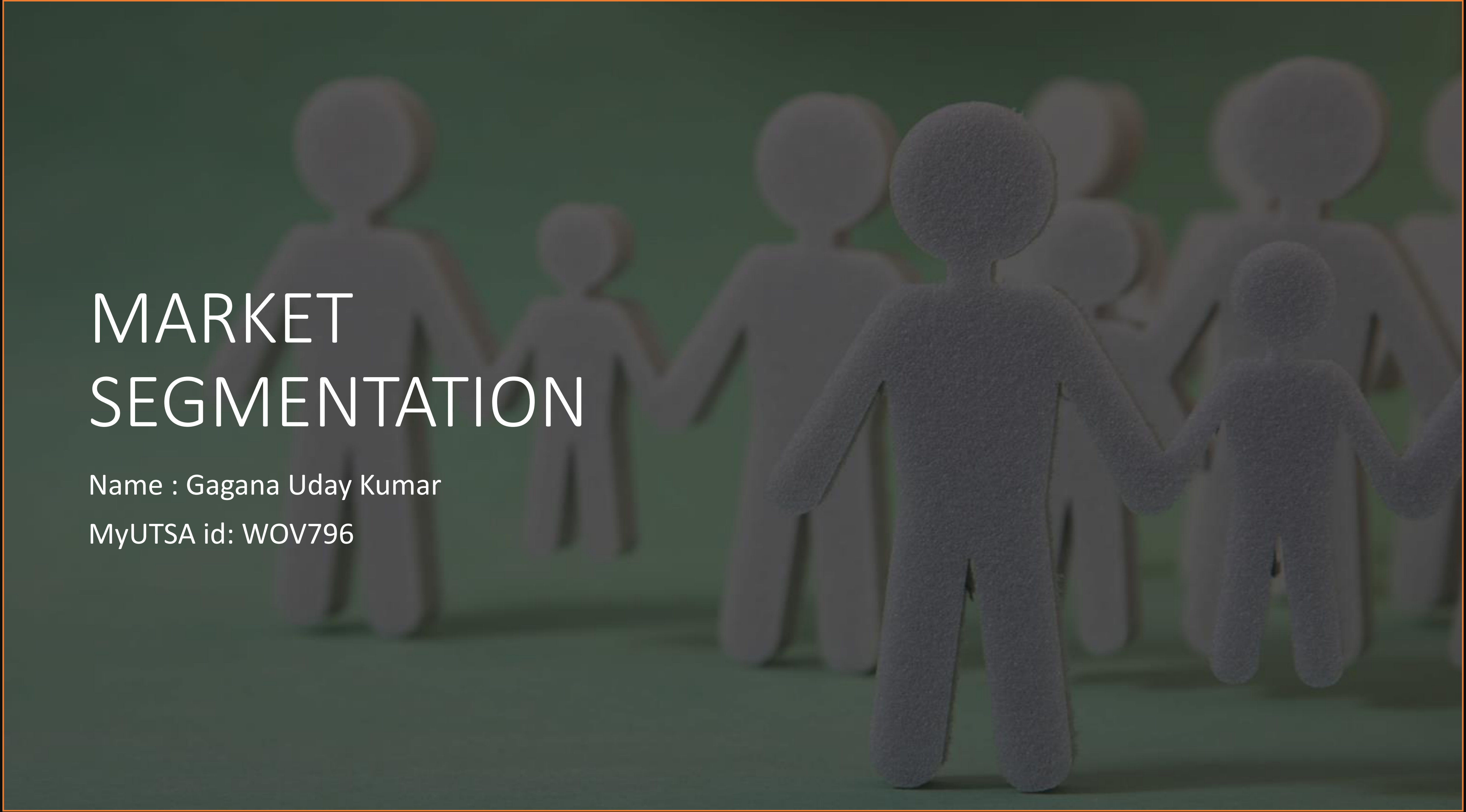
MARKET SEGMENTATION

Name : Gagana Uday Kumar

MyUTSA id: WOV796



Design the Market Segmentation

Target Variable

- DONUTS / DOUGHNUTS - BRANDS MO KRISPY KREME

Single Driver Variable

- I FEEL GUILTY WHEN I EAT SWEETS
- PREFER FOOD PRESENTED AS AN ART FORM
- BRKFST IS MORE IMPRTNT THN LUNCH OR DNNR
- EATING FAST FOOD HELPS ME STAY IN BUDGET

Abstract construct #1 - Travel

- I LOVE THE IDEA OF TRAVELING ABROAD
- WILLING MAKE TRVL PLAN WITH UNKNWN COMP
- RATHER TAKE TWO/THREE SHRT QUICK VACATNS
- VAC. SOMEWHERE DIFFERENT EVERY TIME

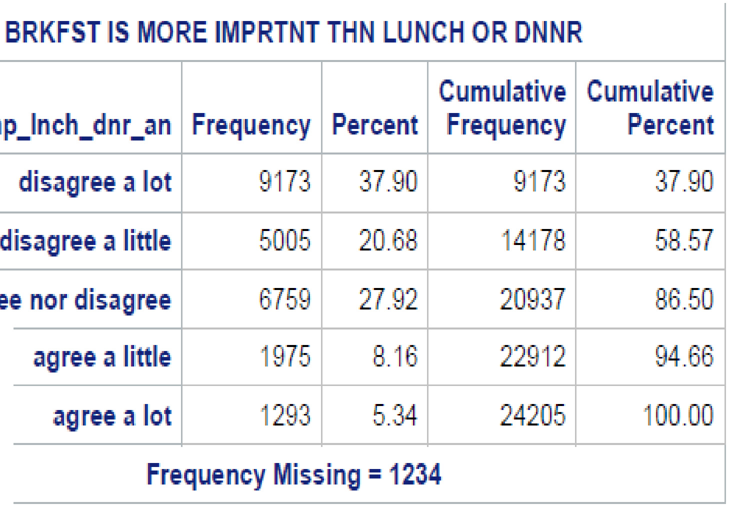
Abstract construct #2 – Social Interaction

- I MAKE FRIENDS EASILY
- PEOPLE SAY MY ENTHUSIASM IS CONTAGIOUS
- I LIKE TO INTRODUCE PEOPLE TO EACH OTHER
- GOOD AT CONVINCING OTHERS TRY NEW THINGS

Descriptor variables

- Major competitors - DUNKIN' DONUTS
- Advertised Channel - YOUTUBE.COM
- Personal Information-Gender MALE , FEMALE
- Origin/ Race - RESPNDNT-SPANISH/HISPANIC/LATINO ORIGIN? YES , NO

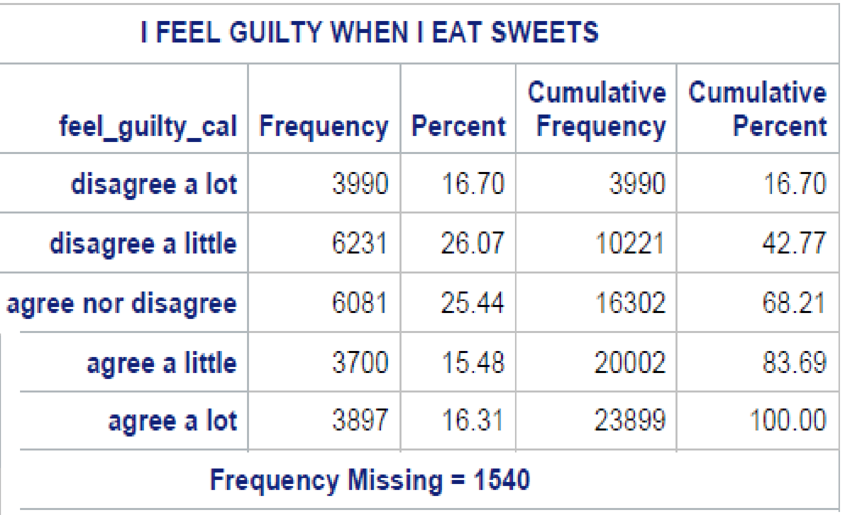
Reading in Raw Data and Creating New Variables



BRKFST IS MORE IMPRTNT THN LUNCH OR DNNR

brft_imp_lnch_dnr_an	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	9173	37.90	9173	37.90
disagree a little	5005	20.68	14178	58.57
neither agree nor disagree	6759	27.92	20937	86.50
agree a little	1975	8.16	22912	94.66
agree a lot	1293	5.34	24205	100.00

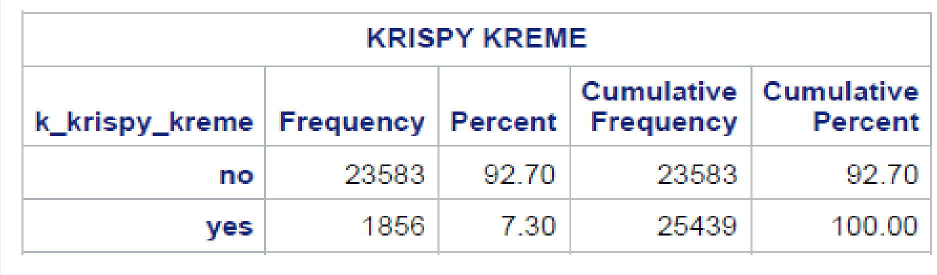
Frequency Missing = 1234



I FEEL GUILTY WHEN I EAT SWEETS

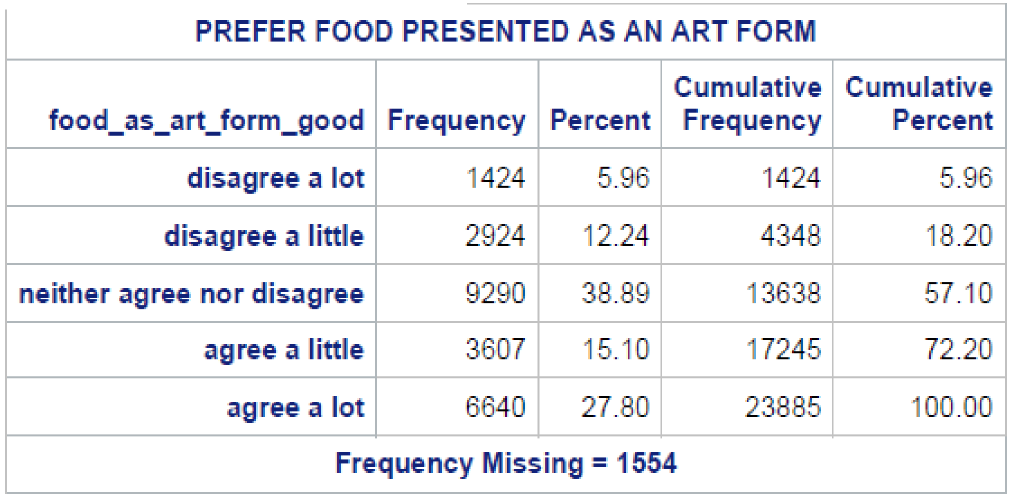
feel_guilty_cal	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	3990	16.70	3990	16.70
disagree a little	6231	26.07	10221	42.77
neither agree nor disagree	6081	25.44	16302	68.21
agree a little	3700	15.48	20002	83.69
agree a lot	3897	16.31	23899	100.00

Frequency Missing = 1540



KRISPY KREME

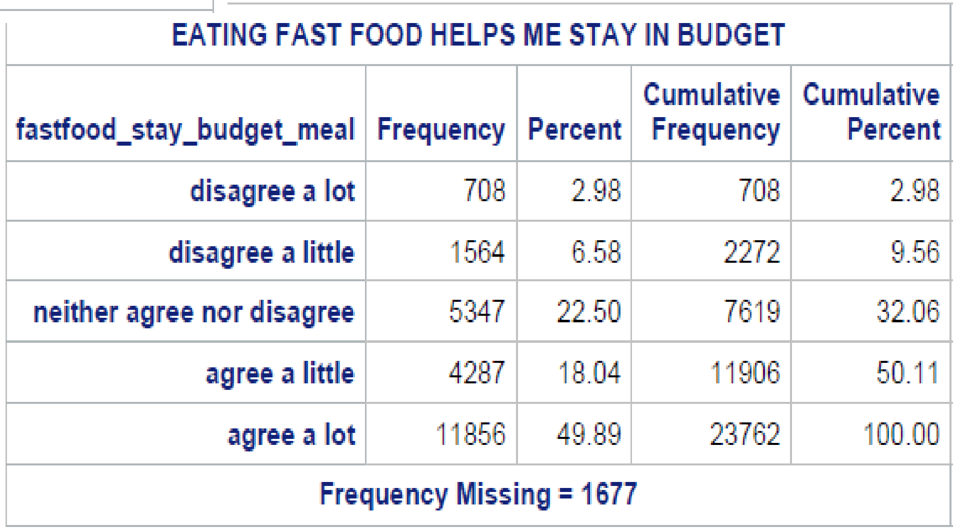
k_krispy_kreme	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	23583	92.70	23583	92.70
yes	1856	7.30	25439	100.00



PREFER FOOD PRESENTED AS AN ART FORM

food_as_art_form_good	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	1424	5.96	1424	5.96
disagree a little	2924	12.24	4348	18.20
neither agree nor disagree	9290	38.89	13638	57.10
agree a little	3607	15.10	17245	72.20
agree a lot	6640	27.80	23885	100.00

Frequency Missing = 1554



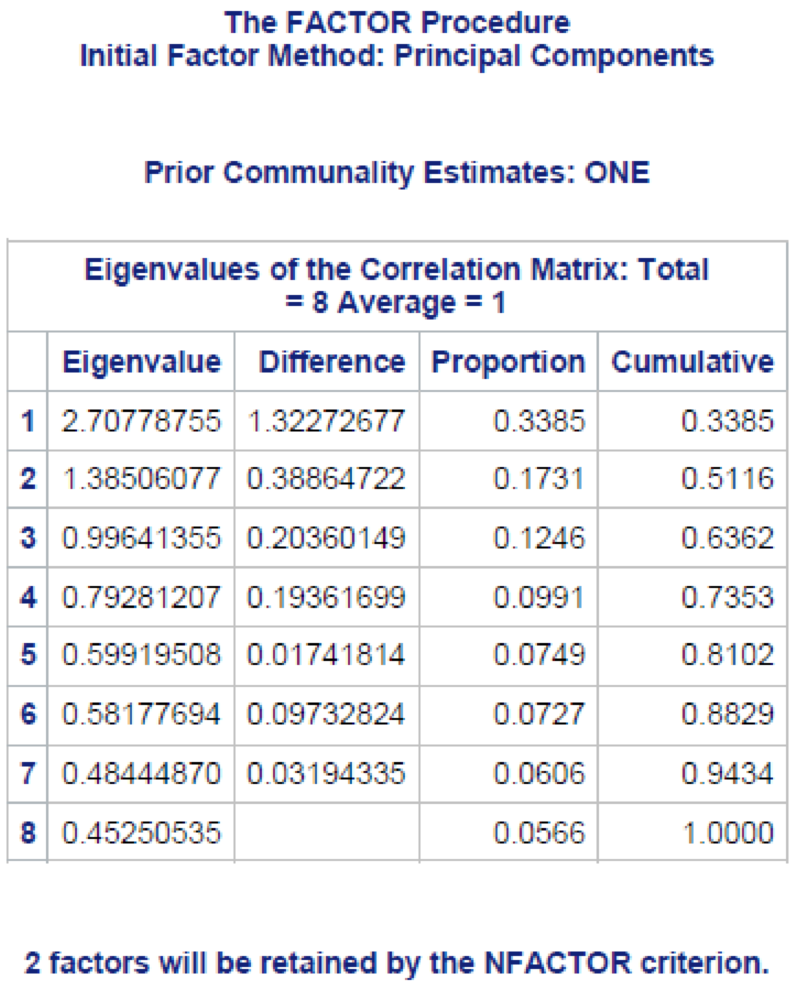
EATING FAST FOOD HELPS ME STAY IN BUDGET

fastfood_stay_budget_meal	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	708	2.98	708	2.98
disagree a little	1564	6.58	2272	9.56
neither agree nor disagree	5347	22.50	7619	32.06
agree a little	4287	18.04	11906	50.11
agree a lot	11856	49.89	23762	100.00

Frequency Missing = 1677

Principle Component Analysis

- Extraction Technique – Principle Component Analysis
- Rotation Method – Varimax
- criteria for determining that a factor was extracted : Kaiser Criterion(eigen value=>1)
- Factor extracted – 2
- Variance explained - 51.16%



The FACTOR Procedure
Initial Factor Method: Principal Components

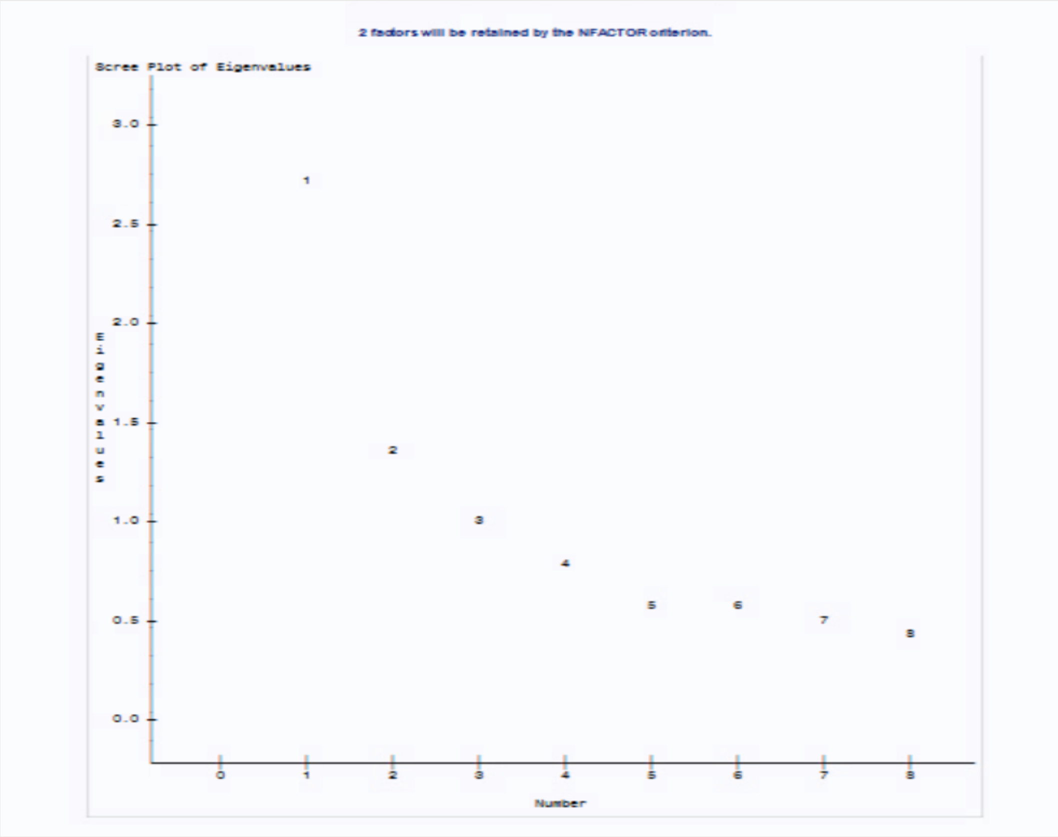
Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 8 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.70778755	1.32272677	0.3385	0.3385
2	1.38506077	0.38864722	0.1731	0.5116
3	0.99641355	0.20360149	0.1246	0.6362
4	0.79281207	0.19361699	0.0991	0.7353
5	0.59919508	0.01741814	0.0749	0.8102
6	0.58177694	0.09732824	0.0727	0.8829
7	0.48444870	0.03194335	0.0606	0.9434
8	0.45250535		0.0566	1.0000

2 factors will be retained by the NFACTOR criterion.

Principle Component Analysis – Scree Plot

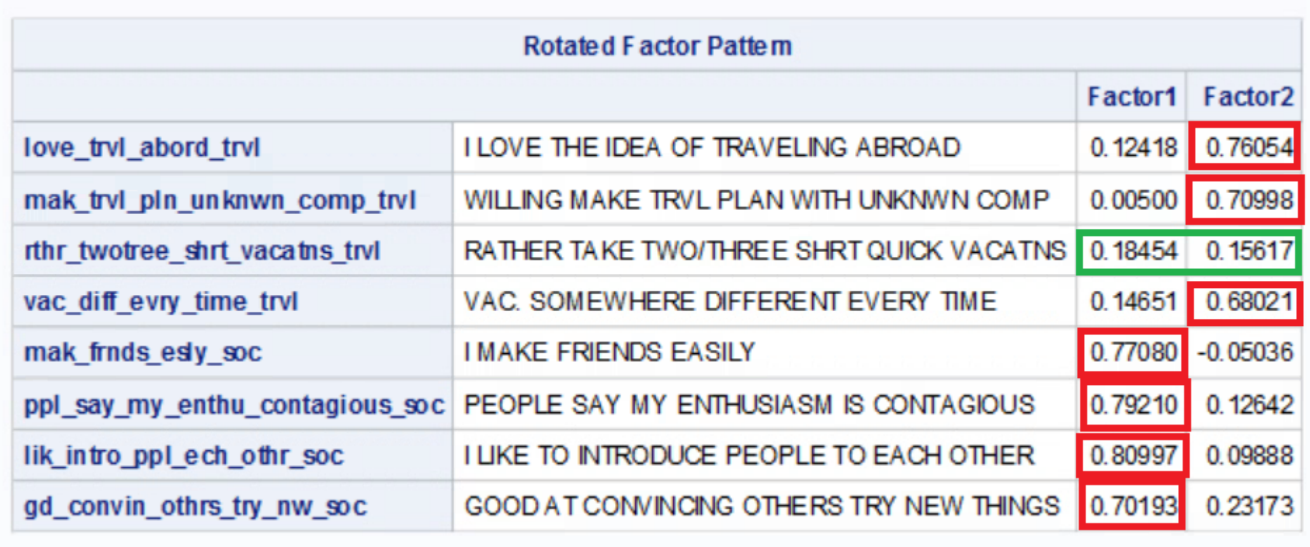


Scree Plot

- The gradient slope are the eigenvalues
- n-1 factor that is 3-1=2 factors are represented on the scree plot

Factor Pattern matrix

- The Variable rthr_twotree_shrt_vacatns_trvl has no suitable factors
- Eliminate Variable and repeat steps
- The two factors extracted are:
  - Travel: People who eat Krispy kreme travel a lot
  - Social Interaction: People who eat krispy kreme are extroverts.



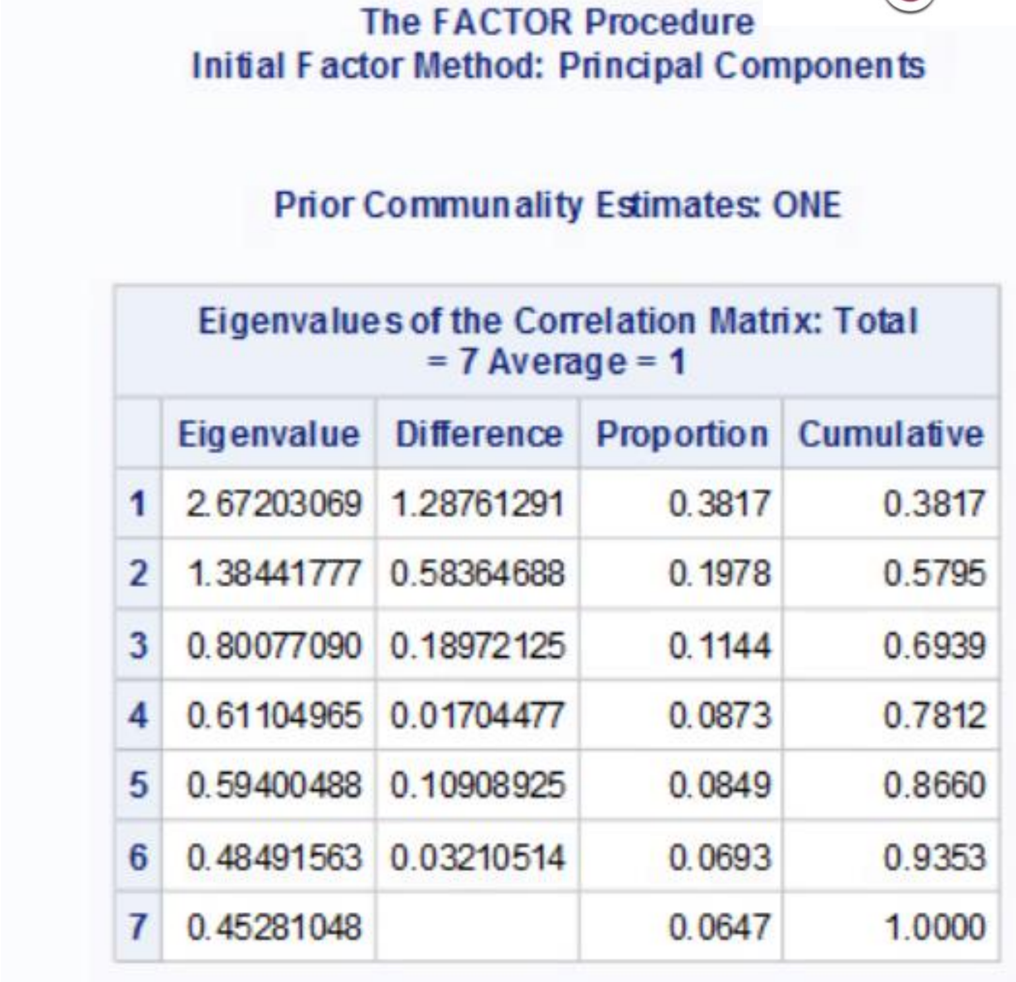
Rotated Factor Pattern

		Factor1	Factor2
love_trvl_abord_trvl	I LOVE THE IDEA OF TRAVELING ABROAD	0.12418	0.76054
mak_trvl_pln_unknwn_comp_trvl	WILLING MAKE TRVL PLAN WITH UNKNWN COMP	0.00500	0.70998
rthr_twotree_shrt_vacatns_trvl	RATHER TAKE TWO/THREE SHRT QUICK VACATNS	0.18454	0.15617
vac_diff_evry_time_trvl	VAC. SOMEWHERE DIFFERENT EVERY TIME	0.14651	0.68021
mak_frnds_esly_soc	I MAKE FRIENDS EASILY	0.77080	-0.05036
ppl_say_my_enthu_contagious_soc	PEOPLE SAY MY ENTHUSIASM IS CONTAGIOUS	0.79210	0.12642
lik_intro_ppl_ech_othr_soc	I LIKE TO INTRODUCE PEOPLE TO EACH OTHER	0.80997	0.09888
gd_convin_othrs_try_nw_soc	GOOD AT CONVINCING OTHERS TRY NEW THINGS	0.70193	0.23173

PCA

Removing the variable
rthr_twotree_shrt_vacatns_trvl

- Extraction Technique – Principle Component Analysis
- Rotation Method – Varimax
- criteria for determining that a factor was extracted : Kaiser Criterion(eigen value=>1)
- Factor extracted – 2
- Variance explained - 57.95%



The FACTOR Procedure
Initial Factor Method: Principal Components

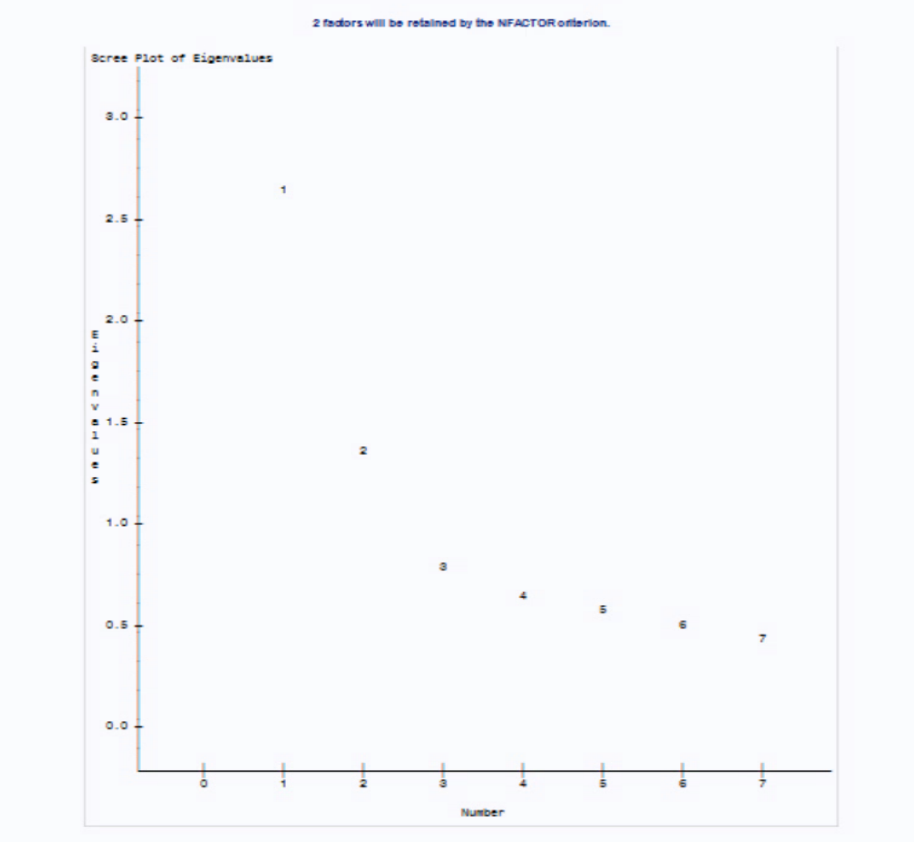
Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.67203069	1.28761291	0.3817	0.3817
2	1.38441777	0.58364688	0.1978	0.5795
3	0.80077090	0.18972125	0.1144	0.6939
4	0.61104965	0.01704477	0.0873	0.7812
5	0.59400488	0.10908925	0.0849	0.8660
6	0.48491563	0.03210514	0.0693	0.9353
7	0.45281048		0.0647	1.0000

PCA

Removing the variable rthr_twotree_shrt_vacatns_trvl



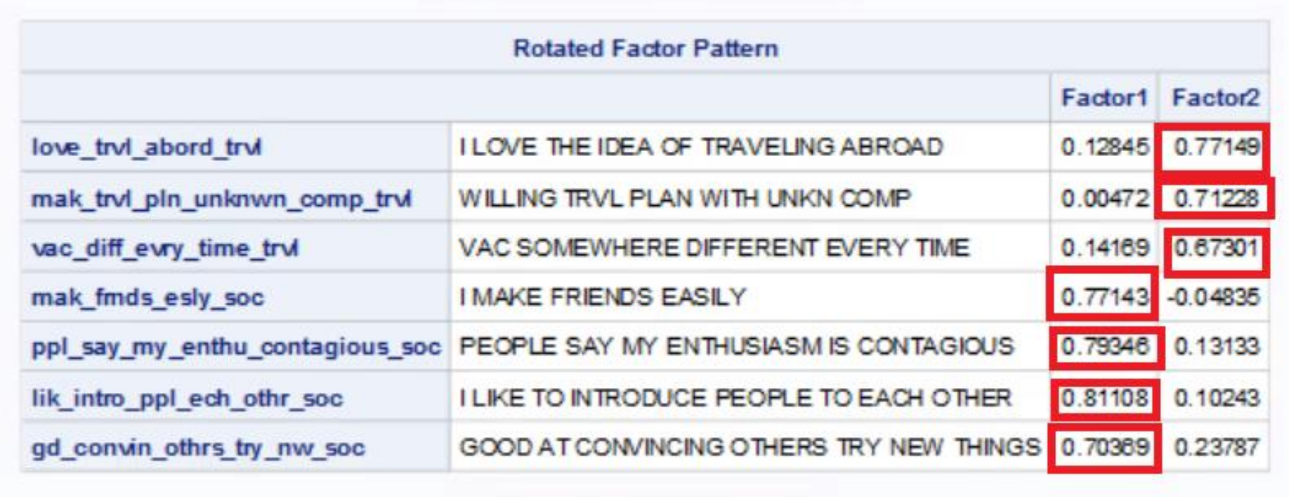
2 factors will be retained by the NFACTOR criterion.

Scree Plot

- The gradient slope are the eigenvalues
- n-1 factor that is 3-1=2 factors are represented on the scree plot

Factor Pattern matrix

- All the variables now have factors which can be chosen.
- The two factors extracted are:
  - Travel: People who eat Krispy kreme travel a lot
  - Social Interaction: People who eat krispy kreme are extroverts.



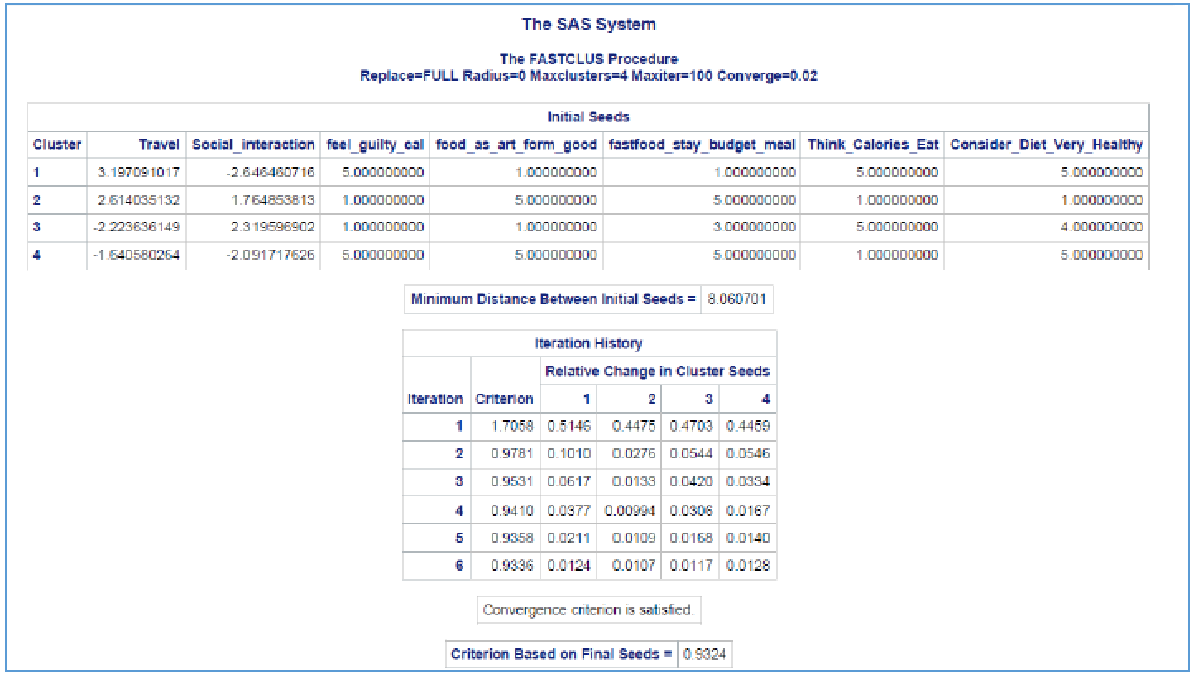
Rotated Factor Pattern

		Factor1	Factor2
love_trvl_abord_trvl	I LOVE THE IDEA OF TRAVELING ABROAD	0.12845	0.77149
mak_trvl_pln_unknwn_comp_trvl	WILLING TRVL PLAN WITH UNKN COMP	0.00472	0.71228
vac_diff_evry_time_trvl	VAC SOMEWHERE DIFFERENT EVERY TIME	0.14169	0.67301
mak_frnds_esly_soc	I MAKE FRIENDS EASILY	0.77143	-0.04835
ppl_say_my_enthu_contagious_soc	PEOPLE SAY MY ENTHUSIASM IS CONTAGIOUS	0.79346	0.13133
lik_intro_ppl_ech_othr_soc	I LIKE TO INTRODUCE PEOPLE TO EACH OTHER	0.81108	0.10243
gd_convin_othrs_try_nw_soc	GOOD AT CONVINCING OTHERS TRY NEW THINGS	0.70369	0.23787

K Means Clustering

- K means cluster analysis using PROC CLUSTER
- Dropped single variable – BRKFST IS MORE IMPRTNT THN LUNCH OR DNNR (as CCC values below -20 for every variable)
- Replaced it with 2 other single Variables
    - I THINK OF THE CALORIES IN WHAT I EAT
    - CONSIDER MY DIET TO BE VERY HEALTHY

For K=4:



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=4 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	3.197091017	-2.646460716	5.000000000	1.000000000	1.000000000	5.000000000	5.000000000
2	2.614035132	1.764853813	1.000000000	5.000000000	5.000000000	1.000000000	1.000000000
3	-2.223636149	2.319596902	1.000000000	1.000000000	3.000000000	5.000000000	4.000000000
4	-1.640580264	-2.091717626	5.000000000	5.000000000	5.000000000	1.000000000	5.000000000

Minimum Distance Between Initial Seeds = 8.060701

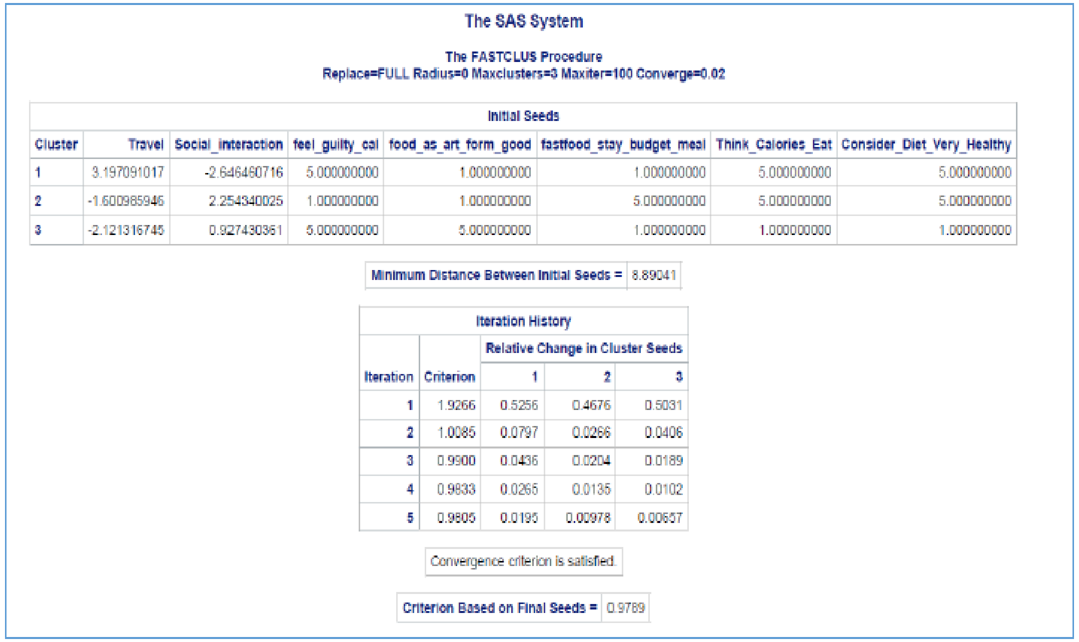
Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds			
		1	2	3	4
1	1.7058	0.5146	0.4475	0.4703	0.4469
2	0.9781	0.1010	0.0276	0.0544	0.0546
3	0.9531	0.0617	0.0133	0.0420	0.0334
4	0.9410	0.0377	0.00994	0.0306	0.0167
5	0.9358	0.0211	0.0109	0.0168	0.0140
6	0.9336	0.0124	0.0107	0.0117	0.0128

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 0.9324

For K=3:



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=3 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	3.197091017	-2.646460716	5.000000000	1.000000000	1.000000000	5.000000000	5.000000000
2	-1.600985946	2.254340025	1.000000000	1.000000000	5.000000000	5.000000000	5.000000000
3	-2.121316745	0.927430361	5.000000000	5.000000000	1.000000000	1.000000000	1.000000000

Minimum Distance Between Initial Seeds = 8.89041

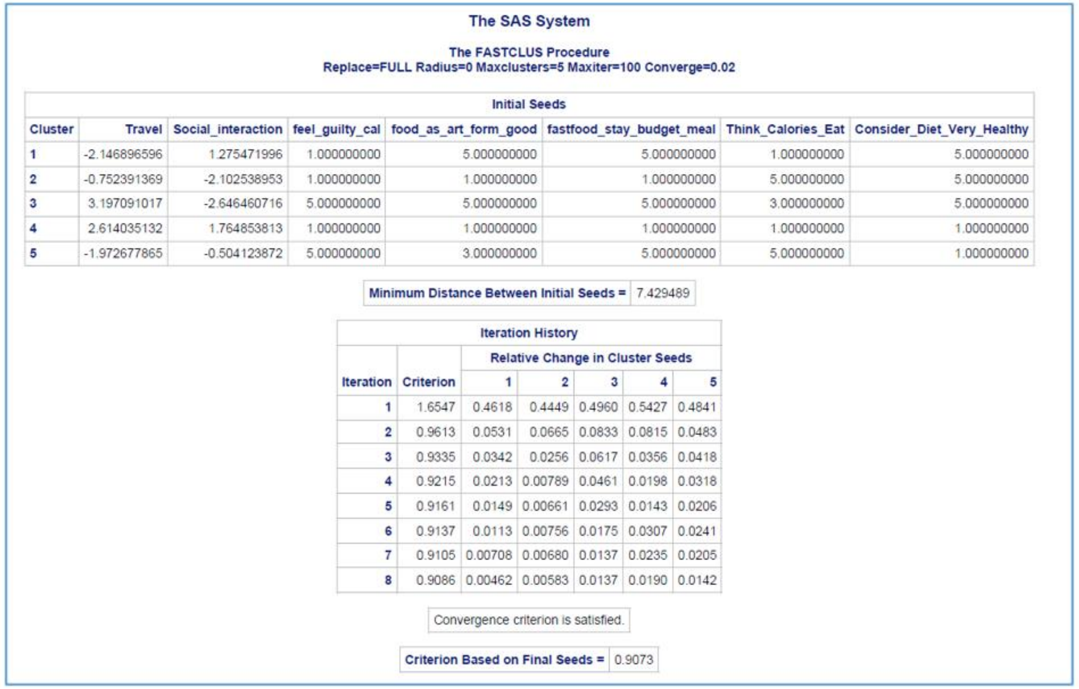
Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds		
		1	2	3
1	1.9266	0.5256	0.4676	0.5031
2	1.0085	0.0797	0.0266	0.0406
3	0.9900	0.0436	0.0204	0.0189
4	0.9833	0.0265	0.0135	0.0102
5	0.9805	0.0195	0.00978	0.00657

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 0.9789

For K=5:



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=5 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	-2.146896596	1.275471996	1.000000000	5.000000000	5.000000000	1.000000000	5.000000000
2	-0.752391369	-2.102538953	1.000000000	1.000000000	1.000000000	5.000000000	5.000000000
3	3.197091017	-2.646460716	5.000000000	5.000000000	5.000000000	3.000000000	5.000000000
4	2.614035132	1.764853813	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000
5	-1.972677865	-0.504123872	5.000000000	3.000000000	5.000000000	5.000000000	1.000000000

Minimum Distance Between Initial Seeds = 7.429489

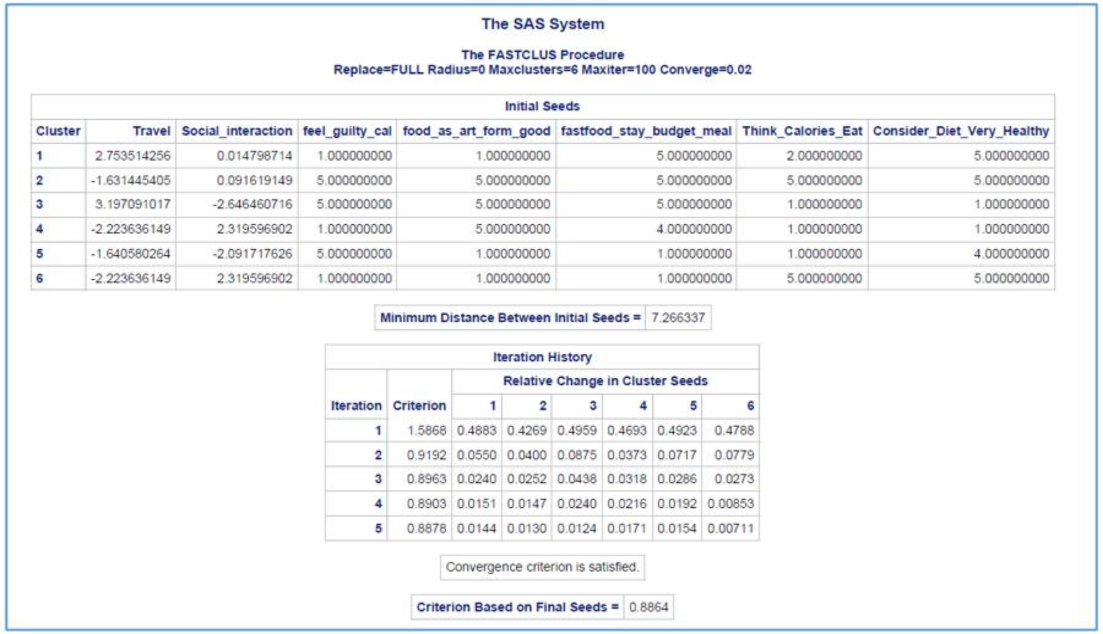
Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds				
		1	2	3	4	5
1	1.6547	0.4618	0.4449	0.4960	0.5427	0.4841
2	0.9613	0.0531	0.0665	0.0833	0.0815	0.0483
3	0.9335	0.0342	0.0256	0.0617	0.0356	0.0418
4	0.9215	0.0213	0.00789	0.0461	0.0198	0.0318
5	0.9161	0.0149	0.00661	0.0293	0.0143	0.0206
6	0.9137	0.0113	0.00756	0.0175	0.0307	0.0241
7	0.9105	0.00708	0.00680	0.0137	0.0235	0.0205
8	0.9086	0.00462	0.00583	0.0137	0.0190	0.0142

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 0.9073

For K=6



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=6 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	2.753514256	0.014798714	1.000000000	1.000000000	5.000000000	2.000000000	5.000000000
2	-1.631445405	0.091619149	5.000000000	5.000000000	5.000000000	5.000000000	5.000000000
3	3.197091017	-2.646460716	5.000000000	5.000000000	5.000000000	1.000000000	1.000000000
4	-2.223636149	2.319596902	1.000000000	5.000000000	4.000000000	1.000000000	1.000000000
5	-1.640580264	-2.091717626	5.000000000	1.000000000	1.000000000	1.000000000	4.000000000
6	-2.223636149	2.319596902	1.000000000	1.000000000	1.000000000	5.000000000	5.000000000

Minimum Distance Between Initial Seeds = 7.266337

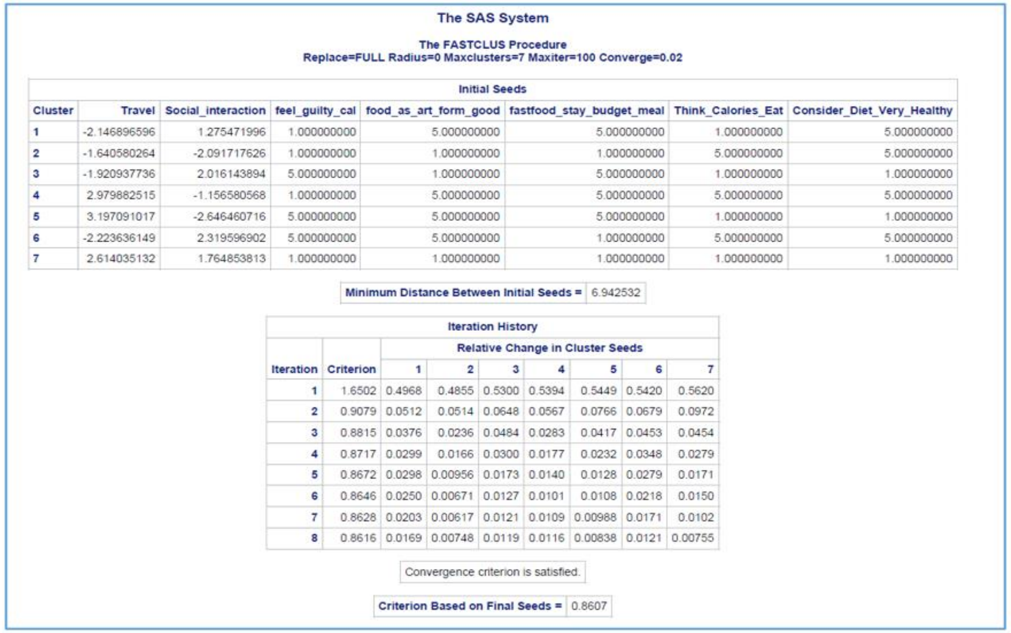
Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds					
		1	2	3	4	5	6
1	1.5868	0.4883	0.4269	0.4959	0.4693	0.4923	0.4788
2	0.9192	0.0550	0.0400	0.0875	0.0373	0.0717	0.0779
3	0.8963	0.0240	0.0252	0.0438	0.0318	0.0286	0.0273
4	0.8903	0.0151	0.0147	0.0240	0.0216	0.0192	0.00853
5	0.8878	0.0144	0.0130	0.0124	0.0171	0.0154	0.00711

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 0.8864

For K=7



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=7 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	-2.146896596	1.275471996	1.000000000	5.000000000	5.000000000	1.000000000	5.000000000
2	-1.640580264	-2.091717626	1.000000000	1.000000000	1.000000000	5.000000000	5.000000000
3	-1.920937736	2.016143894	5.000000000	1.000000000	5.000000000	1.000000000	1.000000000
4	2.979882515	-1.156580568	1.000000000	5.000000000	5.000000000	5.000000000	5.000000000
5	3.197091017	-2.646460716	5.000000000	5.000000000	5.000000000	1.000000000	1.000000000
6	-2.223636149	2.319596902	5.000000000	5.000000000	1.000000000	5.000000000	5.000000000
7	2.614035132	1.764853813	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000

Minimum Distance Between Initial Seeds = 6.942532

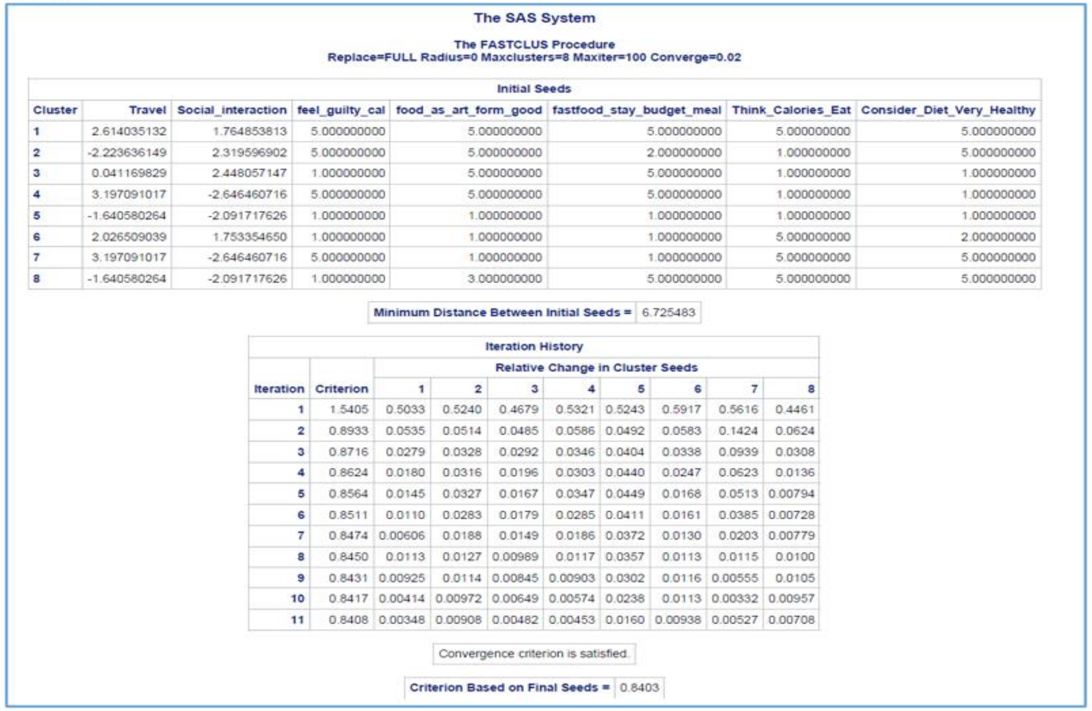
Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds						
		1	2	3	4	5	6	7
1	1.6502	0.4968	0.4855	0.5300	0.5394	0.5449	0.5420	0.5620
2	0.9079	0.0512	0.0514	0.0648	0.0567	0.0766	0.0679	0.0972
3	0.8815	0.0376	0.0236	0.0484	0.0283	0.0417	0.0453	0.0454
4	0.8717	0.0299	0.0166	0.0300	0.0177	0.0232	0.0348	0.0279
5	0.8672	0.0298	0.00956	0.0173	0.0140	0.0128	0.0279	0.0171
6	0.8646	0.0250	0.00671	0.0127	0.0101	0.0108	0.0218	0.0150
7	0.8628	0.0203	0.00617	0.0121	0.0109	0.00988	0.0171	0.0102
8	0.8616	0.0169	0.00748	0.0119	0.0116	0.00838	0.0121	0.00755

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 0.8607

For K=8



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=8 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	2.614035132	1.764853813	5.000000000	5.000000000	5.000000000	5.000000000	5.000000000
2	-2.223636149	2.319596902	5.000000000	5.000000000	2.000000000	1.000000000	5.000000000
3	0.041169829	2.448057147	1.000000000	5.000000000	5.000000000	1.000000000	1.000000000
4	3.197091017	-2.646460716	5.000000000	5.000000000	5.000000000	1.000000000	1.000000000
5	-1.640580264	-2.091717626	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000
6	2.026509039	1.753354650	1.000000000	1.000000000	1.000000000	5.000000000	2.000000000
7	3.197091017	-2.646460716	5.000000000	1.000000000	1.000000000	5.000000000	5.000000000
8	-1.640580264	-2.091717626	1.000000000	3.000000000	5.000000000	5.000000000	5.000000000

Minimum Distance Between Initial Seeds = 6.725483

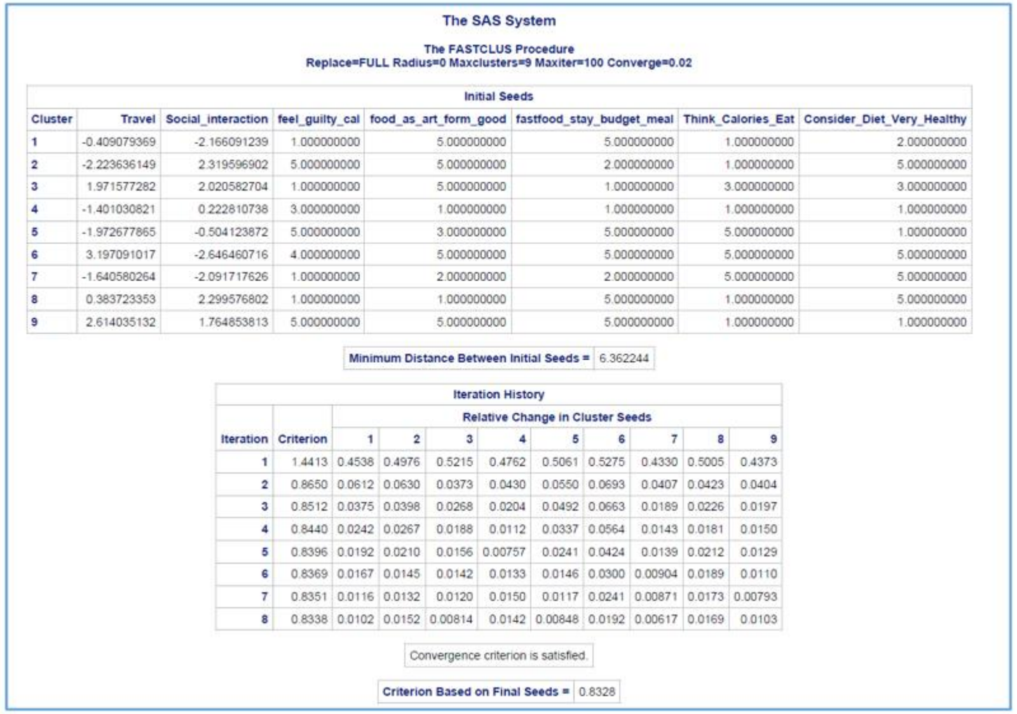
Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds							
		1	2	3	4	5	6	7	8
1	1.5405	0.5033	0.5240	0.4679	0.5321	0.5243	0.5917	0.5616	0.4461
2	0.8933	0.0535	0.0514	0.0485	0.0586	0.0492	0.0583	0.1424	0.0624
3	0.8716	0.0279	0.0328	0.0292	0.0346	0.0404	0.0338	0.0939	0.0308
4	0.8624	0.0180	0.0316	0.0196	0.0303	0.0440	0.0247	0.0623	0.0136
5	0.8564	0.0145	0.0327	0.0167	0.0347	0.0449	0.0168	0.0513	0.00794
6	0.8511	0.0110	0.0283	0.0179	0.0285	0.0411	0.0161	0.0385	0.00728
7	0.8474	0.00606	0.0188	0.0149	0.0186	0.0372	0.0130	0.0203	0.00779
8	0.8450	0.0113	0.0127	0.00989	0.0117	0.0357	0.0113	0.0115	0.0100
9	0.8431	0.00925	0.0114	0.00845	0.00903	0.0302	0.0116	0.00555	0.0105
10	0.8417	0.00414	0.00972	0.00649	0.00574	0.0238	0.0113	0.00332	0.00957
11	0.8408	0.00348	0.00908	0.00482	0.00453	0.0160	0.00938	0.00527	0.00708

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 0.8403

For K=9



The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=9 Maxiter=100 Converge=0.02

Initial Seeds

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	-0.409079369	-2.166091239	1.000000000	5.000000000	5.000000000	1.000000000	2.000000000
2	-2.223636149	2.319596902	5.000000000	5.000000000	2.000000000	1.000000000	5.000000000
3	1.971577282	2.020582704	1.000000000	5.000000000	1.000000000	3.000000000	3.000000000
4	-1.401030821	0.222810738	3.000000000	1.000000000	1.000000000	1.000000000	1.000000000
5	-1.972677865	-0.504123872	5.000000000	3.000000000	5.000000000	5.000000000	1.000000000
6	3.197091017	-2.646460716	4.000000000	5.000000000	5.000000000	5.000000000	5.000000000
7	-1.640580264	-2.091717626	1.000000000	2.000000000	2.000000000	5.000000000	5.000000000
8	0.383723353	2.299576802	1.000000000	1.000000000	5.000000000	1.000000000	5.000000000
9	2.614035132	1.764853813	5.000000000	5.000000000	5.000000000	1.000000000	1.000000000

Minimum Distance Between Initial Seeds = 6.362244

Iteration History

Iteration	Criterion	Relative Change in Cluster Seeds								
		1	2	3	4	5	6	7	8	9
1	1.4413	0.4538	0.4976	0.5215	0.4762	0.5061	0.5275	0.4330	0.5005	0.4373
2	0.8650	0.0612	0.0630	0.0373	0.0430	0.0550	0.0693	0.0407	0.0423	0.0404
3	0.8512	0.0375	0.0398	0.0268	0.0204	0.0492	0.0663	0.0189	0.0226	0.0197
4	0.8440	0.0242	0.0267	0.0188	0.0112	0.0337	0.0564	0.0143	0.0181	0.0150
5	0.8396	0.0192	0.0210	0.0156	0.00757	0.0241	0.0424	0.0139	0.0212	0.0129
6	0.8369	0.0167	0.0145	0.0142	0.0133	0.0146	0.0300	0.00904	0.0189	0.0110
7	0.8351	0.0116	0.0132	0.0120	0.0150	0.0117	0.0241	0.00871	0.0173	0.00793
8	0.8338	0.0102	0.0152	0.00814	0.0142	0.00848	0.0192	0.00617	0.0169	0.0103

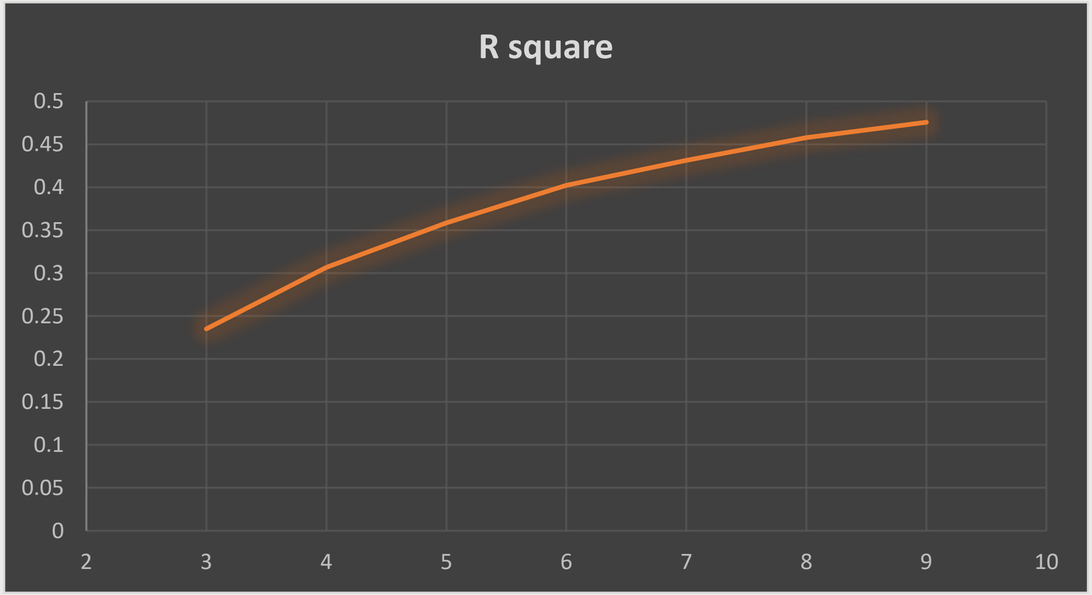
Convergence criterion is satisfied.

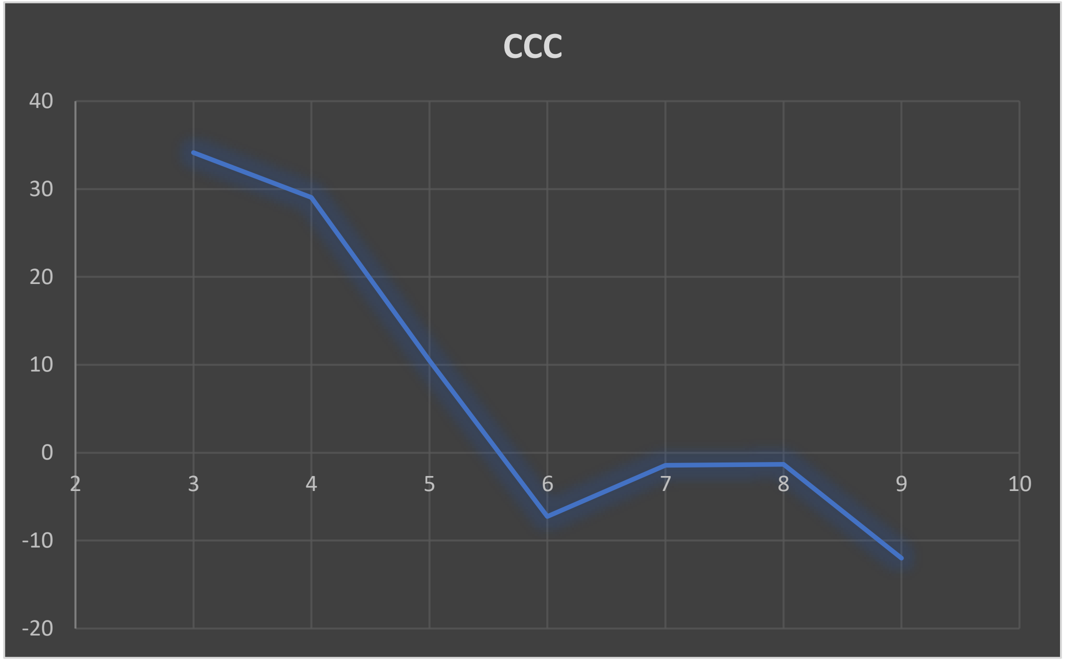
Criterion Based on Final Seeds = 0.8328

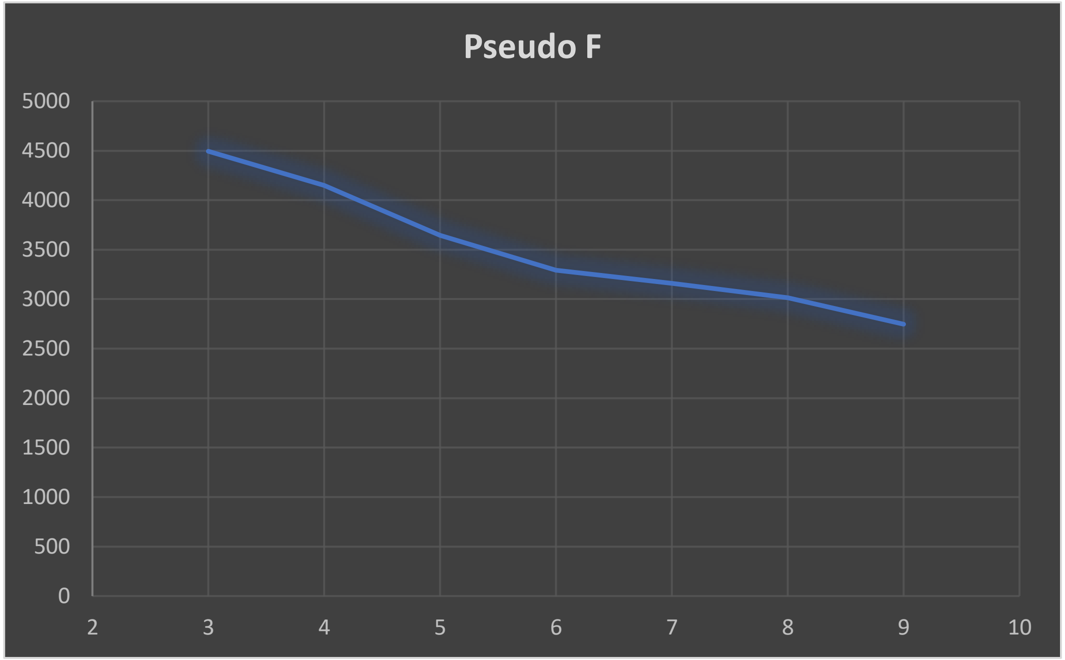
K Means Clustering

- CCC : The first local maximum number of is k=7(cluster 7)
- The Pseudo F plot does not exhibit a clear first local maximum number of clusters as it shows a gradual slope.

K	Number of clusters	R square	CCC	Pseudo F
3	3	0.23506	34.146	4494.41
4	4	0.30671	29.029	4148.73
5	5	0.35866	10.482	3644.44
6	6	0.40204	-7.262	3293.81
7	7	0.43147	-1.428	3160.71
8	8	0.45767	-1.313	3014.21
9	9	0.47566	-11.991	2747.19



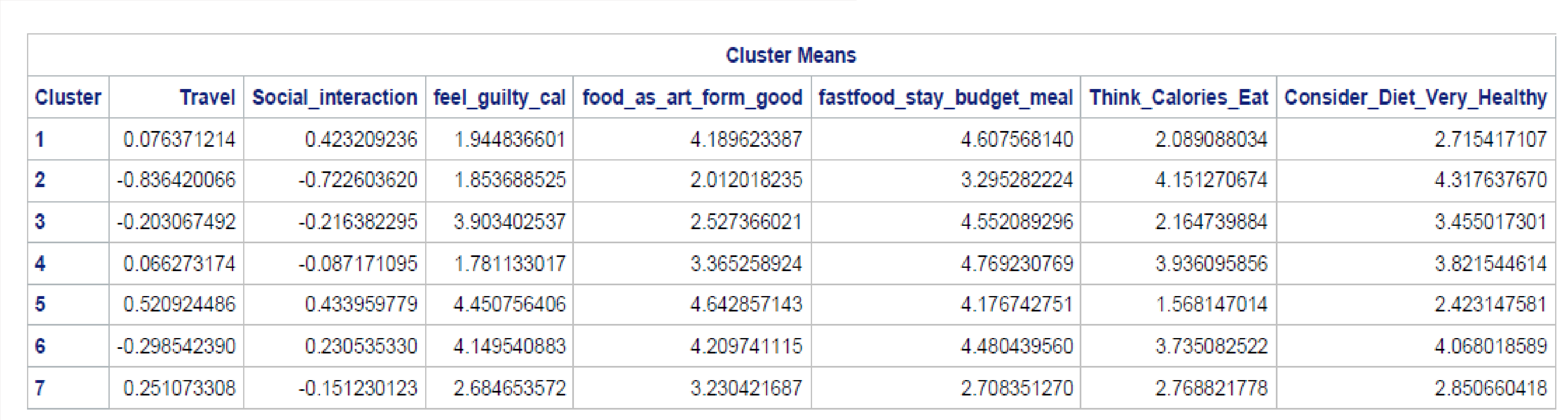




K Means

Means of the driver variables

- This looks like a good solution because the single driver variables have a decent to good spread (difference)
- After Analyzing the difference within the cluster variable’s (0.1 as difference) I found the below :
  - n! / r!(n-r)! = 7! / 2!(7-2)! = 5040/240 = 21
  - 21*7 (single driver+ abstract constant) = 147
  - Found 10 Ties after differencing the means
  - 10/147 = 0.06803 *100 = 6.8%
- Good solution as the percentage does not cross 15% and it is 6.8% of ties

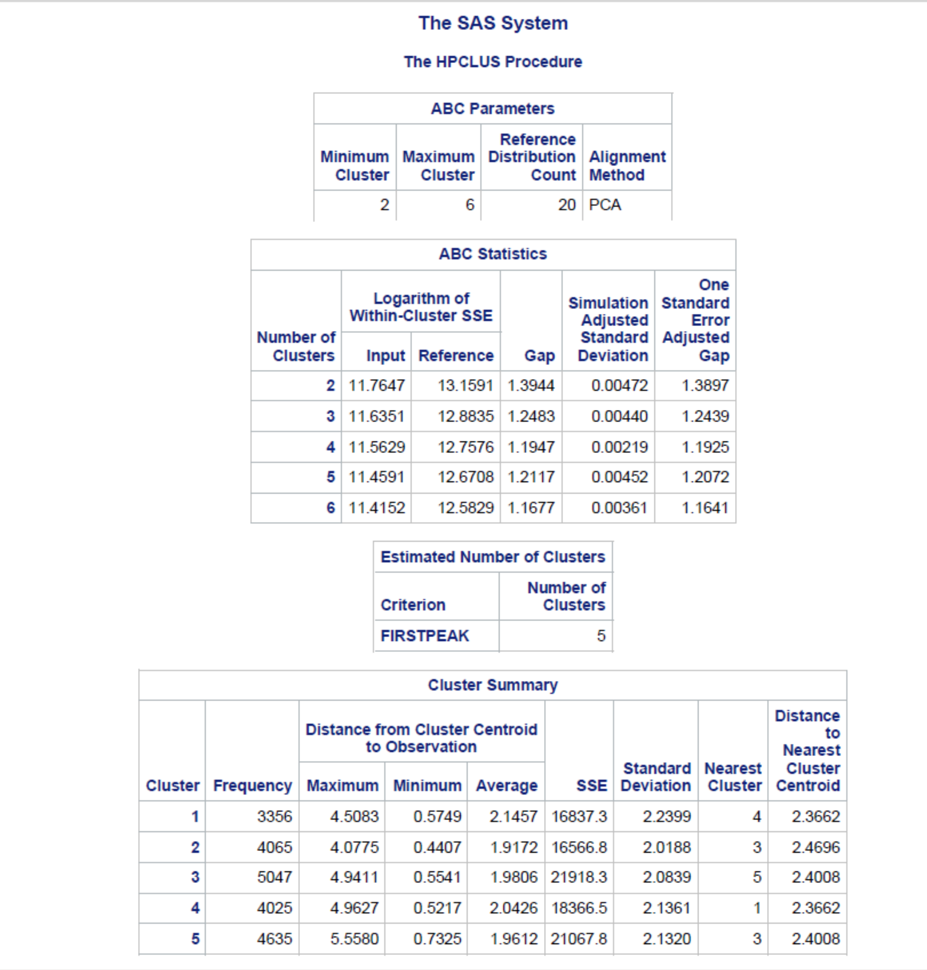


Cluster Means

Cluster	Travel	Social_interaction	feel_guilty_cal	food_as_art_form_good	fastfood_stay_budget_meal	Think_Calories_Eat	Consider_Diet_Very_Healthy
1	0.076371214	0.423209236	1.944836601	4.189623387	4.607568140	2.089088034	2.715417107
2	-0.836420066	-0.722603620	1.853688525	2.012018235	3.295282224	4.151270674	4.317637670
3	-0.203067492	-0.216382295	3.903402537	2.527366021	4.552089296	2.164739884	3.455017301
4	0.066273174	-0.087171095	1.781133017	3.365258924	4.769230769	3.936095856	3.821544614
5	0.520924486	0.433959779	4.450756406	4.642857143	4.176742751	1.568147014	2.423147581
6	-0.298542390	0.230535330	4.149540883	4.209741115	4.480439560	3.735082522	4.068018589
7	0.251073308	-0.151230123	2.684653572	3.230421687	2.708351270	2.768821778	2.850660418

Gap Analysis

- Removed one single variable as the clusters I got was 2 with the previous single variables which was not very optimum
- Number of clusters for the firstpeak to be 5 which is optimum
- Number of clusters for the globalpeak to be 5 which is optimum



The SAS System

The HPCLUS Procedure

ABC Parameters

Minimum Cluster	Maximum Cluster	Reference Distribution Count	Alignment Method
2	6	20	PCA

ABC Statistics

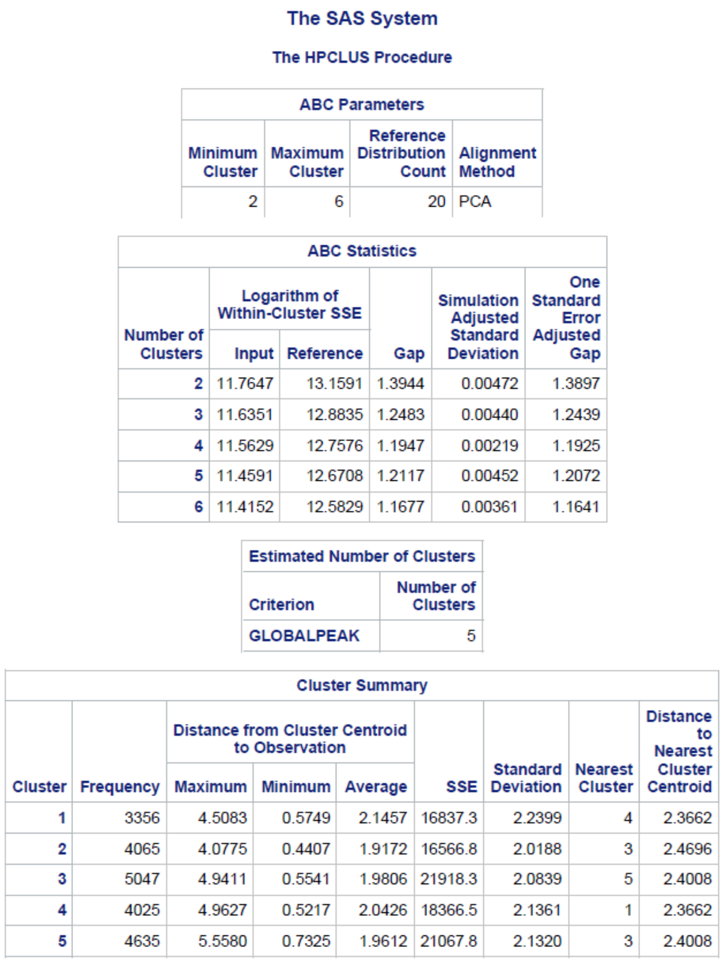
	Logarithm of Within-Cluster SSE				
Number of Clusters	Input	Reference	Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
2	11.7647	13.1591	1.3944	0.00472	1.3897
3	11.6351	12.8835	1.2483	0.00440	1.2439
4	11.5629	12.7576	1.1947	0.00219	1.1925
5	11.4591	12.6708	1.2117	0.00452	1.2072
6	11.4152	12.5829	1.1677	0.00361	1.1641

Estimated Number of Clusters

Criterion	Number of Clusters
FIRSTPEAK	5

Cluster Summary

		Distance from Cluster Centroid to Observation						
Cluster	Frequency	Maximum	Minimum	Average	SSE	Standard Deviation	Nearest Cluster	Distance to Nearest Cluster Centroid
1	3356	4.5083	0.5749	2.1457	16837.3	2.2399	4	2.3662
2	4065	4.0775	0.4407	1.9172	16566.8	2.0188	3	2.4696
3	5047	4.9411	0.5541	1.9806	21918.3	2.0839	5	2.4008
4	4025	4.9627	0.5217	2.0426	18366.5	2.1361	1	2.3662
5	4635	5.5580	0.7325	1.9612	21067.8	2.1320	3	2.4008



The SAS System

The HPCLUS Procedure

ABC Parameters

Minimum Cluster	Maximum Cluster	Reference Distribution Count	Alignment Method
2	6	20	PCA

ABC Statistics

	Logarithm of Within-Cluster SSE				
Number of Clusters	Input	Reference	Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
2	11.7647	13.1591	1.3944	0.00472	1.3897
3	11.6351	12.8835	1.2483	0.00440	1.2439
4	11.5629	12.7576	1.1947	0.00219	1.1925
5	11.4591	12.6708	1.2117	0.00452	1.2072
6	11.4152	12.5829	1.1677	0.00361	1.1641

Estimated Number of Clusters

Criterion	Number of Clusters
GLOBALPEAK	5

Cluster Summary

		Distance from Cluster Centroid to Observation						
Cluster	Frequency	Maximum	Minimum	Average	SSE	Standard Deviation	Nearest Cluster	Distance to Nearest Cluster Centroid
1	3356	4.5083	0.5749	2.1457	16837.3	2.2399	4	2.3662
2	4065	4.0775	0.4407	1.9172	16566.8	2.0188	3	2.4696
3	5047	4.9411	0.5541	1.9806	21918.3	2.0839	5	2.4008
4	4025	4.9627	0.5217	2.0426	18366.5	2.1361	1	2.3662
5	4635	5.5580	0.7325	1.9612	21067.8	2.1320	3	2.4008

Gap Analysis

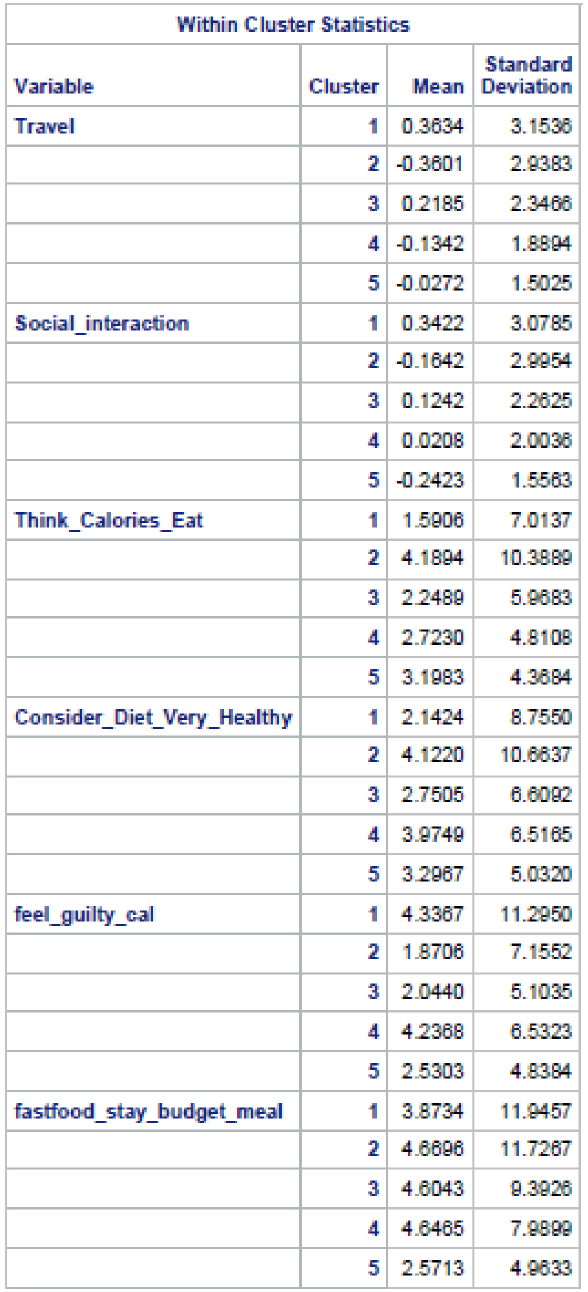
Cluster means for the drivers

- Considered the means from the Firstpeak and it seems like there is decent discrimination.
  - n! / r! (n-r)! = 5! / 2! (5-2)! = 120/12 = 10
  - 10*6 = 60
  - 9 Ties (Working show in the excel attached)
  - 9/60 = 0.15 *100 = 15%

- This looks like a decent discrimination among the clusters for the driver variables as the percentage does not cross 15% and it has exactly 15% of ties

K means Vs HPCLUS suggested number of clusters

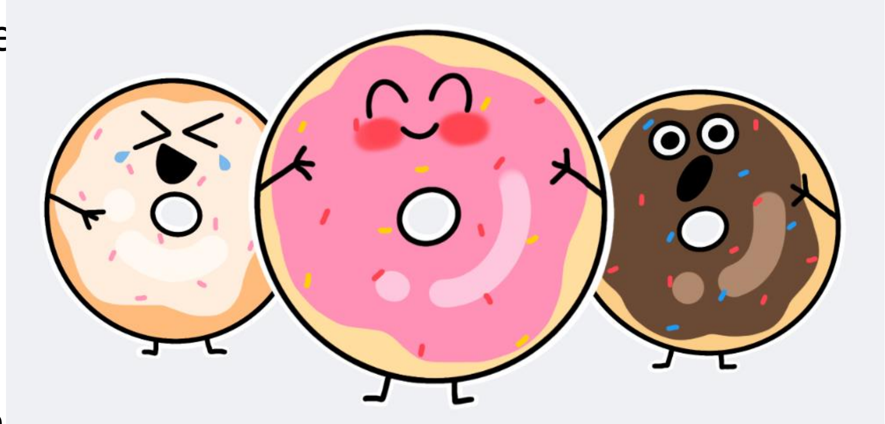
- K means: The number of cluster I got through CCC is 7.
- HPCLUS, the number of clusters I got was 5
- I would select the K means cluster as it has 7 clusters, and it seems to be more optimum than the HPCLUS clusters



Within Cluster Statistics

Variable	Cluster	Mean	Standard Deviation
Travel	1	0.3634	3.1536
	2	-0.3601	2.9383
	3	0.2185	2.3466
	4	-0.1342	1.8894
	5	-0.0272	1.5025
Social_interaction	1	0.3422	3.0785
	2	-0.1642	2.9954
	3	0.1242	2.2625
	4	0.0208	2.0036
	5	-0.2423	1.5563
Think_Calories_Eat	1	1.5906	7.0137
	2	4.1894	10.3889
	3	2.2489	5.9683
	4	2.7230	4.8108
	5	3.1983	4.3684
Consider_Diet_Very_Healthy	1	2.1424	8.7550
	2	4.1220	10.6637
	3	2.7505	6.6092
	4	3.9749	6.5165
	5	3.2967	5.0320
feel_guilty_cal	1	4.3367	11.2950
	2	1.8706	7.1552
	3	2.0440	5.1035
	4	4.2368	6.5323
	5	2.5303	4.8384
fastfood_stay_budget_meal	1	3.8734	11.9457
	2	4.6696	11.7267
	3	4.6043	9.3926
	4	4.6465	7.9899
	5	2.5713	4.9633

Cluster Analysis Across Descriptor Variables



- Used K=7 cluster solution as it worked best in the Previous exercise
- Considered 0.1 as the difference
  - n! / r! (n-r)! = 7! / 2! (7-2)! = 5040/240 = 21
  - 21*5 (descriptor variables) = 105
  - By considering 1% difference, we have:
  - 13 Ties -> 13/105 = 0.1238*100 = 12.38%
- We have 13 ties which is 12.38% of the overall ties
- Above output says the variable means are far apart from each other
- This has good number of ties as this does not exceed 15%. Therefore, the clustering solution discriminates on the descriptor variable well.
- Hence, cluster(k)=7 is an appropriate cluster choice for the market segmentation analysis.

	Cluster=.	Cluster=1	Cluster=2	Cluster=3	Cluster=4	Cluster=5	Cluster=6	Cluster=7
K_Krispy_Kreme	0.033033	0.082511	0.050963	0.085295	0.057314	0.096478	0.059162	0.074715
D_Dunkin_Donuts	0.258258	0.181726	0.227256	0.195514	0.18777	0.182007	0.202136	0.219259
Y_Youtube	0.159159	0.389269	0.380899	0.430906	0.404317	0.362829	0.322104	0.396173
gender_resp	0.468469	0.386484	0.383541	0.453337	0.335731	0.572359	0.437962	0.498371
RESPNDNT_ORIGIN	0.543544	0.243736	0.454134	0.282747	0.303118	0.23794	0.271159	0.34202

Cluster Analysis Across Descriptor Variables

	Highest	Lowest
K_Krispy_Kreme	Cluster 5 has the highest mean value of 9.64%, indicating a high preference for Krispy Kreme	Clusters 2 has lowest mean value with 5.09%, suggesting a lower preference for Krispy Kreme
D_Dunkin_Donuts	Cluster 2 has the highest mean value of 22.72%, indicating a strong preference for Dunkin Donuts over Krispy Kreme.	Clusters 1 has the lowest mean value of 18.17% indicating a strong preference for Krispy Kreme over Dunkin Donuts
Y_Youtube	Cluster 3 has the highest mean value of 43.09%, indicating they have higher chance of watching ads over Youtube	Clusters 6 has lowest mean values of 32.21% indicating they have lower chance of watching ads over Youtube
Gender_resp	Cluster 5 has the highest mean value of 57.23%, suggesting a higher representation of a male gender	Clusters 4 has lowest mean values of 33.57% suggesting a lower representation of a male gender
RESPNDNT_ORIGIN	Cluster 2 has the highest mean value of 45.41%, indicating a higher proportion from a SPANISH/HISPANIC/LATINO origin	Clusters 5 has lowest mean values of 23.79% indicating a lower proportion from a SPANISH/HISPANIC/LATINO origin