Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 91 (2025) 139–146

Transportation Research Procedia

www.elsevier.com/locate/procedia

The Science and Development of Transport - TRANSCODE 2025

Profiling Socially-Structured Vanpooling Users in Oman: A Data-Driven Approach

Amal Almurfadi[a], Hedi Haddad[b], Zied Bouyahia[c], Ansar-Ul-Haque Yasar[a], Li Li[d], Youssef El Hansali[a]

[a]*Transportation Research Institute (IMOB), Hasselt University, Hasselt, Belgium*
[b]*Universite´ Jean Monnet Saint-Etienne, IUT de Roanne, LASPI, UR, F-42300, ROANNE, FRANCE*
[c]*Ecole Centrale de Lyon, Universite´ de Lyon, LIRIS, CNRS UMR 5205, Ecully, France*
[d]*Tongji University, Shanghai, China*

Abstract

Understanding the social preferences of ridesharing users is very important for the implementation of user-oriented transportation services in several countries throughout the Middle East. This paper reports preliminary results on profiling socially-structured vanpooling users in Oman. We adopted a three-step “cluster-then-classify” approach to analyze a dataset of 3,615 current and potential ridesharing users that we collected from the various regions of Oman. In the first step, we applied the K-Modes clustering (an unsupervised labeling algorithm of categorical data) to identify distinct clusters of riders. Five profiles of riders were identified: “reserved students”, “broad-minded students”, “independent workers”, “dependent workers”, and the “unemployed”. In the second step, we compared the performance of four algorithms (Decision Tree, Random Forest, CatBoost, and Logistic Regression) to classify riders into the five identified classes, leading to an accuracy of 0.91. In the third step, we performed an additional test using a new dataset, and compared the performance of the four classifiers against the modes of each cluster. A best similarity percentage of 90.24% was obtained, suggesting that the five clusters satisfactorily partition riders into distinguishable and interpretable classes. To the best of our knowledge, this is the first research work that applies such an approach to profile shared mobility users in the MENA (Middle East and North Africa) region. The obtained results are expected to help transportation services better address the preferences and needs of riders in Oman and other MENA countries with similar socio-cultural contexts.

Keywords: Ridesharing; User profiling; Clustering and classification; Case study

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: *amal.almurfadi@uhasselt.be*

10.1016/j.trpro.2025.10.019

1. Introduction

The development of reliable transportation systems has a considerable impact on the daily lives and well-being of citizens. Understanding users' preferences is vital to implementing user-oriented transportation planning and services, as this awareness helps service providers collect insights about users' needs and choices of transportation modes. Therefore, analyzing users' preferences has been extensively studied in different areas of the transportation sector, including shared mobility. Ridesharing is one form of shared mobility that has been gaining a lot of popularity around the world from both industrial and academic communities, especially in Europe, North America, and Asia. In the Middle East and North Africa (MENA) region, however, ridesharing has been explored to a limited extent Delatte et al. (2018). As a result, additional research is needed to better understand different aspects of ridesharing Hossain and Mozahem (2022). To fill this gap, in this paper, we are interested in exploring the general concept of social ridesharing in Arab nations through the example of the Sultanate of Oman Shaheen and Cohen (2021) Machado et al. (2018). We are particularly interested in the socially-structured shared mobility, which refers to a transportation service used by a group of socially-compatible riders who share the same mode of transportation for daily and occasional commuting activities Haddad et al. (2022). The work in Haddad et al. (2022) presented a new framework as well as an algorithm for matching and pooling riders based on their social preferences, while seeking to improve their satisfaction with the shared mobility services. We also explore the social ridesharing matching problem by answering two main questions: 1) Are there different profiles or typical groups of Omani ridesharing users? If yes, 2) how predictable are these groups?. For the purpose of this study, we used a dataset of 3,615 current and potential ridesharing users from all over Oman. We used a three-step "cluster-then-classify" approach to analyze the dataset and answer the two questions. In the first step, we used an unsupervised labeling algorithm (K-Modes clustering) to identify the clusters and label data points. Five profiles of riders were identified: "reserved students", "broad-minded students", "independent workers", "dependent workers", and the "unemployed". In the second step, we compared the performance of four algorithms (Decision Tree, Random Forest, CatBoost, and Logistic Regression) to classify riders into the five identified classes, leading to a best accuracy of 0.91. In the third step, we performed an additional test using a new dataset of 124 respondents, and compared the performance of the four classifiers against the modes of each cluster. A best similarity percentage of 90.24% was obtained, suggesting that the five clusters satisfactorily partition riders into distinguishable and interpretable classes. The remainder of this paper is structured as follows. In Section 2, we present a brief summary of the related work on profiling ridesharing users. In Section 3, we present the context, methodology and dataset used in this study. We present our preliminary results of the clustering and classification of riders' profiles in Section 4 and Section 5, respectively. Finally, we outline the conclusion and limitations of the study in Section 6.

2. Related work

Many research works have studied the factors affecting riders' perceptions and attitudes toward social ridesharing Zhang and Zhao (2019); Sarriera et al. (2017); Soltani et al. (2021), but our literature review indicated that only a limited number of studies have addressed the topic of profiling ridesharing users. A "user profile" is defined as the concise representation of the user's interests, characteristics, behaviors, and preferences Eke et al. (2019). Meanwhile, "profiling" is the process of gathering, structuring, and drawing inferences from a user's profile information Eke et al. (2019). Loa et al. (2023) identified three ride-sourcing profiles: traditional users, commute and social users, and mixed users. The traditional users use ride-sourcing for social trips and trips from and to the airport. They are mostly old people with high levels of vehicle ownership, who have income of 100,000 annually Loa et al. (2023). More recently, Dolins et al. (2025) used cross-tabulations and Decision Tree cluster analyses to identify the profiles of users and individuals who refused to use shared autonomous vehicles in Swedish cities. These researchers reported that potential users tend to be "progressive, environmentally conscious men with public transport habits and positive experiences with AVs", while refusers are "often women with traditional values, less formal education, and a preference for private cars, exhibiting concerns about safety and privacy" Dolins et al. (2025). We believe that researchers are continuing to explore the complexities of riders' profiles based on their social preferences to improve user experiences, and address specific needs.

3. Context, methodology and dataset

3.1. Context: ridesharing in Oman

Shared vans and mini-buses represent the most commonly used mode of shared mobility in Oman. Due to its flexibility, affordable prices, and widespread social acceptance, this mode of transportation is widely used among Omanis, either for long-term agreement or occasional transportation services Haddad et al. (2022). Long-term service is usually up to several months, where the occasional service is a trip-based agreements. Both long-term and occasional transportation services are subject to prior agreements between drivers and riders. Currently, there is no dedicated platform for this ridesharing mode, and supply-demand matching is done using word of mouth, social media, and WhatsApp groups. As a result, riders often face difficulties finding compatible co-riders, buses, and prices that correspond to their requirements. To solve this problem, a socially-structured vanpooling framework was proposed by Haddad et al. (2022) to provide a formal formulation of the shared van transportation service in Oman. The authors of this study have proposed a three-step clustering algorithm that uses social connections and the preferences of riders to form pools of socially-compatible and comfortable co-riders. In this paper, we improve the work presented in Haddad et al. (2022) by profiling current and potential users. The identification of different profiles is expected to improve our understanding of riders' behaviors and needs, which will help improve the quality of services. In the remainder of this paper, when we mention "ridesharing in Oman", we are referring to the socially-structured vanpooling transportation mode.

3.2. Methodology

In this study, we are following a data-driven approach to identify the current and potential classes of ridesharing users in Oman. For this purpose, we are using a "cluster-then- classify" approach consisting of three main steps. These steps were conducted using Python 3.10 in a Jupyter Notebook environment and main libraries like numpy (v2.0.2), pandas (v2.2.2), scikit-learn (v6.1.6) matplotlib (v3.10.0) and seaborn (v0.13.2). The first step uses a clustering algorithm to identify the groups of riders. The second step uses machine learning classification to predict those riders' profiles. The classification models allow for an "on-the-fly" classification of new riders based on their preferences. The "cluster-then-classify" approach has been widely used for classification tasks when labeled data is unavailable or very difficult to manually label Mei et al. (2025) Vrhovac et al. (2024). To the best of our knowledge, this approach has not previously been used for profiling ridesharing users.

3.3. Dataset

For the purpose of this study, we designed a questionnaire to collect data covering six dimensions: the riders, the reasons they use or will use the service, their preferences for co-riders, buses, drivers, and temporal aspects of the service. The questionnaire was reviewed, validated, and approved by the National Center for Statistics and Information (NCSI) in the Sultanate of Oman. It is a specialized national center that grants research approvals and analyzes the efficiency of research studies before publication. In addition, the questionnaire was reviewed and approved by the Dhofar University Research Ethics and Biosafety Committee. The target population in this study includes individuals who are using or willing to use ridesharing services in Oman. Based on the defined population of our study, a sampling framework was developed, and the questionnaire was distributed to a sample of over than 4,000 anonymous people across Oman, during fall 2022 and spring 2023. A majority of male respondents who are not currently using and not interested in using ridesharing services were discarded, and only data about current and potential ridesharing users was considered for data analysis. The collected data was cleaned and preprocessed, then a total of 3,615 data entries were kept for analysis. Table 1 presents a summary of the dataset. We can observe that females represent a majority with 78.73%. This high percentage is due to the fact that ridesharing is typically used by females, for many reasons, one of them is the fact that 75% of women in Oman don't have driving licenses and cars (National Center for Statistics and Information (2024)). Compared to (National Center for Statistics and Information (2023)), the distribution among age groups indicates that the dataset reflects the target population, where the largest segment is between 19 and 35 years, with the smallest segment above 45 years. Therefore, users aged above 45 were included in the questionnaire but due to their low representation, they did not form a significant portion of any cluster.

Table 1. Summary of Dataset Demographics

Category	Count	%	Category	Count	%	Category	Count	%
Governorate			Age			Income (EUR)		
Al Buraimi	77	2.13	18–35	2962	81.94	€2,400–€1,200	663	18.34
Al Dakhiliyah	639	17.68	36–55	335	9.26	€1,200–€720	529	14.63
Al Dhahirah	235	6.50	More than 56	47	1.30	Less than €720	277	7.66
ALWista	58	1.60	Less than 18	271	7.50	More than €2,400	287	7.94
Dhofar	843	23.32	Total	3615	100	No income	1859	51.43
Musandam	74	2.04	Employment			Total	3615	100
Muscat	170	4.70	Full-time	699	19.33	Education		
North Al Batinah	1218	33.70	Self-employed	314	8.69	Bachelor's	2275	62.93
North Sharqiyah	89	2.47	Job Seekers	411	11.37	Diploma	636	17.60
South Al Batinah	110	3.04	Part-time	166	4.60	Postgrad	156	4.31
South Sharqiyah	102	2.82	Student	2025	56.01	Prof. Diploma	468	12.95
Total	3615	100	Total	3615	100	Not applicable	80	2.21
Gender						Total	3615	100
Female	2846	78.73						
Male	769	21.27						
Total	3615	100						

4. Unsupervised labeling for profiling riders

The objective of this step is to answer the first question of this study. We aim to partition respondents into groups of riders according to their preferences collected in the dataset. Clustering algorithms are very useful machine-learning-based tools for this purpose. They are particularly suitable to support our effort to get better insights into the unknown patterns in the dataset. In this study, we selected the K-Modes clustering algorithm Huang (1998) which is commonly used for categorical data analysis. In the following sub-sections, we present the clustering process and the obtained results.

4.1. K-Modes clustering

The k-Modes algorithm is commonly used to cluster categorical data, and scales well to large datasets. However, apart from the requirement to define the number of clusters as input, it is very sensitive to the initialization of clusters' centroids. In order to choose the best number of clusters, we used the Elbow method based on the cost function values produced by the K-Modes with two initialization algorithms: Cao and Huang. As it is shown in Figure 1, the curve flattens more smoothly, but the diminishing decrease is noticeable around k = 4 to k = 5. So 4 or 5 clusters seem to be a good choice. We applied the k-Modes algorithm with groups of four and five clusters, respectively. Using four clusters led to unbalanced groups with 51% of the riders belonging to the same group, while the choice of five clusters resulted in a more balanced distribution of riders among the groups. Also, we implemented Chi-Square Test to identify significant features for K-Modes Clustering. For the ordinal features such as income level, age group, and education level, we ensured they were appropriately encoded and compared across clusters. Table 2 presents top 10 significant variables ordered by decreasing contribution score. We can see that the three most influencing variables are the purpose of using ridesharing, types of co-riders and Governorate. This suggests that preferences for ridesharing services may vary according to the Governorates of the Sultanate.

4.2. Identified clusters

To simplify the visualization of the five clusters, we used the Multiple Corresponding Analysis (MCA) to reduce the dimensionality of the dataset to 2 dimensions as MCA1 and MCA2, which do not represent any specific variables, but they help in interpreting the similarities between the clusters, as illustrated in Figure 2.

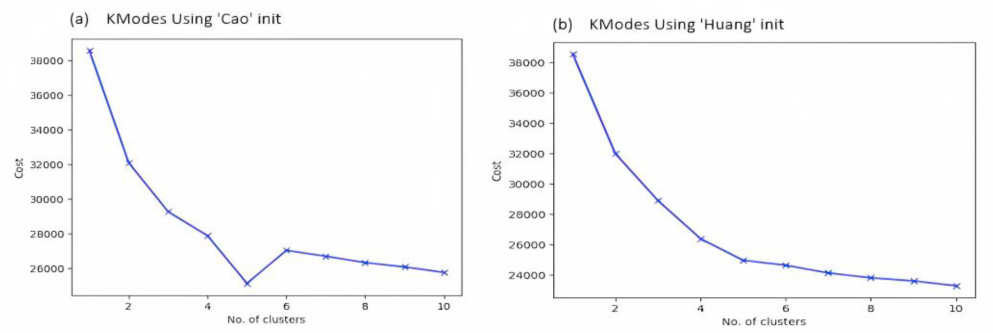

Fig. 1. Elbow method's score using (a) 'Cao' init and (b) 'Huang' init

Table 2. Top 10 Significant Variables from Chi-Square Test

Variable	Chi²	p-value	DoF
Purposes to use ridesharing services for	5342.32	<0.001	796
Co-riders you are using ridesharing with	5315.35	<0.001	216
Governorate	4857.73	<0.001	40
Driver nationality	3727.65	<0.001	8
Functional status	2741.00	<0.001	16
Number of co-passengers	2423.85	<0.001	8
Having a driving license	2041.70	<0.001	4
Income level	1855.12	<0.001	16
Co-riders with the same educational level	1855.05	<0.001	8
Co-riders from the same family	1704.54	0.000000	8

Fig. 2. 2D Clusters Plot

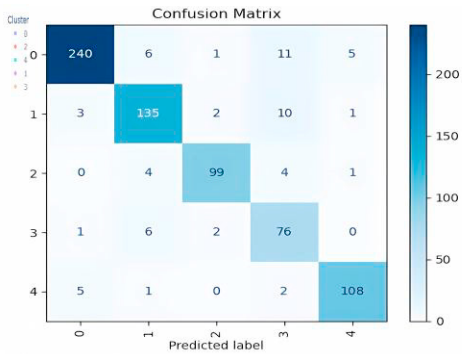

Fig. 3. Confusion Matrix for Logistic Regression

Reserved students (cluster 0): The majority of riders in this cluster are females between the ages of 18 and 35 and living in North Al Batinah Governorate (77.3 %). They are mostly students with no income or job-seekers. These riders have driver's licenses, but opt to use ridesharing with family members, classmates, and friends, mainly for home-study commuting, visiting immediate family, and attending medical appointments. Most reserved students require buses with air conditioning (AC) and are in favor of music being played during rides. They preferred co-riders from the same age group, interests and area. Reserved students prefer both long-term and occasional service agreements with an experienced Omani driver (regardless of gender) who speaks Arabic. We call these riders "reserved" because they use ridesharing services for basic mobility purposes, generally with members of their own social network.

Independent workers (cluster 1): The majority of riders in this cluster are males between the ages of 18 and 35, who are currently working and having a good income. Most of these riders are from Dhofar Governorate (35,5 %), North Al Batinah Governorate (16,3 %) and Al Dakhiliyah Governoarte (13,4 %). These riders have driver's licenses, which is why they are currently not using ridesharing services. The members of this cluster who are using ridesharing tend to use the services alone and less commonly with family members or friends. They used ridesharing to travel outside of their regions, but use the services less for going to work or for other purposes. They generally don't tolerate music being played during rides, require buses with AC and Internet, and have the strongest social preferences for their co-riders. They prefer an Omani driver who speaks Arabic, though the experience of the driver is not important. Independent workers use both long-term and occasional ridesharing services.

Dependent workers (cluster 2): The majority of riders in this cluster are females between the ages of 18 and 35, who are working and earn an income between €726 and €2,420. Riders from this cluster belong to different Governorates. They don't have driver's licenses and use ridesharing with co-workers or family members, mainly to commute to work, visit immediate family members, and for leisure activities. These riders prefers Internet connection and don't prefer music to be played during rides, but they are fine forgoing AC. They prefer to share rides with more than five co-riders, but have no strong preferences regarding having the same interests, age group or educational level. However, they strongly prefer to share rides with co-riders from the same family or region. Dependent workers prefer to have occasional service agreements with experienced non-Omani drivers (regardless of gender), who can speak Arabic.

Unemployed (cluster 3): The majority of riders in this cluster are females between the ages of 18 and 35. Most of riders are from Dhofar Governorate (72,2 %), and are either students or job seekers with a general diploma qualification earning no or low income. These riders don't have driver's licenses and use ridesharing with family members, mainly to attend hospital appointments, visit immediate family, or commute to school, work, etc. They require AC and tolerate music being played during rides. In addition, unemployed riders don't mind if there is no Internet connection in shared buses. They prefer to share rides with co-riders from the same family or area, who have the same interests and are from the same age group, but they have no strong preferences regarding educational level. Furthermore, the number of co-passengers is unimportant to these riders. They also prefer to have occasional service agreements with experienced female Omani drivers, who can speak Arabic.

Broad minded students (cluster 4): The majority of riders in this cluster are females between the ages of 18 and 35. They are mostly students with no income and from Al Dakhiliyah Governorate (80,2 %). These riders have driver's licenses, but use ridesharing with friends for all kinds of purposes, such as going to hospital appointments, school, restaurants, shopping centers, etc. They don't mind if there is no Internet connection in shared buses, and are fine with music being played or having no AC during rides. The number of co-passengers is unimportant to this cluster, and they don't have strong preferences for sharing rides with co-riders from the same family, area, age group, or educational level, and same interests. They prefer to have occasional service agreements with experienced Omani drivers (regardless of gender), who can speak Arabic.

The identified clusters partition current and potential riders into profiles that are distinguishable from one another, but still have certain similarities. On the one side, the independent workers can easily be distinguished from the other groups. They correspond mainly to male riders (and a minority of females) who are independent (in the sense that they have a job, an income, and a driver's license). These riders commonly don't use ridesharing services because they use their own cars. They use it only to travel outside the city and commute to work. Members of this cluster have the strongest social preferences among the overall population of riders. This could be intuitively explained by the fact that traveling outside the city for long distances or commuting to work require more social compatibility with co-riders in order to have a good traveling experience. On the other side, the most distinctive cluster is the dependent workers, where they rely on ridesharing services for their mobility. They tend to look for a safe traveling experience with more than five co-riders who are either family members or neighbors. They also prefer a quiet traveling experience (no music) over comfort (AC), but having an Internet connection is very important to this cluster. Intuitively, riders from this cluster don't look to build new social connections during their commute. Similarly unemployed cluster, but they appear to consider ridesharing as a social networking opportunity to make new connections in a more relaxed and comfortable atmosphere (music and AC). The "reserved" and "broad-minded" students are very similar profiles. They mainly differ in their attitudes toward ridesharing. Reserved female students use the ridesharing services for limited mobility purposes with a preference for co-riders who have same age and interests. Broad-minded female students, however, use the ridesharing service for all kinds of mobility services with all types of co-riders. In fact, members of this cluster have the weakest social preferences for co-riders among all the clusters.

5. Classification of riders

The clustering step allowed for the identification of five groups that potentially represent distinct major attitudes of riders toward shared mobility services in Oman, particularly socially-structured vanpooling. In this section, we examine how predictable the clusters are and the possibility to identify a new rider's cluster based on the survey's questions. We applied four machine learning classification algorithms to predict the cluster of riders. The four selected algorithms are: the 1) Decision Tree, 2) Logistic Regression, 3) Random Forest, and 4) CatBoost classification algorithms, respectively. We used the clustered dataset obtained in the previous step to train and evaluate the performance of the selected algorithms. The hyperparameters of the models were selected using the Optuna library which can automatically search the hyperparameter space and return the combination that maximizes cross-validated accuracy. In the third step, we perform a cross-validation test for the classification models by evaluating their performance on a new, randomly collected dataset. The performance of the four classification algorithms is detailed in Tables 3-6. Given the imbalanced clusters distribution, a random oversampling has been applied on the dataset prior to classification in order to improve the prediction of small classes. In addition, a multiple correspondence analysis (MCA) was applied to the dataset prior to classification to retain only the features that capture 95% of the variance, while reducing the dimensionality.

Table 3. Decision Tree Classification Report

Class	Prec.	Recall	F1-Score	Sup.
0	0.88	0.90	0.89	263
1	0.79	0.79	0.79	151
2	0.89	0.81	0.85	108
3	0.59	0.58	0.58	85
4	0.86	0.88	0.87	116
Acc.			0.82	723
Macro Avg	0.80	0.79	0.80	723
Weighted Avg	0.82	0.82	0.82	723

Table 4. Logistic Regression Classification Report

Class	Prec.	Recall	F1-Score	Sup.
0	0.96	0.91	0.94	263
1	0.89	0.89	0.89	151
2	0.95	0.92	0.93	108
3	0.74	0.89	0.81	85
4	0.94	0.93	0.94	116
Acc.			0.91	723
Macro Avg	0.90	0.91	0.90	723
Weighted Avg	0.92	0.91	0.91	723

Table 5. Random Forest Classification Report

Class	Prec.	Recall	F1-Score	Sup.
0	0.97	0.86	0.91	263
1	0.75	0.88	0.81	151
2	0.90	0.94	0.92	108
3	0.65	0.68	0.67	85
4	0.95	0.91	0.93	116
Acc.			0.86	723
Macro Avg	0.84	0.85	0.85	723
Weighted Avg	0.87	0.86	0.87	723

Table 6. CatBoost Classification Report

Class	Prec.	Recall	F1-Score	Sup.
0	0.97	0.94	0.95	263
1	0.84	0.93	0.89	151
2	0.93	0.90	0.92	108
3	0.84	0.75	0.80	85
4	0.93	0.97	0.95	116
Acc.			0.91	723
Macro Avg	0.90	0.90	0.90	723
Weighted Avg	0.92	0.91	0.91	723

The classification results show that Logistic Regression and CatBoost classifiers led to the best results with an overall accuracy of 0.91 each. However, they have slightly different performances with respect to class-based performance. The f1-scores show that CatBoost classifier performs better for clusters 0 and 4, Logistic Regression performs better for clusters 2 and 3, and they have similar performances for cluster 1. Moreover, the Logistic Regression showed a balanced performance across all clusters, regardless of their sizes. Further more, the Confusion Matrix was applied for the classifiers. As an example, we present the results of the Logistic Regression in terms of correctly and incorrectly classified instances. As shown in Figure 3, the model shows a balanced classification pattern with the highest accuracy for cluster 0 and 1, where the highest confusion occurs in cluster 3. Cluster 2 and 4 also stands out with minimal confusion, reflecting the model's ability to perform effectively. Across the models, the features that contributed most to the classification of riders are Governorate, purpose of using the service (go to work, to study, etc.) and co-riders to use the service with (friends, co-workers, etc.).

To further explore the classification performance, we have collected an additional 124 data entries to be used as a supplementary testing dataset. We performed a Hamming distance clustering on the 124 new data points and assigned the data point to the cluster with the closest centroid. We used the four classifier models that we trained to classify the newly collected dataset, and we calculated the similarity percentages of their results with the Hamming distance classification (ground truth) (Table 7). The Logistic Regression led to the best similarity percentage of 90.24% as shown in Table 7 which detailed the performance of the four classification algorithms.

Table 7. Confusion Matrix for Logistic Regression

Classifier	Accuracy	Similarity (%)
Logistic Regression	0.91	90.24%
CatBoost	0.91	88.62%
Random Forest	0.86	81.3%
Decision Tree	0.82	78.85%

The high level of agreement between the two algorithms suggests that the Logistic Regression is effectively capturing the underlying cluster structure defined by the original categorical data. At this stage, our objective is to explore how predictable the clusters are, and the obtained classification results show that they can be predictable with up to 91% accuracy. We believe that these results suggest that the five identified clusters satisfactory capture the different profiles of ridesharing users within Oman.

6. Conclusion

This paper reports preliminary results on the profiling of current and potential ridesharing users in Oman. We applied a "cluster then classify" approach on a dataset of 3615 respondents, and the obtained results suggest that Omani ridesharing users can be partitioned into five distinguishable, interpretable and predictable profiles. In this article, only the riders in Oman are considered, generalizability of the results to other cultural or geographic contexts may be limited. Also, implementing the same clustering and classification algorithms in studying social preference parameters among the drivers can be a potential direction for future study. Moreover, the sample showed gender differences, which may have affected the representativeness of the results, especially with regard to gender-specific preferences. Also, one of the limitations of this study is that some factors which influence ride-sharing behavior may not have been captured in the dataset. In addition, the study reliance on self-reported data, which is limiting the ability to infer changes over time. By considering these aspects, efficient social-profiling will enhance the quality of ride-sharing services and improve the transportation planning in the Sultanate of Oman, towards a tailored mobility experience for users. User profiles identified through clustering the preferences can be used to tailor app features for Ridesharing service, providing a personalized matching filter or preference-based routing.

References

Delatte, A., Baouni, T., Belwal, R., Daou, L., Gourram, D., Imam, R., Zitoun, M.S., Smadi, A., 2018. Understanding the needs of mena public transport customers: culture of service and gender responsive recommendations. TeMA Journal of Land Use, Mobility and Environment, 7–30.

Dolins, S., Karlsson, M., Smith, G., Stro¨mberg, H., 2025. Riding with strangers: profiling potential users and refusers of shared autonomous vehicles in swedish cities. Transportation , 1–22.

Eke, C.I., Norman, A.A., Shuib, L., Nweke, H.F., 2019. A survey of user profiling: State-of-the-art, challenges, and solutions. IEEE Access 7, 144907–144924.

Haddad, H., Bouyahia, Z., Jabeur, N., Yasar, A.U.H., 2022. Socially structured vanpooling: A case study in salalah, oman. IEEE Intelligent Transportation Systems Magazine 14.

Hossain, M., Mozahem, N.A., 2022. Drivers' perceptions of the sharing economy for transport services. Technological Forecasting and Social Change 179, 121668.

Huang, Z., 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data mining and knowledge discovery 2, 283–304.

Loa, P., Ong, F., Habib, K.N., 2023. Identifying profiles of ride-sourcing users in the metro vancouver region for a better understanding of ride-sourcing behaviour. Travel Behaviour and Society 33, 100637.

Machado, C.A.S., de Salles Hue, N.P.M., Berssaneti, F.T., Quintanilha, J.A., 2018. An overview of shared mobility. Sustainability 10, 4342. Mei, P., Karimi, H.R., Ou, L., Xie, H., Zhan, C., Li, G., Yang, S., 2025. Driving style classification and recognition methods for connected vehicle control in intelligent transportation systems: A review. ISA transactions .

National Center for Statistics and Information, 2023. Population, housing, and household statistics 2022. September 2023 Edition. National Center for Statistics and Information, 2024. Omani women observatory: Special edition 2024. Based on data from 2023.

Sarriera, J.M., A´ lvarez, G.E., Blynn, K., Alesbury, A., Scully, T., Zhao, J., 2017. To share or not to share: Investigating the social aspects of dynamic ridesharing. Transportation Research Record 2605, 109–117.

Shaheen, S., Cohen, A., 2021. Shared mobility: An overview of definitions, current practices, and its relationship to mobility on demand and mobility as a service. International encyclopedia of transportation , 155–159.

Soltani, A., Allan, A., Khalaj, F., Pojani, D., Mehdizadeh, M., 2021. Ridesharing in adelaide: Segmentation of users. Journal of Transport Geography 92, 103030.

Vrhovac, V., Orosˇnjak, M., Ristic´, K., Sremcˇev, N., Jocanovic´, M., Spajic´, J., Brkljacˇ, N., 2024. Unsupervised modelling of e-customers' profiles: Multiple correspondence analysis with hierarchical clustering of principal components and machine learning classifiers. Mathematics 12, 3794.

Zhang, H., Zhao, J., 2019. Mobility sharing as a preference matching problem. IEEE Transactions on Intelligent Transportation Systems 20,2584–25