

Python 3.6.4 |Anaconda custom (64-bit)| (default, Jan 16 2018, 10:22:32) [MSC v.1900 64 bit (AMD64)]

Type "copyright", "credits" or "license" for more information.

IPython 6.2.1 -- An enhanced Interactive Python.

```
In [1]: runfile('MarketSegmentation.py')
```

FACTOR ANALYSIS

Cleaning & Compiling Variables for Factor Analysis

Question: PEOPLE HVE RESPONS TO USE RECYCLD PRDCTS

Frequency tables:

agree_lot

0.0 19088

1.0 6476

Name: V4361, dtype: int64

agree_little

0.0 18087

1.0 7477

Name: V4442, dtype: int64

neither

0.0 17617

1.0 7947

Name: V4604, dtype: int64

disagree_little

0.0 24134

1.0 1430

Name: V4685, dtype: int64

disagree_lot

0.0 24752

1.0 812

Name: V4766, dtype: int64

Check if more than one answer selected:

1.0 24142

0.0 1422

Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5

6476 = 6476

correct: column 2 , value 4

7477 = 7477

correct: column 3 , value 3

7947 = 7947

correct: column 4 , value 2

1430 = 1430

correct: column 5 , value 1

812 = 812

Finished: PEOPLE HVE RESPNS TO USE RECYCLD PRDCTS

Question: COMPANIES/HELP CONSUMERS ENVRNMNT RESPNS

Frequency tables:

agree_lot

0.0 17590

1.0 7974

Name: V3929, dtype: int64

agree_little

0.0 17327

1.0 8237

Name: V3938, dtype: int64

neither

0.0 19027

1.0 6537

Name: V3956, dtype: int64

disagree_little

0.0 24805

1.0 759

Name: V3965, dtype: int64

disagree_lot

0.0 24922

1.0 642

Name: V3974, dtype: int64

Check if more than one answer selected:

1.0 24149

0.0 1415

Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5

7974 = 7974

correct: column 2 , value 4

8237 = 8237

correct: column 3 , value 3

6537 = 6537

correct: column 4 , value 2

759 = 759

correct: column 5 , value 1

642 = 642

Finished: COMPANIES/HELP CONSUMERS ENVRNMNT RESPNS

Question: MORE LIKELY PRCH/ENVRNMNTLLY-FRNDLY COMP

Frequency tables:

agree_lot

0.0 18635

1.0 6929

Name: V3927, dtype: int64

agree_little

0.0 18029

1.0 7535
Name: V3936, dtype: int64

neither
0.0 17319
1.0 8245
Name: V3954, dtype: int64

disagree_little
0.0 24669
1.0 895
Name: V3963, dtype: int64

disagree_lot
0.0 24683
1.0 881
Name: V3972, dtype: int64

Check if more than one answer selected:

1.0 24485
0.0 1079
Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5
6929 = 6929

correct: column 2 , value 4
7535 = 7535

correct: column 3 , value 3
8245 = 8245

correct: column 4 , value 2
895 = 895

correct: column 5 , value 1
881 = 881

Finished: MORE LIKELY PRCH/ENVRNMNTLLY-FRNDLY COMP

Question: MORE LKLY BUY/COMP W/ENVRNMNT FRNDLY ADS

Frequency tables:

agree_lot
0.0 20972
1.0 4592
Name: V3934, dtype: int64

agree_little
0.0 19786
1.0 5778
Name: V3943, dtype: int64

neither
0.0 14822
1.0 10742
Name: V3961, dtype: int64

disagree_little
0.0 23833
1.0 1731
Name: V3970, dtype: int64

disagree_lot
0.0 24385
1.0 1179
Name: V3979, dtype: int64

Check if more than one answer selected:
1.0 24022
0.0 1542
Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:
correct: column 1 , value 5
4592 = 4592

correct: column 2 , value 4
5778 = 5778

correct: column 3 , value 3
10742 = 10742

correct: column 4 , value 2
1731 = 1731

correct: column 5 , value 1
1179 = 1179

Finished: MORE LKLY BUY/COMP W/ENVRNMNT FRNDLY ADS

Question: USUALLY FRST AMNG FRNDS SHOP NEW STORE

Frequency tables:
agree_lot
0.0 24647
1.0 917
Name: V6106, dtype: int64

agree_little
0.0 23768
1.0 1796
Name: V6148, dtype: int64

neither
0.0 17543
1.0 8021
Name: V6232, dtype: int64

disagree_little
0.0 20683
1.0 4881
Name: V6274, dtype: int64

disagree_lot
0.0 17353
1.0 8211
Name: V6316, dtype: int64

Check if more than one answer selected:
1.0 23826
0.0 1738
Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:
correct: column 1 , value 5
917 = 917

correct: column 2 , value 4
1796 = 1796

correct: column 3 , value 3
8021 = 8021

correct: column 4 , value 2
4881 = 4881

correct: column 5 , value 1
8211 = 8211

Finished: USUALLY FRST AMNG FRNDS SHOP NEW STORE

Question: I'M 1ST OF FRNDS HAVE NEW ELCTRNC EQUIP

Frequency tables:

agree_lot

0.0 24714

1.0 850

Name: V6677, dtype: int64

agree_little

0.0 23549

1.0 2015

Name: V6694, dtype: int64

neither

0.0 19718

1.0 5846

Name: V6728, dtype: int64

disagree_little

0.0 20688

1.0 4876

Name: V6745, dtype: int64

disagree_lot

0.0 14947

1.0 10617

Name: V6762, dtype: int64

Check if more than one answer selected:

1.0 24204

0.0 1360

Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5

850 = 850

correct: column 2 , value 4

2015 = 2015

correct: column 3 , value 3

5846 = 5846

correct: column 4 , value 2

4876 = 4876

correct: column 5 , value 1

10617 = 10617

Finished: I'M 1ST OF FRNDS HAVE NEW ELCTRNC EQUIP

Question: I'M USUALLY FIRST TO TRY NEW HEALTH FOOD

Frequency tables:

agree_lot

0.0 23672

1.0 1892

Name: V3784, dtype: int64

agree_little

0.0 22712

1.0 2852

Name: V3802, dtype: int64

neither

0.0 17706

1.0 7858

Name: V3838, dtype: int64

disagree_little

0.0 20535

1.0 5029

Name: V3856, dtype: int64

disagree_lot

0.0 19406

1.0 6158

Name: V3874, dtype: int64

Check if more than one answer selected:

1.0 23789

0.0 1775

Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5

1892 = 1892

correct: column 2 , value 4

2852 = 2852

correct: column 3 , value 3

7858 = 7858

correct: column 4 , value 2

5029 = 5029

correct: column 5 , value 1

6158 = 6158

Finished: I'M USUALLY FIRST TO TRY NEW HEALTH FOOD

Question: I AM FRST AMNG MY FRIENDS TRY NEW STYLES

Frequency tables:

agree_lot

0.0 24642

1.0 922

Name: V3177, dtype: int64

agree_little

0.0 23763

1.0 1801
Name: V3204, dtype: int64

neither
0.0 19024
1.0 6540
Name: V3258, dtype: int64

disagree_little
0.0 20598
1.0 4966
Name: V3285, dtype: int64

disagree_lot
0.0 15641
1.0 9923
Name: V3312, dtype: int64

Check if more than one answer selected:
1.0 24152
0.0 1412
Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:
correct: column 1 , value 5
922 = 922

correct: column 2 , value 4
1801 = 1801

correct: column 3 , value 3
6540 = 6540

correct: column 4 , value 2
4966 = 4966

correct: column 5 , value 1
9923 = 9923

Finished: I AM FRST AMNG MY FRIENDS TRY NEW STYLES

Question sets for factor analysis

Set 1: Environmentally conscious
Set 2: Early adopters

KMO:

| | KMO |
|----------------------------------|----------|
| environRespRecycldProd | 0.874262 |
| enrironCompHelpConsumEnvironResp | 0.785257 |
| environPurchEnvironFriendlyComp | 0.782759 |
| environFriendlyAdPurchase | 0.812069 |
| firstShopNewStore | 0.710641 |
| firstElecEquip | 0.769311 |
| firstNewHealthFood | 0.825986 |
| firstNewStyles | 0.708634 |

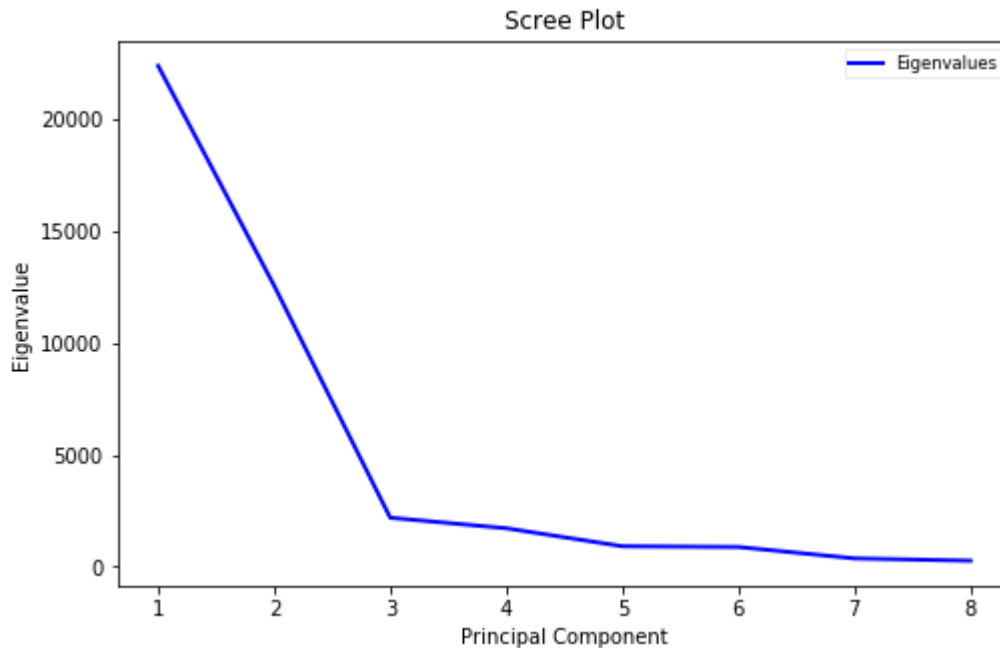
KMO overall score: 0.7860927597199712

Bartlett Spercicity:

Chi-Square value: 47560.22853092119
P-value: 0.0

PCA results (all components):

| | eigenvalues | prop_var | cumm_var |
|---|-------------|----------|----------|
| 0 | 3.127319 | 0.335984 | 0.335984 |
| 1 | 2.342190 | 0.251633 | 0.587617 |
| 2 | 0.977821 | 0.105052 | 0.692669 |
| 3 | 0.866053 | 0.093044 | 0.785714 |
| 4 | 0.631598 | 0.067856 | 0.853569 |
| 5 | 0.619806 | 0.066589 | 0.920158 |
| 6 | 0.402103 | 0.043200 | 0.963358 |
| 7 | 0.341060 | 0.036642 | 1.000000 |



PCA results (2 components):

| | eigenvalues | prop_var | cumm_var |
|---|-------------|----------|----------|
| 0 | 3.127319 | 0.335984 | 0.335984 |
| 1 | 2.342190 | 0.251633 | 0.587617 |

Extraction Communality:

| | factor1 | factor2 | communality |
|----------------------------------|----------|-----------|-------------|
| environRespRecycldProd | 0.530314 | -0.511031 | 0.542385 |
| enrironCompHelpConsumEnvironResp | 0.558388 | -0.589094 | 0.658829 |
| environPurchEnvironFriendlyComp | 0.623003 | -0.571018 | 0.714195 |
| environFriendlyAdPurchase | 0.692266 | -0.502012 | 0.731248 |
| firstShopNewStore | 0.622640 | 0.573944 | 0.717092 |
| firstElecEquip | 0.525002 | 0.613033 | 0.651437 |
| firstNewHealthFood | 0.769224 | 0.317849 | 0.692734 |
| firstNewStyles | 0.641803 | 0.591335 | 0.761589 |

Communality Estimates Total: 5.469508939223212

Original loadings:

| | 0 | 1 |
|----------------------------------|----------|-----------|
| environRespRecycldProd | 0.530314 | -0.511031 |
| enrironCompHelpConsumEnvironResp | 0.558388 | -0.589094 |
| environPurchEnvironFriendlyComp | 0.623003 | -0.571018 |

| | | |
|---------------------------|----------|-----------|
| environFriendlyAdPurchase | 0.692266 | -0.502012 |
| firstShopNewStore | 0.622640 | 0.573944 |
| firstElecEquip | 0.525002 | 0.613033 |
| firstNewHealthFood | 0.769224 | 0.317849 |
| firstNewStyles | 0.641803 | 0.591335 |

Rotated loadings:

| | 0 | 1 |
|----------------------------------|-----------|-----------|
| environRespRecycldProd | 0.009924 | -0.736401 |
| environCompHelpConsumEnvironResp | -0.025801 | -0.811273 |
| environPurchEnvironFriendlyComp | 0.032504 | -0.844475 |
| environFriendlyAdPurchase | 0.130272 | -0.845149 |
| firstShopNewStore | 0.845928 | -0.038696 |
| firstElecEquip | 0.805016 | 0.058192 |
| firstNewHealthFood | 0.767059 | -0.323040 |
| firstNewStyles | 0.871770 | -0.040080 |

Saved Factors

CLUSTER ANALYSIS

Cleaning & Compiling Variables for Cluster Analysis

Question: WILLING TO VOLUNTEER MY TIME/GOOD CAUSE

Frequency tables:

agree_lot
0.0 16865
1.0 8699
Name: V4375, dtype: int64

agree_little
0.0 16283
1.0 9281
Name: V4456, dtype: int64

neither
0.0 20707
1.0 4857
Name: V4618, dtype: int64

disagree_little
0.0 24725
1.0 839
Name: V4699, dtype: int64

disagree_lot
0.0 25043
1.0 521
Name: V4780, dtype: int64

Check if more than one answer selected:

1.0 24197
0.0 1367
Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5
8699 = 8699

correct: column 2 , value 4
9281 = 9281

correct: column 3 , value 3
4857 = 4857

correct: column 4 , value 2
839 = 839

correct: column 5 , value 1
521 = 521

Finished: WILLING TO VOLUNTEER MY TIME/GOOD CAUSE

Question: USUALLY READ INFO ON LABEL

Frequency tables:

agree_lot
0.0 21334
1.0 4230
Name: V7337, dtype: int64

agree_little
0.0 17950
1.0 7614
Name: V7361, dtype: int64

neither
0.0 18993
1.0 6571
Name: V7409, dtype: int64

disagree_little
0.0 22285
1.0 3279
Name: V7433, dtype: int64

disagree_lot
0.0 23474
1.0 2090
Name: V7457, dtype: int64

Check if more than one answer selected:

1.0 23784
0.0 1780
Name: sum, dtype: int64

Check if sum of columns equal to counts in combined column:

correct: column 1 , value 5
4230 = 4230

correct: column 2 , value 4
7614 = 7614

correct: column 3 , value 3
6571 = 6571

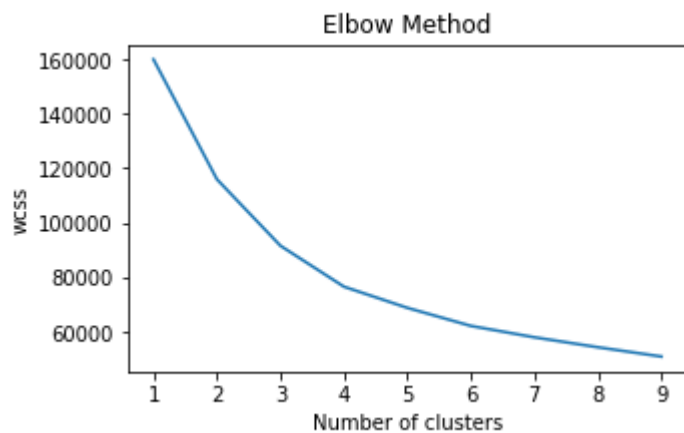
correct: column 4 , value 2
3279 = 3279

correct: column 5 , value 1
2090 = 2090

Finished: USUALLY READ INFO ON LABEL

Questions added as statistical/cluster drivers:

WILLING TO VOLUNTEER MY TIME/GOOD CAUSE
USUALLY READ INFO ON LABEL

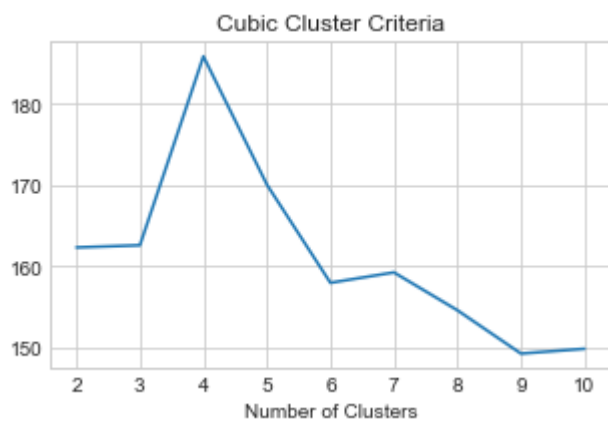


Best number of clusters (first kmeans run):

CCC PseudoT2

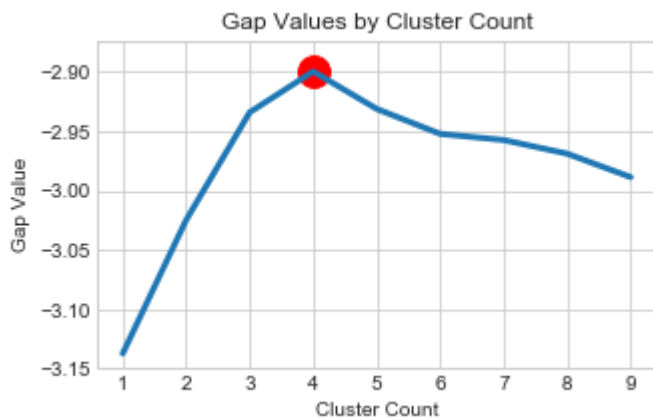
Number_clusters 4.0000 2.0000

Value_Index 185.8826 -607.5756



Gap Method

Optimal clusters: 4



Cluster Centers:

| clust_assign | 1.0 | 2.0 | 3.0 | 4.0 |
|---------------|-----------|----------|-----------|----------|
| environ | -2.157865 | 0.039401 | 0.102789 | 2.482217 |
| earlyad | -0.037010 | 1.477401 | -1.573452 | 0.384109 |
| volunteerTime | 3.742952 | 3.690411 | 4.286394 | 4.439581 |
| readLabels | 2.506633 | 3.189041 | 3.715350 | 4.040891 |

PROFILE VARIABLES

Cleaning & Compiling Profile Variables

Question: SEX - HOUSEHOLD HEAD 2506 FEMALE
Name: level3, dtype: object

Frequency tables:

col_1
1.0 17071
Name: V2506, dtype: int64

col_2
1.0 8493
Name: V2507, dtype: int64

Check if more than one answer selected:

1.0 25564
Name: sum, dtype: int64

Question: RESPONDENT MOST INF IN HH PURCH DECISIONS 43197 YES
Name: level3, dtype: object

Frequency tables:

col_1
1.0 14852
Name: V43198, dtype: int64

col_2
1.0 10112
Name: V43199, dtype: int64

Check if more than one answer selected:

```
1.0      24964
0.0       600
Name: sum, dtype: int64
Finished: SEX - HOUSEHOLD HEAD
```

Finished: RESPONDENT MOST INF IN HH PURCH DECSIONS

CBS THIS MORNING- FREQ OF VIEWING

```
0.0      25165
1.0       399
Name: mornCBS, dtype: int64
```

GOOD MORNING AMERICA (ABC)-FREQ OF VIEW

```
0.0      24831
1.0       733
Name: mornABC, dtype: int64
```

TODAY SHOW (NBC) - FREQUENCY OF VIEWING

```
0.0      25029
1.0       535
Name: mornNBC, dtype: int64
```

Questions added as profile variables:

```
SEX - HOUSEHOLD HEAD
RESPONDENT MOST INF IN HH PURCH DECSIONS
CBS THIS MORNING- FREQ OF VIEWING (1/week)
GOOD MORNING AMERICA (ABC)-FREQ OF VIEW (1/week)
TODAY SHOW (NBC) - FREQUENCY OF VIEWING (1/week)
```

Final Cluster Centers:

| clust_assign | 1.0 | 2.0 | 3.0 | 4.0 |
|---------------|-----------|----------|-----------|----------|
| environ | -2.157865 | 0.039401 | 0.102789 | 2.482217 |
| earlyad | -0.037010 | 1.477401 | -1.573452 | 0.384109 |
| volunteerTime | 3.742952 | 3.690411 | 4.286394 | 4.439581 |
| readLabels | 2.506633 | 3.189041 | 3.715350 | 4.040891 |
| mornCBS | 0.009536 | 0.017295 | 0.011576 | 0.020183 |
| mornABC | 0.018864 | 0.028767 | 0.028227 | 0.037221 |
| mornNBC | 0.018035 | 0.023116 | 0.018871 | 0.026212 |
| sex | 0.725954 | 0.667808 | 0.671741 | 0.647444 |
| purchDecs | 0.567371 | 0.552397 | 0.616714 | 0.606815 |

In [2]: