Mathematics and Statistics 10(5): 971-980, 2022 http://www.hrpub.org
DOI: 10.13189/ms.2022.100508

Using Clustering Methods to Detect the Revealed Preferences of Moroccans towards the Electric Vehicles: Latent Class Analysis (LCA) and K-Modes Algorithm (K-MA)

Taoufiq El Harrouti[1,*], **Mourad Azhari**[2], **Hajar Deqqaq**[1], **Abdellah Abouabdellah**[1], **Sanaa El Aidi**[3], **Habiba Chaoui**[3]

[1]Engineering Science Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco

[2]Center of Guidance and Planning of Education, MEN, Rabat, Morocco

[3] Laboratory of Advanced Systems Engineering, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco

Received May 28, 2022; Revised July 15, 2022; Accepted August 20, 2022

Cite This Paper in the following Citation Styles

(a)*: [1] Taoufiq El Harrouti, Mourad Azhari, Hajar Deqqaq, Abdellah Abouabdellah, Sanaa El Aidi, Habiba Chaoui, "Using Clustering Methods to Detect the Revealed Preferences of Moroccans towards the Electric Vehicles: Latent Class Analysis (LCA) and K-Modes Algorithm (K-MA)," Mathematics and Statistics, Vol.10, No.5, pp. 971-980, 2022. DOI: 10.13189/ms.2022.100508*

(b)*: Taoufiq El Harrouti, Mourad Azhari, Hajar Deqqaq, Abdellah Abouabdellah, Sanaa El Aidi, Habiba Chaoui, (2022). Using Clustering Methods to Detect the Revealed Preferences of Moroccans towards the Electric Vehicles: Latent Class Analysis (LCA) and K-Modes Algorithm (K-MA). Mathematics and Statistics, 10(5), 971-980. DOI: 10.13189/ms.2022.100508*

Abstract Latent Class Analysis (LCA) and k-Mode Algorithm (K-MA) are two unsupervised machine learning techniques. These methods aim to identify individuals on the basis of their shared traits. They are utilized in the context of categorical data and can be used to detect people's opinions toward green forms of transportation, especially Electric Vehicles (EV) as an alternative to conventional internal combustion engine vehicles. The LCA approach discovers group profiles (clusters) based on observed variables, whereas the K-MA technique is an adaptation of the k-means algorithm for categorical variables. In this study, we apply these two methods to identify Moroccans' preferences for the electrification of their means of transportation. Both algorithms are able to divide the analyzed sample into two groups, with the first group being more interested in EV. The second group consists of individuals who are less concerned about ecologically sustainable transportation. In addition, we conclude that the LCA algorithm performs well and is superior to the K-MA, and that its discrimination power (65% vs 35%) is more than that of the K-MA (52% vs 48%).

Keywords Clustering, Latent Class Analysis (LCA), K-Mode Algorithm (K-MA), Electric Vehicles (EV), Survey

1 Introduction

According to the report published by Intergovernmental Panel on Climate Change (IPCC), the transportation sector was responsible for causing a significant amount of damage to the environment. It was responsible for 23% of the total of CO2 emissions related to energy and produced 7 Gt CO2eq worth of direct greenhouse gas emissions, which is a massive amount of pollution [1, 2]. To make the transition to low-carbon modes of transportation requires a significant behavioral shift and the readiness of people to adapt to new transportation technologies. [3] asserted that the choice of transportation method is significantly impacted by a number of major drivers, including factors related to costs, social and cultural factors, and rise in disposable income.
There is a significant opportunity for the transportation industry to play a part in the development of a green economy. The term "sustainable mobility" refers to a transportation policy that makes an effort to strike a balance between factors such as accessibility, economic development, and the impacts of various modes of transportation on the environment. Researchers and transportation companies face a huge hurdle when trying

to determine how to best use EV as a solution to the problem of polluting transportation [4, 5].
These challenges encouraged us to conduct this study to investigate the expectations and attitudes of Moroccan consumers about the use of EV as opposed to conventional ones. In this regard, a survey was conducted to determine the revealed preferences of Moroccans regarding EV, and it was entirely appropriate to explore the possibility of adopting this green mode of transportation and the extent to which EV are perceived as the optimal solution for Moroccan consumers [6, 7].
This research paper aims to:

- identify the profiles/classes of people who are likely to use EV as a future alternative to sustainable transport and who may be the target of a strategy to replace conventional vehicles with EV;

- compare the discrimination power between the profiles detected by three methods: K-MA, MCA, and LCA.

In order to answer these questions, a questionnaire was administered to a population restricted to the Rabat-Salé metropolitan area. This decision is bolstered by the fact that people of this region are already confronted with the issue of electric mobility, since the tramway has been one of the modes of transportation there since 2011, forming a network of electric transport that spans around 30 kilometers.
This study is divided into five sections. After an introduction and the related work, we will present the proposed methods, which comprises LCA and the K-MA, the sampling and the setting. In the following section, we will discuss our application results. We will conclude with a conclusion and perspectives.

2 Related work

LCA has been applied in a significant number of research studies, and projects in order to classify individuals. According to [8], LCA, as a form of structural equation modeling, is based on the identification of structure in data. Despite the fact that LCA has been around for a while, recent breakthroughs in mathematics and software design have made it far more powerful and flexible than it was previously. The authors came to the conclusion in their research that LCA offers a more robust statistical foundation than K-means and Hierarchical Clustering for both exploratory and predictive purposes.
Krantz et al.[9] argued that K-means Clustering is a practical and useful tool for exploring differences among museum visitors. However, in their study, the authors identified pitfalls that stem from the use of the K-means method, they stated that there are, in reality, a great deal of issues with the K-means Clustering approach, as well as with clustering algorithms in general. Among the potentially problematic aspects is a sensitivity to outliers. In the field of psychology, LCA has been utilized to classify children into groups according to their level of behavioral disturbance [10, 11]. LCA was also employed by Klonsky et al. [12] in order to identify clinically distinct types of people who self-inflict harm. According to the findings, there are four distinct subgroups of people who self-inflict injuries, and these subgroups differ on measures of suicidality, anxiety, borderline personality disorder, and depression. In the medical field, Lui Z. et al. [13] used LCA to identify physician competency subgroups based on eight different competency dimensions. They found that a four class model provided the best fit for the data. These four classes are as follows: excellent competency groups, lack of professionalism competency group, individual competency-driven group, and lack of competency cognitive group. In addition, Rossi et al. [14] performed LCA to investigate the association between psychological discomfort and cancer-related complications. College administrators can use LCA to identify new freshmen with a pluralistic attitude and use that information in program design and targeted interventions focused on diversity issues, according to research by Denson et al. [15]. A four-class model fits best: high pluralistic orientation, high-disposition, low-skill, low-disposition, high-skill, and low pluralistic orientation. Using data from the Longitudinal Study of American Youth, Young and Nylund-Gibson [16] used LCA to investigate student attitudes about mathematics and science. Rhead et al [17] used LCA in the social sciences to explore UK public commitment to the environment and its relationship with pro-environmental behavior. Based on their environmental commitment, the study's findings reveal four types of people: pro-environment, neutral majority, disengaged, and paradoxical. This finding was also more predictive of environmental behavior than most previous research with environmental attitude groups. Fox B. et al [18] performed an LCA on a nationwide sample of 702 American adults to evaluate the different categories of people who favor or oppose police militarization. The findings highlight the complexities of public opinion on this contentious subject in contemporary American police. [19] employed LCA to identify unobservable types of cybercrime victims and propose a measure of cybercrime risk based on LCA conditional probabilities. Many research in politics have emphasized the use of LCA as a method for detecting latent classes in a population. For example, [20] utilized LCA to quantify the number of latent classes in each of Taiwan's two aspects of political identity. [21] used LCA to examine democratic principles among Europeans based on data from the 2012 European Social Survey. According to the findings, the majority of Europeans value both political and social rights equally, whereas some people prioritize either political or social rights. In contrast to previous research, Alvarez et al. [22] apply LCA to define the four faces of political engagement, which cover both standard and unorthodox forms of participation.

3 Materials and Methods

This section aims to set the theoretical bases of the proposed methods (LCA and K-MA) and present the sampling and setting of the survey.

3.1 Latent Class Analysis (LCA)

LCA is an unsupervised machine learning approach for discovering qualitative subgroups of populations that share common characteristics. It entails creating latent (non-observable) classes based on people' observed (manifest) reactions [23]. It consists in constructing latent (non-observable) classes based on observed (manifest) responses of individuals.
The model includes a latent variable X with K categories, each category represents a class. Each class comprises a homogeneous group of individuals sharing the same interests, values, characteristics, and behaviors. The belonging to a class C_k is calculated from the probability of realization of a response Y_j, knowing the classes C_k and the determination of the number of classes, is calculated using the selection criteria (AIC, BIC, ...) [24, 25]
The mathematical model for LCA can be expressed as follows [26, 27]:
Let p represent dichotomous observed variables $\mathrm{X}_1, \mathrm{X}_2, .., \mathrm{X}_\mathrm{p}$ taking values 0 or 1, and Y the latent variable with k classes. Let $\mathrm{p_i}$ denote the probability that $\mathrm{X_i} = 1$ for an individual of the latent class j.
If π_j is the prior probability of belonging to the latent class j, the hypothesis of conditional independence for the vector of observed variable x returns:

$$f(x)=\sum_{j=1}^{k} \pi_j \prod_{i=1}^{p} p_{ij}^{x_i}\left(1-p_{ij}\right)^{1-x_i} \tag{1}$$

We deduce that the a posteriori probability that an individual of vector x belongs to the latent class j is:

$$h(j/\mathrm{x})=\pi_j \prod_{i=1}^{p} p_{ij}^{x_i}\left(1-p_{ij}\right)^{1-x_i} / f(\mathrm{x}) \tag{2}$$

This formula is used to assign an individual to the most probable latent class.

3.2 K-Modes Algorithm(K-MA)

K-MA is an adaptation of the k-means algorithm for categorical data [28, 29]. It employs a basic matching dissimilarity measure for categorical items, substituting the mode for the cluster mean [30]. Let $D=\{X_1, X_2, \ldots, X_n\}$ be a set of n objects that is to be clustered. Object X_i is represented as $[x_{i1}, x_{i2}, \ldots, x_{id}]$ and the cluster centers are represented by $Z_l=\{z_{l1}, z_{l2}, \ldots, z_{ld}\}$ for $1 \leq l \leq k$ where k is the number of clusters. Assume X and Y are two categorical data objects in D, the simple matching distance measure between X and Y is defined as [31] :

$$d_c(X, Y)=\sum_{j=1}^{d} \delta\left(x_j, y_j\right) \tag{3}$$

where x_j and y_j are the j^{th} component of X and Y respectively and

$$\delta\left(x_j, y_j\right)=\begin{cases}0 & \text{if } x_j=y_j \\ 1 & \text{otherwise}\end{cases} \tag{4}$$

This algorithm identifies the modes using a frequency-based technique by solving the following mathematical problem [31]. Then the objective of K-Modes clustering is to find W and Z that minimize

$$F_c(W, Z)=\sum_{l=1}^{\kappa} \sum_{i=1}^{n} w_{li} d_c\left(X_i, Z_l\right) \tag{5}$$

subject to (6), (7), and (8)

$$w_{li} \in \{0,1\}, \quad 1 \leq l \leq k, \quad 1 \leq i \leq n \tag{6}$$

$$\sum_{l=1}^{k} w_{li}=1, \quad 1 \leq i \leq n \tag{7}$$

$$0<\sum_{i=1}^{n} w_{li}<n, \quad 1 \leq l \leq k \tag{8}$$

The K-MA operates in the following steps [32, 33] :

1. set k instances at random and assign them leaders/cluster;
2. identify the leader that is closer to instance and assign;
3. calculate the mode of each cluster;
4. assign the new mode to be the new leader.
5. repeat the 2 to 4 steps until no reallocation is necessary.

3.3 Sampling and setting

Because reducing bias and increasing the extent to which findings may be generalized are not the major focuses of qualitative research [34], we want to make sure this study's sample is as representative and feasible of the population in terms of identifying the classes that exist within it. Nylund et al. [16] propose 300 or more cases as the optimal sample size for a research study to apply a LCA. Because the sample universe for this study is restricted in terms of demographics and geography, preventing incorrect generalizations may be accomplished by restricting generalization to the sample universe level rather than attempting to make abstract conclusions [35]. We used convenience sampling in this research study. It uncovers any convenient individual that fits the requisite criteria and selects those who reply on a first-come-first-served basis until the sample size quotient is reached. We use data from a cross-sectional survey conducted within the Rabat-Sale metropolitan area and with a sample selected by convenience sampling. The gathering of data was carried out with consideration given to the two crireria listed below:
• Inclusion criteria: Anyone who falls into the category of the target population, whether they are of Moroccan or another nationality, is required to be at least 18 years old, in possession of a valid driver's license, and present during the study.
• Exclusion Criteria: Anyone who was not present when the data were collected or who declined to take part in the study.
A questionnaire was elaborated constituting a guided interview with the respondents via a platform on the web. This questionnaire is divided into three main sections: the first is the demographic characteristics of the participants, the second section

is dedicated to the current mobility properties, and the third is the features of the EV if chosen. The survey questionnaire was distributed to respondents via a web-based platform.The respondents provided their responses to the following 18 questions, which are listed in table 1. The processing and analyzing of the responses are done with the programming language R. The data was collected between January and May of 2022. We received 340 responses to the questionnaire, which represents a participation rate that is exceptionally high. Following a review of the questionnaire's response rate, we retained 209 entries after removing those that were incomplete or inconsistently answered.

Table 1. questions submitted to the participants in the survey

N	question	N	question
Q.1	sex	Q.10	fiscal power
Q.2	age	Q.11	EV information
Q.3	residence	Q.12	attitude on the environement
Q.4	education level	Q.13	estimated route
Q.5	profession	Q.14	purchase decision
Q.6	income category	Q.15	EV recharge location of EV
Q.7	personal vehicle	Q.16	recharge time of EV
Q.8	fuel type	Q.17	charging frequency
Q.9	vehicle type	Q.18	purchase criteria

4 Results and discussion

To test whether LCA and K-MA can be applied to the collected data, we propose using the Multiple Correspondence Analysis (MCA). When dealing with nominal categorical data, MCA is a data analysis approach for identifying and representing underlying classes, using a low-dimensional Euclidean space to represent data points.

4.1 Multiple Correspondence Analysis application

We applied MCA to characterize the relationships between variables/modalities and individuals in our data [36]. The figure 1 shows that the cumulative projected inertia suggests that the first five axes account for 29.84 % of the information, while the first two axes alone account for 14.20% of the observed fluctuations in our sample.

```
> print(ve.acm$eig)
       eigenvalue percentage of variance cumulative percentage of variance
dim 1  0.15826535              8.3787540                          8.378754
dim 2  0.11001145              5.8241355                         14.202889
dim 3  0.10406853              5.5095102                         19.712400
dim 4  0.09845248              5.2121900                         24.924590
dim 5  0.09295035              4.9209011                         29.845491
```

Figure 1. variance explained with first five dimensions

The biplot in figure 2 showcases the MCA results that produced the correlations between the modalities that can be used to construct various profiles. Using two dimensions, fewer than 15% of the information can be conveyed. As a result, we may discern two distinct profiles :

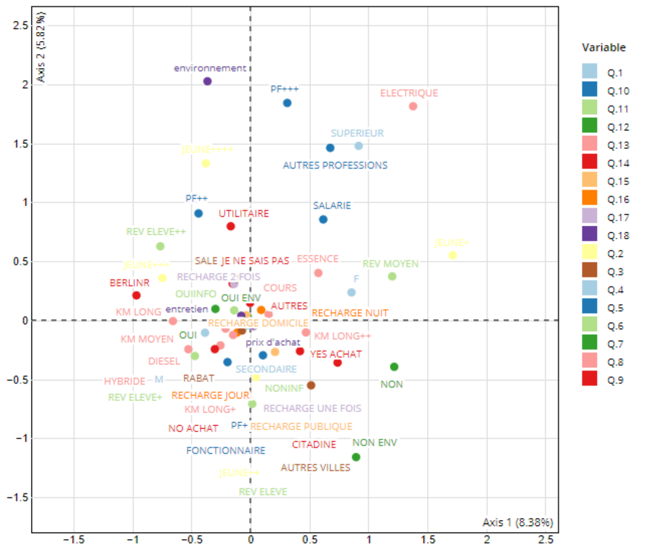

Figure 2. The biplot illustrates the associations between individuals and variables/modalities.

4.2 Interpretation of MCA results

The explained information using two dimensions is less than 15%. Moreover, we can identify two profiles:

4.2.1 Dimension 1

The contributions concern young (18-30 years:15%) and old (60 years and over:7.1%) respondents. The participants belonging to this class have an average income (ranging from 601 to 1000 Euros: 12.5 %), or a high salary (ranging from 1001 to 1200 Euros: 5.5%).Those who do not own a vehicle account for only 2.5 % of the respondents, while those who do constitute 7.5% of the total. Women make up 8 % of this group, while men make up 4 %. Some of these individuals work in liberal professions, government positions and employees.

At first glance, we find that axis 1 , on the Figure 2 above, shows a clear distinction between young (aged 18-30) and old respondents (aged 60 and over). This axis also compares young and old, along with vehicle owners with non-vehicle owners, women versus men, average wage versus higher income, and liberal professions versus employees and government officials.

4.2.2 Dimension 2

The liberal professions make up 18 % of the total, followed by government officials at 5 %, and employees at 3.75 % . Vehicles with a range of horsepower (8-11 hp), (11 hp and over), and (5-7 hp) are all included in this group (3.9 %). These individuals have revenues varying from high (7.4%) to very high (5.1%). Young adults (30-45 years old) weigh environmental concerns while deciding whether or not to buy an electric vehicle.

Axis 2 in figure 2 divides the fiscal power into two categories: those with 8 hp or more and those with 5 hp to 7 hp. In summary, the respondents/individuals who decide to buy an EV are close to the modalities contributed to axis 1. Those opposed to purchasing the EV are close to the modalities contributed to axis 2. The MCA method describes the profiles of the groups formed by the qualitative variables, but the information explained by the first two axes is insufficient to draw meaningful conclusions. An unsupervised machine learning method, such as the LCA method, can be used to solve this problem.

4.3 LCA application

Table 2 displays the outcomes of the LCA method applied to the collected data. The survey's goal is to learn about respondents' characteristics, their thoughts on EV, and their likelihood of purchasing one. After that, it looks for noteworthy relationships, issues connected to different types of travel, as well as the opportunities and limitations presented by this method of transportation.

4.3.1 Class 1: Individuals who have a lower level of interest in EV

The profile of these individuals according to the manifest variables is depicted in the figure 3. People belonging to Class 1 comprises 65% of the surveyed population, with a disproportionate number of men (84 %). This class includes Moroccans aged between 31 and 45 (prob = 0.47), 46 and 60 (prob = 0.49) from two cities: Rabat (prob = 0.54) and Salé (prob = 0.35). Most of the people in this group are civil servants (prob=0.99), earning between the middle and upper-class salaries (prob=0.28 and 0.38). This group of individuals owns diesel-powered vehicles (prob = 0.96) and uses city cars (prob= 0.41) and sedans (0.33). They are informated about EV and believe that these types of technology will improve environmental quality with a high probability (prob=0.98). Vehicle owners in this class are opposed to purchasing an EV (prob=0.36). However, some respondents will be able to purchase an EV (prob = 0.10) and prefer once-a-day home charging (prob = 0.88). These individuals own vehicles with between 5 and 7 (prob = 0.71) and 8 to 10 (prob = 0.27) horsepower. These vehicles travel an estimated daily distance of 15 to 30 kilometers (probability = 0.70). This study concludes that the purchase price of an EV is a significant factor (prob=0.85).

4.3.2 Class 2: Individuals who have a higher level of interest in EV

The figure 3 depicts the manifest variables-based profile of individuals of Class 2. This group accounts for 35% of all respondents. Unlike class 1, this class has a female representation of 59 %. These young people (prob = 0.9), who live in two cities: Rabat (prob = 0.43) and Salé (prob = 35), are civil servants (prob = 0.63), have a high school diploma (prob = 0.93), and have an average income (prob = 0.60). In contrast to class 1, this group of people owns diesel vehicles (prob = 0.52) of the city car type (prob = 0.77), has a fiscal power between 5 and 7 (prob = 0.89), and travels an estimated distance of 15 km per day (prob = 0.72).

Figure 3. profiles of the participants using LCA method (lower and higher interest in EV

This group's respondents are knowledgeable about EV (prob = 0.6), and they believe that using this mode of transportation will help the environment by improving air quality (prob=0.43). In contrast to individuals in class 1, people who fall into this category make the decision to accept an offer to purchase an EV (prob=0.42) after taking into consideration whether or not the purchase price is adequate (prob=0.85). These EV drivers prefer to charge their vehicles at home (prob = 0.82).

Table 2. Conditional item response probabilities

		class 1	class 2			class 1	class 2
		0,65	0,35			0,65	0,35
Questions	**Modalies**			**Questions**	**Modalies**		
Q.1	**M**	**0,84**	**0,41**	**Q.10**	**0-600 Euro :1**	**0,05**	**0,60**
	F	**0,16**	**0,59**		**601-1000 Euro:2**	**0,28**	**0,28**
Q.2	**18-30 :1**	**0,00**	**0,40**		**1001-1200 Euro:3**	**0,29**	**0,06**
	31-45: 2	**0,47**	**0,50**		**1201 and more :4**	**0,38**	**0,06**
	46-60:3	**0,49**	**0,08**	**Q.11**	**YES INFO: 1**	**0,73**	**0,60**
	60 and more : 4	**0,04**	**0,00**		**NO INFO : 2**	**0,27**	**0,40**
Q.3	**RABAT: 1**	**0,54**	**0,43**	**Q.12**	**YES ENV :1**	**0,98**	**0,96**
	SALE:2	**0,35**	**0,35**		**NO ENV :2**	**0,02**	**0,04**
	others:3	**0,11**	**0,22**	**Q.13**	**0-15 km:1**	**0,59**	**0,72**
Q.4	**YES:1**	**0,96**	**0,52**		**16-30 km:2**	**0,13**	**0,04**
	NO:2	**0,04**	**0,48**		**31-45 km:3**	**0,16**	**0,14**
Q.5	**diesel: 1**	**0,87**	**0,65**		**46 -60 km : 4**	**0,09**	**0,06**
	Essence :2	**0,10**	**0,30**		**60 and more : 4**	**0,03**	**0,03**
	Hybrid:3	**0,03**	**0,00**	**Q.14**	**YES purchase 1**	**0,10**	**0,42**
	others:4	**0,00**	**0,05**		**NO purchase:2**	**0,36**	**0,20**
Q.6	**city car:1**	**0,41**	**0,77**		**i don't know**	**0,27**	**0,29**
	BERLINE:2	**0,33**	**0,01**		**others: 3**	**0,27**	**0,09**
	useful:3	**0,09**	**0,09**	**Q.15**	**home charging:1**	**0,88**	**0,82**
	others:4	**0,17**	**0,13**		**Public charging station:2**	**0,12**	**0,18**
Q.7	**5-7 :1**	**0,71**	**0,89**	**Q.16**	**during the night with a reduced rate:1**	**0,51**	**0,62**
	8-10: 2	**0,27**	**0,08**		**during the day:2**	**0,49**	**0,38**
	11 and more: 3	**0,02**	**0,03**	**Q.17**	**one liver:1**	**0,82**	**0,90**
Q.8	**secondary level :1**	**0,99**	**0,93**		**twice:2**	**0,18**	**0,10**
	high level :2	**0,01**	**0,07**	**Q.18**	**purchase price:1**	**0,85**	**0,85**
Q.9	**official : 1**	**0,85**	**0,63**		**maintenance price:2**	**0,13**	**0,15**
	employee:2	**0,06**	**0,17**		**price of consumption:3**	**0,00**	**0,00**
	others:3	**0,09**	**0,20**		**respect for the environment :4**	**0,02**	**0,00**

4.4 K-Mode Algorithm application

In this application, we use Python programming languages (see figure 4) and Elbow curve to find optimal K. To determine the optimal number of clusters, we have to select the value of k at the "Elbow": the point after which the distortion/inertia starts decreasing in a linear fashion (see the figure 5). Thus for the given data, we conclude that the optimal number of clusters for the data is two.

```
# Elbow curve to find optimal K
cost = []
K = range(1,5)
for num_clusters in list(K):
kmode = KModes(n_clusters=num_clusters, init = "random", n_init = 5, verbose=1)
kmode.fit_predict(df)
cost.append(kmode.cost_)
```

```
# Building the model with 3 clusters
kmode = KModes(n_clusters=3, init = "random", n_init = 10, verbose=1)
clusters = kmode.fit_predict(df)
clusters
```

```
df.insert(0, "cluster",clusters, True)
df
```

Figure 4. code used with Python programming language

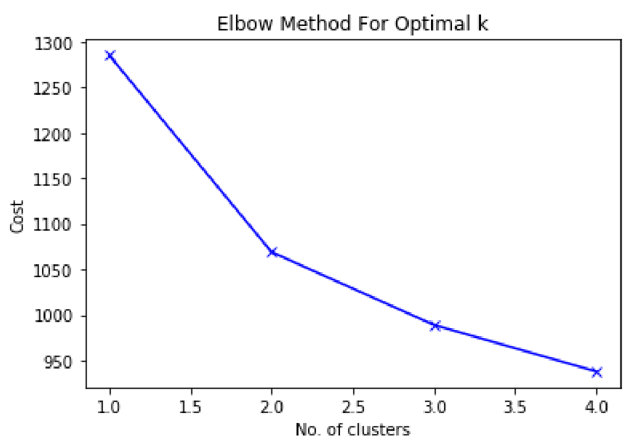

Figure 5. Elbow method for optimal K

Table 3 summarizes the accuracy using K-MA. It distincts two different profiles.

Table 3. Accuracy using K-MA

		class 1	**class 2**			**class 1**	**class 2**
		0,52	**0,48**			**0,52**	**0,48**
questions	**Modalities**			**questions**	**Modalities**		
Q.1	M	0,68	0,70	Q.10	0-600 Euro : 1	0,27	0,23
	F	0,32	0,30		601-1000 Euro: 2	0,10	0,47
Q.2	18-30 :1	0,15	0,14		1001-1200 Euro: 3	0,19	0,22
	31-45: 2	0,30	0,69		1201 and more: 4	0,44	0,44
	46-60:3	0,52	0,15	Q.11	YES INFO: 1	0,73	0,63
	60 and more : 4	0,03	0,02		NO INFO : 2	0,27	0,37
Q.3	RABAT: 1	0,30	0,72	Q.12	YES ENV :1	0,99	0,96
	SALE:2	0,53	0,15		NO ENV :2	0,01	0,04
	others:3	0,17	0,13	Q.13	0-15 km:1	0,65	0,62
Q.4	YES:1	0,78	0,83		16-30 km:2	0,09	0,11
	NO:2	0,22	0,17		31-45 km:3	0,17	0,14
Q.5	diesel: 1	0,82	0,76		46 -60 km : 4	0,07	0,08
	essence :2	0,14	0,21		60 and more : 4	0,02	0,05
	Hybrid:3	0,02	0,02	Q.14	YES purchase 1	0,19	0,24
	others:4	0,02	0,01		NO purchase:2	0,19	0,41
Q.6	city car:1	0,46	0,61		I don't know	0,42	0,13
	berline:2	0,27	0,17		others: 3	0,20	0,22
	useful:3	0,10	0,08	Q.15	home charging:1	0,89	0,83
	others:4	0,17	0,14		Public charging station:2	0,11	0,17
Q.7	5-7 :1	0,72	0,82	Q.16	during the night with a reduced rate:1	0,75	0,32
	8-10: 2	0,24	0,17		during the day:2	0,25	0,68
	11 and more: 3	0,04	0,01	Q.17	one liver:1	0,86	0,83
2[4]*Q.8	secondary level :1	0,95	0,99		twice:2	0,14	0,17
	high level :2	0,05	0,01	Q.18	purchase price:1	0,81	0,90
Q.9	official : 1	0,72	0,82		maintenance price:2	0,17	0,09
	employee:2	0,11	0,09		price of consumption:3	0,00	0,00
	others:3	0,17	0,09		respect for the environment :4	0,02	0,01

4.4.1 Class 1: 52% of people are less interest in EV

This class is comprised of 68 % men and 32 % women, with more than 80 % of its members being active young adults (aged 31 to 60). More than half of the population (53%) reside in Salé, while 30% reside in Rabat. 95 % of those interrogated have a secondary education. 72% of them are civil servants, 44% have a monthly income greater than 1200 euros, while 27% have a monthly income of less than 600 euros. Moreover, 78 % of respondents own a vehicle for transportation, of which 82 % use diesel, 14% gasoline, and 2% hybrids. This group uses city cars at a rate of 46%, Berlines at a rate of 27%, with a fiscal power ranging between 5 and 7 horsepower (72%), and 8 to 10 horsepower (24%). Furthermore, 73% of respondents are knowledgeable about EV, and 99% are aware of their positive environmental impact. 62% of this group commute less than 15 kilometers per day, while 17%. commute between 31 and 45 kilometers. 19% of this group are in favor of purchasing an EV, 19% are opposed, and 62% are hesitant to purchase. 89% of respondents prefer the purchase option and prefer to charge their EV overnight at home with a discount rate in a single charge, while 11% prefer public charging throughout the day. 81% of people who want to purchase an EV consider the purchase price, compared to 17% who prefer EV for their lower maintenance costs and 2% for their environmental impact. The figure 6 illustrates the Individuals who are less and more interest in EV using K-MA.

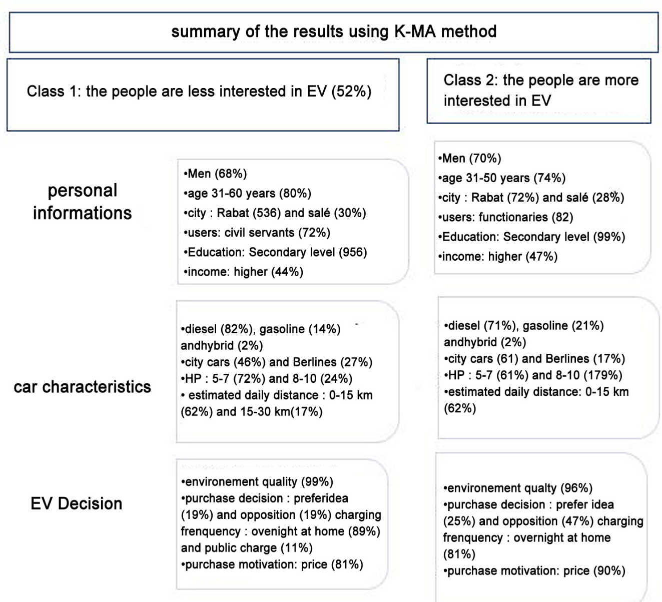

Figure 6. profiles of the participants using K-MA (lower and higher interest in EV)

4.4.2 Class 2: 48% of people are more interested in EV

This group consists of 70% men and 30% women, with 74% of them being active young adults (aged 31 to 60). In contrast to class 1, there are more people living in Salé (28%) and Rabat (72%). 99 % of those interviewed have a secondary education. 82% of them are functionaries, with 47% earning more than 1200 Euros per month and 23% earning less than 600 Euros per month. In addition, 83 % of respondents own a vehicle for transportation, of which 76% use diesel, 21% gasoline, and 2% hybrids. 61% of this group use city cars, 17% use Berlines with a fiscal power between 5 and 7 horsepower, and 17% use vehicles with a fiscal power between 8 and 10 horsepower. We can also conclude that 63% of respondents are knowledgeable about EVs and 96% are aware of their positive environmental impact. 62% of this group commute less than 15 kilometers per day, while 14% commute between 31 and 45 kilometers. In contrast to class 1, 23% of this group prefer to purchase an EV, while 47% oppose the idea and 25% are hesitant to purchase. 83% of those who prefer the purchase option prefer to charge their EV overnight at home at a discounted rate, while only 17% prefer public charging during the day. The purchase price is a deciding factor for 90% of EV buyers, compared to 9% who prefer EVs for their lower maintenance costs and 1% for their reduced environmental impact.
Finally, we can conclude that the LCA method's discrimination power (class 1 (prob=0.65) and class 2 (prob= 0.35) is superior to that of the K-Mode method (class 1 : 48% and class 2 : 52 %). Moreover, we can observe that the first method identifies two profiles significantly, whereas the second method classifies individuals into two distinct groups only slightly.

5 Conclusions

It is essential for those in charge of making decisions and formulating policies in the field of environmentally sustainable transportation to be able to recognize subgroups of individuals who have the same preferences and viewpoints regarding electric vehicles. This will allow them to fully understand and anticipate the impact that this emerging industry will have not only on the automotive industry but also on the environment. This study seeks to identify the profiles of Moroccans who are likely to use Electric vehicles as a future alternative to sustainable transportation and who may be the focus of a strategy to replace conventional vehicles. To answer the research question, a questionnaire was administered to a population restricted to the Rabat-Sale metropolitan area.
After that, MCA was applied to the sample in order to describe the relationships that existed between the variables and the individuals in the data that was obtained. According to MCA results, the first five axes account for 29.84 % of the information, whereas the first two axes alone account for 14.20 % of the observed fluctuations in our sample. We can conclude that the information provided by the first two axes is insufficient to allow us to draw meaningful judgments. This problem can be solved using an unsupervised machine learning method, such as the LCA method or KMA.
In this research, we use two techniques: LCA and KMA to determine the preferences of Moroccans with regard to the EV. We discovered that the Moroccan context is suitable for the implementation of this policy. Both algorithms are able to distinguish the sample being evaluated into two distinct groups, with the first group displaying a greater level of enthusiasm for EVs. The second category includes people who care less about using environmentally friendly modes of transportation:

• class 1 of both algorithms is made of individuals who are less interested in switching to EV. This group accounts for 52 % of the sample for KMA Vs 65 % for LCA. The majority of these individuals own a vehicle, which may explain why they are less enthusiastic about purchasing an EV. Surprisingly, 99 % for KMA and (prob=0.98) for LCA, of those in this group are aware of the environmental benefits of driving an EV. However, the price constitutes a deciding factor for this class since few participants (prob=0.10) for LCA are considering purchasing an EV.

• class 2 for both algorithms correspond to one-third of the population sampled. These individuals are more inclined to switch to EV. However, the price of an EV may discourage its purchase. Consequently, decision-makers are obligated to meet the needs of this demographic to ensure a promising start for EV sales, putting the nation on a promising path of electric mobility and contributing to the reduction of CO2 emissions. These individuals are considered to be active young adults, which may explain their increased interest in EV. The average/high income of this group of people is an additional factor that merits mention. In contrast to their reduced environmental impact, this group considers the lower maintenance costs to be the most compelling reason to purchase an EV.
In this study, we discovered that the LCA method performs better and is more efficient than the KMA, and that the LCA method has a higher discrimination power (65 % vs. 35%) than the KMA (52 % vs 48 %). In addition, rather than ratios, LCA provides information on the likelihood that an individual belongs to a specific class.

In terms of perspectives, we intend to expand the scope of this research to include the entire Moroccan territory and use other clustering techniques such as Fuzzy Clustering.

Acknowledgements

We are very grateful to experts for their appropriate and constructive suggestions to improve this paper.

REFERENCES

[1] Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.), Climate Change 2014:

Mitigation of Climate Change, Contribution of Working Group III to the Fifth Assessment Report of the Intergouvernemental Panel on Climate Change . Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

[2] Acharya S., and S. Morichi, Motorization and Role of Mass Rapid Transit in East Asian Megacities, IATSS Research, Vol.31, 6–16, 2007.

[3] Schäfer A., The Future of Energy for Urban Transport, Urban Transport in the Developing World: A Handbook of Policy and Practice, Edward Elgar, Northampton, 113–136, 2011.

[4] Chachdi, A., Rahmouni, B., and Aniba, G, Socio-economic analysis of ELectric Vehicles in Morocco, Energy Procedia, Vol.141, 644-653, 2017.

[5] Liberto, C., Valenti, G., Orchi, S., Lelli, M., Nigro, M., and Ferrara, M, The impact of electric mobility scenarios in large urban areas: The rome case study, IEEE Transactions on Intelligent Transportation Systems,Vol.19, No.11, 3540-3549, 2018.

[6] Dombrowski, U., and Engel, C, Impact of electric mobility on the after sales service in the automotive industry, Procedia CIRP, Vol.16, 152-157, 2014.

[7] El Harrouti, T., Abouabdellah, A., and Serrou, D, Impact of electric mobility on the sustainable development of the country, Case study in Morocco, In 2020 IEEE 13th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA), IEEE, 1-6, 2020.

[8] James B. Schreiber and Andrew, echnical Note: Using Latent Class Analysis versus K-means or Hierarchical Clustering to Understand Museum Visitors, Curator The Museum Journal, Vol.57, 2014.

[9] Krantz, A., R. Korn, and M. Menninger., Rethinking museum visitors: Using K-means Cluster Analysis to explore a museum's audience, Curator: The Museum Journal, vol.52, No.4, 363–374, 2009.

[10] Lacourse, E., Baillargeon, R., Dupéré, V., Vitaro, F., Romano, E., & Tremblay, R, Two-year predictive validity of conduct disorder subtypes in early adolescence: a latent class analysis of a Canadian longitudinal sample, Journal of Child Psychology and Psychiatry, Vol.51, No.12, 1386-1394, 2010.

[11] Burke, J. D, An affective dimension within oppositional defiant disorder symptoms among boys: personality and psychopathology outcomes into early adulthood, Journal of Child Psychology and Psychiatry, Vol.53, No.11, 1176-1183, 2012.

[12] Klonsky, E. D., and Olino, T. M, Identifying clinically distinct subgroups of self-injurers among young adults: a latent class analysis? Journal of consulting and clinical psychology, Vol. 76, No.1, 22, 2008.

[13] Liu, Z., Zhang, Y., Tian, L., Sun, B., Chang, Q., & Zhao, Y. Application of latent class analysis in assessing the competency of physicians in China, BMC medical education, Vol.17, No.1, 1-8, 2017.

[14] Rossi, A., Marconi, M., Mannarini, S., Minelli, I., Rossini, C., Morena, R., and Verusio, C, Profiling cancer-related needs and the role of physical, practical and psychological difficulties: A latent class analysis approach, 2018.

[15] Denson, N, Latent class analysis in higher education: An illustrative example of pluralistic orientation. Research in Higher Education, Vol.55, No.5, 508-526, 2014.

[16] Ing, M., & Nylund-Gibson, K. , Linking early science and mathematics attitudes to long-term science, technology, engineering, and mathematics career attainment: Latent class analysis with proximal and distal outcomes, Educational Research and Evaluation, Vol.19, No.6, 510-524, 2013.

[17] Rhead, R., Elliot, M., and Upham, P, Using latent class analysis to produce a typology of environmental concern in the UK, Social Science Research, Vol.74, 210-222, 2018.

[18] Fox, B., Moule, R. K., & Parry, M. M., Categorically complex: A latent class analysis of public perceptions of police militarization, Journal of Criminal Justice, Vol.58, 33–46, 2018.

[19] Bentaleb, Y., Abarda, A., Mharzi, H., & El Hajji, S., Application of latent class analysis to identify the youth population who risk being cybercrime victim on social networks, Contemporary Engineering Sciences, Vol.8, No.32, 1529-1534, 2015.

[20] Huang, C., Dimensions of Taiwanese/Chinese identity and national identity in Taiwan: a latent class analysis. Journal of Asian and African Studies, Vol.40, No.1-2, 51-70, 2005.

[21] Oser, J., and Hooghe, M, Democratic ideals and levels of political participation: The role of political and social conceptualisations of democracy, The British Journal of Politics and International Relations, Vol.20, No.3, 711-730, 2018.

[22] Alvarez, R. M., Levin, I., and Núñez, L, The four faces of political participation in Argentina: Using latent class analysis to study political behavior, The Journal of Politics, Vol.79, No.4, 1386-1402, 2017.

[23] Hagenaars, J. A., and McCutcheon, A. L. (Eds.), Applied latent class analysis, Cambridge University Press, 2002.

[24] Vermunt, J. K., and Magidson, J, Latent class models for classification, Computational Statistics and Data Analysis, Vol.41, No. 3-4 , 531-537, 2003.

[25] Abarda, A., Bentaleb, Y., El Moudden, M., Dakkon, M., Azhari, M., Zerouaoui, J., and Ettaki, B, Solving the problem of latent class selection, In Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications, 1-6, 2018.

[26] Acharoui, Z., Alaoui, A., Azhari, M., Abarda, A., Ettaki, B., and Zerouaoui, J, Using Latent Class Analysis to Identify Political Behavior of Moroccan Citizens on Social Media, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.9, No.6, 2020.

[27] Abarda, A., Dakkon, M., Azhari, M., Zaaloul, A., and Khabouze, M, Latent transition analysis (LTA): A method for identifying differences in longitudinal change among unobserved groups, Procedia Computer Science, Vol.170, 1116-1121, 2020.

[28] Huang, J. Z., Clustering categorical data with k-Modes, In Encyclopedia of Data Warehousing and Mining, Second Edition, IGI Global, 246-250, 2009.

[29] Zhou, H., Zhang, Y., Liu, Y.,, A global-relationship dissimilarity measure for the k-modes clustering algorithm, Computational Intelligence and Neuroscience, 2017.

[30] Chaturvedi, A., Green, P. E., Caroll, J. D., K-modes clustering, Journal of classification, vol.18, No.1 , 35-55, 2001.

[31] Izakian, H., Abraham, A., & Snášel, V., Clustering categorical data using a swarm-based method, In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), 1720-1724, IEEE, 2009.

[32] Khan, S. S., Ahmad, A., Cluster center initialization algorithm for K-modes clustering, Expert Systems with Applications, Vol.40, No.18, 7444-7456, 2013.

[33] Sharma, N., Gaud, N. , K-modes clustering algorithm for categorical data, International Journal of Computer Applications, Vol.127, No.17 , 1-6, 2015.

[34] Higginbottom, G. M. A, Sampling issues in qualitative research, Nurse Researcher, Vol.12, No.1, 2004.

[35] Oliver C. Robinson, Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide, Qualitative Research in Psychology, Vol.11, No.1, 25-41, 2014.

[36] Choulakian, V. , Analyse factorielle des correspondances de tableaux multiples, Revue de statistique appliquée, Vol.36, No.4, 33-41, 1988.