

Lineamientos para la gestión y analítica de datos en sistemas IoT de servicios hídricos urbanos

Mariana Marín Ruiz

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría Gestión de Información
Bogotá D.C., 20 de enero 2021 Colombia
2021**

Lineamientos para la gestión y analítica de datos en sistemas IoT de servicios hídricos urbanos

Mariana Marín Ruiz

Trabajo de investigación para optar al título de
Magíster en Gestión de Información

Director

Victoria Eugenia Ospina Becerra
Doctor en Gestión de Conocimiento y Organizaciones

Codirector

Dante Conti
PhD en Dirección de Proyectos

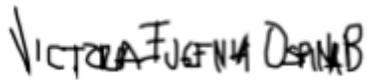
**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría en Gestión de Información
Bogotá D.C., 20 de enero Colombia
2021**

© Únicamente se puede usar el contenido de las publicaciones para propósitos de información. No se debe copiar, enviar, recortar, transmitir o redistribuir este material para propósitos comerciales sin la autorización de la Escuela Colombiana de Ingeniería. Cuando se use el material de la Escuela se debe incluir la siguiente nota “Derechos reservados a Escuela Colombiana de Ingeniería” en cualquier copia en un lugar visible. Y el material no se debe notificar sin el permiso de la Escuela.

Publicado en 2020 por la Escuela Colombiana de Ingeniería Julio Garavito. Avenida 13 No 205-59 Bogotá. Colombia
TEL: +57 – 1 668 36 00

Aceptación del jurado

El Trabajo de grado de maestría titulada “Lineamientos para la gestión y analítica de datos en sistemas IoT de servicios hídricos urbanos”, presentada por Mariana Marín Ruiz, cumple con los requisitos establecidos para optar al título de Magíster en Gestión de Información.



Victoria Eugenia Ospina Becerra
Director trabajo de grado



Dante Conti
Codirector trabajo de grado



Daniela De La Rosa Pérez
Jurado



Oswaldo Castillo Navetty
Jurado

Bogotá, D.C., 20 de enero de 2021.

Reconocimiento o Agradecimientos

Quiero agradecer a Dios por brindarme la oportunidad de alcanzar un logro más y haberme puesto en el lugar y momento indicado para hacerlo. Gracias a Dante Conti, codirector de este trabajo de grado quien trabajo de la mano conmigo para lograr concluir esta investigación de manera satisfactoria, a Victoria Eugenia Ospina Becerra, directora de este trabajo de grado por su apoyo de manera personal e institucional, a mi amada Escuela Colombiana de Ingeniería Julio Garavito por abrirme de nuevo sus puertas para alcanzar mis metas como siempre lo he soñado, a mi esposo James Alfonso Muñoz Espitia por su apoyo profesional y moral, que me alentaron a concluir este trabajo, a mi hijo Martín Muñoz Marín por ser mi motivación más grande, a mi hermano Camilo Marín Ruiz, que siempre ha sido mi ejemplo profesional y ese modelo de esfuerzo y dedicación para lograr todo aquello que me proponga, a mi padre Pedro Nel Marín Ortiz (Q.E.P.D) quien me enseñó su gran ejemplo de vida y me demostró que todo es posible mientras hayan ganas de hacer las cosas y que estaría orgulloso por verme alcanzar un logro más, a mi madre Deisy Ruiz Zuluaga por su amor incondicional y por enseñarme el camino a seguir para lograr mi propósito y por último a mis amigos y familia que siempre estuvieron a mi alrededor para motivarme y demostrarme que todo con amor se puede alcanzar.

Resumen

En la búsqueda de optimizar, monitorear y gestionar el proceso de abastecimiento de agua en el ámbito urbano, los datos derivados del Internet de las Cosas (IoT) se posicionan como una herramienta clave. El procesamiento y análisis de dichos datos permite conocer los patrones de consumo de agua de los usuarios, la curva de demanda en diferentes granularidades temporales y la tipología de los usuarios en base a su consumo. Esto sustenta a su vez, procesos decisionales y operativos relacionados con la detección de conexiones fraudulentas, fugas y políticas de tarificación dinámica por consumo y tipología de usuario. Esta investigación presenta los lineamientos necesarios de gestión y analítica de datos derivados de dispositivos IoT en sistemas hídricos urbanos. La propuesta se lleva a la práctica en 3 bases de datos. Se presenta un apartado especial para la ciudad de Bogotá (sin IoT), analizando su situación actual con las 2 bases de datos restantes (con IoT). Los resultados obtenidos son determinantes para promocionar el uso de IoT en la ciudad y avanzar con ello en una mejora continua de la gestión hídrica urbana de Bogotá.

Abstract

In the search to optimize, monitor and manage the water supply process in urban areas, data derived from the Internet of Things (IoT) is positioned as a key tool. The processing and analysis of said data allows to know the water consumption patterns of the users, the demand curve in different temporal granularities and the typology of the users based on their consumption. This, in turn, supports decision-making and operational processes related to the detection of fraudulent connections, leaks, and dynamic pricing policies by consumption and user typology. This research presents the necessary guidelines for the management and analytics of data derived from IoT devices in urban water systems. The proposal is carried out in 3 databases. A special section is presented for the city of Bogotá (without IoT), analyzing its current situation with the 2 remaining databases (with IoT). The results obtained are decisive to promote the use of IoT in the city and thereby advance in a continuous improvement of urban water management in Bogotá.

Tabla de contenido

Lista de Figuras

Lista de Tablas

1	INTRODUCCIÓN	1
1.1	PROBLEMÁTICA	1
1.2	OBJETIVOS	2
1.3	ALCANCE Y LIMITACIONES.....	2
1.4	METODOLOGÍA.....	4
1.5	DESCRIPCIÓN	4
2	ESTADO DEL ARTE	7
2.1	CUARTA REVOLUCIÓN INDUSTRIAL.....	7
2.2	CIUDADES INTELIGENTES	8
2.3	DISTRIBUCIÓN DE AGUA EN BOGOTÁ	10
2.4	INTERNET DE LAS COSAS EN EL MONITOREO DE AGUA EN EL HOGAR.....	11
2.5	ESTUDIO DE LA DEMANDA Y PREDICCIÓN CON EL USO DE LOS DATOS TOMADOS CON LECTORES DE MEDICIÓN AUTOMÁTICA (AMR - AUTOMATED METER READING)	12
3	MODELO DE DATOS DERIVADOS DEL INTERNET DE LAS COSAS	14
3.1	PASO 1: EXTRACCIÓN DE DATOS ORIGEN	16
	<i>Primera capa de dispositivos</i>	16
	<i>Segunda capa de comunicación</i>	17
	<i>Tercera capa IoT</i>	17
3.2	PASO 2: TRANSFORMACIÓN DE LOS DATOS	19
3.3	PASO 3: CARGUE DE LA INFORMACIÓN	25
4	ANÁLISIS EXPLORATORIO Y PERFILAMIENTO DE LOS DATOS	27
4.1	SERIES TEMPORALES POR FECHA	28
	<i>Londres</i> 28	
	<i>Bogotá</i> 32	
4.2	SERIES TEMPORALES POR HORA.....	41
	<i>Londres</i> 41	
	<i>Alicante</i> 44	
5	ARQUITECTURA DE IOT MÍNIMA PARA RECOLECCIÓN DE DATOS DE AGUA EN HOGARES	51
5.1	ARQUITECTURA PARA MONITOREO DE AGUA.....	51
6	IMPACTO SOBRE LA CUARTA REVOLUCIÓN INDUSTRIAL	57
6.1	GOBERNANZA ÁGIL	57
6.2	DISRUPCIÓN EMPRESARIAL.....	58
6.3	SEGURIDAD Y CONFLICTO	58
7	CONCLUSIONES Y RECOMENDACIONES	60

7.1	CONCLUSIONES.....	60
7.2	TRABAJOS FUTUROS.....	60
	BIBLIOGRAFÍA	62
	ABREVIACIONES	64
	ANEXOS.....	65

Lista de Figuras

Figura 1. Metodología para el desarrollo del proyecto	4
Figura 2. Definición de una Smart City	8
Figura 3. Contribuciones de una Smart City	9
Figura 4. Fase de abastecimiento y distribución (EAB).....	10
Figura 5. Proceso ETL.....	15
Figura 6. Arquitectura IoT para toma de datos de consumo de agua en la casa de Canadá.16	
Figura 7. Base de datos AMR configurado por primera vez	18
Figura 8. Base de datos con fechas y horas exactas de las mediciones.	20
Figura 9. Intervalos de horas.....	21
Figura 10. Base de datos agrupada por consumo total cada media hora.....	22
Figura 11. Gráfica del comportamiento de los valores de z para identificar valores atípicos. 23	
Figura 12. Gráfico Boxplot.....	25
Figura 13. Serie temporal edificio industrial Londres.....	29
Figura 14. Consumo de agua febrero 2012.....	30
Figura 15. Consumo total mes de marzo de 2012.....	32
Figura 16. Consumo total en Bogotá por bimestre 2016-2019.....	32
Figura 17. Consumo total de agua por localidad.....	36
Figura 18. Consumo promedio de agua por localidad.....	37
Figura 19. Consumo promedio total localidad Puente Aranda.....	38
Figura 20. Consumo promedio total localidad Chapinero.....	41
Figura 21. Consumo promedio de agua por hora Londres	42
Figura 22. Consumo total por día de la semana Londres.	43
Figura 23. Consumo promedio total de agua por hora Alicante.....	44
Figura 24. Consumo promedio total de agua por hora hogar householi.....	45
Figura 25 Consumo promedio total de agua hogar householj.	46
Figura 26. Consumo promedio total de agua casa householw.....	47
Figura 27. Consumo promedio total de agua hogar househomg.....	48

Figura 28. Consumo promedio total de agua hogar househong.....	49
Figura 29. Arquitectura IoT para Monitoreo de Agua	51
Figura 30. MinoConnectRadio Marca Zenner.....	55
Figura 31. Cuadrante Mágico de Gartner Plataformas IoT.....	55
Figura 32. Fuerzas de impulso para las grandes economías mundiales.	57
Figura 33. Figura 1. Anexo 1: Tipificación de la curva horaria de consumo según el tipo de día 66	
<i>Figura 34. Figura 2 Anexo 1: Dendograma para extraer los clústeres (tipo de día).....</i>	<i>67</i>
<i>Figura 35. Figura 3 Anexo 1: Tipificación de la curva horaria de consumo según la estación del año</i>	<i>68</i>
<i>Figura 36. Figura 4 Anexo 1: Tipificación de la curva horaria de consumo según la estación del año</i>	<i>68</i>
<i>Figura 37. Figura 4 Anexo 1: Tipificación del consumo según el trimestre del año.....</i>	<i>69</i>

Lista de Tablas

Tabla 1. Identificación de valores. desviación estándar.....	23
Tabla 2. Datos para diagramar un Boxplot	24
Tabla 3. Bases de datos de insumo para análisis.....	27
Tabla 4. Consumo promedio febrero 2012.....	30
Tabla 5. Consumo promedio por estaciones del año	31
Tabla 6. Clasificación habitacional por localidad.	33
Tabla 7. Consumo total de agua por localidad.	35
Tabla 8. Consumo promedio de agua por localidad.....	36
Tabla 9. Comparativo entre localidades de mayor y menor consumo por cuenta contrato.	37
<i>Tabla 10. Consumo promedio de agua localidad Puente Aranda.</i>	<i>38</i>
<i>Tabla 11. Consumo promedio total de agua localidad Chapinero.</i>	<i>39</i>
Tabla 12. Intervalos de horas de la jornada laboral.	42
Tabla 13. Inventario de Dispositivos Tecnología AMR.....	53
Tabla 14. Software con Mayor Prestigio para Análisis de Datos en Servicios Públicos.....	55

1 Introducción

1.1 Problemática

Según el último reporte de la Empresa de Acueducto y Alcantarillado de Bogotá [1] referido al consumo de agua potable en la ciudad desde el año 2018 y hasta junio de 2019, se consumen en promedio 11.33 metros cúbicos al mes, y teniendo en cuenta que la población aproximada es de 7.2 millones según el censo realizado por el DANE en 2018 [2], la ciudad de Bogotá se enfrenta a una tendencia alcista relacionada al consumo de agua para los próximos años. Ante tal situación, la gestión del recurso hídrico, al ser este no infinito, requiere de soluciones que permitan un manejo eficiente del recurso y una planificación adecuada ante los retos futuro.

En plena era de las TIC's, las arquitecturas tecnológicas basadas en la Internet de las Cosas (IoT) se fundamentan como herramientas claves para monitorear, gestionar y mejorar los procesos asociados a las redes urbanas de agua potable. En Colombia, se está promocionando la implementación del IoT bajo un esquema de ciudades inteligentes y sustentables. Bogotá, no escapa a esta realidad y necesita una solución para que el agua se use de manera eficiente y responsable. Con IOT se pueden reportar consumos en tiempo real para que a partir de estos datos con una granularidad temporal muy detallada (minutos, horas, días, etc.), se puedan reportar consumos dinámicos, posibles fugas y detección de consumos fraudulentos. Asimismo, esta fuente de datos puede usarse para implementar políticas de consumo responsable asociados a tarifas dinámicas y/o políticas de penalización, para de alguna forma desincentivar el desperdicio del agua. “Las tarifas más altas de agua durante los períodos de escasez proporcionan un incentivo para reducir el consumo y conciencian al usuario”, concluye el investigador Antonio López [3], de acuerdo con el estudio y proyecto que se desarrolló en Valencia - España para monitorear el consumo de agua en los hogares que cuentan con contadores inteligentes, conocidos como AMR.

Bajo estas premisas, este artículo presenta una propuesta con los lineamientos necesarios para la gestión y analítica de datos provenientes de dispositivos AMR aplicados al IoT en sistemas hídricos urbanos en el siguiente esquema: algunos conceptos teóricos y su ámbito en la realidad Colombiana y del Acueducto de Bogotá, la metodología abordada para explotar la data aplicándose a tres bases de datos de acceso público: dos de ellas desde dispositivos AMR y la base de datos de Bogotá tomada únicamente con el consumo bimensual (sin IoT, es decir recopilada con lectura manual) y se cierra con las conclusiones y reflexiones finales surgidas de la experimentación.

Por lo tanto, esta investigación se enfoca en descubrir en términos prácticos, los hallazgos relacionados con la identificación de evolutivos de demanda, perfiles horarios de consumo y

su conexión con la tipología de usuarios y sus patrones de consumo. Esto con el propósito de demostrarle a la Empresa de Acueducto y Alcantarillado de Bogotá hasta donde se puede llegar con datos tomados con tecnología IoT versus las limitaciones de los datos actuales leídos aún de forma manual.

1.2 Objetivos

Objetivo General

Diseñar un modelo de datos para analizar el consumo de agua en hogares, derivados del uso de tecnologías como el Internet de las Cosas (IoT) para la toma acertada de decisiones que favorezcan el uso eficiente del agua en las ciudades.

Objetivos Específicos

1. Potencializar el uso de la analítica avanzada para diferenciar aspectos relacionados al consumo de agua como evolutivos de la demanda, perfiles horarios/usuarios y patrones de consumo.
2. Analizar datos de consumo de agua y presentarlos en una propuesta de visualización para la toma acertada de decisiones en el uso eficiente del recurso hídrico.
3. Proponer la arquitectura mínima del Internet de las Cosas que permitirá la generación de datos del consumo de agua en hogares.
4. Presentar un benchmarking comparativo entre bases de datos de consumo de agua incluyendo la ciudad de Bogotá.

1.3 Alcance y Limitaciones

El presente trabajo de grado se basa en tres entregables, el primero va relacionado con un modelo de datos tomando como insumo información del consumo de agua derivados del uso de Internet de las Cosas cuyo propósito es identificar gráficamente a través de series temporales y matemáticamente a través de clusterización, el comportamiento de las personas al utilizar el recurso hídrico cotidianamente y la tipología familiar en la que se encuentran los hogares. Por otra parte, demostrar que las mediciones pueden ser tomadas de manera

automática, en tiempo real y que será completamente transparente para los clientes conocer su consumo y tomar decisiones de manera responsable. Para este primer entregable, los datos con los que se trabajará son de tres bases de datos derivados del uso de Internet de las Cosas con AMR (lectores de medición automática).

El segundo entregable es una propuesta de visualización para la toma acertada de decisiones respecto al uso de agua en la capital colombiana. La presentación de los datos se hará por medio de una herramienta tecnológica que permita la visualización de los datos entregados por la Empresa de Acueducto y Alcantarillado de Bogotá, de tal forma que sea amigable con el usuario final y presente información que le permita a la Empresa la toma de decisiones basada en la realidad de la ciudad. Los datos con los que se trabajará en este entregable fueron facilitados por la Empresa de Acueducto y Alcantarillado de Bogotá, la cual contiene información histórica del consumo de agua de por lo menos 500 hogares en los 3 últimos años (2017, 2018 y 2019).

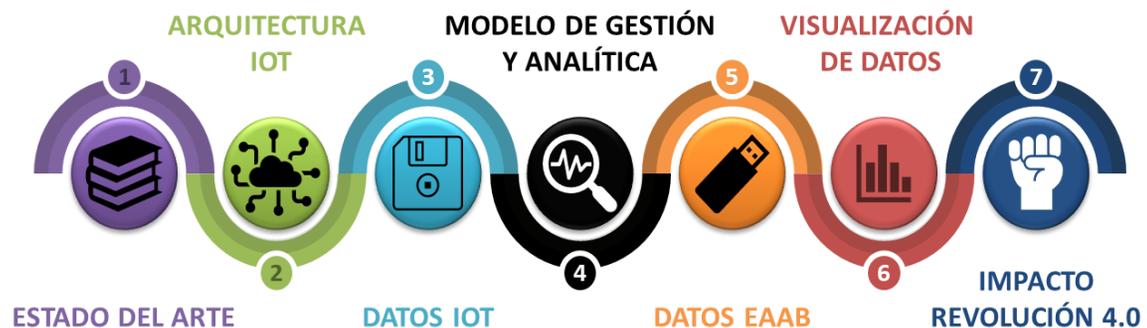
El tercer entregable hace relación a la arquitectura mínima del Internet de las Cosas (IoT), que permita tomar las medidas de consumo de los hogares en Bogotá o en cualquier ciudad que lo quiera implementar a nivel Colombia. La arquitectura propuesta se construirá teniendo en cuenta un proveedor local que cuente con los dispositivos tecnológicos con conexión estándar y soporten conexión inalámbrica sin tener restricciones de marca o tecnología. La arquitectura no será implementada ni será la fuente de los datos que se analizan, es una propuesta basada en la teoría y estudios realizados en otros países del mundo como Italia, España y Londres quienes ya han implementado estrategias basadas en IoT para construir ciudades inteligentes que toman medidas de servicios públicos con esta tecnología. El propósito es que en Colombia se pueda implementar entregando como primer insumo una arquitectura que tome mediciones acertadas, eficientes y eficaces para una toma de decisiones en tiempo real, ahorrando tiempo y mejorando la calidad de los datos recolectados por la empresa de servicios públicos. Esta arquitectura además podrá ser implementada no solo para monitorear agua sino luz y gas.

Como resultados finales se hará la presentación de las ventajas de implementar la tecnología en al menos una ciudad del país dando alcance a la revolución 4.0 que abrió las puertas desde el 2015 en Colombia para que se vean los frutos a nivel mundial de trabajar con este tipo de tecnologías y promover el crecimiento económico de las ciudades y el reconocimiento en el resto del mundo al convertir una parte de su sector en inteligente. Así como la comparación de las bases de datos utilizadas en este estudio, ya que al ser datos tomados con IoT y sin IoT permite visualizar las limitantes que se encuentran entre estos dos paralelismos, lo cual le permitirá a la EAAB entender porque la importancia de implementar esta tecnología en los servicios públicos.

1.4 Metodología

Para desarrollar la investigación y la propuesta de solución en la problemática presentada, se creó un proceso metodológico (ver figura 1) compuesto por siete pasos que fueran el mapa de ruta que de manera progresiva entregará los resultados acordes con los objetivos planteados. A lo largo del desarrollo del trabajo de grado, se podrá relacionar directamente cada uno de estos pasos durante la lectura de cada capítulo.

Figura 1. Metodología para el desarrollo del proyecto



1.5 Descripción

1. Estado del arte: Se hará la investigación basada en cuatro temáticas centrales, las cuales se enfocan directamente con la solución planteada y que pretenden resaltar los estudios desarrollados en cada ámbito. Estas cuatro temáticas son: cuarta revolución industrial, ciudades inteligentes, distribución de agua en Bogotá, Internet de las Cosas en el monitoreo de agua en el hogar, estudio de la demanda y predicción con el uso de los datos tomados con lectores de medición automática (AMR – Automated Meter Reading).
2. Arquitectura IoT: Seleccionar una arquitectura mínima como propuesta para ser implementada en la ciudad de Bogotá y para el monitoreo de agua doméstico, que permitirá obtener los datos que posteriormente soportarían la toma de decisiones de un proceso de tarificación dinámica.

3. Datos IoT: Recolectar las bases de datos que hayan sido tomadas con tecnología IoT para demostrar los beneficios que tienen este tipo de soluciones a la hora de analizarlos y tomar decisiones frente a esto.
4. Modelo de gestión y analítica: Se realizará un estudio de los datos derivados del Internet de las Cosas de tal forma que permita identificar el perfilamiento de los hogares de acuerdo con el consumo de agua, tipificación de usuarios a nivel hora, día y estacionalidad y pronóstico de curvas.
5. Datos EAAB: Solicitar a la Empresa de Acueducto y Alcantarillado de Bogotá, la base de datos de la solución que actualmente se tiene implementada en la ciudad, para identificar el alcance de estos y las soluciones que se pueden llegar a desarrollar con estos datos y así mismo presentar las limitaciones al no contar con niveles de granularidad más bajos (hora, día, semana, mes, año).
6. Propuesta de visualización: Con el uso de los datos entregados por la Empresa de Acueducto y Alcantarillado de Bogotá se realizará una propuesta de visualización para la toma de decisiones relacionados con el consumo de agua de hogares en Bogotá. La herramienta tecnológica con la que se presentará la visualización será una herramienta que soporte el uso para Inteligencia de Negocios (BI).
7. Impacto a la revolución 4.0: Presentar los impactos que se generan por implementar este tipo de soluciones en Bogotá, teniendo en cuenta las 8 fuerzas de las grandes potencias mundiales, según lo planteado por el Foro Económico Mundial.

2 Estado del arte

2.1 Cuarta Revolución Industrial

Con la llegada de la Cuarta Revolución Industrial, los países buscan el mejoramiento de sus sistemas económicos, políticos y sociales a través de la transformación y fortalecimiento de la tecnología digital en la sociedad. Según Klaus Schwab fundador del Foro Económico Mundial la cuarta revolución es “la combinación de sistemas digitales, físicos y biológicos en pro de la transformación de la humanidad”.

Dado lo anterior y comprendiendo la importancia de que los países se sumen a las garantías y ventajas que trae hacer parte de la Cuarta Revolución, Colombia participó el pasado 23 de enero de 2019 en el Foro Económico Mundial para hacer parte de los países que cuentan con un “Centro para la cuarta Revolución Industrial” quinto en el mundo y primero en América Latina ubicado en la ciudad de Medellín. Este centro permitirá que Colombia comience a hacer parte de la ciencia, tecnología e innovación a través del uso de tecnologías como el Internet de las Cosas. Soportado por el documento CONPES 3968 “declaración de importancia estratégica del proyecto de desarrollo, masificación y acceso a internet nacional, a través de la fase II de la iniciativa de incentivos a la demanda de acceso a internet” (Lucía et al., 2019), se impulsa la conexión a Internet de toda la población en el territorio colombiano, eliminando barreras sociales y económicas para el crecimiento económico del país.

La iniciativa de tener un centro especializado en el desarrollo de nuevas tecnologías consiguió que varios grupos de interés se motivaran a crear centros dedicados a la Inteligencia Artificial, al Internet de las Cosas y al blockchain, como: Ruta N, Alianza Caoba, ViveLab Bogotá, la Universidad Nacional y en especial el Centro de Excelencia y Apropiación en Internet de las Cosas el cual centra su esfuerzo en la formación y acompañamiento para proyectos independientes y aquellos que surjan al interior de las compañías existentes. Este centro fundamenta la propuesta de investigación que se desarrolla en esta investigación dado que en sus líneas de trabajo incluyen a Gobierno, cuyo enfoque es el de las SmartCities, motivando a Colombia a hacer parte del grupo selecto de países que se suman a esta iniciativa y permita desarrollo social y económico garantizando sostenimiento y calidad de vida.

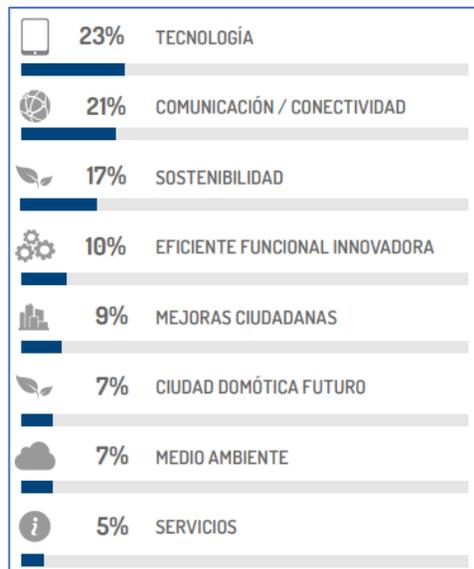
2.2 Ciudades Inteligentes

A pesar de que no existe un consenso entre los diferentes autores para definir que es una ciudad inteligente, la más acertada y que define en su conjunto sus objetivos primarios es “Una smart city utiliza la tecnología para prestar de forma más eficiente los servicios urbanos, mejorar la calidad de vida de los ciudadanos y transformar la relación entre entidades locales, empresas y ciudadanos, facilitando una nueva forma de vivir la ciudad.” (Telefónica, 2015).

Una ciudad inteligente debe tener presente el crecimiento de su población, hacer proyecciones que le permitan definir sus planes de desarrollo para tomar decisiones a tiempo. España se destaca por estar a la vanguardia de las ciudades inteligentes puesto que proyectó para el 2020 el crecimiento de su población urbana en 8.4 millones de personas, desde el 2015, permitiendo que todo su gobierno se coordinará para generar presupuesto enfocado en la transformación de las ciudades.

Cabe destacar que, para transformar una ciudad en inteligente derivado del estudio desarrollado por Telefónica, se deben tener en cuenta diferentes ámbitos o campos a mejorar en una ciudad como:

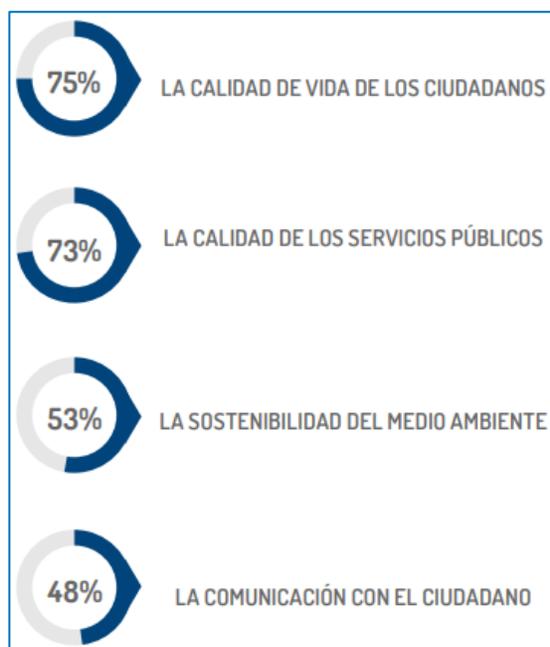
Figura 2. Definición de una Smart City



Dadas las estadísticas que se hicieron a nivel España ciudad denominada como Smart City, se puede visualizar que su objetivo es aportar a la calidad de vida de las personas. Es el caso de tecnologías como el Internet de las Cosas, en el cual se centra este estudio, que le garantiza al ciudadano un aporte en cada uno de los componentes desde tecnología hasta servicios puesto que, al garantizar transparencia y efectividad en la toma de datos del consumo de agua doméstico, el ciudadano tendrá la certeza de que aquello que consume es lo que realmente paga.

Lo anterior no solo se puede ver a nivel personal sino a nivel gobierno pues al transformarse a una ciudad inteligente se entrega información en tiempo real, asertiva y efectiva, garantizando a las ciudades transformarse y aportar a cada uno de los sectores de mayor interés para el Estado (Celino & Kotoulas, 2013).

Figura 3. Contribuciones de una Smart City



Más allá de los pasos o las estrategias para transformarse en un ciudad inteligente, lo más importante es el compromiso del Gobierno en la inclusión de diferentes proyectos e iniciativas que impulsen la inclusión de programas de cooperación en temas de tecnología, disponer presupuesto y capacitar a sus ciudadanos en las tendencias tecnológicas a nivel mundial y permitir la introducción de los diferentes modelos de conectividad que garanticen una infraestructura adecuada y sofisticada para las tecnologías que se quieran implementar en campos de estudio como en el que este estudio se plantea, monitoreo de agua en los hogares.

2.3 Distribución de Agua en Bogotá

Colombia está interesada en hacer parte de un cambio tecnológico y replicar sus estrategias y soluciones a toda América Latina, impulsando ideas que promuevan la sostenibilidad y aprovechamiento de los recursos que hoy por hoy en temas de consumo de agua doméstico no se han implementado. En la actualidad la Empresa de Acueducto y Alcantarillado se encuentra con una problemática de abastecimiento de agua a los hogares Bogotanos, cuyo estudio le ha llevado a determinar que para el 2025 se presentarán problemas de abastecimiento que incluye problemas de filtraciones y conexiones fraudulentas de las cuales no se trata este estudio pero que si contribuirá con una detección temprana de fugas y que hasta el momento no han diseñado propuesta que solucione la problemática.

El foco de estudio de esta investigación está centrado en la zona urbana de Bogotá para la cual, la Empresa de Acueducto, Alcantarillado y Aseo de Bogotá (EAAB) es quien realiza el abastecimiento y distribución a toda la zona, abarcando el 100% de la población. En la actualidad la EAAB entrega el recurso hídrico a través de dos fases, abastecimiento y distribución.

Figura 4. Fase de abastecimiento y distribución (EAB)



La EAAB ha unido sus esfuerzos desde el 2006 para renovar el Centro de Control Operativo incorporando tecnología de última generación con el objetivo de controlar los cambios de presión en los embalses, plantas de tratamiento, tanques, estructuras de control y estaciones de bombeo. Lograron automatizar la operación y obtener información en tiempo real a nivel

macro de toda la infraestructura, esto contribuye a que la capital colombiana logre abastecerse al 100%, sin embargo, hace falta robustecer la infraestructura y pensar a futuro en el uso ineficiente que se le está dando al recurso hídrico (Peña-Guzmán, Melgarejo, & Prats, 2016). A nivel hogar no se está haciendo un monitoreo al mismo nivel que el del Centro de Control operativo es decir la información se obtiene bimestralmente y a pesar de que se haga control desde la fuente del recurso no se está controlando como las personas hacen uso del mismo lo que permitiría generar programas de incentivos y multas por uso inadecuado, un hogar que conozca y tenga claridad del uso, tendría conciencia y contribuiría en reducir costos de factura y por lo tanto ayudaría de manera inmediata con la preservación del recurso.

La administración pública no solo debe centrar sus esfuerzos en el abastecimiento y distribución del recurso hídrico, es fundamental que se empiecen a generar iniciativas y presupuesto que mejore el monitoreo del agua a nivel micro es decir a nivel hogar y de esta manera detectar infiltraciones en el sistema, daños, fugas y demás afectaciones que a hoy no se controlan de manera instantánea (Peña-Guzmán et al., 2016).

2.4 Internet de las Cosas en el Monitoreo de Agua en el Hogar

Una propuesta que motive el uso de dispositivos inteligentes como lo es el Internet de las Cosas o Internet of Things (IoT) por sus siglas en inglés, garantiza un monitoreo de consumo de agua por hogar de manera innovadora y a la vanguardia de las ciudades inteligentes, adicionalmente permitirá una posible clusterización de clientes para generar una tarificación dinámica que son los desarrollos que han causado grandes implementaciones en cuanto a medidores inteligentes en Europa (Roth, 2001). Adicionalmente generará impacto ambiental dado que una vez los usuarios perciban un valor menor en su factura de servicio, esto les hará tomar conciencia y adquirir un patrón de consumo que se mantenga en el tiempo.

Por otra parte, cabe resaltar la importancia que hoy por hoy está tomando el Internet de las Cosas y como Colombia al abrir la posibilidad de hacer usos de esta tecnología se está permitiendo generar ideas innovadoras entregando a los ciudadanos la posibilidad de controlar sus dispositivos y conocer de manera transparente cómo funcionan las cosas y cómo esto permite tomar decisiones de consumo a manera personal, contribuyendo con su calidad de vida (Alaa, Zaidan, Zaidan, Talal, & Kiah, 2017).

Implementar sensores, contadores y georreferencias GPS conectados por medio de una adecuada arquitectura de IoT, garantiza que se obtenga la mayor cantidad de datos en tiempo real, sin embargo, el desafío está en el análisis que se haga de los mismo y para esto existen varios estudios realizados y usados a nivel mundial para lograr un sistema de abastecimiento inteligente en una ciudad. Dicho esto, los métodos que se destacan por ser los más utilizados una vez se implementa la arquitectura IoT son el clustering y sistema multi-agente para la

detección georreferenciada de fugas y el spectral clustering para la eficiencia en la red de abastecimiento (Polit, 2012).

2.5 Estudio de la demanda y predicción con el uso de los datos tomados con lectores de medición automática (AMR - Automated Meter Reading)

Uno de los retos más grandes para instalar redes de IoT, es seleccionar tecnología que permita analizar datos de manera detallada y con precisión, es por esto que, a la hora de seleccionar sensores para tomar los datos, estudios recomiendan hacer uso de lectores de medición automática por sus siglas en inglés AMR (Automated Meter Reading), ya que este tipo de sensores permite hacer estudio con los pronósticos y los datos reales para detectar fallas de los medidores inteligentes, posibles fraudes o ataques ciberfísicos (Antonio Candelieri, 2017).

La innovación al hacer uso de AMR para monitorea en este caso agua de uso doméstico es que la toma se hace directamente por cada uno de los hogares. En la actualidad se usa tecnología SCADA Supervisory Control And Data Acquisition (Supervisión, Control y Adquisición de Datos), este sistema monitorea y controla los datos desde un centro de control a distancia de la fuente, lo que no permite detallar el consumo por hogar minuto a minuto y tener los datos en tiempo real. Al tener datos con ese detalle es posible hacer estudios de demanda y determinar el comportamiento de los clientes en cuanto al consumo, de esta forma hacer clusterización y tarificación dinámica.

Las empresas de servicios públicos que opten por hacer uso de redes de Internet de las Cosas tendrán como insumo el éxito rotundo de los resultados de estudios anteriores, puesto que se sabe que hacer uso de estas tecnologías le garantiza al operador programar de manera eficaz el abastecimiento de los servicios en la ciudad, reducir costos y tendrá a la mano toda la información que se requiera para diseñar estrategias enfocadas en el consumidor para el ahorro y uso adecuado de los recursos (A. Candelieri & Archetti, 2014) .

Por otra parte, y teniendo en cuenta métodos de analítica avanzada cabe la pena resaltar a la inteligencia de negocios como relevante en este tipo de estudios pues a través de esta metodología combinada con Big Data es posible que las empresas tomen sus decisiones basada en la realidad y definan la ruta para aumentar sus ingresos y garantizar que la operación está o no generando alta rentabilidad. Para el caso de una empresa que presta el servicio de distribución de agua, su enfoque está basado en garantizar que el recurso hídrico se entregue milimétricamente a sus usuarios, que casi lo que consume sea lo que llegue a su hogar, esto se resume en costos e ingresos para la empresa y por lo tanto beneficios a futuros a la hora de hacer predicciones frente a su demanda y que mejor técnica que la inteligencia

de negocios para detectar este tipo de problemas a tiempo. Por eso hoy por hoy la EAAB debe concentrarse en manejar adecuadamente su información e implementar nuevas estrategias y soluciones que le permita optimizar sus recursos, para este caso, sus datos, saber manejarlos y convertirlos en información que soporte la toma de decisiones, que se apoye en ese 96% de empresas que colombianas que ya implementan métodos como la inteligencia de negocios para llevar sus operaciones al éxito basado en decisiones a través de los datos. (Gutiérrez Camelo, Devia Llanos, & Tarazona Bermúdez, 2016)

3 Modelo de datos derivados del Internet de las Cosas

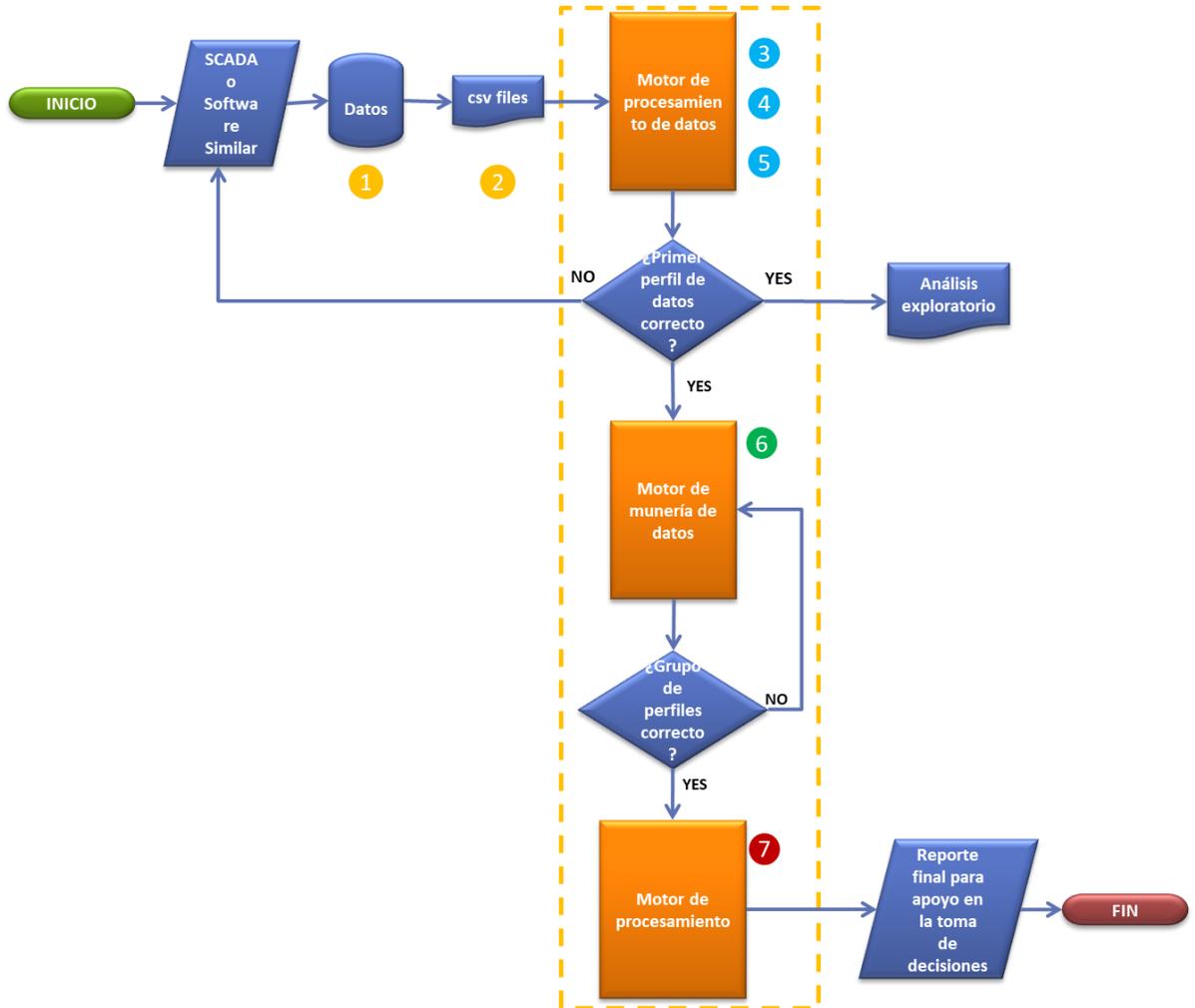
Este modelo de datos da a conocer una serie de pasos para extraer, transformar y cargar los datos de consumo de agua en hogares tomados de un AMR y perfilar consumo, tipificación de usuarios a nivel hora, día, estacionalidad y pronóstico de curvas. Este apartado pretende demostrar como a través del uso de Internet de las Cosas, se puede obtener información detallada y en tiempo real del consumo de agua y planear de manera óptima el abastecimiento de una ciudad de acuerdo con los patrones de consumo de sus usuarios. Por otra parte, este modelo de datos busca que la Empresa de Acueducto y Alcantarillado de Bogotá vea la necesidad de trabajar con tecnología AMR.

El modelo se hará con el uso de tres bases de datos, la primera corresponde al origen de los datos tomados de una casa ubicada en Vancouver, Columbia Británica (Canadá)¹. Esta base de datos visualiza el cómo son recogidos los datos una vez se configura el AMR por primera vez, en un periodo de tiempo que el analista determine según el estudio a realizar, para luego transformar los datos y trabajar con datos de alta calidad y confiabilidad; la segunda base de datos corresponde a la ciudad de Londres, cuyo AMR se instaló en una industria, con esta información se pretende identificar patrones de consumo, perfil de usuario por hora/día, estacionalidad y jornada laboral; y la tercera base son datos tomados de 822 hogares en Alicante – España, para identificar evolutivos de demanda, perfiles y patrones de consumo por tipología familiar.

A continuación, se presenta el flujograma de los pasos que todo analista debe seguir para llegar a una base de datos confiable para ser analizada y generar propuestas o ideas de negocio. Esta metodología busca dar al lector las técnicas que se deben implementar para realizar todo el proceso de un ETL (Extracción, Transformación y Cargue) y llegar a una base de datos de calidad para su uso.

¹ Para mayor información consultar la URL: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FIE0S4>, base de datos de libre descarga.

Figura 5. Proceso ETL

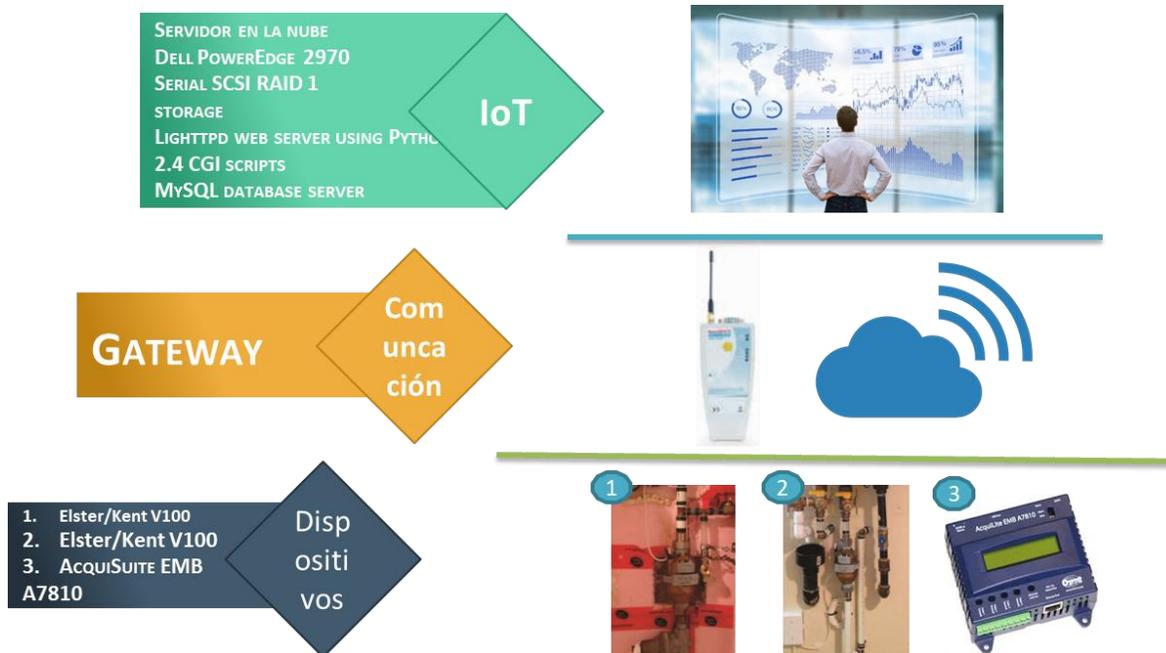


3.1 Paso 1: Extracción de datos origen

Las mediciones relacionadas con el consumo de agua en la casa de Canadá fueron tomadas con una arquitectura de Internet de las Cosas como la que aquí se relaciona. Cabe aclarar que los equipos que se usen en las diferentes arquitecturas van de acuerdo con la capacidad del analista en cuanto a infraestructura tecnológica y costos, sin embargo, debe cumplir con las tres capas relacionadas con el propósito de recopilar los datos.

A continuación, se presenta gráficamente la arquitectura utilizada por el analista para la toma de las mediciones para el estudio:

Figura 6. Arquitectura IoT para toma de datos de consumo de agua en la casa de Canadá.



Primera capa de dispositivos

- 1 y 2 Elster/Kent V100: Estos son los dispositivos que se instalaron para la toma de datos del agua directamente, trabajan a través de pulso. Estos dispositivos fueron usados a consideración del analista dado que permitían tomar la medida con mayor frecuencia (pulso/galón y pulso/0.5L).

- 3 AcquiSuite EMB A7810: Dispositivo AMR para la adquisición de datos inteligente y flexible que permite a los usuarios recopilar datos de energía de medidores de salida de pulso. Este dispositivo tiene la particularidad que no requiere software para acceder a la información, ya que se puede consultar desde cualquier navegador web.

La frecuencia en la toma de las mediciones es por minuto. Para el analista esto es suficiente y apropiado para el estudio dado que se adquiere a un bajo costo para la frecuencia con la que permite trabajar.

Segunda capa de comunicación

La comunicación que se usó para la recolección de los datos está basada en conexión Ethernet (LAN), que permitió el envío de POST HTTP (enviados 1/min) desde ambas unidades de adquisición de datos a un servidor remoto en la nube.

Tercera capa IoT

Los datos se almacenaron fuera del sitio en un servidor de base de datos alojado en una instalación de ubicación conjunta con respaldo de energía adecuado y redundancia de conexión de red. Los datos técnicos del servidor corresponden al serial SCSI RAID 1 almacenamiento servidor web Lighttpd usando scripts Python 2.4 CGI servidor de base de datos MySQL.

Una vez instalada y configurada la arquitectura de IoT para la recolección de datos, se obtiene la primera base de datos. Es importante tener en cuenta que de la configuración inicial depende como se quieran recibir los datos, para este caso de estudio, los datos fueron recolectados minuto a minuto. A continuación, se presenta una muestra de la base de datos que se configuró inicialmente:

Figura 7. Base de datos AMR configurado por primera vez

	A	B	C	D
1	counter	unix_ts	avg_rate	inst_rate
2	221663.0	1356998400	0.0	0.0
3	221663.0	1356998460	0.0	0.0
4	221663.0	1356998520	0.0	0.0
5	221663.0	1356998580	0.0	0.0
6	221663.0	1356998640	0.0	0.0
7	221663.0	1356998700	0.0	0.0
8	221663.0	1356998760	0.0	0.0
9	221663.0	1356998820	0.0	0.0
10	221663.0	1356998880	0.0	0.0
11	221663.0	1356998940	0.0	0.0
12	221663.0	1356999000	0.0	0.0
13	221663.0	1356999060	0.0	0.0
14	221663.0	1356999120	0.0	0.0
15	221664.0	1356999180	1.0	0.0
16	221664.0	1356999240	0.0	0.0
17	221664.0	1356999300	0.0	0.0
18	221664.0	1356999360	0.0	0.0
19	221664.0	1356999420	0.0	0.0
20	221664.0	1356999480	0.0	0.0
21	221664.0	1356999540	0.0	0.0
22	221664.0	1356999600	0.0	0.0

La base de datos resultante del ejercicio presenta la información sin tratamiento, es decir que puede tener valores nulos, espacios en blanco o valores perdidos (que no son consistentes con las mediciones tomadas). En este sentido, es necesario implementar el modelo de base de datos y el formato de datos que se requiere para poder procesar la información (paso 1, ver figura 8), extrayendo los datos en un formato de archivos planos (txt, csv, entre otros), para poder procesarlos fácilmente en herramientas R, RapidMiner, MySQL o Weka. (paso 2, ver figura 8). Para este caso, se toma una muestra de 525.600 correspondientes al año 2013, de la base de datos, con el fin de presentar un ejercicio para el tratamiento de los datos y convertirlos en información para la toma de decisiones.

3.2 Paso 2: Transformación de los datos

Ya con la base de datos se procede a transformarla de acuerdo con el nivel de granularidad (hora, minuto, segundo, etc.) que el analista requiera (paso 3, ver figura 8), a continuación, se deben efectuar procesos que permite limpiar los datos (paso 4, 5 y 6 ver figura 8) encontrando: valores faltantes, valores atípicos.

La mejor forma de trabajar los datos para identificar valores faltantes o atípicos es trabajar con herramientas de análisis estadístico como media, mediana y/o desviación estándar, así como, el uso de gráficos, histograma, diagramas de caja y curvas de densidad, con el fin de identificar valores faltantes o atípicos y corregir estos valores. Se usa la estrategia del vecino más cercano, la cual consiste en rellenar los espacios con valores que se hayan presentado durante la misma hora y en el mismo tipo de día (día entre semana, fin de semana o vacaciones), esto garantiza que el valor de relleno no se desvíe tanto del promedio tomado por el AMR.

Con el fin de presentarle al lector el proceso de transformación de los datos para limpiar la base de datos, se tomó una muestra de la base de datos de Canadá. El archivo se trabajó en Excel.

Paso 1. Conversión de unidades

Dado que la base de datos contiene las fechas en formato Unix², es necesario convertir estos valores en fecha real con la hora en la que fue tomada la medición. Para transformar estas unidades se toma:

$$\frac{\textit{Tiempo en Unix}}{86400 + 25569}$$

Donde:

- Tiempo en unix: Valor dado a la fecha y hora correspondiente de la toma.
- 86400: Número de segundos que contiene un día.
- 25569: Son los días transcurridos entre las 00:00:00 del 1 de enero de 1900 y las 00:00:00 del 1 de enero de 1970.

² Sistema para la descripción de instantes de tiempo: se define como la cantidad de segundos transcurridos desde la medianoche UTC del 1 de enero de 1970, sin contar segundos intercalares.

Una vez convertida toda la base en las unidades de la fecha y hora exacta en la que fueron tomadas las mediciones, se deben separar tanto la fecha como la hora para manejo adecuado de los datos. La base de datos que se obtiene deberá ser algo como:

Figura 8. Base de datos con fechas y horas exactas de las mediciones.

	A	B	C	D	E	F	G
1	counter	unix_ts	avg_rate	inst_rate	Conversión Unix	Fecha	Hora
2	221663.0	1356998400	0.0	0,00	1/01/2013 12:00 AM	1/01/2013	12:00:00 a. m.
3	221663.0	1356998460	0.0	0,00	1/01/2013 12:01 AM	1/01/2013	12:01:00 a. m.
4	221663.0	1356998520	0.0	0,00	1/01/2013 12:02 AM	1/01/2013	12:02:00 a. m.
5	221663.0	1356998580	0.0	0,00	1/01/2013 12:03 AM	1/01/2013	12:03:00 a. m.
6	221663.0	1356998640	0.0	0,00	1/01/2013 12:04 AM	1/01/2013	12:04:00 a. m.
7	221663.0	1356998700	0.0	0,00	1/01/2013 12:05 AM	1/01/2013	12:05:00 a. m.
8	221663.0	1356998760	0.0	0,00	1/01/2013 12:06 AM	1/01/2013	12:06:00 a. m.
9	221663.0	1356998820	0.0	0,00	1/01/2013 12:07 AM	1/01/2013	12:07:00 a. m.
10	221663.0	1356998880	0.0	0,00	1/01/2013 12:08 AM	1/01/2013	12:08:00 a. m.
11	221663.0	1356998940	0.0	0,00	1/01/2013 12:09 AM	1/01/2013	12:09:00 a. m.
12	221663.0	1356999000	0.0	0,00	1/01/2013 12:10 AM	1/01/2013	12:10:00 a. m.
13	221663.0	1356999060	0.0	0,00	1/01/2013 12:11 AM	1/01/2013	12:11:00 a. m.
14	221663.0	1356999120	0.0	0,00	1/01/2013 12:12 AM	1/01/2013	12:12:00 a. m.
15	221664.0	1356999180	1.0	0,00	1/01/2013 12:13 AM	1/01/2013	12:13:00 a. m.
16	221664.0	1356999240	0.0	0,00	1/01/2013 12:14 AM	1/01/2013	12:14:00 a. m.
17	221664.0	1356999300	0.0	0,00	1/01/2013 12:15 AM	1/01/2013	12:15:00 a. m.
18	221664.0	1356999360	0.0	0,00	1/01/2013 12:16 AM	1/01/2013	12:16:00 a. m.
19	221664.0	1356999420	0.0	0,00	1/01/2013 12:17 AM	1/01/2013	12:17:00 a. m.
20	221664.0	1356999480	0.0	0,00	1/01/2013 12:18 AM	1/01/2013	12:18:00 a. m.
21	221664.0	1356999540	0.0	0,00	1/01/2013 12:19 AM	1/01/2013	12:19:00 a. m.
22	221664.0	1356999600	0.0	0,00	1/01/2013 12:20 AM	1/01/2013	12:20:00 a. m.

Paso 2. Creación de intervalos por hora

Al tener las fechas y horas separadas, es importante agrupar los valores en rangos de media hora, es decir el primer grupo de las 00:00 horas hasta las 00:30 horas, así hasta completar las 23:30 horas. Por otra parte, se deben transformar los valores en número dado que en este caso Excel no los identifica como tal y al hacer las operaciones toma los “.” Como miles y no como “,”, es decir, décimas. Obteniendo un resultado como el que se presenta a continuación:

Figura 9. Intervalos de boras.

	A	B	C	D	E	F	G	H
1	counter	unix_ts	avg_rate	inst_rate	Conversión Unix	Fecha	Hora	Intervalo
2	221663.0	1356998400	0.0	0,00	1/01/2013 12:00 AM	1/01/2013	12:00:00 a. m.	00.00
3	221663.0	1356998460	0.0	0,00	1/01/2013 12:01 AM	1/01/2013	12:01:00 a. m.	00.00
4	221663.0	1356998520	0.0	0,00	1/01/2013 12:02 AM	1/01/2013	12:02:00 a. m.	00.00
5	221663.0	1356998580	0.0	0,00	1/01/2013 12:03 AM	1/01/2013	12:03:00 a. m.	00.00
6	221663.0	1356998640	0.0	0,00	1/01/2013 12:04 AM	1/01/2013	12:04:00 a. m.	00.00
7	221663.0	1356998700	0.0	0,00	1/01/2013 12:05 AM	1/01/2013	12:05:00 a. m.	00.00
8	221663.0	1356998760	0.0	0,00	1/01/2013 12:06 AM	1/01/2013	12:06:00 a. m.	00.00
9	221663.0	1356998820	0.0	0,00	1/01/2013 12:07 AM	1/01/2013	12:07:00 a. m.	00.00
10	221663.0	1356998880	0.0	0,00	1/01/2013 12:08 AM	1/01/2013	12:08:00 a. m.	00.00
11	221663.0	1356998940	0.0	0,00	1/01/2013 12:09 AM	1/01/2013	12:09:00 a. m.	00.00
12	221663.0	1356999000	0.0	0,00	1/01/2013 12:10 AM	1/01/2013	12:10:00 a. m.	00.00
13	221663.0	1356999060	0.0	0,00	1/01/2013 12:11 AM	1/01/2013	12:11:00 a. m.	00.00
14	221663.0	1356999120	0.0	0,00	1/01/2013 12:12 AM	1/01/2013	12:12:00 a. m.	00.00
15	221664.0	1356999180	1.0	0,00	1/01/2013 12:13 AM	1/01/2013	12:13:00 a. m.	00.00
16	221664.0	1356999240	0.0	0,00	1/01/2013 12:14 AM	1/01/2013	12:14:00 a. m.	00.00
17	221664.0	1356999300	0.0	0,00	1/01/2013 12:15 AM	1/01/2013	12:15:00 a. m.	00.00
18	221664.0	1356999360	0.0	0,00	1/01/2013 12:16 AM	1/01/2013	12:16:00 a. m.	00.00
19	221664.0	1356999420	0.0	0,00	1/01/2013 12:17 AM	1/01/2013	12:17:00 a. m.	00.00
20	221664.0	1356999480	0.0	0,00	1/01/2013 12:18 AM	1/01/2013	12:18:00 a. m.	00.00
21	221664.0	1356999540	0.0	0,00	1/01/2013 12:19 AM	1/01/2013	12:19:00 a. m.	00.00
22	221664.0	1356999600	0.0	0,00	1/01/2013 12:20 AM	1/01/2013	12:20:00 a. m.	00.00

	A	B	C	D	E	F	G	H
1	counter	unix_ts	avg_rate	inst_rate	Conversión Unix	Fecha	Hora	Intervalo
2	221663.0	1356998400	0.0	0,00	1/01/2013 12:00 AM			
3	221663.0	1356998460	0.0	0,00	1/01/2013 12:01 AM			
4	221663.0	1356998520	0.0	0,00	1/01/2013 12:02 AM			
5	221663.0	1356998580	0.0	0,00	1/01/2013 12:03 AM			
6	221663.0	1356998640	0.0	0,00	1/01/2013 12:04 AM			
7	221663.0	1356998700	0.0	0,00	1/01/2013 12:05 AM			
8	221663.0	1356998760	0.0	0,00	1/01/2013 12:06 AM			
9	221663.0	1356998820	0.0	0,00	1/01/2013 12:07 AM			
10	221663.0	1356998880	0.0	0,00	1/01/2013 12:08 AM			
11	221663.0	1356998940	0.0	0,00	1/01/2013 12:09 AM			
12	221663.0	1356999000	0.0	0,00	1/01/2013 12:10 AM			
13	221663.0	1356999060	0.0	0,00	1/01/2013 12:11 AM			
14	221663.0	1356999120	0.0	0,00	1/01/2013 12:12 AM			
15	221664.0	1356999180	1.0	0,00	1/01/2013 12:13 AM			
16	221664.0	1356999240	0.0	0,00	1/01/2013 12:14 AM			
17	221664.0	1356999300	0.0	0,00	1/01/2013 12:15 AM			
18	221664.0	1356999360	0.0	0,00	1/01/2013 12:16 AM			
19	221664.0	1356999420	0.0	0,00	1/01/2013 12:17 AM			
20	221664.0	1356999480	0.0	0,00	1/01/2013 12:18 AM			
21	221664.0	1356999540	0.0	0,00	1/01/2013 12:19 AM			
22	221664.0	1356999600	0.0	0,00	1/01/2013 12:20 AM			

Ordenar de A a Z

Ordenar de Z a A

Ordenar por color >

Borrar filtro de "Intervalos"

Filtrar por color >

Filtros de texto >

Buscar

- (Seleccionar todo)
- 00.00
- 00.30
- 01.00
- 01.30
- 02.00
- 02.30
- 03.00
- 03.30

ACEPTAR Cancelar

Paso 3: Consumo total por cada media hora de cada día

A continuación, es necesario agrupar los consumos por cada media hora para trabajar los datos cómodamente y realizar las estadísticas necesarias para continuar con la limpieza de datos. Se deben obtener una base de datos más pequeña y ajustada a la necesidad del estudio. Adicionalmente las mediciones están en kl(kilolitros) y deben convertirse a m³ (metros cúbicos), esto se hizo dado que en Colombia se maneja normalmente esta unidad de medida.

Figura 10. Base de datos agrupada por consumo total cada media hora.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Fecha	00.00	00.30	01.00	01.30	02.00	02.30	03.00	03.30	04.00	04.30	05.00	05.30	06.00	06.30	07.00
2	1/01/2013	8,50	17,00	47,40	63,60	84,90	140,80	143,20	41,40	0,00	19,30	0,00	0,00	18,90	0,00	0,00
3	2/01/2013	4,60	12,50	20,50	57,60	43,90	51,60	116,30	4,10	36,40	51,00	152,70	137,70	0,00	11,60	0,00
4	3/01/2013	96,30	27,20	55,65	110,60	33,50	16,50	15,00	20,50	12,90	118,60	6,00	93,20	143,20	20,40	6,00
5	4/01/2013	45,20	6,00	49,40	0,00	24,50	81,00	81,20	9,20	4,60	121,20	92,40	0,00	0,00	14,40	107,10
6	5/01/2013	0,00	66,90	43,80	8,50	57,90	32,60	13,50	75,60	5,40	163,60	26,50	3,00	0,00	26,40	80,30
7	6/01/2013	8,50	121,00	29,90	11,20	51,40	40,60	0,00	6,00	6,00	63,00	131,60	4,60	15,40	0,00	10,00
8	7/01/2013	0,00	24,00	159,50	83,10	22,40	86,60	131,20	7,60	153,60	26,40	26,20	0,00	12,00	5,00	0,00
9	8/01/2013	24,50	72,60	23,20	76,45	68,20	82,60	11,60	8,90	42,10	125,10	114,80	8,50	11,10	11,20	15,00
10	9/01/2013	22,40	29,40	136,00	16,20	16,00	42,00	182,20	192,10	17,40	0,00	0,00	0,00	0,00	116,50	42,30
11	10/01/2013	6,00	41,40	109,10	11,20	32,90	108,80	30,70	36,50	25,30	4,60	18,00	150,40	122,90	8,40	0,00
12	11/01/2013	6,60	12,00	75,50	64,50	44,20	5,50	114,60	64,50	139,20	35,90	0,00	0,00	0,00	0,00	47,60
13	12/01/2013	0,00	37,10	44,50	0,00	0,00	12,00	112,50	7,70	36,30	94,80	6,00	121,50	0,00	24,00	35,60
14	13/01/2013	23,00	6,00	11,40	30,50	15,15	0,00	90,10	10,80	106,40	53,70	0,00	0,00	0,00	0,00	104,10
15	14/01/2013	26,70	64,80	0,00	90,20	109,50	204,30	3,55	100,40	21,90	105,35	14,50	0,00	12,00	37,75	13,90
16	15/01/2013	17,00	5,40	64,20	108,55	26,00	39,70	109,55	29,50	46,10	160,20	53,80	6,00	0,00	0,00	17,00
17	16/01/2013	17,00	35,20	56,80	4,60	173,10	114,10	212,60	105,20	0,00	100,90	9,75	20,10	17,60	0,00	44,00
18	17/01/2013	60,75	53,90	48,80	18,50	9,00	4,00	73,60	8,50	50,90	140,60	12,00	0,00	94,80	19,80	36,30
19	18/01/2013	0,00	8,50	32,90	50,50	26,10	7,70	80,80	17,40	67,20	48,40	0,00	0,00	114,90	0,00	60,70
20	19/01/2013	17,00	6,00	0,00	0,00	0,00	98,50	21,70	119,00	46,15	142,10	72,80	0,00	107,40	5,40	2,40
21	20/01/2013	13,90	24,00	57,70	0,00	47,10	12,00	17,40	121,30	122,60	104,90	11,60	0,00	5,00	39,60	78,45
22	21/01/2013	39,80	0,00	0,00	0,00	0,00	10,90	22,40	124,85	255,50	0,00	2,00	13,50	0,00	0,00	30,50

Paso 4. Identificación de valores faltantes o valores atípicos

Para la muestra que se tomó del año 2013, no se identificaron valores faltantes. Sin embargo, los analistas usan la estrategia del vecino más cercano para corregir estos valores, la cual consiste en rellenar los espacios con valores que se hayan presentado durante la misma hora y en el mismo tipo de día (día entre semana, fin de semana o vacaciones), esto garantiza que el valor de relleno no se desvíe tanto del promedio tomado por el AMR.

Por otra parte, se identificaron los valores atípicos de la muestra de dos formas:

1. Desviación estándar calculando el valor de z.

Se calcula el valor de z para el consumo diario durante cada día del año 2013, a través del cálculo de la media y la desviación estándar de estos valores, obteniendo con la siguiente ecuación, la puntuación de los valores z de cada día.

$$z = \frac{(\text{Consumo del día} - \text{media})}{\text{Desviación estándar}}$$

Para identificar los valores atípicos, se deben seleccionar aquellos datos cuya puntuación de z este por fuera del rango de -3 y 3. Para este caso, se identificaron en total los siguientes valores atípicos.

Tabla 1. Identificación de valores. desviación estándar.

Consumo día (m ³)	Puntuación z	Fecha	Día de la semana
3,35	6,67	08/07/2013	Lunes (martes festivo 09/07/2013)
2,95	5,49	19/07/2013	Viernes
2,49	4,14	21/07/2013	Domingo
0,00	-3,14	03/08/2013	Sábado
0,00	-3,14	04/08/2013	Domingo

Figura 11. Gráfica del comportamiento de los valores de z para identificar valores atípicos.

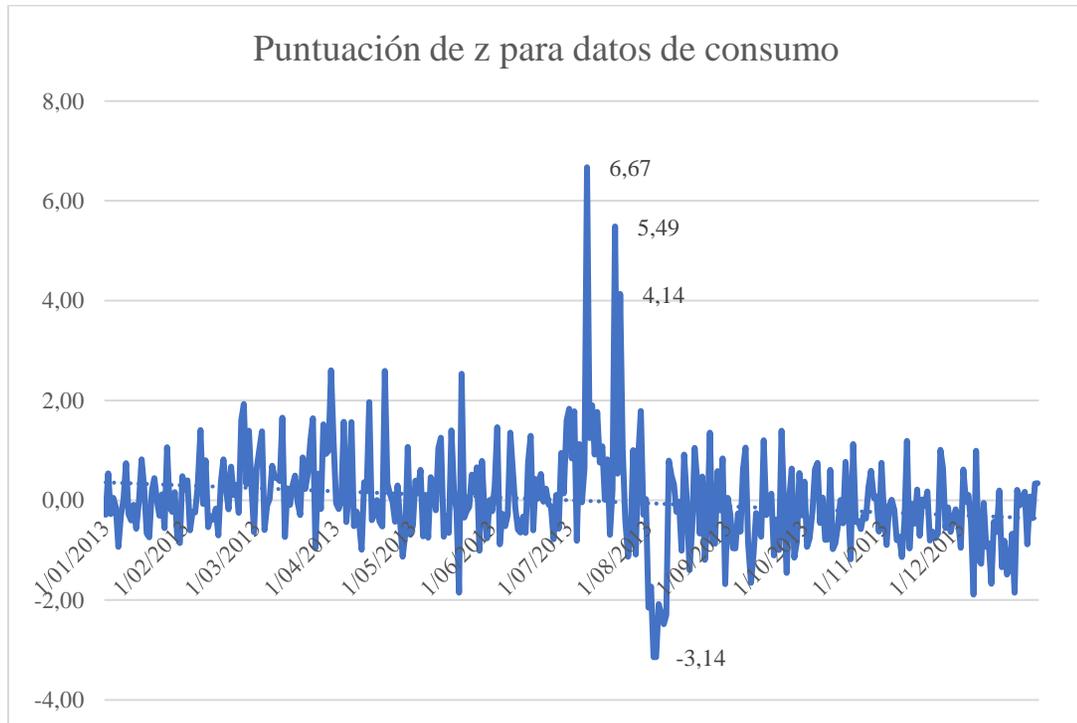
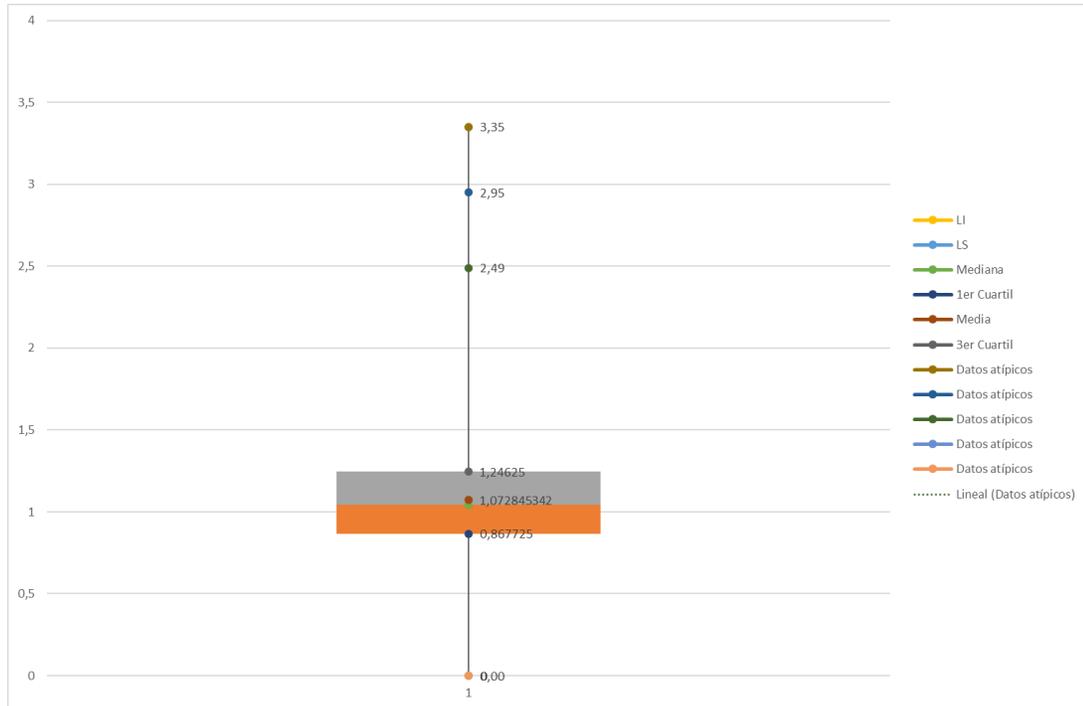


Figura 12. Gráfico Boxplot



Los datos atípicos son valores que se encuentran por fuera de los rangos en los que se tienen las mediciones, para tomar la decisión de qué hacer con estos datos, se debe validar su comportamiento analizando por ejemplo el día de la semana en que se presentó y comparándolo con otros días de la semana similar como fechas especiales. Teniendo en cuenta que aquí se presentan solo 5 valores atípicos, es posible que en estos días haya habido algo especial en el hogar como viajes en el caso en los que las mediciones fueron 0 o alguna reunión dado el consumo en horarios nocturnos.

3.3 Paso 3: Cargue de la información

Finalmente se tiene la base de datos depurada y lista para hacer uso de esta en cualquiera de las herramientas de analítica avanzada que permitirá identificar los diferentes patrones de comportamiento de acuerdo con la granularidad del tiempo (hora, día, semana y mes). Dado lo anterior, la base de datos debe dividirse por grupos de días de la semana, diferenciando los días laborales de los fines de semana y festivos.

Por otra parte, al implementar estrategias de agrupamiento, es posible identificar los grupos de clientes según las épocas del año, semana santa, vacaciones de mitad de año y fiestas decembrinas, pudiendo identificar patrones de consumo dependiendo de la actividad que

tienen los clientes. Todo esto con el fin de orientar los resultados a modelar y pronosticar la demanda de agua y suplir necesidades de abastecimiento que hoy sufren los países por la escasez del recurso hídrico, capítulo que se desarrolla más adelante, con las diferentes técnicas de clusterización de usuarios.

4 Análisis exploratorio y perfilamiento de los datos

En este capítulo se desarrollará un análisis exploratorio de los datos con el fin de encontrar patrones de comportamiento que permita perfilar a los usuarios de acuerdo con el consumo de agua. Para realizar este análisis, es importante haber ejecutado cada uno de los pasos descritos en el capítulo anterior y trabajar con una base de datos confiable.

Es importante tener en cuenta que las bases de datos que a continuación se relacionan y que harán parte de todo el ejercicio de analítica, son bases de datos que previamente fueron transformadas tal y como se ejemplificó en el capítulo 3, para que la información allí contenida contemplara su completitud y veracidad para un análisis eficiente.

Para este apartado se usaron tres bases de datos:

Tabla 3. Bases de datos de insumo para análisis

Lugar	Cantidad de datos	Intervalo de fechas	Tipo
Londres	32.796	29/07/2010 -14/06/2012	Industria
Alicante	138.096 datos 822 casas	1/1/2015 - 20/1/2017	Hogar
Bogotá	361.896 datos 15.079 casas	Bimestre 1 al 6 de 2016 a 2019	Hogar

El análisis exploratorio de los datos se realizó con la herramienta Power BI. Con este análisis se busca determinar para cada una de las bases de datos, aspectos como:

1. Londres:

- Serie temporal por fecha: gracias a los datos obtenidos de esta base de datos es factible graficarlos mes a mes para encontrar patrones de comportamiento en el consumo de agua relacionados con picos de alto consumo, estacionalidad (primavera, verano, otoño. Invierno) y jornada laboral (por ser una industria).
- Serie temporal por hora: los datos tomados del consumo tienen una ventaja y es que se detallaron por cada media hora, gracias a que fueron tomados con AMR. Al tener datos con este nivel de detalle, se procedió a determinar patrones de consumo en la jornada laboral (horarios de entrada, descansos y salida).

- Para este caso específico se realizó una comprobación de las hipótesis que aquí se van a presentar, a través del uso de analítica avanzada para realizar clusterización, este desarrollo se encuentra generalizado en el Anexo 1.

2. Bogotá:

- Serie temporal por fecha: los datos obtenidos de esta base de datos tienen una limitante dado que los consumos se tomaron bimestralmente por año y no tiene el nivel de detalle por día, lo cual dificultó el análisis con la primera gráfica y por lo tanto se tuvo que realizar un estudio más profundo para encontrar algún tipo de patrón de comportamiento a nivel localidades. Con Bogotá se descartó la estacionalidad por festividades (semana santa, vacaciones de mitad y fin de año, verano, navidad, entre otros) porque los datos demuestran un comportamiento constante en cada uno de los años.

3. Alicante:

- Serie temporal por hora: la base de datos de Alicante contiene el consumo por cada hora, y día de la semana. Esto permitió identificar patrones de consumo de jornada horaria (mañana, tarde, noche) y tipología por familia (familia con hijos, adultos mayores, familia típica (hijos, adultos mayores, padres empleados), familia con local comercial (habitan en el mismo lugar de trabajo) y persona soltera.

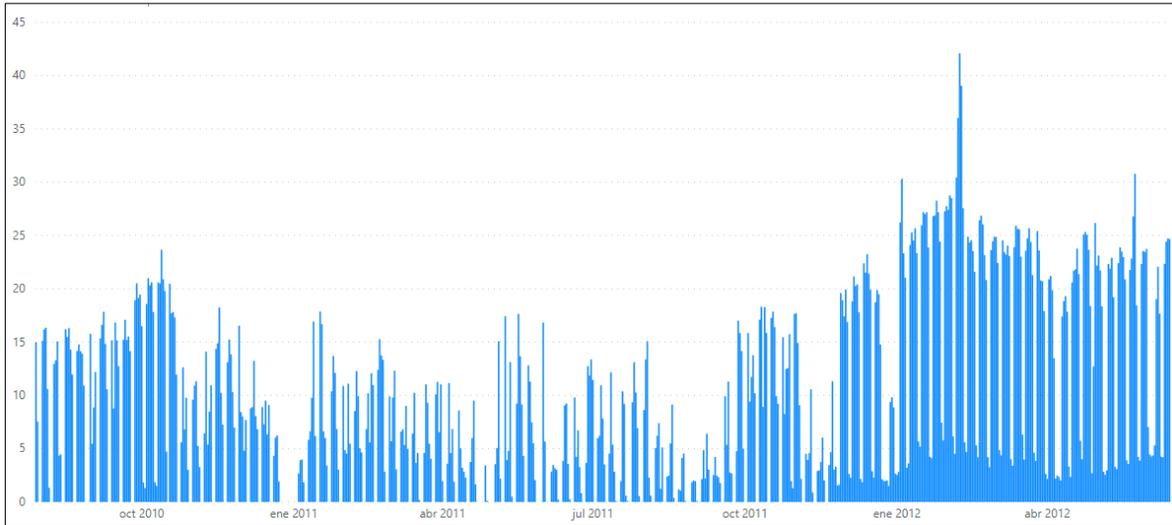
4.1 Series temporales por fecha

En este primer perfilamiento, se tomaron la base de datos de Londres y la de Bogotá, con el fin de que gráficamente y aplicando los filtros necesarios se identificaran patrones de comportamiento en el consumo de agua relacionados con la estacionalidad, actividades o festividades de acuerdo con el desarrollo anual de cada uno de los países.

Londres

La gráfica de serie temporal del consumo por cada día desde el 29/07/2010 -14/06/2012 del edificio industrial en la ciudad de Londres se realizó con la suma de los consumos totales por cada uno de los días del periodo medido para identificar comportamientos anómalos y patrones de consumo de acuerdo con las estacionalidades o actividades que se realizan en la ciudad.

Figura 13. Serie temporal edificio industrial Londres



Una vez graficados los datos se realizaron tres tipos de análisis:

1. Picos de alto consumo:

Al graficar los datos es más fácil visualizar el comportamiento de estos para identificar anomalías. Para el caso de Londres, se encontró que a partir de enero de 2012 los consumos aumentan drásticamente superando los 25 m³ consumo que no se superó en los años 2010 y 2011. Al hallar este tipo de comportamientos se tuvo que ampliar a mayor detalle los puntos anómalos y determinar si se seguía trabajando con ellos, eliminarlos o reemplazarlos con la estrategia del vecino más cercano.

En este caso y de acuerdo con el análisis realizado se tomó una foto de este comportamiento para validarlo, encontrando lo siguiente:

Figura 14. Consumo de agua febrero 2012

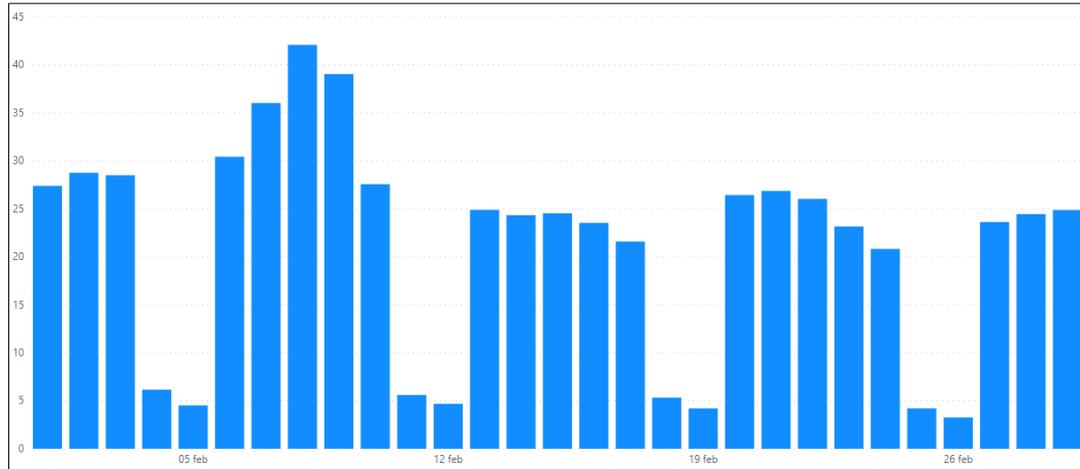


Tabla 4. Consumo promedio febrero 2012

Semana	Fecha	Consumo promedio (m ³)
1	1-3 febrero	28.21
2	6-10 febrero	35.03
3	13-17 febrero	23.78
4	20-24 febrero	24.66
5	27 febrero – 02 marzo	24.04

En la semana 2 se encuentra un aumento en el consumo de la operación de manera gradual iniciando el lunes y bajando su consumo el jueves, el martes, miércoles y jueves con un promedio de 39 m³ demuestran que posiblemente para este mes tuvieron que doblar los turnos para cumplir con el objetivo de la operación. Y se determina así porque en mayo nuevamente se presenta un pico similar lo que quiere decir que la empresa aumenta su consumo de agua en dado caso que tenga que cumplir con algo específico, lo cual se ratifica con el aumento del consumo al principio del año.

2. Estacionalidad:

Al tener la gráfica de consumo por fecha es fácil interpretar si se encuentran patrones de consumo por estacionalidad (primavera, verano, otoño, invierno) en este caso se tomaron de los 3 años, el promedio de consumo por estación para determinar si hay o no un patrón, aunque a simple vista se visualiza un comportamiento similar durante todo

el año por ser industria, lo que quiere decir que para esta empresa en particular, no existe afectación por época del año.

Tabla 5. Consumo promedio por estaciones del año

Estación	Fechas	Año	Consumo Promedio (m³)
Primavera	21 marzo - 21 de junio	2011	4.13
		2012	15.18
Verano	21 junio - 23 septiembre	2011	3.74
Otoño	23 septiembre - 21 diciembre	2010	9.05
		2011	9.35
Invierno	21 diciembre - 21 de marzo	2010-2011	4.97
		2011-2012	18.70

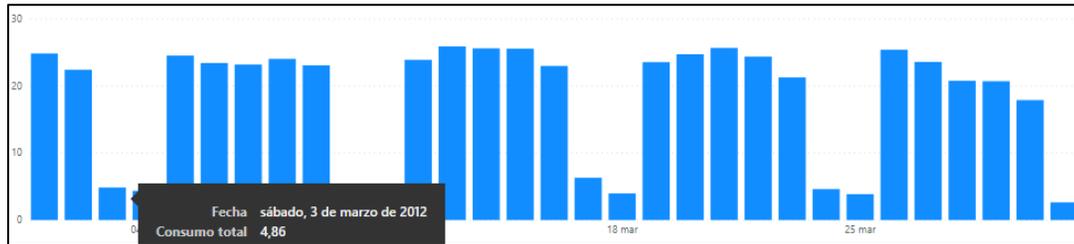
Al validar los consumos promedio por cada una de las estaciones del año, no se evidencia que se produzca un aumento o una disminución por estacionalidad, los aumentos como ya se mencionó se produjeron a finales de 2011 y principios 2012 por un posible doble turno de la empresa no por ser temporada de fin de año, ya que finalizando 2010 e iniciando 2011 se hace una disminución por vacaciones de fin de año. Al no tener más información de la empresa es posible deducir que la producción de la empresa de acuerdo con su sector tuvo un aumento por demanda.

En verano donde se espera alto consumo por aumento en la temperatura ambiente no se evidencian novedades e igualmente en primavera, para otoño se mantiene el consumo en 2010 y 2011, si se tuviera todo el consumo de esta estación para el 2012 es posible que el consumo haya aumentado ya que para este año es donde se evidencia aumentos altamente elevados por su producción.

3. Jornada laboral

Dado que la base de datos hace referencia a una industria, con las gráficas se puede identificar el patrón relacionado con la jornada laboral, por lo tanto, se hace una ampliación en un mes aleatorio, en este caso marzo de 2012 para determinar los días de la semana que más consumen agua y en cuales se mantiene o se reduce su consumo y de esta forma identificar la jornada laboral que maneja la empresa.

Figura 15. Consumo total mes de marzo de 2012

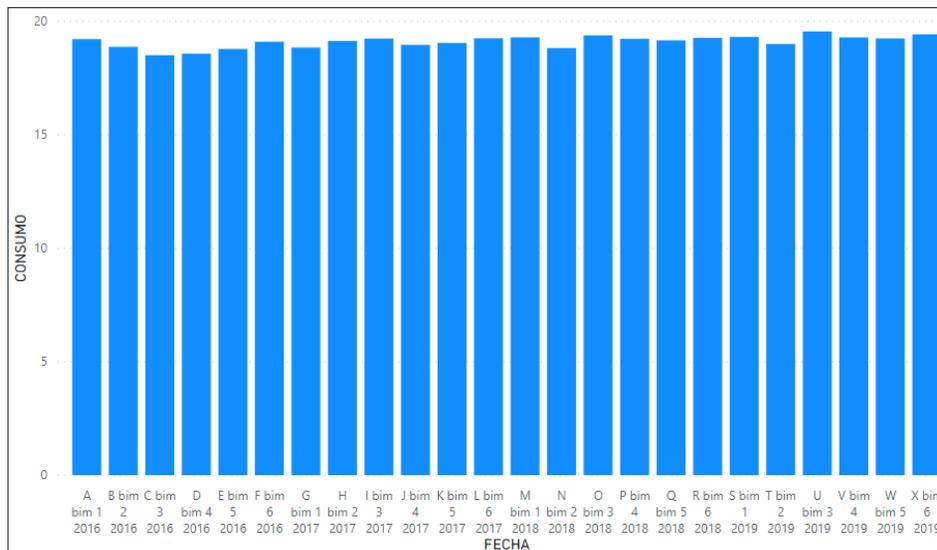


Dado lo anterior se evidencia que los sábados y domingos el consumo se reduce en un 20% del consumo entre semana. Detectando entonces que la semana laboral que maneja esta empresa es de lunes a viernes con descansos de sábados y domingos y el consumo mínimo que se ve los fines de semana da lugar al cuidado que deben dar los vigilantes a las instalaciones de la empresa.

Bogotá

La gráfica para la ciudad de Bogotá se realizó con el consumo total por cada uno de los bimestres de 2016 a 2019 de 11 localidades y 15.079 hogares, distribuidos entre las localidades. Al realizar la gráfica de la serie temporal del consumo por bimestre de los cuatro años, no se evidencian patrones de consumo por valores atípicos o determinantes, estacionalidad, actividad por fecha (vacaciones, escolaridad, festividades entre otros).

Figura 16. Consumo total en Bogotá por bimestre 2016-2019



Para una ciudad como Bogotá que cuenta con 20 localidades en total, la Secretaría de Cultura, Recreación y Deporte generó un resumen de las características de cada una de ellas, clasificándolas según sus actividades en cultural, turística, industrial, entre otros. Al tener este tipo de característica, una curva de consumo debería detectar esos bimestres en donde se presenta mayor consumo o menor de acuerdo con la actividad. Dado que no se evidencia a simple vista en la gráfica general, se estudiaron cada una de las 11 localidades para detectar anomalías o compartimientos que puedan identificar patrones de consumo.

1. Patrones de consumo por localidad

La Secretaria de Cultura, Recreación y Deporte detalló por cada una de las localidades de la ciudad de Bogotá sus diferentes actividades por zona, lo que para este estudio permite determinar por qué entre una u otra localidad existen diferencias en sus consumos de agua.

Tabla 6. Clasificación habitacional por localidad.

LOCALIDAD	CARACTERÍSTICA	CLASIFICACIÓN
Puente Aranda	Se caracteriza por su actividad industrial y por sus amplias zonas residenciales.	Centro Industrial Amplia zona residencial
Ciudad Bolívar	En este sentido, la oferta cultural tiene una variedad de espacios de encuentros para la formación y el estímulo, con la opción de la expresión de inquietudes artísticas, culturales, patrimoniales y sociales. El abanico de posibilidades va desde festivales y encuentros artísticos, escuelas de formación, hasta proyectos de investigación y de patrimonio.	Cultural, Artístico, Social
Barrios Unidos	Cuenta con una gran oferta de bienes y servicios en zonas y barrios, tales como las zonas de servicios de autopartes y reparación de	Clúster económico y cultural.

	automotores en el Siete de Agosto; los talleres metalmecánicos en el barrio Rio Negro, los de manufacturas de cuero en el Muequetá, las marqueterías en el barrio San Fernando o los talleres y almacenes de muebles de madera en el Doce de Octubre.	
Candelaria	Legado arquitectónico de sus calles, ha sido y continúa siendo epicentro de hechos y testigo de la historia política colombiana.	Centro histórico
Bosa	Bosa es considerado como el segundo poblado Chibcha después de Bacatá, territorio gobernado por el Cacique Techotiva.	Cultural, Artístico, Social
Engativá	Cuenta con tres humedales: (La Florida, Jaboque y Santa María del Lago)	Residencial
Los Mártires	Territorio principalmente urbano distribuido	Residencial
Kennedy	Diversidad de expresiones artísticas y poblacionales evidencia una riqueza cultural local artística y poblacional que se destaca en el Distrito Capital.	Cultural, Artístico, Social
Antonio Nariño	Debe su nombre al precursor de la Independencia y traductor de los Derechos del Hombre quien vivió gran parte de su vida en una	Residencial

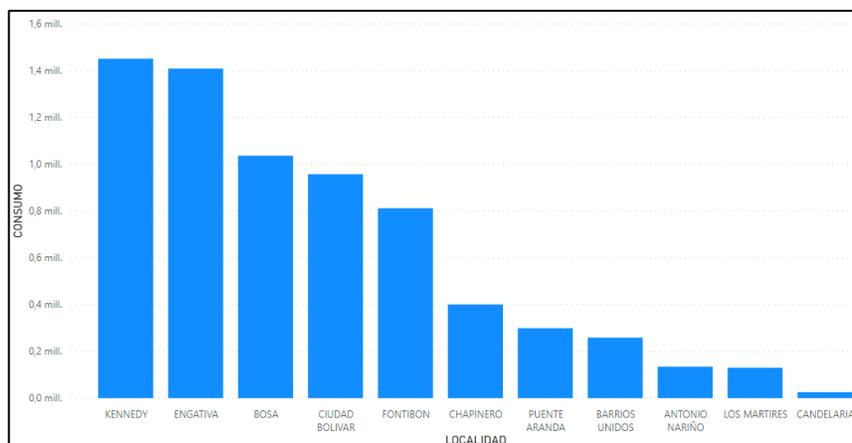
	hacienda localizada en esta zona	
Fontibón	Es el principal eje articulador del desarrollo industrial dada la presencia de una importante zona industrial, la zona franca y su ubicación estratégica regional al ser la conexión del distrito con los municipios de Mosquera, Funza, Madrid y Facatativá.	Industrial Residencial
Chapinero	La oferta cultural de Chapinero es amplia, declarada –zona de interés cultural- cuenta con un buen número de reconocidos artistas, academias de artes y aficionados que viven o trabajan en la localidad.	Residencial Universitario

Se tomaron las 11 localidades y se verificó en detalle por cada una su comportamiento en el consumo de agua por cada uno de los bimestres, encontrando lo siguiente.

Tabla 7. Consumo total de agua por localidad.

Localidad	Consumo Total (m³)
KENNEDY	1.451.336
ENGATIVA	1.409.177
BOSA	1.036.777
CIUDAD BOLIVAR	957.473
FONTIBON	811.957
CHAPINERO	400.598
PUENTE ARANDA	299.052
BARRIOS UNIDOS	258.588
ANTONIO NARIÑO	134.580
LOS MARTIRES	129.774
CANDELARIA	25.428
Total	6.914.740

Figura 17. Consumo total de agua por localidad.

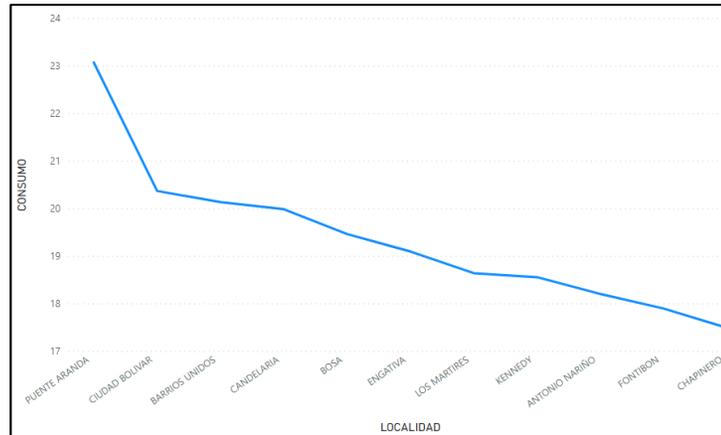


Es importante tener en cuenta que Kennedy junto a Suba y Engativá son las localidades más pobladas y La Candelaria, Los Mártires y Santa Fe las menos pobladas de Bogotá, por lo tanto, según la cantidad de habitantes y según su actividad socioeconómica se determinaría que a mayor o menor cantidad poblacional, más o menos consumo de agua, respectivamente, tal y como se evidencia en el gráfico anterior. Sin embargo, es importante visualizar la relación de consumo que se tiene por localidad y la cuenta contrato (denominada por la EAAB para identificar el contador de agua por cada hogar) y de esta manera evidenciar realmente cuál de estas localidades están consumiendo mayor cantidad de agua.

Tabla 8. Consumo promedio de agua por localidad.

Localidad	Cuenta Contrato	Consumo Total	Consumo Promedio
PUENTE ARANDA	12.960	299.052	23,08
CIUDAD BOLIVAR	46.992	957.473	20,38
BARRIOS UNIDOS	12.840	258.588	20,14
CANDELARIA	1.272	25.428	19,99
BOSA	53.256	1.036.777	19,47
ENGATIVA	73.776	1.409.177	19,10
LOS MARTIRES	6.960	129.774	18,65
KENNEDY	78.192	1.451.336	18,56
ANTONIO NARIÑO	7.392	134.580	18,21
FONTIBON	45.360	811.957	17,90
CHAPINERO	22.896	400.598	17,50
Total	361.896	6.914.740	19,11

Figura 18. Consumo promedio de agua por localidad.



Según lo que se esperaba, se encontró una gran diferencia de consumo pasando Puente Aranda a ser la localidad que consume más y Chapinero la que menos así:

Tabla 9. Comparativo entre localidades de mayor y menor consumo por cuenta contrato.

Localidad	Cuenta Contrato	Consumo Total (m³)	Consumo Promedio (m³)	Localidad	Cuenta Contrato	Consumo Total (m³)	Consumo Promedio (m³)
PUENTE ARANDA	12.960	299.052	23,08	KENNEDY	78.192	1.451.336	18,56
CHAPINERO	22.896	400.598	17,50	CANDELARIA	1.272	25.428	19,99

Dado lo anterior, Puente Aranda a pesar de tener menor cantidad de cuentas contrato en comparación con Kennedy que tiene la mayor cantidad de habitantes en Bogotá, consume en promedio más que Kennedy y más que el promedio de todas las localidades que está en un 19.11 m³ y Chapinero aun teniendo más cuentas contrato que Candelaria que tiene menor habitantes a nivel Bogotá, es la localidad que en promedio consume menos.

Es importante revisar estas dos localidades (Puente Aranda y Chapinero) para validar posibles valores atípicos que estén generando este comportamiento.

Puente Aranda

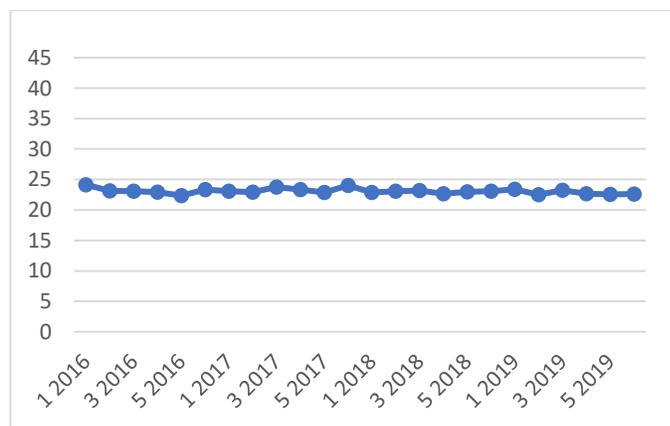
Al validar el consumo de cada una de las cuentas contrato de la localidad, se evidencia que los valores son constantes y al graficarlos se visualiza una línea recta que no permite detectar a simple vista si alguna cuenta contrato este presentando un alto

consumo. Dado lo anterior se debe validar que características tiene la localidad y las posibles causas del alto consumo, pero por el detalle de los datos no es posible determinar un patrón de consumo. Habría que tener un nivel de detalle más descriptivo que permita validar a tiempo anomalías o categorizar los clientes para estudiar la posibilidad de hacer una tarificación dinámica.

Tabla 10. Consumo promedio de agua localidad Puente Aranda.

Bimestre/Año	Consumo promedio
1 2016	24,15
2 2016	23,11
3 2016	23,10
4 2016	22,91
5 2016	22,36
6 2016	23,32
1 2017	23,10
2 2017	22,92
3 2017	23,77
4 2017	23,33
5 2017	22,86
6 2017	24,03
1 2018	22,89
2 2018	23,06
3 2018	23,17
4 2018	22,67
5 2018	23,00
6 2018	23,07
1 2019	23,38
2 2019	22,52
3 2019	23,25
4 2019	22,65
5 2019	22,55
6 2019	22,63
Total	23,08

Figura 19. Consumo promedio total localidad Puente Aranda.



Ahora, como bien se dio a conocer cada localidad está clasificada por su actividad socioeconómica y por lo tanto puede que esta localidad es la que más consume agua por ser un sector industrial y altamente residencial. Si las personas que allí habitan

trabajan cerca del sector, es posible que retornen a sus hogares con mayor frecuencia en horas de almuerzo, descanso o puedan llegar más temprano después de su jornada laboral que en un sector como Kennedy donde la mayoría de las personas laboran a distancias muy retiradas de su hogar y por tanto salen muy temprano y regresen al anochecer.

Estas conclusiones solo se pueden decir dadas las estadísticas en la ciudad, pero si se tuviera un mayor nivel de granularidad en las mediciones podría identificarse realmente qué está produciendo este alto o bajo consumo en ciertos meses del año y evidenciar por hora y día de la semana, horarios en los que permanecen las personas con mayor frecuencia y si hay cambios drásticos determinar algún patrón de comportamiento.

En este caso la EAAB tendría que verificar una por una las cuentas contrato de la localidad para validar cuales están presentando alto consumo de agua e ir a hacer una visita técnica para verificar la causa. En la actualidad la EAAB se basa en los consumos bimensuales que toman, al detectar que una medición con respecto a la otra varía demasiado, piden cita en la casa para identificar el problema.

Chapinero

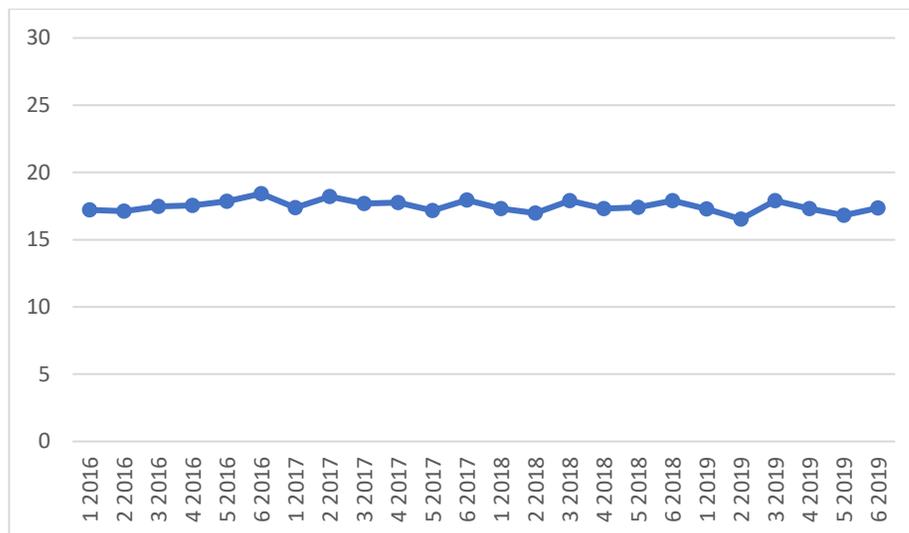
En el caso de Chapinero se validó el consumo promedio por cada bimestre de los cuatro años, evidenciando una constante que gráficamente se representa como una línea recta, que al igual que Puente Aranda, no permite detectar patrones ni anomalías.

Tabla 11. Consumo promedio total de agua localidad Chapinero.

Bimestre/Año	Consumo promedio
1 2016	17,22
2 2016	17,12
3 2016	17,48
4 2016	17,55
5 2016	17,86
6 2016	18,42
1 2017	17,39
2 2017	18,21
3 2017	17,69
4 2017	17,77
5 2017	17,16
6 2017	17,96

1 2018	17,31
2 2018	16,99
3 2018	17,91
4 2018	17,32
5 2018	17,40
6 2018	17,91
1 2019	17,28
2 2019	16,53
3 2019	17,90
4 2019	17,32
5 2019	16,82
6 2019	17,36

Figura 20. Consumo promedio total localidad Chapinero.



Los datos anteriores no representan un factor determinante o un posible patrón de consumo, por lo tanto, se deduce que al ser Chapinero una localidad de alta concentración de estudiantes y zona de bares y discotecas, las personas dueren la mayor parte del tiempo fuera de sus hogares, adicionalmente que por ciertas épocas del año los apartamentos estén desocupados por la temporada de vacaciones y cambio de recinto por finalización de actividades académicas. Sin embargo, estas siguen siendo deducciones debido a las estadísticas de la ciudad lo cual obliga a entrar al detalle al igual que en Puente Aranda y estudiar cada cuenta contrato de la localidad para detectar si en alguna de estas casas se están presentando anomalías.

4.2 Series temporales por hora

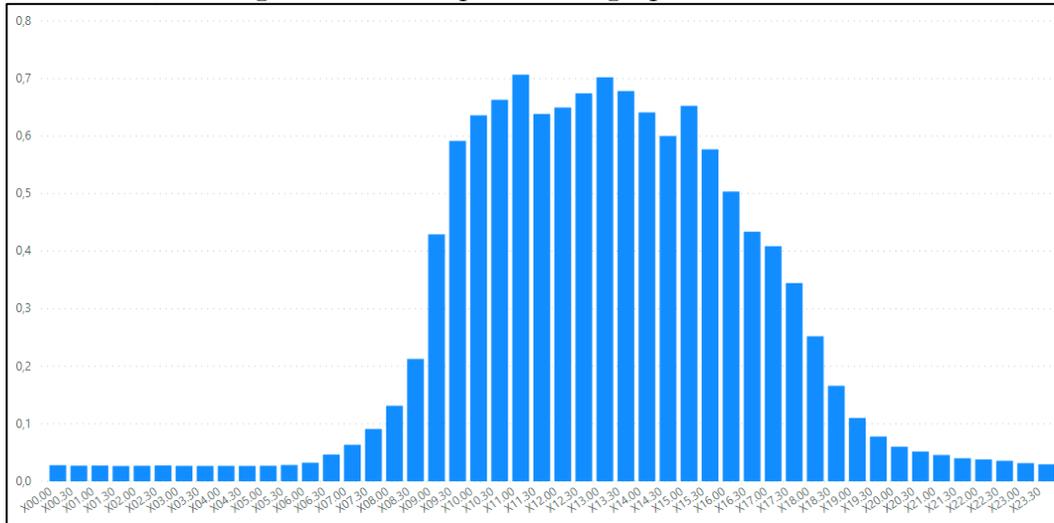
Para el segundo perfilamiento se usaron las bases de datos de Londres y Alicante, pero esta vez para determinar patrones de consumo por horarios y determinar, jornadas laborales e identificar picos de demanda que generen una línea base de tarificación dinámica.

Londres

La gráfica de serie temporal del consumo de agua por hora se realizó con el promedio de consumo por cada media hora, iniciando a las 00:00 horas y finalizando a las 23:00 horas de

lunes a viernes, con el fin de detectar patrones de comportamiento a nivel horario y definir las posibles causas en altos o bajos consumos.

Figura 21. Consumo promedio de agua por hora Londres



Con la gráfica anterior se realizaron los siguientes análisis:

1. Jornada laboral

Al tener la gráfica por hora y conociendo que la base de datos fue tomada a una empresa, se facilita la identificación de una jornada laboral desde las 6:00 am donde se observa un ligero aumento en el consumo y su elevación de manera progresiva hasta las 7:30pm que es donde se normaliza nuevamente la curva.

Por otra parte, es factible encontrar la hora de almuerzo, las horas donde la empresa está en constante operación, pudiendo establecer patrones de horas en los que más o menos consumen agua.

Tabla 12. Intervalos de horas de la jornada laboral.

Hora	Consumo promedio (m ³)	Característica
6:30 am	0,04	Ingreso a la empresa
7:00 am – 11:30 am	3,88	Operación constante
11:30 am – 12:30 pm	1,96	Descanso
1:00 pm – 7:00 pm	6,87	Operación constante
7:30 pm	0,08	Salida de la empresa

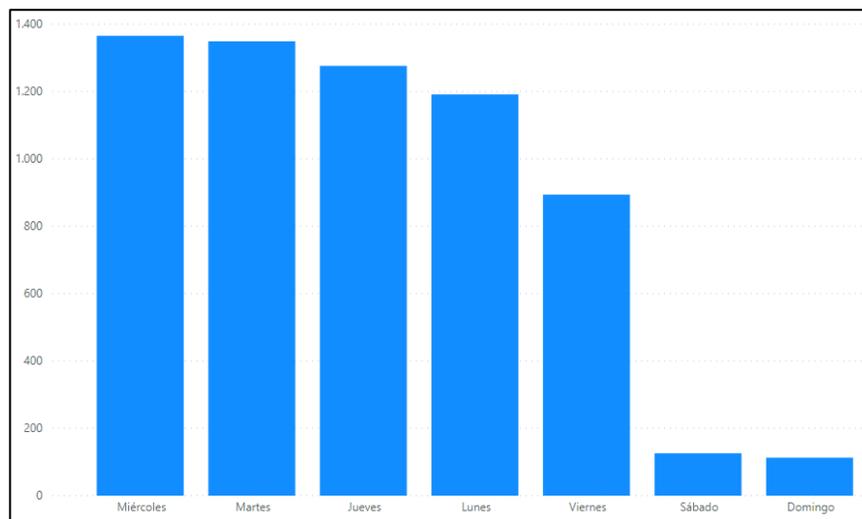
Los picos de mayor consumo en esta fábrica están agrupados de acuerdo con los valores cuya característica relaciona operación constante, lo que quiere decir que la empresa en estos periodos es donde tiene activa su producción.

Al tener una base de datos tan detallada permite visualizar qué pasa cada media hora con el comportamiento de consumo, es fácil detectar anomalías a tiempo, puesto que su patrón de consumo lo determina la jornada laboral y los horarios de operación constante. Si se presentan cambios drásticos en el consumo de agua resulta eficiente tener datos históricos con qué comparar no tan alejados en el tiempo, es decir, poder comparar un día con otro o una semana con otra para detectar las causas posibles del consumo elevado o mínimo de agua.

2. Días de la semana

Ahora bien, como ya se logró identificar el patrón por horas de la jornada laboral, es indispensable identificar en que días de la semana ejecutan sus labores la empresa, por lo tanto, se grafica el consumo total de agua por los días de la semana.

Figura 22. Consumo total por día de la semana Londres.

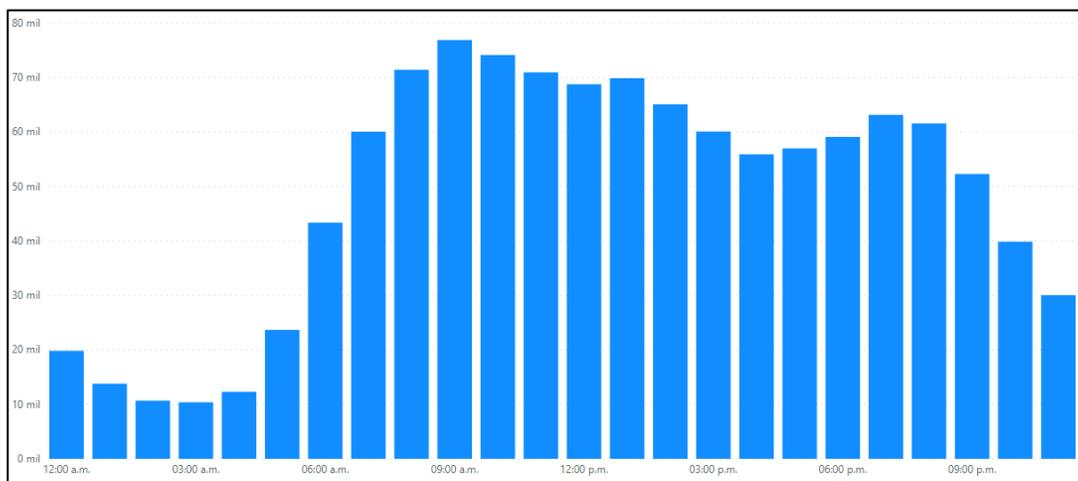


Con esta gráfica se identifica que la jornada laboral transcurre de lunes a viernes y los fines de semana descansan. Por otra parte, los miércoles y jueves es donde más agua se consume posiblemente por su tipo de actividad, el martes y lunes la producción se mantiene estable y a partir del viernes y hasta el domingo se reduce completamente el consumo para dar comienzo al fin de semana.

Alicante

Para graficar la serie temporal de la ciudad de Alicante, se tomó el consumo promedio total de agua por cada una de las horas del día desde las 12:00 am hasta las 11:00 pm con el propósito de visualizar patrones de compartimiento en el consumo. Para esto se realizaron los siguientes análisis.

Figura 23. Consumo promedio total de agua por hora Alicante.



1. Jornada horaria:

Una vez se graficó el consumo total de agua por hora en la ciudad de Alicante, se pudo determinar que manejan una jornada de horarios cotidiana y similar a cualquier residente de una ciudad, donde su jornada laboral en promedio inicia muy temprano en la mañana, evidenciado en el horario de 5 am a 9 am que es el tiempo en el que los miembros de una casa salen para sus respectivos lugares de estudio o trabajo, a la 1pm se presenta nuevamente un aumento donde posiblemente algunos hogares retornan por hora de almuerzo y a las 7pm donde todos retornan nuevamente. Se evidencia también que durante el día en la gran mayoría de los hogares alguien permanece en casa por el consumo constante de agua, sin embargo, esto se verá reflejado en el próximo análisis que se hizo por hogar para validar si todos los hogares manejan este patrón de consumo o uno diferente al de toda la ciudad.

2. Consumo por hogar

Con la gráfica general se puede asegurar que con el patrón de consumo que se presenta, es factible crear incentivos o penalidades en el uso de agua de acuerdo con la hora, sin

embargo, se hizo un análisis profundo tomando aleatoriamente 5 hogares de la ciudad para validar si hacer una tarificación dinámica basado en el consumo de toda la ciudad, es justificable.

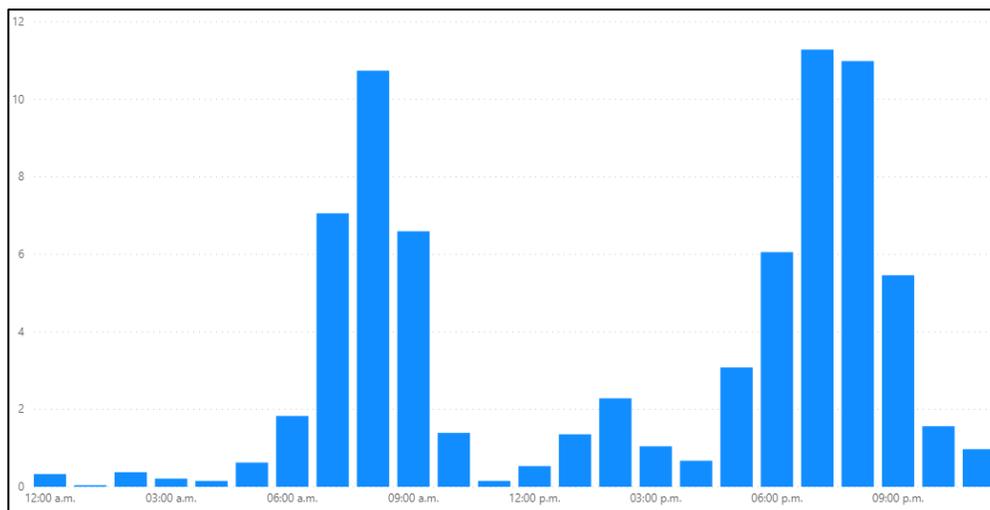
Todo esto será demostrado por medio de la clusterización de los datos para determinar los grupos de hogares que consumen agua de manera similar y categorizarlos para generar una propuesta de tarificación adaptable a toda una ciudad.

2.1. Tipología por hogares

2.1.1. Familia con hijos

En el hogar “householi”, se observa un comportamiento muy diferente al de la curva general en la ciudad.

Figura 24. Consumo promedio total de agua por hora hogar householi.

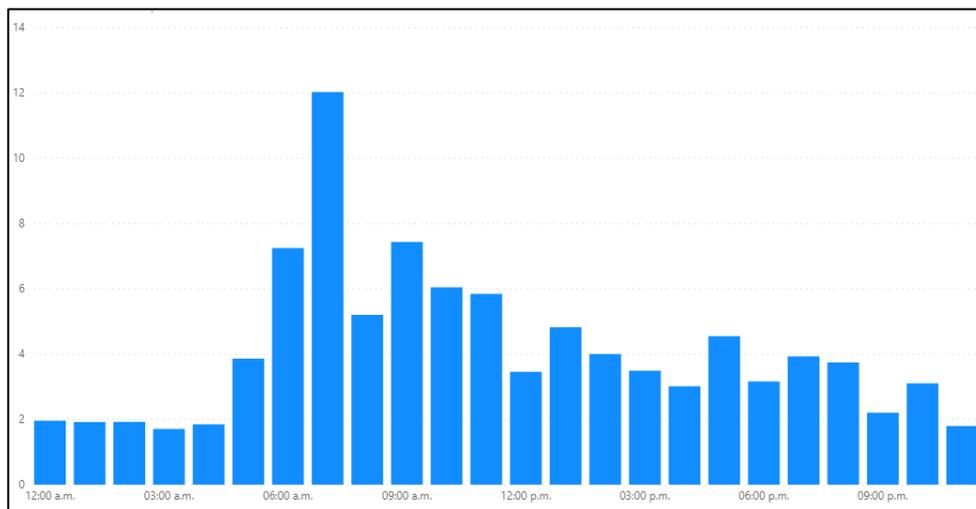


En esta casa se observa un patrón de consumo similar al de la gráfica de serie temporal de toda la ciudad de Alicante, en el horario de 6am a 8am que es donde habitualmente las personas salen de sus hogares, sin embargo, sobre medio día no se presenta un consumo constante como el que representa la gráfica general de la ciudad. Sobre la 1pm y 2pm es posible que algún miembro de la familia retorne a casa por temas académicos o laborales y ya sobre las 7 y 8pm los habitantes retornan a la vivienda. En casas de este tipo de comportamientos se podría categorizar dentro de las familias con hijos, que luego de su jornada escolar retornan a casa y permanecen allí hasta que sus padres regresan en la noche.

2.1.2. Familias con servicios de cuidador para adultos mayores o niños pequeños

El hogar “householj” se evidencia que las personas inician su día a las 5am, salen de casa sobre las 7am, pero alguien permanece en casa durante el día y cerca de las 7pm van retornando gradualmente al hogar. Este tipo de hogares deben estar conformados por familias con niños pequeños o adultos mayores, quienes permanecen en casa con algún cuidador.

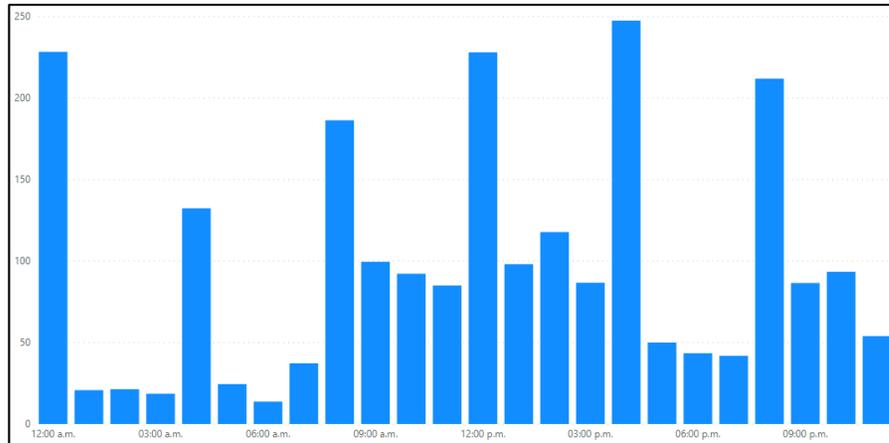
Figura 25 Consumo promedio total de agua hogar householj.



2.1.3. Familia con local residencial en el mismo lugar de la vivienda

El hogar “householw” tiene un comportamiento muy particular respecto al promedio general de la ciudad puesto que los patrones de consumo son difíciles de interpretar por su consumo en horas no habituales, lo cual se evidencia en la siguiente gráfica.

Figura 26. Consumo promedio total de agua casa household.



Dado que para estos datos solo se cuenta con el consumo por horas durante un periodo de tiempo sin especificar la fecha exacta en la que fue tomado el dato, no es posible deducir que esta casa este consumiendo agua en horarios no habituales por algún evento especial, pero si se observa un patrón de consumo de cada 4 horas, como por ejemplo: a las 12 am se presenta un alto consumo que se reduce a partir de la 1am y vuelve a aumentar a las 4am, luego a las 8am, 12pm, 4pm, 8pm.

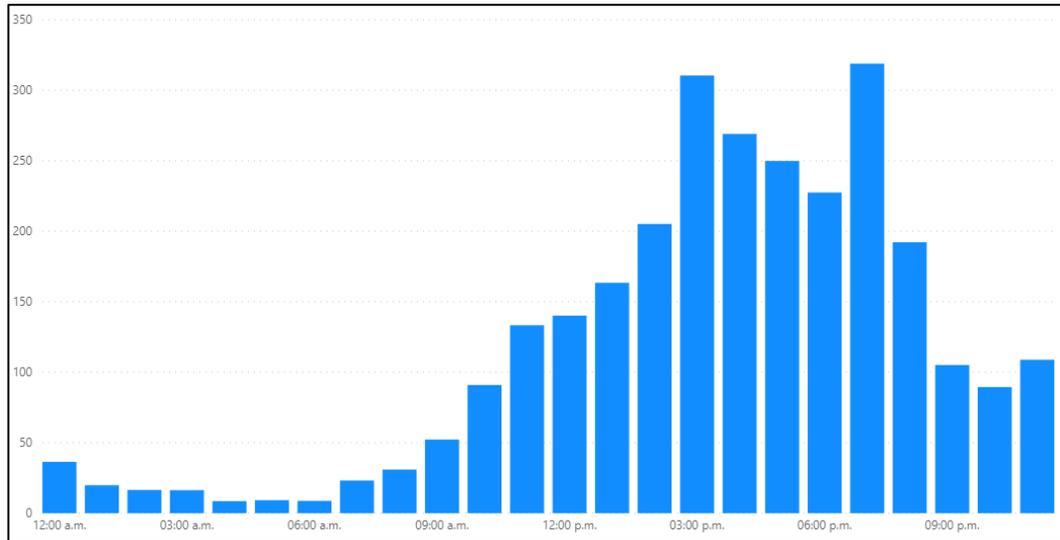
Para entender que puede estar pasando en este horario, fue indispensable investigar un poco más de la ciudad y entender su actividad económica. En este caso, Alicante es una ciudad donde la actividad económica de los servicios representa el 75,29%, la cual incluye al consumo de comidas y bebidas (Ministerio de Empleo y Seguridad Social, 2019).

Al contar con información más precisa de las actividades económicas relevantes en la ciudad, es factible atribuir este patrón de consumo a una casa donde desarrollan actividades comerciales como un bar restaurante y residen allí mismo sin haberlo notificado a la empresa de acueducto de la ciudad. Es una casa que por sus horarios coincide con el horario en que los ciudadanos salen de sus lugares de trabajo y consumen alguna comida y bebida.

2.1.4. Familia con actividad económica independiente o jubilados.

En el hogar “househomg” se visualiza un comportamiento particular y muy diferente a una jornada habitual de una familia que sale en la mañana y regresa en la noche como se representa en la gráfica de la ciudad.

Figura 27. Consumo promedio total de agua hogar househomg.

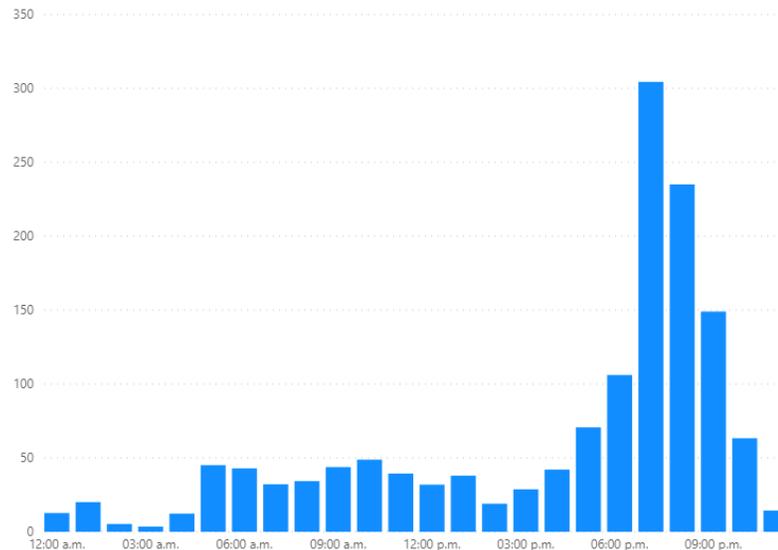


Este hogar en la jornada de la mañana no presenta actividad hasta llegadas las nueve de la mañana, consumo que se mantiene en aumento y constante, se eleva nuevamente sobre las tres de la tarde, horario en el que realizan alguna actividad en particular relacionado posiblemente con servicios generales, cena o almuerzo, al igual que a las siete de la noche en la última cena del día para disponerse a descansar. Este hogar tiene un comportamiento relacionado con personas que permanecen en casa la gran mayoría del tiempo y no indica que puedan tener una jornada laboral marcada, siendo personas jubiladas o con un trabajo independiente desde casa.

2.1.5. Pareja

En el hogar “househong” el patrón de consumo de agua es muy diferente al de la gráfica general de la ciudad pues en esta casa se evidencia que se sale muy temprano de casa, alguien permanece en ella durante el día y ya en la noche se reúnen nuevamente para finalizar la jornada.

Figura 28. Consumo promedio total de agua hogar housebong



Dadas las condiciones de la gráfica, el patrón de consumo es de una familia de una pareja donde uno de los dos sale a trabajar y el otro desarrolla sus actividades desde casa.

Para poder realizar todos los análisis e interpretaciones de los datos que aquí se detallaron, es importante contar con bases de datos confiables y eficientes, que sea información con la cual se pueda trabajar y emitir posibles juicios de valor que le den respuesta a las diferentes problemáticas de distribución de agua en las ciudades.

Para casos como los de Londres y Alicante donde el nivel de detalle de los datos era tal que se podía visualizar por hora y por fecha el comportamiento del consumo de agua se pudo concluir gráficamente lo que estaba pasando y el por qué de los consumos. Al tener datos de esta manera, las empresas de servicios de agua de las ciudades pueden tener los datos en tiempo real y tomar decisiones a tiempo y no esperar a hacer estudios como este para luego si entender lo que está pasando y garantizar que el agua que se distribuye en la ciudad este acorde con el comportamiento de los usuarios.

En el caso de Bogotá al tener un nivel de granularidad tan alto (bimensual) no se logra determinar un comportamiento de la ciudad, es decir, no es posible identificar en que épocas del año se consume más o menos agua. Para llegar a encontrar anomalías en el consumo de agua es necesario entrar a validar más a fondo los datos y estudiar por localidad, barrios y cuentas contrato. Si Bogotá tuviera un sistema que le permitiera recoger los datos de manera automática a través de tecnología como el Internet de las Cosas, podría distribuir de manera más eficiente el agua en la ciudad y solucionar problemas de cortes de agua por no identificar

a tiempo las fugas que se presentan en los barrios o las conexiones fraudulentas que realizan los usuarios. Bogotá necesita un sistema más eficiente para cuidar y valorar el recurso hídrico, concientizar a los usuarios en hacer uso eficiente del agua desde sus casas, hacer que el usuario participe de su propio consumo al implementar una tecnología que permite la transparencia con los clientes al conocer en detalle cómo está usando el agua a diario, permitirá disminuir el número de quejas por cobros excesivos, errores en la toma de mediciones bimensuales y cortes de agua por fugas detectadas a destiempo.

Ahora bien, una vez se hace todo el análisis exploratorio de los datos, es importante validar matemáticamente que todas esas deducciones gráficas que se hicieron sean verídicas por medio de la clusterización, garantizando agrupar los usuarios por consumo y de esta forma estudiar la posibilidad de crear una tarificación dinámica que incentive o penalice a los hogares por un bajo o alto consumo de agua respectivamente. En el siguiente capítulo se trabajará con clústeres para determinar que las agrupaciones anteriormente descritas sean verídicas.

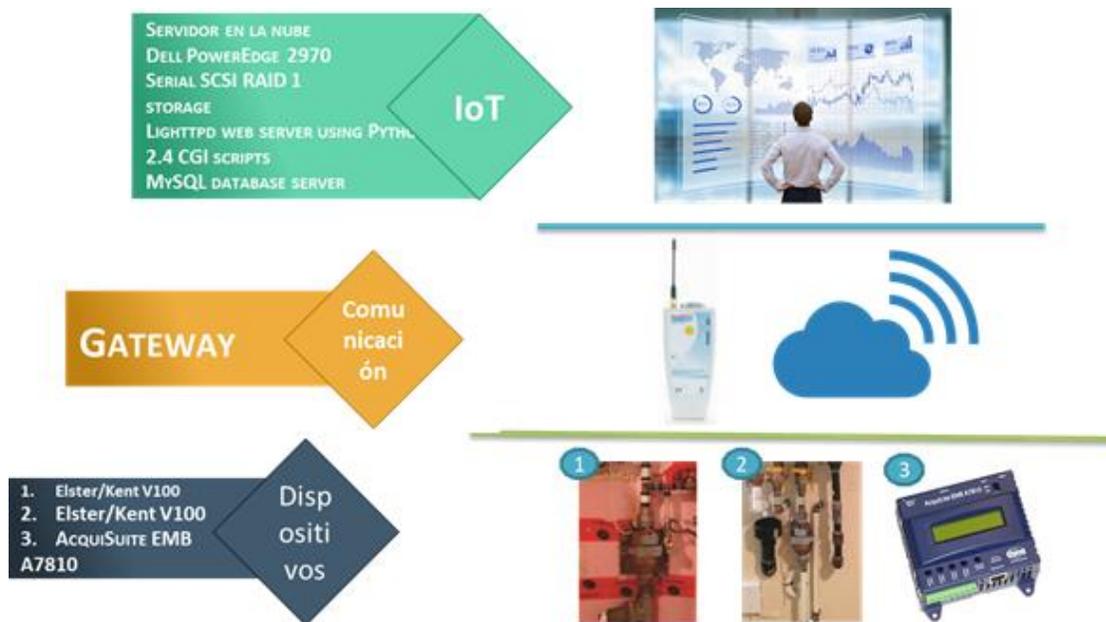
5 Arquitectura de IoT mínima para recolección de datos de agua en hogares

5.1 Arquitectura para monitoreo de agua

Para seleccionar una arquitectura idónea para el monitoreo de agua, se tomó como referencia la Norma ISO / IEC / IEEE 42010, la cual contiene el estándar IEEE P2413 con el marco de referencia para diseñar una arquitectura IoT en una Smart City, este marco fue aprobado el 14 de junio de 2018 y hasta el momento es el estándar más reciente que da una guía de los componentes básicos que se deben tener en cuenta a la hora de implementar proyectos con tecnología IoT.

Este marco de referencia menciona que una arquitectura IoT para una Smart City cuya aplicación está enfocada en la gestión del agua, debe estar definida por capas como lo son: capa de dispositivos o sensores que harán las mediciones esperadas, capa de red de comunicación por la cual se hará la transferencia de los datos recolectados por los dispositivos, la capa de la plataforma IoT que define el software que consumirá los datos recolectados para transformarlos en información y entrega la información para ser visualizada por el usuario final.

Figura 29. Arquitectura IoT para Monitoreo de Agua



Capa de dispositivos

Con el propósito de dar solución a la problemática que en la actualidad la Empresa de Acueducto y Alcantarillado de Bogotá presenta en cuanto a monitoreo y consumo en tiempo real por hogar en la ciudad, se propone hacer uso de lectores de medición automática en inglés Automated Meter Reading, abreviadamente AMR o medidores inteligentes.

Los medidores inteligentes permiten que desde el centro de control de la Empresa de Acueducto y Alcantarillado se tomen las mediciones por hogar de manera remota, garantiza la confiabilidad de los datos y que en dado caso la factura no haya sido cancelada se suspenda el servicio.

La ventaja de instalar medidores inteligentes en Bogotá es que le ayudará a la Empresa de Acueducto y Alcantarillado tomar decisiones en tiempo real, generando patrones de consumo de cada uno de los hogares y de esta manera incentivar a los usuarios al ahorro del recurso hídrico.

a. Ventajas de instalar medidores inteligentes:

- ✓ Empoderamiento del usuario: Le permite al usuario final acceder desde su teléfono móvil a los datos reales de consumo y tomar medidas de comportamiento para hacer uso adecuado del recurso.
- ✓ Tarifación intra-diaria: Garantiza la medición en tiempo real del consumo por hogar en Bogotá.
- ✓ Eficiencia operativa: Garantiza eficiencia en la toma de medidas ya que ésta se realiza de forma remota.
- ✓ Integrar recursos hídricos distribuidos: Crear una red de lectura remota de todos los hogares en la ciudad de Bogotá por barrio, zona y localidad.
- ✓ Desarrollar la respuesta de la demanda: Optimizar el proceso de abastecimiento en cada una de las zonas de la ciudad.
- ✓ Habilitar nuevos modelos de negocio: Generación de empleos e ideas de negocio en el manejo de tecnologías IoT.

b. Dispositivos:

Para este estudio se tomará como referencia la marca Zenner marca líder a nivel mundial con sedes en Europa, América del Norte, Central y Sur, Este Medio y África y con una de sus sedes en Colombia. Especialista en sistemas de medición de agua bajo la tecnología AMR.

El sistema que será propuesto como solución a la medición inteligente de agua en la ciudad de Bogotá funciona con una transmisión de datos unidireccional. Los dispositivos de medición envían de forma autónoma un protocolo de datos a intervalos determinados. Los protocolos de radio están codificados con los medios más avanzados por motivos de seguridad. Se procesan con el receptor de radio MinoConnectRadio y con un portátil o tableta con software de lectura y se guardan. Los datos de lectura registrados y los parámetros de contador correspondientes se transfieren al centro de control desde el terminal portátil al PC y se encuentran disponibles para la facturación o la monitorización de agua (ZENNER International GmbH & Co. KG, 2019). Se entrega además un valor aproximado de los dispositivos, teniendo en cuenta que estos valores son unitarios y pueden medir hasta una zona residencial, los costos dependerán netamente de la cantidad de dispositivos que se requieran y de la arquitectura que finalmente se elija para hacer la toma de mediciones (KG, 2019)

Tabla 13. Inventario de Dispositivos Tecnología AMR.

Dispositivo	Descripción	Valor aproximado
 <p>EDC-wM-Bus</p>	Permite la lectura a distancia de contadores de agua. Los datos de contador se transfieren a un terminal portátil a través del receptor de radio MinoConnect Radio.	29.32 USD
Zelsius serie C5	Dispone de un módulo M-Bus integrado que transfiere los datos directamente al dispositivo de lectura.	205 USD

		
<p>Contador de agua convencional</p> 	<p>Contador convencional de agua con salida de impulsos.</p>	<p>75,80 USD</p>
<p>PulseRadioGateway con interfaz M-Bus</p> 	<p>Módulo alternativo para el conteo de impulsos o es posible integrar contadores con salida de impulsos.</p>	<p>97.12USD</p>

Fuente: Realización propia del autor.

Capa de red de comunicación

Basado en la misma tecnología desarrollada por Zenner se propone hacer uso del MinoConnectRadio, encargado de recibir los datos que se generan de los contadores de agua Zelsius serie C5 y los transmite por Bluetooth al terminal portátil. El dispositivo de lectura móvil a batería procesa los modos y puede utilizarse para la lectura de todos los dispositivos de medición en conformidad con OMS (Open Metering System), incluso de otros fabricantes. Además, el dispositivo también puede utilizarse en combinación con un cable de conexión adecuado para la lectura de equipos de M-Bus, dado que dispone de una interfaz RS232 adicional. Este dispositivo maneja un valor aproximado de 2219,52 USD.

Figura 30. MinoConnectRadio Marca Zenner.



Capa de plataforma IoT

Según el cuadrante mágico de Gartner dentro de las mejores plataformas de IoT que existen en el mercado a 2019 se ubican con liderazgo tres compañías así:

Figura 31. Cuadrante Mágico de Gartner Plataformas IoT



Tabla 14. Software con Mayor Prestigio para Análisis de Datos en Servicios Públicos.

Compañía	Software	Descripción
Software AG	Cumulocity IoT	Administración de dispositivos, reglas inteligentes y aplicaciones IoT preconfiguradas, y componentes opcionalmente integrados como análisis en tiempo real, integración empresarial y en la nube, y administración de datos y API.
PTC	ThingWorx	Monitoreo de activos y el mantenimiento predictivo de múltiples activos en el campo, como flotas automotrices y productos industriales conectados, y operaciones de fabricación, como bienes de consumo y productos electrónicos
Hitachi	Lumada	Enfocada en ciudades inteligentes e industrias pesadas. Ofrece una solución completa de extremo a extremo que satisface los tipos de instalación local y en la nube para industrias intensivas en activos como la fabricación, el transporte, la energía y los servicios públicos.

Fuente: Realización propia del autor.

Como parte de esta propuesta se recomienda utilizar la plataforma Hitachi Visualization Suite con tecnología de nube híbrida y cómputo perimetral de Lumada para clientes de ciudades inteligentes, ya que a través de la conexión de los diferentes dispositivos IoT permite tener una visión amplia de las conexiones, registrando alertas para la toma de decisiones a tiempo. Por otra parte está focalizada en servicios públicos lo que permitirá no solo integrar servicio de agua sino los demás servicios que se prestan en la ciudad como lo son gas y luz.

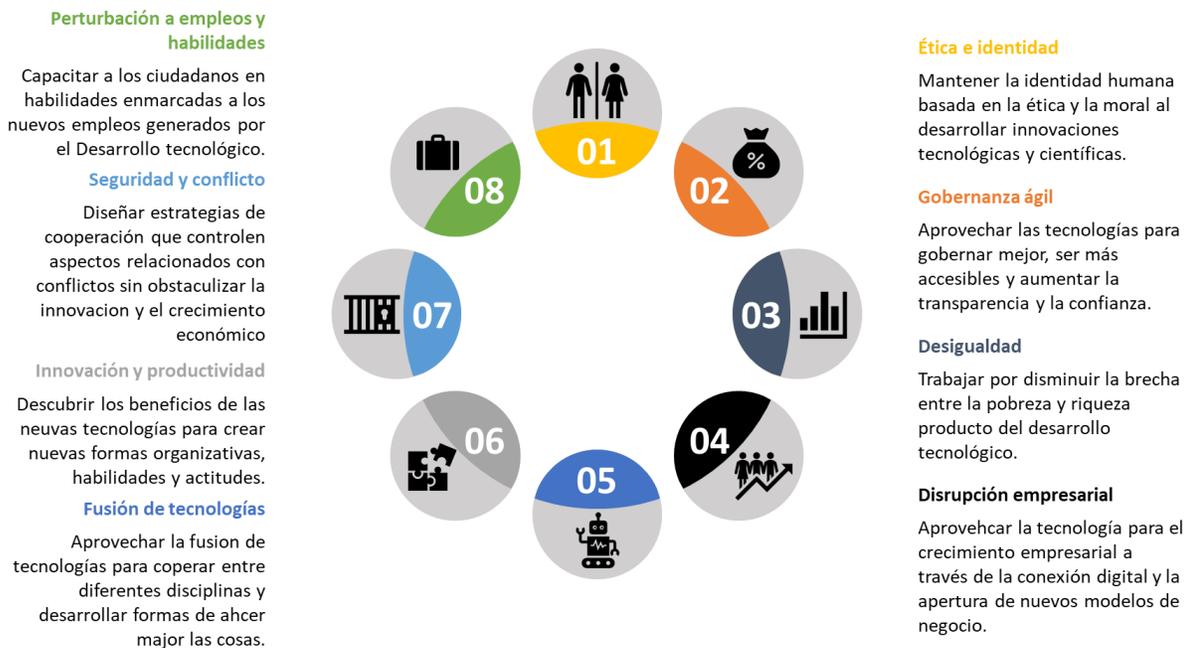
La idea de implementar tecnologías es hacerlo de manera global, hacer una inversión que garantice la optimización de los recursos en Bogotá, pensar en una solución que abarque más de un servicio conlleva a que Colombia se enfoque en convertirse en una ciudad inteligente a futuro, que a través de una plataforma como la escogida se integre no solo servicios públicos sino el monitoreo de transporte, seguridad y demás campos de interés de las ciudades que garanticen a los ciudadanos confianza y seguridad.

6 Impacto Sobre la Cuarta Revolución Industrial

De acuerdo con el foro económico mundial la cuarta revolución industrial abarca 8 fuerzas a nivel mundial en las cuales deben centrarse las grandes economías del mundo para generar impacto en el desarrollo de sus países y en los lineamientos a seguir para que sus economías se centren en la nueva era digital, produciendo crecimiento económico, bienestar a sus habitantes e innovación y emprendimiento.

Dentro de estas ocho fuerzas, tres de ellas enfatizan en el uso del Internet de las cosas y cómo este genera nuevos productos, servicios y modelos de negocio, y, además, está creando formas para que los Gobiernos aumenten la participación pública y entreguen mejores servicios.

Figura 32. Fuerzas de impulso para las grandes economías mundiales.



6.1 Gobernanza ágil

Un gobierno debe ser transparente ante sus ciudadanos, hacer que sus operaciones sean rápidas y no generen tanto bloqueo a la hora de que la población requiera realizar algún trámite, queja o reclamo por alguna problemática que se presente a nivel de los servicios públicos, por ejemplo. Un gobierno que facilite la comprensión en la lectura de los consumos en sus hogares, al

implementar tecnologías que garanticen mediciones en tiempo real, detección de fugas a tiempo y perfilamiento de usuarios con el fin de generar campañas educativas en el manejo adecuado del agua, incentivos para los ciudadanos que hagan uso de agua en horas valle y a nivel ambiental, reduciendo el Índice de Agua No Contabilizada (IAND) que según la Comisión de Regulación de Agua Potable y Saneamiento Básico (CRA) para Bogotá está en un 30%, es un gobierno ágil.

Si la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB) inicia su proyecto basado en el Internet de las Cosas con la instalación de medidores inteligentes (AMR) logrará contribuir con:

- Detección de fugas.
- Detección de conexiones fraudulentas.
- Protección de ingresos tanto de hogares como de la EAAB.
- Mediciones exactas y en tiempo real.
- Transparencia en los cobros, porque el cliente paga solo lo que consume.
- Concientización de los clientes, ya que ahorran y a menor consumo menor cobro.
- Mejora en la distribución de agua al conocer los valores exactos de la demanda de agua por hogares.

6.2 Disrupción empresarial

La evolución de las tecnologías ha traído consigo que las empresas que no se involucran con nuevas tecnologías, se queden atrás y por ende salgan del mercado puesto que todas aquellas empresas que se suman a la era digital y la innovación tecnológica hacen apertura de mercados y de la economía, entregando productos y servicios que faciliten la vida de quien los adquiere.

Una vez la EAAB implemente una solución basada en internet de las cosas, va a necesitar quien provee de estos equipos, el software que le ayude a analizar los datos, personal con alta capacidad de análisis, por lo tanto, es una gran oportunidad para los sectores de abrir su mercado y desplazarse hacia la era digital, capacitar a su personal y por lo tanto desarrollar nuevos productos que garanticen una alta capacidad de procesamiento de datos y emprender hacia la toma de decisiones para mejorar en la distribución del agua en Bogotá.

6.3 Seguridad y conflicto

Un gobierno que desarrolle soluciones basadas en tecnologías de la información debe garantizar que los datos personales de las personas sean privados y que no genere una oportunidad para los ciberdelincuentes de aprovecharla para cometer actos de vandalismo.

En el caso de que el gobierno de Colombia y la alcaldía de la ciudad de Bogotá decida implementar una solución basada en el Internet de las Cosas para optimizar el proceso de distribución de agua y facturación del servicio en la ciudad, deberá garantizar que los datos que esta tecnología recoja, estén salvaguardados y firmar acuerdos de confidencialidad con las empresas que hagan parte del equipo de desarrollo de esta solución puesto que las bases de datos que de aquí se desprendan, contendrán información detallada de la ubicación de cada hogar y los datos de contacto de los clientes a quienes se les presta el servicio de agua, lo que podría ser un producto de alto costo para venderlo a delincuentes que trabajen para atacar a personas de acuerdo con su estrato socioeconómico.

7 Conclusiones y Recomendaciones

7.1 Conclusiones

- a) Los resultados de esta investigación ratifican el poder de la información que se puede extraer desde dispositivos IoT, en especial con AMR conexos como contadores automáticos de lectura de agua.
- b) La investigación detalla los lineamientos a seguir para extraer, tratar y analizar la data para que los hallazgos obtenidos puedan utilizarse como soporte a la toma de decisiones en un sector tan vital como el agua potable.
- c) En casos como Londres y Alicante con un nivel de detalle de los datos por hora y por fecha se identificaron gráficamente los evolutivos de la demanda y el perfilamiento de los usuarios con una granularidad temporal tan detallada que facilita la gestión no solo a nivel del usuario, sino que en el agregado general se pueden usar para planificar la distribución del agua, programar el consumo y, por tanto, mejorar el desempeño de acueductos urbanos.
- d) En el caso de Bogotá, al tener un nivel de granularidad muy alto (bimensual) no se logra explorar un comportamiento pormenorizado ni de la ciudad ni de sus suscriptores, es decir, no es posible identificar tipología de consumo ni por días, semanas, meses y mucho menos, llevar un reporte en tiempo real del consumo de agua de sus usuarios.
- e) Se sugiere entonces al Acueducto de Bogotá comenzar a participar en este tipo de arquitecturas, con fines de iniciar un proceso de mejora continua en la gestión del agua potable ante los desafíos del futuro cercano.

7.2 Trabajos futuros

- a) Implementar la arquitectura IoT propuesta y hacer la toma de mediciones de agua en un grupo seleccionado para la ciudad de Bogotá con el fin de agregar valor al estudio realizado, demostrando con datos reales la veracidad de toda la investigación.
- b) Implementar las técnicas de clusterización aquí estructurada en la base de datos tomada con AMR en la ciudad de Bogotá y evidenciar los patrones de comportamiento de los usuarios, con el fin de agruparlos según su consumo y

comenzar a desarrollar políticas de precios que incentive a los usuarios en el uso adecuado del agua.

- c) Realizar no solo el estudio a nivel hogar sino industrial para ampliar el estudio en la ciudad de Bogotá y validar que sectores son aquellos que mayor consumo de agua generan en la ciudad con el fin de aplicar políticas ambientales para el uso del recurso a gran escala y motivar a los sectores a que implementen alternativas para el reciclaje de agua.
- d) Participar con esta investigación en proyectos relacionados con la revolución 4.0 que está implementando en la actualidad el gobierno para llegar a más sectores que puedan patrocinar la idea y la creación de una empresa dedicada al desarrollo de este tipo de ideas con el fin de que Colombia comience a encabezar la lista de los países con ciudades inteligentes.

Bibliografía

- Alaa, M., Zaidan, A. A., Zaidan, B. B., Talal, M., & Kiah, M. L. M. (2017, noviembre 1). A review of smart home applications based on Internet of Things. *Journal of Network and Computer Applications*, Vol. 97, pp. 48-65. <https://doi.org/10.1016/j.jnca.2017.08.017>
- Candelieri, A., & Archetti, F. (2014). Smart water in urban distribution networks: Limited financial capacity and Big Data analytics. *WIT Transactions on the Built Environment*, 139, 63-73. <https://doi.org/10.2495/UW140061>
- Candelieri, Antonio. (2017). Clustering and support vector regression for water demand forecasting and anomaly detection. *Water* (Switzerland), 9(3). <https://doi.org/10.3390/w9030224>
- Celino, I., & Kotoulas, S. (2013). Smart cities. *IEEE Internet Computing*, 17(6), 8-11. <https://doi.org/10.1109/MIC.2013.117>
- Everitt, B. S., Landau, S. and Leese, M. (2001), *Cluster Analysis*, 4th Edition, Oxford University Press, Inc., New York; Arnold, London. ISBN 0-340-76119-9
- Gutierrez Camelo, A. A., Devia Llanos, M. A., & Tarazona Bermudez, G. M. (2016). Research inteligencia de negocios: estudio de caso sector tecnológico colombiano. *Redes de Ingeniería*,
- KG, Z. I. G. & C. (2019). TARIFA PVP DE CONTADORES ZENNER.
- Lucía, M., Blanco, R., Patricia, N., Castañeda, G., Holmes, C., García, T., ... Montaña, G. (2019). CONSEJO NACIONAL DE POLÍTICA ECONÓMICA Y SOCIAL CONPES Iván Duque Márquez Presidente de la República. Recuperado de <https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/3968.pdf>
- Ministerio de Empleo y Seguridad Social. (2019). Informe del Mercado de Trabajo de Trabajo Estatal. 19. Recuperado de https://www.sepe.es/contenidos/que_es_el_sepe/publicaciones/pdf/pdf_mercado_trabajo/imt2017_datos2016_estatal_general.pdf
- Polit, I. U. (2012). El sistema de abastecimiento de agua en ciudades inteligentes.
- Peña-Guzmán, C. A., Melgarejo, J., & Prats, D. (2016). El ciclo urbano del agua en Bogotá, Colombia: Estado actual y desafíos para la sostenibilidad. *Tecnología y Ciencias del Agua*, 7(6), 57-71.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Roth, E. (2001). Water Pricing in the EU A Review. Recuperado de www.eeb.org

Sidorov, Grigori; Gelbukh, Alexander; Gómez-Adorno, Helena; Pinto, David. «Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model». (2014) *Computación y Sistemas*, 18 (3): 491-504.

Ward, J. H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 58, 236–244.

ZENNER International GmbH & Co. KG. (2019). Tecnología de sistemas AMR.

Abreviaciones

IOT Internet of Things (Internet de las cosas)

AMR Automated Meter Reading (lectores de medición automática)

Anexos

Anexo 1/Apéndice Capítulo 4. Comprobación analítica de los resultados obtenidos con el Análisis Exploratorio de Datos.

Preliminares

El presente apartado constituye un apéndice de la investigación realizada como extensión del capítulo 4. Permite mostrar al lector del documento algunas aplicaciones sencillas sobre analítica avanzada aplicadas a datos IoT y como sustentación para validar resultados del análisis exploratorio cubierto en esta investigación. La idea consiste en corroborar los resultados obtenidos mediante la metodología de análisis exploratorio y visual sustentada con los datos obtenidos a través de las plataformas IoT conexas al consumo de agua urbano.

Si bien, en la investigación se trataron diferentes bases de datos, aquí se testean algunas hipótesis (comprobación de los resultados obtenidos) para el caso de la ciudad de Londres, ya que el mismo procedimiento o técnicas pueden aplicarse a cualquier base de datos, siempre y cuando dichas bases de datos posean la granularidad temporal necesaria para establecer patrones y perfiles en el eje tiempo. Por tanto, a mayor granularidad, mayor será la profundización de los resultados en términos de la variable tiempo (tipología de días, horas, estaciones, etc.).

Base de datos Londres

Recordemos que la base de datos etiquetada como Londres recopila el consumo de agua en una instalación industrial en la ciudad de Londres a lo largo de casi 2 años y en lecturas cada media hora durante todo el día. Los datos comienzan con registros diarios desde el 29 de julio de 2010 hasta 14 de junio de 2012 y cada día reporta el consumo de agua cada media hora en variables etiquetadas como X00.00 hasta X23.30.

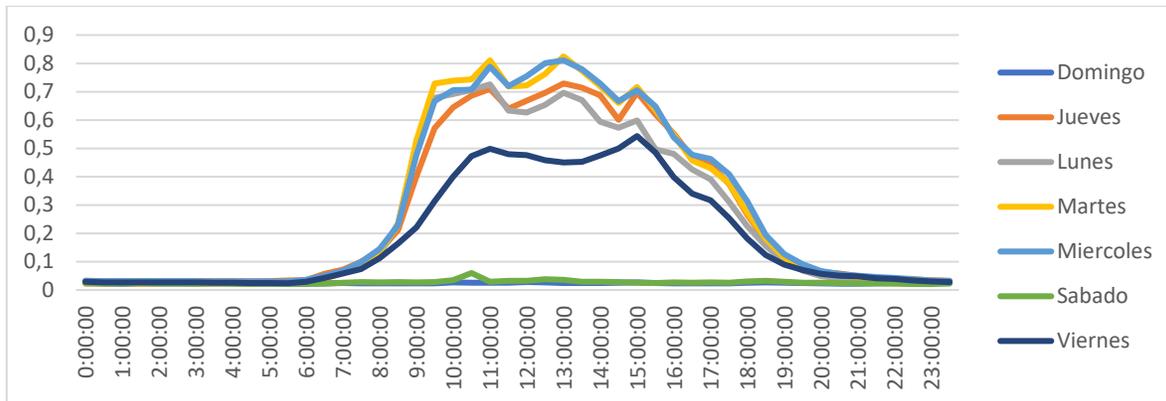
1. Aplicación de Analítica Avanzada para el caso de Londres – Post-procesamiento de curvas horarias de consumo de agua para varios bloques temporales y aplicación de algoritmos no supervisados (Clustering)

a. Tipificación del consumo – Nivel “Días”

Considerando el consumo promedio cada media hora y separando cada día según su día calendario correspondiente (lunes, martes hasta domingo) se procedió a construir la curva horaria (cada media hora) del consumo durante los 2 años de datos recopilados en Londres.

Los resultados se muestran en el gráfico adjunto:

Figura 33. Figura 1. Anexo 1: Tipificación de la curva horaria de consumo según el tipo de día



Se aprecia a través de las curvas específicas para cada día que existe un posible agrupamiento o perfilamiento de tipología de día en base a la similitud de las curvas temporales. Dicho perfilamiento podría denotarse estableciendo que los martes y miércoles son muy similares en su consumo diario; lunes y jueves parecen tener alta relación, el viernes forma un perfil aislado al igual que los sábados y domingos, respectivamente. Parece que existe un comportamiento que varía no en forma sino en magnitud (consumo) según sea el día de la semana: martes y miércoles (días de mayor consumo), seguidos por el grupo de lunes y jueves (días de consumo estándar) y ya, a partir del viernes el consumo baja drásticamente, efecto que se verifica con consumo casi nulo durante los fines de semana. Esta conclusión puede sustentar que el consumo a nivel diario en dicha instalación tiene comportamientos similares durante los días laborables (nótese que el contorno de las curvas es muy parecido) y que los fines de semana existe un cierre de actividades ya que el consumo es prácticamente cero en promedio.

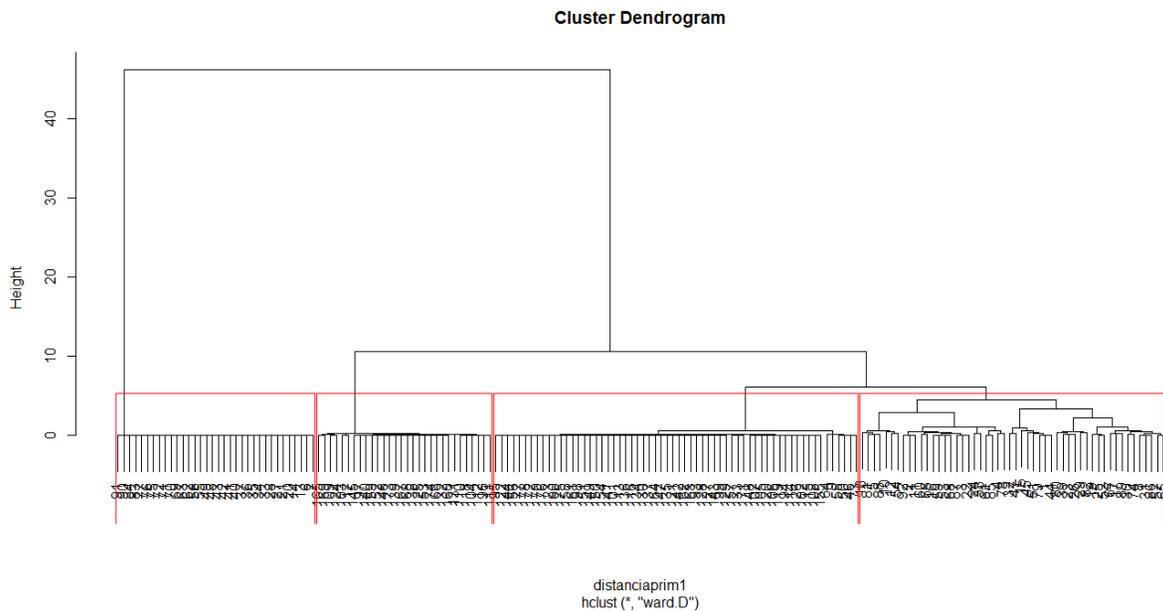
Este análisis realizado con las llamadas “curvas de consumo horario según día” (ver figura 1) permite verificar la hipótesis de la “jornada laboral” establecida con el análisis exploratorio de datos que estableció una clara diferencia entre el consumo para días laborables y los fines de semana (sábados, domingos e incluso días festivos).

Nótese que un simple gráfico con técnicas de analítica de visualización y post-procesamiento es capaz de detectar mayores detalles a los obtenidos con una exploración inicial de datos.

A los fines de sustentar aún más dicha hipótesis, se procedió a realizar un clustering jerárquico (Everitt et al., 2001) con la data. Los registros representan los casi 700 días y las variables discriminantes son los diferentes consumos horarios (cada media hora). Dicho clustering usa el método “Ward” (Ward, 1963) y distancia matemática del coseno (Sidorov et al., 2014). Su ejecución se hizo en lenguaje en R (R, 2020) y los resultados (ver figura 2) permitió agrupar los días en 2 o 4 grupos factibles. Si cortamos para 2 grupos se llega a la conclusión que el consumo es diferenciado entre los días laborables (lunes a viernes) y los fines de semana incluidos también días festivos según el calendario británico. Si cortamos en 4 el resultado nos lleva al mismo

perfilamiento descrito anteriormente: martes y miércoles (grupo 1), lunes y jueves (grupo 2), viernes (grupo 3) y sábados, domingos y festivos (grupo 4).

Figura 34. Figura 2 Anexo 1: Dendrograma para extraer los clústeres (tipo de día)



b. Tipificación del consumo – “Diferenciación Estacional y Mensual”

Siguiendo las mismas líneas del apartado anterior se procedió a tipificar el consumo a nivel estacional para detectar posibles diferencias según la estación climática del año (consumo promedio horario según las fechas acordes a cada estación del año). La idea es identificar cambios o influencias a lo largo del año a un nivel superior al anterior (de días a meses o estaciones) que permitan monitorear mejor el consumo, establecer políticas de manejo adecuado del agua e incluso poder de negociación del cliente hacia el proveedor del servicio a la hora de establecer políticas de contratación de tarifas dinámicas basadas en esos bloques mensuales “estacionales”.

Figura 35. Figura 3 Anexo 1: Tipificación de la curva horaria de consumo según la estación del año

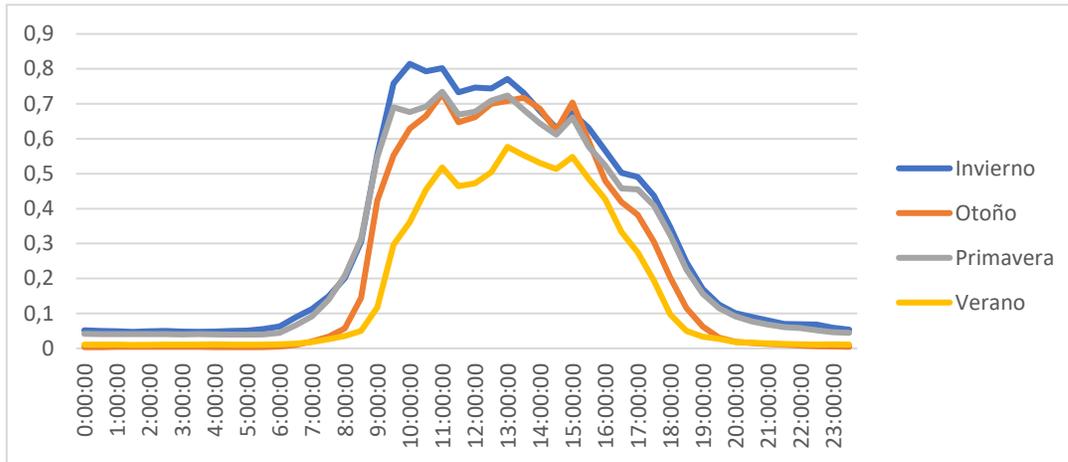
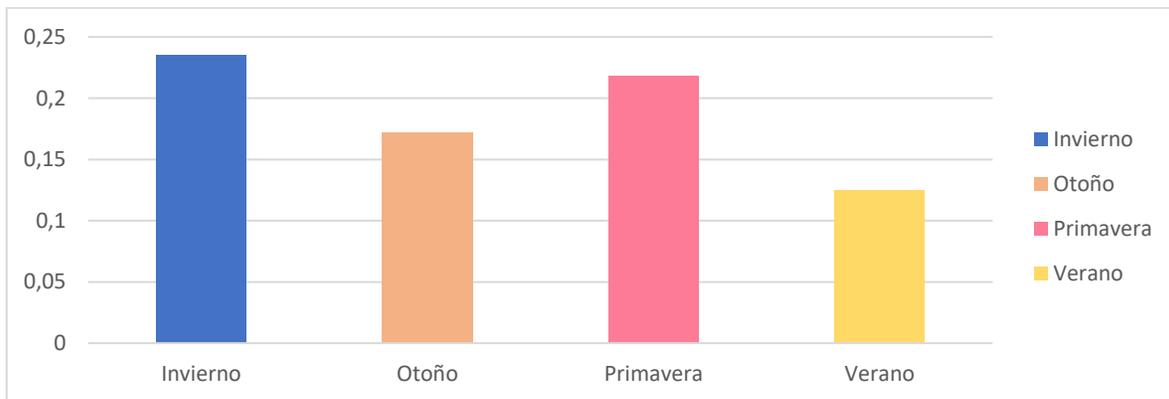
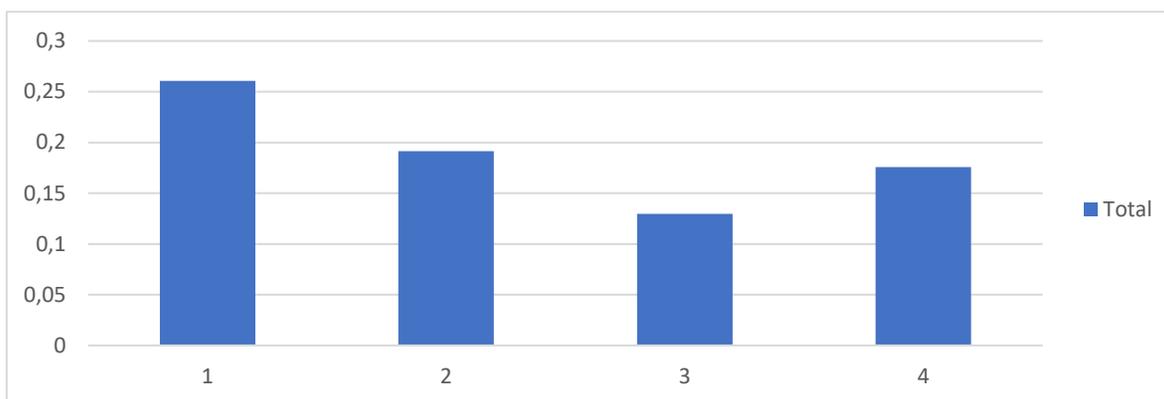


Figura 36. Figura 4 Anexo 1: Tipificación de la curva horaria de consumo según la estación del año



Del gráfico adjunto (ver figuras 3 y 4) puede verse que existen cuatro o tres grupos bien diferenciados en cuanto a la tipificación de la curva estacional: El invierno es la época de mayor consumo, seguida por la primavera y el Otoño (los cuales pueden fusionarse en un solo grupo dada su altísima similitud) y finalmente, se tiene al verano como la época de menor consumo, quizás debido al hecho que es el mes de vacaciones colectivas comunes en Europa. El algoritmo de clustering también arrojó los mismos resultados, pero expresados por trimestres bien diferenciados al año (que van en línea) con el análisis de curvas mostrado por estaciones. Seguidamente se muestra el gráfico resultante al agrupar los días en sus respectivos trimestres:

Figura 37. Figura 4 Anexo 1: Tipificación del consumo según el trimestre del año



Puede apreciarse (ver figura 4) que el primer trimestre del año (enero a marzo) es cuando mayor se consume agua (que va en línea con la curva del invierno), seguida por los trimestres 4 y 2 (octubre a diciembre junto a abril a junio que corresponden al otoño y primavera, respectivamente). El menor consumo se evidencia en el tercer trimestre del año (julio, agosto y septiembre) que coincide con el verano del hemisferio norte.

Conclusiones

Resulta interesante denotar por una parte que si bien el análisis exploratorio y visual sugerido para esta investigación, permitió establecer y obtener resultados de perfiles muy útiles para monitorear el consumo de agua, el uso adecuado del mismo e incluso, para tipificar clientes según su consumo en diferentes niveles temporales (hora, días, meses, estación del año), hallazgos que constituyen ejes fundamentales para la toma de decisiones en el manejo urbano de recursos hídricos; dichos hallazgos por otra parte, pueden validarse y profundizarse aún más con técnicas de analítica avanzada para garantizar la contundencia de los resultados y para demostrar empíricamente los mismos, evitando cualquier sesgo que pueda presentarse por parte del analista que decida únicamente abordar la gestión del conocimiento a un nivel de Inteligencia de Negocios. El poder que representan esos datos para abordar temáticas de clusterización e incluso de predicción son de extrema utilidad, ya que con data a ese nivel de granularidad temporal se pueden diseñar algoritmos de “Machine Learning” e Inteligencia Artificial para predecir curvas de consumo a futuro con esos mismos datos y variables exógenas del sistema subyacente del consumo de agua como clima, elasticidad ante subida o bajada de precios, estratos económicos, etc.

Es por ello por lo que los autores recomiendan como líneas futuras de investigación aplicar técnicas de analítica avanzada que permita enriquecer los estudios como el abordado en esta investigación. Es indudable que el modelado basado en datos y el poder de la data obtenida a través de IoT puede usarse para crear plataformas inteligentes, interactivas y sustentables

orientadas en la optimización de las redes hídricas urbanas en un entorno de ciudades sostenibles e inteligentes “Smart-Cities”.