



UNIVERSIDAD DE GRANADA

Análisis relacional mediante segmentación

Teresa Cabrera del Moral

7 de diciembre de 2023

Inteligencia de Negocio
teresacdm@correo.ugr.es
Grupo 2

Índice

Índice	2
1. Introducción	3
2. Casos de Estudio	5
2.1. Caso de estudio 1: Los bandos del conflicto palestino-isarelí en España	5
2.1.1. Epígrafe Caso 1	5
2.1.2. Análisis Caso 1	6
2.1.3. Interpretación de la segmentación	11
2.2. Caso de estudio 2: Jóvenes y Tercera Edad, y sus respectivas amenazas	12
2.2.1. Epígrafe Caso 2	12
2.2.2. Análisis Caso 2	12
2.2.3. Interpretación de la segmentación	23
2.3. Caso 3: ¿Qué partido está perdiendo votos?	25
2.3.1. Epígrafe Caso 3	25
2.3.2. Análisis Caso 3	25
2.3.3. Interpretación de la segmentación	30
3. Bibliografía	31

1. Introducción

Esta práctica consiste en un estudio de técnicas de aprendizaje no supervisado para análisis relacional mediante segmentación. He aplicado distintos algoritmos de clustering sobre el conjunto de datos proporcionado por el profesor para realizar un estudio en profundidad a raíz de los resultados obtenidos y poder sacar conclusiones sobre las 2000 personas encuestadas.

La base de datos proviene de la empresa 40db, la cual elabora mensualmente una encuesta para el Grupo Prisa donde recoge la intención de voto de la población y otros asuntos de actualidad. Son un total de 2.000 respuestas y más de 50 variables que recogen datos demográficos, socioeconómicos, intención de voto, ideología y opinión sobre temas de actualidad, como el conflicto palestino-israelí, el cambio climático o la crisis inflacionaria. Existen algunas variables naturalmente numéricas como la edad (en años) o la posición ideológica en una escala 0 a 10, y otras que, al ser ordinales (como por ejemplo, el nivel de educación) se han pasado a numéricas para tratarlas con medidas de distancia tradicionales. Otras variables, sin embargo, son nominales sin orden (por ejemplo, la comunidad autónoma de residencia) y no nos sirven para aplicar clustering, pero las emplearemos para definir casos de uso y comparar resultados entre distintos grupos.

Realizaremos 3 casos de estudio a partir del conjunto de datos haciendo uso de los algoritmos de clustering siguientes:

- **K-means:** Comienza seleccionando aleatoriamente k centroides, como representantes iniciales de cada cluster. Luego, asigna cada punto al cluster cuyo centroide es el más cercano, basándose en la distancia euclidiana. Posteriormente, recalcula los centroides de los clusters tomando el promedio de los puntos asignados a cada uno. Este proceso de asignación y actualización se repite iterativamente hasta que los centroides convergen o hasta alcanzar un criterio de detención predefinido. Toma como parámetro el número de clusters por lo que, para averiguar el agrupamiento óptimo, deberemos probar con distintos valores y elegir el que mejor métricas nos devuelva.
- **Birch:** Emplea una estructura jerárquica basada en árboles, llamados Clustering Feature Trees (árboles CF), que representan los datos de manera compacta. Los datos se agrupan iterativamente mediante la construcción y fusión de los árboles, manteniendo centroides y estadísticas para realizar una partición progresiva de los datos. También toma como parámetro el número de clusters además de, el branching factor (máxima cantidad de subclusters que se pueden tener en un nodo del árbol), threshold (límite para la distancia entre puntos) y compute labels (lo ponemos a True para que el modelo asigne etiquetas de cluster a cada punto una vez que se haya realizado el ajuste).
- **MeanShift:** Funciona sin la necesidad de especificar previamente el número de clusters y se centra en encontrar regiones densas en el espacio de características. Identifica y mueve iterativamente centros hacia zonas con mayor acumulación de ejemplos. Lo hace mediante el cálculo del gradiente de la densidad de puntos. A medida que los centros convergen hacia las regiones de alta densidad, los puntos cercanos se asignan a los clusters asociados con estos centros, generando así los clusters finales.
- **DBSCAN:** Identifica clusters basados en la densidad de los puntos en un espacio de características. Utiliza dos parámetros clave: epsilon (eps) y minPts. Comienza seleccionando un punto aleatorio y encuentra todos los puntos vecinos dentro de una distancia eps. Si el número de puntos vecinos es igual o mayor a minPts, se considera un punto central y se forma un cluster al expandir iterativamente a través de sus vecinos.
- **Agglomerative Cluster:** Técnica jerárquica que comienza tratando cada punto de datos como un cluster individual y, en cada paso, fusiona los clusters más similares hasta que todos los puntos están en un solo cluster o en un número predeterminado de clusters. En cada iteración, el algoritmo une los clusters más cercanos basados en una medida de distancia, que en este caso he escogido 'ward', un método de enlace que minimiza la varianza intra-cluster cuando se combinan dos clusters.

Analizaremos los resultados obtenidos para cada uno de ellos comparando tiempos de ejecución, número de clusters y valor de métricas como las siguientes que nos ayudaran a evaluar qué tan bien están separados los clusters y cómo de homogéneos son:

- **Coefficiente Silhouette:** Toma valores entre -1 y +1, donde un alto valor indica que el objeto es bastante similar a su grupo y muy diferente a los de otros clusters.
- **Índice de Calinski-Harabasz:** Es un cociente entre la dispersión intracluster e intercluster. Un valor alto nos indica que los grupos son densos y están bien separados.
- **Índice de Davies-Bouldin:** Compara la distancia entre los clusters con el tamaño de los propios clusters. Un menor valor de este índice indica una mejor separación entre los clusters.

Para generar todas estas métricas utilizo el siguiente código:

```
metric_CH = metrics.calinski_harabasz_score(X, cluster_predict)
print("Calinski-Harabaz Index: {:.3f}, ".format(metric_CH), end='')

if len(X) > 10000:
    m_sil = 0.2
else:
    m_sil = 1.0
metric_SC = metrics.silhouette_score(X, cluster_predict, metric='euclidean',
                                     sample_size=floor(m_sil*len(X)), random_state=123456)
print("Silhouette Coefficient: {:.5f}".format(metric_SC))

metric_DB = metrics.davies_bouldin_score(X, cluster_predict)
print("Davies-Bouldin Score: {:.3f}, ".format(metric_DB), end='')
```

Tomaremos aquel algoritmo que nos cree clusters válidos con valores de métricas adecuados y, a partir de éste, crearemos gráficos que puedan resultar interesantes e interpretaremos la segmentación obtenida para cada caso.

Trataré de sacar conclusiones y deducir características de los clusters de la forma más veraz posible apoyándome en artículos que puedan corroborar estas observaciones.

2. Casos de Estudio

2.1. Caso de estudio 1: Los bandos del conflicto palestino-israelí en España

2.1.1. Epígrafe Caso 1

Desde que empezó el conflicto bélico entre Palestina e Israel el 7 de octubre de 2023, hemos oído a miembros políticos de los diferentes partidos dar su opinión al respecto e incluso posicionarse en tal guerra. Existen diversos artículos como [este](#) que describen la opinión de cada uno. Mientras que los partidos de derecha como PP y VOX declaran una mayor simpatía hacia el bando israelí, los partidos de izquierdas, Sumar, Podemos y PSOE abogan por el reconocimiento de Palestina como estado.



Figura 1: Irene Montero e Ione Belarra, miembros de Podemos, con unos pañuelos típicos Kufiya mostrando así su apoyo a Palestina.

Por tanto, aprovechando que se han hecho preguntas sobre este conflicto en la encuesta de 40db, he querido comprobar si los ciudadanos opinan igual que sus líderes políticos. Es decir, si, efectivamente, aquellos de izquierdas están más de acuerdo con afirmaciones que defienden a Palestina que los de derechas y viceversa.

He tratado de escoger los apartados de la pregunta 11 que más pueden significar un posicionamiento en el conflicto, según el nivel de acuerdo o desacuerdo. El clustering se efectuará a partir de las siguientes variables:

- **ideología:** en una escala del 0 al 10, en la que '0' representa la extrema izquierda y '10' representa la extrema derecha, los encuestados han escogido dónde se posicionan ellos según su ideología.
- **p11.1:** Israel tiene derecho a defenderse de los ataques del grupo terrorista Hamás de la forma en que lo está haciendo.
- **p11.2:** El uso de la violencia por parte de Hamás es consecuencia de las condiciones en las que se encuentran los palestinos en Gaza
- **p11.3:** Mientras exista Hamás, Palestina no debería tener un estado propio
- **p11.5:** En los territorios ocupados de Gaza y Cisjordania, Israel no respeta los derechos de los palestinos

```
subset=datos
subset=subset.rename(columns={
    "p7": "ideologia",
    "p11_1": "Israel dcho defenderse",
    "p11_2": "Violencia es consecuencia",
    "p11_3": "Con Hamas nada",
    "p11_5": "Israel no respeta dchos palestinos"
})
usadas = ['ideologia', 'Israel dcho defenderse', 'Violencia es consecuencia',
          'Con Hamas nada', 'Israel no respeta dchos palestinos']
```

Trabajo con el conjunto de datos completo por lo que serán un total de 2000 instancias. Las cuales normalizaremos con valores de 0 a 1.

```
def norm_to_zero_one(df):
    return (df - df.min()) * 1.0 / (df.max() - df.min())
```

Para aquellas preguntas que algunos encuestados han dejado sin respuesta, en este caso he decidido completarlas mediante el cálculo de la media.

```
for col in datos:
    datos[col].fillna(datos[col].mean(), inplace=True)
```

2.1.2. Análisis Caso 1

Estos son los resultados obtenidos:

Algoritmo	Nº clusters	Calinski-Harabasz	Silhouette	Davies-Bouldin	Tiempo
K-Means	2	731.845723	0.276711	1.381416	0.187004
Birch	2	627.896913	0.282851	1.409476	0.048998
MeanShift	3	380.952960	0.258020	1.337777	0.640467
DBSCAN	2	102.705827	0.198006	3.909752	0.056892
AggCluster	2	453.920597	0.283189	1.011681	0.063993

Cuadro 1: Métricas Caso 1

Elegir un algoritmo como 'el mejor' haciendo uso de distintas métricas puede ser difícil pues alguno puede destacar más en una y otro en otra. Como se puede apreciar en el [Cuadro 1](#) el algoritmo que mejor resultado nos da en cuanto a Silhouette y Davies-Bouldin es el algoritmo jerárquico Agglomerative Clustering, sin embargo, es K-means el que más destaca de todos en Calinski-Harabasz.

Analizaremos el comportamiento del Agglomerative Clustering cambiando su parámetro num_clusters y también estudiaremos si, cambiando el método de cálculo de distancia, hallamos números mejores.

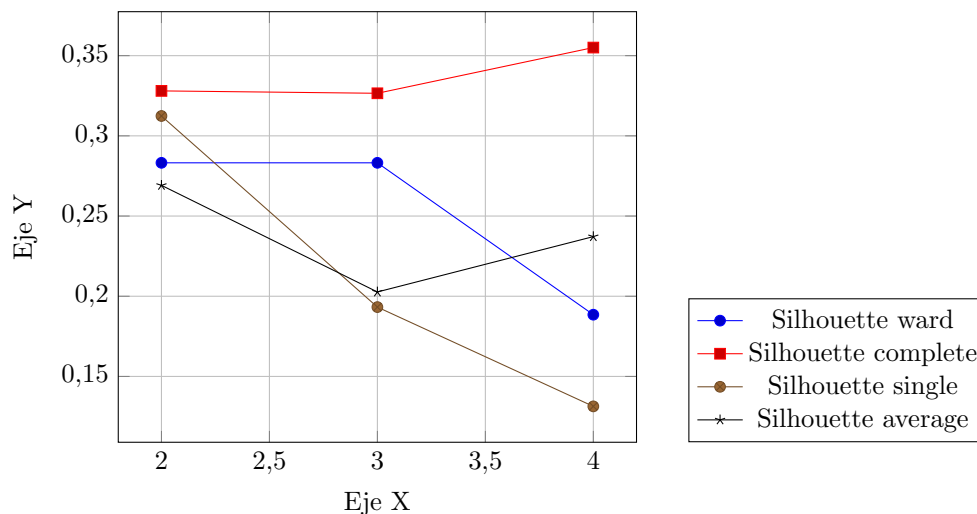
Nº clusters	linkage	Calinski-Harabasz	Silhouette	Davies-Bouldin	Tiempo
2	ward	453.920597	0.283189	1.011681	0.080898
2	complete	421.810154	0.217229	1.826229	0.081357
2	single	5.603179	0.312388	0.698452	0.024004
2	average	132.774128	0.269151	1.355089	0.078891
3	ward	454.226075	0.0283189	1.646894	0.063995
3	complete	313.490473	0.187	1.719197	0.079898
3	single	3.960828	0.193235	0.662889	0.039992
3	average	81.29498	0.202667	1.223897	0.172202
4	ward	413.835826	0.188491	1.634409	0.172843
4	complete	406.698185	0.195431	1.738797	0.188204
4	single	3.388808	0.131328	0.647535	0.047687
4	average	291.037539	0.237175	1.178241	0.20917

Cuadro 2: Métricas Caso 1: Agglomerative Clustering

De nuevo, se puede apreciar en el [Cuadro 2](#) que nos encontramos ante un caso en el que es algo difícil divisar la mejor elección de parámetros pues pueden destacar más o menos según la métrica. Observamos como el cálculo de distancia ward nos da mejores resultados en los tres casos de k en cuanto al índice Calinski-Harabasz. Mientras, es single el que mejores valores de Davies-Bouldin nos proporciona, no obstante, sus valores Calinski-Harabasz son mucho peores que el resto. El método de enlace 'single' conecta clústeres basándose en las distancias mínimas entre puntos de diferentes clústeres. De esta forma, tiende a fusionar clústeres de forma más 'alargada', lo que resulta en clústeres conectados en cadena. Esta estructura no es ideal para la métrica Calinski-Harabasz, que favorece clústeres densos y bien separados. En general, comparando todos los valores, se puede ver que eligiendo 2 clusters y el método ward es cómo obtenemos los resultados globalmente mejores.

Para comenzar a analizar las características de nuestros clusters podemos fijarnos en el Kplot de la [Figura 2](#) el cual nos sirve de utilidad para observar la distribución de datos para cada cluster por cada variable individual. El KPlot lo he creado con el siguiente método:

Gráfico de Líneas: Agglomerative Clustering



```
def KPlot(X, name, k, usadas):
    print("\nGenerando kplot...")
    n_var = len(usadas)
    fig, axes = plt.subplots(k, n_var, sharex='col', figsize=(20,20))
    fig.subplots_adjust(wspace=0.2)
    colors = sns.color_palette(palette=None, n_colors=k, desat=None)

    for i in range(k):
        dat_filt = X.loc[X['cluster']==i]
        for j in range(n_var):
            sns.kdeplot(dat_filt[usadas[j]], shade=True, color=colors[i], ax=axes[i,j])

    plt.savefig("kdeplot_case1_"+name+".pdf")
    plt.figure()
    plt.show()
    plt.clf()
```

En color azul tenemos el cluster 1 y en naranja el 2. Es cierto que el cluster 1 parece tener respuestas de todos los tipos en cada pregunta, siendo la variable 'ideología' donde más se aprecia que ha agrupado a población que se posiciona en el centro político. Sólo con esta gráfica podemos hacernos una idea de por qué el coeficiente silhouette de este cluster es más bajo (0,248), pues es menos homogéneo.

Por otro lado, el cluster 2, con un coeficiente silhouette de 0,5398, sí que tiene mayores picos en cada variable. Contiene en su mayoría ciudadanos de izquierdas que están más de acuerdo con que la violencia de Hamás es consecuencia de las condiciones en las que se encuentran los palestinos en Gaza y con que Israel no respeta los derechos de los palestinos en Gaza y Cisjordania. Por el contrario parece que muestran un nivel alto de desacuerdo con el derecho de Israel para defenderse de la forma en la que lo está haciendo y con que mientras exista Hamás, Palestina no debería tener un estado propio.

A partir de esta gráfica ya podemos empezar a deducir que el algoritmo Agglomerative Clustering ha encontrado un grupo claro en la base de datos que, podemos asumir por sus respuestas, muestra apoyo al bando palestino y, como hemos comentado anteriormente, comparte opinión con partidos de izquierdas como Sumar y Podemos que así lo han exteriorizado.

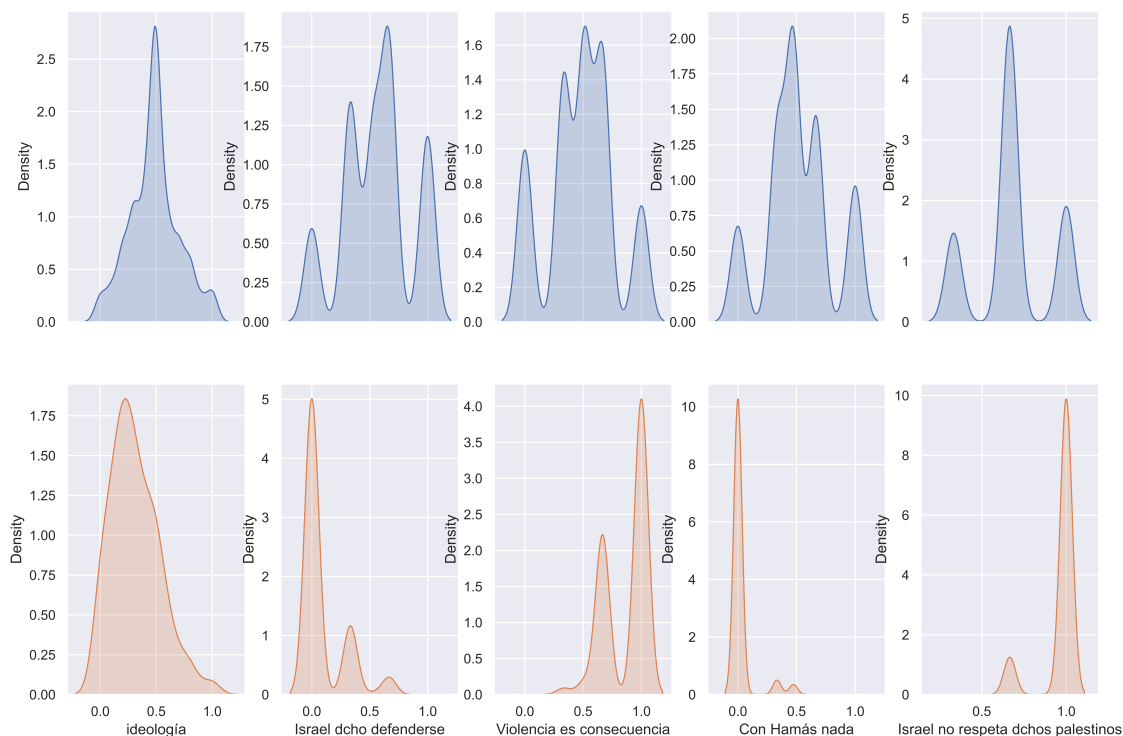


Figura 2: KPlot para el caso 1 con Agglomerative Clustering.

En la [Figura 3](#) vemos una Scatter Matrix creada a partir del Agglomerative Cluster. El código para esta gráfica es el siguiente:

```
def ScatterMatrix(X, name):
    print("\nGenerando scatter matrix...")
    plt.figure()
    sns.set()
    variables = list(X)
    variables.remove('cluster')
    sns_plot = sns.pairplot(X, vars=variables, hue="cluster", palette='Set2',
                           plot_kws={"s": 25}, diag_kind="hist")
    sns_plot.fig.subplots_adjust(wspace=.03, hspace=.03);

    plt.savefig("scatmatrix_case1_"+name+".pdf")
    plt.show()
    plt.clf()
```

En verde tenemos el cluster 1 y en naranja el 2. Esta representación nos permite comparar las variables de las preguntas sobre el conflicto bélico entre sí. Por ejemplo, podemos ver que, dentro del cluster 2, aquellos que opinan que la violencia de Hamás es consecuencia de las condiciones en las que se encuentran los palestinos, también son aquellos que rechazan la afirmación de que mientras Hamás exista, no puede haber estado palestino. Este cluster, como veníamos comentando, está comprendido de personas que tienden a estar de acuerdo con declaraciones que defienden al pueblo palestino y discrepan con aquellas que lo acusan.



Figura 3: Scatter Matrix para el caso 1 del algoritmo Agglomerative Clustering.

He creado un dendrograma, que se pueda observar en la [Figura 4](#) para ayudarnos a distinguir la agrupación de datos y las relaciones de similitud entre ellos. El dendrograma se ha formado a partir de esta función:

```
def Dendrograms(X, name, usadas):
    print("\nGenerando dendogramas...")
    if len(X)>1000:
        X = X.sample(1000, random_state=123456)
    X_filtrado_normal = preprocessing.normalize(X, norm='l2')
    linkage_array = hierarchy.ward(X_filtrado_normal)
    plt.figure(1)
    plt.clf()
    hierarchy.dendrogram(linkage_array, orientation='left', p=10, truncate_mode='lastp')
    X_filtrado_normal_DF = pd.DataFrame(X_filtrado_normal, index=X.index, columns=usadas)
    sns.clustermap(X_filtrado_normal_DF, method='ward', col_cluster=False,
        figsize=(20, 10), cmap="YlGnBu", yticklabels=False)
    plt.savefig("dendrogram_case1"+name+".pdf")
```

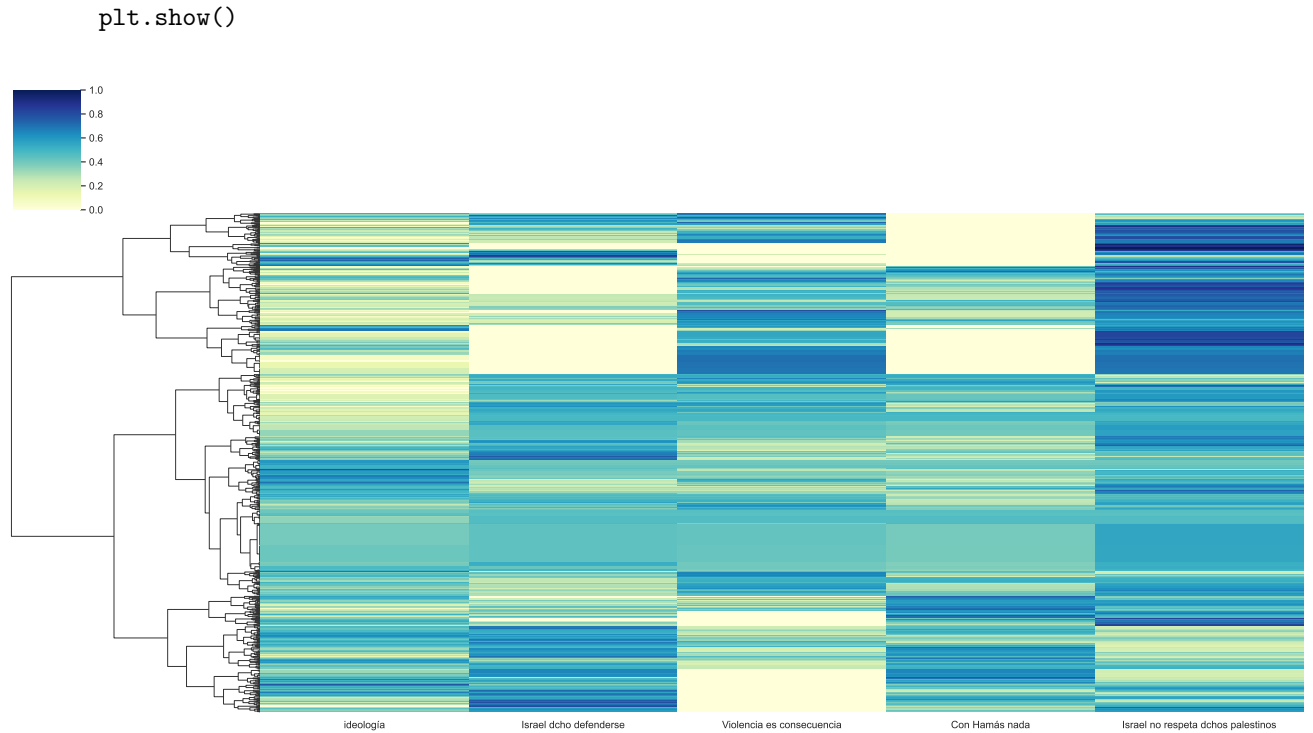


Figura 4: Dendrograma para el caso 1 del algoritmo Agglomerative Clustering.

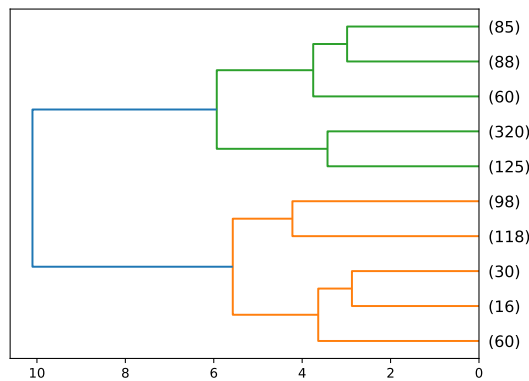


Figura 5: Dendrograma simplificado para el caso 1 del algoritmo Agglomerative Clustering.

La altura en el eje vertical de un dendrograma indica cuán diferentes o similares son dos grupos de datos cuando se combinan en un nuevo clúster. Los colores representan la similitud entre los diferentes elementos. Cuanto más similares sean los colores entre sí, más similares serán los elementos y cuantas más ramas haya al final del dendrograma, más diversos serán los elementos.

Observando con un vistazo rápido el dendrograma cabe destacar una franja de color turquesa (que corresponde aproximadamente a un 0,5) en el cluster inferior que nos dice que en esta agrupación un número notable de personas se posicionan en el centro de cada variable. No tienen una ideología muy radical y parecen no estar completamente de acuerdo o en desacuerdo con las afirmaciones sobre la guerra palestino-israelí.

Por otro lado, en el cluster superior, que correspondería al 2 del que hemos venido hablando, destacan grandes cuadros de color amarillo claro en las afirmaciones 'Israel tiene derecho a defenderse[...]' y 'Mien-

tras exista Hamás, Palestina no debería tener estado propio'. Este color indica que están totalmente en desacuerdo con estas afirmaciones. Mientras que con las otras dos, podemos encontrar más diferencias de opiniones.

También podemos apreciar con esta representación que, a pesar de la diversidad de opiniones, parece que, por lo general en ambos clusters, la población considera que en Gaza y Cisjordania, Israel no respeta los derechos de los palestinos. Aunque se aprecia que es el cluster 2 el que está más de acuerdo con dicha afirmación, como habíamos podido concluir ya en previos gráficos.

2.1.3. Interpretación de la segmentación

Es posible que el algoritmo jerárquico Agglomerative Cluster no haya podido discernir muy bien aquellas personas que ha incluido en el cluster 1, puesto que contenían gran diversidad de opiniones y posturas ideológicas, sin embargo, nos ha revelado un segmento de la población cuya ideología tiende hacia la izquierda y que muestra una inclinación más marcada hacia opiniones que cuestionan la actuación de Israel y respaldan al pueblo palestino. Se trata del cluster 2, un 12,05 % de nuestros encuestados. Todas las representaciones visuales que hemos empleado nos han ayudado a conocer a los integrantes de este cluster y sus opiniones con respecto a la guerra.

De todas formas, me gustaría añadir la conclusión que sacó un estudio del CIS en 2011 llamado "PREFERENCIAS SOBRE LOS PROCESOS DE TOMA DE DECISIONES POLÍTICAS". El estudio incluye una pregunta sobre la ubicación ideológica de los encuestados en una escala de 0 a 10, como la de nuestro cuestionario. Según los resultados, la media de la ubicación ideológica de los encuestados fue de 4,9 %, lo que sugiere que estos se ubican en posiciones centrales. Sin embargo, también muestra que la mayoría de los encuestados se ubican en los extremos de la escala, con un 27,7 % de los encuestados en la extrema izquierda y un 22,3 % en la extrema derecha. Por lo tanto, parece que, por lo general, los encuestados tienden a posicionarse en posiciones centrales, pero la realidad es que son algo más extremos.

Quiero que se tengan en cuenta las conclusiones de este estudio, porque evidencian que cada persona tiene una percepción distinta del 'centro' ideológico y por tanto, no todas seguirán un mismo criterio contestando esta pregunta. Es por ello que, realizar clusters a partir de esta escala puede confundir en cierta medida a los algoritmos y puede que no se obtengan las conclusiones esperadas, al no tener una respuesta del todo objetiva, sincera y clara.

De todas formas, considero que Agglomerative Clustering ha hecho un buen trabajo detectando esa similitud de opiniones entre las 2000 personas encuestadas, que han acabados conformando el cluster 2 y que nos ha mostrado un sector de la población que no duda en manifestar su apoyo a Palestina.



Figura 6: Imagen de RTVE de una manifestación en España en defensa de Palestina

2.2. Caso de estudio 2: Jóvenes y Tercera Edad, y sus respectivas amenazas

2.2.1. Epígrafe Caso 2

Existen numerosos asuntos de gravedad en la actualidad que pueden hacer sentir amenazada a la población. No obstante, según la edad es posible que ese nivel de amenaza sea de mayor o menor gravedad y, que las causas sean distintas. He decidido aprovechar la pregunta número 8 de la encuesta: '*¿En qué medida sientes como una amenaza cada uno de los siguientes asuntos de carácter global?*' para analizar cuáles son los asuntos de actualidad que preocupan más a los jóvenes a diferencia de a los más mayores y, qué nivel de preocupación tienen (Nada/Poco/Bastante/Mucho). Para ello he decidido hacer dos estudios diferentes, uno tomando como muestra el rango de edad de 18 a 24 años y el otro, tomando a los mayores de 65 años. Para cada una de las muestras he empleado las siguientes variables:

- p8. 1: El cambio climático y los desastres naturales.
- p8.2: Los conflictos bélicos (Israel/Hamás, Ucrania...).
- p8.3: Los flujos migratorios.
- p8.4: El terrorismo internacional.
- p8.5: La crisis de inflación que encarece el coste de la vida.
- p8.8: El aumento de los populismos en el mundo.
- p8.9: Las pandemias

He elegido aquellos asuntos que, o nos han tocado personalmente, como pandemias, o son muy recurrentes en los telediarios y, de los que, por tanto, cualquier ha podido oír hablar, como atentados terroristas.

```
subset=datos.loc[(datos['edad_r']==0)] #18 a 24 años
# Seleccionar variables de interés para clustering
subset=subset.rename(columns={
    "p8_1": "cambio climático",
    "p8_2": "conflicto bélico",
    "p8_3": "flujo migratorio",
    "p8_4": "terrorismo internacional",
    "p8_5": "crisis inflación",
    "p8_8": "populismos",
    "p8_9": "pandemias",
})
usadas = ['cambio climático', 'conflicto bélico', 'flujo migratorio',
          'terrorismo internacional', 'crisis inflación',
          'populismos', 'pandemias']
```

A diferencia del caso de estudio anterior, esta vez trato los '*missing values*' imputándolos por sus vecinos más cercanos:

```
imputer = KNNImputer(n_neighbors=3)
datos_norm = datos.apply(norm_to_zero_one)
datos_imputados_array = imputer.fit_transform(datos_norm)
datos_norm2 = pd.DataFrame(datos_imputados_array, columns = datos.columns)
datos = datos_norm2
```

2.2.2. Análisis Caso 2

Estos son los resultados del estudio sobre los encuestados de entre 18 y 24 años (total datos = 151), con los algoritmos ya conocidos:

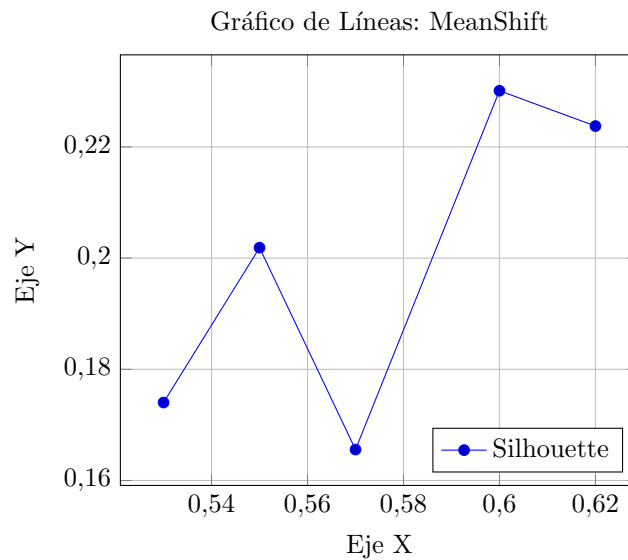
Algoritmo	Nº clusters	Calinski-Harabasz	Silhouette	Davies-Bouldin	Tiempo
K-Means	2	46.797106	0.223667	1.704065	0.205981
MeanShift	2	46.595868	0.223751	1.702539	0.252978
AggCluster	2	36.018123	0.171997	1.944978	0.001000
DBSCAN	2	18.145265	0.190422	2.369131	0.004999
Birch	2	31.649271	0.197917	1.932834	0.007001

Cuadro 3: Métricas Caso 2, 18-24 años

A pesar de que las más eficientes son Agglomerative Clustering y DBSCAN, la que mejor resultado nos da, en cuanto a Silhouette y Davis-Bouldin, es MeanShift. Siendo un algoritmo tan sensible a su parámetro bandwidth, veamos cómo evoluciona al cambiarlo sólo unas centésimas, y cuántos clusters crea, teniendo también en cuenta la proporción.

Bandwidth	Nº clusters	Proporción	Calinski-Harabasz	Silhouette	Davies-Bouldin
0,62	2	60/40	46.595868	0.223751	1.702539
0,6	2	64/36	46.281203	0.230105	1.681655
0,57	4	58/34/7/1	21.573597	0.165564	1.486498
0,55	3	62/37/1	26.26821	0.201878	1.447084
0,53	4	62/31/5/2	21.373569	0.174025	1.475858

Cuadro 4: Métricas Caso 2: MeanShift



Observando el [Cuadro 4](#), apreciamos que unas pocas centésimas de variación en el parámetro bandwidth pueden significar agrupamientos de clusters muy distintos. Se evidencia la consecuencia de minimizar o aumentar el valor de bandwidth:

- **Bandwidth alto:** Cuando aumentas el bandwidth, el algoritmo MeanShift tiende a generar clusters más grandes, ya que cada punto busca atraer más muestras hacia sí mismo. Esto nos ha llevado a la formación de un cluster muy grande que abarca la mayoría de las muestras, lo que puede resultar en un desbalance en la distribución de los datos entre los clusters.
- **Bandwidth bajo:** Por otro lado, si el bandwidth es demasiado bajo, el algoritmo divide los datos en clusters muy pequeños, lo que lleva a una alta fragmentación de los datos y a la formación de muchos clusters de tamaños desiguales.

Se puede advertir que el mejor valor de cada métrica difiere mucho, siendo el valor óptimo de bandwidth para el índice Calinski-Harabasz 0,62, pero 0,6 para Silhouette y 0,55 para Davies-Bouldin. Es-

cogiendo entre estos tres valores para bandwidth creo que optaré por 0,6 pues, generalmente, es el que mejor métricas tiene de los tres.

Veamos si MeanShift también nos da buenas métricas cuando tomamos los datos de aquellas personas mayores de 65 años.

Algoritmo	Nº clusters	Calinski-Harabasz	Silhouette	Davies-Bouldin	Tiempo
K-Means	2	175.789399	0.261074	1.501135	0.220994
MeanShift	2	100.279724	0.227068	1.687392	0.365989
AggCluster	2	138.233690	0.233652	1.575771	0.005000
DBSCAN	2	23.939001	0.317249	1.262513	0.013002
Birch	2	123.483596	0.255387	1.517145	0.024768

Cuadro 5: Métricas Caso 2, mayor de 65 años

Para mi sorpresa, MeanShift no sólo no ha destacado como el mejor algoritmo en el caso de la tercera edad, sino que ha destacado como el peor. Esto se puede deber, no sólo a una variedad de respuestas distintas en esta muestra, sino también a que el tamaño es mayor (439), por lo que facilita que el resto de algoritmos pueda rendir mejor.

Destaca DBSCAN con un silhouette de 0,317, no obstante, se debe a que tiene una distribución de clusters 98/2, debido a que simplemente deja a un lado los outliers, por lo que, no lo tendremos en cuenta.

Con el objetivo de analizar este caso de la forma más fidedigna posible, considero que lo mejor será optar por K-means en vez de por MeanShift, ya que, en el caso de los mayores de 65 años, resulta ser el que mejor métricas nos devuelve con una proporción balanceada, y en el caso de los jóvenes tenía prácticamente los mismos valores que MeanShift por lo que no 'perdemos' mucho.

```
k_means = KMeans(init='k-means++', n_clusters=2, n_init=5,
                  random_state=123456)
```

Con el propósito de tratar de hacernos una idea de cómo están separados nuestros clusters, en la [Figura 7](#) tenemos una gráfica de *Multidimensional Scaling* (MDS). Se trata de una técnica de análisis multivariante utilizada para analizar la similitud o disimilitud entre los clusters en su espacio multidimensional. Gracias a la gráfica, podemos representar los datos de los clusters en un espacio bidimensional mientras se conservan, en la medida de lo posible, las distancias o similitudes originales. Lo he creado con el siguiente código:

```
cluster_predict = alg.fit_predict(X,subset[peso])
labels = alg.labels_
X_labeled = X.copy()
X_labeled['Cluster'] = labels
mds = MDS(n_components=2, random_state=123456)
resultados_mds = mds.fit_transform(X_labeled)
plt.figure(figsize=(10, 8))
plt.scatter(resultados_mds[:, 0], resultados_mds[:, 1], c=labels, cmap='viridis')
plt.title('Multidimensional Scaling (MDS) con K-Means')
plt.colorbar(label='Cluster')
plt.savefig("mds_caso2_kmeans.pdf")
plt.show()
```

En el caso de los jóvenes vemos que los clusters están muy bien diferenciados sin valores en la frontera. El cluster 1 (morado) se compone del 60,26 % de la muestra y tiene un coeficiente silhouette de 0.1601. Se puede ver reflejado este valor algo bajo de silhouette pues se aprecian los puntos algo más dispersos que los del cluster 2.

El cluster 2 (amarillo) tiene el 39,74 % restante con un silhouette algo más alto de 0,2655.

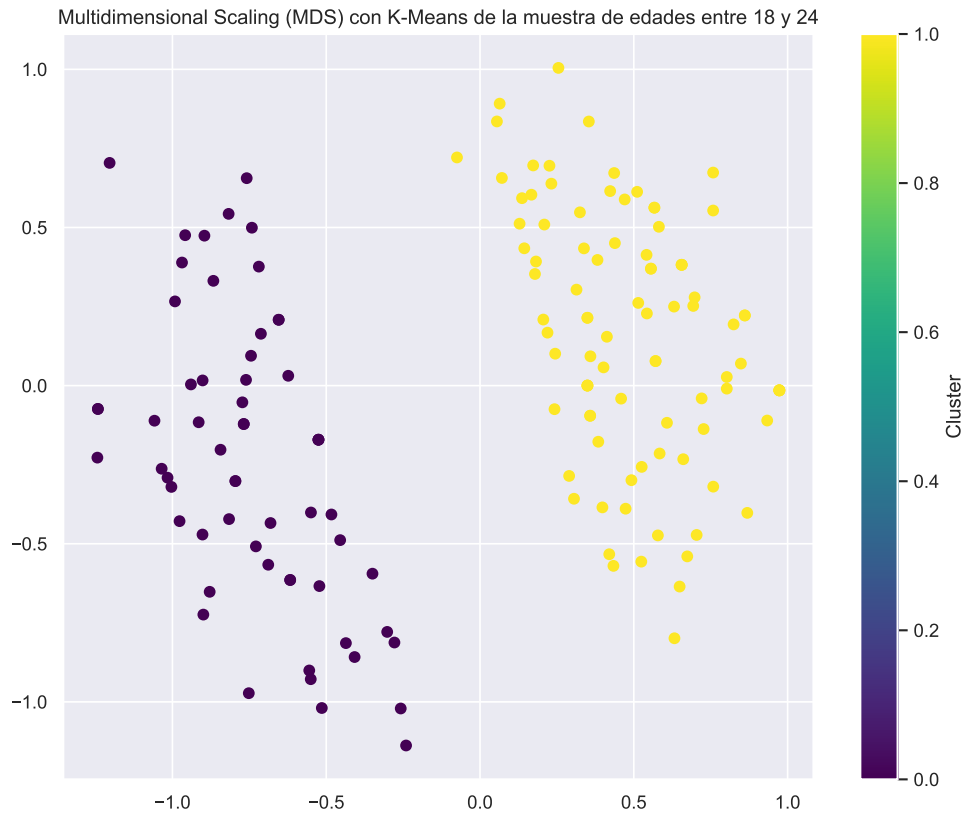


Figura 7: MDS, edades entre 18 y 24

De igual forma, pero para la muestra de mayores de 65, tenemos la [Figura 8](#) donde podemos visualizar una mayor cantidad de puntos pues, hay más encuestados que cumplen con este criterio de edad, en total 439. Se puede ver que se trata de un caso más balanceado en el que el cluster 1 (morado) tiene el 53.76 % de la muestra con un silhouette de 0,3223, y el 2 (amarillo) tiene el otro 46.24 % con silhouette algo más bajo de 0,1898.

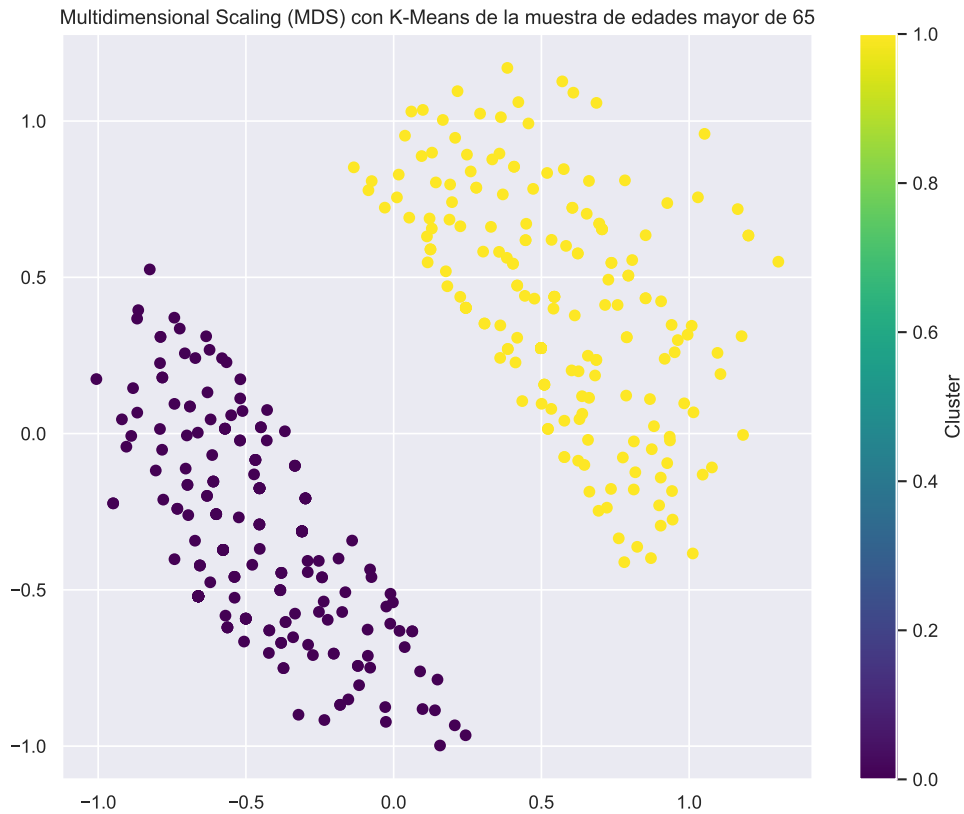


Figura 8: MDS, edades mayores de 65

En este heatmap de la [Figura 9](#) podemos distinguir a primera vista observando los colores, que los dos clusters se han formado según el nivel de preocupación general por las amenazas. Es decir, podríamos decir que los jóvenes se dividen 60/40 entre los más preocupados por las amenazas globales y los que menos se preocupan. No obstante, estos dos grupos tienen algunos puntos de encuentro como en el flujo migratorio o los populismos que parecen ser los asuntos que menos inquietan a la juventud.

Si lo comparamos con el heatmap de los mayores de 65 años de la [Figura 10](#) podemos ver que de forma equivalente, esta muestra de la población también se puede dividir entre los más amenazados generalmente y los que menos, con una proporción 54/46. Pero, fijándonos en los colores vemos que, globalmente, son todos más oscuros. Esto es porque los mayores de 65 años sienten un grado de preocupación mucho mayor por estas cuestiones de alerta ciudadana.

Llaman la atención en ambos casos los colores tan oscuros que vemos en conflictos bélicos y terrorismo internacional. Probablemente se deba a las noticias que escuchamos cada día sobre guerras en [Ucrania](#), [Gaza](#) y atentados terroristas como el de [París](#), el pasado 2 de Noviembre. Este tipo de noticias de forma recurrente puede hacer pensar a la población que esta clase de amenazas resulta cada vez más inminente.

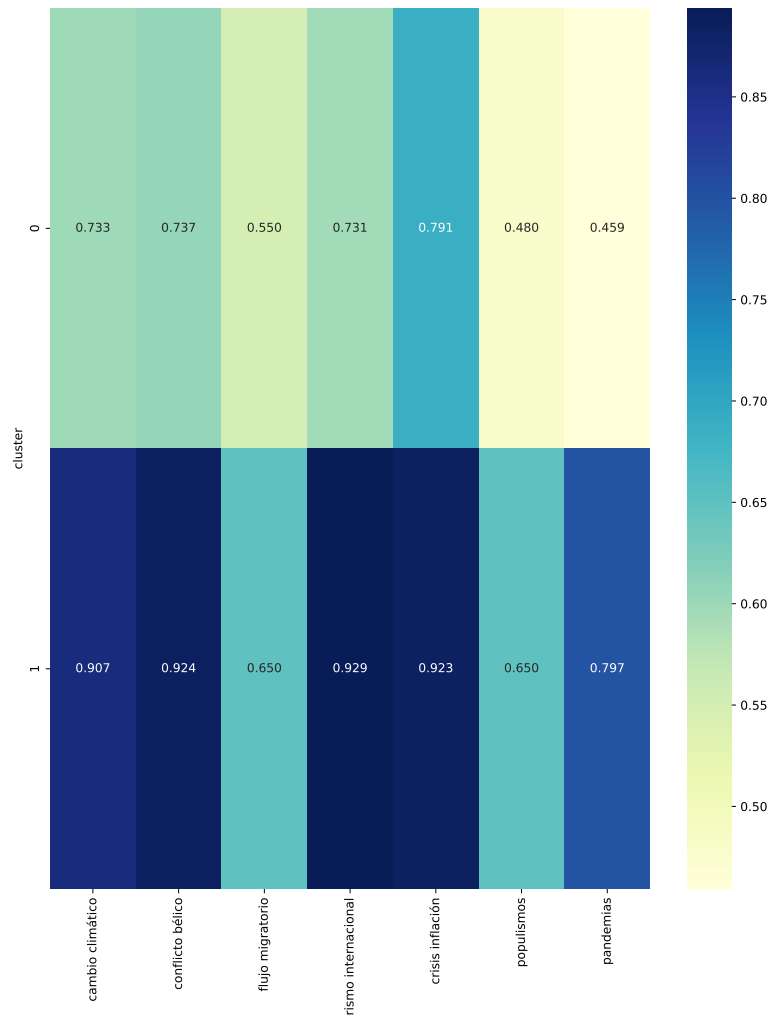


Figura 9: Heatmap para edades entre 18 y 24

Asimismo, en ambos casos, la variable de la inflación destaca por ser uno de los asuntos más preocupantes. No es de extrañar pues afecta gravemente tanto a los [jóvenes](#) que aún no tienen sueldo y deben pagarse su vida estudiantil, como a los [pensionistas](#) que tienen complicaciones para llegar a fin de mes. Todos hemos notado ese aumento de precios en el carro de la [compra](#) que nos sigue sorprendiendo y que tememos que no vuelva a ser como antes.

Estas representaciones gráficas evidencian lo amenazados que se ven estos sectores de la población, pues los números llegan a equipararse con situaciones tan adversas como conflictos bélicos comentados anteriormente.

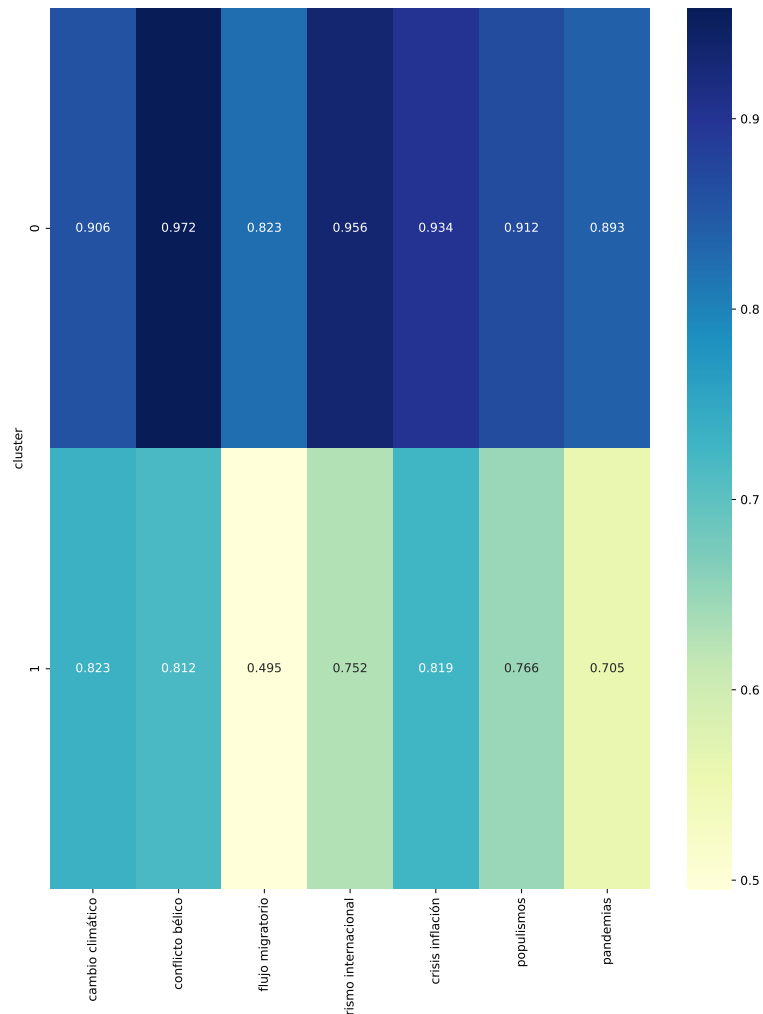


Figura 10: Heatmap para edades mayores de 65

También se pueden apreciar los distintos niveles de preocupación del caso de jóvenes en el siguiente KPlot de la [Figura 11](#). En este gráfico podemos analizar mejor qué tipo de jóvenes conforman cada grupo. Vemos cómo en el cluster naranja todos los picos están próximos al extremo derecho del eje x, demostrando así que este grupo se siente amenazado con mayor gravedad por todos estos asuntos, especialmente, como habíamos comentado, por el terrorismo internacional y los conflictos bélicos.

Además, también muestran gran preocupación por el cambio climático. De hecho, existen numerosos estudios como [este](#) que así lo confirman, mencionando también que dichos jóvenes suelen sufrir 'ecoansiedad'. La American Psychology Association (APA) describe la ecoansiedad como "el temor crónico a sufrir un cataclismo ambiental que se produce al observar el impacto aparentemente irrevocable del cambio climático y la preocupación asociada por el futuro de uno mismo y de las próximas generaciones". Podemos comprobar esta afirmación fijándonos en el histograma de la [Figura 12](#) sobre la pregunta 9 del cuestionario: "¿Y qué sientes o has sentido a raíz de estos asuntos", tomando el apartado del cambio climático. Lo he hecho con el siguiente código:

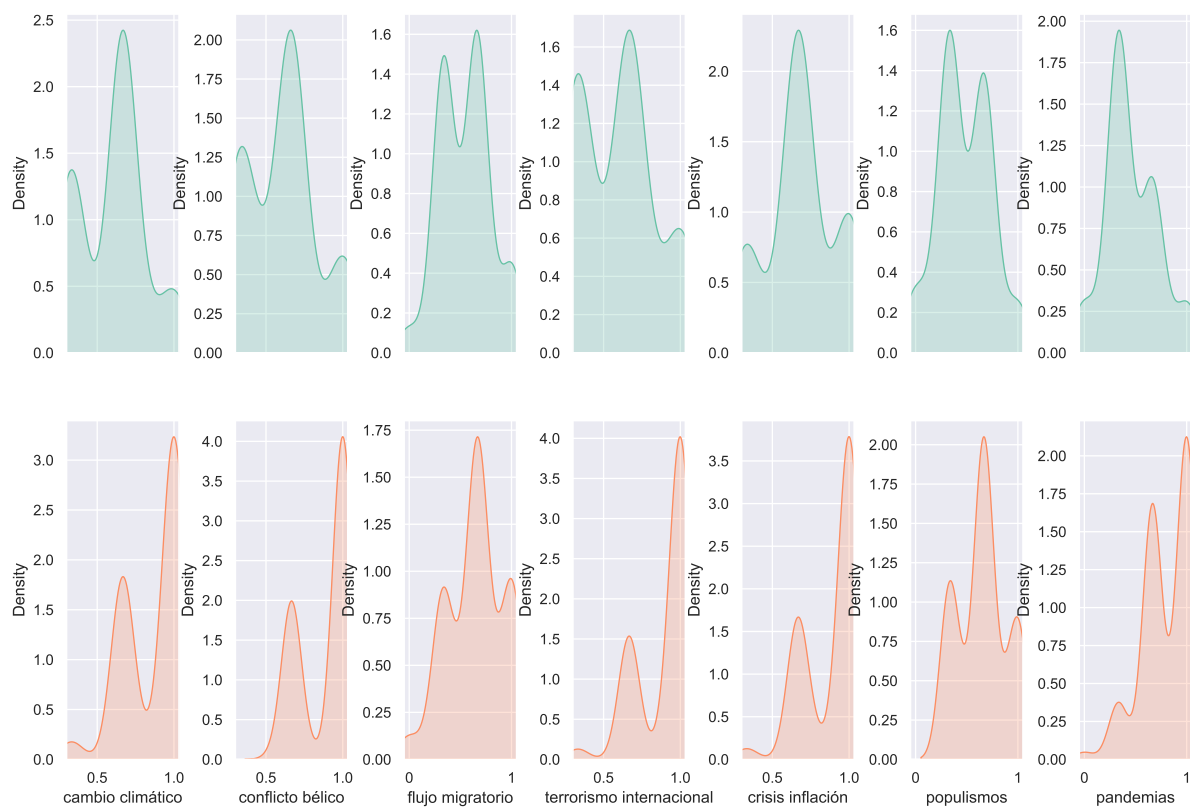


Figura 11: KPlot para edades entre 18 y 24

```
subset=datos.loc[(datos['edad_r']==0)]
respuestas_cambio_climatico = subset['p9_1']
plt.figure(figsize=(10, 8))
plt.hist(respuestas_cambio_climatico, bins=10, color='skyblue', edgecolor='black')
plt.xlabel('Respuestas sobre Cambio Climático')
plt.title('Histograma de Respuestas sobre Cambio Climático (18-24 años)')
plt.grid(True)
plt.show()
plt.savefig("histograma_cambio_climatico.pdf")
```

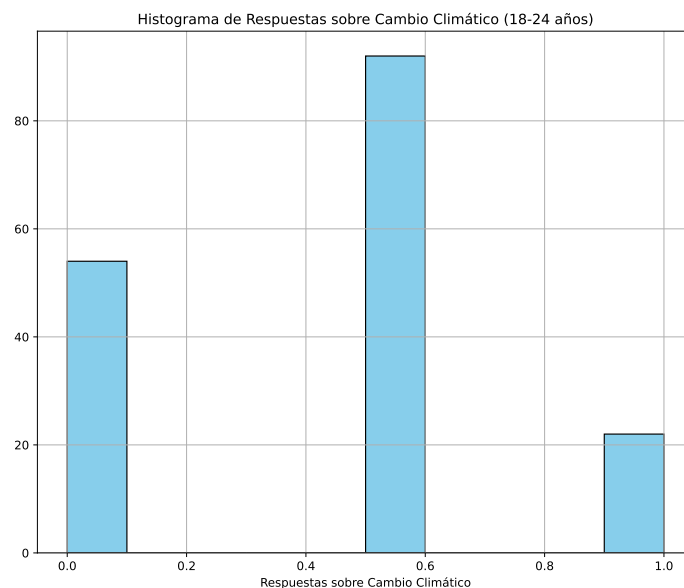


Figura 12: Histograma de la pregunta 9.1)

Esta pregunta tiene tres posibles respuestas: 'No he sentido nada de esto', 'Siento o he sentido desánimo y/o pesimismo' o 'Sufro o he sufrido ansiedad, depresión'. Vemos cómo la mayoría de los encuestados sí han sentido algo al respecto; gran parte de ellos desánimo o pesimismo pero hay más de 20 casos de entre 151 en los que sufrieron ansiedad o depresión. La urgencia climática es inminente y todos nosotros somos testigos de ella en nuestro día a día.

En la [Figura 13](#) vemos, por otra parte, el KPlot de los mayores de 65 años. En este caso, es el cluster verde el que siente una mayor amenaza comparado con el naranja. Fijándonos en el eje 'y' de cada variable vemos que se alcanzan niveles mucho más altos que en el caso de los jóvenes. Demostrando, una vez más, que son los más mayores los que más se preocupan, y que están especialmente turbados también por el terrorismo y las guerras, pues un mayor número de personas en cada cluster eligió la opción más extrema de la pregunta: "Mucho".

Comparando los dos KPlots, notamos que a ambos clusters de los mayores de 65 años les preocupan en mayor medida las pandemias que a los clusters de los jóvenes. La causa puede resultarnos fácil de deducir. Según [datos](#) del Ministerio de Sanidad "De 145.358 casos confirmados, fueron hospitalizados 67.451 y de ellos el 72,3% tenían más de 70 años, 48.760 personas, de las que fallecieron 10.576 que corresponden al 21,7% de los mayores de 70 años hospitalizados". Como todos presenciamos, los mayores fueron los más castigados durante la pandemia y los que más precauciones tuvieron que tomar durante la misma. En esta gráfica se ve reflejado ese creciente miedo que ha provocado el COVID, ante la posibilidad de que aparezca algún otro virus que pueda volver a amenazar al planeta en igual, o incluso, mayor medida.

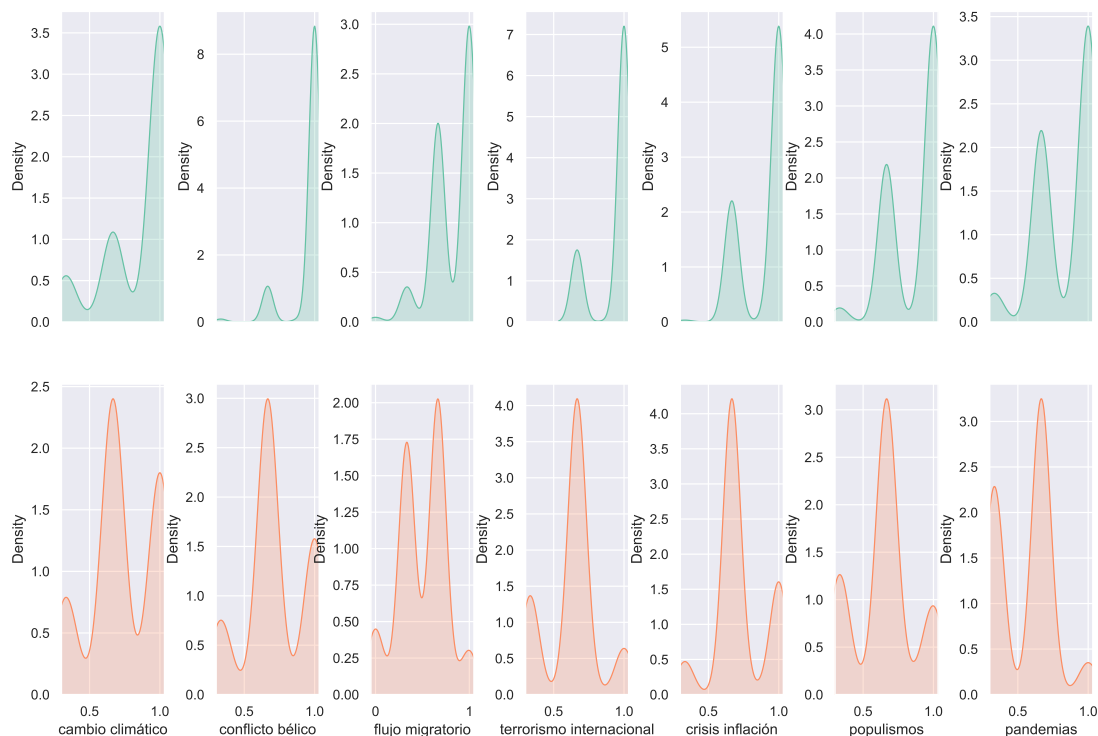


Figura 13: KPlot para edades mayores de 65

Por último, quiero dejar representadas las gráficas de dispersión de ambos casos. En las Figuras 14 y 15 podemos distinguir las relaciones entre las variables de la pregunta 8. En ambos casos vemos como los clusters 'preocupados' de ambos casos se posicionan en las esquinas superiores derechas de cada cuadro; en el caso de los jóvenes sería el naranja y en el de los mayores, el verde.

Parece que en ambos casos, la columna del flujo migratorio es la que más mezcla ejemplos de cada cluster. Es un tema controvertido y difícil que según de donde provenga cada encuestado, puede diferir su enfoque hacia este asunto desde sus dos perspectivas: tanto en inmigración, como se está viendo implicada [Canarias](#) últimamente, como en emigración, en los numerosos pueblos de la '[España vaciada](#)'.

Llama la atención la relación entre las variables 'crisis de inflación que encarece el coste vida' y 'conflicto bélico' en el caso de los mayores de 65, pues parece que el cluster verde se reparte de diversas formas según qué variables pero, en relación a estas dos, todos sus integrantes guardan gran preocupación por ambas. Una posible causa podría ser los efectos que provoca un conflicto bélico como el de Ucrania. Según [este artículo](#) de accenture, el impacto de las sanciones impuestas a Putin sobre su exportación de petróleo agravó la presión inflacionaria y, como todos notamos repostando nuestro coche, encareció la gasolina. Por lo que, en varias ocasiones, un evento puede desembocar en el otro.

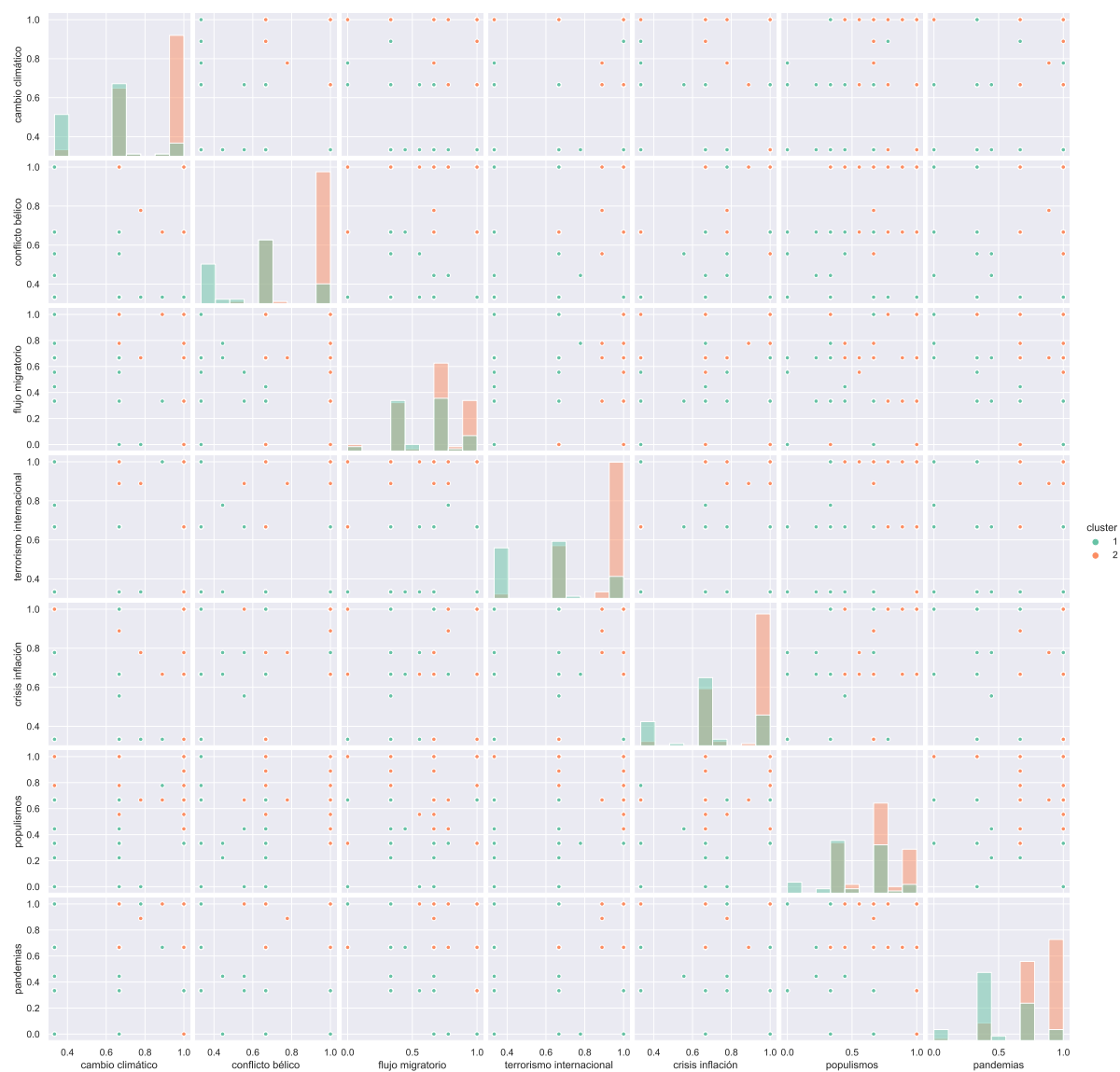


Figura 14: Scatter Matrix para edades entre 18 y 24

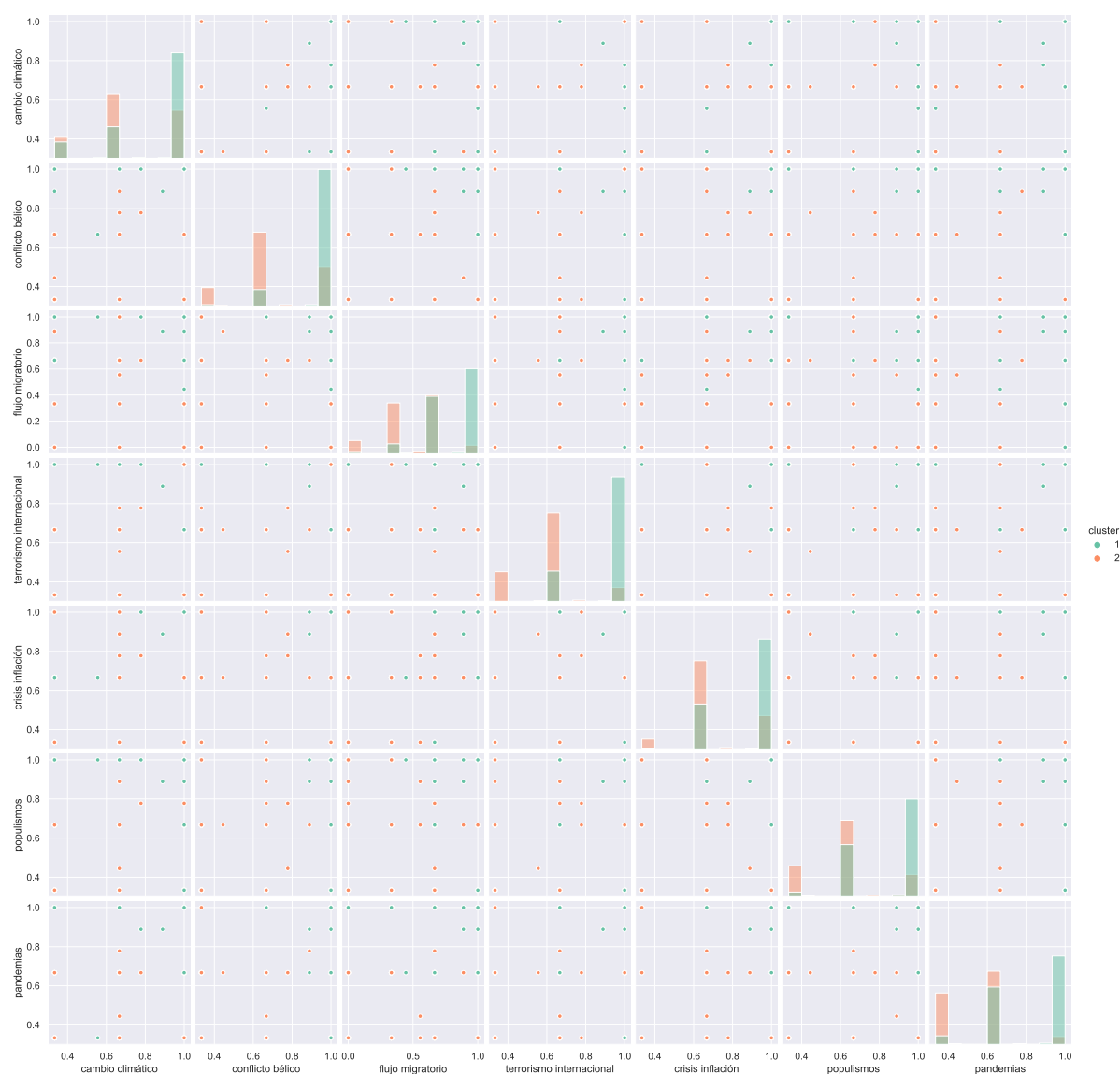


Figura 15: Scatter Matrix para edades mayores de 65

2.2.3. Interpretación de la segmentación

Realizando este estudio sobre los más jóvenes y los más mayores de nuestro grupo de encuestados nos hemos podido dar cuenta de que, en ambos casos, podemos dividir estos grupos entre los 'más y menos preocupados' por cualquier asunto en general. Pero, sin embargo, es el grupo sénior el que más amenazado se siente ante todas estas cuestiones políticas, económicas y sociales.

Destacar que los asuntos de mayor gravedad según estos dos grupos son, sin duda alguna:

- Los conflictos bélicos: Teniendo el de Ucrania y el de Gaza muy presentes ahora mismo, estando tan cerca de nuestro país.
- El terrorismo internacional: Posiblemente al oír noticias recurrentes de atentados en países occidentales.
- La crisis de la inflación que encarece el coste de la vida: Cada año la cesta de la compra, la gasolina y demás bienes del día a día salen un poco más caros.

- El cambio climático: En España notamos cómo cada año hace más calor, llueve menos y sufrimos más incendios forestales; por lo que es un asunto muy presente en nuestra actualidad.

Comparar ambos grupos ha sido interesante para demostrar que diferentes asuntos afectan de forma distinta a diferentes generaciones. Como hemos comentado, es comprensible que las pandemias sean un asunto más importante para los mayores de 65 años que para aquellos entre 18 o 24 años, pues pueden llegar a ser más vulnerables ante las mismas.

Al igual que las pandemias, ellos también se ven más comprometidos por el auge de los populismos. Este puede ser un tema controvertido pues, no toda la población tendrá la misma definición y percepción de 'populismo'. No obstante, es importante mencionar que es la cuestión que menos parece preocupar a la juventud, de manera general.

2.3. Caso 3: ¿Qué partido está perdiendo votos?

2.3.1. Epígrafe Caso 3

Como último caso, he optado por un análisis algo más sencillo, pues requerirá menos variables, pero que puede resultar de gran interés a los líderes políticos.

He aprovechado la pregunta 2 del cuestionario: 'Si mañana se celebraran unas nuevas elecciones generales, ¿cuál sería la probabilidad de que acudieras a votar?' en la cual se puede resolver con una escala del 0 al 10, en la que '0' representa 'con toda seguridad, no iría a votar' y '10' 'con toda seguridad, iría a votar', para filtrar la base de datos por aquellos que, dentro de esta escala, se posicionan entre el 0 y el 3, es decir, lo más probable es que no voten.

```
subset = datos.loc[(datos['p2']==0) | (datos['p2']==1) | (datos['p2']==2) | (datos['p2']==3)]
```

Compararé estos datos con aquellos encuestados que tienen un nivel 9 o 10 de seguridad de ir a votar.

```
subset = datos.loc[(datos['p2']==9) | (datos['p2']==10)]
```

Dentro de este análisis comparativo utilizaré las variables de la pregunta 5: "¿cuál sería la probabilidad de que votases a los siguientes partidos?", que también cuenta con una escala del 0 al 10:

- p5.1: PSOE
- p5.1: PP
- p5.1: Vox
- p5.1: Sumar

Con el siguiente caso de estudio me gustaría estudiar los perfiles de aquellos que rechazan votar en las próximas elecciones y los perfiles de quienes están dispuestos a votar con toda seguridad. Creo que puede ser interesante comprobar qué partido político tiene los simpatizantes más dispuestos a votar. Asimismo, analizar a qué partido político están más abiertos los 'no votantes' pues significaría que este partido está perdiendo posibles votos, ya que esta persona, dentro de no querer votar, puede estar más abierta a un partido que a otro.

2.3.2. Análisis Caso 3

Procedemos al estudio del tercer caso con las cinco variables que ya conocemos.

Algoritmo	Nº clusters	Calinski-Harabasz	Silhouette	Davies-Bouldin	Tiempo
K-Means	2	101.149249	0.404675	1.189722	0.214834
MeanShift	2	82.086397	0.406449	1.288634	0.218383
AggCluster	2	77.199140	0.337715	1.494682	0.004975
DBSCAN	2	84.459931	0.400834	1.439001	0.006507
Birch	2	89.377875	0.388795	1.143831	0.0070008

Cuadro 6: Métricas Caso 3, No Votantes

Algoritmo	Nº clusters	Calinski-Harabasz	Silhouette	Davies-Bouldin	Tiempo
K-Means	2	1556.151078	0.463946	0.869172	0.177224
MeanShift	2	1533.557902	0.465349	0.867470	0.180339
AggCluster	2	1436.714510	0.454702	0.875841	0.038537
DBSCAN	2	12.642949	0.133744	2.058817	0.03104
Birch	2	1049.372759	0.383248	1.056953	0.030766

Cuadro 7: Métricas Caso 3, Votantes

En ambos casos, el mejor algoritmo es MeanShift, que, aunque no tenga el mejor Calinski-Harabasz, sí que cuenta con los mejores Silhouettes. Aprovecharemos estas buenas métricas para imprimir representaciones gráficas y analizar en más profundidad este caso de estudio con MeanShift.

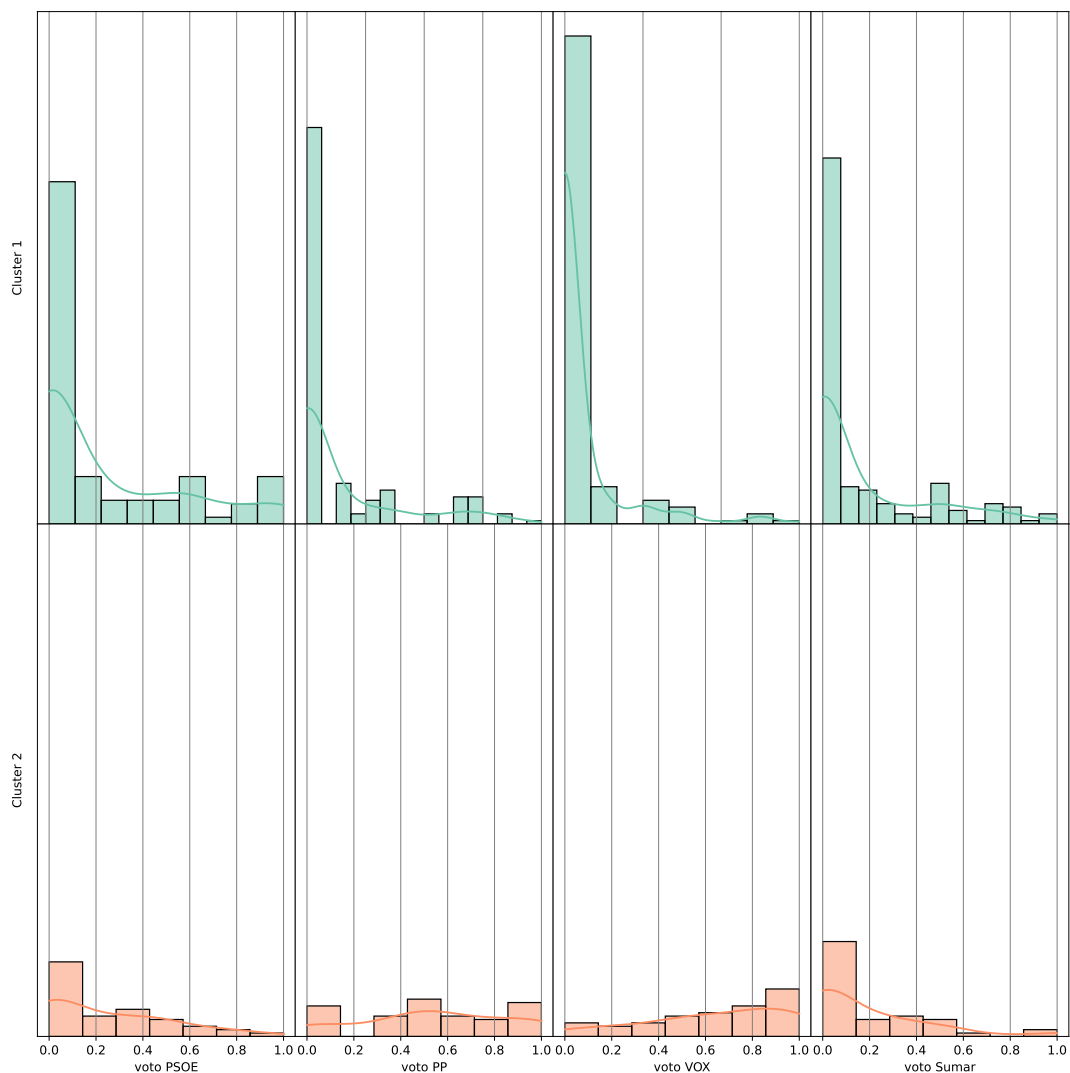


Figura 16: Gráfica de distribución: No Votantes

Comenzamos por el caso de aquellos cuya probabilidad de votar está entre 0 y 3. Los llamaremos los 'No Votantes'. MeanShift ha dividido esta muestra en dos clusters. El cluster 1 está formado por el 78,54 % de los 'No Votantes' y presenta un coeficiente silhouette de 0,4526 y el cluster 2 tiene el 21.46 % restante con un silhouette de 0,2373.

En la [Figura 16](#) se pueden ver las probabilidades de que los integrantes de cada cluster vote a los cuatros principales partidos políticos en España. El cluster 1 se caracteriza por ser más 'radical' en cuanto a sus respuestas. La mayoría de personas de este cluster no votarían a ninguno de estos partidos. Notamos que existe más rechazo hacia Vox y menos hacia el PSOE. De hecho, es en la variable del PSOE donde vemos una mayor posibilidad de votarles con probabilidades relativamente altas.

Por otro lado, el cluster naranja, con menos integrantes, parece opinar lo contrario. Tiene más rechazo hacia PSOE y Sumar, y menos hacia PP y Vox.

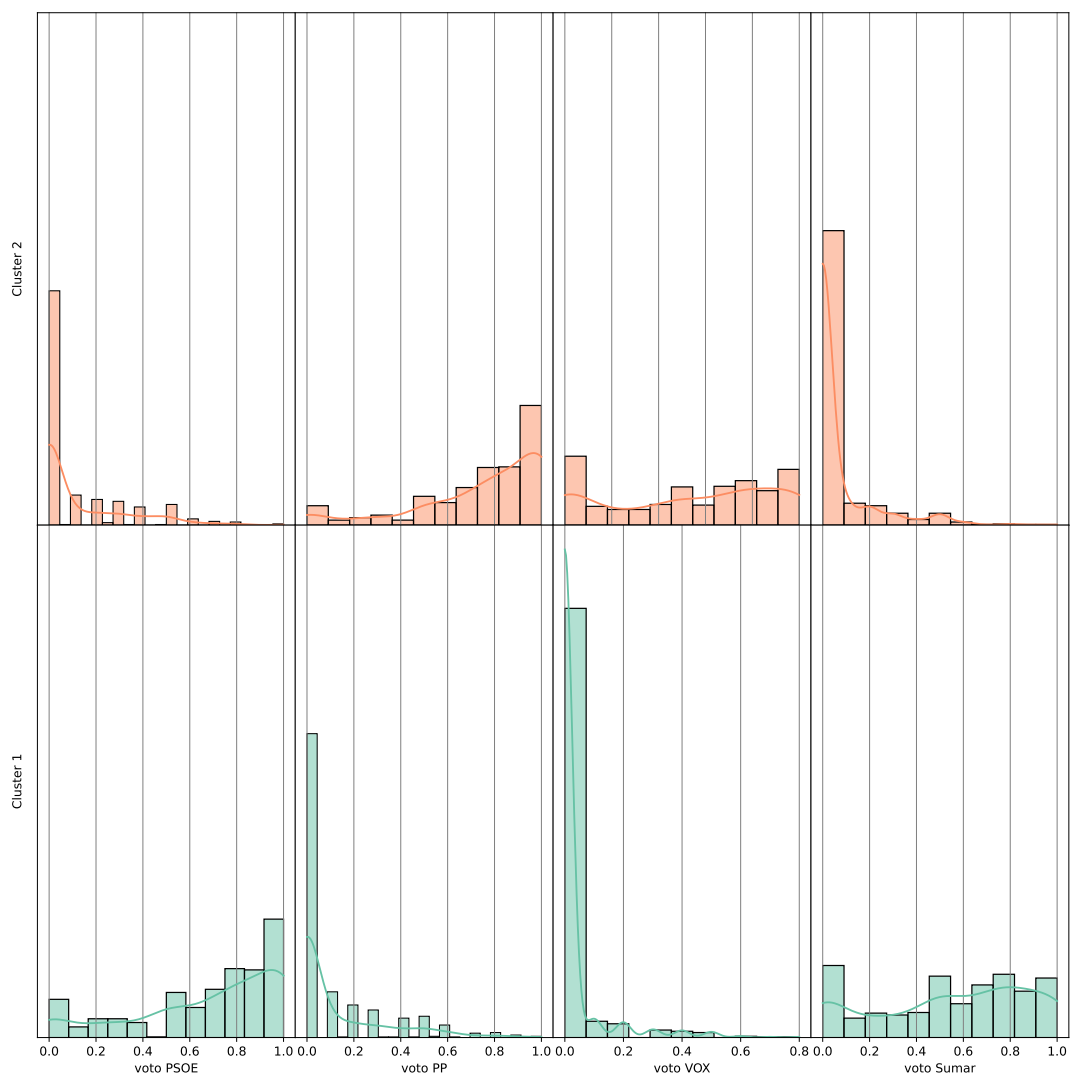


Figura 17: Gráfica de distribución: Votantes

En cuanto a aquellos que votarán con probabilidad de 9 o 10, a los que podemos llamar 'Votantes' para simplificar, MeanShift también los ha dividido en dos clusters. El cluster 1 lo conforman 767 encuestados de esta muestra, lo que supone el 56,40 %, y tiene un coeficiente silhouette de 0,4664. El cluster 2 tiene el 43,60 % de la muestra, que implica 593 votantes, y un silhouette de 0,4639. Llama la atención que ambos clusters están muy balanceados, suponiendo una división casi de 50/50 con unos coeficientes de silhouette prácticamente idénticos.

Podemos ver en la [Figura 17](#) que, al igual que en el caso de los 'No Votantes', los clusters siguen el criterio de división de 'aquellos que están abiertos a votar a PSOE y Sumar y cerrados a PP y Vox' y lo contrario, 'aquellos abiertos a PP y Vox y cerrados a PSOE y Sumar'.

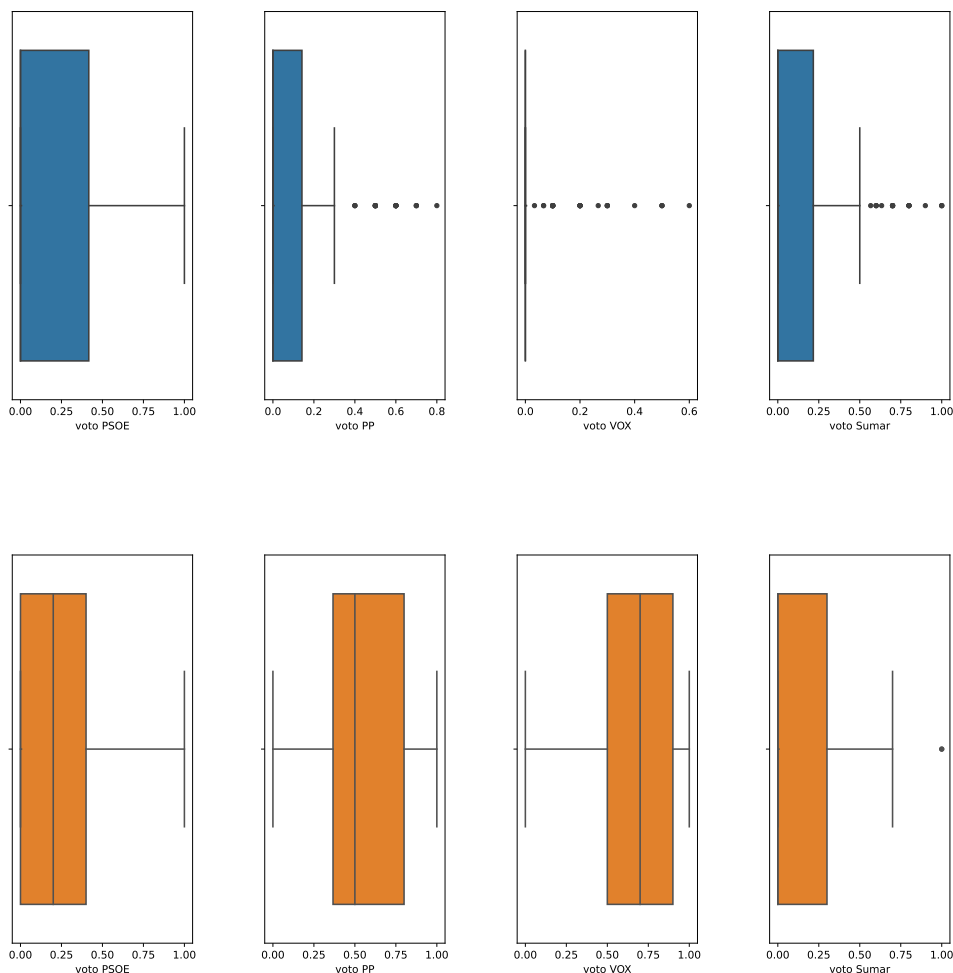


Figura 18: Boxplot de los No Votantes

Para entender mejor estos clusters, podemos utilizar otro tipo de visualización como el Boxplot de la [Figura 18](#). El rectángulo central, o caja, en el boxplot representa el rango intercuartílico (IQR) y la línea central dentro del rectángulo representa la mediana.

Destaca cómo en el cluster azul, el 1, todas las cajas están pegadas al 0 y, en todas menos en la del PSOE, se ven puntos individuales más allá de a donde alcanza el extremo del 'bigote' reflejando que, aquellos casos dentro del cluster en los que se responde una probabilidad mayor de votar a estos partidos, son casos aislados que no representan el conjunto del cluster. En este gráfico se puede distinguir muy bien cómo la variable del PSOE es la que más puntuaciones 'probables' tiene, representada con un tamaño de caja mucho más ancho.

Por otro lado, la mediana del cluster naranja nos revela el punto medio de la distribución. También se evidencian las conclusiones sacadas anteriormente. Es un cluster formado por aquellos más simpatizantes del PP y de Vox, y con prácticamente ningún caso en el que se plantee votar al PSOE, o especialmente a Sumar.

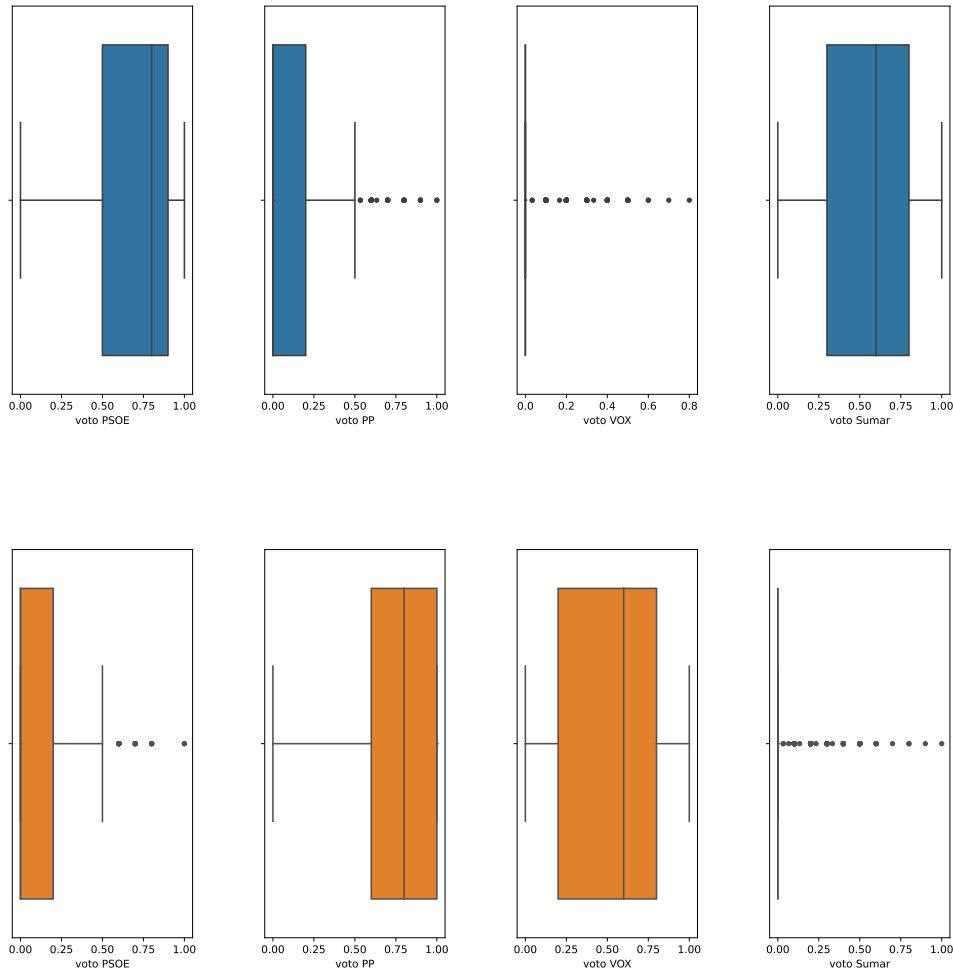


Figura 19: Boxplot de los Votantes

El Boxplot de los Votantes de la [Figura 19](#) parece que tiene las decisiones mucho más claras pues, a primera vista ya vemos más cantidad de puntos individuales. En el caso del cluster azul, estos puntos se encuentran en las mayores probabilidades de PP y Vox. Destaca Vox por no llegar a tener ni la forma de la caja, representándose con una única línea a la izquierda. Hay una opinión más diversa sobre PP, con una caja más ancha; pero ambos Sumar y PSOE se llevan la mayoría de votos en este cluster.

Por otro lado, en el naranja, donde vemos los puntos es en los partidos de izquierdas. Siendo ahora el que no llega a tener caja Sumar. Es evidente que Sumar y Vox tienen simpatizantes completamente opuestos. Este cluster tiende hacia una ideología de derecha que mantiene una probabilidad media de voto hacia Vox, pero muy alta hacia el PP.

2.3.3. Interpretación de la segmentación

Cualquiera que conozca un poco sobre la situación política actual de nuestro país, sabe que la derecha y la izquierda cada vez son más opuestas y cada vez se repelen más. Especialmente los partidos más extremistas como Sumar o Vox. Es por ello que no es sorpresa que hasta MeanShift se haya dado cuenta, habiendo dividido ambos clusters entre 'derecha' e 'izquierda'.

No obstante, creo que ha sido un caso de estudio interesante para analizar a aquellas personas 'No Votantes'. El hecho de que no voten, no significa que no tengan una ideología definida pues, se ha demostrado que prefieren antes una situación política a otra. Con un cluster casi cuatro veces más grande que el otro, parece que los 'No Votantes' proceden más de la izquierda que de la derecha y estarían más abiertos a votar al PSOE.

Teniendo en cuenta las últimas elecciones, el hecho de que ninguno de estos partidos llegó a mayoría absoluta, y todas las negociaciones que se tienen que dar para llegar a formar gobierno, al PSOE le interesaría convencer a ese sector de la población que no está dispuesta a votar pues, parece que si lo hiciera, sería a ellos, y cada voto cuenta.

En la otra cara de la moneda están los más dispuestos a votar, que, sorprendentemente están divididos prácticamente a la mitad, y de nuevo, entre izquierda y derecha. Los votantes más fieles parecen ser los del PP, seguidos muy de cerca del PSOE. Esto como todos sabemos, se hizo notar en las elecciones del [23J](#) donde consiguieron prácticamente el mismo número de escaños.

En ambos casos no destacan tanto las respuestas rotundas positivas, sino más bien las negativas. Podríamos concluir que los españoles tenemos más claro a quien no votar que a quien sí.

3. Bibliografía

- Model Differently: Modelos de clustering
- PGConocimiento: Diagramas Boxplot
- Público: Así se posiciona cada partido en el conflicto entre Israel y Palestina
- RTVE: Centenares de personas protestan en distintas ciudades españolas al grito de 'Palestina libertad'
- RTVE: Resumen de la guerra entre Rusia y Ucrania el 6 de diciembre de 2023
- Público: La ofensiva israelí en el sur de Gaza arrincona a la población palestina en una trampa mortal
- El Mundo: Un turista muerto y dos heridos en un ataque terrorista con cuchillo al lado de la Torre Eiffel
- Cinco Días: El coste de la vida es la mayor preocupación de los jóvenes
- Newtral: Los pensionistas han perdido poder adquisitivo en ocho de los últimos 15 años
- Statista: Ranking de supermercados en función de la subida de los precios de la cesta de la compra producida en España en septiembre de 2023 en comparación con septiembre de 2021
- Europa Press: El 97% de jóvenes españoles, preocupado por la emergencia climática y un 82% ha sufrido ecoansiedad, según un estudio
- Libertad Digital: La llegada de inmigrantes a Canarias duplica las cifras de 2022 y bate el récord de la crisis de los cayucos
- El Español: Los 1.943 'pueblos' de España con un solo vecino
- Accenture: Guerra en Ucrania: la hora de la reflexión en la industria del petróleo y gas
- Resultados 23J