

ESTUDIO DE LA AGRUPACIÓN SOCIOECONÓMICA DE LOS MUNICIPIOS DEL ÁREA METROPOLITANA DE VALENCIA

MINERÍA DE DATOS EN NEGOCIOS

Ágatha del Olmo Tirado | 2ºBIA | 28/10/2023



VNIVERSITAT
DE VALÈNCIA

INTELIGENCIA Y ANALÍTICA DE NEGOCIOS

ÍNDICE

1. Introducción	1-2
2. El proceso de clustering en Weka	2-3
2.1. Tareas previas	2
2.2. Configuración de las agrupaciones	2
2.3. Visualización y atributos de clustering	3
3. El proceso de clustering en Rstudio	3-10
3.1. Tareas previas	3
3.2. Configuración de la agrupación	3-4
3.3. Elementos de cada agrupación	4-6
3.4. Análisis de medias	6
3.5. Análisis ANOVA	6-8
3.6. Análisis múltiple	8-10
4. Interpretación de los grupos	10-13
4. Conclusiones	13-14

INTRODUCCIÓN

Este informe presenta un análisis detallado de la agrupación de municipios en el Área Metropolitana de Valencia, basado en datos socioeconómicos recopilados en el archivo "amValencia1.csv". El objetivo de este estudio es utilizar técnicas de clustering para agrupar los municipios en grupos homogéneos pero heterogéneos entre sí, lo que nos permitirá comprender mejor las similitudes y diferencias entre las diversas regiones valencianas.

Hemos empleado técnicas de "farthest first" y clustering jerárquico, incluyendo encadenamiento simple y encadenamiento completo, así como el método de Ward, para obtener cuatro agrupaciones tanto en RStudio como en Weka.

Hemos realizado un análisis detallado de las variables socioeconómicas que podrían estar influyendo en la formación de grupos, y hemos explorado las diferencias en las medias de las variables entre los grupos. Tras identificar las medias más diferenciadoras a primera vista, hemos realizado un análisis de la varianza (ANOVA) para identificar diferencias significativas entre los grupos en las variables más

relevantes, y un análisis múltiple de estas para poder interpretar más eficazmente los resultados.

Los resultados y las conclusiones derivados de este análisis ofrecen una comprensión más profunda de la diversidad dentro de la región y pueden ser de utilidad para la toma de decisiones y la planificación estratégica de la zona.

EL PROCESO DE CLUSTERING EN WEKA

Empezamos usando Weka, entorno gráfico que nos permite crear y analizar experimentos sobre tareas de clustering, y en concreto su interfaz Explorer. Aquí realizamos las agrupaciones con los métodos FarthestFirst, Jerárquico Ward y Jerárquico Simple.

TAREAS PREVIAS

El punto de partida es el archivo "amValencia1.csv", que contiene información socioeconómica de los 59 municipios en el área metropolitana valenciana. Al trabajar con este archivo en formato CSV, es necesario realizar ciertos ajustes iniciales. Para ello, abrimos la ventana de diálogo "Invoke Options Dialog" y cambiamos el separador de campo de "," a ";" para asegurarnos de que los datos se importen correctamente.

Uno de los primeros pasos iniciales importantes es la eliminación de atributos innecesarios. En este caso, eliminamos los atributos de "Cercanía" e "Implantación" utilizando la opción "Remove". Además, identificamos una fila vacía adicional en el archivo CSV, lo que resultaba en un total de 60 instancias en lugar de las 59 esperadas. Esta fila adicional se eliminó utilizando la opción "Edit", haciendo clic derecho sobre ella y seleccionando "Delete Selected Instance", garantizando así que tengamos 59 instancias.

Y, por último, dado que estamos buscando agrupar los municipios y no clasificarlos, configuramos la clase en "No Class" para que no se incluya un atributo de clase.

CONFIGURACIÓN DE LAS AGRUPACIONES

Como ya hemos mencionado antes, las agrupaciones que realizamos en Weka fueron FarthestFirst, y el jerárquico Ward y simple, y la forma de agrupar de cada clustering es la siguiente:

- FarthestFirst: busca crear grupos tomando como punto de partida el punto más alejado del centroide del conjunto de datos. Este enfoque asegura que los grupos estén lo más separados posible entre sí, es decir, que sean lo más heterogéneos posible.

- Jerárquico Ward: se centra en minimizar la varianza dentro de los grupos, es decir, hacerlos lo más homogéneos en sentido intragrupal. Comienza considerando cada punto de datos como un grupo individual y luego fusiona gradualmente los grupos más cercanos entre sí, de manera que se minimice la variabilidad en los grupos.

-Jerárquico Simple: se basa en la idea de fusionar gradualmente los grupos eligiendo los dos grupos más cercanos entre sí en función de la distancia euclídea entre sus elementos más cercanos.

VISUALIZACIÓN Y ATRIBUTOS DE CLUSTERING

Una vez completado el proceso de clustering, es crucial entender la composición de los grupos resultantes. Para ello, utilizamos la opción "Visualize Cluster Assignments", lo que nos permite ver las instancias que conforman cada clúster.

Además, vamos creando atributos de clustering que reflejan la pertenencia de cada municipio a un grupo específico según el método usado. Esto facilita el análisis y la interpretación de los resultados.

En resumen, este proceso de clustering y las tareas previas realizadas en Weka nos proporcionan una base sólida para explorar y comprender la diversidad de los municipios en el área metropolitana, lo que nos permitirá tomar decisiones informadas y planificar estrategias futuras.

EL PROCESO DE CLUSTERING EN RSTUDIO

TAREAS PREVIAS

Después, realizamos la agrupación restante, la del jerárquico Complete con RStudio, entorno de lenguaje de programación dedicado a la computación estadística entre otras labores. [Link al script de Rstudio.](#)

Lo primero fue importar el archivo que guardamos desde Weka con las agrupaciones para ya tenerlas para su posterior estudio. Con este archivo importado podemos crear la matriz que llamamos Z sin incluir ni "muni" ni las agrupaciones en las variables, y usamos la variable "muni" para nombrar cada una de las observaciones y que así sea más fácil su comprensión.

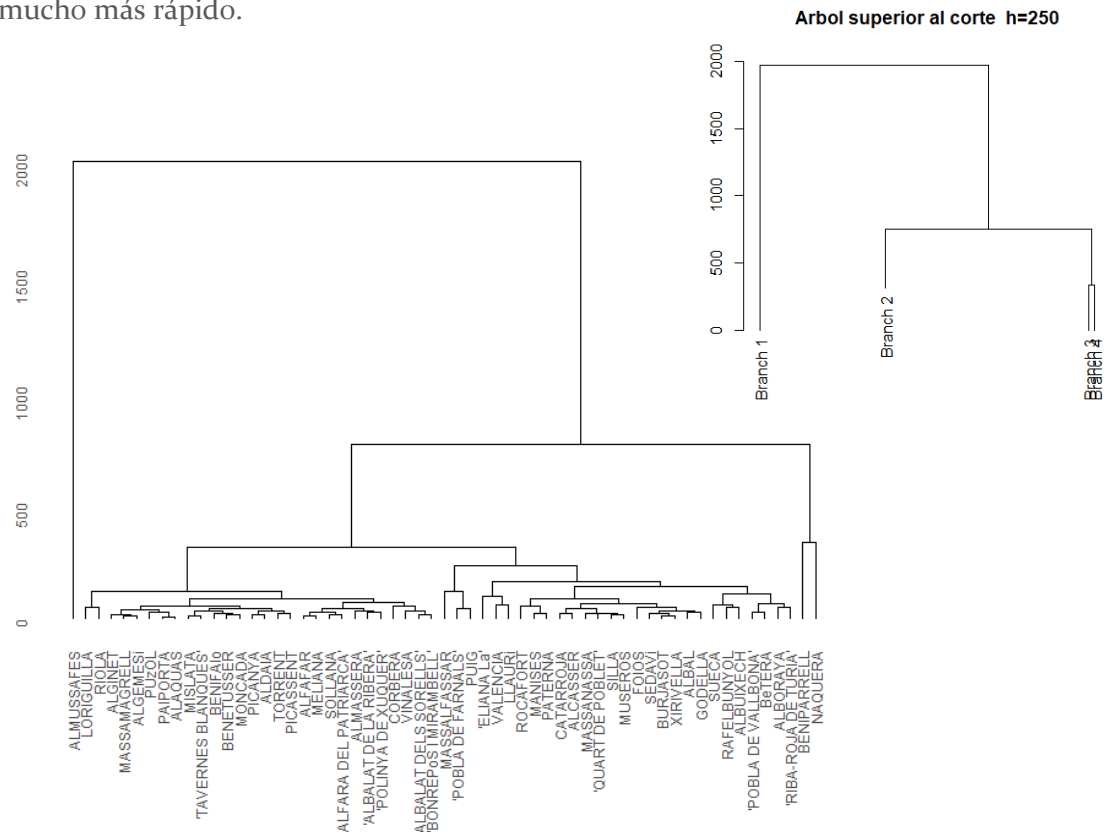
CONFIGURACIÓN DE LA AGRUPACIÓN

Ya montada la matriz Z podemos crear el cluster jerárquico complete, que consiste en lo siguiente:

-Jerárquico Complete: busca fusionar los grupos seleccionando los dos que tienen la mayor distancia entre sus elementos más lejanos.

Para hacerlo en R usamos la función `hclust(method="complete")`, y para visualizarlo usamos el `plot(cut(h=320))` y así podemos verlo con los 4 grupos formados.

Solo observándolo se puede ver cómo los grupos 3 y 4 son grupos individuales, y si vemos el árbol completo podemos detectar cuáles son estas ciudades mucho más rápido.



Aquí ya vemos que los pueblos “marginados” son Beniparrell y Náquera, y luego estudiaremos por qué están apartados de los demás.

Ya habiendo visualizado el cluster podemos añadir a nuestro archivo que ya contenía tres agrupaciones, una cuarta. Para esto, creamos el cluster con un `cutree(k=4)` y lo añadimos.

ELEMENTOS DE CADA AGRUPACIÓN

FARTHEST FIRST:

GRUPO 1: Mislata, Tavernes Blanques, Burjassot, Valencia, Alboraya, Pobl de Farnals, La Eliana, Massamagrell, Picanya, Almassera, Meliana, Torrent, Rocafort, Quart de Poblet, Manises, Paterna, Museros, Foios, Godella, Puig, Albalat dels Sorells, Rafelbunyol, Moncada, Pobl de Vallbona, Puzol, Betera, 'Riba-Roja de Turia, Bonrepos i Mirambell, Naquera, Vinalesa.

GRUPO 2: Benifaió, Algemesí, Sueca, Alginet, Sollana, Albalat de la Ribera, Albuixech, Alfara del Patriarca, Corbera, Llaurí, Loriguilla, Massalfassar, Polinya de Xuquer, Riola.

GRUPO 3: Silla, Aldaia, Almussafes, Picassent, Beniparrell.

GRUPO 4: Benetússer, Paiporta, Xirivella, Alfafar, Sedaví, Albal, Alaquàs, Massanassa, Catarroja, Alcàsser.

JERÁRQUICO COMPLETE:

GRUPO 1: Benetússer, Mislata, Tavernes Blanques, Paiporta, Burjassot, Xirivella, Valencia, Alfafar, Sedaví, Albal, Alboraya, Alaquàs, Massanassa, Pobla de Farnals, Eliana La, Massamagrell, Picanya, Catarroja, Silla, Almassera, Aldaia, Meliana, Torrent, Rocafort, Alcàsser, Quart de Poblet, Manises, Paterna, Museros, Foios, Godella, Puig, Albalat dels Sorells, Rafelbunyol, Benifaió, Moncada, Algemesí, Pobla de Vallbona, Puzol, Picassent, Sueca, Bétera, Riba-roja de Túria, Alginet, Sollana, Albalat de la Ribera, Albuixech, Alfara del Patriarca, Bonrepòs i Mirambell, Corbera, Llaurí, Loriguilla, Massalfassar, Polinyà de Xuquer, Riola, Vinalesa.

GRUPO 2: Almussafes

GRUPO 3: Beniparrell

GRUPO 4: Naquera.

JERÁRQUICO SIMPLE:

GRUPO 1: Benetússer, Mislata, Tavernes Blanques, Paiporta, Burjassot, Xirivella, Valencia, Alfafar, Sedaví, Albal, Alboraya, Alaquàs, Massanassa, Pobla de Farnals, Eliana La, Massamagrell, Picanya, Catarroja, Silla, Almassera, Aldaia, Meliana, Torrent, Rocafort, Alcàsser, Quart de Poblet, Manises, Paterna, Museros, Foios, Godella, Puig, Albalat dels Sorells, Rafelbunyol, Benifaió, Moncada, Algemesí, Pobla de Vallbona, Puzol, Picassent, Sueca, Bétera, Riba-roja de Túria, Alginet, Sollana, Albalat de la Ribera, Albuixech, Alfara del Patriarca, Bonrepòs i Mirambell, Corbera, Llaurí, Loriguilla, Massalfassar, Polinyà de Xuquer, Riola, Vinalesa.

GRUPO 2: Almussafes

GRUPO 3: Beniparrell

GRUPO 4: Naquera.

JERÁRQUICO WARD

GRUPO 1: Benetússer, Paiporta, Xirivella, Alfafar, Sedaví, Albal, Alaquàs, Massanassa, Picanya, Catarroja, Silla, Aldaia, Torrent, Alcàsser, Benifaió, Algemesí, Almussafes, Picassent, Alginet, Beniparrell.

GRUPO 2: Mislata, Tavernes Blanques, Burjassot, Valencia, Massamagrell, Almassera, Meliana, Quart de Poblet, Manises, Paterna, Museros, Foios, Albalat dels Sorells, Rafelbunyol, Moncada, Puçol, Albuixech, Alfara del Patriarca, Bonrepòs i Mirambell, Massalfassar, Vinalea.

GRUPO 3: Alboraya, Pobl de Farnals, Eliana La, Rocafort, Godella, Puig, Pobl de Vallbona, Bétera, Riba-roja de Túria, Náquera.

GRUPO 4: Sueca, Sollana, Albalat de la Ribera, Corbera, Llaurí, Loriguilla, Polinyà de Xuquer, Riola.

Como podemos ver, lo más característico es que el jerárquico complete y el jerárquico simple han acabado siendo exactamente iguales en sentido de elementos que componen sus grupos, a pesar de haber seguido un proceso distinto. Otra cosa a tener en cuenta en estas agrupaciones es la marginación de tres municipios en tres grupos, así que podemos tener en cuenta que al analizarlos estaremos hablando de unos grupos concretos y lo suficientemente diferentes o extraños como para que se hayan exiliado en diferentes grupos.

ANÁLISIS DE LAS MEDIAS

Para poder estudiar qué diferencia a cada grupo, miramos las medias y detectamos las que tengan una diferencia mayor entre grupos dentro de una misma agrupación. Para esto usamos la función `aggregate()` con el modelo previo al que llamamos Z y el modelo correspondiente y con `FUN=mean`. Al analizarlas, vemos que el Jerárquico Complete y el Single son exactamente iguales, así que los trataremos como tal. Y detectamos diferencias en las medias de las siguientes variables:

-FarthestFirst: Tmortalidad, Tcrecimiento, Tmigración, envejecimiento, nvehi, segvivienda, taf, iacomercial, iaindus y cenergindus.

-Jerárquico simple y complete: Tnatalidad, Tcrecimiento, Tmigracion, envejecimiento, nlineastelef, nvehi, segvivienda, vivdesocu, taf, iafinan, iaindus, cenerdomest, cenergindus, iparlabindustria e iparlabconstr.

.Jerárquico Ward: Tmortalidad, Tcrecimiento, Tmigración, envejecimiento, segvivienda, iacomercial, iaindus, cenerdomest, cenergindus, formacad, iparlabindustria e iparlabconstr.

ANÁLISIS ANOVA

El análisis de la varianza (ANOVA) es una técnica estadística utilizada para evaluar si existen diferencias significativas entre los grupos en función de una variable o factor. En concreto, en este caso, detectaremos diferencias significativas entre los grupos de municipios en Valencia en las variables en las que hemos encontrado aparentes diferencias entre las medias de los grupos con la función `aov(variable~cluster)`. Para hacerlo antes hemos convertido las variables de los cluster a factor con `as.factor()` porque en este caso se refiere a categorías discretas que representan distintos grupos, y no tiene un significado numérico continuo. Ahora, ANOVA lo tratará como niveles y

no como variables que tienen una interpretación numérica. Los resultados han sido, en resumen, los siguientes:

FARTHEST FIRST

- Tasa de mortalidad: se encontraron diferencias significativas a todos los niveles de significación.
- Tasa de crecimiento: se encontraron diferencias significativas a cualquier nivel de significación.
- Tasa de migración: no se encontraron diferencias significativas.
- Envejecimiento: se encontraron diferencias significativas entre los grupos a cualquier nivel de significación.
- Número de vehículos: se encontraron diferencias significativas entre los grupos a partir de un nivel de significación del 0.01.
- Segunda vivienda: no se encontraron pruebas suficientes para demostrar que había significantes diferencias de medias.
- Tasa de actividad femenina: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Índice de Actividad Comercial: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Índice de Actividad Industrial: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Consumo de Energía Eléctrica Industrial: se encontraron diferencias significativas a cualquier nivel de significación.

JERÁRQUICO SIMPLE/JERÁRQUICO COMPLETE

- Tasa de migración: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.05.
- Número de líneas de teléfono: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Segunda vivienda: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Viviendas desocupadas: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.01.
- Consumo de energía eléctrica doméstica: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.01.
- Consumo de energía eléctrica industrial: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Índice de participación laboral industrial: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.05.
- Índice de participación laboral en la construcción: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.05.

JERÁRQUICO WARD

- Tasa de mortalidad: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Tasa de crecimiento: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.

- Tasa de migración: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Envejecimiento: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Segunda vivienda: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Tasa de actividad femenina: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Índice de actividad comercial: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.01.
- Índice de actividad industrial: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Consumo de energía eléctrica doméstica: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Formación académica: se encontraron diferencias significativas entre los grupos a un nivel de significación del 0.01.
- Índice de participación laboral industrial: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.
- Índice de participación laboral de construcción: se encontraron diferencias significativas entre los grupos a un nivel de significación cualquiera.

ANÁLISIS MÚLTIPLE

Para acabar, realizamos el análisis de Tukey con la función `TukeyHSD(anova)` sobre las variables en las que hemos detectado diferencias significativas. Este análisis es útil cuando tienes más de dos grupos y deseas determinar cuáles de ellos son significativamente diferentes entre sí, cosa que con el ANOVA no podemos conseguir. De esta forma podremos entender qué diferencia a los grupos en cada agrupación realmente.

FARTHEST FIRST

- Tasa de mortalidad: Las pruebas de análisis múltiple revelaron que los grupos 2 y 4, 3 y 2 y 2 y 1 mostraron diferencias significativas entre sí, y en concreto, analizando la diff (si es negativa, el primer número es menor, y si es positiva, es mayor), el grupo 2 es significativamente mayor al 1, 3 y 4.
- Tasa de crecimiento: El análisis reveló que el grupo 2 tenía una tasa de crecimiento significativamente diferente a los otros grupos. Siendo 2 significativamente menor que el 1, 3 y 4.
- Envejecimiento: Las pruebas revelaron que el grupo 2 tenía un envejecimiento significativamente diferente a los otros grupos. Siendo el grupo 2 significativamente mayor que el 1, 3 y 4.
- Número de vehículos: Las pruebas indicaron que el grupo 3 tenía un número de vehículos censados significativamente diferente en comparación con los otros grupos. Siendo el 3 significativamente mayor que el 1, el 2 y el 4.
- Tasa de actividad femenina: Las pruebas post hoc revelaron que el grupo 2 tenía una tasa de actividad femenina significativamente diferente a los otros grupos. Siendo el 2 significativamente menor al 1, 3 y 4.

- Índice de Actividad Comercial: Las pruebas post hoc indicaron que el grupo 4 tenía un número de locales de actividad comercial significativamente diferente en comparación con los otros grupos. Siendo el grupo 4 significativamente mayor al 1, 2 y 3.
- Índice de Actividad Industrial: Las pruebas post hoc revelaron que el grupo 4 tenía un número de locales de actividad industrial significativamente diferentes en comparación con los otros grupos. Siendo el grupo 4 significativamente mayor al 1, 2 y 3.
- Consumo de Energía Eléctrica Industrial: Las pruebas post hoc indicaron que el grupo 3 tenía un consumo significativamente diferente en comparación con los otros grupos. Siendo el grupo 3 significativamente mayor al 1, 2 y 4.

JERÁRQUICO SIMPLE/JERÁRQUICO COMPLETE

- Tasa de migración: Las pruebas revelaron que el grupo 4 tenía una tasa de migración significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente mayor que el 1, 2 y 3.
- Número de líneas de teléfono: Las pruebas revelaron que el grupo 4 eran significativamente mayor que el 1, 2 y 3 (solo si escoge el nivel de significación del 0.05) y el grupo 3 era significativamente mayor al 1 y al 2.
- Segunda vivienda: Las pruebas revelaron que el grupo 4 tenía un número de segundas viviendas significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente mayor que el 1, 2 y 3.
- Viviendas desocupadas: Las pruebas revelaron que el grupo 4 tenía un número de viviendas desocupadas significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente menor que el 1, 2 y 3 (a un nivel de significación del 0.01).
- Consumo de energía eléctrica doméstica: Las pruebas revelaron que el grupo 4 tenía un consumo de energía doméstica significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente mayor que el 1, 2 y 3.
- Consumo de energía eléctrica industrial: Las pruebas revelaron que el grupo 1 era significativamente menor que el 2 y el 3, y el grupo 4 significativamente menor que el 2 y el 3.
- Índice de participación laboral industrial: Las pruebas revelaron que el grupo 4 era significativamente menor que el 2 y el 3.
- Índice de participación laboral en la construcción: Las pruebas revelaron que el grupo 4 tenía un índice de participación laboral en la construcción significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente mayor que el 1, 2 y 3.

JERÁRQUICO WARD

- Tasa de mortalidad: Las pruebas revelaron que el grupo 4 tenía una tasa de mortalidad significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente mayor que el 1, 2 y 3.

- Tasa de crecimiento: Las pruebas revelaron que el grupo 4 era significativamente menor que el 1 y el 3, y el grupo 2 significativamente menor que el 1 y el 3.
- Tasa de migración: Las pruebas revelaron que el grupo 3 tenía una tasa de migración significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente mayor que el 1, 2 y 4.
- Envejecimiento: Las pruebas revelaron que el grupo 4 era significativamente mayor que el 1, 2 y 3, el grupo 3 significativamente menor que el 1 y el 2, y el grupo 2 significativamente mayor al 1.
- Segunda vivienda: Las pruebas revelaron que el grupo 3 era significativamente mayor que el 1, 2 y 4, y el grupo 4 significativamente mayor que el 1 y el 2.
- Tasa de actividad femenina: Las pruebas revelaron que el grupo 4 tenía una tasa de actividad femenina significativamente diferente a los otros grupos. Siendo el grupo 4 significativamente menor que el 1, 2 y 3.
- Índice de actividad comercial: Las pruebas revelaron que el grupo 1 era significativamente mayor que el 3 y 4.
- Índice de actividad industrial: Las pruebas revelaron que el grupo 1 era significativamente mayor que el 3 y 4.
- Consumo de energía eléctrica doméstica: Las pruebas revelaron que el grupo 3 era significativamente mayor que el 1, 2 y 4.
- Formación académica: Las pruebas revelaron que el grupo 3 era significativamente mayor que el 1 y 2.
- Índice de participación laboral industrial: Las pruebas revelaron que el grupo 1 era significativamente mayor que el 2, 3 y 4, y el grupo 2 significativamente mayor que el 3 y 4.
- Índice de participación laboral de construcción: Las pruebas revelaron que el grupo 4 era significativamente mayor que el 1 y 2.

INTERPRETACIÓN DE LOS GRUPOS

En el último apartado de nuestro informe, centramos nuestra atención en la interpretación de las agrupaciones resultantes de nuestros análisis de clustering. Hasta este punto, hemos aplicado diversas técnicas estadísticas para agrupar los municipios de la región metropolitana de Valencia en conjuntos homogéneos. Sin embargo, agrupar los datos es solo el primer paso. Ahora, analizamos el significado y las características distintivas de cada uno de estos grupos. Para ello, utilizamos las comparaciones múltiples, específicamente la prueba de Tukey, que acabamos de proporcionar. A través de este análisis, desvelaremos las particularidades que caracterizan a cada grupo y, con ello, contribuiremos a una comprensión más profunda de la estructura socioeconómica y demográfica de la región.

FARTHEST FIRST:

GRUPO 1: El Grupo 1 se perfila como un conjunto de municipios con una dinámica demográfica más activa en comparación con el Grupo 2, menores índices de actividad industrial y comercial en contraste con el Grupo 4, y una mayor participación laboral

femenina que el 2. Estas características sugieren un perfil demográfico y económico específico que lo distingue de los demás grupos, pero con matices menos diferenciadores que en los otros.

GRUPO 2: Este grupo se caracteriza por tener la mayor tasa de mortalidad y envejecimiento, y la menor tasa de crecimiento, lo que nos lleva a pensar que son ciudades envejecidas y más vacías, esto explicaría por qué la tasa de actividad femenina es menor a todos los demás ya que en estas ciudades la actividad laboral suele estar más masculinizada mientras que las mujeres se quedan en casa por unos roles de género más tradicionales.

GRUPO 3: Este grupo se caracteriza por tener más consumo de energía industrial, aunque no el mayor índice de actividad industrial, y más vehículos censados que los demás, esto se puede dar porque su actividad industrial es muy potente en sentido de uso de energía eléctrica, y necesita el uso de mucho transporte para llevar a cabo su actividad.

GRUPO 4: Este grupo se caracteriza por tener más índice de actividad comercial y de actividad industrial, esto se puede deber a que son ciudades con una gran concentración de empresas y negocios en los sectores comercial e industrial. Esto podría ser resultado de su ubicación estratégica, infraestructura favorable, acceso a recursos o mercados específicos, políticas de desarrollo económico, o tradiciones comerciales e industriales arraigadas en la región.

JERÁRQUICO SIMPLE/COMPLETE:

GRUPO1: el Grupo 1 muestra un perfil que se caracteriza por una menor conectividad telefónica, la cual puede darse por no ser una zona muy comercial o industrial, un menor consumo de energía industria y menor migración, segundas viviendas, consumo de energía doméstica y construcción en comparación con el Grupo 4 a pesar de tener más viviendas desocupadas que este. Estas diferencias pueden indicar que estas zonas, así agrupadas, son poco interesantes en sentido del estudio ya que no arrojan una interpretación muy coherente.

GRUPO 2: el Grupo 2 se distingue por su mayor consumo de energía eléctrica industrial y un mayor índice de participación laboral en la industria, indicando una posible mayor actividad industrial o residencial. A pesar de esto, tiene menos líneas telefónicas que otros grupos, lo que sugiere una infraestructura de comunicación menos desarrollada. Además, el Grupo 2 tiene menos migración, menos segundas viviendas, un menor consumo de energía doméstica y una menor participación laboral en la construcción en comparación con el Grupo 4. Estas diferencias podrían reflejar una población más estable y menos actividad en la construcción en estas áreas, teniendo además un mayor número de viviendas desocupadas en comparación con el Grupo 4.

GRUPO 3: el Grupo 3 se distingue por tener más líneas telefónicas que el Grupo 1 y 2, aunque todavía menos que el Grupo 4, muestra un mayor consumo de energía industrial en comparación con los Grupos 1 y 4, y una mayor participación laboral en la industria en comparación con el Grupo 4, lo que podría indicar una mayor actividad industrial en estas áreas que necesita de líneas telefónicas. Por otro lado, este grupo presenta menos migración, menos segundas viviendas, un menor consumo de energía doméstica y una menor participación laboral en la construcción en comparación con el Grupo 4, pero presenta más viviendas desocupadas que este (aunque con menos significación que en los grupos 1 y 2).

GRUPO 4: el Grupo 4 es el más interesante de todos en términos de diferencias claras e interpretables, se caracteriza por tener un alto nivel de migración, lo que sugiere una mayor movilidad de la población en estas áreas. Además, este grupo cuenta con más líneas telefónicas y segundas viviendas en comparación con los Grupos 2 y 3, lo que podría indicar una mayor infraestructura de comunicaciones y una mayor presencia de propiedades de uso vacacional. Por otro lado, el Grupo 4 tiene menos viviendas desocupadas, un mayor consumo de energía doméstica pero un menor consumo de energía industrial en comparación con los Grupos 2 y 3. Esto podría reflejar un mayor uso de energía en los hogares, posiblemente debido a un mayor número de viviendas ocupadas, pero una menor actividad industrial en estas áreas. Además, el Grupo 4 tiene un índice de participación laboral industrial más bajo y un índice de participación en la construcción más alto en comparación con los otros grupos. Esto podría sugerir una menor dependencia de la industria y una mayor participación en el sector de la construcción en estas áreas por la construcción de viviendas vacacionales.

JERÁRQUICO WARD:

GRUPO 1: el Grupo 1 se caracteriza por tener un crecimiento poblacional superior en comparación con otros grupos, mostrando también un envejecimiento moderado. Presenta una menor presencia de segundas viviendas en la región y se destaca por un mayor índice de actividad comercial e industrial. Sin embargo, la formación académica es menor en este grupo en comparación con el Grupo 3. Además, el Grupo 1 tiene un mayor índice de participación laboral en el sector industrial, pero una menor participación en el sector de la construcción en comparación con el Grupo 4. Estas características sugieren un mayor desarrollo económico comercial e industrial, pero no en el sector constructor, y una población en crecimiento, pero con niveles educativos ligeramente inferiores en comparación con otros grupos.

GRUPO 2: Menos crecimiento que el 3 y 1, mas envejecimiento que el 1, pero menos que el 4 y el 3, menos segundas viviendas que el 4 y el 3, menos formación académica que el 3, mas índice de participación laboral industrial que el 3 y el 4, pero menos que el 1, y menos índice de participación laboral constructora que el 4.

GRUPO 3: el Grupo 2 se caracteriza por tener una tasa de crecimiento poblacional menor en comparación con los Grupos 1 y 3, y aunque presenta un envejecimiento poblacional, este es moderado y se encuentra en un rango intermedio entre los grupos

1 y 4. Además, muestra una menor presencia de segundas viviendas en la región y niveles de formación académica inferiores en comparación con el Grupo 3. Este grupo destaca por un índice de participación laboral en el sector industrial, aunque es superado por el Grupo 1 en este aspecto. Por otro lado, tiene una participación laboral en el sector de la construcción menor que el Grupo 4. Estas características sugieren un grupo con un crecimiento poblacional y un envejecimiento moderados y una mayor presencia en actividades industriales en comparación con otros sectores económicos.

GRUPO 3: el Grupo 3 se caracteriza por tener una tasa de crecimiento poblacional mayor en comparación con los Grupos 2 y 4, lo que indica un crecimiento demográfico más dinámico. Además, este grupo presenta una mayor tasa de migración, lo que sugiere un mayor flujo de población. Además, el envejecimiento poblacional en este grupo es más bajo que en los Grupos 1, 2 y 4, lo que refleja una población más joven. Este grupo también muestra una mayor presencia de segundas viviendas en la región, lo que podría indicar una mayor afluencia de turismo o zonas residenciales, lo cual se suma a que tiene un mayor consumo de energía doméstica y una mayor formación académica en comparación con los Grupos 1 y 2. En cambio, su índice de participación laboral en el sector industrial es menor que el de los Grupos 1 y 2, lo que sugiere una economía menos centrada en la industria. En resumen, el Grupo 3 parece estar experimentando un crecimiento demográfico sostenido, con una población más joven, una economía diversificada y zonas más residenciales.

GRUPO 4: el Grupo 4 se caracteriza por tener una tasa de mortalidad más alta, una tasa de crecimiento poblacional más baja en comparación con los Grupos 1 y 3, y un mayor envejecimiento, lo que sugiere un perfil demográfico menos dinámico y una población más envejecida. La presencia de segundas viviendas en este grupo es mayor que en los Grupos 1 y 2, lo que podría indicar una afluencia de turismo o un mayor número de zonas residenciales. La tasa de actividad femenina en este grupo es más baja en comparación con los Grupos 1 y 2, lo que podría reflejar una menor participación de las mujeres en la fuerza laboral por unos roles más tradicionales. A nivel económico, el Grupo 4 tiene un índice de actividad comercial e industrial menor que el Grupo 1, lo que sugiere una economía menos orientada hacia estos sectores. Sin embargo, su índice de participación laboral en el sector de la construcción es mayor que el de los Grupos 1 y 2. En resumen, el Grupo 4 parece estar experimentando una población envejecida con un crecimiento más lento, una economía menos industrializada y una mayor presencia de segundas viviendas.

CONCLUSIONES

Tras analizar las distintas agrupaciones generadas, hemos observado que presentan muchas diferencias entre sí. Sin embargo, nuestra elección sería la agrupación obtenida mediante el algoritmo Farthest First. Esta selección se justifica por la mayor claridad y distinción de los grupos, con excepción del Grupo 1, que es menos diferenciado. Por otro lado, las agrupaciones jerárquicas Complete y Single se centran demasiado en el Grupo 4 como el más distintivo y además tiene los tres municipios “extraños” separados en tres grupos distintos únicos, mientras que la agrupación Ward

genera conjuntos más complejos de explicar debido a las relaciones de diferencia entre grupos. En resumen, consideramos que la agrupación Farthest First presenta la interpretación más interesante y clara para nuestro estudio.