**Logotipo

Descripción generada automáticamente**

**Elaboración de un modelo de clustering de clientes del servicio de energía en Antioquia**

Alejandro Angel Ceballos y Pedro Jesus Tarazona Cote

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
David Manuel Villanueva Valdés, MSc

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

|  |  |
| --- | --- |
| **Cita** | (Angel Ceballos & Tarazona Cote, 2024) |
| **Referencia**  **Estilo APA 7 (2020)** | Angel Ceballos, A., & Tarazona Cote, P. (2024). *Elaboración de un modelo de clustering de clientes del servicio de energía en Antioquia* [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteVIII.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Tabla de contenido**

[1.](#_35nkun2) Descripción del problema 7

[1.1.](#_1ksv4uv) Problema de negocio 7

[1.2.](#_44sinio) Aproximación desde la analítica de datos 9

[1.3.](#_z337ya) Origen de los datos 9

[1.4.](#_3j2qqm3) Métricas de desempeño 11

[2.](#_4i7ojhp) Objetivos 12

[2.1.](#_2xcytpi) Objetivo general 12

[2.2.](#_3whwml4) Objetivos específicos 12

[3.](#_2bn6wsx) Datos 12

[3.1.](#_qsh70q) Datos originales 13

[3.2.](#_3as4poj) Datasets 16

[3.3.](#_1pxezwc) Analítica descriptiva 18

[Referencias](#_1jlao46) 24

[Anexos](#_43ky6rz) 26

[Anexo 1. Autoarchivo en Repositorio y documentos de interés 26](#_xvir7l)

**Lista de tablas**

**Tabla 1** Riesgos asociados al proyecto10

**Tabla 2** Descripción de los archivos de consumo de servicios13

[**Tabla 3** Diccionario de datos para el archivo de consumo base: BaseEnergia\_202408.csv](#_19c6y18) 14

[**Tabla 4**](#_3tbugp1) Descripción de los archivos de PQR para el servicio de energía.15

**Tabla 5** Diccionario de datos de los archivos de PQR del servicio de energía16

**Lista de figuras**

[**Figura 1** Representación gráfica del proceso de preparación de los datos.](#_2r0uhxc) 16

[**Figura 2** Vista del DataFrame resultante de la unión de los archivos de consumo y PQR procesados](#_1664s55) 18

[**Figura 3** Conteo de PPS por Tipo de Medida](#_1664s55) 19

[**Figura 4** Conteo de PPS por Ubicación](#_1664s55) 19

[**Figura 5** Conteo de PPS por TipoAreaEspecial](#_1664s55) 20

[**Figura 6** Conteo de PPS por Uso - Estrato](#_1664s55) 21

[**Figura 7** Conteo de PPS por Región 2](#_1664s55)1

[**Figura 8** Describe de consumo de energía por Uso - Estrato 22](#_1664s55)

[**Figura 9** Describe de consumo de energía por Región 2](#_1664s55)3

**Siglas, acrónimos y abreviaturas**

**EPM** Empresas Públicas de Medellín

**PPS** Puntos de Prestación de Servicio

**MME** Ministerio de Minas y Energía

**SUI** Sistema Único de Información

**SSPD** Superintendencia de Servicios Públicos Domiciliarios

**PQR** Peticiones Quejas y Reclamos

**CVI** Cluster Validity Index

# Descripción del problema

El direccionamiento estratégico y el informe de gestión del año 2023 de Empresas Públicas de Medellín (EPM), establece brindar soluciones ágiles e innovadoras para sus más de 2,8 millones de clientes en el departamento de Antioquia ubicando al cliente como pilar principal [[1]](https://www.zotero.org/google-docs/?4o6vlf) [[2]](https://www.zotero.org/google-docs/?01Envy). La percepción de los clientes frente a EPM, las cuales se detallan en el **Anexo 1**, realizadas de forma interna y externa tales como las elaboradas por la firma merco o el informe “Medellín cómo vamos”, indican una disminución en la que la experiencia del cliente juega un papel fundamental, por lo que se hace relevante generar estrategias encaminadas en mejorar dicha percepción. Adicionalmente, también puede evidenciarse que la experiencia del cliente es importante en empresas del sector de *utilities,* como por ejemplo, en la empresa de servicios públicos de Ocaña ESPO S.A. ESP, en donde se realizó una medición de experiencia del cliente con los servicios prestados por la entidad, se identificaron falencias en la calidad de los servicios derivando en una propuesta de solución basado en dicha medición [[3]](https://www.zotero.org/google-docs/?SXTlHa).

A través de la elaboración de un modelo de *clustering,* analizando los patrones de consumo de los Puntos de Prestación de Servicio (PPS) de EPM correspondientes al servicio de energía, con datos históricos de 3 años de consumo de los servicios públicos domiciliarios de los diferentes PPS con sus respectivas características, se pretende suministrar insumos a diferentes dependencias de EPM con el fin de que estas enfoquen estrategias de servicios personalizados acorde a la necesidad de cada cliente contribuyendo al mejoramiento de los indicadores referenciados anteriormente.

## Problema de negocio

En la actualidad, en las empresas prestadoras de servicios públicos de energía en Colombia, la experiencia del cliente viene tomando relevancia en sus estrategias de relacionamiento con sus clientes debido al aumento del ámbito competitivo que se viene dando en la comercialización del servicio de energía motivado, entre otros, por la resolución 40072 de 2018 del Ministerio de Minas y Energía (MME) en donde se establecen sistemas de medición avanzadas en el que dentro de sus objetivos se encuentra “Dinamizar la competencia en la comercialización minorista de energía eléctrica y generar nuevos modelos de negocio y servicios.” [[4]](https://www.zotero.org/google-docs/?AfhVlc).

Como se expuso en el anterior apartado, la experiencia del cliente juega un papel fundamental, debido a que la generación de valor en la reputación depende de cómo sus grupos de interés la califican, dicha calificación determina el nivel de reputación de esta y un factor que incide dentro de ella es la experiencia del cliente (*Customer Experience*); que desde el ámbito empresarial, se asocia a las emociones que se generan en los clientes como consecuencia de las decisiones y acciones que se realizan desde dicha compañía [[5]](https://www.zotero.org/google-docs/?quOZCb). Además, el hecho de que exista competencia en el mercado de la comercialización de energía eléctrica, conduce a las organizaciones cuyos clientes pueden elegir a cualquier comercializadora, tal como se realiza hoy en día en las empresas de telecomunicaciones en Colombia. Por lo tanto, las grandes empresas han entendido que una organización que se enfoque al cliente tendrá éxito si logra tomar decisiones estratégicas de marketing, operativas y organizacionales, basado en mantener un permanente conocimiento de su cliente bajo sus motivaciones, preferencias y comportamientos [[6]](https://www.zotero.org/google-docs/?F3hAAk).

Para el caso de EPM, teniendo en cuenta el establecimiento del cliente como pilar fundamental, hace necesario que desde la compañía se tenga una asistencia proactiva de los clientes en donde se ofrecen soluciones antes de que el cliente tenga que contactar el servicio al cliente de forma personalizada y lo más consecuente posible a las necesidades de los clientes.

Debido a la diversidad de los clientes de EPM por circunstancias tales como: el alto volumen de clientes, la cobertura amplia del servicio en diferentes territorios, tipos de comportamientos de consumo, condiciones económicas, usos del servicio, entre otros, se hace complejo ofrecer dichas soluciones o estrategias con la finalidad de responder a las necesidades de los clientes de forma personalizada, por lo tanto, se plantea la elaboración de una caracterización de los clientes a través de la creación de un modelo de *clustering* para analizar los patrones de consumo en los PPS de EPM, específicamente en relación con el servicio de energía, en donde se busca utilizar datos históricos de consumo de los servicios públicos domiciliarios registrados en los últimos tres años. Este análisis se desarrollará considerando las características particulares de cada PPS y tiene como objetivo proporcionar información valiosa a las distintas áreas de la compañía. Con esta información, se espera que dichas áreas puedan diseñar estrategias de servicios personalizados que respondan mejor a las necesidades de cada cliente, contribuyendo así a la mejora de la medición de la percepción de los clientes frente a EPM.

## Aproximación desde la analítica de datos

Se busca utilizar datos históricos de consumo en un marco de tiempo de 3 años correspondiente a los servicios públicos domiciliarios, para la creación de un modelo de *clustering* que permita analizar los patrones de consumo en los PPS de EPM e identificar información que pueda servir como insumo para el desarrollo de futuras estrategias de comunicación, orientadas a educar a los usuarios sobre el uso racional de la energía o a advertir sobre posibles fallos técnicos, sin embargo, cabe resaltar que el proyecto no contempla la implementación de dichas estrategias de comunicación, ni propondrá sugerencias específicas para mejorar el consumo energético en caso de detectar anomalías.

Es conocido que el potencial analítico de los datos de consumo energético son comúnmente desaprovechados [[7]](https://www.zotero.org/google-docs/?A2BKqU). Según Liu y Yen [[8]](https://www.zotero.org/google-docs/?hoOwSX), el desarrollo y gestión de servicios en el sector público puede ser mejorado por medio de herramientas de analítica de datos, por lo que el estudio de estos podría representar un valor agregado al brindar conocimiento sobre los clientes, potencialmente reducir los tiempos de reacción, incentivar el consumo sostenible de energía [[9]](https://www.zotero.org/google-docs/?NryGbd) y por consiguiente, reforzar su experiencia positiva.

Por ejemplo, desde el *Laboratory for Open Systems and Networks*, se han implementado modelos de segmentación sobre datos de consumo energético [[10]](https://www.zotero.org/google-docs/?u7XmUY), con el fin de caracterizar los diferentes patrones de consumo y se resalta el valor de los grupos generados por su utilidad para la formulación de estrategias de gestión del consumo energético y planeación de red, detección de posibles robos de energía, fallas en medidores o consumos irregulares.

En Colombia, Celsia, empresa dedicada a la generación y comercialización de energías renovables, cuenta con un sistema de monitoreo inteligente de energía, el cual ha permitido reducir consumos energéticos de sus clientes hasta un 30% [[11]](https://www.zotero.org/google-docs/?GnOuZX). Por otro lado, EnelX, es una empresa que impulsa la transformación energética de sus clientes ofreciendo servicios de gestión del consumo energético basado en análisis de los mismos [[12]](https://www.zotero.org/google-docs/?VBVPnH).

Este enfoque basado en datos podría incrementar el conocimiento de los clientes y de sus posibles necesidades de acuerdo a sus patrones de consumo. De igual forma, se podría esperar la detección de consumos inusuales y anomalías que no podrían ser vistos por medio de los métodos convencionales.

Es importante considerar los riesgos asociados a la ejecución y desarrollo de este proyecto, consideramos que estos pueden estar agrupados por categorías como se presenta en la siguiente tabla.

**Tabla 1**

*Riesgos asociados al proyecto*

|  |  |
| --- | --- |
| **Tipos de Riesgos** | **Riesgos** |
| **Técnicos** | Resolución temporal insuficiente, al no tener disponibilidad de una mayor granularidad de los datos, podrían perderse patrones importantes. |
| Los algoritmos de *clustering*, especialmente en *datasets* grandes, pueden ser computacionalmente intensivos, lo que dificulta experimentos exhaustivos [[13]](https://www.zotero.org/google-docs/?NKdPjQ). |
| Recursos de cómputo, al trabajar con múltiples fuentes de datos de gran volumen, es posible que el mismo procesamiento de los datos sea dispendioso. |
| **Éticos y Sociales** | Los *clusters* podrían identificar indirectamente características sensibles como nivel socioeconómico, ubicación geográfica, o patrones que podrían ser interpretados como discriminatorios. |
| Las estrategias a implementar por las dependencias derivadas de los *clusters,* podrían favorecer a ciertos grupos de clientes en detrimento de otros. |

## 

## Origen de los datos

El conjunto de datos para el servicio de energía fue obtenido de los formatos CSV certificados o pendientes por certificar por EPM al Sistema Único de Información (SUI) de la Superintendencia de Servicios Públicos Domiciliarios (SSPD), correspondiente al inventario de suscriptores y facturación, adicionalmente, fueron complementados por medio de consultas en SQL Server con información de la bodega comercial de datos ALFA de EPM.

Para el caso de las Peticiones Quejas y Reclamos (PQR), los servicios de acueducto, alcantarillado y gas, la información fue obtenida mediante consultas en SQL Server a tablas que se encuentran consignadas en la bodega comercial de datos ALFA de EPM.

Todos los conjuntos de datos fueron obtenidos para un periodo de tiempo de enero de 2021 a agosto de 2024, teniendo como resultado 5 archivos (4 de servicios públicos y 1 de PQR) para cada mes del periodo, resultando en un total de 225 archivos CSV.

La información se encuentra debidamente anonimizada para cada PPS, garantizando la protección de los datos personales de los clientes, sin embargo, la información puede considerarse clasificada y reservada por lo que el conjunto de datos será utilizado exclusivamente para la finalidad de este proyecto y el detalle de la información no será compartido a terceros.

## Métricas de desempeño

Se busca seleccionar un modelo de *clustering* que agrupe los usuarios en grupos de acuerdo a patrones de consumo similares entre sí, por lo que se debe evaluar la capacidad de cada variación del modelo de agrupar los clientes a partir de unos lineamientos predefinidos. Es conocido que distintos modelos se ajustan mejor a diferentes conjuntos de datos [[14]](https://www.zotero.org/google-docs/?2j9u7d) y para este caso, se pretende seleccionar un número fijo de *clusters* que puedan ser fácilmente atendidos por las dependencias respectivas dentro de EPM, con esta restricción, es necesario seleccionar un *Cluster Validity Index* (CVI) que priorice la métrica que mejor describa el caso de estudio [[15]](https://www.zotero.org/google-docs/?eJCNpn). Para este caso se pretende usar índices internos, ya que no se tiene conocimiento previo de las etiquetas de los *clusters* a los que pertenece cada usuario, específicamente, se utilizará el Coeficiente de Silueta, que permite medir qué tan similar es un elemento a su propio *cluster* comparado con los demás *clusters* y para la selección del número óptimo de grupos a generar se implementa el método del codo; en general, se pretende escoger el modelo que entregue las segmentaciones más compactas entre sus propios elementos y más separadas entre segmentos distintos. Estas propiedades garantizan que cada segmento generado tenga usuarios con características y patrones de consumo similares, por lo que estos usuarios compartirán necesidades similares entre sí, que podrán ser atendidas por medio de estrategias enfocadas a sus requerimientos específicos.

# Objetivos

## Objetivo general

Caracterizar los PPS del servicio de energía asociado al mercado regulado de EPM en el departamento de Antioquia, basado en información adquirida en el periodo de tiempo de enero de 2021 a agosto de 2024, con el fin de proveer un insumo a las diferentes dependencias de EPM de modo que estas enfoquen esfuerzos de gestión en pro de mejorar la experiencia del cliente.

## Objetivos específicos

* Consolidar la información histórica en el marco de tiempo definido, con el fin de generar un registro único por cada PPS, identificando las variables relevantes que inciden en la caracterización del patrón de consumo de estos.
* Determinar las características de tamaño y densidad de los segmentos a generar, con base al insumo que se desee suministrar para seleccionar un *Cluster Validity Index* (CVI) que priorice las características deseadas.
* Desarrollar y comparar diferentes modelos de segmentación con base al CVI seleccionado.
* Describir los clusters generados desde la perspectiva de negocio, para proveer como insumo a las dependencias respectivas.

# Datos

En esta sección, se presentan las características de procedencia de los datos empleados durante este proyecto, descripción de las variables y tipos de datos de las mismas, transformaciones y particularidades de cada una de ellas, como también el análisis descriptivo con el fin de entender las propiedades y utilidad de la información para el modelo a desarrollar.

## Datos originales

Cada archivo consigna la información de las características y consumos de los servicios de acueducto, alcantarillado, energía, gas, y adicionalmente información de PQR, para cada uno de los PPS.

Se dispone de datos de consumo de energía eléctrica de clientes anonimizados entre los años 2021 y 2024 del departamento de Antioquia, esta información, se compone de varios archivos CSV, como se menciona en la **Sección 1.3**, producto de consultas SQL realizadas a la base de datos comercial de EPM asociada a reportes sobre la facturación de consumo energético, registro de PQRs y consumos de los servicios de acueducto, alcantarillado y gas.

A continuación, se presenta la descripción de los archivos utilizados para los consumos, estos son 4 archivos mensuales correspondientes a cada servicio considerado.

**Tabla 2**  
*Descripción de los archivos de consumo de servicios.*

|  |  |  |
| --- | --- | --- |
| **Elemento** | **Descripción** | **Ejemplo** |
| Formato de nombre | Base[Servicio]\_YYYYMM.csv | BaseEnergia\_202408.csv, BaseAcueducto\_202304.csv |
| Servicios incluídos | Acueducto, Alcantarillado, Gas, Energía |  |
| Periodicidad | Mensual |  |
| Formato de archivo | CSV |  |
| Tamaño promedio (MB) | 118.1 - 665.1 por archivo dependiendo del servicio |  |
| Registros por archivo | 1,401,539 - 2,855,490 dependiendo del servicio y del mes |  |
| Columnas comunes | IDD\_PPS\_Anonimo, Consumo[Servicio]\_YYYYMM | ConsumoEnergia\_202408 |

De los archivos de consumo de energía se utiliza el del mes más reciente como base – BaseEnergia\_202408.csv – para establecer las características de cada PPS. En la siguiente tabla se presenta el diccionario de datos del archivo base, el cual contiene además de las columnas comunes que contienen los demás archivos de consumo presentadas en la **Tabla 2**, las características adicionales que se toman como referencia para cada PPS durante todo el análisis.

**Tabla 3**  
*Diccionario de datos para el archivo de consumo base: BaseEnergia\_202408.csv.*

|  |  |  |
| --- | --- | --- |
| **Columna** | **Descripción** | **Tipo de dato** |
| IDD\_PPS\_Anonimo | Identificador anónimo del Punto de Prestación de Servicio | String |
| Tipo de medida | Puede tomar el valor "Prepago" o "Pospago" | String |
| NivelTension | Indicador del nivel de tensión en kilovoltios, toma valores entre 0 - 5 y entre más alto el valor, mayor el nivel de tensión | Int64 |
| CodigoDane | Código estándar indicativo del departamento y municipio | String |
| Ubicacion | Valor numérico que indica; 1: Rural disperso, 2: Urbano y 3: Centro poblado | Int64 |
| TipoAreaEspecial | Valor numérico que indica; 0: Ninguna, 1: Barrio subnormal, 2: Área rural de menor desarrollo, 3: Zonas de difícil gestión | Int64 |
| Uso - Estrato | Estrato socioeconómico del 1 al 6 y de 7 a 11 indica usos Industrial, Comercial, Oficial, Provisional o Alumbrado público, respectivamente | Int64 |
| Altitud | Metros sobre el nivel del mar | Int64 |
| AutoGenerador | Valor numérico que indica si es o no autogenerador de la siguiente forma; 1: Autogenerador pequeña escala, 2: Autogenerador gran escala, 3: No | Int64 |
| ValorFacturacionConsumoUsuario | Valor facturado en pesos colombianos para el consumo respectivo del mes | Int64 |
| ValorSubsidioUsuario | Corresponde al valor facturado en pesos colombianos de subsidios aplicados en la factura a los usuarios que se vean beneficiados por estos | Int64 |
| ValorContribucion | Corresponde al valor facturado en pesos colombianos por concepto de "contribución de solidaridad" de acuerdo con la normatividad vigente | Int64 |
| ValorTotalFacturado | Corresponde al valor total facturado en pesos colombianos al usuario en el mes, el cual incluye conceptos inherentes a la prestación del servicio de energía | Int64 |
| VCVeces | Número de eventos de compensación al usuario en el mes | Int64 |
| ConexionRed | 1 si el tipo de conexión es aérea o 2 si es subterránea | Int64 |
| ConsumoEnergia\_202408 | Consumo energético en kWh para el mes | Int64 |

Se puede observar que hay columnas numéricas y categóricas*,* por lo que se tendrá que codificar estas últimas con el fin de poder hacer uso de las características en el modelo a desarrollar.

De igual forma, se presenta la descripción de los archivos de PQR, los cuales son 1 por cada mes y únicamente corresponden a PQRs del servicio de energía.

**Tabla 4**

*Descripción de los archivos de PQR para el servicio de energía.*

|  |  |  |
| --- | --- | --- |
| **Elemento** | **Descripción** | **Ejemplo** |
| Formato de nombre | Base\_PQR\_Energia\_YYYYMM.csv | Base\_PQR\_Energia\_202408.csv |
| Periodicidad | Mensual |  |
| Formato de archivo | CSV |  |
| Tamaño promedio (MB) | ~10 MB por archivo |  |
| Registros por archivo | ~41,000 dependiendo del mes |  |
| Columnas | IDD\_PPS\_Anonimo, Tipo\_Atencion |  |

El diccionario de datos para los archivos de PQR descritos en la **Tabla 4** se presenta a continuación, aquí se describe el contenido de cada una de las columnas incluídas en los archivos y su tipo de dato.

**Tabla 5**

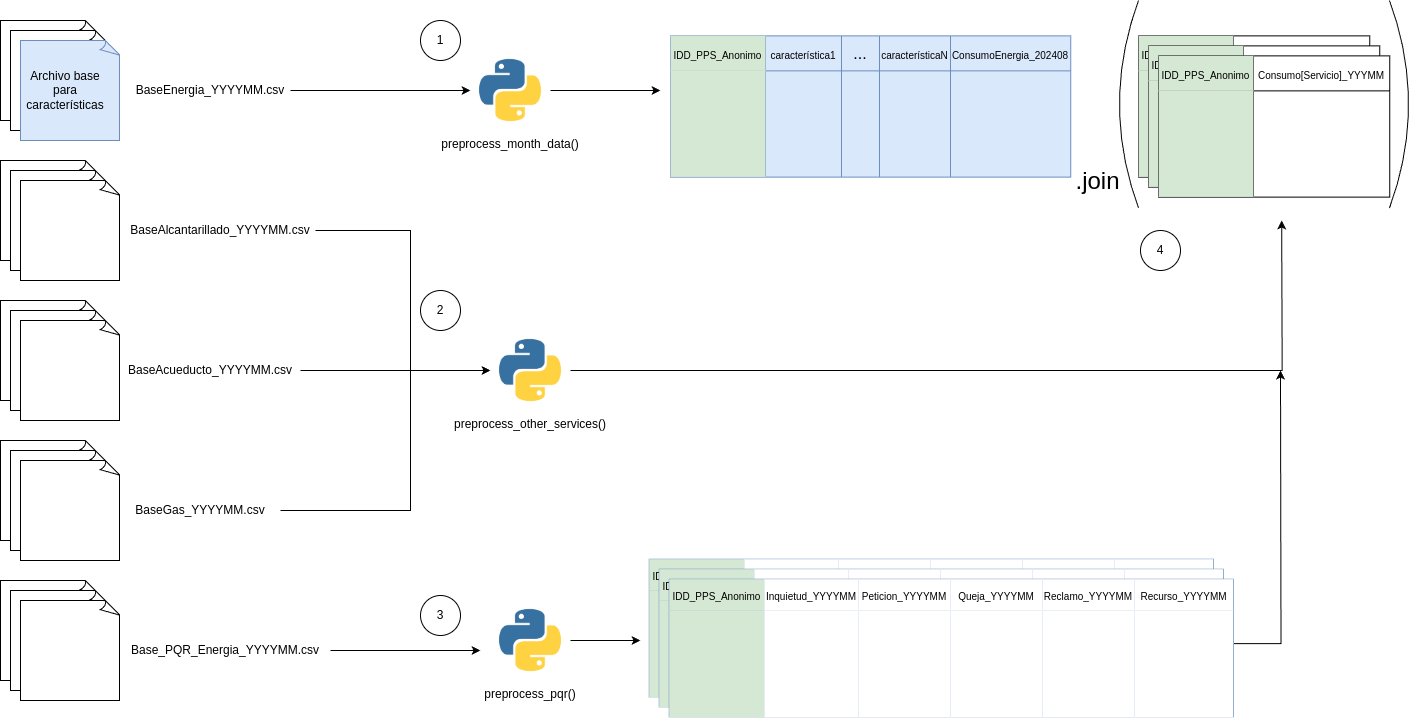
*Diccionario de datos de los archivos de PQR del servicio de energía.*

|  |  |  |
| --- | --- | --- |
| **Columna** | **Descripción** | **Tipo de dato** |
| IDD\_PPS\_Anonimo | Identificador anónimo del Punto de Pestación de Servicio | String |
| Tipo de atención | Hace referencia a la atención solicitada por el usuario, esta puede ser una de los siguientes: Inquietud, Petición, Queja, Reclamo o Recurso | String |

## Datsets

El proceso de limpieza y preparación de los datos se encuentra detallado en el notebook “ME03 - G10 - [1017261217]-[1098625287].ipynb” del repositorio de GitHub asociado al proyecto (<https://bit.ly/4fEmkKc>). A continuación, se presenta un diagrama que sintetiza los pasos de preparación y limpieza de los datos.

**Figura 1**

*Representación gráfica del proceso de preparación de los datos*

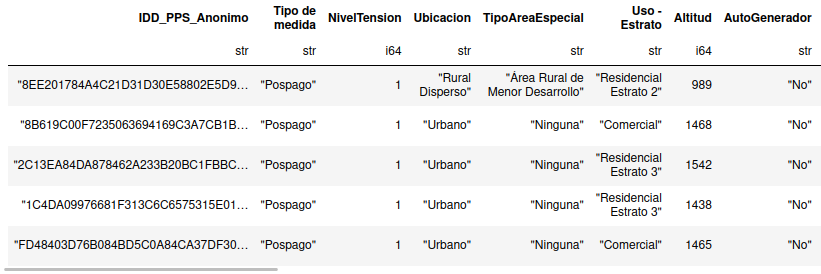
En el proceso ilustrado en la **Figura 1** se detallan tres funciones principales encargadas de la limpieza y preparación de cada uno de los archivos de consumo y PQR, y se muestra una simplificación de los *datasets* obtenidos con sus respectivas columnas después de cada transformación, finalmente se muestra cómo es el proceso de unión de todos los archivos procesados para obtener el *dataset* final.

La carga y preparación de los archivos fue implementada en *Polars*, lo cual permitió aprovechar su capacidad de procesamiento en paralelo. A continuación se describe el proceso señalado en cada punto de la **Figura 1.**

* En el punto 1 se utiliza la función *preprocess\_month\_data(),* encargada de recibir únicamente los archivos BaseEnergia\_YYYYMM.csv, aquí se ajustan los tipos de datos y se remueven las columnas innecesarias, posteriormente se agregan las características numéricas por cada usuario identificado por la columna “IDD\_PPS\_Anonimo”, se remueven registros duplicados y para el caso del archivo base descrito en la **Sección 3.1** (BaseEnergia\_202408.csv), se determinan las características del PPS que son utilizadas durante todo el desarrollo del proyecto. Para conocer el detalle de las características, ver la **Tabla 3**.
* En el punto 2 se utiliza la función *preprocess\_other\_services()*, la cual se encarga de procesar los archivos Base[Servicio]\_YYYYMM.csv de los servicios restantes; alcantarillado, acueducto y gas, también se encarga de rellenar los posibles consumos inexistentes para cada mes y renombrar las columnas de consumo para una fácil identificación de estas a la hora de unir todos los *DataFrames*.
* El proceso del punto 3, se encarga de preparar los datos de los archivos Base\_PQR\_Energia\_YYYYMM.csv. Se hace un conteo de cada uno de los tipos de atención por cada usuario y se asignan estos conteos a una nueva columna.
* Finalmente, en el paso 4, por medio de la columna “IDD\_PPS\_Anonimo”, se une el dataset del archivo base procesado junto con los datasets de los demás servicios y las PQR.

Después de ejecutar los procesos listados anteriormente, se obtiene el *dataset* resultante el cual cumple con las condiciones de un registro por identificador (IDD\_PPS\_Anonimo), y donde cada columna corresponde a características del PPS y consumos de servicios en cada mes. En la **Figura 2** se pueden ver 5 registros del dataset resultante y las 7 primeras características de cada PPS.

**Figura 2**

*Vista del DataFrame resultante de la unión de los archivos de consumo y PQR procesados*

## 

## Analítica descriptiva

Después del proceso de limpieza y preparación de los datos descritos en el numeral anterior, el análisis de la información se encuentra detallada en el notebook “ME03 - G10 - [1017261217]-[1098625287].ipynb” del repositorio de GitHub asociado al proyecto (<https://bit.ly/4fEmkKc>). El análisis se realiza por cada tipo de variable consignado en el *DataFrame* *df\_merge:*

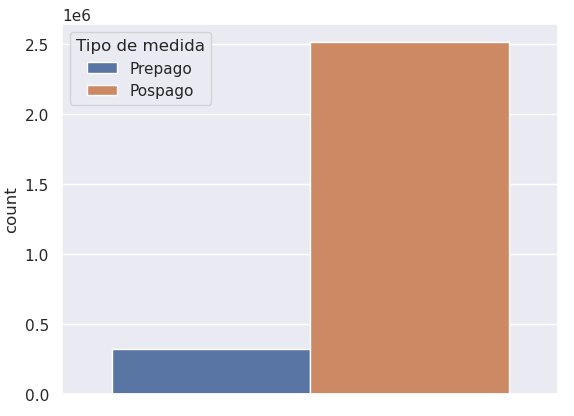
***Variables Categóricas:***

En el caso de las variables categóricas, el análisis se realiza de forma univariada con el fin de validar el comportamiento de los PPS y cómo cada variable aporta en datos de valor para el modelo:

* **Tipo de Medida:** Indica si los clientes se encuentran con tipo de medida prepago (realiza recarga para obtener el servicio de energía) o pospago (el consumo se mide con lecturas con una periodicidad específica cuya liquidación se emite con factura a suministrar a cada cliente). En la **Figura 3** se relaciona el conteo de PPS segmentados por la variable “Tipo Medida” observando consistencia en el comportamiento de los datos toda vez que el tipo de medida preponderante es el pospago:

**Figura 3**

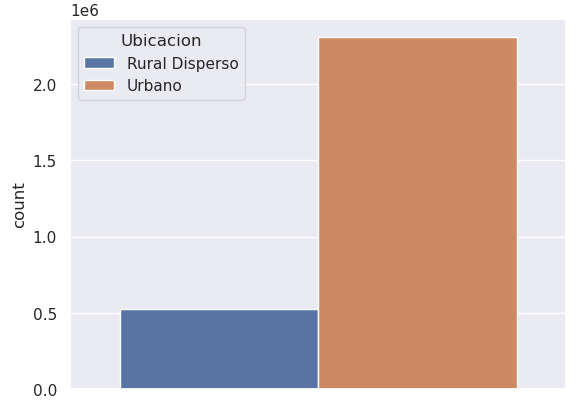
*Conteo de PPS por Tipo de medida*

**

* **Ubicación:** Indica la zona geográfica de donde se encuentran los clientes, en la **Figura 4** se observa que la concentración de los PPS en el departamento de Antioquia se encuentra en las zonas urbanas de los diferentes municipios:

**Figura 4**

*Conteo de PPS por Ubicación*

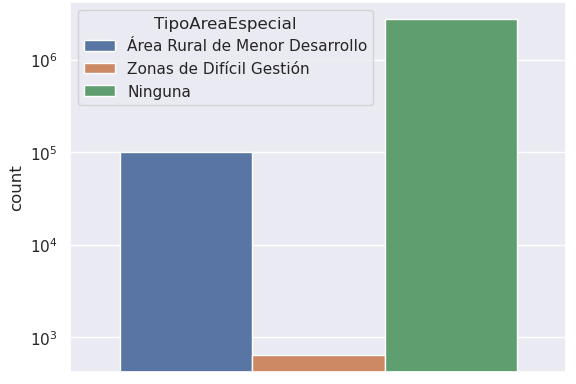


# 

# Tipo Area Especial y Autogenerador: Al realizar el análisis de dichas variables se encuentra que las mismas tienen concentrados la gran mayoría de sus PPS en solo uno de sus atributos. Incluso se debió aplicar escala logarítmica con el fin de poder visualizar la información. En la Figura 5 se relaciona la visualización de conteo de PPS con escala logarítmica para la variable “Tipo Area Especial”, al visualizar dicha participación se decide eliminar dichas variables debido a que no aportarían valor al modelo.

**Figura 5**

*Conteo de PPS por TipoAreaEspecial:*

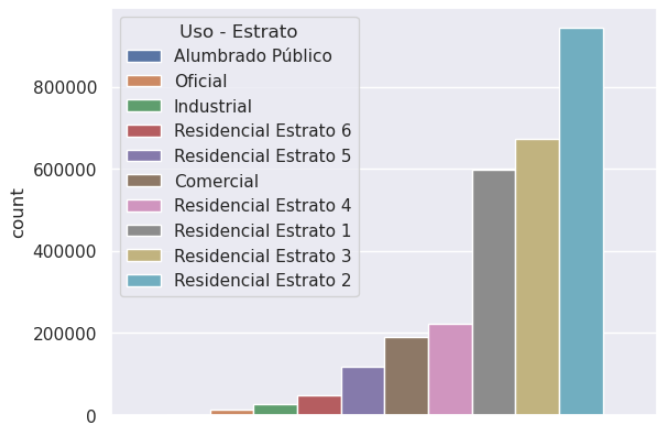
**

# 

* **Uso Estrato:** Corresponde a una de las variables más importantes debido a que indica el uso del servicio que un PPS le da a la energía eléctrica y es uno de los campos más relevantes para establecer estrategias de soluciones a los clientes. En la **Figura 6** se relaciona el conteo de PPS por dicha variable

**Figura 6**

*Conteo de PPS por Uso - Estrato:*

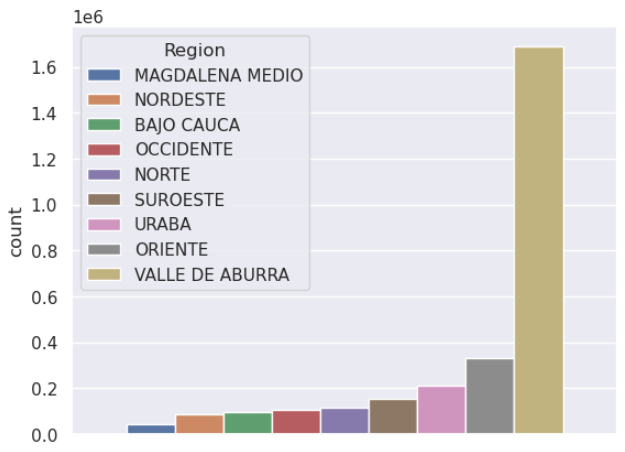
**

* **Región:** Dicha variable indica la distribución de los clientes por cada una de las regiones del departamento de Antioquia. En la **Figura 7** se evidencia que la distribución de los clientes se concentra en gran medida en el Valle de Aburrá

# 

**Figura 7**

*Conteo de PPS por Región*

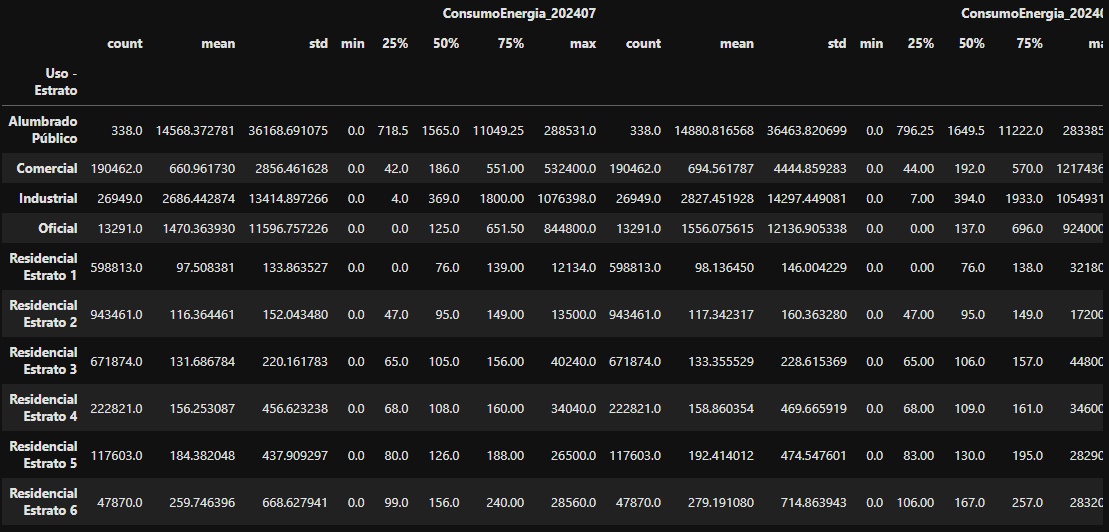
**

***Variables Numéricas:***

En el caso de las variables numéricas, el análisis de la información se realiza mediante un *describe* del consumo de energía tomando como referencia las variables categóricas “uso-estrato” y “Región” teniendo en cuenta que son las que más aportan a visualizar la información de forma consolidada. Cabe señalar que la información de las variables numéricas se encuentra principalmente conformada por el consumo de los diferentes servicios públicos y valores facturados. En la Figura 8 y Figura 9 se evidencia el describe realizado con respecto al “uso-estrato” y “Región” respectivamente:

**Figura 8**

*Describe de consumo de energía por Uso - Estrato*

**

**Figura 9**

*Describe de consumo de energía por Región*

**

Como se puede observar en los *describe* anteriormente referenciados, se evidencian valores máximos muy elevados comparados con la media y la tercera desviación estándar de cada uno de los uso-estratos y regiones. Para los valores mínimos es normal que se encuentren valores de 0 debido al no consumo por diferentes motivos en los PPS.

También se observa que la media de consumo de las regiones que se encuentran ubicadas en zonas de baja altitud sobre el nivel del mar tales como Urabá y Magdalena Medio, es más alta que en regiones con mayor altitud.

## 

## Referencias

[[1] Empresas Públicas de Medellín, “Direccionamiento estratégico”. [En línea]. Disponible en: https://www.epm.com.co/institucional/sobre-epm/quienes-somos/direccionamiento-estrategico.html#carousel-33422cf2c8-item-a6bdfc4d40-tabpanel](https://www.zotero.org/google-docs/?JHfFWG)

[[2] Empresas Públicas de Medellín, “Informe de Gestión EPM 2023”. [En línea]. Disponible en: https://www.epm.com.co/content/dam/epm/institucional/transparencia/rendicion-de-cuentas/rendicion-de-cuentas-2023/informe-de-gestion-version-final.pdf](https://www.zotero.org/google-docs/?JHfFWG)

[[3] S. Yaruro y P. Shazkia, “Evaluación de la satisfacción de los clientes externo e interno en la empresa de servicios públicos de Ocaña ESPO S.A, ESP, con el fin de proponer estrategias que mejoren la situación actual”. [En línea]. Disponible en: https://repositorioinstitucional.ufpso.edu.co/xmlui/handle/20.500.14167/2553](https://www.zotero.org/google-docs/?JHfFWG)

[[4] Comisión de Regulación Energía y Gas (CREG), “Resolución 40072 de 2018 Ministerio Minas y Energía”, Gestor Normativo. [En línea]. Disponible en: https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_minminas\_40072\_2018.htm](https://www.zotero.org/google-docs/?JHfFWG)

[[5] S. Ferruz-González, “Conceptualización de la Reputación Corporativa. Nuevo enfoque y propuesta.”, *Rev. Asoc. Esp. Investig. Comun.*, vol. 4, núm. 7, pp. 130–137, mar. 2017, doi: 10.24137/raeic.4.7.13.](https://www.zotero.org/google-docs/?JHfFWG)

[[6] N. L. Heredia y F. Fonseca, “La nueva tendencia de sostenibilidad y crecimiento de la Mipymes de actividad comercial en Colombia”. noviembre de 2021. [En línea]. Disponible en: http://scielo.sld.cu/pdf/rus/v13n6/2218-3620-rus-13-06-123.pdf](https://www.zotero.org/google-docs/?JHfFWG)

[[7] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, y P. Siano, “A comparative study of clustering techniques for electrical load pattern segmentation”, *Renew. Sustain. Energy Rev.*, vol. 120, p. 109628, mar. 2020, doi: 10.1016/j.rser.2019.109628.](https://www.zotero.org/google-docs/?JHfFWG)

[[8] W.-K. Liu y C.-C. Yen, “Optimizing Bus Passenger Complaint Service through Big Data Analysis: Systematized Analysis for Improved Public Sector Management”, *Sustainability*, vol. 8, núm. 12, p. 1319, dic. 2016, doi: 10.3390/su8121319.](https://www.zotero.org/google-docs/?JHfFWG)

[[9] S. Chandra y S. Verma, “Big Data and Sustainable Consumption: A Review and Research Agenda”, *Vis. J. Bus. Perspect.*, vol. 27, núm. 1, pp. 11–23, feb. 2023, doi: 10.1177/09722629211022520.](https://www.zotero.org/google-docs/?JHfFWG)

[[10] R. Kaur y D. Gabrijelčič, “Behavior segmentation of electricity consumption patterns: A cluster analytical approach”, *Knowl.-Based Syst.*, vol. 251, p. 109236, sep. 2022, doi: 10.1016/j.knosys.2022.109236.](https://www.zotero.org/google-docs/?JHfFWG)

[[11] J. C. González, “Monitoreo Inteligente de Energía de Celsia: Optimiza tus energéticos y reduce costos”. 2023. [En línea]. Disponible en: https://www.celsia.com/es/blog-celsia/monitoreo-inteligente-de-energia-de-celsia-optimiza-tus-energeticos-y-reduce-costos/](https://www.zotero.org/google-docs/?JHfFWG)

[[12] EnelX, “Democratización de la información energética”. 2024. [En línea]. Disponible en: https://www.enelxenergy.com/Home/blog/energia-para-empresas/democratizacion-de-informacion-en-consumo-de-energia-con-EnelX.html](https://www.zotero.org/google-docs/?JHfFWG)

[[13] M. A. Mahdi, K. M. Hosny, y I. Elhenawy, “Scalable Clustering Algorithms for Big Data: A Review”, *IEEE Access*, vol. 9, pp. 80015–80027, 2021, doi: 10.1109/ACCESS.2021.3084057.](https://www.zotero.org/google-docs/?JHfFWG)

[[14] M. Jain, T. AlSkaif, y S. Dev, “Validating Clustering Frameworks for Electric Load Demand Profiles”, *IEEE Trans. Ind. Inform.*, vol. 17, núm. 12, pp. 8057–8065, dic. 2021, doi: 10.1109/TII.2021.3061470.](https://www.zotero.org/google-docs/?JHfFWG)

[[15] M. Halkidi, Y. Batistakis, y M. Vazirgiannis, “On Clustering Validation Techniques”, *J. Intell. Inf. Syst.*, vol. 17, núm. 2/3, pp. 107–145, 2001, doi: 10.1023/A:1012801612483.](https://www.zotero.org/google-docs/?JHfFWG)

[[16] merco, “Monitor empresarial de reputación corporativa.” [En línea]. Disponible en: https://www.merco.info/co/](https://www.zotero.org/google-docs/?JHfFWG)

[[17] Empresas Públicas de Medellín, “Informe de Sostenibilidad”. [En línea]. Disponible en: https://sostenibilidadgrupoepm.com.co/wp-content/uploads/2024/08/5\_INFORME-SOSTENIBILIDAD-ESP\_2023\_unif.pdf](https://www.zotero.org/google-docs/?JHfFWG)

[[18] Medellín Cómo Vamos, “Informe Calidad de Vida 2020 - 2023”. [En línea]. Disponible en: https://www.medellincomovamos.org/system/files/2024-06/docuprivados/20240604\_INFORME CUATRIENIO\_3.pdf](https://www.zotero.org/google-docs/?JHfFWG)

[[19] Comisión de Regulación de Energía y Gas, “Indicadores de calidad de la atención al usuario del servicio público domiciliario de energía eléctrica.”](https://www.zotero.org/google-docs/?JHfFWG)

# Anexos

## Anexo 1. Mediciones de percepción de clientes frente a EPM.

Empresas Públicas de Medellín tiene varias formas de medir su reputación y el nivel de satisfacción de los clientes, de forma interna y externa. A continuación, se relacionan algunas de las mediciones más considerables:

***Medición Merco:***

La empresa Merco que realiza un monitoreo empresarial de reputación corporativa, la cual realiza dichas mediciones desde el año 2000, es un instrumento de evaluación reputacional basado en una metodología multistakeholder compuesta por seis evaluaciones y más de veinte fuentes de información [[16]](https://www.zotero.org/google-docs/?vIGo2B). De acuerdo con lo anterior, EPM presenta el siguiente ranking a lo largo de los últimos 7 años:

|  |  |
| --- | --- |
| Año | Posición |
| 2023 | 29 |
| 2022 | 28 |
| 2021 | 27 |
| 2020 | 14 |
| 2019 | 10 |
| 2018 | 8 |
| 2017 | 5 |

Diseño propio. Tomado de la búsqueda del término epm en<https://www.merco.info/co/buscar>

Como se puede evidenciar en la anterior tabla, EPM ha presentado disminución en el ranking, esto se debe a diversos factores, pero indica que se deben realizar acciones en pro de mejorar la reputación.

***Medición relacional EPM:***

EPM realiza medición relacional la cual monitorea seis fuerzas claves: trato justo, certeza control, estatus, pertinencia y disfrute, permitiendo una construcción con los clientes usuarios. El resultado para el 2023 corresponde a 48 (de 0 sobre 100), entre los factores que fueron determinantes en la medición, se encuentran la percepción de alza en las tarifas, falta de claridad de las facturas, entre otros [[17]](https://www.zotero.org/google-docs/?TruNWS).

Lo anterior, denota que los atributos asociados a consumos y facturación influyen en la medición que indica cómo recuerda el cliente la relación que tiene con EPM y las emociones que le genera.

***Medición transaccional EPM:***

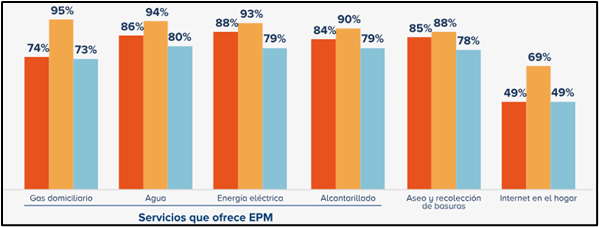
La medición transaccional o Net Promoter Score (NPS), el cual corresponde a un indicador internacional que determina qué tan dispuesto está un cliente a recomendar a EPM a partir de las experiencias que vivió con las transacciones realizadas, tiene una calificación de 39 para EPM en el año 2023 [[17]](https://www.zotero.org/google-docs/?M1RHbi)

La calificación en EPM fue aceptable teniendo en cuenta que una buena calificación se considera entre 30 y 60, sin embargo, EPM se encuentra cerca del límite inferior lo que genera la necesidad de movilizar recursos para mejorar dicho puntaje en la medición NPS.

***Encuesta Informe Medellín Cómo Vamos:***

Medellín Cómo Vamos, institución privada con 9 socios que realiza análisis y seguimiento a la calidad de vida en Medellín, realiza un informe anual [[18]](https://www.zotero.org/google-docs/?jpajeG), en el cual se tiene en cuenta aspectos asociados a la reputación que tiene EPM frente a los servicios que ofrece a la población del distrito de Medellín.

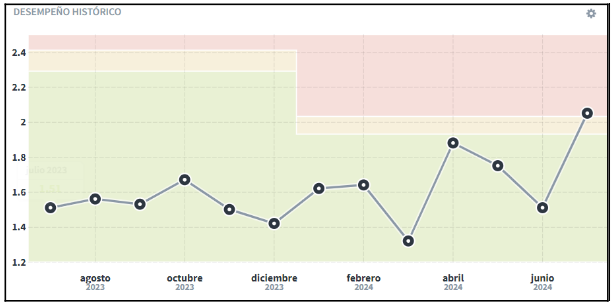
La imagen que se relaciona a continuación muestra el nivel de satisfacción con los servicios públicos recibidos en los años 2015, 2019 y 2023, donde un mayor porcentaje indica una mayor satisfacción de la población. De aquí se puede evidenciar un deterioro en la percepción asociada a los servicios de energía eléctrica, gas domiciliario, acueducto y alcantarillado con el paso del tiempo:

******

Satisfacción con los servicios públicos recibidos

***Informe indicador de reclamos EPM:***

Dentro de los indicadores que se trabajan en EPM se tiene una medición asociada a reclamos que interponen los clientes, dicho indicador se calcula teniendo en cuenta la cantidad de reclamos por cada 10000 facturas emitidas, dicha medida se sustenta de acuerdo a lo establecido en la resolución CREG-126 [[19]](https://www.zotero.org/google-docs/?WLi2Vk), en la gráfica que se muestra a continuación se evidencia una tendencia de aumento en el último año asociado a dicho indicador:



Tomado de<https://epm-wapp53.corp.epm.com.co/scorecards/124276>

Los reclamos indican inconformidades de los clientes sobre los valores facturados lo que indica una experiencia negativa por parte del cliente.