

KPrototype

of Clusters: 5

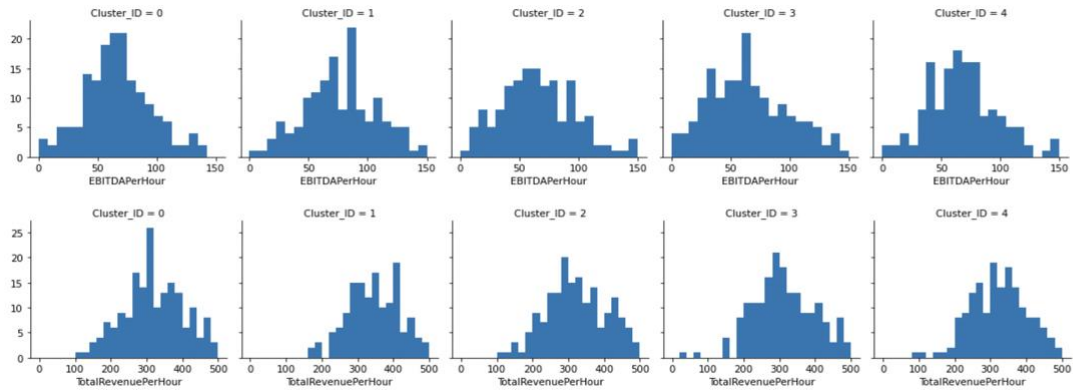
Steps:

1. Capped at 99% percentile to handle outliers
2. PowerTransformer() to normalize the numerical data
3. Observed the data distribution by reducing the dimenions to 2D (UMAP)

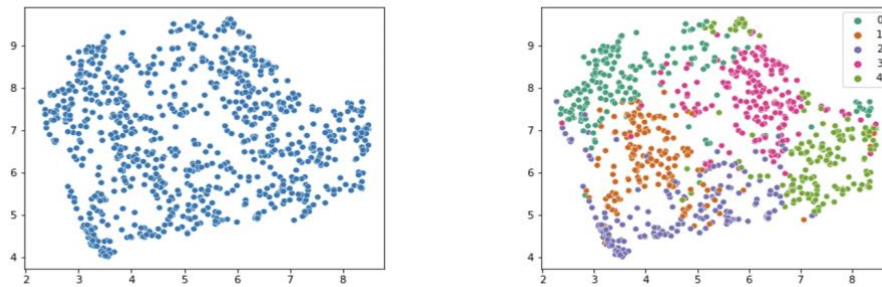
Cluster	Count of Offices	Total Revenue Per Hour				EBITDA Per Hour			
		Minimum	Maximum	Median	Mean	Minimum	Maximum	Median	Mean
0	187	\$118.17	\$786.32	\$316.57	\$332.26	-\$2,797.22	\$203.19	\$63.47	\$38.53
1	156	\$172.71	\$764.12	\$349.32	\$362.88	-\$14.80	\$253.99	\$79.32	\$83.29
2	176	\$110.11	\$554.63	\$320.60	\$326.67	-\$132.23	\$581.12	\$62.66	\$67.88
3	189	\$23.43	\$621.82	\$308.20	\$326.06	-\$3,826.76	\$267.89	\$60.75	\$42.56
4	169	\$87.92	\$571.81	\$323.01	\$330.87	-\$136.75	\$208.98	\$62.70	\$58.93

F1-weighted score: 87.6 %

Revenue and EBITDA Distribution across Clusters



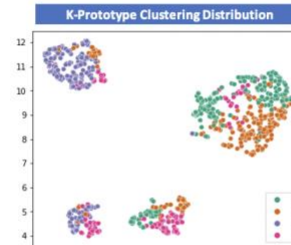
Data projected on to 2D plane - UMAP



Model Name	Variable clustering results	K-Prototype model
type with variable clustering	Variables considered 34, eigenval = 1, no. of clusters = 12	No. of clusters = 4, Init = 'Huang'

Variable clustering results						
Cluster	Variable	RS_Own	RS_NC	RS_Ratio	1-RS_Ratio	
0	0	% Hispanic	0.490304	0.199906	0.637045	0.362955
1	0	% No Health Insur. (10mi)	0.933040	0.237085	0.087769	0.912231
2	0	% Uninsured	0.933040	0.237085	0.087769	0.912231
3	1	% Speaking Only English (20mi)	0.509942	0.266090	0.667736	0.332264
4	1	% Veterans (20mi)	0.317571	0.064902	0.729794	0.270206
5	1	Aspen Within 20 Miles	0.591754	0.127185	0.467735	0.532265
6	1	Employees (3mi)	0.428836	0.122566	0.650948	0.349052
7	1	Ext Comp Within 20 Miles	0.798381	0.160566	0.240185	0.759815
8	2	Clinic	0.771709	0.058464	0.242466	0.757534
9	2	Primary Chair	0.416071	0.038761	0.807475	0.392525
10	2	Services	0.390225	0.020532	0.622558	0.377442
11	2	Total Employee	0.825005	0.194602	0.093116	0.906884
12	2	Big Box Retailers (1.5mi)	0.057720	0.126219	1.078393	-0.078393
13	3	% AfricanAmerican	0.900766	0.055855	0.105104	0.894896
14	3	% Black (10mi)	0.899777	0.054888	0.106043	0.893957
15	3	Delta Dental per 10K (10mi)	0.209639	0.083636	0.862496	0.137504
16	3	NPA Benchmarks	0.467834	0.140554	0.619080	0.380920
17	3	Median Age (20mi)	0.135327	0.046139	0.906498	0.093502
18	4	% HH Income \$20K+ (5mi)	0.529703	0.221756	0.604305	0.395695
19	4	% High School or Less (15mi)	0.672604	0.147762	0.384161	0.615839
20	4	% Medicaid (10mi)	0.559803	0.021562	0.449907	0.550093
21	4	Median Normalised Income	0.809028	0.121699	0.217433	0.782567
22	4	% Rural Industries (10mi)	0.144632	0.209708	1.082344	-0.082344
23	5	Fast Cas. Mexican Burger (25mi)	0.607174	0.033993	0.406649	0.593351
24	5	Shop. Center GLA (1.5mi)	0.607174	0.066859	0.420972	0.579028
25	6	Current Working Year	0.738469	0.157738	0.310511	0.689489
26	6	Secondary Chair	0.358515	0.030369	0.661576	0.338424
27	6	Years in DMA	0.826281	0.196959	0.216326	0.783674
28	7	Technicians	1.000000	0.117386	0.000000	1.000000
29	8	CVS Walgreens (25mi)	0.547242	0.005015	0.455040	0.544960
30	8	Highway Exit (.5mi)	0.547242	0.011282	0.457924	0.542076
31	9	Soft	1.000000	0.003699	0.000000	1.000000
32	10	Housing Value Ratio (1/15mi)	1.000000	0.009858	0.000000	1.000000
33	11	Dental Chains (1mi)	1.000000	0.032731	0.000000	1.000000

Selected Variables	
8	% Hispanic
4	% Veterans (20mi)
12	Big Box Retailers (1.5mi)
17	Median Age (20mi)
22	% Rural Industries (10mi)
24	Shop. Center GLA (1.5mi)
26	Secondary Chair
28	Technicians
30	Highway Exit (.5mi)
31	Soft
32	Housing Value Ratio (1/15mi)
33	Dental Chains (1mi)



F1-weighted score: 90 %

Cluster Distribution	
Cluster Id 0	248
Cluster Id 2	244
Cluster Id 1	231
Cluster Id 3	154

Cluster_ID		EBITDAPerHour				TotalRevenuePerHour			
		min	max	median	mean	min	max	median	mean
0	0	-3826.76	208.98	67.70	48.65	23.43	786.32	333.28	346.27
1	1	-2797.22	183.50	62.80	48.54	110.11	577.76	316.57	326.50
2	2	-104.72	581.12	67.42	77.27	121.80	764.12	340.90	344.54
3	3	-139.91	202.18	60.41	52.05	87.92	571.66	311.28	314.38

GMM

of Clusters : 5

Steps:

1. Capped at 99% percentile to handle outliers
2. PowerTransformer() to normalize the numerical data
3. Observed the data distribution by reducing the dimenstions to 2D (UMAP)

Cluster	Count of	Total Revenue Per Hour				EBITDA Per Hour			
		Minimum	Maximum	Median	Mean	Minimum	Maximum	Median	Mean
0	0	-\$132.23	\$581.12	\$54.64	\$59.14	\$110.11	\$536.81	\$301.76	\$317.38
1	1	-2797.22	\$303.34	\$73.37	\$59.56	\$121.80	\$786.32	\$346.31	\$357.91
2	2	-\$139.91	\$183.53	\$70.59	\$66.57	\$118.17	\$641.33	\$330.62	\$338.40
3	3	-\$117.10	\$202.18	\$62.70	\$59.69	\$87.92	\$571.81	\$322.11	\$328.94
4	4	-3826.76	\$267.89	\$60.80	\$41.73	\$23.43	\$619.42	\$308.06	\$322.91

F1-weighted score: 81.2 %

Revenue and EBITDA Distribution across Clusters

