

Módulo: ADM1929 - Business understanding: Pensamiento analítico basado en datos. - (A51)

**Actividad: Reto de aprendizaje 17.
Evaluación de clúster**

Nombre: Roberto Mora Balderas

Asesor: José Carlos Soto Monterrubio

Fecha: 24 de julio de 2023

Objetivo:

Analizar cada uno de los elementos visuales que genera el auto model de Rapid miner con la finalidad de comprender a fondo el proceso de segmentación.

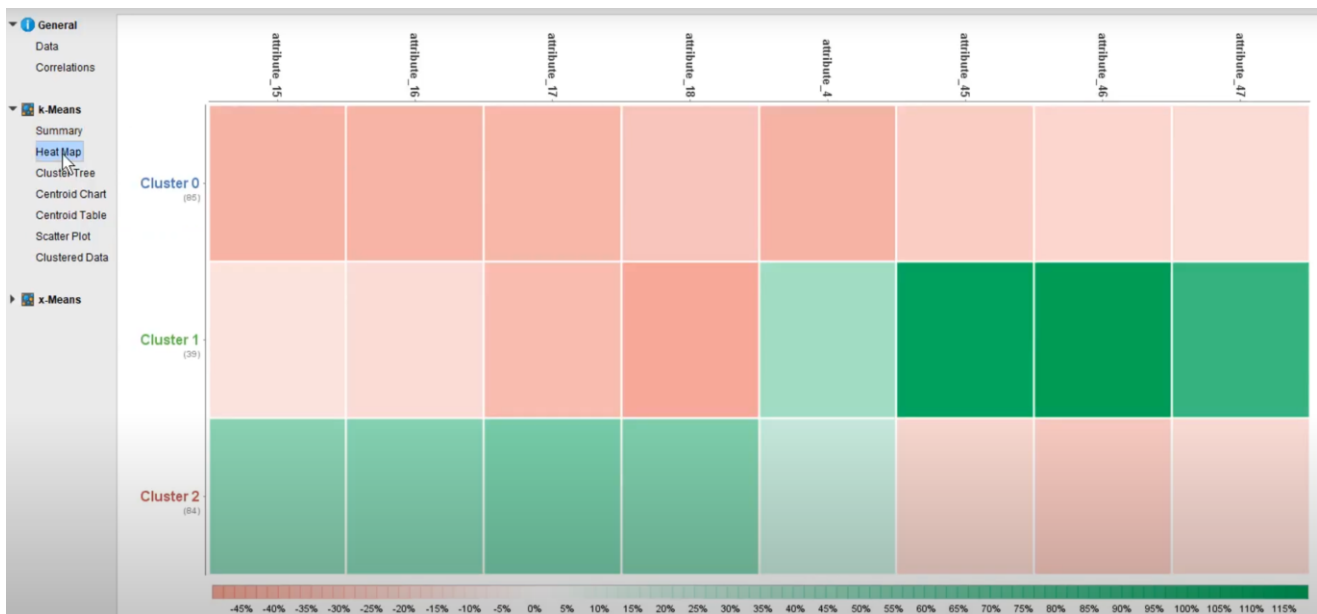
Instrucciones:

1. Desarrolla un documento en el que expliques detalladamente cada uno de los siguientes elementos. Toma de base el video del último tema donde explico cada rubro.
 - a) Heat Map (se refiere a que atributo es más importante en cada clúster)
 - b) Centroid Chart (muestra los centroides en una gráfica paralela)
 - c) Centroid Table (muestra los valores de los centroides)
 - d) Scatter Plot (gráfica de dispersión de los dos atributos más importantes)

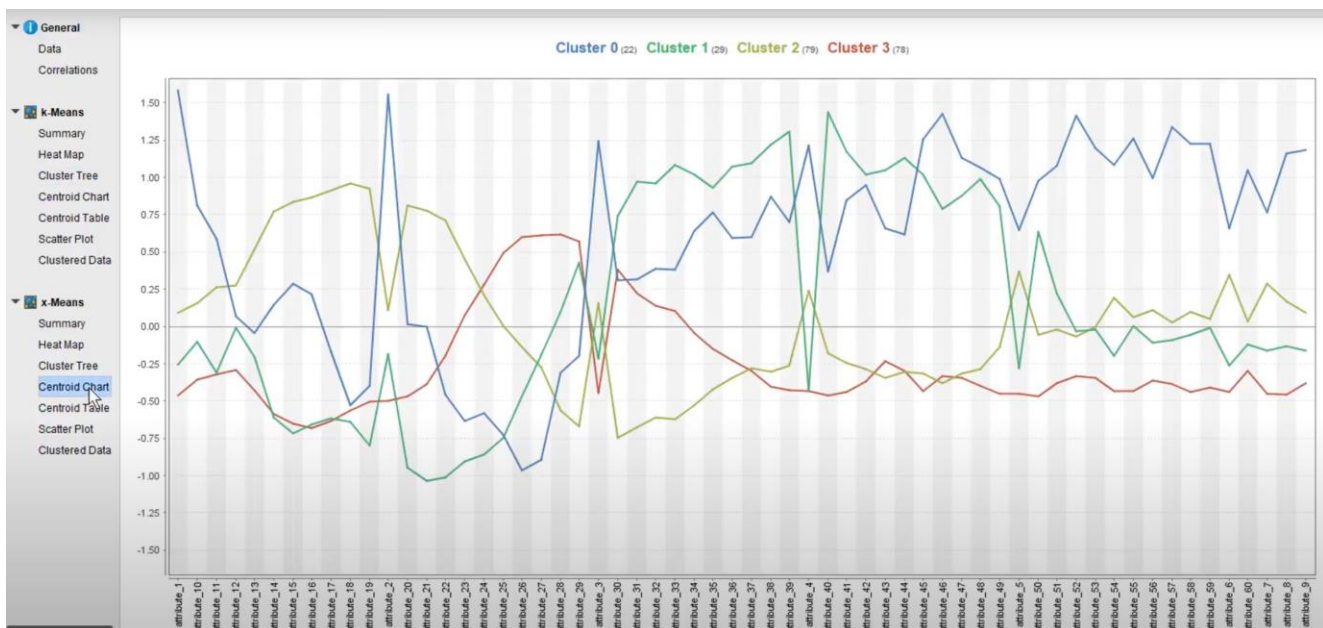
Desarrollo:

En RapidMiner, los términos mencionados están relacionados con la visualización y análisis de los resultados obtenidos mediante el proceso de clustering. A continuación, se explica cada uno de ellos:

a) Un Heat Map o mapa de calor es una visualización que muestra la importancia o relevancia de cada atributo en cada clúster obtenido a través del análisis de clustering. La importancia se representa mediante colores o tonos diferentes, donde los colores más intensos indican una mayor importancia. Esta herramienta ayuda a identificar qué atributos tienen mayor influencia en la formación de cada clúster, lo que proporciona una mejor comprensión de las características distintivas de cada grupo.



b) El Centroid Chart es una gráfica paralela que muestra la ubicación de los centroides de los clústeres en un espacio multidimensional. Un centroeide representa el punto medio de un clúster y está determinado por el promedio de los valores de atributos de los datos pertenecientes a ese clúster. En la gráfica paralela, cada atributo se representa como una línea vertical y el valor del atributo para el centroeide se muestra en esa línea. Al observar esta visualización, se pueden comparar las ubicaciones relativas de los centroides y entender cómo se distribuyen en el espacio de atributos.



c) La Centroid Table o tabla de centroides es una representación tabular que muestra los valores específicos de cada atributo para los centroides de los clústeres. Cada fila de la tabla representa un clúster y cada columna corresponde a un atributo. Al ver la tabla, se puede obtener información detallada sobre las características numéricas de cada clúster y cómo se diferencian en términos de valores de atributos.

General

Data

Correlations

k-Means

Summary

Heat Map

Cluster Tree

Centroid Chart

Centroid Table

Scatter Plot

Clustered Data

x-Means

Summary

Heat Map

Cluster Tree

Centroid Chart

Centroid Table

Cluster	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...	attribut...
Cluster 0	1.584	0.810	0.587	0.067	-0.042	0.142	0.284	0.214	-0.167	-0.526	-0.398	1.556	0.017	-0.004	-0.459	-0.635	-0.579
Cluster 1	-0.257	-0.101	-0.308	-0.007	-0.202	-0.608	-0.715	-0.659	-0.617	-0.641	-0.796	-0.187	-0.947	-1.034	-1.010	-0.903	-0.858
Cluster 2	0.089	0.158	0.265	0.276	0.514	0.769	0.836	0.865	0.913	0.957	0.920	0.110	0.814	0.773	0.711	0.452	0.212
Cluster 3	-0.461	-0.359	-0.323	-0.293	-0.427	-0.585	-0.654	-0.683	-0.635	-0.563	-0.507	-0.499	-0.467	-0.388	-0.200	0.071	0.278

d) Un Scatter Plot o gráfica de dispersión es una visualización que muestra la relación entre dos atributos en un gráfico de dos dimensiones. En el contexto del análisis de clúster, el Scatter Plot se utiliza para visualizar la distribución de los datos según los dos atributos más importantes obtenidos del proceso de clustering. Cada punto en el gráfico representa un dato, y su posición está determinada por los valores de los dos atributos seleccionados. Esto permite identificar patrones y tendencias en la distribución de los datos y cómo se agrupan en función de sus valores en esos dos atributos clave.



En resumen, en RapidMiner, estas herramientas de visualización son valiosas para comprender y analizar los resultados del análisis de clustering, lo que permite identificar patrones, características distintivas de los clústeres y cómo se distribuyen los datos en el espacio de atributos. Estas representaciones gráficas facilitan la toma de decisiones informadas y ayudan a obtener una comprensión más profunda de los datos.

Referencias

- Anáhuac Online. (2019). *Algoritmo clave* [Archivo de video]. [Contenido creado para Anáhuac Online].