


# ***Los sistemas de abastecimiento de agua potable: prevención y detección temprana de fugas***

*Santiago Linder Rubiños Jiménez  
Abilio Bernardino Cuzcano Rivas  
Adán Almircar Tejada Cabanillas  
Fernando Mendoza Apaza  
Antenor Leva Apaza  
Juan Antonio Apesteguía Infantes*

**CIDE**  
EDITORIAL



*Los sistemas de abastecimiento de  
agua potable: prevención y  
detección temprana de fugas*

*Los sistemas de abastecimiento de  
agua potable: prevención y  
detección temprana de fugas*

**Autores:**

**Santiago Linder Rubiños Jiménez**

**Abilio Bernardino Cuzcano Rivas**

**Adán Almircar Tejada Cabanillas**

**Fernando Mendoza Apaza**

**Antenor Leva Apaza**

**Juan Antonio Apesteguía Infantes**

## Los sistemas de abastecimiento de agua potable: Prevención y detección temprana de fugas

Reservados todos los derechos. Está prohibido, bajo las sanciones penales y el resarcimiento civil previstos en las leyes, reproducir, registrar o transmitir esta publicación, íntegra o parcialmente, por cualquier sistema de recuperación y por cualquier medio, sea mecánico, electrónico, magnético, electroóptico, por fotocopia o por cualquiera otro, sin la autorización previa por escrito al Centro de Investigación y Desarrollo Ecuador (CIDE).

Copyright © 2023

Centro de Investigación y Desarrollo Ecuador

Tel.: + (593) 04 2037524

<http://www.cidecuador.org>

ISBN: 978-9942-636-36-2

<https://doi.org/10.33996/cide.ecuador.AP2636362>

**Dirección editorial:** Lic. Pedro Misacc Naranjo, Msc.

**Coordinación técnica:** Lic. María J. Delgado

**Diseño gráfico:** Lic. Danissa Colmenares

**Diagramación:** Lic. Alba Gil

**Fecha de publicación:** noviembre, 2023



Guayaquil - Ecuador

La presente obra fue evaluada por pares académicos  
experimentados en el área

### **Catalogación en la Fuente**

Los sistemas de abastecimiento de agua potable: Prevención y detección temprana de fugas/ Santiago Linder Rubiños Jiménez, Abilio Bernardino Cuzcano Rivas, Adán Almircar Tejada Cabanillas, Fernando Mendoza Apaza, Antenor Leva Apaza, Juan Antonio Apestegua Infantes. - Ecuador: Editorial CIDE, 2023.

149 p.: incluye tablas, figuras; 21,6 x 29,7 cm.

ISBN: 978-9942-636-36-2

1. Agua potable 2. Algoritmos predictivos

## *Acerca de los autores*

### **Santiago Linder Rubiños Jiménez**

Correo: [r.rservice@yahoo.es](mailto:r.rservice@yahoo.es); [slrubinosj@unac.edu.pe](mailto:slrubinosj@unac.edu.pe)

<https://orcid.org/0000-0003-0095-6988>

Doctor en Ingeniería Eléctrica, Doctor en Gestión Pública y Gobernabilidad, Maestro en Gestión Pública, Maestro en Ingeniería Eléctrica con mención en Gestión de sistemas de Energía Eléctrica, Bachiller de la carrera profesional en ingeniería eléctrica. Titulado con experiencia laboral en el área de docencia, Proyectos de Ingeniería e Investigación.



Persona formada humanista, científica y tecnológicamente, caracterizada por cualidades, como: Responsable, emprendedor, puntual, comunicativo con capacidad de trabajar en equipo, colaborador, respetuoso y principalmente honrado. Actualmente se desempeña como Director del Instituto de Investigación del Vicerrectorado de Investigación de la Universidad Nacional del Callao.

## **Abilio Bernardino Cuzcano Rivas**

Correo: acuzcano75@gmail.com; abcuzcanor@unac.edu.pe

<https://orcid.org/0000-0002-9823-8740>

Doctor en Ingeniería Eléctrica, Maestro en Ingeniería electrónica con mención en biomédica, Bachiller de la carrera profesional en Ingeniería Electrónica. Titulado con experiencia laboral en el área de docencia, proyectos de Ingeniería e investigación. Persona formada humanista, científica y tecnológicamente, caracterizada por



cualidades, como: responsable, emprendedor, puntual, comunicativo con capacidad de trabajar en equipo, colaborador, respetuoso y principalmente honrado. Actualmente se desempeña como director en la Unidad de Posgrado de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao.

## Adán Almirar Tejada Cabanillas

Correo: [atejadacabanillas@yahoo.es](mailto:atejadacabanillas@yahoo.es); [aatejadac@unac.edu.pe](mailto:aatejadac@unac.edu.pe)

<https://orcid.org/0000-0002-9736-3654>

Doctor en Ingeniería Eléctrica, Maestro en Economía con mención en Métodos Cuantitativos de la Economía y en Ingeniería Eléctrica con mención en Gestión de Sistemas de Energía Eléctrica, Bachiller y Licenciado en Estadística. Directivo académico con una sólida experiencia en liderar y gestionar una Facultad de



Ingeniería. 30 años de dedicación al fomento de la excelencia educativa y el desarrollo de programas innovadores en ingeniería. Destacado por la mejora de la calidad de la enseñanza-aprendizaje y el aumento de la investigación. Experto en la administración de recursos y la planificación estratégica para el crecimiento sostenible de la Facultad. Comunicador eficaz y líder inspirador, capaz de promover la colaboración entre profesores, estudiantes y personal administrativo. Comprometido con el avance continuo de la Facultad y la preparación de futuros líderes en ingeniería. Actualmente se desempeña como director del departamento de Ingeniería Eléctrica y director de escuela de Ingeniería Eléctrica de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao.



## Fernando Mendoza Apaza

Correo: mastermetric@gmail.com; fmendozaa@unac.edu.pe

<https://orcid.org/0000-0001-7981-8291>

Ingeniero Electrónico de la Universidad Nacional del Callao. Con estudios en Doctorado en Educación; Maestría en Ingeniería Eléctrica, a través del Ministerio de Energía y Minas; Maestría en Telecomunicaciones; Maestría en Ingeniería de Sistemas, con especializaciones y certificaciones en Normas de Calidad ISO9001 Quality Management Systems Associate, ISO/IEC2000 IT Service Management Associate, ISO/IEC27001 Information Security Associate, SCRUM Foundation Professional Certificate y certificación Lean Foundations Professional Certification, Help Desk Professional Certificate HDPC y capacitación en IT Service Management Basado en ITIL 4. Con experiencia en diferentes empresas e instituciones desempeñando funciones en las áreas de Gestión de proyectos y licitación, construcción y ejecución; Adquisición y puesta en marcha de equipos de telecomunicaciones; Control y automatización; Networking; Servidores; Redes Cisco; Radioenlaces Wireless y CATV (Diseño y optimización). En el área Automotriz con conocimientos en Electrónica automotriz, Diagnóstico y mantenimiento con equipos LAUNC. En el área Eléctrica con desempeño en Planeamiento eléctrico, Planeamiento y soporte. En planta, Sistemas de control y automatización neumáticos e hidráulicos, Scada y PLC. En el ámbito Industrial Minero e Impacto Ambiental, mantenimiento de planta e instrumentación. Capacidad en la planificación y organización de actividades dirigidas al cumplimiento de objetivos, manteniendo y estableciendo adecuadas relaciones a todo nivel, conocimientos bajo las normas con efectividad y dedicación a la labor profesional.



## **Antenor Leva Apaza**

Correo: [aleva116@gmail.com](mailto:aleva116@gmail.com); [alevaa@unac.edu.pe](mailto:alevaa@unac.edu.pe)

<https://orcid.org/0000-0002-0973-0240>

Maestro en Investigación y docencia universitaria de la Universidad Peruana de Ciencias e Informática. Matemático de la Universidad Nacional de San Antonio de Abad del Cusco. Actualmente desempeña labores en la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao. Perfilado como



investigador activo mediante la producción de artículos científicos de conferencias internacionales y revistas de alto impacto.

## **Juan Antonio Apesteguía Infantes**

Correo: [apesteguia@outlook.com](mailto:apesteguia@outlook.com); [jaapestegui@unac.edu.pe](mailto:jaapestegui@unac.edu.pe)

<https://orcid.org/0000-0003-2545-0764>

Maestro en Control y Automatización en la Universidad Nacional del Callao. Ingeniero Electrónico de la Universidad Nacional del Callao. Actualmente es parte de la plana docente en la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao. Profesional en constante capacitación en el uso de nuevas tecnologías para la Docencia. Con amplia experiencia en el campo laboral en el rubro de Control y Automatización. Docente investigador con publicaciones de artículos científicos en congresos internacionales.



# Contenido

Acerca de los autores .....	5
Introducción .....	14

## **Capítulo 1** **Medio ambiente y los recursos hídricos**

1.1 Introducción al capítulo .....	18
1.2 Desarrollo sostenible .....	19
1.3 Desarrollo sostenible y agua .....	22
1.4 Recursos hídricos .....	23
1.5 Políticas de manejo del agua potable .....	25

## **Capítulo 2** **Fugas en sistemas de agua potable**

2.1 Introducción al capítulo .....	30
2.2 Servicio de detección de fugas no visibles en redes primarias (ECRF) .....	31
2.3 Equipos utilizados por el ECRF .....	32
2.4 Consideraciones finales .....	43

## **Capítulo 3**

# **Aplicaciones de la Inteligencia Artificial a la detección de fallas en tuberías**

3.1 Introducción al capítulo .....	45
3.2 Fugas programadas y de emergencia .....	48
3.3 Filtrado y discriminación de datos .....	49
3.4 Objetivos de la sectorización .....	54
3.5 Tipo de fuga .....	55
3.6 Diámetro de la tubería .....	57
3.7 Tipo de tubería .....	58
3.8 Presión .....	58
3.9 Caudal .....	58
3.10 Vehículo .....	59
3.11 Fugas detectadas .....	59
3.12 Observaciones .....	59
3.13 Modelo predictivo .....	59

## **Capítulo 4**

# **Aplicación**

4.1 Introducción al capítulo .....	64
4.2 Aplicación de la machine learning en la detección de fallas en tuberías de agua potable en la ciudad de Lima .....	65
4.3 Carga de datos al software R .....	66
4.4 Librerías a utilizar software R .....	70
4.5 Preprocesamiento de datos software R .....	72
4.6 Uso del comando SPLIT .....	79
4.7 Uso del comando tapply .....	84
4.8 Conversión de un archivo csv a data.frame .....	90
4.9 Tratamiento de los datos usando ggplot .....	92

4.10 Procesamiento de la información mediante algoritmos .....	97
4.11. Población y muestra utilizada para la simulación.....	124
4.11.1 Unidad de análisis .....	124
4.11.2 Población escogida de trabajo.....	124
4.11.3 Tamaño de la muestra trabajado.....	125
4.11.4 Prueba piloto planteada.....	125
4.12 Procedimiento de recolección de datos .....	125

## **Capítulo 5**

### **Resultados**

5.1 Introducción al capítulo .....	128
5.2 Generalización de los resultados .....	129
5.3 Implementación .....	138
Conclusiones finales .....	146
Referencias .....	148

## *Introducción*

El análisis estocástico y numérico con fines predictivos, ha representado una estilo innovador y avanzado para prevenir la aparición de posibles fallas, que tenga una alta probabilidad de suceder en una determinada condición o situación de operación, de equipos, sistemas o estructuras de cuyo podemos citar sobre el particular lo siguiente:

La analítica predictiva es una forma de análisis avanzado que utiliza datos nuevos e históricos para predecir una actividad futura de un evento determinado, identificando comportamientos y las tendencias, lo cual implica la aplicación de técnicas de manejo del dato, entre ellas análisis estadístico, consultas analíticas y algoritmos de aprendizaje automático a conjuntos de datos para crear modelos predictivos que sitúen un valor numérico o puntuación en la probabilidad de que ocurra un evento en particular. (Rubiños & Apesteguia, 2020, p. 8)

En esta cita notamos que las ciencias de la ingeniería le dedican una parte importante de sus estudios al manejo del dato y el producto de este estudio histórico es básicamente la producción de una herramienta para predecir y prevenir la ocurrencia de eventos no deseados que acarreen pérdidas materiales o personales, en el manejo del dato es importante hacer en secuencia correcta la aplicación de los pasos para la adecuada determinación del problema asociado al manejo de datos en (Rubiños &

Apesteagua, 2020) encontramos, en primer lugar obtención y organización de los datos, como segundo momento el procesamiento de los datos, el tercer paso el desarrollo de modelos predictivos y el cuarto paso es validar en última instancia los resultados y aplicar las modificaciones que le permitan al algoritmo en su último paso corregir y tomar nueva información.

Por otro lado, el abordaje que se hace en esta obra está centrado en la distribución de agua potable, la cual en el siempre esta sometidas a pérdidas durante el proceso de trasegó del agua, lo cual implica perdidas monetarias ya que el esfuerzo económico para lograr el abastecimiento de ciudades enteras, se ve perjudicado en la cantidad de agua que no puede ser facturado debidamente al no llegar al equipo de medición de los clientes finales. Estas pérdidas tienen diversos orígenes, por ejemplo: tuberías rotas y/o desgastadas, que fallan ya sea por el tipo de material utilizado, por la antigüedad, por la ubicación, entre otras causas (Rubiños & Apesteagua, 2020). Por los motivos hasta aquí expresados se desarrolla de un algoritmo predictivo, que permite la detección de fugas en las redes de agua potable de la ciudad de la que se disponga la información histórica requerida.

Es entonces este libro una guía para a aquellos lectores interesados en la aplicación de análisis de fallas en sistemas de abastecimiento de agua y de sistemas algorítmicos con la facilidad de convertirse, elementos de aprendizaje, monitoreo y predicción en la detección temprana de fallas. Se debe indicar que los datos que se utilizarán en el desarrollo del algoritmo fueron proporcionados por empresa de agua potable de la ciudad de Lima,



Perú, para los fines declarados en este documento escrito. Sin embargo, el algoritmo y sus procedimientos en su propia versatilidad pueden ser extrapolado a otras ciudades en aplicaciones similares, queda entonces servida la mesa para que nos acompañen a través de la lectura en una hermosa aventura que nos adentra en un tema que tiene alta relevancia por lo importante del agua potable como recurso fundamental para la subsistencia de la humanidad.

# 1

## Capítulo 1

*Medio ambiente y los recursos hídricos*



## Medio ambiente y los recursos hídricos

### 1.1 Introducción al capítulo

El agua es un recurso limitado e insustituible, clave para el bienestar y la sobrevivencia de la especie humana, que puede sólo ser considerado como recurso renovable si está adecuadamente gestionado. Hoy en día, más de 1.700 millones de personas viven en cuencas fluviales, superando la capacidad de recarga natural de las propias cuencas, una tendencia que indica que dos tercios de la población mundial podría vivir en un países con escasez de agua para 2025 (NU, 2014). El agua puede suponer un serio desafío para el desarrollo sostenible, pero gestionada de manera eficiente y equitativa, puede jugar un papel facilitador clave en el fortalecimiento de la resiliencia de los sistemas sociales, económicos y ambientales, a la luz de unos cambios rápidos y adecuados para su mejor manejo y gestión como recurso natural renovable bajo ciertas condiciones.

El presente capítulo, considerado introductorio al tema, está dedicado al conocimiento sobre el Desarrollo Sostenible, su relación con el agua y los recursos hídricos. En el mismo, se hace énfasis en lo que dispone la ciudad de Lima, en Perú, que es una de las pocas en el mundo con la característica particular de estar asentada sobre un desierto, lo cual hace evidente que sus recursos hídricos son escasos.

## **1.2 Desarrollo sostenible**

El desarrollo sostenible se popularizó de manera explícita y contextualizada por la Comisión Brundtland en el documento “Nuestro Futuro Común” donde se define como “el desarrollo que satisface las necesidades del presente sin comprometer la capacidad de las generaciones futuras para atender sus propias necesidades” (ONU, 1987). Esta se centró en tres (3) pilares del bienestar humano: las condiciones económicas, sociopolíticas y ecológicas/ambientales, como se muestra en la Figura 1.

Este concepto básico fue desarrollado como apoyo a la implementación de medidas sólidas dirigidas a impulsar el desarrollo económico y social, en particular para las personas de los países en vías de desarrollo y, al mismo tiempo, garantizar que la integridad del medio ambiente se mantenga para las generaciones futuras.

## Figura 1.

### *Pilares del desarrollo sostenible*



*Nota.* Adaptado de (Cepei, 2021)

El desarrollo sostenible es importante porque equilibra las consideraciones económicas, sociales y ambientales. Ayuda a garantizar que las personas no se vean desfavorecidas de ninguna manera y que se mantenga la salud de la sociedad. Además, asegura un futuro para todos, independientemente de quiénes somos, dónde vivimos y qué hacemos.

El uso reducido de energía, un hábitat saludable para todos y el impacto social son algunos de los beneficios del desarrollo sostenible. El desarrollo económico sostenible puede ayudar a las naciones a hacer crecer sus economías de forma equitativa y respetuosa con el medio ambiente. Los objetivos de desarrollo sostenible son el plan maestro para conseguir un

futuro sostenible para todos. Es importante destacar además que no sólo es beneficioso para el medio ambiente, sino también para la economía y la sociedad en general.

En este sentido, el desarrollo sostenible y el agua están estrechamente relacionados. El agua es un recurso natural vital para la vida y el desarrollo humano y su gestión sostenible es un componente clave del desarrollo. La gestión sostenible del agua ayuda a garantizar que las personas no se vean desfavorecidas de ninguna manera y que se mantenga la salud de la sociedad. Además, el uso reducido de energía, un hábitat saludable para todos y el impacto social son algunos de los beneficios del desarrollo sostenible.

En este sentido, pertinente mencionar el caso de Aqualia, una empresa certificada por AENOR por su contribución a los Objetivos de Desarrollo Sostenible (ODS), apuesta por un modelo de crecimiento centrado en el cuidado del medioambiente y el bienestar de las personas. Esta compañía invirtió 16,8 millones de euros en tecnología durante 2022, un 216% más que el año anterior. La digitalización como meta en sus gestiones, es una de las líneas estratégicas de la empresa, lo que la condujo a activar durante 2022 diferentes proyectos para la detección temprana de fugas en la red de abastecimiento, gracias a estos proyectos, se evitaron pérdidas innecesarias de agua y daños en la red (Aqualia, 2023).

### 1.3 Desarrollo sostenible y agua

La agricultura es, con diferencia, el mayor consumidor de agua a nivel mundial representando el 70% de las extracciones en todo el mundo, aunque esta cifra varía considerablemente entre países (NU, 2014). La agricultura de secano es el sistema de producción agrícola predominante en todo el mundo y su productividad actual es, en promedio, un poco más de la mitad del potencial a obtener sobre una gestión agrícola óptima. Para 2050, la agricultura tendrá que producir un 60% más de alimentos a nivel mundial y un 100% más en los países en vías de desarrollo (FAO, 2009).

La industria y la energía juntas representan el 20% de la demanda de agua (NU, 2014). Los países más desarrollados tienen una proporción mucho mayor de extracciones de agua dulce para la industria que los restantes, donde predomina la agricultura. El equilibrio entre los requisitos de sostenibilidad frente a la visión convencional de la producción industrial en masa crea una serie de interrogantes para la industria. A gran escala, la globalización y la forma de extender los beneficios de la industrialización a todo el mundo equitativamente y sin impactos insostenibles sobre el agua y otros recursos naturales es la cuestión clave.

El sector doméstico representa el 10% del uso total de agua (NU, 2014). Y, en todo el mundo, se estima que 748 millones de personas siguen sin tener acceso a una fuente mejorada de agua y que 2.500 millones siguen sin acceso a unos servicios de saneamiento mejorados (NU, 2014).

## 1.4 Recursos hídricos

Lo primero que hay que precisar antes de entrar en la definición de recursos hídricos es conocer su origen etimológico. Recursos es una palabra cuya raíz es el latín "*recursus*", que viene a hacer referencia, de hacer uso de los medios o bienes de los que dispone alguien para acometer algo en concreto (DiccionarioActual, s.f.). Es decir, un recurso es una materia prima o un bien del que se dispone y se le da una utilidad en pos de un objetivo. Ese objetivo por lo general, se trata de hacer algo que satisface una necesidad o que permite la subsistencia. La palabra hídrico, por su parte, es aquello que está vinculado al agua.

Los recursos hídricos son los cuerpos de agua que existen en el planeta, desde los océanos hasta los ríos pasando por los lagos, los arroyos y las lagunas. Estos recursos deben preservarse y utilizarse de forma racional, ya que son indispensables para la existencia de la vida. El problema es que, aunque en su mayoría son recursos renovables, la sobreexplotación y la contaminación que provocan diversas actividades humanas hacen que los recursos hídricos estén en riesgo. Su capacidad de regeneración muchas veces no resulta suficiente ante el ritmo de uso.

Una de las grandes dificultades que enfrenta la humanidad es la falta de agua dulce. Más del 97% del agua de la Tierra es agua salada (Funcagua, 2020), cuyo aprovechamiento es complejo. Por eso el agua dulce, que se

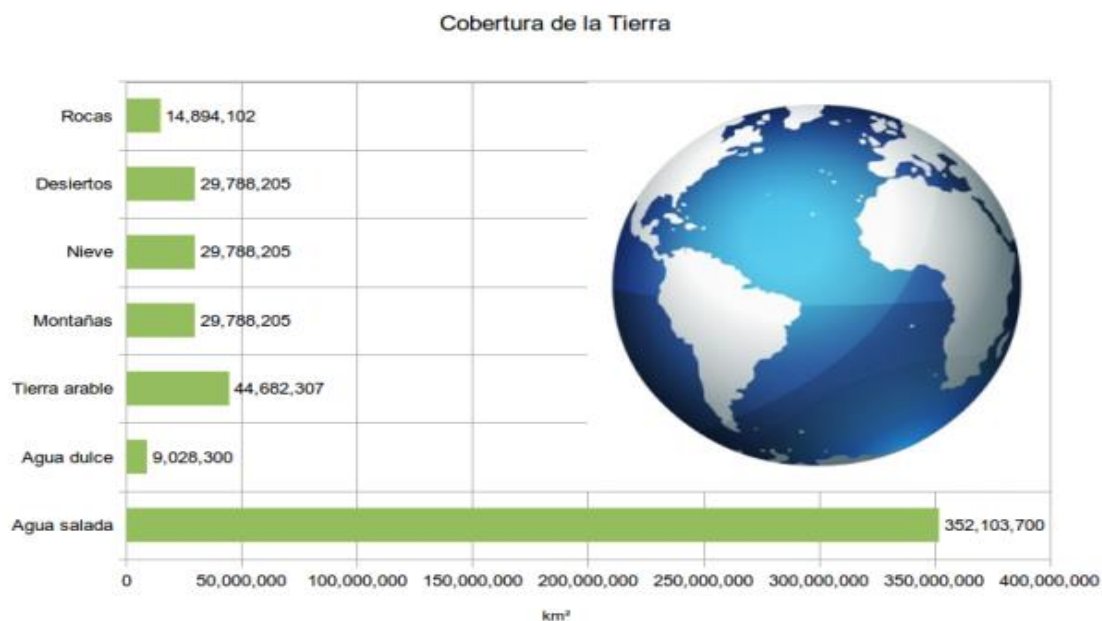


utiliza para el consumo humano y un sinnfín de actividades, es tan importante.

En concreto, las estimaciones llevadas a cabo vienen a establecer que el 100% del agua total del planeta se distribuye de la siguiente manera: 97,47% de agua salina, 2,53% de agua dulce, 1,76% de glaciares y capas polares, 0,76% de agua subterránea y 0,01% de lagos, ríos y atmósfera, como se muestra en la Figura 2.

## Figura 2.

*Distribución del agua a nivel mundial*



*Nota.* Adaptado de (De León, 2020)

Uno de los principales problemas que enfrenta el mundo en un futuro cercano es la disminución del suministro de agua hasta en un 40% para el año 2030 (NU, 2018), por lo que es importante mejorar considerablemente la

gestión de este recurso. Según datos del Ministerio del Ambiente, en la vertiente amazónica reside el 26% de la población, que cuenta con el 97,7% de agua, mientras que en la vertiente del Pacífico reside el 70 % de la población, que cuenta tan solo con el 1,8% de agua (Palermo, 2015).

En una nota importante para destacar en este punto es que los recursos y servicios relacionados con el agua son esenciales para el logro de la sostenibilidad global, puesto que estos ayudan al crecimiento económico, la reducción de la pobreza y la sostenibilidad ambiental.

### **1.5 Políticas de manejo del agua potable**

El agua potable es agua que es segura para beber y cocinar. Es importante tener acceso a agua potable para mantener una buena salud y prevenir enfermedades. Su acceso es un derecho humano fundamental. La falta de acceso puede tener consecuencias graves para la salud, la nutrición, la educación y otros aspectos de la vida (UNICEF, s.f.).

La Organización Mundial de la Salud (OMS) establece que el agua potable debe cumplir con ciertos estándares de calidad para ser segura para el consumo humano. Estos estándares incluyen niveles seguros de bacterias, virus, químicos y otros contaminantes (Aguasdevalencia, 2023).

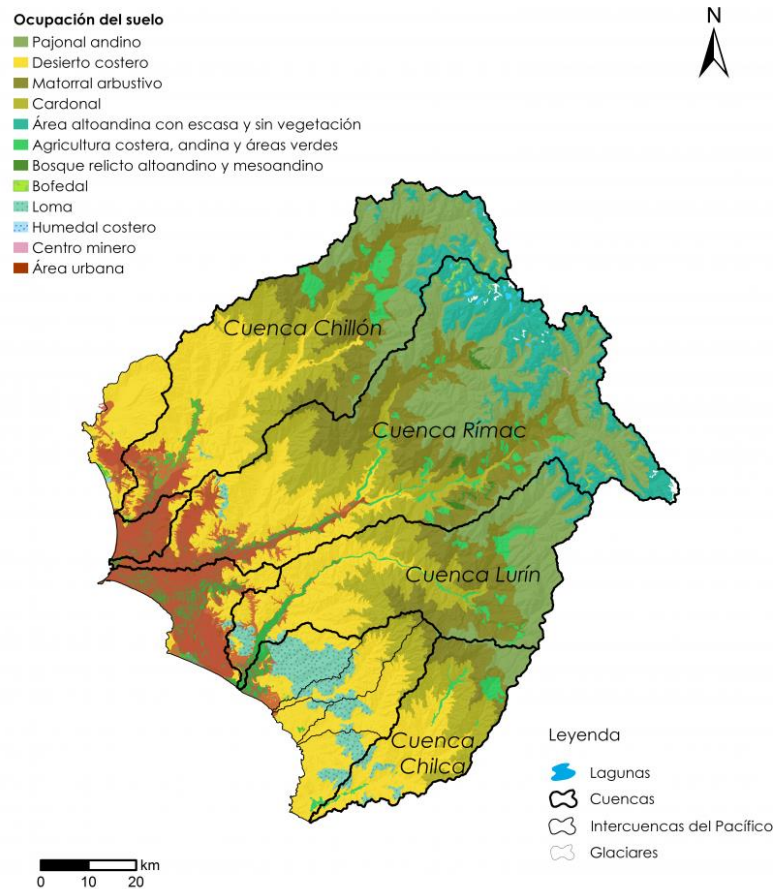
Es importante destacar que el desarrollo sostenible y el agua están estrechamente relacionados. La gestión sostenible del agua ayuda a

garantizar que las personas no se vean desfavorecidas de ninguna manera y que se mantenga la salud de la sociedad. Además, el uso reducido de energía, un hábitat saludable para todos y el impacto social son algunos de los beneficios del desarrollo sostenible.

Lima es la segunda ciudad en el mundo afincada en un desierto (como muestra la Figura 3) después del Cairo en Egipto, por lo que los recursos hídricos que posee son insuficientes y lo serán más en los próximos años. Es necesario por ello hacer un uso adecuado de las fuentes de agua.

**Figura 3.**

*Lima sobre el desierto Sechura*



*Nota.* Adaptado de (ANA, s.f.)

El Decreto Supremo No. 006-2015-MINAGRI (MINAGRI, 2015) y 013-2015-MINAGRI (MINAGRI, Decreto Supremo No. 013-2015-MINAGRI, 2015) aprobaron la Política y Estrategia Nacional de Recurso Hídricos y el Plan Nacional de Recursos Hídricos, respectivamente.

En el Perú se cuenta con una empresa para todo lo que tiene que ver con servicio de agua potable y alcantarillado para la colección de aguas servidas, se le conoce como SEDAPAL, es una empresa estatal peruana creada en 1981 la cual atiende a más de 9 millones de habitantes solo en la ciudad de Lima.

Entre los objetivos que persigue la SEDAPAL están: Prestar servicios de agua potable y alcantarillado sanitario, la ejecución de política del sector en la operación, mantenimiento, control y desarrollo de los servicios básicos, con funciones específicas en aspectos de normatividad, planeamiento, programación entre otras actividades conexas inherentes a su naturaleza de servicio.

Este capítulo se ha dedicado a tratar dos temas por demás importantes como lo es desarrollo sostenible y por otro lado los recursos hídricos y el agua. En este sentido, es importante destacar que estos están estrechamente relacionados entre sí y con la calidad de vida de los seres humanos. Es decir, la gestión sostenible del agua ayuda a garantizar que las personas no se vean desfavorecidas de ninguna manera por las carencias

del vital líquido y que se mantenga la salud de la sociedad por el consumo de agua realmente potable.

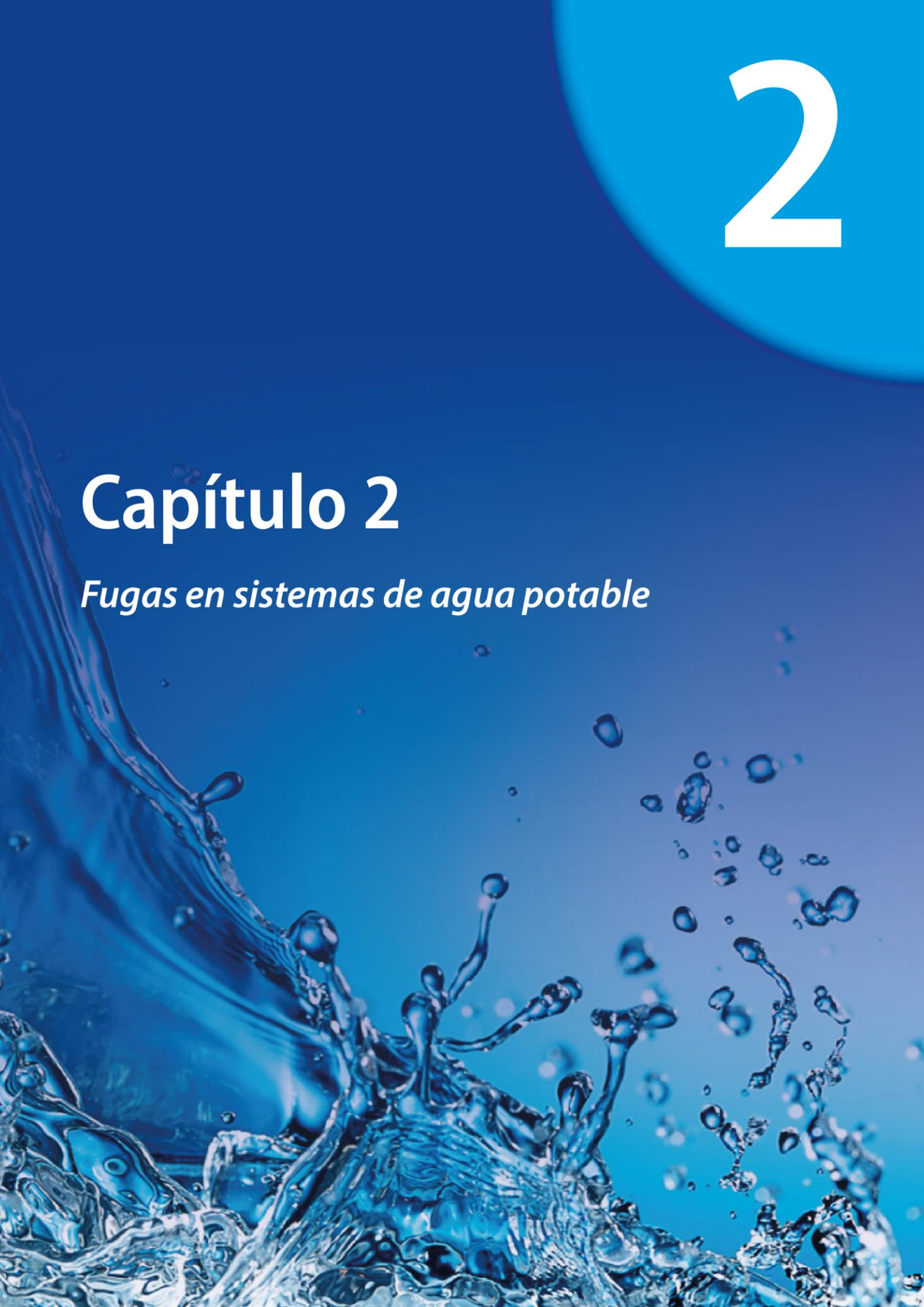
Si se enumeran las bondades del desarrollo sostenible, se tendrá en primera condición el uso reducido de energía, alcanzar la garantía de un hábitat saludable para todos y el impacto social igualitario son algunos de los beneficios del desarrollo sostenible.

En las ciudades enclavadas en el desierto, como Lima, el agua es un recurso escaso y valioso. La gestión adecuada del agua puede ayudar a garantizar que las personas no se vean desfavorecidas y que mantengan la salud de la sociedad. Además, el uso reducido de energía, un hábitat saludable para todos y el impacto social son algunos de los beneficios del desarrollo sostenible para este tipo de ciudades.

# 2

## Capítulo 2

*Fugas en sistemas de agua potable*



## Fugas en sistemas de agua potable

### 2.1 Introducción al capítulo

La detección de averías que producen posteriormente fugas de agua, en las redes de agua potable urbanas, es un proceso importante para garantizar el uso sostenible del agua y reducir las pérdidas tanto financieras como del vital líquido así lo manifiestan (Fuentes-Mariles, Palma-Nava, & Rodríguez-Vázquez, 2011). Las fugas de agua en las redes de distribución pueden producir desperdicio, reducir la eficiencia de las redes y generar una pérdida económica a los organismos operadores del sistema de distribución (Hidrotec, 2016).

El proceso para llegar a la detección de fugas en una red de tuberías es complicado, por múltiples situaciones, aunado al hecho de que en su mayoría las redes no se encuentran visibles (normativamente son

enterradas). Para reducir las fugas es necesario contar con procedimientos e instrumentos especiales para localizarlas y eliminarlas (Fuentes-Mariles, Palma-Nava, & Rodríguez-Vázquez, 2011). En la actualidad, existen diferentes tecnologías para detectar fugas en las redes de distribución, como sensores acústicos, correladores, geófonos, entre otros (Hidrotec, 2016).

## **2.2 Servicio de detección de fugas no visibles en redes primarias (ECRF)**

La finalidad pública del presente servicio está relacionada con el objetivo de primer nivel “Disminuir el agua no facturada” y el objetivo de segundo nivel “Reducir los volúmenes de pérdidas de agua potable”; lo que permite reducir las pérdidas de agua y brindar un mejor servicio. En la ciudad de Lima este servicio lo ejecuta el Equipo Control y Reducción de Fugas (ECRF) de la Gerencia de Producción y Distribución Primaria de SEDAPAL. Para la realización de estas labores se cuentan con 5 unidades móviles las cuales se desplazan por la ciudad de Lima, realizando labores diversas como:

- Inspección de Redes Primarias de Agua Potable según programación establecida con anticipación.
- Detección de fugas imprevistas en calles y avenidas.
- Detección de conexiones clandestinas de agua.
- Monitoreo de la presión de agua (psi) por cuadrantes y/o zonas.



Para la realización de las actividades diarias, las unidades cuentan con instrumental especializado tanto digital como analógico. Este equipo se utiliza para las labores de detección de fugas de agua, obstrucción de tuberías, tuberías clandestinas y así como para controlar la presión de agua en la red de tuberías.

El personal de cada unidad ha recibido entrenamiento especial para el manejo y la operación de estos equipos (por lo general para el servicio se cuenta con personal con una experiencia no menor a 5 años dado el tipo de labor que se realiza). Cada unidad móvil cuenta con 5 personas (Técnico, Operario especializado, chofer y 2 operarios).

### **2.3 Equipos utilizados por el ECRF**

En la actualidad, para realizar su labor el ECRF cuenta con equipos electrónicos portátiles de última generación los cuales les permiten realizar su labor en forma eficiente.

Evidentemente que es de suma importancia contar con las herramientas y equipos necesarios para acometer la difícil tarea de ubicar la avería en la red, por los múltiples factores que influyen en esta condición, la idea pasa por una premisa de mantenimiento que reducir los tiempos de reparación y prolongar los tiempos entre las averías, es decir disminuir los tiempos que toman realizar las reparaciones y alargar los tiempos de ocurrencias de las fallas, con planes de mantenimientos centrados en la

prevención, la idea central es reparar las tuberías en cuestión de horas, a partir de la localización exacta del tramo de tubería dañada. Para lograr tomar datos para esta empresa se requiere de equipos y se cuenta con:

a) Manómetro

El manómetro (del griego *μανός*, ligero y *μέτρον*, medida) es un instrumento de medición diseñado para obtener la presión de fluidos contenidos en recipientes cerrados. Se distinguen dos tipos de manómetros, según el modo en que se empleen para medir la presión del fluido que recordemos pueden ser líquidos o de gases. La Figura 4 muestra un manómetro de fluidos, para nuestro caso.

**Figura 4**

*Manómetro de glicerina para fluidos presurizados.*



*Nota.* Adaptado de Rubiños & Apesteguia (2020)

El instrumento es sumergido en glicerina como se muestra en la Figura 6, lo cual no altera su funcionamiento, ya que realiza toma de la presión del mismo modo que un manómetro común; sin embargo, la adición de la glicerina permite proteger la funcionabilidad del equipo, es un estabilizador de la sensibilidad del equipo cuando funciona en elementos de tubería o maquinarias sometidas a ciclos de carga que producen vibraciones, además la glicerina lubrica las partes internas del mismo; es por ello que se recomienda para esta actividad en específico, el uso de este modelo manométrico lo cual reduce la posibilidad de lecturas erróneas por falta de lubricación o por vibraciones externas.

En la Tabla 1, se muestran las unidades de medida que más se utilizan en las lecturas de manómetros para fluidos, las unidades según el sistema de medición son PSI, Mpa y Bar.

**Tabla 1**

*Unidades de presión utilizadas por los manómetros.*

Unidades	Definición
PSI	Libras por pulgada cuadrada
Mpa	Megapascal
Bar	Bar
Conversiones	
Bar = 1 Atmósfera = 1000000 de barias	
1 Bar	= 0.1 Megapascal = 14.5038 psi

*Nota.* Adaptado de Rubiños & Apestegua (2020)

Un dato adicional con respecto a los manómetros es el siguiente:

se rellenan por lo general también con glicerina para trabajar a baja temperatura, incluso con una base de silicona, cualesquiera de las dos opciones brindarán la protección adecuada al equipo. Este modelo manométrico lo hace ideal para fluidos de cualquier sistema hidráulico que posean líquidos de trabajo como el aire, agua, aceite, entre otros. En maquinarias térmicas tipo turbinas, motores, etc. Según Rubiños & Apesteguía (2020):

Existen diferentes aplicaciones y modelos de manómetros; los utilizados en ECRF son del tipo de metal con relleno de glicerina y se utilizan para las labores de toma de presión en los medidores domésticos. La presión para la ciudad de Lima oscila entre 20 y 65 psi. Es decir, el rango mínimo de presión que se necesita para poder realizar trabajos de detección de fugas es de 10 psi, debajo de ese valor no se puede realizar trabajos de detección de fugas.

Para satisfacción del proceso de simulaciones algorítmicas ejecutadas en la ciudad de Lima, se cumple con los rangos para poder realizar la detección de fugas; a continuación, se presenta al lector la hoja de datos técnicos del manómetro específico utilizado por el ECRF y se muestra en la Figura 5.

## Figura 5

Manómetro de glicerina especificaciones técnicas.

### Manómetro con muelle tubular Modelo 213.53, Líquido de relleno, Caja acero inoxidable

Hoja técnica WIKA PM 02.12

 otras homologaciones véase página 2

#### Aplicaciones

- Para puntos de medida con elevadas cargas dinámicas y vibraciones
- Para medios gaseosos, líquidos, no viscosos y no cristalizantes, compatibles con aleaciones de cobre
- Hidráulica
- Compresores, industria naval

#### Características

- Resistente contra vibraciones y golpes
- Construcción de extrema robustez
- DN 63 y 100 con homologación Germanischer Lloyd y Gost
- Rangos de indicación hasta 0 ... 1.000 bar



**Manómetro con muelle tubular, modelo 213.53.100, conexión inferior**

#### Descripción

<b>Versión</b> EN 837-1	<b>Carga de presión máxima</b> DN 50, 63: Carga estática: 3/4 x valor final de escala Carga dinámica: 2/3 x valor final de escala Puntual: valor final de escala DN 100: Carga estática: valor final de escala Carga dinámica: 0,9 x valor final de escala Puntual: 1,3 x valor final de escala
<b>Diámetro en mm</b> 50, 63, 100	<b>Temperatura admisible</b> Ambiente: -20 ... +60 °C Medio: máx. +60 °C
<b>Clase de exactitud</b> DN 50, 63: 1,6 DN100: 1,0	<b>Influencia de temperatura</b> En caso de desviación de la temperatura de referencia en el sistema de medición (+20 °C): máx. ±0,4 %/10 K de la gama de indicación
<b>Rangos de indicación</b> DN 50: 0 ... 1 a 0 ... 400 bar DN 63, 100: 0 ... 0,6 a 0 ... 1.000 bar así como todas las gamas correspondientes para presión negativa y sobrepresión negativa y positiva	<b>Tipo de protección</b> IP 65 según EN 60529 / IEC 60529

Nota. Adaptado de Rubiños & Apesteguia (2020)

### b) Geófono

Dentro de una gamma más sofisticada para la detección de averías en tuberías subterráneas tenemos un instrumento de medición, como lo es el

geófono, es un equipo no invasivo que se utiliza para lograr la detección a distancia de la falla en la tubería averiada, es decir sin realizar excavaciones el detector indica donde está la pérdidas causadas por la fugas de agua y su principio de funcionamiento es similar al de un sismógrafo, localiza con alta precisión las fugas subterráneas mediante la captación de una señal de pulso ultrasónico. Las ventajas del uso de este ingenio de la detección de fallas en tuberías es que reduce a su mínima expresión los tiempos de localización y los costos asociados a excavaciones de aproximación y descarte, debido a que es muy preciso en sus detecciones. Comercialmente existen dos tipos de geófonos: los mecánicos y los electrónicos.

- El geófono mecánico

El principio de funcionamiento del geófono mecánico es netamente acústico y por eso su principio copia principalmente a un estetoscopio; en este equipo se logra escuchar con claridad el sonido que genera el agua al salir por la grieta y dicho sonido se propaga en forma de vibración hasta la superficie y es captado por este estetoscopio el cual amplifica y se las transmite a través de sus auriculares al operador, el cual va realizando el registro de la zona hasta el punto focal de partida del sonido. Es un equipo muy fiable el cual se muestra en la Figura 6.

## Figura 6

*Geófono de piso mecánico.*



*Nota.* Adaptado de Rubiños & Apestequia (2020)

- Geófono Electrónico o Electro acústico

Otro modelo de geófono que es de interesante manejo y de mayor precisión que los de principio mecánico, son los geófonos electroacústicas o electrónicos, cumple la misma función que el de principio mecánico, transforma la vibración en una señal detectable para la escucha de las fugas en las tuberías, comercialmente se pueden encontrar de 2 tipos: Lmic y Xmic.

El modelo de geófono electro acústico LMIC® está provisto de un sensor de amplificación, incrementando la capacidad del oído humano para detectar el ruido producido por una fuga subterránea, adicionalmente agregan un control de volumen adicional en los auriculares diferenciado del control de volumen general del propio sensor, con otros adicamentos piezoeléctricos que permiten realizar mediciones en varios tipos de suelo y condiciones de dureza y humedad.

**Figura 7**

*Geófono electrónico LMIC.*



*Nota.* Adaptado de Rubiños & Apestequia (2020)

El modelo Xmic es un geófono, superior en prestaciones al Lmic, el mismo incorpora mejoras en la amplificación de la señal, lo cual mejora la calidad de lo que escucha el operador, al igual que una mayor cantidad de funciones que colaboran con el operador para lograr una ubicación muy precisa de la avería en la tubería. En las figuras a continuación se muestra un geófono de estas características en operación.

**Figura 8**

*Geófono electrónico XMIC.*



*Nota.* Cortesía PALMER INC



## Figura 9

*Correcto uso de los Geófonos electrónicos XMIC y Lmic.*



*Nota.* Cortesía PALMER INC

- Correlador

En la detección de fugas de agua hay equipos de mayores dimensiones que requieren contacto directo con la tubería para dar la ubicación de la grieta o de la pérdida de presión que implique una fuga; estos equipos son los Correladores, dispositivos que se aprovecha de la dinámica vibracional que genera el flujo de fluidos en tuberías o embases a presión, lo cual genera patrones de ruido que son detectados por estos dispositivos y por lo general son anulares que trabajan como un mínimo en pares, aun cuando se pueden disponer mayores cantidades de los mismos a lo largo de la red de distribución.

Sus características principales son: el uso para detectar fugas en largos tramos de tubería no importando el diámetro de la misma y puede adaptarse a los patrones de ruido de diferentes materiales, es un equipo que su medición se realiza valorando el tiempo de tránsito de las ondas de presión y las correlaciona con patrones geométricos propios de la tubería con la propagación del sonido, y el diferencial de ruido en el tiempo como se muestra en la ecuación (1):

$$L = \frac{1}{2} (D - V \times Td) \quad (1)$$

Donde:

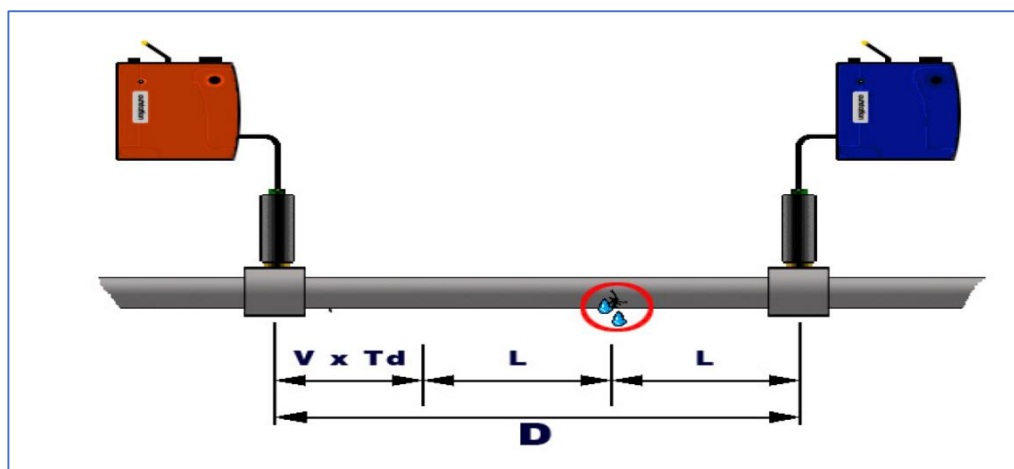
V: Velocidad del sonido dentro de la tubería (m/s)

Td: Diferencia de la velocidad del sonido entre uno y otro sensor

Su mayor desventaja es que requiere estar en contacto con la tubería en estudio (Figuras 10 y 11).

### Figura 10

*Funcionamiento Correlador Palmer.*



Nota. Cortesía PALMER INC

**Figura 11.**

*Funcionamiento Correlador Palmer.*

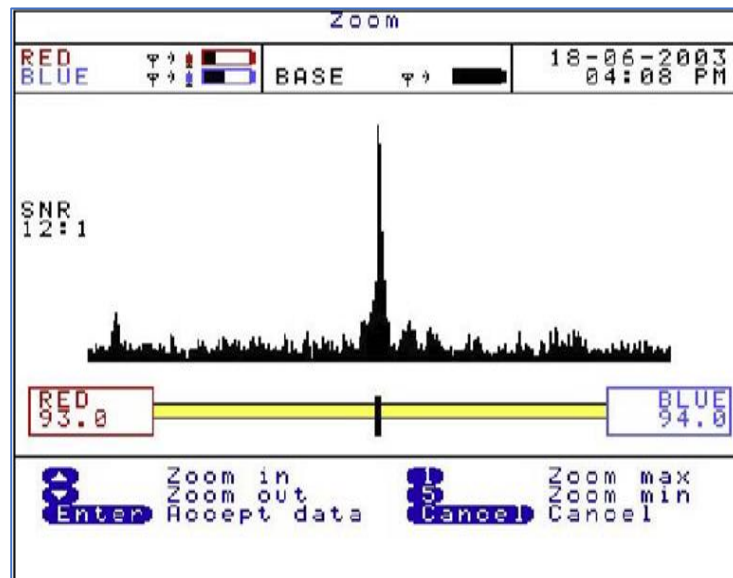


*Nota.* Cortesía PALMER INC

En la Figura 11 se muestran (1) Correlador Palmer; (2,3,4) Radios Transmisoras de Señal; Audífonos Estéreo HI-FI y Sensores de Audio (6 y 7) Se utiliza para detectar fugas de agua No visibles haciendo uso de sus 2 o 3 sensores, instalados intercaladamente en las conexiones domiciliarias. Se conectan a la unidad central vía ondas de radio (FM) y esta da la posición aproximada de la tubería rota, como se muestra en la Figura 12.

**Figura 12.**

*Correlador Palmer en funcionamiento.*



*Nota.* Cortesía PALMER INC

## 2.4 Consideraciones finales

En este apartado se ha mantenido la idea de la supervivencia de la especie y se sabe que el agua es fundamental para la misma y por estas razones, es importante detectar las fugas de agua potable en las tuberías principales o matrices de las ciudades lo antes posible y repararlas a la brevedad para evitar lo desperdicios de la misma. Para ello, se pueden utilizar diferentes métodos y tecnologías, como la inspección visual, la medición de caudales y presiones, la detección acústica, la termografía, el gas trazador, la correlación cruzada y otros. Estos métodos permiten localizar las fugas con mayor precisión y eficiencia, reduciendo el tiempo y el costo de las intervenciones.

# 3

## Capítulo 3

*Aplicaciones de la Inteligencia Artificial  
a la detección de fallas en tuberías*



## **Aplicaciones de la Inteligencia Artificial a la detección de fallas en tuberías**

### **3.1 Introducción al capítulo**

Las fugas en las tuberías de agua potable son las salidas no controladas en cualquiera de los componentes del sistema, se producen en las uniones, elementos como válvulas, ventosas, té, collarines, hidrantes, etc., y la menos deseada de las fugas es la que se produce como consecuencia de las roturas de las tuberías generales y acometidas que pueden ser circunferenciales, poros, longitudinales o reventones (Iagua, 2019). Las causas de las fugas pueden atender a agentes externos a la tubería, también influyen las características intrínsecas del material, condiciones de operación de la red, circunstancias de la instalación y otras causas.

Fundamentalmente son los agentes externos los que provocan fugas que pueden atender a la presencia de raíces de árboles, con obras cercanas a la red que provocan averías, temperatura y condiciones climáticas extremas,

corrientes parásitas externas, sobrecargas externas como el tráfico pesado y otras como roedores, incendios y terremotos entre los elementos de índole natural.

Desde el punto de vista de las características propias del material son motivo de cuidado ya que se suele producir por fallas propias del material, que permiten la aparición de averías. Situaciones que se pueden incluir que también afectan a las tuberías y accesorios de las redes debido a su composición, como el caso de la corrosión, fallas en tornillos, defectos en la manufactura del material, el tiempo de operación, entre tantos otros motivos como problemas en la fabricación en el material o revestimientos internos y externos, disminución de espesores, falta de un adecuado revestimiento, deformación por sobre presión, reacciones químicas debido al tipo de agua y terreno (Iagua, 2019).

La operación de la red también puede provocar roturas y fugas, como las elevadas presiones continuas o momentáneas debidas a golpes de ariete por cierres rápidos, problemas con el aire no evacuado o que no entra en los vaciados, vibraciones o fuerzas continuadas que producen el colapso por agotamiento de los materiales.

Finalmente, las condiciones de instalación provocan roturas y fugas por asentamientos del terreno cercano a arquetas, asentamientos del terreno en las zanjas, no protección a corrosión externa e interna, mala ejecución de la zanja para distribuir cargas externas e internas, no adecuación al tipo,

velocidad y presión del agua a transportar y otros defectos en la instalación como mala ejecución del montaje, malos materiales de unión o no previsión de contracciones y dilataciones por cambios térmicos.

Ante la cantidad de posibilidades por las cuales puede presentarse una fuga, se hace necesario detectar las fugas de agua potable en las tuberías principales o matrices de las ciudades considerando estos aspectos:

- Preservar el recurso hídrico: las fugas de agua potable implican un desperdicio de un recurso vital y escaso, que podría ser utilizado para otros fines o para abastecer a más población.
- Reducir el costo económico: las fugas de agua potable generan pérdidas económicas tanto para las empresas prestadoras del servicio como para los usuarios.
- Mejorar la calidad del servicio: las fugas de agua potable afectan la presión y el caudal del agua en la red, lo que puede provocar problemas de suministro, interrupciones, contaminación y deterioro de la infraestructura.
- Prevenir daños ambientales y sociales: las fugas de agua potable pueden causar inundaciones, erosión del suelo, daños a las propiedades, riesgos sanitarios y molestias a los habitantes de las zonas afectadas.



### 3.2 Fugas programadas y de emergencia

El servicio de detección de fugas no visibles de fugas programadas, como su nombre lo indica, es el que se encarga de realizar una programación previamente establecida de cuadrantes y sectores dentro de la ciudad. En el caso de Lima y Callao, esta programación se hace un recorrido o barrido de calles tratando de detectar fugas no visibles para su posterior notificación y/o reparación.

Una vez culminada la programación se repite en forma de ciclos variando el orden de los distritos cada vez, por lo general se repite los ciclos cada 2 años aproximadamente. Los 50 distritos de la ciudad de Lima que se distribuyen son parte integrante de la Provincia de Lima (43 distritos) sumada a la Provincia Constitucional del Callao (7 distritos).

El servicio de detección de fugas de emergencia (es decir, no visibles) como su nombre lo indica, son reportadas por las mismas zonales de SEDAPAL (Oficinas de Centro de Servicio y Cobro ubicadas estratégicamente en varios distritos de la capital). Esta información es recolectada por diversos medios: llamadas telefónicas, inspecciones realizadas por el organismo, denuncias presenciales, rotura de pistas por paso de vehículos pesados, etc.

Las fugas emergencia son atendidas por las unidades móviles de acuerdo con su importancia y/o urgencia y se determina si efectivamente

existe una fuga de agua. Posteriormente, se procede a hacer su informe respectivo catalogando y clasificando la fuga y su informe es ingresado a la base de datos del sistema para su posterior reparación y/o archivamiento en caso la fuga no haya sido detectada.

### **3.3 Filtrado y discriminación de datos**

No todos los campos son necesarios para hacer el diseño del algoritmo, por lo tanto, se eliminarán aquellos que no sean relevantes para el estudio, tales como nombre dirección, código interno y códigos de fecha duplicados, entre otros. Se detallarán los campos que se han conservado para su posterior análisis y estudio.

#### **a) Número de fugas**

La cantidad de fugas registradas en el sistema en estos años incluyen las fugas detectadas y las que no eran fugas también para el archivo.

#### **b) Número de Identificación del Suministro (NIS)**

Número de Identificación del Suministro (NIS) es un código de nueve dígitos. De estos, siete son invariables y identifican el suministro y dos variables que identifican el contrato vigente. Este se encuentra en todos los recibos emitidos por la empresa prestadora del servicio y es único, como se muestra en la Figura 13.

**Figura 13**

Detalle de un recibo de facturación de agua y el número de NIS.

**sedapal**  
SERVICIO DE AGUA POTABLE Y ALCANTARILLADO DE LIMA  
Av. Ramiro Priate 210 El Agustino  
RUC : 20100152356  
www.sedapal.com.pe

**OFICINA COMERCIAL: AV. TINGO MARIA 600**

**Para Consultas Suministro N°**  
**NIS: 3196104-8**

RECIBO N°  
06388472-13111200906  
Referencia de Cobro  
31961041279

0239716  
9138-311-0001-0001-19M2

Frecuencia de Facturación	Mes Facturado	Emisión	Categoría	Vencimiento	Tarifa
Mensual	Junio 2009	03/06/2009	RESIDENCIAL	20/06/2009	DOMESTICO

Titular de la Conexión: \_\_\_\_\_ Dirección del Suministro: \_\_\_\_\_

Distrito	Actividad	Unid. uso	Período de Consumo	Tipo de Facturación
LIMA (CERCADO)	PREDIO UNIFAMILIAR	1	02/05/2009 - 02/06/2009	LECTURA

Información de medidores			
Medidor	Lectura Anterior	Lectura Actual	Consumo M3
E206150052	903	937	34

Detalle de Facturación					
Concepto					Importe
Consumo de agua	34.00	m3			62.75
Cargo Fijo					4.44
I.G.V.	67.19	x	19%	12.77	
Mora					1.65
Redondeo del mes anterior					0.07
Redondeo del mes actual					-0.18
Consumo de agua					81.50
<b>IMPORTE TOTAL</b>					<b>S/.******81.50</b>

**Información Complementaria**

Estructura Tarifaria (01/11/2008)

Tarifa	Rango	(S/.)	m3	(S/.)
DOMESTICO	20 a 30	1.735	30.00	52.05
	30 a 50	2.675	4.00	10.70
			34.00	62.75

Horario de abastecimiento

Código : CER002 00

Frecuencia : DIARIO

De : 00 : 00 hrs.

Hasta : 24 : 00 hrs.

Diámetro Conex.: \_\_\_\_\_

x Enotria S.A.RUC 20100117526

**A partir del vencimiento del presente recibo se efectuara el cierre del servicio.**

Nota. Adaptado de Rubiños & Apesteagua (2020)

**c) Fecha**

Fecha de la detección de la fuga.

**d) Distrito**

Donde se realizó la auscultación de la fuga o fugas, como se muestra en la Figura 14.

## Figura 14.

*Detalle de los 50 distritos de la ciudad de Lima y Callao.*



*Nota.* Adaptado de Rubiños & Apesteguía (2020)

### e) Urbanización

Donde se realizó la auscultación de la fuga o fugas, como se muestra en la Figura 15.

**Figura 15**

*Detalle de un NIS ubicado en el plano geo referenciado.*



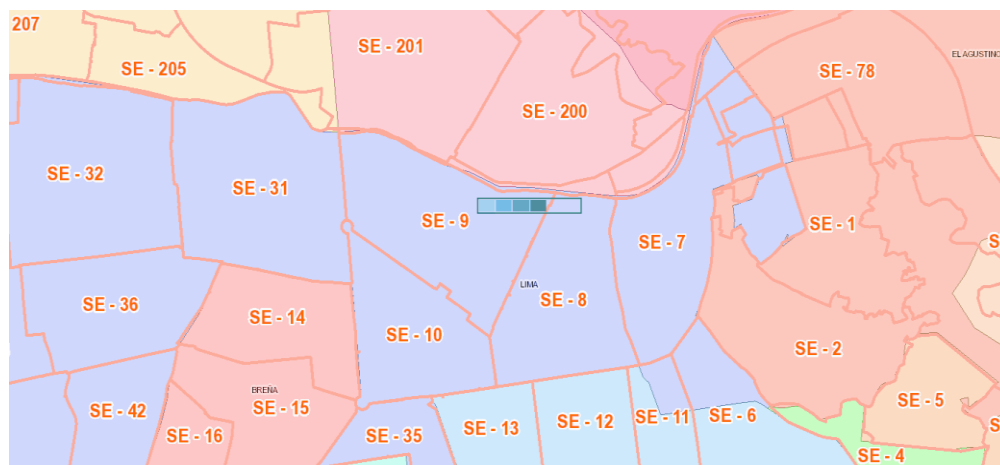
*Nota.* Adaptado de Rubiños & Apestequia (2020)

## f) Código de sector

Como ya se indicó anteriormente, los 50 distritos de Lima están subdivididos a su vez en sectores para que sea más rápida la ubicación de un suministro o NIS para ello se cuenta con una codificación propia, como se muestra en la Figura 16.

## Figura 16

Detalle del distrito de Lima y sus sectores.



Nota. Adaptado de Rubiños & Apestequia (2020)

El proyecto de sectorización de Lima y Callao tuvo en cuenta los siguientes parámetros:

- El área debe ser menor a 3 km<sup>2</sup>
- Las presiones deben estar entre 15 PSI y 50 PSI.
- Utilizar de preferencia las avenidas como límite de sector, en especial aquellas donde el tendido de tubería es por ambos lados de la calle.
- Definir anillos o circuitos conformados por tuberías de gran capacidad.
- Se debe evitar en lo posible dejar puntos muertos en la red, considerando redes secundarias complementarias que los anule.
- Las tuberías mayores e iguales a 6" (0,1524 m) de diámetro que crucen el límite de un sector deben cerrarse por medio de una válvula, mientras que las tuberías menores o iguales a 4" (0,1016 m)

deberán ser cortadas o taponadas. Asimismo, se dejará disponible por lo menos un pase de emergencia.

- Los sectores, de ser posible, respetarán los límites de separación de zonas de presión.

En el caso por ejemplo del distrito de Lima Cercado, este cuenta con diez (10) sectores enumerados SE-1 hasta SE-10.

### **3.4 Objetivos de la sectorización**

Los objetivos que persigue la sectorización son:

- Permitir controlar, en un área definida, parámetros importantes para el buen funcionamiento del Sistema de Distribución de Agua Potable. Estos parámetros son: caudal de ingreso al sector y presión en la red (que debe ser entre de 15 a 50 psi).
- Permitir la aplicación de una justa política de racionamiento de agua, en épocas de escasez, mediante la correcta utilización de fuentes superficiales y subterráneas, en lo que se denomina uso conjuntivo.
- Determinar la cantidad de agua no facturada, obtenida como la diferencia del volumen de agua que ingresa al sector y el volumen facturado, obtenido a través de la micro medición.
- Permitir el aislamiento de un sector con respecto al resto del sistema a fin de realizar trabajos de mantenimiento y reparación por problemas de emergencia en una zona definida de la red de agua. Con ello se

reducirá las molestias a los usuarios por falta de agua, pasando una gran área del Sistema de Distribución afectada hacia un pequeño sector en el futuro.

### **3.5 Tipo de fuga**

#### **a) La fuga en la caja**

Se refiere a las fugas de agua que se suscitan en el mismo buzón de la conexión domiciliaria (medidor de agua) y las llaves de paso, llamadas llaves telescópicas, son las más comunes y las más fáciles de detectar.

#### **b) La fuga en línea**

Se refiere a la tubería de PVC que normalmente es de  $\frac{1}{2}$  pulgada (0,0127 m) que sale del medidor de agua y se prolonga hasta la tubería matriz y puede tener una longitud variable que llegan hasta los 12 m, en algunos casos, y la cual se encuentra a una profundidad de hasta 2 m. Su sonido es característico, pero no son visibles a simple vista y dependerá de la presión del agua y del tipo de terreno donde se encuentra la tubería para que sea fácil su escucha.



### **c) La fuga en corporation**

Se refiere al tramo de la tubería matriz generalmente de 4 (0,1016 m) a 32 pulgadas (0,8128) donde va colocada una abrazadera que servirá de conexión entre esta y la tubería de ½ pulgada (0,0127 m) de PVC.

### **d) La fuga en tubería**

Se refiere a la tubería matriz generalmente de 4 (0,1016 m) a 32 pulgadas (0,8128) que alimenta a las conexiones domiciliarias y que puede estar construida de PVC, FOFO, Hierro Ductil, etc .

### **e) Válvula de red**

Se refiere a las válvulas colocadas en las tuberías principales para controlar y regular el paso y la presión de agua de un sector a otro.

### **f) Válvula grifo / CL**

Se refiere a los hidrantes colocados en las esquinas para apagar los incendios. Un hidrante de incendio o boca de incendio es una toma de agua diseñada para proporcionar un caudal considerable en caso de incendio. El agua puede obtenerla de la red urbana de abastecimiento o de un depósito, mediante una bomba

### 3.6 Diámetro de la tubería

Se refiere al diámetro de la tubería en donde se presenta la fuga no visible, la cual puede variar desde los 25 mm<sup>2</sup> (1/2 pulgada) hasta las 1600 mm<sup>2</sup> (32 pulgadas), como se muestran en la Tabla 3.

**Tabla 3.**

*Tuberías comerciales para agua potable y sus diámetros interior y exterior.*

Diámetro de la Tubería (in)	Diámetro Exterior (in)	Pared Mínima (in)	Diámetro Interior Promedio (in)	Peso de la Tubería (lbs/ft)	Presión Máxima de Agua a 73 °F (psi)
¼	0.540	0.119	0.288	0.110	1130
⅜	0.675	0.126	0.407	0.153	920
½	0.840	0.147	0.528	0.225	850
¾	1.050	0.154	0.724	0.305	690
1	1.315	0.179	0.935	0.450	630
1 ¼	1.660	0.191	1.256	0.621	520
1 ½	1.900	0.200	1.476	0.754	470
2	2.375	0.218	1.913	1.043	400
2 ½	2.875	0.276	2.289	1.594	420
3	3.500	0.300	2.864	2.132	370
4	4.500	0.337	3.786	3.116	320
6	6.625	0.432	5.709	5.951	280
8	8.625	0.500	7.565	9.040	250
10	10.750	0.593	9.492	13.413	230
12	12.750	0.687	11.294	18.440	230
14	14.000	0.750	12.410	22.119	220
16	16.000	0.843	14.214	28.424	220

*Nota.* Adaptado de Rubiños & Apesteguía (2020)

### **3.7 Tipo de tubería**

Se refiere al material del cual está fabricada la tubería de agua y puede influir en el tiempo de vida de esta, por el material y la marca. Se clasifican en:

- PVC
- Hierro Ductil
- Fierro Fundido
- Asbesto Cemento

### **3.8 Presión**

La presión del agua es un factor muy importante porque incidirá en la cantidad de agua desperdiciada por la fuga en un lapso de tiempo, ya que, a mayor presión de agua, mayor será el volumen de agua no facturada. Las presiones en la ciudad de Lima no son uniformes registrándose valores que van desde cero (0) hasta los 110 PSI en algunos casos.

### **3.9 Caudal**

Es la cantidad de agua no facturada que se pierde por una fuga no visible y dependerá de diversos factores, tales como la presión de agua, el diámetro de la tubería, el tipo de tubería, etc. Se calcula en función de la cantidad de litros por día perdidos por dicha fuga los cuales son por lo general aproximados.

### **3.10 Vehículo**

Se refiere a la unidad móvil encargada de realizar la detección de fugas identificada con un código único.

### **3.11 Fugas detectadas**

La cantidad de fugas detectadas por inspección pueden variar en número.

### **3.12 Observaciones**

Algunas observaciones o comentarios respecto a la fuga o fugas encontradas.

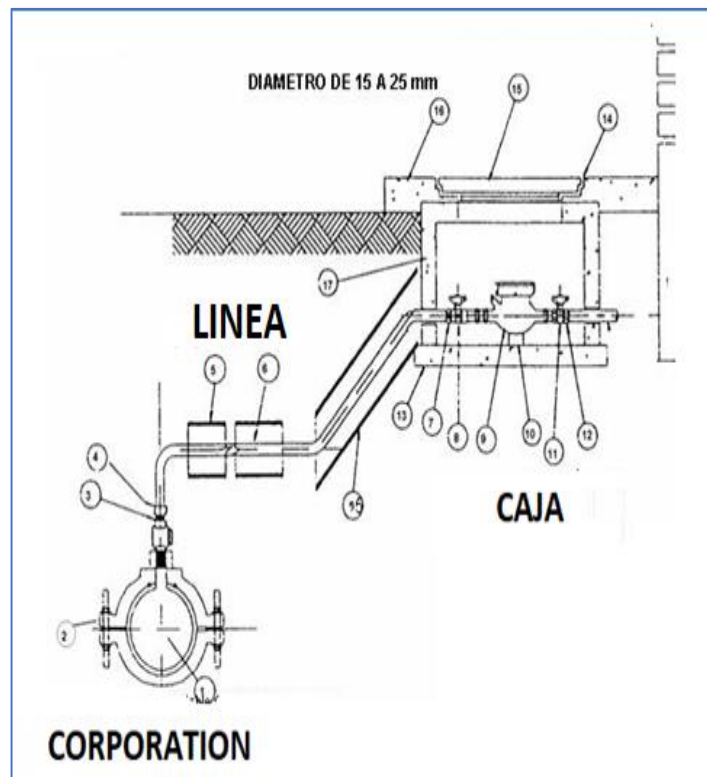
### **3.13 Modelo predictivo**

El modelo predictivo es un modelo de datos, basado en estadísticas inferenciales, que se utiliza para predecir la respuesta a un determinado evento. El modelo predictivo utiliza estadísticas para predecir los resultados. Con el propósito de desarrollar un algoritmo predictivo para la detección temprana de fugas de agua en las redes de agua potable de la ciudad de Lima, se necesita recopilar la información con la cual desarrollar el pretendido algoritmo. Los tipos de Fugas que se pueden encontrar son de los siguientes tipos y que muestra la Figura 17:

- Corporation
- Tubería
- Caja A/Med
- Caja D/Med
- línea A/Caja
- Línea D/Caja
- Válvula Red
- Válvula Grifo C/L
- Medidor

**Figura 17**

*Corte Transversal de un sistema de conexión domiciliar de agua.*



*Nota.* Adaptado de Rubiños & Apestequia (2020)

Para ello, luego de detectada la fuga, se hace un reporte con fecha y hora, el cual contiene información con respecto a la fuga encontrada, el caudal, el tipo, presión, así como la dirección completa, el distrito y el número de suministro del usuario llamado NIS, como muestra la Figura 18

**Figura 18.**

*Software de Gestión de Incidencias Operativas y Comerciales SGIOC.*

MANUAL Cerrar

## gestión de incidencias operativas & comerciales GIOC

PRODUCCION - DBPROD  
23/May/2018 04:57 am  
Versión: B.02 - 01-05-2018

Usuario:

Contraseña:

**Iniciar sesión**

[¿ Se olvido la clave? Ingresar su Correo](#)

Planos y  
Tableros GIS  
SGIOC

Equipo Tecnologías de la Información  
y Comunicaciones

*Nota.* SEDAPAL

Toda esa información llamada BASE DE DATOS DE FUGAS que contiene la información histórica, es la que se usará para el desarrollo del algoritmo, como se muestra en la Figura 19.

**Figura 19**

*Base de Datos de Fugas 2014 – 2019.*

**Inspecciones de Fugas Programadas**

Fecha Inicio: 01/12/2017 00:00 Fecha Termino: 31/12/2017 23:59

Relacion

N° Registros: 697

Tipo Fuga	Codigo	NIS	NIS Origen	Tipologia	Estado Fuga	F Registro	F Programacion	F Revision	Tiempo Aten.	Distrito
PROGRAMADA	27579	5164742	0	LINEA D/CAJA	OT FINALIZADA	29/12/2017 11:05:31	28/12/2017 00:00:00	28/12/2017 14:50:00	1	SAN MARTIN DE PORRES
PROGRAMADA	27581	5152604	0	CAJA D/MED	GENERAR OT	30/12/2017 10:16:20	29/12/2017 00:00:00	29/12/2017 10:20:00	1	SAN MARTIN DE PORRES
PROGRAMADA	27583	3732280	0	CAJA D/MED	GENERAR OT	30/12/2017 10:24:52	29/12/2017 00:00:00	29/12/2017 14:40:00	1	SAN MARTIN DE PORRES
PROGRAMADA	27586	4128811	0	CORP	OT FINALIZADA	30/12/2017 11:13:07	29/12/2017 00:00:00	29/12/2017 09:40:00	1	LA MOLINA
PROGRAMADA	27587	4086505	0	CAJA A/MED	GENERAR OT	30/12/2017 11:13:29	29/12/2017 00:00:00	29/12/2017 10:50:00	1	LA MOLINA
PROGRAMADA	27588	4281298	0	CAJA A/MED	GENERAR OT	30/12/2017 11:15:51	29/12/2017 00:00:00	29/12/2017 11:21:00	1	LA MOLINA
PROGRAMADA	27589	4118267	0	CAJA A/MED	GENERAR OT	30/12/2017 11:16:14	29/12/2017 00:00:00	29/12/2017 14:31:00	1	LA MOLINA
PROGRAMADA	27590	4092172	0	LINEA A/CAJA	OT FINALIZADA	30/12/2017 11:25:56	29/12/2017 00:00:00	29/12/2017 09:20:00	1	LA MOLINA
PROGRAMADA	27591	5820687	0	VALVULA GRIFO	REVISADO CON FUGA	30/12/2017 11:26:41	29/12/2017 00:00:00	29/12/2017 10:00:00	1	LA MOLINA
PROGRAMADA	27592	5395088	0	CAJA A/MED	GENERAR OT	30/12/2017 11:27:54	29/12/2017 00:00:00	29/12/2017 10:25:00	1	LA MOLINA
PROGRAMADA	27593	4117434	0	CAJA A/MED	GENERAR OT	30/12/2017 11:28:27	29/12/2017 00:00:00	29/12/2017 11:25:00	1	LA MOLINA
PROGRAMADA	27594	4176365	0	CAJA A/MED	GENERAR OT	30/12/2017 11:29:02	29/12/2017 00:00:00	29/12/2017 12:00:00	1	LA MOLINA
PROGRAMADA	27595	4117436	0	CAJA D/MED	GENERAR OT	30/12/2017 11:29:45	29/12/2017 00:00:00	29/12/2017 12:15:00	1	LA MOLINA
PROGRAMADA	27596	5394819	0	CORP	OT FINALIZADA	30/12/2017 11:32:07	29/12/2017 00:00:00	29/12/2017 14:10:00	1	LA MOLINA
PROGRAMADA	27597	6760533	0	CAJA D/MED	GENERAR OT	30/12/2017 11:33:03	29/12/2017 00:00:00	29/12/2017 14:30:00	1	LA MOLINA
PROGRAMADA	27598	4141504	0	CAJA D/MED	GENERAR OT	30/12/2017 11:35:51	29/12/2017 00:00:00	29/12/2017 14:45:00	1	LA MOLINA
PROGRAMADA	27599	4190091	0	CAJA D/MED	GENERAR OT	30/12/2017 11:36:25	29/12/2017 00:00:00	29/12/2017 15:00:00	1	LA MOLINA
PROGRAMADA	27600	5335742	0	LINEA A/CAJA	OT FINALIZADA	30/12/2017 11:59:34	29/12/2017 00:00:00	29/12/2017 08:51:00	1	VILLA MARIA DEL TRIUNF
PROGRAMADA	27602	2654684	0	CAJA A/MED	GENERAR OT	30/12/2017 12:20:54	29/12/2017 00:00:00	29/12/2017 09:25:00	1	VILLA MARIA DEL TRIUNF
PROGRAMADA	27603	2654684	0	CAJA D/MED	GENERAR OT	30/12/2017 12:21:26	29/12/2017 00:00:00	29/12/2017 09:25:00	1	VILLA MARIA DEL TRIUNF
PROGRAMADA	27604	2654683	0	LINEA A/CAJA	OT FINALIZADA	30/12/2017 12:22:08	29/12/2017 00:00:00	29/12/2017 09:37:00	1	VILLA MARIA DEL TRIUNF
PROGRAMADA	27605	2654207	0	CAJA A/MED	GENERAR OT	30/12/2017 12:22:35	29/12/2017 00:00:00	29/12/2017 10:25:00	1	VILLA MARIA DEL TRIUNF
PROGRAMADA	27606	2654712	0	CAJA A/MED	GENERAR OT	30/12/2017 12:23:12	29/12/2017 00:00:00	29/12/2017 11:40:00	1	VILLA MARIA DEL TRIUNF

Nota. SEDAPAL

# 4

## Capítulo 4

*Aplicación*





## Aplicación

### 4.1 Introducción al capítulo

La aplicación IA en la detección de fallas es un campo activo de investigación. Sus técnicas pueden ayudar a detectar y diagnosticar fallas en sistemas dinámicos complejos, como los sistemas eléctricos, mecánicos y químicos. Estas técnicas pueden ayudar a mejorar la eficiencia y la precisión de la detección de fallas, lo que puede reducir el tiempo de inactividad y los costos asociados con las fallas del sistema.

En resumen, la aplicación de la IA a la detección de fallas puede mejorar significativamente la eficiencia y la precisión del proceso de detección de fallas, lo que puede reducir el tiempo de inactividad y los costos asociados con las fallas del sistema.

## **4.2 Aplicación de la machine learning en la detección de fallas en tuberías de agua potable en la ciudad de Lima**

Se tienen en total 14 campos los cuales contienen la información más resaltante con respecto al histórico de las fugas detectadas. De estas, se utilizarán para realizar un análisis y obtener patrones que permitan hacer más fácil el algoritmo predictivo. Se utiliza para el desarrollo del algoritmo data histórica recopilada entre los años 2014 al 2019, y desde el punto de vista de la comprobación de la utilidad de modelo desarrollado se tomará data perteneciente al año 2020.

Como dato adicional, sólo se considerarán los caudales con valores iguales o mayores a 1000 L por ser más relevantes para esta investigación, ya que pérdidas a menores a este valor no se pueden asumir como significativas. Se encuentra que la fuga de mayor caudal fue una de 150.000 L/día, la cual estaba ubicada en el distrito de San Borja y que ocurrió el año 2017 registrada con NIS 2717633 en el sector SE-72. Además, la fuga de menor caudal fue una de 1000 L/día, la cual estaba ubicada en el distrito del Agustino en el año 2018 registrado con NIS 4200016 en el sector SE-2.

En total, 31 fugas fueron registradas en el Callao y las restantes en Lima. Se tienen en total 2858 registros de fugas registradas en ese periodo de tiempo con diferentes presiones y diferentes diámetros de tuberías, los cuales se tienen que filtrar por distrito y por caudales para realizar una mejor aproximación del modelo.

Las unidades móviles han ido aumentando a lo largo de los años para poder dar seguimiento a todas las fugas reportadas como emergencias; en el año 2014 eran sólo 3 unidades móviles, ya en 2019 son 6 unidades móviles que recorren los 50 distritos de la gran Lima. Por eso, la cantidad de fugas procesadas ha crecido en forma ascendente a lo largo de estos años ya que se cuentan con más unidades móviles para dicho fin.

Se tiene por año la cantidad de fugas atendidas:

- El año 2014 se atendieron 62 reportes de Fugas de emergencia.
- El año 2015 se atendieron 77 reportes de Fugas de emergencia.
- El año 2016 se atendieron 86 reportes de Fugas de emergencia.
- El año 2017 se atendieron 563 reportes de Fugas de emergencia.
- El año 2018 se atendieron 984 reportes de Fugas de emergencia.
- El año 2019 se atendieron 1150 reportes de Fugas de emergencia.

### **4.3 Carga de datos al software R**

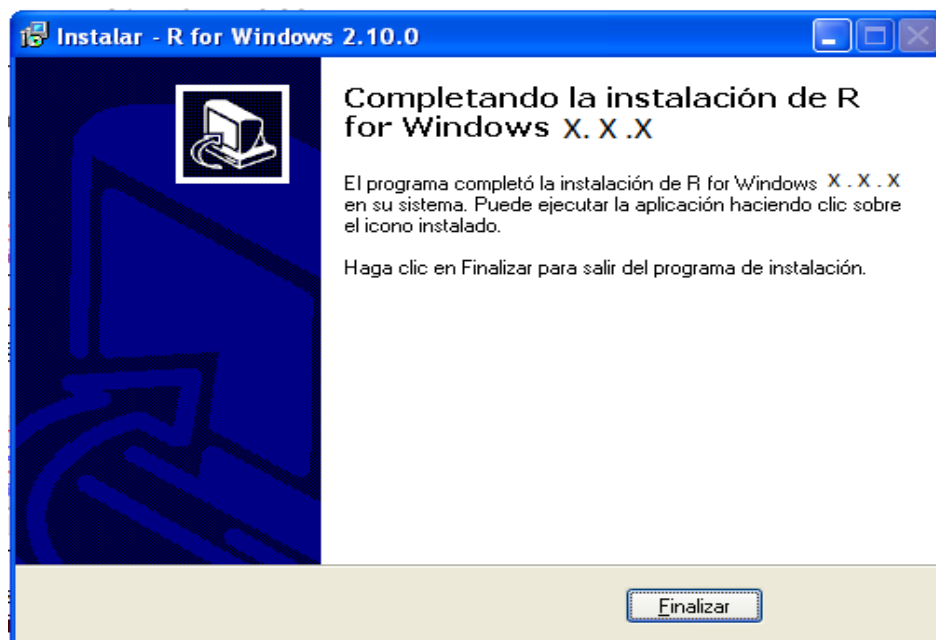
El software libre R en su versión 3.6.3 permite realizar análisis estadísticos y es uno de los más utilizado en la comunidad científica. Este programa está disponible en la página web: <http://www.r-project.org> y consta de una aplicación central y de librerías de multitud de temas que se pueden instalar según necesidad.

Asimismo, R es un programa de instrucciones y, por tanto, no resulta del todo “amigable” para los usuarios que no están acostumbrados a este tipo de manejo. En la actualidad, existe una interfaz que permite el manejo del programa R mediante una ventana de menús, este interfaz se llama RCommander.

Como primer paso, se debe instalar el software R y cargar algunas librerías, sin las cuales, no es posible hacer el tratamiento de la información, como se muestra en la Figura 20 :

**Figura 20.**

*Instalación del Software R.*



Para instalar las librerías se debe ingresar a la opción Instalar paquetes, como muestra la Figura 21:

**Figura 21.**

*Instalación del Librerías del Software R.*



A continuación, se selecciona la región que se quiere descargar del paquete y se pulsa OK, como muestra la Figura 22:

**Figura 22.**

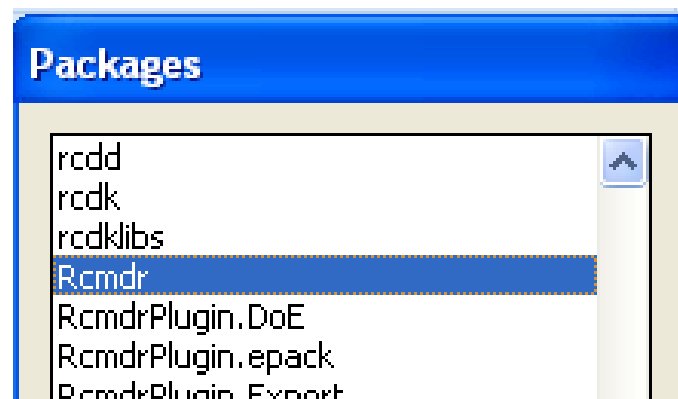
*Servidores Disponibles Librerías Software R.*



En la ventana Packages que aparece, se busca desplazándose con la barra lateral hacia abajo buscando las librerías que se necesitan para el proyecto de investigación y se le da OK, como muestra la Figura 23:

**Figura 23**

*Instalación del Packages Software R.*



Comenzará entonces la descarga e instalación de esta librería; al terminar la descarga se debe instalar con el comando `install` seguido del nombre de la librería, así por ejemplo, para la librería `dplyr`, desde la línea de comandos.

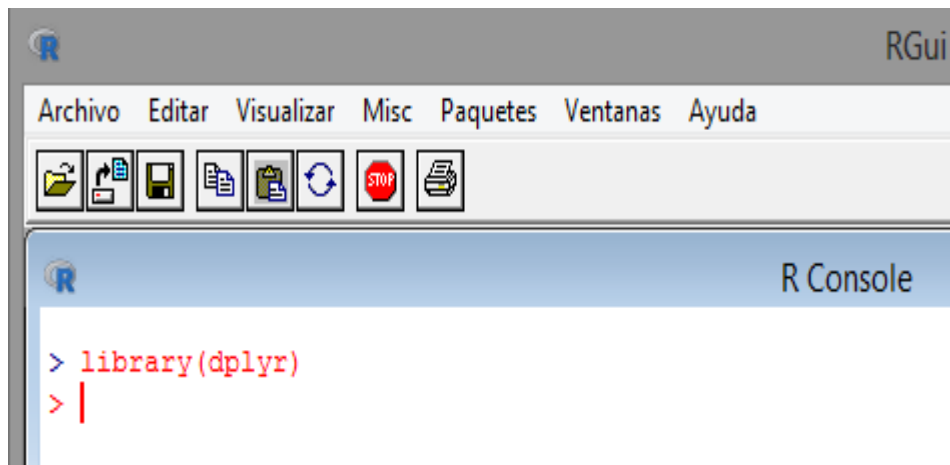
*`install.packages("dplyr")`*

Finalmente, para poder usar la librería, se debe llamarla cada vez que se inicie una sesión en R con el comando `library` seguido del nombre de la librería, que muestra la Figura 24:

*`library(dplyr)`*

**Figura 24**

*Llamar a una Librería en Software R.*



Las librerías son útiles para múltiples propósitos y facilitan mucho el tratamiento de la información y el consiguiente diseño algorítmico.

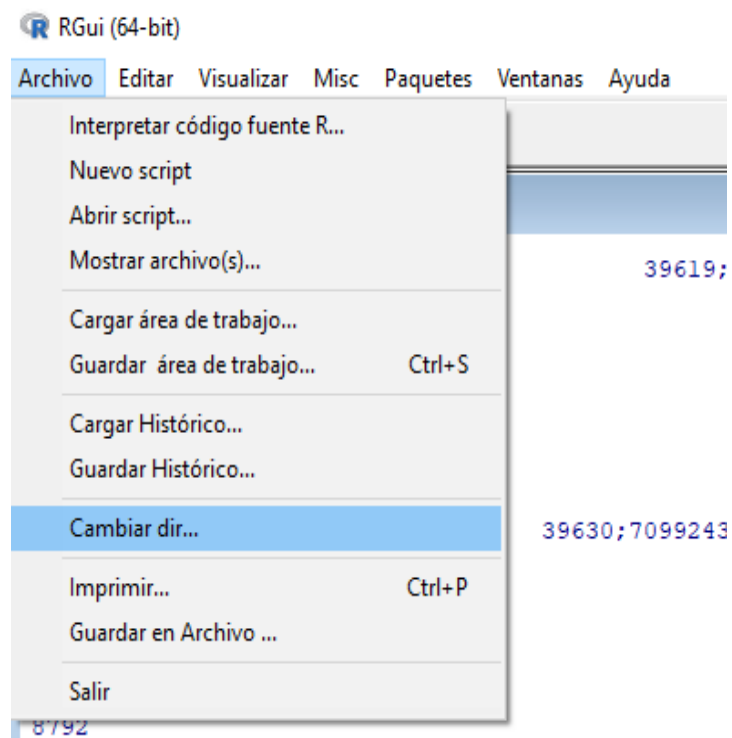
#### **4.4 Librerías a utilizar software R**

- `library(tidyverse)` : utilizado para Data Science.
- `library(AppliedPredictiveModeling)`: Modelos Predictivos.
- `library(ggplot)`: Para diseño de Tablas.
- `library(dplyr)`: Para trabajar con Data.Frames.

Se debe seleccionar al inicio la ruta de trabajo para el proyecto con la opción *Cambiar dir...* ya que si no podría ocurrir errores con los archivos y/o librerías; por defecto se debe configurar la carpeta *mis documentos* o en su defecto, otra que ya se tenga con los datos, como muestra la Figura 25:

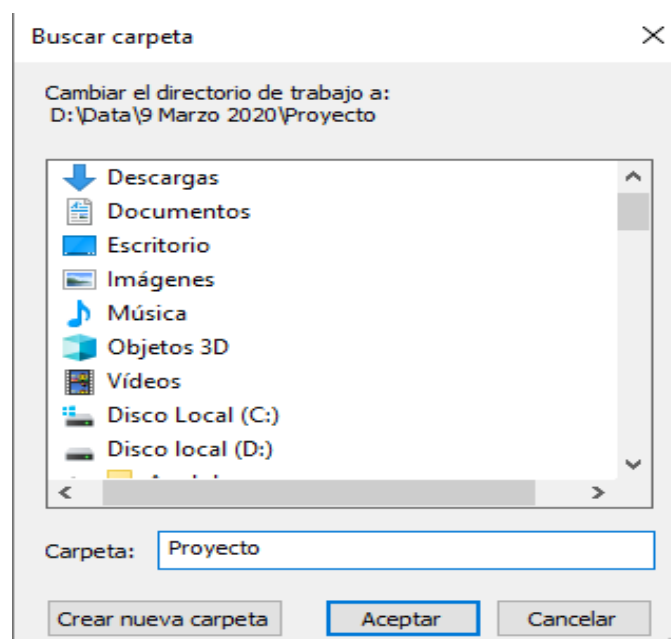
**Figura 25**

*Cambio de Ruta de Trabajo.*



**Figura 26**

*Selección de carpeta llamada Proyecto.*





Para comenzar a trabajar se debe cargar la información donde se encuentran almacenados los datos a procesar. En este sentido, el Software R acepta multitud de formatos, entre los más conocidos se encuentran:

- CSV: csv
- EXCEL: .xls y .xlsx
- SPSS: .sav y .por
- STATA: .dta
- SAS: .sas
- R objects: .RData o .rda
- Serialized R objects: .rds
- JSON
- XML
- Bases de Datos
- Etc

#### **4.5 Preprocesamiento de datos software R**

Para ello, como se indicó anteriormente, se dispone de una base de datos de fugas histórica, de la cual se toma el periodo comprendido entre los años 2014 - 2019 y de estos, se consideran las fugas más resaltantes y con mayor pérdida y de ahí se genera el modelo predictivo.

Como dato adicional, sólo se considerarán los caudales con valores iguales o mayores a 1000 L/día, por ser más relevantes para esta

investigación, ya que pérdidas a menores a este valor no se pueden considerar como pérdidas significativas.

En total, 23 fugas fueron registradas en el Callao y las restantes en Lima; se tienen en total 2858 registros de fugas registradas en ese periodo de tiempo con diferentes presiones y diferentes diámetros de tuberías, los cuales se tienen que filtrar por distrito y por caudales para realizar una mejor aproximación del modelo.

Lo primero es hacer el filtrado de la información y centrarse en un tipo de fuga ya que será más conveniente para el análisis predictivo; se procede a exportar el archivo a Excel para que sea más fácil el filtrado de la información por tipo y por fecha y por tipo de fuga y caudal y los demás datos que sean importantes para el filtrado.

Como se mencionó, las fugas dependen de donde estén ubicadas en el tendido de la red de agua potable y dependerá de la presión del agua y del tipo de terreno donde se encuentre la tubería para que sea fácil su escucha.

Toda esa información denominada BASE DE DATOS DE FUGAS que contiene la información histórica, es la que se usará para el desarrollo del algoritmo. Los tipos de fuga pueden ser:

- Caja A/Medidor
- Caja D/Medidor

- Línea A/Caja
- Línea A/Caja
- Corporation
- Tubería
- Válvula De Red
- Válvula Grifo C/L

Lo primero es hacer el filtrado de la información y centrarse en los campos más relevantes de la Tabla para ejecutar análisis predictivo; se procede a exportar el archivo a Excel para que sea más fácil el filtrado de la información por distrito, por fecha y por tipo de fuga y los demás datos que sean importantes para el filtrado.

La base de datos original fue extraída de una aplicación en entorno SQL, y pueden ser exportados en Excel. Una vez que se tienen los datos exportados a Excel, se procede a hacer la discriminación de los mismos, para esto se escogerán los datos de las fugas llamadas de emergencia, las cuales son incidencias en tuberías y redes primarias de un periodo comprendido entre los años 2012 al año 2019; es un periodo de 7 años y se usará el año 2020 como dato de control para el algoritmo.

También se puede enviar los datos directamente de Excel a software R y desde ahí hacer la discriminación de datos, pero para fines didácticos se mostrará desde Excel, para visualizar los campos y algunos tipos de datos en ellos y su posterior conversión a CSV, como muestra la Figura 27:

**Figura 27**

*Base de datos exportados a Excel.*

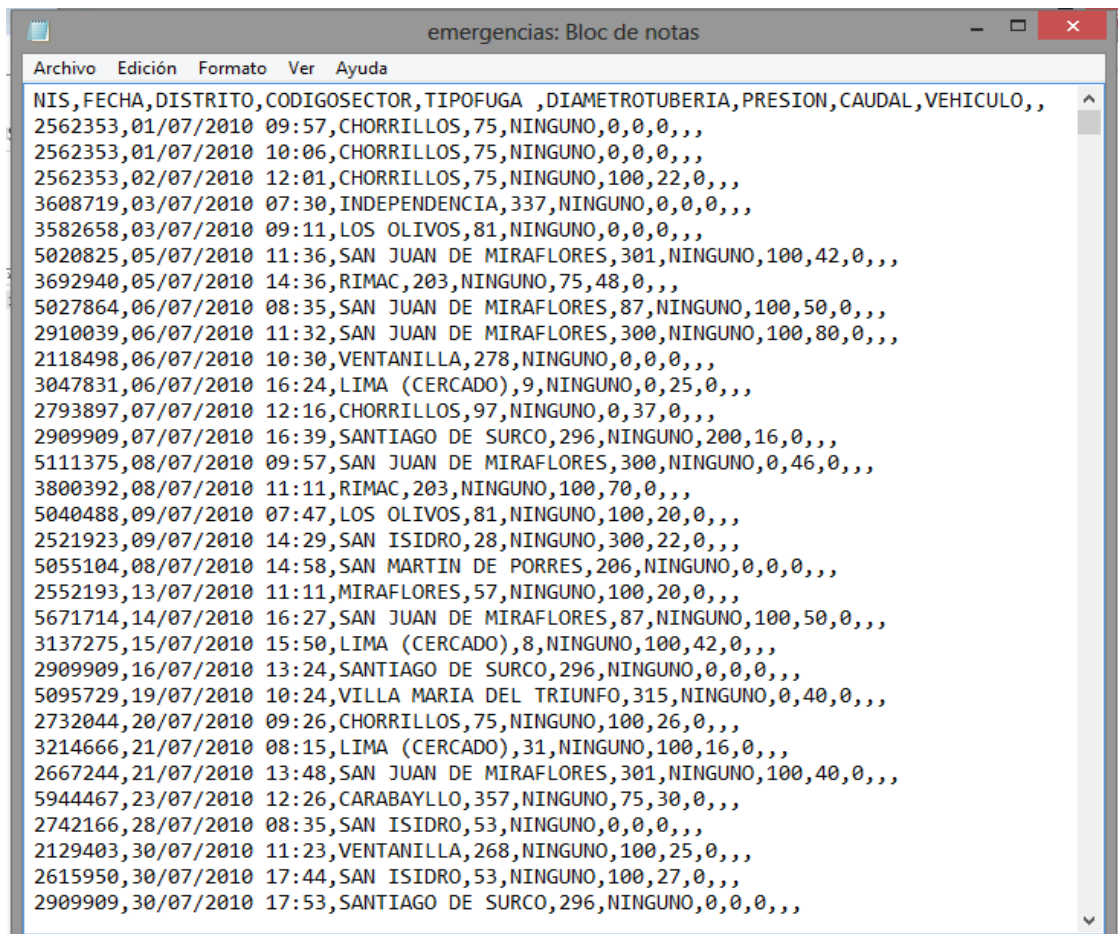
	A	B	C	D	E	F	G	H	I	J	K
1	Nº DE FUGA	NIS	FECHA	DISTRITO	URBANIZACION	CODIGO SECTOR	TIPO DE FUGA	DIAMETRO TUBERIA	PRESION (PSI)	CAUDAL (L)	VEHICUL
420	23265	2717633	20/07/2017 15:19	SAN BORJA	URB CHACARILLA DEL ESTANQUE		72 TUBERIA	600	15	150000	4072
426	24532	2533457	9/09/2017 15:04	MIRAFLORES	URB BARBONCITO		56 TUBERIA	100	14	150000	4072
430	30958	3045311	9/05/2018 12:04	LIMA (CERCADO)	CERCADO		9 TUBERIA	300	40	150000	4073
433	33310	3821039	22/08/2018 12:09	RIMAC	P.J VILLA FATIMA		200 TUBERIA	100	55	150000	4073
434	27408	3003841	21/12/2017 09:44	LIMA (CERCADO)	CERCADO		8 CORP	15	0	120000	4074
436	7391	3573654	15/02/2015 18:32	COMAS	URB SAN FELIPE 2DA ET.		350 TUBERIA	150	10	100000	4075
437	8710	4201997	21/05/2015 07:50	EL AGUSTINO	P.J EL AGUSTINO 1 ZONA		2 TUBERIA	110	40	100000	771
439	14305	3186841	13/04/2016 16:46	LA VICTORIA	URB SANTA CATALINA		18 TUBERIA	100	36	100000	771
442	22380	3530577	8/06/2017 09:46	SAN MARTIN DE PORRES	URB PERU		205 TUBERIA	250	6	100000	4071
443	22600	2565919	19/06/2017 10:40	CHORRILLOS	URB CHORRILLOS		75 TUBERIA	150	20	100000	771
447	22879	6059692	1/07/2017 14:24	CARABAYLLO	A.H SAN GRABIEL		357 TUBERIA	60	60	100000	4072
451	23163	3633227	15/07/2017 11:01	PUEENTE PIEDRA	CERCADO		386 TUBERIA	200	42	100000	4071
461	27912	2596083	13/01/2018 15:24	SANTIAGO DE SURCO	URB DERBY EL		297 TUBERIA	100	23	100000	4072
462	30144	2742059	10/04/2018 09:27	SAN ISIDRO	URB CHACARILLA DE SANTA CRUZ		52 TUBERIA	150	26	100000	4073
463	32894	3726145	10/08/2018 08:27	SAN MARTIN DE PORRES	URB PALAO 1RA ET.		206 TUBERIA	200	38	100000	4074
470	32966	5880178	10/08/2018 15:49	CARABAYLLO	A.H VILLA ESPERANZA AMPL. DEL		357 TUBERIA	100	80	100000	4074
472	34473	3800631	16/10/2018 09:40	RIMAC	A.H MUNICIPAL 3		203 CORP	15	10	100000	4074
612	21526	2611609	29/04/2017 13:45	SAN BORJA	URB SAN BORJA		71 TUBERIA	100	20	90000	4072
624	25618	6433321	21/10/2017 13:47	PUEENTE PIEDRA	A.H TWINZA		389 TUBERIA	100	50	85000	836
647	2815	5008163	9/07/2014 15:54	PUCUSANA	ASOC ISLA GALAPAGOS		541 TUBERIA	75	40	80000	
648	2960	2801997	6/08/2014 14:02	VILLA MARIA DEL TRIUNFO	P.J GALVEZ, POETA JOSE		315 CORP	15	20	80000	4071
653	9514	2543813	7/07/2015 07:20	BARRANCO	CERCADO		74 TUBERIA	350	28	80000	771
655	13333	2534961	15/02/2016 09:11	SAN ISIDRO	URB SAN GABRIEL		50 TUBERIA	100	20	80000	4071
679	13742	3204606	10/03/2016 13:04	PUEBLO LIBRE	CERC CERCADO		30 TUBERIA	200	20	80000	771
680	15276	5115567	9/06/2016 06:45	SAN MARTIN DE PORRES	A.H SEÑOR DE LOS MILAGROS		206 TUBERIA	90	65	80000	4073
681	17888	6797976	11/10/2016 16:55	SAN MIGUEL	URB MARANGA 1RA ET.		45 TUBERIA	100	6	80000	771
688	18719	2901250	18/11/2016 14:07	VILLA MARIA DEL TRIUNFO	P.J GALVEZ, POETA JOSE		315 TUBERIA	100	40	80000	4074
692	25738	5288420	27/10/2017 09:26	SAN MARTIN DE PORRES	A.H NUEVO AMANECER		206 TUBERIA	160	47	80000	4072
698	27644	5087320	4/01/2018 10:10	EL AGUSTINO	A.H TERRAZAS DE CATALINA HUAN		1 CORP	15	40	80000	4073
612	27962	5767593	16/01/2018 08:28	COMAS	A.H PRIMAVERA Y MANCO INCA 11		339 TUBERIA	90	40	80000	4072
625	29001	3719844	24/02/2018 12:25	COMAS	A.H COLLIQUE		349 TUBERIA	75	40	80000	4074
627	29052	3150544	27/02/2018 10:46	MAGDALENA	URB PRIMAVERA		48 CORP	15	30	80000	836
636	29419	5155776	13/03/2018 13:51	LOS OLIVOS	P.J JULIO, 28 DE		80 TUBERIA	75	9	80000	841
643	29979	2590977	3/04/2018 10:20	SAN ISIDRO	URB CORPAC		53 TUBERIA	100	28	80000	4073
654	30206	3590244	12/04/2018 10:03	LOS OLIVOS	URB PARQUE EL NARANJAL 1RA ET		83 LINEA A/CAJA	15	17	80000	4072
673	31474	5370507	1/06/2018 09:34	CARABAYLLO	A.H FLOR II ET. SECT. I Y II, LA		357 TUBERIA	110	100	80000	4073
675	31735	3813933	14/06/2018 11:16	RIMAC	A.H ZEVALLOS, HORACIO		203 CORP	15	30	80000	836
676	33958	5185092	19/09/2018 10:58	LA MOLINA	URB CAPILLA, LA		199 CORP	15	15	80000	4074
680	34750	2888821	26/10/2018 09:44	SAN BORJA	A.H SAN JUAN MASIAS		20 TUBERIA	250	30	80000	4072
684	35968	3130907	17/12/2018 17:59	PUEBLO LIBRE	CERC CERCADO		30 TUBERIA	100	37	80000	4072

*Nota.* Adaptado de Rubiños & Apesteagua (2020)

La Base de datos cuenta con 67 columnas y 8226 filas de los cuales varios son vacíos y/o duplicados; se procede a eliminar y a dejar solo los campos más importantes. Una vez terminado, se exporta el archivo a formato CSV separado por comas quedando como se muestra en la Figura 28:

Figura 28.

Datos de Excel exportados a CSV.



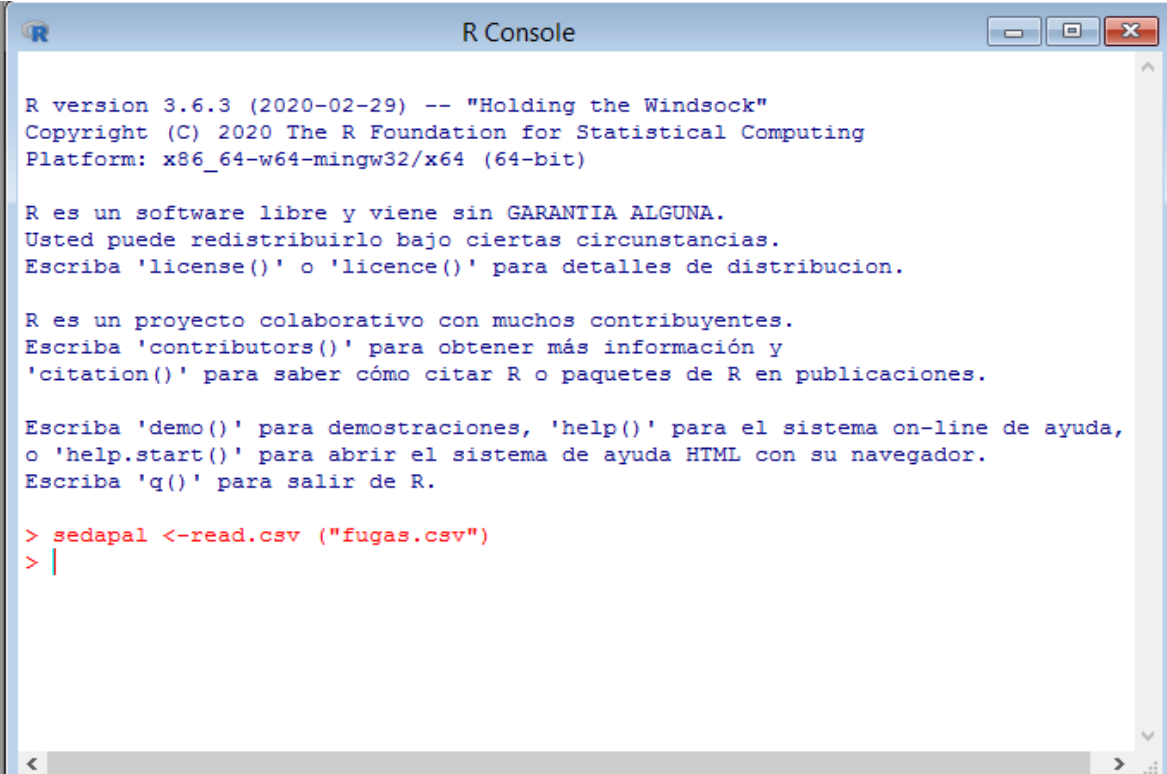
Luego para cargar el archivo se ejecuta el comando *read.csv*

```
sedapal <-read.csv ("fugas.csv")
```

El formato se escribe así porque primero se le asigna un nombre al archivo **fugas.csv** para que sea cargado en memoria en un arreglo de datos, en este caso, se le llamará **sedapal** y se da **enter** y no debe aparecer ningún mensaje de error tal como se observa en la Figura 29:

**Figura 29**

*Cargar archivo fugas.csv en Memoria.*



```
R Console

R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

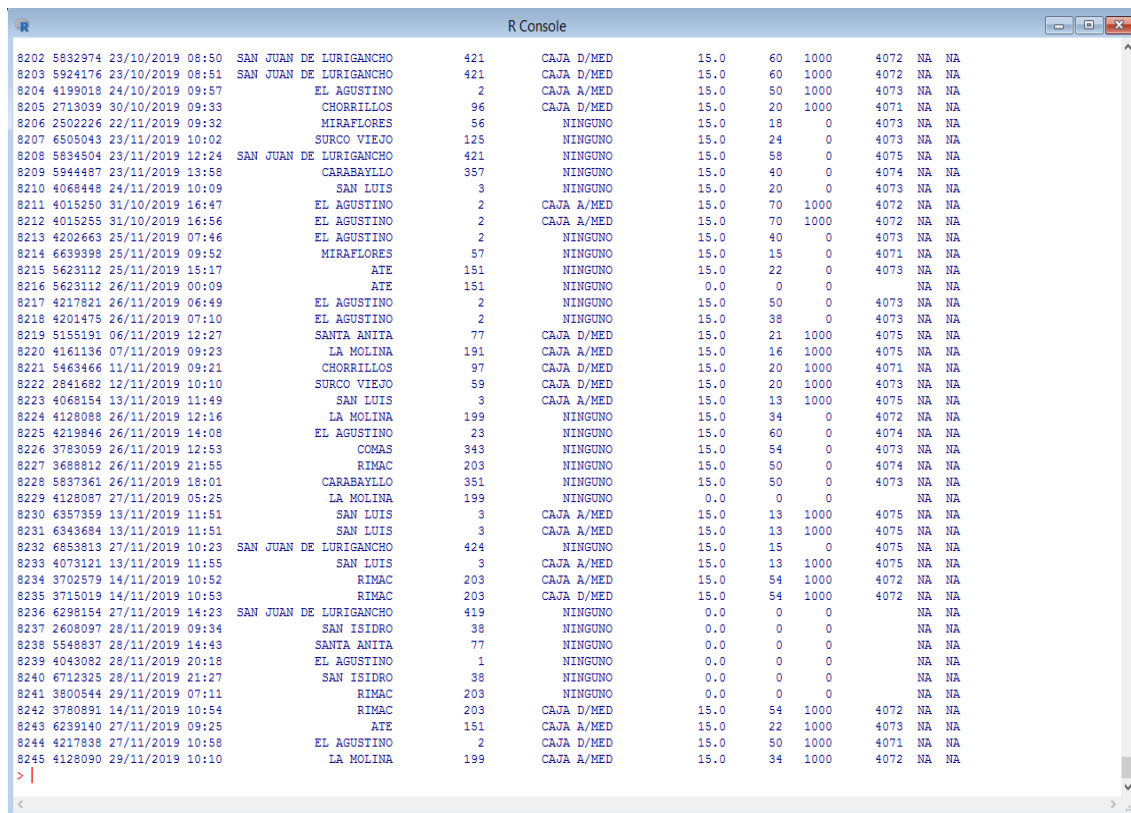
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> sedapal <-read.csv ("fugas.csv")
> |
```

Finalmente, se escribe el nombre con el cual se ha designado al archivo, en este caso **sedapal** y se presiona **enter**, dando el siguiente resultado tal como se observa en la Figura 30:

Figura 30.

Archivo fugas.csv cargado en R.



```
R Console
8202 5832974 23/10/2019 08:50 SAN JUAN DE LURIGANCHO 421 CAJA D/MED 15.0 60 1000 4072 NA NA
8203 5924176 23/10/2019 08:51 SAN JUAN DE LURIGANCHO 421 CAJA D/MED 15.0 60 1000 4072 NA NA
8204 4199018 24/10/2019 09:57 EL AGUSTINO 2 CAJA A/MED 15.0 50 1000 4073 NA NA
8205 2713039 30/10/2019 09:33 CHORRILLOS 96 CAJA D/MED 15.0 20 1000 4071 NA NA
8206 2502226 22/11/2019 09:32 MIRAFLORES 56 NINGUNO 15.0 18 0 4073 NA NA
8207 6505043 23/11/2019 10:02 SURCO VIEJO 125 NINGUNO 15.0 24 0 4073 NA NA
8208 5834504 23/11/2019 12:24 SAN JUAN DE LURIGANCHO 421 NINGUNO 15.0 58 0 4075 NA NA
8209 5944487 23/11/2019 13:58 CARABAYLLO 357 NINGUNO 15.0 40 0 4074 NA NA
8210 4068448 24/11/2019 10:09 SAN LUIS 3 NINGUNO 15.0 20 0 4073 NA NA
8211 4015250 31/10/2019 16:47 EL AGUSTINO 2 CAJA A/MED 15.0 70 1000 4072 NA NA
8212 4015255 31/10/2019 16:56 EL AGUSTINO 2 CAJA A/MED 15.0 70 1000 4072 NA NA
8213 4202663 25/11/2019 07:46 EL AGUSTINO 2 NINGUNO 15.0 40 0 4073 NA NA
8214 6639398 25/11/2019 09:52 MIRAFLORES 57 NINGUNO 15.0 15 0 4071 NA NA
8215 5623112 25/11/2019 15:17 ATE 151 NINGUNO 15.0 22 0 4073 NA NA
8216 5623112 26/11/2019 00:09 ATE 151 NINGUNO 0.0 0 0 NA NA
8217 4217821 26/11/2019 06:49 EL AGUSTINO 2 NINGUNO 15.0 50 0 4073 NA NA
8218 4201475 26/11/2019 07:10 EL AGUSTINO 2 NINGUNO 15.0 38 0 4073 NA NA
8219 5155191 06/11/2019 12:27 SANTA ANITA 77 CAJA D/MED 15.0 21 1000 4075 NA NA
8220 4161136 07/11/2019 09:23 LA MOLINA 191 CAJA A/MED 15.0 16 1000 4075 NA NA
8221 5463466 11/11/2019 09:21 CHORRILLOS 97 CAJA D/MED 15.0 20 1000 4071 NA NA
8222 2841682 12/11/2019 10:10 SURCO VIEJO 59 CAJA D/MED 15.0 20 1000 4073 NA NA
8223 4068154 13/11/2019 11:49 SAN LUIS 3 CAJA A/MED 15.0 13 1000 4075 NA NA
8224 4128088 26/11/2019 12:16 LA MOLINA 199 NINGUNO 15.0 34 0 4072 NA NA
8225 4219846 26/11/2019 14:08 EL AGUSTINO 23 NINGUNO 15.0 60 0 4074 NA NA
8226 3783059 26/11/2019 12:53 COMAS 343 NINGUNO 15.0 54 0 4073 NA NA
8227 3688812 26/11/2019 21:55 RIMAC 203 NINGUNO 15.0 50 0 4074 NA NA
8228 5837361 26/11/2019 18:01 CARABAYLLO 351 NINGUNO 15.0 50 0 4073 NA NA
8229 4128087 27/11/2019 05:25 LA MOLINA 199 NINGUNO 0.0 0 0 NA NA
8230 6357359 13/11/2019 11:51 SAN LUIS 3 CAJA A/MED 15.0 13 1000 4075 NA NA
8231 6343684 13/11/2019 11:51 SAN LUIS 3 CAJA A/MED 15.0 13 1000 4075 NA NA
8232 6853813 27/11/2019 10:23 SAN JUAN DE LURIGANCHO 424 NINGUNO 15.0 15 0 4075 NA NA
8233 4073121 13/11/2019 11:55 SAN LUIS 3 CAJA A/MED 15.0 13 1000 4075 NA NA
8234 3702579 14/11/2019 10:52 RIMAC 203 CAJA A/MED 15.0 54 1000 4072 NA NA
8235 3715019 14/11/2019 10:53 RIMAC 203 CAJA D/MED 15.0 54 1000 4072 NA NA
8236 6298154 27/11/2019 14:23 SAN JUAN DE LURIGANCHO 419 NINGUNO 0.0 0 0 NA NA
8237 2608097 28/11/2019 09:34 SAN ISIDRO 38 NINGUNO 0.0 0 0 NA NA
8238 5548837 28/11/2019 14:43 SANTA ANITA 77 NINGUNO 0.0 0 0 NA NA
8239 4043082 28/11/2019 20:18 EL AGUSTINO 1 NINGUNO 0.0 0 0 NA NA
8240 6712325 28/11/2019 21:27 SAN ISIDRO 38 NINGUNO 0.0 0 0 NA NA
8241 3800544 29/11/2019 07:11 RIMAC 203 NINGUNO 0.0 0 0 NA NA
8242 3780891 14/11/2019 10:54 RIMAC 203 CAJA D/MED 15.0 54 1000 4072 NA NA
8243 6239140 27/11/2019 09:25 ATE 151 CAJA A/MED 15.0 22 1000 4073 NA NA
8244 4217838 27/11/2019 10:58 EL AGUSTINO 2 CAJA D/MED 15.0 50 1000 4071 NA NA
8245 4128090 29/11/2019 10:10 LA MOLINA 199 CAJA A/MED 15.0 34 1000 4072 NA NA
> |
```

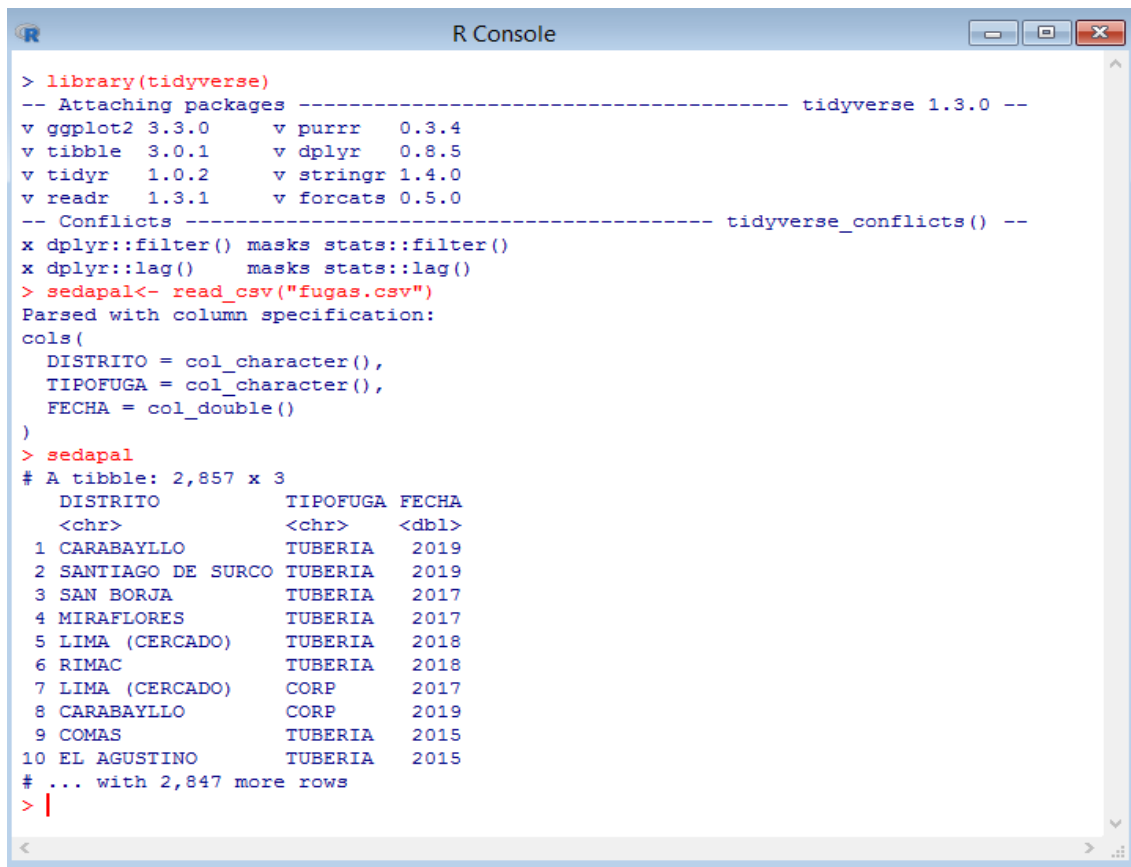
Otra forma de llamar al archivo es con el comando `read_csv` pero se necesita de la librería `library(tidyverse)` lo cual no muestra de otra forma, los datos en pantalla con información sobre los campos y el tipo de dato.

```
library(tidyverse)
sedapal<- read_csv("fugas.csv")
```

Esto es muy útil para conocer cómo hacer el tratamiento de la información y su posterior análisis en detalle. Finalmente, se escribe el nombre con el cual se ha designado el archivo, en este caso `sedapal` y se presiona **enter**, dando el siguiente resultado.

**Figura 31.**

*Archivo CSV usando la librería tidyverse.*



```
> library(tidyverse)
-- Attaching packages ----- tidyverse 1.3.0 --
v ggplot2 3.3.0      v purrr  0.3.4
v tibble  3.0.1      v dplyr  0.8.5
v tidyr   1.0.2      v stringr 1.4.0
v readr   1.3.1      v forcats 0.5.0
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
> sedapal<- read_csv("fugas.csv")
Parsed with column specification:
cols(
  DISTRIITO = col_character(),
  TIPOFUGA = col_character(),
  FECHA = col_double()
)
> sedapal
# A tibble: 2,857 x 3
  DISTRIITO      TIPOFUGA FECHA
  <chr>          <chr>   <dbl>
1 CARABAYLLO    TUBERIA  2019
2 SANTIAGO DE SURCO TUBERIA  2019
3 SAN BORJA     TUBERIA  2017
4 MIRAFLORES   TUBERIA  2017
5 LIMA (CERCADO) TUBERIA  2018
6 RIMAC        TUBERIA  2018
7 LIMA (CERCADO) CORP     2017
8 CARABAYLLO    CORP     2019
9 COMAS        TUBERIA  2015
10 EL AGUSTINO   TUBERIA  2015
# ... with 2,847 more rows
> |
```

#### 4.6 Uso del comando SPLIT

Si se desea tener la misma información pero clasificada por años, se puede emplear la función **split()** que recibe básicamente dos argumentos principales:

- el objeto a ser clasificado (típicamente, un data frame o un vector) y,
- una serie de datos que servirán para la clasificación (típicamente, un factor o un vector numérico o de cadenas de caracteres).



Este último argumento debe tener la misma longitud que una columna del data frame, dado como primer argumento, o que el vector dado como argumento. Según el caso, y no es necesario que este argumento sea parte del objeto dado como primer argumento. Sí, como primer argumento de la función.

#### a) Uso del comando `split` por vehículo

Ahora bien, si se tiene la misma información, pero clasificada por vehículos, se puede emplear la función `split()`, que recibe básicamente dos argumentos principales:

- el objeto a ser clasificado (típicamente, un data frame o un vector) y,
- una serie de datos que servirán para la clasificación (típicamente, un factor o un vector numérico o de cadenas de caracteres).

Este último argumento debe tener la misma longitud que una columna del data frame, dado como primer argumento, o que el vector dado como argumento. Según el caso, y no es necesario que este argumento sea parte del objeto dado como primer argumento. Si, como primer argumento de la función pasamos una porción de la tabla contenida en el data frame justamente leído -sólo las columnas 3 a la 5 del data frame-, y como segundo argumento pasamos la columna correspondiente a los VEHICULOS, la función operaría como se muestra y en este caso el resultado de la función, en el caso del ejemplo, es una lista de tablas, cada

una correspondiente a cada uno de los años registrados en la columna `sedapal$VEHICULO`, como se muestra a continuación:

```
sedapal_vehiculo <- split(sedapal[, 1:3], sedapal$DISTRITO)  
sedapal_vehiculo
```

Al ejecutar el comando se pueden clasificar los datos, en este caso por vehículo:

```
$`4073`
```

DISTRITO	.SECTOR	TIPOFUGA
1604 SAN MARTIN DE PORRES	208	CORP
2299 SAN MARTIN DE PORRES	205	NINGUNO

```
$`4074`
```

DISTRITO	SECTOR	TIPOFUGA
1401 SANTIAGO DE SURCO	123	TUBERIA

## **b) Uso del comando `split` por código del sector**

Si se desea tener la misma información, pero clasificada por `CODIGO.SECTOR` se puede emplear la función `split()`, que recibe básicamente dos argumentos principales: el objeto a ser clasificado - típicamente, un data frame o un vector-, y una serie de datos que servirán para la clasificación -típicamente, un factor o un vector numérico o de cadenas de caracteres-.

Este último argumento debe tener la misma longitud que una columna del data frame, dado como primer argumento, o que el vector dado como argumento según el caso, y no es necesario que este argumento sea parte del objeto dado como primer argumento.

Si, como primer argumento de la función pasamos una porción de la tabla contenida en el data frame justamente leído -sólo las columnas 1 a la 8 del data frame-, y como segundo argumento pasamos la columna correspondiente a los CÓDIGO DE SECTOR, la función operaría como se muestra y en este caso el resultado de la función, en el caso del ejemplo, es una lista de tablas, cada una correspondiente a cada uno de los años registrados en la columna `sedapal$CODIGO.SECTOR`, como se muestra a continuación:

```
sedapal_sector <- split(sedapal[, 1:3], sedapal$CODIGO.SECTOR)
```

**sedapal\_sector**

\$`467`

	FECHA	DISTRITO	TIPOFUGA	DIAMETRO	PRESION	CAUDAL
1400	21/03/2014 15:04	LURIN	NINGUNO	0	40	0
1422	4/04/2014 10:58	LURIN	NINGUNO	0	0	0

\$`501`

	FECHA	DISTRITO	TIPOFUGA	DIAMETRO	PRESION	CAUDAL
521	9/04/2012 11:41	PUNTA HERMOSA	NINGUNO	0	0	0
532	16/04/2012 11:47	PUNTA HERMOSA	NINGUNO	50	15	0

### c) Uso del comando `split` por caudal

Ahora se desea tener la misma información, pero clasificada por CAUDAL. Con este propósito se puede emplear la función `split()`, que recibe básicamente dos argumentos principales:

- el objeto a ser clasificado (típicamente, un data frame o un vector) y,
- una serie de datos que servirán para la clasificación (típicamente, un factor o un vector numérico o de cadenas de caracteres).

Este último argumento debe tener la misma longitud que una columna del data frame, dado como primer argumento, o que el vector dado como argumento. según el caso, y no es necesario que este argumento sea parte del objeto dado como primer argumento.

Si, como primer argumento de la función pasamos una porción de la tabla contenida en el data frame justamente leído -sólo las columnas 1 a la 8 del data frame-, y como segundo argumento pasamos la columna correspondiente a los CAUDAL, la función operaría como se muestra y en este caso el resultado de la función, en el caso del ejemplo, es una lista de tablas, cada una correspondiente a cada uno de los años registrados en la columna `sedapal$CAUDAL`, como se muestra a continuación:

```
sedapal_caudal <- split(sedapal[, 1:3], sedapal$CAUDAL)  
sedapal_caudal
```

`150000`

NIS	FECHA	DISTRITO	CODIGO.SECTOR	TIPOFUGA	DIAMETRO	TUBERIA
1419	2717633	20/07/2017	15:19	SAN BORJA	72	TUBERIA
1425	2533457	9/09/2017	15:04	MIRAFLORES	55	TUBERIA

150000

\$`180000`

NIS	FECHA	DISTRITO	CODIGO.SECTOR	TIPOFUGA	DIAMETRO	TUBERIA
1401	5968591	17/05/2019	10:56	SANTIAGO DE SURCO	123	TUBERIA

8 180000

\$`200000`

NIS	FECHA	DISTRITO	CODIGO.SECTOR	TIPOFUGA	DIAMETRO	TUBERIA
1399	5183985	11/10/2019	11:10	CARABAYLLO	357	TUBERIA

200000

Como se observa, el comando `SPLIT` es uno de los comandos más útiles del R porque permite ver en forma ordenada, lo cual es útil para agrupar los datos y hacer más fácilmente el análisis respectivo de los datos.

#### 4.7 Uso del comando `tapply`

Finalmente, en este tipo de funciones está la función `tapply()`, que efectúa el mismo tipo de operaciones de clasificación y operación, pero actúa sobre un vector y no sobre un data frame. Cada una de las columnas

del data frame leído, es o bien un vector numérico, o un factor. Debido a la organización de la tabla anterior, para obtener mismo el resultado que las funciones usadas anteriormente, `by()` y `aggregate()`, la función `tapply()` tiene que sacar provecho de que el segundo argumento, que sirve para clasificar los datos, puede ser una lista de uno o más objetos, interpretables como factores, y que en el caso del ejemplo estará constituida por las columnas correspondiente a los TIPOSDEFUGA y a los CAUDALES. Así, el resultado deseado se obtiene como sigue:

```

sedapal <- read.csv("FUGAS.csv")

sedapal

> (rr <- tapply(sedapal$DISTRITO, list(sedapal$TIPOFUGA,
sedapal$CAUDAL), mean, na.rm = T))

```

Así, el resultado deseado se obtiene como sigue:

```

0 15 32  50 100  200 500  1000  2000  3000  4000  5000  6000  7000  8000
9000 10000 12000 15000 16000
CAJA A/MED      140.6667 NA NA 275.2  2  NA 88.3 132.39394 150.10323 108.83929
194.7500 187.5185 343.0000 165.4000 92.5000 NA  NA  NA 203.0000  NA
CAJA D/MED      NA NA NA  NA  NA 188.5 79.2 115.42647 137.66364 149.90244
186.6364 181.0625  NA 177.5000 332.0000  NA 178.0000  NA  NA  NA
CORP            132.0000 NA NA  NA  NA  NA  NA 145.00000 132.00000 155.00000
128.5556 148.5338 159.4348 165.3448 148.3786 203 153.5420 138 142.6364 18
LINEA A/CAJA    2.0000 NA NA 203.0  NA  NA 9.5 147.38462 174.17647 98.43333
139.3750 161.8431 103.0000 138.4615 127.9333  NA 139.3750  NA 120.5000  NA

```

LINEA D/CAJA	NA	NA	NA	NA	NA	NA	NA	98.18182	137.32353	145.20000
	271.7778	138.2692	272.5000	188.2500	243.8571	NA	167.0000	NA	NA	NA
MEDIDOR	NA	NA	NA	NA	NA	NA	NA	90.00000	NA	NA
	90.0000	NA	NA	NA	NA	NA	NA	NA	NA	NA
NINGUNO	173.4030	27	NA	NA	NA	NA	NA	NA	228.00000	203.00000
NA	NA	NA	2.0000	NA	NA	223.5000	NA	NA	NA	NA

Se están evaluando y comparando varios parámetros a la vez utilizando el comando TAPPLY para obtener los datos agrupados de una mejor manera en este caso, con respecto al SECTOR junto con los valores de TIPODE FUGA y CAUDAL respectivamente.

Como se puede ver por lo expuesto, mucho del trabajo que otros lenguajes resuelven con ciclos, decisiones y secuencias de instrucciones, en R, alternativamente, se puede hacer por medio de este elegante conjunto de funciones de clasificación, transformación y agregación de datos, lo que junto con las operaciones para manipular porciones de los datos estructurados, hacen del lenguaje una poderosa herramienta para el procesamiento de información.

Se puede observar que R es versátil para el manejo y procesamiento de la información y la forma de representarla, nuestra base de datos de las fugas obtenidas a lo largo de varios años, va ser exportada y procesada filtrando los datos más relevantes para con ellos hacer el diseño del algoritmo. En este sentido, interesan 3 campos, los más importantes para poder parametrizar mejor los datos y estos son:

- Distrito
- Tipo de Fuga
- Fecha

Una vez que se importan datos a R es conveniente limpiarlos, esto implica almacenarlos de una manera consistente que permita enfocarse en responder preguntas de los datos en lugar de estar luchando con los datos. Entonces, datos limpios son datos que facilitan las tareas del análisis de datos:

- Manipulación: manipulación de variables como agregar, filtrar, reordenar, transformar.
- Visualización: resúmenes de datos usando gráficas, análisis exploratorio, o presentación de resultados.
- Modelación: ajustar modelos es sencillo si los datos están en la forma correcta.

Los principios de datos limpios proveen una manera estándar de organizar la información:

- Cada variable forma una columna.
- Cada observación forma un renglón.
- Cada tipo de unidad observacional forma una tabla.

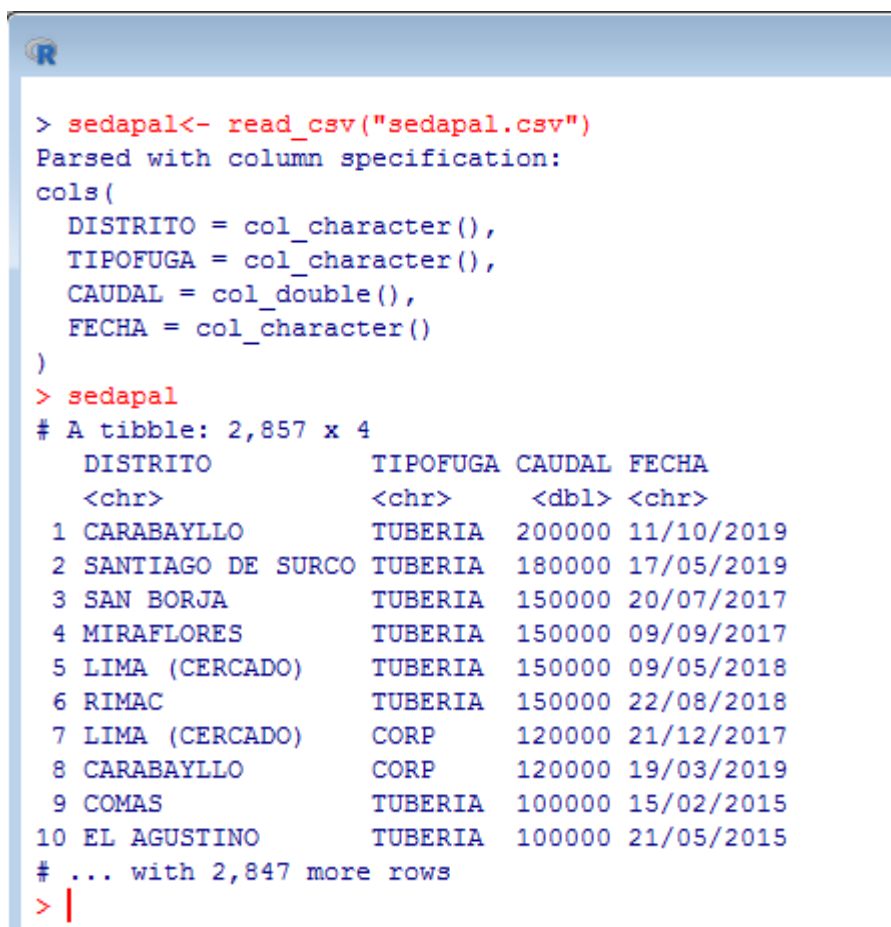


Vale la pena notar que los principios de los datos limpios se pueden ver como teoría de álgebra relacional para estadísticos, estos principios equivalen a la tercera forma normal de Codd con enfoque en una sola tabla de datos en lugar de muchas conectadas en bases de datos relacionales.

Una vez que importados los datos a R es conveniente limpiarlos, y como se ve el dato FECHA sería más conveniente que solo estuviera como año (2019) y no como formato con día y mes, ejemplo el día (11/10/2019), tal como se observa en la Figura 32:

**Figura 32**

*Tabla con datos fecha.*



```
> sedapal<- read_csv("sedapal.csv")
Parsed with column specification:
cols(
  DISTRITO = col_character(),
  TIPOFUGA = col_character(),
  CAUDAL = col_double(),
  FECHA = col_character()
)
> sedapal
# A tibble: 2,857 x 4
  DISTRITO      TIPOFUGA CAUDAL FECHA
  <chr>         <chr>    <dbl> <chr>
1 CARABAYLLO    TUBERIA  200000 11/10/2019
2 SANTIAGO DE SURCO TUBERIA  180000 17/05/2019
3 SAN BORJA     TUBERIA  150000 20/07/2017
4 MIRAFLORES    TUBERIA  150000 09/09/2017
5 LIMA (CERCADO) TUBERIA  150000 09/05/2018
6 RIMAC         TUBERIA  150000 22/08/2018
7 LIMA (CERCADO) CORP     120000 21/12/2017
8 CARABAYLLO    CORP     120000 19/03/2019
9 COMAS         TUBERIA  100000 15/02/2015
10 EL AGUSTINO   TUBERIA  100000 21/05/2015
# ... with 2,847 more rows
> |
```

```

sedapal <- fugas %>%
mutate(
  FECHA = parse_number(week),
  date = date.entered * (week - 1),
  rank = as.numeric(rank)
) %>%
select(-date.entered)

```

### Figura 33

*Tabla con datos fecha limpios.*



```

> sedapal<- read_csv("fugas.csv")
Parsed with column specification:
cols(
  DISTRITO = col_character(),
  TIPOFUGA = col_character(),
  CAUDAL = col_double(),
  FECHA = col_double()
)
> sedapal
# A tibble: 2,857 x 4
  DISTRITO      TIPOFUGA CAUDAL FECHA
  <chr>         <chr>    <dbl> <dbl>
1 CARABAYLLO   TUBERIA  200000  2019
2 SANTIAGO DE SURCO TUBERIA  180000  2019
3 SAN BORJA    TUBERIA  150000  2017
4 MIRAFLORES   TUBERIA  150000  2017
5 LIMA (CERCADO) TUBERIA  150000  2018
6 RIMAC        TUBERIA  150000  2018
7 LIMA (CERCADO) CORP     120000  2017
8 CARABAYLLO   CORP     120000  2019
9 COMAS        TUBERIA  100000  2015
10 EL AGUSTINO TUBERIA  100000  2015
# ... with 2,847 more rows
> |

```

## 4.8 Conversión de un archivo CSV a data.frame

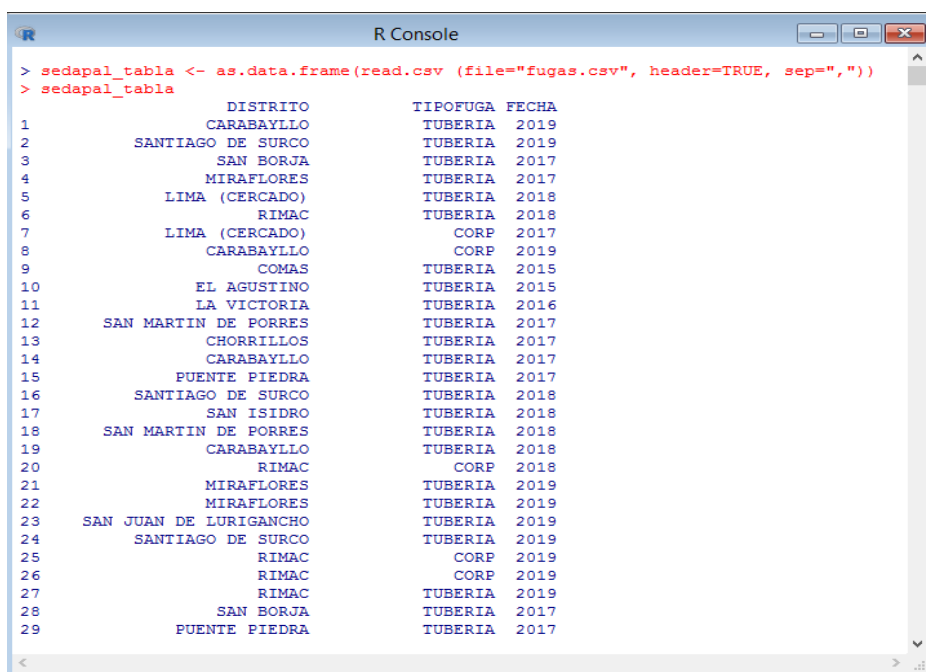
Un *data.frame* es una lista, cuyos componentes pueden ser vectores, matrices o factores, con la única salvedad de que las longitudes, o número de renglones, en el caso de matrices, deben coincidir en todos los componentes. La apariencia de un *data.frame* es la de una tabla y una forma de crearlos es mediante la función `data.frame()`.

Para este caso puntual es recomendable trabajar los datos como *data.frame* para poder ordenarlos y averiguar los niveles y demás información.

```
sedapal_tabla <- as.data.frame(read.csv  
(file="fugas.csv", header=TRUE, sep=","))
```

Figura 34

Tabla como *data.frame*.



```
> sedapal_tabla <- as.data.frame(read.csv (file="fugas.csv", header=TRUE, sep=","))  
> sedapal_tabla
```

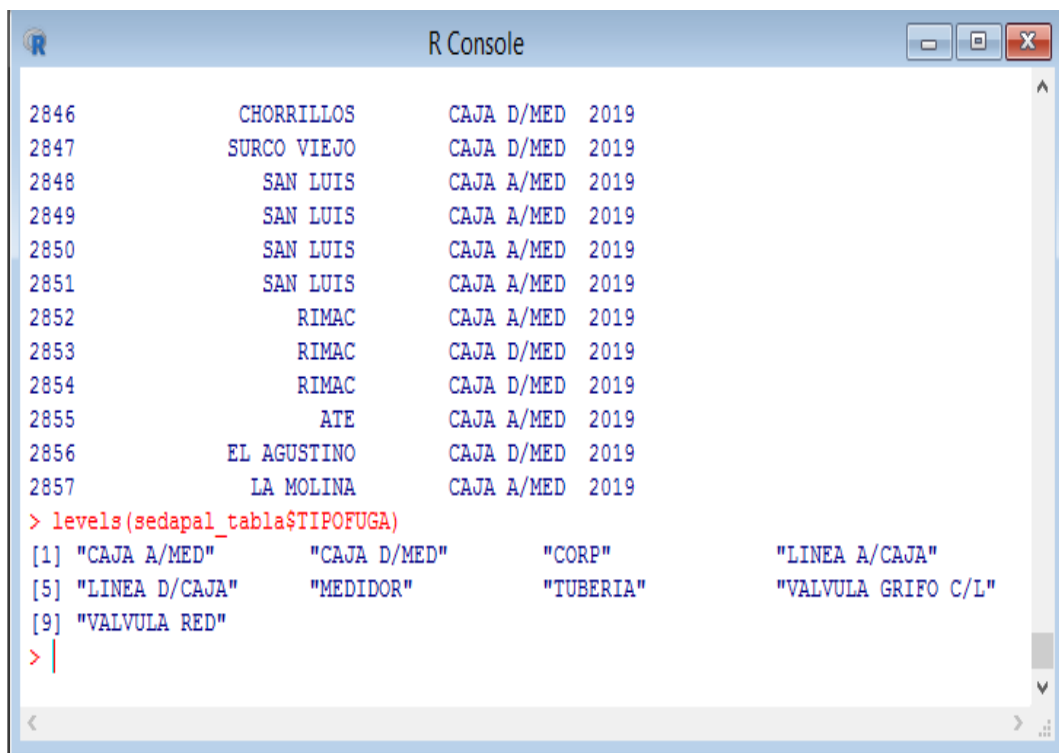
	DISTRITO	TIPOFUGA	FECHA
1	CARABAYLLO	TUBERIA	2019
2	SANTIAGO DE SURCO	TUBERIA	2019
3	SAN BORJA	TUBERIA	2017
4	MIRAFLORES	TUBERIA	2017
5	LIMA (CERCADO)	TUBERIA	2018
6	RIMAC	TUBERIA	2018
7	LIMA (CERCADO)	CORP	2017
8	CARABAYLLO	CORP	2019
9	COMAS	TUBERIA	2015
10	EL AGUSTINO	TUBERIA	2015
11	LA VICTORIA	TUBERIA	2016
12	SAN MARTIN DE PORRES	TUBERIA	2017
13	CHORRILLOS	TUBERIA	2017
14	CARABAYLLO	TUBERIA	2017
15	PUENTE PIEDRA	TUBERIA	2017
16	SANTIAGO DE SURCO	TUBERIA	2018
17	SAN ISIDRO	TUBERIA	2018
18	SAN MARTIN DE PORRES	TUBERIA	2018
19	CARABAYLLO	TUBERIA	2018
20	RIMAC	CORP	2018
21	MIRAFLORES	TUBERIA	2019
22	MIRAFLORES	TUBERIA	2019
23	SAN JUAN DE LURIGANCHO	TUBERIA	2019
24	SANTIAGO DE SURCO	TUBERIA	2019
25	RIMAC	CORP	2019
26	RIMAC	CORP	2019
27	RIMAC	TUBERIA	2019
28	SAN BORJA	TUBERIA	2017
29	PUENTE PIEDRA	TUBERIA	2017

Ya se tiene el CSV como un *data.frame* y se puede observar lo que son los niveles (valores) de un campo *CHAR* como TIPOFUGA.

```
levels(sedapal_tabla$TIPOFUGA)
```

Figura 35

LEVELS TIPOFUGA.



The screenshot shows an R console window with the following content:

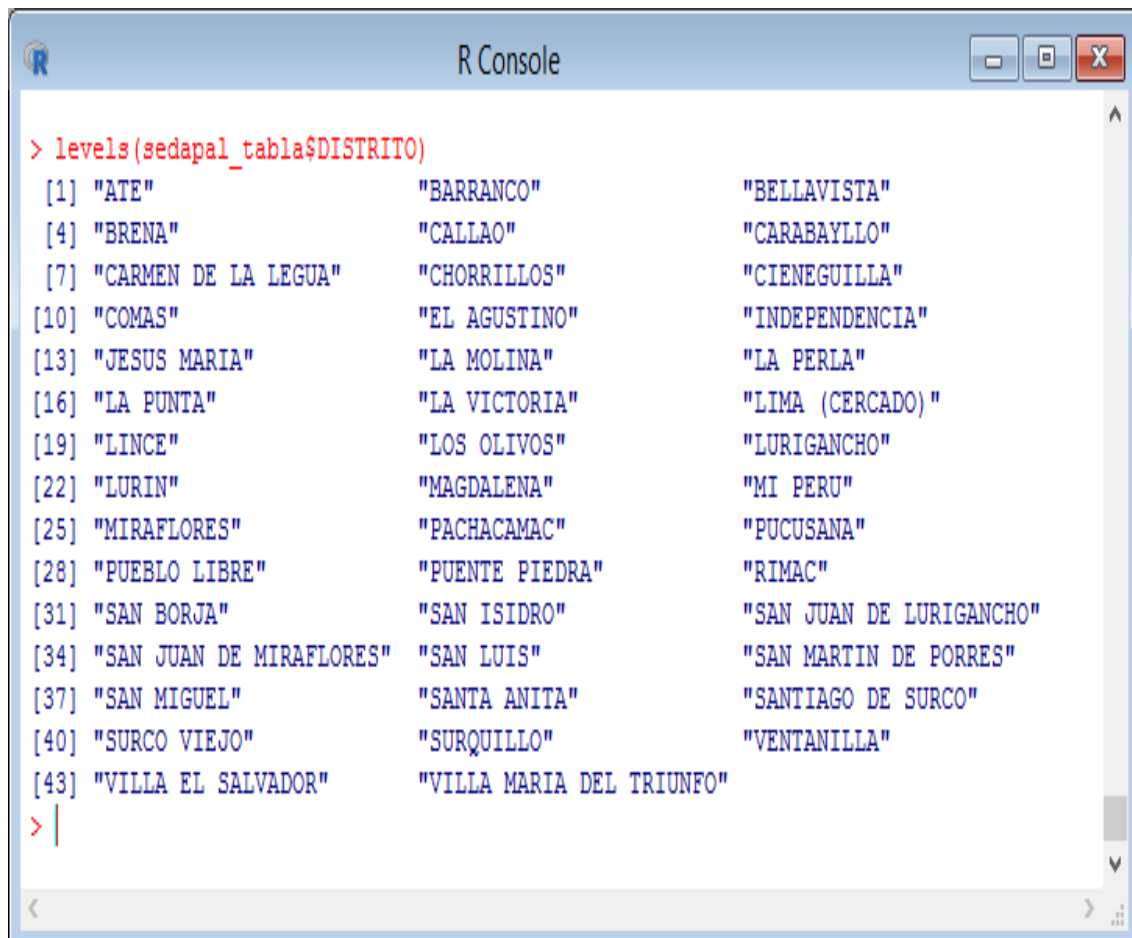
```
2846      CHORRILLOS      CAJA D/MED  2019
2847      SURCO VIEJO     CAJA D/MED  2019
2848      SAN LUIS        CAJA A/MED  2019
2849      SAN LUIS        CAJA A/MED  2019
2850      SAN LUIS        CAJA A/MED  2019
2851      SAN LUIS        CAJA A/MED  2019
2852      RIMAC          CAJA A/MED  2019
2853      RIMAC          CAJA D/MED  2019
2854      RIMAC          CAJA D/MED  2019
2855      ATE            CAJA A/MED  2019
2856      EL AGUSTINO     CAJA D/MED  2019
2857      LA MOLINA       CAJA A/MED  2019
> levels(sedapal_tabla$TIPOFUGA)
[1] "CAJA A/MED"      "CAJA D/MED"      "CORP"             "LINEA A/CAJA"
[5] "LINEA D/CAJA"    "MEDIDOR"         "TUBERIA"          "VALVULA GRIFO C/L"
[9] "VALVULA RED"
> |
```

Por ejemplo, para el campo DISTRITO

```
levels(sedapal_tabla$DISTRITO)
```

**Figura 36**

*LEVELS TIPOFUGA.*



```
> levels(sedapal_tabla$DISTRITO)
[1] "ATE" "BARRANCO" "BELLAVISTA"
[4] "BRENA" "CALLAO" "CARABAYLLO"
[7] "CARMEN DE LA LEGUA" "CHORRILLOS" "CIENEGUILLA"
[10] "COMAS" "EL AGUSTINO" "INDEPENDENCIA"
[13] "JESUS MARIA" "LA MOLINA" "LA PERLA"
[16] "LA PUNTA" "LA VICTORIA" "LIMA (CERCADO)"
[19] "LINCE" "LOS OLIVOS" "LURIGANCHO"
[22] "LURIN" "MAGDALENA" "MI PERU"
[25] "MIRAFLORES" "PACHACAMAC" "PUCUSANA"
[28] "PUEBLO LIBRE" "PUENTE PIEDRA" "RIMAC"
[31] "SAN BORJA" "SAN ISIDRO" "SAN JUAN DE LURIGANCHO"
[34] "SAN JUAN DE MIRAFLORES" "SAN LUIS" "SAN MARTIN DE PORRES"
[37] "SAN MIGUEL" "SANTA ANITA" "SANTIAGO DE SURCO"
[40] "SURCO VIEJO" "SURQUILLO" "VENTANILLA"
[43] "VILLA EL SALVADOR" "VILLA MARIA DEL TRIUNFO"
> |
```

## 4.9 Tratamiento de los datos usando ggplot

### a) Matriz de gráficos de dispersión

Una matriz de gráficos de dispersión es una cuadrícula (o matriz) de gráficos de dispersión que se utiliza para visualizar relaciones bivariantes entre combinaciones de variables. Cada gráfico de dispersión de la matriz muestra la relación entre un par de variables, lo que permite explorar muchas relaciones en un solo gráfico.

Una matriz de gráficos de dispersión se compone de una cuadrícula de pequeños gráficos y un gráfico de vista previa más grande que muestra un gráfico pequeño seleccionado con mayor detalle. Además, se puede agregar a la matriz un histograma que muestra la distribución de cada variable numérica.

Se ingresa el código en R para visualizar la matriz digitando el siguiente algoritmo en la ventana de comandos.

```
library(AppliedPredictiveModeling)  
transparentTheme(trans = .4)  
library(caret)  
featurePlot(x = sedapal_tabla[, 1:3],  
  
          y = sedapal_tabla$TIPOFUGA,  
          plot = "pairs",  
          ## Add a key at the top  
          auto.key = list(columns = 3))
```



Como se puede apreciar se tiene ya un primer acercamiento a una tendencia utilizando la matriz de dispersión entre 3 variables como son TIPOFUGA, DISTRITO y FECHA ya que se puede observar una concentración de datos para el caso de TIPOFUGA en 2 de sus variables TUBERIA y CORP notándose un incremento de estas en el 2019. Por tanto, el software R permite agrupar en este caso datos CUALITATIVOS y NUMERICOS y mostrar la tendencia de un modelo predictivo a partir de las gráficas obtenidas.

El modelo predictivo debe indicar la cantidad de datos dispersos y sin relación varios parámetros a saber, tales como:

- Distritos con más incidencia de emergencias reportadas.
- Tipos de Fugas más incidencia de emergencias reportadas.
- Variación de los datos en función de una fecha específica.
- Relacionar estas variables en un solo cuadro.
- Relacionar las variables entre ellas.

Para ellos se utilizará la información con la que se dispone para hallar una relación y un modelo predictivo, teniendo en cuenta que estos datos no son necesariamente concluyentes y que a mayor cantidad de datos se obtienen modelos y tendencias más precisos. Luego se procede a compararlos con los datos de control que no se han procesado del periodo 2020 para ver su eficacia.



Cabe aclarar que se fundamenta esta investigación en datos reales obtenidos en un periodo de 7 años y con algunas variaciones en la tomas de muestras debido a que se utilizaron cada vez más personal para ello (vehículos) y que esto en el futuro puede seguir aumentando.

Como se indicó, el servicio inició con 4 vehículos y los primeros registros no son datos completamente homogéneos y luego entre los años 2017, 2018 y 2019 ya con 6 vehículos los datos se hicieron más fiables y aumentaron en número la cantidad de muestras, en una forma dramática como se aprecia en los resultados.

En los últimos años estamos viviendo una gran explosión de aplicaciones y servicios que giran alrededor del big data y los sistemas predictivos. Estamos rodeados de sistemas que, apoyándose en algoritmos, son capaces de procesar grandes volúmenes de información, hasta el punto de realizar predicciones o, directamente, indicarnos qué tenemos que hacer ante determinada situación.

Lo que se intenta con esta investigación es a partir de los datos obtenidos profundizar en modelos predictivos aplicados a no solo este caso en particular sino a muchos otros, donde se tengan información la cual se requiere analizar y encontrar un patrón que permita su modelamiento matemático.

**Figura 38**

*Grafica del modelo predictivo.*



Se dispone de una gran cantidad de datos por analizar desde diversos frentes ya que se cuenta con 3 variables 2 de ellas son datos cualitativos y uno cuantitativo, es por ello que se le da un enfoque tratando de encontrar la relación que mejor defina la tendencia de la información con la que se dispone. Nuestro algoritmo se aplica al tratamiento de la información y mostrarla en forma gráfica para su mejor comprensión y predicción de la tendencia, haciendo uso de las herramientas y librerías que el software R nos brinda.

#### **4.10 Procesamiento de la información mediante algoritmos**

##### **a) Relación entre variables DISTRITO Y FECHA**

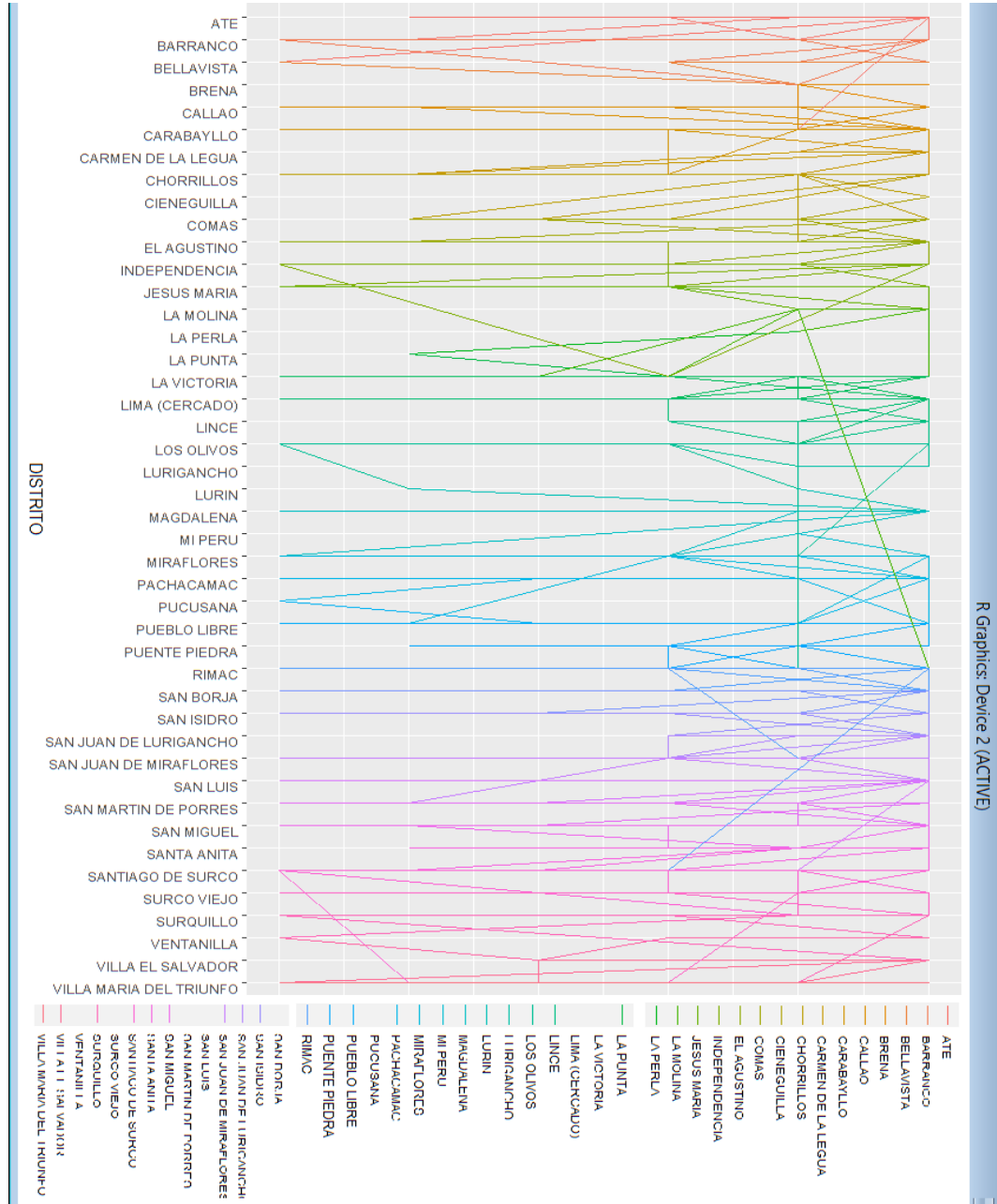
```
ggplot(sedapal, aes(x = DISTRITO, y = FECHA, color =  
DISTRITO, group = TIPOFUGA)) + geom_line() +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

Al ejecutar nuestro algoritmo, como se puede apreciar, este tipo de dato obtenido no aporta mucha información sobre las tendencias de las fugas en función de los distritos y de los años, se ha aplicado un tipo de gráfico del tipo polilínea el cual no es por lo visto el más adecuado para nuestro modelo en cuestión, es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar .

Como se observa tenemos en el eje **Y** los años desde el 2014 al 2019 y en el eje **X** la relación de distritos en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

**Figura 39**

*Datos obtenidos modelo del tipo polilineal.*



## Datos obtenidos modelo del tipo Polilineal

Como se puede observar, hay que darle un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.

```
ggplot(data=sedapal_tabla, aes(x=reorder  
  
(FECHA,DISTRITO), y=1, fill=DISTRITO)) +  
  geom_bar(stat="identity", position="stack")
```

Al ejecutar nuestro algoritmo, se puede apreciar, que este tipo de dato obtenido aporta algo de información sobre las tendencias de las fugas en función de los distritos y de los años, se aplicó un modelo de gráfico del tipo barras el cual es por lo visto más adecuado para nuestro modelo en cuestión; es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se aprecia, existe un aumento en las fugas desde los años 2017 al 2019 y en el eje **X** la relación de distritos en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible pero se debe modelar aún más para encontrar la tendencia en los datos obtenidos para utilizarlo como modelo.



```
ggplot(sedapal, aes(x = DISTRITO, y = 1, group =  
DISTRITO)) +  
facet_wrap(~ FECHA, nrow = 1) +  
geom_bar(stat = "identity", fill = "darkgray") +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

Al ejecutar nuestro algoritmo, se puede apreciar, que este tipo de dato obtenido aporta algo de información sobre las tendencias de las fugas en función de los distritos y de los años, se aplicó un modelo de gráfico del tipo barras el cual es por lo visto más adecuado para nuestro modelo en cuestión, es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se aprecia, hay un aumento en las fugas en los años desde el 2017 al 2019 y en el eje X la relación de distritos en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible, pero se debe modelar aún más para encontrar la tendencia en los datos obtenidos para utilizarlo como modelo.





**theme(axis.text.x = element\_text(angle = 90, hjust = 1))**

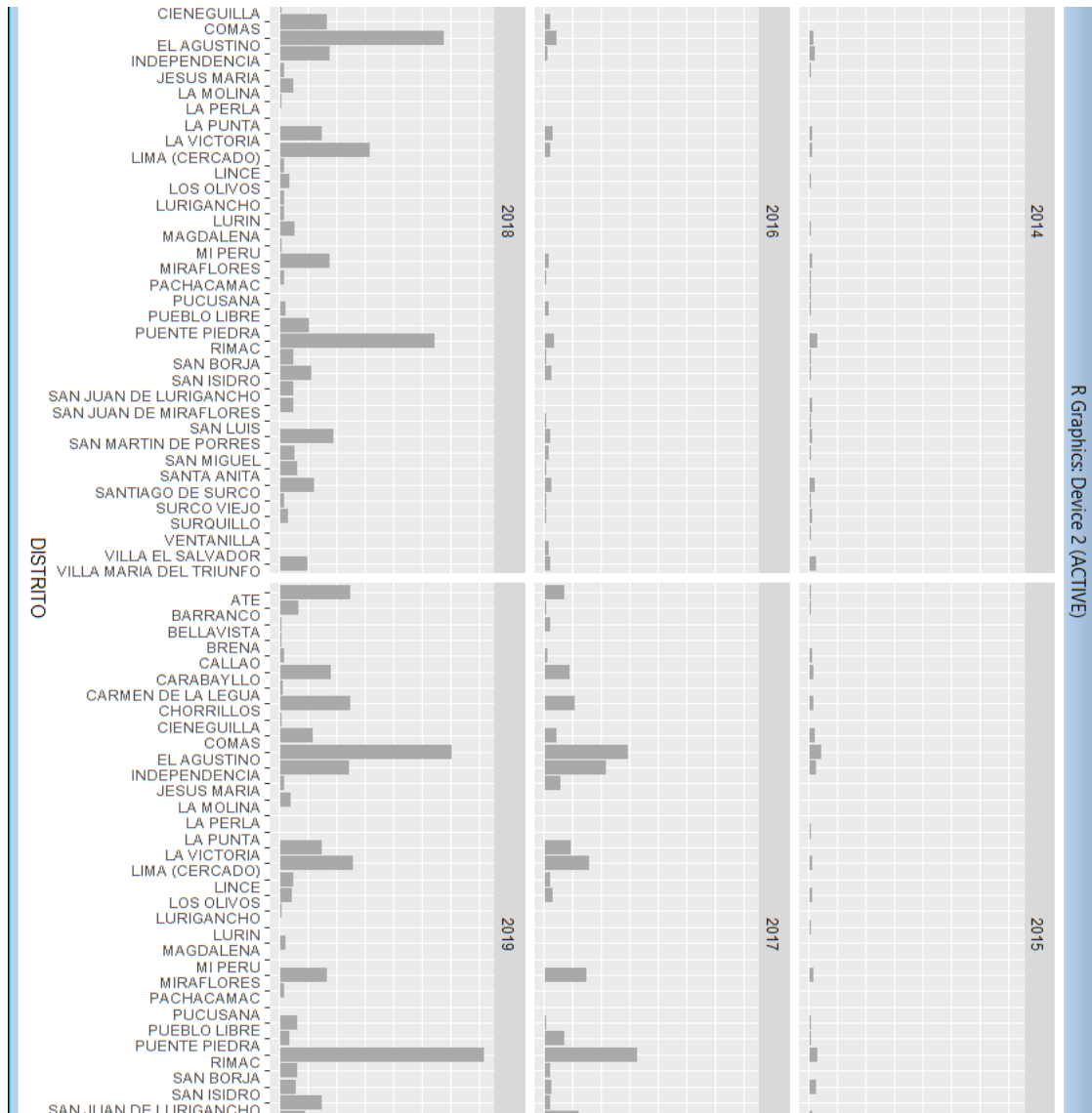
Al ejecutar nuestro algoritmo modificado, como se puede apreciar, este tipo de gráfico obtenido da mucha más información al mostrarla en formato de columnas superpuestas mostrando distritos y los años, se aplicó el gráfico tipo barras el cual es por lo visto, el más adecuado para nuestro modelo en cuestión; es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se aprecia, existe un aumento en las fugas en los años desde el 2017 al 2019 y en el eje **X** la relación de distritos en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible y se nota la tendencia al crecimiento de las fugas según el año y también ya es posible apreciar qué distritos son los que más servicios de fugas de emergencia presentan.

**Figura 42**

*Datos obtenidos modelo del tipo barras.*



Se dispone con una gran cantidad de datos por analizar desde diversos frentes ya que se cuenta con 3 variables, 2 de ellas son datos cualitativos y uno cuantitativo, es por ello que se le da un enfoque tratando de encontrar la relación que mejor defina la tendencia de la información con la que se tiene.

Nuestro algoritmo se aplica al tratamiento de la información y mostrarla en forma gráfica para su mejor comprensión y predicción de la tendencia, haciendo uso de las herramientas y librerías que el software R nos brinda.

## **b) Relación entre variables DISTRITO Y TIPOFUGA**

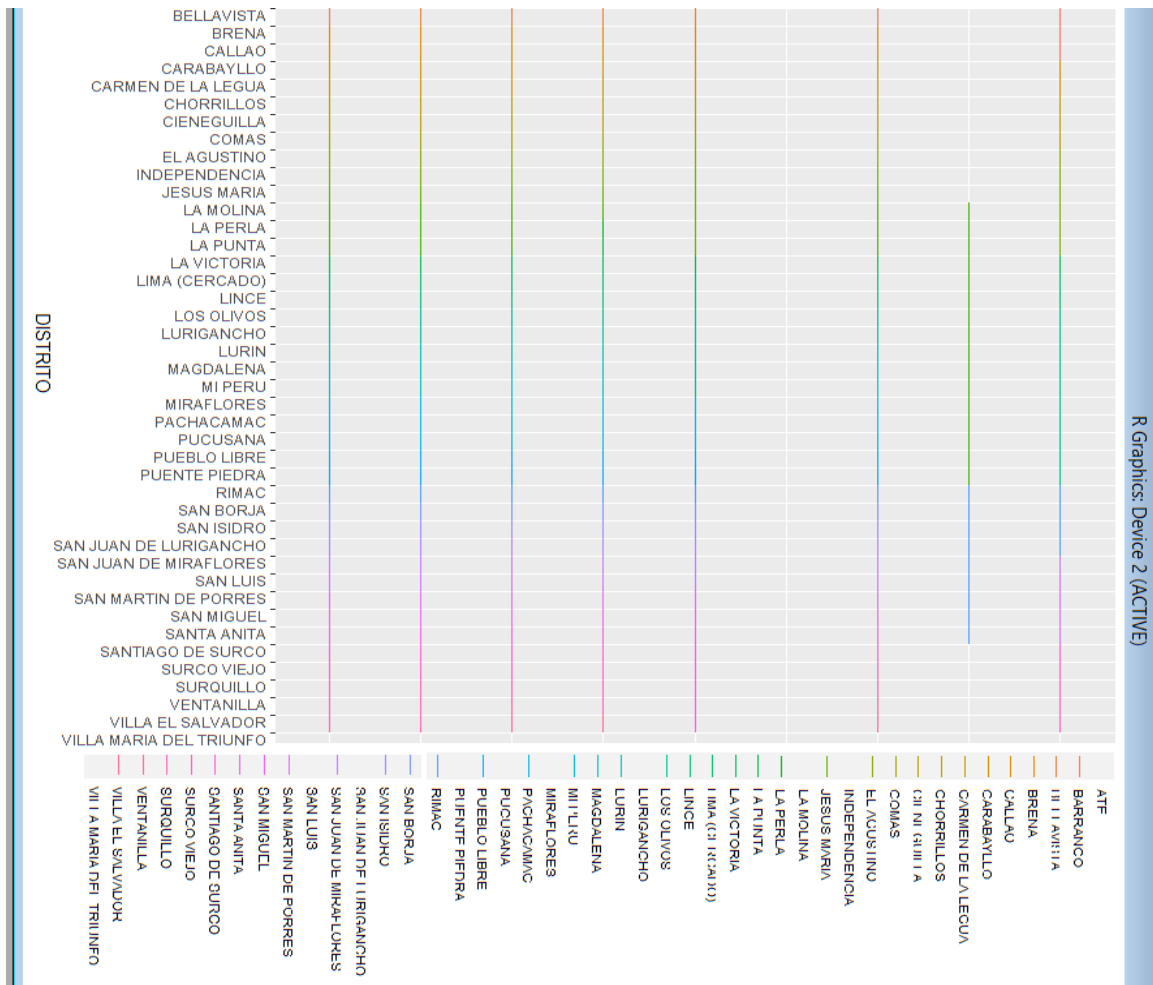
```
ggplot(sedapal, aes(x = DISTRITO, y = TIPOFUGA, color =  
DISTRITO, group = TIPOFUGA)) + geom_line() +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

Al ejecutar nuestro algoritmo, como se puede apreciar, este tipo de dato obtenido no aporta mucha información sobre las tendencias de las fugas en función de los distritos y de los años, se aplicó un tipo de gráfico del tipo polilínea el cual no es por lo visto el más adecuado para nuestro modelo en cuestión, es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar .

Como se observa, tenemos en el eje **Y** los tipos de fugas y en el eje **X** la relación de distritos en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

**Figura 43**

*Datos obtenidos modelo del tipo Polilineal.*



Como se aprecia, se debe dar un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.

```
ggplot(data=sedapal_tabla, aes(x=reorder
(TIPOFUGA,DISTRITO), y=1, fill=DISTRITO)) +
  geom_bar(stat="identity", position="stack")
```

Al ejecutar nuestro algoritmo, como se puede apreciar este tipo de dato obtenido aporta algo de información sobre las tendencias de las fugas en función de los distritos y de los años, se aplicó un modelo de gráfico del tipo barras el cual es por lo visto el más adecuado para nuestro modelo en cuestión; es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

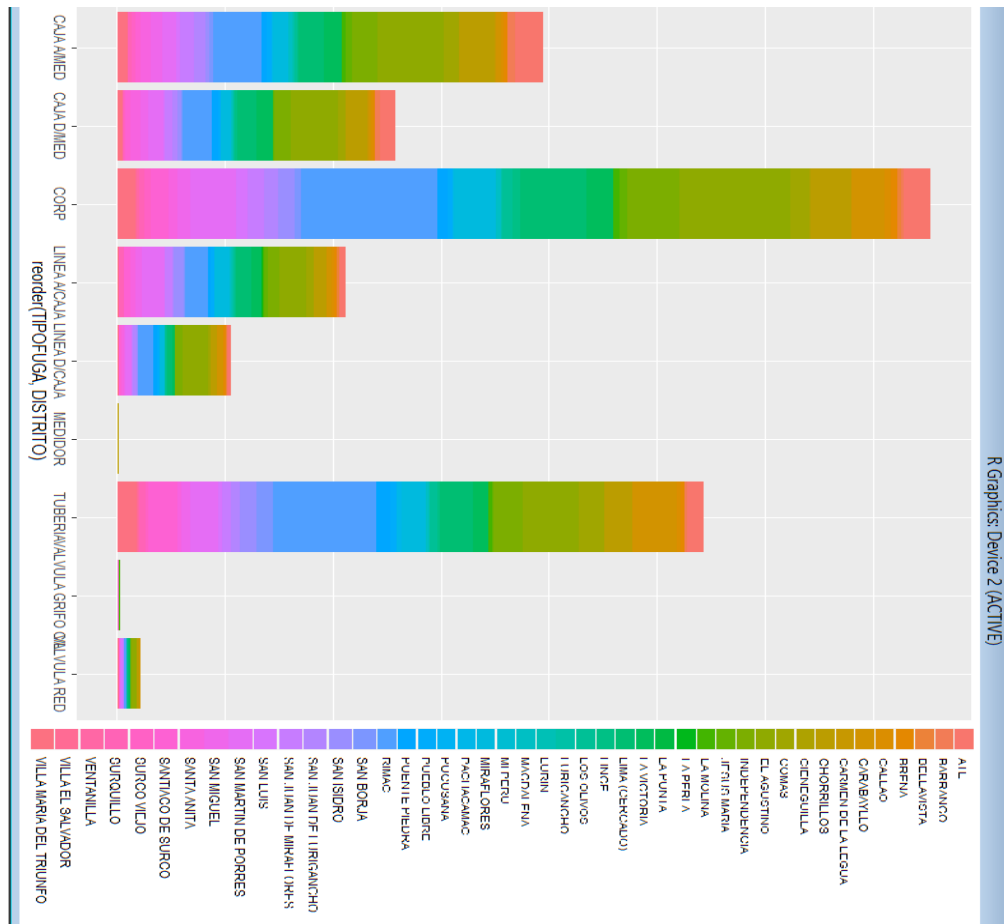
Como se aprecia, tenemos los distintos tipos de fugas encontradas en las emergencias y distritos en el eje **X** las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible, pero se debe detallar aún más para encontrar la tendencia en los datos obtenidos para utilizarlo como modelo.

Se hace uso del comando **reorder** para ordenar los datos y también está utilizando colores para diferenciar los distritos, sin embargo no se logra aun información concluyente.

**Figura 44**

Datos obtenidos modelo del tipo barras.



Como se observa, se debe dar un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.

```
ggplot(sedapal, aes(x = DISTRITO, y = 1, group =  
DISTRITO)) +  
facet_wrap(~ TIPOFUGA, nrow = 1) +  
geom_bar(stat = "identity", fill = "darkgray") +
```

```
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

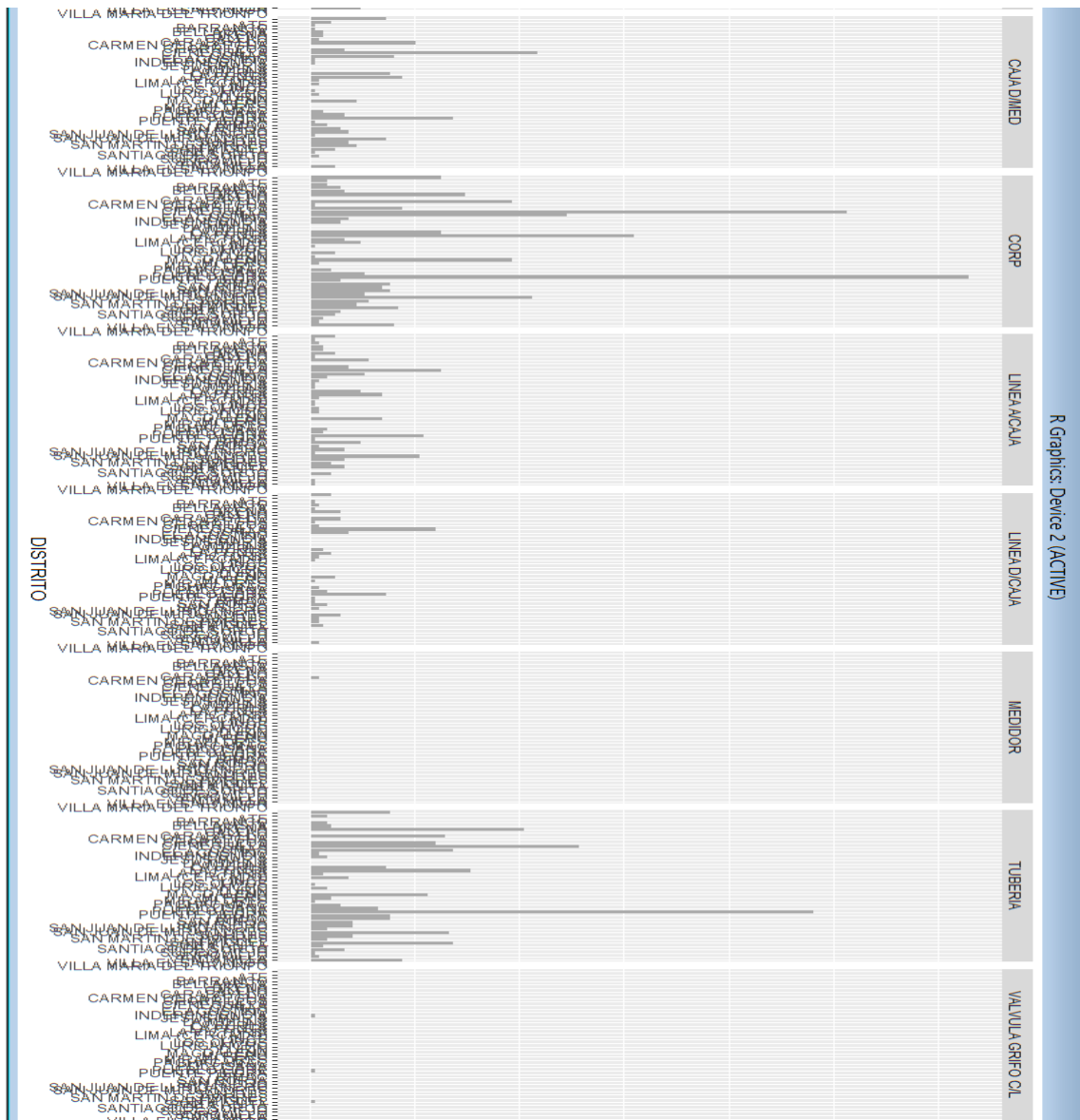
Al ejecutar nuestro algoritmo, como se puede apreciar este tipo de dato obtenido aporta algo de información sobre las tendencias de las fugas en función de los distritos y de los años, se aplicó un modelo de gráfico del tipo barras el cual es por lo visto el más adecuado para nuestro modelo en cuestión, es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se aprecia, tenemos los distintos tipos de fugas encontradas en las emergencias y distritos en el eje X las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible, pero se debe detallar aún más para encontrar la tendencia en los datos obtenidos para utilizarlo como modelo.

Figura 45

Datos obtenidos modelo del tipo barras.



Como se observa, tenemos que darle un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.



```
ggplot(sedapal, aes(x = DISTRITO, y = 1, group =  
DISTRITO)) +  
facet_wrap(~ TIPOFUGA, nrow = 3) +  
geom_bar(stat = "identity", fill = "darkgray") +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

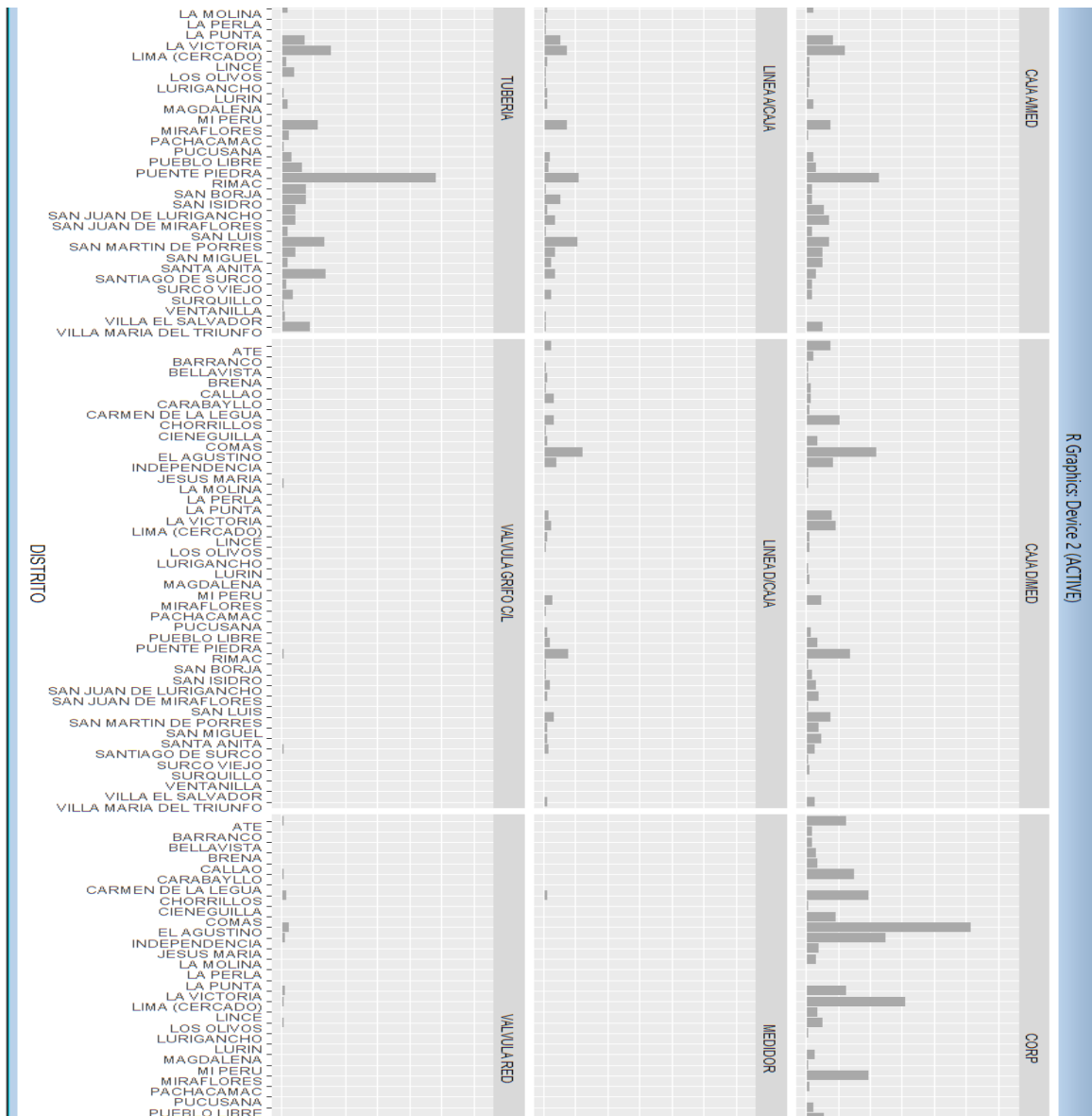
Al ejecutar nuestro algoritmo modificado, como se puede apreciar este tipo de gráfico obtenido da mucha más información al mostrarla en formato de columnas superpuestas mostrando distritos y los tipos de fugas, se aplicó el gráfico tipo barras el cual es por lo visto, el más adecuado para nuestro modelo en cuestión, es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se aprecia, tenemos fugas con mucha incidencia y algunos tipos son casi inexistentes y en el eje X la relación de distritos en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible y se nota la tendencia al crecimiento de las fugas y también ya es posible apreciar qué distritos son los que más servicios de fugas de emergencia presentan.

**Figura 46**

*Datos obtenidos modelo del tipo barras.*



Se dispone de una gran cantidad de datos por analizar desde diversos frentes ya que se cuenta con 3 variables, 2 de ellas son datos cualitativos y uno cuantitativo, es por ello que se le da un enfoque tratando de encontrar la relación que mejor defina la tendencia de la información con la que se tiene.

Nuestro algoritmo se aplica al tratamiento de la información y mostrarla en forma gráfica para su mejor comprensión y predicción de la tendencia, haciendo uso de las herramientas y librerías que el software R nos brinda.

**d) Relación entre variables FECHA Y TIPOFUGA**

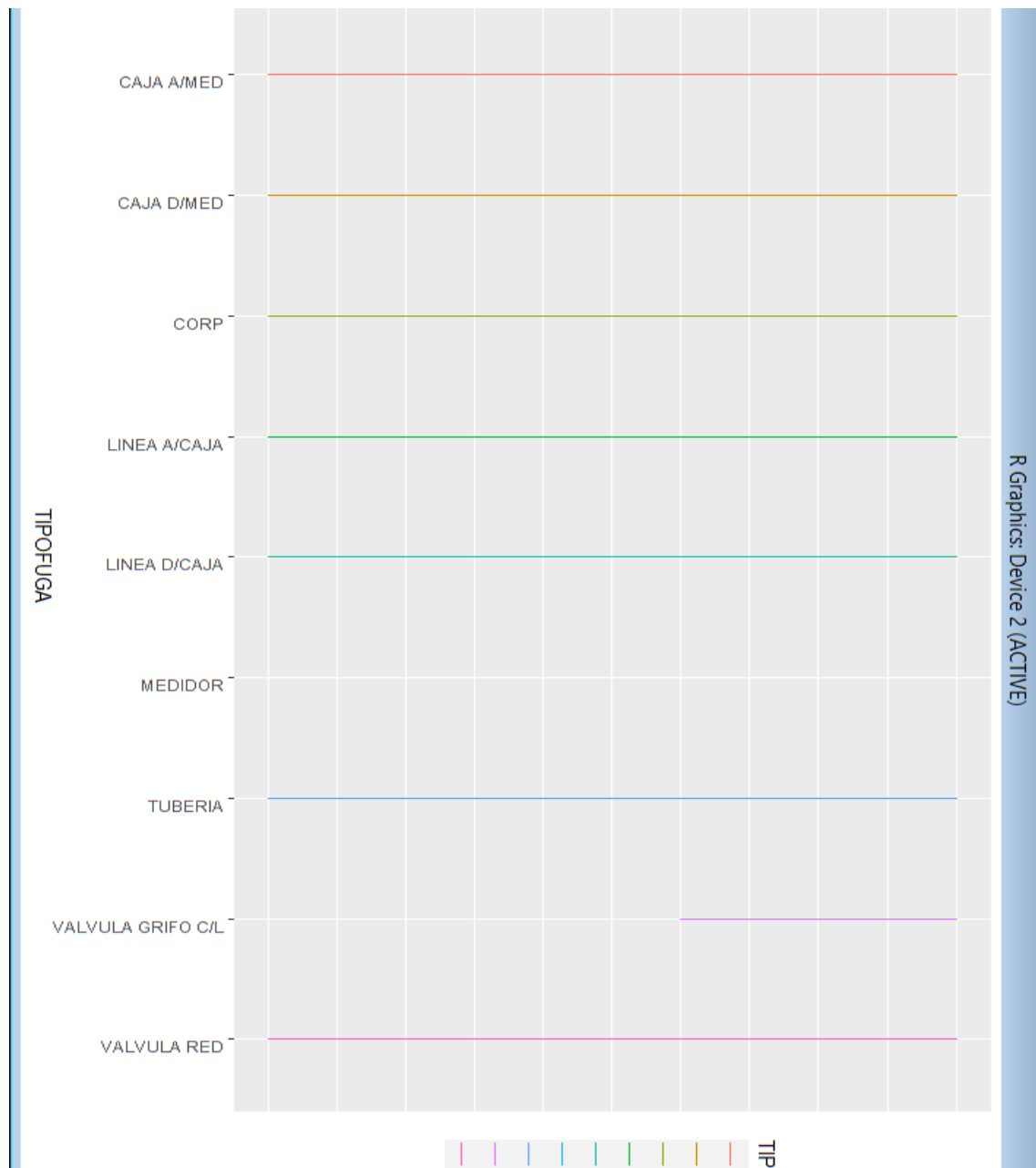
```
ggplot(sedapal, aes(x = TIPOFUGA, y = FECHA, color =  
TIPOFUGA, group = TIPOFUGA)) + geom_line() +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

Al ejecutar nuestro algoritmo, como se puede apreciar este tipo de dato obtenido no aporta mucha información sobre las tendencias de las fugas en función de los tipos de fugas y de los años, se aplicó un modelo de gráfico del tipo polilínea el cual no es por lo visto el más adecuado para nuestro modelo en cuestión, es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar .

Como se puede observar se tiene en el eje Y los tipos años y en el eje X los tipos de fugas los cuales se presentaron de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

**Figura 47**

*Datos obtenidos modelo del tipo Polilineal*



Como se puede observar es importante darle un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.

```
ggplot(data=sedapal_tabla, aes(x=reorder  
  
(TIPOFUGA,FECHA), y=1, fill=FECHA)) +  
  geom_bar(stat="identity", position="stack")
```

Al ejecutar nuestro algoritmo, como se puede apreciar, este tipo de dato obtenido aporta algo de información sobre las tendencias de las fugas en función de los años, se aplicó un tipo de gráfico del modelo de barras el cual es por lo visto más adecuado para nuestro modelo en cuestión; es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

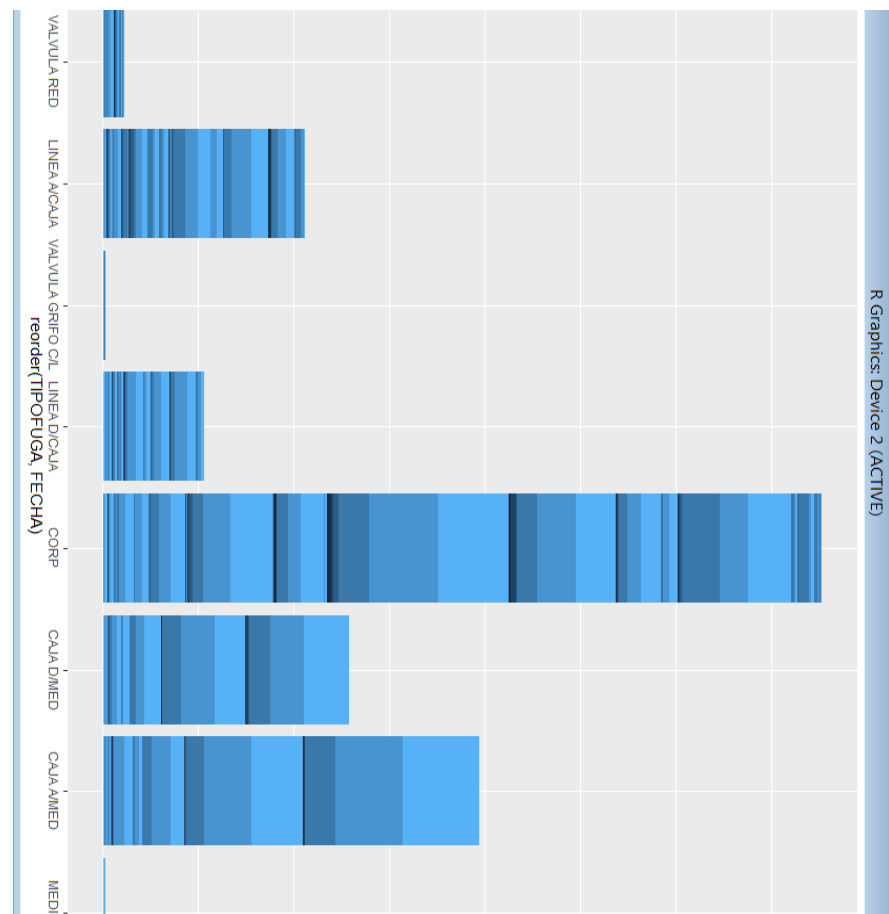
Como se aprecia, tenemos los distintos tipos de fugas encontradas en las emergencias en el eje **X** las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible pero se debe detallar aún más para encontrar la tendencia en los datos obtenidos para utilizarlo como modelo.

Se procede a utilizar el comando **reorder** para ordenar los datos y también se utilizan colores para diferenciar los distritos, sin embargo no se logró aun información concluyente.

**Figura 48**

*Datos obtenidos modelo del tipo barras.*



Como se puede observar hay que darle un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.

```
ggplot(sedapal, aes(x = FECHA, y = 1, group =  
FECHA)) +  
facet_wrap(~ TIPOFUGA, nrow = 1) +  
geom_bar(stat = "identity", fill = "darkgray") +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

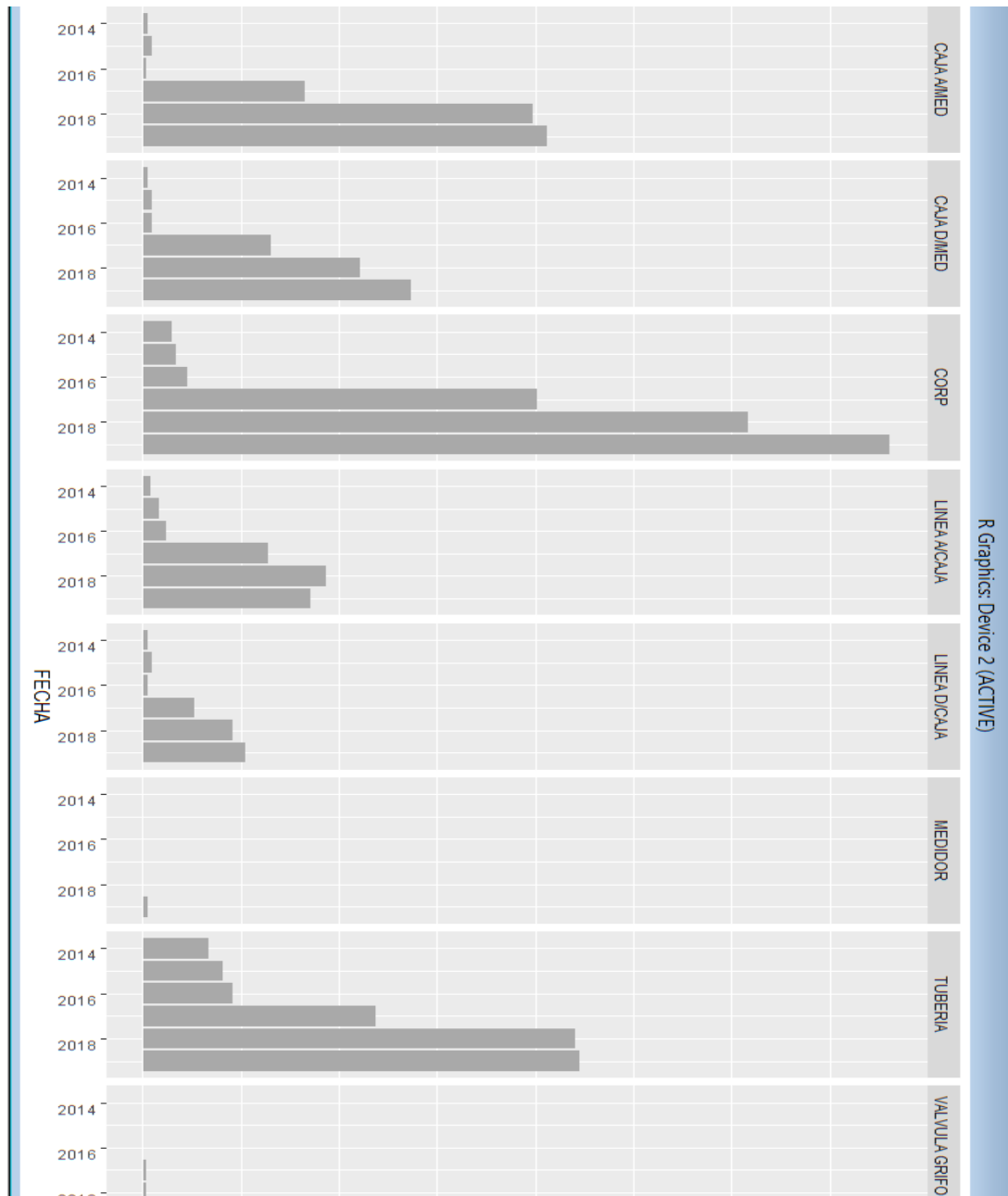
Al ejecutar nuestro algoritmo, como se puede apreciar, este tipo de dato obtenido aporta algo de información sobre las tendencias de las fugas en función de los distritos y de los años, se aplicó un modelo de gráfico del tipo barras el cual es por lo visto el más adecuado para nuestro modelo en cuestión; es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se puede apreciar tenemos los distintos tipos de fugas encontradas en las emergencias y distritos en el eje **X** las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible, pero se debe detallar aún más para encontrar la tendencia en los datos obtenidos para utilizarlo como modelo.

**Figura 49**

*Datos obtenidos modelo del tipo barras.*



Como se puede observar hay que darle un enfoque diferente al análisis empezando por cómo ordenar las variables y cómo graficarlas de una mejor manera.



```
ggplot(sedapal, aes(x = FECHA, y = 1, group =  
FECHA)) +  
facet_wrap(~ TIPOFUGA, nrow = 3) +  
geom_bar(stat = "identity", fill = "darkgray") +  
theme(axis.text.x = element_text(angle = 90, hjust =  
1))
```

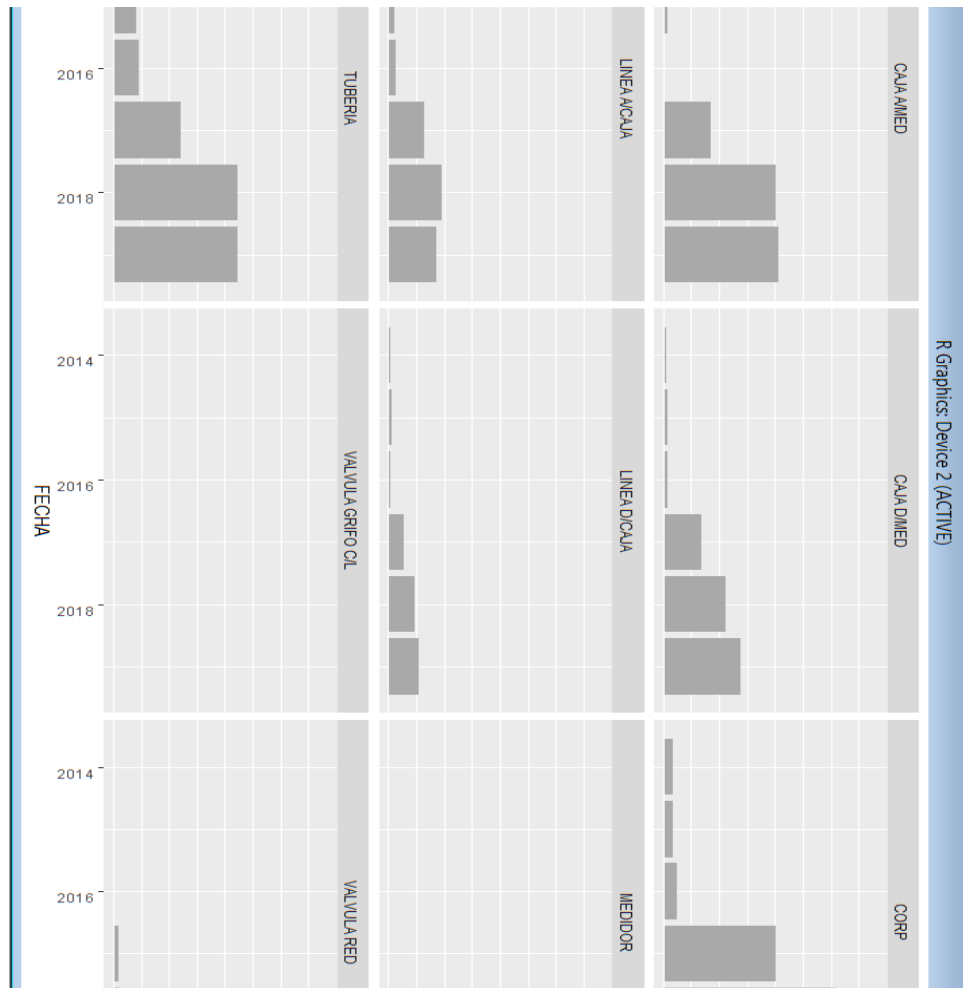
Al ejecutar nuestro algoritmo modificado, como se puede apreciar, este tipo de gráfico obtenido da mucha más información al mostrarla en formato de columnas superpuestas mostrando años y los tipos de fugas, se aplicó el modelo de barras el cual es por lo visto, el más adecuado para nuestro modelo en cuestión; es muy importante tener en cuenta el tipo de gráfico a utilizar, así como también el tipo de datos que se va a analizar.

Como se observa, tenemos fugas con mucha incidencia y algunos tipos son casi inexistentes y en el eje X la relación de años en los cuales se presentaron las fugas de agua potable de emergencia (tubería, corporation, fuga en caja, etc.) las cuales forman parte de esta base de datos 2014 -2019.

La información así presentada es un poco más legible y se nota la tendencia al crecimiento de las fugas y también ya es posible apreciar que distritos son los que más servicios de fugas de emergencia presentan.

**Figura 50**

*Datos obtenidos modelo del tipo barras.*

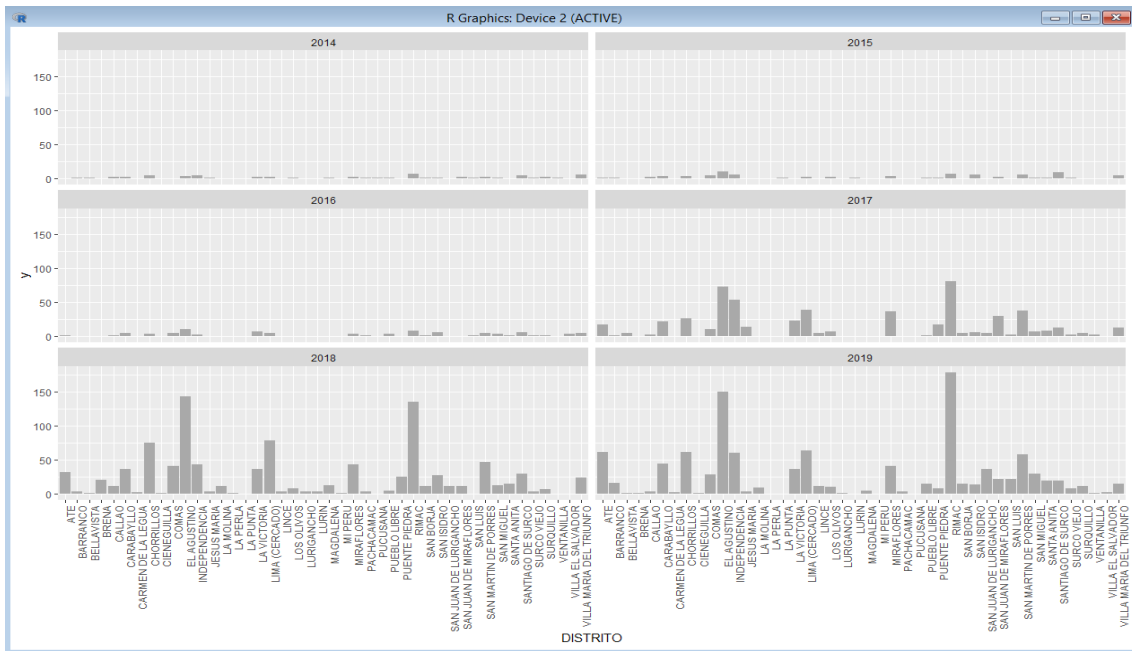


La información así presentada es un poco más legible y se nota la tendencia al crecimiento de las fugas y también ya es posible apreciar que distritos y en que años se dieron más estas emergencia, por lo que se utilizará este algoritmo como fuente para hacer la proyección estadística general.

```
ggplot(sedapal, aes(x = CAMPO1, y = 1, group =
CAMPO1)) +
facet_wrap(~ CAMPO2, nrow = 3) +
geom_bar(stat = "identity", fill = "darkgray") +
theme(axis.text.x = element_text(angle = 90, hjust =
1))
```

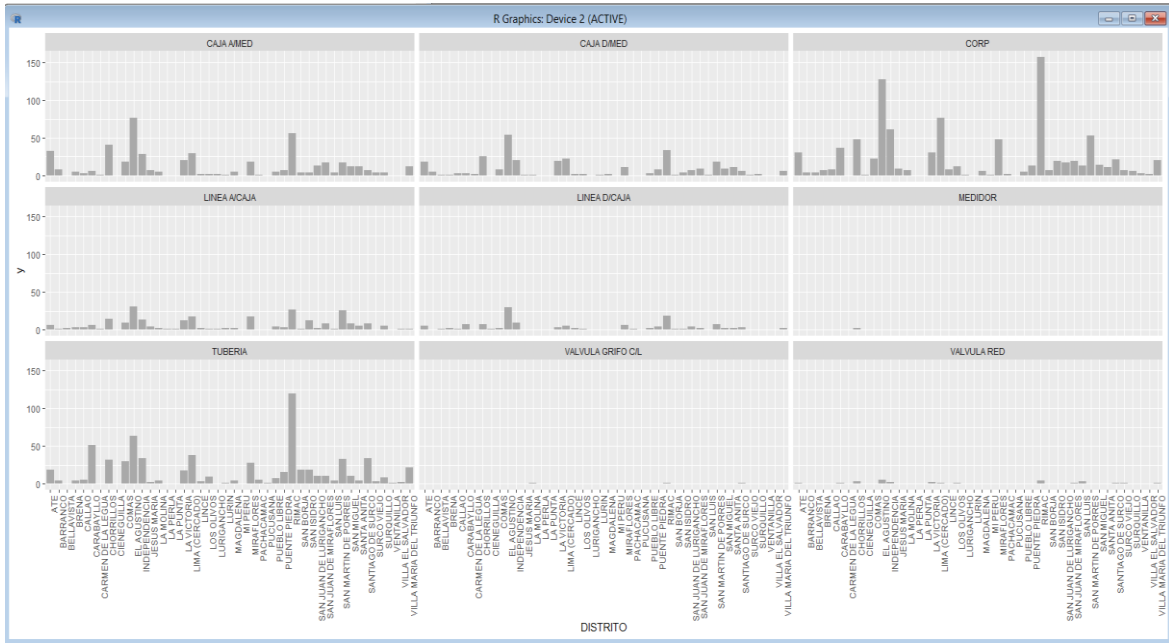
Figura 51

Modelo del tipo barras: Distritos & Fecha.



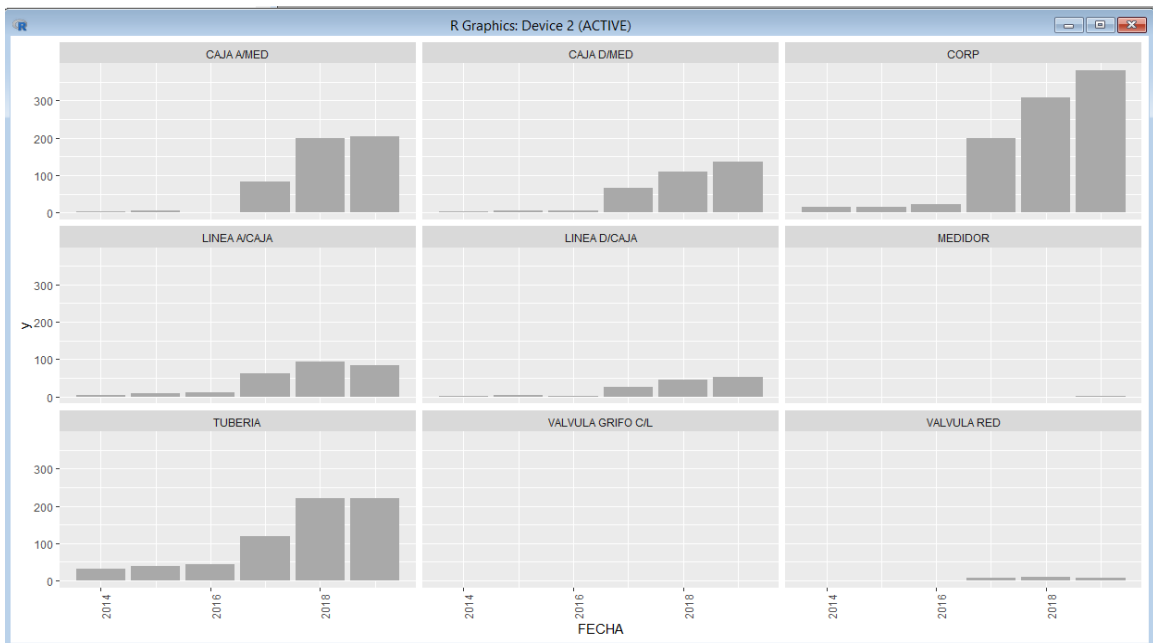
**Figura 52**

*Modelo del tipo barras: Distritos & Tipofugas.*



**Figura 53**

*Modelo del tipo barras: Tipofugas & Fecha*



#### **4.11. Población y muestra utilizada para la simulación**

En esta investigación predominantemente tecnológica, transversal, intitulada: “Diseño de un algoritmo predictivo para monitoreo temprano de redes de agua potable en la ciudad de Lima, 2019” es importante delimitar sus alcances para su análisis correspondiente.

##### **4.11.1 Unidad de análisis**

La delimitación de la Unidad de Análisis del problema objeto de investigación, para diseño de un algoritmo predictivo para monitoreo temprano de redes de agua potable en la ciudad de Lima, se ha determinado geográficamente como “unidad de análisis” a la ciudad de Lima en el periodo comprendido entre enero del 2014 a diciembre del 2019.

##### **4.11.2 Población escogida de trabajo**

En esta investigación se considera como población a las fugas de agua potable detectadas ya sean del tipo: industriales, comerciales y residenciales, ubicados en la ciudad de Lima.

De esta gran muestra que son aproximadamente 8700 fugas detectadas e ingresadas al sistema SGIOC se procesaron algo más de 2500 fugas bajo la premisa de una presión de agua constante de entre 30 y 40 libras aproximadamente.

#### **4.11.3 Tamaño de la muestra trabajado**

Para determinar probabilísticamente el “Tamaño de la Muestra” es necesario definir la característica principal de la población constituido por las fugas de agua potable detectadas en ese periodo de tiempo, que corresponde a una población “finita”, de ( $N = 2500$ ) que es menor igual a el tamaño de la muestra de 8700 ( $N \leq 8700$ ).

#### **4.11.4 Prueba piloto planteada**

Debido a que no existen antecedentes de estudio correspondiente a esta investigación para determinar el tamaño de la muestra, se ha optado por descargar los datos obtenidos durante ese periodo de tiempo (5 años) y procesarlos para verificar qué porcentaje de las fugas seleccionadas fueron correctamente procesadas y cuáles no y los tipos de fugas.

El procesamiento arrojó que solo el 60% de las fugas fue detectado y catalogado correctamente en ese periodo de tiempo.

#### **4.12 Procedimiento de recolección de datos**

##### **a) Algoritmo Predictivo usando Software Estadístico R**

Para ello, ha sido necesario desarrollar las siguientes actividades:

- Justificar el uso del software R.
- Procesar la información de la base de datos SGI OC.
- Seleccionar la muestra a utilizar del periodo 2014 al 2019.

- Implementar el algoritmo predictivo para el tratamiento de la información.
- Finalmente hacer pruebas en campo con ambos sistemas y ver la tasa de eficiencia de ambos.

b) La información obtenida por el algoritmo repercutirá en la mejora del servicio.

Se implementó usando una laptop i5 con el software Excel y Software R además de la data que se encuentran en la base de datos SGIO de Sedapal.

c) El crecimiento de la población con servicio de agua potable influirá en los resultados obtenidos.

d) Procesamiento estadístico y análisis de datos.

El procedimiento estadístico a ser aplicado en esta investigación para explicar, demostrar y verificar lo planteado en la hipótesis, consistirá primeramente en las lecturas de la data obtenida llámese las fugas propiamente dichas obtenidas por el software de gestión de Fugas SGIOC durante el periodo comprendido entre enero del 2014 a diciembre del 2019.



5

# Capítulo 5

*Resultados*





## Resultados

### 5.1 Introducción al capítulo

La detección temprana de fugas en sistemas de acueductos es crucial para garantizar la eficiencia y la sostenibilidad del suministro de agua potable. Las fugas en los sistemas de acueductos pueden causar una serie de problemas, como la pérdida de agua, la disminución de la presión del agua, el aumento del costo del tratamiento del agua y la contaminación del agua. Además, las fugas pueden dañar la infraestructura y los edificios cercanos, lo que puede resultar en costosas reparaciones y daños a la propiedad.

La detección temprana de fugas puede ayudar a reducir el desperdicio de agua y los costos asociados con las reparaciones y el mantenimiento del sistema. Las técnicas modernas de detección de fugas, como el análisis acústico, el análisis de imágenes térmicas y el análisis de datos, pueden ayudar a detectar fugas en los sistemas de acueductos. Estas técnicas pueden ayudar a identificar las fugas antes de que se conviertan en

un problema mayor, lo que puede reducir el costo y el tiempo necesarios para repararlas.

En resumen, la detección temprana de fugas en sistemas de acueductos es esencial para garantizar la eficiencia y la sostenibilidad del suministro de agua potable. Las técnicas modernas de detección pueden ayudar a identificar las fugas antes de que se conviertan en un problema mayor, lo que puede reducir los costos asociados con las reparaciones y el mantenimiento del sistema.

## **5.2 Generalización de los resultados**

Como se puede apreciar en la Tabla 6, los tipos de fugas: medidor, válvula grifo cl y valvula grifo, no se detectaron o fueron casi nulos en el periodo 2014 -2016. Se dispone de un mejor modelo de tratamiento de los datos y se puede generalizar aún más para estimar las proyecciones. Es por eso que agruparemos la variable FECHA para tener datos globales de los 7 años, entonces modificamos el algoritmo y obviamos el campo FECHA y nos centramos en las variables TIPOFUGA en función de la variable DISTRITO para tener el consolidado de las proyecciones.

Modificamos nuestro algoritmo en función de la cantidad de TIPOFUGA encontradas y lo ejecutamos en la venta de comandos del Software R.

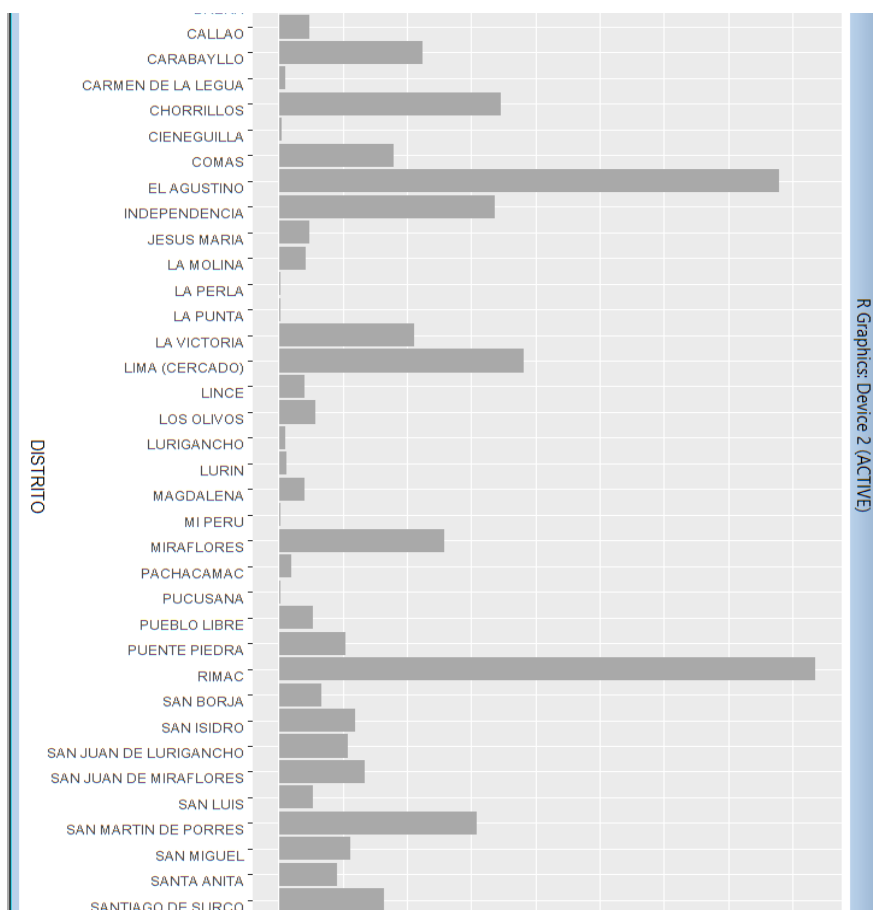
```

GGPLOT(DATA=SEDAPAL_TABLA, AES(X=DISTRITO, Y =
1, )) +
  GEOM_BAR(STAT="IDENTITY",FILL = "DARKGRAY",
POSITION="STACK")+
THEME(AXIS.TEXT.X = ELEMENT_TEXT(ANGLE = 90,
HJUST = 1))

```

**Figura 54**

*Modelo del tipo barras: Tipofugas & Distrito 2014-2019.*



Tenemos un mejor modelo de tratamiento de los datos y se puede generalizar aún más para estimar las proyecciones. Es por eso que agruparemos la variable FECHA para tener datos globales de los 7 años,

entonces modificamos el algoritmo y obviamos el campo FECHA y nos centramos en las variables TIPOFUGA en función de la variable DISTRITO para tener el consolidado de las proyecciones.

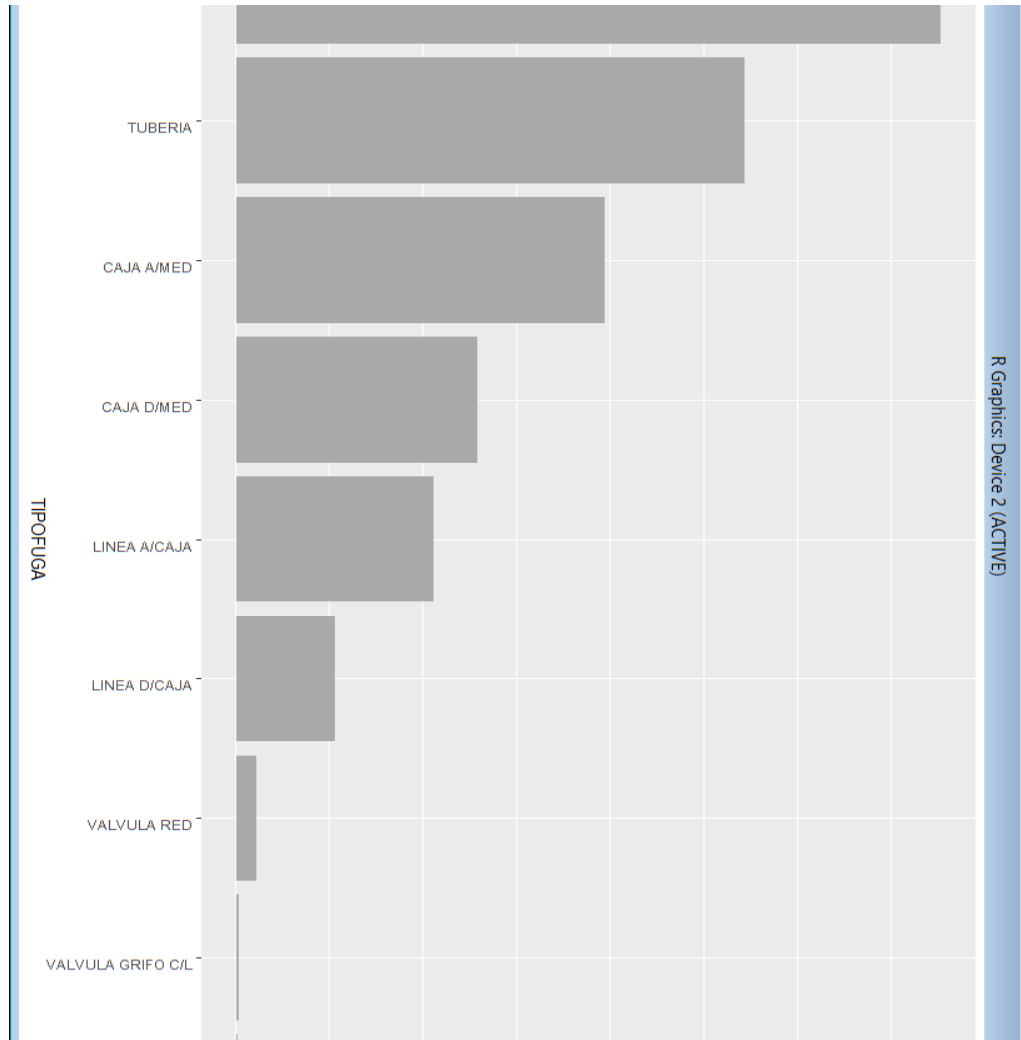
Luego procedemos a ordenar nuestra variable TIPOFUGA haciendo uso del comando **factor**

```
sedapal_tabla$TIPOFUGA = factor(sedapal_tabla$TIPOFUGA,
levels=c("CORP","TUBERIA", "CAJA A/MED", "CAJA D/MED", "LINEA
A/CAJA", "LINEA D/CAJA", "VALVULA RED", "VALVULA GRIFO C/L",
"MEDIDOR" ))
levels(sedapal_tabla$TIPOFUGA)
"CORP"          "TUBERIA"          "CAJA A/MED"      "CAJA D/MED"
"LINEA A/CAJA"      "LINEA D/CAJA"      "VALVULA RED"
"VALVULA GRIFO C/L" "MEDIDOR"
ggplot(data=sedapal_tabla, aes(x=DISTRITO, y = 1, )) +
  geom_bar(stat="identity",fill = "darkgray",
position="stack")+
theme(axis.text.x = element_text(angle = 90, hjust =
1))
```

Como se puede observar, ya tenemos ordenados los tipos de fugas TIPOFUGA por su número a los largo del periodo 2014-2019 desde CORP hasta MEDIDOR.

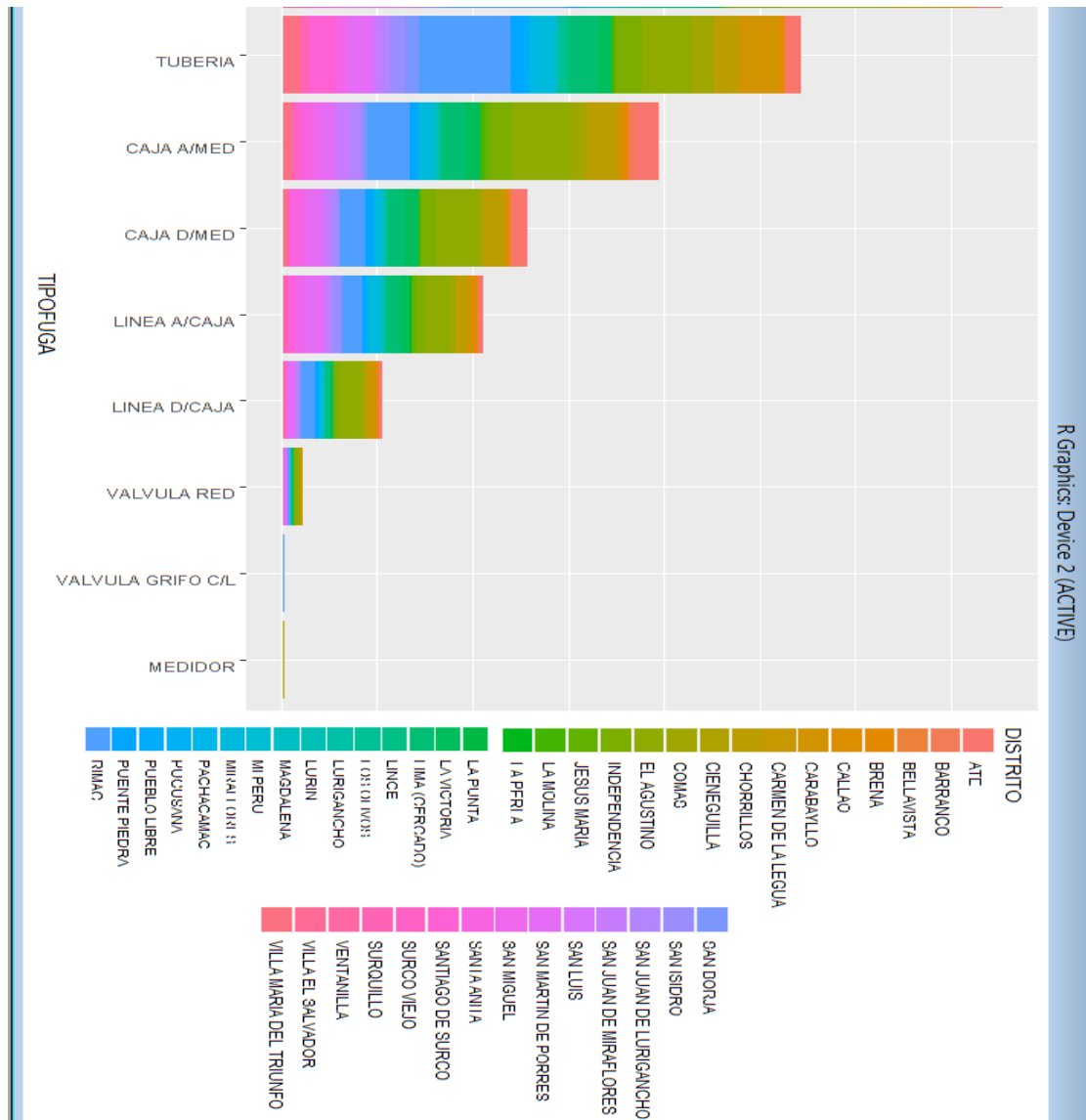
**Figura 55**

*Modelo del tipo barras ordenado por Tipofugas 2014-2019.*



**Figura 56**

*Modelo del tipo barras ordenado por Tipofugas & Distrito 2014-2019.*



Tenemos un mejor modelo de tratamiento de los datos y se puede generalizar aún más para estimar las proyecciones. Es por eso que agruparemos la variable FECHA para tener datos globales de los 7 años, entonces modificamos el algoritmo y obviamos el campo FECHA y nos centramos en las variables DISTRITO en función de la variable TIPOFUGA para tener el consolidado de las proyecciones.

Luego procedemos a ordenar nuestra variable DISTRITO haciendo uso del comando **factor**

```
sedapal_tabla$DISTRITO = factor(sedapal_tabla$DISTRITO,  
levels=c(  
"RIMAC","EL AGUSTINO","LIMA (CERCADO)","CHORRILLOS",  
"INDEPENDENCIA","SAN MARTIN DE PORRES","MIRAFLORES",  
"ATE","CARABAYLLO","LA VICTORIA","COMAS", "SANTIAGO DE  
SURCO", "SAN JUAN DE MIRAFLORES","VILLA MARIA DEL  
TRIUNFO","SAN ISIDRO","SAN MIGUEL","SAN JUAN DE  
LURIGANCHO","PUENTE PIEDRA","SANTA ANITA","SAN BORJA",  
"LOS OLIVOS","PUEBLO LIBRE","SAN LUIS","SURQUILLO",  
"CALLAO","BRENA","BARRANCO","JESUS MARIA","LA MOLINA",  
"MAGDALENA","LINCE","SURCO VIEJO","PACHACAMAC",  
"BELLAVISTA","LURIN","VILLA EL SALVADOR","VENTANILLA",  
"LURIGANCHO","CARMEN DE LA LEGUA","CIENEGUILLA", "MI  
PERU","PUCUSANA","LA PERLA","LA PUNTA"  
))  
ggplot(data=sedapal_tabla, aes(x=DISTRITO, y = 1, )) +  
  geom_bar(stat="identity",fill = "darkgray",  
position="stack")+  
theme(axis.text.x = element_text(angle = 90, hjust = 1))  
levels(sedapal_tabla$DISTRITO)
```

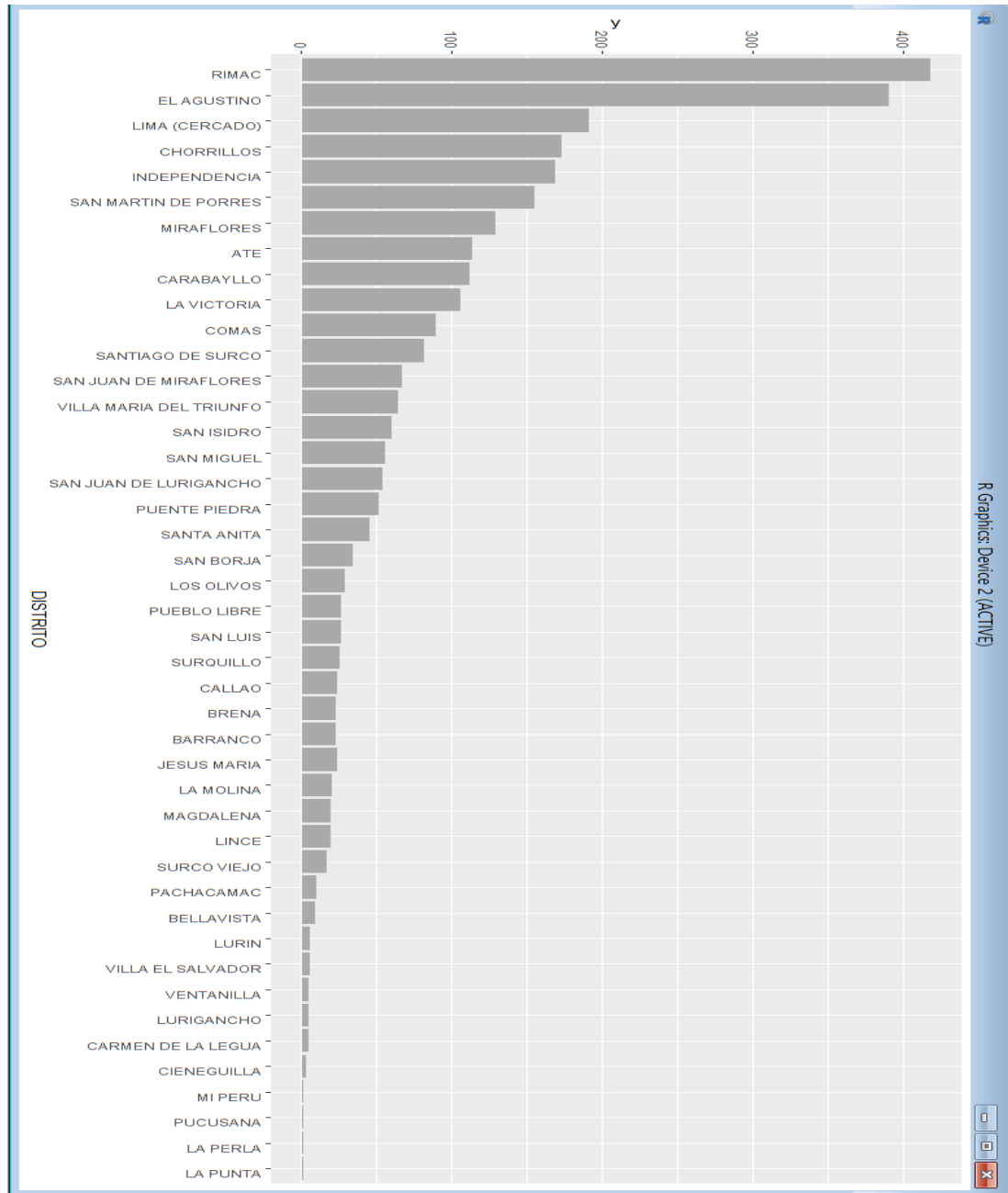
"RIMAC"	"EL AGUSTINO"	"LIMA (CERCADO)"
"CHORRILLOS"	"INDEPENDENCIA"	"SAN MARTIN DE PORRES" "MIRAFLORES"
"ATE"	"CARABAYLLO"	"LA VICTORIA"
"COMAS"	"SANTIAGO DE SURCO"	"SAN JUAN DE MIRAFLORES" "VILLA MARIA DEL TRIUNFO"
"SAN ISIDRO"	"SAN MIGUEL"	"SAN JUAN DE LURIGANCHO" "PUENTE PIEDRA"
"SAN BORJA"	"LOS OLIVOS"	"SANTA ANITA"
"PUEBLO LIBRE"	"SAN LUIS"	"SURQUILLO"
"CALLAO"	"BRENA"	"BARRANCO" "JESUS MARIA"
"LA MOLINA"	"MAGDALENA"	"LINCE"
"SURCO VIEJO"	"PACHACAMAC"	"BELLAVISTA"
"LURIN"		
"VILLA EL SALVADOR"	"VENTANILLA"	"LURIGANCHO"
"CARMEN DE LA LEGUA"	"CIENEGUILLA"	"MI PERU"
"PUCUSANA"		
"LA PERLA"	"LA PUNTA"	

Como vemos ya tenemos ordenados los distritos DISTRITO por su número de fugas detectadas a los largo del periodo 2014-2019 desde RIMAC hasta LA PUNTA (44 distritos).



**Figura 57**

*Modelo del tipo barras ordenado por distrito 2014-2019.*





### 5.3 Implementación

Tenemos la información de los 44 distritos y de los diferentes tipos de fugas a lo largo del periodo 2014-2019 y al haberla procesado con nuestro algoritmo obtuvimos los siguientes resultados luego de aplicar el algoritmo.

Los distritos con más Apoyos de Emergencia en el periodo 2014-2019 fueron:

- RIMAC
- EL AGUSTINO
- LIMA (CERCADO)
- CHORRILLOS
- INDEPENDENCIA
- SAN MARTÍN DE PORRES
- MIRAFLORES
- ATE
- CARABAYLLO
- LA VICTORIA

Las Fugas de Agua Potable que más se encontraron en el periodo 2014-2019 fueron:

- CORP
- TUBERÍA
- CAJA A/MED

- CAJA D/MED
- LINEA A/CAJA

Como se puede observar, el servicio de detección fugas depende en gran medida de la cantidad de unidades móviles que hay para el servicio, así como usuarios del servicio de agua potable (usuarios con medidor). En los últimos 3 años, 6 unidades móviles han sido destinadas para dicho fin por lo que su número se mantiene constante, no siendo así el caso de los usuarios del servicio; se estima que solo un 94,5% de las viviendas cuentan con medidor de agua<sup>(1)</sup>, es por eso que las emergencias en el servicio de detección de fugas de agua potable subirá aún más los próximos años al aumentar la cantidad de usuarios con medidor de agua en sus hogares, así se mantengan constante el número de unidades.

Se puede observar, el resumen de distritos por emergencias reportadas, figurando el Rímac, El Agustino y Lima Cercado como los distritos con más fugas reportadas a lo largo de esos 4 últimos años.

**Tabla 5.**

*Distrito con más emergencias reportadas 2016-2019.*

DISTRITOS CON MAS EMERGENCIAS REPORTADAS			
2016	2017	2018	2019
RIMAC	RIMAC	EL AGUSTINO	RIMAC
EL AGUSTINO	EL AGUSTINO	RIMAC	EL AGUSTINO
LIMA CERCADO	INDEPENDENCIA	LIMA CERCADO	LIMA CERCADO
LA VICTORIA	LIMA CERCADO	CHORRILLOS	INDEPENDENCIA
CHORRILLOS	SAN MARTIN	INDEPENDENCIA	CHORRILLOS

<sup>1</sup> Nota. SEDAPAL, 2018.

Se puede apreciar, el resumen de tipo de fugas por emergencias reportadas figurando como la fuga más reportada la de rotura de corporación, tubería y caja a/ medidor como las más recurrentes en los últimos 4 años.

**Tabla 6.**

*Tipo de fugas en emergencia más reportadas 2016-2019.*

TIPOS DE FUGAS DE EMERGENCIAS REPORTADAS			
2016	2017	2018	2019
TUBERIA	CORPORATION	CORPORATION	CORPORATION
CORPORATION	TUBERIA	TUBERIA	TUBERIA
CAJA A/ MEDIDOR	CAJA A/ MEDIDOR	CAJA A/ MEDIDOR	CAJA A/ MEDIDOR
CAJA D/MEDIDOR	CAJA D/MEDIDOR	CAJA D/MEDIDOR	CAJA D/MEDIDOR
LINEA A/CAJA	LINEA A/CAJA	LINEA A/CAJA	LINEA A/CAJA

Se disponen de los resultados de nuestro algoritmo predictivo y como se indicó en un principio, se tiene información la cual no se ha utilizado para este proyecto, que guardamos como información de control con la cual procederemos a hacer un modelo con el periodo de esta información que data de Enero-Marzo del 2020.

Lo primero es hacer el filtrado de la información y centrarnos en los campos más relevantes de la tabla para ejecutar análisis predictivo, procedemos a exportar el archivo a Excel para que sea más fácil el filtrado de la información por distrito, por fecha y por tipo de fuga y los demás datos que sean importantes para el filtrado.

La información procesada data de 01/01/2020 al 15/03/2020 correspondiente a los 3 primeros meses del año.

Luego de correr nuestro modelo obtuvimos los siguientes resultados en donde se nota cierta variación en los distritos del 3ro al 5to que puede deberse a que es solo información concerniente a los 3 primeros meses del 2020, así como a factores debido a que luego de la inspección no se encontró fuga (sin reporte de fuga) por lo que debe recabarse más información en un futuro.

**Tabla 7.**

*Distrito con más emergencias reportadas 2017-2020.*

DISTRITOS CON MAS EMERGENCIAS REPORTADAS			
2017	2018	2019	2020
RIMAC	EL AGUSTINO	RIMAC	RIMAC
EL AGUSTINO	RIMAC	EL AGUSTINO	EL AGUSTINO
INDEPENDENCIA	LIMA CERCADO	LIMA CERCADO	SAN JUAN DE LURIGANCHO
LIMA CERCADO	CHORRILLOS	INDEPENDENCIA	SAN LUIS
SAN MARTIN	INDEPENDENCIA	CHORRILLOS	CHORRILLOS

Luego de correr nuestro modelo obtuvimos los siguientes resultados en donde se nota que no hay variación con respecto a los tipos de fugas reportados en los años anteriores por lo que el algoritmo de tendencia predictiva es válido en este caso.

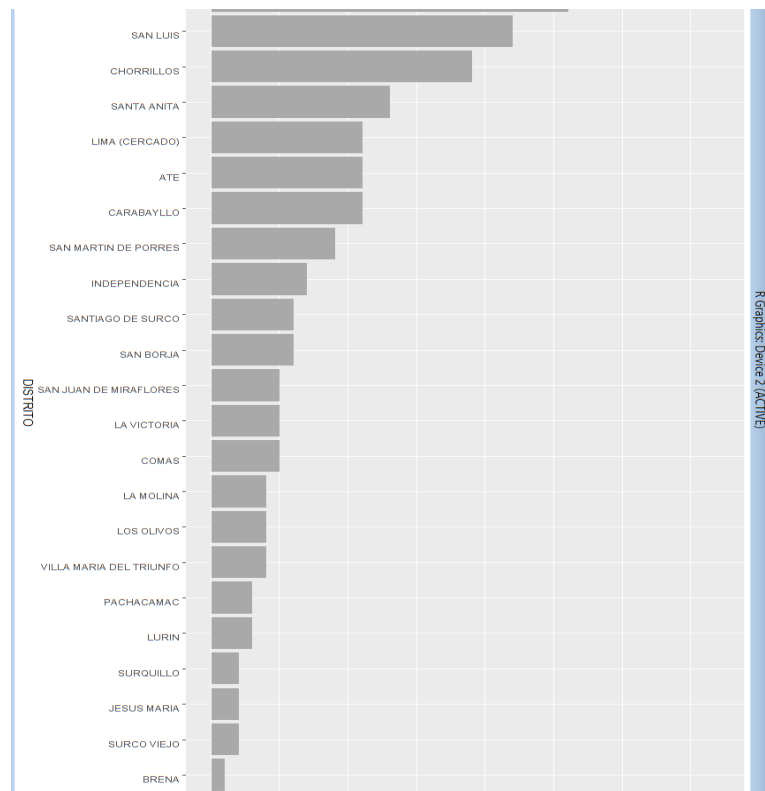
**Tabla 8.**

*Tipo de fugas en emergencia más reportadas 2017-2020.*

TIPOS DE FUGAS DE EMERGENCIAS REPORTADAS			
2017	2018	2019	2020
TUBERIA	CORPORATION	CORPORATION	CORPORATION
CORPORATION	TUBERIA	TUBERIA	TUBERIA
CAJA A/ MEDIDOR	CAJA A/ MEDIDOR	CAJA A/ MEDIDOR	CAJA A/ MEDIDOR
CAJA D/MEDIDOR	CAJA D/MEDIDOR	CAJA D/MEDIDOR	CAJA D/MEDIDOR
LINEA A/CAJA	LINEA A/CAJA	LINEA A/CAJA	LINEA A/CAJA

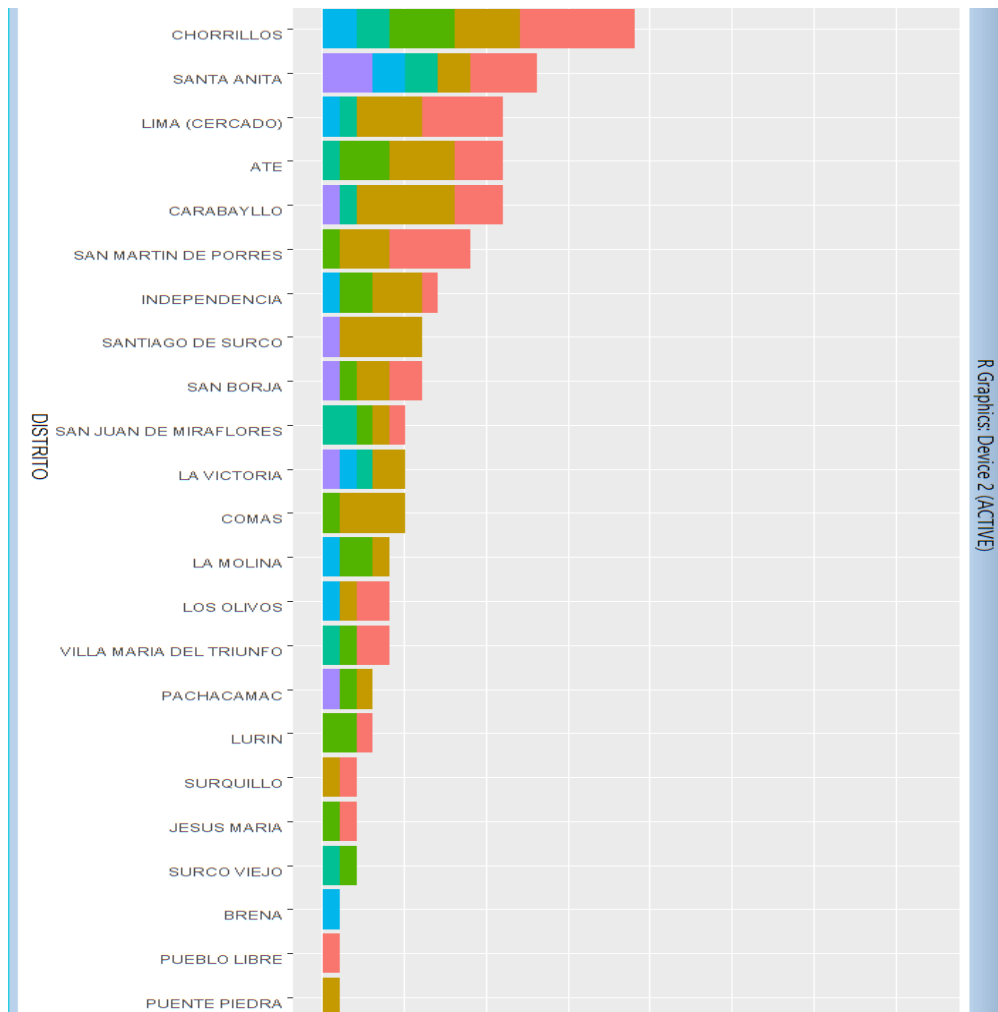
**Figura 59**

*Modelo del tipo barras ordenado por distrito 2020.*



**Figura 60**

*Modelo del tipo barras ordenado por Distrito & Tipofuga 2020.*

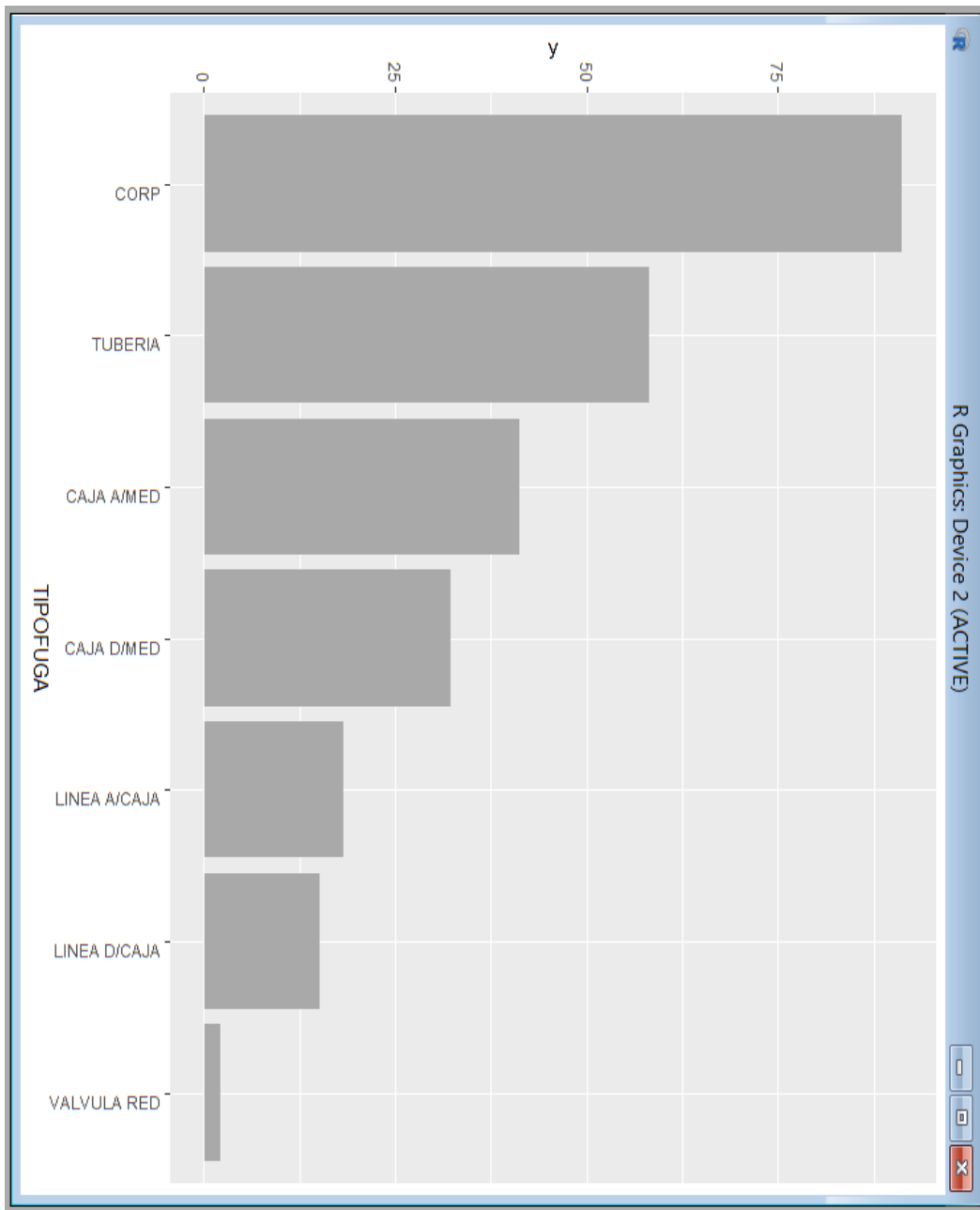


R Graphics Device 2 (ACTIVE)



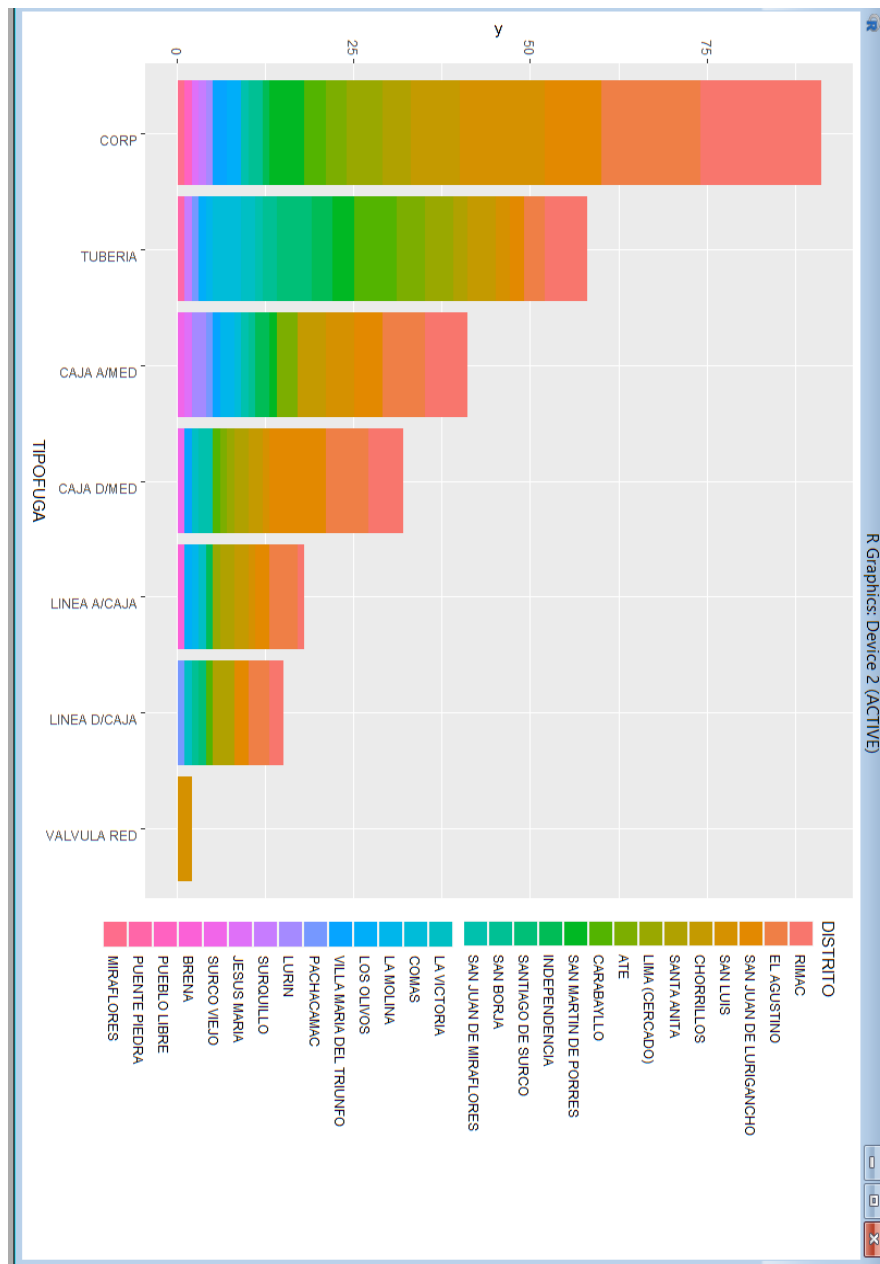
**Figura 61**

*Modelo del tipo barras ordenado por Tipofugas 2020.*



**Figura 62**

*Modelo del tipo barras ordenado por Tipofuga & Distrito 2020.*



## *Conclusiones finales*

Se desarrolló un “modelo algorítmico predictivo en el contexto del Software R para realizar el monitoreo de las redes de agua potable en la ciudad de Lima, 2019” obteniendo resultados prometedores en lo que es la predicción de fugas. Se comprobó que es posible, utilizando el análisis de Big Data y el árbol de toma de decisiones, para generar un Algoritmo que pueda predecir en qué distritos se producirán nuevas fugas de agua potable por emergencia y de qué tipos serán. Este es un desarrollo de suprema valía por estas época en que el agua es un recurso básico para la supervivencia de la raza humana predecir dónde se producirán fugas de agua ayudará a su control y reparación inmediata permitiendo reducir el alto porcentaje que bordea el 50% de agua no facturada en la actualidad.

SEDAPAL utilizó la información obtenida como un mapa de monitoreo del servicio y esto permitió organizar y optimizar de una mejor manera el uso de los recursos en el servicio de detección de fugas. Este estudio se puede ampliar y mejorar si se amplía la data y el historial de alimentación al software, mientras más datos se tendrá mayor calidad de proyecciones; como, por ejemplo, la cantidad de medidores de agua instalados por año en la ciudad de Lima para estimar la curva de crecimiento del servicio. Este tipo de algoritmos predictivos requieren del insumo de grandes bloques de datos con la cual trabajar para obtener resultados más confiables.

Finalmente, es para nosotros como autores un logro este interesante desarrollo que ha permitido esperanzarnos en producir mayores avances en el manejo de big data asociado a la prevención y detección temprana de fugas en los sistemas de abastecimiento de agua potable por todo lo que significa el agua en sí misma como recurso y como garantía de vida, por tanto, es nuestro aporte por la subsistencia de la humanidad.

## Referencias

- Aguasdevalencia. (2023). *Quieres beber agua de grifo.*  
<https://www.aguasdevalencia.es/Grupo/Inicio/>
- ANA. (s.f.). *Observatorio del Agua.*  
<https://crhc.ana.gob.pe/caracterizaci%C3%B3n-de-cuencas/ocupacion-del-suelo>
- Aqualia. (20 de 10 de 2023). *Sostenibilidad y progreso: como dos gotas de agua.*  
<https://www.eldia.es/sociedad/2023/10/20/sostenibilidad-progreso-gotas-agua-93565754.html>
- Cepei. (24 de 5 de 2021). *El desarrollo sostenible: pensar integralmente, hacer en común.* <https://cepei.org/documents/desarrollo-sostenible/>
- De León, M. (2020). *El agua que malgastas hoy.*  
<https://burica.wordpress.com/2007/03/22/25-es-dulce-y-975-es-agua-salada-en-el-mundo/>
- DiccionarioActual. (s.f.). *Recurso.* <https://diccionarioactual.com/recursos-humanos/>
- FAO. (13 de 10 de 2009). *La perspectiva mundial de la agricultura en el año 2050.*  
[https://www.fao.org/fileadmin/templates/wsfs/docs/Issues\\_papers/Issues\\_papers\\_SP/La\\_agricultura\\_mundial.pdf](https://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/Issues_papers_SP/La_agricultura_mundial.pdf)
- Fuentes-Mariles, O., Palma-Nava, A., & Rodríguez-Vázquez, K. (2011). Estimación y localización de fugas en una red de tuberías de agua potable usando algoritmos genéticos. *Ingeniería, investigación y tecnología*, 12(2), 235-242.  
[https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1405-77432011000200012](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-77432011000200012)

- Funcagua. (2020). *Agua en el planeta*. <https://funcagua.org.gt/agua-en-el-planeta/#:~:text=Aunque%20existe%20mucho%20agua%20en,menos%20del%201%25%20del%20planeta>.
- Hidrotec. (2016). *Cómo detectar fugas de agua en tuberías enterradas y soluciones (Actualizado 2021)*. <https://www.hidrotec.com/blog/detectar-fugas-agua-tuberias-enterradas-soluciones/>
- Hurtado-Cortés, L., Villarreal-López, E., & Villarreal-López, L. (2016). Detección y diagnóstico de fallas mediante técnicas de inteligencia artificial, un estado del arte. *Dyna*, 19-28.
- Iagua. (9 de 9 de 2019). *¿Por qué se producen las fugas de agua en las tuberías?* <https://www.iagua.es/blogs/arturo-albaladejo-ruiz/que-se-producen-fugas-agua-tuberias>
- MINAGRI. (2015). *Decreto Supremo No. 006-2015-MINAGRI*. <https://www.midagri.gob.pe/portal/decreto-supremo/ds-2015/12601-decreto-supremo-n-006-2015-minagri>
- MINAGRI. (2015). *Decreto Supremo No. 013-2015-MINAGRI*. <https://www.midagri.gob.pe/portal/decreto-supremo/ds-2015/13248-decreto-supremo-n-013-2015-minagri>
- NU. (2014). *Decenio Internacional para la Acción de "El agua frente a la vida" 2005-2015*. [https://www.un.org/spanish/waterforlifedecade/water\\_and\\_sustainable\\_development.shtml](https://www.un.org/spanish/waterforlifedecade/water_and_sustainable_development.shtml)
- NU. (2018). *Decenio por el agua 2018-2028*. <https://www.un.org/sustainabledevelopment/es/>
- Palermo, J. (12 de 11 de 2015). *Agua y desarrollo sostenible para el Perú*. <https://www.ulima.edu.pe/en/node/7142>
- Rubiños Jiménez, S.L. y Apesteguía Infantes, J.A. (2020). Diseño de un algoritmo predictivo para monitoreo temprano de redes de agua potable en la ciudad de Lima, 2019. [Tesis, Universidad Nacional del Callao].
- UNICEF. (s.f.). *Agua*. <https://www.unicef.org/es/agua>

ISBN: 978-9942-636-36-2



9789942636362