

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

**Actividad: Reto de aprendizaje 23. Integración del proyecto**

**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.-** Genera un solo documento a partir de todas las actividades desarrolladas en esta unidad. Tu documento final debe contener:

1. Introducción a la organización y el negocio estudiado
2. Análisis del Planteamiento del problema
3. Definición de preguntas clave
4. Análisis de la información disponible
5. Justificación de elección de método de clúster
6. Resultados
7. Evaluación de resultados

**Título: Uso del análisis de clúster para abordar la variabilidad en los costos de insumos en una empresa de manufactura de refacciones automotrices.**

### **Introducción:**

El presente documento tiene como objetivo argumentar el uso del análisis de clúster para abordar el problema de la alta variabilidad en los costos de insumos en una empresa de manufactura que produce 22 partes diferentes para la industria automotriz. La alta variación en los costos de los insumos ha generado dificultades en la gestión de abastecimiento y rentabilidad, por lo que es fundamental encontrar una solución que permita identificar patrones y agrupar los insumos con comportamientos similares. El análisis de clúster se presenta como una herramienta adecuada para comprender la estructura y relación entre los datos, permitiendo así la toma de decisiones informadas y la predicción de costos más precisa.

### **Análisis del Planteamiento del problema:**

La gerencia general enfrenta el problema de la alta variación en los costos de insumos utilizados en la producción de las 22 partes de la industria automotriz. Esta variabilidad errática afecta la rentabilidad, la planificación financiera y la gestión de abastecimiento de manera significativa. Para resolver este problema, se requiere una solución inteligente que permita comprender los factores que influyen en los cambios de costos y agrupar los insumos con comportamientos similares para tomar decisiones informadas y prever costos de manera más efectiva.

### **Definición de preguntas clave:**

- a) ¿Cómo predecir de manera más precisa y confiable los costos de los insumos para la producción de cada una de las 22 partes?
- b) ¿Qué factores o variables están influyendo en las fluctuaciones de los costos de los insumos?
- c) ¿Es posible agrupar las partes en categorías que compartan patrones de comportamiento de costos de insumos similares?

## Análisis de la información disponible:

Para nuestro análisis contamos con un conjunto de registros de diferentes características para cada producto. El data set que tenemos cuenta con distintas variables, la mayoría de los datos son del tipo numérico, salvo la primera columna que es el id del producto el cual posee un valor alfanumérico.

A continuación, se proporcionan los significados de cada una de las variables involucradas en el problema:

- a) Product ID: Es un identificador único o código que se asigna a cada producto en el sistema de la empresa. Se utiliza para distinguir y rastrear individualmente los diferentes productos que se fabrican o comercializan.
- b) Fixed Costs ratio: Es la proporción o porcentaje de costos fijos en relación con los costos totales de producción u operación de un producto o proceso. Los costos fijos son aquellos que no varían con la cantidad producida o vendida, como alquiler, salarios fijos, entre otros.
- c) Equated Margins: También conocido como "margen equitativo" o "margen igualado", se refiere a igualar o ajustar los márgenes de beneficio entre diferentes productos o servicios para obtener una distribución más uniforme de las ganancias.
- d) Process Costs: Son los costos asociados con el proceso de fabricación o producción de un producto. Esto incluye el costo de materias primas, mano de obra directa, costos de maquinaria y otros costos directos relacionados con el proceso de producción.
- e) Utilization Rate: Es la proporción del tiempo o capacidad de producción utilizada en relación con la capacidad total disponible. Por lo general, se expresa en porcentaje y representa la eficiencia con la que se está utilizando la capacidad de producción.
- f) Demand\_growth: Es el crecimiento de la demanda o la tasa de cambio en la cantidad requerida de un producto o servicio durante un período específico. Indica cómo la demanda está aumentando o disminuyendo con el tiempo.
- g) Revenue: Es el ingreso total generado por la venta de productos o servicios durante un período específico. Representa la cantidad de dinero que ingresa a la empresa por sus ventas.

- h) **RoboticAssy:** Es un término que sugiere la presencia o uso de sistemas robóticos o automatizados en el ensamblaje o fabricación de productos. Indica que el proceso de ensamblaje puede estar parcial o totalmente automatizado mediante el uso de robots o maquinaria automatizada.
- i) **Rawmatl Costs:** Son los costos asociados con la adquisición o compra de materias primas utilizadas en el proceso de producción. Estos costos representan el gasto en los insumos básicos necesarios para fabricar un producto o proporcionar un servicio.

### **Justificación de elección de método de clúster:**

#### **1. Uso del análisis de clúster para resolver el problema:**

El análisis de clúster es una técnica de aprendizaje no supervisado que tiene como objetivo agrupar datos similares en conjuntos llamados clústeres. En el caso de la empresa de manufactura, aplicar el análisis de clúster a los datos de costos de insumos permitirá identificar grupos de insumos que comparten patrones de comportamiento en sus costos. Al agrupar los insumos en clústeres, se puede obtener una visión más clara de los comportamientos erráticos de los costos y de las relaciones entre los diferentes insumos.

#### **2. Susceptibilidad de la información para el análisis de clúster:**

Los datos proporcionados sobre los costos de insumos de las 22 partes son susceptibles de ser analizados con el análisis de clúster debido a la naturaleza de los datos y el problema en cuestión. Los datos de costos representan múltiples variables que pueden presentar patrones subyacentes y relaciones no lineales. Al aplicar el análisis de clúster, podemos encontrar agrupaciones de insumos con características y comportamientos similares en términos de sus costos, lo que proporciona una visión más profunda de la variabilidad y posibles factores influyentes.

## Resultados y Evaluación:

Aplicando el método x-means al data set, se obtuvieron los siguientes resultados:

### x-Means - Summary

Number of Clusters: 2

#### Cluster 0

55

**equated margins** is on average **42.77%** smaller, **RoboticAssy** is on average **36.18%** smaller, **Rawmatl Costs** is on average **27.87%** larger

#### Cluster 1

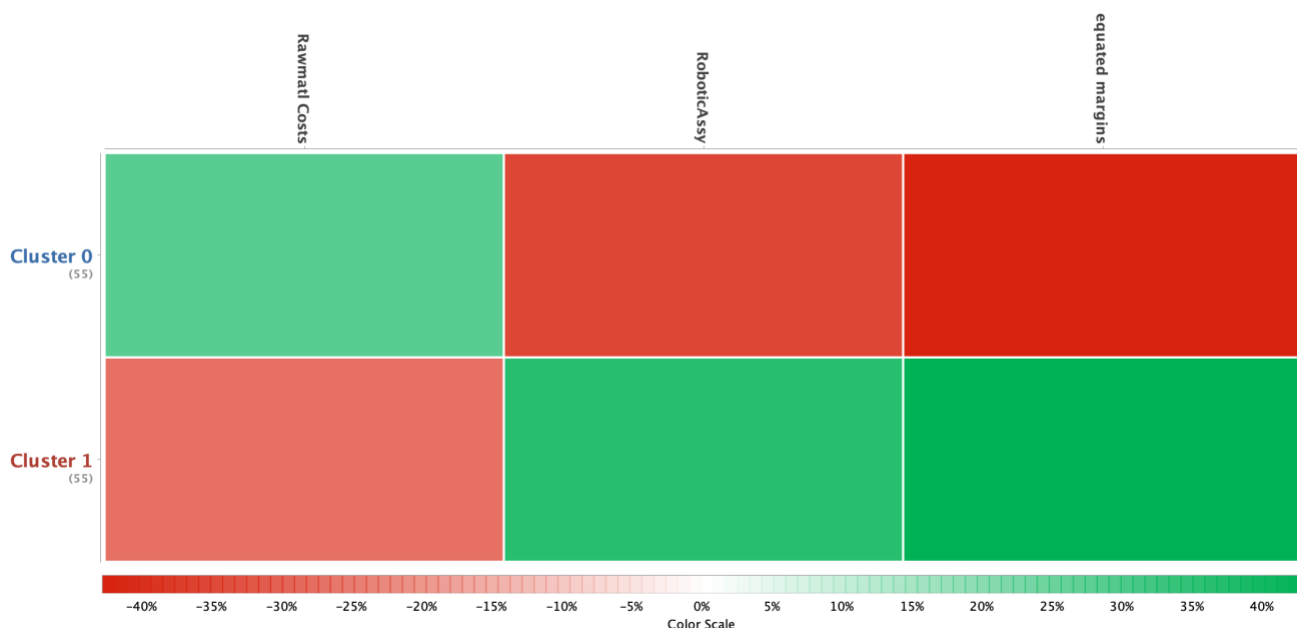
55

**equated margins** is on average **42.77%** larger, **RoboticAssy** is on average **36.18%** larger, **Rawmatl Costs** is on average **27.87%** smaller

Se obtuvieron dos clusters, que separan exactamente a la mitad los datos proporcionados y de manera totalmente inversa. Esto no es un resultado típico pero habla bastante de los datos empleados, profundizaremos al respecto mas adelante.

Otro factor importante a considerar son las correlaciones entre las variables y para esto podemos utilizar la visualización llamada heatmap.

### x-Means - Heat Map

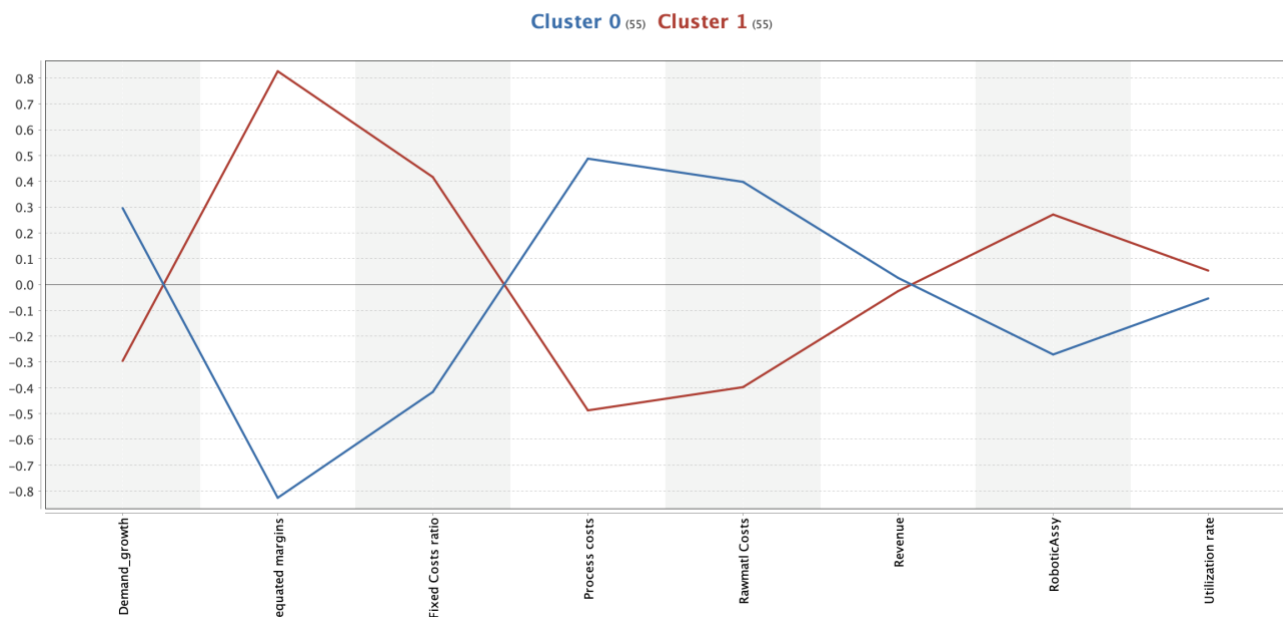


Un heatmap (mapa de calor) se utiliza para visualizar la distribución y similitud de los datos entre los diferentes clústeres. Un heat map es una representación gráfica en forma de matriz de colores que muestra la relación entre las filas y las columnas de los datos.

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 algun cluster.

Siguiendo con nuestro análisis, se mostrará el gráfico de centroides para poder visualizar la distribución espacial de los centroides del cluster con respecto a las distintas variables.

#### 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”.

Otra forma de analizar los clusters es mediante la tabla de centroides, la cual es la fuente de datos para el gráfico anterior.

### 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, y yendo mas a profundidad 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.

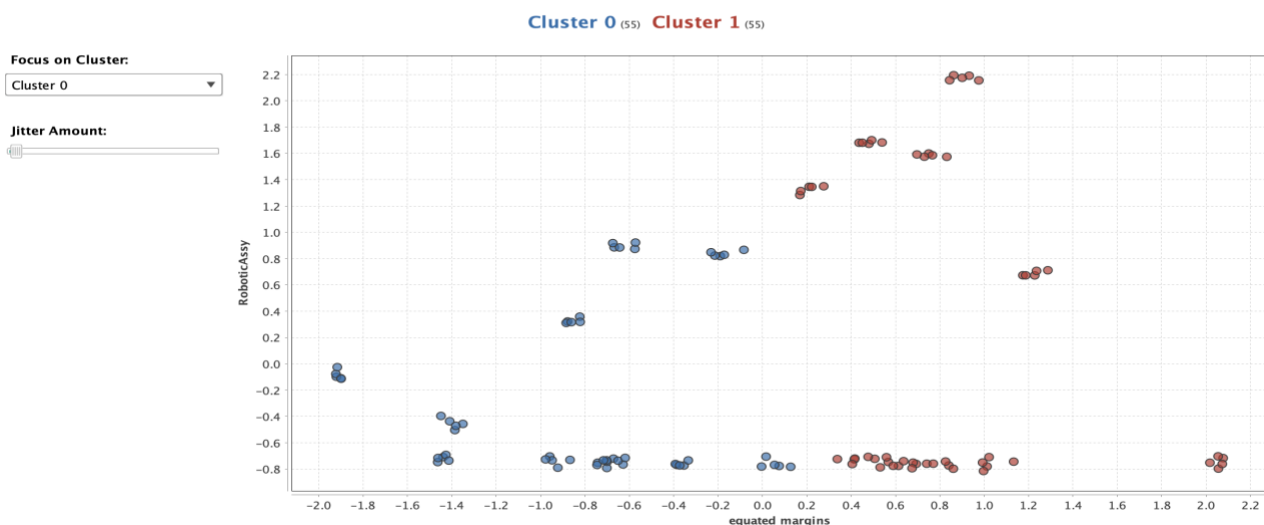


## Cluster 1:

- 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.

Finalmente como última pieza de análisis podemos ver un scatter plot.

### x-Means – Scatter Plot



El scatter plot (gráfico de dispersión) es una herramienta visual ampliamente utilizada en el análisis de clustering con el propósito de visualizar y comprender la distribución y agrupación de los datos. 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.

### Conclusión:

El análisis de clúster es una herramienta poderosa para abordar el problema de la alta variabilidad en los costos de insumos en la empresa de manufactura de refacciones automotrices. Al agrupar los insumos con comportamientos similares, se puede obtener una visión más clara de las tendencias y patrones en los costos, lo que facilita la toma de decisiones informadas y la predicción de costos más precisa. El análisis de clúster brinda una solución inteligente para comprender la estructura de los datos y encontrar una segmentación efectiva que permita mejorar la gestión de costos y la rentabilidad de la empresa. Finalmente, como pudimos observar los datos empleados arrojaron la separación perfecta en dos clusters, los cuales eran completamente “idénticos” de manera inversa, esto puede ocurrir por la procedencia y la finalidad de los datos proporcionados, ya que poseen un fin meramente académico.

### Referencias:

- Anáhuac Online. (2019). *Problema de negocio* [Contenido creado para Anáhuac Online].
- Vohra, G. (2018). *Cluster Analysis For Business*. <https://bit.ly/3plCgmq>
- 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>
- Anáhuac Online. (2019). *Problema de negocio* [Contenido creado para Anáhuac Online].