Team 3 – Brendan Pernecky, Huan Chong, Jiawei Huang, Lily Botueva, Sia Savla, William Knibbs

1. It is critical to determine whether there is a relationship between the 30 attribute variables and the target mvliking variable. To do so, run a regression with mvliking as the dependent variable and all thirty independent variables as predictors. What do you find?

While the model is significant overall, we see that only Leather Seats and Shared Carpool are the only significant independent variables (with 0.043 and 0.02 p_value).

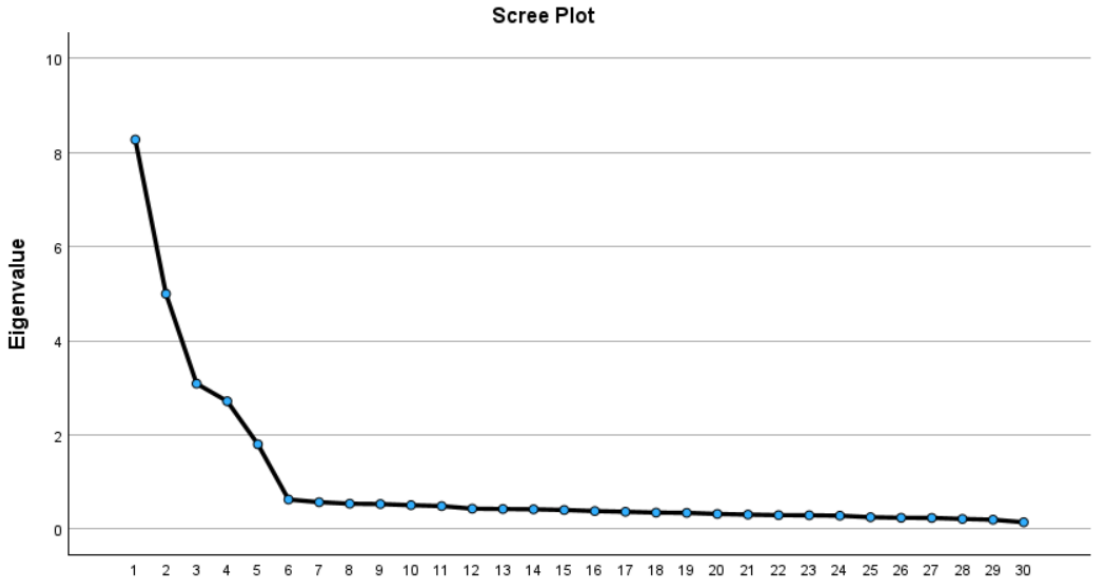
This means that with other factors constant, an increase in a unit of Leather Seats increase Concept Liking by 0.248, and an increase in a unit of Shared Carpool decreases Concept Liking by 0.287.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.381	2.959		.129	.898
	kidtrans	.241	.165	.136	1.463	.144
	miniboxy	.178	.129	.098	1.376	.170
	lthrbetr	.248	.122	.140	2.030	.043
	secbiggr	-.105	.106	-.059	-.990	.323
	safeimpt	-.019	.134	-.010	-.139	.890
	buyhghnd	.113	.116	.065	.969	.333
	pricqual	.105	.105	.059	1.005	.316
	prmsound	.010	.108	.006	.093	.926
	perfimpt	.233	.128	.129	1.815	.070
	tkvacatn	.166	.125	.094	1.333	.183
	noparkrm	.178	.116	.101	1.538	.125
	homlrgst	-.209	.122	-.116	-1.705	.089
	envrminr	-.033	.123	-.019	-.271	.787
	needbetw	.128	.103	.074	1.252	.211
	suvcmpct	.215	.123	.120	1.754	.080
	next2str	.024	.107	.014	.227	.820
	carefmny	-.243	.134	-.138	-1.809	.071
	shdcarpl	-.287	.122	-.160	-2.343	.020
	imprtapp	.059	.104	.033	.567	.571
	lk4whldr	-.064	.127	-.035	-.506	.613
	kidsbulk	-.097	.122	-.053	-.794	.428
	wntguzlr	-.029	.116	-.016	-.250	.803
	nordtrps	.073	.127	.041	.573	.567
	stylclth	.016	.114	.009	.139	.890
	strngwrn	-.197	.113	-.110	-1.735	.084
	passnimp	.162	.119	.091	1.360	.175
	twoincom	.170	.096	.099	1.767	.078
	nohummer	.009	.096	.005	.095	.925
	aftrschl	-.026	.117	-.014	-.221	.825
	accesfun	-.003	.122	-.002	-.028	.977

a. Dependent Variable: mvliking

2. The analyses run thus far address how well the 30 attribute variables explain the mvliking variable. These results may be more interpretable by uncovering the degree of redundancy in the 30 attribute variables. How would you go about doing this? Conduct all of the necessary steps in this analysis. After identifying the underlying factors, determine their meaning and name them accordingly.



Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.275	27.585	27.585	8.275	27.585	27.585
2	4.998	16.660	44.244	4.998	16.660	44.244
3	3.088	10.294	54.539	3.088	10.294	54.539
4	2.715	9.049	63.588	2.715	9.049	63.588
5	1.800	6.001	69.589	1.800	6.001	69.589
6	.626	2.088	71.676			
7	.569	1.896	73.572			

	Component				
	1	2	3	4	5
kidtrans	.123	.002	.933	-.019	.007
miniboxy	.123	.842	-.107	.051	.013
lthrbetr	.710	-.186	.247	.292	.067
secbiggr	-.076	.759	.059	.033	-.082
safeimpt	.026	.047	.054	-.020	.907
buyhghnd	.815	.176	.023	.057	.098
pricqual	.783	-.190	-.081	-.133	.004
prmsound	.681	-.017	.168	.292	.073
perfimpt	.111	-.081	-.075	.026	-.884
tkvacatn	.652	-.031	.258	.460	.023
noparkrm	.174	.807	.012	-.087	-.016
homlrgst	.330	-.679	.154	.318	.085
envrminr	-.171	-.035	.086	-.867	-.014
needbetw	.126	.758	-.011	.044	.042
suvcmpct	.083	.819	.202	.037	-.004
next2str	.258	-.743	.106	-.114	-.007
carefmny	-.762	-.145	-.198	-.310	-.080
shdcarpl	.161	-.028	-.055	.867	.078
imprtapp	.508	-.010	.346	.351	.200
lk4whldr	.167	.025	.032	.105	.856
kidsbulk	.179	.017	.825	.058	.022
wntguzlr	-.355	.032	-.008	-.764	.015
nordtrps	-.063	-.102	-.867	-.014	-.041
stylclth	.602	.236	.183	.428	-.027
strngwrn	.273	-.259	.083	.062	.735
passnimp	-.646	-.020	-.400	-.279	.014
twoincom	.756	.117	-.093	-.072	.095
nohummer	.060	.706	.046	-.037	.038
aftrschl	.195	-.109	.776	-.112	.182
accesfun	.677	-.039	.301	.372	.003

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

On conducting Factor Analysis, we see that there are 5 factors. They can be categorized as the following:

FACTOR	NAME	VARIABLES	MEANING
Factor 1	Affluent & Premium Experience	buyhghnd (.815) pricqual (.783) prmsound (.681) tkvacatn (.652) twoincom (.756) accesfun (.677)	People falling under this value high end features, are willing to pay for quality, want stylish and equipped cars, enjoy taking vacations, lead an affluent lifestyle, and are not frugal with

		Imprtapp (.508) lthrbetr (.710) stylcith (.602) passnimp (-.646) carefmny (-.762)	money, as they have higher disposable income due to 2 income sources.
Factor 2	Compact and Simple	miniboxy (.842) secbiggr (.759) noparkrm (.807) needbetw (.758) suvcmpct (.819) nohummer (.706) next2str (-.743) homlrgst (-.679)	They want a smaller, more manageable car, appreciate size of an SUV, and have parking restraints. They also might not have a high disposable income.
Factor 3	Family Transport	nordtrps (-.867) kidtrans (.933) kidsbulk (.825) aftrschl (.776)	They need it for practical needs of families. They need it to move bulky things associated with children, and are unlikely to take road trips.
Factor 4	Environmental	envrminr (-.867) shdcarpl (.867) wntguzlr (-.764)	They are environmentally conscious.
Factor 5	Safety First	safeimpt (.907) lk4whldr (.856) strngwrn (.735)	They care and prioritize safety more than performance of the car.

		perfimpt (-.884)	

3. Run another regression, this time using the saved factor scores as the independent variables and again using concept liking (mvliking) as the dependent variable. What does this new regression suggest compared with the prior regression results?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	967.512	5	193.502	39.717	<.001[b]
	Residual	1919.565	394	4.872		
	Total	2887.078	399			

a. Dependent Variable: mvliking

b. Predictors: (Constant), Safety, Environmental, Family, Compact, Affluent

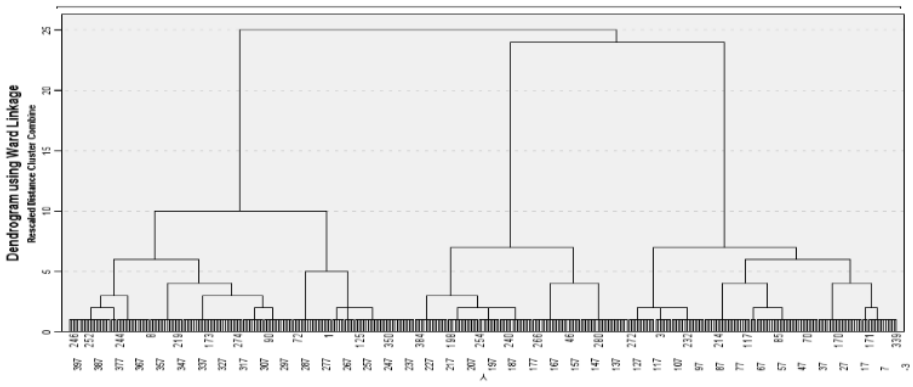
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.843	.110		43.878	<.001	4.626	5.059
	Affluent	1.044	.111	.388	9.448	<.001	.827	1.261
	Compact	.985	.111	.366	8.918	<.001	.768	1.203
	Family	.185	.111	.069	1.674	.095	-.032	.402
	Environmental	-.139	.111	-.052	-1.255	.210	-.356	.079
	Safety	-.557	.111	-.207	-5.041	<.001	-.774	-.340

a. Dependent Variable: mvliking

This new model suggests that the significant categories are Affluent, Compact, and Safety. Therefore, this means that a change in any of these categories would change the Concept Liking. Compared to the first regression we see that the underlying variables in these 3 factors do affect the liking, whereas the first one suggested only 2 variables.

4. **Segment the market based on the factors that you identified, by going through all of the necessary steps. Do the clusters make sense? On what factors do they differ? Name the clusters based on their identifying characteristics.**



Final Cluster Centers

	Cluster 1	Cluster 2	Cluster 3
Affluent	-.46927	.44551	.30047
Compact	-.66717	1.07399	-.05497
Family	.12110	.40452	-.64604
Environmental	.33374	.44372	-1.04601
Safety	.07630	-.17404	.06234

Report

Cluster Number of Case		Affluent	Compact	Family	Environmental	Safety
1	Mean	-.4692656	-.6671686	.1211022	.3337398	.0762960
	N	178	178	178	178	178
2	Mean	.4455104	1.0739869	.4045162	.4437226	-.1740403
	N	116	116	116	116	116
3	Mean	.3004724	-.0549667	-.6460384	-1.0460142	.0623396
	N	106	106	106	106	106
Total	Mean	.0000000	.0000000	.0000000	.0000000	.0000000
	N	400	400	400	400	400

Cluster Name	Description	Key Factor Characteristics
Practical & Environmentally Aware	Values practicality and environmental responsibility.	Affluent Lifestyle & Premium Experience: Negative (-.469) Compact Size Preference: Negative

		(-.667) Environmental Concern: Positive (.333)
Compact Prioritizers	Prioritizes a compact vehicle above all else.	Compact Size Preference: Positive (1.073)
Unbothered	Is concerned about any of the factor segments.	Compact Size Preference: Negative (-.054) Family Transportation Needs: Negative (-.646) Environmental Concern: Negative (-1.046)

5. **Determine how the clusters vary on the concept liking (mvliking) variable. This can be done in many different ways (e.g., regression of mvliking on the cluster id categorical variable, t-tests of the mean of mvliking for the different clusters, cross tabulation of cluster membership and discrete levels of mvliking, etc.). How do they differ?**

One-way ANOVA analysis tool shows that there is significant difference between at least one pair of clusters. Linear regression analysis with independent variable of cluster number also shows clusters significantly predict mvliking.

Oneway

ANOVA

mvliking

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	615.103	2	307.552	53.741	<.001
Within Groups	2271.974	397	5.723		
Total	2887.077	399			

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	226.297	1	226.297	33.850	<.001[b]
	Residual	2660.781	398	6.685		
	Total	2887.078	399			

a. Dependent Variable: mvliking

b. Predictors: (Constant), Cluster Number of Case

Cross tabulation analysis in SPSS also supports the result is statistically significant, meaning

mvliking scores differ significantly across the three clusters.

Cluster Number of Case * mvliking Crosstabulation

			mvliking									
			1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	Total
Cluster Number of Case	1	Count	49	18	27	26	22	11	13	5	7	178
		Expected Count	30.7	11.6	18.7	21.8	23.1	18.7	14.7	14.2	24.5	178.0
		Standardized Residual	3.3	1.9	1.9	.9	-.2	-1.8	-.4	-2.4	-3.5	
	2	Count	6	3	5	7	15	16	12	20	32	116
		Expected Count	20.0	7.5	12.2	14.2	15.1	12.2	9.6	9.3	16.0	116.0
		Standardized Residual	-3.1	-1.7	-2.1	-1.9	.0	1.1	.8	3.5	4.0	
	3	Count	14	5	10	16	15	15	8	7	16	106
		Expected Count	18.3	6.9	11.1	13.0	13.8	11.1	8.7	8.5	14.6	106.0
		Standardized Residual	-1.0	-.7	-.3	.8	.3	1.2	-.3	-.5	.4	
Total		Count	69	26	42	49	52	42	33	32	55	400
		Expected Count	69.0	26.0	42.0	49.0	52.0	42.0	33.0	32.0	55.0	400.0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	95.898[a]	16	<.001
Likelihood Ratio	100.391	16	<.001

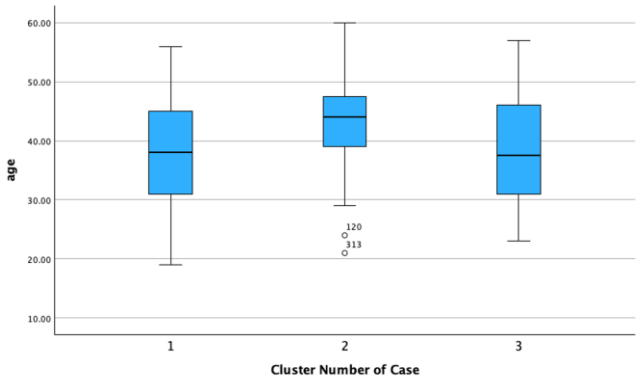
The standardized residuals for mvliking = 1 in Cluster 1 are positive, meaning there are more low-liking scores than expected. Our findings suggest that Cluster 1 contains people who are less favorable toward the concept.

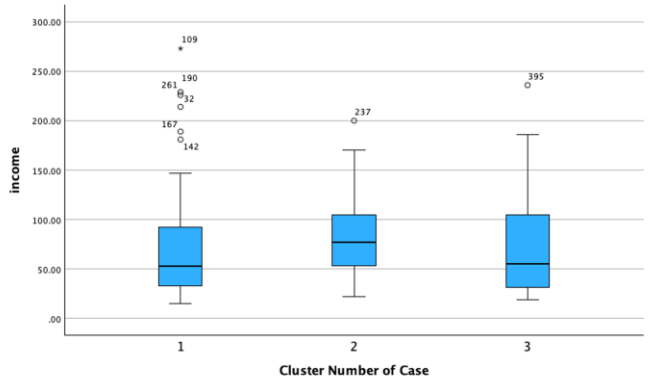
The standardized residuals are highly positive for mvliking = 9 in Cluster 2, meaning people in this cluster have a strong preference for the concept.

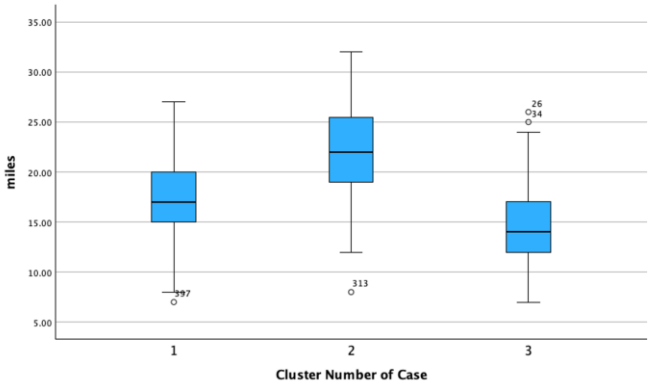
The distribution of mvliking in Cluster 3 is closer to expected counts, with somewhat moderate levels of liking.

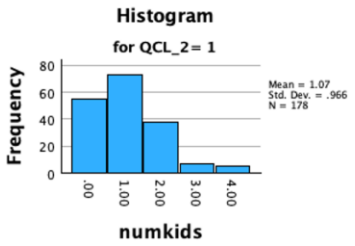
Since the difference is significant, this means that cluster membership plays an important role in determining mvliking levels. It is best to focus more on Cluster 2 “Compact Prioritizers“ as a key target audience while investigating why Cluster 1 “Practical & Environmentally Aware” is less engaged.

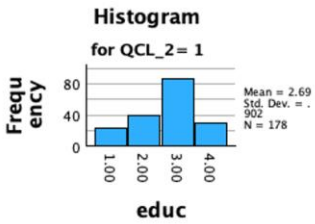
6. **As a final step, examine the demographic profile of each of the clusters. Given your findings about the “needs and wants” of each of the clusters, do their demographics make sense? Who are these people? Be sure to examine all available demographic variables.**

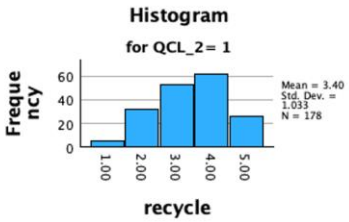

60.00
50.00
40.00
30.00
20.00
10.00
age
120
313
1
2
3
Cluster Number of Case

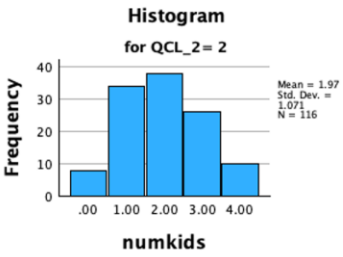

300.00
250.00
200.00
150.00
100.00
50.00
.00
income
109
190
261
32
167
142
237
395
1
2
3
Cluster Number of Case

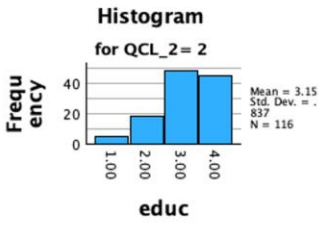

35.00
30.00
25.00
20.00
15.00
10.00
5.00
miles
26
34
313
197
1
2
3
Cluster Number of Case

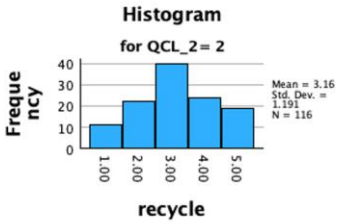


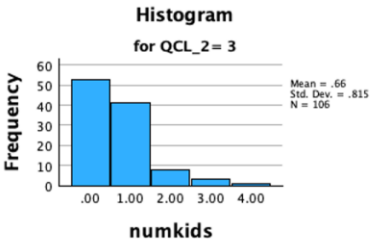


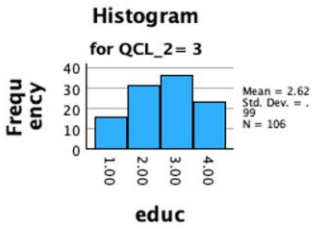


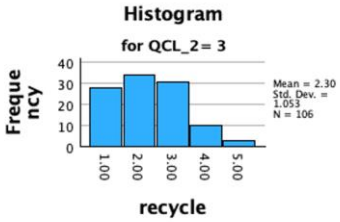












Cluster Number of Case * female

Crosstab

			female		
			.00	1.00	Total
Cluster Number of Case	1	Count	64	114	178
		Expected Count	81.9	96.1	178.0
		Standardized Residual	-2.0	1.8	
	2	Count	49	67	116
		Expected Count	53.4	62.6	116.0
		Standardized Residual	-.6	.6	
	3	Count	71	35	106
		Expected Count	48.8	57.2	106.0
		Standardized Residual	3.2	-2.9	
Total		Count	184	216	400
		Expected Count	184.0	216.0	400.0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	26.675[a]	2	<.001
Likelihood Ratio	26.950	2	<.001
Linear-by-Linear Association	23.848	1	<.001
N of Valid Cases	400		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 48.76.

	Age	Income	Miles	Numkids	Female	Educ	Recycle
Practical & Environmentally Aware	30-45	<$50k	15-20k	0-1	1:2 (more woman)	Undergraduate degree	Most
Compact Prioritizers	40+	$50-$100k	15-30k	1-3	1:1	Undergraduate or higher	Moderate
Unbothered	25-40	<$50k-$150k	10-17k	0-1	2:1 (less woman)	Less formal education	Less

Cluster 1: This group is environmentally conscious and practical. They likely value sustainability and eco-friendly practices. They may prioritize cost-effective and fuel-efficient transportation. A higher proportion of women suggests possible preferences for vehicles with safety and environmental features.

Cluster 2: This group is older and more financially stable, likely balancing work, family, and practicality. Their moderate recycling habits indicate that while they care about sustainability, it is not their top priority. Since they drive more miles (15–30k), they might prioritize comfort, efficiency, and reliability in their vehicle choices. The presence of children (1–3) suggests they may value family-friendly features such as safety, space, and fuel economy.

Cluster 3: This group appears less concerned with environmental factors and more individualistic. A wider income range ($50k–$150k) suggests diverse financial backgrounds, from young professionals to higher earners. They drive fewer miles (10–17k), indicating less reliance on personal vehicles—possibly using alternative transportation or remote work. Less formal education might correlate with a preference for cost-effective, functional, and possibly performance-oriented vehicles. Since they recycle the least, sustainability is likely not a key factor in their decision-making.

7. **Make a recommendation to GPA. Which segment(s) should they target? How can car manufacturers position the microvan to attract new customers?**

Based on our analysis, GPA should target Cluster 2 (Compact Prioritizers) for the microvan market. This segment shows the strongest interest in the microvan concept, making them the most promising audience for adoption.

1. The standardized residuals are highly positive for mvliking = 9 in Cluster 2, indicating that people in this group strongly favor the microvan idea compared to other clusters.
2. With incomes between $50k–$100k, Compact Prioritizers have the financial means to afford a new vehicle.
3. They are likely to consider practical, high-value, and family-friendly options over luxury or performance-driven cars.
4. They drive 15k–30k miles per year, meaning they need reliable, fuel-efficient vehicles that balance daily commutes and family trips.
5. Many have 1–3 children, making space, safety, and practicality top considerations—all key advantages of microvans.

To attract this segment, car manufacturers should position the microvan as:
A Smart, Compact Family Vehicle: Highlight spaciousness, fuel efficiency, and advanced safety features for family appeal. Offer flexible seating and storage solutions to accommodate both kids and work-life balance.
Affordable & Value-Driven: Competitive pricing with good financing options to appeal to middle-income buyers.
Practical for Everyday Life: Position microvans as ideal for both urban and suburban lifestyles—easy to park, fuel-efficient, and adaptable for work and family needs. Market it as a "smart alternative" to SUVs, offering more space with better fuel economy.