Supplementary Materials

List of Content (chronological order based on the manuscript content):

Table S1. Corresponding meaning of each feature.

Feature	Measure
AC1	During the pandemic, public utility buses are accessible on weekdays.
AC2	During the pandemic, public utility buses are accessible on weekends.
AC3	During the pandemic, public utility buses are accessible within daytime working hours (7 AM to 7 PM).
AC4	During the pandemic, public utility buses are accessible outside normal working hours (7 PM to 7 AM).
AC5	During the pandemic, the waiting time at the bus stop is short.
AC6	There is a sufficient number of public utility buses despite the pandemic.
AC7	Bus stops are located at a reasonable distance.
AC8	There is a variety of payment methods.
SA1	During rush hour, there is an order of getting on or off the public utility bus.
SA2	Accidents seldom happen when riding a public utility bus.
SA3	Crimes seldom happen when riding a public utility bus.
SA4	Safety measures are followed by drivers and bus operators.
SA5	I feel safe every time I ride a public utility bus.
SA6	I feel comfortable sharing the public utility bus space with fellow passengers.
EB1	Public utility bus fares are fair and affordable.
EB2	I only allot a small percentage of my income to travel expenses.
EB3	I can save money when riding a public utility bus.
EB4	I ride a public utility bus due to its affordability.
EB5	Through public utility buses, both I and transport operators obtain financial gains.
EB6	Increasing the number of public utility buses would result in positive economic gain.
CM1	Passengers of public utility buses follow COVID-19 precautions.
CM2	Drivers and bus operators follow COVID-19 precautions.
CM3	Drivers and bus operators value COVID-19 precautions to protect commuters.
CM4	I believe that the government values the health of commuters.
CM5	The government ensures that appropriate COVID-19 precautions are followed in public utility buses.
CM6	I appreciate the COVID-19 precautions mandated by the public utility bus system and the government.
CM7	I acknowledge the effectiveness of COVID-19 precautions followed in public utility buses.
TR1	Drivers and bus operators have good manners.
TR2	Frequent use of public utility buses would create employment opportunities.
TR3	I think that public utility buses are reliable during the pandemic.
TR4	I think that public utility buses are essential during the pandemic.
TR5	I feel comfortable riding public utility buses during the pandemic.
TR6	I trust public utility buses for my daily commute despite the pandemic.
TR7	I feel satisfied with the implemented COVID-19 precautions practiced in public utility buses.
AT1	I prefer to use a public utility bus than a private car despite the pandemic.
AT2	I feel content with traveling by public utility bus despite the pandemic.
AT3	I trust that everyone on the public utility bus follows appropriate COVID-19 precautions.
AT4	I feel that I have a smaller risk of getting infected with COVID-19 when riding public utility buses.
AT5	I feel satisfied with the current public utility bus system.
AT6	It is highly acceptable to use public utility buses despite the pandemic.

SN1	People around me often use public utility buses despite the pandemic.
SN2	If people around me choose to ride public utility buses, I feel the need to go along with them despite the pandemic.
SN3	If I choose to ride a public utility bus, people around me would also do the same despite the pandemic.
SN4	People whose opinions I value affect my decision to use a public utility bus.
SN5	During the pandemic, my friends and family expect me to use a public utility bus.
SN6	During the pandemic, everyone supports me in using a public utility bus.
PBC1	The choice to take public utility buses is dependent on my decision.
PBC2	I find it easy to ride public utility buses despite the pandemic.
PBC3	I find it acceptable to actively use public utility buses despite the pandemic.
PBC4	I am equipped with knowledge of COVID-19 precautions followed in public utility buses.
PBC5	I believe that COVID-19 precautions in public utility buses protect me from getting infected with COVID-19.
PBC6	I am confident that I won't get infected with COVID-19 when riding public utility buses.
IU1	I will make an effort to use a public utility bus despite the pandemic.
IU2	I will ride a public utility bus during rush hour despite the pandemic.
IU3	I will ride a public utility bus to carry out essential duties despite the pandemic.
IU4	I will ride a public utility bus for leisure purposes despite the pandemic.
IU5	I plan to increase my travel frequency using a public utility bus despite the pandemic.
IU6	My intention to use a public utility bus is strong despite the pandemic.

Feature Selection Equations:

In performing one of the feature selection techniques, the following filter method's correlation equation was applied:

$$\frac{\sum_{x=1}^{n}(f_{i,x}-\bar{f}_i)(f_{j,x}-\bar{f}_j)}{\sqrt{\sum_{x=1}^{n}(f_{i,x}-\bar{f}_i)^2}\sqrt{\sum_{x=1}^{n}(f_{j,x}-\bar{f}_j)^2}} \quad \text{(S1)}$$

where n, $f_{i,x}$, $f_{j,x}$, $\bar{f}_i$, $\bar{f}_j$ are the number of observations, values of features from independent variables, values from class/dependent variable, mean values of features from independent variables, and mean values from class, respectively.

Next, the filter method's univariate selection score was calculated using Eq. (S2):

$$\frac{\sum_{i=1}^{l} n_i \frac{(\bar{f}_i^k-\bar{f}_j^k)^2}{l-1}}{\sum_{i=1}^{l}\sum_{j=1}^{n_i}\frac{(\bar{f}_{ij}^k-\bar{f}_i^k)^2}{n-1}} \quad \text{(S2)}$$

where l, n, $\bar{f}_i^k$, $\bar{f}_j^k$ pertain to the number of features under the class, features under independent variables, the mean value of f_k in class i, and overall mean value, respectively.

The corresponding equation for the wrapper method's backward elimination was utilized as follows:

$$f_0 + Xf_1 + \varepsilon \quad \text{(S3)}$$

where f_0, f_1, ε are y-intercept, X coefficient of features under independent variables, and the error term.

For the embedded-LASSO method, Eq. (S4) signifies its formula:

$$\sum_{i=1}^{n}(y_i - \sum_{j=0}^{p}(w_j)(x_{ij}))^2 + \lambda \sum_{j=0}^{p} w_j^2 \quad \text{(S4)}$$

where $(y_i - \sum_{j=0}^{p}(w_j)(x_{ij}))^2$ refers to the sum of squared residuals, while $\lambda \sum_{j=0}^{p} w_j^2$ pertains to penalty.

PSO Initialization Equations:

Other necessary formulas to prepare the dataset of combined K-means and PSO were presented in the following equations. The initial number of clusters *(M)* was calculated using Eq. (S5):

$$M = \sqrt{TCD} \quad \text{(S5)}$$

where TCD is the total collected data.

For the normalization process, the following equations were used:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \text{(S6)}$$

$$Mean\ Data = \frac{\sum X_i}{K} \quad \text{(S7)}$$

where X_{min} is the minimum value, X_{max} is the maximum value in the data set, and K is the number of data points.

Table S2. Summary of *gbest* fitness descriptive statistics.

Parameter	Mean	Standard Deviation	Minimum	Maximum
1	6432.42	10.66	6413.1	6449.9
2	6424.26	15.63	6387.3	6445.7
3	6423.10	8.57	6411.2	6434.9
4	6702.46	361.79	6408.5	7202.9
5	6420.91	13.09	6405.6	6440.0
6	6422.55	8.65	6405.7	6434.3
7	6434.57	14.08	6421.8	6461.6
8	6422.19	8.3	6404.4	6434.1
9	6426.99	12.14	6404.0	6439.3
10	6692.55	362.59	6398.1	7188.1
11	6423.46	7.82	6414.9	6436.1
12	6415.86	15.89	6389.7	6434.2

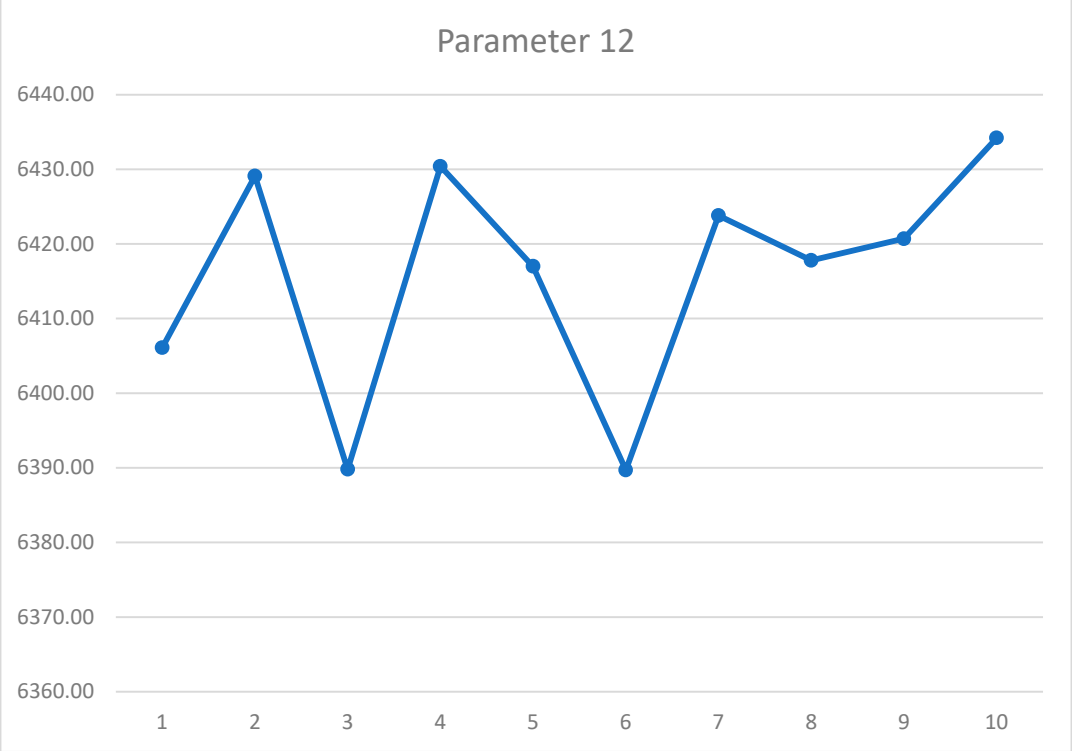

Figure S1. The learning curve of the optimal solution.

Figure S2 reflects four possibilities of cluster combinations. Since K-means and PSO's optimal parameter setting supported the convergence to the global optimum, participants were clustered within the same groups regardless of the number of runs. As illustrated in Figure S2, participants' cluster numbers varied for each plot, but the respective groupings were constant. Clusters 1 to 5 and 18 to 23 had similar participant cluster groups for all plots. Meanwhile, they had varying cluster combinations from clusters 6 to 17. Nevertheless, participants still belonged to the same groupings. For instance, cluster 6 in plots 1, 3, and 4 comprised one 1 participant (Participant #129), while cluster 6 in plot 2 had 69 participants. Looking into the figure meticulously, participant #129 was under plot 2's cluster 7. Then, plot 2's cluster 6 with the same set of participant numbers was found in the 10th cluster of plot 1, 17th cluster of plot 3, and 8th cluster of plot 4. A similar principle was applied across all cluster plots regardless of the number of runs. This result supported that parameter 12 is the best combination as it yielded constant groupings for all participants. Furthermore, only one plot was needed because the four plots generated the same results. In this case, plot 1 was analyzed in the succeeding processes.

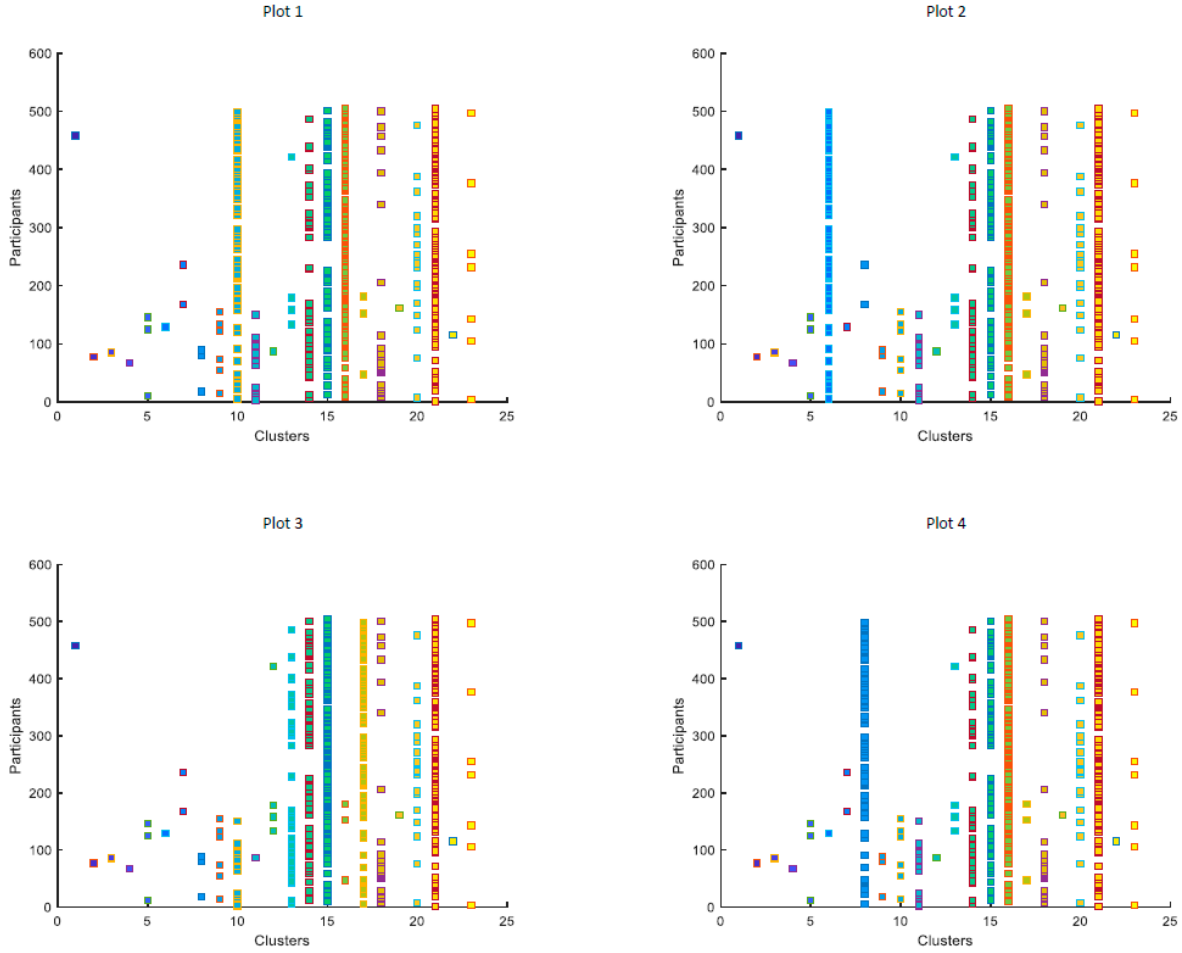

Figure S2. Cluster groups of participants using the optimal parameter settings.

Table S3. Descriptive analysis of PUB passengers' clusters.

Cluster Number	Participant Number
C16	9,10,23,31,37,40,58,76,84,96,99,100,102,104,112,116,127,130,132,136,139,141,151,160,163,164, 167,175,177,178,180,182,184,185,186,188,193,194,199,200,202,204,207,211,215,219,222,233,234, 242,243,244,248,252,253,256,257,258,259,263,267,268,270,274,275,277,279,282,286,287,295,298, 301,304,307,310,311,312,313,316,323,325,331,332,333,336,337,339,344,347,363,365,366,367,368, 369,370,378,382,383,384,386,390,399,406,407,409,413,414,420,422,425,427,429,436,443,452,453, 454,456,462,465,466,469,478,480,483,489,490,491,494,495,502,505
C21	1,19,30,33,34,35,36,38,39,49,52,71,95,98,110,113,117,131,135,142,144,145,147,156,159,173,187, 189,192,195,197,201,208,209,212,220,224,226,187,189,192,195,197,201,208,209,212,220,224,226, 269,273,278,281,289,294,315,317,326,335,341,342,350,351,354,356,357,359,374,380,381,387,391, 395,397,401,403,404,405,408,411,418,419,426,428,430,431,434,445,447,451,460,461,471,477,479, 493,498,503,504
C10	5,22,32,41,43,46,48,70,91,120,128,157,165,174,176,190,196,213,214,216,217,218,229,240,245,262, 266,272,276,280,285,292,296,297,321,322,328,329,334,349,353,360,361,371,375,377,385,392,396, 400,410,412,417,432,435,448, 449,450,455,464,470,474,484,485,487,488,492,496,499
C15	13,27,28,44,59,64,68,74,88,90,103,106,118,126,138,162,171,172,183,191,205,210,221,223,225,284, 288,291,293,302,306,309,314,327,330,338,343,345,346,348,355,358,372,379,389,393,415,416,423, 424,438,440,441,442,444,446,459,463,467,468,472,475,481,482,501
C14	6,12,42,45,56,65,69,79,93,101,108,119,121,137,140,148,153,154,166,169,228,230,283,300,303,305, 308,318,319,324,352,364,373,398,402,437,439,486
C18	8,15,17,21,26,29,50,51,53,54,57,60,61,62,66,78,82,92,114,206,340,394,433,457,473,500
C20	7,75,123,124,149,170,198,203,231,237,238,250,254,271,290,299,320,362,388,476
C11	2,3,16,20,24,25,63,72,80,83,94,97,107,109,111,150
C23	4, 105, 143, 232, 255, 376, 497
C9	14,55,73,122,134,155
C13	133,158,179,421
C5	11,125,146
C8	18,81,89
C17	47,152,181
C3	85,86
C7	168,236
C1	458
C2	77
C4	67
C6	129
C12	87
C19	161
C22	115

Table S4. Pareto analysis.

Cluster Number	Frequency	Cumulative Frequency	Cumulative Frequency Percentage
C16*	134	134	26.53%
C21*	100	234	46.34%
C10*	69	303	60.00%
C15*	65	368	72.87%
C14*	38	406	80.40%
C18	26	432	85.54%
C20	20	452	89.50%
C11	16	468	92.67%
C23	7	475	94.06%
C9	6	481	95.25%
C13	4	485	96.04%
C5	3	488	96.63%
C8	3	491	97.23%
C17	3	494	97.82%
C3	2	496	98.22%
C7	2	498	98.61%
C1	1	499	98.81%
C2	1	500	99.01%
C4	1	501	99.21%
C6	1	502	99.41%
C12	1	503	99.60%
C19	1	504	99.80%
C22	1	505	100.00%

*Vital clusters that met the 80% cut-off.

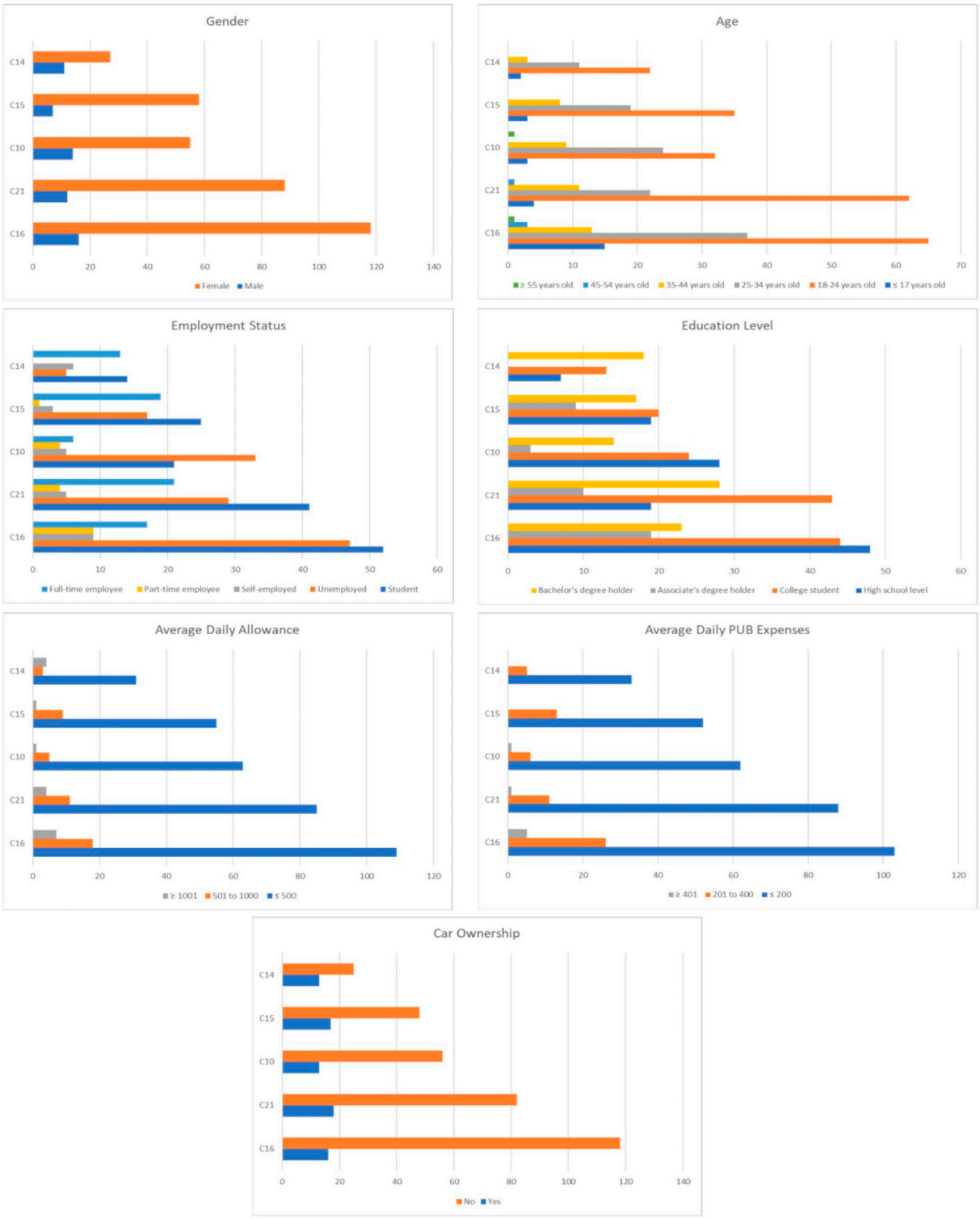

Figure S3. Cluster comparison according to demographics.

Vital Clusters' Demographic Profiles

Cluster 16 comprised all age ranges compared to other clusters, which lacked one or two age ranges (45-54 years old and ≥ 55 years old). Thus, cluster 16 results were more inclusive than other clusters. Cluster 21 comprised teenagers to middle-aged adults (≤ 17 years old to 54 years old), while Cluster 10 contained teenagers, early adults, and seniors (≤ 17 years old to 44 years old and ≥ 55 years old). Clusters 15 and 14 comprised the younger generation in the mid-40s since they encompassed ≤ 17 to 44 years old.

Next, clusters 16, 21, 15, and 14 mostly comprised students, and cluster 10 was dominated by unemployed individuals. It was also noted that the 2nd highest employment status for clusters 15 and 14 was full-time employees. Thus, clusters 16, 21, and 10 were mostly comprised of students and unemployed individuals., while clusters 15 and 14 had a combination of students and full-time employees.

Based on the participants' education level, clusters 16 and 10 were dominated by high school and college. Cluster 21 had more college students than bachelor's degree holders. Cluster 15 yielded the opposite result to clusters 16 and 10, where college students dominated the clusters instead of high school students or graduates. Also, cluster 14 generated an opposite result to cluster 21, where it had more bachelor's degree holders than college students.

The overall allowance and PUB expenses had analogous results to clusters 16, 21, and 10. Thus, participants of these clusters preferred the least allowance and expenses, followed by the middle allowance and expenses and high allowance and expenses. However, the clustering method revealed a different result for clusters 15 and 14. Although some participants had a high allowance, they only preferred the least and middle expenses. Interestingly, none of the participants from clusters 15 and 14 spent the highest PUB expenses.

The aforementioned cluster description was explained directly without the cross-study comparisons due to the lack of relevant studies. The researchers proposed new sets of features and methods. Hence, these novel findings could be used by future scholars as a benchmark in understanding PUB passengers' clusters during COVID-19.

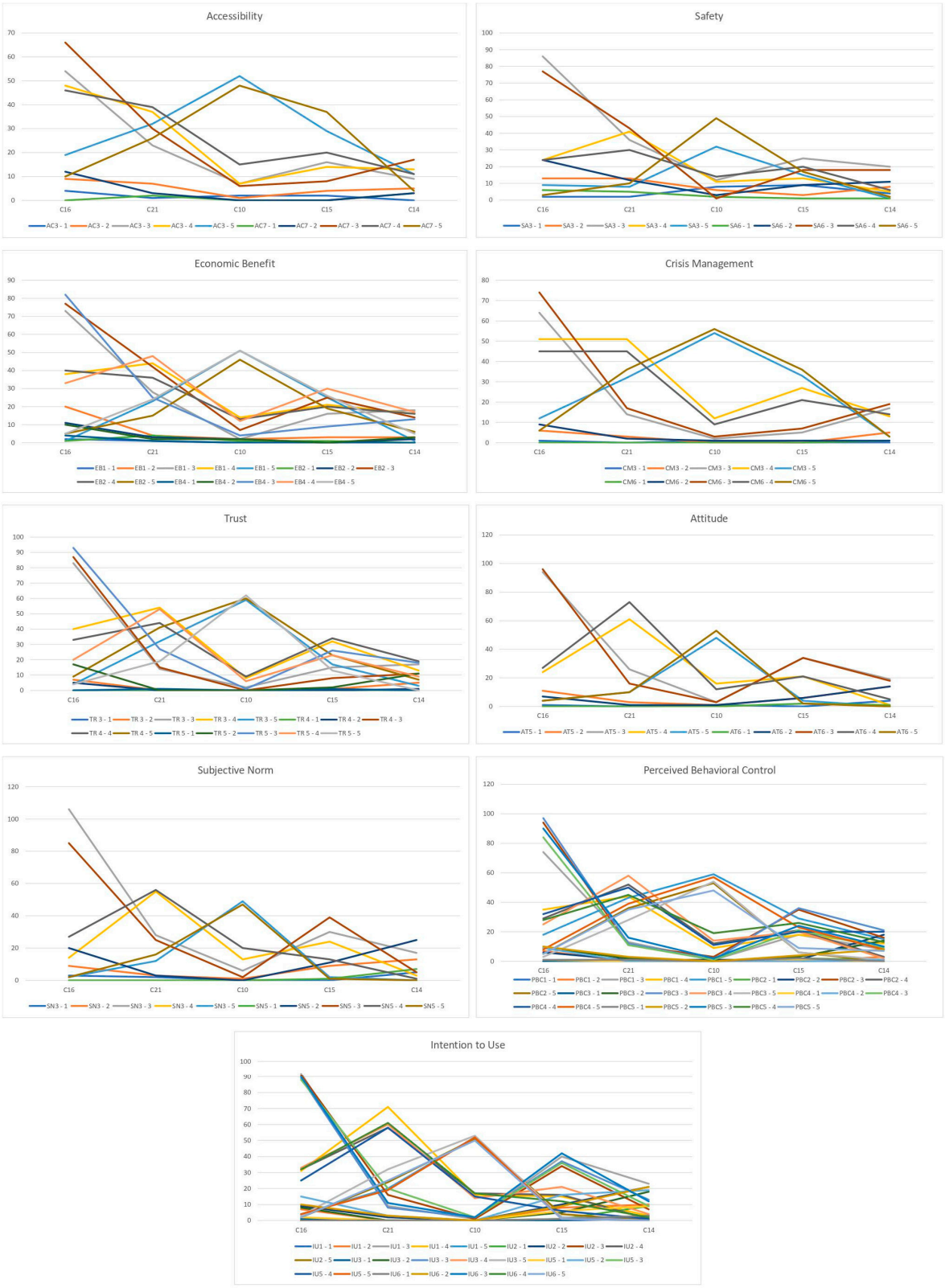

Figure S4. Cluster comparison according to features.

Five Clusters' Comprehensive Details

Cluster 16 comprised all age ranges but was dominated by high school graduates and high school/college students who had the least allowances and expenses. Also, a lot of PUB passengers were unemployed. Most importantly, PUB passengers responded neutrally to all 26 optimal features. Therefore, PUB passengers under cluster 16 neither agree nor disagree with the current PUB system and services based on the 26 features' characteristics.

Cluster 21 PUB passengers were teenagers to middle-aged adults but most of them were college students with limited financial resources. Although other PUB passengers of cluster 21 were bachelor's degree holders, they were most likely fresh graduates due to unemployed employment status, resulting in preferring the least amount for both PUB allowances and expenses. Cluster 21 agreed with 22 optimal features except for responding neutrally to SA6 and EB2 and extremely agreed with SN3 and SN5.

Cluster 10 had participants who were teenagers, early adults, and seniors. Most of them attained high school level, were still college students, and were unemployed. Given their status and inability to produce finances, they chose the least allowances and expenses next to middle and high ranges. Moreover, most PUB passengers extremely agreed with all 26 optimal features. Thus, these 26 features were strong indicators of PUB passengers' behaviors, and the majority of the passengers were satisfied with the Philippines' PUB system during the pandemic.

Meanwhile, cluster 15's passengers consisted of the younger generation to mid-40s. While most of them were college students, there were a few who had a high allowance but only preferred least to middle expenses. Cluster 15 had varying results of neutral, agree, and strongly agreeing responses. Most participants extremely agreed with optimal features under accessibility and crisis management and responded neutrally to optimal features under attitude, subjective norm, and intention to use. However, cluster 15 had varying types of responses for safety, economic benefit, trust, and perceived behavioral control features.

Finally, cluster 14's participants were the younger generation to mid-40s, who mostly acquired a bachelor's degree. Since this cluster had many students though they had already completed college, it could be noted that other participants pursued further studies. Similar to cluster 15 wherein some participants had a high allowance, but they only preferred the least to middle expenses. Cluster 14 is more diverse since their responses ranged from disagree to strongly agree. Cluster 14 was the sole cluster to have disagreeable participants, while clusters 16, 21, 10, and 15 had neutral to strongly agree on responses. Therefore, cluster 14 should be prioritized by the PUB stakeholders as some of them disagree with the current system. This instance supported the need to improve external and internal variables affecting PUB passenger behaviors.

Table S5. Summarized analysis of the vital clusters.

Cluster Number	Feature	Clustered Feature Response
16	All 26 features (AC3, AC7, SA3, SA6, EB1, EB2, EB4, CM3, CM6, TR3, TR4, TR5, AT5, AT6, SN3, SN5, PBC1, PBC2, PBC3, PBC4, PBC5, IU1, IU2, IU3, IU5, and IU6)	Neutral
21	AC3 and AC7	Agree
	SA3	Agree
	SA6	Neutral
	EB1 and EB4	Agree
	EB2	Neutral
	CM3 and CM6	Agree
	TR3, TR4, and TR5	Agree
	AT5 and AT6	Agree
	SN3 and SN5	Extremely Agree
	PBC1, PBC2, PBC3, PBC4, and PBC5	Agree
	IU1, IU2, IU3, IU5, and IU6	Agree
10	All 26 features (AC3, AC7, SA3, SA6, EB1, EB2, EB4, CM3, CM6, TR3, TR4, TR5, AT5, AT6, SN3, SN5, PBC1, PBC2, PBC3, PBC4, PBC5, IU1, IU2, IU3, IU5, and IU6)	Extremely Agree
15	AC3 and AC7	Extremely Agree
	SA3	Neutral
	SA6	Agree
	EB1	Extremely Agree
	EB2	Neutral
	EB4	Agree
	CM3 and CM6	Extremely Agree
	TR3 and TR4	Agree
	TR5	Neutral
	AT5 and AT6	Neutral
	SN3 and SN5	Neutral
	PBC1 and PBC4	Extremely Agree
	PBC2 and PBC3	Neutral
	PBC5	Agree
	IU1, IU2, IU3, IU5, and IU6	Neutral

	AC3	Agree
	AC7	Neutral
	SA3 and SA6	Neutral
	EB1	Neutral
	EB2 and EB4	Agree
	CM3 and CM6	Neutral
	TR3 and TR5	Neutral
	TR4	Agree
14	AT5 and AT6	Neutral
	SN3	Neutral
	SN5	Disagree
	PBC1	Extremely Agree
	PBC2	Disagree
	PBC3	Neutral
	PBC4 and PBC5	Agree
	IU1	Neutral
	IU2, IU3, IU5, and IU6	Disagree