

ANÁLISIS DE LA INFRAESTRUCTURA EN SALUD DE LAS PRINCIPALES CIUDADES DEL PAÍS

Javier Carrillo Martínez

Aleidali Macías Ángeles

19 de agosto de 2020

Importancia y Objetivo del Análisis

La Salud juega un papel importante para lograr el bienestar humano. Una muestra de esto, son los recientes sucesos relacionados a la pandemia por COVID-19. La falta de equipo médico, unidades de consulta e incluso personal de la salud, ha evidenciado una escasa infraestructura en ciertas ciudades, lo cual ha traído consigo un incremento en la velocidad de contagios, mayores pérdidas humanas y, en consecuencia, el deterioro de la economía mexicana.

Con base en lo anterior, resulta interesante realizar un estudio sobre la infraestructura en salud de las distintas ciudades del país, de tal forma que se cuente con una herramienta que agrupe a las ciudades de acuerdo a su nivel de infraestructura en salud y se alerte a las autoridades correspondientes sobre los grupos con mayor vulnerabilidad.

El presente estudio basa su desarrollo en las técnicas estadísticas de Análisis de Componentes Principales, Escalamiento Multidimensional y Clusterización Jerárquica.

Metodología y justificación

Fase 1

La primer fase para poder llevar a cabo el presente análisis, fue recopilar las información sobre la infraestructura de salud por ciudad. En este caso se trabajó con la base de Recursos por clínicas públicas a nivel localidad que pone a disposición la Secretaría de Salud de forma anual.

La base de datos más recientes corresponden al año 2018 y cuenta con el registro de 22,544 clínicas públicas en todo el país, las cuales son caracterizadas a través de 233 variables.

De las 233 variables se eliminaron las que correspondían a totales generales o a valores dicotómicos y se conservan 174, las cuales miden distintos aspectos de infraestructura médica por especialidad médica, por ejemplo, el Número de quirófanos, el Número de equipos de resonancia magnética, el Número de médicos, consultorios y camas por especialidad, etc.

Posteriormente, se agrupó la información por municipio, dando una total de 2,057 municipios en todo el país.

Por otra parte, para escoger las ciudades más importantes del país se utilizaron dos criterios:

- **Base 1.** Selección de los 50 municipios o alcaldías con mayor densidad poblacional del país.
- **Base 2.** Selección del top 2 de municipios con mayor densidad poblacional por Estado, obteniendo así 64 municipios o alcaldías del país.

Después de realizar el pegado correspondiente de la base de Recursos de salud a la Base 1 y Base 2, se consideró conveniente que las variables estuvieran medidas en unidades por cada 10 mil habitantes para evitar que ciudades más grandes dieran la impresión de que están mejor equipadas que las ciudades pequeñas, por lo tanto, se dividieron los datos entre la población total por municipio y se multiplicó por 10,000. Obteniendo una base de datos en unidades por cada 10,000 habitantes, por ejemplo: Número de quirófanos por 10,000 habitantes.

De las 174 variables se observó que algunas representan características muy específicas de una clínica, de modo que muchas observaciones presentan ceros unidades en estas columnas, para encontrar una mejor representación de los datos se realizó una reducción de dimensiones mediante la técnica de Componentes Principales (PCA) sobre la base de datos estandarizada ya que las magnitudes entre las variables eran variadas.

En la figura 1 se observa que con a partir de 30 componentes principales se logrará explicar más del 90 %, en este caso se utilizaron los primeros 50 componentes que explican el 92 % de la varianza total.

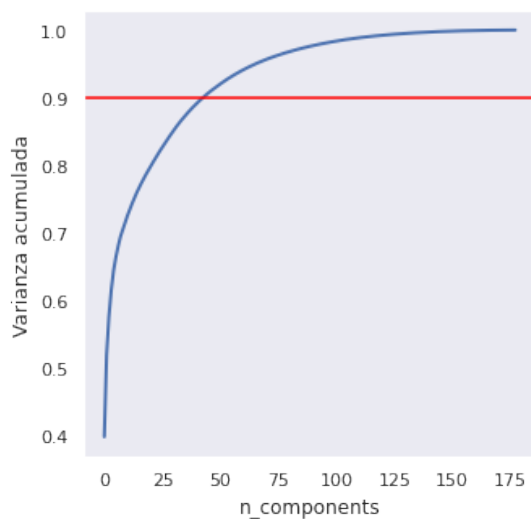


Figura 1: Varianza explicada por número de componentes principales.

Fase 2

Debido a que la intención de este análisis es agrupar a las principales ciudades de acuerdo a su infraestructura en salud, la técnica de Escalamiento Multidimensional(MDS) es una buena herramienta para lograrlo, ya que permite segmentar las observaciones con base en las disimilaridades de las variables en cada observación, de manera que se obtiene una representación gráfica de los grupos en baja dimensión conservando la mayor información posible.

En este caso, las distancias utilizadas para construir el modelo MDS se calcularon mediante cuatro distintas métricas: Euclideana, Manhattan, Canberra y Bray Curtis, posteriormente se evalúan con el Modelo MDS con enfoque Clásico y el Modelo MDS con enfoque de Mínimos Cuadrados. Finalmente se selecciona el modelo que proporciona la mayor varianza explicada o el menor estrés, según sea el caso. En la sección Resultados el lector encontrará una comparación del desempeño de las distancias y los distintos análisis MDS a fin de formar un criterio sobre el modelo final a utilizar.

Fase 3

Después de haber reducido la dimensión de los datos mediante PCA y posteriormente haber aplicado MDS se obtienen las *configuraciones*, la última etapa consiste en definir grupos mediante distintos métodos de clusterización sobre la base de datos en su representación en sus 50 componentes princi-

pales, la respectiva matriz de distancias y sobre las configuraciones MDS.

En la sección de resultados se expone con detalle los resultados obtenido de este análisis y en la sección de Interpretación se comparan los tipos de clusterización visualizándolos en la representación de configuraciones MDS.

Resultados

Para calcular las matrices de distancias de la Base 1 (50 observaciones) y la Base 2 (64 observaciones) a través de sus 50 componentes principales, se utilizó la función *pairwise distances* de la paquetería de *sklearn metrics*. Las distancias calculadas fueron la Euclidiana, Manhattan, Canberra y Bray Curtis.

A partir de estas matrices, se aplicó el modelo MDS Clásico en dos dimensiones utilizando la función *mdscale()* del lenguaje R y se obtuvieron los porcentajes de varianza explicada para cada tipo de distancia. En el Cuadro 1 y Cuadro 2 se puede apreciar que tanto para la Base 1 como para la Base 2 la distancia Euclideana presentó el mejor performance, sin embargo, en ambos casos el porcentaje fue menor al 80 %, lo cual se considera un ajuste de regular.

Para explorar el número de dimensiones necesarias para lograr al menos el 90 % de la varianza explicada se realiza la gráfica de varianza acumulada respecto al número de dimensiones, como se muestra en la figura 2 y 3, se observa que para la Base 1 al menos se necesitan 4 dimensiones y para la Base 2 se necesitan al menos 5 dimensiones, con lo la visualización sería muy complicada, por lo que se decide probar el Modelo MDS por Mínimos Cuadrados.

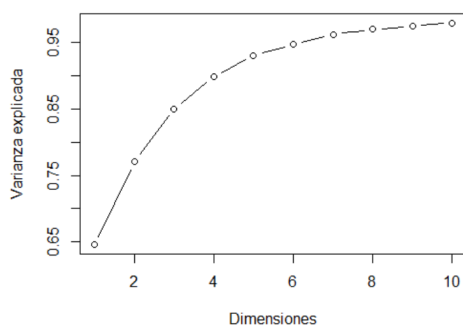


Figura 2: Base 1. Varianza explicada por número de dimensiones. Modelo Clásico distancia Euclidiana

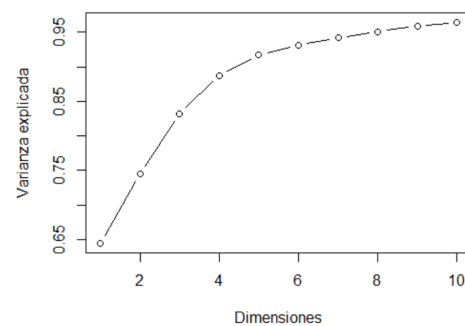


Figura 3: Base 2. Varianza explicada por número de dimensiones. Modelo Clásico distancia Euclidiana

Distancia	MDS clásico % Varianza Acumulada	MDS SMACOF % estrés-1	MDS SMACOF % estrés-1 transf. ordinal
Euclidiana	77.02	9.42	4.52
Manhattan	47.59	14.5	6.5
Canberra	26.45	32.66	18.19
Bray Curtis	27.75	28.4	21.82

Cuadro 1: Performance Base 1

Distancia	MDS clásico % Varianza Acumulada	MDS SMACOF estrés-1	MDS SMACOF estrés-1 transf. ordinal	MDS SMACOF estrés-1 transf. spline
Euclidiana	74.28	10.12	0.91	5.76
Manhattan	44.3	14.71	0.7	7.93
Canberra	20.97	33.46	21.02	21.54
Bray Curtis	22.34	28.61	21.44	22.23

Cuadro 2: Performance Base 2

Los modelos por Mínimos Cuadrados se obtuvieron mediante la función `mds()` de la librería `SMA-COF`, nativa del lenguaje R. Se obtienen los modelos en dos dimensiones para cada tipo de distancia y se elige el que presente el menor porcentaje de estrés-1.

Como se puede observar en el Cuadro 1 y Cuadro 2, los modelos con menor porcentaje de estrés son con la distancia Euclidea. No obstante, a este tipo de modelos es posible aplicar transformaciones sobre las disimilaridades con la finalidad de mejorar el ajuste.

A continuación se muestra el nivel de estrés de acuerdo al número de dimensiones para cada tipo de transformación para la Base 1 y Base 2 (usando distancia Euclidea). La transformación que mejor minimiza el estrés es la Transformación Ordinal.

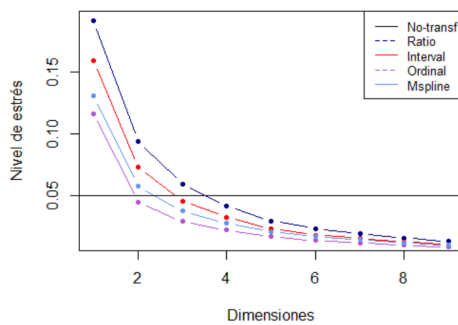


Figura 4: Base 1. Nivel de estrés para distintas transformaciones MDS por Mínimos Cuadrados distancia euclidea.

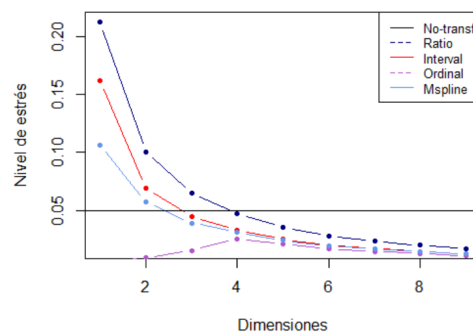


Figura 5: Base 2. Nivel de estrés para distintas transformaciones MDS por Mínimos Cuadrados distancia euclidea.

En las tablas y gráficos anteriores se puede observar que la distancia Euclidea es la que presenta valores de varianza acumulada y estrés-1 más convenientes y de igual forma se observa que al realizar

el escalamiento multidimensional con transformación ordinal se obtiene un menor valor de estrés, de modo que en adelante sólo se muestran los resultados asociados a esta distancia y análisis de MDS.

A continuación se analizan las dispersiones de las configuraciones obtenidas para cada grupo de ciudades.

Base 1: 50 ciudades más densamente pobladas

Los resultados del MDS por mínimos cuadrados con transformación ordinal usando la métrica euclidiana se muestran en la figura 6 a), se puede observar que la mayoría de datos se encuentran agrupados y que hay al menos 5 datos que se encuentran más alejados de este grupo de forma dispersa. Para una mejor visualización del grupo que contiene a la mayoría de los datos se quitan los datos Cuauhtémoc, Azcapotzalco y Miguel Hidalgo (3, 9 y 17) y se muestran en la figura 6 b).

Cabe señalar que la razón por la cual algunos puntos se encuentran muy separados del resto tiene una razón de ser, esto y el significado de cada dimensión se abordarán en la siguiente sección.

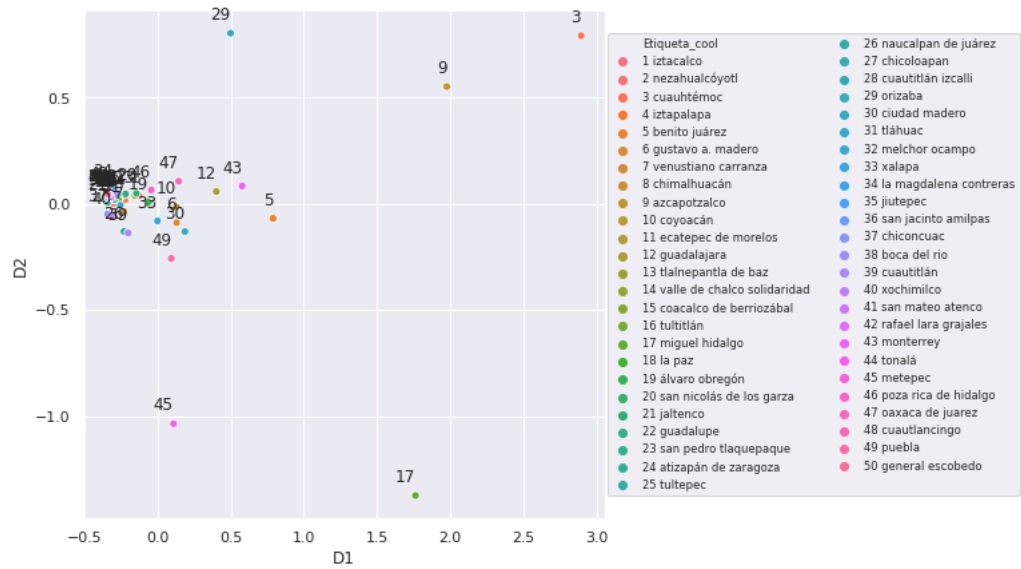
Posteriormente, a fin de encontrar agrupaciones interesantes, se realizaron tres métodos de clustering, espectral, jerárquico y K-means, aplicado a: las configuraciones obtenidas por escalamiento multidimensional (Figura 7), los datos originales (50 componentes principales) (Figura 8) y la matriz de distancias euclidianas (Figura 9).

Para poder visualizarlos se utilizó la representación bidimensional obtenida por MDS mediante mínimos cuadrados.

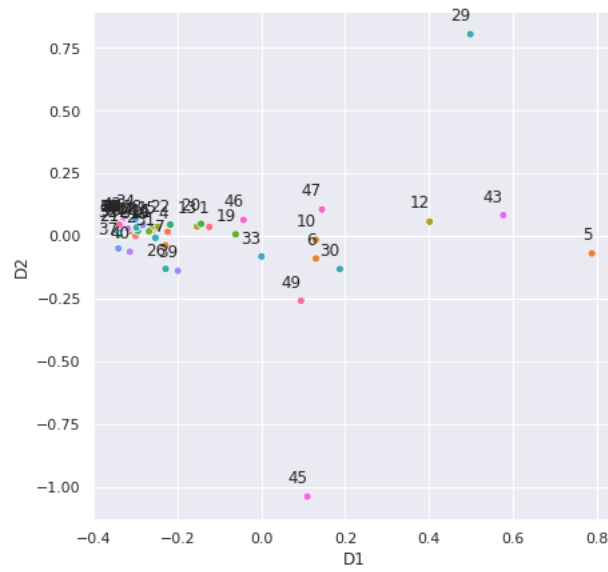
En la figura 7 se pueden observar que el agrupamiento por clustering jerárquico b) y K-means c) presentan algunas diferencias, pues la clase azul en el jerárquico contiene un sólo punto mientras que la equivalente en K-means tiene 5.

Obsérvese que en ambos casos la clasificación parece realizarse prácticamente sobre los valores de D1, siendo la diferencia en estos casos los intervalos de cada clase, en el jerárquico estos se encuentran ligeramente más cargados a la izquierda que en K-means. En esto es distinta la agrupación realizada con clustering espectral a) que muestra los valores 47, 10 y 6 con un valor muy similar en su componente D1, pero distinto D2 y distinta clasificación, el punto 47 pertenece a la clase azul, mientras los otros dos a la clase amarilla.

En la figura 8 se puede observar que el agrupamiento por clustering jerárquico b) y K-means c) son prácticamente el mismo, salvo el punto 1. A su vez también es remarcable que la agrupación sobre las proyecciones es la misma que al usar K-means sobre la configuración MDS, lo cual quiere decir



a)



b)

Figura 6: Configuración bidimensional resultante de un escalamiento multidimensional por mínimos cuadrados y transformación ordinal, para las distancias euclidianas de la base 1.

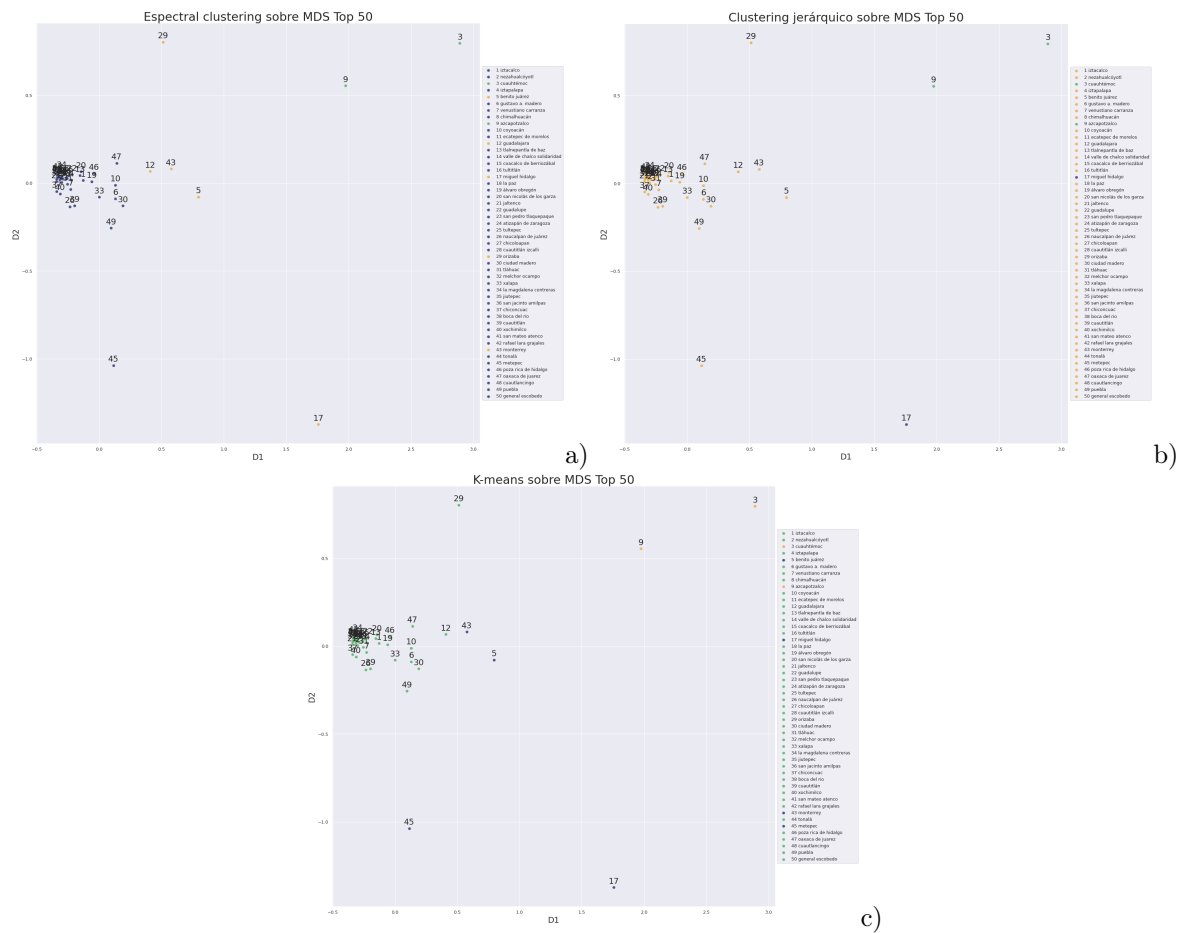


Figura 7: Diferentes métodos de clustering sobre las configuraciones a) Espectral, b) Jerárquico y c) Kmeans.

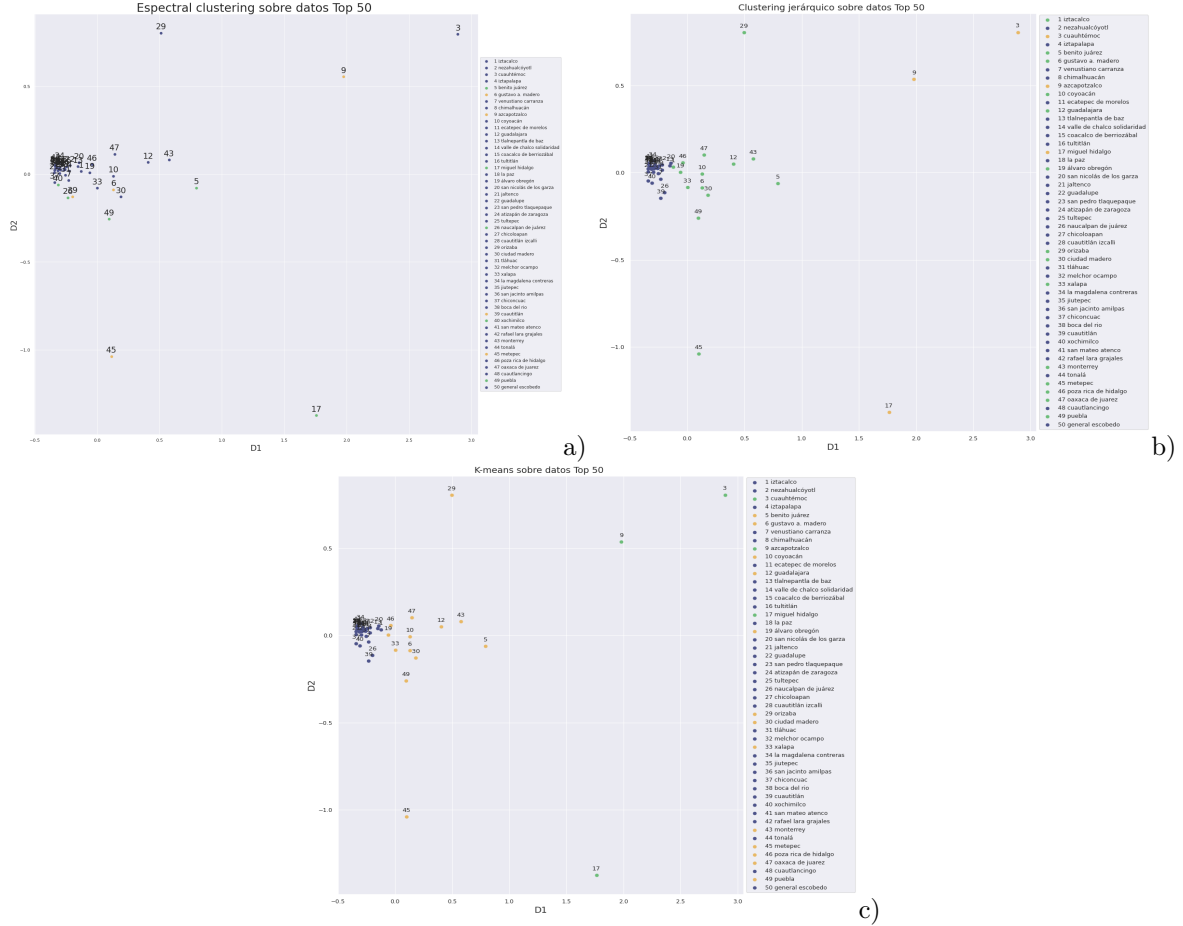


Figura 8: Diferentes métodos de clustering sobre la proyección en compontes principales a) Espectral, b) Jerárquico y c) Kmeans.

que la agrupación sobre la configuración MDS es intrínseca a los datos, más que a la representación. Estas observaciones no coinciden con el agrupamiento por clustering espectral que al observarlas en esta representación se muestran muy mezcladas las clases.

Por último, en la figura 9 se pueden observar que el agrupamiento por clustering jerárquico b) y K-means c) es exáctamente el mismo, que a su vez es prácticamente igual al agrupamiento por K-means de los dos conjuntos anteriores, siendo la única diferencia los puntos 46, 19 y 33, lo que refuerza la idea de que el agrupamiento es algo intrínseco a los datos y no a la representación. Por otro lado, observamos que el clustering espectral nuevamente muestra clases muy mezcladas, de modo que podemos concluir que este método no es el adecuado para el agrupamiento.

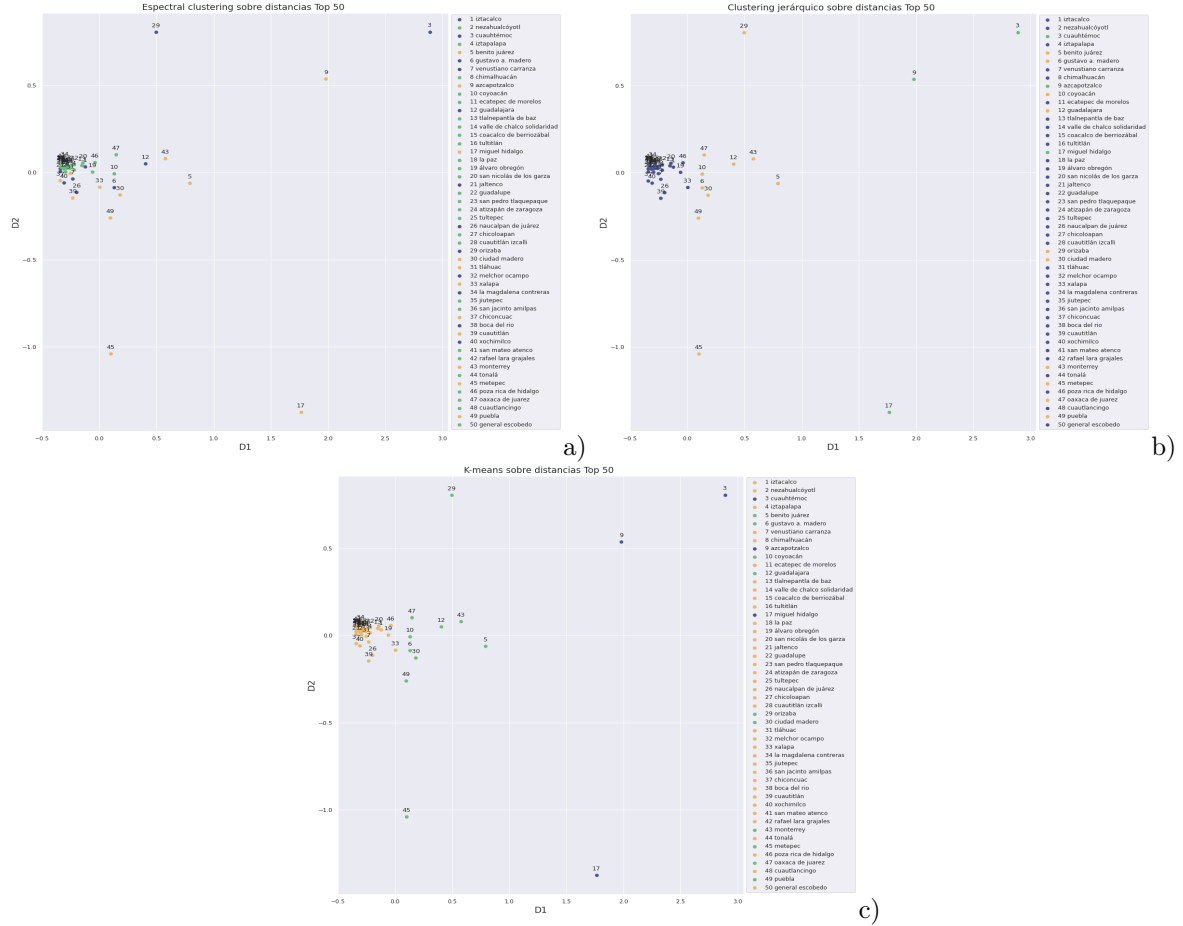


Figura 9: Diferentes métodos de clustering sobre las distancias euclidianas a) Espectral, b) Jerárquico y c) K-means, representadas sobre la configuración obtenida por MDS.

Top 2 ciudades más densamente pobladas por estado

En la figura 3 se observa que usando la base 2 para recuperar algo cercano al 90 % de varianza acumulada con un MDS clásico se requieren al menos cuatro componentes, y como lo muestran la tabla 2 y la figura 5 el MDS por mínimos cuadrados con menor stress asociado es el realizado con una transformación ordinal y distancia Manhattan, de modo que se realizó el análisis con esta distancia y modelo por ser el más conveniente para la base 2.

En la figura 10 a) se puede observar que el punto 52 asociado a la delegación Cuauhtémoc está diametralmente opuesto al resto de los municipios, de modo que para visualizar la observación e interpretación removeremos este punto, obteniendo la figura 10 b) donde notamos que la mayoría de los puntos se aglomera en un punto con valores negativos de D1 y positivos cercanos a cero en la dimensión D2, salvo unos puntos que se encuentran dispersos al rededor del mayor conglomerado como el 1, 5 y 8 que corresponden a Zacatecas, Orizaba y Tlaxcala.

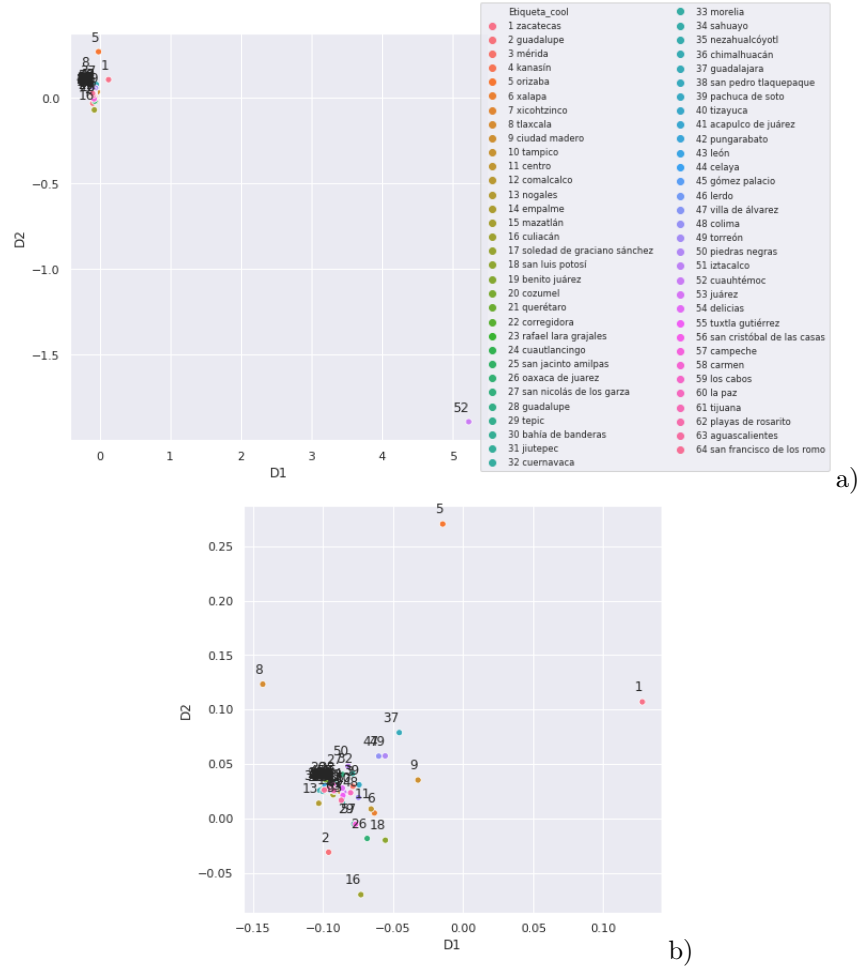


Figura 10: Configuración bidimensional resultante de un escalamiento multidimensional por mínimos cuadrados y transformación ordinal, para las distancias manhattan de la base 2.

De acuerdo al análisis realizado sobre los métodos de clusterización en la base 1 se decidió presentar sólo el análisis de agrupamiento mediante K-means con 3 Clusters, este método se utilizó sobre la representación MDS, la proyección de los datos en sus componentes principales y la matriz de distancias calculadas mediante la métrica Manhattan, los resultados se muestran en la figura 11.

En esta figura se puede observar que el agrupamiento sobre las configuraciones MDS, la proyección de los datos y las distancias es similar, sin embargo presenta algunas diferencias interesantes, como el punto 8 correspondiente a Tlaxcala que en los tres casos se agrupa distinto, a excepción de este punto se puede notar que la agrupación sobre la representación MDS se realiza sobre su valor en D2, mientras que sobre los datos se agrupan conforme a una combinación de su posición en D1 y D2, destacando que marca como una sola clase al grupo más colapsado de puntos y las otras dos clases se pueden describir como las que se encuentran por debajo de 0.1 y por encima de este valor.

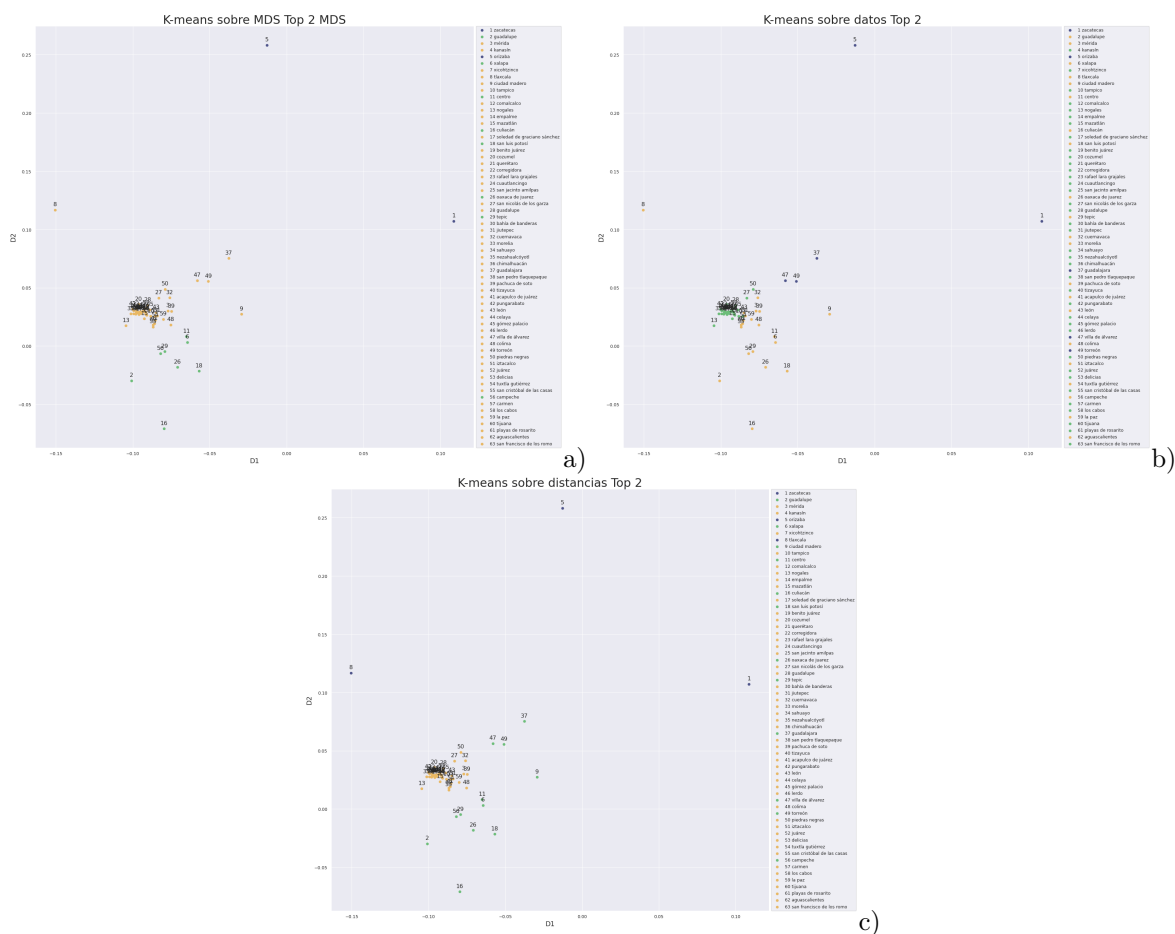


Figura 11: Clustering mediante 3 clusters K-means sobre las configuraciones MDS a), proyección de la base 2 en sus 50 componentes principales b) y distancias manhattan c), representadas sobre las configuración obtenida por MDS.

Por otro lado el agrupamiento sobre la matriz de distancias parece estar representando un agrupamiento respecto a las distancias de los puntos respecto al cúmulo de puntos más denso, este tipo de agrupación tiene sentido si pensamos que hay por ejemplo un número bajo de D2 compensa un bajo valor de D1 y viceversa.

Interpretación y conclusiones

Después de realizar las distintas combinaciones de distancias y análisis MDS bajo el enfoque Clásico y el de Mínimos Cuadrados, se llega a la conclusión de que el modelo más conveniente para la Base 1 es un Modelo MDS en dos dimensiones por Mínimos Cuadrados con distancia euclídeana y transformación ordinal, mientras que para la Base 2 es un Modelo MDS en dos dimensiones por Mínimos Cuadrados con distancia Manhattan y transformación ordinal. De acuerdo a lo interpretado de las tablas 1 y 2.

Recordando que el análisis de escalamiento multidimensional se realizó sobre la proyección en componentes principales de los datos, la interpretación de las dos dimensiones de la representación mostrada en la figura 6 se complica un poco, de modo que se realizó un mapa de calor mostrado en la figura 12 y 15 a fin de identificar correlaciones entre las configuraciones y las variables originales, para observar con qué variables está asociada cada dimensión.

Interpretación Base 1: 50 ciudades más pobladas.

Para la primer base, es posible observar en la figura 12, que la primer dimensión presenta una fuerte relación positiva a la mayoría de las variables médicas, sólo en 22 de las 178 variables se tiene una correlación positiva menor a 0.4, entre ellas se encuentra las relacionadas a la labor de parto, los consultorios de Valoración y Medicina General, así como algunas especialidades tales como la Proctología, Psiquiatría, Traumatología, Radioterapia y Transplante de médula ósea. Por lo tanto, para la primer dimensión, entre mayor sea el valor o más a la derecha esté en el mapa, mayor infraestructura por habitante tiene esa ciudad.

Asimismo, en la figura 12 se identifica que para la dimensión 2 existe correlación negativa (-0.6 en promedio) con variables como camas para quemados, camas en unidades de transplante de médula ósea, camas de Gineco-Obstetricia, etc. Esto sugiere que entre mayor sea el valor en la dimensión, el municipio tendrá mayores deficiencias en las variables antes mencionadas. Por lo tanto, valores negativos de esta dimensión, presentan mejor infraestructura de salud.

Esto nos permite identificar que el caso ideal se encuentra en la esquina inferior derecha y el peor de los casos sería un punto en la esquina superior izquierda.

Cabe señalar que, de acuerdo al Modelo Clásico, la primer dimensión aporta el 65 % y la segunda el 12 %, de modo que es consistente con lo observado en la figura 12 sobre la mayor relevancia de

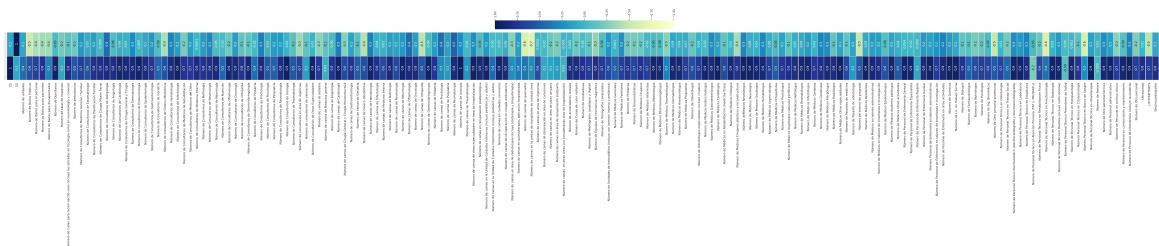


Figura 12: Correlaciones Configuraciones vs. variables originales

la dimensión uno, respecto a las dos. De modo que la posición sobre la dimensión 1 nos da mayor información sobre la infraestructura en salud del municipio.

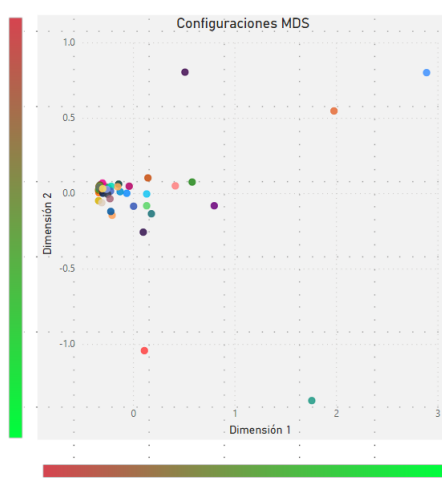


Figura 13: Mapa intepretativo basado en las configuraciones MDS.

Como se pudo apreciar en la sección de resultados para la base 1, las gráficas de las configuraciones MDS prsentan ciertas observaciones muy sobresalientes sobre el resto, tal el es caso de los municipios Cuauhtémoc, Azcapotzalco y Miguel Hidalgo. Se revisaron los valores de estas 3 observaciones y se observó que sus variables de infraestructra médica por cada 10 mil habitantes era muy superior con respecto al resto de los 50 municipios de mayor densidad poblacional, en la figura 14 podemos verificar que estos 3 minicipios encabezan la lista en el mayor número de camas, consultorios, médicos, enfermeras y demás personal, por cada 10 mil habitantes.

Interpretación Base 2: Top 2 ciudades más densamente poblabas por estado.

Para la segunda base es posible observar en la figura 15, que al igual que la base 1, la primer dimensión presenta una fuerte relación positiva a la mayoría de las variables médicas, sin embargo, en ésta se castiga mucho más el hecho de no contar con la infraestructura relacionada a la labor de

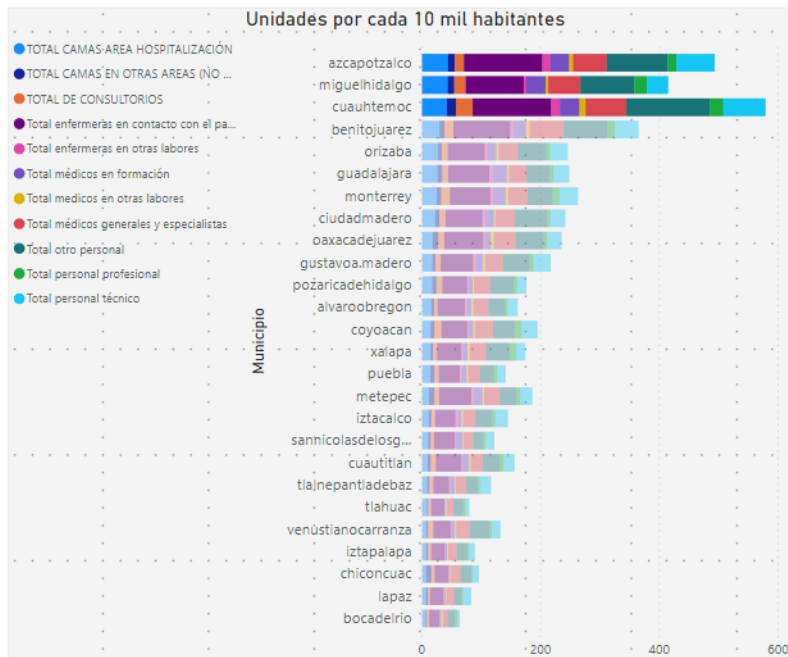


Figura 14: Comparativo total de unidades por cada 10 mil habitantes.

parto, los consultorios de Valoración y Medicina General, así como algunas especialidades tales como la Proctología, Psiquiatría, Traumatología, Radioterapia y Transplante de médula ósea, debido a que se observan correlaciones negativas.

Por lo tanto, para la primer dimensión, entre mayor sea el valor o más a la derecha esté en el mapa, mayor infraestructura por habitante tiene esa ciudad.

Además, en la figura 15 se puede observar que las variables que presentan correlación positiva fuerte con la componente D1 tienen asociada una correlación negativa de aproximadamente la misma magnitud en D2. De modo que un mayor valor en D1 es señal de una mejor infraestructura en salud, mientras que un menor valor en D2 es el reflejo de esta situación.

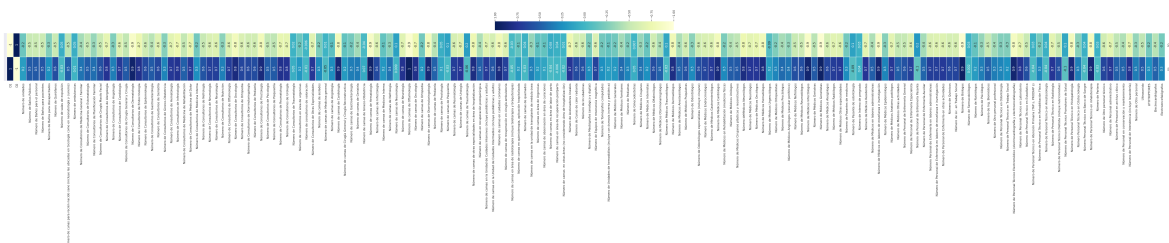


Figura 15: Correlaciones Configuraciones vs. variables originales

Al igual que en con la base 1, el caso ideal está ubicado en la esquina inferior derecha, mientras el peor de los casos se encuentra en la esquina superior izquierda, con la única diferencia de que en este

caso, ambas componentes parecen tener la misma relevancia.

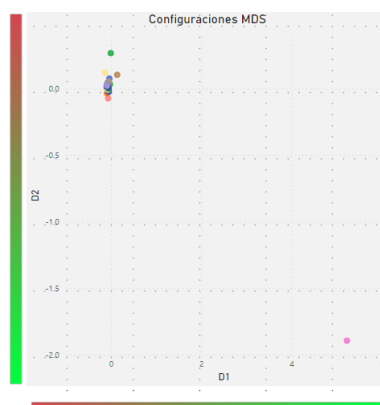


Figura 16: Mapa interpretativo basado en las configuraciones MDS.

Por último, en la figura 16 se observa un punto muy extremo que corresponde al municipio Cuauhtémoc de la Ciudad de México, no obstante se revisaron los valores de sus variables de infraestructura médica por cada 10 mil habitantes y como se puede observar en la figura 17 se encuentra muy por encima del resto de municipios en número de camas, consultorios, médicos, enfermeras y demás personal.

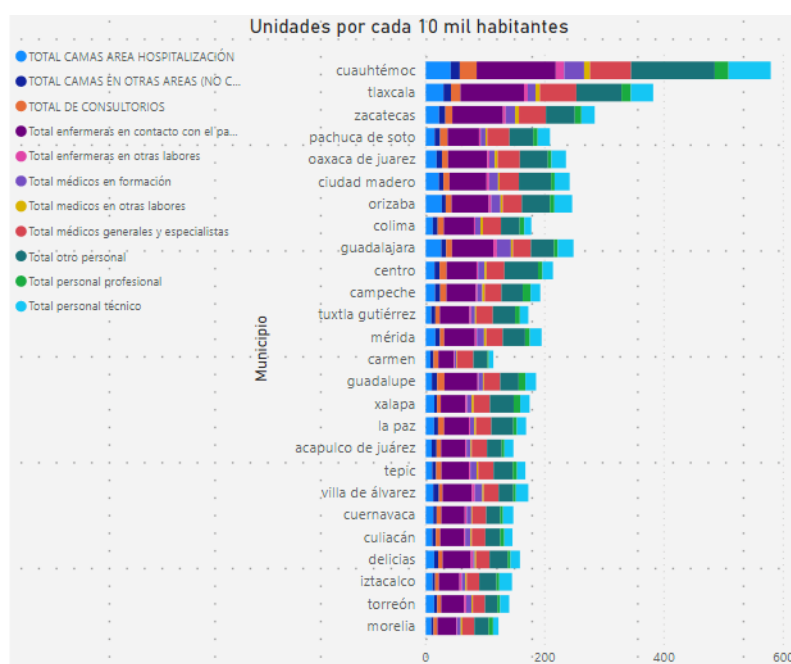


Figura 17: Comparativo total de unidades por cada 10 mil habitantes.

Por lo anterior, el hecho de tener un valor muy lejano hace que la interpretación de los clusters obtenidos en la figura 11 sean más difíciles de interpretar. La opción más viables es formar dos clusters,

la alcaldía Cuauhtémoc y el resto de los municipios, de manera que se evidencía la desigualdad tan grande entre los municipios más densos por estado, lo cual no era tan evidente en el análisis de la base 1.

Conclusiones

Tras observar el análisis realizado para la base de los 50 municipios más densamente poblados y comparando con los 2 más densamente poblados por estado se pueden observar dos puntos principales:

- La alcaldía Cuauhtémoc de la Ciudad de México cuenta con la mejor infraestructura en salud a nivel país.
- Es evidente que existe una gran centralización de la infraestructura en salud.

Revisando las clínicas y hospitales que se encuentran en la alcaldía Cuauhtémoc, cobra sentido que sea un punto muy importante, pues entre estos se encuentra el Hospital General Gregorio Salas y el Centro Médico Siglo XXI, los cuales son dos de los hospitales más importantes del país por la cantidad de personas que reciben y la cantidad de especialidades que practica.

Una posible línea de investigación como trabajo a futuro es realzar el análisis sin considerar las alcaldías que presentan un comportamiento muy superior al resto, esto con la finalidad de encontrar más información sobre la infraestructura en el resto del país.

Por otro lado, la centralización de infraestructura en el país es un tema socialmente conocido, y este análisis brinda argumentos cuantitativos sólidos para determinar qué tan centralizada está la infraestructura en salud, si revisamos los 10 municipios con mejor infraestructura basándonos en la componente D1 del análisis de la base 1 como se muestra en la tabla 3 podemos observar que la mitad de estos están en la Ciudad de México, en particular los 4 primeros se encuentran en este lugar.

Estado	Municipio
Ciudad de México	Cuauhtémoc
Ciudad de México	Azcapotzalco
Ciudad de México	Miguel Hidalgo
Ciudad de México	Benito Juárez
Nuevo León	Monterrey
Veracruz	Orizaba
Jalisco	Guadalajara
Tamaulipas	Ciudad Madero
Oaxaca	Oaxaca de Juárez
Ciudad de México	Coyoacán

Cuadro 3: 10 municipios con mejor infraestructura en salud de acuerdo a la componente D1 del análisis MDS de la base 1.

Bibliografía

- Daniel Peña (2002). Análisis de datos multivariantes. McGraw Hill.
- Johnson, R.A. y Wichern, D.W. (2007) . Applied multivariate statistical analysis, 6th Ed. Prentice Hall.
- Distance, Similarity, and Multidimensional Scaling Recuperado el 2 de Junio de 2020 de: <https://pages.mtu.edu/~shanem/psy5220/daily/Day16/MDS.html>
- Recursos en Salud 2018 Recuperado el 25 de Mayo de 2020 de: <https://datos.gob.mx/busca/dataset/recursos-en-salud-nivel-central>
- Datos de densidad de población por municipio 2018 Recuperado el 25 de Mayo de 2020 de: <https://www.inegi.org.mx/app/indicadores/1>
- Datos de población por municipio 2018 Recuperado el 25 de Mayo de 2020 de: <http://cuentame.inegi.org.mx/monografias/informacion/nl/poblacion/default.aspx?tema=m>