

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

**Actividad: Reto de aprendizaje 21.
Documento sobre aplicación
(método)**

Nombre: Roberto Mora Balderas

Asesor: José Carlos Soto Monterrubio

Fecha: 01 de Agosto de 2023

Objetivo:

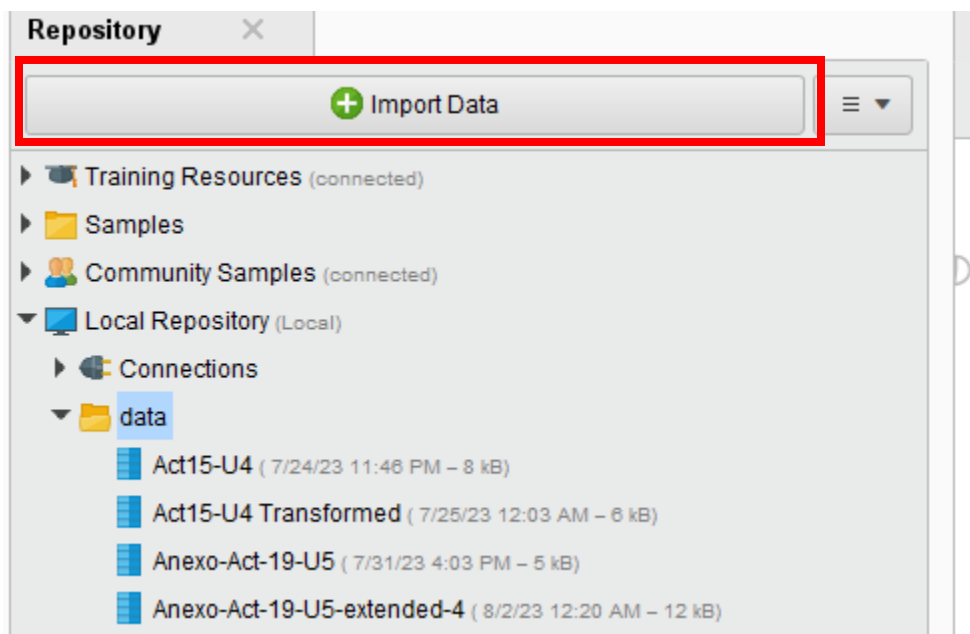
Aplicar el análisis de clúster con la información presentada con la finalidad de realizar un análisis que optimice la modelación de costos.

Instrucciones:

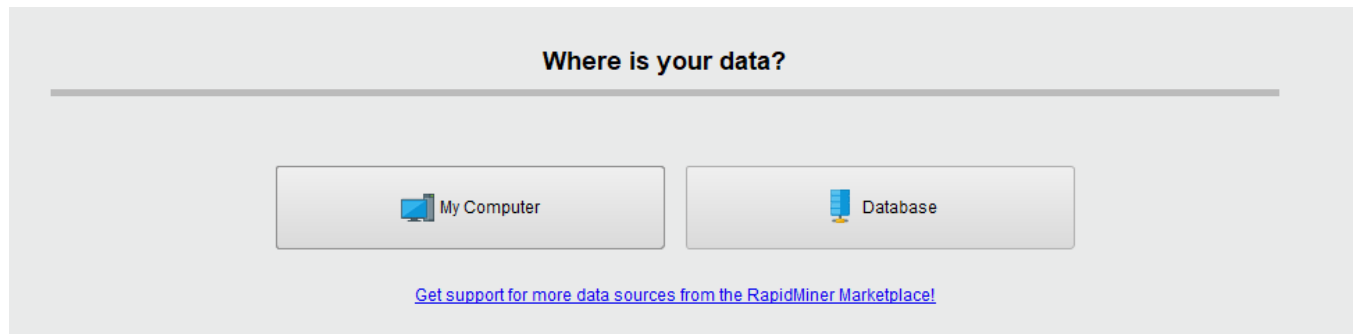
1. Aplica el análisis de clúster X-means con los datos proporcionados, y genera un documento explicando paso por paso las etapas para aplicarlo. Incluye los siguientes aspectos:
2. Importación de data.
3. Aplicación en Auto Model.
4. Presentación del número de clúster y la tabla de centroides.
5. Análisis preliminar de los resultados con respecto a los clústeres generados. ¿Qué caracteriza a cada clúster?

Desarrollo:

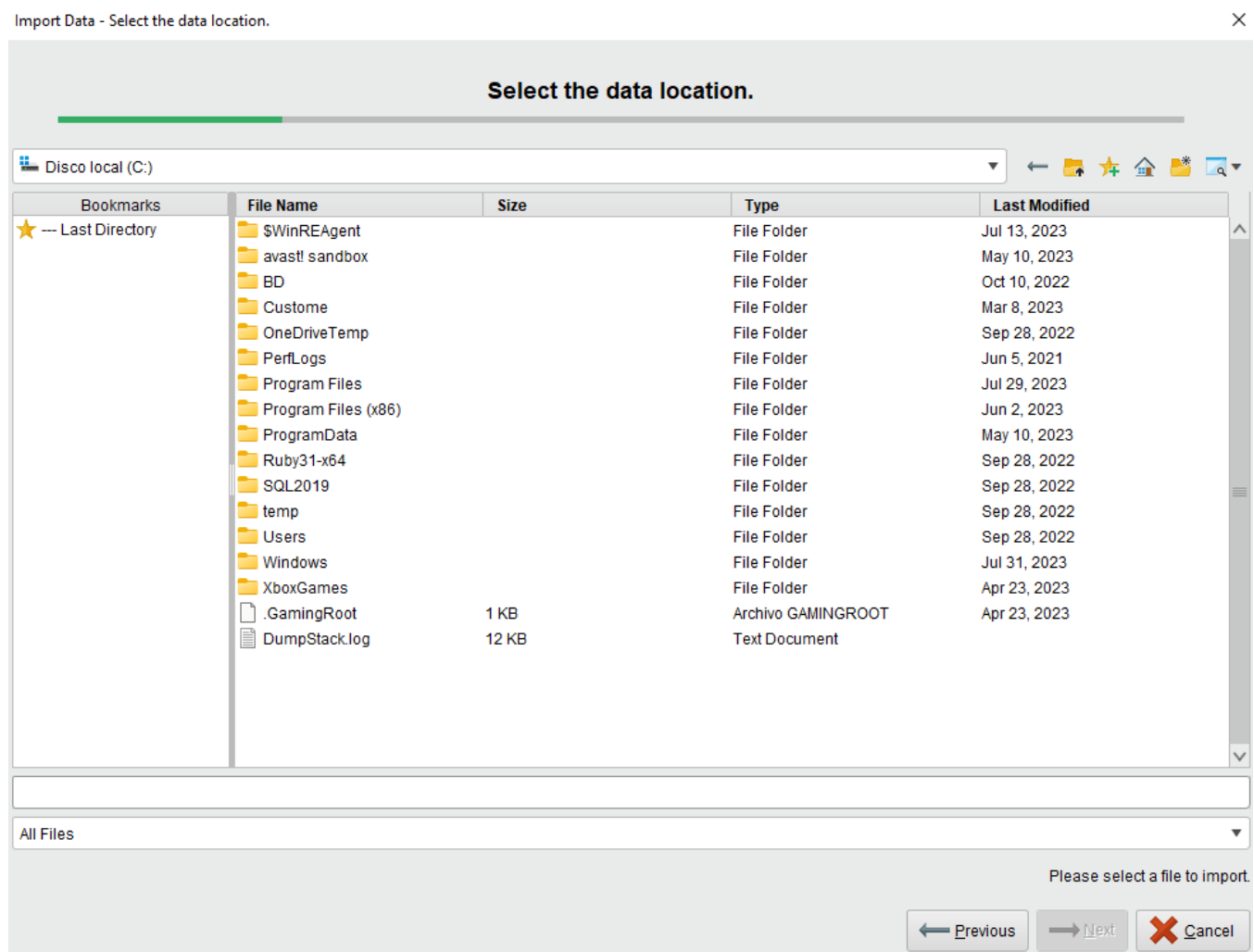
1. Importar los datos, se inicia dando clic en el icono de importación de datos.



2. Elegir de donde se importará el data set.



3. Elegir la ruta donde se cargará el archivo.



4. Elegir opciones con respecto a la importación y dar siguiente.

Import Data - Select the cells to import.

×

Select the cells to import.

Sheet: Sheet1 Cell range: A:I Select All ☒ Define header row: 1

| | A | B | C | D | E | F | G | H | I |
|----|--------------|------------------|----------------|---------------|------------------|---------------|---------|-------------|---------------|
| 1 | Product ID | Fixed Costs r... | equated mar... | Process costs | Utilization rate | Demand_gro... | Revenue | RoboticAssy | Rawmatl Costs |
| 2 | Firpn14280 | -0.293 | -0.685 | -0.417 | 0.578 | -0.526 | 0.046 | -0.75 | -0.854 |
| 3 | Firpn12457 | -1.215 | -0.194 | 0.821 | -0.207 | -0.334 | -1.078 | 0.84 | 0.813 |
| 4 | Firpn31567 | 1.712 | 2.078 | -1.34 | 0.892 | 0.051 | 0.084 | -0.75 | -0.08 |
| 5 | Firpn411052 | -0.51 | 0.207 | -0.004 | 0.219 | -0.943 | -0.702 | 1.326 | -0.724 |
| 6 | Firpn517563 | 2.037 | -0.863 | 0.578 | 1.295 | -0.719 | -1.581 | 0.317 | 1.693 |
| 7 | Firpn612256 | 1.116 | 1.232 | -1.388 | -0.678 | -1.745 | 0.623 | 0.689 | 0.249 |
| 8 | HdlInr113157 | 0.574 | 0.652 | 0.166 | -2.381 | -0.334 | -0.358 | -0.75 | 0.988 |
| 9 | HdlInr214464 | -0.076 | -0.685 | 1.865 | -0.005 | 0.019 | 1.174 | -0.75 | -1.427 |
| 10 | HdlInr319242 | 1.224 | 1.009 | -0.004 | -0.767 | 1.27 | -0.143 | -0.75 | -0.433 |
| 11 | HdlInr410525 | 0.032 | 0.741 | 0.7 | 0.892 | -0.173 | -0.693 | 1.59 | -0.863 |
| 12 | HdlInr514212 | -1.973 | -1.442 | 0.117 | 1.228 | 1.045 | 2.402 | -0.75 | -0.602 |
| 13 | HdlInr615432 | 0.086 | 0.073 | 0.238 | -1.126 | 0.147 | -0.777 | -0.75 | 1.428 |
| 14 | Stfm119999 | 0.195 | 0.875 | 0.748 | 0.735 | 1.013 | -0.489 | 2.183 | -1.035 |
| 15 | Stfm217039 | -0.131 | 0.563 | -1.752 | 1.609 | -0.59 | 0.214 | -0.75 | -0.926 |
| 16 | Stfm318825 | -0.835 | -1.398 | -0.102 | -1.171 | -1.071 | -0.689 | -0.476 | 0.535 |
| 17 | Stfm41689 | 0.249 | -0.373 | 2.035 | 0.219 | 1.911 | 1.994 | -0.75 | -0.868 |
| 18 | Stfm415723 | -1.919 | -1.932 | -0.781 | -1.103 | 1.847 | -0.901 | -0.077 | 1.47 |
| 19 | Stfm518979 | -0.347 | 0.83 | -0.441 | 0.062 | -0.173 | 0.345 | -0.75 | 0.009 |
| 20 | Stfm513647 | 0.249 | 0.429 | -1.558 | 0.667 | -1.713 | 1.294 | -0.75 | -0.839 |

← Previous
Next →
✖ Cancel

5. Elegir la dirección del repositorio donde se guardará el data set.

Import Data - Where to store the data?

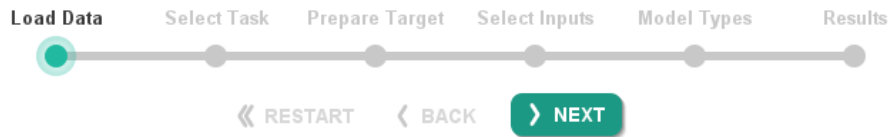
×

Where to store the data?

Local Repository (Local)

- data
 - Act15-U4 (7/24/23 11:46 PM - 8 kB)
 - Act15-U4 Transformed (7/25/23 12:03 AM - 6 kB)
 - Anexo-Act-19-U5 (7/31/23 4:03 PM - 5 kB)
 - Anexo-Act-19-U5-extended-4 (8/2/23 12:20 AM - 12 kB)
- processes
 - A12-U3 (7/21/23 5:55 PM - 168 kB)

6. Irse a la pestaña automodel y elegir la data con la que se desea trabajar.



Select Data for a New Model

Training Resources (connected)

- Samples
- Community Samples (connected)
- Local Repository (Local)
 - Connections
 - data
 - Act15-U4 (7/24/23 11:46 PM – 8 kB)
 - Act15-U4 Transformed (7/25/23 12:03 AM – 6 kB)
 - Anexo-Act-19-U5 (7/31/23 4:03 PM – 5 kB)
 - Anexo-Act-19-U5-extended-4 (8/2/23 12:20 AM – 12 kB)**
 - processes
 - A12-U3 (7/21/23 5:55 PM – 168 kB)
 - DB (Legacy)

Information

Name: **Anexo-Act-19-U5-extended-4**
Number of rows: **110**
Number of columns: **9**

Attributes / Columns

Product ID, Fixed Costs ratio, equated margins, Process costs, Utilization rate, Demand_growth, Revenue, RoboticAssy, Rawmatt Costs

7. Elegir modelo a utilizar, en este caso utilizaremos el modelo de clustering para aplicar X-means.

Auto Model

Load Data **Select Task** Prepare Target Select Inputs Model Types Results

RESTART BACK NEXT

Predict
Want to predict the values of a column?

Clusters
Want to identify groups in your data?

Outliers
Want to detect outliers in your data?

8. Seleccionar la variable utilizar en el modelo, como recomendación se utilizan solo las que posean un estatus de color verde, en nuestro caso eso no incluye el ProductID. El estatus nos brinda información de la relevancia y de la correlación entre las variables y el modelo.

Auto Model

Load Data

Select Task

Prepare Target

Select Inputs

Model Types

Results

RESTART

BACK

NEXT

Selected: 8 / Total: 9

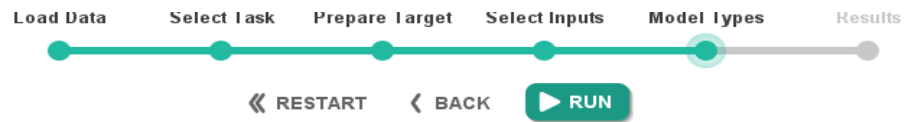
Deselect Red

Select All

Deselect All

| Selected | Status ↑ | Quality | Name | Correlation | ID-ness | Stability | Missing | Text-ness |
|-------------------------------------|--------------------------------------|--|-------------------|-------------|---------|-----------|---------|-----------|
| <input type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | ProductID | ? | 100.00% | 0.91% | 0.00% | 37.45% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | Fixed Costs ratio | ? | ? | 9.09% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | equated margins | ? | ? | 9.09% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | Process costs | ? | ? | 9.09% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | Utilization rate | ? | ? | 9.09% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | Demand_growth | ? | ? | 9.09% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | Revenue | ? | ? | 4.55% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | RoboticAssy | ? | ? | 54.55% | 0.00% | 0.00% |
| <input checked="" type="checkbox"/> | ● | <div><div></div><div></div><div></div><div></div><div></div></div> | Rawmatl Costs | ? | ? | 4.55% | 0.00% | 0.00% |

9. Elegir el modelo, en este caso x-means y correr el modelo.



Models

☐ k-Means Clustering

Number of Clusters:

☒ x-Means Clustering

Maximal Number of Clusters:

Data Preparation

☐ Remove Columns with Too Many Values

Maximum Number of Values:

☐ Extract Date Information

☐ Extract Text Information

Select Text Columns (0)...

Number of Extracted Features:

☐ Automatic Feature Selection

Additional Time (in Minutes):

Final Feature Set should be

Column Analysis

☒ Correlations between Columns

10. Generado el modelo podemos observar como nos arroja dos clusters como modelo optimo.

Model Summary

x-Means - Summary

Number of Clusters: 2

Cluster 0

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

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

11. 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 |

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.

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>