DOM 207: IDSBRP
MINI PROJECT 2

INTRODUCTION

We, Aditya Arora (2010110038) and Sanjana Nair (2010110764) under Dr. Jaideep Ghosh as part of the DS Project Team for the Consulting firm CX, are presenting our findings in our project contracted out by The World Bank.The data provided to us contains 9 different individual issues of workers in the ICT sector scored via two different measures of 37 countries holding varied positions in the economic spectrum. However, through powerful clustering algorithms this seemingly complicated task with diverse parameters becomes much more tangible and allows us to come up with interesting criterias to group these countries to make our data far more comprehensible.

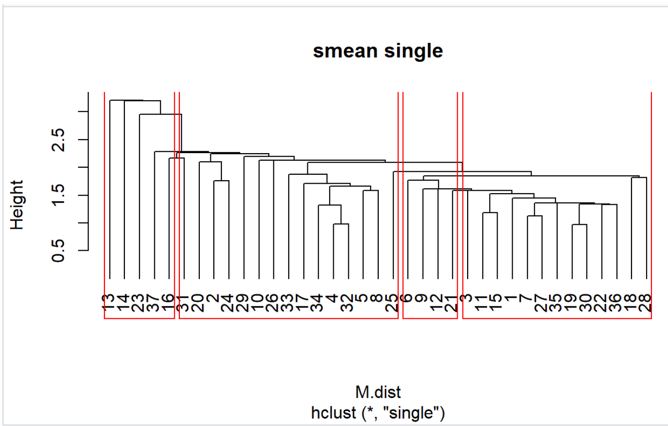
METHODOLOGY

We separate the excel data file provided into 2 files FSMEAN ICT INDIV and SMEAN ICT INDIV according to the different scores namely fs_mean and s_mean.
Also, we fix the distance metric to euclidean. It is the most generally applicable metric and puts higher values together and lower values together which works well with the ward.D method(which minimizes intracluster variance) which we will see is used by us later.

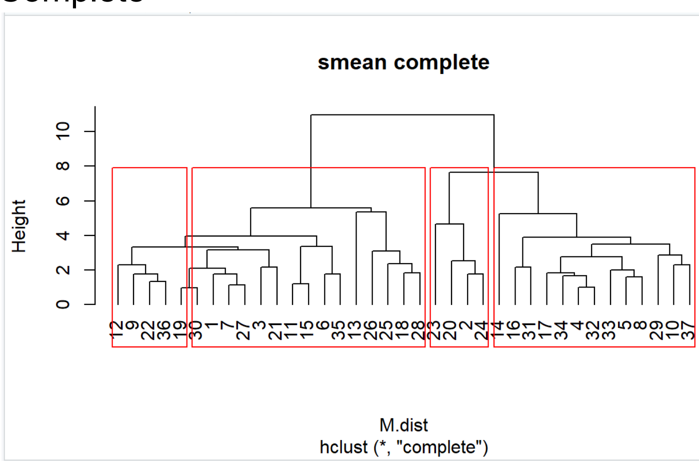
Now it's time for us to observe various dendrograms and decide which method of choosing distances between clusters works best for us.
We form dendrograms using R through all 5 methods to see which method for each score yields us the most well separated and identifiable clusters. We also keep in mind the fact that we are able to keep our clusters between 3 to 5.

SMEAN

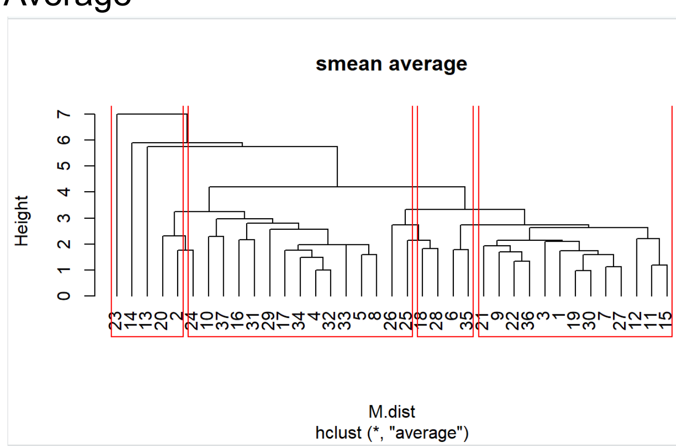
Single



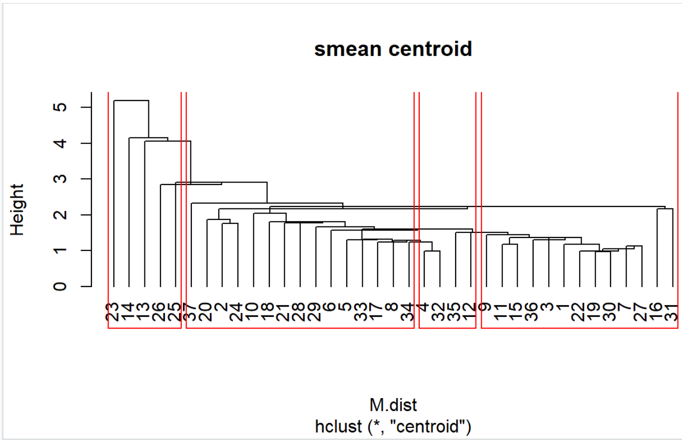
Complete



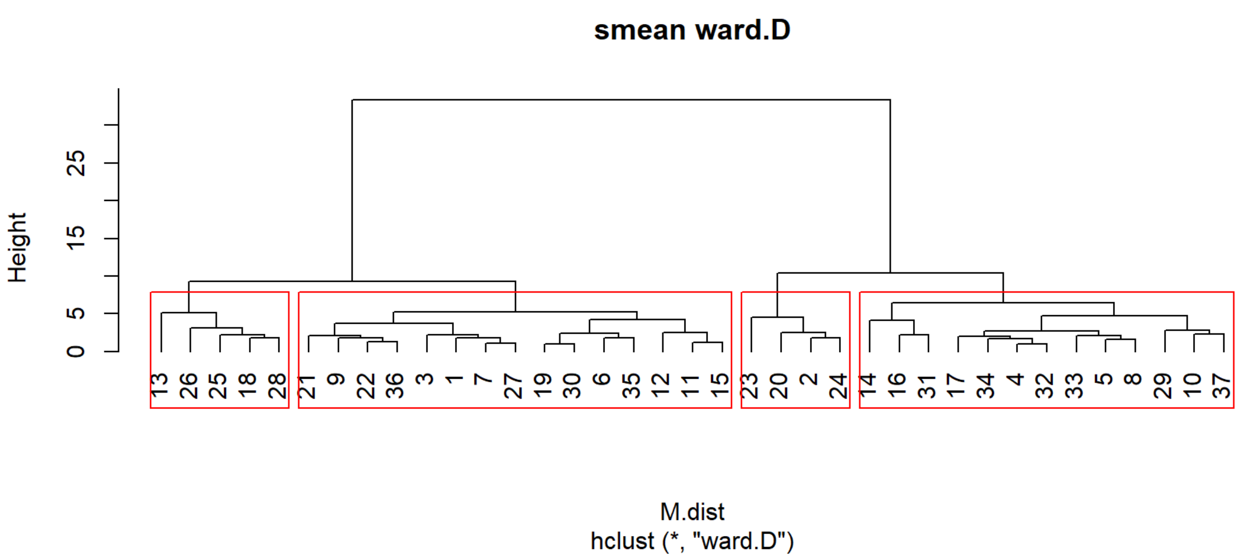
Average



Centroid

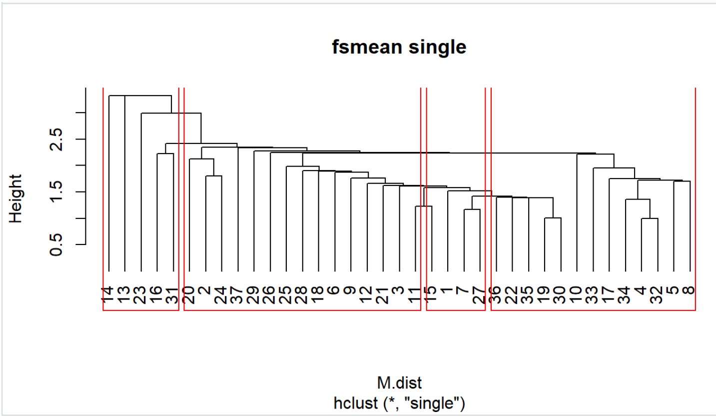


Ward.D

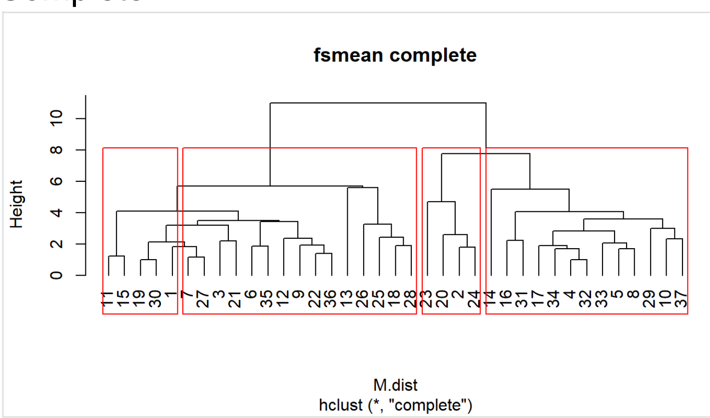


FSMEAN

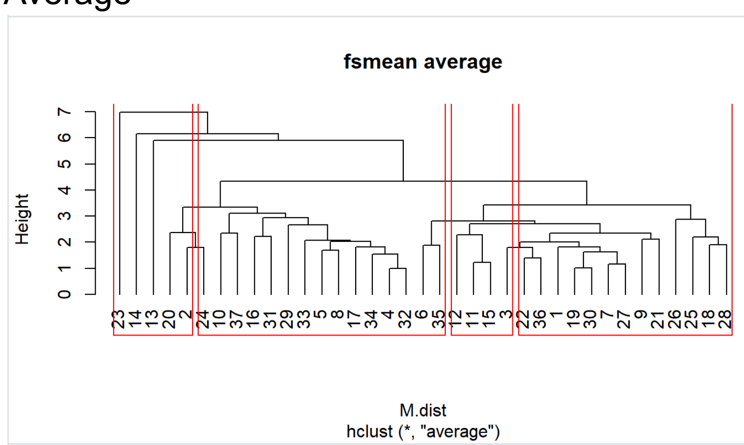
Single



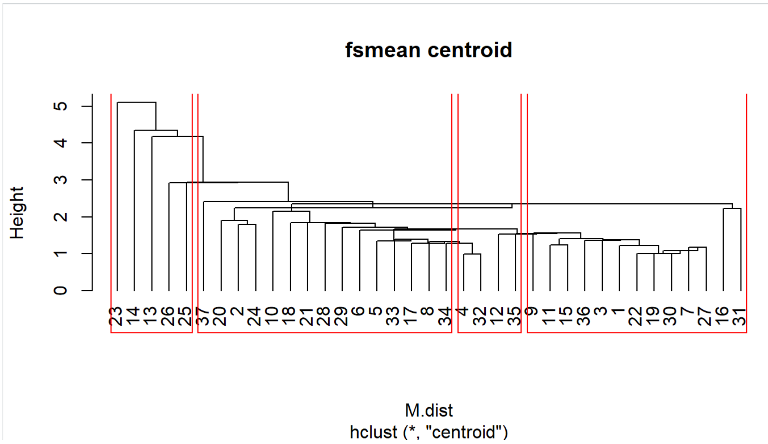
Complete



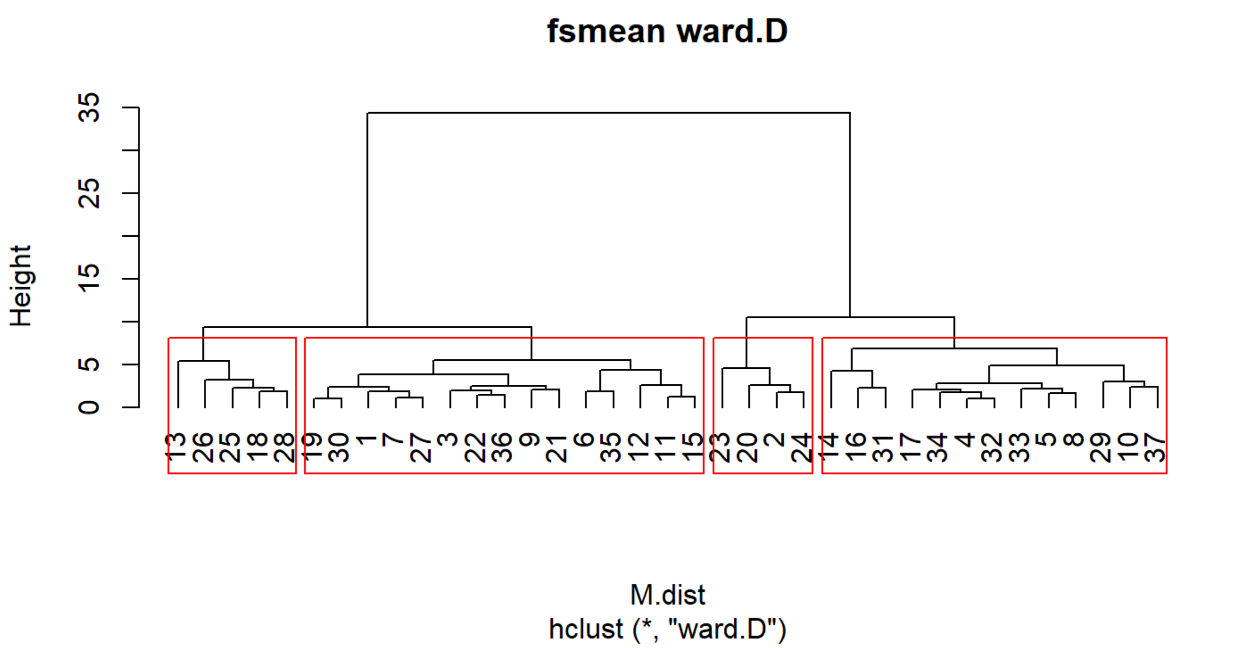
Average



Centroid



WARD.D



Post observing the dendrograms for SMean and FSMean with different methods i.e, Single, Complete, Average, Centroid and Ward.D, knowing that the optimum number of clusters must be 3-5, we deduce that Ward.D is the most efficient way of plotting our dendrogram because it provides the most clear, well separated and well defined clusters followed by Complete, however we deice to stick with ward.D method.

We want to give preference that we keep our clusters within the range of 3 to 5 over the values at which they divide into clusters we put our trust in the hierarchical clustering algorithm to stop when it finds 4 clusters since 4 is the average of our range 3 to 5 and not use the height criteria as one is more gullible to find higher number of clusters through height when clusters are mixed together.
We also observe that we have obtained the same clusters in the ward.D method for both s_mean and fs_mean.

After using R to form dendrograms we manipulate the excel files FSMEAN ICT INDIV and SMEAN ICT INDIV. We have put each cluster entries together and color coded the clusters so it becomes easier to interpret them.
Also,the ward.D clustering method gave us the most clear clusters and it focuses on minimizing the within cluster variance of the scores of all the issues . We choose to bank on this advantage and use the average of each of the 9 issue values for a given type of score and then use the average values for a cluster as a representative of all the entries in it, mentioned at the end of each cluster in the excel files

We also observe that the relation between the average values for each issue between the clusters remains the same , but the average values for both types of scores have changed.However, if for each fixed issue if the relation between the average values has been preserved our conclusion for both the scores will come out the same.

The different clusters for WARD.D (SMEAN)

CLUSTER 1

13 - India
26 - Poland
25 - Peru upper
18 - Lithuania
28 - Romania

1	Country	FC_Smea	JI_Smea	SE_Smea	WS_Smea	WO_Smea	WH_Smea	WE_Sme	TI_Smea	TA_Sme
14	India	5.034286	4.588572	6.037143	15.76571	13.33429	11.22571	14.85143	11.73429	12.88
19	Lithuania	6.047945	5.623288	8.59589	14.25342	13.30822	11.19863	14.00685	9.993151	11.97945
26	Peru	5.786163	5.672956	7.012578	14.43396	13.30818	11.66038	15.3522	10.25157	11.42138
27	Poland	6.836667	5.626667	9.236667	13.29667	14.1	11.16	14.80667	10.44	11.16
29	Romania	5.615854	5.911585	7.689024	14.64634	12.91159	10.78049	14.13415	10.28049	10.84146

CLUSTER 2

21 - Mexico(Upper middle income)
9 - Germany(High income)
22 - New Zealand
36 - USA
3 - Brazil
1 - Argentina
7 - Finland

27 - Portugal
19 - Macedonia
30 - South Africa
6 - Egypt
35 - UK
12 - Hungary
11 - Greece
15 - Italy

1	Country	FC_Smea	JI_Smea	SE_Smea	WS_Smea	WO_Smea	WH_Smea	WE_Sme	TI_Smea	TA_Sme
2	Argentina	5.50809	5.838188	7.89644	13.96117	11.589	10.71845	13.83819	9.2589	12.96117
4	Brazil	5.316092	5.896552	7.215517	12.69253	11.93678	10.8477	13.67529	10.33908	12.3908
7	Egypt	5.045714	6.171429	6.84	13.72	12.30857	9.611428	11.49714	9.594286	11.72
8	Finland	5.895833	6.194445	7.680555	13.84722	11.33333	10.40972	13.72222	10.02083	12.07639
10	Germany	6.425324	6.032467	7.438312	14.37662	13.09091	10.40909	12.8539	9.685065	12.75
12	Greece	5.584906	6.207547	8.358491	14.5	11.31132	10	12.68868	10.95283	11.56604
13	Hungary	6.124542	5.893773	7.655678	14.68498	13.01832	9.923077	13	11.54212	12.37363
16	Italy	5.551613	6.241935	7.987097	14.26129	11.52581	9.890323	13.00968	11.90323	11.5129
20	Macedonia	5.486394	6.238095	7.778912	14.15986	12.18367	10.31293	13.13946	9.530612	11.79932
22	Mexico	5.678679	5.546546	7.003003	13.31231	13.37538	11.02102	13.36336	9.297297	11.97598
23	New Zealand	5.835271	5.804264	7.387597	13.89535	12.37984	11.06589	13.31202	10.10853	12.12791
28	Portugal	5.714286	5.665179	7.808036	14.375	11.14732	10.57589	13.58482	10.62946	12.36161
31	South Africa	5.664452	6.169435	7.475083	14.34884	11.86379	9.986711	12.7309	10.05316	11.98671
36	UK	6.020833	5.802083	7.28125	14.4375	11.4375	9.583333	11.72917	10	11.94792
37	USA	5.828479	5.430421	7.106796	13.8835	12.67314	10.62136	12.94175	10.91909	12.67961

CLUSTER 3

23 - Nigeria
20 - Malaysia
2 - Bangladesh
24 - Pakistan

1	Country	FC_Smea	JI_Smea	SE_Smea	WS_Smea	WO_Smea	WH_Smea	WE_Sme	TI_Smea	TA_Sme
3	Bangladesh	5.267606	6.285211	7.841549	12.10563	10.15141	8.718309	11.95423	8.950705	10.11268
21	Malaysia	5.194346	6.749116	8.95053	11.39929	10.90106	9.477032	11.80919	9.639576	9.508834
24	Nigeria	5.032258	6.096774	7.268817	11.10753	9.172043	7.129032	9.483871	8.247312	10.11828
25	Pakistan	5.302326	7.016611	7.568106	11.62791	10.63787	7.86711	11.07309	9.415282	10.12292

CLUSTER 4

14 - Iran
16 - Japan
31 - South Korea
17 - Jordan
34 - Turkey
4 - Canada
32 - Taiwan

33 - Thailand
5 - China
8 - France
29 - Russia
10 - Ghana
37 - Vietnam

1	Country	FC_Smea	JI_Smea	SE_Smea	WS_Smea	WO_Smea	WH_Smea	WE_Sme	TI_Smea	TA_Sme
5	Canada	6.064309	6.694534	8.401929	12.88424	11.9164	10.02572	12.34084	10.14791	11.07074
6	China	5.861953	7.154882	9.178452	12.07071	11.57912	9.084175	12.90236	9.582492	10.77778
9	France	6.535836	6.955631	8.139932	12.56314	11.12287	9.177474	12.58362	9.372014	11.05461
11	Ghana	5.546052	6.855263	8.253289	13.84868	12.21382	9.713816	12.67763	9.6875	9.223684
15	Iran	6.960784	8.140056	11.72269	13.78431	11.28011	9.887955	11.59944	8.803922	9.238095
17	Japan	6.980645	7.909678	9.877419	12.72581	12.01935	9.864516	12.2	10.28387	10.47742
18	Jordan	5.790514	6.237154	7.521739	13.17391	11.51383	9.774704	11.75494	9.470356	10.34783
30	Russia	6.141892	5.986486	9.432432	14.18243	11.58784	10.06081	12.75	9.297297	10.75
32	South Korea	6.770764	6.707641	9.860465	11.75747	11.299	9.318937	11.33223	10.67774	10.20266
33	Taiwan	5.877888	6.79538	8.046205	12.32673	11.72937	10.0429	11.76568	10.12211	11.22112
34	Thailand	5.629337	6.922713	9.089906	12.89117	10.88959	10.01104	11.76814	9.138802	10.84385
35	Turkey	6.087108	6.843205	8.550523	13.0453	12.01742	8.961673	11.96167	10.13937	10.60627
38	Vietnam	5.503356	6.895973	9.362416	12.63423	12.18792	10.34228	13.82886	9.191276	9.325503

The different clusters for WARD.D (FSMEAN)

CLUSTER 1

13 - India(developing)
26 - Poland
25 - Peru
18 - Lithuania
28 - Romania

1	Country	FC_FSmea	JI_FSmea	SE_FSme	WS_FSmea	WO_FSmea	WH_FSme	WE_FSmea	TI_FSmea	TA_FSme
14	India	-0.4863254	-0.7131865	-0.7736582	0.6034991	0.411868	0.3765458	0.5379285	0.6640864	0.6192487
19	Lithuania	0.1089554	-0.3008906	0.1437158	0.3114155	0.3906154	0.3553663	0.2887011	0.0751261	0.2936025
26	Peru	-0.0395868	-0.2410416	-0.4324602	0.2904271	0.3736726	0.503345	0.6331366	0.109727	0.1383834
27	Poland	0.5883378	-0.3120074	0.3537602	-0.1732683	0.6040968	0.3378305	0.501323	0.2198322	0.0064812
29	Romania	-0.1147762	-0.2041972	-0.1994678	0.1869205	0.2608706	0.2151031	0.3231023	0.0986648	-0.1901144

CLUSTER 2

19 - Macedonia
30 - South Africa
1 - Argentina
7 - Finland
27 - Portugal
3 - Brazil
22 - New Zealand

36 - USA
9 - Germany
21 - Mexico
6 - Egypt
35 - UK
12 - Hungary
11 - Greece
15 - Italy

1	Country	FC_FSmea	JI_FSmea	SE_FSme	WS_FSmea	WO_FSmea	WH_FSme	WE_FSmea	TI_FSmea	TA_FSme
2	Argentina	-0.2094778	-0.2077165	-0.1193405	0.3686504	-0.112377	0.1997711	0.2644838	-0.313584	0.637878
4	Brazil	-0.3028045	-0.1937347	-0.3518725	-0.0315953	-0.015879	0.23489	0.2106494	0.1220362	0.4095934
7	Egypt	-0.4966382	-0.0751387	-0.4709443	0.1741985	0.0877883	-0.1647951	-0.3272959	-0.1086336	0.2506225
8	Finland	0.038701	-0.0154287	-0.1303959	0.2729759	-0.1363862	0.1523927	0.292304	0.0260346	0.3194362
10	Germany	0.3213172	-0.0461395	-0.1990594	0.6225027	0.4274112	0.1882129	0.0886402	-0.0441418	0.6680799
12	Greece	-0.1564722	-0.0422948	0.0468805	0.2936675	-0.1952175	-0.0413032	-0.0353891	0.346714	0.0824126
13	Hungary	0.1390544	-0.1966832	-0.1967964	0.4117202	0.307143	-0.0714601	0.041068	0.5522429	0.3703316
16	Italy	-0.1829289	-0.062558	-0.0826637	0.2139314	-0.1382458	-0.0737899	0.0368868	0.7090418	0.0582464
20	Macedonia	-0.2331361	-0.0503789	-0.1587269	0.2665427	0.0554722	0.0619873	0.0769938	-0.1967992	0.2095296
22	Mexico	-0.1014778	-0.3204842	-0.4037302	0.1236416	0.4192901	0.3310312	0.1698307	-0.2310697	0.5009748
23	New Zealand	-0.0119369	-0.223931	-0.2958756	0.1981293	0.1056621	0.3069851	0.1129149	0.006052	0.2845071
28	Portugal	-0.0732581	-0.2565626	-0.1475619	0.4141135	-0.2272283	0.1658715	0.1971452	0.2181586	0.3975018
31	South Africa	-0.0934488	-0.0485879	-0.2513277	0.3450571	-0.0073043	-0.0183602	0.0059889	0.0295029	0.283302
36	UK	0.0898309	-0.2014817	-0.3025106	0.3913471	-0.1269937	-0.1456176	-0.256117	0.0041332	0.2625467
37	USA	-0.0202409	-0.3699629	-0.3879041	0.2213126	0.2001528	0.1718984	0.0361204	0.3502323	0.5305105

CLUSTER 3

23 - Nigeria
20 - Malaysia
2 - Bangladesh
24 - Pakistan

1	Country	FC_FSmea	JI_FSmea	SE_FSme	WS_FSmea	WO_FSmea	WH_FSme	WE_FSmea	TI_FSmea	TA_FSme
3	Bangladesh	-0.3224826	0.0604205	-0.1058339	-0.4245317	-0.450292	-0.4042176	-0.1374162	-0.1499239	-0.422314
21	Malaysia	-0.3748917	0.1704879	0.2618449	-0.6469991	-0.305995	-0.2151088	-0.2600119	-0.1310841	-0.6899559
24	Nigeria	-0.4098476	0.122328	0.0217955	-0.3152314	-0.4042	-0.6084588	-0.5057551	-0.2425787	0.2672582
25	Pakistan	-0.333588	0.2060018	-0.2349786	-0.5624713	-0.3756672	-0.7347329	-0.4414383	-0.187122	-0.3884149

CLUSTER 4

14 - Iran 16 - Japan 31 - South Korea 17 - Jordan 34 - Turkey 4 - Canada 32 - Taiwan	33 - Thailand 5 - China 8 - France 29 - Russia 10 - Ghana 37 - Vietnam

1	Country	FC_FSmea	JI_FSmea	SE_FSme	WS_FSmea	WO_FSmea	WH_FSme	WE_FSmea	TI_FSmea	TA_FSme
5	Canada	0.1182705	0.1218635	0.0565461	-0.213239	-0.0229455	-0.0358087	-0.1293741	0.0742533	-0.0471908
6	China	0.0209906	0.3006831	0.3465708	-0.529495	-0.1259053	-0.3459992	0.0152623	-0.1719987	-0.0960616
9	France	0.4312542	0.2261264	-0.0470081	-0.1818525	-0.247512	-0.31424	-0.0719518	-0.1854289	-0.1018106
1	Ghana	-0.1647507	0.2494081	0.0120916	0.1764504	0.0586393	-0.0192822	-0.0407555	-0.1287587	-0.8453493
5	Iran	0.6813552	0.717529	1.255868	-0.1665573	-0.1991514	-0.081916	-0.3145958	-0.4702049	-0.8141794
7	Japan	0.6974447	0.6308764	0.6007019	-0.2718218	0.0053292	-0.0913381	-0.161772	0.1313387	-0.2238972
8	Jordan	-0.0413375	-0.0190614	-0.2498585	-0.1207685	-0.1413911	-0.1175031	-0.2703012	-0.1998842	-0.367796
0	Russia	0.1807752	-0.1290011	0.4562361	0.2638907	-0.076654	-0.0030887	-0.0008395	-0.1670627	-0.044792
2	South Korea	0.6142066	0.1326254	0.6058709	-0.537681	-0.1999574	-0.2467723	-0.3761268	0.2479572	-0.3759705
3	Taiwan	0.0347258	0.1576029	-0.0552349	-0.3438185	-0.0831942	-0.029349	-0.2711025	0.0240229	-0.0399738
4	Thailand	-0.1223227	0.2164325	0.3296522	-0.256585	-0.3118231	-0.0383886	-0.2696362	-0.3767795	-0.1107657
5	Turkey	0.1397102	0.161402	0.1238105	-0.1311545	0.0102906	-0.385786	-0.2237206	0.0422102	-0.2508285
8	Vietnam	-0.1973073	0.2287616	0.4076424	-0.2647885	0.0515008	0.0723537	0.2501684	-0.3184828	-0.7851575

Results and Discussion

Using the Excel sheets and taking out the averages we make the following observations .(Here green is for a positive impact and red for a negative impact.)

FC: DEPENDENCE /USE OF FRIENDSHIP CIRCLES
CLUSTER3 < CLUSTER2 < CLUSTER1 < CLUSTER4

JI:JOB INSECURITY
CLUSTER1 < CLUSTER2 < CLUSTER3 < CLUSTER4

SE:SELF-EFFICACY
CLUSTER2 < CLUSTER1 < CLUSTER3 < CLUSTER4

WS:WORK SATISFACTION
CLUSTER3<CLUSTER4<CLUSTER2<CLUSTER1

WO:WORK OVERLOAD
CLUSTER3<CLUSTER4<CLUSTER2<CLUSTER1

WH:WORK-HOME CONFLICT
CLUSTER3<CLUSTER4<CLUSTER2<CLUSTER1

WE:WORK EXHAUSTION OR STRAIN
CLUSTER3<CLUSTER4<CLUSTER2<CLUSTER1

TI:TURNOVER INTENTION
CLUSTER3<CLUSTER4<CLUSTER2<CLUSTER1

TA:TURNAWAY INTENTION
CLUSTER3<CLUSTER4<CLUSTER1<CLUSTER2

Analyzing CLUSTER 1

With the above inference we see that even though cluster 1 is the highest in work overload, work exhaustion and work home conflict, the work satisfaction is highest. We also see that cluster 1 does not have the highest turnaway intention value which is indicative also to the fact that the job insecurity is least for cluster 1. So the people in the ICT sector working in countries belonging to cluster 1 have in general secure jobs and are productive but do not do well in matters concerning work life balance.This is showcased even with the 5 countries in cluster 1 i.e. India, Lithuania, Poland, Romania and Peru with the following information.

Peru is recording one of the highest growth rates in e-commerce in Latin America and the Caribbean(ref:https://unctad.org)

Romania is one of the fastest-growing information technology (IT) markets in Central and Eastern Europe(ref:https://www.romanianstartups.com/about-romania)

Lithuania has the largest ICT industry in the Baltic States with outstanding potential both for local as well as foreign expanding businesses(ref:https://b2lithuania.com)

Poland has one of the most progressive iT sectors among 23 countries of Central and Easter Europe(ref:https://scand.com)

India's ICT sector has grown meteorically in the last two decades, making it an important lever of employment for the next generation of workers(ref:https://www.ilo.org)

Analyzing cluster 2

The nature of work in Cluster 2 countries is very similar to the cluster 1 countries but more balanced out as for each issue the value lies in the middle, closer to cluster 1 when compared with other clusters.

Analyzing cluster 3

We observe the polar opposite of attributes in cluster 3 when compared to cluster 1.Cluster 3 has the lowest workload, work home conflict and work exhaustion but it also has least work satisfaction. It also has the highest job insecurity. It has the least turnover and turnaway intention values. Hence, the people in the ICT sector working in countries belonging to cluster 3 are in general happy with their work life balance and jobs overall and the low turn away and turn over values in conjunction with the fact that they have less work load, work home conflict and work exhaustion support this. However, possibly due to this they do not end up performing very well and get stuck in insecure jobs with lesser job satisfaction.This is showcased even with the 4 countries in cluster 1 i.e. Bangladesh,Malaysia,Pakistan and Nigeria .

Analyzing cluster 4

The relation between cluster 1 and cluster 2 is to a large extent imitated by the relationship between cluster 3 and cluster 4 with their roles being in the same order. The nature of work in Cluster 4 countries is very similar to cluster 3 countries but more balanced out as for four out of six issues the value lies in the middle, but closer to cluster 3 when compared with other clusters.

CONCLUSION

In conclusion, after cluster analysis we are able to find four emergent clusters in which we can divide the provided countries. We find out that two of these clusters are polar opposites of each other(Cluster 1 and Cluster 3) while the other two clusters behave in moderate fashion while imitating the other two extreme clusters(Cluster 2 and Cluster 4). The main thing that differentiates the extreme clusters are factors surrounding productivity in Cluster 1 and job contentment in Cluster 3. Our analysis proves how beneficial cluster analysis is in efficiently dealing with data. We are able to arrive at a simplistic and easy to interpret grouping, which is supported by real life as shown for the case of cluster 1 from a very tangled data .