# Resumen

# En la era del big data, la abundancia de información presenta oportunidades sin precedentes para obtener conocimientos valiosos sobre el comportamiento y las actividades humanas. Este estudio se enfoca en aprovechar el poder de los datos recopilados de aplicaciones móviles utilizadas por estudiantes de la Universidad de Azuay, con el objetivo de identificar la ubicación de sus hogares. Al filtrar estas ubicaciones se busca mejorar el estudio de los patrones de movilidad y los puntos de interés más frecuentes entre los estudiantes, con el propósito de obtener información útil para la toma de decisiones en temas relacionados con la movilidad y la planificación del transporte en la ciudad de Cuenca.

# Abstract

In the era of big data, the abundance of information presents unprecedented opportunities to gain valuable insights into human behavior and activities. This study focuses on harnessing the power of data collected from mobile applications used by students at the University of Azuay, with the aim of identifying the location of their homes. By filtering these locations, the study seeks to improve the understanding of mobility patterns and the most frequent points of interest among students, to obtain useful information for decision-making in mobility-related issues and transportation planning in the city of Cuenca.

***Keywords:*** Big data, mobile app data, student homes, mobility patterns, points of interest

Introduction

Gaining insights into the mobility patterns and points of interest among university students is crucial in developing effective strategies to optimize transportation for both students and the entire city. In the scope of this study, our objective is to identify and analyze the precise locations of student residences at the University of Azuay in the city of Cuenca, utilizing mobile app data.

Mobile phone data can be a useful source for official statistics, but there are challenges and uncertainties that need to be addressed before they can be used. Detecting home locations from mobile phone data and analyzes the performance of five home detection algorithms, offering recommendations for more reliable use of mobile phone data in official statistics.

Some studies like [1] demonstrates the potential of using social media data, specifically Twitter, for studying urban mobility patterns. However, the authors note the limitations of the data, such as biases in social media usage and the need for ongoing validation and improvement of the methods used. In other study [2] about Home Detection Algorithms (HDAs) using mobile phone data for official statistics. The authors found that the type of data stream used, and the algorithm choice significantly influenced the accuracy of home detection. Algorithms based on weekdays and/or nighttime records were the most accurate and using Extended Detail Records (XDRs) with specific algorithms yielded the best results. Daytime records and spatial perimeter-based algorithms had low accuracy and should be avoided. XDRs and Cellular Positioning Records (CPRs) were more resilient to data reduction compared to Call Detail Records (CDRs). The study had limitations in sample size and lack of demographic information. In [3] concludes that mobile positioning data can be used to monitor population geography and mobility, particularly in socially unstable or rapidly developing regions. However, further work is needed to standardize the data and model for different sources and conditions. The methodology is promising for geographical research and offers opportunities for real-time monitoring tools, geographical applications, tourism development, traffic management, urban planning, and optimizing network services. Detecting the location of households from mobile phone data is a key challenge for the use of this data source in official statistics. The authors argue that current address detection methods suffer from a lack of consensus on criteria and limited validation capabilities. The article presents an analysis of five address detection algorithms applied to a large French Call Detail Record (CDR) dataset, showing that the choice of criteria in Home Detection Algorithms (HDAs) influences address location detection in up to 40% of users, and that HDAs perform poorly when compared to a validation dataset [4]. In [5] the authors used mobile phone data and metadata, including call detail records (CDRs) enriched with gender, socioeconomic segment, and number of phone lines registered under that number. then analyzed the data to reveal a gender gap in mobility and mapped this mobility gap over administrative divisions to observe the association with lower income and lack of public and private transportation options. The paper concludes that there is a gender gap in urban mobility, where women visit fewer unique locations than men and distribute their time less equally among such locations. This mobility gap is associated with lower income and lack of public and private transportation options. In [6] the paper investigates the performance and capabilities of five popular criteria for home detection based on a very large mobile phone dataset from France. The study shows that most Home Detection Algorithms (HDAs) suffer from "blind" deployment of criteria to define homes and from limited possibilities for validation. The paper introduces a data-driven framework to assess the spatial uncertainty related to the application of HDAs. The findings appropriate spatial uncertainty in HDA and, in extension, for detection of meaningful places. The study shows how spatial uncertainties on the individuals' level can be assessed in absence of ground truth annotation, how they relate to traditional, high-level validation practices and how they can be used to improve results for, e.g., nation-wide population estimation. Therefore, the paper concludes that the proposed framework can be used to improve the accuracy of home detection algorithms and population estimation. In [7] the authors discuss the importance of knowing user location in the digital world, not only in real-time but also predicting future locations. It is necessary to semantically label a place, particularly detecting the most probable home location for a given user. The paper aims to provide insights on the differences among the ways how different types of human digital trails represent actual mobility patterns and the differences between the approaches interpreting those trails for inferring said patterns. The paper starts with an example showing how human mobility patterns described by means of radius of gyration are different for Flickr social network and dataset of bank card transactions. The paper considers several methods for home location definition known from the literature and demonstrates that although for bank card transactions they provide highly consistent results, home location definition detection methods applied to Flickr dataset happen to be way more sensitive to the method selected, stressing the paramount importance of adjusting the method to the specific dataset being used. In universities, there is a need to enhance student mobility, and various strategies have been suggested to achieve this goal. To implement effective policies, it is crucial to have dynamic origin-destination matrices (OD) that represent different scenarios. The widespread use of mobile devices has made it possible to automatically extract daily travel data through tracking apps, eliminating the need for traditional trip surveys. Mendoza et. al [8] presents a new methodology for extracting multi-day mobility demand in universities using logs obtained from dedicated apps regularly used by students. The proposed approach was evaluated using real-life logs from a representative group of students over a five-month period. The results showed that this approach is effective in obtaining average demand data, which can be utilized in planning mobility strategies, as long as continuous tracking of mobility data is feasible through mobile devices.

The primary objective of our study is to filter out student homes from the collected mobile app data and categorize them. By accurately identifying these locations, we can lay the foundation for a more comprehensive analysis of student mobility patterns and points of interest.

This research is part of a larger investigation focused on understanding the unique needs and preferences of University of Azuay students. By identifying their homes, we gain crucial insights into their daily commute patterns, transportation modes, and the places they frequent. Such information enables the university authorities and municipal stakeholders, such as the city of Cuenca, to make data-driven decisions and develop tailored mobility plans that cater to the specific needs of the student population.

In this paper, we present the methodology used to collect and preprocess the mobile app data, with a specific focus on identifying student homes. We discuss the techniques employed to filter and categorize these locations accurately. Additionally, we delve into the analysis of student mobility patterns and points of interest derived from the identified residences. The results and their implications for decision-making are presented and discussed in detail.

The findings from this study will provide valuable to the understanding their mobility patterns, decision-makers can develop targeted interventions, optimize transportation systems, and create a supportive environment for the student community. This research underscores the significance of leveraging mobile app data to gain a comprehensive understanding of student mobility patterns and points of interest, facilitating evidence-based decision-making and the development of effective mobility plans.

1. Joaquín Osorio-Arjona, Juan Carlos García-Palomares: "Social media and urban mobility: Using Twitter to calculate home-work travel matrices." Cities 89, 268-280 (2019).
2. Pappalardo, L., Ferres, L., Sacasa, M. et al.: "Evaluation of home detection algorithms on mobile phone data using individual-level ground truth." EPJ Data Sci. 10, 29 (2021)
3. Rein Ahas, Siiri Silm, Olle Järv, Erki Saluveer, Margus Tiru: "Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones." Journal of Urban Technology 17(1), 3-27 (2010)
4. Vanhoof, M., Reis, F., Ploetz, T., & Smoreda, Z. (2016). Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics. Journal of Official Statistics, 32(4), 873-892.
5. Laetitia Gauvin, Michele Tizzoni, Simone Piaggesi, Andrew Young, Natalia Adler, Stefaan Verhulst, Leo Ferres, Ciro Cattuto: "Gender gaps in urban mobility." EPJ Data Science 8(1), 1-19 (2019).
6. Vanhoof, M., Reis, F., Ploetz, T., & Smoreda, Z. (2016). Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics. Journal of Official Statistics, 32(4), 873-892.
7. Piotr Krawiec, Tomasz Kajdanowicz, Przemyslaw Kazienko: "Choosing the right home location definition method for the given dataset." In International Conference on Computational Science, 3-12. Springer, Cham (2018).
8. Iván Mendoza, Gustavo Alvarez, Mateo Coello, Joaquín López and Pablo Carvallo, "Automatic estimation of demand matrices for universities through mobile devices," 2020 IEEE ANDESCON, Quito, Ecuador, 2020, pp. 1-6, doi: 10.1109/ANDESCON50619.2020.9272071.