Geojournal of Tourism and Geosites
ISSN 2065-0817, E-ISSN 2065-1198

Year XVII, vol. 56, no. 4, 2024, p.1448-1453
DOI 10.30892/gtg.56401-1315

PUBLIC TRANSPORTATION THROUGH THE CLUSTERING ANALYSIS OF BUS STOPS. CASE STUDY MANIZALES, COLOMBIA

Jorge Alberto MONTOYA*
Universidad Nacional de Colombia, Sede Manizales, Facultad de Ingeniería y Arquitectura, Departamento de Ingeniería Civil, Grupo de Investigación en Movilidad Sostenible, Manizales, Colombia, e-mail: joamontoyago@unal.edu.co

Diego Alexander ESCOBAR
Universidad Nacional de Colombia, Sede Manizales. Facultad de Ingeniería y Arquitectura, Departamento de Ingeniería Civil, Grupo de Investigación en Movilidad Sostenible, Manizales, Colombia, e-mail: daescobarga@unal.edu.co

Carlos Alberto MONCADA
Universidad Nacional de Colombia, Sede Bogotá. Facultad de Ingeniería, Departamento de Ingeniería Civil y Agrícola, Programa de Investigación en Tránsito y Transporte – PIT, Bogotá, Colombia, e-mail: camoncadaa@unal.edu.co

Citation: Montoya, J.A., Escobar, D.A., & Moncada, C.A. (2024). PUBLIC TRANSPORTATION THROUGH THE CLUSTERING ANALYSIS OF BUS STOPS. CASE STUDY MANIZALES, COLOMBIA. *Geojournal of Tourism and Geosites*, 56(4), 1448–1453. https://doi.org/10.30892/gtg.56401-1315

Abstract: The objective of this research is to identify the main characteristics and geographical distribution of bus stops, associated with collective public transport systems, through data analysis techniques that can identify clustering patterns based on various characteristics of the stops, such as geographical location, passenger flow, service demand, land use, etc. number of routes available, slope of the road, signage, among other relevant factors. The main result of this study is that the current layout of the stops is distributed according to the main characteristics of their environment, allowing us to conclude that the grouping of transport stops facilitates the interpretation of the location and operability of the stops, thus allowing to improve decision-making in the implementation of stops, guaranteeing an adequate operation of the service without compromising coverage and travel time of users.

Keywords: clustering, public transportation, bus stops, coverage

* * * * * *

INTRODUCTION

The operation of transport systems is closely linked to the layout and operation of bus stops, which serve as modal interchange points and access points to the system for users; Whereas, an adequate location of bus stops allows an operational balance between the distance that a user must travel on foot and the cost in time for the vehicle to perform boarding and disembarking maneuvers of passengers in very close segments. In this sense, understanding the operation and characteristics of each bus stop allows for an adequate structuring of the location and area of influence of the stop, ensuring that users who use the stop have an adequate infrastructure and maximize the use of the transport system.

Not only do these stops serve the basic function of being embarkation and disembarkation points, but they also play a crucial role in creating an accessible, safe and comfortable environment for public transport users. Therefore, well-designed bus stops are critical to optimizing urban mobility and improving the user experience; From the strategic choice of location to the physical infrastructure and information provided, the academic literature has extensively explored the importance of bus stops and their impact on urban mobility. Previous research has examined topics such as inclusive design for people with disabilities (Burton and Mitchell, 2006), the influence of the location of Transit Hubs on an intermodal network (Yuan and Yu, (2018)); however, there are no significant valuations related to coverage, focusing only on the area of action, which is based on radial structures of between 250 and 400 meters (Furth and Rahbee, 2000; Ibeas, 2010), leaving aside the surrounding dynamics and behaviors of the vehicular flow, thus generating some problems related to the variation of costs, behavior and perception of users, among others (Gibson et al., 1989; Giannopoulos, 1990; Ibeas et al., 2010).

On the other hand, with respect to data analysis, there are some evaluations related to clustering and analysis in transport issues; These approaches have facilitated a deeper understanding of mobility patterns (Cats, 2023; Strömblad, 2024), enabling the optimisation of efficiency in services and improving decision-making in the field of public transport. Some of these analyses focus on urban mobility, in order to identify behavior patterns in public transport users, using clustering algorithms to segment users into categories according to their route and schedule preferences, thus contributing to more effective transport planning (de Oña et al., 2014); Another type of analysis involved the evaluation of data in real time to improve the operation, combining data analysis with clustering techniques to improve resource allocation and operational management in public transport systems (Lin et al., 2013). Likewise, route and frequency optimization

* Corresponding author

assessments have been developed, using clustering models to segment public transport demands and design more efficient routes and schedules Shima et al. (2021). This background illustrates how the application of clustering models and data analysis has been fundamental to advance in the understanding and optimization of public transport, however, no analyses related to bus stops are identified, thus leaving a significant gap to be addressed, with great impact today.

For this reason, the present study proposes to carry out an analysis of the grouping of bus stops in urban collective public transport systems, based on clustering statistical models, using the set of stops in Manizales, Colombia, as a basis.

Manizales, the capital of the department of Caldas in Colombia, is strategically located in the central Andes mountain range, at an altitude of around 2,150 meters above sea level. Its mountainous landscape, surrounded by valleys and mountains, distinguishes it as one of the highest and most picturesque cities in the country. Nicknamed the "City of Open Doors," Manizales is characterized by its colonial charm, steep streets, and a lively cultural scene. In addition to being a prominent educational center with various university institutions, the city is famous for its connection to the coffee industry and its contribution to Colombia's coffee cultural heritage. It has an urban collective public transport system consisting of 901 buses and minibuses, 2 overhead cable lines and a total of 1,214 stops (Alcaldía de Manziales, 2017) to serve a population of 434,006 inhabitants (DANE, 2023) in its urban area (Figure 1). The operation of the system focuses on three main corridors: Santander Aveneu, Paralela Aveneu and Kevin Ángel Aveneu, which support more than 90% of the available routes (62 in total). This means that traffic flow issues in any of these corridors have a significant impact on the overall functioning of the system; On average, the routes in operation cover distances ranging from 25 to 27 kilometers per trip.

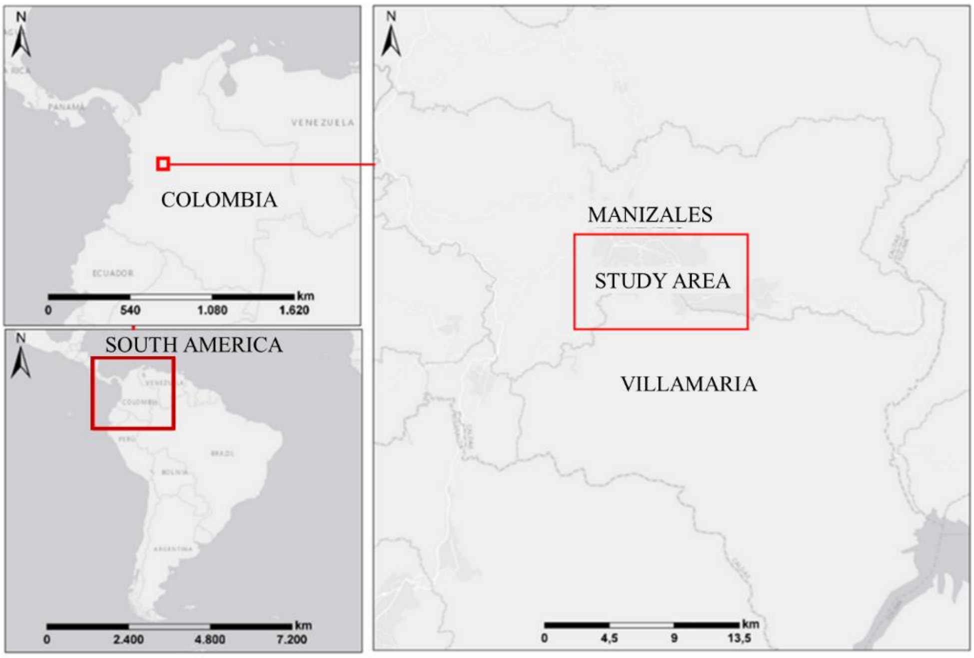

Figure 1. Geographical location of the study area (Source: authors)

METHODOLOGY

The methodological structure used consists of 5 consecutive steps, which are described below.

Phase 1. Collection of Information. As a first methodological phase, we proceed to collect the basic information associated with the set of bus stops of the collective public transport system of Manizales. Within this information base, 1214 stops are contemplated, on which data are collected from 53 variables, product of field inventory and desktop assessment, including characteristics related to the urban collective public transport system (TPCU), the direct functionality of the stop, population covered, road safety, network, land uses and urban facilities. Figure 2 shows the variables considered in the collection of information.

Phase 2. Validation of information and structuring of data for modeling. Once the information on all the stops in the city has been collected, the existence of the stops is validated, as well as the information collected, in order to identify possible inconsistencies and non-representative parameters for the classification. In this order of ideas, a total of 1061 bus stops are obtained, existing in the city, and a total of 49 variables, with non-redundant information, which can be used within the model. Next, the data associated with the continuous variables are normalized, thus allowing the original values to be adjusted to a specific range 0-1, which allows the elimination of different rating scales, improving the convergence of the algorithm, as well as improving the interpretation and analysis of the results.

Phase 3. Quantification of groups. After validating the information, as well as normalizing the data, we proceed to evaluate and estimate the number of groups for the execution of the clustering algorithm; For the process, 2 verification mechanisms are used, as a first assessment we have the elbow method used to identify the optimal number of groups in a data set. The idea behind this method is to observe how the sum of intra-cluster squares varies as a function of the number of clusters (Shi, 2021). The sum of intra-cluster squares measures the dispersion of points within each cluster. The second method of verification is the silhouette method, considered as a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation) (Rousseeuw, 1987). Figures 3 and 4 show the evaluations obtained for the dataset using the 2 methods, identifying that the optimal number of groups for classification is 4.

	DIRECT FUNCTIONS OF BUS STOP											
EXISTENCE OF THE STOP	INFORMATION ON URBAN COLLECTIVE PUBLIC TRANSPORT ROUTES	BAY	AVAILABILITY OF BENCH FOR SERVICE WAITING ACTIONS	EXISTENCE OF COVER FOR PROTECTION AGAINST RAIN OR SUN	ON-STOP BENCH FUNCTIONALITY	FUNCIONALIDAD DE LA CUBIERTA EN LA PARADA	AVAILABILITY OF EMERGENCY COMMUNICATION DEVICE	AVAILABILITY OF ITEM FOR GARBAGE COLLECTION	AVAILABILITY OF MAP FOR LOCATION IN THE CITY	EMERGENCY COMMUNICATION DEVICE FUNCTIONALITY	AVERAGE GRADIENT OF THE ROADS SURROUNDING THE STOP	AVAILABILITY OF ACCESS FOR USERS WITH REDUCED MOBILITY
V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13

POPULATION COVERED			URBAN FACILITIES								
POPULATION COVERED WITHIN A RADIUS OF 100 METERS	POPULATION COVERED WITHIN A RADIUS OF 200 METERS	POPULATION COVERED WITHIN A RADIUS OF 300 METERS	EDUCATIONAL FACILITIES WITHIN A RADIUS OF 100 METERS	EDUCATIONAL FACILITIES WITHIN A RADIUS OF 200 METERS	EDUCATIONAL FACILITIES WITHIN A RADIUS OF 300 METRES	HEALTH FACILITIES WITHIN A RADIUS OF 100 METERS	HEALTH FACILITIES WITHIN A RADIUS OF 200 METERS	HEALTH FACILITIES WITHIN A RADIUS OF 300 METERS	RECREATIONAL FACILITIES WITHIN A RADIUS OF 100 METERS	RECREATIONAL FACILITIES WITHIN A RADIUS OF 200 METERS	RECREATIONAL FACILITIES WITHIN A RADIUS OF 300 METERS
V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25

LAND USE						URBAN COLLECTIVE PUBLIC TRANSPORT SYSTEM (TPCU)				
RESIDENTIAL LAND USE	RECREATIONAL LAND USE	EDUCATIONAL LAND USE	INSTITUTIONAL LAND USE	LAND USE COMMERCE	LAND USE SERVICES	NUMBER OF URBAN COLLECTIVE PUBLIC TRANSPORT ROUTES AVAILABLE	OPERATIONAL SPEED OF URBAN COLLECTIVE PUBLIC TRANSPORT ON THE ROADS THAT CONVERGE AT THE BUS STOP	OPERATIONAL SPEED OF URBAN COLLECTIVE PUBLIC TRANSPORT ON THE ROADS THAT CONVERGE AT THE BUS STOP	PERCENTAGE OF TRIPS GENERATED IN TPCU, IN THE ZAT WHERE THE WHEREABOUTS ARE LOCATED	PERCENTAGE OF TRIPS ATTRACTED AT TPCU, IN THE ZAT WHERE THE WHEREABOUTS ARE LOCATED
V26	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36

ROAD SAFETY						
VERTICAL SIGNAGE IS AVAILABLE?	STATUS OF VERTICAL SIGNAGE	THERE IS A ZEBRA CROSSING?	ZEBRA CROSSING STATUS	THERE IS A DEMARCATION OF BUS STOP	STATUS OF THE BUS STOP DEMARCATION	GENERAL CONDITIONS OF VISIBILITY OF THE BUS STOP OF USERS
V37	V38	V39	V40	V41	V42	V43

NETWORK								
OPERATING SPEED OF THE SYSTEM ON THE TRACKS CONVERGING ON THE BUS STOP	TYPOLOGY	VEHICULAR VOLUME OF THE ROADS THAT CONVERGE AT THE BUS STOP	PERCENTAGE OF LIGHT VEHICLES OF THE TOTAL VOLUME ON ROADS CONVERGING AT THE BUS STOP	AVERAGE VOLUME/CAPACITY RATIO OF ROADS CONVERGING AT THE BUS STOP	NUMBER OF OFF-STREET PARKING SPACES WITHIN A RADIUS OF 100 METRES	NUMBER OF LEGAL ON-ROAD PARKING SPACES WITHIN A RADIUS OF 100 METERS	NUMBER OF ILLEGAL PARKING ON THE ROAD WITHIN A RADIUS OF 100 METERS	NUMBER OF LANES ON ROADS CONVERGING AT THE BUS STOP
V44	V45	V46	V47	V48	V49	V50	V51	V52

Figure 2. Variables used (Source: authors)

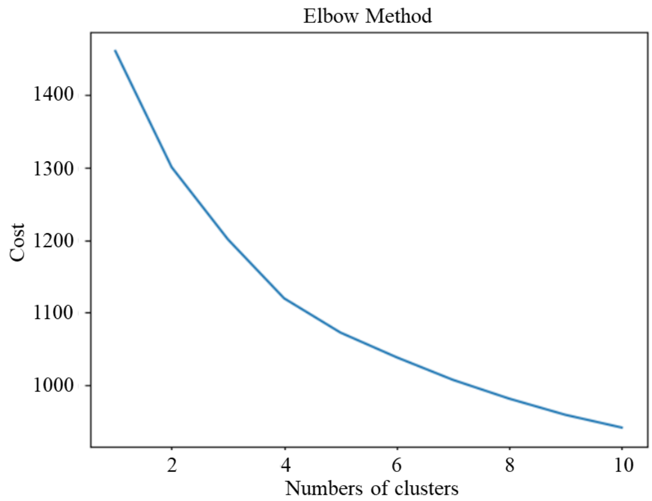

Figure 3. Elbow method (Source: authors)

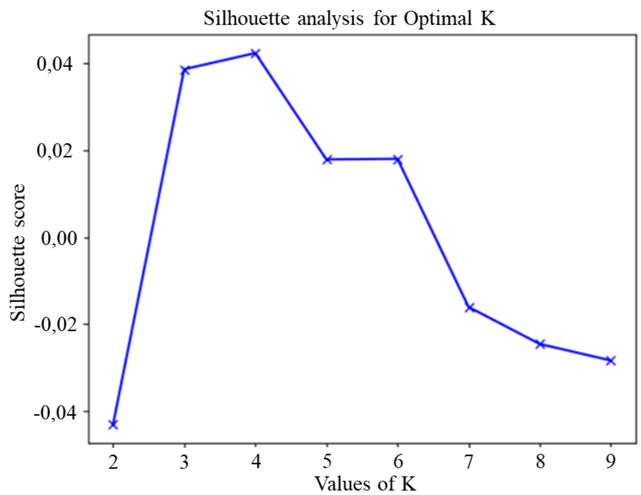

Figure 4. Silhouette method (Source: authors)

Phase 4. Clustering. Once the normalization, characterization and definition of the optimal number of groups for classification have been carried out, the k-prototypes algorithm is used as a grouping mechanism. The K-Prototypes clustering method, an extension of the K-Means algorithm, is used to group data with numerical and categorical variables; By combining numerical centroids and categorical examples, this method, proposed by Z. Huang in 1997, Used in various types of analysis (Szepannek et al., 2024; Wei, 2024), facilitates grouping into mixed data sets, assigning prototypes that represent the diversity of features, allowing a more complete analysis of heterogeneous data (Huang, 1997).

Phase 5. Cluster spatialization. As a final methodological phase, we proceed to specialize the Bus Stop classified by means of geographic information systems, in order to identify and analyze spatial patterns in the current arrangement of the stops. EsRi's ArcMap software is used as a digitization tool.

```
from sklearn.preprocessing import MinMaxScaler

# Calculate the cost for a given number of clusters
def calculate_cost(kproto, X):
    cost = kproto.cost_
    return cost

# Elbow algorithm with data normalization
def elbow_algorithm(X, categorical_features_idx, max_clusters):
    cost = []

    for num_clusters in range(1, max_clusters+1):

        kproto = KPrototypes(n_clusters=num_clusters, verbose=2, max_iter=20)
        kproto.fit(X, categorical=categorical_features_idx)
        cost.append(calculate_cost(kproto, X))

    # Plot the cost against the number of clusters
    plt.plot(range(1, max_clusters+1), cost)
    plt.xlabel('Number of Clusters')
    plt.ylabel('Cost')
    plt.title('Elbow Method')
    plt.show()
```

Figure 5. Base algorithm used (Source: authors)

	V12	V14	V15	V16	V17	V18	V19	Cluster
mean	0.231405	0.246551	0.286915	0.086451	0.041284	0.096636	0.134217	0
	0.224088	0.230951	0.299514	0.227852	0.145238	0.201905	0.258201	1
	0.292706	0.392029	0.461567	0.340723	0.110577	0.194231	0.231303	2
	0.274956	0.444617	0.505788	0.350127	0.292722	0.358861	0.368143	3
	V20	V21	V22	V23	V24	V25		Cluster
mean	0.014526	0.025994	0.042813	0.033202	0.094801	0.169864		0
	0.047619	0.090476	0.134199	0.062585	0.161376	0.260606		1
	0.018029	0.022236	0.041521	0.034341	0.096154	0.17701		2
	0.041139	0.063687	0.094649	0.038879	0.123769	0.217491		3
	V32	V33	V34	V35	V36	V44	V46	Cluster
mean	0.093068	0.663424	0.051308	0.399025	0.381935	0.426997	0.183868	0
	0.204683	0.695506	0.045876	0.373514	0.372679	0.480505	0.246712	1
	0.062099	0.639805	0.888558	0.406692	0.449604	0.287486	0.116181	2
	0.091561	0.659027	0.081699	0.396147	0.440916	0.299219	0.154735	3
	V47	V48	V49	V50	V51	V52		Cluster
mean	0.474175	0.2013	0.0421	0.012997	0.039755	0.365225		0
	0.516802	0.342718	0.041746	0.006667	0.020952	0.454422		1
	0.127232	0.01446	0.007372	0.001923	0.068029	0.317995		2
	0.47254	0.129771	0.015928	0.010918	0.062342	0.322333		3

Figure 6. Mean values in variables evaluated by classification group (Source: authors)

V3	V4	V5	V6	V7	V8	V9	V11	Cluster
NO	NO	NO	VACIO	VACIO	NO	NO	VACIO	0
NO	SI	SI	BUENO	BUENO	NO	NO	VACIO	1
NO	NO	NO	VACIO	VACIO	NO	NO	VACIO	2
NO	NO	NO	VACIO	VACIO	NO	NO	VACIO	3
V13	V26	V27	V28	V29	V30	V31	V37	Cluster
NO	SI	NO	NO	NO	NO	NO	NO	0
NO	SI	NO	NO	NO	NO	NO	SI	1
NO	SI	NO	NO	NO	NO	NO	NO	2
NO	SI	NO	NO	NO	NO	NO	SI	3
V38	V39	V40	V41	V42	V43	V45		Cluster
VACIO	NO	VACIO	NO	VACIO	VISIBLE	Colectora		0
BUENO	SI	BUENO	NO	VACIO	VISIBLE	Principal		1
VACIO	NO	VACIO	NO	VACIO	VISIBLE	Secundaria		2
BUENO	NO	VACIO	SI	VACIO	VISIBLE	Colectora		3

Figure 7. Mode of Categorical Variables Assessed by Classification Group (Source: authors)

RESULTS AND DISCUSSION

As a result of the grouping of the stops using the Kprototypes method, a total of 4 classes are obtained, from the inventoried information, with a general balance of 208 stops associated with cluster 0, 210 associated with cluster 1, 328 stops in cluster 2 and 315 for cluster 3. As a general estimate, there is a significant balance between clusters, with a difference of up to 110 stops in the identified maximum and minimum groupings. Regarding the characteristics identified, we have Figures 6 and 7, in which the mean ratings obtained for the numerical variables and the modes identified in the categorical variables are appreciated. With this information, it is possible to identify that the stops associated with cluster number 1 mainly relate those stops with the existence of physical infrastructure and signaling in their vicinity, which in turn, have a greater availability of equipment in their immediate environment, as well as greater operational characteristics of the road network in which they are located. such as speed, vehicle volume and availability of transport routes, this being a characteristic of stops located in main corridors of the city. In this order of ideas, this grouping refers to stops of great importance and of greater use within the system.

Regarding cluster 0, it is possible to identify that it corresponds to the group of stops with less availability of physical infrastructure, as well as fewer facilities and users covered in the surrounding area, this behavior could refer to stops distributed in an environment of lower demand, such as peripheral areas or access to the city. as well as interurban corridors that connect population centers or areas of greater activity. On the part of cluster 2, some characteristics related to the greater amount of population covered in the nearby area and the lower availability of facilities are identified; Likewise, a lack of physical infrastructure is identified, and less vehicular volume in the operational areas, these assessments could refer to types of stops in residential areas of the city, in which there is a greater population served, but preserving the characteristics of a collector or local road, in which the volumes are low and the proportion of public transport can be higher.

Finally, cluster number 3 relates a high population coverage, as well as an adequate provision of educational facilities in the surrounding area, and moderate existence of other types of facilities; Likewise, it refers to the existence and good condition of physical infrastructure, in addition to a high attraction of trips for a low operational speed; This could be inferred from a group of operational stops in educational and residential environments within the city. Finally, Figure 8 is presented, in which the spatial distribution of the stops is appreciated, according to the group to which they belong, it is possible to identify that the valuations referred to by the model show great agreement with the current urban structure of the city, in which case the stops of cluster 1 are located on main corridors. cluster 0 is located in external areas and inter-neighbourhood corridors; The bus stops associated with type 2 are mainly located in residential areas and finally the bus stops associated with type 3 are located in areas of residential concentration, with some strips of particular interest within the city.

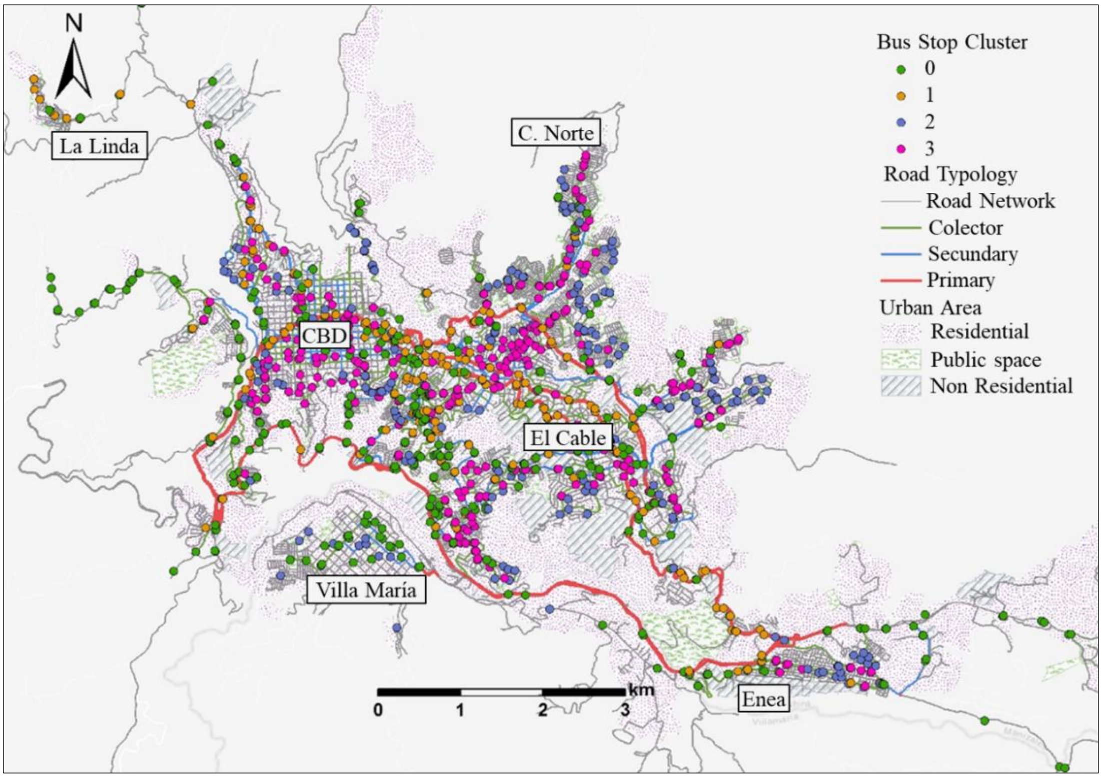

Figure 8. Spatial Distribution of Stops by Classification Group (Source: authors)

CONCLUSION

As conclusions of the evaluation carried out, we can say that the current layout of the collective public transport stops of the city of Manizales has a grouping of characteristics of the existing infrastructure, as well as the conditions of population coverage and surrounding equipment, which allows to identify the basic patterns for the implementation and characterization of new bus stops. On the other hand, the identified evaluations show that, despite having a large number of variables for the grouping of stops, it is not representative to use all of them since it is possible to analyze and group from some variables with the greatest impact on the process.

Finally, it is possible to appreciate how the valuation by classification allows to expand the knowledge related to the analysis of public transport stops, which have not been explored in recent years.

Author Contributions: Conceptualization, J.A.M. and D.A.E.; methodology, J.A.M. and C.A.M.; software, J.A.M.; validation, J.A.M., D.E. and C.A.M.; formal analysis, D.A.E. and J.A.M.; investigation, C.A.M. and D.A.E.; data curation, C.A.M. and D.A.E.; writing - original draft preparation, J.A.M. and D.A.E.; writing - review and editing, J.M, D.A.E. and C.A.M.; visualization, J.A.M. and D.A.E.; supervision, C.A.M.. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study may be obtained on request from the corresponding author.

Acknowledgements: The following study is part of the analysis and development of the Doctoral Thesis of Engineer Jorge Alberto Montoya Gómez of the National University of Colombia, as a requirement for his training and aspiration to the title of Doctor in Engineering, under the analysis project "*Propuesta Metodológica Para El Cálculo De Cobertura Variable Desde Las Paradas De Transporte Público, Mediante La Aplicación De Modelos De Accesibilidad Geográfica"*. The authors express their sincere thanks to the members of the Sustainable Mobility Research Group of the National University of Colombia, Manizales Campus, for their support in the collection of field information, as well as their most sincere thanks to Engineer Daniel Gonzales Núñez, for his support in the analysis and structuring of the classification model.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Alcaldía de Manziales. (2017). *Plan Maestro de Movilidad de Manizales: Línea base y diagnóstico de la situación actual de la movilidad de Manizales [Manizales Mobility Master Plan: Baseline and diagnosis of the current situation of mobility in Manizales]*. (In Spanish). Manizales. https://infimanizales.com/negocio-2/gestion/plan/

Burton, E., & Mitchell, L. (2006). Inclusive urban design: Streets for life. *Architectural Press.* http://dx.doi.org/10.4324/9780080456454

Cats, O. (2023). Identifying human mobility patterns using smart card data. *Transport Reviews*, 44(1), 213–243. https://doi.org/10.1080/01441647.2023.2251688

Departamento Administrativo Nacional De Estadística – DANE (2023). *Proyecciones de población total por sexo y grupos de edad de 0 hasta 80 y más años. [Total population projections by sex and age groups 0-80 and over]*. (In Spanish). http://www.dane.gov.co/files/ investigaciones/poblacion/proyepobla06_20/7Proyeccio nes_poblacion.pdf

De Oña, R., Lopez, G., Diez de los Rios, F., & De Oña, J. (2014). Cluster analysis for diminishing heterogeneous opinions of service quality public transport passengers, *Procedia - Social and Behavioral Sciences* 162, 459 – 466. https://doi.org/10.1016/j.sbspro.2014.12.227

Furth, P., & Rahbee, A. (2000). Optimal bus stop spacing through dynamic programming and geographic modeling. *Transportation Research Record*. 1731(1), 15-22. https://doi.org/10.3141/1731-03

Giannopoulos, G. (1990). Bus Planning and Operation in Urban Areas: A Practical Guide. *Avebury, Gower Publishing Company*. Ltd., UK. https://trid.trb.org/View/357471

Gibson, J., Baeza, I., & Willumsen, L. (1989). Bus-stop, congestion and congested bus-stop. *Traffic Engineering and Control*, 30(6), 291 – 302. https://www.researchgate.net/publication/285765251_Bus-stops_congestion_and_congested_bus-stopssad

Huang, Z. (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining. *CSIRO Mathematical and Information Sciences.* https://www.diag.uniroma1.it/~sassano/STAGE/Fast_Clustering.pdf

Ibeas, Á., Dell´Olio, L., Borja, A., & Zainz, O. (2010). Optimizing bus stop spacing in urban áreas. *Transportation Research Part E*. 46, 446-458. http://dx.doi.org/10.1016/j.tre.2009.11.001

Lin, Y., Yang, X., Zou, N., & Jia, L. (2013). Real-Time Bus Arrival Time Prediction: Case Study for Jinan, China. *Journal of Transportation Engineering* 139(11):1133-1140. http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000589

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *J Wireless Com Network* 2021, 31. https://doi.org/10.1186/s13638-021-01910-w

Strömblad, E. (2024). Identifying mobility segments for leisure travel: A cluster analysis based on a one-month travel survey, *Transportation Research Part A: Policy and Practice* 181. https://doi.org/10.1016/j.tra.2024.104001

Szepannek, G., Aschenbruck, R., & Wilhelm, A. (2024). Clustering large mixed-type data with ordinal variables. *Advances in Data Analysis and Classification*. https://doi.org/10.1007/s11634-024-00595-5

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Shima, Y., Kadir, R., & Ali, F. (2021). A Novel Approach to the Optimization of a Public Bus Schedule Using K-Means and a Genetic Algorithm, *IEEE Access*, 9, 73365-73376. https://doi.org/10.1109/ACCESS.2021.3080508.

Wei, Q. (2024). Analysing university student pension insurance using the K-prototypes algorithm and logistic regression model. *International Journal of Information and Communication Technology*, 24(6), 92 – 102. https://doi.org/10.1504/IJICT.2024.138563

Yuan, Y., & Yu, J. (2018). Locating transit hubs in a multi-modal transportation network: A cluster-based optimization approach, *Transportation Research Part E: Logistics and Transportation Review*, 114, 85-103. https://doi.org/10.1016/j.tre.2018.03.008

Article history: Received: 17.05.2024 Revised: 29.08.2024 Accepted: 16.09.2024 Available online: 01.10.2024