# Table of Contents

## 1. Data Verification

We implemented an exploratory data analysis in order to investigate the basic data characteristics, along with visualization. The purpose of this analysis is to ensure that we have a clean dataset and eliminate any outliers if they exist.

Firstly, we ran a summary statistic to obtain mean, standard deviation as well as distribution kurtosis. Since the results from the survey are scaled at an ordinal level, it is advisable to treat them as both categorical and continuous variables. See *Appendix A* for statistical summary table.

Then, we investigated statistical distribution for each of the independent variables to explore data accuracy. The independent variables are data survey result, ranging from 1 to 9. Based on our qualitative assessment, no user has given the same ranking for all the survey questions, thus we can conclude there is no possibility of random answers. The following histograms demonstrate frequency distribution for each of the independent variables. As we can see, all observations are within the range of 1 to 9 and normal distributed. As a result, no outliers are present in the survey dataset.
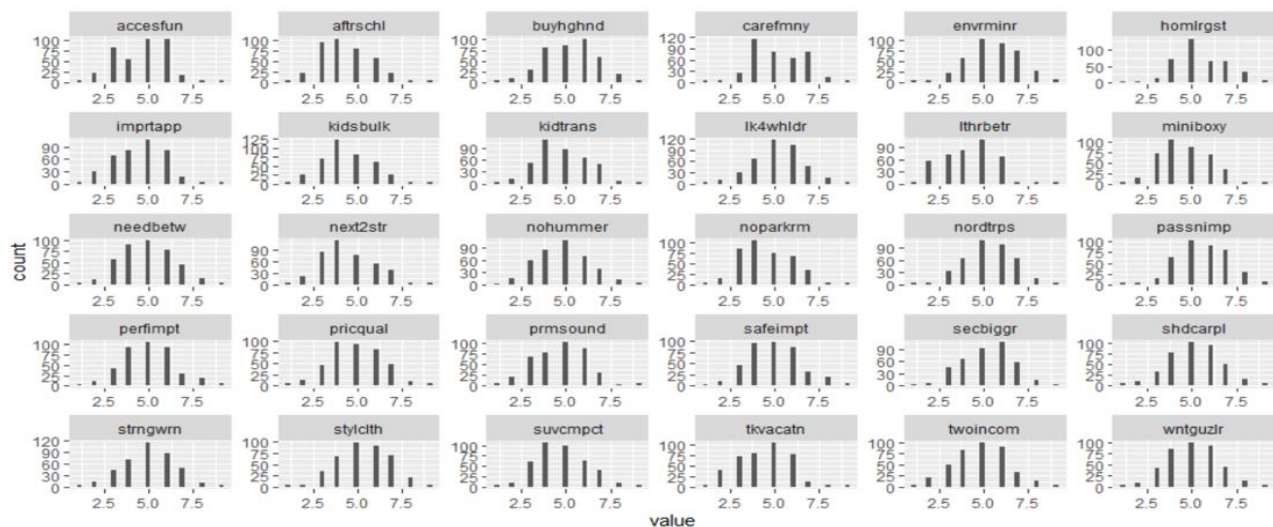


We have also implemented the following density plot to better demonstrate frequency distribution for each of the independent variable.
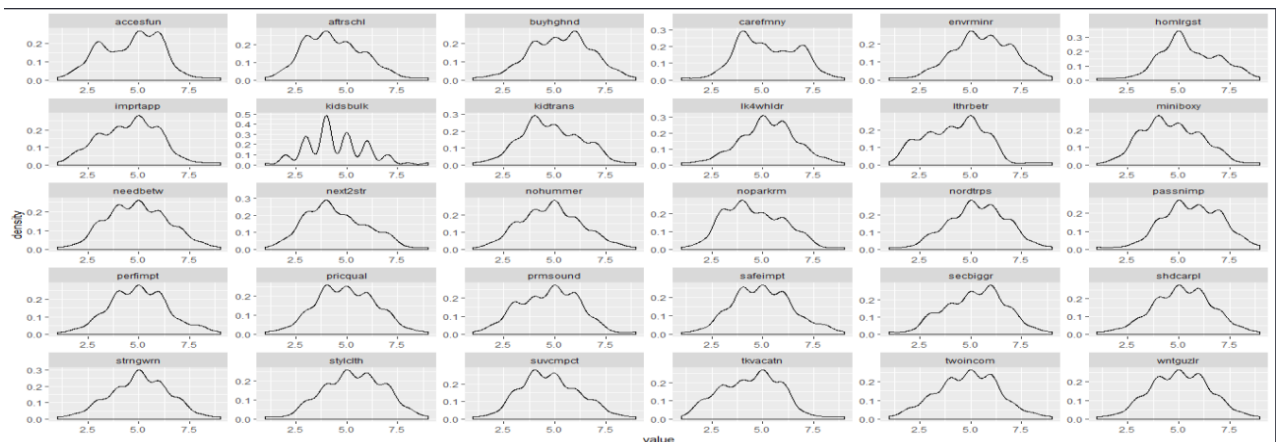
In addition, we also evaluated seven additional demographic and behavior answers for all users, and they all seem reasonable. Since the data range for this set of variables are unbounded, we do not need to investigate underlying distribution as they will all be significantly different. Based on our observation, all answers seem reasonable and are within the acceptable range.

Furthermore, we ran a correlation matrix to further explore underlying multicollinearity among independent variables. As we can see from the following matrix, multicollinearity is present, and the dataset needs further trimming for regression purpose. Overall, from the correlation matrix, we have 8 clusters of variables which are either highly positively correlated or negatively correlated.
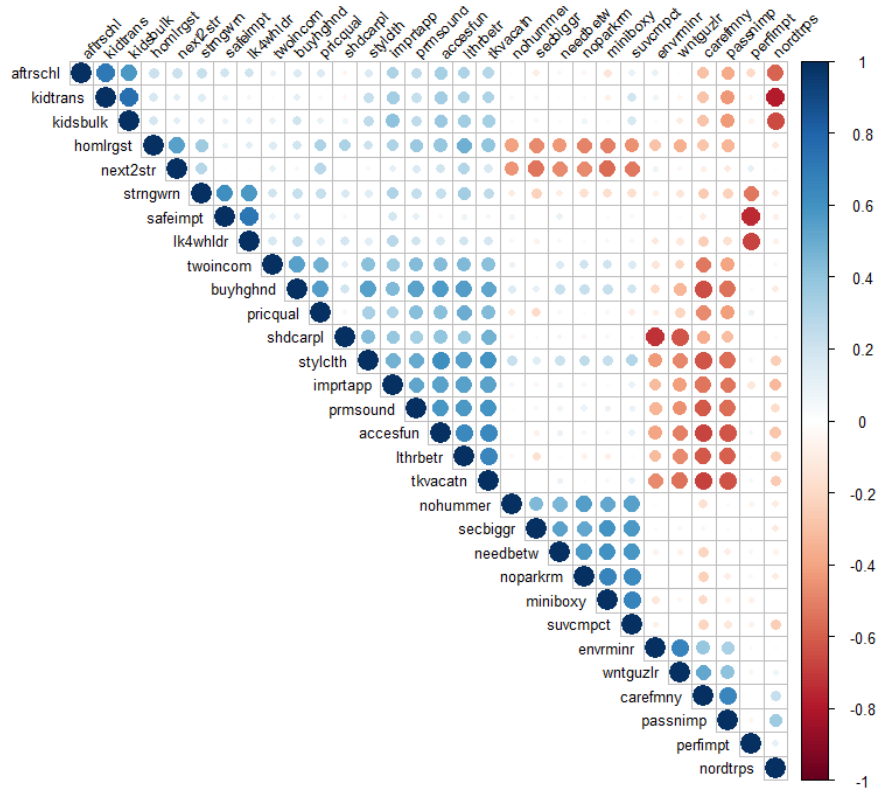


In conclusion, our dataset is absent of outliers and lack of involvement by the respondent.

## 2. Establishing Relationships

First, we wanted to determine the relationship between the 30 attribute (independent) variables and the target (dependent) variable **mvliking**. To do so, we regressed all the survey variables in a simple linear model. Please see *Appendix B* for detailed results.

The summary shows that most of the variables are not statistically significant and that the adjusted $R^2$ is quite low at 0.315. This means, that not all the variables are suited to explain the likeliness of a respondent willing to buy a microvan. To improve the model, we used both stepwise forward and backward regression to find the optimal combination of variables to include in the final model. This resulted in 9 leftover predictor variables which we used in a final model. Although all car-related, these significant variables are not well

suited for forming a pattern we would expect from the respondent's preferences. (See *Appendix C*) Although they include attributes like "our residence doesn't have a lot of parking room", the majority are attributes like "I like SUVs more than minivans" instead of "Current minivans are just too boxy and large". People who care less about money seem to be more inclined to buy a microvan.

The residuals show no systematic arrangement, they seem randomly distributed. There is also no recognizable spread and therefore no indication of heteroskedasticity. However, the adjusted $R^2$ could only be improved by 0.02 and therefore the model still has limited prediction power. Further analysis is necessary.
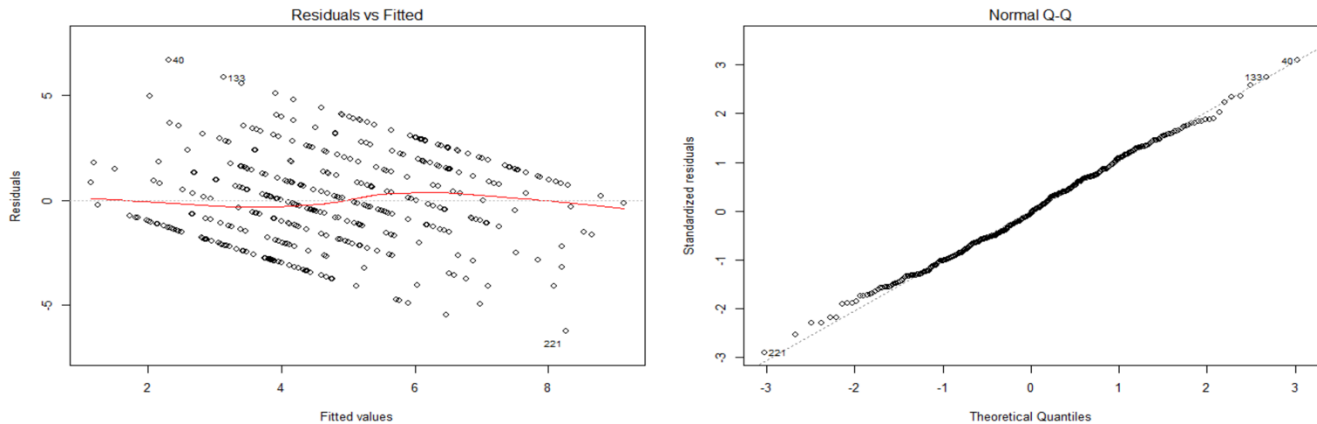


## 3. Data Reduction

Now we dive into the data set by first finding the correlation among our variables of interest. We notice 5 clusters of variables with relatively high positive correlations and 4 clusters with relatively high negative correlation



4
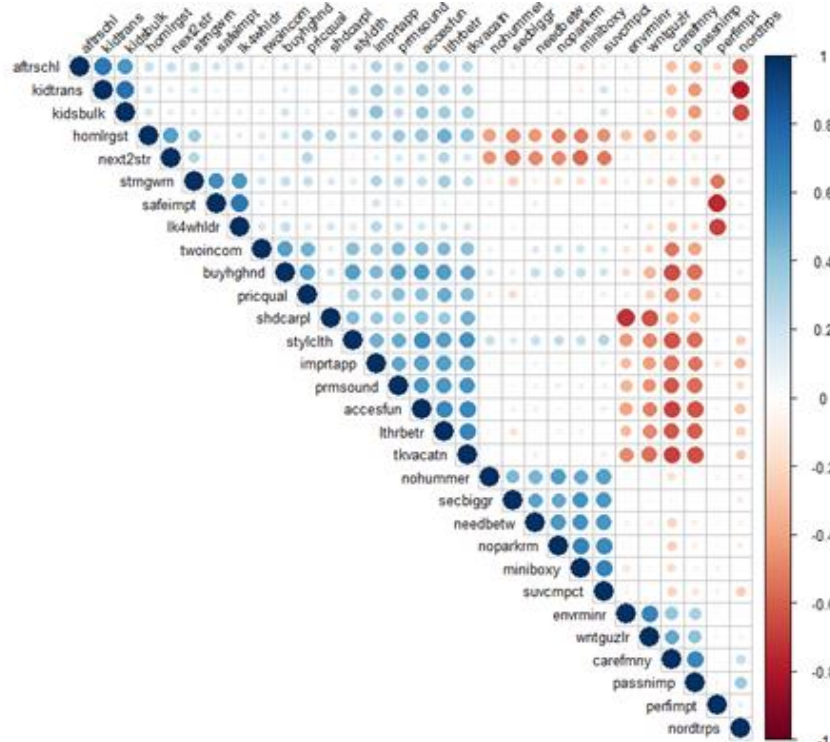
## 3.1 Finding the Eigenvalues

After finding the eigenvalues for all the variables (please refer to *Appendix D* for detailed output), we've noticed that the first 5 dimensions together account for 70 % of the variance in the data. To decide how many components, we need to adequately describe the data, we construct a scree plot as seen on the right side here.

According to the scree plot, we take only those eigenvalues that are greater than one; therefore, five components are retained. And in the plot below, we can see the cumulative variance explained by the five principle components.
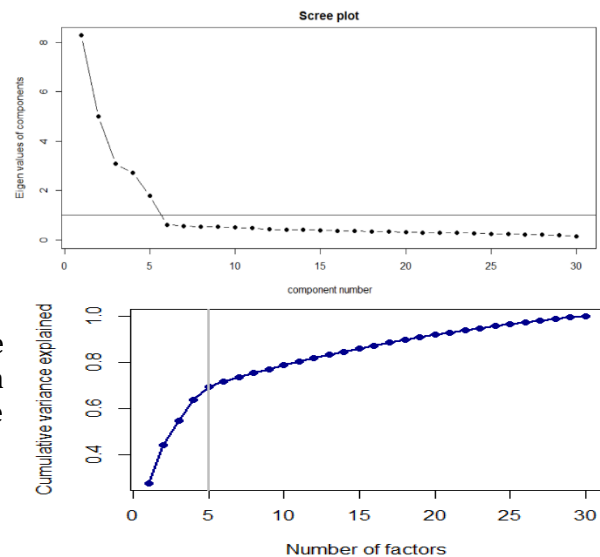


## 3.2 Executing PCA

To simplify the loadings matrix, we can also use the option cut = 0.35 to display only those component loadings that are larger than 0.35.  We conduct a *Varimax* rotation to help with the interpretation of the result. Please refer to *Appendix E* for the detailed result.

## 3.3 Interpreting the five principal components

In the output below, we can see that the communalities are between 0.51 and 0.89, which indicates that between 51% and 89% of the variation in each of the input variables has been retained within these five principle components. In our team's interpretation, we believe this is an enough coverage for this analysis.

```{r}
sort(PCA.survey$communality)
```

```
 nohummer   imprtapp   prmsound   needbetw   secbiggr   twoincom   stylclth   next2str   passnimp   pricqual   accesfun   noparkrm   lthrbetr   strngwrn   aftrschl
0.5074668  0.5411918  0.5828106  0.5935800  0.5936531  0.6089022  0.6357851  0.6430263  0.6560868  0.6733958  0.6888094  0.6889111  0.6889598  0.6928098  0.6970928
 homlrgst   tkvacatn   buyhghnd   wntguzlr   kidsbulk   suvcmpct   miniboxy   carefmny   nordtrps   lk4whldr   shdcarpl   envrminr   perfimpt   safeimpt   kidtrans
0.7021911  0.7051524  0.7077680  0.7107867  0.7167893  0.7202812  0.7381128  0.7442939  0.7670653  0.7725671  0.7873493  0.7898254  0.8060323  0.8296296  0.8862464
```

1) Frist Principal Component - **Premium Life**:

The attributes loading positively high on this component are "buyhghnd", "pricqual","twoincom", "prmsound","accesfun", "tkvacatn", "stylclth", "imprtapp" and negatively high are "carefmny", "passnimp".    This principle component **Premium    Life** consists     of variables that represent a tendency of pursuing high quality lifestyle, with accessories equipped high-end cars, stylish clothes, imported goods. Moreover, this component also indicates a strong tendency of taking vacations, being insensitive with money and a belief of price reflecting quality. In general, all these tendencies reflect

2) Second Principal Component - **Compactness:**

Secondly, "miniboxy", "suvcmpct", "noparkrm", "secbiggr", "needbetw", "nohummer" are loading positively high on this component while "next2str", "homlrgst" are loading negatively high. This component represents a group of variables that indicates certain dissatisfaction of the big size of a

minivan and the insufficiency of a sedan on the market currently, as well as having a smaller space in the household.

3) Third Principal Component - **Child Friendly:**

The attributes that load positively high on this component are "kidtrans", "kidsbulk", "aftrschl" while "nordtrps" load negatively high. This can be explained by that in this component, it contains attributes that showcase a high tendency of family and child -oriented decisions and behavior, such as high frequency of family trips, after-school activities, and bringing kids' toys and transporting kids, friends and family in the car.

4) Forth Principal Component - **Safety:**

The fourth factor is highly correlated with "safeimpt", "lk4whldr", "strngwrn" and "perfimpt". These independent variables reflect consumers' concern over auto safety as well as warranty protection. As a result, we name this principle component "Safety".

5) Fifth Principal Component – **Sustainability:**

The last factor demonstrates high correlation with "shdcarpl", and negative correlation with "wntguzlr" and "envrminr". This group of customers value environmentally friendly automobiles and believe that carpool or public transportation is essential. In addition, they will not purchase automobiles that are gas guzzlers due to $CO_2$ emission. Their main belief reflected in these variables is "Sustainability".

In the graph on the right, we can clearly see what we discuss earlier with the first and second principal component. Some variables that indicates pursuit of a premium life are obviously highly correlated with **Premium Life** while having no significant correlation with **Compactness** and vice versa. This graph indicates the distinct differences between these two principal components.
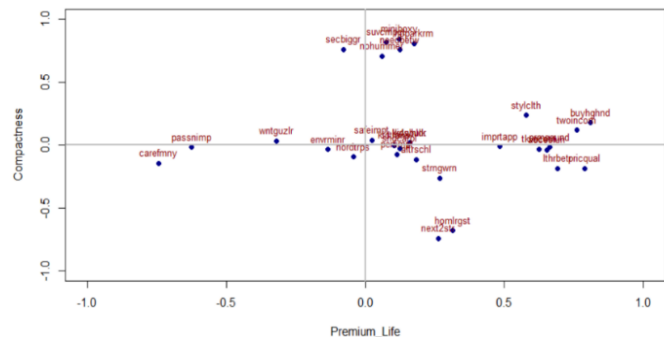


## 4. Explanation: Regression with PCA components

A regression model helps in understanding the underlying relationship between the variables. In step 2 we ran a regression analysis on the 30 variables upon our desired outcome "mvliking". The output of the model and subsequent pruning of the model resulted in a refined model consisting of only 9 variables deeming to have a profound impact on the outcome variable. With the Principal Component Analysis carried out on the 30 variables in step 3 we performed a regression analysis on the factor scores (components) on the outcome variable "mvliking". (Refer to *Appendix F* for complete output of regression)

```
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.84250    0.11036  43.878  < 2e-16 ***
premium_Life        1.04456    0.11050   9.453  < 2e-16 ***
compactness         0.99024    0.11050   8.961  < 2e-16 ***
Child_Friendliness  0.21584    0.11050   1.953   0.0515 .
sustainability     -0.09498    0.11050  -0.860   0.3906
safety             -0.54547    0.11050  -4.936 1.18e-06 ***
```

From the above output of the regression model, we can see that 3 (Premium Life, Compactness and Safety) out of the 5 factors are significant (statistically) at the highest level while one factor (Child Friendliness) is quite significant and the remaining factor (Sustainability) is not significant.

Determining relationship, keeping all other factors as constant, a unit increase in Premium Life associates a positive effect on the "mvliking" implying that customers tend to focus on the Premium aspects of the microvan. This is in line with our previous regression results as we saw that "carefmny" was highly significant. In a similar fashion, unit increases in Compactness and Child Friendliness is associated with a positive effect on the microvan likeliness of customers. On the other hand, the factors Sustainability and Safety are on the opposite track as customers likeliness towards microvan is negatively related. The intention to run a regression on the factor scores is to depict the customer's traits (~ variables compacted into as components) on the likelihood of buying a microvan. The results prove to be a foundation for the segmentation analysis which follows below.

## 5. Segmentation

Leveraging the 5 components from the result of the above factor analysis, we were able to segment the 400 survey respondents into 3 clusters, achieving an 800-fold reduction in dimensionality. The clusters are not wildly dissimilar, but we can notice some key differences amongst them. Below is a graphical representation of the 3 clusters across two dimensions. The y-axis dim2 represents the component compactness and the x-axis dim1 represent the component Price and quality. We can see that although the clusters are not wildly dissimilar, they appear to be no overlaps amongst them hence this is indeed a valid segmentation of the survey respondents on these two dimensions.
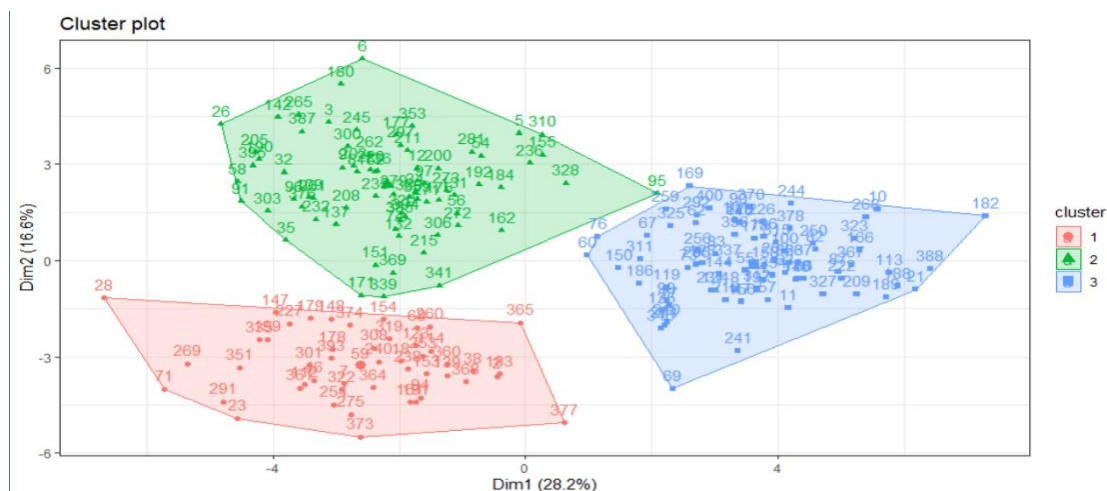
Cluster 1 which we describe as '**Microvan Enthusiasts'** in comparison to other clusters, 'Microvan Enthusiasts' evaluates compactness at a value that is 1.17 standard deviations higher than the original average which means that people in this segment have a strong desire for a microvan that has a size between a sedan and a minivan. They also have Premium Life at a value that's 0.58 standard deviations above the average ranking across all Survey respondents which indicate that they care strongly for high-end products and a good quality life. They are less sensitive to price, care more about quality and children friendliness in comparison to other respondents.

Respondents in cluster 2 which we call **'Hummer Lovers'** share almost the same preferences as the respondents of 'Microvan Enthusiasts'. The only difference is that contrary to 'Microvan Enthusiasts', 'Hummer Lovers' care a lot less for a smaller sized car. In comparison to other clusters, 'Hummer Lovers' evaluate Compactness at a value that is 1.09 standard deviations higher than average and Premium Life at a value that's 0.58 standard deviations above the average ranking across all survey respondents.

Cluster 3 which we call the **'Bargain Customers'** are very sensitive to price and would care less if the car is child-friendly or not. What these respondents want is a cheap car, big or small they could not be bothered, and they do not value sustainability as a key input quality for the proposed microvan. They evaluate Premium Life at value that's 0.9 standard deviations lower than the average which indicates that they don't really have a strong pursuit for high-end or stylish goods as long as its cheap enough. Moreover, the sustainability is at a value that's 0.55 standard deviations below the average across all survey respondents which indicates that they don't care much for the impact that automobiles have on the environment. Below is an heatmap summarizing the wants of the different clusters briefly.
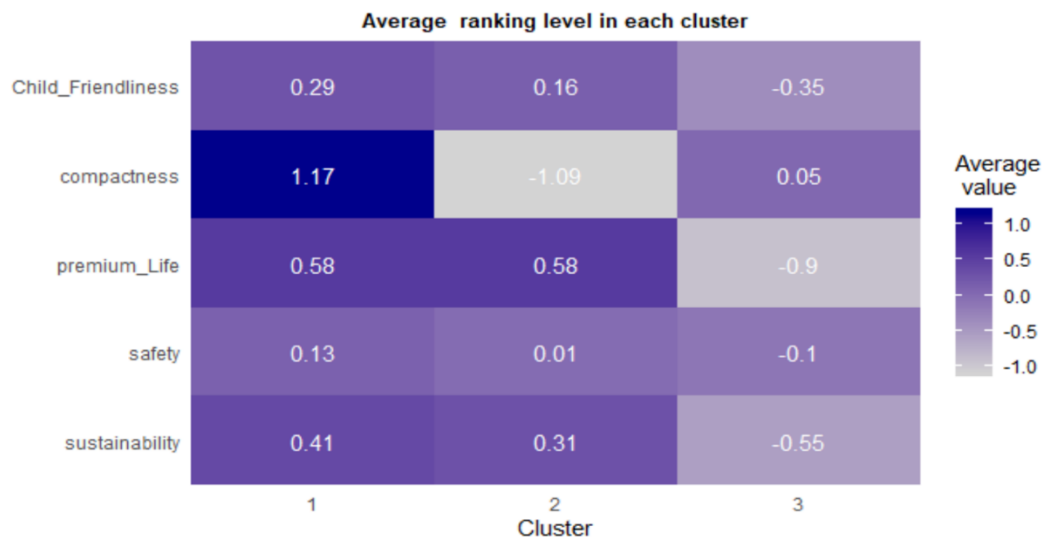


## 6. Reality Check: Relating clusters to demographicsOBJ OBJ OBJ OBJ

So far, we have confirmed the wants of our esteemed respondents but, does their wants align with their needs? To answer this question, we conducted a reality check by checking the correlations between the 5 components and the demographic data we have initially collected. Below is a brief summary of our demography.

```
        age           income          miles          numkids
Min.   :19.00  Min.   : 15.00  Min.   : 7.00  Min.   :0.00
1st Qu.:34.00  1st Qu.: 36.00  1st Qu.:15.00  1st Qu.:0.00
Median :40.00  Median : 62.50  Median :18.00  Median :1.00
Mean   :40.05  Mean   : 71.28  Mean   :18.04  Mean   :1.22
3rd Qu.:46.00  3rd Qu.: 96.00  3rd Qu.:21.00  3rd Qu.:2.00
Max.   :60.00  Max.   :273.00  Max.   :32.00  Max.   :4.00
       female          educ         recycle
Min.   :0.00   Min.   :1.000  Min.   :1.00
1st Qu.:0.00   1st Qu.:2.000  1st Qu.:2.00
Median :1.00   Median :3.000  Median :3.00
Mean   :0.54   Mean   :2.805  Mean   :3.04
3rd Qu.:1.00   3rd Qu.:3.000  3rd Qu.:4.00
Max.   :1.00   Max.   :4.000  Max.   :5.00
```

Moving forward, we ran a correlation analysis which shows us that all of our demographic attributes were positively correlated to our 5 components except for the psychographic attribute(recycle) which is negatively correlated to the components 'Premium Life' and 'Child Friendliness', Find below the correlation matrix. The relationship between the demographic attributes and the 30 survey variables is almost the same as the one above, we notice a lot of positive correlations and just few negative correlations amongst, the correlation matrix is in appendix G.
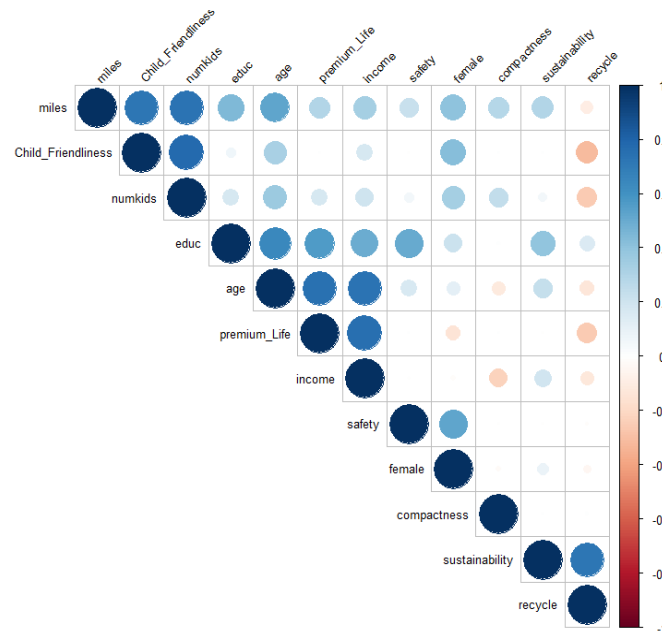


In trying to understand the needs of our demography we conducted a summary analysis. Our findings here are that for **Microvan Enthusiasts** which include respondents who drive an average of 22 thousand miles per annual and are of average age 44, have an average income 82.5, have mostly two kids and have mostly an undergraduate educational level. Based on what we know from research and experience about the needs of these demography, we can conclude that it aligns with our findings as regard their wants and again that the respondents in this cluster are indeed less sensitive to price due to their high incomes and do care for a small car because this may likely be a second car they are purchasing.

For **Hummer Lovers**, which include respondents who drive an average of 18 thousand miles per annual and are of average age 46, have an average income 104.5, have mostly 1 child and have mostly an undergraduate educational level. Based on these statistics, we can conclude that their needs align with our findings as regard their wants and again that the respondents in this cluster are indeed less sensitive to price due to their high incomes and do care for big cars due to the size of their family.

For **Bargain Customers,** which include respondents who drive an average of 5 thousand miles per annual and are of average age 32, have an average income 36, have mostly 1 child, mostly male and have mostly a college level educational degree. Based on these statistics, we can conclude that  their needs align with our findings as regard their wants and again that the respondents in this cluster are probably some men who just graduated from college with entry level positions at work  as a result they are very price sensitive due to  their low incomes and do not care about the size of the cars due to their lack of children. Find below the summary statistics on this demography aggregated by their cluster membership.

| Cluster | Mean age | Mean income | Mean miles | Mode num kid | Mode gender | Mode education | Mode recycle |
|---------|----------|-------------|------------|--------------|-------------|----------------|--------------|
| 1 | 44.2 | 82.5 | 22.3 | 2.0 | 1 | 3 | 3 |
| 2 | 46.0 | 104.0 | 18.4 | 1.0 | 1 | 3 | 3 |
| 3 | 32.1 | 36.2 | 14.7 | 0.0 | 0 | 2 | 3 |

# 7. Conclusion

We could confirm that there is a viability for the microvan concept based on the survey we had conducted on the US market, along with our detailed analyses. Our analyses yielded three distinct segments and one of the segments shows very high potential in pursuing microvans.

The company's target customers would be cluster one, "Microvan enthusiasts" and this group of customers should potentially generate most revenue. This cluster accounts for 29% of the total sample and we believe that they would be able to bring cashflow into the company's operations. As a result, the management should customize microvans in a way that matches with the preferences of this group of customers. Microvan enthusiasts value compactness, which is relatively a new concept of automobiles. Microvans' size should be designed in between a sedan and a minivan. In addition, microvan enthusiasts are less price sensitive and appreciate excellent quality, so we also recommend implementation of several premium features, such as leather seats and high-end entertainment system since this group of customers on average tend to drive the longest distances. Furthermore, engine should not only be powerful but also economical so we should explore the possibility of a hybrid system.

# 8. Appendix

## A.  Summary table

| | vars | n | mean | sd | median | trimmed | mad | min | max | range | skew | kurtosis | se |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| kidtrans | 1 | 400 | 4.83 | 1.52 | 5 | 4.80 | 1.48 | 1 | 9 | 8 | 0.18 | -0.16 | 0.08 |
| miniboxy | 2 | 400 | 4.66 | 1.48 | 5 | 4.60 | 1.48 | 1 | 9 | 8 | 0.32 | -0.10 | 0.07 |
| lthrbetr | 3 | 400 | 4.25 | 1.52 | 4 | 4.24 | 1.48 | 1 | 9 | 8 | 0.23 | 0.07 | 0.08 |
| secbiggr | 4 | 400 | 5.23 | 1.51 | 5 | 5.26 | 1.48 | 1 | 9 | 8 | -0.18 | -0.13 | 0.08 |
| safeimpt | 5 | 400 | 5.00 | 1.51 | 5 | 4.96 | 1.48 | 1 | 9 | 8 | 0.16 | -0.05 | 0.08 |
| buyhghnd | 6 | 400 | 5.28 | 1.54 | 5 | 5.31 | 1.48 | 1 | 9 | 8 | -0.17 | -0.13 | 0.08 |
| pricqual | 7 | 400 | 4.96 | 1.52 | 5 | 4.95 | 1.48 | 1 | 9 | 8 | 0.06 | -0.12 | 0.08 |
| prmsound | 8 | 400 | 4.72 | 1.51 | 5 | 4.70 | 1.48 | 1 | 9 | 8 | 0.04 | -0.14 | 0.08 |
| perfimpt | 9 | 400 | 5.00 | 1.49 | 5 | 4.97 | 1.48 | 1 | 9 | 8 | 0.12 | 0.09 | 0.07 |
| tkvacatn | 10 | 400 | 4.44 | 1.53 | 5 | 4.45 | 1.48 | 1 | 9 | 8 | 0.16 | -0.06 | 0.08 |
| noparkrm | 11 | 400 | 4.58 | 1.52 | 4 | 4.50 | 1.48 | 1 | 9 | 8 | 0.34 | -0.17 | 0.08 |
| homlrgst | 12 | 400 | 5.49 | 1.50 | 5 | 5.46 | 1.48 | 1 | 9 | 8 | -0.03 | 0.15 | 0.08 |
| envrminr | 13 | 400 | 5.53 | 1.52 | 6 | 5.56 | 1.48 | 1 | 9 | 8 | -0.24 | 0.09 | 0.08 |
| needbetw | 14 | 400 | 4.94 | 1.55 | 5 | 4.91 | 1.48 | 1 | 9 | 8 | 0.12 | -0.20 | 0.08 |
| suvcmpct | 15 | 400 | 4.78 | 1.51 | 5 | 4.73 | 1.48 | 1 | 9 | 8 | 0.26 | -0.03 | 0.08 |
| next2str | 16 | 400 | 4.49 | 1.55 | 4 | 4.42 | 1.48 | 1 | 9 | 8 | 0.41 | -0.12 | 0.08 |
| carefmny | 17 | 400 | 5.22 | 1.53 | 5 | 5.24 | 1.48 | 1 | 9 | 8 | 0.02 | -0.35 | 0.08 |
| shdcarpl | 18 | 400 | 5.17 | 1.50 | 5 | 5.20 | 1.48 | 1 | 9 | 8 | -0.13 | 0.05 | 0.08 |
| imprtapp | 19 | 400 | 4.56 | 1.50 | 5 | 4.57 | 1.48 | 1 | 9 | 8 | 0.10 | 0.00 | 0.08 |
| lk4whldr | 20 | 400 | 5.24 | 1.46 | 5 | 5.27 | 1.48 | 1 | 9 | 8 | -0.17 | 0.30 | 0.07 |
| kidsbulk | 21 | 400 | 4.48 | 1.48 | 4 | 4.42 | 1.48 | 1 | 9 | 8 | 0.41 | 0.17 | 0.07 |
| wntguzlr | 22 | 400 | 5.06 | 1.51 | 5 | 5.06 | 1.48 | 1 | 9 | 8 | -0.02 | -0.06 | 0.08 |
| nordtrps | 23 | 400 | 5.32 | 1.49 | 5 | 5.36 | 1.48 | 1 | 9 | 8 | -0.19 | 0.08 | 0.07 |
| stylclth | 24 | 400 | 5.34 | 1.53 | 5 | 5.38 | 1.48 | 1 | 9 | 8 | -0.18 | -0.15 | 0.08 |
| strngwrn | 25 | 400 | 5.04 | 1.50 | 5 | 5.06 | 1.48 | 1 | 9 | 8 | -0.09 | 0.00 | 0.08 |
| passnimp | 26 | 400 | 5.62 | 1.51 | 6 | 5.63 | 1.48 | 1 | 9 | 8 | -0.26 | 0.19 | 0.08 |
| twoincom | 27 | 400 | 4.87 | 1.56 | 5 | 4.87 | 1.48 | 1 | 9 | 8 | 0.01 | -0.13 | 0.08 |
| nohummer | 28 | 400 | 4.85 | 1.54 | 5 | 4.82 | 1.48 | 1 | 9 | 8 | 0.17 | -0.15 | 0.08 |
| aftrschl | 29 | 400 | 4.38 | 1.51 | 4 | 4.31 | 1.48 | 1 | 9 | 8 | 0.46 | 0.11 | 0.08 |
| accesfun | 30 | 400 | 4.64 | 1.52 | 5 | 4.65 | 1.48 | 1 | 9 | 8 | 0.01 | -0.24 | 0.08 |

B. **Regression Output – All Predictors**

```
Residuals:
    Min      1Q  Median      3Q     Max
-5.9061 -1.5569 -0.0689  1.4779  5.9608

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.296759   3.797149  -0.078  0.93775
kidtrans     0.311711   0.212521   1.467  0.14330
miniboxy     0.160529   0.169829   0.945  0.34516
lthrbetr     0.322174   0.166306   1.937  0.05348 .
secbiggr    -0.059844   0.130790  -0.458  0.64754
safeimpt    -0.076560   0.157681  -0.486  0.62758
buyhghnd     0.087437   0.163200   0.536  0.59244
pricqual     0.102529   0.126937   0.808  0.41977
prmsound    -0.004841   0.138582  -0.035  0.97215
perfimpt     0.219916   0.148955   1.476  0.14069
tkvacatn     0.200439   0.169171   1.185  0.23685
noparkrm     0.346774   0.150081   2.311  0.02141 *
homlrgst    -0.035620   0.157742  -0.226  0.82147
envrminr     0.045832   0.156730   0.292  0.77013
needbetw     0.136256   0.125564   1.085  0.27856
suvcmpct     0.224927   0.159667   1.409  0.15976
next2str     0.029335   0.142162   0.206  0.83663
carefmny    -0.323277   0.184834  -1.749  0.08112 .
shdcarpl    -0.405195   0.152174  -2.663  0.00809 **
imprtapp     0.061781   0.131743   0.469  0.63938
lk4whldr     0.013471   0.151045   0.089  0.92898
kidsbulk    -0.055417   0.155412  -0.357  0.72161
wntguzlr    -0.098407   0.149472  -0.658  0.51072
nordtrps     0.095156   0.154168   0.617  0.53747
stylclth    -0.100005   0.149379  -0.669  0.50361
strngwrn    -0.296375   0.131993  -2.245  0.02534 *
passnimp     0.180349   0.149265   1.208  0.22773
twoincom     0.271248   0.119609   2.268  0.02392 *
nohummer     0.003824   0.117813   0.032  0.97412
aftrschl    -0.170563   0.149524  -1.141  0.25473
accesfun    -0.047095   0.174411  -0.270  0.78729
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.226 on 369 degrees of freedom
Multiple R-squared:  0.3664,    Adjusted R-squared:  0.3149
F-statistic: 7.114 on 30 and 369 DF,  p-value: < 2.2e-16
```

## C.   Regression Output – Optimum Predictors

```
Residuals:
   Min     1Q Median     3Q     Max
-6.258 -1.474 -0.112  1.507  6.696

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.98611    1.70123   2.343 0.019628 *
noparkrm     0.22579    0.10382   2.175 0.030246 *
carefmny    -0.38425    0.10667  -3.602 0.000356 ***
perfimpt     0.28086    0.09204   3.052 0.002431 **
suvcmpct     0.29440    0.10056   2.928 0.003617 **
shdcarpl    -0.25000    0.08235  -3.036 0.002561 **
twoincom     0.22706    0.08637   2.629 0.008904 **
lthrbetr     0.34499    0.10529   3.277 0.001145 **
strngwrn    -0.19852    0.10061  -1.973 0.049199 *
homlrgst    -0.22922    0.10718  -2.139 0.033086 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.194 on 390 degrees of freedom
Multiple R-squared:  0.3496,     Adjusted R-squared:  0.3346
F-statistic: 23.29 on 9 and 390 DF,  p-value: < 2.2e-16
```

## D. Eigenvalue & Cumulative Variance

```
 [1] 8.2754529 4.9978929 3.0882938 2.7146499 1.8002828 0.6262671 0.5687380
 [8] 0.5365149 0.5292885 0.5050908 0.4860666 0.4308727 0.4239333 0.4179672
[15] 0.4031693 0.3801746 0.3650650 0.3474459 0.3419102 0.3187956 0.3049603
[22] 0.2929897 0.2904053 0.2813665 0.2493451 0.2372715 0.2341725 0.2125542
[29] 0.1965920 0.1424709
```

```
 [1] 0.2758484 0.4424449 0.5453880 0.6358763 0.6958857 0.7167613 0.7357192
 [8] 0.7536031 0.7712460 0.7880824 0.8042846 0.8186470 0.8327781 0.8467104
[15] 0.8601494 0.8728218 0.8849907 0.8965722 0.9079692 0.9185957 0.9287611
[22] 0.9385274 0.9482076 0.9575865 0.9658980 0.9738070 0.9816128 0.9886979
[29] 0.9952510 1.0000000
```

## E. PCA Output

| item | RC1 | RC2 | RC3 | RC5 | RC4 | h2 | u2 | com |
|---|---|---|---|---|---|---|---|---|---|
| | <S3: AsIs> | <S3: AsIs> | <S3: AsIs> | <S3: AsIs> | <S3: AsIs> | <S3: AsIs> | <dbl> | <dbl> | <dbl> |
| buyhghnd | 6 | 0.81021 | | | | | 0.7077680 | 0.2922320 | 1.159435 |
| pricqual | 7 | 0.78968 | | | | | 0.6733958 | 0.3266042 | 1.161975 |
| twoincom | 27 | 0.76032 | | | | | 0.6089022 | 0.3910978 | 1.108413 |
| carefmny | 17 | -0.74342 | | | | | 0.7442939 | 0.2557061 | 1.721360 |
| lthrbetr | 3 | 0.69108 | | | | | 0.6889598 | 0.3110402 | 1.939982 |
| prmsound | 8 | 0.66381 | | | | | 0.5828106 | 0.4171894 | 1.650002 |
| accesfun | 30 | 0.65385 | | | 0.39923 | | 0.6888094 | 0.3111906 | 2.173992 |
| tkvacatn | 10 | 0.62618 | | | 0.48701 | | 0.7051524 | 0.2948476 | 2.307265 |
| passnimp | 26 | -0.62502 | | -0.41485 | | | 0.6560868 | 0.3439132 | 2.255423 |
| stylclth | 24 | 0.57874 | | | 0.45287 | | 0.6357851 | 0.3642149 | 2.543638 |
| imprtapp | 19 | 0.48456 | | 0.35657 | 0.37159 | | 0.5411918 | 0.4588082 | 3.182086 |
| miniboxy | 2 | | 0.84251 | | | | 0.7381128 | 0.2618872 | 1.080598 |
| suvcmpct | 15 | | 0.81805 | | | | 0.7202812 | 0.2797188 | 1.153486 |
| noparkrm | 11 | | 0.80691 | | | | 0.6889111 | 0.3110889 | 1.116846 |
| secbiggr | 4 | | 0.75947 | | | | 0.5936531 | 0.4063469 | 1.059004 |
| needbetw | 14 | | 0.75736 | | | | 0.5935800 | 0.4064200 | 1.070124 |
| next2str | 16 | | -0.74310 | | | | 0.6430263 | 0.3569737 | 1.334367 |
| nohummer | 28 | | 0.70586 | | | | 0.5074668 | 0.4925332 | 1.037317 |
| homlrgst | 12 | | -0.68055 | | | | 0.7021911 | 0.2978089 | 2.082249 |
| kidtrans | 1 | | | 0.93543 | | | 0.8862464 | 0.1137536 | 1.025630 |
| nordtrps | 23 | | | -0.86803 | | | 0.7670653 | 0.2329347 | 1.036217 |
| kidsbulk | 21 | | | 0.82875 | | | 0.7167893 | 0.2832107 | 1.087676 |
| aftrschl | 29 | | | 0.77750 | | | 0.6970928 | 0.3029072 | 1.320326 |
| envrminr | 13 | | | | -0.87403 | | 0.7898254 | 0.2101746 | 1.068242 |
| shdcarpl | 18 | | | | 0.87340 | | 0.7873493 | 0.2126507 | 1.064794 |
| wntguzlr | 22 | | | | -0.77844 | | 0.7107867 | 0.2892133 | 1.337070 |
| safeimpt | 5 | | | | | 0.90821 | 0.8296296 | 0.1703704 | 1.011627 |
| perfimpt | 9 | | | | | -0.88446 | 0.8060323 | 0.1939677 | 1.061282 |
| lk4whldr | 20 | | | | | 0.85612 | 0.7725671 | 0.2274329 | 1.109379 |
| strngwrn | 25 | | | | | 0.73421 | 0.6928098 | 0.3071902 | 1.596482 |

```
Principal Components Analysis
Call: principal(r = cor, nfactors = 5, rotate = "Varimax", scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix

                          RC1     RC2     RC3     RC5     RC4
SS loadings           5.67079 4.99288 3.71398 3.47914 3.01978
Proportion Var        0.18903 0.16643 0.12380 0.11597 0.10066
Cumulative Var        0.18903 0.35546 0.47926 0.59523 0.69589
Proportion Explained  0.27163 0.23916 0.17790 0.16665 0.14465
Cumulative Proportion 0.27163 0.51080 0.68870 0.85535 1.00000

Mean item complexity =  1.5
Test of the hypothesis that 5 components are sufficient.

The root mean square of the residuals (RMSR) is  0.02898

Fit based upon off diagonal values = 0.99163
```

14

## F. Regression Analysis using the Factor Scores on mvliking variable

```
Call:
lm(formula = mvliking ~ premium_Life + compactness + Child_Friendliness +
    sustainability + safety, data = survey.scores)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9530 -1.5723 -0.0992  1.6137  6.1489

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.84250    0.11036  43.878  < 2e-16 ***
premium_Life        1.04456    0.11050   9.453  < 2e-16 ***
compactness         0.99024    0.11050   8.961  < 2e-16 ***
Child_Friendliness  0.21584    0.11050   1.953   0.0515 .
sustainability     -0.09498    0.11050  -0.860   0.3906
safety             -0.54547    0.11050  -4.936 1.18e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.207 on 394 degrees of freedom
Multiple R-squared:  0.3351, Adjusted R-squared:  0.3267
F-statistic: 39.72 on 5 and 394 DF,  p-value: < 2.2e-16
```

## G. Correlation of Demography attributes and Survey Questions