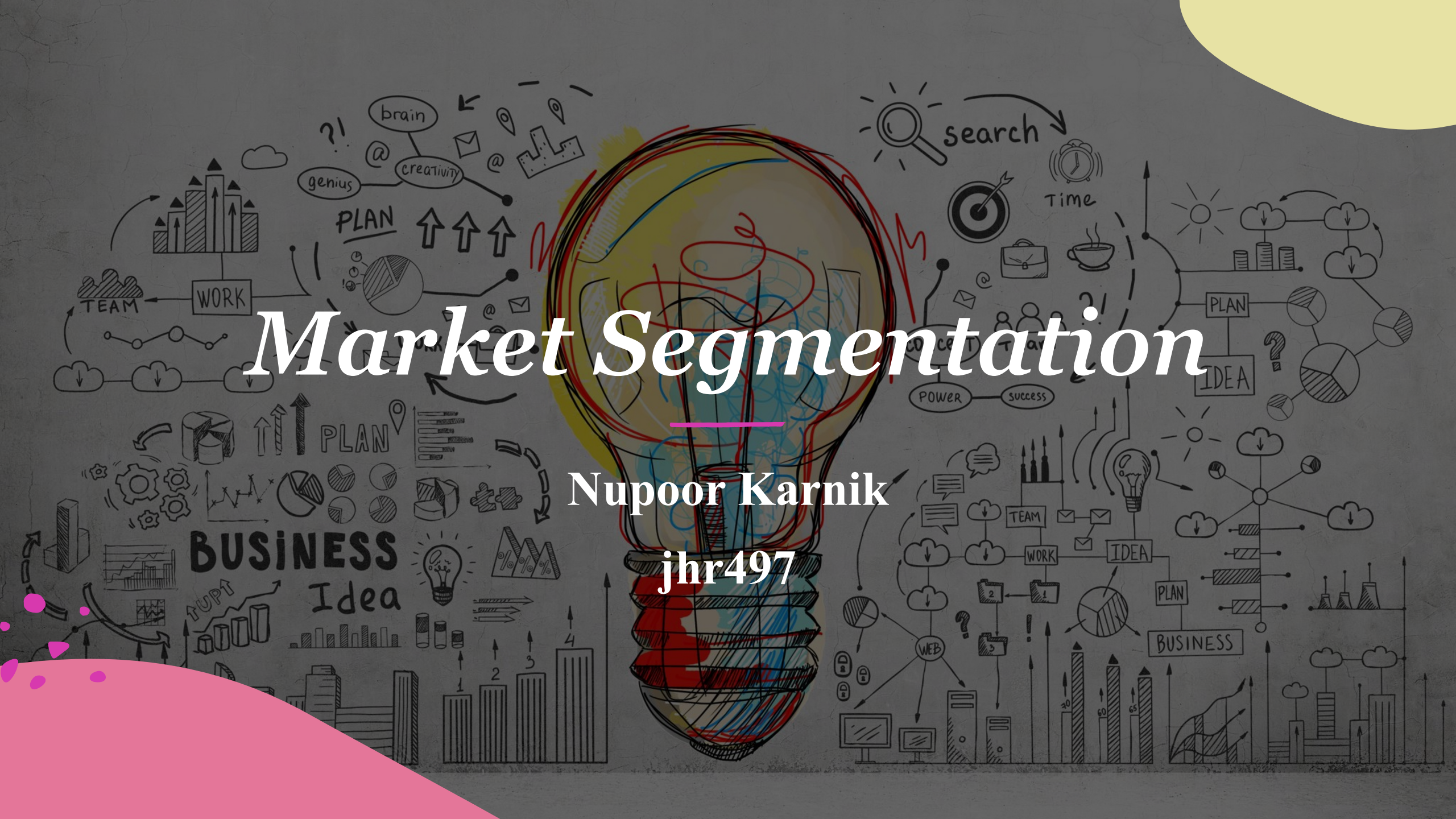


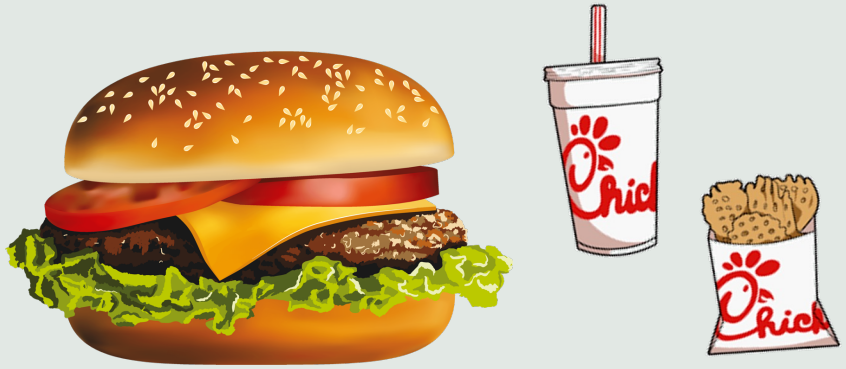
Market Segmentation

Nupoor Karnik

jhr497

Design Of Market Segmentation

Target Variable



Single Variables

- CONSIDER MY DIET TO BE VERY HEALTHY
- TOO BUSY TO TAKE CARE OF MYSELF AS I SHOULD
- OFTEN EAT STORE-MADE, PRE-COOKED MEALS
- IMPORTANT TO BE ATTRACTIVE/OPPOSITE SEX

Abstract Variables 1 & 2

- 1- **APPAREL**-People who eat chick-fil-a won't bother/care about styles
- I MAKE MY CLOTHES LAST A LONG TIME
- COMFORT MST IMPRTNT FCTR IN CLTHS I BUY
- STICK W/ STYLES HAVE STOOD TEST OF TIME
- DRESS TO PLEASE MYSELF

- 2- **TRAVEL-** People who travel will eat more chick-fil-a.
- VAC. SOMEWHERE DIFFERENT EVERY TIME
- I LOVE THE IDEA OF TRAVELING ABROAD
- VACATION EXPRNCES DIFFERENTIATE ME/FRNDS
- WILLING MAKE TRVL PLAN WITH UNKNWN COMP

Descriptive Variables

- McDonald's
- Instagram
- Nickelodeon
- Marital Status
- RESPNDNT-SPANISH/HISPANIC/LATINO ORIGIN?
- I MAKE SURE I EXERCISE REGULARLY

Reading raw data and Creating new Variables

Chick_Fil_A				
N_chick_fil_a	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	22612	88.89	22612	88.89
yes	2827	11.11	25439	100.00

CONSIDER MY DIET TO BE VERY HEALTHY				
healthy_diet_food	Frequency	Percent	Cumulative Frequency	Cumulative Percent
agree a lot	3546	14.72	3546	14.72
agree a little	7100	29.47	10646	44.19
neither agree nor disagree	7787	32.32	18433	76.52
disagree a little	4154	17.24	22587	93.76
disagree a lot	1503	6.24	24090	100.00
Frequency Missing = 1349				

OFTEN EAT STORE-MADE, PRE-COOKED MEALS				
time_saving_meals_quick	Frequency	Percent	Cumulative Frequency	Cumulative Percent
agree a lot	1101	4.57	1101	4.57
agree a little	3225	13.38	4326	17.94
neither agree nor disagree	5082	21.08	9408	39.02
disagree a little	5693	23.61	15101	62.63
disagree a lot	9010	37.37	24111	100.00
Frequency Missing = 1328				

TOO BUSY TO TAKE CARE OF MYSELF AS I SHOULD				
self_care_feel	Frequency	Percent	Cumulative Frequency	Cumulative Percent
agree a lot	2900	12.09	2900	12.09
agree a little	5945	24.78	8845	36.86
neither agree nor disagree	6970	29.05	15815	65.91
disagree a little	3959	16.50	19774	82.41
disagree a lot	4221	17.59	23995	100.00
Frequency Missing = 1444				

IMPORTANT TO BE ATTRACTIVE/OPPOSITE SEX				
attractive_look_date	Frequency	Percent	Cumulative Frequency	Cumulative Percent
agree a lot	3918	16.12	3918	16.12
agree a little	7213	29.69	11131	45.81
neither agree nor disagree	8456	34.80	19587	80.61
disagree a little	2407	9.91	21994	90.52
disagree a lot	2304	9.48	24298	100.00
Frequency Missing = 1141				

Principal Component Analysis

Extraction Technique –Principle Component Analysis

Rotation Method – Varimax

Criteria for determining factor extraction: Kaiser Criterion(eigen value=>1)
Factors extracted – 2
Variance explained – 48.82%

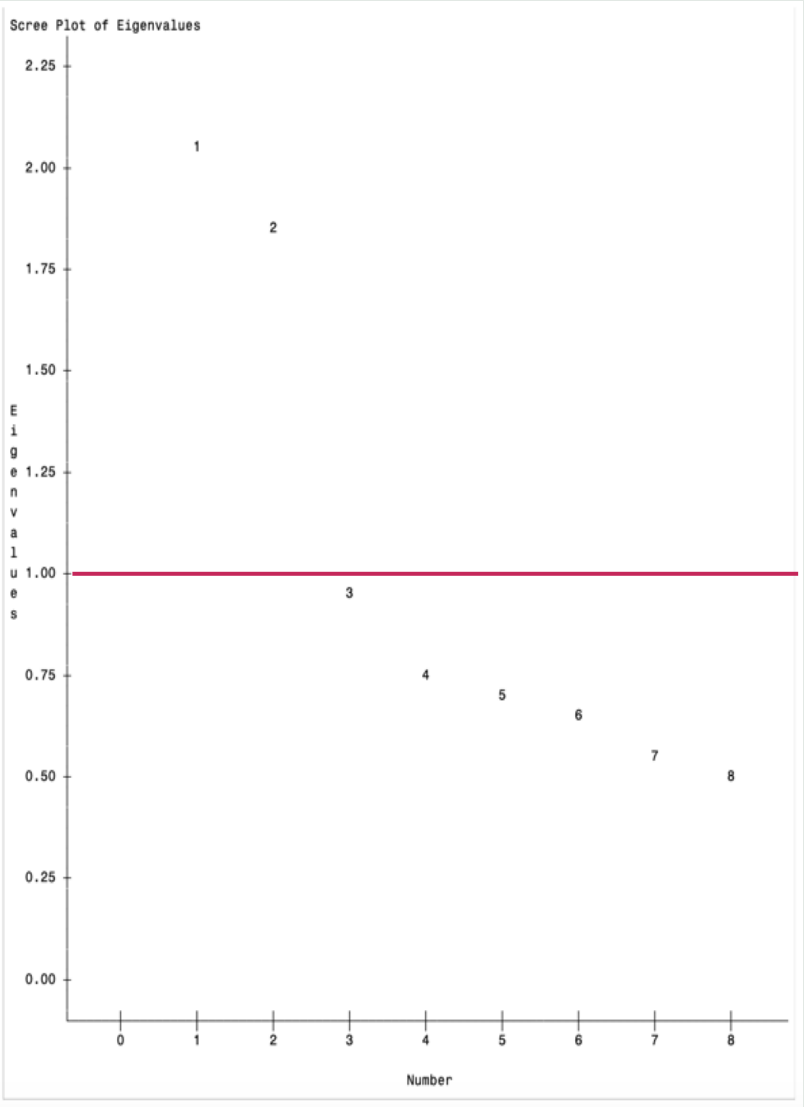
The SAS System

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	2.07436265	0.24273421	0.2593	0.2593
2	1.83162844	0.89378930	0.2290	0.4882
3	0.93783914	0.18780712	0.1172	0.6055
4	0.75003202	0.06110277	0.0938	0.6992
5	0.68892926	0.04652088	0.0861	0.7853
6	0.64240838	0.08340718	0.0803	0.8656
7	0.55900120	0.04320228	0.0699	0.9355
8	0.51579892		0.0645	1.0000



Rotated Factor Pattern			
		Factor1	Factor2
Use_Clothes_Long_Time	'I MAKE MY CLOTHES LAST A LONG TIME'	-0.01878	0.76515
Comfort_Factor	'COMFORT MST IMPRTNT FCTR IN CLTHS I BUY'	-0.04368	0.79904
Stick_To_Styles	'STICK W/STYLES HAVE STOOD TEST OF TIME'	0.05767	0.67984
Dress_To_Please	'DRESS TO PLEASE MYSELF'	-0.07996	0.46985
Vac_Diff_Places	'VAC SOMEHWERE DIFFERENT EVERY TIME'	0.69312	0.09323
Travel_Abroad	'I LOVE THE IDEA OF TRAVELING ABROAD'	0.73721	-0.01158
Diff_Vac_Exp	'VAC EXP DIFFERENTIATE ME AND FRNDS'	0.71544	-0.05139
Travel_Unknown_Comp	'WILLING MAKE TRVL PLAN WIH UNKNWN COMP'	0.64784	-0.14231

However, because one of the variables- **‘DRESS TO PLEASE MYSELF’** had 0.46985 as a value hence I changed it to **‘FUNCTION MST IMPRTNT FCTR IN CLTHS I BUY’** and following are the revised Eigen values table which shows 2 extracted factors and 56.85% of variance which can be explained, scree plot has a gradient slope and varimax rotation looks good as well.

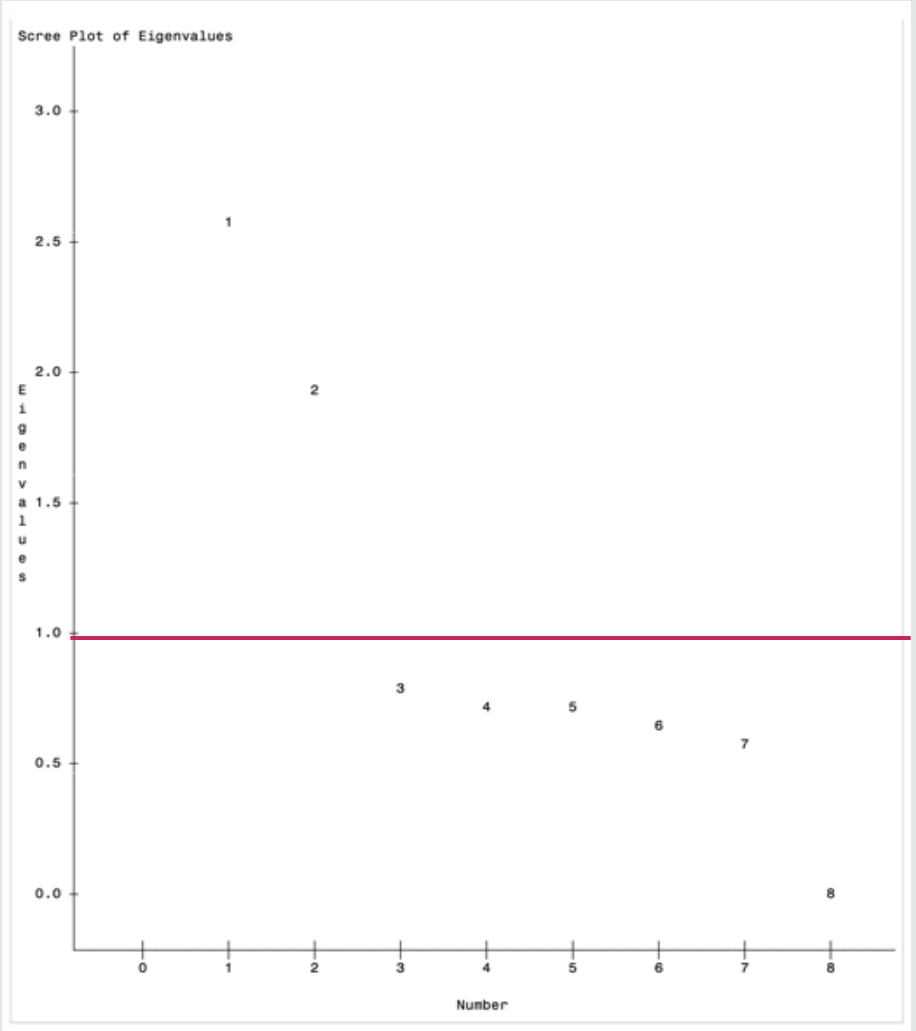
The SAS System

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

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

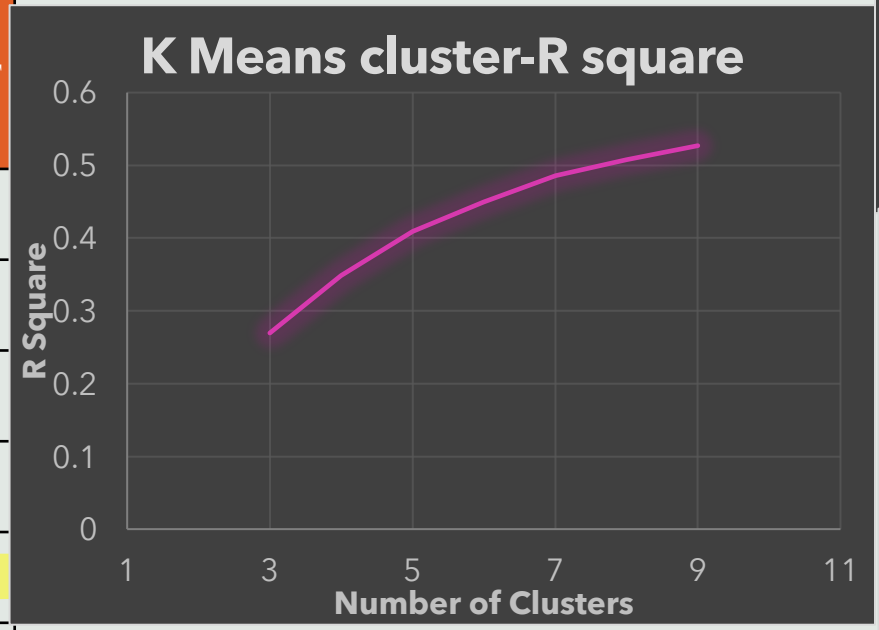
	Eigenvalue	Difference	Proportion	Cumulative
1	2.59900288	0.64990533	0.3249	0.3249
2	1.94909756	1.15106102	0.2436	0.5685
3	0.79803653	0.05636736	0.0998	0.6683
4	0.74166917	0.05709988	0.0927	0.7610
5	0.68456929	0.05104551	0.0856	0.8465
6	0.63352377	0.07424548	0.0792	0.9257
7	0.55927830	0.52445580	0.0699	0.9956
8	0.03482250		0.0044	1.0000

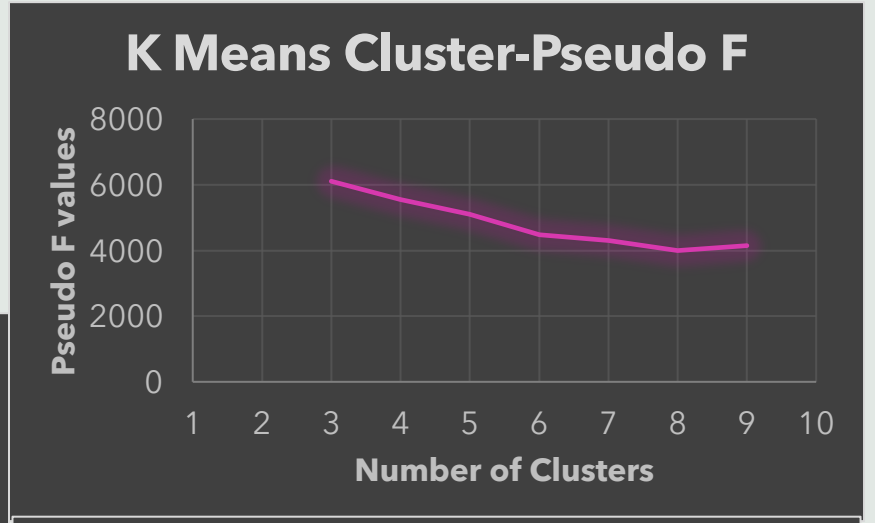


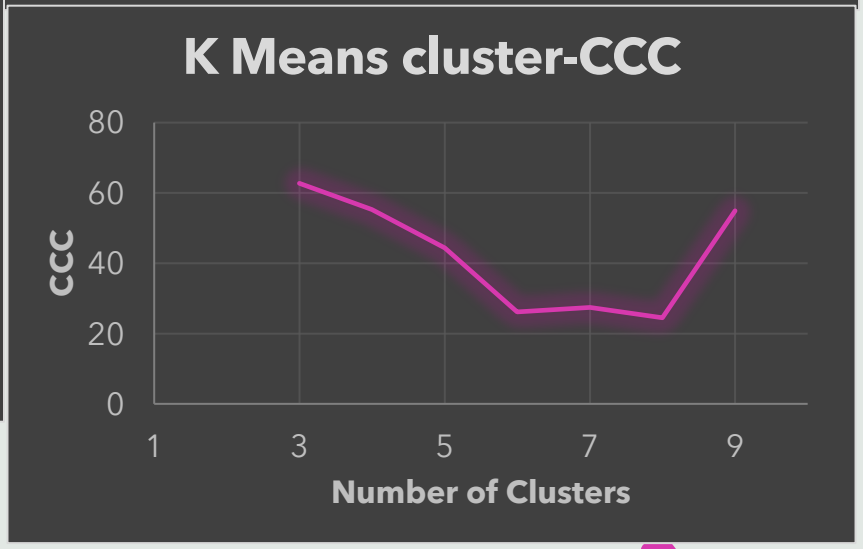
Rotated Factor Pattern			
		Factor1	Factor2
Use_Clothes_Long_Time	'I MAKE MY CLOTHES LAST A LONG TIME'	0.66797	-0.04678
Comfort_Factor	'COMFORT MST IMPRTNT FCTR IN CLTHS I BUY'	0.92715	-0.03395
Stick_To_Styles	'STICK W/STYLES HAVE STOOD TEST OF TIME'	0.63589	0.04100
Function_Important	'FUNCTION MST IMPRTNT FCTR IN CLTHS I BUY'	0.92332	-0.03868
Vac_Diff_Places	'VAC SOMEHWERE DIFFERENT EVERY TIME'	0.06936	0.69160
Travel_Abroad	'I LOVE THE IDEA OF TRAVELING ABROAD'	-0.01902	0.73408
Diff_Vac_Exp	'VAC EXP DIFFERENTIATE ME AND FRNDS'	0.00209	0.72247
Travel_Unknown_Comp	'WILLING MAKE TRVL PLAN WIH UNKNWN COMP'	-0.11107	0.64983

K Means Clustering

K	Number of clusters	R square	CCC	Pseudo F
3	3	0.26954	62.725	6106.28
4	4	0.34856	55.321	5555.32
5	5	0.40932	44.394	5094.47
6	6	0.44995	26.086	4477.51
7	7	0.48563	27.45	4297.81
8	8	0.50816	24.486	3996.41
9	9	0.52704	54.909	4140.61







Pseudo f shows cluster 9 as a first local maxima.
CCC shows the local maxima as cluster 7.
R Square for cluster 7 is 48.56%.

The cluster means show a good solution because the means of single driver variables and abstract constructs have very good spread, and there is discrimination among the cluster means.

Following is the post analysis of difference within the cluster variable with (0.1 difference):
Combinations Formula-n!/r!(n-r)!= 7!/2!(7−2)!=21
21*6(single driver+ abstract construct)=126 with **6 ties** that is 6/126 which is **4.76%.**

Cluster Means

Cluster	Apparel	Travel	healthy_diet_food	self_care_feel	time_saving_meals_quick	attractive_look_date
1	-0.186010805	0.438507476	3.032231405	2.967572062	4.482497900	4.221739130
2	-0.279034835	0.214310008	1.800048123	4.513493800	4.846715328	2.564127290
3	-0.005706585	-0.332014083	1.902720835	3.905294556	3.082156134	2.200819672
4	2.945094845	0.075466092	3.465517241	3.738528139	3.883361921	2.355004277
5	-0.270976415	-0.935627150	2.011769834	1.775197195	2.181818182	1.832614323
6	-0.163811949	-0.116178015	2.964105931	1.908891809	4.593695652	2.025059408
7	-0.018435008	0.246466989	3.480621902	2.592016238	2.407718502	3.033325928

Gap Analysis

For both Firstpeak and Globalpeak the estimated number of clusters are 4.

The SAS System

The HPCLUS Procedure

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

ABC Statistics					
Number of Clusters	Logarithm of Within-Cluster SSE		Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
	Input	Reference			
2	11.1589	12.5177	1.3589	0.0122	1.3467
3	11.0490	12.3600	1.3110	0.00545	1.3056
4	10.9389	12.2596	1.3208	0.00489	1.3159
5	10.8716	12.0656	1.1940	0.00373	1.1903
6	10.7880	12.0034	1.2154	0.00492	1.2105

Estimated Number of Clusters	
Criterion	Number of Clusters
FIRSTPEAK	4

The SAS System

The HPCLUS Procedure

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

ABC Statistics					
Number of Clusters	Logarithm of Within-Cluster SSE		Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
	Input	Reference			
2	11.1589	12.5177	1.3589	0.0122	1.3467
3	11.0490	12.3600	1.3110	0.00545	1.3056
4	10.9389	12.2596	1.3208	0.00489	1.3159
5	10.8716	12.0656	1.1940	0.00373	1.1903
6	10.7880	12.0034	1.2154	0.00492	1.2105

Estimated Number of Clusters	
Criterion	Number of Clusters
GLOBALPEAK	4

Within Cluster Statistics			
Variable	Cluster	Mean	Standard Deviation
Apparel	1	0.8312	3.2050
	2	-0.2895	2.3969
	3	-0.2766	1.8065
	4	-0.2356	1.5794
Travel	1	0.2780	2.3582
	2	0.1032	2.6776
	3	-0.0129	2.1930
	4	-0.3976	1.9999
healthy_diet_food	1	3.2569	7.4427
	2	1.9674	6.5592
	3	2.8660	5.6476
	4	2.6726	4.7594
self_care_feel	1	3.2340	8.4588
	2	4.5247	9.2793
	3	2.1230	5.5878
	4	2.2143	3.8989
time_saving_meals_quick	1	3.2591	9.4091
	2	4.5306	10.1675
	3	4.6361	8.3522
	4	2.1896	5.8301
attractive_look_date	1	3.3608	7.8154
	2	2.5278	7.3273
	3	2.5372	5.0606
	4	2.1840	3.9694

FirstPeak cluster means

Combinations formula:

n! / r!(n-r)! = 4! /2!(4-2)! = 6

6*6 (Driver variables) = 36

0.1 difference:

3 Ties-

3/36 = **8.33%**

Hence, based on the firstpeak cluster statistics and combinations formula solution it is seen that there is decent discrimination amongst the cluster means.

Descriptor Variable Cluster Means

Combinations formula:
n! /r!(n-r)! = 7!/2!(7-2)! = 21
21*7 (descriptor variables) = 147

With 0.1 difference:
3 Ties- **3/147 * 100 = 2.04%**
The clustering solution discriminates on the descriptor variables well.

Note: Removed the factor '**Function Mst Imprtnt Fctr In Clths I Buy'** as it was having the same correlational values as 'Comfort_Factor'.
After removing it everything worked fine.
Changed the '**Nickelodeon**' variable to '**Youtube**' as the mean values for both Nickelodeon and Instagram were very close
and was in turn causing lot of ties in the data which would have caused further issues.

Descriptors	Cluster=.	Cluster=1	Cluster=2	Cluster=3	Cluster=4	Cluster=5	Cluster=6	Cluster=7
Chick_fil	0.0495627	0.0685484	0.0851359	0.08318	0.0840282	0.08723	0.0754251	0.0851454
Mc_D	0.1749271	0.3096591	0.3117163	0.2453297	0.3746995	0.3881128	0.3535118	0.3418319
Insta	0.0145773	0.0589489	0.0983542	0.0859779	0.0817308	0.0873733	0.0845945	0.115203
Youtube	0.1516035	0.4650687	0.4951999	0.4865064	0.3835792	0.416735	0.3697202	0.4042442
RESPNDNT_ORIGIN	0.5685131	0.2613636	0.3373824	0.2606347	0.3010817	0.2648589	0.298409	0.3862134
MaritalStatus	0.6916996	0.7965642	0.7081875	0.8001658	0.6598947	0.7606418	0.6650222	0.6158279
Regular_Exercise	1.7272727	2.8365773	2.6191339	2.0703218	3.1220412	3.3509168	2.9411521	2.5194546

Thank you!