

CS 5433: Bigdata Management
Programming Assignment 3
Task5 – Output

Group 4

Output results

Initial Data:

INSTITUTE ID	NAME	CITY	STATE	PR Score	PR Rank	Score	Year	Rank
241.0	22.0	2.0	4.0	47.27	3.0	61.53	2017.0	2.0
306.0	4.0	39.0	3.0	44.01	4.0	58.92	2017.0	3.0
264.0	19.0	7.0	6.0	28.81	9.0	57.32	2017.0	5.0
284.0	3.0	0.0	0.0	43.94	5.0	56.5	2017.0	6.0
235.0	39.0	5.0	10.0	27.06	11.0	56.3	2017.0	7.0
309.0	38.0	21.0	4.0	30.76	7.0	55.37	2017.0	8.0
282.0	2.0	4.0	0.0	26.12	12.0	54.7	2017.0	9.0
254.0	33.0	3.0	1.0	11.2	35.0	52.81	2017.0	10.0
238.0	21.0	2.0	4.0	9.73	42.0	51.75	2017.0	12.0
315.0	85.0	31.0	13.0	49.96	2.0	51.46	2017.0	13.0
312.0	43.0	40.0	0.0	32.95	6.0	51.36	2017.0	14.0
231.0	133.0	2.0	4.0	11.52	34.0	51.2	2017.0	15.0
316.0	10.0	7.0	6.0	17.15	20.0	48.9	2017.0	16.0
269.0	81.0	4.0	0.0	3.53	86.0	48.84	2017.0	17.0
314.0	44.0	7.0	6.0	4.71	69.0	48.19	2017.0	19.0
255.0	166.0	6.0	12.0	15.57	23.0	46.72	2017.0	20.0
278.0	17.0	1.0	1.0	8.7	47.0	46.45	2017.0	21.0
310.0	6.0	0.0	0.0	3.79	84.0	46.45	2017.0	21.0
298.0	27.0	5.0	10.0	11.16	36.0	45.52	2017.0	23.0
226.0	157.0	47.0	8.0	3.83	83.0	44.99	2017.0	24.0

only showing top 20 rows

Data after performing Scaling:

INSTITUTE ID	NAME	CITY	STATE	PR Score	PR Rank	Score	Year	Rank	PR Score_Scaled	PR Rank_Scaled	Score_Scaled
241.0	22.0	2.0	4.0	47.27	3.0	61.53	2017.0	2.0	0.473	0.008	0.954
306.0	4.0	39.0	3.0	44.01	4.0	58.92	2017.0	3.0	0.44	0.012	0.914
264.0	19.0	7.0	6.0	28.81	9.0	57.32	2017.0	5.0	0.288	0.033	0.89
284.0	3.0	0.0	0.0	43.94	5.0	56.5	2017.0	6.0	0.439	0.017	0.877
235.0	39.0	5.0	10.0	27.06	11.0	56.3	2017.0	7.0	0.271	0.041	0.874
309.0	38.0	21.0	4.0	30.76	7.0	55.37	2017.0	8.0	0.308	0.025	0.86
282.0	2.0	4.0	0.0	26.12	12.0	54.7	2017.0	9.0	0.261	0.046	0.85
254.0	33.0	3.0	1.0	11.2	35.0	52.81	2017.0	10.0	0.112	0.141	0.821
238.0	21.0	2.0	4.0	9.73	42.0	51.75	2017.0	12.0	0.097	0.17	0.805
315.0	85.0	31.0	13.0	49.96	2.0	51.46	2017.0	13.0	0.5	0.004	0.8
312.0	43.0	40.0	0.0	32.95	6.0	51.36	2017.0	14.0	0.33	0.021	0.799
231.0	133.0	2.0	4.0	11.52	34.0	51.2	2017.0	15.0	0.115	0.137	0.796
316.0	10.0	7.0	6.0	17.15	20.0	48.9	2017.0	16.0	0.171	0.079	0.761
269.0	81.0	4.0	0.0	3.53	86.0	48.84	2017.0	17.0	0.035	0.353	0.76
314.0	44.0	7.0	6.0	4.71	69.0	48.19	2017.0	19.0	0.047	0.282	0.75
255.0	166.0	6.0	12.0	15.57	23.0	46.72	2017.0	20.0	0.156	0.091	0.728
278.0	17.0	1.0	1.0	8.7	47.0	46.45	2017.0	21.0	0.087	0.191	0.724
310.0	6.0	0.0	0.0	3.79	84.0	46.45	2017.0	21.0	0.038	0.344	0.724
298.0	27.0	5.0	10.0	11.16	36.0	45.52	2017.0	23.0	0.112	0.145	0.71
226.0	157.0	47.0	8.0	3.83	83.0	44.99	2017.0	24.0	0.038	0.34	0.702

only showing top 20 rows

Found out Variable vector after applying assembler on “Scaled” columns:

INSTITUTE ID	NAME	CITY	STATE	PR Score	PR Rank	Score	Year	Rank	PR Score_Scaled	PR Rank_Scaled	Score_Scaled	variable
241.0	22.0	2.0	4.0	47.27	3.0	61.53	2017.0	2.0	0.473	0.008	0.954	[0.473,0.008,0.954]
306.0	4.0	39.0	3.0	44.01	4.0	58.92	2017.0	3.0	0.44	0.012	0.914	[0.44,0.012,0.914]
264.0	19.0	7.0	6.0	28.81	9.0	57.32	2017.0	5.0	0.288	0.033	0.89	[0.288,0.033,0.89]
284.0	3.0	0.0	0.0	43.94	5.0	56.5	2017.0	6.0	0.439	0.017	0.877	[0.439,0.017,0.877]
235.0	39.0	5.0	10.0	27.06	11.0	56.3	2017.0	7.0	0.271	0.041	0.874	[0.271,0.041,0.874]
309.0	38.0	21.0	4.0	30.76	7.0	55.37	2017.0	8.0	0.308	0.025	0.86	[0.308,0.025,0.86]
282.0	2.0	4.0	0.0	26.12	12.0	54.7	2017.0	9.0	0.261	0.046	0.85	[0.261,0.046,0.85]
254.0	33.0	3.0	1.0	11.2	35.0	52.81	2017.0	10.0	0.112	0.141	0.821	[0.112,0.141,0.821]
238.0	21.0	2.0	4.0	9.73	42.0	51.75	2017.0	12.0	0.097	0.17	0.805	[0.097,0.17,0.805]
315.0	85.0	31.0	13.0	49.96	2.0	51.46	2017.0	13.0	0.5	0.004	0.8	[0.5,0.004,0.8]
312.0	43.0	40.0	0.0	32.95	6.0	51.36	2017.0	14.0	0.33	0.021	0.799	[0.33,0.021,0.799]
231.0	133.0	2.0	4.0	11.52	34.0	51.2	2017.0	15.0	0.115	0.137	0.796	[0.115,0.137,0.796]
316.0	10.0	7.0	6.0	17.15	20.0	48.9	2017.0	16.0	0.171	0.079	0.761	[0.171,0.079,0.761]
269.0	81.0	4.0	0.0	3.53	86.0	48.84	2017.0	17.0	0.035	0.353	0.76	[0.035,0.353,0.76]
314.0	44.0	7.0	6.0	4.71	69.0	48.19	2017.0	19.0	0.047	0.282	0.75	[0.047,0.282,0.75]
255.0	166.0	6.0	12.0	15.57	23.0	46.72	2017.0	20.0	0.156	0.091	0.728	[0.156,0.091,0.728]
278.0	17.0	1.0	1.0	8.7	47.0	46.45	2017.0	21.0	0.087	0.191	0.724	[0.087,0.191,0.724]
310.0	6.0	0.0	0.0	3.79	84.0	46.45	2017.0	21.0	0.038	0.344	0.724	[0.038,0.344,0.724]
298.0	27.0	5.0	10.0	11.16	36.0	45.52	2017.0	23.0	0.112	0.145	0.71	[0.112,0.145,0.71]
226.0	157.0	47.0	8.0	3.83	83.0	44.99	2017.0	24.0	0.038	0.34	0.702	[0.038,0.34,0.702]

only showing top 20 rows

Standardized Variate:

```
+-----+
|Standardized variate|
+-----+
|[1.5963257792085033,-1.0776995573167338,2.2262379596389934]|
|[1.4062059524738282,-1.0586198365465274,1.8619008421467076]|
|[0.5305025081201714,-0.9584513025029443,1.6432985716513355]|
|[1.4004447456030804,-1.0347701855837697,1.5248890084663425]|
|[0.43256199131746004,-0.9202918609625316,1.497563724654421]|
|[0.6457266455351264,-0.9966107440433569,1.3700457335321206]|
|[0.3749499226099826,-0.8964422099997736,1.2789614541590488]|
|[-0.48346990113143085,-0.4432988417073735,1.014817043977141]|
|[-0.569888004192647,-0.3049708661233775,0.8690821969802272]|
|[1.7518783647186924,-1.09677927808694,0.8235400572936913]|
|[0.7724731966915768,-1.0156904648135634,0.8144316293563841]|
|[-0.46618628051918765,-0.4623785624775797,0.7871063455444626]|
|[-0.1435586957573141,-0.7390345136455716,0.4683113677387116]|
|[-0.927082830179007,0.5679263591135619,0.4592029398014044]|
|[-0.8579483477300339,0.22926131544239955,0.36811866042833274]|
|[-0.22997679881853028,-0.6817953513349525,0.16773324580757498]|
|[-0.6275000729001244,-0.20480233207979429,0.13129953405834632]|
|[-0.9097992095667636,0.5249969873805976,0.13129953405834632]|
|[-0.48346990113143085,-0.4242191209371671,0.003781542936045923]|
|[-0.9097992095667636,0.5059172666103916,-0.06908588056241144]|
+-----+
only showing top 20 rows
```

Eigenvectors:

```
##### Eigenvector #####
DenseMatrix([[-0.60496295,  0.22328741,  0.76430528],
              [ 0.54284721,  0.81788602,  0.19073374],
              [-0.58252616,  0.53028783, -0.61600169]])
```

Contribution Rate:

```
##### Contribution rate #####
[0.7016699649623026,0.18292388625092215,0.11540614878677524]
```

Final PCA Score's/Main Component Score:

```
##### FINAL PCA SCOREs #####
+-----+
|Main component score|
+-----+
|[-2.8475860138937548,0.6555509410225909,-0.35683979506912467]|
|[-2.5099772855353377,0.435501074409122,-0.2740779549223128]|
|[-1.7984913915960998,0.2059718424800403,-0.7896178318926204]|
|[-2.297227041952679,0.2750076908267295,-0.066332483062998]|
|[-1.633631899586792,0.13803161617743287,-0.7674230806244504]|
|[-1.7297355452212397,0.05558721220075158,-0.5405054979407649]|
|[-1.4584904739496252,0.028751738701833673,-0.6722479873372307]|
|[-0.5393196384054464,0.06762446283122203,-1.0791996575067304]|
|[-0.32705457063862275,0.08418349046202545,-1.0290927457584742]|
|[-2.134938711810865,-0.06915479280594028,0.62247500112131]|
|[-1.4931101342136794,-0.226352315009137,-0.10501235838986739]|
|[-0.42748552233609305,-0.06487357201970728,-0.9293586660842779]|
|[-0.5871387551295573,-0.3881610288025444,-0.5391620782140139]|
|[0.6016502807444919,0.501002841853035,-0.8831213707334981]|
|[0.4290420817523568,0.1911494100642601,-0.8387683001904124]|
|[-0.3286922645747341,-0.5200349129081956,-0.4091378195806554]|
|[0.19195250966108768,-0.23799128419143845,-0.5995450667319702]|
|[0.7589025536880172,0.2958675372014221,-0.6761104351857921]|
|[0.05999236589841709,-0.4529103247197494,-0.4527609322058134]|
|[0.8652749272124152,0.17400055304060855,-0.5563118272650989]|
+-----+
only showing top 20 rows
```