

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

**Actividad: Reto de aprendizaje 22.
Evaluación de Clúster**

Nombre: Roberto Mora Balderas

Asesor: José Carlos Soto Monterrubio

Fecha: 02 de Agosto 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:

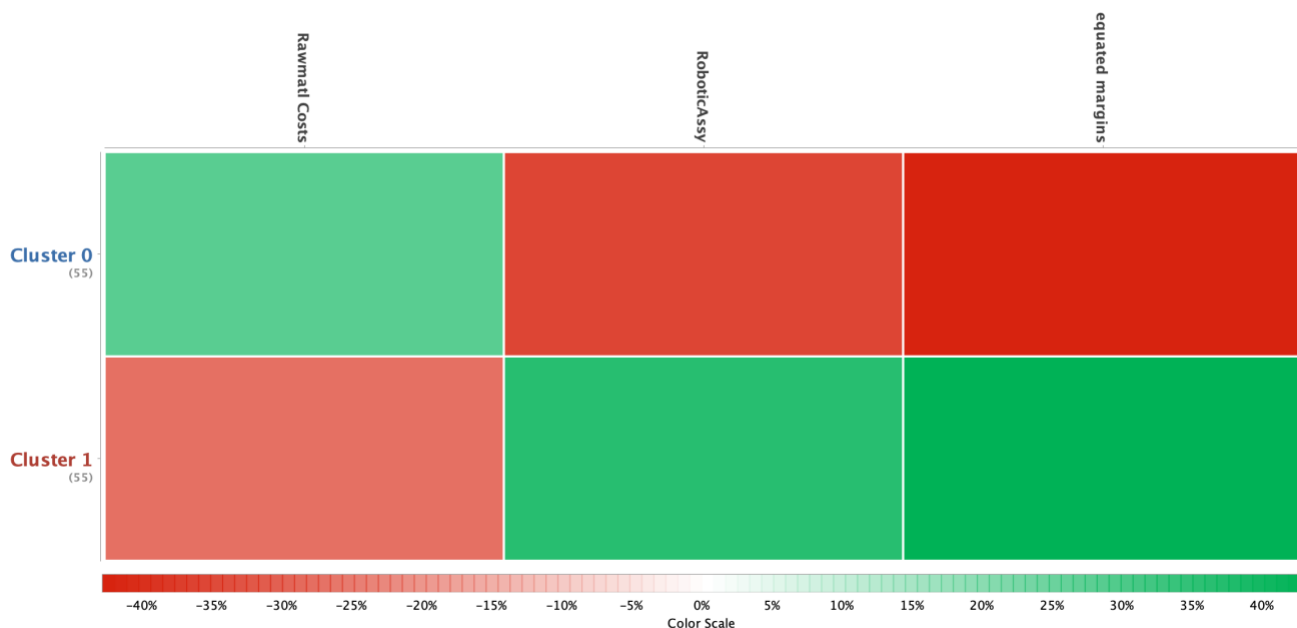
1. Desarrolla la misma evaluación del tema anterior con este resultado. En un documento en el que expliques detalladamente cada uno de los siguientes elementos. Toma de base el video del último tema donde explico cada rubro.

- 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:

- a) A continuacion, se muestra el heat map obtenido.

x-Means – Heat Map

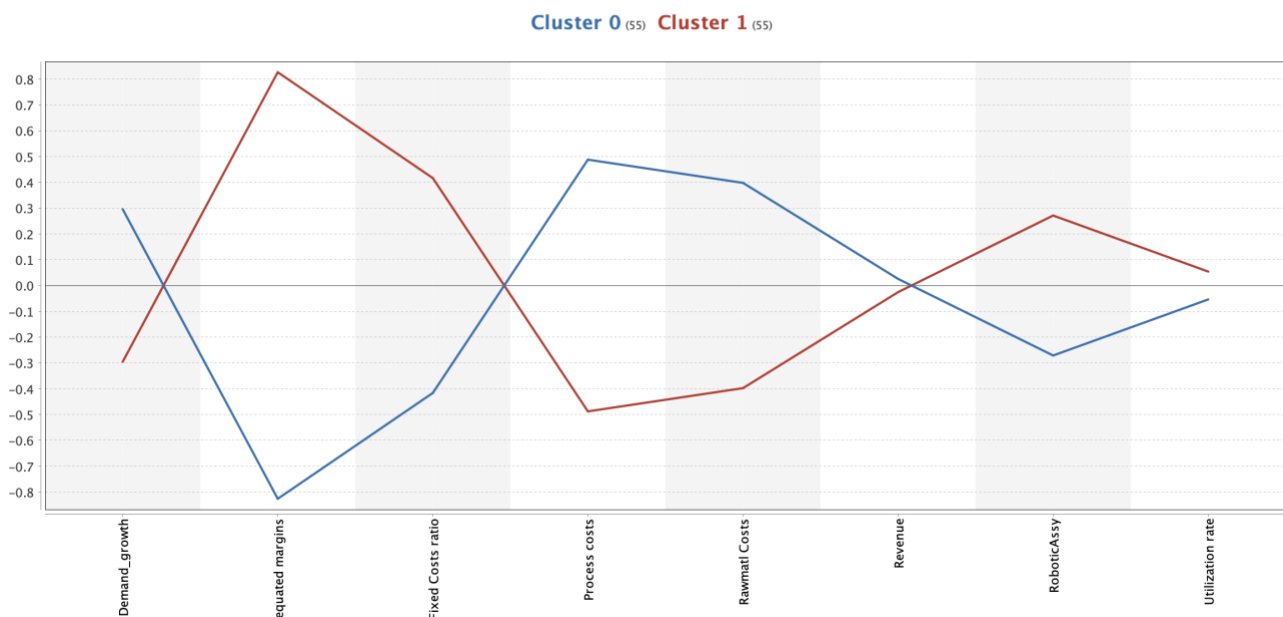


Un heatmap (mapa de calor) se utiliza para visualizar la distribución y similitud de los datos entre los diferentes clústeres. Un heat map es una representación gráfica en forma de matriz de colores que muestra la relación entre las filas y las columnas de los datos.

Como podemos observar, la diferencia de los clusters es muy notoria, siendo que para nuestro primer cluster la relación de las variables Rawmatl Costs, RoboticAssy, Equated Margins, va disminuyendo respectivamente, mientras que para el segundo cluster es totalmente lo contrario. Esto quiere decir que mientras una variable afecta directamente a 1 cluster el otro no se ve tan afectado, haciendo que estas variables definan prácticamente la pertenencia del registro a algún cluster.

b) A continuación, se muestra el gráfico de centroides.

κ-Means – Centroid Chart



Como se menciona en el punto anterior, podemos observar como la diferencia de los clusters es muy notoria, en este caso se aprecia como son inversas en su totalidad, siendo que la posición de los centroides para cada variable tienen valores de “-y”/”y”.

c) A continuación, se muestra la tabla de centroides.

x-Means - Centroid Table

Cluster	Demand_gro...	equated margi...	Fixed Costs ra...	Process costs	Rawmatl Costs	Revenue	RoboticAssy	Utilization rate
Cluster 0	0.296	-0.826	-0.416	0.488	0.398	0.026	-0.271	-0.054
Cluster 1	-0.296	0.826	0.416	-0.488	-0.398	-0.026	0.271	0.054

Aquí se aprecia lo comentado previamente, pero en forma numérica, la caracterización de los clusters puede interpretarse de la siguiente manera:

Cluster 0:

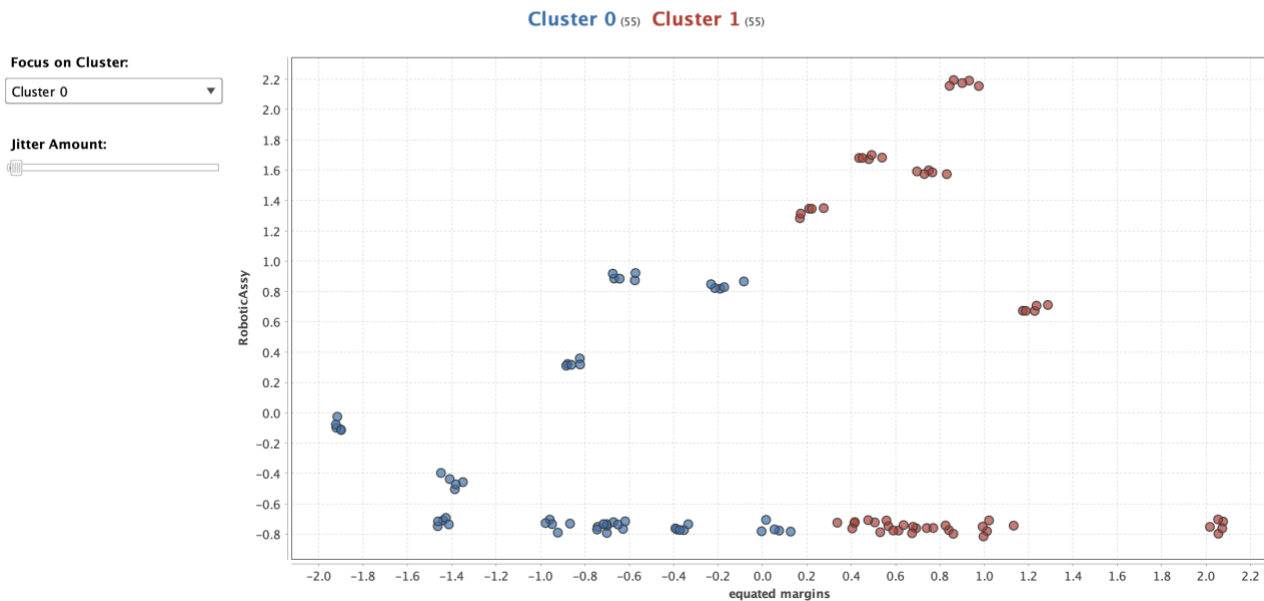
- Demand_growth: Tiene un valor positivo, lo que indica que este cluster tiende a tener un crecimiento de la demanda.
- Equated margins: Tiene un valor negativo, lo que sugiere que este cluster tiende a tener márgenes más bajos.
- Fixed costs ratio: Tiene un valor negativo, lo que indica que este cluster tiende a tener una relación más baja entre los costos fijos y los costos totales.
- Process costs: Tiene un valor positivo, lo que sugiere que este cluster tiende a tener costos de proceso más altos.
- Rawmatl Costs: Tiene un valor positivo, lo que indica que este cluster tiende a tener mayores costos de materia prima.
- Revenue: Tiene un valor positivo, lo que sugiere que este cluster tiende a generar mayores ingresos.
- RoboticAssy: Tiene un valor negativo, lo que indica que este cluster tiende a utilizar menos ensamblaje robótico.
- Utilization rate: Tiene un valor negativo, lo que sugiere que este cluster tiende a tener una tasa de utilización más baja.

Cluster 1:

- Demand_growth: Tiene un valor negativo, lo que indica que este cluster tiende a tener un decrecimiento de la demanda.
- Equated margins: Tiene un valor positivo, lo que sugiere que este cluster tiende a tener márgenes más altos.
- Fixed costs ratio: Tiene un valor positivo, lo que indica que este cluster tiende a tener una relación más alta entre los costos fijos y los costos totales.
- Process costs: Tiene un valor negativo, lo que sugiere que este cluster tiende a tener costos de proceso más bajos.
- Rawmatl Costs: Tiene un valor negativo, lo que indica que este cluster tiende a tener menores costos de materia prima.
- Revenue: Tiene un valor negativo, lo que sugiere que este cluster tiende a generar menores ingresos.
- RoboticAssy: Tiene un valor positivo, lo que indica que este cluster tiende a utilizar más ensamblaje robótico.
- Utilization rate: Tiene un valor positivo, lo que sugiere que este cluster tiende a tener una tasa de utilización más alta.

d) Se presenta el scatter plot.

x-Means – Scatter Plot



Aquí podemos apreciar la relación entre las dos variables más importantes para el modelo, y podemos ver como afectan a la determinación del cluster. Siendo que la variable equated margins, genera la distinción de manera mas marcada en los clusters.

Referencias:

- Anáhuac Online. (2019). *Problema de negocio* [Contenido creado para Anáhuac Online].
- Vohra, G. (2018). *Cluster Analysis For Business*. <https://bit.ly/3pICgmq>
- Whittaker, C. (2019). *7 Innovative Uses of Clustering Algorithms in the Real World*. <https://bit.ly/2KVXP0u>
- Pandey, P. (2020). Evaluating Clustering Methods. <https://tinyurl.com/42termbz>
- Rapid Miner. (7 de agosto de 2018). *20 Auto Model - Clustering & Outliers* [Archivo de video]. YouTube. <https://bit.ly/3dRmUcp>