

Proceso de Clustering basado en Distancias:

Método: Finalmente, decidimos agrupar los departamentos manualmente utilizando la información de las distancias con respecto al departamento "pets". Este método se basó en la observación de saltos significativos en las distancias, lo que sugirió agrupaciones naturales en los datos.

1) Recolección y Organización de Datos

Recolección de Datos: Se obtuvo una matriz de distancias que muestra la proximidad entre diferentes departamentos, incluida la distancia con el departamento "pets".

Organización en Excel: Los datos fueron ingresados en una hoja de cálculo de Excel, donde cada celda representa la distancia entre dos departamentos. La columna correspondiente al departamento "pets" se identificó para su análisis.

2) Ordenación de la Columna "Pets"

Seleccionar la Columna: La columna que contiene las distancias entre cada departamento y "pets" fue seleccionada.

Ordenar de Menor a Mayor: Utilizando la función de ordenación de Excel, la columna fue ordenada de menor a mayor para identificar qué departamentos están más cercanos al departamento "pets". Esto facilitó la visualización de las relaciones entre departamentos basadas en sus distancias.

3) Identificación de Saltos Significativos

Análisis de Distancias: Se examinaron las distancias ordenadas para encontrar cambios abruptos, conocidos como "saltos". Estos saltos indican una separación natural entre grupos de departamentos.

Saltos Identificados:

- Entre "dairy eggs" (375) y "beverages" (739): Salto de 364
- Entre "beverages" (739) y "frozen" (1065): Salto de 326
- Entre "frozen" (1065) y "snacks" (1096): Salto de 31
- Entre "snacks" (1096) y "produce" (1402): Salto de 306
- Entre "produce" (1402) y "pantry" (1872): Salto de 470

Interpretación de Saltos: Se observó que los saltos significativos indican dónde los departamentos están más o menos relacionados con "pets", sugiriendo la formación de clusters.

4) Agrupamiento de Departamentos (Clustering)

Definición de Clusters: Basándonos en los saltos identificados, se decidió formar tres clusters:

- Cluster 1: Departamentos muy relacionados con "pets" (distancias pequeñas).
- Cluster 2: Departamentos moderadamente relacionados con "pets" (distancias intermedias).
- Cluster 3: Departamentos menos relacionados con "pets" (distancias grandes).

Asignación de Departamentos a Clusters:

Cluster 1:

- dairy eggs (375)
- beverages (739)
- frozen (1065)
- snacks (1096)

Cluster 2:

- produce (1402)
- pantry (1872)
- household (1913)
- bakery (2292)
- canned goods (2315)
- dry goods pasta (2556)
- deli (2590)
- personal care (2639)
- meat seafood (2740)
- breakfast (2769)

Cluster 3:

- babies (3040)
- international (3083)
- alcohol (3094)
- missing (3245)
- other (3262)
- bulk (3294)

Este agrupamiento se basó en la observación de saltos significativos en las distancias ordenadas, lo que sugiere agrupaciones naturales en los datos.