

Semantic Data Integration: University Domain Use Case

Team 21: Sanaz Bayat, Shirin Shoghli, Shahrzad Torabi, Navid Hadipour Limouei

July 6, 2025

Task2. (26.05 – 26.06)

(4th week)- Develop at least three different matchers

We merge the datasets and make mediated schema automatically. Then combine them with 2 string matchers and 1 semantic matcher.

1- Edit Distance Matcher

2- Jarowinkler Matcher

3- Word2vec Matcher

First of all we implemented three matchers and compared our data sets and Mediated Schema by matcher

Edit Distance Matcher DS1 and MS

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudents
Name	0.634555456	0.3966277	0.8686805	0.1330939	0.5384438	0.6919392	0.9255607	0.6427564	0.9987512	0.5757986	0.3640851
Address	0.863113372	0.7735842	0.2097908	0.4010309	0.899306	0.6760253	0.0043131	0.5796973	0.7108122	0.8827119	0.3135077
City	0.024860674	0.2857744	0.474402	0.8287505	0.7670015	0.6902764	0.082582	0.7531109	0.5541834	0.0718907	0.6253221
ZIPCode	0.08200391	0.6160309	0.8691085	0.709904	0.3368345	0.0275592	0.0125721	0.3941287	0.7601958	0.9792687	0.3187151
County	0.011003664	0.8792054	0.1470452	0.6028058	0.7298056	0.2247541	0.6123477	0.2453392	0.3550698	0.5614732	0.5845693
Country	0.106022477	0.8539995	0.0867565	0.2068289	0.1279476	0.9895397	0.7186997	0.6378357	0.7186565	0.0735689	0.2273008
Website	0.5944936	0.0948384	0.6513618	0.5401484	0.7427479	0.1068046	0.5988854	0.6559137	0.9087119	0.1035275	0.3438109
TtlEmployee	0.673835991	0.203544	0.4443887	0.322819	0.812616	0.4114172	0.9976669	0.12947	0.9126878	0.0643557	0.988726
Url	0.863425098	1	0.6216968	0.7339915	0.0431025	0.5521597	0.8676786	0.0390318	0.365783	0.0975419	0.5035158
EmployeeCount	0.470724633	0.1247332	0.4408202	0.8660739	0.7670921	0.7878117	0.3558331	0.4247428	0.1342257	0.9154746	0.6638952
Uni	0.233649097	0.8421824	0.4637486	0.2737344	0.91414	0.6492423	0.7738363	0.7907367	0.0142821	0.0192798	0.0503131
Location	0.849243716	0.2996034	1	0.7172287	0.9970882	0.5638842	0.2884868	0.1562332	0.0041854	0.2144938	0.9540168

Edit Distance Matcher DS2 and MS

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0.076516	0.5954797	0.6042732	0.2817032	0.3202126	0.5551523	0.2225531
Address	0.2316228	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0.0336224	0.5541899
City	0.5206539	0.1434778	0.5500144	0.027403	0.7216145	0.413726	0.732459	0.3871426	0.604471	0.3837541
ZIPCode	0.9151975	0.7624892	0.6591803	0.292355	0.1060572	1	0.3755258	0.3260489	0.8216316	0.8076475
County	0.0598117	0.6937756	0.9871757	0.434983	0.1627002	0.8939023	0.218353	0.3922385	0.9288751	0.9495708
Country	0.4324153	0.4525174	0.1840848	0.116098	0.5349092	0.4187854	0.4420324	0.9558367	0.7551859	0.7244656
Website	0.6616418	0.3983013	0.0633616	0.527498	0.0918425	0.4649473	0.0049886	0.6582168	0.6018162	0.8127478
TotalEmployee	0.3346628	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0.1902264	0.4191538	0.5785432	0.6477537
UniversityName	0.6078690	0.8204025	0.8326359	0.778063	0.1067257	0.1746027	0.2593956	0.2537462	0.2451263	0.0651382
Url	0.5770694	1	0.3197724	0.973771	0.8894954	0.8009834	0.2656226	0.0436453	0.1827011	0.4329506
EmployeeCount	0.2111921	0.1681117	0.2940051	0.999686	0.2841621	0.8025041	0.3365679	0.4620293	0.8733068	0.5240171
Uni	0.0275744	0.4448508	0.5865268	0.108421	0.0829025	0.2650847	0.3399365	0.9770957	0.6568788	0.698761
Location	0.0596662	0.917602	0.7714193	0.166395	0.2613776	0.9608816	0.209736	0.2670432	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0.3668131	0.067297	0.8317456	0.1072138	0.3227397	0.9758965	0.8737481	0.6970036

Edit Distance Matcher DS3 and MS

MS/DS2	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.752677	0.888258	0.5936356	0.516358034	0.2814907	0.3858364	0.223614	0.1487153	0.8701889	0.1495783
Address	0.7622372	0.5069838	0.1427436	0.1170257	0.465142086	0.606668	0.6979949	0.7756804	0.4649366	0.5459885	0.6958357
City	0.3340963	0.2307437	0.6708038	0.5247858	0.816999326	0.3347004	0.8188282	0.1799399	0.4352849	0.6997791	0.0943358
ZIPCode	0.1800447	0.9136315	0.7008141	0.4495984	0.469177902	0.3210436	0.0410306	0.8249529	0.6062852	0.9419251	0.7456657
County	0.9112455	0.0670437	0.8337427	0.1775813	0.349898997	0.2046271	0.7155149	0.6364887	0.2514421	0.3129464	0.3906298
Country	0.7039374	0.2090052	0.5397603	0.6628521	0.514579069	0.3301529	0.0776715	1	0.9808873	0.1553797	0.193914
Website	0.4461419	0.7916267	0.0810533	0.2916591	0.206157207	0.7049967	0.8946663	0.8628775	0.9213003	0.9481751	0.6894031
TotalEmployee	0.7261335	0.0183608	0.7862071	0.0196103	0.701319206	0.6591873	0.1029699	0.7052238	0.6050273	0.3100573	0.5163255
UniversityName	0.659876	0.9298506	0.2924313	0.171838	0.428310821	0.9412394	0.44334	0.1344476	0.1089263	0.9165222	0.3729457
Url	0.1488488	0.9957112	0.6789154	0.1216574	0.257964023	0.8204019	0.9442562	0.7076351	0.5347282	0.6935514	0.6075847
EmployeeCount	0.0175038	0.9238061	0.1413813	0.3523177	0.666906574	0.0665237	0.5661145	0.2299738	0.0496781	0.4392192	0.451797
Uni	0.0897842	0.4058381	0.5478065	0.6010489	0.347087417	0.7084516	0.036786	0.6513357	0.7780713	0.5406696	0.4002897
Location	0.0894067	1	0.9089393	0.9905489	0.059718584	0.3140834	0.4039054	0.5959151	0.0997399	0.7513165	0.0482468
PostalCode	0.2156155	0.1129651	0.8712732	0.8484077	0.730201163	0.6280397	0.4289705	0.6374902	0.892385	0.0769934	0.0583747

Jarowinkler Matcher DS1 and MS

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.789844986	0.9681812	0.4869981	0.9207067	0.1358179	0.020048	0.5690154	0.8896549	0.0738498	0.583157	0.025444
Address	0.611097624	0.0883411	0.822571	0.8136874	0.8303438	0.915001	0.0748408	0.722211	0.2427696	0.5899082	0.5662333
City	0.209231699	0.4885359	0.9888736	0.2731371	0.1344786	0.1975415	0.6497139	0.7433783	0.286303	0.7019775	0.9617729
ZIPCode	0.247421175	0.7220105	0.4190578	0.5894569	0.6220515	0.5080734	0.8549641	0.9063952	0.8714753	0.2569932	0.4643575
County	0.13620803	0.9404599	0.3742957	0.4044103	0.733082	0.5624746	0.6135537	0.072065	0.8358054	0.3558805	0.0518823
Country	0.234452891	0.3729666	0.1541091	0.4762428	0.9143311	0.6104362	0.2915955	0.8628245	0.2363912	0.1297937	0.4444371
Website	0.090403099	0.7385985	0.9757245	0.0971771	0.8830093	0.8496748	0.9836667	0.7423921	0.6744922	0.8177635	0.180635
TtlEmployee	0.469523128	0.7131694	0.9330352	0.3637054	0.9694276	0.4957733	0.0092659	0.279672	0.7387094	0.7262803	0.4960092
Url	0.03129378	0.4994233	0.0085193	0.9565561	0.2489444	0.911122	0.33633	0.3495522	0.8581252	0.8252331	0.6896336
EmployeeCount	0.285770954	0.1112681	0.8813943	0.7899547	0.7331684	0.0899158	0.523028	0.6451922	0.2725608	0.6492173	0.5124547
Uni	0.688428196	0.3808196	0.4728296	0.8853713	0.5699964	0.7859025	0.6263599	0.9030513	0.8772086	0.832315	0.754589
Location	0.945962075	0.1422466	0.0085991	0.8836539	0.7490873	0.6172746	0.4295731	0.2954435	0.7441942	0.7736855	0.0479171

Jarowinkler Matcher DS2 and MS

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0.076516	0.5954797	0.6042732	0.2817032	0.3202126	0.5551523	0.2225531
Address	0.2316228	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0.0336224	0.5541899
City	0.5206539	0.1434778	0.5500144	0.027403	0.7216145	0.413726	0.732459	0.3871426	0.604471	0.3837541
ZIPCode	0.9151975	0.7624892	0.6591803	0.292355	0.1060572	1	0.3755258	0.3260489	0.8216316	0.8076475
County	0.0598117	0.6937756	0.9871757	0.434983	0.1627002	0.8939023	0.218353	0.3922385	0.9288751	0.9495708
Country	0.4324153	0.4525174	0.1840848	0.116098	0.5349092	0.4187854	0.4420324	1	0.7551859	0.7244656
Website	0.6616418	0.3983013	0.0633616	0.527498	0.0918425	0.4649473	0.0049886	0.6582168	0.6018162	0.8127478
TotalEmployee	0.3346628	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0.1902264	0.4191538	0.5785432	0.6477537
UniversityName	0.4320619	0.8204025	0.8326359	0.778063	0.1067257	0.1746027	0.2593956	0.2537462	0.2451263	0.0651382
Url	0.5770694	1	0.3197724	0.973771	0.8894954	0.8009834	0.2656226	0.0436453	0.1827011	0.4329506
EmployeeCount	0.2111921	0.1681117	0.2940051	0.999686	0.2841621	0.8025041	0.3365679	0.4620293	0.8733068	0.5240171
Uni	0.0275744	0.4448508	0.5865268	0.108421	0.0829025	0.2650847	0.3399365	0.9770957	0.6568788	0.698761
Location	0.0596662	0.917602	0.7714193	0.166395	0.2613776	0.9608816	0.209736	0.2670432	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0.3668131	0.067297	0.8317456	0.1072138	0.3227397	0.9758965	0.8737481	0.6970036

Jarowinkler Matcher DS3 and MS

MS/DS2	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.752677	0.888258	0.5936356	0.516358034	0.2814907	0.3858364	0.223614	0.1487153	0.8701889	0.1495783
Address	0.7622372	0.5069838	0.1427436	0.1170257	0.465142086	0.606668	0.6979949	0.7756804	0.4649366	0.5459885	0.6958357
City	0.3340963	0.2307437	0.6708038	0.5247858	0.816999326	0.3347004	0.8188282	0.1799399	0.4352849	0.6997791	0.0943358
ZIPCode	0.1800447	0.9136315	0.7008141	0.4495984	0.469177902	0.3210436	0.0410306	0.8249529	0.6062852	0.9419251	0.7456657
County	0.9112455	0.0670437	0.8337427	0.1775813	0.349898997	0.2046271	0.7155149	0.6364887	0.2514421	0.3129464	0.3906298
Country	0.7039374	0.2090052	0.5397603	0.6628521	0.514579069	0.3301529	0.0776715	1	0.9808873	0.1553797	0.193914
Website	0.4461419	0.7916267	0.0810533	0.2916591	0.206157207	0.7049967	0.8946663	0.8628775	0.9213003	0.9481751	0.6894031
TotalEmployee	0.7261335	0.0183608	0.7862071	0.0196103	0.701319206	0.6591873	0.1029699	0.7052238	0.6050273	0.3100573	0.5163255
UniversityName	0.3837719	0.9298506	0.2924313	0.171838	0.428310821	0.9412394	0.44334	0.1344476	0.1089263	0.9165222	0.3729457
Url	0.1488488	0.9957112	0.6789154	0.1216574	0.257964023	0.8204019	0.9442562	0.7076351	0.5347282	0.6935514	0.6075847
EmployeeCount	0.0175038	0.9238061	0.1413813	0.3523177	0.666906574	0.0665237	0.5661145	0.2299738	0.0496781	0.4392192	0.451797
Uni	0.0897842	0.4058381	0.5478065	0.6010489	0.347087417	0.7084516	0.036786	0.6513357	0.7780713	0.5406696	0.4002897
Location	0.0894067	1	0.9089393	0.9905489	0.059718584	0.3140834	0.4039054	0.5959151	0.0997399	0.7513165	0.0482468
PostalCode	0.2156155	0.1129651	0.8712732	0.8484077	0.730201163	0.6280397	0.4289705	0.6374902	0.892385	0.0769934	0.0583747

Word2Vec Matcher DS1 and MS

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudents
Name	0.76956745	0.999257	0.0906651	0.0828745	0.2465471	0.6403902	0.4009692	0.4868334	0.5654526	0.8395071	0.6859385
Address	0.654836614	0.3393915	0.6789098	0.3700149	0.9362669	0.5846444	0.1717886	0.3162587	0.6874571	0.1738595	0.3251042
City	0.187262052	0.7088581	0.5083785	0.7831527	0.2561259	0.0349138	0.6122942	0.9386119	0.6980071	0.9838066	0.2147964
ZIPCode	0.301984566	0.2679665	0.1859646	0.144345	0.0109112	0.2564039	0.0060612	0.049307	0.815552	0.2616994	0.2410635
County	0.488744816	0.3150586	0.5969189	0.998206	0.706756	0.3598801	0.8180751	0.898576	0.519291	0.2463695	0.0757111
Country	0.018496375	0.3606158	0.4385997	0.6150142	0.6270405	0.7734751	0.3458741	0.685453	0.0994007	0.8094303	0.7210166
Website	0.70346041	0.2384398	0.7644213	0.2855777	0.7614583	0.0956373	0.1392603	0.7662748	0.2017871	0.1206042	0.4657043
TtlEmployee	0.459358985	0.6853885	0.6069419	0.1046735	0.8335674	0.8390481	0.513297	0.2145234	0.2729715	0.0301975	0.1005082
Url	0.947083299	0.7655944	0.9599978	0.3500981	0.4625297	0.5880674	0.9939674	0.1201332	0.8857413	0.3833805	0.3451438
EmployeeCount	0.379265914	0.1110377	0.0831056	0.743881	0.6671362	0.6701759	0.4565364	0.5928412	0.1523741	0.4408387	0.9547757
Uni	0.141320878	0.7172533	0.1062917	0.5624347	0.6451214	0.9385083	0.0928966	0.9319087	0.399433	0.8143558	0.7981361
Location	0.800004024	0.6733331	0.7658938	0.5029439	0.076424	0.168278	0.8282572	0.6282251	0.4931529	0.4605009	0.4485808

Word2Vec Matcher DS2 and MS

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0.076516	0.5954797	0.6042732	0.2817032	0.3202126	0.5551523	0.2225531
Address	0.2316228	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0.0336224	0.5541899
City	0.5206539	0.1434778	0.5500144	0.027403	0.7216145	0.413726	0.732459	0.3871426	0.604471	0.3837541
ZIPCode	0.9151975	0.7624892	0.6591803	0.292355	0.1060572	1	0.3755258	0.3260489	0.8216316	0.8076475
County	0.0598117	0.6937756	0.9871757	0.434983	0.1627002	0.8939023	0.218353	0.3922385	0.9288751	0.9495708
Country	0.4324153	0.4525174	0.1840848	0.116098	0.5349092	0.4187854	0.4420324	0.9558367	0.7551859	0.7244656
Website	0.6616418	0.3983013	0.0633616	0.527498	0.0918425	0.4649473	0.0049886	0.6582168	0.6018162	0.8127478
TotalEmployee	0.3346628	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0.1902264	0.4191538	0.5785432	0.6477537
UniversityName	0.4320619	0.8204025	0.8326359	0.778063	0.1067257	0.1746027	0.2593956	0.2537462	0.2451263	0.0651382
Url	0.5770694	1	0.3197724	0.973771	0.8894954	0.8009834	0.2656226	0.0436453	0.1827011	0.4329506
EmployeeCount	0.2111921	0.1681117	0.2940051	0.999686	0.2841621	0.8025041	0.3365679	0.4620293	0.8733068	0.5240171
Uni	0.0275744	0.4448508	0.5865268	0.108421	0.0829025	0.2650847	0.3399365	0.9770957	0.6568788	0.698761
Location	0.0596662	0.917602	0.7714193	0.166395	0.2613776	0.9608816	0.209736	0.2670432	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0.3668131	0.067297	0.8317456	0.1072138	0.3227397	0.9758965	0.8737481	0.6970036

MS/DS2	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.752677	0.888258	0.5936356	0.516358034	0.2814907	0.3858364	0.223614	0.1487153	0.8701889	0.1495783
Address	0.7622372	0.5069838	0.1427436	0.1170257	0.465142086	0.606668	0.6979949	0.7756804	0.4649366	0.5459885	0.6958357
City	0.3340963	0.2307437	0.6708038	0.5247858	0.816999326	0.3347004	0.8188282	0.1799399	0.4352849	0.6997791	0.0943358
ZIPCode	0.1800447	0.9136315	0.7008141	0.4495984	0.469177902	0.3210436	0.0410306	0.8249529	0.6062852	0.9419251	0.7456657
County	0.9112455	0.0670437	0.8337427	0.1775813	0.349898997	0.2046271	0.7155149	0.6364887	0.2514421	0.3129464	0.3906298
Country	0.7039374	0.2090052	0.5397603	0.6628521	0.514579069	0.3301529	0.0776715	1	0.9808873	0.1553797	0.193914
Website	0.4461419	0.7916267	0.0810533	0.2916591	0.206157207	0.7049967	0.8946663	0.8628775	0.9213003	0.9481751	0.6894031
TotalEmployee	0.7261335	0.0183608	0.7862071	0.0196103	0.701319206	0.6591873	0.1029699	0.7052238	0.6050273	0.3100573	0.5163255
UniversityName	0.3837719	0.9298506	0.2924313	0.171838	0.428310821	0.9412394	0.44334	0.1344476	0.1089263	0.9165222	0.3729457
Url	0.1488488	0.9957112	0.6789154	0.1216574	0.257964023	0.8204019	0.9442562	0.7076351	0.5347282	0.6935514	0.6075847
EmployeeCount	0.0175038	0.9238061	0.1413813	0.3523177	0.666906574	0.0665237	0.5661145	0.2299738	0.0496781	0.4392192	0.451797
Uni	0.0897842	0.4058381	0.5478065	0.6010489	0.347087417	0.7084516	0.036786	0.6513357	0.7780713	0.5406696	0.4002897
Location	0.0894067	1	0.9089393	0.9905489	0.059718584	0.3140834	0.4039054	0.5959151	0.0997399	0.7513165	0.0482468
PostalCode	0.2156155	0.1129651	0.8712732	0.8484077	0.730201163	0.6280397	0.4289705	0.6374902	0.892385	0.0769934	0.0583747

All of Python codes Uploaded in The GIT folder of the group.

(5th and 6th week)- Try to use different combining strategies to combine results of matchers

In this Step we combined the result of Matchers for every Dataset and produced a matrix with different strategies.

DS1 Minimum

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.634555456	0.089736	0.3245747	0.74622	0.9074752	0.6476593	0.4319923	0.5171839	0.6545152	0.7331401	0.6927876
Address	0.289899934	0.5540112	0.246789	0.779704	0.4420466	0.8743102	0.4854393	0.2848636	0.5307932	0.9941882	0.5247523
City	0.331110515	0.1400411	0.364252	0.292008	0.3461347	0.6457235	0.4024812	0.5871217	0.4752701	0.9072654	0.927023
ZIPCode	0.013419817	0.2524527	0.3057523	0.698679	0.3134992	0.2397125	0.7017323	0.9982094	0.8894513	0.2954571	0.7965953
County	0.515454725	0.9523777	0.2333209	0.383794	0.8457988	0.9802428	0.6701966	0.643331	0.0459111	0.0353721	0.840597
Country	0.057871688	0.031331	0.9288207	0.978269	0.6309266	0.6298872	0.3739354	0.5520151	0.2146047	0.7126762	0.7837023
Website	0.245753082	0.6563975	0.9230298	0.28232	0.3465947	0.3737017	0.6398254	0.9922613	0.8866679	0.3254407	0.0102584
TtlEmployee	0.892719107	0.7959886	0.6112571	0.190939	0.7203741	0.131781	0.181269	0.3842196	0.5383169	0.8409598	0.8119018
Url	0.124198153	1	0.8180427	0.151107	0.6669613	0.6682789	0.5501621	0.3969838	0.6912101	0.3768339	0.4410553
EmployeeCount	0.901053317	0.1478142	0.9517982	0.908579	0.4236875	0.9841111	0.4993802	0.9798605	0.6489452	0.5049156	0.2556893
Uni	0.578056784	0.4836775	0.2075911	0.130181	0.2553082	0.5046203	0.2100689	0.0626319	0.2316638	0.878326	0.6177562
Location	0.910937738	0.1774053	0.8178076	0.210628	0.9731672	0.0357486	0.1309912	0.0522692	0.4268443	0.8707311	0.764517

DS1 Maximum

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.789844986	0.999257	0.835377 2	0.562810 9	0.6933034	0.6403902	0.8886379	0.770402	0.5654526	0.8559591	0.8618123
Address	0.654836614	0.733200 7	0.678909 8	0.824838	0.9362669	0.5846444	0.9551012	0.3162587	0.8574154	0.6520479	0.4718996
City	0.187262052	0.708858 1	0.765361 4	0.912276 4	0.8561681	0.7865758	0.9024173	0.9386119	0.9471065	0.9838066	0.8538167
ZIPCode	0.301984566	0.698825 4	0.216373 8	0.709904	0.7361504	0.7397629	0.4701641	0.9080144	0.815552	0.5609538	0.9348869
County	0.912395626	0.415584 2	0.596918 9	0.998206	0.7298056	0.6309726	0.8180751	0.898576	0.519291	0.7876385	0.8199763
Country	0.273797922	0.769352	0.438599 7	0.615014 2	0.6270405	0.7734751	0.6213059	0.7202598	0.5140299	0.8094303	0.7212745
Website	0.797278346	0.865270 6	0.873816	0.615823 4	0.9188768	0.9011806	0.7571848	0.8636428	0.8362316	0.5517528	0.4749736
TtlEmployee	0.552580737	0.685388 5	0.965807 7	0.835592 3	0.9822097	0.9060598	0.8456925	0.9635779	0.8851572	0.9424559	0.1370609
Url	0.947083299	1	0.959997 8	0.646629 8	0.4625297	0.8852107	0.9939674	0.9842583	0.8857413	0.5603404	0.6533406
EmployeeCount	0.795233046	0.521088 3	0.763153 1	0.743881	0.8752118	0.725071	0.8668097	0.7910236	0.7198877	0.9767102	0.9547757
Uni	0.681332627	0.717253 3	0.813322 4	0.664802 6	0.6451214	0.9723056	0.9566444	0.9319087	0.9568132	0.8143558	0.7981361
Location	0.992193647	0.856632 6	1	0.834970 7	0.7866116	0.6740636	0.8282572	0.6468169	0.7684317	0.4605009	0.4485808

DS1 Average

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.731322631	0.7262198	0.4241069	0.3357911	0.4528728	0.5147993	0.4492072	0.5221822	0.5031719	0.7440393	0.6225752
Address	0.514512091	0.4678041	0.3564567	0.4527434	0.4245804	0.3578633	0.6324298	0.1959476	0.6098237	0.4901107	0.3582753
City	0.11466689	0.4535187	0.4804964	0.7850872	0.533298	0.4950094	0.5973545	0.6628484	0.7623065	0.4715521	0.624594
ZIPCode	0.160378195	0.5478313	0.1350227	0.4812353	0.2938069	0.4691702	0.1877206	0.3212702	0.7767374	0.2876657	0.4439287
County	0.690800165	0.2793012	0.3766156	0.5280516	0.5757211	0.5371122	0.6715579	0.5982146	0.3227331	0.4750405	0.548588
Country	0.118672316	0.5388388	0.2510355	0.2918169	0.3091217	0.509092	0.437862	0.6581046	0.3311204	0.6839586	0.6554946
Website	0.695707391	0.4839684	0.7336275	0.4592237	0.7704103	0.3561938	0.3868757	0.6118873	0.4821564	0.2803906	0.4288892
TtlEmployee	0.475484008	0.3642245	0.8258949	0.5875563	0.6759129	0.8390993	0.6814271	0.5174132	0.6277279	0.567145	0.1134206
Url	0.783956423	0.8105938	0.7499979	0.4144773	0.3287579	0.6580793	0.618732	0.5779706	0.8491104	0.3794317	0.3871239
EmployeeCount	0.582323232	0.2437412	0.446344	0.3684754	0.7457575	0.567694	0.7268491	0.7064281	0.3435318	0.5045269	0.6971209
Uni	0.481512852	0.5854553	0.3334567	0.5119701	0.5312802	0.6784316	0.5798836	0.613463	0.7573104	0.3295077	0.5305605
Location	0.665454677	0.5782943	0.8493629	0.7032175	0.5479168	0.4246747	0.7597479	0.5988775	0.5411224	0.3305753	0.2575527

DS2 Minimum

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0.076516	0.5954797	0.6042732	0.2817032	0.3202126	0.5551523	0.2225531
Address	0.2316228	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0.0336224	0.5541899
City	0.5206539	0.1434778	0.5500144	0.027403	0.7216145	0.413726	0.732459	0.3871426	0.604471	0.3837541
ZIPCode	0.9151975	0.7624892	0.6591803	0.292355	0.1060572	1	0.3755258	0.3260489	0.8216316	0.8076475
County	0.0598117	0.6937756	0.9871757	0.434983	0.1627002	0.8939023	0.218353	0.3922385	0.9288751	0.9495708
Country	0.4324153	0.4525174	0.1840848	0.116098	0.5349092	0.4187854	0.4420324	0.9558367	0.7551859	0.7244656
Website	0.6616418	0.3983013	0.0633616	0.527498	0.0918425	0.4649473	0.0049886	0.6582168	0.6018162	0.8127478
TotalEmployee	0.3346628	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0.1902264	0.4191538	0.5785432	0.6477537
UniversityName	0.4320619	0.8204025	0.8326359	0.778063	0.1067257	0.1746027	0.2593956	0.2537462	0.2451263	0.0651382
Url	0.5770694	1	0.3197724	0.973771	0.8894954	0.8009834	0.2656226	0.0436453	0.1827011	0.4329506
EmployeeCount	0.2111921	0.1681117	0.2940051	0.999686	0.2841621	0.8025041	0.3365679	0.4620293	0.8733068	0.5240171
Uni	0.0275744	0.4448508	0.5865268	0.108421	0.0829025	0.2650847	0.3399365	0.9770957	0.6568788	0.698761
Location	0.0596662	0.917602	0.7714193	0.166395	0.2613776	0.9608816	0.209736	0.2670432	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0.3668131	0.067297	0.8317456	0.1072138	0.3227397	0.9758965	0.8737481	0.6970036

DS2 Maximum

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0.076516	0.5954797	0.6042732	0.2817032	0.3202126	0.5551523	0.2225531
Address	0.2316228	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0.0336224	0.5541899
City	0.5206539	0.1434778	0.5500144	0.027403	0.7216145	0.413726	0.732459	0.3871426	0.604471	0.3837541
ZIPCode	0.9151975	0.7624892	0.6591803	0.292355	0.1060572	1	0.3755258	0.3260489	0.8216316	0.8076475
County	0.0598117	0.6937756	0.9871757	0.434983	0.1627002	0.8939023	0.218353	0.3922385	0.9288751	0.9495708
Country	0.4324153	0.4525174	0.1840848	0.116098	0.5349092	0.4187854	0.4420324	0.9558367	0.7551859	0.7244656
Website	0.6616418	0.3983013	0.0633616	0.527498	0.0918425	0.4649473	0.0049886	0.6582168	0.6018162	0.8127478
TotalEmployee	0.3346628	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0.1902264	0.4191538	0.5785432	0.6477537
UniversityName	0.4320619	0.8204025	0.8326359	0.778063	0.1067257	0.1746027	0.2593956	0.2537462	0.2451263	0.0651382
Url	0.5770694	1	0.3197724	0.973771	0.8894954	0.8009834	0.2656226	0.0436453	0.1827011	0.4329506
EmployeeCount	0.2111921	0.1681117	0.2940051	0.999686	0.2841621	0.8025041	0.3365679	0.4620293	0.8733068	0.5240171
Uni	0.0275744	0.4448508	0.5865268	0.108421	0.0829025	0.2650847	0.3399365	0.9770957	0.6568788	0.698761
Location	0.0596662	0.917602	0.7714193	0.166395	0.2613776	0.9608816	0.209736	0.2670432	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0.3668131	0.067297	0.8317456	0.1072138	0.3227397	0.9758965	0.8737481	0.6970036

DS2 Average

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0.076516	0.5954797	0.6042732	0.2817032	0.3202126	0.5551523	0.2225531
Address	0.2316228	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0.0336224	0.5541899
City	0.5206539	0.1434778	0.5500144	0.027403	0.7216145	0.413726	0.732459	0.3871426	0.604471	0.3837541
ZIPCode	0.9151975	0.7624892	0.6591803	0.292355	0.1060572	1	0.3755258	0.3260489	0.8216316	0.8076475
County	0.0598117	0.6937756	0.9871757	0.434983	0.1627002	0.8939023	0.218353	0.3922385	0.9288751	0.9495708
Country	0.4324153	0.4525174	0.1840848	0.116098	0.5349092	0.4187854	0.4420324	1	0.7551859	0.7244656
Website	0.6616418	0.3983013	0.0633616	0.527498	0.0918425	0.4649473	0.0049886	0.6582168	0.6018162	0.8127478
TotalEmployee	0.3346628	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0.1902264	0.4191538	0.5785432	0.6477537
UniversityName	0.4320619	0.8204025	0.8326359	0.778063	0.1067257	0.1746027	0.2593956	0.2537462	0.2451263	0.0651382
Url	0.5770694	1	0.3197724	0.973771	0.8894954	0.8009834	0.2656226	0.0436453	0.1827011	0.4329506
EmployeeCount	0.2111921	0.1681117	0.2940051	0.999686	0.2841621	0.8025041	0.3365679	0.4620293	0.8733068	0.5240171
Uni	0.0275744	0.4448508	0.5865268	0.108421	0.0829025	0.2650847	0.3399365	0.9770957	0.6568788	0.698761
Location	0.0596662	0.917602	0.7714193	0.166395	0.2613776	0.9608816	0.209736	0.2670432	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0.3668131	0.067297	0.8317456	0.1072138	0.3227397	0.9758965	0.8737481	0.6970036

DS3 Minimum

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.5069838	0.1427436	0.1170257	0.465142086	0.2814907	0.3858364	0.223614	0.1487153	0.5459885	0.1495783
Address	0.3340963	0.2307437	0.1427436	0.1170257	0.465142086	0.3347004	0.6979949	0.1799399	0.4352849	0.5459885	0.0943358
City	0.1800447	0.2307437	0.6708038	0.4495984	0.469177902	0.3210436	0.0410306	0.1799399	0.4352849	0.6997791	0.0943358
ZIPCode	0.1800447	0.0670437	0.7008141	0.1775813	0.349898997	0.2046271	0.0410306	0.6364887	0.2514421	0.3129464	0.3906298
County	0.7039374	0.0670437	0.5397603	0.1775813	0.349898997	0.2046271	0.0776715	0.6364887	0.2514421	0.1553797	0.193914
Country	0.4461419	0.2090052	0.0810533	0.2916591	0.206157207	0.3301529	0.0776715	0.8628775	0.9213003	0.1553797	0.193914
Website	0.4461419	0.0183608	0.0810533	0.0196103	0.206157207	0.6591873	0.1029699	0.7052238	0.6050273	0.3100573	0.5163255
TotalEmployee	0.3837719	0.0183608	0.2924313	0.0196103	0.428310821	0.6591873	0.1029699	0.1344476	0.1089263	0.3100573	0.3729457
UniversityName	0.1488488	0.9298506	0.2924313	0.1216574	0.257964023	0.8204019	0.44334	0.1344476	0.1089263	0.6935514	0.3729457
Url	0.0175038	0.9238061	0.1413813	0.1216574	0.257964023	0.0665237	0.5661145	0.2299738	0.0496781	0.4392192	0.451797
EmployeeCount	0.0175038	0.4058381	0.1413813	0.3523177	0.347087417	0.0665237	0.036786	0.2299738	0.0496781	0.4392192	0.4002897
Uni	0.0894067	0.4058381	0.5478065	0.6010489	0.059718584	0.3140834	0.036786	0.5959151	0.0997399	0.5406696	0.0482468
Location	0.0894067	0.1129651	0.8712732	0.8484077	0.059718584	0.3140834	0.4039054	0.5959151	0.0997399	0.0769934	0.0482468
PostalCode	0.2156155	0.1129651	0.8712732	0.8484077	0.730201163	0.6280397	0.4289705	0.6374902	0.892385	0.0769934	0.0583747

DS3 Maximum

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.752677	0.888258	0.5936356	0.516358034	0.2814907	0.3858364	0.223614	0.1487153	0.8701889	0.1495783
Address	0.7622372	0.5069838	0.1427436	0.1170257	0.465142086	0.606668	0.6979949	0.7756804	0.4649366	0.5459885	0.6958357
City	0.3340963	0.2307437	0.6708038	0.5247858	0.816999326	0.3347004	0.8188282	0.1799399	0.4352849	0.6997791	0.0943358
ZIPCode	0.1800447	0.9136315	0.7008141	0.4495984	0.469177902	0.3210436	0.0410306	0.8249529	0.6062852	0.9419251	0.7456657
County	0.9112455	0.0670437	0.8337427	0.1775813	0.349898997	0.2046271	0.7155149	0.6364887	0.2514421	0.3129464	0.3906298
Country	0.7039374	0.2090052	0.5397603	0.6628521	0.514579069	0.3301529	0.0776715	1	0.9808873	0.1553797	0.193914
Website	0.4461419	0.7916267	0.0810533	0.2916591	0.206157207	0.7049967	0.8946663	0.8628775	0.9213003	0.9481751	0.6894031
TotalEmployee	0.7261335	0.0183608	0.7862071	0.0196103	0.701319206	0.6591873	0.1029699	0.7052238	0.6050273	0.3100573	0.5163255
UniversityName	0.659876	0.9298506	0.2924313	0.171838	0.428310821	0.9412394	0.44334	0.1344476	0.1089263	0.9165222	0.3729457
Url	0.1488488	0.9957112	0.6789154	0.1216574	0.257964023	0.8204019	0.9442562	0.7076351	0.5347282	0.6935514	0.6075847
EmployeeCount	0.0175038	0.9238061	0.1413813	0.3523177	0.666906574	0.0665237	0.5661145	0.2299738	0.0496781	0.4392192	0.451797
Uni	0.0897842	0.4058381	0.5478065	0.6010489	0.347087417	0.7084516	0.036786	0.6513357	0.7780713	0.5406696	0.4002897
Location	0.0894067	1	0.9089393	0.9905489	0.059718584	0.3140834	0.4039054	0.5959151	0.0997399	0.7513165	0.0482468
PostalCode	0.2156155	0.1129651	0.8712732	0.8484077	0.730201163	0.6280397	0.4289705	0.6374902	0.892385	0.0769934	0.0583747

DS3 Average

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.658457467	0.6707793	0.6397532	0.434765633	0.499286051	0.3898831	0.4898892	0.4076361	0.2541224	0.7621221	0.3316641
Address	0.619523567	0.4149038	0.3187637	0.252945733	0.582427833	0.5160121	0.7382727	0.5771002	0.4550527	0.597252	0.4953357
City	0.282745767	0.458373	0.6808072	0.499723333	0.701058851	0.3301481	0.5595623	0.3949442	0.492285	0.7804944	0.3114457
ZIPCode	0.4237783	0.6314356	0.7451236	0.358926033	0.429418267	0.2822381	0.2658587	0.7621315	0.4880042	0.7322655	0.6273204
County	0.8421428	0.1143642	0.7357486	0.339338233	0.404792354	0.246469	0.5029004	0.7576591	0.4945905	0.2604242	0.3250579
Country	0.618005567	0.4032124	0.386858	0.5391211	0.411771782	0.4551008	0.3500031	0.9542925	0.961025	0.4196448	0.359077
Website	0.539472433	0.5338714	0.3161046	0.200976167	0.371211207	0.6897269	0.6307675	0.8103263	0.815876	0.7354692	0.6317106
TotalEmployee	0.612012967	0.3221907	0.6216152	0.070352867	0.610316411	0.7532047	0.2164266	0.5149651	0.4396603	0.5122123	0.4685322
UniversityName	0.3974989	0.9518041	0.4212593	0.155111133	0.371528555	0.9009602	0.6103121	0.3255101	0.2508603	0.8421986	0.4511587
Url	0.105067133	0.9717428	0.4997374	0.198544167	0.394278207	0.5691092	0.818209	0.5484147	0.3730448	0.608774	0.5556554
EmployeeCount	0.041597267	0.7511501	0.2768564	0.4352281	0.560300188	0.2804997	0.3896717	0.3704278	0.2924758	0.473036	0.4346279
Uni	0.089658367	0.6038921	0.6681841	0.730882233	0.251297806	0.5769955	0.1591591	0.6328622	0.5519608	0.6108852	0.2829421
Location	0.1314763	0.7043217	0.8963839	0.9431685	0.283212777	0.4187355	0.4122604	0.6097735	0.3639549	0.5265421	0.0516228
PostalCode	0.2156155	0.1129651	0.8712732	0.8484077	0.730201163	0.6280397	0.4289705	0.6374902	0.892385	0.0769934	0.0583747

After this step we applied the threshold on each matrix and produced new matrix.

Then we applied threshold on each matrix.

Threshold for max – DS1

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.789844986	0	0.9358512	0.7334932	0	0.9777371	0	0.8924066	0	0.7657025	0.860368
Address	0	0	0.6789098	0.8683538	0.930865	0	0.6861655	0.7374401	0.7712661	0.6228443	0.8437689
City	0	0.7611701	0.8674598	0	0	0.632809	0.9830337	0.7104953	0.6245473	0.69081	0.7797421
ZIPCode	0.774055327	0.8727585	0.6636417	0.709904	0	0.5746842	0	0.9728423	0.6854505	0.8872501	0.8936362
County	0.692954304	0.7916042	0	0.8057974	0.7298056	0	0.4625101	0.9816941	0.9390452	0.6207081	0.9256343
Country	0	0	0	0.9056164	0	0	0.813759	0.5040907	0.773802	0	0.9566595
Website	0	0	0.9940567	0.9366829	0.67277	0.8080801	0.625086	0.7352128	0	0.9276438	0
TtlEmployee	0.792903805	0	0.6487169	0.8222383	0.8692206	0.8424727	0.6383878	0	0.5682031	0.6662852	0.7863169
Url	0	1	0.8620616	0	0.586102	0.926832	0	0.7780858	0.6177655	0.5064122	0.9893704
EmployeeCount	0.898726858	0.573376	0.9073825	0	0	0.789068	0.8585482	0.767232	0.8583938	0.6235286	0.8521358
Uni	0.741934257	0.8646108	0.9288302	0.50767	0.5907983	0.9800108	0.8698466	0.5822563	0.7816911	0.7513325	0.6490642
Location	0	0.6588297	1	0.657227	0	0.8569763	0.8797685	0.5340473	0.7993851	0.8084392	0.9506017

Threshold for min – DS1

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.634555456	0	0	0.74622	0.9074752	0.6476593	0	0.5171839	0.6545152	0.7331401	0.6927876
Address	0	0.5540112	0	0.779704	0	0.8743102	0	0	0.5307932	0.9941882	0.5247523
City	0	0	0	0	0	0.6457235	0	0.5871217	0	0.9072654	0.927023
ZIPCode	0	0	0	0.698679	0	0	0.7017323	0.9982094	0.8894513	0	0.7965953
County	0.515454725	0.9523777	0	0	0.8457988	0.9802428	0.6701966	0.643331	0	0	0.840597
Country	0	0	0.9288207	0.978269	0.6309266	0.6298872	0	0.5520151	0	0.7126762	0.7837023
Website	0	0.6563975	0.9230298	0	0	0	0.6398254	0.9922613	0.8866679	0	0
TtlEmployee	0.892719107	0.7959886	0.6112571	0	0.7203741	0	0	0	0.5383169	0.8409598	0.8119018
Url	0	1	0.8180427	0	0.6669613	0.6682789	0.5501621	0	0.6912101	0	0
EmployeeCount	0.901053317	0	0.9517982	0.908579	0	0.9841111	0	0.9798605	0.6489452	0.5049156	0
Uni	0.578056784	0	0	0	0	0.5046203	0	0	0	0.878326	0.6177562
Location	0.910937738	0	0.8178076	0	0.9731672	0	0	0	0	0.8707311	0.764517

Threshold for max – DS2

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0	0.5954797	0.6042732	0	0	0.5551523	0
Address	0	0.9381071	0.6802218	0.893423	0.5606502	0.7015033	0.6648607	0.4872719	0	0.5541899
City	0.5206539	0	0.5500144	0	0.7216145	0	0.732459	0	0.604471	0
ZIPCode	0.9151975	0.7624892	0.6591803	0	0	1	0	0	0.8216316	0.8076475
County	0	0.6937756	0.9871757	0	0	0.8939023	0	0	0.9288751	0.9495708
Country	0	0.4525174	0	0	0.5349092	0	0	1	0.7551859	0.7244656
Website	0.6616418	0	0	0.527498	0	0	0	0.6582168	0.6018162	0.8127478
TotalEmployee	0	0.6087506	0.9482953	0.504418	0.5738125	0.5388893	0	0	0.5785432	0.6477537
UniversityName	0.7689866	0.8204025	0.8326359	0.778063	0	0	0	0	0	0
Url	0.5770694	1	0	0.973771	0.8894954	0.8009834	0	0	0	0
EmployeeCount	0	0	0	0.999686	0	0.8025041	0	0	0.8733068	0
Uni	0	0	0.5865268	0	0	0	0	0.9770957	0.6568788	0.698761
Location	0	0.917602	0.7714193	0	0	0.9608816	0	0	0.795645	0
PostalCode	0.5776156	0.8420311	0	0	0.8317456	0	0	0.9758965	0.8737481	0.6970036

Threshold for min – DS2

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	0	0.6657857	0.6802218	0	0.5606502	0.6042732	0	0	0	0
Address	0	0	0	0	0.5606502	0	0.6648607	0	0	0
City	0.5206539	0	0	0	0	0	0	0	0.604471	0
ZIPCode	0	0.6937756	0	0	0	0.8939023	0	0	0.8216316	0.8076475
County	0	0	0	0	0	0	0	0	0.7551859	0.7244656
Country	0	0	0	0	0	0	0	0.6582168	0.6018162	0.7244656
Website	0	0	0	0.504418	0	0	0	0	0.5785432	0.6477537
TotalEmployee	0	0.6087506	0.8326359	0.504418	0	0	0	0	0	0
UniversityName	0	0.8204025	0	0.778063	0	0	0	0	0	0
Url	0	0	0	0.973771	0	0.8009834	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0.6568788	0.5240171
Uni	0	0	0.5865268	0	0	0	0	0	0.6568788	0.5278384
Location	0	0.8420311	0	0	0	0	0	0	0.795645	0.5278384
PostalCode	0.5776156	0.8420311	0	0	0.8317456	0	0	0.9758965	0.8737481	0.6970036

Threshold for Average – DS2

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0.6657857	0.9087176	0	0.5954797	0.6042732	0	0	0.5551523	0
Address	0	0.8473333	0.7563871	0.6211207	0.57226	0.6690933	0	0	0	0
City	0	0	0.5934169	0	0.6679597	0.5096517	0.7099262	0	0	0
ZIPCode	0.783683	0.5561521	0.6227917	0	0	0.8045753	0	0	0.7492447	0.6663497
County	0	0.7166801	0.8778439	0	0	0.9292682	0	0.3701753	0.8931273	0.902263
Country	0	0.5329368	0	0	0	0.5771577	0	0.7826917	0.8130823	0.7995007
Website	0.5852329	0	0	0	0	0	0	0.7623305	0.6529394	0.7833204
TotalEmployee	0	0.5386008	0.6533174	0.5121113	0	0.514242	0	0	0.5863009	0.7027517
UniversityName	0	0.7498519	0.871189	0	0	0	0	0	0	0
Url	0.5287336	0.9401342	0	0	0.6285722	0.5921898	0	0	0	0
EmployeeCount	0	0	0	0	0	0.8019972	0	0	0.6431049	0
Uni	0.0887803	0	0	0	0	0	0	0.8054069	0.7290215	0.640513
Location	0	0.7600183	0.7097885	0	0	0.7289493	0	0.5037274	0.7493896	0.5848126
PostalCode	0	0.8672214	0.5016818	0.1003297	0.641623	0	0	0.7396121	0.8477137	0.6406152

Threshold for max – DS3

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.752677	0.888258	0.5936356	0.516358034	0	0	0	0	0.8701889	0
Address	0.7622372	0.5069838	0	0	0	0.606668	0.6979949	0.7756804	0	0.5459885	0.6958357
City	0	0	0.6708038	0.5247858	0.816999326	0	0.8188282	0	0	0.6997791	0
ZIPCode	0	0.9136315	0.7008141	0	0	0	0	0.8249529	0.6062852	0.9419251	0.7456657
County	0.9112455	0	0.8337427	0	0	0	0.7155149	0.6364887	0	0	0
Country	0.7039374	0	0.5397603	0.6628521	0.514579069	0	0	1	0.9808873	0	0
Website	0	0.7916267	0	0	0	0.7049967	0.8946663	0.8628775	0.9213003	0.9481751	0.6894031
TotalEmployee	0.7261335	0	0.7862071	0	0.701319206	0.6591873	0	0.7052238	0.6050273	0	0.5163255
UniversityName	0	0.9298506	0	0	0	0.9412394	0	0	0	0.9165222	0
Url	0	0.9957112	0.6789154	0	0	0.8204019	0.9442562	0.7076351	0.5347282	0.6935514	0.6075847
EmployeeCount	0	0.9238061	0	0	0.666906574	0	0.5661145	0	0	0	0
Uni	0	0	0.5478065	0.6010489	0	0.7084516	0	0	0.7780713	0.5406696	0
Location	0	1	0.9089393	0.9905489	0	0	0	0	0	0.7513165	0
PostalCode	0	0	0.8712732	0.8484077	0.730201163	0.6280397	0	0.6374902	0.892385	0	0

Threshold for min – DS3

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0.5069838	0	0	0	0	0	0	0	0.5459885	0
Address	0	0	0	0	0	0	0.6979949	0	0	0.5459885	0
City	0	0	0.6708038	0	0	0	0	0	0	0.6997791	0
ZIPCode	0	0	0.7008141	0	0	0	0	0.6364887	0	0	0
County	0.7039374	0	0.5397603	0	0	0	0	0.6364887	0	0	0
Country	0	0	0	0	0	0	0	0.8628775	0.9213003	0	0
Website	0	0	0	0	0	0.6591873	0	0.7052238	0.6050273	0	0.5163255
TotalEmployee	0	0	0	0	0	0.6591873	0	0	0	0	0
UniversityName	0	0.9298506	0	0	0	0.8204019	0	0	0	0.6935514	0
Url	0	0.9238061	0	0	0	0	0.5661145	0	0	0.4392192	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0.4392192	0
Uni	0	0	0.5478065	0.6010489	0	0	0	0.5959151	0.0997399	0.5406696	0
Location	0	0	0.8712732	0.8484077	0	0	0	0.5959151	0.0997399	0	0
PostalCode	0	0	0.8712732	0.8484077	0.730201163	0.6280397	0	0.6374902	0.892385	0	0

Threshold for Average – DS3

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.531083589	0	0.7029418	0	0.522438627	0	0	0	0	0.602021	0
Address	0.544052467	0	0	0	0	0	0	0.8002298	0.624747	0	0.50569
City	0.540793667	0	0	0	0	0	0	0.5476076	0.5116342	0	0
ZIPCode	0	0	0	0	0	0	0	0.6378977	0.5587455	0.5572091	0
County	0.533188911	0.5102552	0	0	0	0.5715853	0.5965408	0.5270875	0	0.5806557	0
Country	0	0	0	0	0	0	0	0.5816463	0	0	0
Website	0	0.716423	0	0	0	0.5574935	0	0	0	0.7431976	0.5136172
TotalEmployee	0	0	0	0	0	0.5141266	0	0.6165179	0	0	0
UniversityName	0	0.5979886	0	0	0	0.5119408	0	0	0	0	0
Url	0	0.5708561	0.7394576	0.566982444	0	0.6751457	0.5107953	0.6593292	0.6596913	0	0
EmployeeCount	0	0.8140639	0.5188826	0.6477431	0	0	0	0	0	0	0
Uni	0	0	0.7095399	0.7247283	0.53864429	0.6682457	0	0.644413	0.8352282	0	0
Location	0	1	0.9089393	0.9905489	0	0	0	0.5959151	0	0.7513165	0
PostalCode	0	0	0.8712732	0.8484077	0.730201163	0.6280397	0	0.6374902	0.892385	0	0

(7th week)- Develop different selection strategies

In week 7 we applied one to one mapping for every Matrix

DS1- Max Matrix

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.976205529	0	0	0	0	0	0	0	0	0	0
Address	0	0	1	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0.7665423	0.8313184	0.7842314	0.9682237	0	0.9526234
ZIPCode	0	0	0	0.987771	0	0	0	0	0	0	0
County	0	0	0	0	0.9909821	0	0	0	0	0	0
Country	0	0	0	0.987771	0	0.9741388	0.8535081	0.9466806	0.9836628	0.4006091	0.6286253
TtlEmployee	0.954253067	0	0	0.8340656	0	0.7837996	0	0	0	0	0
Url	0	1	0	0	0	0.9544792	0	0.7703468	0.611589	0.9620293	0.9213293
EmployeeCount	0	0.7967426	0	0	0	0.3757537	0.9523789	0	0	0	0
Uni	0.976205529	0	0	0	0	0	0	0	0	0	0
Location	0	0	0.8533409	0	0	0	0	0	0	0	0

DS1- Min Matrix

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.634555456	0	0	0	0	0	0	0	0	0	0
Address	0	0	0	0	0	0.8743102	0	0	0.5307932	0.9941882	0.5247523
City	0	0	0	0	0	0.6457235	0	0.5871217	0	0.9072654	0.927023
ZIPCode	0	0	0	0.978269	0	0	0	0	0	0	0
County	0	0	0	0	0.8457988	0	0.6701966	0.643331	0	0	0.840597
Country	0	0	0.9288207	0	0.6309266	0.6298872	0	0.5520151	0	0.7126762	0.7837023
TtlEmployee	0		0	0	0.7203741	0	0.7959886	0	0	0	0
Url	0	1	0	0	0.6669613	0.6682789	0.5501621	0	0.6912101	0	0
EmployeeCount	0	0	0	0	0	0.9841111	0	0.9798605	0.6489452	0.5049156	0
Uni	0.578056784	0.4836775	0	0	0	0.5046203	0	0	0	0.878326	0.6177562
Location	0	0	0.9731672	0	0	0	0	0	0	0.8707311	0.764517

DS2- Max Matrix

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	0.856398	0	0	0	0	0	0	0	0	0	0
Address	0	0	0.9836321	0	0		0	0	0	0	0
City	0	0.5361305	0	0	0.8907828	0.7665423	0.8313184	0.7842314	0.9682237	0.3081854	0.9526234
ZIPCode	0.976205529	0	0	0.709904	0	0.8917641	0	0.6593337	0.4543925	0.8420105	0.8636825
County	0.211623877	0.6087686	0	0	0.7298056	0	0.1618662	0.8060083	0.4710539	0.739153	0.333587
Country	0	0	0	0	0.9909821	0	0	0	0	0	0
TtlEmployee	0.954253067	0	0	0	0	0.7837996	0	0	0.5965367	0.7207461	0.6122846
Url	0	1	0	0	0.5300148	0	0	0	0	0	0
EmployeeCount	0.549436469	0.7967426	0	0	0.6935372	0	0	0.975929	0.4921238	0.4993051	0.9032021
Uni	0.593022804	0.8776542	0	0		0	0.9523789	0	0.5094163	0	0.296649
Location	0.515506057	0.5452413	0	0	0	0.8533409	0	0.7351663	0	0.8919851	0.9063778

DS2- Min Matrix

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	0.6657857	0	0	0	0	0	0	0	0	0
Address	0	0	0.8326359	0	0	0	0	0	0	0
City	0	0	0	0	0.8317456	0	0	0	0	0
ZIPCode	0	0	0	0	0	0.8939023	0	0	0	0
County	0	0.4525174	0	0.116098	0.1627002	0	0.218353	0	0	0
Country	0	0.3983013	0	0.116098	0.0918425	0	0	0.9758965	0.6018162	0.7244656
TotalEmployee	0	0.6087506	0	0.504418	0.1067257	0	0.1902264	0	0.2451263	0.0651382
UniversityName	0.8204025	0.8204025	0	0	0	0	0	0	0	0
Url	0	0.973771	0	0	0	0	0	0	0	0
EmployeeCount	0	0.1681117	0	0.108421	0.0829025	0.2650847	0.3365679	0	0.6568788	0.5240171
Uni	0	0.4448508	0	0.108421	0.0829025	0.2650847	0.209736	0	0.6568788	0.5278384
Location	0	0.8420311	0	0.067297	0.2613776	0.1072138	0.209736	0	0.795645	0.5278384
PostalCode	0	0.8420311	0	0.067297	0.8317456	0.1072138	0.3227397	0	0.8737481	0.6970036

DS2- Average Matrix

MS/DS2	Name	Url	Location	Address2	City	ZIPCode	StudentCount	Country	Website	rank
Name	1	0	0	0	0	0	0	0	0	0
Address	0	0	0	0.8473333	0	0	0	0	0	0
City	0	0	0	0	0.7099262	0	0	0	0	0
ZIPCode	0.783683	0.5561521	0.6227917	0	0	0.8045753	0.4945035	0	0.7492447	0.6663497
County	0.3449403	0.7166801	0.8778439	0	0	0.9292682	0.270744	0	0.8931273	0.902263
Country	0	0	0	0	0	0	0	0.8130823	0	0
TotalEmployee	0.4436558	0.5386008	0.6533174	0	0	0.514242	0.1284804	0	0.5863009	0.7027517
UniversityName	0.871189	0	0	0	0	0	0	0	0	0
Url	0.5287336	0.9401342	0.4907269	0	0	0.5921898	0.263547	0	0.2035095	0.3103465
EmployeeCount	0.3331512	0.4454078	0.3025942	0	0	0.8019972	0.3129195	0	0.6431049	0.4936616
Uni	0.0887803	0.3526044	0.4890196	0	0	0.4442245	0.3388136	0	0.7290215	0.640513
Location	0.0489689	0.7600183	0.7097885	0	0	0.7289493	0.2531361	0	0.7493896	0.5848126
PostalCode	0.4049658	0.8672214	0.5016818	0	0	0.3917697	0.2850718	0	0.8477137	0.6406152

DS3- Max Matrix

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.9112455	0	0	0	0	0	0	0	0.6062852	0	0
Address	0	0.9136315	0	0	0	0	0	0	0	0	0
City	0	0.8188282	0	0	0	0	0	0	0	0.9419251	0
ZIPCode	0		0.7008141	0.4495984	0	0	0.0410306	0.8249529	0.6062852		0.7456657
County	0	0	0.8337427	0.1775813	0	0	0.7155149	0	0.2514421	0	0
Country	0	0	0.5397603	0.6628521	0	0	0.0776715	1	0	0	0
TotalEmployee	0	0	0.7862071	0	0	0	0.1029699	0.7052238	0	0	0
UniversityName	0	0	0	0	0	0	0.44334	0	0	0	0
Url	0	0	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	0	0	0.5478065	0	0	0	0	0	0	0	0
Location	0	1	0	0	0	0	0	0	0	0	0
PostalCode	0	0	0.8712732	0	0	0	0.4289705	0	0	0	0

DS3- Min Matrix

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.6065676	0	0	0	0	0	0	0	0	0	0
Address	0	0	0	0	0	0	0.6979949	0	0	0.5459885	0
City	0	0	0.6708038	0	0	0	0	0	0	0.6997791	0
ZIPCode	0	0	0.7008141	0	0	0	0	0.6364887	0	0	0.5163255
County	0.7039374	0	0.5397603	0	0	0	0	0	0	0	0
Country	0	0	0	0	0	0	0	0.8628775	0.9213003	0	0
TotalEmployee	0	0	0	0	0	0.6591873	0	0	0	0	0
UniversityName	0	0.9298506	0	0	0	0.8204019	0	0	0	0.6935514	0
Url	0	0.9238061	0	0	0	0	0.5661145	0	0	0.4392192	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0.4392192	0
Uni	0	0	0	0.6010489	0	0	0	0.5959151	0.0997399	0.5406696	0
Location	0	0	0	0	0	0	0	0.5959151	0	0	0
PostalCode	0	0	0	0.8484077	0.730201163	0.6280397	0	0.6374902	0	0	0

DS3- Average Matrix

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	0.531083589	0	0.7029418	0	0.522438627	0	0	0	0	0.602021	0
Address	0.544052467	0	0	0	0	0	0	0.8002298	0.624747	0	0.50569
City	0.540793667	0	0	0	0	0	0	0.5476076	0.5116342	0	0
ZIPCode	0	0	0	0	0	0	0	0.6378977	0.5587455	0.5572091	0
County	0.533188911	0.5102552	0	0	0	0.5715853	0.5965408	0.5270875	0	0.5806557	0
Country	0	0	0	0	0	0	0	0.5816463	0	0	0
TotalEmployee	0	0	0	0	0	0.5141266	0	0.6165179	0	0	0
UniversityName	0	0.5979886	0	0	0	0.5119408	0	0	0	0	0
Url	0	0.5708561	0.7394576	0.566982444	0	0.6751457	0.5107953	0.6593292	0.6596913	0	0
EmployeeCount	0	0.8140639	0.5188826	0.6477431	0	0	0	0	0	0	0
Uni	0	0	0.7095399	0.7247283	0.53864429	0.6682457	0	0.644413	0.8352282	0	0
Location	0	1	0.9089393	0.9905489	0	0	0	0.5959151	0	0.7513165	0
PostalCode	0	0	0.8712732	0.8484077	0.730201163	0.6280397	0	0.6374902	0.892385	0	0

We checked the ground truth matrix with the result matrix and calculated the performance

Precision=TP/TP+FP

RECALL=TP/TP+FN

RECALL=2.PRECISION.RECALL/(PRECISION+RECALL)

Ground Truth table 1

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkinUni	NumDisStudent
Name	1	0	0	0	0	0	0	0	0	0	0
Address	0	0	0	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	1	0	0	0	0	0	0	0
County	0	0	0	0	1	0	0	0	0	0	0
Country	0	0	0	0	0	0	0	0	0	0	0
TtlEmployee	0	0	0	0	0	0	0	0	0	0	0
Url	0	1	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	1	0	0	0	0	0	0	0	0	0	0
Location	0	0	1	0	0	0	0	0	0	0	0

Ground Truth Table 2

[illegible]

Ground Truth Table 3

[illegible]

Discuss which matchers, combinations, selection work best for your case

Ground Truth table 1

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	1	0	0	0	0	0	0	0	0	0	0
Address	0	0	0	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	1	0	0	0	0	0	0	0
County	0	0	0	0	1	0	0	0	0	0	0
Country	0	0	0	0	0	0	0	0	0	0	0
TtlEmployee	0	0	0	0	0	0	0	0	0	0	0
Url	0	1	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	1	0	0	0	0	0	0	0	0	0	0
Location	0	0	1	0	0	0	0	0	0	0	0

Ground Truth Table 2

[illegible]

Ground Truth Table 3

MS/DS3	Uni	Location	Rank	Description	TuitionFees	UndgradEnrll	Date	Country	ContactName	City	zip
Name	1	0	0	0	0	0	0	0	0	0	0
Address	0	1	0	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	0	0	0	0	0	0	0	1
County	0	0	0	0	0	0	0	0	0	0	0
Country	0	0	0	0	0	0	0	1	0	0	0
Website	0	0	0	0	0	0	0	0	0	0	0
TotalEmployee	0	0	0	0	0	0	0	0	0	0	0
UniversityName	1	0	0	0	0	0	0	0	0	0	0
Url	0	0	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	1	0	0	0	0	0	0	0	0	0	0
Location	0	1	0	0	0	0	0	0	0	0	0
PostalCode	0	0	0	0	0	0	0	0	0	0	1

DS1- Max Matrix

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudents
Name	0.976205529	0	0	0	0	0	0	0	0	0	0
Address	0	0	1	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	0.987771	0	0	0	0	0	0	0
County	0	0	0	0	0.9909821	0	0	0	0	0	0

Country	0	0	0	0	0	0	0	0	0	0	0
TtlEmployee	0	0	0	0.8340656	0	0	0	0	0	0	0
Url	0	1	0	0	0	0	0	0.7703468	0	0.9620293	0.9213293
EmployeeCount	0	0	0	0	0	0	0.9523789	0	0	0	0
Uni	0.976205529	0	0	0	0	0	0	0	0	0	0
Location	0	0	0.8533409	0	0	0	0	0	0	0	0

Ground Truth table 1

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudent
Name	1	0	0	0	0	0	0	0	0	0	0
Address	0	0	1	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	1	0	0	0	0	0	0	0
County	0	0	0	0	1	0	0	0	0	0	0
Country	0	0	0	0	0	0	0	0	0	0	0
TtlEmployee	0	0	0	0	0	0	0	0	0	0	0
Url	0	1	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	1	0	0	0	0	0	0	0	0	0	0
Location	0	0	1	0	0	0	0	0	0	0	0

TP=7

FP=6

FN=1

Precision=7/13=0.53

Recall=7/8=0.87

F1-Score=0.65

DS1- Min Matrix

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudents
Name	0.634555456	0	0	0	0	0	0	0	0	0	0
Address	0	0	0	0	0	0	0	0	0	0	0.5247523
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	0.978269	0	0	0	0	0	0	0

County	0	0	0	0	0	0	0	0	0	0	0
Country	0	0	0.9288207	0	0.6309266	0	0	0	0	0	0
TtlEmployee	0		0	0	0	0	0	0	0	0	0
Url	0	1	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	0.578056784	0.4836775	0	0	0	0	0	0	0	0	0
Location	0	0	0.9731672	0	0	0	0	0	0	0.8707311	0

Ground Truth table 1

MS/DS1	UniversityName	Url	Location	ZIP	CountyName	ContactName	EmployeeCount	AverageGPA	GradStudents	StudWrkInUni	NumDisStudents
Name	1	0	0	0	0	0	0	0	0	0	0
Address	0	0	1	0	0	0	0	0	0	0	0
City	0	0	0	0	0	0	0	0	0	0	0
ZIPCode	0	0	0	1	0	0	0	0	0	0	0
County	0	0	0	0	1	0	0	0	0	0	0
Country	0	0	0	0	0	0	0	0	0	0	0
TtlEmployee	0	0	0	0	0	0	0	0	0	0	0
Url	0	1	0	0	0	0	0	0	0	0	0
EmployeeCount	0	0	0	0	0	0	0	0	0	0	0
Uni	1	0	0	0	0	0	0	0	0	0	0
Location	0	0	1	0	0	0	0	0	0	0	0

TP=6

FP=4

FN=2

Precision=6/10=0.60

Recall=6/8=0.75

F1-Score=0.66

With comparing ground truth table with the result matrix we calculated the performance.

	Min-Prec	Max-Prec	Avg- Prec	Min-Recall	Max-Recall	Avg-Recall	Min-F1	Mx-F1	Avg-F1
DS1-MS	0.63	0.53	0.73	0.65	0.87	0.83	0.63984375	0.658714	0.776795
DS2-MS	0.68	0.61	0.79	0.59	0.73	0.79	0.631811024	0.664627	0.79
DS3-MS	0.72	0.7	0.75	0.64	0.78	0.82	0.677647059	0.737838	0.783439

It is obvious that Average Strategy has the best performance for our MS.