

# Prestadores: Prueba Técnica

Jorge Luis Aquino Olmos

2024-04-06

- 1. Carga de los datos
- 2. Análisis Descriptivo General
- 3. Municipios
- 4. Relación entre la Cantidad de Prestadores y Características Demográficas de los Municipios
- 5. Análisis de Clúster
- 5.1 Limpieza y preparación de los datos
- 6. Agrupamiento por K-Medios (K-Means Clustering)
- Análisis conjunto prestadores y departamentos

## 1. Carga de los datos

Carga de los datos: Se ha encontrado un problema al intentar cargar el archivo `prestadores.csv` debido a un error de codificación, que suele surgir cuando el archivo contiene caracteres que no están en el formato UTF-8. Para resolver este inconveniente, se adoptó la estrategia de cargar el archivo especificando una codificación diferente, como 'ISO-8859-1' o 'latin1', que son comunes en archivos que incluyen caracteres especiales no presentes en el alfabeto inglés. Esta acción permitió sortear el obstáculo inicial relacionado con la codificación del archivo, asegurando que los caracteres especiales se interpreten correctamente durante el proceso de carga.

Posteriormente, se procedió a superar desafíos adicionales relacionados con la codificación y delimitación del archivo `prestadores.csv`. Específicamente, se estableció la codificación como 'ISO-8859-1' y se especificó el delimitador como ';' para asegurar una interpretación precisa de los datos. Este enfoque permitió una carga exitosa del archivo, garantizando que los caracteres especiales se manejen adecuadamente y que la estructura tabular de los datos se preserve correctamente. Con estos ajustes, se allanó el camino para realizar un análisis exploratorio y descriptivo más detallado de los datos contenidos en el archivo `prestadores.csv`.

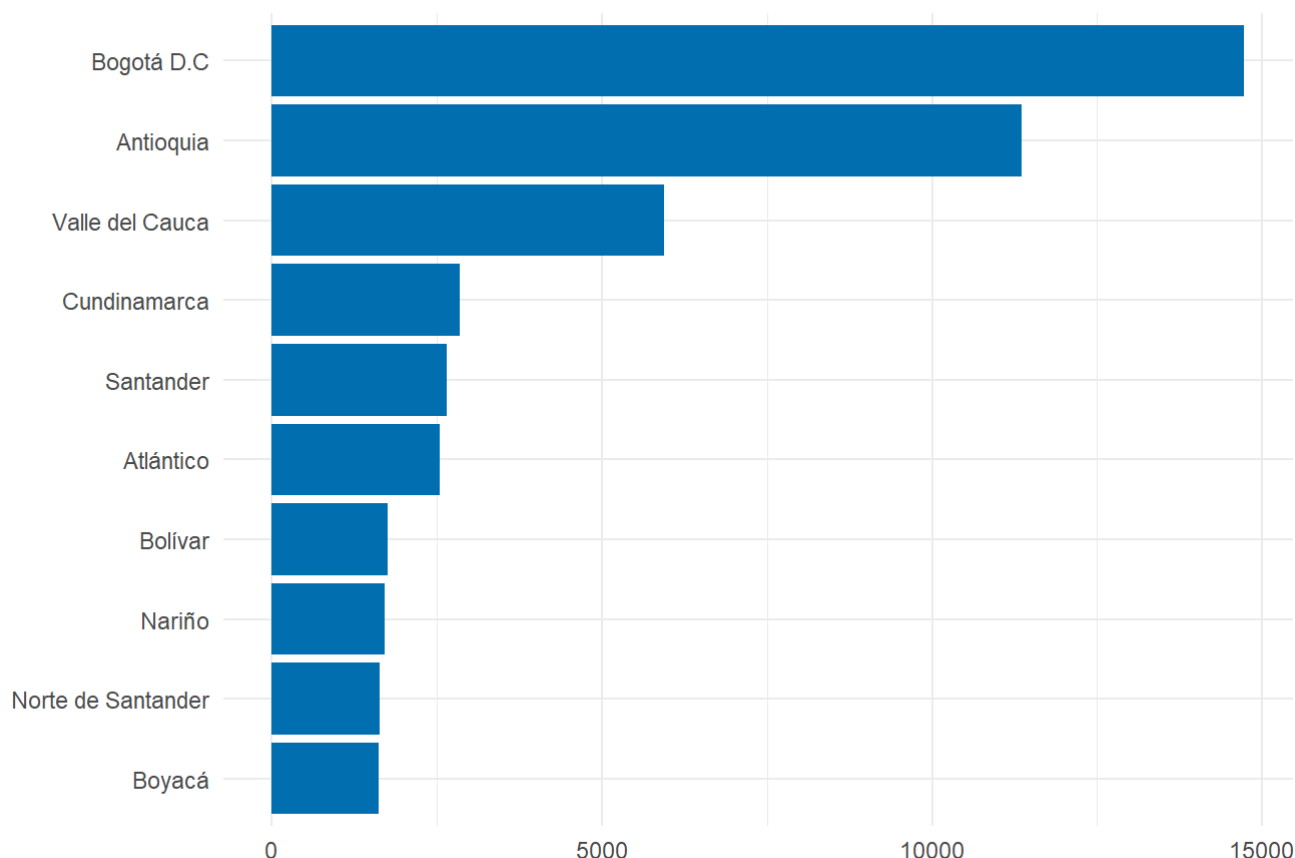
## 2. Análisis Descriptivo General

Comenzaremos con un análisis descriptivo general de ambos datasets para tener una visión global de los datos. Dataset `prestadores`:

El dataset de prestadores de servicios contiene información diversa sobre 60,946 registros únicos, distribuidos en 38 departamentos y 914 municipios diferentes. Algunas observaciones clave incluyen:

La mayoría de los registros provienen de Bogotá D.C., lo que indica una alta concentración de prestadores en la capital del país. Hay una amplia variedad de prestadores, con 58,562 nombres únicos registrados, aunque algunos nombres se repiten, lo que sugiere la presencia de entidades con múltiples registros o sucursales. La columna `clpr_nombre` (clase de prestador) muestra que hay 4 categorías únicas, con 'Profesional Independiente' siendo probablemente la más común. La naturaleza jurídica (`naju_nombre`) y el nivel de atención (`nivel`) son otros aspectos importantes registrados. Notablemente, hay 58,711 NITs únicos y 57,270 razones sociales únicas, lo que indica una diversidad significativa en los tipos de prestadores de servicios. Se observa que la columna `ese` (que parece indicar si son entidades sin ánimo de lucro o algo similar) tiene solo dos valores únicos, pero solo está presente en 10,704 registros.

Prestadores de servicios por departamento



### 3. Municipios

La distribución geográfica y demográfica de los municipios colombianos es un reflejo de la variedad de paisajes y poblaciones que conforman el país, desde las zonas urbanas densamente pobladas hasta las regiones rurales dispersas. En este sentido, nos sumergimos en el estudio de características clave como la población, la superficie y el índice de ruralidad de cada municipio, reconociendo la importancia de estos datos para la planificación y el desarrollo equitativo a nivel local y nacional.

Por ello no podemos pasar por alto la relevancia de los prestadores de servicios en este análisis. La distribución de estos proveedores en los diferentes municipios no solo refleja las dinámicas económicas y sociales de cada región, sino que también tiene un impacto significativo en el acceso y la calidad de los servicios disponibles para la población. Por lo tanto, al explorar la diversidad geográfica y demográfica de Colombia, también nos enfocamos en entender la distribución de los prestadores de servicios y su relación con las características locales.

(Top 10) Población y superficie por Departamento

Departamento	Población	Superficie
Bogotá D.C	7592871	1622.853
Antioquia	6550206	62804.709
Valle del Cauca	4506768	20665.545
Cundinamarca	3085522	22370.489
Atlántico	2638151	3315.752

### (Top 10) Población y superficie por Departamento

Departamento	Población	Superficie
Santander	2237587	30561.498
Bolívar	2120925	26302.810
Córdoba	1741111	24454.516
Nariño	1628981	31497.573
Norte de Santander	1565362	21856.754

## 4. Relación entre la Cantidad de Prestadores y Características Demográficas de los Municipios

Para explorar la relación entre la cantidad de prestadores de servicios y las características demográficas de los municipios (población y superficie), necesitaríamos combinar la información de los dos datasets. Debido a la variabilidad en la codificación de los nombres de los municipios y departamentos, este paso puede requerir una normalización de los datos para asegurar una combinación precisa.

Sin embargo, la inspección inicial sugiere que una mayor población en un municipio podría correlacionarse con una mayor cantidad de prestadores, dado el número elevado de prestadores en Bogotá D.C. Este patrón probablemente se repita en otros centros urbanos principales.

Para realizar un análisis de clusterización con los datos disponibles, primero necesitamos definir las características (variables) que utilizaremos para agrupar los datos. Dado que tenemos dos conjuntos de datos principales, podríamos considerar varias opciones:

**Prestadores de Servicios:** Utilizar características como la naturaleza jurídica, el tipo de prestador, entre otros, para agrupar los prestadores de servicios. Sin embargo, muchas de estas variables son categóricas y requerirían un preprocesamiento adecuado (por ejemplo, codificación One-Hot) antes de la clusterización.

**Municipios:** Utilizar datos demográficos y geográficos, como población, superficie, e índice de ruralidad para agrupar los municipios. Esta opción parece más directa y podría ofrecer insights interesantes sobre cómo se agrupan los municipios según sus características.

```
## Cantidad de municipios en prestadores = 914
```

```
## Cantidad de municipios en Municipios = 1030
```

```
## Cantidad de municipios en prestadores que no están en Municipios = 106
```

```
## Cantidad de municipios en Municipios que no están en prestadores = 222
```

Es evidente que existen discrepancias notables en los nombres de los municipios entre ambos conjuntos de datos. Este hallazgo subraya la importancia de abordar las discrepancias en la fuente de datos para garantizar la integridad y precisión de nuestro análisis. Tras una revisión exhaustiva, he identificado que las discrepancias en los nombres de los departamentos son considerablemente menores y, por ende, más manejables. Además, las diferencias encontradas en esta característica fueron subsanadas mediante el

procedimiento previamente realizado y las correcciones aplicadas. Por consiguiente, hemos decidido enfocarnos en la unificación de características a través de los departamentos, aprovechando su menor número y las diferencias menos significativas encontradas en este nivel de agregación.

## 5. Análisis de Clúster

Utilizaré los datos demográficos y geográficos disponibles, tales como población, superficie e índice de ruralidad, para agrupar los departamentos. Esta estrategia nos brindará valiosos insights sobre cómo se organizan los departamentos en función de sus características, permitiéndonos identificar patrones y tendencias en la agrupación departamental basados en dichos atributos. Al concluir esta agrupación, procederemos a integrar ambos datasets para explorar la relación entre la cantidad de prestadores de servicios y las características demográficas de los municipios, enriqueciendo así nuestra comprensión de la estructura demográfica y geográfica del país, así como de las dinámicas que influyen en la distribución y configuración de los departamentos. Este enfoque integral no solo iluminará aspectos clave de la distribución de servicios en el territorio sino que también facilitará una evaluación más detallada de cómo las características de cada departamento afectan la disponibilidad y el acceso a dichos servicios.

### 5.1 Limpieza y preparación de los datos

```
## 'data.frame': 1118 obs. of 4 variables:
## $ Departamento: chr "Antioquia" "Antioquia" "Antioquia" "Antioquia" ...
## $ Poblacion : int 2483545 20258 2710 4669 30777 26552 43713 5790 11462 18737 ...
## $ Superficie : chr "374.8306248" "507.1341142" "296.9559801" "128.9321532" ...
## $ Irural : int 5 45 58 48 28 54 38 37 47 58 ...
```

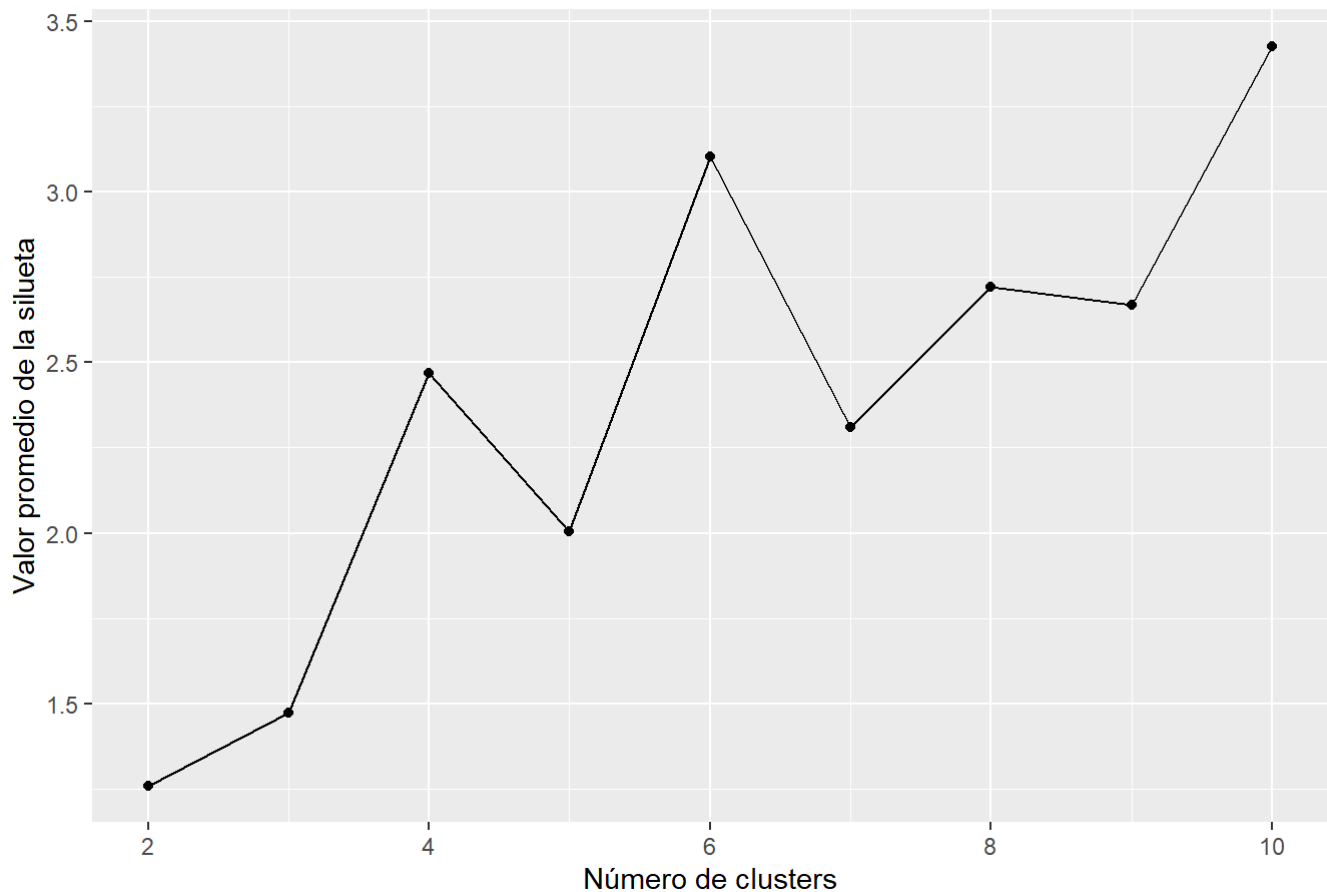
```
## 'data.frame': 1118 obs. of 4 variables:
## $ Departamento: chr "Antioquia" "Antioquia" "Antioquia" "Antioquia" ...
## $ Poblacion : int 2483545 20258 2710 4669 30777 26552 43713 5790 11462 18737 ...
## $ Irural : int 5 45 58 48 28 54 38 37 47 58 ...
## $ Superficie : num 374.8 507.1 297 128.9 84.1 ...
```

## 6. Agrupamiento por K-Medios (K-Means Clustering)

En el método K-means, la cantidad óptima de centroides (clusters) no siempre se conoce de antemano y puede variar dependiendo de la estructura de los datos. Es por eso que es importante utilizar técnicas como el Método del Codo (Elbow Method) para determinar un valor apropiado para el número de clusters.

Se aplicará la función `kmeans` al conjunto de datos, `municipios_normalizados`, variando en cada caso el valor de `k`, y acumulando los valores de WCSS obtenidos:

Gráfico de Siluetas para determinar el número óptimo de clusters



Si bien es cierto que un valor más alto de la métrica de silueta sugiere una mejor separación entre los clusters, la elección del número óptimo de clusters también debe tomar en cuenta el contexto específico y el conocimiento del dominio. En ocasiones, es preferible seleccionar un número de clusters que sea más fácil de interpretar y tenga una mayor aplicabilidad práctica, incluso si esto significa no alcanzar el valor máximo de silueta. En este contexto, considero adecuado elegir  $K = 6$ , basándonos en el punto pico de la curva de silueta y en un punto medio de esta métrica. Esto nos permite lograr una agrupación efectiva de los datos mientras evitamos la sobreagrupación de entidades, lo que facilita la interpretación de los resultados y su aplicación en el contexto específico del problema.

La clusterización de los municipios en seis grupos nos ofrece una interesante perspectiva sobre cómo se diferencian entre sí según las variables seleccionadas (población, superficie e índice de ruralidad):

Cluster	Cardinalidad	Media_Poblacion	Media_Superficie	Media_Irural	Moda_Irural
6	1	7592871.00	1622.8526	6.00000	6
1	19	739644.42	516.8377	14.84211	12
5	450	36854.10	237.8359	37.54444	42
2	23	20057.43	17994.9797	83.91304	81
3	48	18448.98	4846.2359	75.47917	67
4	577	16848.50	651.4211	51.24090	48

La composición de los clústeres revela una diversidad significativa entre ellos. El clúster 6, caracterizado por su elevada población, está compuesto únicamente por un departamento. Por otro lado, el clúster con mayor número de departamentos es el cuarto, aunque su media de población puede asemejarse al clúster 3, la disparidad entre ellos radica en la media de superficie y en la distribución de la ruralidad. Los clústeres 5 y 2

presentan similitudes en términos de población, pero difieren considerablemente en cuanto a la superficie media y la distribución de la ruralidad. En contraste, el clúster 1 se distingue claramente del resto, ya que sus 19 departamentos muestran diferencias significativas en todas las variables analizadas.

Cluster	Cardinalidad	Media_Poblacion	Media_Superficie	Media_Irural	Prestadores
6	1	7592871.00	1622.8526	6.000000	14728
1	13	905612.44	288.4162	9.495442	34236
5	23	35126.08	213.2025	36.838220	42350
2	8	25652.13	16508.6859	78.716768	1781
4	26	17470.60	656.0608	50.469196	43340
3	13	13491.98	2873.3448	67.980272	14944

## Análisis conjunto prestadores y departamentos

El análisis de clusterización ha revelado una diversidad notable en la estructura de los departamentos de Colombia, mostrando variaciones en urbanización, tamaño y accesibilidad de servicios. Los departamentos agrupados en clusters con altas cifras de población y baja ruralidad exhiben una abundancia de prestadores de servicios, un reflejo directo de la demanda intensa característica de zonas urbanizadas con alta densidad de población. Esta correlación subraya la capacidad de las áreas metropolitanas para sostener y atraer una variedad de servicios, satisfaciendo así una gama más amplia de necesidades ciudadanas.

Por otro lado, los clusters identificados por su marcada ruralidad evidencian un número menor de prestadores de servicios. Este patrón plantea interrogantes sobre la accesibilidad y suficiencia de los servicios en áreas rurales, lo que podría significar desafíos en términos de cobertura y calidad. Estos departamentos rurales pueden enfrentar obstáculos sustanciales que afectan la provisión de servicios esenciales, desde la atención de la salud hasta la educación y el comercio.

Este análisis sirve como un punto de partida estratégico para la asignación equitativa de recursos y la elaboración de políticas públicas que se ajusten a las necesidades de cada cluster de departamentos. La creación de políticas adaptadas a estas agrupaciones podría optimizar los resultados y asegurar que los servicios estén disponibles donde más se necesiten.

Para profundizar en este estudio, sería conveniente investigar la calidad y el espectro de los prestadores de servicios en cada cluster de departamentos. Un entendimiento detallado de estos factores contribuiría a una imagen más completa de cómo los servicios actuales satisfacen las demandas de las poblaciones locales y revelaría oportunidades para mejoras y expansiones en sectores donde existen déficits.

En conjunto, el análisis demuestra la importancia de considerar la demografía y la geografía en la planificación de servicios y subraya la necesidad de estrategias diferenciadas que puedan abordar las disparidades en la distribución y calidad de los servicios entre los departamentos urbanos y rurales de Colombia.