

# Market Segmentation Presentation



Partha Sai Madallapalli  
pua528

# Data Source and Business Problem

- Data Source:

Data extracted from NCS 2015 data.

- Business Problem:

1. Identify the target customers.
2. Whether customers are so picky about the diet and health conscious of the products they consume, brand oriented and look for environment friendly products.
3. How to reach customers i.e., advertising/promotion channel.
4. Identify the origin of the customers.

# Variables Used in Market Segmentation

Target Variable

- DIET MOUNTAIN DEW

Single Driver Variables

- I THINK OF THE CALORIES IN WHAT I EAT
- TOO BSY TO TAKE CARE OF MYSELF AS I SHLD
- CONSIDER MY DIET TO BE VERY HEALTHY
- LIKE TO KNW ABOUT INGRDNTS BEFR BUY FOOD

My Abstract Construct 1- ENVIRONMENT

- PERSONAL OBLIGATION/ENVRNMNT RESPONSIBLE
- I MAKE A CONSCIOUS EFFORT TO RECYCLE
- MORE LIKELY PRCH/ENVRNMNTLLY-FRNDLY COMP
- PACKAGING FOR PRDCTS SHOULD BE RECYCLED

My Abstract Construct 2- SHOPPING  
BEHAVIOR

- DON'T BUY UNKNOWN BRANDS TO SAVE MONEY
- ALWAYS LOOK FOR BRAND NAME
- I KEEP UP WITH CHANGES-STYLES/FASHIONS
- PEOPLE COME TO ME/ADVICE BEFORE BUYING

Descriptive Variables

- Major competitors of that product- SPRITE ZERO
- Media Channel- ESPN
- Personal Information- Gender (Male & Female)
- Origin/ Race- RESPNDNT-SPANISH/HISPANIC/LATINO ORIGIN?
- SHOES/ FOOTWEAR – BRANDS- NIKE
- Attitude Variable- USE APP/SITE TO TRACK CALORIES/ EXERCISE

# Frequencies for Single Driver and Target Variables

## TOO BSY TO TAKE CARE OF MYSELF AS I SHLD

Take_Care_Myself_Shld	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	4221	17.59	4221	17.59
disagree a little	3959	16.50	8180	34.09
neither agree nor disagree	6970	29.05	15150	63.14
agree a little	5945	24.78	21095	87.91
agree a lot	2900	12.09	23995	100.00

Frequency Missing = 1444

## CONSIDER MY DIET TO BE VERY HEALTHY

Consider_Diet_Very_Healthy	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	1503	6.24	1503	6.24
disagree a little	4154	17.24	5657	23.48
neither agree nor disagree	7787	32.32	13444	55.81
agree a little	7100	29.47	20544	85.28
agree a lot	3546	14.72	24090	100.00

Frequency Missing = 1349

## LIKE TO KNW ABOUT INGRDNTS BEFR BUY FOOD

Knw_Ingrdnts_Befr_Buy_Food	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	2119	8.81	2119	8.81
disagree a little	3723	15.48	5842	24.29
neither agree nor disagree	7852	32.65	13694	56.94
agree a little	6520	27.11	20214	84.05
agree a lot	3835	15.95	24049	100.00

Frequency Missing = 1390

## I THINK OF THE CALORIES IN WHAT I EAT

Think_Calories_Eat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
disagree a lot	3998	16.61	3998	16.61
disagree a little	5519	22.93	9517	39.54
neither agree nor disagree	7157	29.74	16674	69.28
agree a little	5042	20.95	21716	90.22
agree a lot	2353	9.78	24069	100.00

Frequency Missing = 1370

## DIET MOUNTAIN DEW

D_Mountain_Dew	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	24808	97.52	24808	97.52
yes	631	2.48	25439	100.00

# PCA Analysis - Varimax

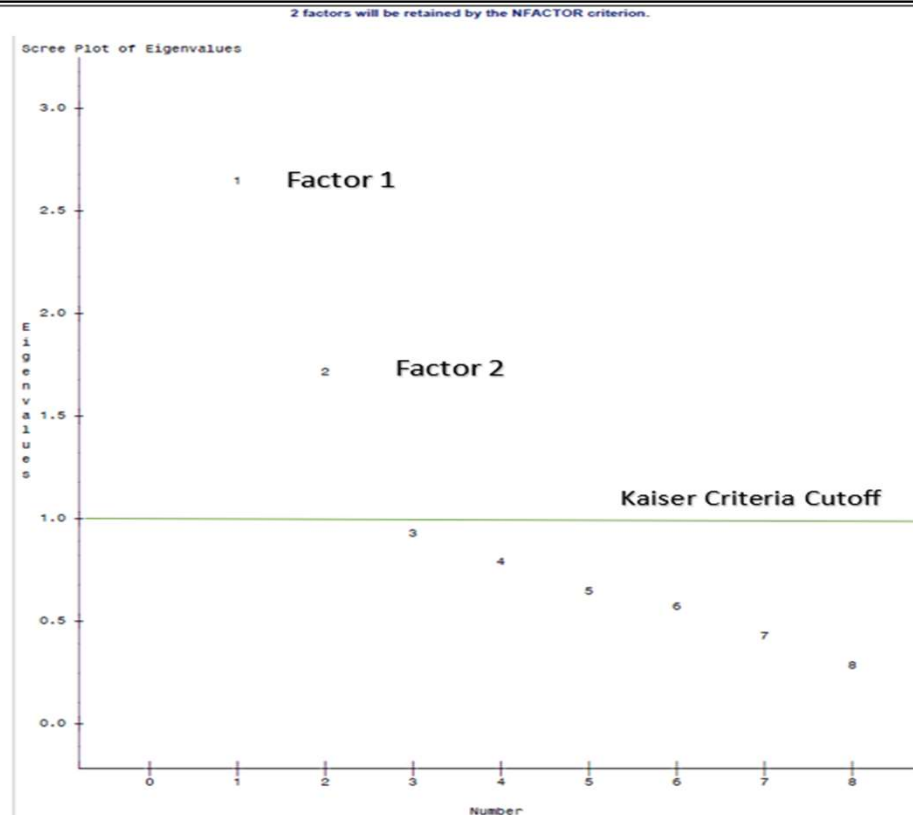
- Variables used Abstract Construct 1 and 2
- kaiser Criteria cutoff  $>1$  (For choosing the Factor)
- Cumulative variance explained - 55.08%
- Scree plot- 2 Factors

## 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	2.66769547	0.92871596	0.3335	0.3335
2	1.73897952	0.80361681	0.2174	0.5508
3	0.93536271	0.17794884	0.1169	0.6678
4	0.75741387	0.12417566	0.0947	0.7624
5	0.63323821	0.04847309	0.0792	0.8416
6	0.58476512	0.16489765	0.0731	0.9147
7	0.41986747	0.15718985	0.0525	0.9672
8	0.26267762		0.0328	1.0000

2 factors will be retained by the NFACTOR criterion.





# Rotated Factor Pattern

Rotated Factor Pattern		Factor1	Factor2
Obligation_rspnse_envnmnt	PERSONAL OBLIGATION/ENVRNMNT RESPONSIBLE	0.79472	0.04993
Effort_recyle	I MAKE A CONSCIOUS EFFORT TO RECYCLE	0.80332	-0.00238
Prch_envrnmntly_frndly_comp	MORE LIKELY PRCH/ENVRNMNTLLY-FRNDLY COMP	0.68493	0.24351
Pck_prdct_recycled	PACKAGING FOR PRDCTS SHOULD BE RECYCLED	0.86918	-0.01915
Dont_buy_unknown_brands	DO Not BUY UNKNOWN BRANDS TO SAVE MONEY	0.13684	0.59323
Brand_name	ALWAYS LOOK FOR BRAND NAME	0.10884	0.69993
Changes_styles_fashions	I KEEP UP WITH CHANGES-STYLES/FASHIONS	0.02699	0.72338
Me_advice_buying	PEOPLE COME TO ME/ADVICE BEFORE BUYING	-0.06242	0.66536

## Factor 1 – ENVIRONMENT

The respondents who drinks diet mountain dew will care for environment pollution.

## Factor 2- SHOPPING BEHAVIOR

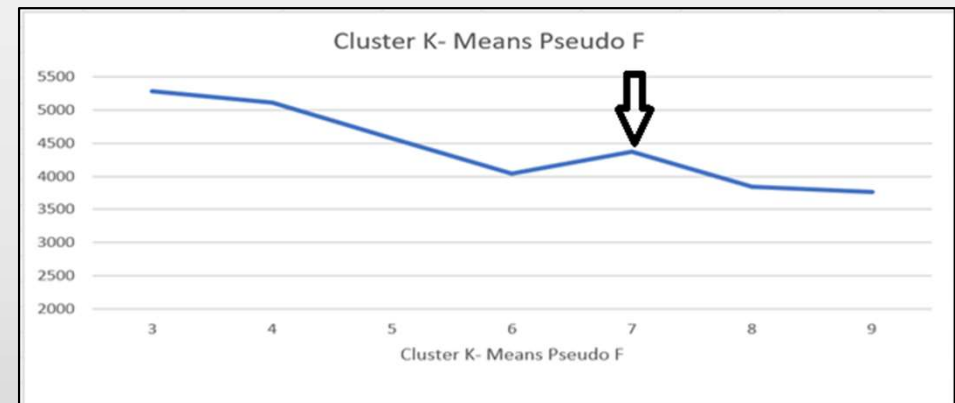
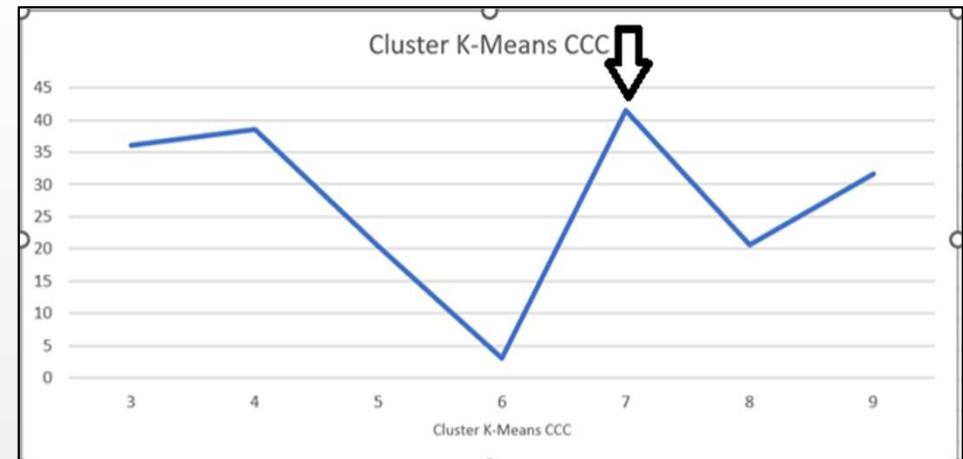
The respondents who drinks diet mountain dew is not very picky about shopping(brand).

**The respondents are more environmentally conscious when compared to the shopping behavior(brand).**

# K Means Clustering

K	Number of clusters	R square	CCC	Pseudo F
3	3	0.26807	36.071	5282.79
4	4	0.34956	38.579	5107.17
5	5	0.40988	20.506	4570.53
6	6	0.45027	2.98	4036.72
7	7	0.48573	41.523	4367.42
8	8	0.50816	20.666	3844.44
9	9	0.52713	31.676	3758.8

- 4 Single-drivers and 2 factor variables from PCA are used in K Means.
- Primarily first local max point is identified at cluster 4. However, I am not taking the first primary since it is starting point of the cluster, hence I am **considering** the secondary local max point which is **cluster 7**.  
In **cluster 7** the means are differed by 0.2 suggests the clustering solution which seems to be a **good solution** as there is a **clear separation** within the clusters and a **13.49% variable combination** explained by the permutation out of 126 total variable combination.



Cluster Means						
Cluster	environment	shopping_behavior_brand	Think_Calories_Eat	Take_Care_Myself_Shld	Consider_Diet_Very_Healthy	Knw_Ingrdnts_Befr_Buy_Food
1	0.523697973	0.138354624	4.127049180	1.536270191	4.200233850	4.188856305
2	0.469389215	-0.301885709	1.908778507	1.663108320	3.828865337	3.350860720
3	-0.769649701	-0.756319000	1.619184377	1.648009233	2.090752441	1.618390805
4	0.571027582	-0.094045251	2.643753015	3.823372853	3.083373610	3.809753742
5	-1.044917530	0.289527091	3.043594646	3.059151143	3.074478269	2.940309929
6	-0.068758154	-0.550913866	1.741622575	4.180854197	2.233873343	2.061284620
7	0.457489312	0.904956688	4.299011251	4.092611508	4.254048583	4.224753485

# GAP Analysis

## 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	11.7896	13.1698	1.3802	0.00266	1.3775
3	11.6806	13.0093	1.3287	0.00882	1.3199
4	11.5735	12.8468	1.2733	0.00745	1.2658
5	11.4613	12.7556	1.2942	0.00683	1.2874
6	11.3890	12.6749	1.2859	0.00541	1.2805

Estimated Number of Clusters	
Criterion	Number of Clusters
FIRSTPEAK	5

## 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	11.7896	13.1698	1.3802	0.00266	1.3775
3	11.6806	13.0093	1.3287	0.00882	1.3199
4	11.5735	12.8468	1.2733	0.00745	1.2658
5	11.4613	12.7556	1.2942	0.00683	1.2874
6	11.3890	12.6749	1.2859	0.00541	1.2805

Estimated Number of Clusters	
Criterion	Number of Clusters
GLOBALPEAK	5

Within Cluster Statistics			
Variable	Cluster	Mean	Standard Deviation

owner/pua528/\_TD13316\_HORRDS12\_/sasht

environment	1	0.5976	2.4598
	2	0.5361	3.0831
	3	-0.2338	2.7264
	4	0.0556	1.8345
	5	-0.7556	1.5849
shopping_behavior_brand	1	0.0282	3.0155
	2	0.5713	2.8852
	3	-0.6875	2.8673
	4	-0.4334	1.9352
	5	0.2393	1.2538
Think_Calories_Eat	1	3.4075	9.8736
	2	3.8919	9.0508
	3	1.6802	6.0874
	4	1.8473	3.3522
	5	2.9926	3.8415
Take_Care_Myself_Shld	1	1.5191	7.8847
	2	4.0214	10.0910
	3	1.6833	6.4141
	4	4.1773	6.2004
	5	2.9629	4.4541
Consider_Diet_Very_Healthy	1	4.1755	10.8182
	2	3.9072	9.6750
	3	2.7703	7.9306
	4	2.3668	4.5379
	5	3.1408	4.2418
Knw_Ingrdnts_Befr_Buy_Food	1	4.1766	10.9098
	2	4.1543	9.8943
	3	1.9764	6.9791
	4	2.6345	4.6045
	5	3.0532	4.4733

- Same drivers are used as in the K-means analysis.
- In **first peak** and **global peak** the optimal number of **clusters** is at **5**.
- Based on the cluster statistics, the means are differed by 0.2 suggests the clustering solution have a good solution as there is a clear separation within the clusters with a **10% variable combination** explained by the permutation out of **60 total variable combination** with 5 clusters.
- Driver variables with these descent differences suggest a good predictors of the clusters.



# Choosing Cluster from K-Means Vs Gap Analysis

- I have picked 7 clusters from the previous K means exercise.
- These 7 clusters are not the same number of clusters which is suggested by HPCLUS, HPCLUS suggested 5 clusters.
- I have decided to use K means as statistical metric to choose clusters because it has higher differences within the clusters and it has maximum number of clusters i.e., 7 clusters, when compared to gap analysis which is only 5 clusters. Also, from business perspective as the number of clusters are high in number it is good solution for them to make better business decisions.

# Cluster Analysis on Descriptor Variables

Variable	Cluster-1	Cluster-2	Cluster-3	Cluster-4	Cluster-5	Cluster-6	Cluster-7
D_Mountain_Dew	0.0249059	0.0197775	0.0226116	0.022961	0.0243442	0.0341307	0.0273927
SPRITEZERO	0.0341732	0.0191595	0.0220464	0.0191342	0.0235894	0.0268378	0.0343234
ESPN_Sport	0.1746308	0.1881953	0.2001131	0.1772303	0.1911681	0.2214119	0.1851485
RESPNDNT_ORIGIN	0.2733855	0.2462917	0.2809497	0.2925138	0.3260993	0.2441657	0.4323432
gender_resp	0.3417318	0.446848	0.5415489	0.4168859	0.4706548	0.5023337	0.3864686
Nike_Brand	0.2151752	0.1724351	0.1978519	0.2018656	0.2321193	0.214119	0.249505
APP_SITE_TRACK_CALORIES	2.0803991	1.5245236	1.400919	1.9198648	2.4687201	1.5296884	2.7960999

Based on the previous analysis(K-Means), I have chosen cluster 7 as it shows significant differences in variables.

Most variables have a difference of 0.1 or more and 5.44% of variable combination explained by the permutations out of 147 total variable combinations with 7 clusters, making **cluster(k)=7** as an appropriate cluster choice for the market segmentation analysis.

Here for cluster 7 the market segmentation strategy would be **“Male Sport Hispanic/latino Enthusiasts”** who would like to pick environment friendly recycle products and maintain a track of their consumption(Diet Products) through application(app).

# Findings & Recommendations

- The target customers for the product **Diet Mountain Dew** is **males** which is an appropriate marketing strategy from the cluster analysis.
- Consumers are health conscious and **track** their **consumption calories** through **application(app)**.
- Also, the products these customers buy of **brand oriented** and **environment friendly** and look for products with ease disposal for recycling and doesn't want to harm the environment by any form of pollution.
- The best mode of channel to effectively communicate these customers is through **ESPN Sports channel**.
- It is identified from the analysis that a concentration of respondents are from specific origin i.e., **spanish/hispanic/latino**.

# Thank You

Partha Sai  
[parthasai.madallapalli@my.utsa.edu](mailto:parthasai.madallapalli@my.utsa.edu)