Final Exam

Jared Stadden

5/7/2020

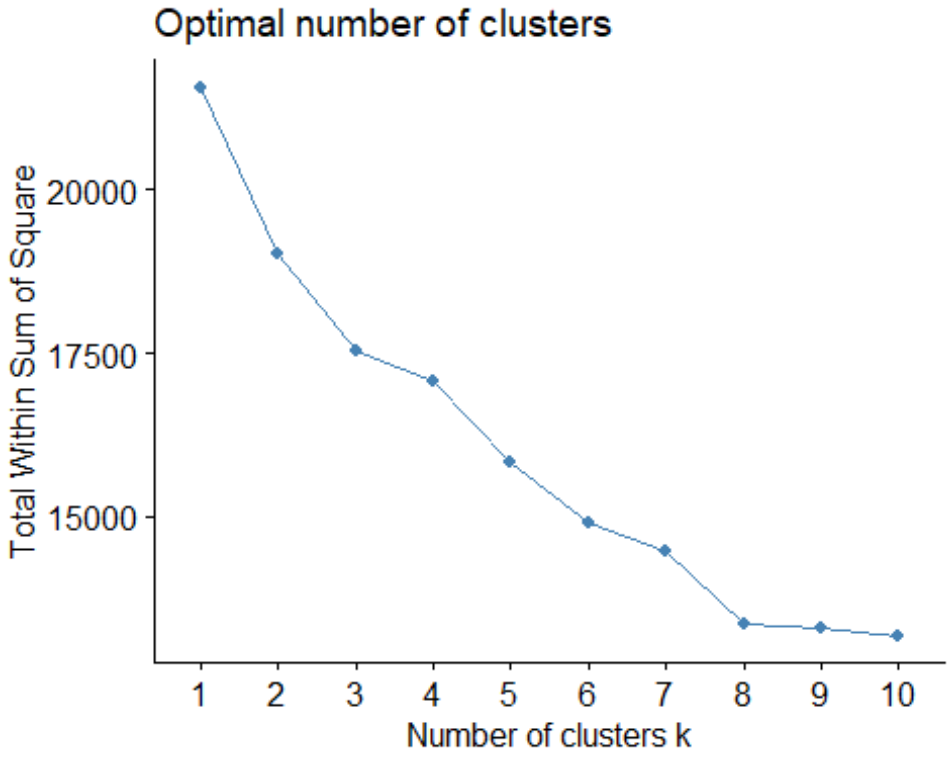
R Markdown

The variables containing percentages were converted to numbers so that they could be included in the k-means analysis.

I have chosen to use a derived variable to represent the customer loyalty aspect of the percent of total purchases by brand variables. I used the maximum percentage of any of the brands for the customer for this measure. A high max would show that the customer is loyal to a brand whereas a low max would indicate that they customer's purchases are spread across many brands.

I have normalized the values of the variables that will be used in the k-means analysis to prevent some variable to become overweighted simply because their inherently have larger values than another variable.

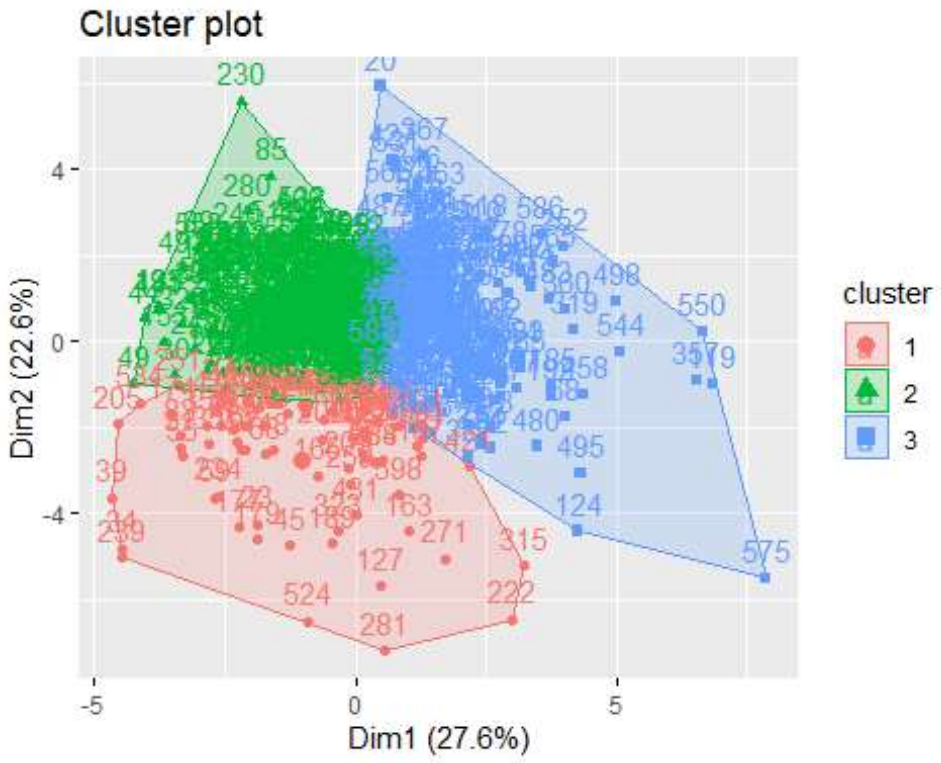
Determing the appropriate k input:



It appears that 3 would be the optimal k from this graph, if we are assume our marketing efforts can only handle between two and five different approaches.

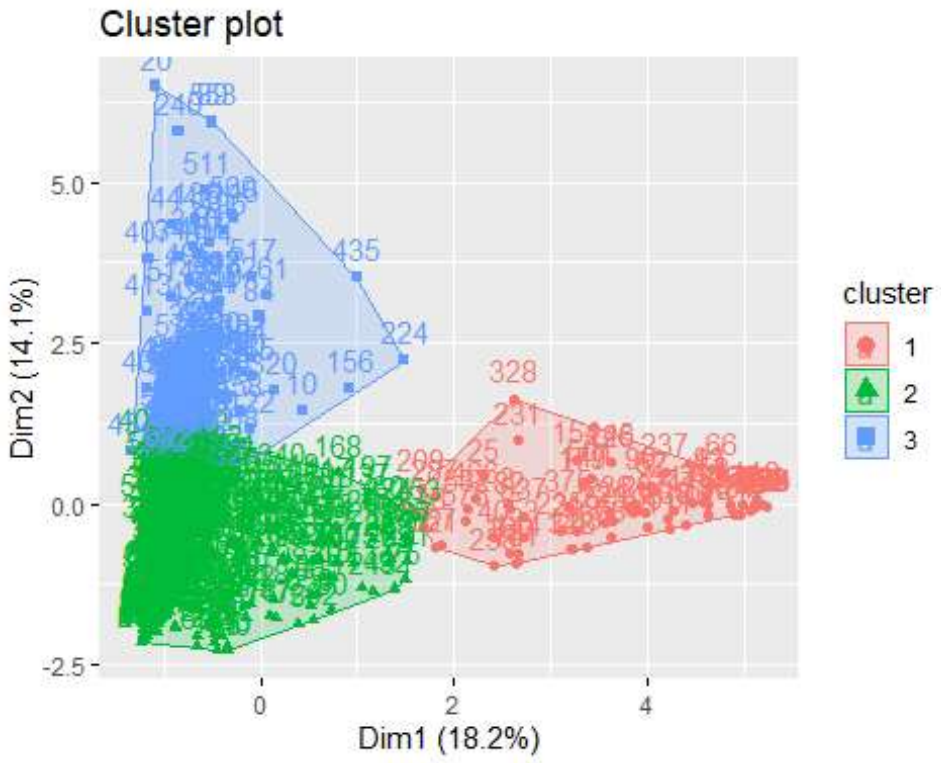
K-means based on Purchase Behavior:

```
##   No..of.Brands Brand.Runs Total.Volume No..of..Trans      Value
## 1    -0.5896898 -0.6407864    1.4208930    -0.1353318  0.9212778
## 2    -0.4390193 -0.5139855   -0.5152088    -0.5530579 -0.5086464
## 3     0.7948994  0.9132870    0.1938017     0.7888798  0.3597561
##   Trans...Brand.Runs   Vol.Tran   Avg..Price Pur.Vol.No.Promo....
## 1         1.18304137  1.7353411 -0.708375663            0.4354231
## 2        -0.09302414 -0.1445887 -0.001885739            0.2345553
## 3        -0.28907178 -0.4131147  0.250301067           -0.4667939
##   Pur.Vol.Promo.6.. Pur.Vol.Other.Promo.. max_loyalty
## 1        -0.4296637           -0.15972375   0.8063602
## 2        -0.2882161           -0.02052368   0.1723942
## 3         0.5367281            0.08338580  -0.5131927
```



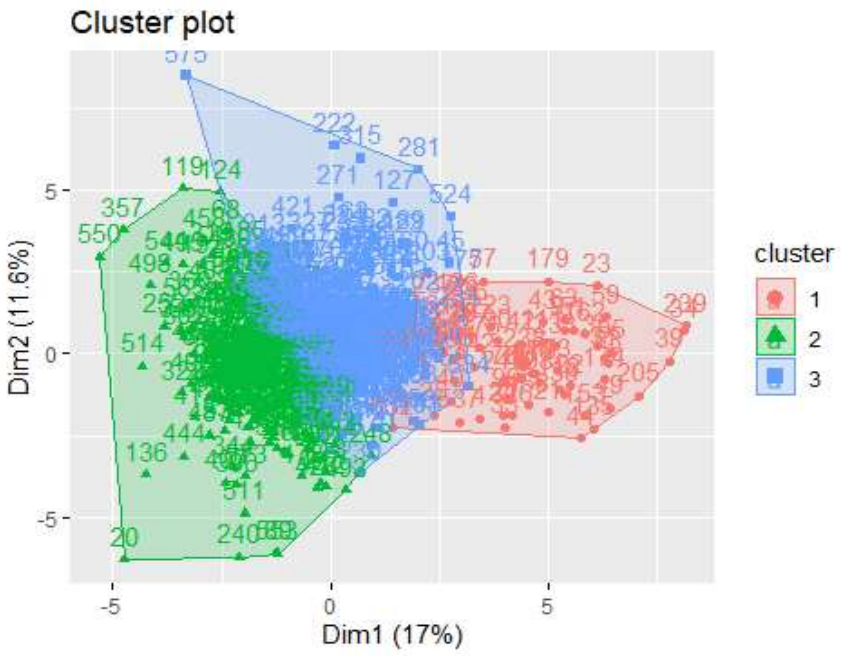
K-means based on Basis for Purchase:

```
##     Pr.Cat.1   Pr.Cat.2   Pr.Cat.3   Pr.Cat.4  PropCat.5  PropCat.6   Prop
Cat.7
## 1 -0.7811336 -1.1192162  2.3533589 -0.3222704 -1.0855323 -0.1711945 -0.443
95337
## 2 -0.4145980  0.5195713 -0.3131789  0.2006057  0.3913062 -0.0626521 -0.046
72354
## 3  1.5006786 -0.7375224 -0.4700697 -0.3446096 -0.4232694  0.2557349  0.363
03702
##      PropCat.8   PropCat.9 PropCat.10 PropCat.11  PropCat.12 PropCat.13
## 1 -0.458786014 -0.16641740 -0.2571876 -0.2304083 -0.16393967 -0.2328841
## 2 -0.008182177  0.01903935 -0.1545182  0.1095443 -0.09946207 -0.2013688
## 3  0.271176507  0.04129779  0.5408044 -0.1585269  0.34723429  0.6490519
##   PropCat.14  PropCat.15
## 1  2.3559150 -0.21596239
## 2 -0.3162083  0.03460623
## 3 -0.4636068  0.02792474
```



K-means based on both Purchase Behavior and Basis for Purchase:

```
##   No..of.Brands Brand.Runs Total.Volume No..of..Trans       Value
## 1  -0.598521651 -0.8120867   0.09190381   -0.42903264 -0.55575082
## 2   0.189532593  0.3915028  -0.34789649    0.22334716  0.02008816
## 3  -0.002581999 -0.0974991   0.22345007   -0.06306329  0.10668291
##   Trans...Brand.Runs   Vol.Tran Avg..Price Pur.Vol.No.Promo....
## 1         1.07696718  0.5271963 -1.3330597            0.2413117
## 2        -0.22834629 -0.5299567  0.8870873           -0.3717739
## 3        -0.07420425  0.2562657 -0.3310673            0.2076974
##   Pur.Vol.Promo.6.. Pur.Vol.Other.Promo..   Pr.Cat.1   Pr.Cat.2   Pr.Cat.3
## 1        -0.4871882            0.23045474 -0.7960149 -1.2358341  2.5300303
## 2         0.4270467            0.06483572  1.0136848 -0.3572921 -0.4477498
## 3        -0.1929535           -0.09543113 -0.5363193  0.5184782 -0.2363734
##     Pr.Cat.4  PropCat.5   PropCat.6   PropCat.7  PropCat.8   PropCat.9
## 1 -0.3579166 -1.1538285 -0.24964584 -0.45522491 -0.4787781 -0.12103046
## 2 -0.2788596 -0.3512634  0.15706153  0.23968178  0.4577418  0.16357708
## 3  0.2728708  0.4964442 -0.05565673 -0.06880197 -0.2162573 -0.08815753
##   PropCat.10    PropCat.11 PropCat.12 PropCat.13 PropCat.14   PropCat.15
## 1 -0.2558533 -0.2835592699 -0.1746531 -0.2404776  2.5320827 -0.252771665
## 2  0.3397891  0.0003590712  0.2614396  0.3908748 -0.4442971  0.067763353
## 3 -0.1821591  0.0613526957 -0.1449804 -0.2212431 -0.2392350  0.007502552
##   max_loyalty
## 1  1.45329573
## 2 -0.47543909
## 3  0.01692349
```



I believe the first clustering based on Purchase Behavior variables is the best segmentation. There didn't seem to be any extremes in the centroids for the percent of volume based on product proposition categories and there are variables addressing the price issue already included within the Purchase Behavior variables, so Basis for Purchase doesn't add much to the analysis.

Charateristics of clusters:

```
##   Group.1 Member.id      SEC      FEH       MT      SEX      AGE      EDU
## 1       1   1060406 3.426471 2.044118 7.691176 1.529412 3.044118 2.264706
## 2       2   1125195 1.908676 1.858447 7.639269 1.703196 3.264840 4.616438
## 3       3   1099001 2.712460 2.182109 8.661342 1.808307 3.214058 4.028754
##         HS    CHILD        CS Affluence.Index No..of.Brands Brand.Runs
## 1 3.897059 3.558824 0.8676471        7.911765  -0.598521651 -0.8120867
## 2 3.675799 3.319635 0.8904110       21.000000   0.189532593  0.3915028
## 3 4.616613 3.102236 0.9744409       16.214058  -0.002581999 -0.0974991
##   Total.Volume No..of..Trans       Value Trans...Brand.Runs   Vol.Tran
## 1   0.09190381   -0.42903264 -0.55575082         1.07696718  0.5271963
## 2  -0.34789649    0.22334716  0.02008816        -0.22834629 -0.5299567
## 3   0.22345007   -0.06306329  0.10668291        -0.07420425  0.2562657
##   Avg..Price Pur.Vol.No.Promo.... Pur.Vol.Promo.6.. Pur.Vol.Other.Promo..
## 1 -1.3330597            0.2413117        -0.4871882            0.23045474
## 2  0.8870873           -0.3717739         0.4270467            0.06483572
## 3 -0.3310673            0.2076974        -0.1929535           -0.09543113
##     Pr.Cat.1   Pr.Cat.2   Pr.Cat.3   Pr.Cat.4  PropCat.5   PropCat.6
## 1 -0.7960149 -1.2358341  2.5300303 -0.3579166 -1.1538285 -0.24964584
## 2  1.0136848 -0.3572921 -0.4477498 -0.2788596 -0.3512634  0.15706153
## 3 -0.5363193  0.5184782 -0.2363734  0.2728708  0.4964442 -0.05565673
##     PropCat.7  PropCat.8   PropCat.9 PropCat.10    PropCat.11 PropCat.12
## 1 -0.45522491 -0.4787781 -0.12103046 -0.2558533 -0.2835592699 -0.1746531
## 2  0.23968178  0.4577418  0.16357708  0.3397891  0.0003590712  0.2614396
## 3 -0.06880197 -0.2162573 -0.08815753 -0.1821591  0.0613526957 -0.1449804
##   PropCat.13 PropCat.14   PropCat.15 max_loyalty
## 1 -0.2404776  2.5320827 -0.252771665  1.45329573
## 2  0.3908748 -0.4442971  0.067763353 -0.47543909
## 3 -0.2212431 -0.2392350  0.007502552  0.01692349
```

Cluster 1: lower socioeconomic class, lower education, least affluent, fewest number of brands purchased, few brand runs, few transactions, low-value transactions, higher volume per transaction, low price, more use of "other" promotions, most purchases from price category 3, high brand loyalty based on percentage purchased by brand

Cluster 2: mid-higher socioeconomic class, higher education, most affluent, higher number of brands purchased, higher number of brand runs, more transactions, lower volume per transaction, high price, high use of "promotion code 6", most purchases from price category 1, lower brand loyalty based on percentage purchased by brand

Cluster 3: middle socioeconomic class, mid-high education, more household members, somewhat affluent, moderate number of brands purchased, moderate number of brand runs, moderate transactions, moderate volume per transaction, moderate price, moderate promo use, most purchases from price category 2, moderate brand loyalty based on percentage purchased by brand

Model to Predict/Assign Data to a Cluster

```
## [1] 3
```

Ex. The sample data was predicted to belong to cluster 3. This means that this customer should be targeted with marketing that is deemed appropriate for similar customers of that cluster.