**Jorge Adrián Padilla Velasco**

**A00570894**

**4 de septiembre de 2018**

**Práctica 1.3**

**Introducción**

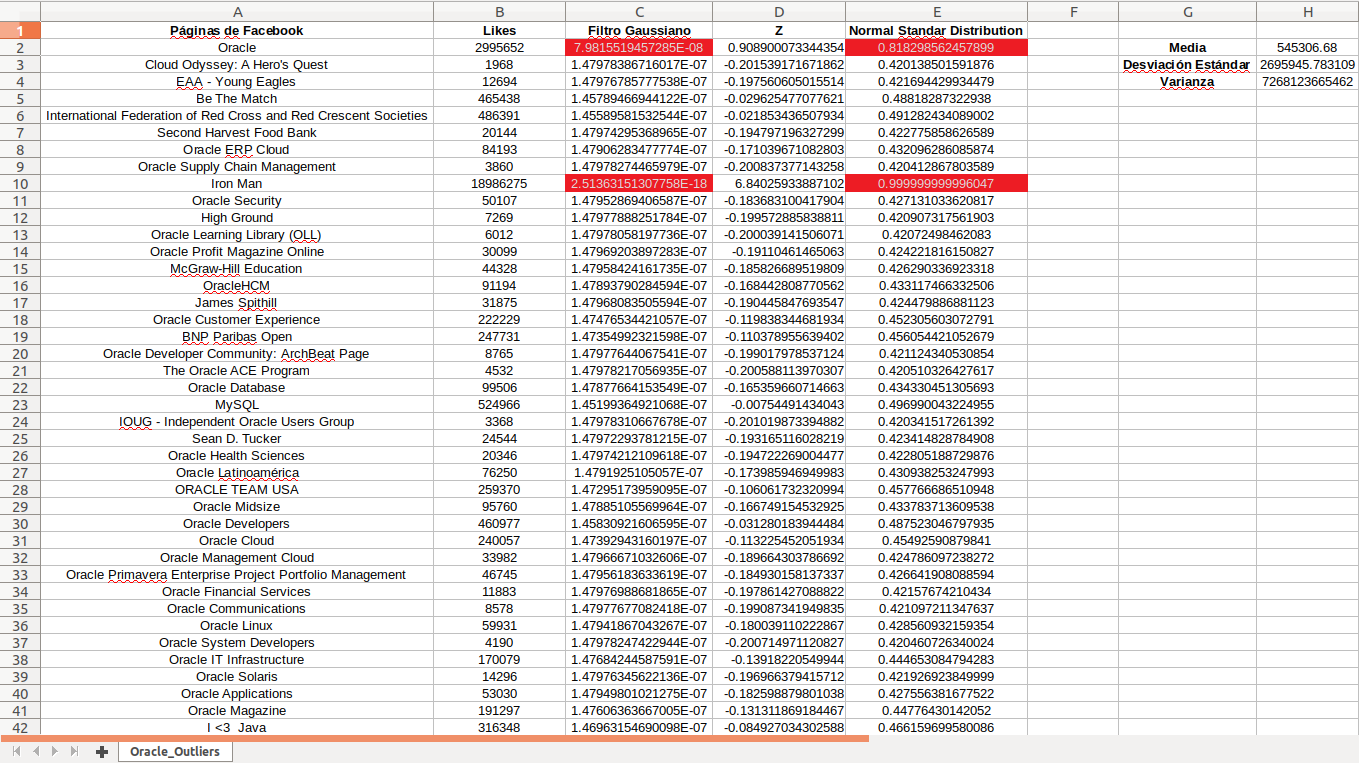
EL objetivo de esta práctica es analizar si la información cumple con las cinco propiedades vistas en clase. Estas son: Validez, precisión, consitencia, uniformidad y por último, que se cuente con los registros completos. Finalmente, la práctica tiene como objetivo identificar posibles outliers, es decir,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



**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.