



The University of Hong Kong
Faculty of Engineering
Department of Computer Science

ICOM6044 Data Science for Business
Instructor: Professor Alan Montgomery

Group Written Assignment 1
Ford Ka Case

By

CHAN Chui Yi, Christy (UID: 3035752145)

KWAN Ho Tin, Joseph (UID:3035751828)

LEE Tin Chak, Daniel (UID: 3035752183)

NG Wah Hay, Jacky (UID: 3035751725)

Date of submission: 17 June 2021

Question 1

Traditionally car industry segmented the car market by size. In Ford, it divided cars into 10 categories (refer to Table 1 below). Ford segmented the overall car market first by product categorization. Type A and Type B were small cars less than 390cm long, in which Type A <360cm, Type B >360cm. Within Type B cars, Ford further segmented by the car price, features, appearance, performance and quality. Type B cars were hence divided into Basic-B (compact car), Trend- B (car with improved driving dynamics), Luxury-B (Luxury car)/ Sports-B (Sports car). The larger a car was, the higher the cost of production and therefore the higher the price. Ford assumed that younger buyers earned lower income than older buyers. Hence, Ford targeted small cars to younger, lower income buyers and large cars to older, wealthier buyers and families.

Ford's typical small car marketing strategy in the past was to classify buyers of Type A and Type B cars into four groups based on average household income per year and their household type (i.e. whether the household has any child, do the children live together with parents as full nesters, or children already depart and establish another household such that parents are empty nesters). Details can refer to Exhibit 3 below. Ford introduced different versions of "New" Fiesta to cater the needs of these different groups of small car buyers.

Segmentation by the size of car is not applicable anymore. On demand side, more people who were not within the traditional segment of small cars (i.e. older and wealthier buyers) favour these cars because of a series of environmental and demographic changes during the 1980s and the early 1990s. Their needs are different from those of the youngers and lower-income buyers. Potential small car buyers' purchase considerations are not solely based on price, but also more features. On supply side, more car manufacturers extend their product lines from large cars to small cars, while existing small car manufacturers launch new models and upgrade existing models. The car market becomes increasingly competitive such that car manufacturers are urged to enhance their segmentation approach to better cater for the needs of car buyers.

Question 2

[Table 2] is the cross-tabulation that illustrates the ratings given by respondents on Question 1 against their preferences (in 3 different groups) in choosing Ford Ka.

Preference Group 1 (Ka Choosers) put Ford Ka as one of their top three choices. Majority in this group (Those giving rating 6-7, contributing to 62% in total) want a car that is trendy. This shows that Ford Ka has strong correlation with "Trendy" and it's likely this group believes Ford Ka can satisfy with this statement. Preference Group 2 (Ka Non-choosers) put Ford Ka as one of their bottom three choices. The rating distribution in this group

is more wide-spread and most gave rating 4 (neutral). This shows that although people in this group refused to choose Ford Ka, they actually have no strong preference on car style. It's possible that there are other reasons why they decided not to choose Ka. The remaining respondents are in Preference Group 3 (Middle). Interestingly, even 76% of them love having a trendy car, Ford Ka is not a car at the top of their list of purchase. This suggests that either Ka is not trendy enough to attract them to buy it, or there are other factors that influence their decisions of which car to buy.

In short, this cross-tabulation reveals that there is no obvious correlation between trendiness and customer preference. It might not be the best approach to market Ford Ka based on trendiness alone.

Question 3

The segmentation scheme is implemented in 2 steps. Firstly, we determine the optimal number of clusters required to form the best k-Means clustering model for the demographic data. Twelve solutions with 2 to 10, 15, 20 and 30 clusters are created and their sum of square errors for points within clusters and between clusters are examined by using scree plots. In the scree plot that shows the R-squared for k-means [Chart 3A], we observe that there is a “kink” at $k=5$ so therefore we use 5 clusters to segment the data. Refer to [Table 3A] for the sum of square errors for the twelve solutions evaluated and [Table 3B] for the size of these 5 clusters.

Secondly, we identify the characteristics of the 5 clusters we generated. Using the centroids of these clusters [Table 3C] we can create a parallel plot [Chart 3B] to understand their relationships with different attributes obtained in the demographics data. [Table 3D] provides a summary of characteristics of different demographic clusters we observe.

Question 4

To conduct the psychographic cluster analysis like Q3, the number of clusters needs to be determined at first. We determine the number of clusters by calculating the sum of squared error (SSE) in which it is calculated by the cluster cohesion (within cluster sum of squares) and cluster separation (between cluster sum of squares) by using scree plots just like what we did in Q3. In the scree plot as shown in [Table 4A] and the numbers in [Table 4E], an “elbow” is observed at $k=6$. Thus, 6 clusters are used to segment the data. [Table 4F] shows the description.

In addition, by deploying the centroids of these 6 clusters [Table 4B], a parallel plot [Chart 4A] is plotted based on the questions we chose [Table 4C]. The questions were chosen based on features of the Ford Ka has

like “Small Cars are more safer nowadays” and “Basic transportation is all I need”. [Table 4D, 4F] provides a brief description based on the parallel plot we plotted.

Question 5

By comparing the two parallel plots of centroid values obtained from demographic/ psychographic cluster analysis in #3 and #4, demographic clusters are more distinctive than psychographic clusters, with far less overlapping of lines in the parallel plot of demographic cluster analysis. The result shows demographic clusters could be easily differentiated. For instance, for cluster 1, 100 % of the respondents are female and not the first-time car buyers. In comparison, all psychographic clusters are statistically similar and difficult to be categorized. In other words, no matter which preference group the 250 respondents belong to, they somehow have similar preference on features and outlooks for choosing cars. As shown in [Chart 4A], clusters 2 to 4 are extremely similar while moving with the same direction across 5 questions.

In [Chart 5A], cluster 1 (Independent & single women in workforce) constitutes 35 % of Preference Group 1 (Ka Chooser) and is clearly the largest group of potential Ka buyers. In fact, among 70 respondents who fall into cluster 1, approximately 6 out of 10 people put Ford Ka as top 3 choices from the 10-small car model list. Demographic cluster 1 is identified to be an effective target segment.

Whereas in [Chart 5B], cluster 5 (Want small cars with practical usage and safety for the basic transportation without concerning the outlook and the privacy for the driving seat) and cluster 6 (Want cars with practical usage for the basic transportation without concerning the outlook and the privacy for the driving seat) are the top 2 groups of potential Ka buyers, respectively constitute 30% and 25% of Preference Group 1 (Ka Chooser). However, looking into cluster 5 or 6 individually, less than half of respondents put Ford Ka as top 3 choices in the corresponding cluster.

As a result, it is reasonable to assume targeting demographic cluster 1 has higher probability to convert consumers' interest into purchase actions than targeting Psychographic cluster 5 or 6. It is more recommended to conduct cluster analysis with demographic data over psychographic data.

[Chart 5C] shows a direct comparison between 2 cluster solutions in a balloon plot. It is observed that nearly 60% of demographic cluster 1 respondents fall into psychographic cluster 5 and 6. All these named clusters are the majority groups in Preference Group 1. The cross-effect of such cluster combination can boost the effectiveness of cluster targeting.

To brainstorm the advertising direction for our targeted demographic cluster 1, two scatter plots for the correlation of psychographic variables and the demographic variables on each demographic cluster are drawn for analysis, respectively.

In [Chart 5D], no insightful pattern could be detected as there are no obvious positive or negative associations between two variables for each cluster. Taking Q1 (I want a car that is trendy) and Q48 (My car must have a very individual interior) as example, this lacks predictability in determining Q1 from a given value of Q48 as there is no relationship for each cluster. The dots literally spread across 1 to 7. This explains the difficulty to identify strong opinions towards specific car attributes hidden behind the 5 selected questions for each demographic cluster.

However, in [Chart 5E], in general, a pattern could be detected as there are positive or negative associations between two variables for each cluster. Taking Age and first time purchase as an example, as we could see for the red cluster, the respondents who are younger tend to be a first time car purchaser. There is a strong relationship between age and first time purchase.

The analysis on two scatter plots provides hints that demographic factors could be a more effective advertising angle than car attributes for our target demographic cluster 1. Hence, we tried to locate distinctive demographics by analysing different balloon plots (see Chart 5F to 5K) with interesting findings summarized below:

1. All demographic cluster 1 respondents are female.
2. All demographic cluster 1 respondents are not first-time purchase.
3. 71% demographic cluster 1 respondents do not have any children.
4. No significant difference is observed in demographic cluster 1 respondent's age, marital status and income.

As such, we suggest to position Ford Ka as a car suitable for women in workforce without any children. Ford Ka is their best accompany for work such that she can decide her way of living with no dependency on others. The advertising message is set to emphasize autonomy of women with a tagline "At Your Own Pace".

Appendix

Question 1:

Table 1:
Categorisation of Cars

Category ¹	Description	Examples
A & B	Small cars less than 390cm long	Ford Fiesta, Fiat Cinquecento, Opel Corsa, Peugeot 106 and 205, Renault Clio, Renault Twingo, VW Polo
C	Lower medium cars	Ford Escort, Opel Astra, Peugeot 306, Renault Mégane, VW Golf
D	Upper medium cars	Ford Mondeo, Audi 4, BMW 3 Series, Peugeot 406, Renault Laguna, VW Passat
E	Executive cars	Ford Scorpio, Audi 6, BMW 5 Series, Mercedes E-Class, Renault Safrane
F	Luxury cars	Jaguar XJ6/12, BMW 7 Series, Mercedes S-Class
G	Cabriolets	BMW Z3, Mazda MX-5, Mercedes SLK, Porsche Boxster
J	Sports utility vehicles (Dual-purpose vehicles)	Ford Maverick/Explorer, Opel Frontera, Range Rover, Toyota RAV4, Toyota Landcruiser
M	Minivans	Ford Galaxy, Fiat Ulysse, Renault Espace, VW Sharan
S	Coupés	Ford Probe, Mercedes CLK, Opel Tigra

Exhibit 3
French B-Category Buyers by Household Type, Age and Income for 1994

Household Description	Age	Average Household Income per year (FF 000)				Total
		<100	100-150	150-250	>250	
Primary Car ^a						
Single Women	<30	3.6 ^b	1.7	1.4	0.9	7.6%
Single Women	30-45	2.7	3.1	0.9	0.2	6.9%
Single Men	<30	2.7	1.6	2.1	1.3	7.7%
Single Men	30-45	1.0	0.9	0.7	0.4	3.0%
Childless Couples	<30	0.9	1.2	1.7	0.3	4.1%
Childless Couples	30-45	0.5	0.8	1.0	0.4	2.7%
Full Nesters	<35	0.9	1.8	2.0	1.0	5.7%
Full Nesters / Single Car	>35	1.1	1.5	1.4	0.5	4.5%
Full Nesters / Multi Car	>35	1.1	1.4	5.8	5.5	13.8%
Empty Nesters	45-60	2.7	2.9	3.1	1.5	10.2%
Empty Nesters	>60	6.9	4.8	3.5	2.1	17.3%
Secondary Car						
Household with Children		0.3	1.2	3.3	4.2	9.0%
Household w/o Children		0.4	0.7	2.5	3.7	7.3%
Total		24.8%	23.6%	29.4%	22.0%	100%

a. Indicates the car driven by the head of the household.

b. To be read as 3.6% of B-category car buyers in 1994 were single women who were less than 30 years old and had an average annual income of less than FF100,000.

Ford's Segmentation of Small Car Buyers for the Ford Fiesta

Ford traditionally grouped the small car buyers (A and B categories) into four target groups based on financial constraints and the presence of children in the household. The resulting four target groups are indicated in the above table by the following colour-coding:

Financially constrained buyers	Average income and/or family constrained buyers
Average income, more youthful/expressive buyers	Higher income, not family constrained buyers

Source: Ford France and New Car Buyer Survey France 1995.

Question 2:

Table 2 – Cross-tabulation of respondent rating against Ford Ka preference groups

Preference Group	Question 1							Sub-Total
	1 (Strongly disagree)	2	3	4 (Neutral)	5	6	7 (Strongly agree)	
Ka Choosers (Group 1)	3 (3%)	1 (1%)	14 (12%)	26 (22%)	17 (15%)	27 (23%)	28 (24%)	116 (100%)
Ka Non-Choosers (Group 2)	1 (1%)	4 (6%)	11 (15%)	22 (31%)	12 (17%)	13 (18%)	9 (13%)	72 (100%)
Middle (Group 3)	0 (0%)	0 (0%)	9 (15%)	6 (10%)	3 (5%)	24 (39%)	20 (32%)	62 (100%)

Question 3:

Chart 3A – Screen Plot to derive the optimal # of k-clusters

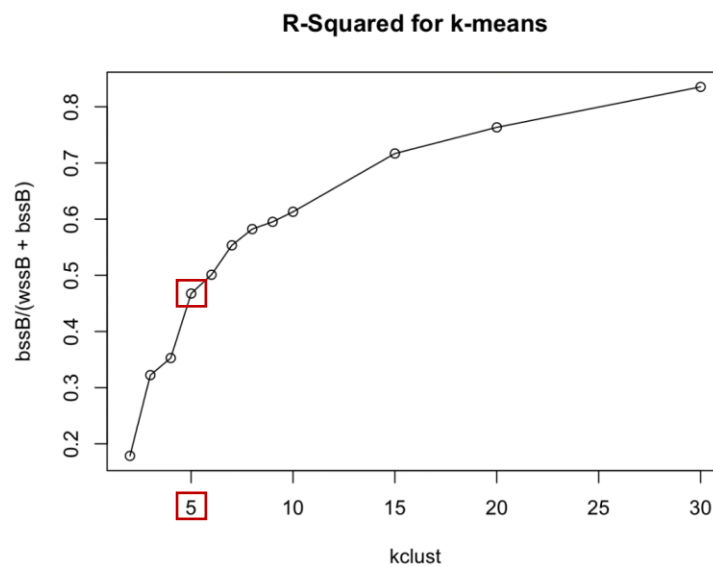


Table 3A – Sum of square errors (SSE) of 12 k-Means clustering solutions

No. of clusters used in k-Means	(A) SSE within cluster	(B) SSE between clusters	R-squared for k-means (A)/(A+B)
2	266.76	1227.24	0.18
3	481.63	1012.37	0.32
4	526.98	967.02	0.35
5	698.51	795.49	0.47
6	748.58	745.42	0.50
7	826.88	667.12	0.55
8	870.00	624.00	0.58
9	889.20	604.80	0.60
10	915.88	578.12	0.61
15	1070.93	423.07	0.72
20	1140.54	353.46	0.76
30	1248.17	245.83	0.84

Table 3B – Cluster size of 5 clusters obtained

Cluster	Cluster Size (No. of respondents)
1	70
2	37
3	50
4	47
5	46
Total	250

Table 3C – Cluster centroids of the 5 clusters generated from k-Means clustering

Cluster	First Time Purchase	Income Category	Number of Children	Gender	Marital Status	Age
1	0.416	-0.078	-0.427	1.039	-0.031	0.004
2	-2.395	-0.226	-0.077	0.337	-0.008	-0.488
3	0.416	0.522	-0.509	-0.959	0.201	0.937
4	0.416	0.190	1.699	0.146	-0.113	0.105
5	0.416	-0.461	-0.472	-0.959	-0.049	-0.739

Chart 3B – Parallel Plot of Cluster centroids against demographics attributes

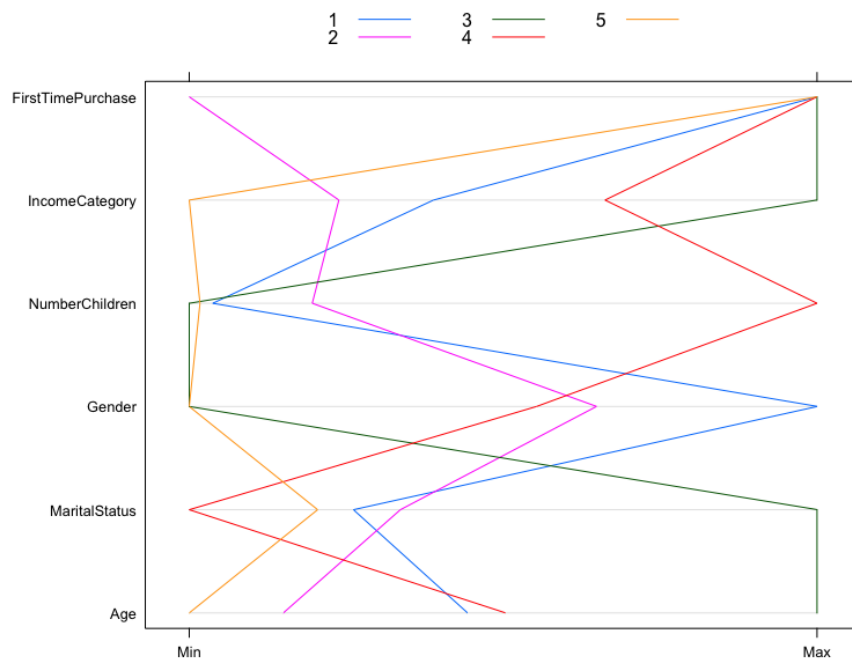


Table 3D – Summary of demographic cluster characteristics

Cluster	Description	Typical Demographic Profile
1	Independent & single women in workforce	100% has bought a car previously; Earned middle range income; Has least number of children; 100% are female; Most are in mid-age group
2	Young adults who just start their career	First time car-buyer; Earned a lower range income; Majority has no children; Comprised both male and female in younger age group
3	Successful wealthy middle-age single business men	100% have bought a car previously; Earned the most income among all clusters; Has least number of children; 100% are male; In oldest age group among all clusters
4	Typical family in workforce with children	Most have bought a car previously; Earned upper range income; Likely to be family that has most number of children
5	Young men car lovers	100% have bought a car previously; Earned the lowest income among all clusters; More interviewees are married than single; 100% are male; In youngest age group among all clusters

Question 4:

Table 4A – Screen Plot to derive the optimal # of k-clusters

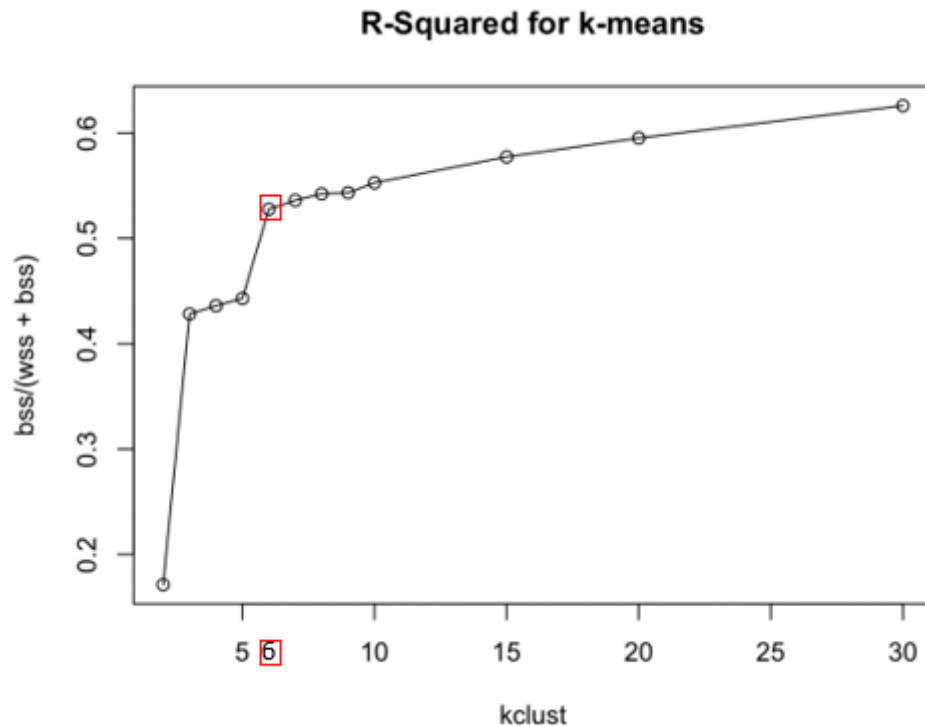


Table 4B – Cluster centroids of the 6 clusters generated from k-Means clustering

Cluster	Q1	Q4	Q21	Q22	Q48
1	0.910656	-1.67915	1.098657	1.150499	-0.50607
2	0.920512	-0.27502	-0.89688	-0.39372	1.153963
3	0.991965	-0.06812	-0.50678	-1.03194	1.12432
4	0.833219	0.030511	-0.86259	-0.65694	1.233013
5	-0.73286	-0.09574	0.77397	0.757964	-0.47129
6	-0.70551	1.092053	-0.51278	-0.64782	-0.60641

Table 4C – Questions chosen as the psychographic variables

Question #	Statement
1	I want a car that is trendy.
4	Basic transportation is all I need.
21	Good aerodynamics help fuel economy.
22	Small cars are much safer nowadays.
48	My car must have a very individual interior.

Chart 4A – Parallel Plot of Cluster centroids against psychographic attributes

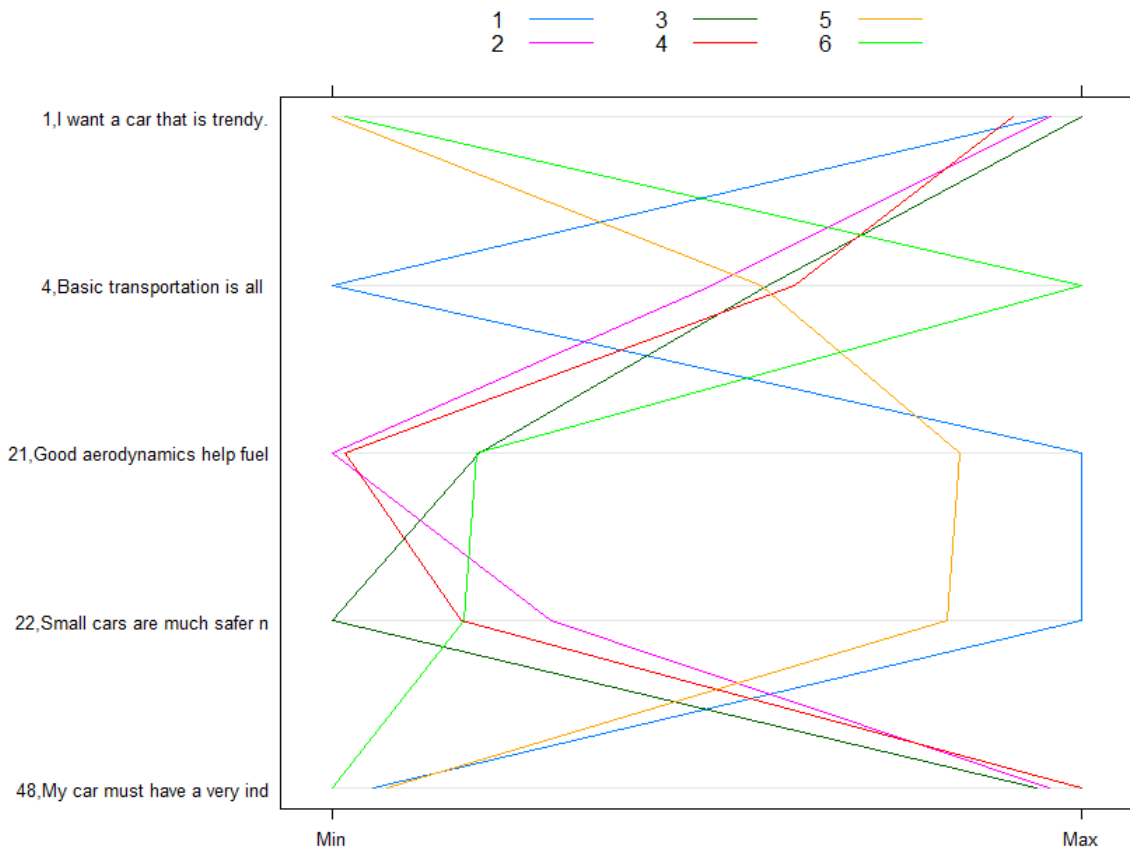


Table 4D – Summary of psychographic cluster characteristic

Cluster	Description	Typical Psychographic Profile
1	Want small trendy cars but not for the basic transportation and not concern the privacy of the driving seat.	Highly concern the outlook of cars; Do not agree a car is a basic transportation. Strongly agree good aerodynamics help fuel economy. Feel that small cars are safer nowadays. Do not need a very individual interior.
2	Want trendy cars for the basic transportation and concern the privacy of the driving seat but do not feel small cars are safer and good aerodynamics help fuel economy.	Highly concern the outlook of cars Somehow agree a car is a basic transportation. Strongly disagree good aerodynamics help fuel economy. Do not feel that small cars are safer nowadays. Highly demand a very individual interior.
3	Want trendy cars for the basic transportation and concern the privacy of the driving seat but do not feel small cars are safer and good aerodynamics help fuel economy.	Highly concern the outlook of cars Somehow agree a car is a basic transportation. Strongly disagree good aerodynamics help fuel economy. Strongly disagree small cars are safer nowadays. Highly demand a very individual interior.
4	Want trendy cars for the basic transportation and concern the privacy of the driving seat but do not feel small cars are safer and good aerodynamics help fuel economy.	Highly concern the outlook of cars Somehow agree a car is a basic transportation. Strongly disagree good aerodynamics help fuel economy. Do not feel that small cars are safer nowadays. Highly demand a very individual interior.
5	Want small cars with practical usage and safety for the basic transportation without concerning the outlook and the privacy for the driving seat	No concern for the outlook of cars Somehow agree a car is a basic transportation. Strongly agree good aerodynamics help fuel economy. Feel that small cars are safer nowadays. Do not need a very individual interior.
6	Want cars with practical usage for the basic transportation without concerning the outlook and the privacy for the driving seat	No concern for the outlook of cars Somehow agree a car is a basic transportation. Do not agree good aerodynamics help fuel economy. Do not feel that small cars are safer nowadays. Do not need a very individual interior.

Table 4E – Sum of square errors (SSE) of 12 k-Means clustering solutions

No. of clusters used in k-Means	(A) SSE within cluster	(B) SSE between clusters	R-squared for k-means
2	12793.10	2644.90	0.83
3	8825.74	6612.26	0.57
4	8706.12	6731.88	0.56
5	8598.72	6839.29	0.56
6	7290.08	8147.92	0.47
7	7161.08	8276.92	0.46
8	7065.54	8372.46	0.46
9	7049.66	8388.34	0.46

10	6902.75	8535.25	0.45
15	6523.35	8914.65	0.42
20	6247.34	9190.66	0.40
30	5772.47	9665.53	0.37

Table 4F – Cluster size of 6 clusters obtained

Cluster	Cluster Size (No. of respondents)
1	32
2	33
3	24
4	21
5	75
6	65
Total	250

Question 5:

Chart 5A – Balloon Plot for Preference Group and Demographic clusters

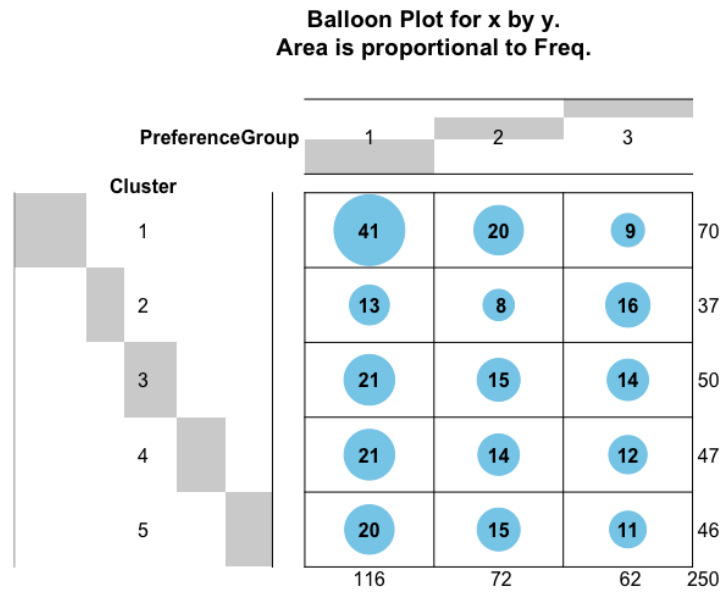


Chart 5B – Balloon Plot for Preference Group and Psychographic clusters

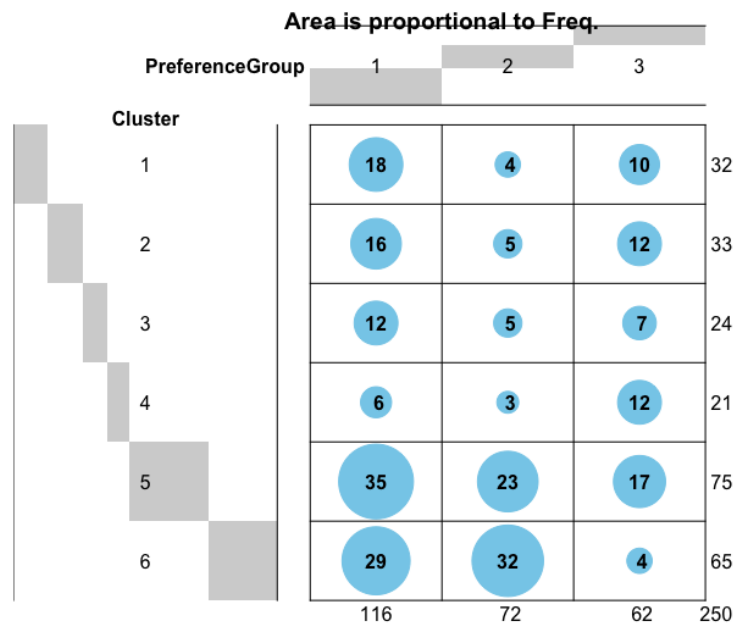


Chart 5C – Balloon Plot for Demographic clusters and Psychographic clusters

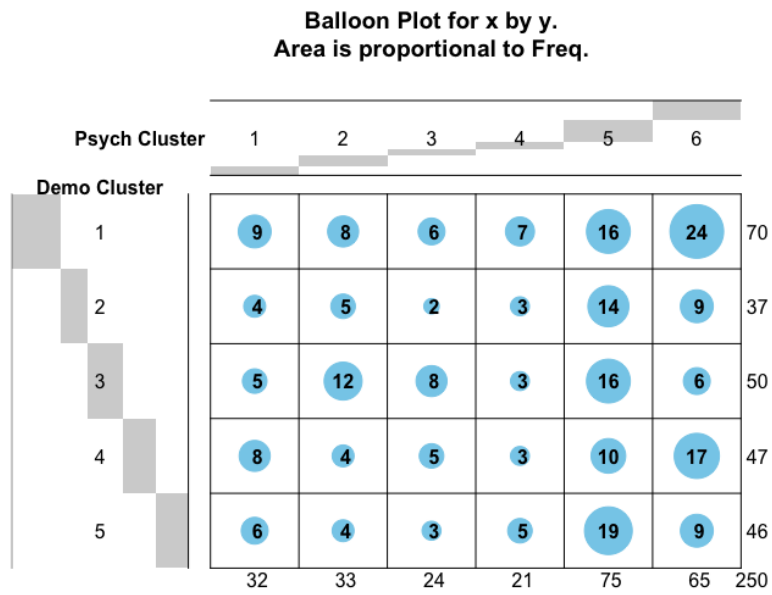


Chart 5D – Pairwise plot of psychographic variables for demographic clusters

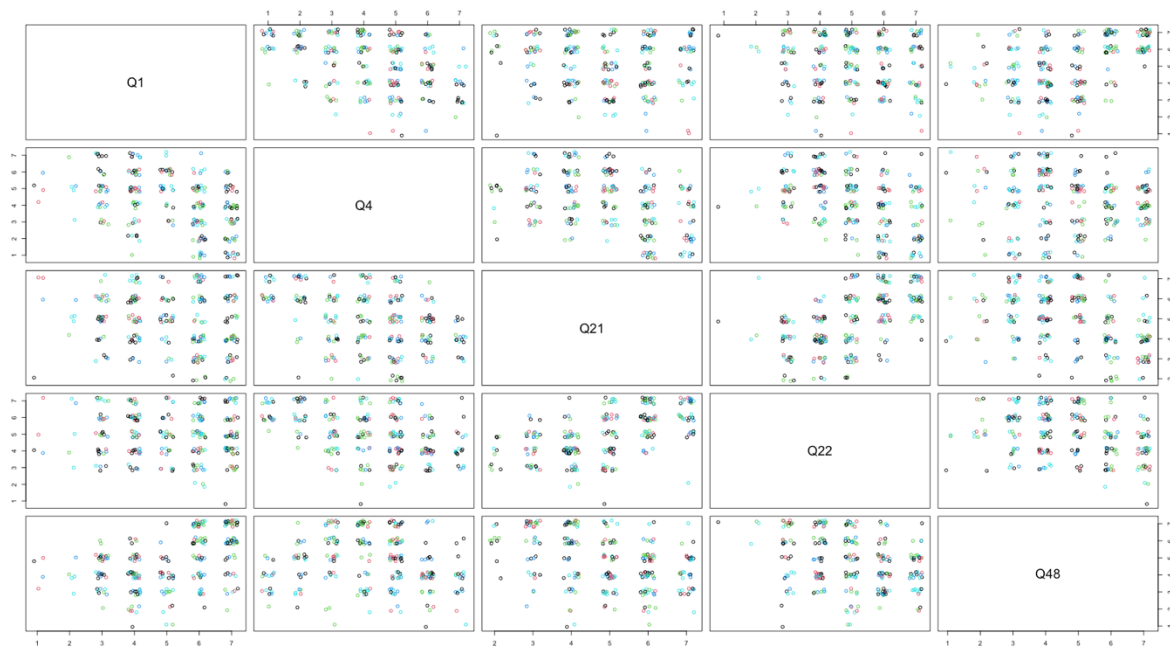


Chart 5E – Pairwise plot of demographic variables for demographic clusters

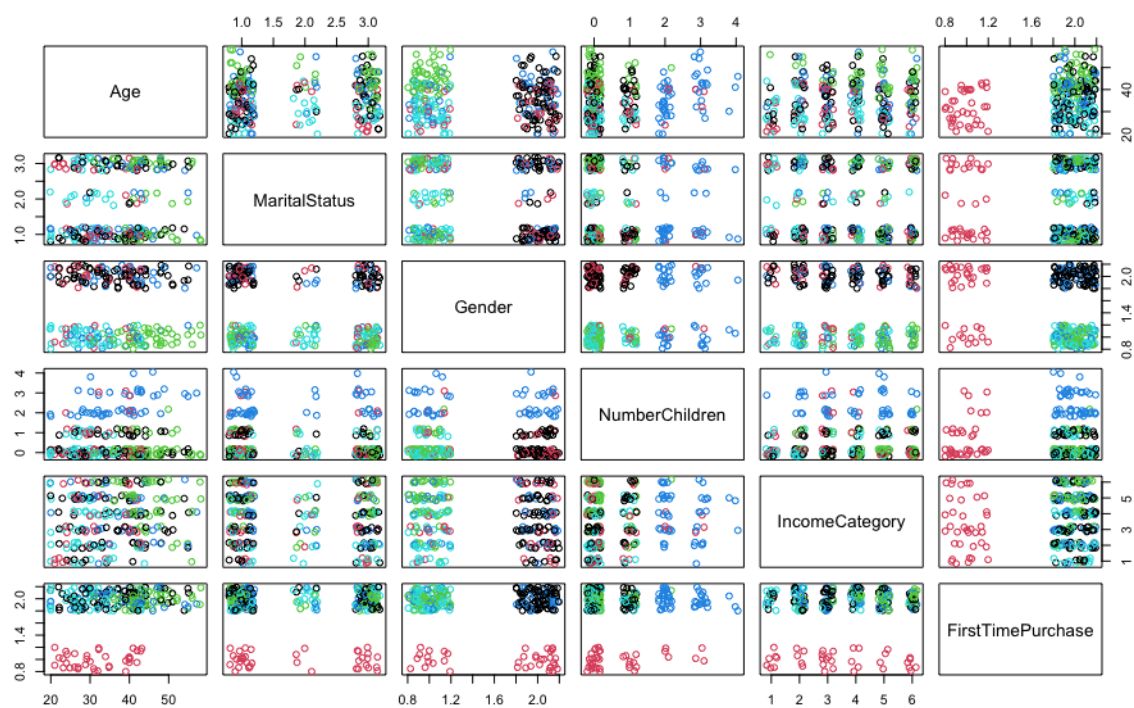


Chart 5F – Balloon plot for Demographic clusters and gender

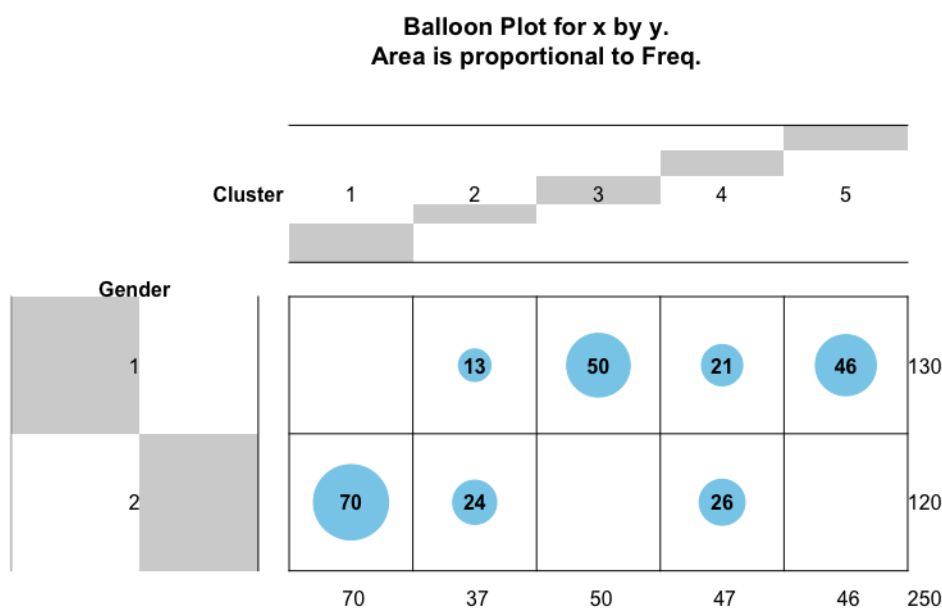


Chart 5G – Balloon plot for Demographic clusters and first time purchase

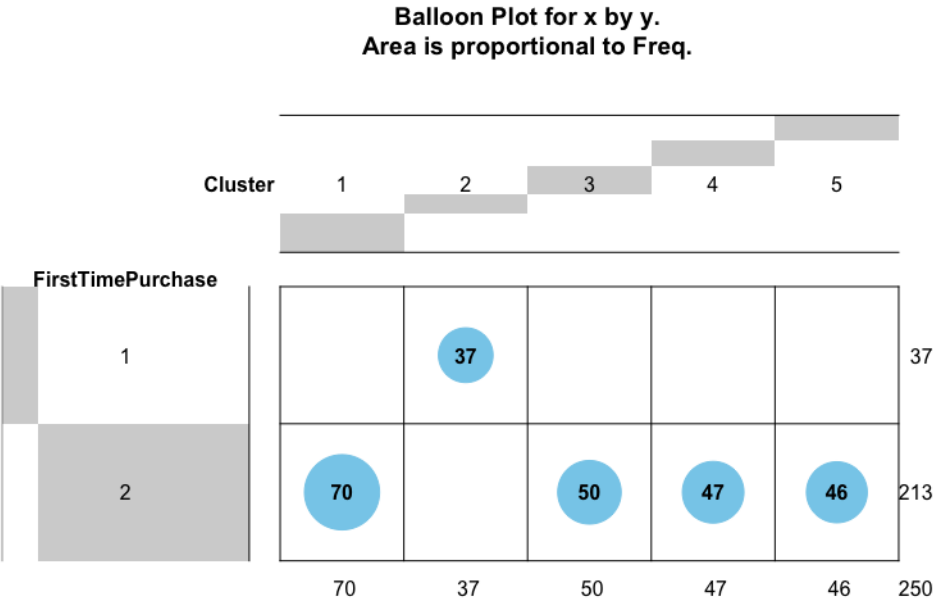


Chart 5H – Balloon plot for Demographic clusters and Children Category

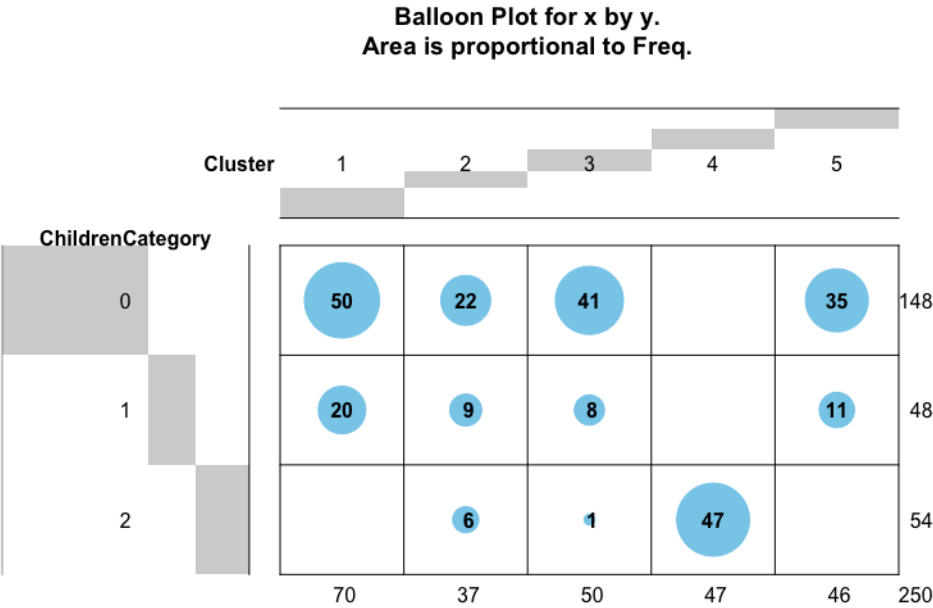


Chart 5I – Balloon plot for Demographic clusters and Age Category

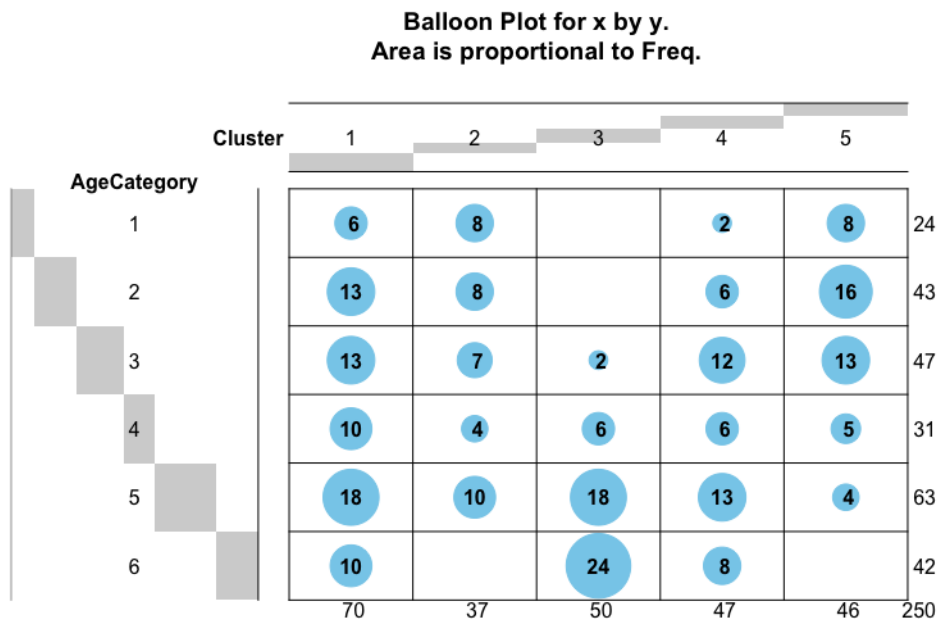


Chart 5J – Balloon plot for Demographic clusters and Marital status

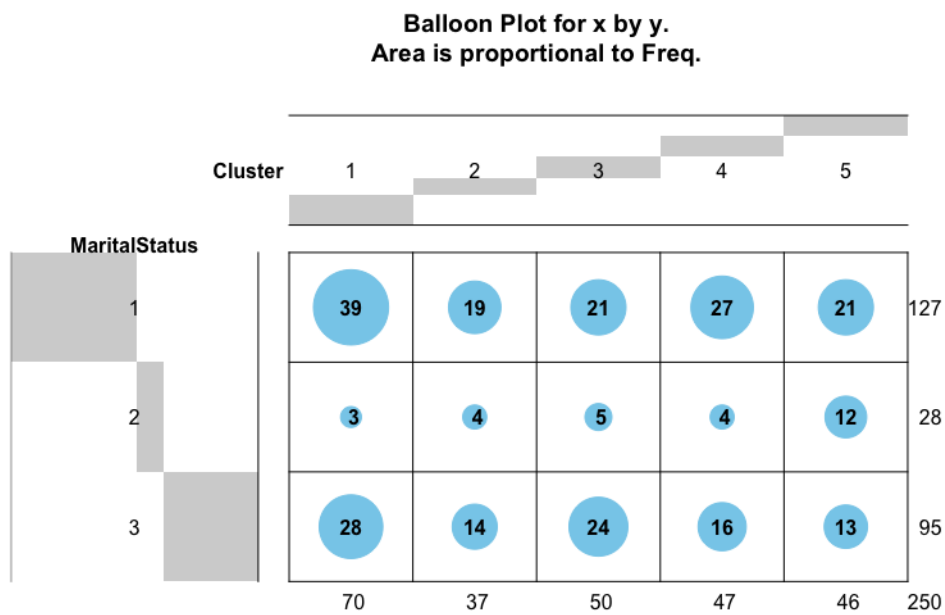


Chart 5K – Balloon plot for Demographic clusters and Income Category

