

UNIVERSIDAD TECNOLÓGICA METROPOLITANA

Ingeniería en Desarrollo y Gestión de Software

Extracción en Conocimiento en Base de Datos

ISC. Mirian Magaly Canché Caamal

Practica 2

Integrantes:

Díaz Sanjuán Brandon Emanuel

Cano Kú David Emmanuel

Urtecho Pedro José

Yama Sonda Fernando Jesús

9no Cuatrimestre.

GRUPO: 9ª

Parcial 3

22 de Julio 2021

**Desarrollo**

Como ya vimos la visualización de datos es un proceso por el cual realizamos una representación de datos de forma visual.

* **Técnicas de visualización y representación de información:**

Existen grupos generales. Hay al menos cuatro de ellos:

Temporal: se refiere al tiempo y la investigación de cambios durante un período específico.

Jerárquico: representa una relación entre diferentes puntos de datos.

De red: se trata de relaciones que se indican mediante líneas que conectan puntos.

Geoespacial: es una categoría que describe áreas geográficas e intenta transmitir una sensación de espacio.

Tipo de representación:

*Gráfico de barras*: utiliza un conjunto de barras para categorizar diferentes puntos de datos;

*Gráfico de área*: contiene una variedad de capas que forman un área debajo de la línea;

*Gráfico circular o pizza*: define partes de un todo utilizando la metáfora de las porciones de pizza;

*Mapa de calor*: es un mapa con diferentes colores para representar diferentes niveles de intensidad.

Cuadros

Tablas

Gráficos

Mapas

Infografía

Dashboards

Existe el gráfico de marcadores, el diagrama de dispersión, el histograma y el gráfico de burbujas.

* **Proceso de Storytelling**

Data Storytelling consiste en un enfoque estructurado sobre cómo comunicamos insights a partir de los datos y son tres los elementos que están relacionados: datos, visualización y narrativa.

Objetivo: Generar un impacto o un cambio en las acciones de las personas basadas en los resultados de un análisis

La conjunción de narrativa y datos permite explicar por qué un insight puede ser tan importante. A su vez, cuando añadimos una visualización a nuestros datos, podemos iluminar (“enlighten”) a la audiencia con insights que no habían visto de otra manera. Por otro lado, la combinación entre narrativa y visualización genera engagement o interés en la audiencia. Si unimos narrativa, datos y visualización logramos influenciar y generar un impacto en las acciones de las personas.

**¿Cómo crear un data storytelling relevante?**

*Ser claro y conciso*: La idea de invertir en storytelling basado en datos es simplificar la lectura de enormes y complejas cantidades de datos. Por lo tanto, es fundamental utilizar un lenguaje claro y conciso.

Piensa en la carga cognitiva de tu buyer persona y elige imágenes que brinden la mayor cantidad de información con el menor esfuerzo posible. Cuanto más conozcas a tu audiencia, mayores serán las posibilidades de acertar en esta elección.

*Resalta los insights*: Como ya hemos mencionado, convertir a los datos en algo más accesible es muy útil para expresar un punto de vista o vender una idea.

Es esencial que identifiques cuál es la información principal que tu data storytelling está tratando de transmitir. Sin él, hay grandes posibilidades de que se trate de un contenido de visualización de datos en lugar de una historia.

*Combina palabras con imágenes:* Contar un data storytelling no significa que debemos contar una historia sin palabras. Al revés, las palabras deben usarse para hacer que las imágenes sean aún más atractivas.

*Hazlo compartible:* Si deseas que tu historia llegue a más personas, debes crearlas de forma que sean simples de compartir. ¿Cómo hacemos eso? Dos características son esenciales aquí:

La primera es el atractivo visual, por supuesto. Descubre los gustos de tu audiencia e identifica formatos y patrones de diseño que sean más adecuados para ellos.

En segundo lugar, no subestimes el contexto. ¿Por qué les cuento esta historia a estas personas a través de estos datos?

**Ejemplo de un StoryTelling:**

*Google Maps*

Google Maps proporciona un informe mensual sobre viajes a los usuarios que activan la función «Historial de ubicaciones» en sus dispositivos móviles. Es posible explorar esta función interactuando con Google Maps.

Puedes saber los lugares y ciudades que más visitaste, ver fotografías tomadas en cada ubicación y adquirir información sobre los medios de transporte más utilizados.

¿Sabes cuánto tiempo pasaste en el transporte público o en tu automóvil durante el último mes? Tal vez anduviste en bicicleta más de lo que caminaste, ¿o no?

Toda esta información se puede rastrear a través de esta herramienta. Como puedes ver, Google Maps cuenta con todas las características principales que debe tener una buena data storytelling: contexto, giros narrativos y personajes.

* **Proceso de elaboración de gráficas con herramientas de visualización de datos.**

En la actualidad el número de herramientas y plataformas de visualización de datos al que se puede acceder es muy extenso. En función de las necesidades del usuario se puede elegir entre un buen número de opciones de vanguardia.

Aunque cada herramienta tiene sus propias características, por lo general funcionan bajo la lógica de importar los datos, elegir las opciones de visualización y luego publicarla.

La mayoría de las herramientas de visualización de datos son capaces de conectarse con fuentes de datos como bases de datos relacionales. Estos datos, que pueden almacenarse en las instalaciones o en la nube, se recuperan para su análisis. Los usuarios pueden seleccionar la mejor manera de presentar los datos de numerosas opciones. Algunas herramientas ofrecen automáticamente recomendaciones de visualización, según el tipo de datos presentados.

La mayoría de herramientas de web de código abierto, su objetivo es proporcionar un enlace faltante entre las aplicaciones de hoja de cálculo (por ejemplo, Microsoft Excel, JSON y Apple Numbers) y los editores de gráficos vectoriales (por ejemplo, Adobe Illustrator y Sketch). Simplemente se insertan datos para personalizar sus gráficos y exportarlos como imágenes vectoriales (SVG) o ráster (PNG). También puede visualizar informes generados a través de escritorio, navegador, dispositivo móvil o incrustado en cualquier aplicación.

* **Proceso de elaboración de gráficas con bibliotecas de visualización de datos.**

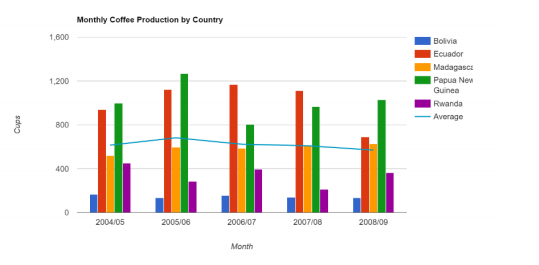
Existen múltiples bibliotecas que facilitan la creación de gráficos 2D para visualizaciones de datos. Dentro de este grupo podemos diferenciar las que están basadas en Canvas, y las que están basadas en SVG.

Tenemos el ejemplo de D3.js D3 ayuda a dar vida a los datos usando HTML, SVG y CSS. El punto clave de D3 es la utilización de estos estándares web para que el editor/desarrollador pueda explotar todas las nuevas capacidades de los navegadores web modernos, facilitando la manipulación del DOM para la creación de poderosas visualizaciones basadas en datos

D3 nos da una capa de abstracción por encima de SVG ofreciendo una serie de utilidades para la manipulación de datos y su posterior visualización. De este modo, D3, nos ofrece la creación de gráficos interactivos que pueden ser personalizados utilizando estándares con CSS3 presentación de información con el proceso de extracción del conocimiento.

Por ello las bibliotecas como NVD3 nos ofrecen clases Javascript para los principales tipos de gráficas, como, por ejemplo: grafico de barras, de líneas, de área, de burbujas, etc., facilitando la creación de este tipo de visualizaciones.

Igualmente, el caso de Google Charts, esta biblioteca también está basada en SVG y dispone de un amplio número de tipos de gráficos y mapas. Dispone también de lo que han denominado datos dinámicos, que nos permiten hacer consultas SQL sobre datos almacenados en Google preadsheets, Google Fusion Tables o SalesForce y visualizarlos. Google ofrece también, aunque de forma limitada, algunos componentes para creación de dashboards interactivos.



Google Charts

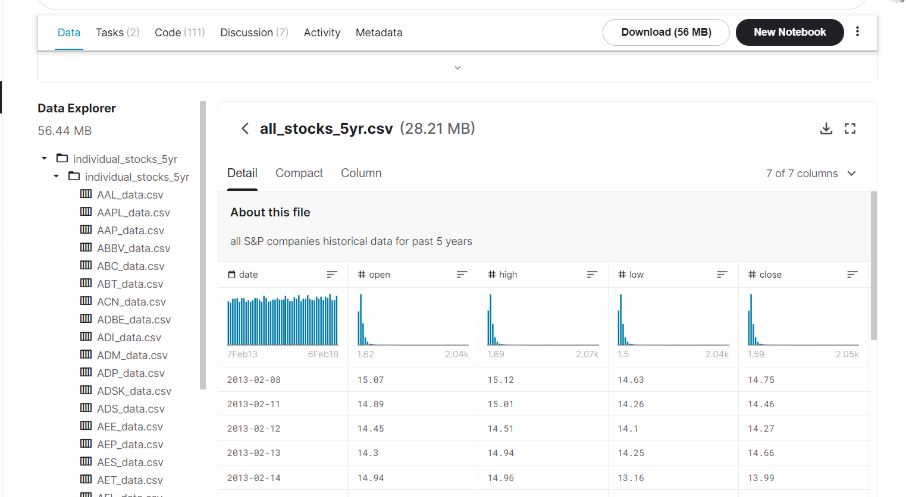
**DASHBOARD A PARTIR DE UN CASO PRÁCTICO**

**Resultados**

Se utilizaron las técnicas de clustering están disponibles en scikit-learn en el paquete sklearn.cluster.

se utilizaron los datos semanales de los **últimos 5 años**. Este extracto muestra qué forma tienen los datos. Cada columna representa a cada empresa.

Las gráficas que se realizaron en la interpretación de resultados, fueron hechas en un software especializado para la interpretación de datos masivos de forma graficas: <https://www.tableau.com/es-mx>

Este software fue instalado y vinculado con una base de datos, en este caso y nuestro caso de uso nos facilitó en que toda la base de datos lo tenía organizado en hojas de Excel, entonces nos dimos la tarea de poder vincular Excel con Tableau. Se descargaron los datos en el siguiente repositorio.

Cada uno de los datos fueron agarrados las siguientes empresas para generar el resultado de las gráficas.

- Apple Inc. (AAPL)

- 3M Company (MMM)

- United Health Group Inc. (UNH)

Caterpillar Inc. (CAT)

- Chevron Corp. (CVX)

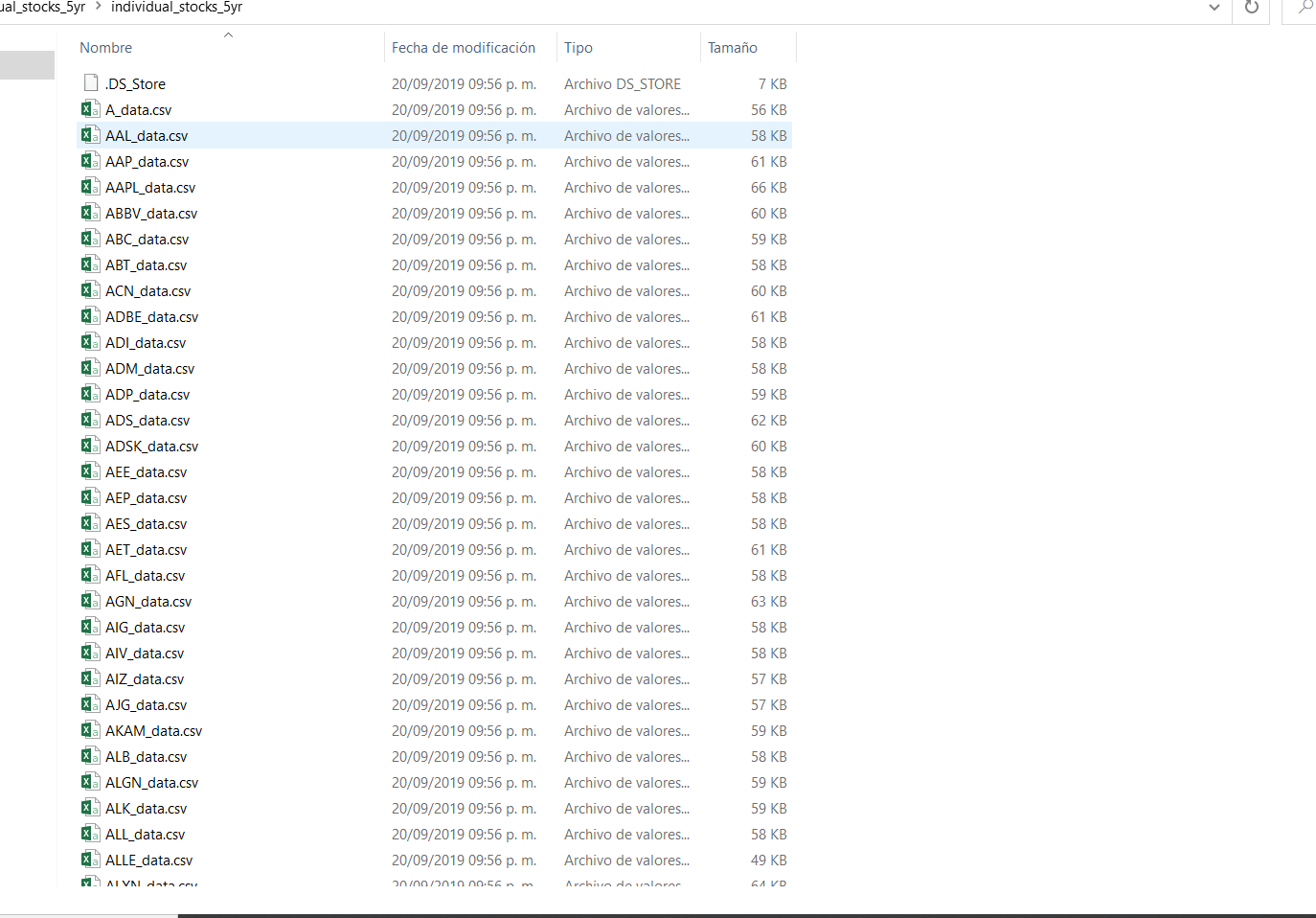
- International Business Machines (IBM)

- Exxon Mobil Corp. (XOM)

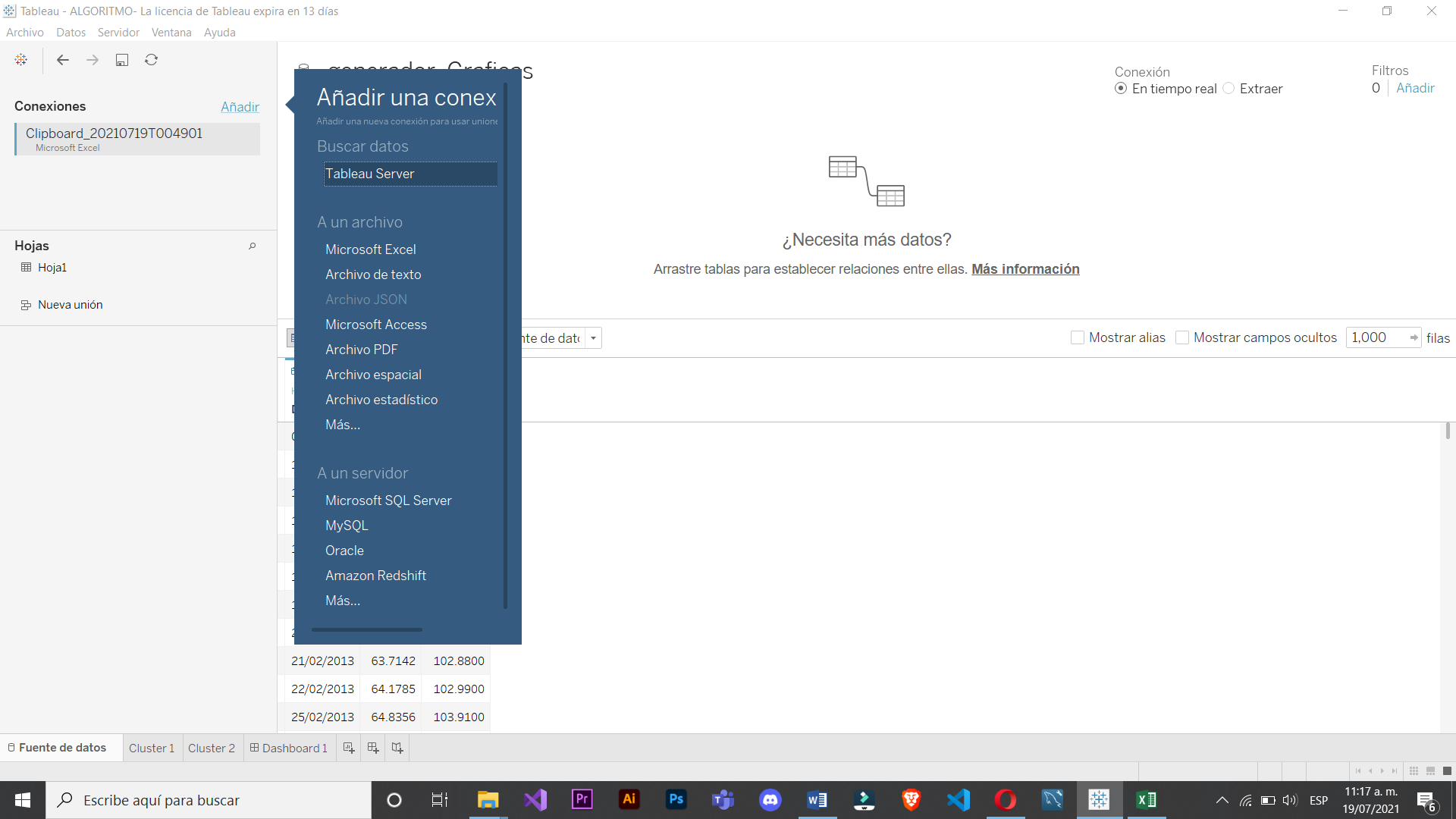
Home Depot (HD)

- Microsoft Corp. (MSFT)

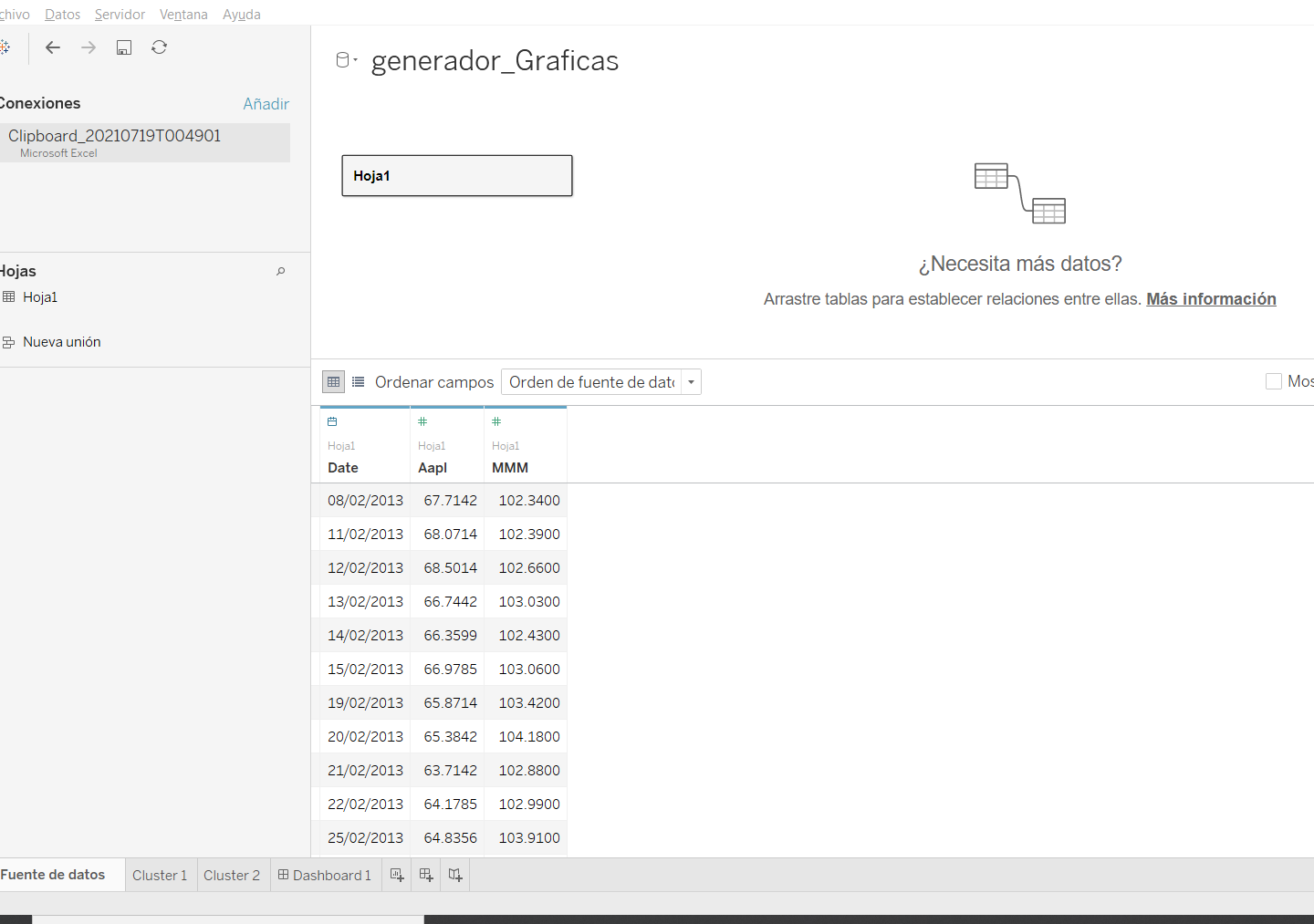
- Visa Inc. (V)



Se realizo cada uno de las comparaciones con los resultados proporcionados por el algoritmo las cuales son el cluster 1 , cluster 2 , cluster 3 y cluster 4 en un excel y fue vinculado con la plataforma de **Tableau** [**https://www.tableau.com/es-mx**](https://www.tableau.com/es-mx)



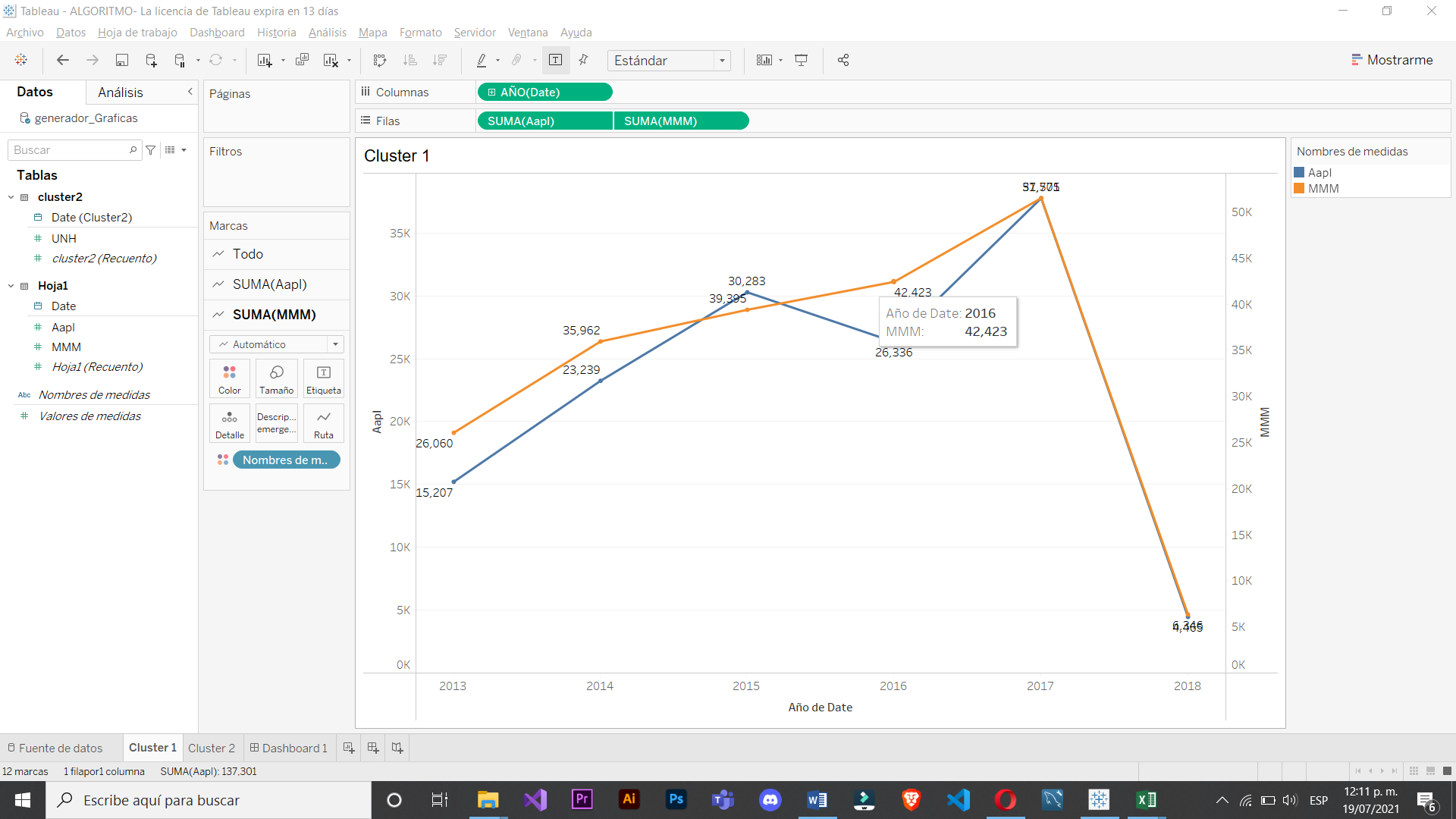
Fue vinculado cada una de las tablas creadas, basandonos en la especificación de cada cluster, aquí mostramos un pequeño ejemplo, de como se vinculo los datos masivos para generar las graficas.

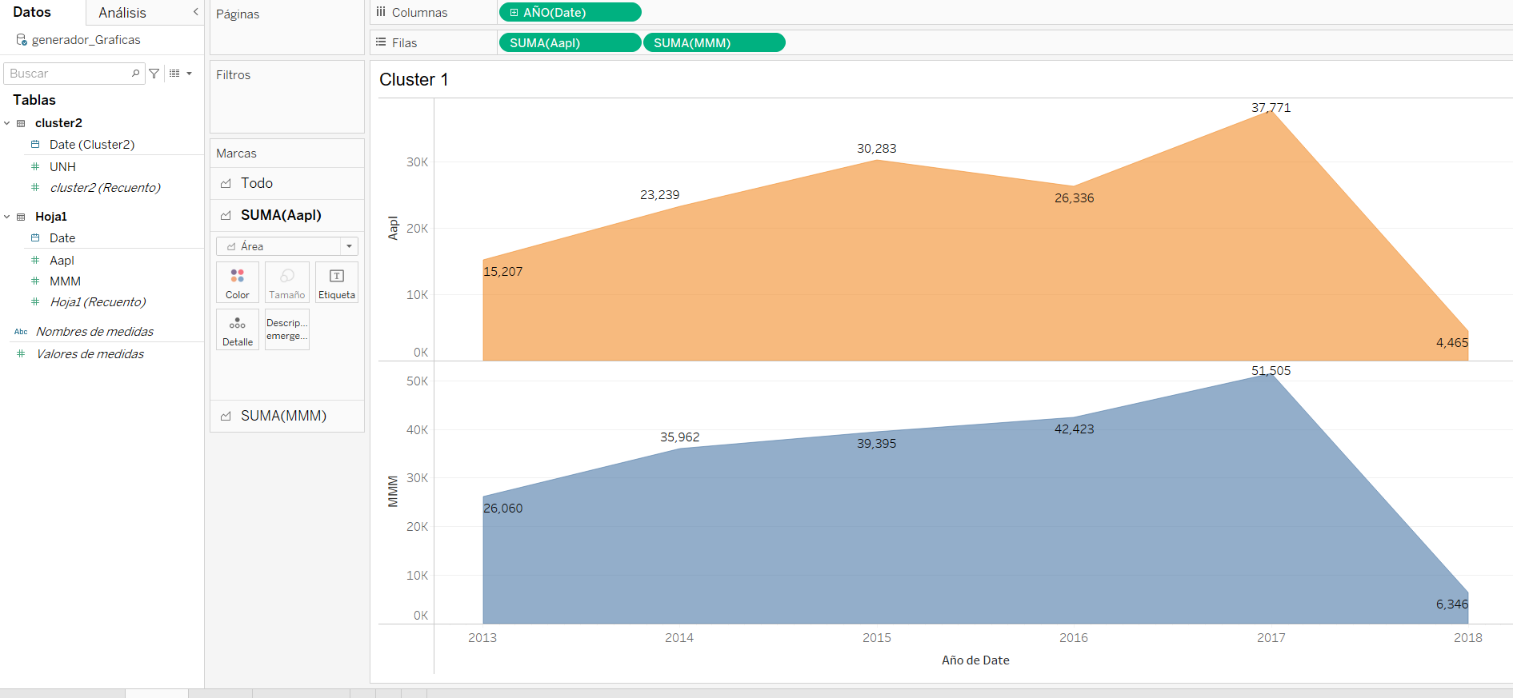


En los siguientes resultados se han comparado el comportamiento de las acciones de cada uno de las empresas en el mercado bursátil estadounidense. Se agruparon cada una de las acciones mediante el comportamiento de las ganancias. De esta forma se compararon el rendimiento de cada acción con las demás empresas.

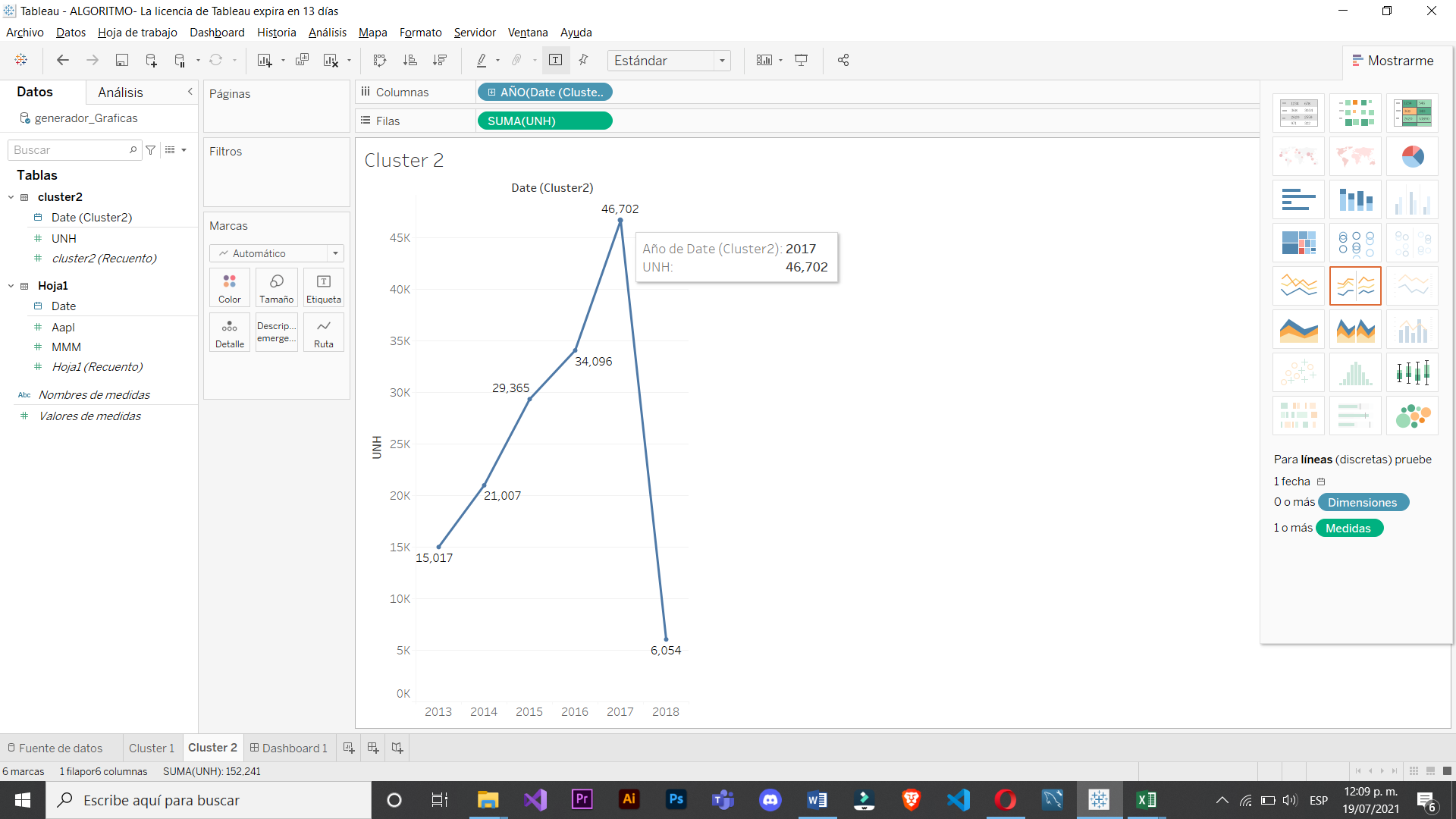
Cluster 1: Apple Inc. (AAPL) vs 3M Company (MMM) ###

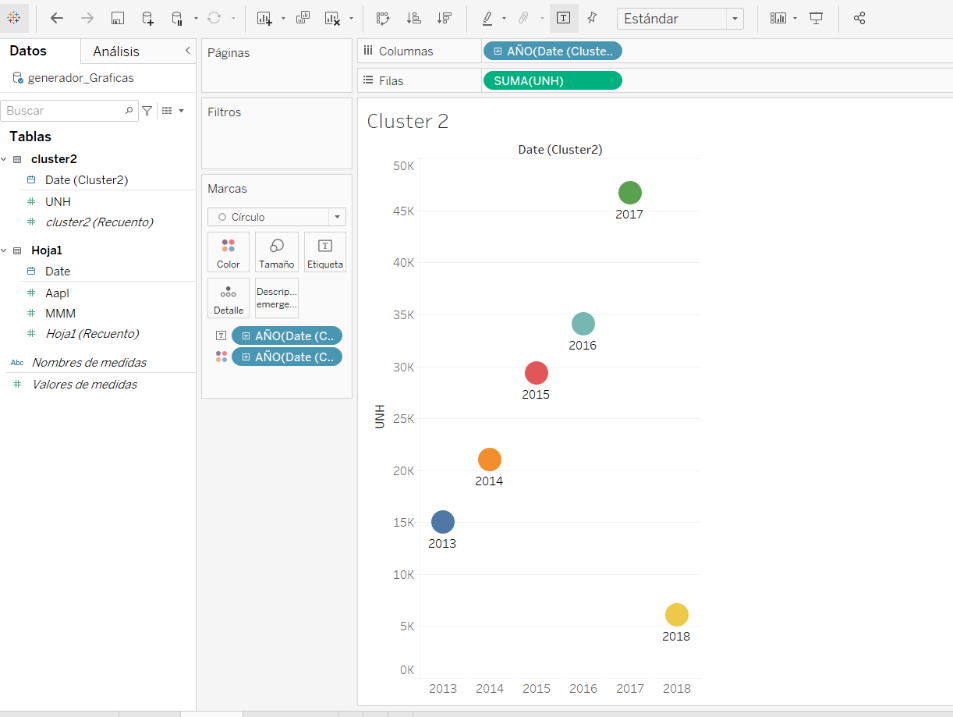
Gráfica 1



Gráfica 2

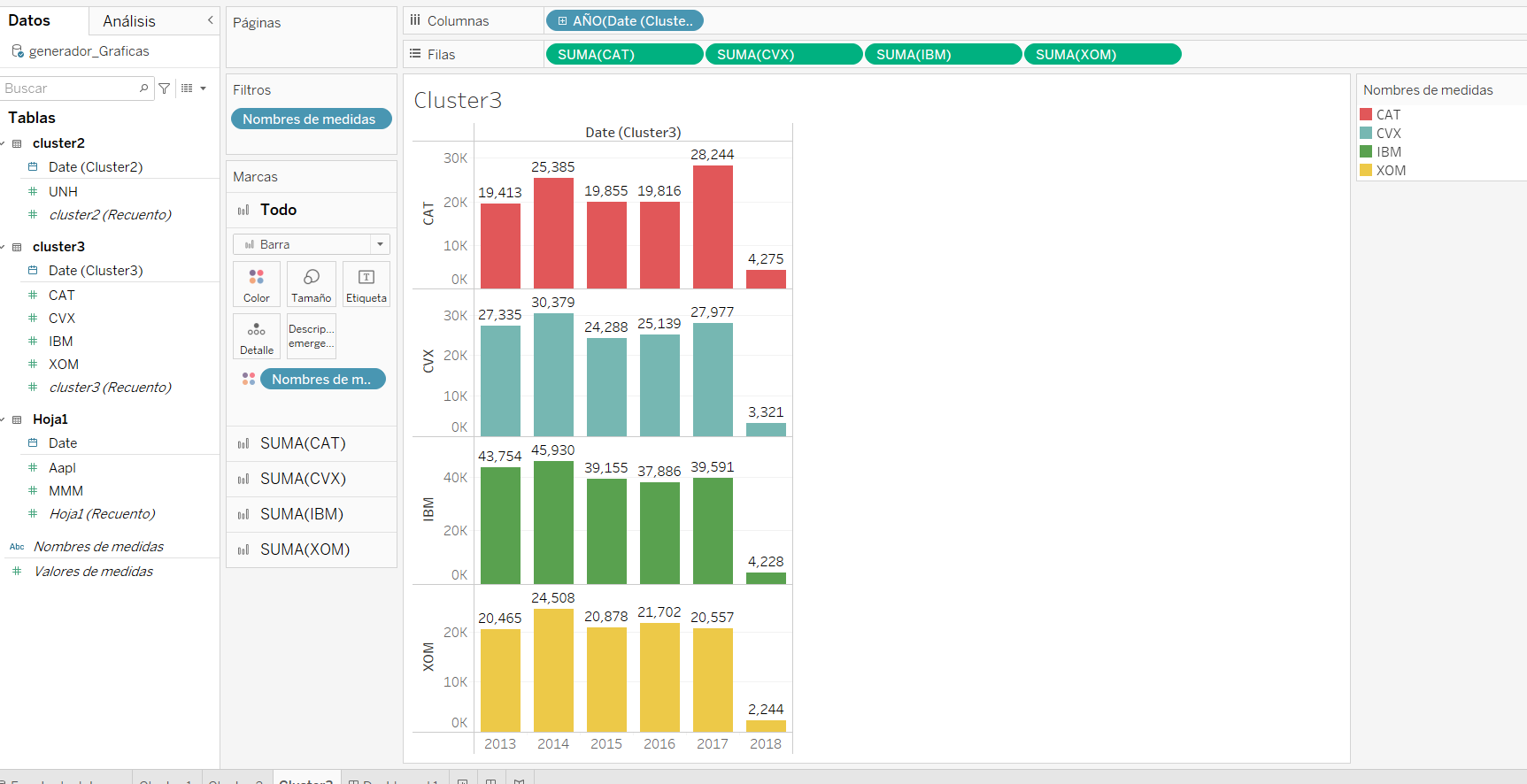
Cluster 2: United Health Group Inc. (UNH)

Gráfica 1

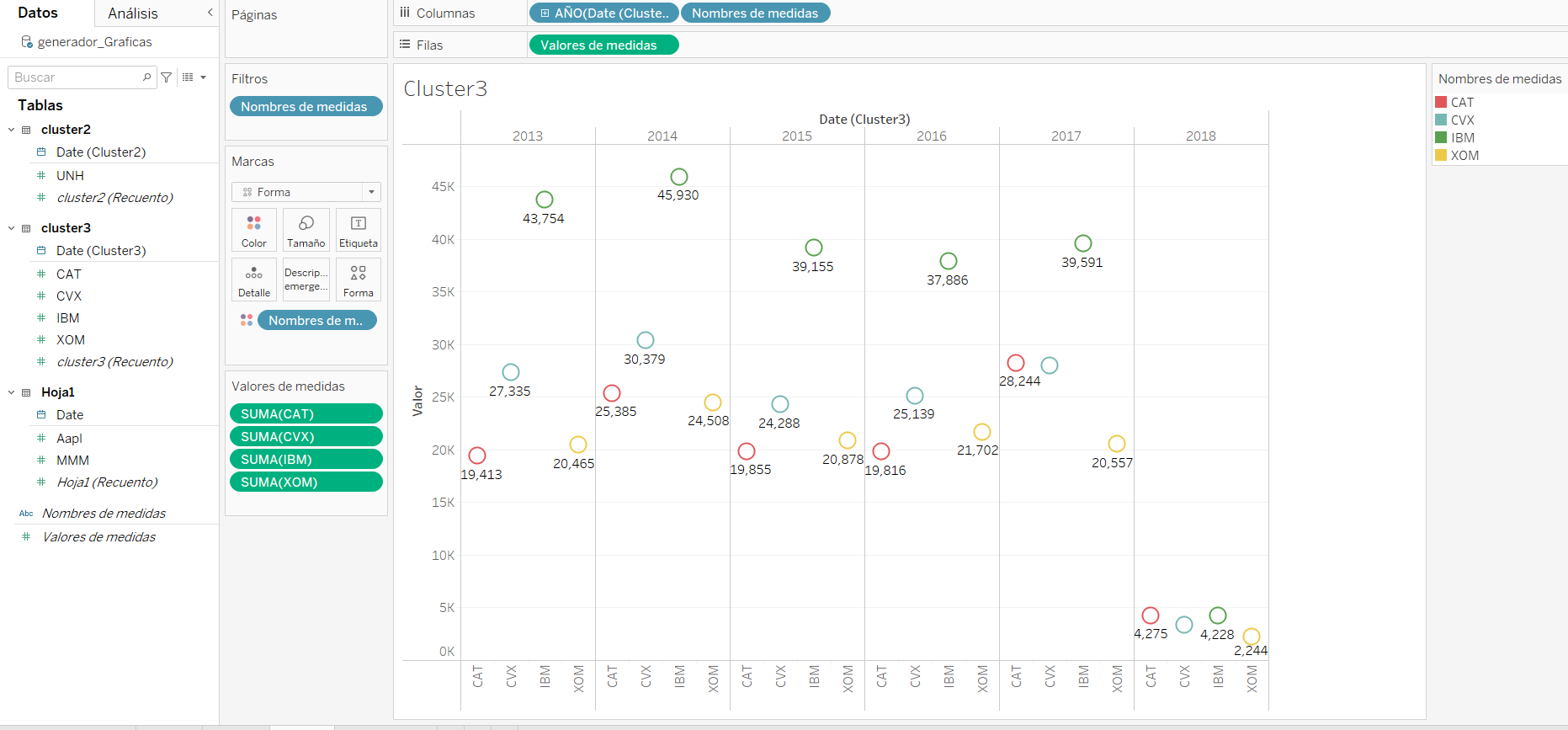
Gráfica 2

Cluster 3: Caterpillar Inc. (CAT) vs Chevron Corp. (CVX) vs International Business Machines (IBM) vs Exxon Mobil Corp. (XOM) ###

Gráfica 1

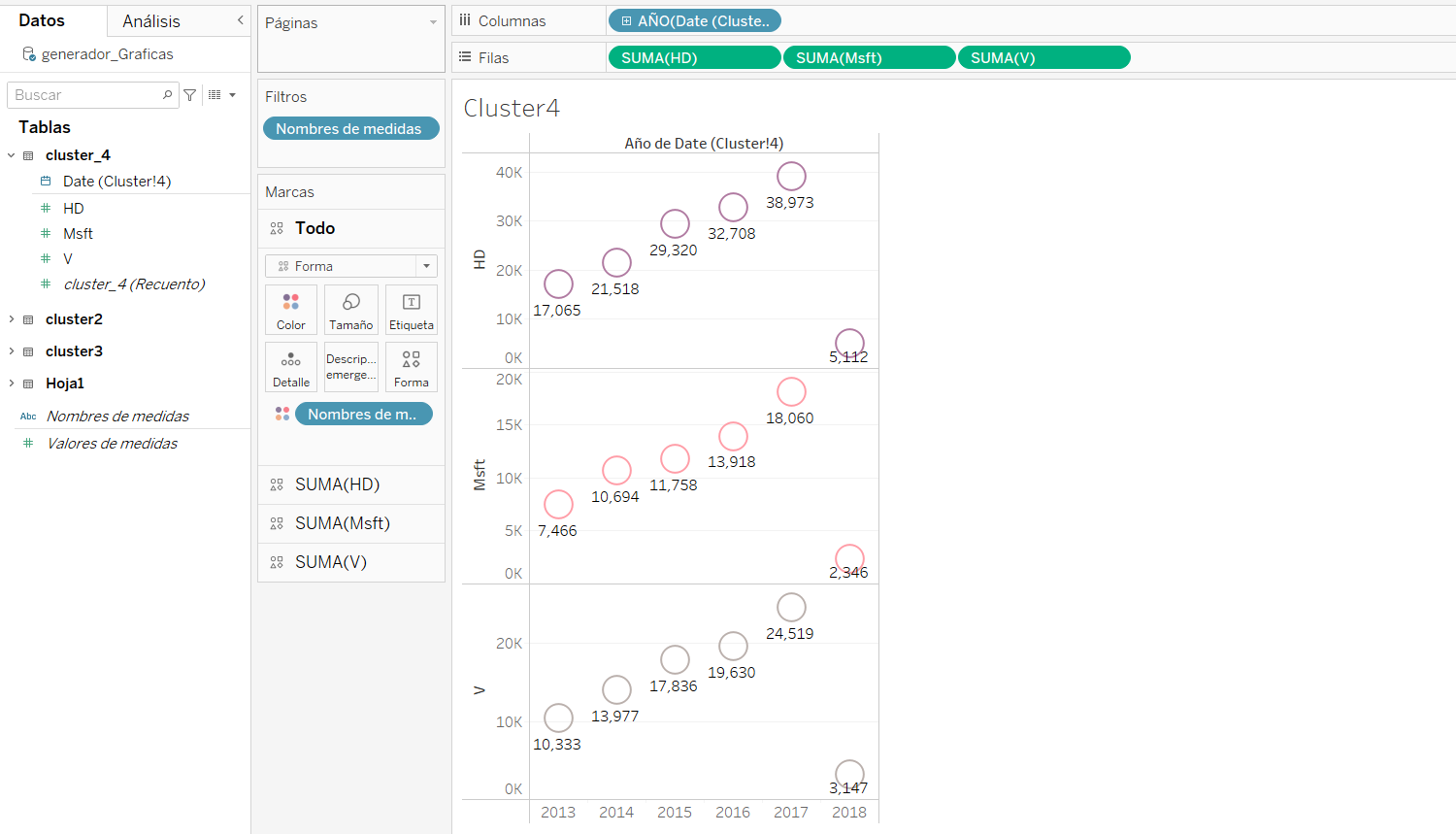


Gráfica 2

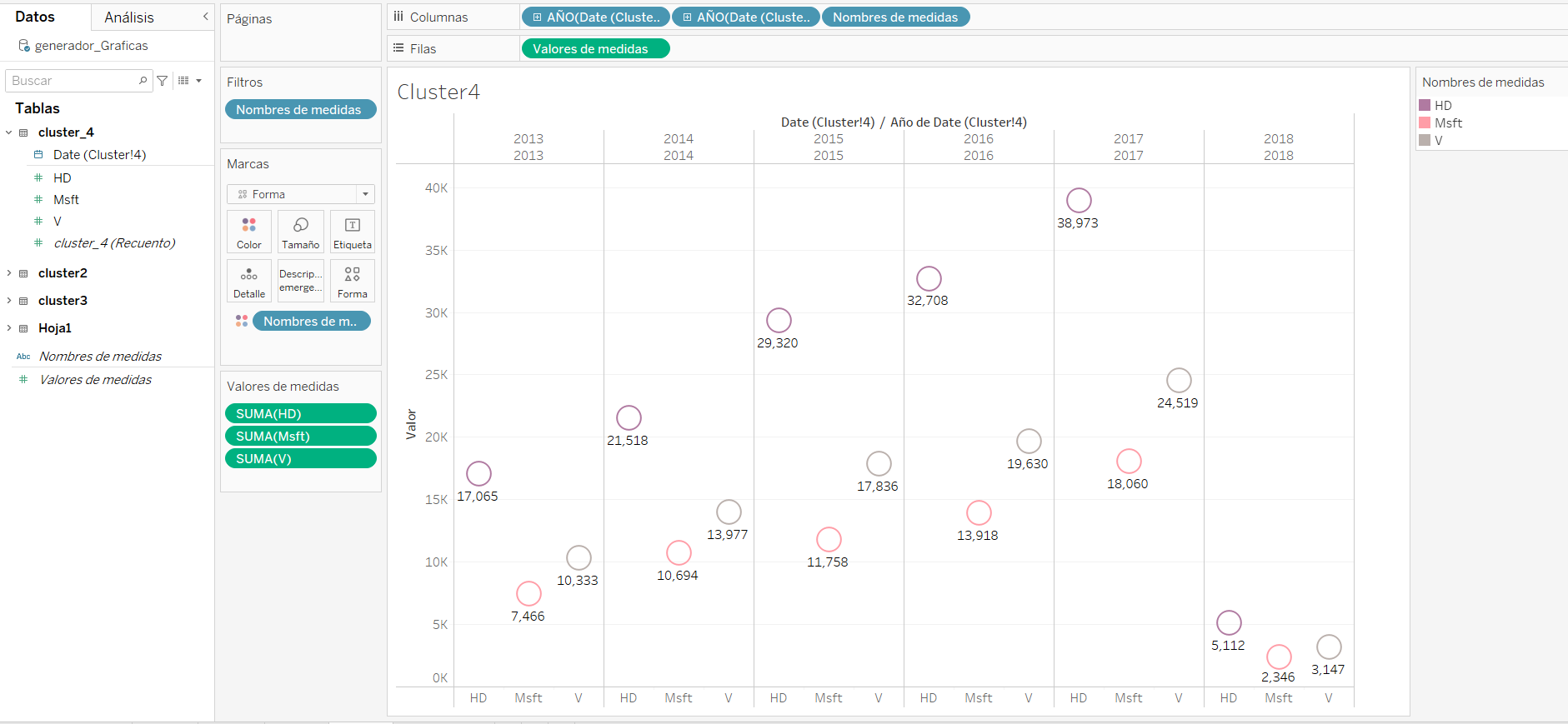


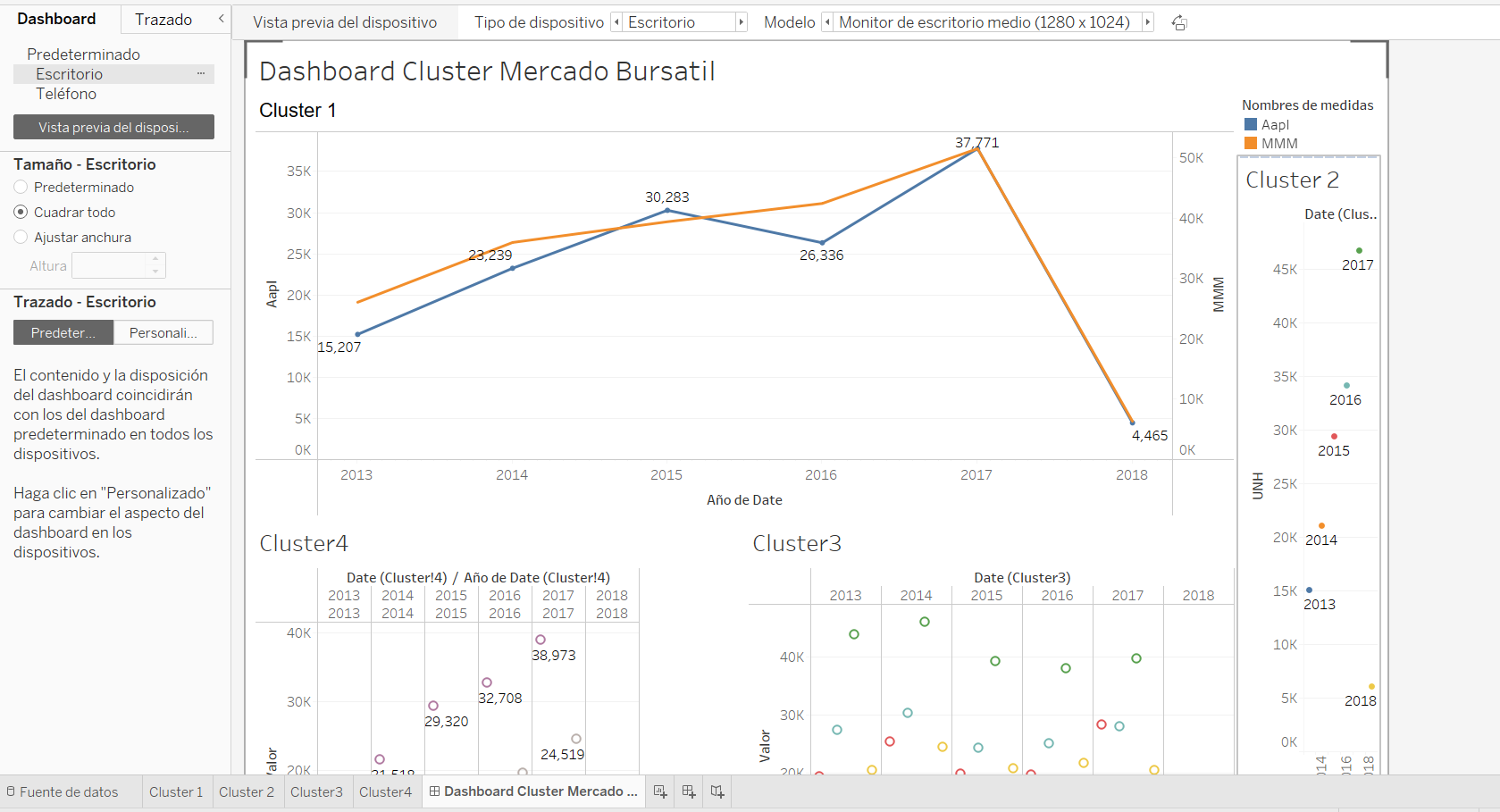
Cluster 4: Home Depot (HD) vs Microsoft Corp. (MSFT) vs Visa Inc. (V) ###

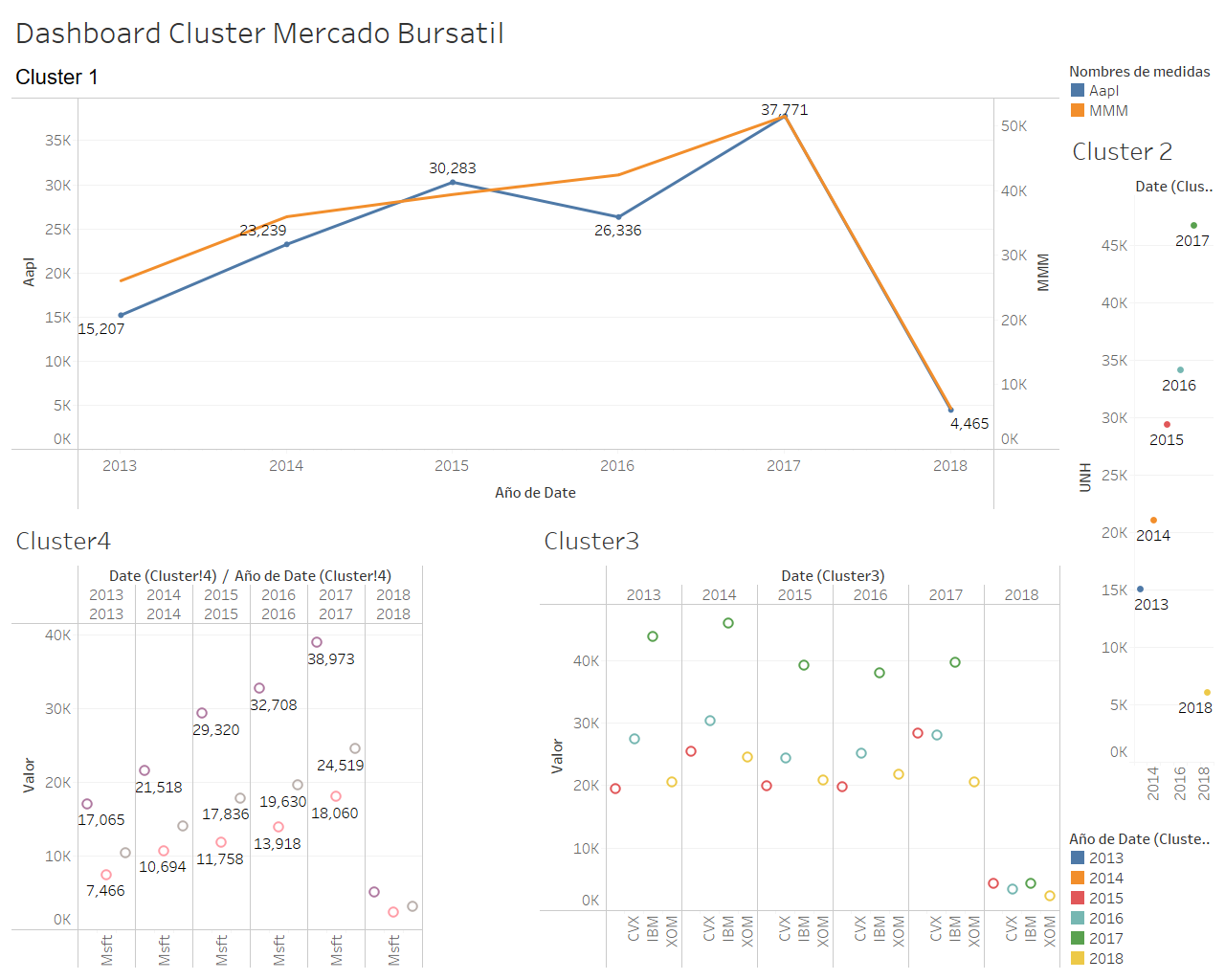
Gráfica 1



Gráfica 2



Podemos generar un Dashboard, en donde aparecerán todas las gráficas de los cluster realizadas.

Resultado Final

Nota: Todos los archivos generados en Excel y el proyecto hecho en Tableau estarán disponibles en el siguiente repositorio:

<https://dataiq.com.ar/blog/lo-que-necesitas-saber-sobre-data-storytelling/#:~:text=Data%20Storytelling%20consiste%20en%20un,los%20resultados%20de%20un%20an%C3%A1lisis>.

<https://datos.gob.es/es/documentacion/visualizacion-de-datos-definicion-tecnologia-y-herramientas>