Segmentación de los organismos Operadores de los servicios de Agua Potable, Alcantarillado y Saneamiento en México

Jessica Briseño

jevabrir@gmail.com

1-8-2024

Contenido

Índice de tablas

Índice de figuras

1. Introducción

Las empresas u organizaciones encargadas de la prestación de los servicios de agua potable, alcantarillado y saneamiento en México se les conoce como OAPAS (Operador de Agua Potable, Alcantarillado y Saneamiento). El INEGI realizó durante el año 2022 el Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México con el que se puede analizar la gestión y desempeño de la administración pública municipal y de las demarcaciones de la Ciudad de México. Este censo en su módulo 5 contiene información sobre la gestión, administración, características técnicas y ambientales de la prestación de los servicios municipales de agua potable y saneamiento (INEGI, 2024).

2. Planteamiento del problema

En el Censo Gubernamental Municipal se reporta un total de **2469** folios que corresponden con los OAPAS en los diferentes municipios de México. Este tipo de organismos presentan diferencias tanto en el tamaño (volumen servido o el número de clientes) como en su estructura, lo que ha dificultado el diseño de un método para clasificarlos. Es por lo que resulta esencial analizar y seleccionar algunos elementos del Censo Gubernamental Municipal, que tienen importancia para agrupar y caracterizar a los OAPAS en México, con el propósito lograr una comprensión de la relación entre los OAPAS clasificados y diseñar políticas públicas acordes a cada grupo.

3. Objetivo

El principal objetivo principal de este trabajo es implementar los algoritmos de aprendizaje no supervisado K-Means, Hierarchical y DBScan para segmentar a los prestadores de los servicios de agua potable, alcantarillado y saneamiento que se encuentran en México, a partir de los datos abiertos del Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México 2022 (INEGI,2022), para lograr una caracterización y comprensión de la relación de las variables que estos prestadores de servicios tienen en común.

El segundo objetivo de este trabajo es generar un modelo matemático empleando el método de regresión lineal múltiple para estimar el ingreso de los OAPAS en función de cuatro variables.

Para lograr estos objetivos se desarrollan los siguientes objetivos específicos:

- Descargar los diversos archivos con formato .csv correspondientes al Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México 2022 de la página de INEGI, para diseñar una dataset con los datos requeridos para el análisis.
- Realizar un análisis exploratorio de los datos para revisar los valores faltantes, identificar datos atípicos, imputar valores y seleccionar las variables para llevar el análisis.
- Realizar el preprocesamiento necesario, así como implementar los algoritmos K-Means, Hierarchical y DBScan para segmentar a los prestadores de los servicios de agua potable y alcantarillado.
- Utilizando la métrica de rendimiento "Silhouette score", identificar el mejor algoritmo de clasificación e implementarlo para clasificar a los OAPAS.
- Realizar la caracterización y descripción de la segmentación de los OAPAS.
- Implementar el método de regresión lineal múltiple para generar un modelo matemático para estimar el ingreso de los OAPAS en función de cuatro variables de estudio.

4. Metodología

Este trabajo se desarrolla utilizando diferentes librerías de Python, numpy, matplotlib, seaborn, scipy y sklearn. La metodología diseñada para cumplir los objetivos de este trabajo se muestra en la siguiente figura:

Figura 1. Metodología

5. Resultados

Para la creación de dataset se descargaron de la página de datos abiertos de INEGI diversos archivos que fueron analizados utilizando las librerías de Pandas, Numpy y Matplotlib de Python. Los siguientes archivos con formatos de archivo separado por comas (csv) o Excel (xls) contienen la información base para la conformación del dataset:

- admncion_cngmd2021.xls
- entidad_cngmd2021.csv
- mnpio_cngmd2021.csv
- servagua_cngmd2021.xls
- servdren_cngmd2021.xls

El archivo admncion_cngmd2021 contiene las variables que caracterizan a los municipios sobre la administración 2469 filas y 73 columnas con información. Las variables de este archivo que conforman el dataset a analizar son:

- **folio**: Indica el identificador de cada cuestionario compuesto por la clave de entidad y por la clave de municipio o delegación
- **total_in**: Indica el ingreso por el suministro de agua potable y saneamiento durante el año 2020 (total de ingresos por suministro de bienes y servicios)
- **totl_tom**: Indica el número total de tomas que cubre el servicio de agua entubada de la red pública

Con base en la descomposición de las claves que componen a los valores de la columna folio, así como a la clave y nombres de la entidad y municipios que contienen los archivos (entidad_cngmd2021.csv y mnpio_cngmd2021.csv), se realizó un diccionario en Python para agregar las siguientes columnas al dataset de estudio:

- **Name_Ent**: Indica el nombre de la entidad federativa
- **Name_Mun**: Indica el nombre del municipio.

Al dataset se le incorporaron las columnas "conx_tot" y "Pob_dren" del archivo que contiene las variables que caracterizan a los municipios sobre el servicio de drenaje y alcantarillado (servdren_cngmd2021.csv), así como las variables "conx_tot" y "Pob_dren"" del del archivo que contiene las variables que caracterizan a los municipios sobre el servicio de agua potable (servagua_cngmd2021.csv). La descripción de las variables "conx_tot", "Pob_dren", "conx_tot" y "Pob_dren" se describe a continuación:

- **conx_tot**: Indica el total de número de conexiones a la red de drenaje y alcantarillado por tipo de usuario.
- **Pob_dren**: Indica el porcentaje de la población del municipio o de la demarcación territorial que tenía acceso al servicio de drenaje y alcantarillado de la red pública.
- **Pobl_aPot**: Indica el porcentaje de población municipal o demarcación territorial que contaba con acceso al servicio de agua de la red pública.
- **Pob_dren:** Indica el porcentaje de la población del municipio o de la demarcación territorial que tenía acceso al servicio de drenaje y alcantarillado de la red pública.

Para la conformación final del dataset de estudio se realiza una selección de las nueve variables anteriormente descritas, de modo que el dataset para el estudio se conforma de 2469 filas y 9 columnas.

5.1 Análisis Exploratorio

Utilizando el dataset se realizó un análisis exploratorio de los datos para revisar los valores faltantes, identificar datos atípicos e los imputar valores. En este sentido la tabla 1 muestra los valores faltantes para cada una de las variables que componen el dataset.

	Valores faltantes
totl_tom	131
conx_tot	708
Pob_dren	563
Pobl_aPot	57
total_in	692
folio	0
cve_ent	0
Name_Ent	0
Name_Mun	0

Tabla 1. Cuantificación de variables nulos.

La variable que presenta la mayor cantidad de datos faltantes del "conx_tot", seguido de "total_in", lo que representa un total de 28.7% y 28.03% respecto al total de datos, respectivamente. La tabla 2 muestra el resumen de los principales estadísticos de las variables de estudio que corresponden las variables número total de tomas de agua potable, número total de conexiones de drenaje, el porcentaje de población servida de agua potable, el porcentaje de población servida de drenaje, así como la variable ingreso total. La tabla muestra en la primera columna la cantidad de datos para cada variable de estudio y las siguientes columnas muestran a la media, la desviación estándar, los cuartiles 1, 2 y 3, así como el valor máximo de cada variable.

Para conocer la distribución de las variables de estudio se llevó a cabo la visualización de los datos mediante graficas de dispersión, así como sus histogramas. En este sentido, la figura 2 muestra una gráfica de dispersión de puntos en las que el eje y corresponde a la variable ingreso total y en el eje x a las otras cuatro variables de estudio. Del análisis de las gráficas de dispersión muestra se puede observar a tres datos que se encuentran lejos del conglomerado de todos los demás datos. La figura A1 presentada en el anexo de este trabajo se muestra el histograma para las cinco variables de estudio utilizando los datos obtenidos del censo sin procesar.

	count	mean	std	min	25%	50%	75%	max
totl_tom	2,338.000	11,617.353	46,772.576	4.000	711.000	2,051.500	6,000.000	1,170,135.000
conx_tot	1,761.000	13,956.785	72,134.443	3.000	680.000	2,051.000	6,726.000	2,250,000.000
Pob_dren	1,906.000	73.816	24.843	0.500	60.000	80.000	94.000	100.000
Pobl_aPot	2,412.000	77.460	24.791	10.000	70.000	89.000	96.000	100.000
total_in	1,777.000	50,885,566.946	414,168,338.008	321.000	160,000.000	1,280,000.000	8,320,415.370	15,094,274,894.000

Tabla 2. Principales variables estadísticas de las variables de estudio con datos del Censo de INEGI.

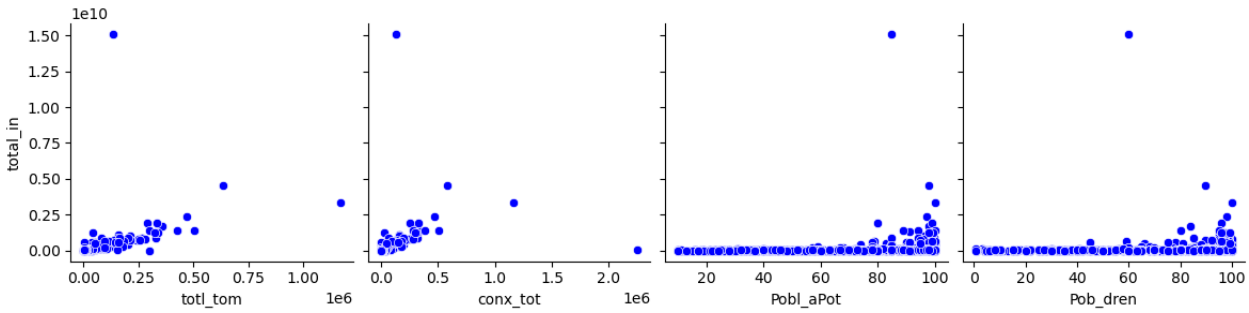

Figura 2. Gráfica de dispersión de las variables de estudio con datos del Censo de INEGI.

5.2 Detección de valores de valores atípicos

La detección de los valores atípicos se llevó a cabo la realización de tablas de información en orden ascendente y descendente para cada una de las variables de estudio. Con base en cada tabla se realizó una identificación de los valores no acordes al tamaño del municipio, o con base en los valores de las demás variables para el mismo folio. En este sentido es importante comentar que la cantidad de conexiones de drenaje tiene una alta correlación con la cantidad de tomas de agua potable y viceversa. Lo mismo ocurre para la población servida con agua potable y drenaje. En el Anexo de este documento se muestran en las tablas A1 a A10, la lista de 10 folios con los valores más altos y bajos de las cinco variables de estudio del dataset con datos crudos de INEGI. En este trabajo se implementó el algoritmo DBScan (Density-Based Spatial Clustering of Applications with Noise) como herramienta para identificar datos atípicos.

El algoritmo DBScan calcula la densidad alrededor de cada punto contando el número de puntos alrededor de un valor especificado (eps) así como por un valor de un mínimo de puntos ('min_samples') para determinar los puntos

centrales, puntos fronterizos y puntos de ruido (Hahsler, Piekenbrock, & Doran, 2024). Los valores del radio eps y 'min_samples' son especificados por el usuario, sin embargo, en este trabajo se implementó la herramienta GridSearchCV para identificarlos, ya que es una técnica de validación cruzada incluida en el paquete de scikit learn.

Los resultados de la implementación con GridSearchCV indicaron que el valor óptimo de 'eps' es de 3.5 y el valor de 'min_samples' es de 7. El algoritmo identificó un clúster y detectó un total de 6 datos atípicos. La figura 3 muestra con puntos de color negro a los datos atípicos encontrados con DBScan y en el Anexo A se muestra la tabla A11 que contiene la lista de los datos atípicos.

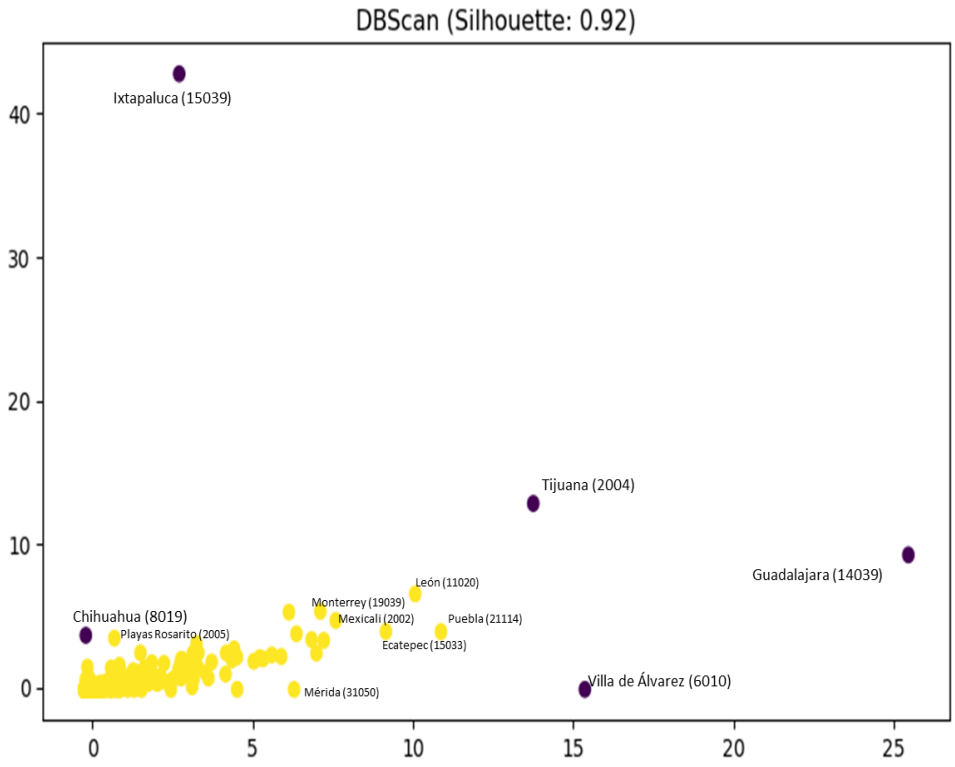

Figura 3. Gráfica de dispersión de los datos atípicos seleccionados con DBScan.

5.3 Eliminación de valores de valores atípicos

Con base en el análisis de cada tabla que contiene los valores máximos y mínimos para cada variable, del análisis de las gráficas de dispersión, así como con los resultados de DBScan se determinó que en los datos abiertos de INEGI se detectaron un total de 75 valores atípicos, 5 de los cuales corresponden a OAPAS en donde se encuentran algunos de los núcleos de población más importantes de México y que corresponden a los municipios de Tijuana, Guadalajara, Chihuahua, Puebla y León. Estos cinco municipios se considerarán como un segmento adicional a los determinados en este análisis. El dataset después de la eliminación se compone de un total de 2394 filas y 9 columnas. La tabla 2 muestra los principales estadísticos de las variables de estudio después de la eliminación de los 75 valores atípicos y la figura 3 muestra la gráfica de dispersión de las cinco variables de estudio, considerando el eje y como el ingreso total.

5.4 Imputación de valores

Unos de los requerimientos necesarios para implementar los algoritmos de segmentación es tener datos sin valores faltantes o nulos. En este trabajo se lleva a cabo la imputación de valores con el objetivo de no perder valores importantes para el análisis. El histograma de las variables de estudio generados con los datos sin valores atípicos (datos limpios) se muestra en la figura A14 del anexo y muestran que las cinco variables de estudio no muestran una distribución normal. Es por lo que la imputación de valores faltantes se llevó a cabo utilizando los siguientes criterios, con el objetivo de que la distribución de las variables no se vea afectada.

- **conx_tot**: se llevó a cabo la imputación de valores considerando los valores de la variable totl_tom.
- **totl_tom:** se llevó a cabo la imputación de valores considerando los valores de la variable conx_tot.
- **Pobl_aPot, Pob_dren** y **total_in:** se imputaron los valores utilizando la mediana de los datos.

La tabla 3 muestra los principales estadísticos de las variables de estudio considerando la eliminación de los 75 datos atípicos y la figura 4 muestras la gráfica de dispersión del ingreso en el eje y con respecto a las cuatro variables de estudio. La figura muestra una distribución más homogénea de los datos después de eliminar los 75 datos atípicos.

	count	mean	std	min	25%	50%	75%	max
totl_tom	2,266.000	9,846.269	31,563.194	4.000	720.000	2,034.000	5,956.250	428,144.000
conx_tot	1,695.000	10,938.087	33,288.860	4.000	746.500	2,113.000	6,690.500	389,221.000
Pob_dren	1,836.000	74.164	24.407	0.600	60.000	80.000	94.000	100.000
Pobl_aPot	2,339.000	77.375	24.871	10.000	70.000	89.000	96.000	100.000
total_in	1,723.000	34,677,231.619	147,400,362.613	321.000	158,370.000	1,271,509.000	7,918,931.150	1,913,074,128.970

Tabla 3. Principales estadísticos de las variables de estudio. Eliminación de 75 datos atípicos.

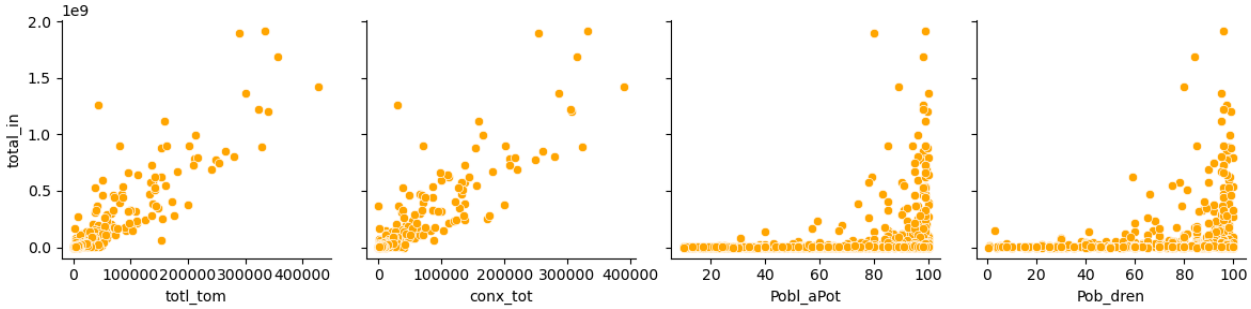

Figura 4. Gráfica de dispersión de las variables de estudio. Eliminación de 75 datos atípicos.

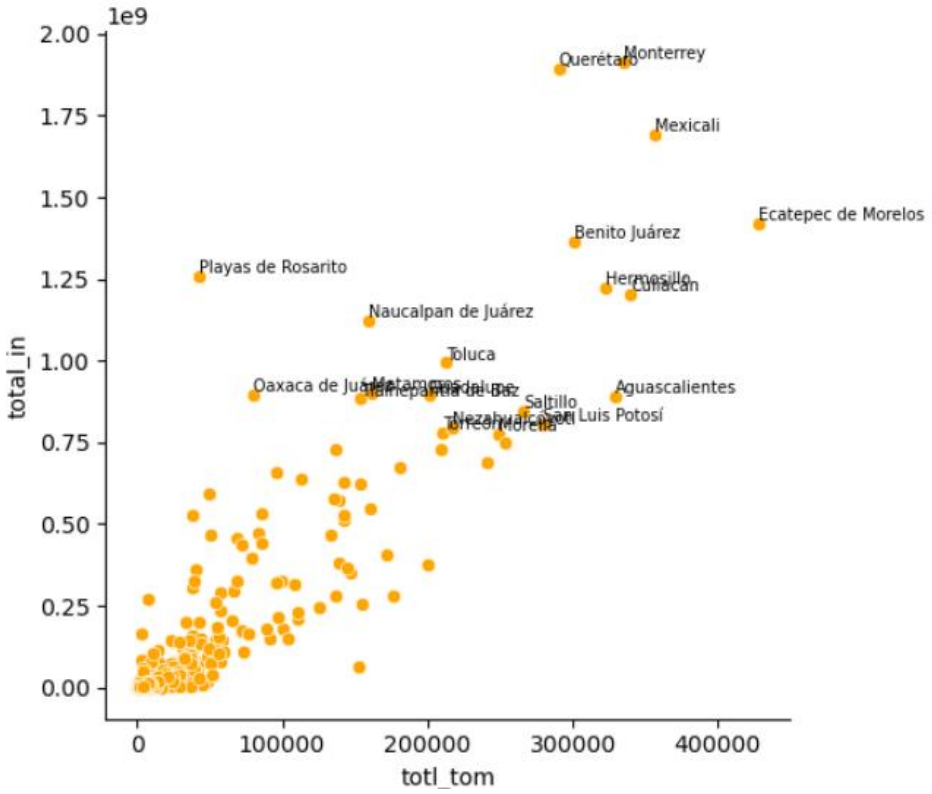

Figura 5. Gráfica de dispersión de los datos atípicos seleccionados con DBScan.

La figura 5 muestra la gráfica de dispersión del ingreso en función de las tomas de agua potables de los datos sin datos atípicos. En la gráfica se puede ver con claridad que los OAPAS de los municipios de Querétaro, Monterrey, Mexicali y Ecatepec tienen valores tanto de ingresos muy altos como de tomas de agua potable.

	totl_tom	conx_tot	Pob_dren	Pobl_aPot	total_in
totl_tom	1.000	0.986	0.139	0.154	0.899
conx_tot	0.986	1.000	0.152	0.149	0.892
Pob_dren	0.139	0.152	1.000	0.361	0.124
Pobl_aPot	0.154	0.149	0.361	1.000	0.124
total_in	0.899	0.892	0.124	0.124	1.000

Tabla 4. Valores de correlación de las variables de estudio.

5.5 Análisis de correlaciones

Con el objetivo de cuantificar el grado de correlación existente entre las cinco variables de estudio se llevó a cabo el cálculo de la matriz de correlación. La tabla 3 muestra los resultados de la matriz de correlación de las variables de estudio y la figura 6 muestra el mapa de colores con los resultados de la correlación. Los resultados muestran que las variables tomas de agua potable y conexiones de drenaje presentan altas correlaciones entre ellas, así como con los ingresos totales.

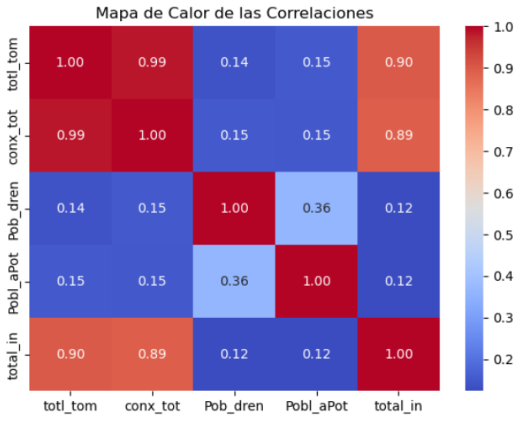

Figura 6. Mapa de correlaciones.

5.6. Modelos de segementación

Con el objetivo de disminuir la diferencia en los valores de los variables, se realizó la estandarización de estas utilizando el StandarScaler de la librería de sklearn, posteriormente se implementaron los algoritmos de clasificación K-Means, Hierarchical y DBScan para segmentar a los prestadores de los servicios de agua potable y alcantarillado.

5.6.1 Segmentación con DBScan

La implementación del algoritmo DBScan se llevó a cabo utilizando la herramienta GridSearch para determinar los mejores parámetros del modelo. En este sentido se determinó que los mejores parámetros son eps= 7, min_samples=40 y el número de agrupamientos calculados es uno. El método DBScan determino que se tienen cuatro valores atípicos que son los municipios de Mexicali, Ecatepec de Morelos, Monterrey y Querétaro. En el anexo de este documento se muestra la figura A4 en los que se observa la segmentación y los valores atípicos y en la Tabla A13 se muestra la lista con los folios de los valores atípicos detectados con el método DBScan.

5.6.3 Segmentación con Hierarchical

El algoritmo de agrupamiento jerárquico Hierarchical clustering busca desarrollar una jerarquía de un conjunto de observaciones que generalmente se representan mediante un dendograma (Nieto-Jeux, 2021). La implementación del algoritmo en este trabajo utilizando la librería hierarchy de scipy. Lo resultados de la implementación se evaluaron utilizando la métrica Silhouette Score en función del número de clústeres y se muestran en la figura 7. En la figura se observa que un número adecuado de clúster es 3. No obstante lo anterior, en este trabajo se decidió realizar la segmentación de los OAPAS en cinco grupos para una mejor caracterización. Los resultados de la agrupación con cinco clústeres con la métrica Silhouette Score es d e 0.50.

5.6.2 Segmentación con KMeans

El algoritmo K-Means busca agrupar observaciones en un número fijo de grupos o clústeres basándose en sus características, asociando observaciones similares en función de la distancia al centroide de cada grupo, de tal forma que esta distancia sea mínima (Prada-Conde, 2022). En este trabajo el algoritmo se implementa utilizando la librería KMeans de sklearn y su rendimiento se evalúa empleando la métrica Silhouette Score (figura 78. Los resultados de la implementación del algoritmo utilizando la métrica Silhouette Score en función del número de clústeres se muestran en la siguiente figura y se puede observar que un número adecuado de clúster es 3. Como se mencionó anteriormente, en este trabajo se decidió realizar la segmentación de los OAPAS en cinco grupos. Los resultados de la agrupación con el método KMeans con cinco clústeres y con la métrica Silhouette Score es de 0.51. Esté resultado es ligeramente mejor que el método de Hierarchical es por lo que el algoritmo seleccionado para realizar la agrupación de los OAPAS es método KMeans.

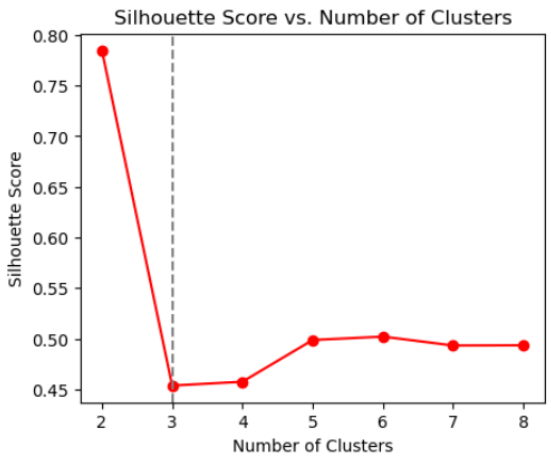

Figura 7. Métrica Silhoutte Score con Hierarchical

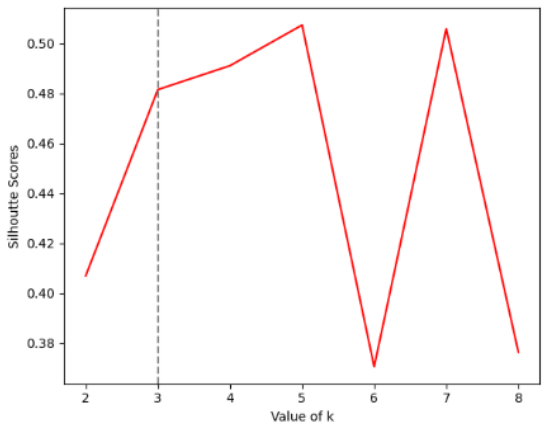

Figura 8. Métrica Silhoutte Score con KMeans

5.6.4 Segmentación de los OAPAS

En esta sección, se muestran los resultados de la agrupación con cinco clústeres de los OAPAS con el método KMeans. La figura 8 muestra la gráfica con la segmentación de los OAPAS, de color rojo se muestra la clase 5, de color verde la clase 4, de color azul la clase 3 y de color morado y amarillo las clases 2 y 1 respectivamente. La figura 9 muestra el ingreso en el eje y con respecto a las otras cuatro variables analizadas en el eje x. Los resultados del agrupamiento de los OAPAS se muestran en la figura10 mediante diagramas de caja para cada una de las variables analizadas. En los diagramas de caga mostrados se muestran simultáneamente información sobre la forma y dispersión de las características en los cinco grupos. La figura muestra en cada caja rellena de color azul al 50% de los OAPAS para cada variable analizada, y con una línea horizontal de color negro que se muestra dentro de cada se muestra al valor de la mediana. El extremo inferior de la caja muestra el primer cuartil mientras que el extremo superior representa el tercer cuartil.

La tabla 5 muestra un resumen de las principales características correspondientes a las cinco clasificaciones de los OAPAS. Del análisis de los resultados de la tabla 4 y de la figura 10 se puede distinguir que los OAPAS que se encuentran en las clases 4 y 5 tienen los valores con más altos de ingresos, en promedio de $391,604,8256 pesos para la clase 4 y de $1,083,648,155 pesos para la clase 5. Este tipo de comportamiento también se asocia con las variables de población servida de agua potable y drenaje con los promedios más altos de cobertura, así como para la cantidad de tomas de agua potable y conexiones de drenaje. Las características de las clases 1,2 y 3 muestran que los ingresos son los más bajos con promedios de $2,654,816 pesos, $4,052,319 pesos y $8,955,772 pesos, respectivamente. En cuanto a la cantidad de tomas de agua potable y conexiones de drenaje, también se observa un comportamiento similar ya que estas clases tienen los valores más bajos, con valores promedio de 2,125 tomas y 2,177 conexiones para la clase 1; 3,653 tomas y 2,656 conexiones para la clase 2 y 5,910 tomas y 5,419 conexiones para la clase 3.

La tabla 5 también muestra los valores de los cinco municipios agrupados y analizados de forma independiente debido a las características que presentan de núcleos de población extremadamente grandes.

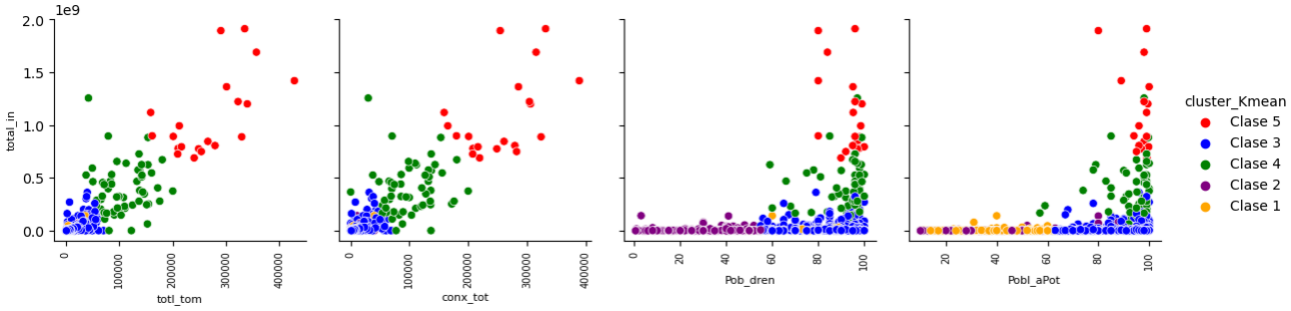

Figura 9. Gráfica de dispersión del ingreso y variables analizadas. OAPAS segmentados.

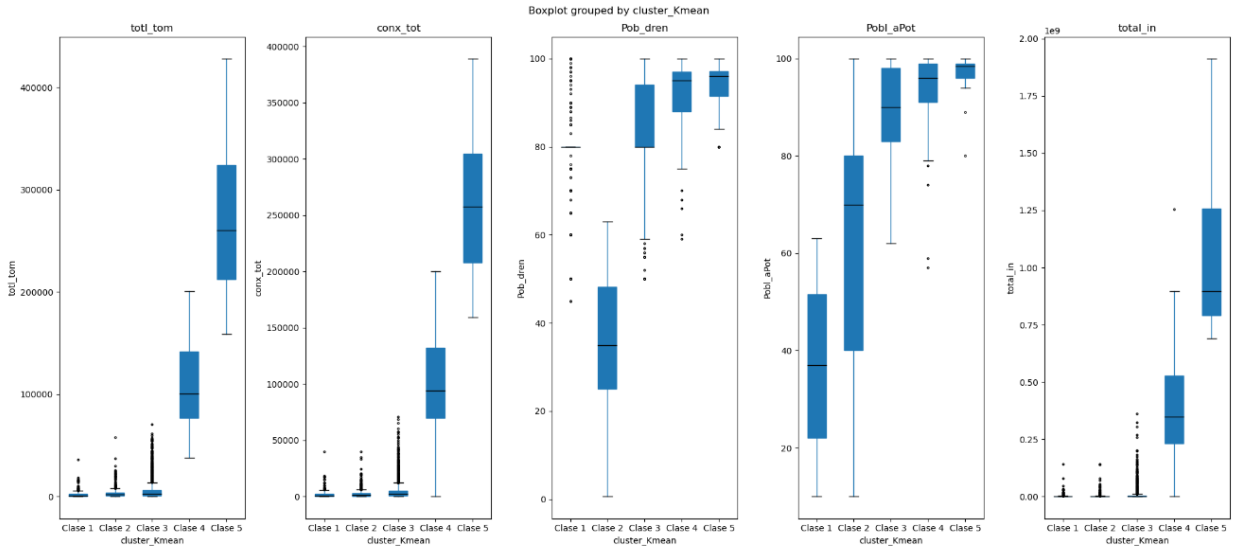

Figura 10. Gráficos de BoxPlot de las variables.

Clúster	Descripción	Núm. de OAPAS	Tomas de agua potable	Conexiones drenaje	% Pob. servida agua potable	% Pob. servida drenaje	Ingresos totales pesos
Clase 1	Muy baja cobertura de agua potable y alcantarillado	347	30 a 36,269 Prom. 2,125	38 a 39,983 Prom. 2,177	10% a 63% Prom. 36%	45% a 100% Prom. 80%	$321 a $141,716,159 Prom. $2,654,816 pesos
Clase 2	Baja cobertura de agua potable y alcantarillado	399	79 a 58,183 Prom. 3,653	11a 40,000 Prom. 2,656	10% a 80% Prom. 62%	0.6% a 63% Prom. 34%	3,390 a $142,660,111 pesos Prom= $4,052,319 pesos
Clase 3	Media cobertura de agua potable y alcantarillado	1,571	4 a 70,645 Prom. 5,910	4 a 70,645 Prom. 5,419	62% a 100% Prom. 90%	50% a 100% Prom. 84%	$1,000 a $363,176,392 Prom. $8,955,772
Clase 4	Alta cobertura de agua potable y alcantarillado	57	38,063 a 200,470 Prom.108,525	74 a 200,040 Prom. 99,790	57% a 100% Prom. 93%	59% a 100% Prom. 91%	$1,271,509 a $1,256,686,180 Prom. $391,604,8256
Clase 5	Muy alta cobertura de agua potable y alcantarillado	20	159,115 a 428,144 Prom. 268,229	159,115 a 389,221 Prom. 257,066	80% a 100% Prom. 97%	80% a 100% Prom. 93%	$689,929,731 a $1,913,074,129 Prom $1,083,648,155
	Tijuana Chihuahua León Guadalajara	5	637,449 469,797 1,170,135 506,375 Prom. 695,939	578,966 326,729 470,367 1,164,305 507,618 Prom. 609,597	89.6% 96% 98% 100% 95% Prom. 96%	98% 91% 97% 100% 94% Prom. 96%	$4,559,694,557 $1,326,602,408 $2,342,206,651 $3,299,290,059 $1,416,260,601 $2,588,810,855

Tabla 5. Características de las 5 clasificaciones de los OAPAS

Es importante mencionar que en la tabla 5 no se muestra ninguna información de las Ciudad de México que es considerada uno de los núcleos de población más importantes de México, debido a que en el Censo de INEGI no se encuentra información. Una descripción más detallada de las clases se muestra en el anexo de este documento.

5.7 Modelo de regresión múltiple

En este apartado se muestran los resultados del modelo matemático calculado utilizando la biblioteca de sklearn para estimar el ingreso de los OAPAS en función de la cantidad de tomas y conexiones de agua potable, así como el porcentaje de población servida de agua potable y drenaje. El proceso para calcular el modelo utilizando el método de regresión lineal múltiple consiste en la estandarización de las variables de estudio utilizando StandarScaler de la librería de sklearn. La ecuación para estimar el Ingreso de los OAPAS y que se calculó entrenado al modelo con las variables estandarizadas es:

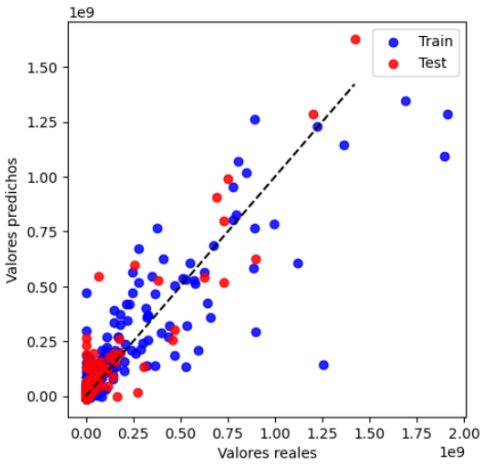

Figura 11. Resultados estimación modelo vs valores reales.

$$I = 3{,}303.4\ totl_{tom} + 573.7\ conx_{tot} - 91{,}795.6\ Pbl_{aPot} + 859.7\ Pob_{dren} - 3{,}670{,}323$$

Donde I es el ingreso total municipal de un Organismo operadore de agua en pesos, $totl_{tom}$ es el número total de tomas de agua potable, $conx_{tot}$ es el número de total de conexiones de drenaje, Pbl_{aPot} es el porcentaje de la población con servicio de agua potable y Pob_{dren} es el porcentaje de la población con servicio de drenaje.

El dataset utilizado para el ajuste del modelo es el mismo que el utilizado para la clasificación de los OAPAS y que tiene un total de 2469 filas y 9 columnas. Como se describe en el punto 5.4 los valores faltantes se imputaron con el valor de la mediana para las variables Pbl_aPot, Pob_dren, mientras que para totl_tom y conx_tot se realizó de otra forma. El modelo se calculó utilizando 80% de las observaciones como datos de entrenamiento y el 20% como datos de prueba. En este sentido se utilizaron un total de 1915 observaciones de entrenamiento y 479 de prueba. Los resultados del coeficiente de determinación R^2 del modelo ajustado en los datos de prueba: 0.84 lo que indica un buen ajuste.

La figura 11 muestra la gráfica de los resultados de la estimación del modelo contra los valores reales. Se puede observar que, aunque existen diferencias los valores estimados (puntos de color rojo) siguen un comportamiento similar a los valores reales (puntos de color azul).

La figura 12 muestra la gráfica de dispersión de los valores estimados con el modelo ajustado con respecto a los valores reales para las cuatro variables independientes. Los puntos de color azul en la gráfica son los valores estimados con el modelo de regresión lineal calculado y los puntos de color gris son los valores reales del Censo. Realizando una comparación visual de figura 11, entre los datos reales y los estimados se puede observar que el modelo realiza estimaciones parecidas a los datos reales.

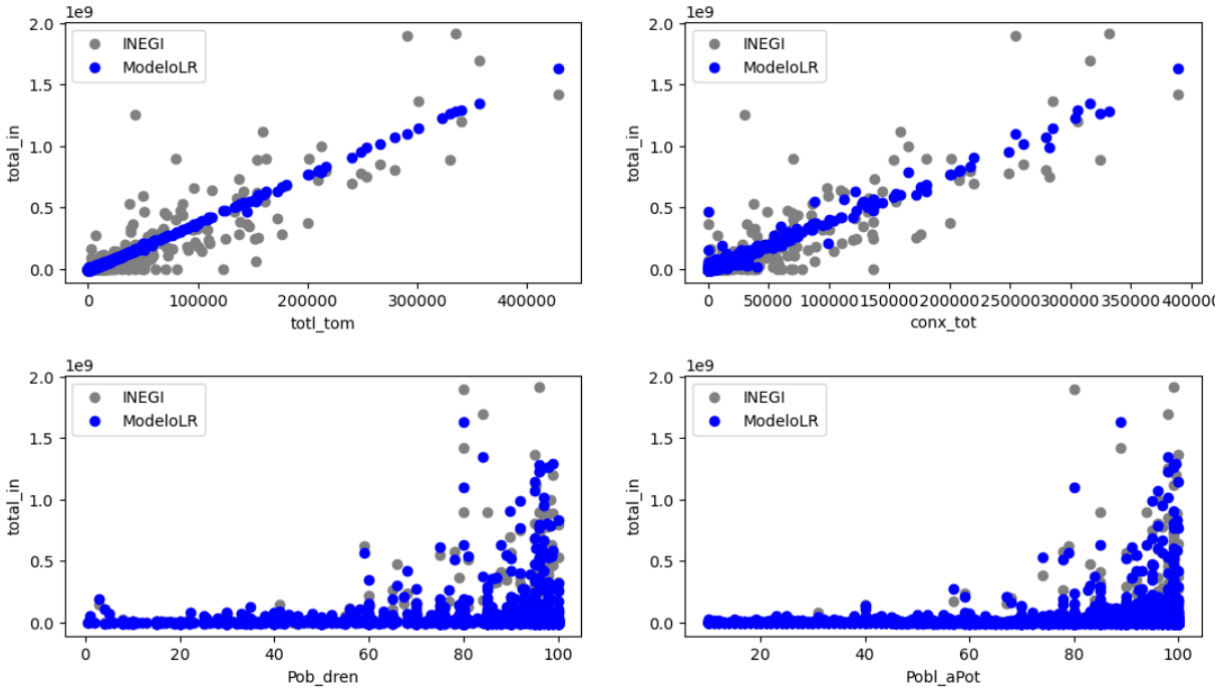

Figura 12. Grafica de dispersión de los valores estimados vs los valores reales

En el conjunto de prueba se calcularon las siguientes métricas para evaluar el desempeño del modelo.

Error absoluto medio (MAE) = 19,742,077 pesos

Error cuadrático medio (MSE)= 2,520,434,325,859,517

Raíz del error cuadrático medio (RMSE) = 50,203,927 pesos

Los resultados del MAE y el RMSE indican que el modelo tiene un error mínimo para estimar el ingreso en los Clase 5, en donde los ingresos tienen un rango de 689,929,731 a 1,913,074,129 pesos. Sin embargo, es recomendable revisar sus estimaciones con mayor detalle para verificar su correcta implementación.

6. Conclusiones

Los resultados de la segmentación de los OAPAS realizada con el método de aprendizaje no supervisado KMeans sugieren que la segmentación con cinco clases permite determinar con éxito algunas de las principales características de estos prestadores de servicios en México. Sin embargo, existen factores importantes como datos con errores presentes en el Censo de INEGI que dificultaron el obtener una segmentación con mayor precisión. Por otro lado, en este trabajo sólo se consideraron para la segmentación el número de tomas de agua potable, el número de conexiones de drenaje, el porcentaje de la población servida con agua potable y alcantarillado, así como el ingreso de los OAPAS, factores que aunque son importantes, se recomienda realizar otros análisis considerando factores como la cantidad de empleados del organismo operador, el número de habitantes, los ingresos de los diferentes rubros, entre otras variables que están presentes en el censo de INEGI.

Otra conclusión importante en este trabajo es que se logró determinar un modelo matemático para estimar el ingreso de los OAPAS en México en función del número de tomas de agua potable, el número de conexiones de drenaje, así como del porcentaje de la población servida con agua potable y alcantarillado con un coeficiente de determinación R^2 de 0.84 calculada en los datos de prueba, lo que indica que el modelo presenta un buen ajuste. Sin embargo, debido a la gran dispersión presente en las observaciones, el modelo podría se útil para estimar el ingreso sólo en los OAPAS de los Clase 5, en los que los ingresos tienen un rango de 689,929,731 a 1,913,074,129 pesos.

Es recomendable revisar las observaciones atípicas detectadas en el Censo de INEGI para mejorar los resultados tanto de la segmentación de los OAPAS, así como el modelo para estimar su ingreso. También se recomienda realizar un dendograma para evaluar los niveles de similitud de las segmentaciones.

7. Referencias Bibliográficas

Hahsler, M., Piekenbrock, M., & Doran, D. (30 de 07 de 2024). *dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms.* . Obtenido de https://cran.r-project.org/web/packages/dbscan/dbscan.pdf

INEGI. (20 de Julio de 2024). *Instituto Nacional de Estadística, Geografía e Informática*. Obtenido de Subsistema de Información de Gobierno, Seguridad Pública e Impartición de Justicia. Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México 2021: https://www.inegi.org.mx/programas/cngmd/2021/#datos_abiertos

Nieto-Jeux, A. (2021). *Algoritmos de Aprendizaje Automático. Un estudio de si difusión y utilización. Trabajo de Fin de Grado.* Madrid, España.: Escuela técnica Superior de Ingenieros Informáticos. Universidad Politécnica de Madrid. Disponible en: https://oa.upm.es/68484/1/TFG_ALEJANDRO_NIETO_JEUX.pdf.

Prada-Conde, L. (2022). *Aplicación de técnicas de clustering como paso previo a la detección de anomalías en redes defnidas por sofware. Trabajo de grado.* España: Facultad de Informática. Universidade da Coruña. Disponible en: https://ruc.udc.es/dspace/bitstream/handle/2183/32837/PradaConde_Luis_TFG_2022.pdf?sequence=3&isAllowed=y.

Anexo

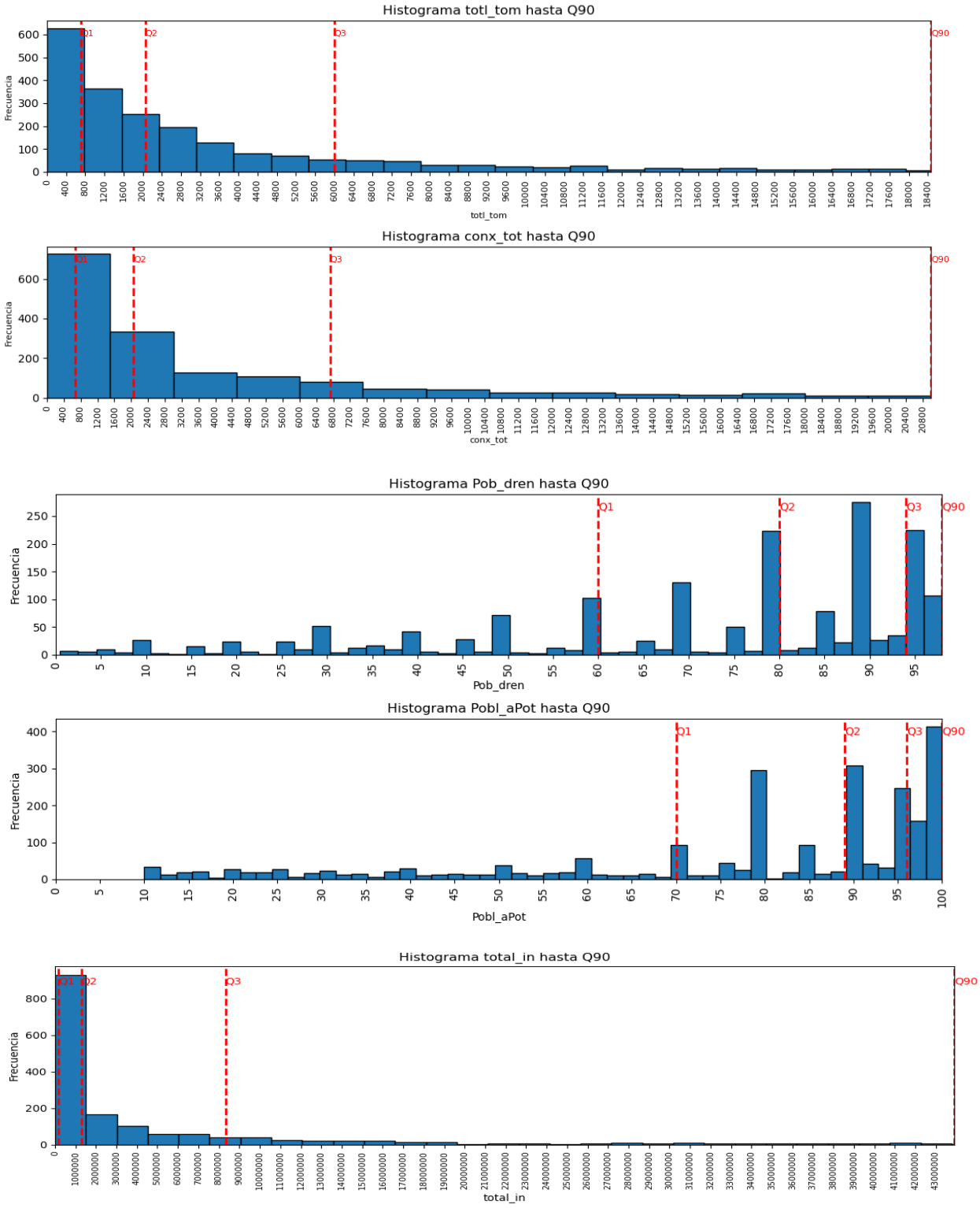

Figura A 1. Gráfica de los histogramas de las variables de estudio con datos del Censo de INEGI.

	folio	Name_Ent	Name_Mun	totl_tom	conx_tot	Pob_dren	Pobl_aPot	total_in
702	15039	Estado de México	Ixtapaluca	134,624.000	136,038.000	60.000	85.000	15,094,274,894.000
14	2004	Baja California	Tijuana	637,449.000	578,966.000	89.600	98.000	4,559,694,557.000
577	14039	Jalisco	Guadalajara	1,170,135.000	1,164,305.000	100.000	100.000	3,299,290,059.000
347	11020	Guanajuato	León	469,797.000	470,367.000	98.000	97.000	2,342,206,650.760
996	19039	Nuevo León	Monterrey	335,097.000	331,865.000	96.000	99.000	1,913,074,128.970
1809	22014	Querétaro	Querétaro	290,449.000	254,564.000	NaN	80.000	1,895,036,863.000
12	2002	Baja California	Mexicali	356,964.000	315,632.000	84.000	98.000	1,690,761,382.000
696	15033	Estado de México	Ecatepec de Morelos	428,144.000	389,221.000	80.000	89.000	1,420,984,395.000
1692	21114	Puebla	Puebla	506,375.000	507,618.000	95.000	94.000	1,416,260,601.090
1818	23005	Quintana Roo	Benito Juárez	301,185.000	285,618.000	95.100	100.000	1,363,658,394.000

Tabla A 1. Lista de 10 folios con ingresos más altos (total_in)

	folio	Name_Ent	Name_Mun	totl_tom	conx_tot	Pob_dren	Pobl_aPot	total_in
1693	21115	Puebla	Quecholac	221.000	NaN	90.000	20.000	321.000
1188	20180	Oaxaca	San Juan Bautista Lo de Soto	540.000	482.000	80.000	90.000	1,000.000
1064	20056	Oaxaca	Mártires de Tacubaya	300.000	NaN	NaN	66.000	1,200.000
1210	20202	Oaxaca	San Juan Lachao	518.000	NaN	NaN	31.000	1,200.000
1438	20430	Oaxaca	Santa María Tataltepec	150.000	NaN	NaN	99.000	1,260.000
1367	20359	Oaxaca	Santa Ana Yareni	260.000	270.000	90.000	100.000	1,800.000
1493	20485	Oaxaca	Santiago Tapextla	350.000	NaN	NaN	47.000	2,000.000
1515	20507	Oaxaca	Santo Domingo Armenta	1,000.000	850.000	79.000	80.000	2,000.000
1522	20514	Oaxaca	Santo Domingo Roayaga	192.000	182.000	90.000	54.000	2,400.000
765	15102	Estado de México	Timilpan	4,000.000	NaN	70.000	90.000	2,600.000

Tabla A 2. Lista de 10 folios con ingresos más altos (total_in)

	folio	Name_Ent	Name_Mun	total_in	Pobl_aPot	Pob_dren	conx_tot	totl_tom
577	14039	Jalisco	Guadalajara	3,299,290,059.000	100.000	100.000	1,164,305.000	1,170,135.000
81	6010	Colima	Villa de Álvarez	NaN	100.000	98.000	65,505.000	710,766.000
14	2004	Baja California	Tijuana	4,559,694,557.000	98.000	89.600	578,966.000	637,449.000
1692	21114	Puebla	Puebla	1,416,260,601.090	94.000	95.000	507,618.000	506,375.000
347	11020	Guanajuato	León	2,342,206,650.760	97.000	98.000	470,367.000	469,797.000
696	15033	Estado de México	Ecatepec de Morelos	1,420,984,395.000	89.000	80.000	389,221.000	428,144.000
12	2002	Baja California	Mexicali	1,690,761,382.000	98.000	84.000	315,632.000	356,964.000
1888	25006	Sinaloa	Culiacán	1,201,402,177.000	99.500	99.000	306,183.000	339,960.000
996	19039	Nuevo León	Monterrey	1,913,074,128.970	99.000	96.000	331,865.000	335,097.000
0	1001	Aguascalientes	Aguascalientes	889,622,823.430	99.000	98.000	324,115.000	329,552.000

Tabla A 3. Lista de 10 folios con los valores más altos de tomas de agua potable (totll_tom)

	folio	Name_Ent	Name_Mun	total_in	Pobl_aPot	Pob_dren	conx_tot	totl_tom
2196	30104	Veracruz de Ignacio de la Llave	Mecayapan	NaN	34.000	70.000	10,000.000	4.000
2175	30083	Veracruz de Ignacio de la Llave	Ixhuatlán de Madero	NaN	85.000	100.000	NaN	4.000
1629	21051	Puebla	Chietla	166,824.260	15.000	1.000	NaN	13.000
2025	28036	Tamaulipas	San Nicolás	NaN	90.000	NaN	NaN	25.000
682	15019	Estado de México	Capulhuac	NaN	25.000	90.000	NaN	30.000
1296	20288	Oaxaca	San Miguel Yotao	NaN	30.000	NaN	NaN	40.000
2291	30199	Veracruz de Ignacio de la Llave	Zaragoza	NaN	80.000	70.000	NaN	40.000
695	15032	Estado de México	Donato Guerra	NaN	15.000	80.000	NaN	41.000
1330	20322	Oaxaca	San Pedro Ocopetatillo	NaN	80.000	70.000	30.000	45.000
1247	20239	Oaxaca	San Martín Huamelúlpam	6,000.000	60.000	NaN	NaN	50.000

Tabla A 4. Lista de 10 folios con los valores más altos de tomas de agua potable (toyl_tom)

	folio	Name_Ent	Name_Mun	total_in	Pob_dren	Pobl_aPot	totl_tom	conx_tot
913	17012	Morelos	Jojutla	19,268,518.000	95.000	98.000	14,467.000	2,250,000.000
577	14039	Jalisco	Guadalajara	3,299,290,059.000	100.000	100.000	1,170,135.000	1,164,305.000
14	2004	Baja California	Tijuana	4,559,694,557.000	89.600	98.000	637,449.000	578,966.000
1692	21114	Puebla	Puebla	1,416,260,601.090	95.000	94.000	506,375.000	507,618.000
347	11020	Guanajuato	León	2,342,206,650.760	98.000	97.000	469,797.000	470,367.000
696	15033	Estado de México	Ecatepec de Morelos	1,420,984,395.000	80.000	89.000	428,144.000	389,221.000
996	19039	Nuevo León	Monterrey	1,913,074,128.970	96.000	99.000	335,097.000	331,865.000
224	8019	Chihuahua	Chihuahua	1,326,602,408.470	96.000	91.000	NaN	326,729.000
0	1001	Aguascalientes	Aguascalientes	889,622,823.430	98.000	99.000	329,552.000	324,115.000
12	2002	Baja California	Mexicali	1,690,761,382.000	84.000	98.000	356,964.000	315,632.000

Tabla A 5. Lista de 10 folios con los valores más altos de total de conexiones de drenaje (conx_tot)

	folio	Name_Ent	Name_Mun	total_in	Pob_dren	Pobl_aPot	totl_tom	conx_tot
1962	26062	Sonora	Suaqui Grande	629,844.750	95.000	95.000	466.000	3.000
491	13037	Hidalgo	Metztitlán	NaN	75.000	90.000	NaN	4.000
779	15116	Estado de México	Zacazonapan	NaN	99.000	100.000	3,500.000	5.000
1000	19043	Nuevo León	Rayones	555,630.150	96.000	99.000	401.000	8.000
2261	30169	Veracruz de Ignacio de la Llave	José Azueta	4,925,008.030	27.000	32.000	2,968.000	8.000
1664	21086	Puebla	Jalpan	NaN	60.000	80.000	200.000	10.000
490	13036	Hidalgo	San Agustín Metzquititlán	NaN	90.000	90.000	NaN	10.000
503	13049	Hidalgo	Pisaflores	28,451.760	80.000	80.000	693.000	10.000
2162	30070	Veracruz de Ignacio de la Llave	Hidalgotitlán	NaN	22.000	22.000	1,139.000	11.000
1763	21185	Puebla	Tlapanalá	139,600.000	90.000	30.000	604.000	11.000

Tabla A 6. Lista de 10 folios con los valores más bajos de total de conexiones de drenaje (conx_tot)

	folio	Name_Ent	Name_Mun	total_in	totl_tom	conx_tot	Pob_dren	Pobl_aPot
1995	28006	Tamaulipas	Bustamante	NaN	900.000	68.000	10.000	10.000
1338	20330	Oaxaca	San Pedro Teutila	NaN	500.000	NaN	NaN	10.000
1111	20103	Oaxaca	San Antonino Castillo Velasco	48,760.000	116.000	1,560.000	60.000	10.000
1647	21069	Puebla	Huaquechula	110,612.000	1,200.000	1,200.000	99.000	10.000
389	12016	Guerrero	Coahuayutla de José María Izazaga	NaN	1,164.000	402.000	10.000	10.000
1488	20480	Oaxaca	Santiago Nundiche	16,950.000	104.000	NaN	NaN	10.000
138	7057	Chiapas	Motozintla	3,575,434.640	7,614.000	6,385.000	80.000	10.000
452	12079	Guerrero	José Joaquín de Herrera	NaN	680.000	1,285.000	10.000	10.000
2276	30184	Veracruz de Ignacio de la Llave	Tlaquilpa	NaN	310.000	400.000	25.000	10.000
1723	21145	Puebla	San Sebastián Tlacotepec	NaN	1,750.000	2,050.000	45.000	10.000

Tabla A 7. Lista de 10 folios con valores más altos de población servida con agua Potable (Pobl_aPot)

	folio	Name_Ent	Name_Mun	total_in	totl_tom	conx_tot	Pob_dren	Pobl_aPot
1176	20168	Oaxaca	San José Estancia Grande	NaN	250.000	NaN	NaN	100.000
1187	20179	Oaxaca	San Juan Bautista Jayacatlán	40,800.000	350.000	NaN	NaN	100.000
1224	20216	Oaxaca	San Juan Tabaá	30,000.000	344.000	344.000	100.000	100.000
1204	20196	Oaxaca	San Juan Evangelista Analco	508,000.000	296.000	147.000	70.000	100.000
1201	20193	Oaxaca	San Juan del Estado	99,000.000	900.000	NaN	NaN	100.000
2073	29041	Tlaxcala	Papalotla de Xicohténcatl	980,000.000	12,200.000	12,200.000	95.000	100.000
1192	20184	Oaxaca	San Juan Bautista Tuxtepec	10,635,921.300	22,400.000	NaN	100.000	100.000
299	10011	Durango	Indé	20,000.000	300.000	142.000	50.000	100.000
298	10010	Durango	Hidalgo	101,079.670	193.000	280.000	80.000	100.000
1189	20181	Oaxaca	San Juan Bautista Suchitepec	37,100.000	243.000	NaN	NaN	100.000

Tabla A 8. Lista de 10 folios con valores más bajos de población servida con agua Potable (Pobl_aPot)

	folio	Name_Ent	Name_Mun	total_in	totl_tom	conx_tot	Pobl_aPot	Pob_dren
773	15110	Estado de México	Valle de Bravo	66,019,070.000	12,738.000	NaN	NaN	0.500
2223	30131	Veracruz de Ignacio de la Llave	Poza Rica de Hidalgo	88,570,285.180	40,805.000	NaN	82.000	0.600
2289	30197	Veracruz de Ignacio de la Llave	Yecuatla	109,885.980	1,117.000	1,177.000	28.000	0.600
2221	30129	Veracruz de Ignacio de la Llave	Platón Sánchez	5,525,193.170	3,675.000	2,147.000	95.000	0.850
1815	23002	Quintana Roo	Felipe Carrillo Puerto	17,478,353.810	18,886.000	116.000	98.000	1.000
1629	21051	Puebla	Chietla	166,824.260	13.000	NaN	15.000	1.000
2093	30001	Veracruz de Ignacio de la Llave	Acajete	NaN	1,200.000	550.000	19.000	2.000
24	4003	Campeche	Carmen	142,660,111.250	58,183.000	11,000.000	85.000	3.000
1030	20022	Oaxaca	Cosoltepec	146,940.000	375.000	35.000	54.000	4.000
414	12041	Guerrero	Malinaltepec	18,000.000	686.000	500.000	80.000	4.000

Tabla A 9. Lista de 10 folios con valores más altos de población servida con drenaje (Pob_dren)

	folio	Name_Ent	Name_Mun	total_in	totl_tom	conx_tot	Pobl_aPot	Pob_dren
244	8039	Chihuahua	López	1,748,999.000	1,163.000	1,163.000	100.000	100.000
2061	29029	Tlaxcala	Tepeyanco	261,223.000	3,223.000	2,922.000	100.000	100.000
1307	20299	Oaxaca	San Pablo Yaganiza	NaN	250.000	250.000	100.000	100.000
1255	20247	Oaxaca	Capulálpam de Méndez	65,270.000	812.000	812.000	100.000	100.000
2040	29008	Tlaxcala	Cuapiaxtla	1,440,931.000	2,628.000	57.000	100.000	100.000
2041	29009	Tlaxcala	Cuaxomulco	432,000.000	1,500.000	1,600.000	100.000	100.000
1224	20216	Oaxaca	San Juan Tabaá	30,000.000	344.000	344.000	100.000	100.000
134	7053	Chiapas	Mazapa de Madero	NaN	489.000	489.000	80.000	100.000
2043	29011	Tlaxcala	Muñoz de Domingo Arenas	305,128.000	789.000	789.000	98.000	100.000
1192	20184	Oaxaca	San Juan Bautista Tuxtepec	10,635,921.300	22,400.000	NaN	100.000	100.000

Tabla A 10. Lista de 10 folios con valores más bajos de población servida con drenaje (Pob_dren)

totl_tom	conx_tot	ext_sani_km	ext_co_km	Pob_dren	Pobl_aPot	total_in	folio	cve_ent	Name_Ent	Name_Mun	cluster_db
637,449.000	578,966.000	4,013.000	4,046.000	89.600	98.000	4,559,694,557.000	2004	2	Baja California	Tijuana	-1
710,766.000	65,505.000	487.480	579.390	98.000	100.000	1,280,000.000	6010	6	Colima	Villa de Álvarez	-1
2,051.500	326,729.000	3,108.000	3,806.090	96.000	91.000	1,326,602,408.470	8019	8	Chihuahua	Chihuahua	-1
1,170,135.000	1,164,305.000	8,577.000	8,570.890	100.000	100.000	3,299,290,059.000	14039	14	Jalisco	Guadalajara	-1
134,624.000	136,038.000	14,709.000	972.000	60.000	85.000	15,094,274,894.000	15039	15	Estado de México	Ixtapaluca	-1
14,467.000	2,250,000.000	120.000	169.830	95.000	98.000	19,268,518.000	17012	17	Morelos	Jojutla	-1

Tabla A 11. Lista de folios seleccionados con DBScan detectados como datos atípicos.

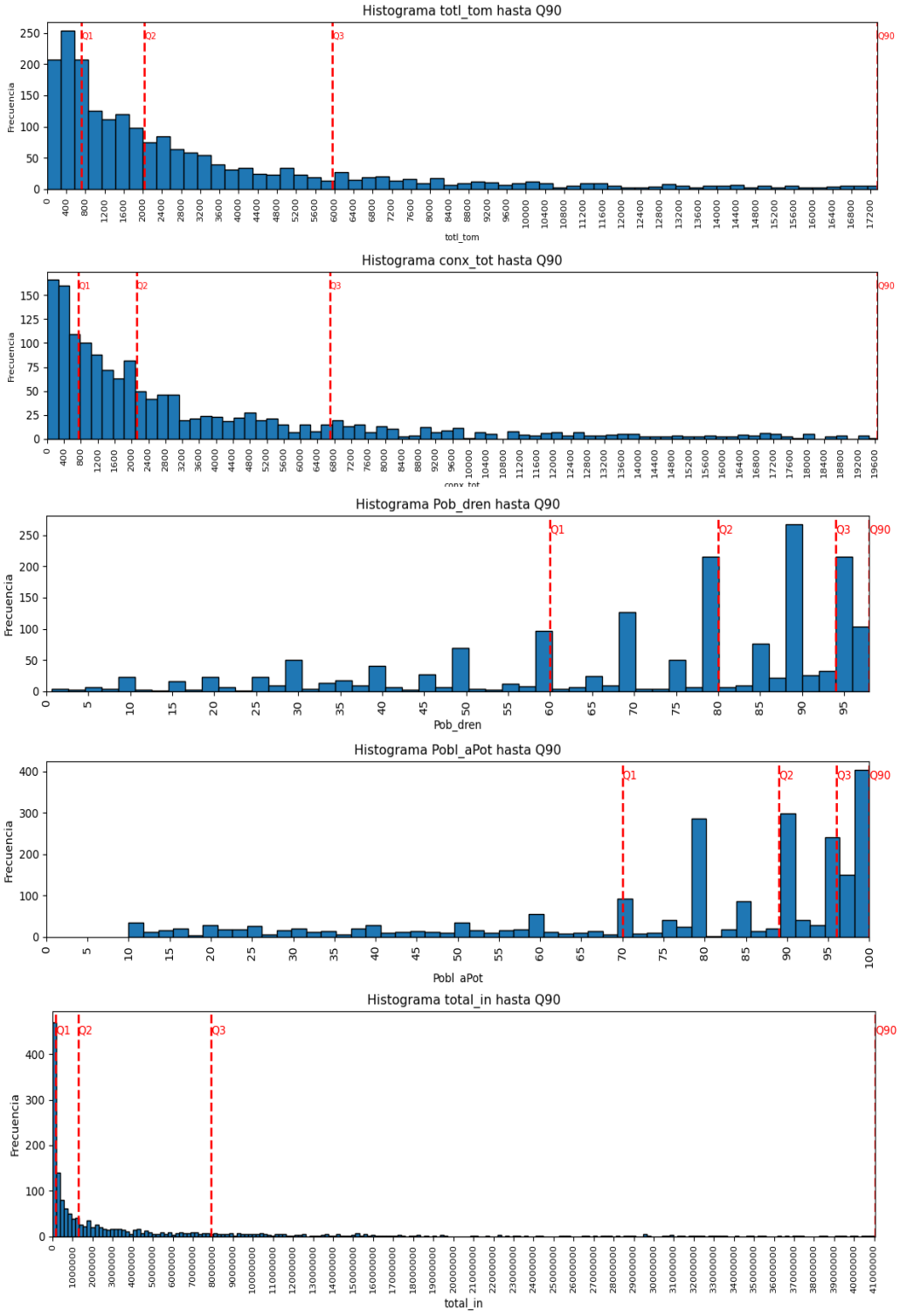

Figura A 2. Histogramas de las variables de estudio. Datos limpios.

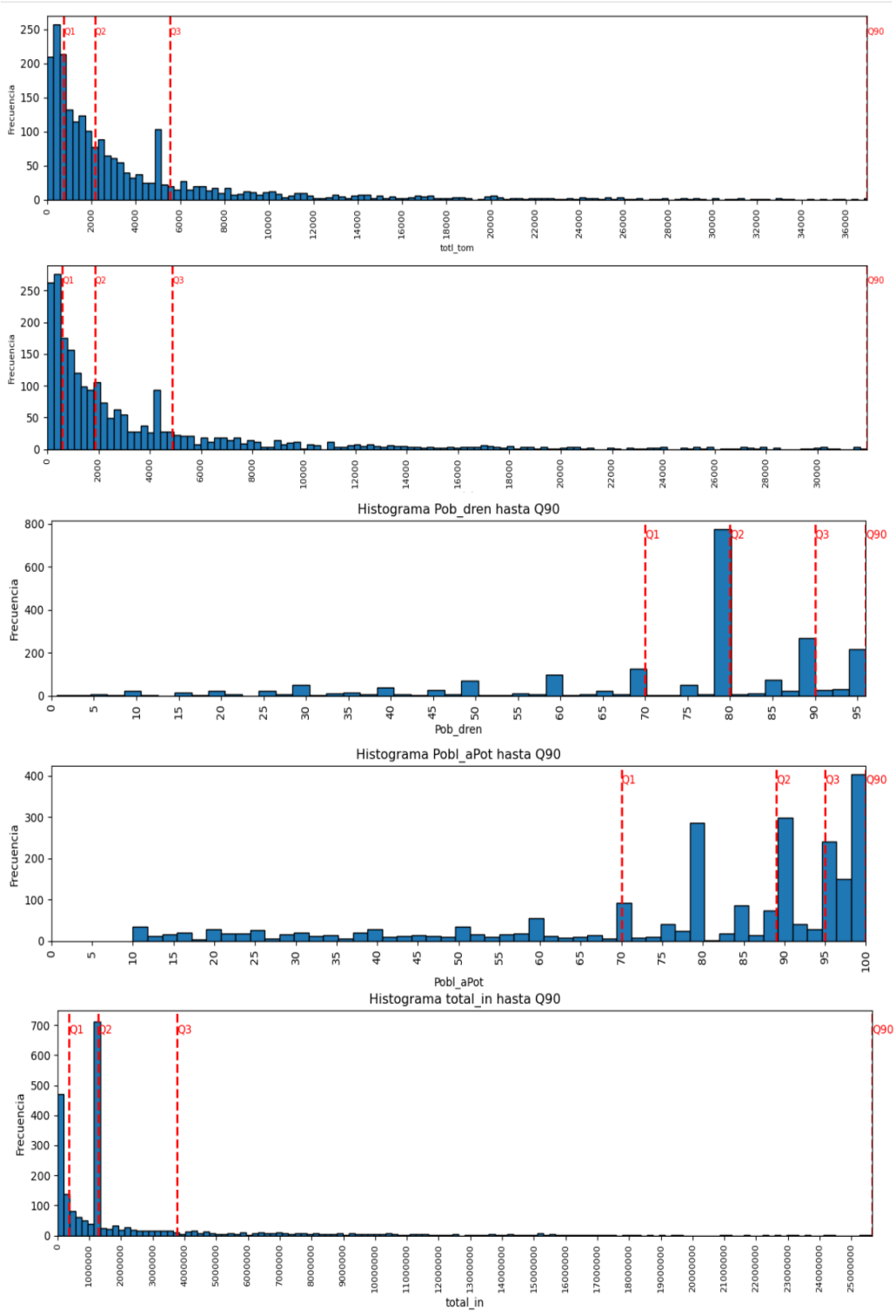

Figura A 3. Histogramas de las variables de estudio. Datos limpios e imputados.

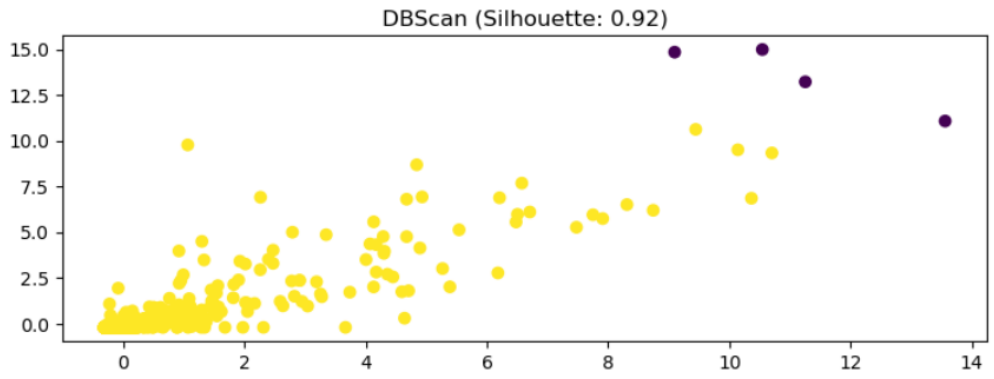

Figura A 4. Grafica de clúster con DBScan.

totl_tom	conx_tot	Pob_dren	Pobl_aPot	total_in	folio	cve_ent	Name_Ent	Name_Mun	cluster_db
356,964.000	315,632.000	84.000	98.000	1,690,761,382.000	2002	2	Baja California	Mexicali	-1
428,144.000	389,221.000	80.000	89.000	1,420,984,395.000	15033	15	Estado de México	Ecatepec de Morelos	-1
335,097.000	331,865.000	96.000	99.000	1,913,074,128.970	19039	19	Nuevo León	Monterrey	-1
290,449.000	254,564.000	80.000	80.000	1,895,036,863.000	22014	22	Querétaro	Querétaro	-1

Tabla A 12. Lista de folios seleccionados con DBScan detectados como datos atípicos.

	count	mean	std	min	25%	50%	75%	max
totl_tom	2,394.000	9,619.802	30,878.242	4.000	753.250	2,147.500	5,527.500	428,144.000
conx_tot	2,394.000	8,837.775	29,300.103	4.000	600.000	1,859.000	4,869.250	389,221.000
Pob_dren	2,394.000	75.524	21.514	0.600	70.000	80.000	90.000	100.000
Pobl_aPot	2,394.000	77.642	24.645	10.000	70.000	89.000	95.000	100.000
total_in	2,394.000	25,314,140.609	125,935,838.495	321.000	365,280.000	1,271,509.000	3,755,738.155	1,913,074,128.970

Tabla A 13. Principales estadísticos de las variables de estudio. Imputación de datos nulos.

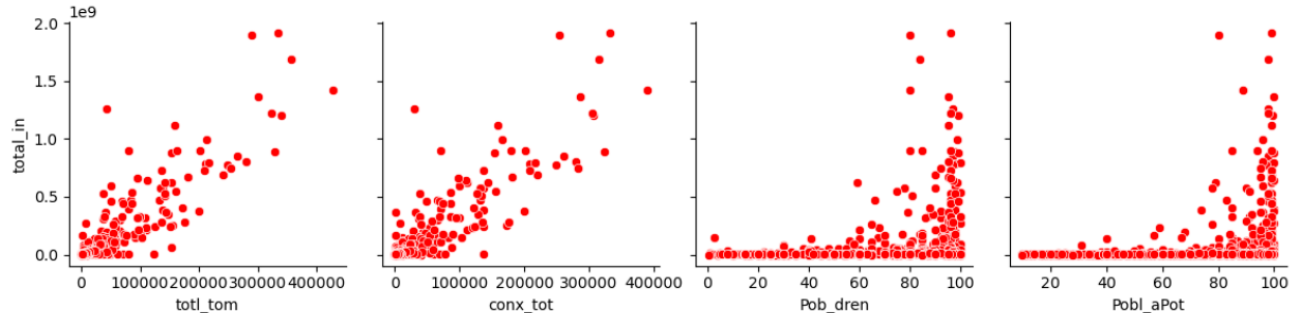

Figura A 5. Gráfica de dispersión de las variables de estudio. Datos limpios e imputados

	count	mean	std	min	25%	50%	75%	max
totl_tom	20.000	268,229.100	69,984.273	159,115.000	212,089.750	259,987.500	324,366.500	428,144.000
conx_tot	20.000	257,065.522	61,816.588	159,115.000	208,119.250	257,824.000	304,789.500	389,221.000
Pob_dren	20.000	93.124	6.661	80.000	91.468	96.000	97.200	100.000
Pobl_aPot	20.000	96.671	4.704	80.000	96.000	98.500	99.000	100.000
total_in	20.000	1,083,648,154.624	387,682,696.940	689,929,731.810	791,167,595.438	896,004,707.675	1,258,350,764.348	1,913,074,128.970

Tabla A 14. Principales estadísticos de las variables de estudio de la Clase 5.

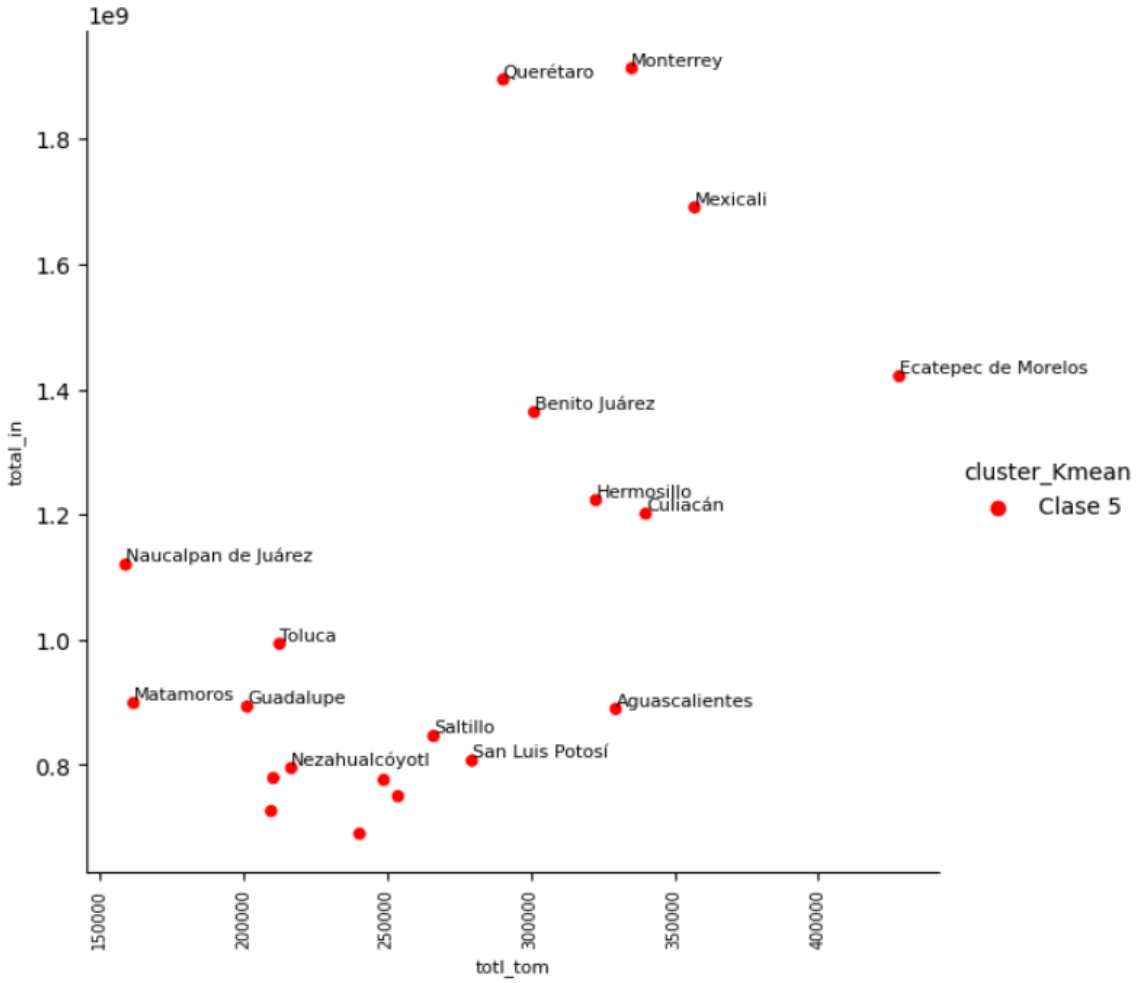

Figura A 6. Gráfica de dispersión del Ingreso con respecto a las tomas de la Clase 5.

count	mean	std	min	25%	50%	75%	max
57.000	108,525.258	39,279.414	38,063.000	76,602.000	100,736.000	142,048.000	200,470.000
57.000	99,790.023	40,786.046	74.000	69,862.000	94,363.000	132,397.000	200,040.000
57.000	90.703	10.095	59.000	88.000	95.000	97.000	100.000
57.000	93.058	9.094	57.000	91.000	96.000	99.000	100.000
57.000	391,604,825.901	232,751,734.675	1,271,509.000	231,225,911.000	348,488,763.900	527,094,536.390	1,256,686,180.000

Tabla A 15. Principales estadísticos de las variables de estudio de la Clase 4.

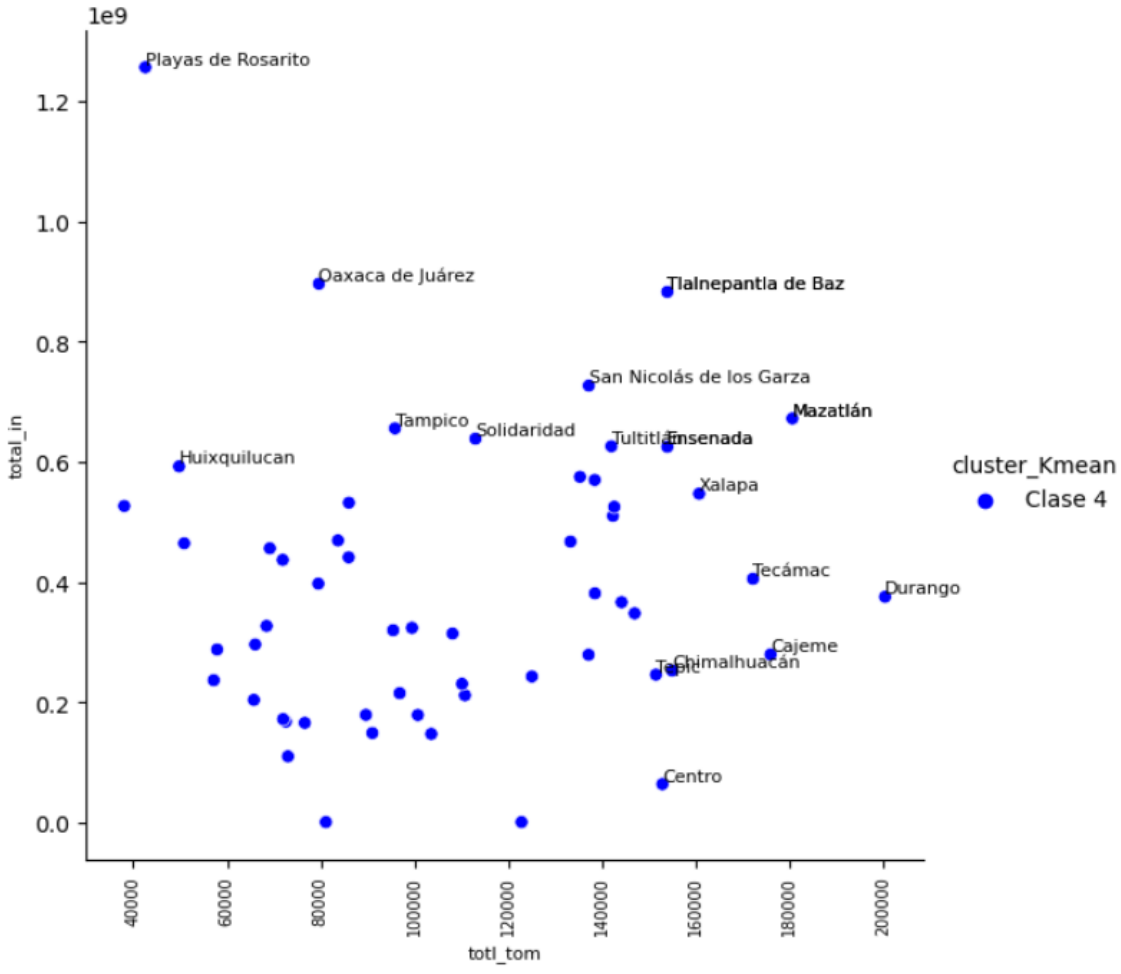

Figura A 7. Gráfica de dispersión del Ingreso con respecto a las tomas de la Clase 4.

	count	mean	std	min	25%	50%	75%	max
totl_tom	1,571.000	5,909.569	9,224.634	4.000	910.500	2,516.000	6,099.000	70,645.000
conx_tot	1,571.000	5,418.881	8,835.853	4.000	797.680	2,213.000	5,362.500	70,645.000
Pob_dren	1,571.000	84.143	10.482	50.000	80.000	80.000	94.000	100.000
Pobl_aPot	1,571.000	89.932	9.072	62.000	83.000	90.000	98.000	100.000
total_in	1,571.000	8,955,772.170	26,852,362.221	1,000.000	481,000.000	1,271,509.000	4,587,818.585	363,176,392.800

Tabla A 16. Principales estadísticos de las variables de estudio de la Clase 3.

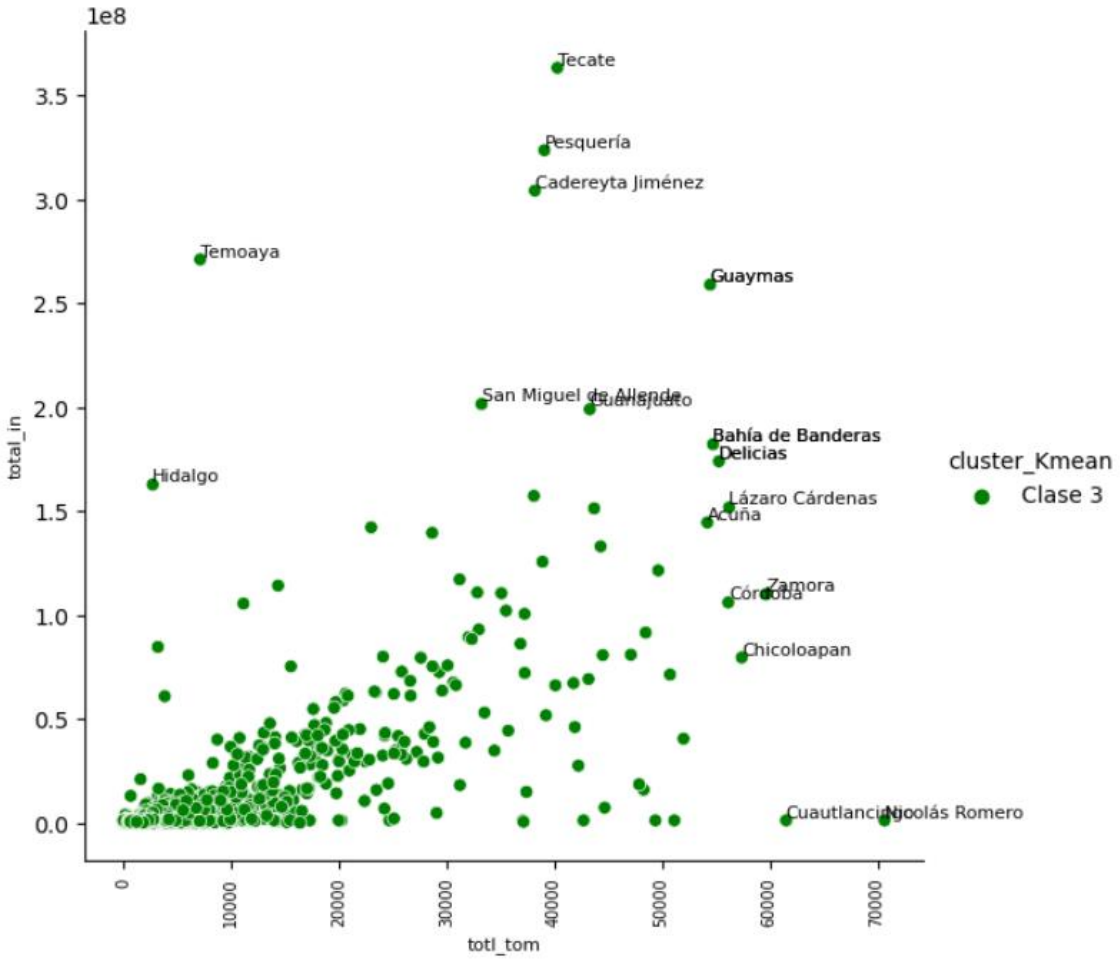

Figura A 8. Gráfica de dispersión del Ingreso con respecto a las tomas de la Clase 3

	count	mean	std	min	25%	50%	75%	max
totl_tom	399.000	3,653.638	5,906.286	79.000	690.000	1,762.000	3,679.000	58,183.000
conx_tot	399.000	2,655.520	4,617.663	11.000	400.000	1,177.000	2,686.500	40,000.000
Pob_dren	399.000	34.827	15.286	0.600	25.000	35.000	48.215	63.000
Pobl_aPot	399.000	61.986	26.547	10.000	40.000	70.000	80.000	100.000
total_in	399.000	4,052,318.967	13,136,540.625	3,390.000	315,455.035	1,271,509.000	1,584,605.560	142,660,111.250

Tabla A 17. Principales estadísticos de las variables de estudio de la Clase 2.

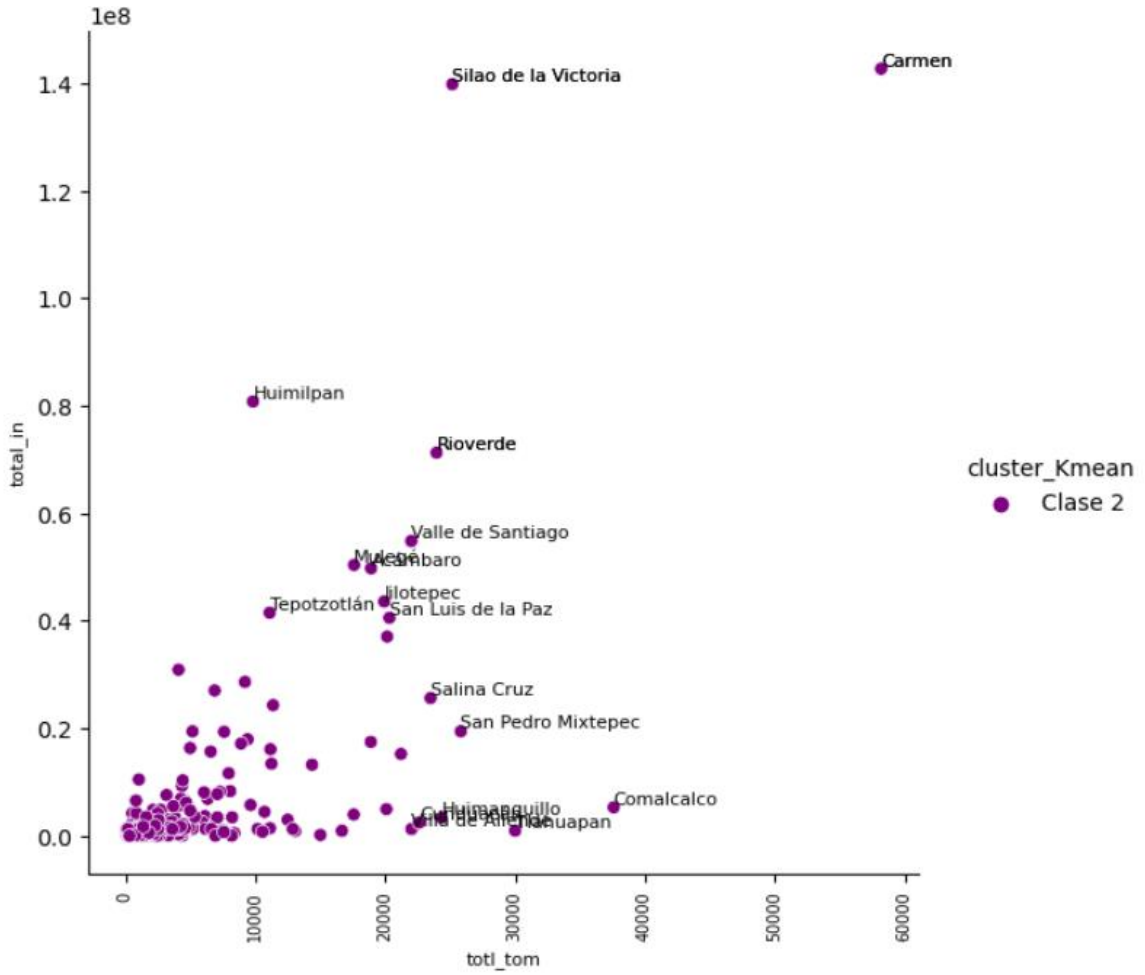

Figura A 9. Gráfica de dispersión del Ingreso con respecto a las tomas de la Clase 2.

	count	mean	std	min	25%	50%	75%	max
totl_tom	347.000	2,125.500	3,492.313	30.000	343.500	850.000	2,475.000	36,269.000
conx_tot	347.000	2,177.743	3,683.194	38.587	321.820	800.000	2,476.927	39,983.000
Pob_dren	347.000	79.791	9.679	45.000	80.000	80.000	80.000	100.000
Pobl_aPot	347.000	36.376	16.304	10.000	22.000	37.000	51.500	63.000
total_in	347.000	2,654,815.858	9,987,260.579	321.000	93,400.000	1,271,509.000	1,271,509.000	141,761,159.290

Tabla A 18. Principales estadísticos de las variables de estudio de la Clase 1.

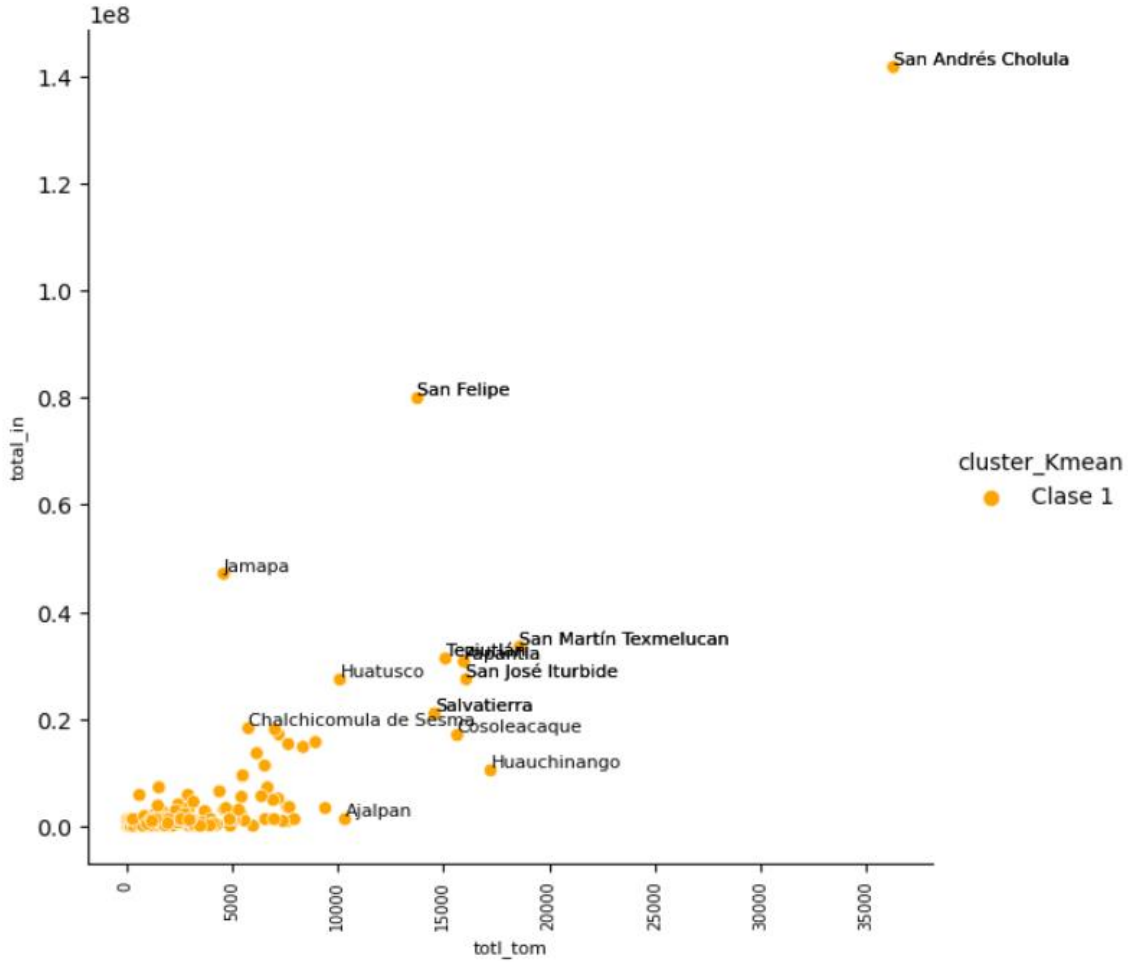

Figura A 10. Gráfica de dispersión del Ingreso con respecto a las tomas de la Clase 1