SJD Sant Joan de Déu València

18th European Research Conference Homelessness

VNIVERSITAT DE VALÈNCIA

Potentialities of Data Science for improving homelessness service provision

Authors

Puchol [A,B], G.; Ben Ghazala[B], R.; Rodilla[A], J. M.; Montalba[B], C.

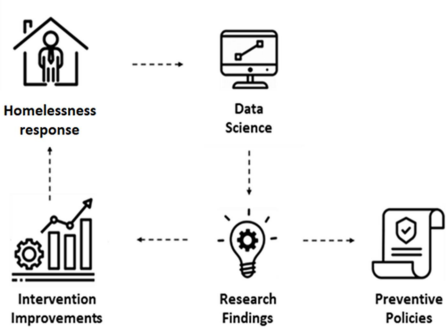
A. St. John of God Valencia B. University of Valencia

Context

Data science studies on homelessness have been used to inform homelessness prevention efforts (Gao, Das, & Fowler, 2017), identify critical aspects where data science can make significant advances in homelessness services (Chelmis, Qi, & Lee, 2021), or to evaluate service allocation methods for homelessness services (Qi & Chelmis, 2023).

Proposal

To explore data science techniques in social care research by using administrative data to identify effective variables, treatments, and pathways for various subgroups. Support decision-making and provide probabilistic success rates for interventions guaranteeing ethical guidance and authorization throughout the process.



Objectives: Explore potentialities of using data science techniques to records from homeless services to support decision-making in social care:

1. **Ethical Use of Algorithms**: integrate ethical perspectives in the process.
2. **Data Processing**: assess data records quality (from data science perspective).
3. **Exploratory Analysis:** conduct univariate, bivariate, & evaluate potential uses of clustering analyses.
4. **Supervised Learning**: evaluate potential uses for predictive modelling.

Data collection

Source: Non-profit organization in Valencia, Spain, providing housing services. Intervention records saved in internal software, shared with the Department of Data Science at the University of Valencia, through collaborative agreements.

Programs:
Transitional Supportive Housing
Transitional Supportive Shelter

Data Coverage: intervention records for participants joining the program between 2011 and 2023.

Sample characteristics

Users:
1,180 individuals, 1,521 records (23% re-entry rate).

Participation in housing programs:
137 used transitional housing only
1,202 used transitional shelter only
168 used both programs

Characteristics:
Housing Program: Avg. Age 44, 34% women, 73% national.
Shelter Program: Avg. Age 47, 100% men, 53% national.

Ethical principles

Privacy: Pseudo-anonymization and data protection.

Equity: Achieve 'differential privacy' by seeking to replace individual fields without significantly altering the statistical results (Camacho, 2023).

Responsibility: Supervision and bias mitigation in predictive models (Sadin, 2020).

Analysis

Univariate and bivariate analysis: the variable "stay duration" has a positive effect on residential outcomes.

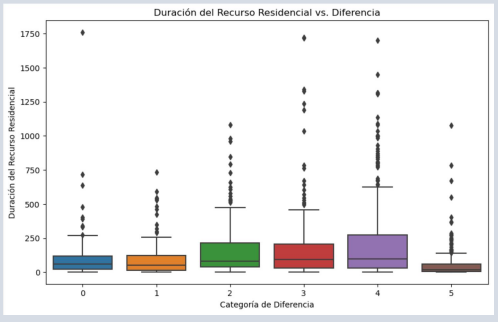


Figure 1: Boxplots of residential service duration based on difference in the residential situation.

Clustering methods: K-means offers best option with two distinct clusters that are consistent with current literature (Culhane & Kuhn, 1998; Muñoz et al., 2005; Aubry, et al., 2013; Bairéad & Norris, 2022).

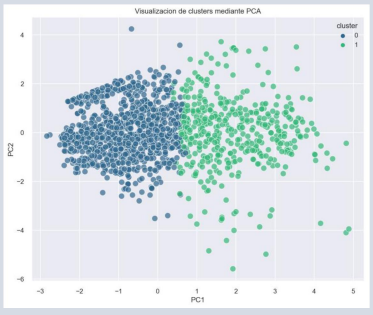


Figure 2: Cluster visualization using PCA, with K-Means (2 clusters)

Predictive Modelling: Random Forests models are the best option for predicting differences in residential outcomes and stay duration.

Prediction of residential difference based on Intervention data:
- Accuracy: 40.54%
- Precision: 39.23%
- Recall: 40.54%
- F1-Score: 30.54%

Prediction of residential difference based on Entry data:
- Accuracy: 56.76%
- Precision: 54.31%
- Recall: 56.76%
- F1-Score: 54.69%

Conclusion

Innovative data science approaches could be adopted, but requires improvements on data quality (Chelmis et al., 2021): Improve data collection, missing values and data inconsistencies to reduce data bias and automate data processes and monitor quality. The study identified k-Means and Random Forest as the best alternatives in the use of data science tools for analysing intervention records on homeless population.

From an ethical perspective, recommendations address the need to work by clusters, not individually, to ensure data protection and privacy, to consider carefully which variables are given more weight and to avoid using personal characteristics as causal variables. Future research should enhance predictive analysis with longitudinal data and the need to refine the definition of "successful exits".

Main references

Aubry, T., Farrell, S., Hwang, S. W., & Calhoun, M. (2013). Identifying the Patterns of Emergency Shelter Stays of Single Individuals in Canadian Cities of Different Sizes. Housing Studies, 28(6), 910–927.

Chelmis, C., Qi, W., & Lee, W. (2021). Challenges and opportunities in using data science for homelessness service provision. Paper presented at the pp. 128–135.

Culhane, D. P., & Kuhn, R. (1998). Patterns and determinants of public shelter utilization among homeless adults in New York City and Philadelphia. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 17(1), 23–43.

Gao, Y., Das, S., & Fowler, P. J. (2017). Homelessness service provision: A data science perspective. Paper presented at the , WS-17-01 - WS-17-15. pp. 20–24. Retrieved from Scopus database. Retrieved from

Muñoz, M., Panadero, S., Santos, E. P., & Quiroga, M. Á. (2005). Role of Stressful Life Events in Homelessness: An Intragroup Analysis. American Journal of Community Psychology, 35(1–2), 35–47.

Qi, W., & Chelmis, C. (2023). Evaluating algorithmic homeless service allocation. Journal of Computational Social Science, 6(1), 59–89.

Contact
Gloria Puchol Ros
gloria.puchol@sjd.es

St. John of God Valencia
www.sanjuandediosvalencia.org
+0034 963 66 50 70

c/ Luz Casanova 8
46009 Valencia
Spain