Jorge Adrián Padilla Velasco
A00570894
4 de septiembre de 2018

Práctica 1.3

Introducción

EL objetivo de esta práctica es analizar si la información cumple con las cinco propiedades vistas en clase. Estas son: Validez, precisión, consitencia, uniformidad y por último, que se cuente con los registros completos. Finalmente, la práctica tiene como objetivo identificar posibles outliers, es decir, cualquier información que es numéricamente distante del resto de los datos.

Para esta práctica, trabajé con el archivo CSV obtenido tras descargar el grafo de profundidad 1 de la página de Facebook de Oracle, con ayuda de la aplicación Gephi.

Procedimiento

Con la profundidad 1, obtuve 50 páginas que están relacionadas con Oracle en Facebook. Trabajé directamente con los likes de cada página, y con esa información obtuve la media, desviación estándar y varianza con ayuda de excel. Con dicha información, apliqué la fórmula del filtro gaussiano para los likes de cada página. A continuación, normalicé los datos, es decir, obtuve la Z para cada página, para así poder aplicar la fórmula de distribución normal estandarizada.

Aquí anexo podemos apreciar el trabajo hecho en Excel:

	A	B	C	D	E	F	G	H
1	**Páginas de Facebook**	**Likes**	**Filtro Gaussiano**	**Z**	**Normal Standar Distribution**			
2	Oracle	2995652	7.9815519457285E-08	0.908900073344354	0.818298562457899		**Media**	545306.68
3	Cloud Odyssey: A Hero's Quest	1968	1.47978386716017E-07	-0.201539171671862	0.420138501591876		**Desviación Estándar**	2695945.783109
4	EAA - Young Eagles	12694	1.47976785777538E-07	-0.197560605015514	0.421694429934479		**Varianza**	7268123665462
5	Be The Match	465438	1.45789466944122E-07	-0.029625477077621	0.48818287322938			
6	International Federation of Red Cross and Red Crescent Societies	486391	1.45589581532544E-07	-0.021853436507934	0.491282434089002			
7	Second Harvest Food Bank	20144	1.47974295368965E-07	-0.194797196327299	0.422775858626589			
8	Oracle ERP Cloud	84193	1.47906283477774E-07	-0.171039671082803	0.432096286085874			
9	Oracle Supply Chain Management	3860	1.47978274465979E-07	-0.200837377143258	0.420412867803589			
10	Iron Man	18986275	2.51363151307758E-18	6.84025933887102	0.999999999996047			
11	Oracle Security	50107	1.47952869406587E-07	-0.183683100417904	0.427131033620817			
12	High Ground	7269	1.47977888251784E-07	-0.199572885838811	0.420907317561903			
13	Oracle Learning Library (OLL)	6012	1.47978058197736E-07	-0.200039141506071	0.42072498462083			
14	Oracle Profit Magazine Online	30099	1.47969203897283E-07	-0.19110461465063	0.424221816150827			
15	McGraw-Hill Education	44328	1.47958424161735E-07	-0.185826689519809	0.426290336923318			
16	OracleHCM	91194	1.47893790284594E-07	-0.168442808770562	0.433117466332506			
17	James Spithill	31875	1.47968083505594E-07	-0.190445847693547	0.424479886881123			
18	Oracle Customer Experience	222229	1.47476534421057E-07	-0.119838344681934	0.452305603072791			
19	BNP Paribas Open	247731	1.47354992321598E-07	-0.110378955639402	0.456054421052679			
20	Oracle Developer Community: ArchBeat Page	8765	1.47977644067541E-07	-0.199017978537124	0.421124340530854			
21	The Oracle ACE Program	4532	1.47978217056935E-07	-0.200588113970307	0.420510326427617			
22	Oracle Database	99506	1.47877664153549E-07	-0.165359660714663	0.434330451305693			
23	MySQL	524966	1.45199364921068E-07	-0.00754491434043	0.496990043224955			
24	IOUG - Independent Oracle Users Group	3368	1.47978310667678E-07	-0.201019873394882	0.420341517261392			
25	Sean D. Tucker	24544	1.47972293781215E-07	-0.193165116028219	0.423414828784908			
26	Oracle Health Sciences	20346	1.47974212109618E-07	-0.194722269004477	0.422805188729876			
27	Oracle Latinoamérica	76250	1.4791925105057E-07	-0.173985946949983	0.430938253247993			
28	ORACLE TEAM USA	259370	1.47295173959095E-07	-0.106061732320994	0.457766686510948			
29	Oracle Midsize	95760	1.47885105569964E-07	-0.166749154532925	0.433783713609538			
30	Oracle Developers	460977	1.45830921606595E-07	-0.031280183944484	0.487523046797935			
31	Oracle Cloud	240057	1.47392943160197E-07	-0.113225452051934	0.45492590879841			
32	Oracle Management Cloud	33982	1.47966671032606E-07	-0.189664303786692	0.424786097238272			
33	Oracle Primavera Enterprise Project Portfolio Management	46745	1.47956183633619E-07	-0.184930158137337	0.426641908088594			
34	Oracle Financial Services	11883	1.47976988681865E-07	-0.197861427088822	0.42157674210434			
35	Oracle Communications	8578	1.47977677082418E-07	-0.199087341949835	0.421097211347637			
36	Oracle Linux	59931	1.47941867043267E-07	-0.180039110222867	0.428560932159354			
37	Oracle System Developers	4190	1.47978247422944E-07	-0.200714971120827	0.420460726340024			
38	Oracle IT Infrastructure	170079	1.47684244587591E-07	-0.13918220549944	0.444653084794283			
39	Oracle Solaris	14296	1.47976345622136E-07	-0.196966379415712	0.421926923849999			
40	Oracle Applications	53030	1.47949801021275E-07	-0.182598879801038	0.427556381677522			
41	Oracle Magazine	191297	1.47606363667005E-07	-0.131311869184467	0.44776430142052			
42	I <3 Java	316348	1.46963154690098E-07	-0.084927034302588	0.466159699580086			

Oracle_Outliers

Conclusión

Analizando los datos y el trabajo realizado, podemos decir que los datos son:

1. Válidos: todos los datos siguen el mismo esquema.

2. Precisos: Las páginas tienen cierta cantidad de “likes”.

3. Consistentes: se ajustan o se igualan a otros datos.

4. Uniformidad: se usa la misma unidad de medición para los datos, la cual es la cantidad de “likes”.

5. Completos: se tienen todos los datos necesarios para el procedimiento. En este caso, fueron todas las páginas de Facebook que se relacionan directamente con Oracle.

Por último, identifiqué dos outliers, los cuales fueron la misma página de Oracle y la página de Iron Man. Esto se debe a que al aplicar el filtro gaussiano y la distribución normal estandarizada, los resultados obtenidos de estas páginas son los únicos resultados que resaltan de los demás. Esto se puede apreciar fácilmente en la imagen anexa en la página anterior, donde las celdas de los outliers se resaltaron en color rojo.