

Beyond The Bundled Strings of GPS Traces: Classifying City Bike Users through Pseudonymous Identifiers

M. Tsaqif Wismadi

In the domain of transport data science, GPS traces have emerged as a prominent data source for understanding mobility behaviour. These spatial records provide invaluable insights into the geography of movement, offering the ability to identify hotspots of activity, observe temporal dynamics over hours, days, or seasons, and explore behavioural correlations through spatial and statistical analysis. When enriched with additional spatial parameters—such as proximity to amenities, land use types, or transportation infrastructure—these traces facilitate deeper inferential modelling. However, such analyses are often limited to aggregate trends. Understanding individual-level patterns remains a methodological challenge, especially in the absence of personal user attributes such as age, gender, or occupation.



INTRODUCTION

This study addresses that gap by proposing a classification method based solely on user pseudonyms identifiers assigned to each user that ensure privacy while maintaining uniqueness. Using GPS data from Trondheim's public bike-sharing system, this research develops a typology of user behaviour grounded in trip frequency, thereby enabling analysis beyond aggregate-level patterns without compromising user anonymity.

Although pseudonymous identifiers do not reveal demographic details, they can be leveraged to assess usage frequency—a behavioural indicator with considerable analytical value. By introducing a derived variable termed '*user_trip_count*', each pseudonymous user is assigned a numerical value reflecting the total number of trips completed within a defined period (e.g., a calendar year). Based on the distribution of these counts, users

are then classified into four distinct categories: *frequent* (more than 36 trips per year), *occasional* (5–36 trips), *seldom* (2–5 trips), and *one-time* (exactly 1 trip). This classification is stored in a new attribute called '*user_classification*'.

The rationale for this approach lies in the assumption that frequency reflects intent and behaviour. Users in the *frequent* and *occasional* categories are likely to be regular commuters, utilising the service for daily travel. In contrast, *seldom* users may represent individuals engaging in leisure cycling or irregular trips. *One-time* users, meanwhile, are plausibly tourists or temporary visitors. These behavioural groupings enable a more granular understanding of how different user types interact with the bike-sharing system across space and time.

METHOD

When visualised, spatial distinctions and different intensity emerge between user groups (Figures 1–2). This foundational classification opens pathways for more advanced behavioural analyses. The next phase of the research involves integrating external variables—such as elevation change, land use, proximity to public facilities, and weather conditions—to statistically assess their influence on route choice and trip frequency among each user group. By conducting inferential tests across these dimensions, it becomes possible to identify which urban or infrastructural attributes are more relevant for commuter cyclists as opposed to leisure users.

Ultimately, this approach offers a privacy-preserving yet behaviourally meaningful method to distinguish between types of bike-share users. The implications extend to both research and policy. Urban planners and transport authorities can utilise such classifications to tailor service provision, whether by prioritising infrastructural improvements along commuting routes or by enhancing access and amenities in areas used by casual or tourist cyclists. Depending on a city's strategic goals—be it increasing active commuting, supporting tourism, or promoting inclusive mobility—bike-sharing schemes can be adapted to better serve diverse user needs.

RESULT

When visualised, spatial distinctions and different intensity emerge between user groups (Figures 1–2). This foundational classification opens pathways for more advanced behavioural analyses. The next phase of the research involves integrating external variables—such as elevation change, land use, proximity to public facilities, and weather conditions—to statistically assess their influence on route choice and trip frequency among each user group. By conducting inferential tests across these dimensions, it becomes possible to identify which urban or infrastructural attributes are more relevant for commuter cyclists as opposed to leisure users.

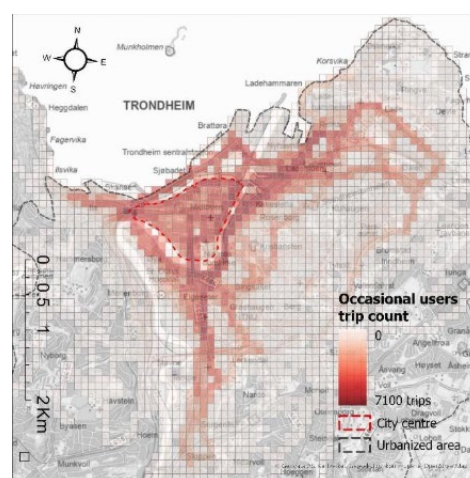
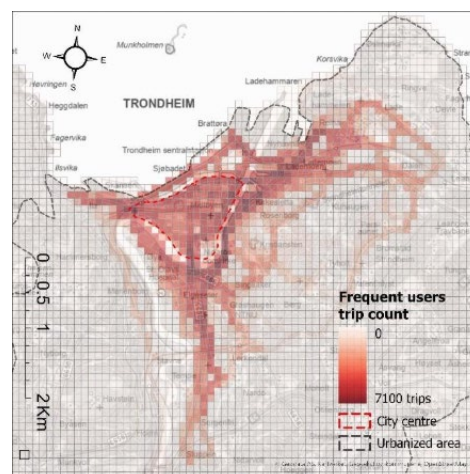
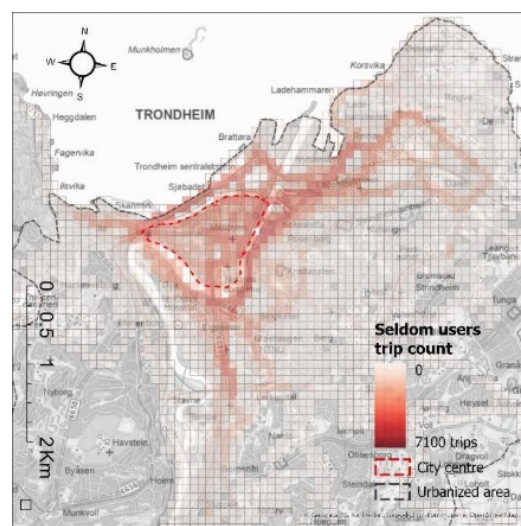


Figure 1: Frequent (left) and occasional (right) users' routes spreads



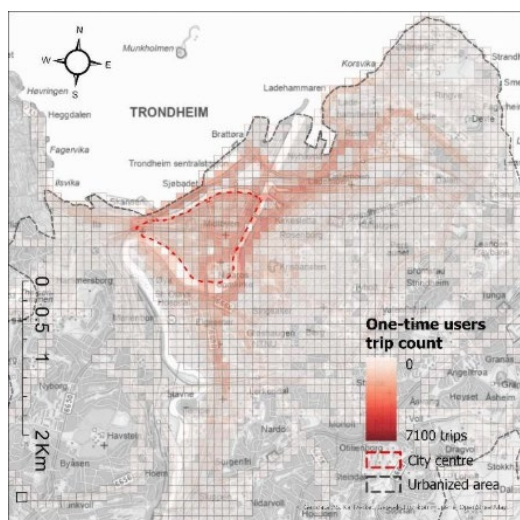


Figure 2: Seldom (left) and one-time (right) users' routes spreads

M. Tsaqif Wismadi received funding from the PERSEUS project, a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101034240.

Ultimately, this approach offers a privacy-preserving yet behaviourally meaningful method to distinguish between types of bike-share users. The implications extend to both research and policy. Urban planners and transport authorities can utilise such classifications to tailor service provision, whether by prioritising infrastructural improvements along commuting routes or by enhancing access and amenities in areas used by casual or tourist cyclists. Depending on a city's strategic goals—be it increasing active commuting, supporting tourism, or promoting inclusive mobility—bike-sharing schemes can be adapted to better serve diverse user needs.



CONCLUSION

In summary, the study demonstrates that even in the absence of demographic data, meaningful user classification is possible through trip behaviour derived from pseudonymous ID number. This method provides a replicable and scalable framework for cities seeking to move beyond aggregate usage statistics toward a more nuanced understanding of mobility behaviour within their bike-sharing systems.