

# Analisis de Conglomerados

Saúl &amp; Leobardo

2024-05-14

El objetivo en este análisis es identificar grupos de clientes que en un futuro se puedan usar para focalizar la publicidad de la empresa.

Como primer paso vamos a considerar que las variables son del tipo continuas, entonces dado ese supuesto obtendremos algunos grupos considerando el método k-means.

```
##
##      1      2      3
## 110 359 470
```

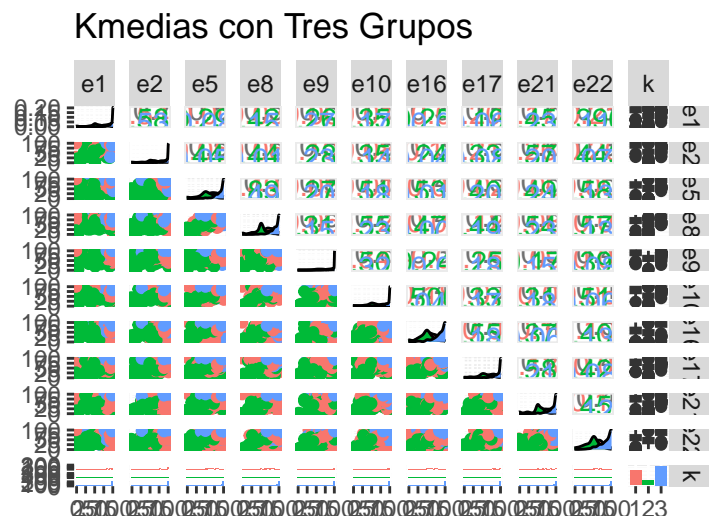
De las preguntas que se nos proporcionaron podemos inferir que hacen referencia a la experiencia que la Aerolínea le brinda al usuario con cosas como la limpieza del avión, la comodidad, los alimentos y la atención que le brinda el personal, además de hacerle su experiencia algo fácil y entretenido. Dados estos puntos podemos focalizar la publicidad de la empresa en 3 grupos de clientes que van a estar divididos en:

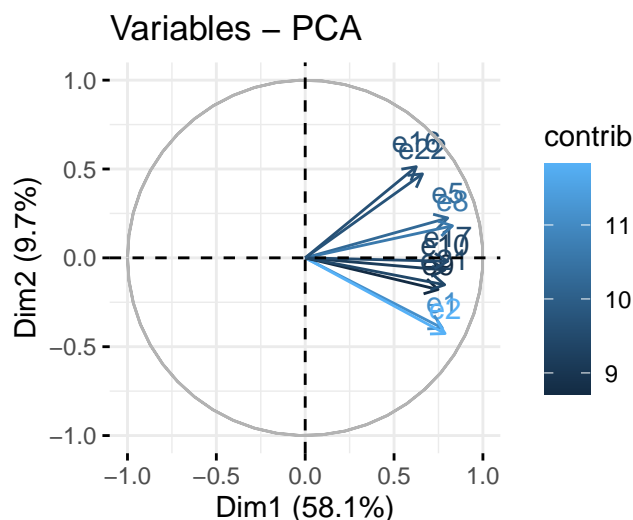
1. Los clientes que esperan que el servicio sea puntual y se lleve conforme a lo planeado y una buena seguridad.
2. Los que quieren recibir un servicio seguro y con una gran experiencia de servicio por parte del personal.
3. Los que esperan un servicio puntual, una gran experiencia por parte del personal y que sea seguro para ellos.

Dadas esas clasificaciones que tomamos vamos a diseñar modelos que nos sirvan para clasificar a los clientes en estas 3 categorías en función de las expectativas que tienen del servicio que ofrece la Aerolínea.

De esta primer grafica podemos ver que el grupo se concentra mas que el segundo grupo, es decir, muestran mas expectativas altas por el servicio que la aerolinea les ofrece mientras que el segundo grupo parece no tener expectativas tan grandes.

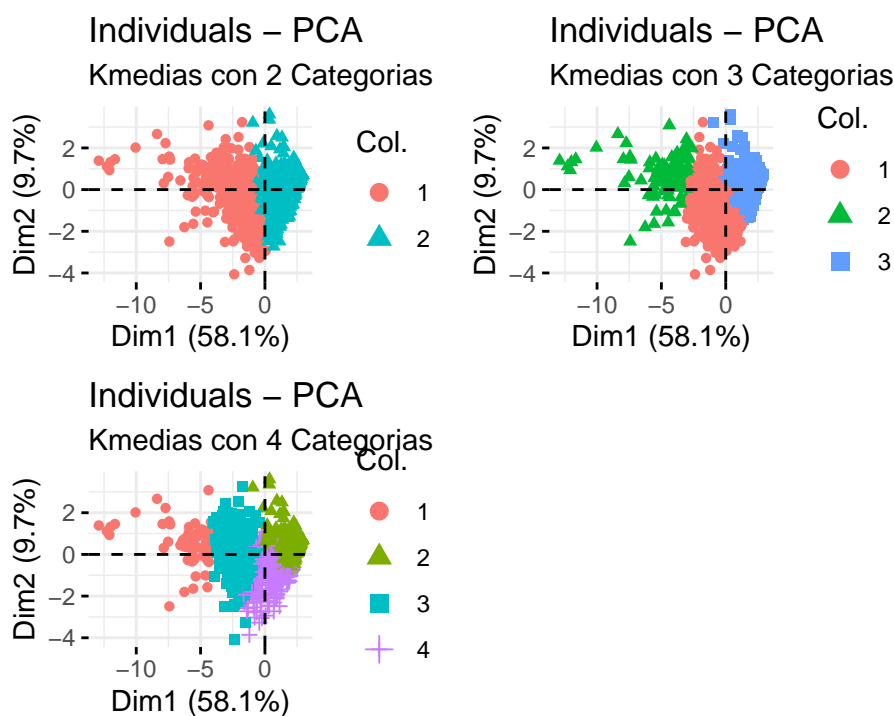
```
## named list()
```





En esta nueva gráfica con 3 grupos vemos que el mas alto ahora fue el tercero con expectativas promedio mas altas, seguido de el grupo 1 y el grupo 2 respectivamente.

En la siguiente gráfica tenemos que la primera y segunda componente principal conservan una varianza de 58.1% y 9.7% respectivamente.



En la primer gráfica, vemos que a la derecha se encuentran los clientes potenciales con buenas expectativas en general en todas las preguntas del cuestionario, mientras que a la izquierda los de regular y mala, según la primera componente principal.

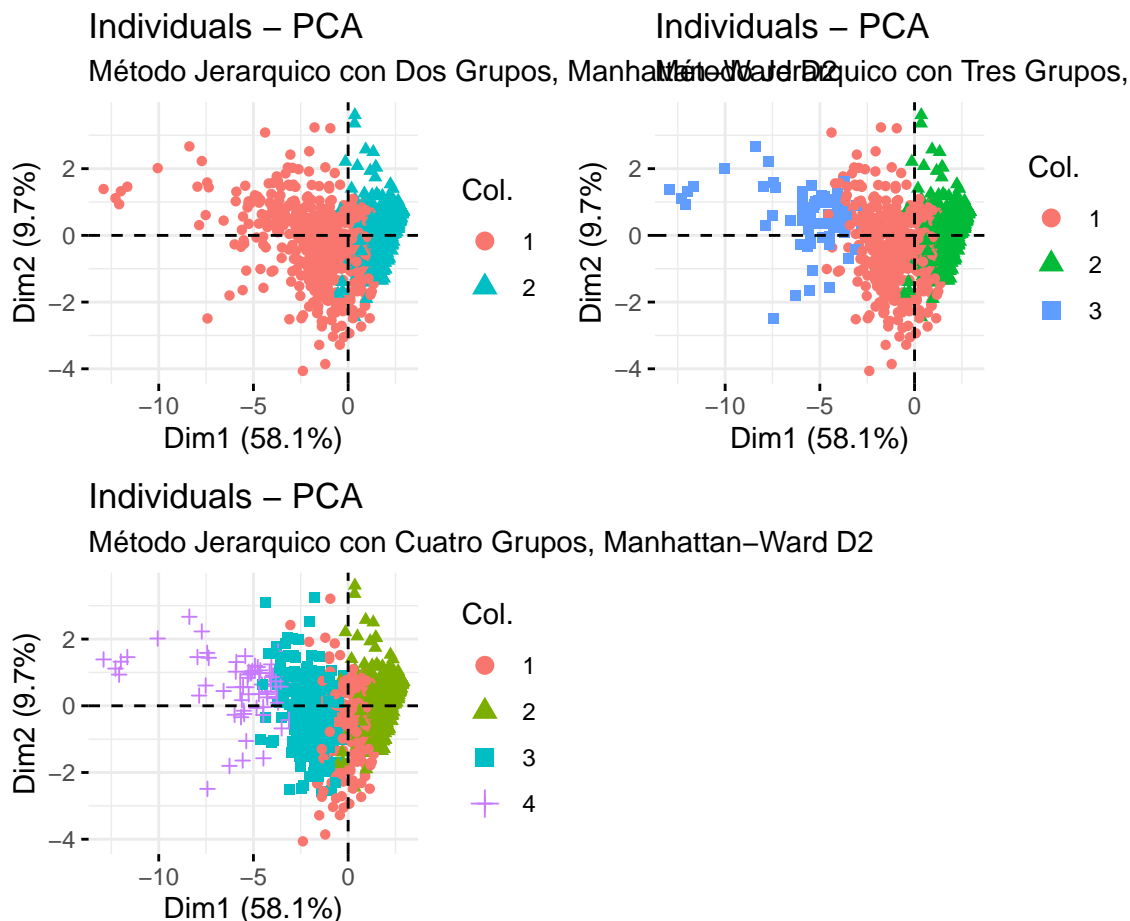
De acuerdo al método silhouette utilizado, se nos sugiere tomar dos grupos, pero en nuestro caso para mejorar la interpretación decidimos que es recomendable usar más de 2. Adicionalmente podemos ver que no cambia mucho la interpretación si nos quedamos con 3 grupos o con 4, pues cuando agrupamos en 4 la concentración de personas en el grupo 3 es muy pequeña.

Observando los componentes principales, podemos decir que es mejor focalizar la publicidad en 3 grupos de

clientes: los que esperan puntualidad y seguridad, los que esperan seguridad y los que esperan calidad de servicio, puntualidad y seguridad. Tal como se decidió agrupar desde un inicio.

## Metodo Jerarquico Aglomerativo

Para esto vamos a tomar que las variables son continuas como se hizo anteriormente y tomando tanto las escalas dadas como haciendo transformaciones. Además agregaremos las disimilaridades entre clientes y clusters.



En esta figura podemos observar las mismas comparaciones que realizamos en el ejercicio 1 donde se puede ver que el resultado obtenido en este caso aplicando el método aglomerativo resulto ser muy similar al obtenido con K-means. En esta ocasión los 3 grupos de clientes son: los que esperan puntualidad y seguridad, los que esperan calidad de servicio, puntualidad y seguridad, y los que esperan seguridad solamente. Algo que notamos que cambio fue en los promedio de expectativas de los clientes donde los clientes que se enfocan mas en la seguridad ya se encuentran mas cerca de los que tiene mas expectativa tanto en la puntualidad como en la seguridad.

## Modificaciones y uso de Componentes principales

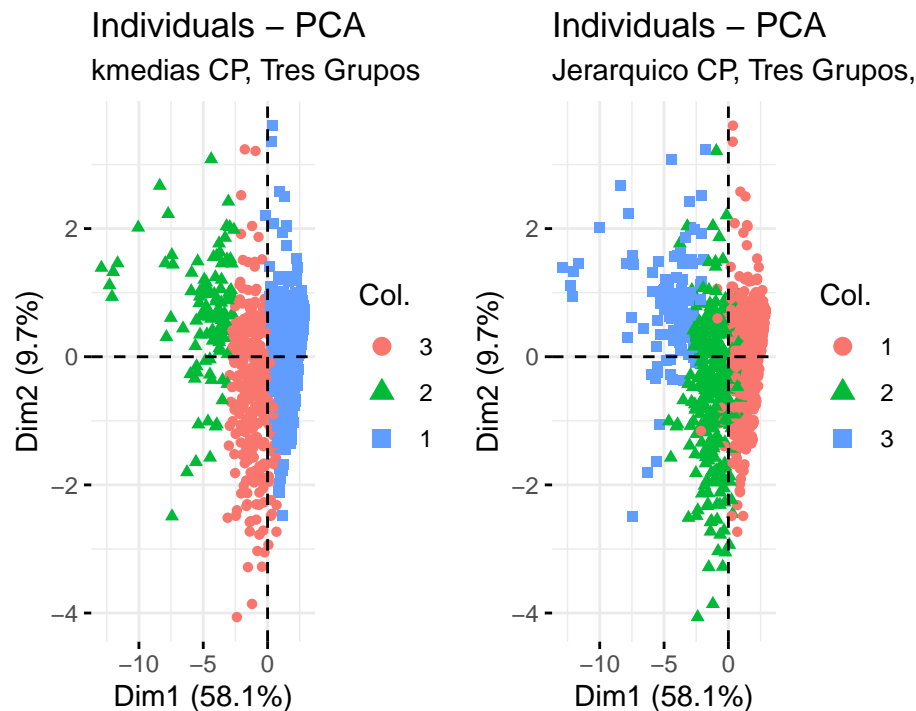
```
## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  2.4101  0.98298  0.79440  0.75653  0.70253  0.65595  0.58946
## Proportion of Variance 0.5808  0.09663  0.06311  0.05723  0.04935  0.04303  0.03475
## Cumulative Proportion 0.5808  0.67747  0.74057  0.79781  0.84716  0.89019  0.92493
##          PC8      PC9      PC10
```

```
## Standard deviation      0.57595 0.47730 0.43717
## Proportion of Variance 0.03317 0.02278 0.01911
## Cumulative Proportion  0.95811 0.98089 1.00000
```

Primero vamos a hacer el proceso de K-means con las 4 componentes principales que se escogieron.

Análogamente a los ejercicios anteriores vamos a probar usando clusters jerarquicos y conservando las disimilaridades que se usaron en el caso anterior

Obtuvimos que los mejores modelos a usar para 3 clusters fueron Euclidean, Minkowski y Ward D2.



## Conclusiones

Como pudimos ver a lo largo de todo este análisis y aplicando distintos métodos de evaluación como fue usar K-means, algoritmos de jerarquía y componentes principales decidimos conservar el de Componentes principales ya que además de permitirnos conservar las variables que conservan mayor información dadas las originales y así poder reducir el estudio a estas los resultados obtenidos fueron mas cercanos a lo que deseamos, por ejemplo, la clusterización que obtuvimos con la primer componente fue mejor, lo mismo paso para la segunda componente. Hablando en términos mas generales tenemos que el primer grupo tiene mayor promedio en todas las respuestas, seguido por el segundo grupo y por ultimo se queda el tercer grupo.

Finalmente, creemos que el modelo a utilizar para focalizar la publicidad al publico siempre dependerá en gran medida de el numero de la cantidad de publico que quiera alcanzar la empresa y conforme a esto lanzar los distintos tipos de publicidad, ya que nosotros concluimos en tomar 3 clasificaciones sobre 2 o 4, esto con el fin de mantener un equilibrio entre las preferencias de todos los clientes que buscan seguridad, puntualidad y un buen trato por parte de los trabajadores, cosas que sin duda son fundamentales para que la empresa logre atraer nuevos clientes potenciales que le den una gran importancia a estos criterios.