Sneha Potlapally
U76253166

Profiling Internet Users

Objective:

The main objective of this project is to distinguish the internet usage of one user is statistically distinguishable when compared to the internet usage of the other users. We compared the results for each time window in order to check how time window affects profiling.

About project:

We consider the dockets and duration of each user from the given files and then we will give the path of these files in the code to calculate the spearman coefficient and P values of these 3 time windows for week 1 and 2 where one user is compared with every other user for all the 54 users is attached in the zip file as excel sheets and find the value of P(z) and the output must be written in the excel format.

Here, We have to caluculate the P(z) values in the interval of 10sec, 227sec and 300sec.We are given the data and the data is splitted in the given intervals. The date we need to take are in the range of date between 4^{th} to 15^{th} and we need to exclude the weekends i.e Saturdays and Sundays. The data must be executed in the time frame between 8am to 5pm in the given intervals by calculating with the given function .

Implementation and Functions used:

Here, Three different files are created for three time windows. The data is readed from the given excel file and this data is splitted according to the time intervals and the functions are calculated to this data.

similar way.

Here we calculate the average doctets/durations in a particular day. For calculating Spearman's correlation coefficient, the difference between the respective ranks (according to the r value) is calculated and applied in the below formula to get three correlation values r1a2a, r1a2b,r2a2b where numbers show weeks and characters are showing subjects.

Sneha Potlapally
U76253166

I have used spearmanr(variable1, variable2) to calculate the spearman coefficient of the two users.

r_{1a2a} =internet usage of subject a for week1 to subject a with week2

r_{1a2b}=internet usage of subject a for week1 to subject b with week 2

r_{2a2b}=internet usage of subject a for week2 to subject b with week2

These values r1a2a, r1a2b,r2a2b are used in below formulas to get Z value.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

d_i: The difference between the ranks of corresponding variables

n: Number of observations (Number of windows in a week)

$$Z = [Z_{1a2a} - Z_{1a2b}] * \frac{\sqrt{[N-3]}}{2 * [1 - r_{2a2b}] * h}$$

$$Z_{1a2a} = \frac{1}{2} log \frac{1 + r_{1a2a}}{1 - r_{1a2a}}$$

$$Z_{1a2b} = \frac{1}{2} log \frac{1 + r_{1a2b}}{1 - r_{1a2b}}$$

$$h = \frac{1 - [f * rm^2]}{1 - rm^2}$$

$$f = \frac{1 - r_{2a2b}}{2 * [1 - rm^2]}$$

$$rm^2 = \frac{r_{1a2a}^2 + r_{1a2b}^2}{2}$$

N: Sample size of the data set (Number of windows in a week)

P values for these Z values are calculated:

$$P = 1 - \Phi(Z)$$

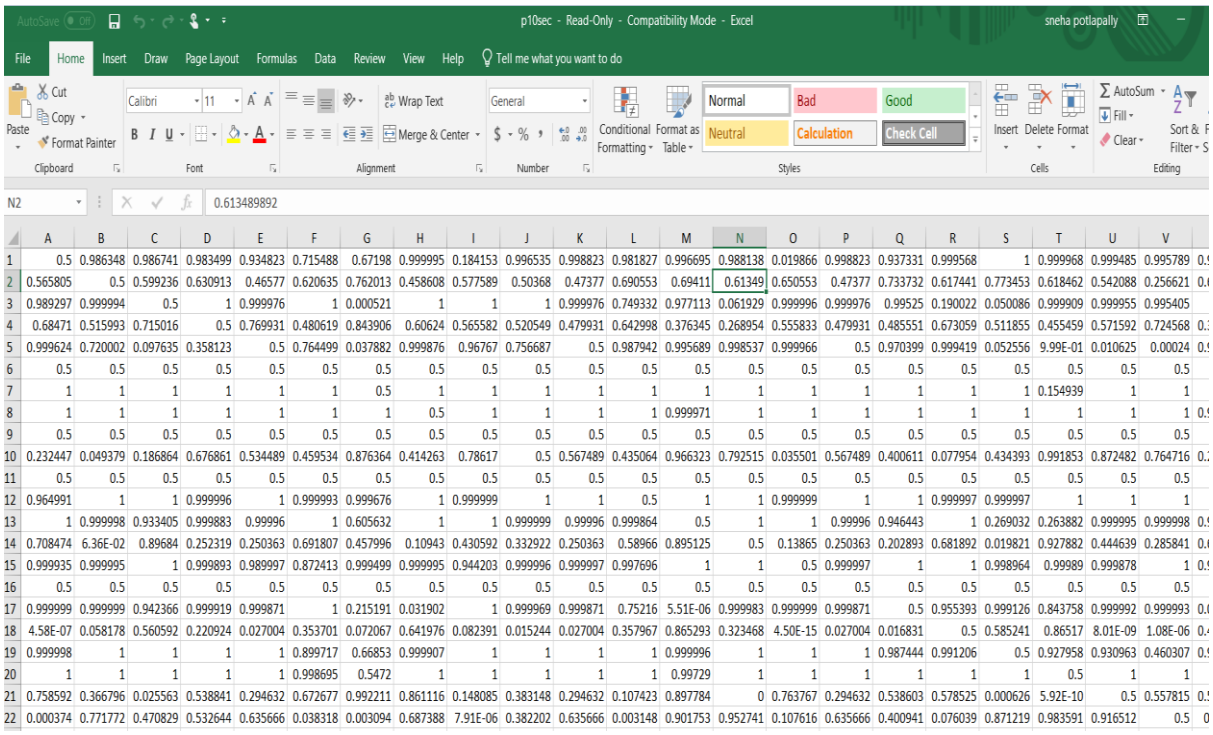
This will be the correlation coefficient.

Sneha Potlapally
U76253166

RESULTS:

Output1:

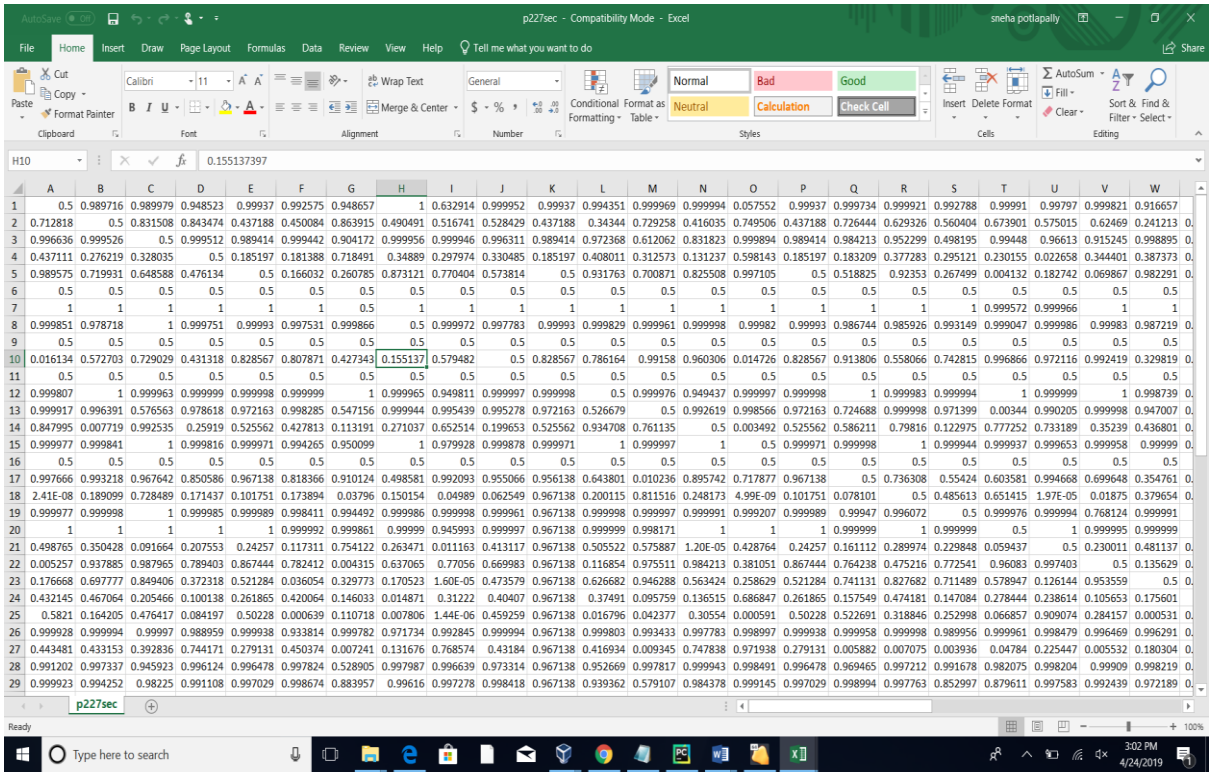
Time interval =10 sec



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	0.5	0.986348	0.986741	0.983499	0.934823	0.715488	0.67198	0.999995	0.184153	0.996535	0.998823	0.981827	0.996695	0.988138	0.019866	0.998823	0.937331	0.999568	1	0.999968	0.999485	0.995789
2	0.565805	0.5	0.599236	0.630913	0.46577	0.620635	0.762013	0.458608	0.577589	0.50368	0.47377	0.690553	0.69411	0.61349	0.650553	0.47377	0.733732	0.617441	0.773453	0.618462	0.542088	0.256621
3	0.989297	0.999994	0.5	1	0.999976	1	0.000521	1	1	1	0.999976	0.749332	0.977113	0.061929	0.999996	0.999976	0.99525	0.190022	0.050086	0.999909	0.999955	0.995405
4	0.68471	0.515993	0.715016	0.5	0.769931	0.480619	0.843906	0.60624	0.565582	0.520549	0.479931	0.642998	0.376345	0.268954	0.555833	0.479931	0.485551	0.673059	0.511855	0.455459	0.571592	0.724568
5	0.999624	0.720002	0.097635	0.358123	0.5	0.764499	0.037882	0.999876	0.96767	0.756687	0.5	0.987942	0.995689	0.998537	0.999966	0.5	0.970399	0.999419	0.052556	9.99E-01	0.010625	0.00024
6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
7	1	1	1	1	1	1	0.5	1	1	1	1	1	1	1	1	1	1	1	1	0.154939	1	1
8	1	1	1	1	1	1	1	0.5	1	1	1	1	0.999971	1	1	1	1	1	1	1	1	1
9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
10	0.232447	0.049379	0.186864	0.676861	0.534489	0.459534	0.876364	0.414263	0.78617	0.5	0.567489	0.435064	0.966323	0.792515	0.035501	0.567489	0.400611	0.077954	0.434393	0.991853	0.872482	0.764716
11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
12	0.964991	1	1	0.999996	1	0.999993	0.999676	1	0.999999	1	1	0.5	1	1	0.999999	1	1	0.999997	0.999997	1	1	1
13	1	0.999998	0.933405	0.999883	0.99996	1	0.605632	1	1	0.999999	0.99996	0.999864	0.5	1	1	0.99996	0.946443	1	0.269032	0.263882	0.999995	0.999998
14	0.708474	6.36E-02	0.89684	0.252319	0.250363	0.691807	0.457996	0.10943	0.430592	0.332922	0.250363	0.58966	0.895125	0.5	0.13865	0.250363	0.202893	0.681892	0.019821	0.927882	0.444639	0.285841
15	0.999935	0.999995	1	0.999893	0.989997	0.872413	0.999499	0.999995	0.944203	0.999996	0.999997	0.997696	1	1	0.5	0.999997	1	1	0.998964	0.99989	0.999878	1
16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
17	0.999999	0.999999	0.942366	0.999919	0.999871	1	0.215191	0.031902	1	0.999969	0.999871	0.75216	5.51E-06	0.999983	0.999999	0.999871	0.5	0.955393	0.999126	0.843758	0.999992	0.999993
18	4.58E-07	0.058178	0.560592	0.220924	0.027004	0.353701	0.072067	0.641976	0.082391	0.015244	0.027004	0.357967	0.865293	0.323468	4.50E-15	0.027004	0.016831	0.5	0.585241	0.86517	8.01E-09	1.08E-06
19	0.999998	1	1	1	1	0.899717	0.66853	0.999907	1	1	1	1	0.999996	1	1	1	0.987444	0.991206	0.5	0.927958	0.930963	0.460307
20	1	1	1	1	1	0.998695	0.5472	1	1	1	1	1	0.99729	1	1	1	1	1	1	0.5	1	1
21	0.758592	0.366796	0.025563	0.538841	0.294632	0.672677	0.992211	0.861116	0.148085	0.383148	0.294632	0.107423	0.897784	0	0.763767	0.294632	0.538603	0.578525	0.000626	5.92E-10	0.5	0.557815
22	0.000374	0.771772	0.470829	0.532644	0.635666	0.038318	0.003094	0.687388	7.91E-06	0.382202	0.635666	0.003148	0.901753	0.952741	0.107616	0.635666	0.400941	0.076039	0.871219	0.983591	0.916512	0.5

Output 2:

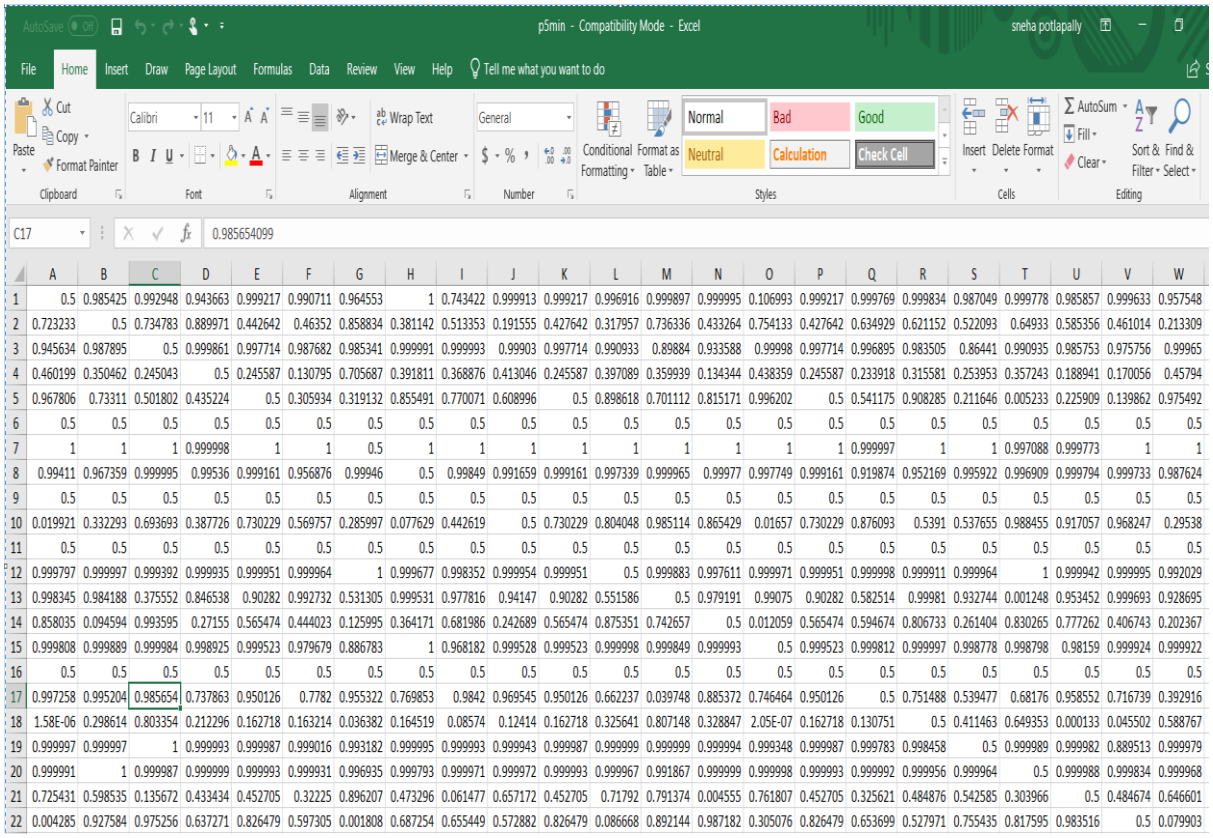
Time interval= 227sec



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	0.5	0.989716	0.989979	0.948523	0.99937	0.992575	0.948657	1	0.632914	0.999952	0.99937	0.994351	0.999969	0.999994	0.057552	0.99937	0.999734	0.999921	0.992788	0.99991	0.99797	0.999821	0.916657
2	0.712818	0.5	0.831508	0.843474	0.437188	0.450084	0.863915	0.490491	0.516741	0.528429	0.437188	0.34344	0.729258	0.416035	0.749506	0.437188	0.726444	0.629326	0.560404	0.673901	0.575015	0.62469	0.241213
3	0.996636	0.999526	0.5	0.999512	0.989414	0.999442	0.904172	0.999956	0.999946	0.996311	0.989414	0.972368	0.612062	0.831823	0.999894	0.989414	0.984213	0.952299	0.498195	0.99448	0.96613	0.915245	0.998895
4	0.437111	0.276219	0.328035	0.5	0.185197	0.181388	0.718491	0.34889	0.297974	0.330485	0.185197	0.408011	0.312573	0.131237	0.598143	0.185197	0.183209	0.377283	0.295121	0.230155	0.022658	0.344401	0.387373
5	0.989575	0.719931	0.648588	0.476134	0.5	0.166032	0.260785	0.873121	0.770404	0.573814	0.5	0.931763	0.700871	0.825508	0.997105	0.5	0.518825	0.92353	0.267499	0.004132	0.182742	0.069867	0.982291
6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
7	1	1	1	1	1	1	1	0.5	1	1	1	1	1	1	1	1	1	1	1	0.999572	0.999966	1	1
8	0.999851	0.978718	1	0.999751	0.99993	0.997531	0.999866	0.5	0.999972	0.997783	0.99993	0.999829	0.999961	0.999998	0.99982	0.99993	0.986744	0.985926	0.993149	0.999047	0.999986	0.99983	0.987219
9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
10	0.016134	0.572703	0.729029	0.431318	0.828567	0.807871	0.427343	0.155137	0.579482	0.5	0.828567	0.786164	0.99158	0.960306	0.014726	0.828567	0.913806	0.558066	0.742815	0.996866	0.972116	0.992419	0.329819
11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
12	0.999807	1	0.999963	0.999999	0.999998	0.999999	1	0.999965	0.949811	0.999997	0.999998	0.5	0.999976	0.949437	0.999997	0.999998	1	0.999983	0.999994	1	0.999999	1	0.998739
13	0.999917	0.996391	0.576563	0.978618	0.972163	0.998285	0.547156	0.999944	0.995439	0.995278	0.972163	0.526679	0.5	0.992619	0.998566	0.972163	0.724688	0.999998	0.971399	0.00344	0.990205	0.999998	0.947007
14	0.847995	0.007719	0.992535	0.25919	0.525562	0.427813	0.113191	0.271037	0.652514	0.199653	0.525562	0.934708	0.761135	0.5	0.003492	0.525562	0.586211	0.79816	0.122975	0.777252	0.733189	0.35239	0.436801
15	0.999977	0.999841	1	0.999816	0.999971	0.994265	0.950099	1	0.979928	0.999878	0.999971	1	0.999997	1	0.5	0.999971	0.999998	1	0.999944	0.999937	0.999653	0.999958	0.99999
16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
17	0.997666	0.993218	0.967642	0.850586	0.967138	0.818366	0.910124	0.498581	0.992093	0.955066	0.956138	0.643801	0.010236	0.895742	0.717877	0.967138	0.5	0.736308	0.55424	0.603581	0.994668	0.699648	0.354761
18	2.41E-08	0.189099	0.728489	0.171437	0.101751	0.173894	0.03796	0.150154	0.04989	0.062549	0.967138	0.200115	0.811516	0.248173	4.99E-09	0.101751	0.078101	0.5	0.485613	0.651415	1.97E-05	0.01875	0.379654
19	0.999977	0.999998	1	0.999985	0.999989	0.998411	0.994492	0.999986	0.999998	0.999961	0.967138	0.999998	0.999997	0.999991	0.999207	0.999989	0.99947	0.996072	0.5	0.999976	0.999994	0.768124	0.999991
20	1	1	1	1	1	0.999992	0.999861	0.99999	0.945993	0.999997	0.967138	0.999999	0.998171	1	1	1	0.999999	1	0.999999	0.5	1	0.999995	0.999999
21	0.498765	0.350428	0.091664	0.207553	0.24257	0.117311	0.754122	0.263471	0.011163	0.413117	0.967138	0.505522	0.575887	1.20E-05	0.428764	0.24257	0.161112	0.289974	0.229848	0.059437	0.5	0.230011	0.481137
22	0.005257	0.937885	0.987965	0.789403	0.867444	0.782412	0.004315	0.637065	0.77056	0.669983	0.967138	0.116854	0.975511	0.984213	0.381051	0.867444	0.764238	0.475216	0.772541	0.96083	0.997403	0.5	0.135629
23	0.176668	0.697777	0.849406	0.372318	0.521284	0.036054	0.329773	0.170523	1.60E-05	0.473579	0.967138	0.626682	0.946288	0.563424	0.258629	0.521284	0.741131	0.827682	0.711489	0.578947	0.126144	0.953559	0.5
24	0.432145	0.467064	0.205466	0.100138	0.261865	0.420064	0.146033	0.014871	0.31222	0.40407	0.967138	0.37491	0.095759	0.136515	0.686847	0.261865	0.157549	0.474181	0.147084	0.278444	0.238614	0.105653	0.175601
25	0.5821	0.164205	0.476417	0.084197	0.50228	0.000639	0.110718	0.007806	1.44E-06	0.459259	0.967138	0.016796	0.042377	0.30554	0.000591	0.50228	0.522691	0.318846	0.252998	0.066857	0.909074	0.284157	0.000531
26	0.999928	0.999994	0.99997	0.988959	0.999938	0.933814	0.999782	0.971734	0.992845	0.999994	0.967138	0.999803	0.993433	0.997783	0.998997	0.999938	0.999958	0.999998	0.989956	0.999961	0.998479	0.996469	0.996291
27	0.443481	0.433153	0.392836	0.744171	0.279131	0.450374	0.007241	0.131676	0.768574	0.43184	0.967138	0.416934	0.009345	0.747838	0.971938	0.279131	0.005882	0.007075	0.003936	0.04784	0.225447	0.005532	0.180304
28	0.991202	0.997337	0.945923	0.996124	0.996478	0.997824	0.528905	0.997987	0.996639	0.973314	0.967138	0.952669	0.997817	0.999943	0.998491	0.996478	0.969465	0.997212	0.991678	0.982075	0.998204	0.99909	0.998219
29	0.999923	0.994252	0.98225	0.991108	0.997029	0.998674	0.883957	0.99616	0.997278	0.998418	0.967138	0.939362	0.579107	0.984378	0.999145	0.997029	0.998994	0.997763	0.852997	0.879611	0.997583	0.992439	0.972189

Output3:

Time interval =300sec



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	0.5	0.985425	0.992948	0.943663	0.999217	0.990711	0.964553	1	0.743422	0.999913	0.999217	0.996916	0.999897	0.999995	0.106993	0.999217	0.999769	0.999834	0.987049	0.999778	0.985857	0.999633	0.957548
2	0.723233	0.5	0.734783	0.889971	0.442642	0.46352	0.858834	0.381142	0.513353	0.191555	0.427642	0.317957	0.736336	0.433264	0.754133	0.427642	0.634929	0.621152	0.522093	0.64933	0.585356	0.461014	0.213309
3	0.945634	0.987895	0.5	0.999861	0.997714	0.987682	0.985341	0.999991	0.999993	0.99903	0.997714	0.990933	0.89884	0.933588	0.99998	0.997714	0.996895	0.983505	0.86441	0.990935	0.985753	0.975756	0.99965
4	0.460199	0.350462	0.245043	0.5	0.245587	0.130795	0.705687	0.391811	0.368876	0.413046	0.245587	0.397089	0.359939	0.134344	0.438359	0.245587	0.233918	0.315581	0.253953	0.357243	0.188941	0.170056	0.45794
5	0.967806	0.73311	0.501802	0.435224	0.5	0.305934	0.319132	0.855491	0.770071	0.608996	0.5	0.898618	0.701112	0.815171	0.996202	0.5	0.541175	0.908285	0.211646	0.005233	0.225909	0.139862	0.975492
6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
7	1	1	1	0.999998	1	1	0.5	1	1	1	1	1	1	1	1	1	0.999997	1	1	0.997088	0.999773	1	1
8	0.99411	0.967359	0.999995	0.99536	0.999161	0.956876	0.99946	0.5	0.99849	0.991659	0.999161	0.997339	0.999965	0.99977	0.997749	0.999161	0.919874	0.952169	0.995922	0.996909	0.999794	0.999733	0.987624
9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
10	0.019921	0.332293	0.693693	0.387726	0.730229	0.569757	0.285997	0.077629	0.442619	0.5	0.730229	0.804048	0.985114	0.865429	0.01657	0.730229	0.876093	0.5391	0.537655	0.988455	0.917057	0.968247	0.29538
11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
12	0.999797	0.999997	0.999392	0.999935	0.999951	0.999964	1	0.999677	0.998352	0.999954	0.999951	0.5	0.999883	0.997611	0.999971	0.999951	0.999998	0.999911	0.999964	1	0.999942	0.999995	0.992029
13	0.998345	0.984188	0.375552	0.846538	0.90282	0.992732	0.531305	0.999531	0.977816	0.94147	0.90282	0.551586	0.5	0.979191	0.99075	0.90282	0.582514	0.99981	0.932744	0.001248	0.953452	0.999693	0.928695
14	0.858035	0.094594	0.993595	0.27155	0.565474	0.444023	0.125995	0.364171	0.681986	0.242689	0.565474	0.875351	0.742657	0.5	0.012059	0.565474	0.594674	0.806733	0.261404	0.830265	0.777262	0.406743	0.202367
15	0.999808	0.999889	0.999984	0.998925	0.999523	0.979679	0.886783	1	0.968182	0.999528	0.999523	0.999998	0.999849	0.999993	0.5	0.999523	0.999812	0.999997	0.998778	0.998798	0.98159	0.999924	0.999922
16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
17	0.997258	0.995204	0.985654	0.737863	0.950126	0.7782	0.955322	0.769853	0.9842	0.969545	0.950126	0.662237	0.039748	0.885372	0.746464	0.950126	0.5	0.751488	0.539477	0.68176	0.958552	0.716739	0.392916
18	1.58E-06	0.298614	0.803354	0.212296	0.162718	0.163214	0.036382	0.164519	0.08574	0.12414	0.162718	0.325641	0.807148	0.328847	2.05E-07	0.162718	0.130751	0.5	0.411463	0.649353	0.000133	0.045502	0.588767
19	0.999997	0.999997	1	0.999993	0.999987	0.999016	0.993182	0.999995	0.999993	0.999943	0.999987	0.999999	0.999999	0.999994	0.999348	0.999987	0.999783	0.998458	0.5	0.999989	0.999982	0.889513	0.999979
20	0.999991	1	0.999987	0.999999	0.999993	0.999931	0.996935	0.999793	0.999971	0.999972	0.999993	0.999967	0.991867	0.999999	0.999998	0.999993	0.999992	0.999956	0.999964	0.5	0.999988	0.999834	0.999968
21	0.725431	0.598535	0.135672	0.433434	0.452705	0.32225	0.896207	0.473296	0.061477	0.657172	0.452705	0.71792	0.791374	0.004555	0.761807	0.452705	0.325621	0.484876	0.542585	0.303966	0.5	0.484674	0.646601
22	0.004285	0.927584	0.975256	0.637271	0.826479	0.597305	0.001808	0.687254	0.655449	0.572882	0.826479	0.086668	0.892144	0.987182	0.305076	0.826479	0.653699	0.527971	0.755435	0.817595	0.983516	0.5	0.079903

Observation:

When P≤0.05 means that correlation coefficient calculated for Internet usage patterns for an unknown subject (say b) is significantly smaller than that for a known subject (say a) and as such "subject b" will be identified as a subject distinct from "subject a". On the contrary, when P> 0.05, indicates that correlation coefficient calculated for Internet usage patterns for an unknown subject (say b) is not significantly smaller than that for a known subject (say a), and as such "subject b" will be identified as indistinguishable from "subject a".

The number of user combinations that are indistinguishable from each other (i.e, with p value > 0.05) and the number of user combinations distinguishable (i.e, P value <=0.05) is calculated.

Also as per my observation, when compared to 5 minutes and 227 sec time intervals I would suggest to take 10 seconds which has reasonably good amount of similarities between same

users across weeks i.e there are less statistically indistinguishable when compared with the other users and at 10sec authentication is present compare to 300sec and 227 sec.