via HCAI's secure data collection system.'

<https://hcai.ca.gov/data-and-reports/submit-data/patient-data/>

Data Overview

Rows and Columns:

Original Dataset (Rows: 25,450 Columns: 17)

Updated Dataset after cleaning (Rows: 25,450 Columns: 16)

Columns and Descriptions:

Column	Description
Encounter Type	Inpatient (IP) Hospitalization or Emergency Department (ED) Visit
Hospital County	The county where the treating hospital is located
OSHPD ID	The 6-digit hospital identifier
Facility Name	The name of the treating hospital
Health System	The name of common ownership and/or association for a group of facilities. For this product, only health systems which operated three or more hospitals are listed
Hospital Ownership Type	The hospital's ownership type (Government, Investor, or Non-Profit)
Urban Rural Designation	Indicates if hospital has an Urban or Rural (includes Rural/Frontier) designation
Teaching Hospital Designation	Indicates if the hospital is designated as a teaching hospital
Licensed Bed Size	Indicates the hospital's number of licensed beds

Primary Care Shortage Area	Indicates if the hospital is located in a Health Professional Shortage Area (HPSA) for Primary Care
Mental Health Shortage Area	Indicates if the hospital is located in a Health Professional Shortage Area (HPSA) for Mental Health Care
Homeless Indicator	Indicates if the data is for Homeless or Non-Homeless encounters
Demographic Category	Age, Race, Sex, or (Expected) Payer Note: Unknown age and Unknown sex are not included. Other Payer includes Workersâ€™ Compensation, Other Government, Title V, Disability, VA Plan, Other Payer, invalid, and missing
Demographic Category Value	Value for demographic category
Encounters	Count of inpatient hospitalizations (i.e., discharges) or emergency department visits
Total Hospital Encounters	Total inpatient hospitalizations or emergency department visits per hospital. Note: the total hospital encounters may be slightly less for age and sex (versus race and expected payer) due to the exclusion of "Unknown" from age and sex
Percent	Calculation: Encounters/Total Hospital Encounters x 100
Non_Homeless_Pop	Derived column that counts accounts for the Non-Homeless Population. (Total Encounters - Encounters)

Data Profile:

● Cleaning the Data

This dataset was very clean upon download. Very little needed to be changed. Please see below for details.

- Missing Data - none
- Duplicates - none
- Inaccurate or inconsistent data
 - **Licensedbedsize** has Jan 1999 for bed size. Upon reviewing the data this should reflect 01-99 beds. Updated value in variable.
- Mixed-type columns
 - **System** had a mixed data type. Changed it to string.

- Outliers - there are outliers as the median is lower than the mean. Further analysis needed.
- Removal of Columns
 - **OSHPD_ID** - removed column as it was not needed for this project.
- Column Renaming
 - Renamed **Encounters** to: Homeless_Pop to be more intuitive

● Column Derivations:

- Created **Non_Homeless_Pop** to obtain the count of Non_Homeless population; however, after further review this column calculation was represented of of that, and therefore

Descriptive Statistics / Understanding the Data:

● Descriptive Statistics

	Encounters	TotalEncounters	Percent
count	25449.000000	25449.000000	25449.000000
mean	4737.598373	20135.015718	23.529211
std	11630.923563	35862.090590	23.181105
min	0.000000	1.000000	0.000000
25%	24.000000	472.000000	2.800000
50%	303.000000	2406.000000	17.100000
75%	3233.000000	24756.000000	39.200000
max	159647.000000	262076.000000	100.000000

● Checking the Average and Variance

```
# checking the average
average = df[["Encounters", "TotalEncounters", "Percent"]].mean()
print(average)
```

```
Encounters      4737.598373
TotalEncounters 20135.015718
Percent         23.529211
dtype: float64
```

```
# checking the median due to outlier sensitivity with the average/mean
median = df[["Encounters", "TotalEncounters", "Percent"]].median()
print(median)
```

```
Encounters      303.0
TotalEncounters 2406.0
Percent         17.1
dtype: float64
```

```
# checking the variance
var = df[["Encounters", "TotalEncounters", "Percent"]].var()
print(var)
```

```
Encounters      1.352784e+08
TotalEncounters 1.286090e+09
Percent         5.373636e+02
dtype: float64
```

● Understanding the Data

- This is a medium size dataset with around 25,000 rows and 16 columns.
- Most of the variables are categorical data in nature. There are three columns that have continuous data that allow us to pinpoint the number of homeless visits, total visits, and the percentage of the homeless visits compared to total visits.
- Mean/Median: We have a large difference between the mean(s) and median(s). The largest difference occurs in the variable Homeless Encounters. Due to the large difference and the mean quite a bit higher as it's sensitive to outliers indicates there are probably quite a few outliers that should be looked into for further analysis.

- Variance: There is a very large variance, which tells us that the spread of the data is very large.
- Std. Deviation: 1 standard deviation away from the mean is 11,630, which will account for about 68% of the homeless encounters at the hospitals in California.

Limitations & Ethics:

- Bias

Overall there is not much bias to this data set; however, one area to consider is “how was the homelessness question calculated?” If there was not a standard definition to define homelessness, the question could be up for interpretation therefore possibly skewing results in a direction.

It is important to keep in mind that these are numbers of encounters at the hospital and are not an accurate representation of the homeless population count as one person could have 1 visit, while someone else may have 10+ encounters

- Timely

This dataset is from 2019-2010 and is considered timely.

- Manual Errors:

Data set is prone to manual errors depending due to the imputation of inputting the information into the hospital system.

Exploratory Questions:

The Story: Homeless Hospital Visits in California

Who:

- Homeless Patients in California

What:

- What department are the homeless more likely to end up in (ER or Inpatient Hospitalization)?
- What age group has the highest encounters
- What are the forecasting trends for homeless visits?

Where:

- Are there certain counties that are hit harder with homeless encounters
- Are these in an area where there are Primary or mental health shortage areas?

When:

- 2019-2020 and past years.
- When was the highest rate of homelessness? 2010

How:

- Do certain demographics have higher visits, factors such as:
 - Age
 - Race
 - Sex
 - Payer
- How has homelessness visits changed throughout the years?

Why:

- Why are the homeless at the hospital (information unknown). Limitation

Results and Findings:

County Demographics:

- Southern California and the San Francisco Bay area were the top areas with the highest encounters.
- Los Angeles was an extreme outlier with the highest number of encounters. It was determined to keep LA in the dataset as Los Angeles (LA) does have a significantly larger population than all other California counties.
 - LA has almost 10 million in total population, while the second most populated county is San Diego with 3.2 million. This stands to reason that LA would have higher findings than other counties.

- I did try to remove LA county and it did not affect the southern area as it was still an area with the highest encounters.

Hospital Demographics:

- Primary Care/Mental Health Care Shortage Area: Los Angeles was the only county to be within the top 5 of the highest encounters and also have primary/mental health care shortages. LA topped the list with 13 facilities in a shortage area.
- ER/Inpatient Hospitalizations: There were many more homeless encounters in the ER. This could mean that homeless are not able to be proactive with their health and therefore that is why the ER visits are higher.

Patient Demographics:

- Age Group: The age group with the highest number of instances was Age Group 40-59. After doing Linear Regression we found there was a significant correlation/relationship between Age Group 40-59 and the total encounters of all the other Homeless Age Groups. Age 0-18 had the fewest instances.
- Race: The White/Caucasian race had the most encounters.
- Sex: Males were 70% of the encounters while Females were 30%

Cluster Analysis:

- I did perform a cluster analysis on the data set to see if I could try to find more relationships; however, due to the limited continuous data I was not able to get any additional insight.

Time Series Analysis:

- Time Series Analysis was performed on the Homeless Population in the USA from 2007-2018. Because this dataset was not broken down by state, I could not pull out California numbers. Therefore, these were standalone observations that were not included in the final analysis.
- When doing the time series analysis on the total population of homeless throughout the year, the data was non-stationary. In order to determine it was non-stationary I



performed the Dickey Fuller Test and Autocorrelation. Then I stationarized the data and reviewed it and noticed that homeless rates increased from 2013 - 2018.

Recommendations:

- Focus on Southern California (especially LA county) and the San Francisco Bay area as they have the greatest numbers of incidents. This includes increasing primary care and mental health staffing.
- Begin tracking overarching themes of what brings the patient to the hospital to in order to find out if there are any trends where age-appropriate resources can be provided.
- Further explore why males have more encounters than females.
- Discuss how hospitals can be more proactive about healthcare needs for homeless patients, as this may help to reduce the amount of ER visits.
- Extra Thought: Ensure that staff, administrators, and clinicians are physically, mentally, and emotionally trained to handle difficult homeless patient encounters. This will help staff to feel equipped to handle the situations if they arise, and thus providing a better staff and patient experience.