Name: Chinedu Onyeka, Idongesit Nkanang, Eric George Mons
Course: Data1202 - Data Analytics Tools

This dataset contains an estimates of homeless people in each U.S. state in the year 2017. The individual column is the number of homeless individuals not part of a family with children. The family_members column is the number of homeless individuals part of a family with children. The state_pop column is the state's total population.

Descriptive statistics

Descriptive statistics on the homeless dataset and insights based on these descriptive statistics indicate that for the variable in focus (individuals, family_members and state_pop) the parameter counts is 510 for all the variables , thus indicating that there is no null values in the data set.
The maximum value for individuals is 109008, minimum value of 434 and a mean value of $7.225784e+03$. However, the family_members had a maximum value of 52070.0 , a minimum value of 75.0 and a mean of $3.504882e+03$

1. Using the isnull().sum() function, the homeless data set was checked for any missing values and and no missing values was identified in the data set, hence no further action was required to remove the null values.

2. Using the .unique() and the .unique() function, 9 different unique regions where identified.

3. Also Using the .unique() and the .unique() function, 51 different states where identified.

   Using the .groupby('region').sum() function. The dataset was grouped into identified inique groups: region, and state.

4. Using the sns.heatmap() function, a heat map of the correlation matrix for the homeless dataset was generated. Using the indicated scale, the box with the lightest color indicated that a correlation exist between the two associated parameters while the very dark boxes indicated no correlation between the parameters.

INSIGHTS FROM DATA SET:

QUESTIONS FOR ANALYSIS:

1. Find region with highest homeless people?
2. Get the region with the maximum population?
3. Get the region with the maximum homeless proportion?
4. Get region with highest total homeless?
5. find state with highest homeless people?

6.   Prepare a complete word document/report on the above five steps.

Note: The word document should be very detailed. It should clearly depict your thought process, the questions asked to the dataset and the insights received from the datasets on those questions.