Grosse Pointe Associates and the 'Microvan' Group Project Report

Team 32

Shiven Ahuja (ska21), Joyce Gao(lg251), Benazir Khurshid (bk190), Zhiyang Ou (zo11), Jason Yang(zy145)

2/22/2022

1. Introduction

Through this project we tried to deep dive into the automotive market to see what factors affect consumers' buying decisions. Given the survey data collected by Grosse Pointe Associates (GPA), we explored the dataset, analysed the automobile attributes and survey responders' psychographic factors that lead us to segment the market and identify our target customer base.

2. Initial Model

Before building the initial model, we explored the dataset for better data understanding. First starting from checking that there is no null value, we then examined the correlation between variables. By making a correlation matrix visualisation (Figure 1), one could find that certain dependent variables have strong correlation with each other.

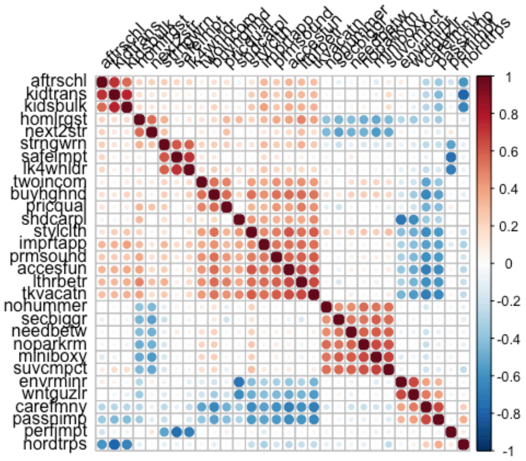

Figure 1 Correlation Matrix

After that we delved into the distribution of variables by composing the histograms and checking the summary data. Dependent variables have similar normal distribution trends

(Figure 2) while the independent variable *mvliking* has a different shape of distribution (Figure 3).

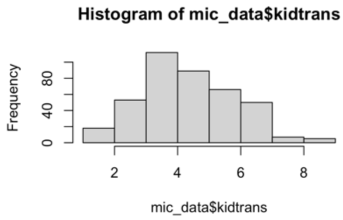

Figure 2

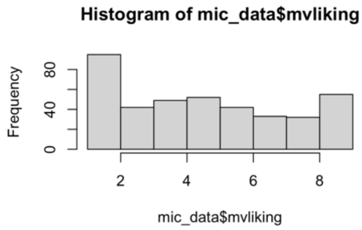

Figure 3

Next, we regressed the target variable against all the attribute variables. There are eight variables with p-value around 0.05 and among those eight variables, *lthrbetr* and *shdcarpl* show statistically significance as their p-value are lower than 0.05. As having so many variables and strong correlation within the predictors, we feel the need to do the factor analysis to reduce the redundancy of our attribute variables.

3. Factor Analysis

<u>Step 1: run Bartlett's Test of Sphericity and KMO-test</u>

```
Overall MSA =  0.92
MSA for each item =
kidtrans miniboxy lthrbetr secbiggr safeimpt buyhghnd pricqual prmsound perfimpt tkvacatn
    0.81     0.92     0.97     0.92     0.78     0.94     0.91     0.97     0.78     0.97
noparkrm homlrgst envrminr needbetw suvcmpct next2str carefmny shdcarpl imprtapp lk4whldr
    0.93     0.94     0.85     0.93     0.93     0.92     0.96     0.86     0.97     0.86
kidsbulk wntguzlr nordtrps stylclth strngwrn passnimp twoincom nohummer aftrschl accesfun
    0.90     0.93     0.88     0.96     0.90     0.96     0.94     0.93     0.91     0.97
```

Figure 4

Here we set the null hypothesis that the correlation matrix is an identity matrix; in other words, the null hypothesis suggests that the variables are unrelated and not ideal for factor analysis. After running Bartlett's Test with a p-value < 2.22e-16, we reject the null hypothesis. The result is statistically significant, and it suggests that the correlation matrix can be used to do the factor analysis.

The KMO-test is used to examine how the factors explain each other. KMO values closer to 1.0 are considered ideal, while values less than 0.5 are unacceptable. As we can see in Figure 4, most variables have KMO values greater than 0.9, and all the variables are considered good enough for factor analysis.

<u>Step 2: Determine the number of factors</u>

We use two approaches to determine how many factors explain as much variance in the data. First, the variables with eigenvalues less than 1 suggest that factors do not explain even as much variance in the data as an 'average' variable. As we can see on Figure 5, only the first five components have eigenvalues greater than 1. The same result can be discovered from the Scree-plot (Figure 5); the 'elbow' is visible in the curve; it suggests that 5 is the right number of clusters.

```
> pca$eig
       eigenvalue percentage of variance cumulative percentage of variance
comp 1  8.2754529             27.5848431                          27.58484
comp 2  4.9978929             16.6596429                          44.24449
comp 3  3.0882938             10.2943126                          54.53880
comp 4  2.7146499              9.0488330                          63.58763
comp 5  1.8002828              6.0009426                          69.58857
```

Figure 5

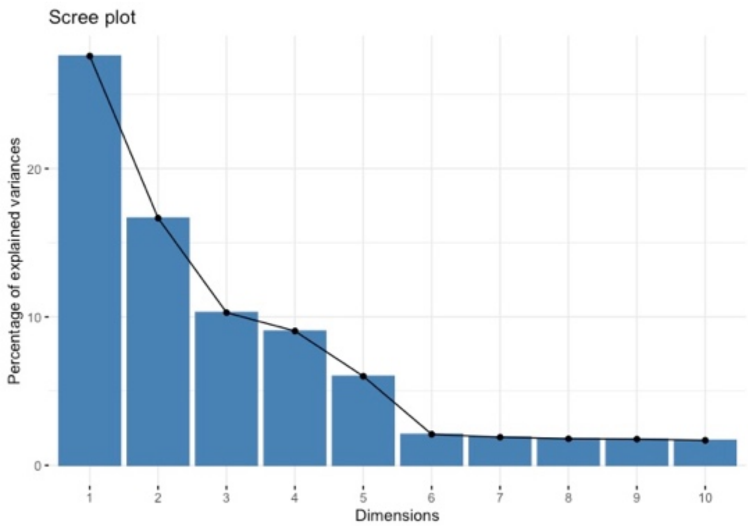

Figure 6

<u>Step 3: Extract the factor solution (varimax rotation)</u>

The varimax rotation enables us to extract and rotate factors to generate a solution. The outputs include the loading plots and the rotated component matric, which help reveal the meaning of the underlying factors. However, as we can see from the loading plot (Figure 7), the variables overlap, making it much harder to interpret based on it.

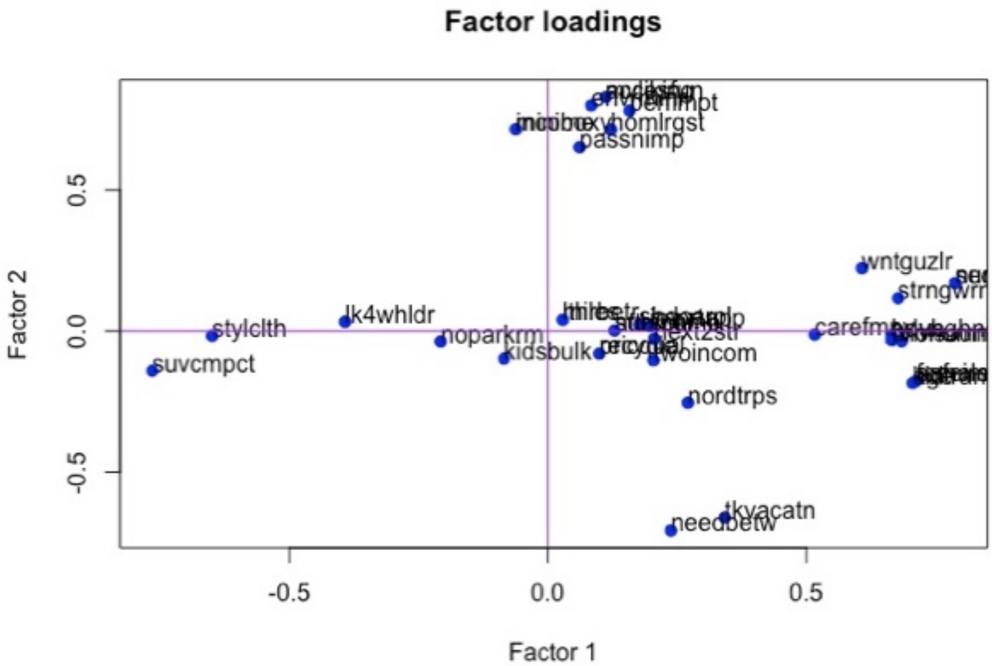

Figure 7

Therefore, we analyze the Loading tables(Figure 8) to identify the underlying factors and determine their meanings.

After considering the positive and negative correlations, we decided to name these five factors: quality matters, the volume of the car, family car, environmentally aware, and safety concerns.

1. **Premium Quality:** Consumers in this factor group care about the quality of cars. The positively correlated variables suggest that consumers prefer to spend more on their car purchases; they care about the cars' product quality and features – leather seats, car accessories & premium sound systems.
2. **Capacity:** Consumers in this group care about the no. of seats in the car. They want mid segment, affordable utility vehicle

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5
kidtrans	0.130		0.955		
miniboxy	0.112	0.829			
lthrbetr	0.707	-0.184	0.236	0.257	
secbiggr		0.715			
safeimpt					0.884
buyhghnd	0.789	0.170			
pricqual	0.715	-0.176		-0.106	
prmsound	0.666		0.159	0.257	
perfimpt	0.100				-0.843
tkvacatn	0.665		0.244	0.419	
noparkrm	0.159	0.781			
homlrgst	0.342	-0.661	0.144	0.287	
envrminr	-0.208			-0.829	
needbetw	0.122	0.714			
suvcmpct		0.800	0.191		
next2str	0.238	-0.707			
carefmny	-0.767	-0.141	-0.195	-0.277	
shdcarpl	0.206			0.818	
imprtapp	0.516		0.304	0.305	0.185
lk4whldr	0.178				0.810
kidsbulk	0.198		0.763		
wntguzlr	-0.393			-0.689	
nordtrps			-0.816		
stylclth	0.608	0.223	0.171	0.380	
strngwrn	0.271	-0.254			0.672
passnimp	-0.650		-0.375	-0.246	
twoincom	0.679	0.117			
nohummer		0.652			
aftrschl	0.205	-0.103	0.717	-0.100	0.177
accesfun	0.686		0.283	0.331	

Figure 8

3. **Affordability & Safety (Good Quality Material Used in car):** Consumers in this segment are families which use their cars a lot to drive around their children
4. **Environmentally Friendly**: This factor group consists of people who take environmental factors into account when purchasing cars. They prefer a car with environmentally friendly features and prefer carpools
5. **Durability & Safety (Driver Protection i.e., Airbags):** This factor group consists of people who put safety as their priority concern when purchasing cars. Variables such as ‘Auto safety is very important to me,’ and ‘Four-wheel drive is a very attractive option’ explain a lot about this group of consumers.

Step 4: Create and saved the factor scores

During the last step, we created names for these five factors (quality, volume, capacity, environment, and safety) based on the analysis above, and then saved the factor scores for further use.

4. Factor Model

We discovered the new regression model (with the saved factor score as the independent variable – Figure 11). Firstly, the factor model is statistically significant with a p-value less than 0.05. Secondly, the quality, the volume, and the safety variables are significant, with p-values less than 0.05. However, the capacity and the environment variables have slightly larger p-values (0.0589, 0.1181 respectively) which means they are statistically insignificant here. Moreover, the quality’s positive coefficient has suggested that as the quality of the car increases; consumer’s attitude toward the new concept of the car go up, and it seems to be the most impactful factor. Thirdly, we use AIC and BIC in model selections, and it suggests that the Factor model is a better model than the initial one since the factor model have lower AIC and BIC compared with the initial model (Figure 9). Lastly, the adjusted R^2 is slightly higher for Factor Model than the Initial Model (Figure 10), which means the variables in Factor Model can explain the model better than the variables in Initial Model.

```
> AIC(initial.model)
[1] 1803.114
> BIC(initial.model)
[1] 1930.841
> AIC(factor.model)
[1] 1775.654
> BIC(factor.model)
[1] 1803.595
```

Figure 9

```
lm(formula = mvliking ~ quality + volume + capacity + environment +
    safety, data = mic_data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.1124 -1.5471 -0.1443  1.5854  6.3664

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.8425     0.1102  43.925  < 2e-16 ***
quality       1.0277     0.1104   9.311  < 2e-16 ***
volume        0.9989     0.1104   9.049  < 2e-16 ***
capacity      0.2091     0.1104   1.894   0.0589 .
environment  -0.1729     0.1104  -1.566   0.1181
safety       -0.5545     0.1104  -5.023 7.72e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.205 on 394 degrees of freedom
Multiple R-squared:  0.3365,	Adjusted R-squared:  0.3281
F-statistic: 39.97 on 5 and 394 DF,  p-value: < 2.2e-16
```

Figure 11

```
lm(formula = mvliking ~ ., data = initial.model.data)

Residuals:
   Min     1Q Median     3Q    Max
-6.023 -1.605 -0.018  1.475  6.508

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.381266   2.958551   0.129   0.8975
kidtrans     0.241114   0.164793   1.463   0.1443
miniboxy     0.177881   0.129258   1.376   0.1696
lthrbetr     0.247630   0.121971   2.030   0.0430 *
secbiggr    -0.104796   0.105833  -0.990   0.3227
safeimpt    -0.018525   0.133630  -0.139   0.8898
buyhghnd     0.112578   0.116162   0.969   0.3331
pricqual     0.105322   0.104796   1.005   0.3155
prmsound     0.010118   0.108312   0.093   0.9256
perfimpt     0.232663   0.128198   1.815   0.0704 .
tkvacatn     0.166171   0.124671   1.333   0.1834
noparkrm     0.178143   0.115804   1.538   0.1248
homlrgst    -0.208684   0.122418  -1.705   0.0891 .
envrminr    -0.033245   0.122777  -0.271   0.7867
needbetw     0.128468   0.102636   1.252   0.2115
suvcmpct     0.215136   0.122643   1.754   0.0802 .
next2str     0.024294   0.106843   0.227   0.8203
carefmny    -0.243143   0.134373  -1.809   0.0712 .
shdcarpl    -0.286783   0.122413  -2.343   0.0197 *
imprtapp     0.059086   0.104214   0.567   0.5711
lk4whldr    -0.064119   0.126739  -0.506   0.6132
kidsbulk    -0.096959   0.122063  -0.794   0.4275
wntguzlr    -0.028943   0.115689  -0.250   0.8026
nordtrps     0.073056   0.127473   0.573   0.5669
stylclth     0.015757   0.113597   0.139   0.8898
strngwrn    -0.196806   0.113448  -1.735   0.0836 .
passnimp     0.161975   0.119056   1.360   0.1745
twoincom     0.170419   0.096469   1.767   0.0781 .
nohummer     0.009052   0.095697   0.095   0.9247
aftrschl    -0.025716   0.116551  -0.221   0.8255
accesfun    -0.003458   0.122112  -0.028   0.9774
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.215 on 369 degrees of freedom
Multiple R-squared:  0.3729,	Adjusted R-squared:  0.3219
F-statistic: 7.314 on 30 and 369 DF,  p-value: < 2.2e-16
```

Figure 10

5. Clustering Analysis

Based on the five factors we defined, we started a cluster analysis of the data to find our target group. Firstly, to determine the number of our clusters, we first applied hierarchical clustering using Ward's method to conduct preliminary clustering of our data. The result of clustering is shown in Figure 12.

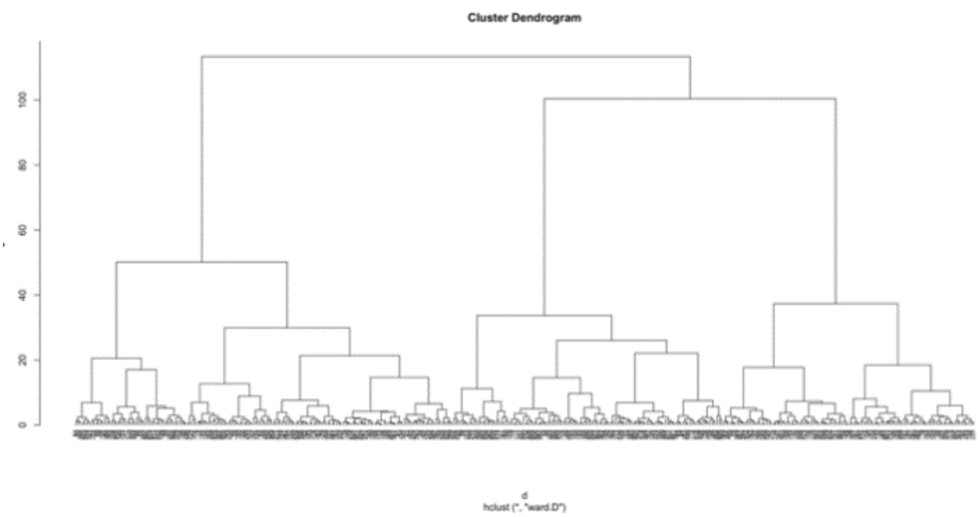

Figure 12

Based on the results, we set our number of clusters to 3. Then we used the k-means clustering method to cluster our data. The clustering results are shown in the following table. (Figure 13)

	premium_quality	Capacity_Seating	safe_family	environment	durability_safety
1	0.6065553	-1.11256505	0.1264856	0.2575779	0.02343056
2	-0.9582770	0.04618334	-0.3015591	-0.4757075	-0.09524731
3	0.6329049	1.15556015	0.2696647	0.3618421	0.10325988

Figure 13

To more clearly observe whether the K-means clustering method well separated the data into three groups, we used 3D scatter graph to visualize our results. The visualization results are as follows. (Figures 14 and 15)

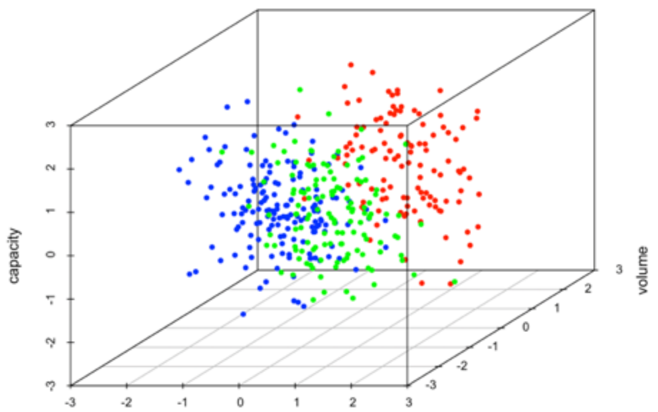

Figure 14

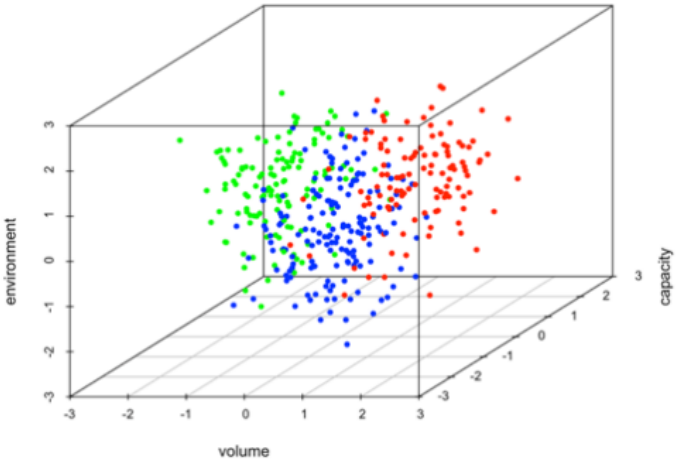

Figure 15

As can be seen from the results, although some data overlapped, we divided the data into three groups with differences overall. According to the results of the centre of each group shown by k-means clustering, we can well summarize the characteristics of these three groups. Cluster 1 cares about quality of life. They don't go for big cars. Instead, they prefer small cars like 2-seaters cars or racing cars. They don't care much about car capacity, durability, environmental protection and safety. Cluster 2, they are very concerned about economic expenditure. They usually don’t care about the capacity of the car. They don’t have strong sense of environment awareness. Cluster 3 has a high pursuit of life quality. They like larger vehicles the size of which is better between microvan and sedan. They need a vehicle that can accommodate more people like a group of family members safely. They relatively care more about environmental protection, and durability and safety of the car compared with the other two segments. Therefore, we define our three clusters as: Cluster 1 -- Quality Racing Enthusiasts, Cluster 2 -- Economical and Applicable, Cluster 3 -- Wealthy Large Families. (Table 1)

Cluster 1	Cluster 2	Cluster 3
Affluent Families	Middle Class Families	**Wealthy Large Families**
Inelastic Higher End Consumers	Less Features	Environmentally Friendly
Not environmentally conscious	**Economical Option**	Durable and Safe
Quality Racing Enthusiasts		High-Capacity Option

Table 1

6. Exploring the Clusters

To determine how the clusters vary on the concept liking, we tried three different methods.

First, we conducted the regression of mvliking on the cluster categorical variable. As it is categorical data, we are bound to lose the explanation of one cluster in the regression process. To this end, we changed the categories' order to carry out two regressions to the model. The result of the regression is as follows. (Figures 16 and 17)

```
Call:
lm(formula = mvliking ~ factor(segment.label), data = mic_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.6034 -1.8981  0.1019  2.1019  5.1019 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                            3.8981     0.1947  20.022   <2e-16 ***
factor(segment.label)QualityRacing     0.5035     0.2911   1.729   0.0845 .  
factor(segment.label)WealthyLargeFam   2.7054     0.2987   9.058   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.44 on 397 degrees of freedom
Multiple R-squared:  0.1816,    Adjusted R-squared:  0.1775 
F-statistic: 44.06 on 2 and 397 DF,  p-value: < 2.2e-16
```

Figure 16

```
Call:
lm(formula = mvliking ~ relevel(factor(segment.label), ref = "QualityRacing"),
    data = mic_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.6034 -1.8981  0.1019  2.1019  5.1019 

Coefficients:
                                                                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                                          4.4016     0.2165  20.333  < 2e-16 ***
relevel(factor(segment.label), ref = "QualityRacing")Economical     -0.5035     0.2911  -1.729   0.0845 .  
relevel(factor(segment.label), ref = "QualityRacing")WealthyLargeFam  2.2019     0.3133   7.028  9.2e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.44 on 397 degrees of freedom
Multiple R-squared:  0.1816,	Adjusted R-squared:  0.1775 
F-statistic: 44.06 on 2 and 397 DF,  p-value: < 2.2e-16
```

Figure 17

According to the regression results, only wealthy large families cluster has a significant positive impact on the attitude towards the new concept of car. Quality racing has a positive but insignificant impact on the attitude. Economical and applicable not only has a negative impact on our attitudes toward the new concept but also is not significant. This means that wealthy large families are very likely to be our target group.

To further verify our conclusion, we conducted t-test on the mean values of different clusters' attitudes to our new concept. Our hypothesis is that the average attitudes of the three clusters to the new concept is greater than the overall mean. Our results are shown in Figures 18, 19 and 20. It was proved that only the mean of wealthy large families passed the T-test. This validates what we learned in the first step.

```
	One Sample t-test

data:  wealthy.large.fam$mvliking
t = 8.3328, df = 115, p-value = 9.548e-14
alternative hypothesis: true mean is greater than 4.8425
95 percent confidence interval:
 6.253022      Inf
sample estimates:
mean of x 
 6.603448 
```

Figure 18

```
        One Sample t-test

data:  quality.racing$mvliking
t = -1.952, df = 126, p-value = 0.9734
alternative hypothesis: true mean is greater than 4.8425
95 percent confidence interval:
 4.027267      Inf
sample estimates:
mean of x
 4.401575
```

Figure 19

```
        One Sample t-test

data:  economical$mvliking
t = -4.7943, df = 156, p-value = 1
alternative hypothesis: true mean is greater than 4.8425
95 percent confidence interval:
 3.572141      Inf
sample estimates:
mean of x
 3.898089
```

Figure 20

Finally, we did cross tabulation and Chi-Squared analysis of segment label and MVliking to further verify our conclusion. The results of the analysis are in Figure 21 (See appendix for detailed results of analysis).

```
> crosstab
        segment.label
mvliking Economical QualityRacing WealthyLargeFam
       1         39            25               5
       2         15             8               3
       3         20            17               5
       4         26            17               6
       5         16            21              15
       6         15            10              17
       7         10            10              13
       8          5             7              20
       9         11            12              32
> # show overall chi-squared analysis
> summary(crosstab)
Call: xtabs(formula = ~mvliking + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
        Chisq = 83.12, df = 16, p-value = 4.527e-11
```

Figure 21

As can be seen from the results, wealthy large families obviously have more positive attitudes and an upward trend than the two clusters of Economical and quality racing. And, overall, it passed the Chi-Squared test. Then we can conclude that our target group is wealthy large families.

7. Demographics Analysis

Finally, we analysed the demographic characteristics of three different clusters to test whether our conclusion is in line with the reality. To accomplish this step, cross Tabulation and Chi-Squared analysis were performed on the demographic data and clusters.

First, for the convenience of analysis, continuous variables (age, income and miles) in demographics are transformed into categorical variables. We divided these three variables into three groups based on the minimum value, 1st quantile, 3rd quantile and maximum value. The grouping results are as follows. (Figure 22)

```
  age_group    income_group miles_group
 young :107    low    :103   low    :119
 middle:197    medium:198    medium:185
 old    : 96   high   : 99   high   : 96
```

Figure 22

Then we performed cross Tabulation and Chi-Squared analysis on cluster and each demographic variable respectively (Figures 23, 24, 25 and 26).

```
> crosstab_age
        segment.label
age_group Economical QualityRacing WealthyLargeFam
   young         103             1               3
   middle         53            72              72
   old             1            54              41
> crosstab_income <- xtabs(~income_group+segment.label, mic_data)
> crosstab_income
           segment.label
income_group Economical QualityRacing WealthyLargeFam
      low            97             0               6
      medium         59            66              73
      high            1            61              37
> crosstab_miles <- xtabs(~miles_group+segment.label, mic_data)
> crosstab_miles
          segment.label
miles_group Economical QualityRacing WealthyLargeFam
     low            88            28               3
     medium         66            73              46
     high            3            26              67
> crosstab_numkids <- xtabs(~numkids+segment.label, mic_data)
> crosstab_numkids
        segment.label
numkids Economical QualityRacing WealthyLargeFam
      0         71            33              12
      1         68            50              30
      2         14            33              37
      3          3             7              26
      4          1             4              11
```

Figure 2310

```
> crosstab_gender <- xtabs(~female+segment.label, mic_data)
> crosstab_gender
      segment.label
female Economical QualityRacing WealthyLargeFam
     0         80            57              47
     1         77            70              69
> crosstab_educ <- xtabs(~educ+segment.label, mic_data)
> crosstab_educ
    segment.label
educ Economical QualityRacing WealthyLargeFam
   1         43             1               0
   2         63            13              12
   3         51            66              53
   4          0            47              51
> crosstab_recycle <- xtabs(~recycle+segment.label, mic_data)
> crosstab_recycle
       segment.label
recycle Economical QualityRacing WealthyLargeFam
      1         15            16              13
      2         39            24              25
      3         51            38              35
      4         37            35              24
      5         15            14              19
```

Figure 24

```
> summary(crosstab_gender)
Call: xtabs(formula = ~female + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 3.0198, df = 2, p-value = 0.2209
> summary(crosstab_educ)
Call: xtabs(formula = ~educ + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 172.91, df = 6, p-value = 1.083e-34
> summary(crosstab_recycle)
Call: xtabs(formula = ~recycle + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 5.875, df = 8, p-value = 0.6613
```

Figure 25

```
> summary(crosstab_age)
Call: xtabs(formula = ~age_group + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 218.96, df = 4, p-value = 3.145e-46
> summary(crosstab_income)
Call: xtabs(formula = ~income_group + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 206.92, df = 4, p-value = 1.219e-43
> summary(crosstab_miles)
Call: xtabs(formula = ~miles_group + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 160.63, df = 4, p-value = 1.074e-33
> summary(crosstab_numkids)
Call: xtabs(formula = ~numkids + segment.label, data = mic_data)
Number of cases in table: 400
Number of factors: 2
Test for independence of all factors:
	Chisq = 100.38, df = 8, p-value = 3.575e-18
	Chi-squared approximation may be incorrect
```

Figure 26

According to the results of Chi-Squared, except gender and recycle, all other variables have passed the test. By observing other variables, we summarized the following group characteristics for different clusters:

- **Wealthy Large Families**: most of them are middle-aged, have moderate income, drive high mileage, have at least two children and have received three or four levels of education.

- **Quality Racing:** Most of them are middle-aged or old-aged, with high income, moderate driving mileage, and only one child or so, and have received 3 levels of education.
- **Economical and Applicable**: The vast majority of them are young, have low incomes, drive low miles, have no children and have received second-grade education.

These demographic characteristics fit our definition of three clusters which means that our segmentation makes sense. Based on the above conclusions, we believe that our target group is wealthy large family members.

8. Business Action/ Recommendation

Based on the different market segments we created, microvan is most likely to charm the educated large families. We highly recommend GPA target this segment that consists of middle aged individuals, having moderate to high income with 2+ kids. This segment values quality, capacity, environment, and safety. GPA should position the microvan as an automobile that combines elite style with high functionality. We suggest they use multiple channels for marketing such as social media, prints, OTT or broadcast advertisements or even onboard mom influencers. These marketing techniques should depict families enjoying a weekend road trip with their kids, highlighting the stylish features of the car, enough room for the kids, their toys and baggage while keeping in mind kids safety, and the long mileage run per gallon. As it is larger than a sedan but smaller than a minivan, the advertisements can position the microvan as the perfect compact van that can easily fit in the driveway.

Additionally, we would recommend conducting a similar focus group survey to position microvans as commercial vehicles such as for logistics supply, airport cabs, school vans etc.

However, having the same vehicle for commercial use might lead to a smaller intent to purchase by the above-mentioned segment as the microvan might lose its charm as a stylish family automobile but conducting a survey could help us understand if there is a potential market in the logistics industry for similar vehicles and the manufacturer could plan to launch a modified model.

9. Appendix

Cross Table Results of mvliking and segment.label

```
   Cell Contents
|-------------------------|
|                       N |
|              Expected N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  400
```

mic_data$mvliking	mic_data$segment.label Economical	QualityRacing	WealthyLargeFam	Row Total
1	39	25	5	69
	27.082	21.907	20.010	
	5.244	0.437	11.259	
	0.565	0.362	0.072	0.172
	0.248	0.197	0.043	
	0.098	0.062	0.013	
2	15	8	3	26
	10.205	8.255	7.540	
	2.253	0.008	2.734	
	0.577	0.308	0.115	0.065
	0.096	0.063	0.026	
	0.037	0.020	0.007	
3	20	17	5	42
	16.485	13.335	12.180	
	0.749	1.007	4.233	
	0.476	0.405	0.119	0.105
	0.127	0.134	0.043	
	0.050	0.043	0.013	
4	26	17	6	49
	19.233	15.557	14.210	
	2.381	0.134	4.743	
	0.531	0.347	0.122	0.122
	0.166	0.134	0.052	
	0.065	0.043	0.015	
5	16	21	15	52
	20.410	16.510	15.080	
	0.953	1.221	0.000	
	0.308	0.404	0.288	0.130
	0.102	0.165	0.129	
	0.040	0.052	0.037	
6	15	10	17	42
	16.485	13.335	12.180	
	0.134	0.834	1.907	
	0.357	0.238	0.405	0.105
	0.096	0.079	0.147	
	0.037	0.025	0.043	
7	10	10	13	33
	12.953	10.477	9.570	
	0.673	0.022	1.229	
	0.303	0.303	0.394	0.083
	0.064	0.079	0.112	
	0.025	0.025	0.033	
8	5	7	20	32
	12.560	10.160	9.280	
	4.550	0.983	12.383	
	0.156	0.219	0.625	0.080
	0.032	0.055	0.172	
	0.013	0.018	0.050	
9	11	12	32	55
	21.587	17.462	15.950	
	5.193	1.709	16.151	
	0.200	0.218	0.582	0.138
	0.070	0.094	0.276	
	0.028	0.030	0.080	
Column Total	157	127	116	400
	0.393	0.318	0.290	

Cross Table Results of Demographic Variables and Segments

Age Groups vs. Segments

```
   Cell Contents
|-------------------------|
|                       N |
|              Expected N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  400


                  | mic_data$segment.label
mic_data$age_group |     Economical |   QualityRacing | WealthyLargeFam |       Row Total |
------------------|----------------|----------------|----------------|----------------|
            young |            103 |              1 |              3 |            107 |
                  |         41.998 |         33.972 |         31.030 |                |
                  |         88.608 |         32.002 |         25.320 |                |
                  |          0.963 |          0.009 |          0.028 |          0.268 |
                  |          0.656 |          0.008 |          0.026 |                |
                  |          0.258 |          0.003 |          0.007 |                |
------------------|----------------|----------------|----------------|----------------|
           middle |             53 |             72 |             72 |            197 |
                  |         77.323 |         62.547 |         57.130 |                |
                  |          7.651 |          1.429 |          3.870 |                |
                  |          0.269 |          0.365 |          0.365 |          0.492 |
                  |          0.338 |          0.567 |          0.621 |                |
                  |          0.133 |          0.180 |          0.180 |                |
------------------|----------------|----------------|----------------|----------------|
              old |              1 |             54 |             41 |             96 |
                  |         37.680 |         30.480 |         27.840 |                |
                  |         35.707 |         18.149 |          6.221 |                |
                  |          0.010 |          0.562 |          0.427 |          0.240 |
                  |          0.006 |          0.425 |          0.353 |                |
                  |          0.003 |          0.135 |          0.102 |                |
------------------|----------------|----------------|----------------|----------------|
     Column Total |            157 |            127 |            116 |            400 |
                  |          0.393 |          0.318 |          0.290 |                |
------------------|----------------|----------------|----------------|----------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  218.9561     d.f. =  4     p =  3.144549e-46
```

Income Groups vs. Segments

```
                     | mic_data$segment.label
mic_data$income_group |     Economical |   QualityRacing | WealthyLargeFam |       Row Total |
---------------------|----------------|----------------|----------------|----------------|
                 low |             97 |              0 |              6 |            103 |
                     |         40.428 |         32.703 |         29.870 |                |
                     |         79.165 |         32.703 |         19.075 |                |
                     |          0.942 |          0.000 |          0.058 |          0.258 |
                     |          0.618 |          0.000 |          0.052 |                |
                     |          0.242 |          0.000 |          0.015 |                |
---------------------|----------------|----------------|----------------|----------------|
              medium |             59 |             66 |             73 |            198 |
                     |         77.715 |         62.865 |         57.420 |                |
                     |          4.507 |          0.156 |          4.227 |                |
                     |          0.298 |          0.333 |          0.369 |          0.495 |
                     |          0.376 |          0.520 |          0.629 |                |
                     |          0.147 |          0.165 |          0.182 |                |
---------------------|----------------|----------------|----------------|----------------|
                high |              1 |             61 |             37 |             99 |
                     |         38.858 |         31.433 |         28.710 |                |
                     |         36.883 |         27.813 |          2.394 |                |
                     |          0.010 |          0.616 |          0.374 |          0.247 |
                     |          0.006 |          0.480 |          0.319 |                |
                     |          0.003 |          0.152 |          0.092 |                |
---------------------|----------------|----------------|----------------|----------------|
        Column Total |            157 |            127 |            116 |            400 |
                     |          0.393 |          0.318 |          0.290 |                |
---------------------|----------------|----------------|----------------|----------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  206.9236     d.f. =  4     p =  1.219209e-43
```

Miles Groups vs. Segments

```
                     | mic_data$segment.label
mic_data$miles_group |      Economical |    QualityRacing | WealthyLargeFam |       Row Total |
---------------------|-----------------|-----------------|-----------------|-----------------|
                 low |              88 |              28 |               3 |             119 |
                     |          46.708 |          37.782 |          34.510 |                 |
                     |          36.505 |           2.533 |          28.771 |                 |
                     |           0.739 |           0.235 |           0.025 |           0.297 |
                     |           0.561 |           0.220 |           0.026 |                 |
                     |           0.220 |           0.070 |           0.007 |                 |
---------------------|-----------------|-----------------|-----------------|-----------------|
              medium |              66 |              73 |              46 |             185 |
                     |          72.612 |          58.737 |          53.650 |                 |
                     |           0.602 |           3.463 |           1.091 |                 |
                     |           0.357 |           0.395 |           0.249 |           0.463 |
                     |           0.420 |           0.575 |           0.397 |                 |
                     |           0.165 |           0.182 |           0.115 |                 |
---------------------|-----------------|-----------------|-----------------|-----------------|
                high |               3 |              26 |              67 |              96 |
                     |          37.680 |          30.480 |          27.840 |                 |
                     |          31.919 |           0.658 |          55.083 |                 |
                     |           0.031 |           0.271 |           0.698 |           0.240 |
                     |           0.019 |           0.205 |           0.578 |                 |
                     |           0.007 |           0.065 |           0.168 |                 |
---------------------|-----------------|-----------------|-----------------|-----------------|
        Column Total |             157 |             127 |             116 |             400 |
                     |           0.393 |           0.318 |           0.290 |                 |
---------------------|-----------------|-----------------|-----------------|-----------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  160.6253     d.f. =  4     p =  1.073564e-33
```

Number of kids vs. Segments

```
                 | mic_data$segment.label
mic_data$numkids |      Economical |    QualityRacing | WealthyLargeFam |       Row Total |
-----------------|-----------------|-----------------|-----------------|-----------------|
               0 |              71 |              33 |              12 |             116 |
                 |          45.530 |          36.830 |          33.640 |                 |
                 |          14.248 |           0.398 |          13.921 |                 |
                 |           0.612 |           0.284 |           0.103 |           0.290 |
                 |           0.452 |           0.260 |           0.103 |                 |
                 |           0.177 |           0.083 |           0.030 |                 |
-----------------|-----------------|-----------------|-----------------|-----------------|
               1 |              68 |              50 |              30 |             148 |
                 |          58.090 |          46.990 |          42.920 |                 |
                 |           1.691 |           0.193 |           3.889 |                 |
                 |           0.459 |           0.338 |           0.203 |           0.370 |
                 |           0.433 |           0.394 |           0.259 |                 |
                 |           0.170 |           0.125 |           0.075 |                 |
-----------------|-----------------|-----------------|-----------------|-----------------|
               2 |              14 |              33 |              37 |              84 |
                 |          32.970 |          26.670 |          24.360 |                 |
                 |          10.915 |           1.502 |           6.559 |                 |
                 |           0.167 |           0.393 |           0.440 |           0.210 |
                 |           0.089 |           0.260 |           0.319 |                 |
                 |           0.035 |           0.083 |           0.092 |                 |
-----------------|-----------------|-----------------|-----------------|-----------------|
               3 |               3 |               7 |              26 |              36 |
                 |          14.130 |          11.430 |          10.440 |                 |
                 |           8.767 |           1.717 |          23.191 |                 |
                 |           0.083 |           0.194 |           0.722 |           0.090 |
                 |           0.019 |           0.055 |           0.224 |                 |
                 |           0.007 |           0.018 |           0.065 |                 |
-----------------|-----------------|-----------------|-----------------|-----------------|
               4 |               1 |               4 |              11 |              16 |
                 |           6.280 |           5.080 |           4.640 |                 |
                 |           4.439 |           0.230 |           8.718 |                 |
                 |           0.062 |           0.250 |           0.688 |           0.040 |
                 |           0.006 |           0.031 |           0.095 |                 |
                 |           0.003 |           0.010 |           0.028 |                 |
-----------------|-----------------|-----------------|-----------------|-----------------|
    Column Total |             157 |             127 |             116 |             400 |
                 |           0.393 |           0.318 |           0.290 |                 |
-----------------|-----------------|-----------------|-----------------|-----------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  100.377     d.f. =  8     p =  3.575106e-18
```

Genders vs. Segments

```
                 | mic_data$segment.label
mic_data$female |      Economical |   QualityRacing | WealthyLargeFam |       Row Total |
----------------|-----------------|-----------------|-----------------|-----------------|
              0 |              80 |              57 |              47 |             184 |
                |          72.220 |          58.420 |          53.360 |                 |
                |           0.838 |           0.035 |           0.758 |                 |
                |           0.435 |           0.310 |           0.255 |           0.460 |
                |           0.510 |           0.449 |           0.405 |                 |
                |           0.200 |           0.142 |           0.117 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|
              1 |              77 |              70 |              69 |             216 |
                |          84.780 |          68.580 |          62.640 |                 |
                |           0.714 |           0.029 |           0.646 |                 |
                |           0.356 |           0.324 |           0.319 |           0.540 |
                |           0.490 |           0.551 |           0.595 |                 |
                |           0.193 |           0.175 |           0.172 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|
   Column Total |             157 |             127 |             116 |             400 |
                |           0.393 |           0.318 |           0.290 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  3.019774     d.f. =  2     p =  0.220935
```

Education Levels vs. Segments

```
              | mic_data$segment.label
mic_data$educ |      Economical |   QualityRacing | WealthyLargeFam |       Row Total |
--------------|-----------------|-----------------|-----------------|-----------------|
            1 |              43 |               1 |               0 |              44 |
              |          17.270 |          13.970 |          12.760 |                 |
              |          38.334 |          12.042 |          12.760 |                 |
              |           0.977 |           0.023 |           0.000 |           0.110 |
              |           0.274 |           0.008 |           0.000 |                 |
              |           0.107 |           0.003 |           0.000 |                 |
--------------|-----------------|-----------------|-----------------|-----------------|
            2 |              63 |              13 |              12 |              88 |
              |          34.540 |          27.940 |          25.520 |                 |
              |          23.450 |           7.989 |           7.163 |                 |
              |           0.716 |           0.148 |           0.136 |           0.220 |
              |           0.401 |           0.102 |           0.103 |                 |
              |           0.158 |           0.033 |           0.030 |                 |
--------------|-----------------|-----------------|-----------------|-----------------|
            3 |              51 |              66 |              53 |             170 |
              |          66.725 |          53.975 |          49.300 |                 |
              |           3.706 |           2.679 |           0.278 |                 |
              |           0.300 |           0.388 |           0.312 |           0.425 |
              |           0.325 |           0.520 |           0.457 |                 |
              |           0.128 |           0.165 |           0.133 |                 |
--------------|-----------------|-----------------|-----------------|-----------------|
            4 |               0 |              47 |              51 |              98 |
              |          38.465 |          31.115 |          28.420 |                 |
              |          38.465 |           8.110 |          17.940 |                 |
              |           0.000 |           0.480 |           0.520 |           0.245 |
              |           0.000 |           0.370 |           0.440 |                 |
              |           0.000 |           0.117 |           0.128 |                 |
--------------|-----------------|-----------------|-----------------|-----------------|
 Column Total |             157 |             127 |             116 |             400 |
              |           0.393 |           0.318 |           0.290 |                 |
--------------|-----------------|-----------------|-----------------|-----------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  172.9148     d.f. =  6     p =  1.083062e-34
```

Number of Recycle Reports vs. Segments

	mic_data$segment.label			
mic_data$recycle	Economical	QualityRacing	WealthyLargeFam	Row Total
1	15	16	13	44
	17.270	13.970	12.760	
	0.298	0.295	0.005	
	0.341	0.364	0.295	0.110
	0.096	0.126	0.112	
	0.037	0.040	0.033	
2	39	24	25	88
	34.540	27.940	25.520	
	0.576	0.556	0.011	
	0.443	0.273	0.284	0.220
	0.248	0.189	0.216	
	0.098	0.060	0.062	
3	51	38	35	124
	48.670	39.370	35.960	
	0.112	0.048	0.026	
	0.411	0.306	0.282	0.310
	0.325	0.299	0.302	
	0.128	0.095	0.087	
4	37	35	24	96
	37.680	30.480	27.840	
	0.012	0.670	0.530	
	0.385	0.365	0.250	0.240
	0.236	0.276	0.207	
	0.092	0.087	0.060	
5	15	14	19	48
	18.840	15.240	13.920	
	0.783	0.101	1.854	
	0.312	0.292	0.396	0.120
	0.096	0.110	0.164	
	0.037	0.035	0.048	
Column Total	157	127	116	400
	0.393	0.318	0.290	

Statistics for All Table Factors

Pearson's Chi-squared test

```
------------------------------------------------------------
Chi^2 =  5.874508     d.f. =  8     p =  0.6612865
```