BU Spark Project (DS701)

Predictive Model Based on Homelessness

Client Team

- **Dr. Tom Byrne**
  Associate Professor, School of Social Work, BU
- **Dr. Molly Richard**
  Postdoctoral Associate, Center for Innovation in Social Science, BU

Instructor: Prof. Thomas Gardos

Technical Project Manager (TPM): Dhruv Shah
Project Manager (PM): Jasmine Dong

Team Members

1. Samritha Aadhi Ravikumar
2. Syeda Shehrbano Aqeel
3. Kunshu Yang
4. Renjie Fan
5. Shiheng Xu

Project Overview:

- **Goal**: Develop a predictive model of homelessness at the community level using data from 2007-2023.
- Focuses on approximately **400 Continuums of Care (CoC)** that receive federal homeless assistance funding from the U.S. Department of Housing and Urban Development (HUD).
- **Unique Focus**: Unlike previous studies that predict homelessness at the individual level, this project centers on community-level factors.

Data Sources:

- **Primary Data**: Annual homelessness counts from HUD across CoC units.
- **Additional Data**: Publicly available community-level factors such as rent rates, demographic and economic conditions, aggregated by CoC.
- **Timeframe**: 2007 to 2023.

Key Research Questions:

- Moving beyond simply identifying associations between community-level factors and homelessness.
- **Objective**: Predict the number or rate of homelessness in each CoC based on structural determinants like rent levels and economic conditions.

Methodology and Tools:

- **Modeling Approach**: Regression models or other predictive machine learning techniques.
- **Required Skills**: Familiarity with regression models, feature engineering, and experience with the **pandas** and **scikit-learn** packages in Python.

Preliminary Analysis:

1. Data Cleaning

- **Completed by a Team Member**: The dataset was cleaned by one of the team members to ensure it was free from inconsistencies and ready for analysis.

2. Data Preprocessing

- **Basic Checks**: Verified dataset's basic information, statistical summary, and data types to ensure consistency.
- **Validation**: Checked for missing values, duplicates, and inconsistencies to confirm the data was clean and ready for analysis.

3. Data Visualization

- **Trend Analysis**: Used line plots to illustrate homelessness trends over time across different Continuums of Care (CoCs).
- **Comparison**: Created bar charts to compare types of homelessness (e.g., family vs. individual) by CoC category.
- **Correlation**: Utilized scatter plots to explore the relationships between factors like rent rates, poverty levels, and homelessness counts.

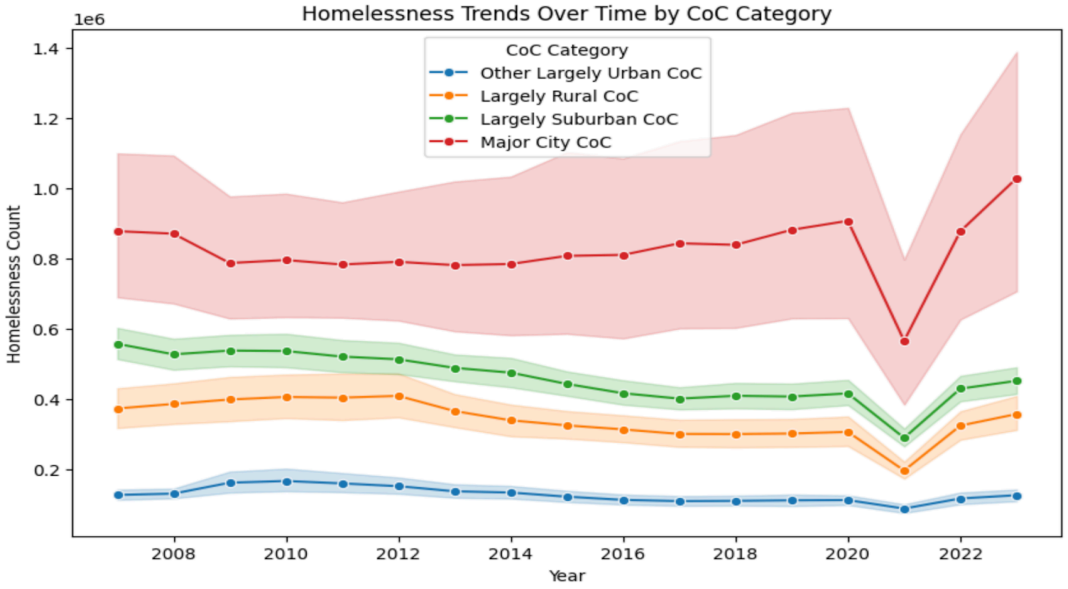
Questions that Can be Answered:

- **Total Count of Homeless Individuals**: The dataset provided the total number of homeless individuals across each Continuum of Care (CoC) region.
- **Sheltered vs. Unsheltered Counts**: The data differentiated between sheltered and unsheltered homeless individuals, offering insights into the living conditions of the homeless population.
- **Homeless Individuals vs. Families**: It included distinctions between homeless individuals and those experiencing homelessness within families, highlighting the varying needs of different groups.
- **Geographic Comparisons**: Geographic assessments were made to compare homelessness rates across different CoC categories, including urban, suburban, and rural areas.
- **Regional Differences**: The data enabled comparisons of homelessness rates between specific regions, such as New York and Alabama, to identify patterns and disparities.
- **Demographic Breakdown**: It highlighted demographic statistics, showcasing the proportion of individuals versus families experiencing homelessness within each CoC, aiding in understanding the affected populations.

Here are some additional questions along with detailed analyses:

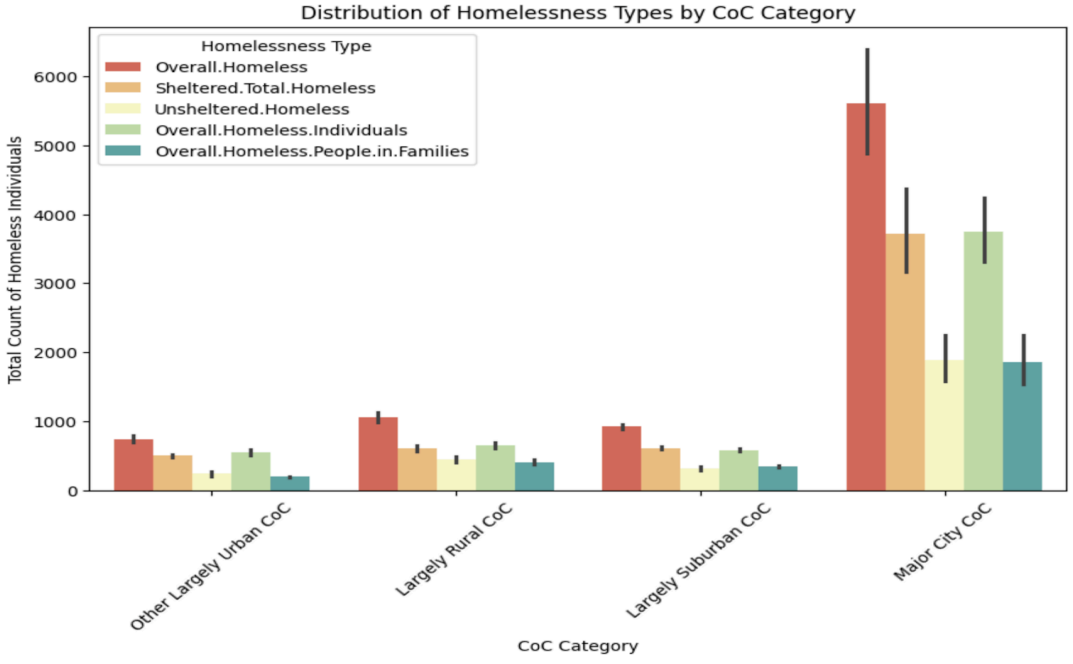
1. ***How Has the no. of Homeless Individuals Changed Over Time Across Different CoCs?***

- Aggregated data by **CoC Category** and **Year**.
- Visualized trends using a **line plot** to show homelessness changes over time for each category.
- **Key observation**: Among all CoC categories, **Major City CoCs** consistently report the highest number of homeless individuals. Notably, in **2020**, there was a decline in homelessness across all categories, including Major City CoCs. However, in **2021**, there was a significant increase in homelessness, reversing the downward trend from the previous year.



2. ***Does Homelessness Type Vary Significantly by CoC Category?***

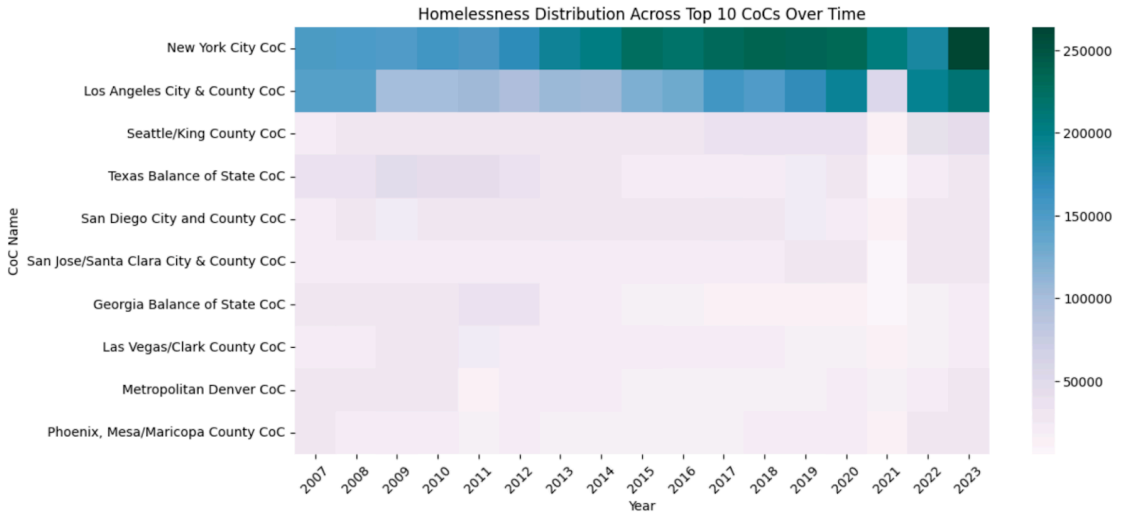
- Aggregated data by **CoC Category** and **Homelessness Type**.
- Visualized the distribution using a **bar plot** to compare homelessness types (e.g., family vs. individual).
- **Key observation**: Among both CoC Categories and Homelessness Types, **Major City CoCs** and **Overall Homelessness** consistently report the highest numbers, respectively.



3. ***Does Homelessness Distribution Vary Across Top 10 CoCs Over Time?***

- Aggregated data by **CoC Name** and **Year**.
- Visualized the findings using a **heatmap** to depict variations in homelessness counts across the top 10 CoCs over the years.

- **Key observation**: The heatmap reveals **New York City** and **Los Angeles** consistently show the highest levels of homelessness among the top 10 CoCs over time, indicating a significant concentration of homelessness in these major urban areas.



Questions that Cannot be Answered:.

- **Key demographic information**, such as the age, gender, race, or ethnicity of homeless individuals, as well as veteran status or chronic homelessness, was missing.
- Additionally, the data did not cover the **causes of homelessness**, such as job loss, eviction, or domestic violence, nor did it provide information on the length of time individuals have been homeless.
- **Service utilization data**, including the types of services accessed by homeless individuals and the effectiveness of interventions, was also not available.
- Other critical gaps included information on housing inventory, such as available shelter beds and permanent supportive housing units, as well as local economic indicators like unemployment rates, housing costs, and poverty levels, which are essential for contextualizing the homelessness data.

Additional Data Needed:

- Integrating **economic data** from the U.S. Census Bureau, American Community Survey, or the Bureau of Labor Statistics would help provide crucial economic context to homelessness trends, such as local unemployment rates, housing costs, and poverty levels.
- **Geographic Information System data** of Continuum of Care (CoC) regions from HUD Exchange and Fair Market Rent data from HUD USER would also be useful in analyzing geographic patterns and housing affordability.
- Furthermore, accessing detailed **client-level data** from the Homeless Management Information Systems could provide valuable insights into service utilization and demographics, such as age, gender, race, veteran status, and chronic homelessness.

Conclusion:

Combining these datasets would allow for a more thorough analysis of the factors influencing homelessness and the effectiveness of interventions.