Name: Nithin Das   CWID: 10422784   Date: 03/23/2020

BIA 672 Marketing Analytics

Assignment 1

HW 01:
Use income tax (income_byzip_pct) data and the following criteria to create ten (10) clusters of zip
codes for Virginia and identify clusters where "Kebab Shop" can open a new store
●
Pct of income in each category and hierarchical clustering methodology
●
Pct of income in each category   and k-means clustering methodology

1. **KMeans clustering on Pct of Income**

   List of Zip codes and corresponding clusters is shown in '*Part2 VA_CLUSTERS.pdf*'

   Exploratory Data Analysis on the obtained clusters is shown in attached Python file '*Assignment
1_Clustering.ipynb*'.

   The mean *Return_pct* values for each cluster shown below:

CLUSTER	STATEFIPS	zipcode	Returns_pct1	Returns_pct2	Returns_pct3	Returns_pct4	Returns_pct5	Returns_pct6	total	DISTANCE
1	51.0	23504.400000	38.213333	40.787333	0.702000	20.296667	0.000000	0.000000	125.333333	8.563161
2	51.0	23826.576923	38.488846	28.303077	25.700000	0.000000	7.508077	0.000000	185.769231	10.280285
3	51.0	23614.121693	36.968148	26.607725	15.408942	10.512381	9.442434	1.060106	3488.624339	6.278148
4	51.0	23723.500000	74.037500	25.962500	0.000000	0.000000	0.000000	0.000000	252.500000	8.731590
5	51.0	22975.094421	30.801888	22.562918	14.948798	10.676094	17.338670	3.672146	6298.969957	6.250085
6	51.0	23806.923077	55.626923	26.788846	11.616154	3.231923	2.538462	0.196923	3667.307692	9.042888
7	51.0	21807.107143	22.861905	15.310952	13.020952	10.175357	25.404048	13.227738	9505.000000	7.921080
8	51.0	20139.000000	0.000000	41.670000	25.000000	0.000000	33.330000	0.000000	120.000000	0.000000
9	51.0	21764.705882	20.620588	10.454118	9.112941	6.940000	23.489412	29.384118	7223.529412	8.070405
10	51.0	23975.618421	43.748728	28.577105	13.577982	7.404912	6.199912	0.492105	2988.815789	5.351473

We want to select cluster where Returns_pct6 is higher, since the income is higher for the group. Higher income is more likely to use Fast food restaurants.

Returns_pct6 is higher for cluster 9, therefore all zip codes under cluster 9 is a good choice starting The Kebab

2. **Hclust on Pct of Income**

List of Zip codes and corresponding clusters is shown in '*Part1 HCLUST10.pdf*'

Cluster Summary is shown below:

Cluster Summary

Cluster	Frequency	RMS Std Deviation	Maximum Distance from Seed to Observation	Radius Exceeded	Nearest Cluster	Distance Between Cluster Centroids
1	15	3.7300	11.8327		10	23.4606
2	26	4.3607	15.4060		3	15.0489
3	189	2.6703	15.7420		10	8.5849
4	4	4.1626	13.8923		6	22.1697
5	233	2.8410	17.4901		3	11.1263
6	26	3.5237	17.8237		10	13.3808
7	84	3.5133	22.5789		5	16.6121
8	1	.	0		7	41.2589
9	17	3.5291	16.3179		7	17.8619
10	228	2.4547	16.0755		3	8.5849

Cluster mean and standard deviation is shown below:

Cluster Means

Cluster	Returns_pct1	Returns_pct2	Returns_pct3	Returns_pct4	Returns_pct5	Returns_pct6
1	38.21333333	40.78733333	0.70200000	20.29666667	0.00000000	0.00000000
2	38.48884615	28.30307692	25.70000000	0.00000000	7.50807692	0.00000000
3	36.96814815	26.60772487	15.40894180	10.51238095	9.44243386	1.06010582
4	74.03750000	25.96250000	0.00000000	0.00000000	0.00000000	0.00000000
5	30.80188841	22.56291845	14.94879828	10.67609442	17.33866953	3.67214592
6	55.62692308	26.78884615	11.61615385	3.23192308	2.53846154	0.19692308
7	22.86190476	15.31095238	13.02095238	10.17535714	25.40404762	13.22773810
8	0.00000000	41.67000000	25.00000000	0.00000000	33.33000000	0.00000000
9	20.62058824	10.45411765	9.11294118	6.94000000	23.48941176	29.38411765
10	43.74872807	28.57710526	13.57798246	7.40491228	6.19991228	0.49210526

Cluster Standard Deviations

Cluster	Returns_pct1	Returns_pct2	Returns_pct3	Returns_pct4	Returns_pct5	Returns_pct6
1	6.172792453	5.036494624	2.718834309	3.552034963	0.000000000	0.000000000
2	5.840509791	4.102475125	3.827777423	0.000000000	6.964241248	0.000000000
3	2.592386615	2.869830896	1.867473245	2.702678566	3.889118809	1.381843122
4	7.209842231	7.209842231	0.000000000	0.000000000	0.000000000	0.000000000
5	3.408726990	2.576289111	1.946998423	2.406942138	3.475176929	2.917063745
6	4.278688368	4.170646307	4.734482930	2.952048806	2.746262467	0.354680354
7	3.848731124	2.498006250	2.757333975	2.894590071	4.020745036	4.567276046
8	.	.	.	.	.	.
9	3.551824303	2.191460868	2.828976769	2.231143205	4.340678908	5.048496136
10	3.192132489	2.777959579	1.907585221	2.381254157	2.846359448	0.913721127

All the zip codes under cluster 9 is a good choice starting The Kebab