

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

DISEÑO E IMPLEMENTACIÓN DE UNA HERRAMIENTA PARA GESTIÓN DE INFORMACIÓN DE CAUDALES DE CAPTACIÓN Y PRODUCCIÓN EN LA (EPMAPS) APLICANDO INTELIGENCIA DE NEGOCIOS.

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

ERIKA ESTEFANIA ANATOA CÁRDENAS

erika.anatoa@epn.edu.ec

MARLON BRYAN VERNAZA TELLO

marlon.vernaza@epn.edu.ec

DIRECTOR: Marco Antonio Segura Morales, PhD.

marco.segura@epn.edu.ec

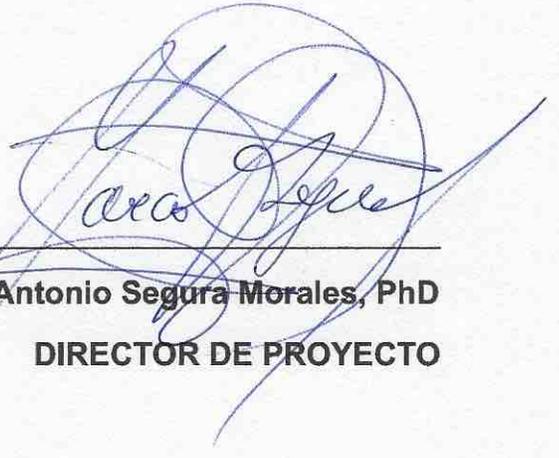
CODIRECTOR: Edison Fernando Loza Aguirre, PhD.

edison.loza@epn.edu.ec

Quito, enero 2020

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Erika Estefanía Anatoa Cárdenas y Marlon Bryan Vernaza Tello, bajo mi supervisión.

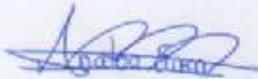


Marco Antonio Segura Morales, PhD
DIRECTOR DE PROYECTO

DECLARACIÓN DE AUTORÍA

Nosotros, Erika Estefanía Anatoa Cárdenas y Marlon Bryan Vernaza Tello, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentada previamente para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedemos nuestros derechos de propiedad intelectual correspondiente a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



Erika Estefanía Anatoa Cárdenas



Marlon Bryan Vernaza Tello

DEDICATORIA

Este trabajo se lo dedico a toda mi familia, pero sobre todo a dos mujeres que han marcado mi vida, las cuales son mi madre Estela Cárdenas quien ha estado apoyándome siempre y cuidado de manera incondicional, así también a la memoria de mi abuelita Rosario Collaguazo quien ha sido un ejemplo de lucha y perseverancia quien me enseñó que uno lo puede lograr todo con esfuerzo.

Erika Anatoa

AGRADECIMIENTO

Agradezco a Dios y mi Virgencita del Quinche quienes han guiado mis pasos y me han dado la fortaleza para continuar.

A mis padres: Estela y José por ser mi apoyo alentándome a ser mejor cada día dándome el mejor ejemplo de ser humano, sobre todo por su amor, paciencia y apoyo incondicional que me ha brindado durante toda mi vida.

A mis hermanos: Iván, Vero, Carlos y Vane quienes han creído en mí y han estado apoyándome y cuidándome. Pero sobre todo a mi hermana Vero quien ha sido la que se ha sacrificado para que no nos falte nada a mis papas y a mí.

A Oscar quien siempre ha creído en mí y me ha brindado su apoyo de forma incondicional, gracias por todo su esfuerzo por su amor y por no dejar que me rinda, te amo.

A mis mejores amigos, Susana y Cristian con quienes hemos compartido buenos y malos momentos en tantos años y han estado siempre para brindarme una amistad sincera.

A Marlon que más que mi compañero de tesis es un gran amigo y una excelente persona quien me apoyado en tantos años y en este último logro.

A Evelyn, Anita y Jacinto quienes han sido amigos excepcionales brindándome sus consejos y sobre todo agradecerles por todos los buenos momentos que hemos compartido juntos.

Erika Anatoa

DEDICATORIA

Este Proyecto de Titulación le dedico primeramente a toda mi familia y en especial a una persona que ya no está entre nosotros mi hermano Edgar Omar Arias Tello, que mediante su fe y dedicación infundieron en mí el afán de responsabilidad y superación además han estado conmigo en las buenas y malas apoyándome en mis estudios y para ellos es lo mejor que les puedo dar. Por último, se las dedico a todos mis compañeros que aún no llegan a este momento ya que es una trayectoria con muchos retos en la cual me di cuenta que sí, hay tropiezos, pero de ellos se aprende, rectifica y supera.

Marlon Vernaza

AGRADECIMIENTO

Mi agradecimiento es primeramente para Dios porque me ha dado la vida para llegar a este momento, a mis padres Welinton y Bertha ya que con su esfuerzo y sacrificio me han dado su ayuda para que yo me encuentre hasta ahora estudiando, a mis hermanos Edgar, Alba y Tatiana quienes me han apoyado también en cada uno de los pasos que he dado. Y en general agradezco a toda mi familia porque nunca perdieron su fe en mí.

Agradezco a cada uno de mis maestros que día a día me impartieron su sabiduría y conocimiento y supieron enrumbarme por el camino del bien para que pueda desenvolverme en la vida social y profesional.

Agradezco a mi director de tesis PhD. Marco Segura porque sin su ayuda, conocimiento y paciencia no estaríamos presentando este proyecto. A mis amigos Luis y Anita porque siempre estuvieron conmigo en los momentos de dificultad, a Evelyn y Edison con los que he compartido muchas vivencias. A mi compañera de Tesis Erika porque además de pasar conmigo casi toda la carrera, me ha tenido mucha paciencia en todo este proceso.

Finalmente, agradezco al personal de la Empresa Pública Metropolitana de Agua Potable y Saneamiento Quito (EPMAPS) quienes facilitaron la información y su conocimiento en cada proceso de este proyecto de titulación.

Marlon Vernaza

ÍNDICE DE CONTENIDO

CERTIFICACIÓN	I
DECLARACIÓN DE AUTORÍA.....	II
DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
DEDICATORIA.....	V
AGRADECIMIENTO.....	VI
ÍNDICE DE CONTENIDO.....	VII
ÍNDICE DE TABLAS	IX
ÍNDICE DE FIGURAS	X
RESUMEN	XII
ABSTRACT.....	XIII
1 INTRODUCCIÓN	1
1.1 FORMULACIÓN DEL PROBLEMA	1
1.2 JUSTIFICACIÓN	1
1.3 OBJETIVOS.....	2
1.3.1 Objetivo General	2
1.3.2 Objetivos Específicos.....	2
1.4 ALCANCE	3
1.5 DESCRIPCIÓN DE LA EPMAPS	3
1.5.1 Información general de la EPMAPS	3
1.5.2 Historia de la EPMAPS	4
1.5.3 Organigrama de la EPMAPS.....	4
1.6 CAUDAL	6
1.6.1 Definición de Caudal	6
1.6.2 Medición de caudales.....	6
1.6.3 Importancia del Caudal	6
1.6.4 Toma de Decisiones en la EPMAPS basada en caudales.....	7

1.7	INTELIGENCIA DE NEGOCIOS	8
1.7.1	¿Qué es Inteligencia de Negocios?.....	8
1.7.2	Arquitectura de la Inteligencia de Negocios.....	8
1.7.3	Beneficios de BI	9
1.7.4	Alcance de BI.....	10
1.7.5	Etapas para el proceso de Toma de Decisiones	10
1.8	DATA WAREHOUSE	11
1.8.1	Definición de Data Warehouse.....	11
1.8.2	Beneficios de un DW.....	12
1.9	DATA MINING.....	12
1.9.1	Definición de Data Mining	12
1.9.2	Etapas principales de Data Mining	13
2	METODOLOGIA	14
2.1	HEFESTO	14
2.1.1	Descripción	14
2.1.2	Características de Hefesto	15
2.2	KIMBALL.....	16
2.2.1	Descripción	16
2.2.2	Características	17
2.3	BILL INMON.....	17
2.3.1	Descripción	17
2.3.2	Características	18
2.4	CRISP-DM	18
2.4.1	Descripción	18
2.4.2	Características	19
2.4.3	Fases de CRISP-DM.....	20
2.5	JUSTIFICACIÓN DE LA METODOLOGÍA.....	21
2.5.1	Comparativa de las metodologías	21

2.6	METODOLOGIA CRISP-DM.....	23
2.6.1	Fases de CRISP- DM.....	24
2.7	DESARROLLO.....	33
2.7.1	Comprensión del negocio.....	33
2.7.2	Comprensión de los datos.....	36
2.7.3	Preparación de los datos.....	46
2.7.4	Modelado	48
2.7.5	Evaluación	59
2.7.6	Implementación.....	70
3	DISCUSIÓN DE RESULTADOS	71
4	CONCLUSIONES Y RECOMENDACIONES.....	73
4.1	Conclusiones.....	73
4.2	Recomendaciones	75
5	BIBLIOGRAFIA	76
6	ANEXOS	79
6.1	Anexo 1 Actas de Reunión.....	79
6.2	Anexo 2 Scripts de Creación de Base de Datos y Consultas SQL.....	79
6.2.1	Creación de la Base de Datos.....	79
6.2.2	Consultas empleadas durante la realización del proyecto:	80
6.3	Anexo 3 Codificación UT EPMAPS	82
6.4	Anexo 4 Manual de Usuario	82
6.5	Anexo 5 Correo de Satisfacción EPMAPS	82

ÍNDICE DE TABLAS

Tabla 2.1	Criterios de Metodologías	22
Tabla 2.2	Comparativa de Metodología.....	23
Tabla 3.1	Mackup vs Versión 1 (Captación)	64
Tabla 3.2	Mackup vs Versión 1 (Producción)	65

Tabla 3.3 Versión 1 vs Versión 2 (Captación).....	67
Tabla 3.4 Versión 1 vs Versión 2 (Producción).....	69
Tabla 6.1 Codificación UT Nivel 1	82
Tabla 6.2 Codificación UT Nivel 2	82

ÍNDICE DE FIGURAS

Figura 1.1 Organigrama EPMAPS [2].....	5
Figura 1.2 Arquitectura de BI [7].....	9
Figura 2.1 Metodología HEFESTO [14].....	15
Figura 2.2 Fases Metodología Ralph Kimball [15]	16
Figura 2.3 Arquitectura de un Data Warehouse según Bill Inmon [16].....	18
Figura 2.4 Esquema de los 4 niveles de CRISP-DM [17].....	19
Figura 2.5 Modelo de proceso CRISP-DM [18].....	20
Figura 2.6 Fase de comprensión del negocio CRISP-DM [20].....	25
Figura 2.7 Fase de comprensión de datos CRISP-DM [20]	26
Figura 2.8 Fase de preparación de datos CRISP-DM [20].....	28
Figura 2.9 Fase de Modelado CRISP-DM [20]	30
Figura 2.10 Fase de evaluación CRISP-DM [20]	31
Figura 2.11 Fase de implementación CRISP-DM [20]	32
Figura 2.12 Esquema relaciona de la base de datos	37
Figura 2.13 Perfilamiento de datos Tabla Plantas	40
Figura 2.14 Perfilamiento de datos Tabla Captación	40
Figura 2.15 Perfilamiento de datos Tabla Producción	40
Figura 2.16 Volumen de captación por planta en la fecha 31/01/2017.....	41
Figura 2.17 Volumen de captación de la planta El Troje en la fecha 31/01/2017	41
Figura 2.18 Volumen de captación de la planta El Troje de todos los años	42
Figura 2.19 Volumen de captación total de todas las plantas en la fecha 31/01/2017	42
Figura 2.20 Porcentaje del volumen de captación de todas las plantas en un rango de fechas	43
Figura 2.21 Volumen de producción por planta en la fecha 30/04/2018	43
Figura 2.22 Volumen de producción de la planta Bellavista en la fecha 31/04/2018.....	44
Figura 2.23 Volumen de producción de la planta Bellavista de todos los años	44
Figura 2.24 Volumen de producción total de todas las plantas en la fecha 30/04/2018	45
Figura 2.25 Porcentaje del volumen de producción de todas las plantas en un rango de fechas	45

Figura 2.26 Comparativa entre los volúmenes de producción y captación de la planta El Placer en un rango de fechas.....	46
Figura 2.27 Integración de nuevos datos.....	48
Figura 2.28 Creación de nuevo proyecto en Power BI.....	50
Figura 2.29 Obtención de datos en Power BI	51
Figura 2.30 Interfaces de Power BI	51
Figura 2.31 Componente Filtros	52
Figura 2.32 Componente Visualizaciones.....	52
Figura 2.33 Componente Campos.....	53
Figura 2.34 Atributos presentados en la interfaz Datos	53
Figura 2.35 Visualización del Modelo de la base de datos	54
Figura 2.36 Dashboard "Captación EPMAPS"	55
Figura 2.37 Dashboard "Comparativa Captación EPMAPS"	56
Figura 2.38 Dashboard "Producción EPMAPS"	57
Figura 2.39 Dashboard "Comparativa Producción EPMAPS"	58
Figura 2.40 Dashboard "Comparativa Captación vs Producción"	59
Figura 2.41 Mockup Captación 1	60
Figura 2.42 Mockup Captación 2.....	60
Figura 2.43 Mockup Producción 1	61
Figura 2.44 Mockup Producción 2	62
Figura 2.45 Primera versión Captación 1	63
Figura 2.46 Primera versión Captación 2	63
Figura 2.47 Primera versión Producción 1.....	64
Figura 2.48 Primera versión Producción 2.....	65
Figura 2.49 Segunda versión Captación 1.....	66
Figura 2.50 Segunda versión Captación 2.....	67
Figura 2.51 Segunda versión Producción 1	68
Figura 2.52 Segunda versión Producción 2.....	68
Figura 2.53 Segunda versión Comparativa Captación vs Producción.....	69

RESUMEN

El presente proyecto de titulación es un aporte para la Empresa Publica Metropolitana de Agua Potable y Saneamiento Quito, empresa que se encuentra conformada por varios departamentos encargados, día con día de tomar importantes decisiones referente a los procesos de captación y producción de agua. Esta información es manejada de forma semi-automatizada, ya que es extraída y procesada en hojas de cálculo antes de ser manualmente archivada, lo que genera problemas en la búsqueda de información histórica.

Para dar solución a estos problemas, en el presente proyecto se diseñó e implementó una solución de Inteligencia de Negocios aplicando la metodología CRISP – DM realizando una adaptación a las necesidades del proyecto. Las perspectivas y los indicadores que se utilizaron para analizar los datos se tomaron en base a la información obtenida en el proceso de especificación de requerimientos.

La solución resultante permite que la empresa genere datos estadísticos comparativos de los procesos de captación y producción de volumen de agua. Estos datos son presentados al usuario final por medio de gráficos y tablas mediante reportes que permite analizar de mejor manera la información obtenida sobre el volumen de los procesos de captación y producción, con el fin que se pueda realizar una toma de decisiones óptima.

Palabras clave: Procesos captación y producción, Metodología CRISP- DM, datos estadísticos comparativos, reportes.

ABSTRACT

This integration project is a report for the Quito Public Metropolitan Water and Sanitation Company, a company that is made up of several departments in charge, day after day, of making important decisions regarding the water capture and production processes. This information is handled in a semi-automated way, since it is extracted and processed in spreadsheets before being manually archived, which generates problems in the search for historical information.

To solve these problems, in this project was designed and implemented a Business Intelligence solution applying the CRISP - DM methodology, adapting it to the needs of the project. The perspectives and indicators that will be used to analyze the data were taken based on the information obtained in the needs specification process.

The resulting solution allows the company to generate comparative statistical data of the processes of capture and production of water volume. This data is sent to the end user by means of graphs and tables through reports that allow better analysis of the information obtained on the volume of the capture and production processes, in order to make an optimal decision-making.

Key words: Capture and production processes, CRISP-DM Methodology, comparative statistical data, reports.

1 INTRODUCCIÓN

1.1 FORMULACIÓN DEL PROBLEMA

La Empresa Publica Metropolitana de Agua Potable y Saneamiento Quito EPMAPS es la encargada del “abastecimiento de agua potable en el Distrito Metropolitano de Quito (DMQ)”, gestionando el funcionamiento de varios sistemas integrados que proveen el servicio a la ciudad, a las parroquias suburbanas y a las rurales, los cuales se encuentran complementados con otros sistemas independientes. El caudal de agua que actualmente está disponible en las fuentes del DMQ alcanza los 8,6 m³/s, y este caudal proviene de cuencas con ciclos hidrológicos suplementarios. “Para su tratamiento existe una capacidad instalada de 8,5 m³/s (considerando pozos y vertientes), de los cuales en la actualidad se producen 7,3 m³/s de agua” [1].

En este proyecto se toman en cuenta las Plantas de Tratamiento más grandes de la EPMAPS, con el fin que los datos que son generados por los caudales de los procesos de captación y producción puedan ser extraídos desde el Sistema SCADA, los cuales servirán para la toma de decisiones de la empresa. A fin de tener una correcta gestión de la información se toman los datos de los últimos tres años para luego proceder a la implementación del ciclo completo de analítica de datos, el cual implica “explorar datos, preparar datos, planificar un modelo, elaborar el modelo, comunicar resultados y utilizar los hallazgos para la toma de decisiones” [2].

Este proyecto propone el diseño y la implementación de reportes y dashboards utilizando el ciclo de vida de proyectos de Inteligencia de Negocios (BI), y aplicando tecnología informática para gestionar de manera ágil y segura toda la información que servirá en el proceso de gestión de caudales en la EPMAPS.

1.2 JUSTIFICACIÓN

Las organizaciones todos los días toman decisiones de gran importancia, las herramientas de BI permiten a los usuarios finales una forma sencilla de acceder a la información, creando requerimientos específicos de reportes al departamento de tecnología de la empresa.

En el presente proyecto se implementan las herramientas de BI para la toma de decisiones de la empresa, ya que, por medio del análisis de datos existentes, estas herramientas

proporcionan tácticas de mejora continua para la empresa, además las mismas ofrecen una visión global de la situación actual de la empresa sobre los caudales, facilitando la identificación de fortalezas y debilidades (aspectos de mejora) acerca del tema, brindando el apoyo necesario a los encargados para que la toma de decisiones sea oportuna.

Este proyecto es de importancia para la EPMAPS ya que los resultados obtenidos del agua que proviene de ríos o quebradas y es sometida a los procesos captación y producción, se basarán en datos de calidad, oportunos y con un nivel de exactitud alta acerca de los caudales, con el fin que la empresa tenga una visión más clara sobre los caudales y los impactos generados en el ambiente con su uso y a su vez tomar las decisiones oportunas para controlar o evitar dichos impactos.

Para la accesibilidad a los datos se firmará un acuerdo de cooperación estratégico con la empresa EPMAPS, en el cual la misma se responsabiliza en facilitar la información consolidada en el sistema SCADA y por otra parte nosotros nos comprometemos en implementar las herramientas de BI en las plataformas que la empresa tenga implementadas.

1.3 OBJETIVOS

1.3.1 Objetivo General

Diseñar y construir un módulo de Inteligencia de Negocios para la EMPRESA PUBLICA METROPOLITANA DE AGUA POTABLE Y SANEAMIENTO QUITO enfocado en la data de Caudales, usando un ciclo de vida de proyectos de Inteligencia de Negocios.

1.3.2 Objetivos Específicos

- Realizar una exploración y documentar los datos disponibles en el sistema SCADA sobre caudales de los procesos captación y producción de la EPMAPS.
- Realizar un proceso de selección de atributos y preparación de datos para asegurar la disponibilidad, formato apropiado, y calidad de los datos relevantes para el negocio.
- Realizar un análisis de los posibles modelos para de esta forma definir indicadores asociados a la data de caudales.
- Implementación de los procesos de carga, limpieza, transformación y presentación de la información.

- Comunicar apropiadamente los resultados mediante dashboards, reportes y visualizaciones adecuadas que hayan sido validadas por los interesados del proyecto.

1.4 ALCANCE

Se realiza el diseño e implementación de una herramienta de Inteligencia de Negocios, para los procesos de captación y producción de volumen de agua, en la empresa EPMAPS.

Para implementar esta herramienta, se crea una base de datos local con SQL Server 2016 y con Microsoft Power BI se desarrollan cinco dashboards. Se realiza la implementación con este software debido a que la empresa actualmente está trabajando con estas herramientas.

En el Anexo I se presentan las actas correspondientes a las reuniones que se efectuaron con los Departamentos encargados de la toma de decisiones. Dentro de estas, se detalla cuáles son sus requerimientos y si éstos están dentro del alcance del proyecto.

El proyecto concluye con la generación de cinco dashboards que contienen toda la información e indicadores que responden a las necesidades del negocio en relación con los procesos de captación y producción de volumen de agua. Este proyecto facilita a los departamentos correspondientes de la EPMAPS la toma de decisiones a nivel estratégico y táctico.

1.5 DESCRIPCIÓN DE LA EPMAPS

1.5.1 Información general de la EPMAPS

“La EPMAPS AGUA de QUITO es una dependencia municipal encargada de la gestión del agua en el Distrito Metropolitano de Quito” [3], que empieza en la captación en las fuentes que incluye el manejo de aguas residuales urbanas, las mismas que atraviesan por los procesos de conducción, potabilización, distribución y recolección de las aguas residuales, con el fin de ofrecer los servicios de agua potable y saneamiento a la ciudad, haciéndose responsable de la descontaminación de aguas residuales devolviendo agua más limpia a la naturaleza [3].

1.5.2 Historia de la EPMAPS

“Desde la época colonial se comienzan a establecer normas regulatorias sobre el manejo del agua en Quito” [3]. Por lo que, en el año de 1535, el Cabildo de Quito tuvo la autoridad de administrar las aguas que descendían del Pichincha y las embalsadas en las lagunas para de esta forma quitar el uso arbitrario del agua a los dueños de las residencias cercanas.

Hoy, con más de 2000 trabajadores y tras cincuenta y cinco años de arduo trabajo han ido mejorando los diferentes sistemas de la Empresa para renovar el compromiso de servicio y calidad.

El objetivo de la empresa es ofrecer los servicios de agua potable y saneamiento a la ciudad, asumiendo también la responsabilidad en la descontaminación las aguas residuales devolviendo así agua más limpia a la naturaleza [3].

1.5.3 Organigrama de la EPMAPS

“La Empresa Pública Metropolitana de Agua Potable y Saneamiento de Quito trabaja con un esquema directivo compuesto por profesionales y servidores altamente capacitados en sus funciones” [4]. La Figura 1.1 representa el organigrama de la EPMAPS y sus diferentes procesos [4].

ESPACIO DEJADO EN BLANCO POR LOS AUTORES

1.6 CAUDAL

1.6.1 Definición de Caudal

Se define al caudal en la hidrografía, la hidrología y la geografía física, como “el volumen de agua que pasa por el cauce de un río en un lugar y tiempo específico” [5]. Se lo define también como “el volumen hidráulico de la corriente de una cuenca hidrográfica agrupada en el río principal de esta cuenca” [5]. La unidad de medida principal del caudal es m^3/seg , generando así un valor anual que es medido en m^3 o en Hm^3 (hectómetros cúbicos: un Hm^3 equivale a un millón de m^3) la cual es utilizada para realizar planificaciones de los recursos hidrológicos y su uso en las obras de canalización. Al pasar los años al caudal de un río se lo ha constituido como el régimen fluvial de dicho río [5].

1.6.2 Medición de caudales

Para la medición de caudales que se realiza en las plantas de tratamiento de aguas residuales aerobios que cuentan con sistemas electromecánicos se utilizan los caudalímetros portátiles equipados con sensores ultrasónicos. Mientras que en los sistemas anaerobios se utiliza el método volumétrico. Para los sistemas en los cuales no puede valerse del método antes mencionado se emplea el método del flotador que consiste en una medición directa en base a la distancia, tiempo y área [6].

1.6.3 Importancia del Caudal

Manteniendo la idea de la recuperación de los ríos se han enfocado en dos metas específicas:

- Mejorar la calidad del agua.
- Establecer un caudal mínimo para que ríos y riachuelos no queden completamente secos.

Sin embargo, en los ecosistemas que tienen aguas terrestres y han sido transformados, no se ha logrado una recuperación total de las funciones y procesos que ayudan a mantener la integridad de estos ríos. El principal inconveniente para la restauración de los ríos es que, más allá del caudal total varios estudios han demostrado que la estabilidad y funcionamiento de estos ecosistemas dependen del régimen temporal que caracteriza al flujo de ese caudal. El cambio en el régimen natural de caudales afecta la salud de los ecosistemas, sugiriendo que un enfoque integral de manejo de los ríos también debería incorporar el mantenimiento del régimen de caudales [7].

1.6.4 Toma de Decisiones en la EPMAPS basada en caudales

Tomando en cuenta los objetivos definidos en la empresa, se han establecido indicadores para la toma de decisiones, considerando a los caudales como punto clave para alcanzar las metas de mejora estratégica y operativa. A continuación, se describen los principales indicadores [8]:

- Satisfacción de la Demanda de Agua Potable en el DMQ. - Lo que se pretende es dotar a la población con agua potable, con el fin de garantizar la demanda, a medio o largo plazo, por lo que se debe tomar en cuenta el agua superficial y el agua subterránea. Entre los más significativos están los Ríos Orientales, siendo éstos la única solución para Quito, logrando así satisfacer la demanda de agua potable a mediano y largo plazo.
- Incremento y optimización de la capacidad de potabilización. - Lo que se pretende es extender la capacidad en las plantas de tratamiento, sin olvidar la demanda de la población. Para la relación del agua subterránea, se realizan estudios con el fin de efectuar un análisis sobre la disponibilidad y calidad en el acuífero de Quito, el cual es considerado como reserva estratégica para el Distrito.
- Nuevas líneas de conducción de agua tratada. - Lo que se pretende es extender la capacidad de conducción de agua tratada, que será analizada en función de un aumento previsto en cada uno de los componentes en los sistemas de agua potable.
- Universalización de la cobertura de los servicios de agua potable y alcantarillado. - Se pretende llegar a cada uno de los clientes del DMQ a través de catorce redes y conexiones domiciliarias de agua potable y alcantarillado, aprovechando la infraestructura con la que ya se cuenta en la Empresa, sin dejar de lado la calidad y políticas existentes. Sin embargo no hay que olvidar los servicios existentes en las áreas de expansión urbana, los centros poblados de zonas de expiación y las parroquias rurales.
- Descontaminación de Ríos de Quito. - Trabajar de una manera integral y responsable el ciclo del agua, que parte en la captación del recurso hasta que termina en la devolución a la naturaleza, teniendo el menor impacto posible, y para esto se ha elaborado el “Plan de Descontaminación de los ríos y quebradas del DMQ” [6].
- Reducción de Consumos. - El agua es necesaria e indispensable para la vida en nuestro planeta y es responsabilidad de cada uno de nosotros cuidarla, ya que distribuirla y consumirla es primordial, y salvaguardando este recurso es la empresa quien maneja la búsqueda de nuevas fuentes de captación de agua para el DMQ.

- Agua No Contabilizada. - Se trabajará en una mejora acerca del manejo de agua no contabilizada.
- Gestión ambiental. - Gestionar el agua para consumo de los seres humanos (generada en cuencas y micro cuencas que suministran los sistemas de agua potable), y a su vez manejar y prevenir cada uno de los impactos ambientales producidos por la Empresa durante los procesos constructivos y los de operación.

1.7 INTELIGENCIA DE NEGOCIOS

1.7.1 ¿Qué es Inteligencia de Negocios?

Se define a la inteligencia de negocios como el conjunto de tecnologías, procesos y herramientas, los mismos que transforman los datos que han sido almacenados en información, y luego convertir esta información en conocimiento, con el fin que el conocimiento sea dirigido a lo que se denomina un plan o estrategia comercial. La inteligencia de negocios es una parte fundamental en la estrategia empresarial, ya que esta optimiza el manejo de los recursos, proporcionando el respectivo seguimiento para el cumplimiento de los objetivos del negocio y la toma de optimas decisiones para obtener mejores resultados [9].

1.7.2 Arquitectura de la Inteligencia de Negocios

“Lo primordial son las fuentes de datos (Cubos essbase, bases de datos Oracle, Sql Server, mainframe, archivos planos, archivos xml, hojas de Excel, etc.) de las que se pueden extraer los datos de múltiples fuentes simultáneamente” [9]. Estas fuentes permitirán realizar una definición de las fuentes heterogéneas y los campos con los que se va a trabajar, en el proceso de extracción, transformación y carga (ETL) [9].

Luego se tiene el repositorio de datos donde se encuentran todos los datos que han sido transformados en información, los cuales son representados de una manera visual por medio de tablas de datos, dimensiones y modelos multidimensionales. Existe un modelo intermedio entre este repositorio y la interfaz de acceso al usuario que toma el nombre de motor de BI en dónde se habilitan los componentes, administra las consultas, monitorea los procesos, cálculos y métricas [9].

Finalmente, la interfaz de acceso a los usuarios permite una interacción con los datos, esta interfaz se muestra de forma gráfica, todo esto se realiza con los resultados de los indicadores

y las consultas de gestión que fueron construidos. Todos estos componentes forman parte de la arquitectura de BI tal como se muestra en la Figura 1.2 [9].



Figura 1.2 Arquitectura de BI [9]

1.7.3 Beneficios de BI

Al utilizar BI en una empresa existen ciertos beneficios que se detallan a continuación [10]:

- Incremento de la eficiencia. - Si los datos fueron registrados de una forma ágil y accesible, estos generan información centralizada valiosa lo que permitirá su análisis y toma de decisiones en tiempos óptimos.
- Respuestas rápidas a situaciones de negocio. - Al trabajar con BI se tienen las respuestas en minutos de forma clara y concisa, la cual se presenta por medio de indicadores, tableros de datos y reportes.
- Control de las áreas funcionales de la empresa. - Todas las áreas de la empresa están generando información muy valiosa día a día, la misma que se puede aprovechar para conocer más acerca de las tendencias, de esta forma proyectar los datos y realizar un análisis de escenarios, etc.
- Mejora el servicio al cliente. - Al disponer de la información de mayor relevancia y en tiempo real se puede ofrecer al cliente un servicio de alto nivel desde la gestión hasta el servicio post-venta ya que se conoce las necesidades del mismo. Donde se realiza un análisis de sus hábitos de consumo, etc.
- Presenta información por medio de tableros de indicadores. - La clave de una buena comunicación es que la misma sea lo más simple y directa en base al entorno de la

empresa. Lo que permite la posibilidad de crear tableros de control, los mismos que se basan en los datos relevantes.

1.7.4 Alcance de BI

BI es una parte fundamental y estratégica en la organización ya que esta proporciona las pautas para establecer la toma de decisiones de la manera más oportuna y confiable, respondiendo a cualquier situación que pueda darse en la empresa; como la adquisición de nuevos productos o servicios, el análisis de costos y la rentabilidad que tendrá el producto o servicio. Los alcances de BI pueden ser [10]:

- Nivel operativo. - Sirve para efectuar la toma de decisiones diarias en base de las transacciones acerca de las operaciones de la empresa.
- Nivel táctico. - Brinda información para efectuar los análisis y decisiones mensuales, las que sirven para realizar el respectivo seguimiento y toma de acciones.
- Nivel estratégico. - En este punto las decisiones tienen gran impacto en la empresa.

1.7.5 Etapas para el proceso de Toma de Decisiones

Hoy en día, la toma de decisiones es fundamental para una empresa que quiera lograr un mejor rendimiento a nivel empresarial. Mediante BI es posible realizar un tratamiento adecuado de la información siguiendo un proceso de 8 pasos con el fin de tomar las decisiones más acertadas para la empresa. A continuación se describen las características de cada paso [11]:

1. Identificación de un problema

Se debe conocer si existe algún problema. Identificar las diferencias del escenario actual y la situación que se tiene como objetivo. Ya que con este conocimiento se puede empezar con el proceso de la toma de decisiones con la recopilación de la información. Es muy importante tener en cuenta que la diferencia identificada se considere como problema.

2. Identificación de los criterios de la toma de decisiones

Seleccionar el método o los puntos notables para dar solución a un problema encontrado. Una parte significativa es que cada miembro sea responsable de tomar cualquier decisión internamente en la empresa, para que se pueda indicar los criterios con los que se rigen y también los que no se toman en cuenta o se omiten.

3. Asignación de ponderaciones a los criterios seleccionados

Es importante que se prioricen los criterios identificados en la etapa preliminar. Esto debe ir acorde con la relevancia que se tenga en relación con la decisión final. Suele haber varios predominantes, los cuales se utilizan como referencia para realizar la ponderación y comparación con el resto de los criterios.

4. Desarrollo de diferentes opciones

El identificar el problema y la ponderación de los criterios determinados, son la pauta necesaria para la generación de distintas alternativas u opciones en la resolución del problema que se encontró inicialmente.

5. Análisis de las alternativas

Las opciones se analizan minuciosamente por la persona encargada de tomar la decisión. Para esto se debe conocer las fortalezas y debilidades de cada una, enfrentando cada criterio que ha sido ponderado en la tercera etapa.

6. Selección de la opción más viable

Después de realizar el análisis y el estudio de las distintas alternativas, la comparación con los criterios ponderados y jerarquizados, es el momento oportuno para seleccionar la mejor opción.

7. Puesta en marcha de la opción elegida

Una vez seleccionada la opción más conveniente para resolver el problema detectado, esta debe ser implementada. Ejecutar la decisión es vital. En primer lugar, la decisión debe ser informada a todos los departamentos que se vean afectados, con el fin de lograr que el personal se comprometa a realizar la implementación de la opción seleccionada.

8. Evaluación y seguimiento de la efectividad de las decisiones

Finalmente se realizará la evaluación del resultado que se obtuvo tras ejecutar la decisión y opciones que se pusieron en marcha, verificando si, definitivamente, se ha logrado corregir el problema.

1.8 DATA WAREHOUSE

1.8.1 Definición de Data Warehouse

Data Warehouse (DW) [12]:

- Es un repositorio unificado donde se recolectan todos los datos que almacenan los diversos sistemas de una empresa. Este repositorio puede ser lógico o físico haciendo hincapié en la recolección de datos de diversas fuentes que son útiles para fines analíticos y de acceso.

- Se aloja en la nube o un servidor corporativo. Los datos de las diversas aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes las que se extraen de forma selectiva para usarlas en aplicaciones analíticas y de consultas realizadas por los usuarios.
- Es una arquitectura que se basa en el almacenamiento de datos, la cual permite a los ejecutivos del negocio: la organización, comprensión y uso de los datos para tomar decisiones estratégicas.

1.8.2 Beneficios de un DW

Actualmente un DW proporciona información accesible, correcta, uniforme y actualizada, a un menor coste para la toma de decisiones, lo cual beneficia a las empresas que manejan grandes volúmenes de datos. A continuación, se detallan algunos beneficios del DW [13]:

- Brinda una herramienta para la toma de decisiones en varias áreas funcionales, en base a la información integrada y total del negocio.
- Proporciona la aplicación de técnicas y estadísticas de modelización y análisis para de esta forma hallar relaciones ocultas entre datos almacenados; obteniendo un valor agregado para la empresa.
- Provee la capacidad de aprender de los datos más antiguos prediciendo situaciones que puedan suceder en un futuro en diversos escenarios.
- Brinda a la empresa la implantación de los sistemas de gestión integral que dependen del cliente.
- Permite la optimización financiera y tecnológica en los Centros de Información para la generación de reportes, informes y estadísticas.

1.9 DATA MINING

1.9.1 Definición de Data Mining

Data Mining llamada también minería de datos, es la combinación de tecnologías y técnicas que permiten analizar bases de datos de gran magnitud, de forma automática o semiautomática, teniendo como objetivo identificar patrones repetitivos, tendencias o normas que manifiesten la conducta de los datos en un determinado contexto [14].

El propósito de Data Mining es ayudar en la comprensión del contenido de un repositorio de datos. Haciendo uso de prácticas estadísticas y, en algunos casos de algoritmos de búsqueda acercándose a la Inteligencia Artificial y a las redes neuronales [14].

Generalmente, la materia prima de un Data Mining son los datos. Al momento que un usuario les asigna un valor específico pasan a convertirse en información. Si los especialistas elaboran o crean un modelo, haciendo que la interpretación que se presenta entre la información y el modelo brindan un valor agregado, entonces se refiere al conocimiento [14].

1.9.2 Etapas principales de Data Mining

En el Data Mining cada caso determinado puede ser absolutamente diferente al anterior, el proceso que tienen en común todos ellos se componen de cuatro etapas que se detallan a continuación [14]:

- Determinación de los objetivos. – Esta etapa trata del establecimiento de los objetivos que el cliente solicita bajo la orientación de un especialista en Data Mining.
- Preprocesamiento de los datos. – Esta etapa se refiere a la selección, limpieza, reducción y transformación de las bases de datos.
- Determinación del modelo. – En esta etapa se comienza realizando algunos análisis estadísticos de los datos, y posteriormente se realiza una visualización gráfica de dichos datos para tener la primera aproximación. De acuerdo con los objetivos propuestos, y la tarea que debe realizarse pueden emplearse algoritmos desarrollados en las distintas áreas de la Inteligencia Artificial.
- Análisis de los resultados. – En esta etapa se comprueba si los resultados que se han obtenido son coherentes y se los coteja con los datos obtenidos por los análisis estadísticos y de visualización gráfica. El cliente estipula si los datos son novedosos y si proporcionan un nuevo conocimiento que le permita considerar sus decisiones.

2 METODOLOGIA

BI es un conjunto de técnicas y métodos que las empresas implementan para la toma de decisiones tácticas y estratégicas. Ya que trabajan de la mano con las tecnologías que se centran en afirmaciones, estadísticas y en los objetivos comerciales para de esta forma mejorar el rendimiento comercial [15]. Existen varias metodologías de BI para la toma de decisiones:

2.1 HEFESTO

2.1.1 Descripción

Hefesto es una metodología, “que está basada en la investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos” [16]. A continuación se detalla el proceso que sigue la metodología Hefesto [16].

- El primer paso es la recolección de necesidades de información de los usuarios, con el fin de obtener las preguntas que son primordiales para el negocio.
- Es primordial Identificar los indicadores resultantes de las preguntas y el análisis, ya que con estos se construirá el respectivo modelo conceptual de datos del DW.
- Después, se procede a analizar los OLTP, con el fin de determinar cómo se construirán los indicadores, los cuales mostrarán las correspondencias con los datos originales para elegir los campos de estudio de cada perspectiva.
- En este punto, se pasa a construir el modelo lógico del repositorio, donde se concretará cuál será el tipo de esquema que se va a implementar.
- Se crearán tablas de hechos y dimensionales, realizando así las uniones correspondientes.
- Finalmente, aplicando las técnicas de calidad y limpieza de datos, procesos ETL, etc, se definen políticas y estrategias para realizar la primera Carga del DW y su actualización.

La metodología Hefesto se resume a través de la figura 2.1:



Figura 2.1 Metodología HEFESTO [14]

2.1.2 Características de Hefesto

Esta metodología tiene las siguientes características [16]:

- Los objetivos y resultados esperados en cada una de las fases se diferencian de una manera rápida y fácil de entender.
- La base son los requerimientos del usuario, por consiguiente, la estructura es capaz de adaptarse con mayor rapidez y facilidad frente a los cambios que existan en el negocio.
- Reduce la resistencia al cambio, ya que esta implica a los usuarios finales en cada una de las etapas para la toma de decisiones con respecto a la funcionalidad y comportamiento del DW.
- Trabaja también con modelos lógicos y conceptuales, ya que estos se los puede interpretar y analizar fácilmente.

- Es independiente del ciclo de vida que se vaya a implementar.
- Es independiente del tipo de herramientas que se emplee en la implementación.
- Es independiente del tipo de estructuras físicas que contenga el DW y de su distribución.
- Cuando una fase es culminada, los resultados obtenidos pasan a ser el punto de partida para llevar a cabo los siguientes pasos.
- Esta es aplicada tanto en DW como en Data Mart.

2.2 KIMBALL

2.2.1 Descripción

Kimball, conocida como Modelo Dimensional (Dimensional Modeling), es una metodología que se basa en el Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). El Modelo Dimensional constituye los modelos de tablas y relaciones con el propósito de mejorar el proceso de toma de decisiones, basándose en las consultas realizadas en la base de datos relacional, la misma que puede estar ligada a la medición o conjunto de mediciones de los resultados en los procesos del negocio [17].

En la Figura 2.2 se presenta las diferentes fases de la metodología Kimball.

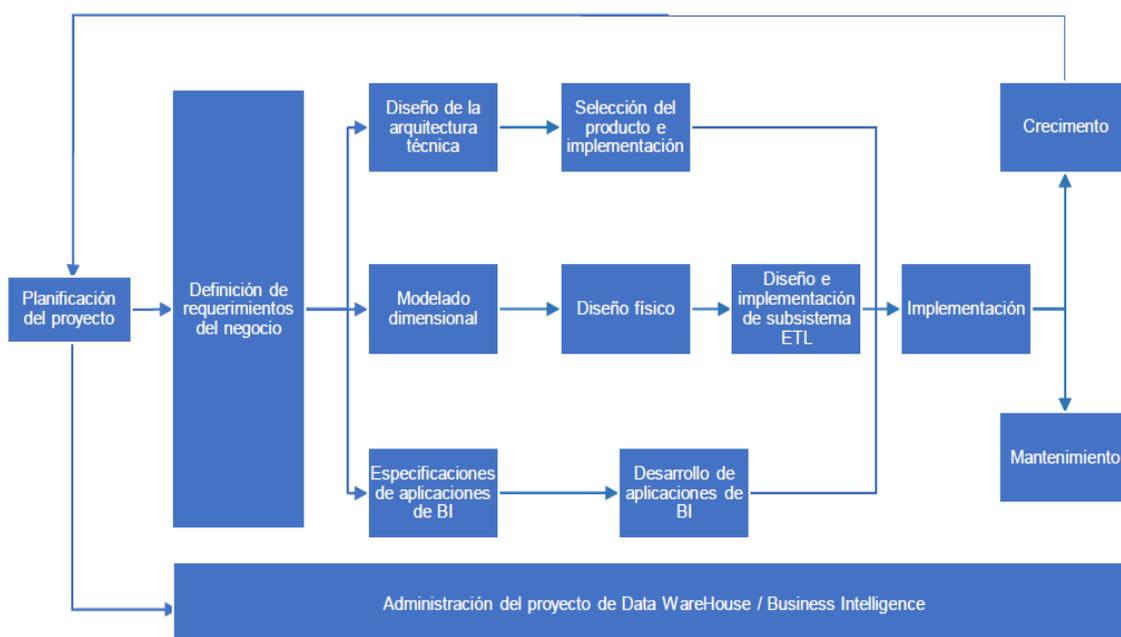


Figura 2.2 Fases Metodología Ralph Kimball [15]

2.2.2 Características

Esta metodología tiene las siguientes características [17]:

- Centrarse en el negocio. - Esta característica permite identificar los requerimientos que existen dentro del negocio y su valor asociado, los mismos que serán implementados al momento de desarrollar relaciones concretas con el negocio.
- Construir una infraestructura de información adecuada. - Esta característica permite conocer si al momento de diseñar una base de información la misma sea integra, única, de fácil uso y lo más importante de alto rendimiento donde se mostrarán los distintos requerimientos del negocio que se identificaron en la empresa.
- Realizar entregas en incrementos significativos. - Esta característica da el paso a la creación del almacén de datos (DW) con incrementos entregables en plazos de 6 a 12 meses. Sin dejar de lado el valor que tiene cada elemento identificado del negocio para de esta forma establecer el orden de ejecución de los incrementos.
- Ofrecer la solución completa. - Esta característica permite proporcionar los elementos que son necesarios para brindar un valor agregado a los usuarios del negocio. Por esto es indispensable contar con un almacén de datos sólido, bien diseñado, con accesibilidad y calidad probada.

2.3 BILL INMON

2.3.1 Descripción

En esta metodología se crea primero la data warehouse de la organización y luego se crean los data marts, considerando así todos los datos corporativos. [16]. En esta metodología existe la necesidad de transferir la información de los distintos OLTP (Sistemas Transaccionales) que tienen las organizaciones en un lugar centralizado en donde se puedan analizar los datos (sería el CIF o Corporate Information Factory)[18] .

En la figura 2.3 se presenta la Arquitectura de un Data WareHouse según Bill Inmon.

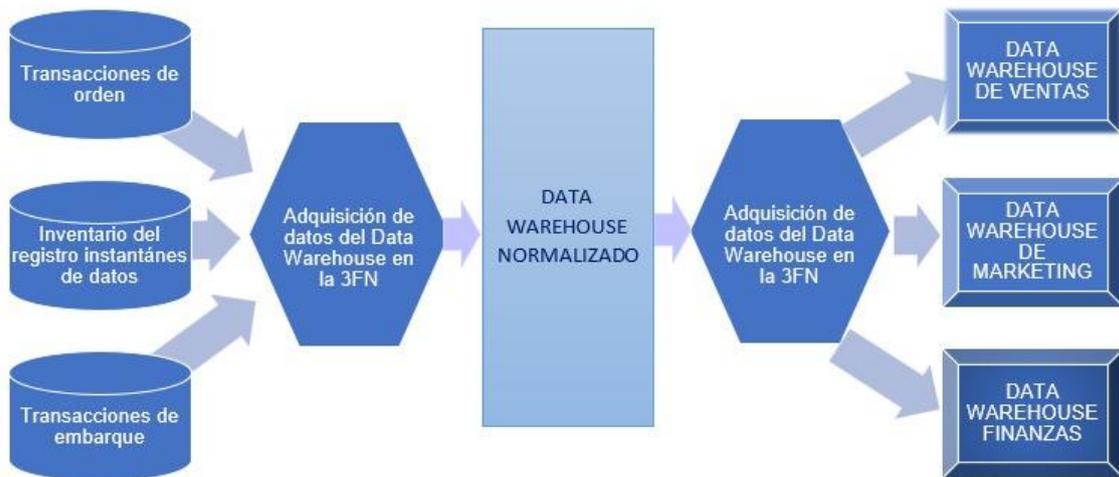


Figura 2.3 Arquitectura de un Data Warehouse según Bill Inmon [16]

2.3.2 Características

Esta metodología tiene las siguientes características [17]:

- Orientado a temas. - Los elementos y datos correspondientes a un mismo tema deben estar organizados con el fin que permanezcan unidos entre si dentro del almacén de datos.
- Variante en el tiempo. - Los datos expuestos a cambios en un periodo de tiempo se registrarán de tal manera que cuando se generen informes, los mismos reflejen dichas variaciones.
- No volátil. - No se modificará o eliminará la información, una vez que el dato haya sido almacenado, más bien el mismo se trasforma en información de sólo lectura, y se guardará para las consultas que se puedan realizar en un futuro.
- Integrado. - Los datos de los sistemas operacionales que existen en la organización, están contenidos en la base de datos, por lo cual estos deben ser consistentes [18].

2.4 CRISP-DM

2.4.1 Descripción

“CRISP - DM (Cross – Industry Standard Process for Data Mining)” es una metodología permite el monitoreo del cumplimiento de los objetivos a nivel empresarial sustentándose en la comprensión del negocio. Esta metodología se caracteriza por ser un modelo que trabaja con situaciones reales en las empresas [18].

Se describe a la metodología CRISP – DM como un modelo de proceso jerárquico distribuido en cuatro niveles [19]:

- Fase: Etapa dentro de la metodología.
- Tareas Genéricas: Descripción de las actividades que se realizan dentro de cada fase.
- Tarea Especializada: Descripción de cómo se pueden llevar a cabo las tareas genéricas en situaciones específicas.
- Instancias de Proceso: Son las acciones y resultados de las actividades realizadas en cada fase.

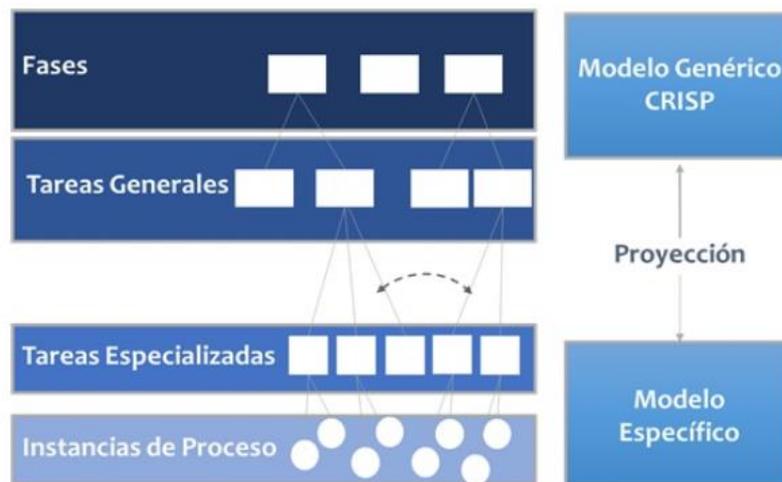


Figura 2.4 Esquema de los 4 niveles de CRISP-DM [17]

2.4.2 Características

Esta metodología tiene las siguientes características [19]:

- Metodología enfocada en Data Mining.
- Metodología basada en experiencias empresariales.
- Orientado a la comprensión del negocio.
- Basado en un modelo de procesos jerárquicos.
- Brinda confiabilidad, robustez y estandarización de los procesos.
- Facilita la creación de proyectos de BI/DM, de planificación y dirección, logrando así un correcto y buen seguimiento del proyecto.

2.4.3 Fases de CRISP-DM

Esta metodología cuenta con seis fases que se describen a continuación [20]:

- Comprensión del negocio (Fase donde son identificados los requerimientos y objetivos desde un enfoque no técnico).
- Comprensión de los datos (Fase donde se relaciona los datos con los objetivos del negocio).
- Preparación de los datos (Fase donde se obtiene el dataset o la vista minable).
- Modelado (Fase donde se aplica a los dataset las diferentes técnicas de minería de datos).
- Evaluación (Fase donde se determina si las técnicas de la fase anterior son útiles para solventar las necesidades del negocio).
- Implantación o Despliegue (Fase donde se explota la utilidad de los modelos, haciendo una integración en las tareas de toma de decisiones de la organización).

En la Figura 2.5 se muestra el proceso que sigue la metodología CRISP-DM.

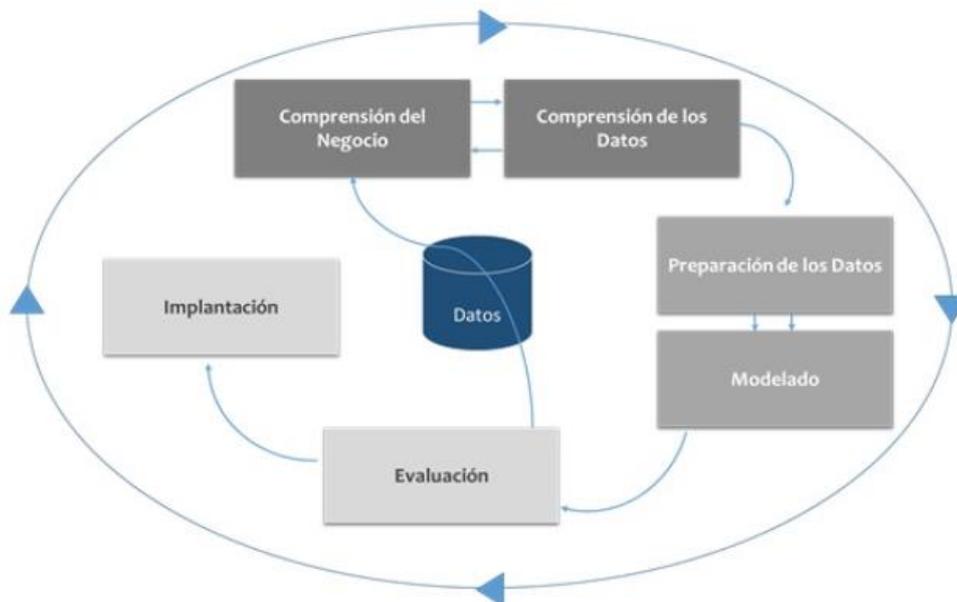


Figura 2.5 Modelo de proceso CRISP-DM [20]

2.5 JUSTIFICACIÓN DE LA METODOLOGÍA

2.5.1 Comparativa de las metodologías

Elegir una metodología que satisfaga las necesidades del negocio y de los usuarios es una tarea complicada, no existen parámetros específicos para evaluarlos, pues cada metodología tiene sus fortalezas y debilidades, y cada proyecto se atiene a los requerimientos del negocio.

Existen diferentes soluciones para cada requerimiento obtenido en la empresa, donde son analizados con objetividad algunos criterios como: el costo, la tecnología, la funcionalidad, la integración y otros, al momento de realizar una comparativa.

Para seleccionar la metodología adecuada, el equipo de proyecto (Anatoa Erika – Vernaza Marlon) plantea varios criterios con sus respectivos porcentajes, en base a la experiencia que se tiene con las metodologías y al entendimiento del tipo de proyecto a realizar, apoyados también en expertos del medio quienes realizaron la debida retroalimentación de cada una de las metodologías para su comprensión. En la tabla 2.1, se detalla los criterios que el equipo del proyecto definió y consideró más importantes los cuales permitirán el análisis y la selección de la metodología de BI:

Criterio	Definición	Porcentaje
C1: Funcionalidades que cumple	La metodología debe tener funcionalidades completamente detalladas, para que al aplicarlas den como resultado la completitud del proyecto.	40%
C2: Facilidad de uso	La metodología permite que su implementación se realice de forma más fácil, cómoda e intuitiva posible.	20%
C3: Conocimiento	Los tesistas deben tener una amplia información de la metodología; ya sea ésta adquirida a través de la experiencia o la educación.	20%

C4: Ajuste al tipo del proyecto	La metodología debe cumplir con el alcance del proyecto. (Tiempo, personal implicado, planeación estratégica de la organización)	20%
----------------------------------------	----------------------------------------------------------------------------------------------------------------------------------	-----

Tabla 2.1 Criterios de Metodologías

Es necesario precisar que la metodología va ligada a la tecnología con la que se va a trabajar, en BI la metodología seleccionada puede demandar aspectos esenciales relacionados a las características propias de la tecnología como: mejores prácticas, documentación, definición de componentes, etc.

Al momento de realizar la selección de una metodología es prescindible identificar la que mejor se adapte a las necesidades del proyecto, tomando en cuenta el alcance del mismo.

A continuación, se detallan los mecanismos de cuantificación de cada criterio en cada una de las metodologías:

- **Criterio 1:** El puntaje que se les asigna a las cuatro metodologías es diez ya que todas cumplen con el análisis del ciclo de vida de los datos lo que permitirá satisfacer las necesidades del negocio.
- **Criterio 2:** Todas las metodologías son genéricas lo que implica que son multiplataforma, pero se les asigna diferente puntaje por la siguiente razón:
 - **Kimball (8):** Esta metodología permite regresar al inicio del proceso, pero siempre que este haya terminado.
 - **Inmon (8):** Esta metodología es unidireccional lo cual no permite regresar a una fase anterior.
 - **Hefesto (8):** En esta metodología cuando una fase es culminada los resultados que fueron obtenidos, pasan a ser el punto de partida para iniciar la siguiente fase.
 - **CRISP-DM (10):** En esta metodología no es necesario seguir el orden de las fases, ya que la misma es bidireccional lo cual permite retroceder entre diferentes fases en caso de haber cometido un posible error.

- **Criterio 3:** El puntaje que se les asigna a las cuatro metodologías es ocho ya que la información adquirida de estas ha sido mediante la educación, pero no se tiene experiencia con ninguna.
- **Criterio 4:** El puntaje que se les asigna a las cuatro metodologías es diez debido que todo proyecto no finaliza una vez que la solución ha sido desplegada. La información obtenida en el proceso y solución desplegada pueden producir nuevas iteraciones del proyecto.

En la tabla 2.2 se observa la cuantificación de cada uno de los criterios.

Metodología	Criterio				Total
	C1	C2	C3	C4	
Kimball	10	8	8	10	9.2
Inmon	10	8	8	10	9.2
Hefesto	10	8	8	10	9.2
Crisp -DM	10	10	8	10	9.6

Tabla 2.2 Comparativa de Metodología

Después de analizar las Tablas 2.1 y 2.2 se obtiene como resultado que la metodología que se adapta mejor a los requerimientos del proyecto de titulación: “Diseño e implementación de una herramienta para gestión de información de caudales de captación y producción en la (EPMAPS) aplicando inteligencia de negocios.” es “CRIS SP-DM”.

2.6 METODOLOGIA CRISP-DM

CRISP-DM es una metodología que engloba una guía y contiene un modelo, que están organizados en seis fases, las mismas que pueden ser bidireccionales o no, por lo que de una fase se puede volver a una fase anterior para poder revisarla o corregirla en caso de ser necesario, permitiendo que el ciclo de las fases no sea ordenado [19].

2.6.1 Fases de CRISP- DM

A continuación, se detallan cada fase contenida en CRISP-DM.

1. Fase de comprensión del negocio o problema

Esta fase de CRISP-DM, es seguramente la más significativa ya que reúne las actividades que existen en la comprensión de los objetivos y los requisitos o requerimientos del proyecto a nivel empresarial o institucional, el propósito de esto es transformar los mismos en objetivos netamente técnicos dando así el plan de proyecto. Es necesario entender de forma más detallada cual es el problema que se quiere solventar, lo cual permitirá recoger los datos apropiados e interpretar de manera correcta los resultados.

En esta primera fase, es prescindible la posibilidad de poder transformar el conocimiento que se ha obtenido del negocio, en un plan previo cuyo fin sea el lograr los objetivos del negocio. A continuación, se detalla cada una de las tareas que forman parte del proceso de esta fase [20]:

- **Determinar los objetivos del negocio.** En esta tarea los objetivos son la determinación del problema que se desea solventar, el por qué surge la necesidad de utilizar Data Mining y realizar una definición de los criterios de éxito.
- **Evaluación de la situación.** Esta tarea califica el estado del entorno antes de empezar con el proceso de DM, tomando en cuenta algunos aspectos como: ¿Qué conocimiento previo existe acerca del problema?, ¿Existe la cantidad de datos suficiente para solventar el problema?, ¿se tiene una relación coste beneficio en la implementación de DM?, etc. En esta tarea se detallan los requisitos o requerimientos de las necesidades del negocio.
- **Determinación de los objetivos de DM.** Esta última tarea tiene como objetivo representar las metas del proyecto relacionadas a los objetivos del negocio, posteriormente en esta tarea hay que desarrollar un plan para el proyecto, donde se detalle tanto el proceso que se va a seguir como las técnicas que se aplicarán en dicho proceso.

La Figura 2.6 muestra el proceso que cumple la primera fase de CRISP-DM

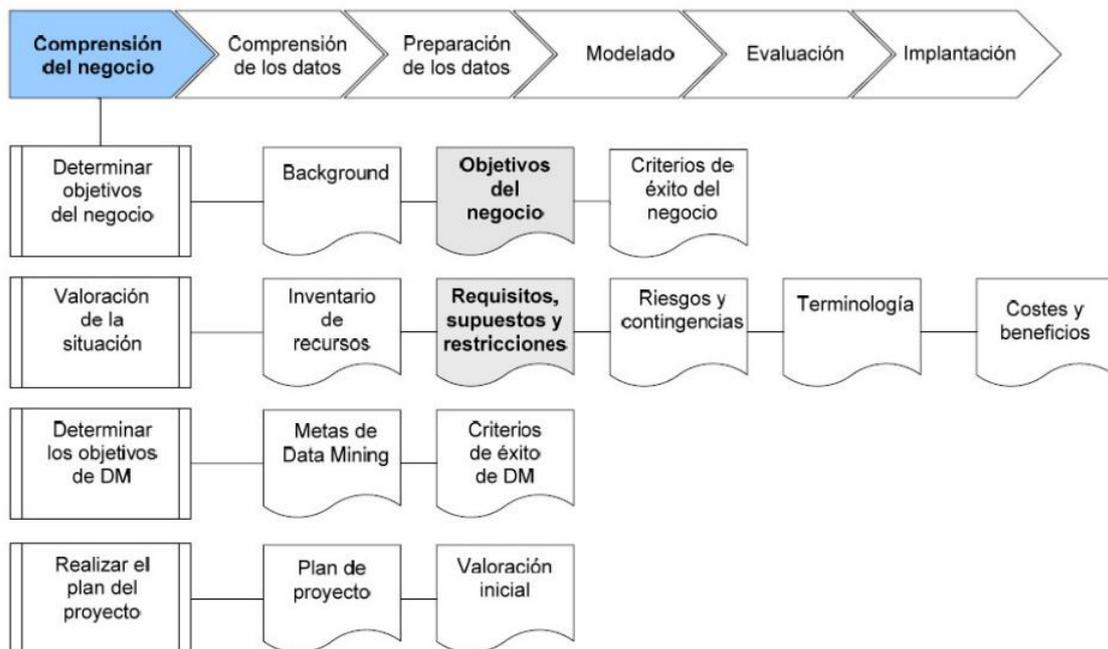


Figura 2.6 Fase de comprensión del negocio CRISP-DM [20]

2. Fase de comprensión de los datos

Esta fase de CRISP-DM, engloba la recolección de datos inicial, teniendo como propósito fijar un primer acercamiento con el problema a solventar, identificando la calidad para de esta manera establecer las relaciones más notables que permitirán determinar las hipótesis iniciales. Si en la empresa o institución existe una base de datos corporativa, es necesario generar una base de datos ad-hoc al proyecto de DM, pues durante el progreso de este, existe la posibilidad que se produzcan varios accesos a la base de datos para realizar consultas y eventualmente modificar la misma, y esto podría ocasionar varios inconvenientes en la base de datos corporativa. A continuación, se detalla cada una de las tareas que forman parte del proceso de esta fase [20].

- **Recolección de datos iniciales.** En esta primera tarea se recolectan los primeros datos y se los adecúa para procesamientos futuros. El objetivo de esta tarea es realizar informes listando los datos obtenidos, la localización, las técnicas que se emplearon para la recolección de estos, así como también los problemas y las soluciones adjuntos a dicho proceso.
- **Descripción de los datos.** Después que los datos iniciales han sido adquiridos estos se deben describir. En este proceso se establecen los volúmenes de datos, el

significado y la identificación de cada uno de sus atributos y también una descripción de su formato inicial.

- **Exploración de datos.** Posteriormente, surge la necesidad de realizar una exploración en los datos. En este proceso se debe hallar una estructura que sea general para todos los datos. Esto implica aplicar tests estadísticos básicos, los cuales revelarán características importantes de los datos que se han obtenido, y con esto se van creando tablas de distribución de frecuencias y construyendo los respectivos gráficos estadísticos. El entregable de esta tarea es un informe de la exploración realizada en los datos.
- **Verificación de la calidad de los datos.** A continuación, se ejecutan las verificaciones respectivas en los datos que se han obtenido, con el fin de establecer si los valores de los campos individuales son consistentes, si existen o no valores nulos, y a su vez identificando los valores que se encuentren fuera de rango, los que se pueden producir en el proceso si existe ruido. El objetivo en esta última tarea es establecer si los datos están completos y correctos.

La Figura 2.7 muestra el proceso que cumple la segunda fase de CRISP-DM.

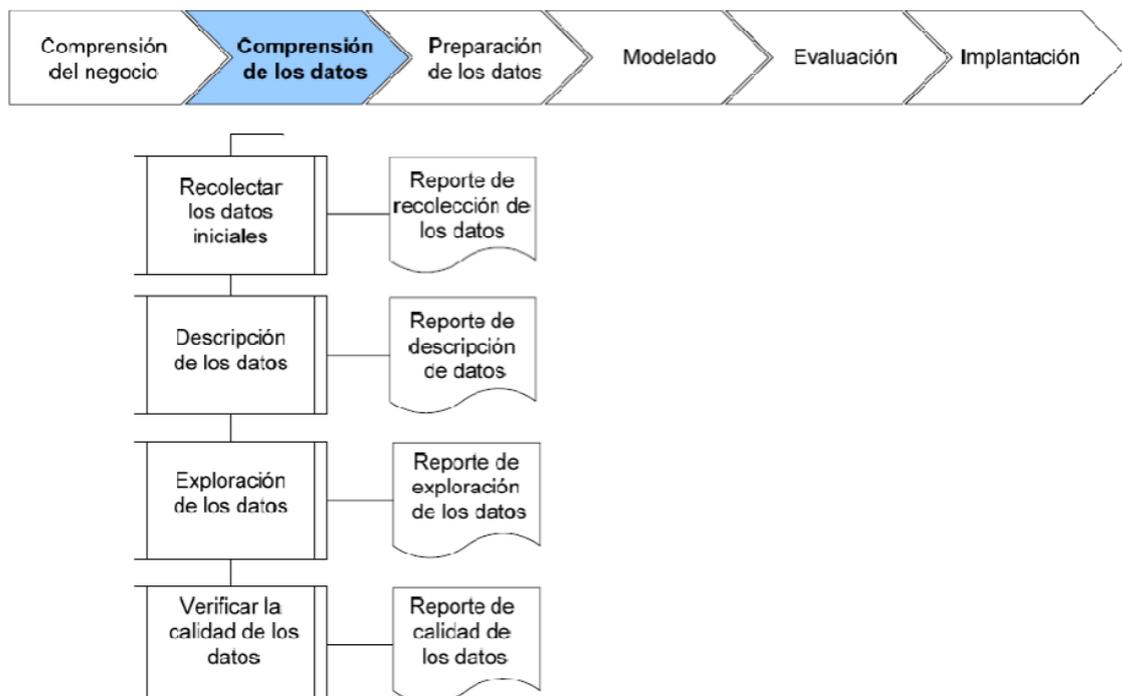


Figura 2.7 Fase de comprensión de datos CRISP-DM [20]

3. Fase de preparación de los datos

Esta fase de CRISP-DM después que se ha efectuado el proceso de recolección de datos, los prepara para adecuarlos a las distintas técnicas de DM, las mismas que serán utilizadas posteriormente, estas técnicas pueden ser: “de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos”. Esta tercera fase comprende diferentes tareas con respecto a la selección de datos en los que se aplicará una técnica de modelado específica, una limpieza, adicionar variables en caso de ser necesario, integrar datos de distintos orígenes y cambios en el formato de los datos.

Esta fase es conexas con la siguiente fase (Modelado), ya que, en base a la técnica de modelado seleccionada, los datos necesitan procesarse de distintas maneras. Por lo que estas fases están interactuando permanente. A continuación, se detalla cada una de las tareas que forman parte del proceso de esta fase [20]:

- **Selección de datos.** El propósito de esta tarea es seleccionar una parte específica de los datos que se han obtenido de la fase antes realizada, guiándose también en los criterios que se han establecido en las fases previas, como, por ejemplo: calidad de datos comprendiendo la corrección y completitud, y especificaciones en la cantidad y tipos de datos los mismos que tienen relación con las diferentes técnicas de DM seleccionadas.
- **Limpieza de los datos.** Esta segunda tarea está complementada con la tarea anterior, además de que la misma emplea el mayor tiempo y esfuerzo de esta fase, debido a que existen diversas técnicas de optimización para validar la calidad de los datos a fin de preparar los mismos para la fase de modelado.
Ciertas técnicas que son empleadas para efectuar la limpieza de datos son: normalizar los datos, discretizar campos numéricos, realizar tratamientos en los valores ausentes, reducir la cantidad de datos, etc.
- **Estructuración de los datos.** En esta tarea se realizan las operaciones de preparación de datos descritas a continuación: crear atributos nuevos en base a los atributos que ya existían, integrar registros nuevos o transformar valores para los atributos que ya existían.
- **Integración de los datos.** En esta tarea se crean estructuras nuevas, tomando como referencia datos específicos, con el propósito que se pueda crear nuevos campos tomando en cuenta los existentes, crear registros nuevos, realizar uniones en las tablas, en los campos o en tablas nuevas donde se especificarán las características de los diversos registros o campos de las tablas nuevas de resumen.

- **Formateo de los datos.** Esta tarea realiza las transformaciones sintácticas de datos sin alterar su significado, teniendo como meta facilitar o permitir la ejecución de una técnica específica de DM.

La Figura 2.8 muestra el proceso que cumple la tercera fase de CRISP-DM

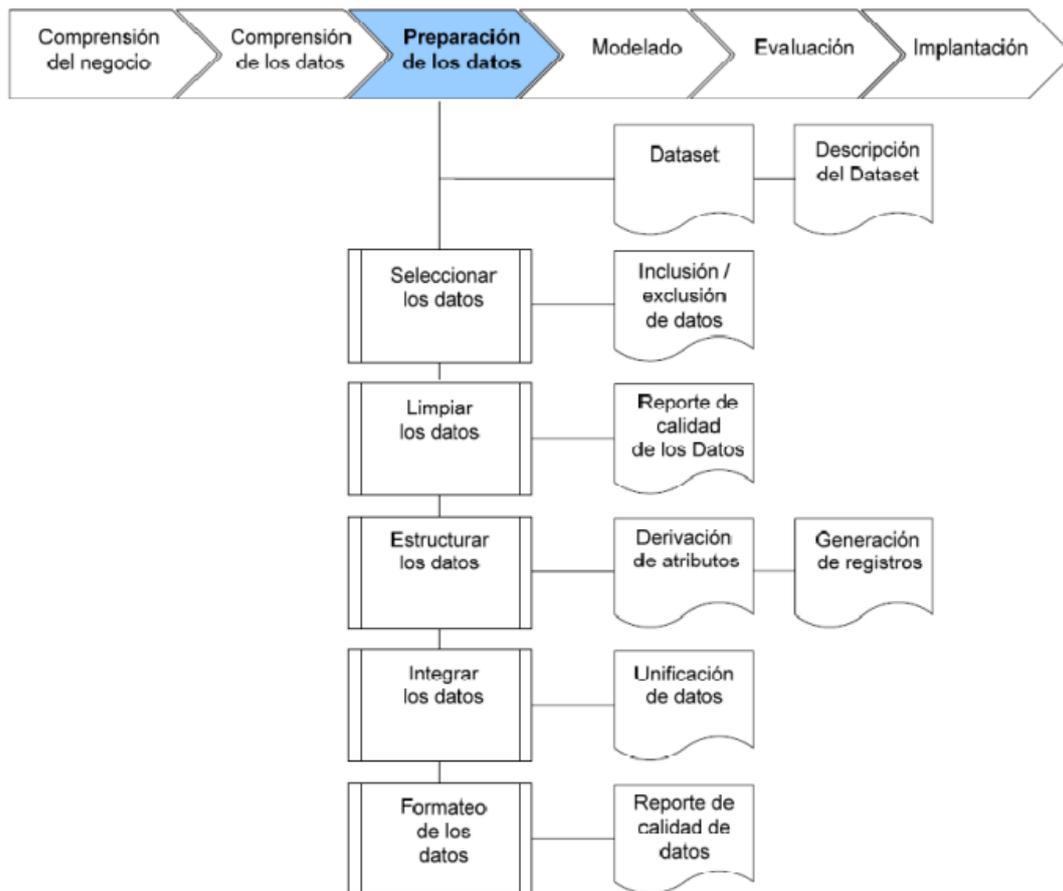


Figura 2.8 Fase de preparación de datos CRISP-DM [20]

4. Fase de modelado

Esta fase de CRISP-DM, realiza una selección de las técnicas de modelado más oportunas para el proyecto de DM. Los criterios que se toman en cuenta para determinar las técnicas con las que se trabajarán en esta fase son:

- “Ser apropiada al problema”.
- “Disponer de datos adecuados”.
- “Cumplir los requisitos del problema”.
- “Tiempo adecuado para obtener un modelo”.

- “Conocimiento de la técnica”.

Se debe establecer un método para evaluar los modelos permitiendo especificar el grado de utilidad de estos. Luego de finalizar con las tareas genéricas anteriores, es oportuno generar y evaluar el modelo. Las características de los datos y de precisión con el modelo influyen en los parámetros empleados para generar el modelo. El propósito de esta tarea es seleccionar la técnica más adecuada de DM para el tipo de problema a solventar. En dicha selección, es importante contemplar si el objetivo central del proyecto se relaciona con las herramientas de DM. A continuación, se detalla cada una de las tareas que forman parte del proceso de esta fase [20]:

- **Generación del plan de prueba.** Una vez que se construye un modelo, es necesaria la generación de un procedimiento que pruebe la calidad y validez del modelo. Por lo tanto, se divide en dos conjuntos de datos, uno de estos servirá para entrenamiento y el otro que servirá para pruebas, con la finalidad de construir un modelo el mismo que se basará en el primer conjunto de entrenamiento, midiendo la calidad del modelo que se generará con el conjunto que es de prueba.
- **Construcción del Modelo.** Después que se ha seleccionado la técnica, ésta se aplicará en los datos que previamente se prepararon para la generación de varios modelos. Estas técnicas de modelado poseen varios parámetros los cuales especifican las propiedades del modelo que se va a generar. Para seleccionar los parámetros más adecuados se realiza un proceso iterativo basándose únicamente en los resultados que han sido generados. Los mismos que en un futuro serán interpretados justificando también su rendimiento.
- **Evaluación del modelo.** En la última tarea de esta fase, los modelos son interpretados en base al conocimiento previo de los criterios de éxito y dominio preestablecidos anteriormente. Los expertos en DM evalúan los modelos dentro del entorno del dominio del problema aplicando sus propios criterios.

La Figura 2.9 muestra el proceso que cumple la cuarta fase de CRISP-DM.

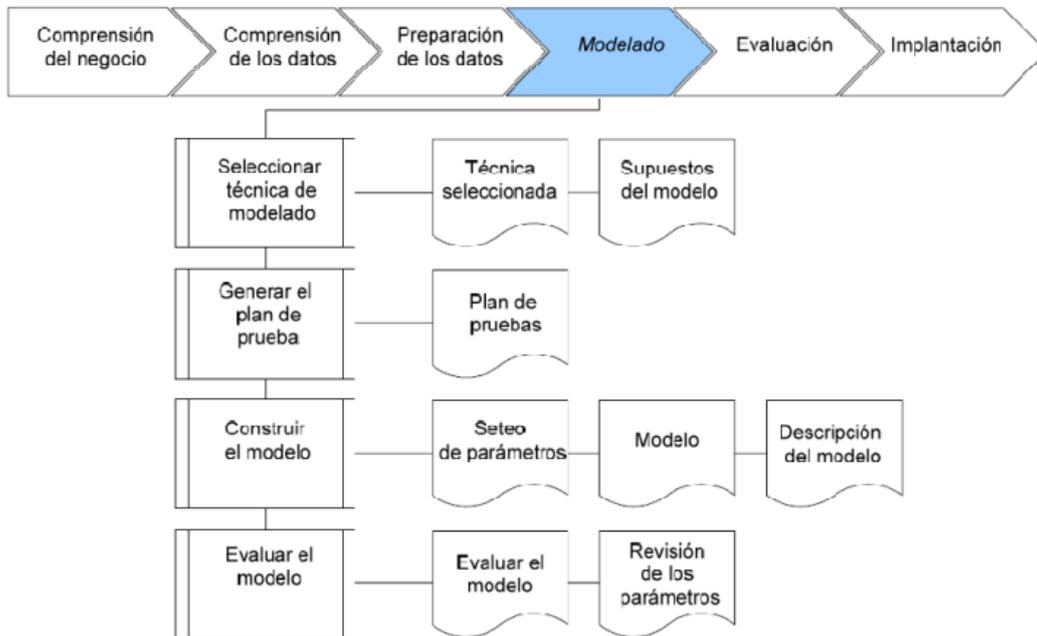


Figura 2.9 Fase de Modelado CRISP-DM [20]

5. Fase de evaluación

Esta fase de CRISP-DM es la encargada de la evaluación del modelo, donde el desempeño de los criterios de éxito del problema son la base primordial. Es necesario considerar, que el modelo es fiable únicamente para los datos analizados.

Es conveniente ir revisando el proceso, considerando los resultados que han sido obtenidos, ya que con esto existe la posibilidad de repetir pasos anteriores, donde se hayan cometido posibles errores. Hay que considerar que existe la posibilidad de aplicar diversas herramientas para interpretar los resultados obtenidos. Si se valida el modelo que ha sido generado en base a los criterios de éxito que fueron establecidos en la fase anterior, se comienza a explotarlo. A continuación, se detalla cada una de las tareas que forman parte del proceso de esta fase [20]:

- **Evaluación de los resultados.** Esta primera tarea engloba el proceso de evaluación del modelo basándose en los objetivos del negocio y analiza las razones en las que el modelo es de deficiente en el cumplimiento de las necesidades del negocio, o si por el contrario este modelo se puede probar, en una situación real considerando el tiempo y las restricciones.

- **Proceso de revisión.** Esta tarea califica completamente el proceso de DM, para de esta forma identificar si puede existir una mejora en los elementos.
- **Determinación de futuras fases.** En esta tarea, si se han obtenido los resultados esperados en las fases anteriores, se puede pasar a la siguiente fase, si este no es el caso entonces se puede decidir realizar una nueva iteración en las fases tres (Preparación de datos) y cuatro (Modelado) con diferentes parámetros. Inclusive es posible que se pueda decidir, empezar un nuevo proyecto de DM desde cero.

La Figura 2.10 muestra el proceso que cumple la quinta fase de CRISP-DM.

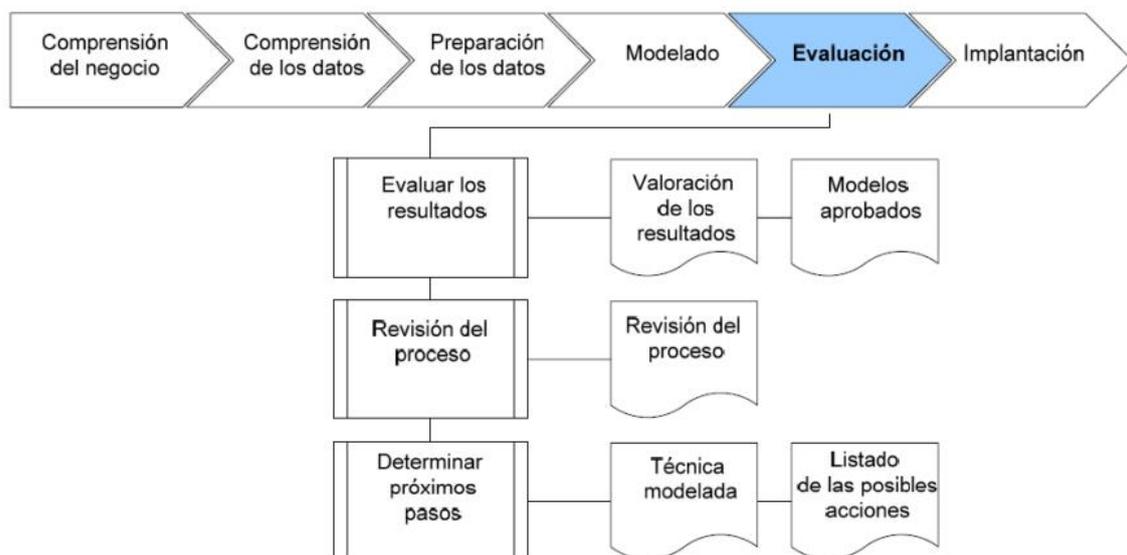


Figura 2.10 Fase de evaluación CRISP-DM [20]

6. Fase de implementación

En esta fase de CRISP-DM se toman acciones dentro del negocio a partir del conocimiento adquirido, recomendando acciones basadas en el análisis del modelo y en los resultados obtenidos del mismo. En general, cualquier proyecto de DM no finaliza en la implementación del modelo, ya que es necesario realizar la documentación y presentación de los resultados tomando en cuenta que estos deben estar detallados claramente para los usuarios finales, a fin de lograr un mejor conocimiento de los usuarios. Además, en esta fase es necesario que se asegure el mantenimiento de la aplicación y una posible comunicación de los resultados obtenidos. A continuación, se detalla cada una de las tareas que forman parte del proceso de esta fase [20]:

- **Plan de implementación.** En esta primera tarea con los resultados obtenidos de la evaluación se concluye la selección de la estrategia a ser implementada. En el momento que se identifica un procedimiento para creación del modelo, este se debe documentar para ser implementado posteriormente, y poder realizar su respectivo mantenimiento y monitorización. La retroalimentación que se genera al realizar el monitoreo y mantenimiento indicará si se está empleando adecuadamente el modelo.
- **Informe Final.** Es la parte final del proyecto de DM, y dependiendo del plan que se implementó, el informe puede contener los aspectos relevantes del proyecto y las experiencias aprendidas en el mismo, o también se puede realizar una presentación final donde se incluya los resultados que se han logrado durante el proyecto y se explique los mismos. Revisión del proyecto: En este punto se realiza una evaluación de los pros y los contras del proyecto, y de lo que se puede mejorar en el mismo.

La Figura 2.11 muestra el proceso que cumple la sexta fase de CRISP-DM

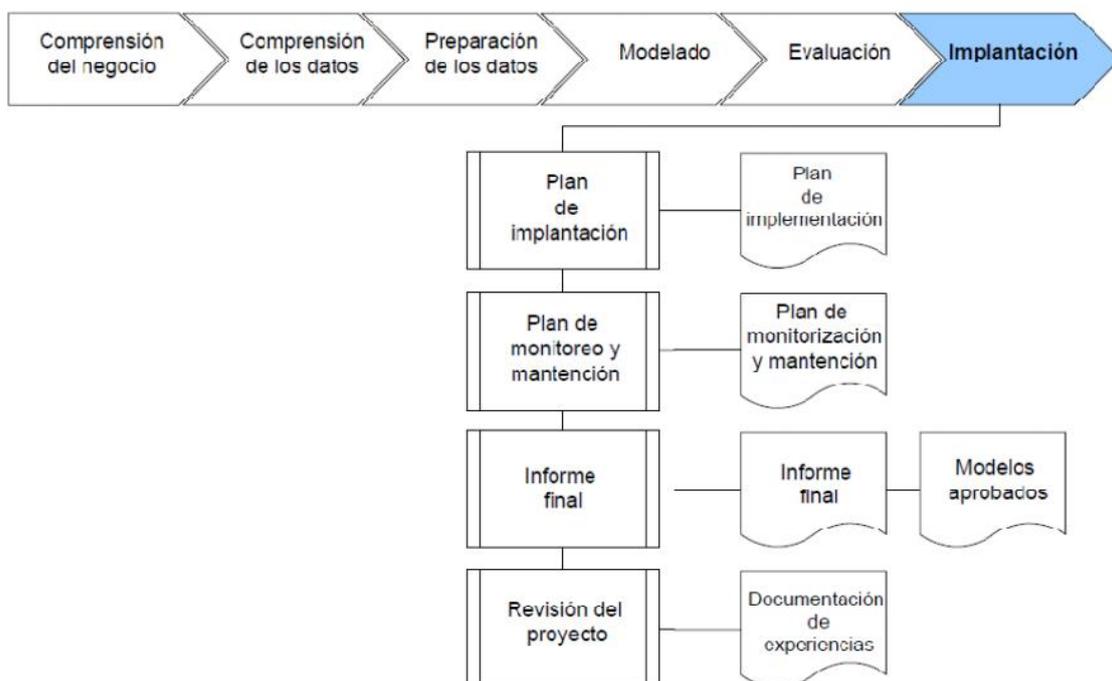


Figura 2.11 Fase de implementación CRISP-DM [20]

2.7 DESARROLLO

2.7.1 Comprensión del negocio

2.7.1.1 Determinación de los objetivos del negocio

Los objetivos de la EPMAPS son los siguientes [21]:

1. “Mejorar la satisfacción de la comunidad superando estándares regionales de servicio”.
2. “Asegurar la disponibilidad de los servicios”.
3. “Alcanzar la sostenibilidad de la empresa con Equidad Social”.
4. “Optimizar la operación y mantenimiento de la infraestructura”.
5. “Aumentar y mejorar la infraestructura de los servicios”.
6. “Reducir el impacto de los riesgos naturales y antrópicos”.
7. “Optimizar los procesos empresariales y la efectividad laboral”.
8. “Descontaminar de agua residuales los cauces naturales”.
9. “Mejorar la atención al cliente”.
10. “Incrementar la eficacia de la comunicación”.
11. “Generar ingresos/ahorros por la innovación del modelo de servicio”.
12. “Asegurar la disponibilidad del recurso hídrico en el mediano y largo plazo”.
13. “Mejorar las competencias del talento humano”.
14. “Mejorar el clima laboral”.
15. “Implementar una cultura de seguridad y salud”.
16. “Contar con información empresarial integrada y oportuna”.
17. “Promover una cultura de gestión organizacional alineado a la estrategia”.

2.7.1.2 Evaluación de la situación

Para la conocer los requerimientos del usuario final se realiza una serie de entrevistas en las que se busca identificar los objetivos de la organización. Con las respuestas obtenidas, se realiza un análisis con el fin de identificar los indicadores y perspectivas que son tomadas en cuenta para la construcción de los dashboards. El resultado final de esta fase es la elaboración del modelo conceptual.

Entrevista a nivel estratégico

El 27 de diciembre del 2019 se llevó acabo la reunión con el Departamento de Investigación Desarrollo e Innovación. El principal objetivo de esta reunión es obtener e identificar las principales necesidades de la organización, así como los aspectos necesarios para llevar a cabo la toma de decisiones a nivel estratégico.

Actividades:

1. Identificar la situación actual de la EPMAPS a nivel estratégico.

Problemática:

- Actualmente la EPMAPS trabaja con procesos manuales (Excel) para identificar el volumen de la producción de agua en cada una de las plantas de tratamiento.
- Al momento no cuentan con reportes de la producción de agua de cada de una de las plantas.

Necesidad:

- El requerimiento del Departamento de Investigación Desarrollo e Innovación es tener un reporte del proceso de producción de agua con periodicidad mensual con el fin de asegurar la cobertura del servicio de agua potable.

Identificación de Indicadores y Perspectivas

Cantidad de volumen (m³) del caudal de producción de agua con periodicidad mensual.

Perspectiva: Caudal.

Indicador: Volumen de producción de agua (m³).

Perspectiva: Periodicidad mensual.

Entrevista a nivel Táctico

El 03 de enero del 2019 se llevó acabo la reunión con el Departamento de Sistemas de Información Geográfica y SCADA. El principal objetivo de esta reunión es obtener e identificar las principales necesidades de la organización, así como los aspectos necesarios para llevar a cabo la toma de decisiones a nivel Táctico.

Actividades:

1. Identificar la situación actual de la EPMAPS a nivel táctico.

Problemática:

- Actualmente la EPMAPS no cuenta con reportes del proceso de captación que permitan conocer cuánta agua se capta en cada planta por lo que solo se tiene un sensor en el ingreso principal de estas; y como también cuentan con entradas secundarias, el departamento correspondiente realizará el análisis del porcentaje de dichas entradas con la comparativa de producción.

Necesidad:

- El requerimiento del Departamento de Sistemas de Información Geográfica y SCADA es tener un reporte del proceso de captación llevando un monitoreo y control mensual con el fin que el departamento correspondiente tome las decisiones necesarias (repotenciación o cierre de la planta).

Identificación de Indicadores y Perspectivas

Cantidad de volumen (m³) del caudal de captación de agua con periodicidad mensual.

Perspectiva: Caudal

Indicador: Volumen de captación de agua (m³).

Perspectiva: periodicidad mensual.

En síntesis, los indicadores resultantes de este proceso de abstracción son:

- Volumen de captación de agua
- Volumen de producción de agua

A su vez, las perspectivas de análisis abstraídas son:

- Periodicidad mensual

2.7.1.3 Determinación de los objetivos

- Conocer el volumen de agua que ingresa en m³ cada mes y analizar los diferentes factores que influyen en el proceso de captación de agua.
- Conocer el volumen de agua que se produce en m³ cada mes y analizar si las plantas están produciendo lo adecuado para de esta manera saber si cada planta necesita repotenciarse o cerrar.

- Realizar una comparativa entre los procesos de captación y la producción de agua para que de esta forma la EPMAPS internamente realice el análisis del ingreso por otras entradas en las que no se encuentran los sensores.

2.7.2 Comprensión de los datos

2.7.2.1 Recolección de datos iniciales

La información que se va a utilizar en este proyecto, son datos referentes a caudales de los procesos de captación y producción. Por lo que en una primera instancia se iba a trabajar con la base de datos interna de la empresa, en conjunto con el personal de administración de base de datos; pero se presentaron ciertos inconvenientes:

- Las plantas no tenían conectividad con la matriz por lo cual la base de datos no estaba recibiendo la información requerida.
- Se analizó los datos históricos de las plantas, cuando éstas tenían conectividad con la matriz; pero al parecer la comunicación no era eficiente y se presentó datos repetidos o nulos.
- Adicional a esto, no se tenía la misma nomenclatura de los procesos de captación y producción en las plantas, y al parecer este es un proyecto que se está implementando actualmente.

Analizando los inconvenientes mencionados anteriormente, se optó por crear una base de datos local con la misma versión con la que trabaja la empresa, además se utiliza la nomenclatura implementada para los procesos de captación y producción. Los datos serán insertados manualmente con la información que emite el sistema SCADA a través de un Excel.

A continuación, listamos los datos adquiridos:

Plantas

- Codificación UT (Ubicaciones Técnicas) de cada planta.
- Nombres específicos de las plantas de tratamiento.
- Dirección
- Sector
- Zona

Captación

- Codificación UT para captación
- Volumen de agua
- Unidad del volumen
- Fecha

Producción

- Codificación UT para captación
- Volumen de agua
- Unidad del volumen
- Fecha

2.7.2.2 Descripción de los datos

Los datos se encuentran en una base local SQL Server 2016 ya que es la versión con la que trabaja la empresa. Se crea un esquema dimensional en estrella. En la figura 2.12 se presenta el esquema relacional de esta base de datos, la misma que fue generada con Case Studio 2.

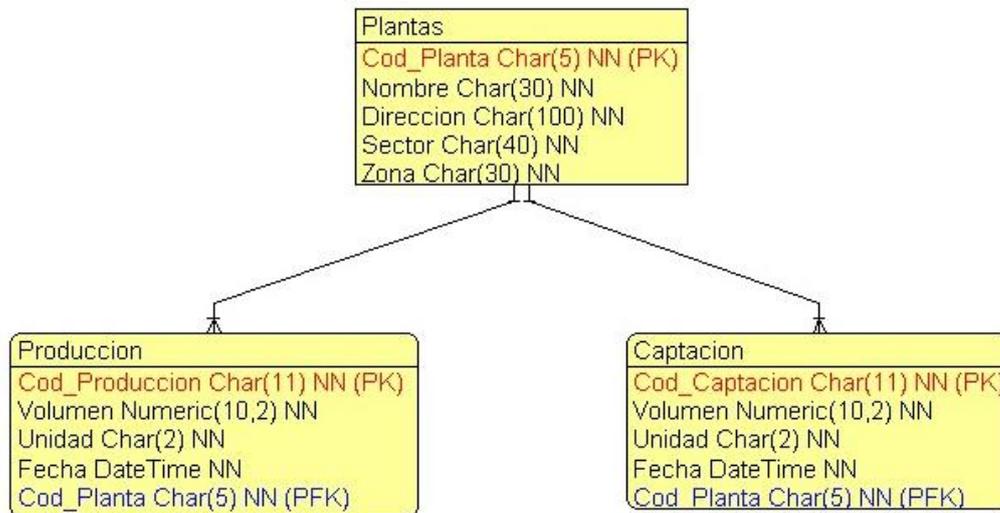


Figura 2.12 Esquema relaciona de la base de datos

En esta figura se muestra claramente que el almacén de datos consta de tres tablas:

- Plantas
- Captación
- Producción

A continuación, se describe cada una de ellas detallando los campos que contienen las mismas.

Tabla dbo.Plantas

Esta es la tabla central del almacén de datos, también llamada “dbo.Plantas”, ya que es en esta tabla en la que se registra la información acerca de las plantas. Esta tabla al ser la tabla central tiene como clave primaria Cod_Planta del resto de las tablas, llamadas tablas dimensionales.

Esta clave es a la vez la clave foránea (foreign key). Esta tabla tiene un total de 4 registros. Los campos de cada registro de esta tabla son:

- **Cod_Planta.** Tipo char. Este campo identifica a cada planta y es único teniendo una longitud máxima de 5 caracteres.
- **Nombre.** Tipo char. Este campo indica el nombre de la planta este campo y tiene una longitud máxima de 30 caracteres.
- **Direccion.** Tipo char. Este campo indica la dirección de la planta con una longitud máxima de 100 caracteres.
- **Sector.** Tipo char. Este campo indica el sector donde está ubicada la planta con una longitud máxima de 40 caracteres.
- **Zona.** Tipo char. Este campo indica la zona específica de la planta con una longitud máxima de 30 caracteres.

Tabla dbo.Captacion

Esta tabla contiene la información acerca del volumen de agua que ingresa en el proceso de captación. Su clave primaria es:

- Cod_Captacion

Tiene un total de 38 registros. Los campos de cada registro de esta tabla son:

- **Cod_Captacion.** Tipo char. Este campo identifica a cada registro con un código donde se encuentra el código de la planta, el código de captación y la numeración secuencial, teniendo una longitud máxima de 11 caracteres.
- **Cod_Planta.** Tipo char. Este campo identifica a cada planta ya que es la clave foránea que se extrae de la tabla Plantas tiene una longitud máxima de 11 caracteres.
- **Volumen.** Tipo numérico. Este campo indica un número decimal con 10 enteros y 2 decimales y representa el volumen de agua captada en m³ de cada planta.

- **Unidad.** Tipo char. Este campo identifica la unidad de medida en la que se encuentra el volumen de captación de agua, teniendo una longitud máxima de 2 caracteres.
- **Fecha.** Tipo fecha. Este campo indica la fecha de captación de agua en la planta de tratamiento. El formato de la fecha es año-mes-día.

Tabla dbo.Produccion

Esta tabla contiene la información acerca del volumen de agua que se produce. Su clave primaria es:

- Cod_Produccion

Tiene un total de 38 registros. Los campos de cada registro de esta tabla son:

- **Cod_Produccion.** Tipo char. Este campo identifica a cada registro con un código donde se encuentra el código de la planta, el código de producción y la numeración secuencial, teniendo una longitud máxima de 11 caracteres.
- **Cod_Planta.** Tipo char. Este campo identifica a cada planta ya que es la clave foránea que se extrae de la tabla Plantas tiene una longitud máxima de 11 caracteres.
- **Volumen.** Tipo numérico. Este campo indica un número decimal con 10 enteros y 2 decimales y representa el volumen de agua producida en m³ de cada planta.
- **Unidad.** Tipo char. Este campo identifica la unidad de medida en la que se encuentra el volumen de producción de agua, teniendo una longitud máxima de 2 caracteres.
- **Fecha.** Tipo fecha. Este campo indica la fecha de producción de agua en la planta de tratamiento. El formato de la fecha es año-mes-día.

2.7.2.3 Exploración de los datos

Una vez que los datos han sido descritos, se procede a explorarlos, aplicando pruebas estadísticas básicas, ya que de esta forma se podrán obtener las propiedades de dichos datos, estas pueden ser tablas de frecuencia, gráficos estadísticos, etc. Este proceso servirá para establecer la consistencia y completitud de los datos. Mediante la herramienta Power BI se realiza un perfilamiento de datos en el cual se verifica la calidad de estos (existencia de datos nulos o repetidos). A continuación, se presentan las imágenes del perfilamiento de datos de cada tabla:

Las Figuras 2.13, 2.14, 2.15 muestran el perfilamiento de las diferentes tablas de la base de datos.

- **Tabla dbo.Plantas**

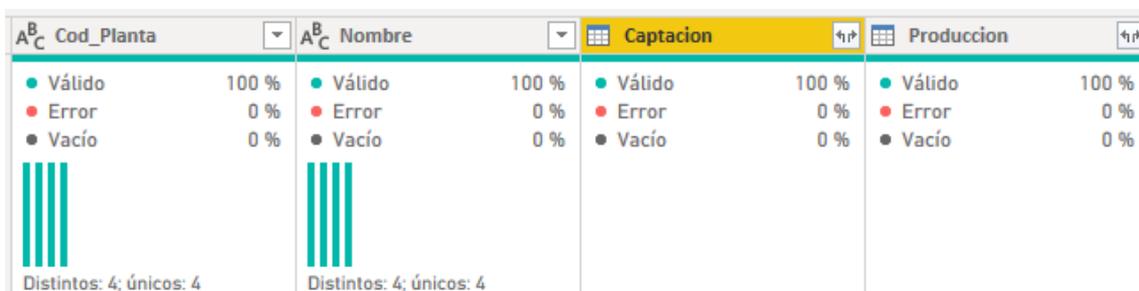


Figura 2.13 Perfilamiento de datos Tabla Plantas

- **Tabla dbo.Captacion**



Figura 2.14 Perfilamiento de datos Tabla Captación

- **Tabla dbo.Produccion**

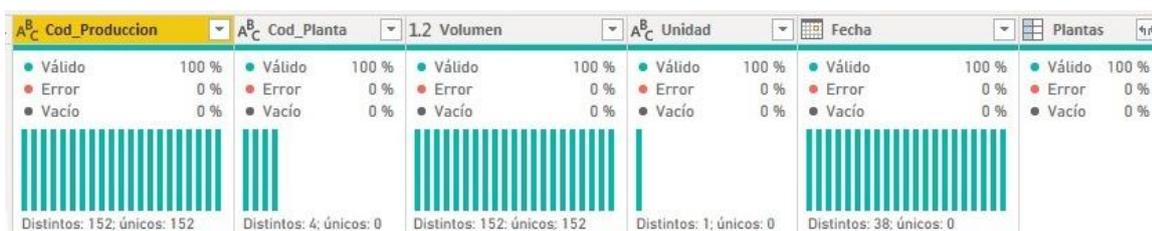


Figura 2.15 Perfilamiento de datos Tabla Producción

Además, mediante consultas SQL que se encuentran en el Anexo 3, se han obtenido los datos necesarios para realizar las gráficas que se explican a continuación las mismas que fueron multiplicadas por una variable para salvaguardar la confidencialidad de la empresa.

La figura 2.16 muestra la obtención de volumen de captación de todas las plantas por fecha, en esta figura se aprecia el volumen de captación de cada una de las plantas tomando como referencia la fecha 31 de enero del 2017.

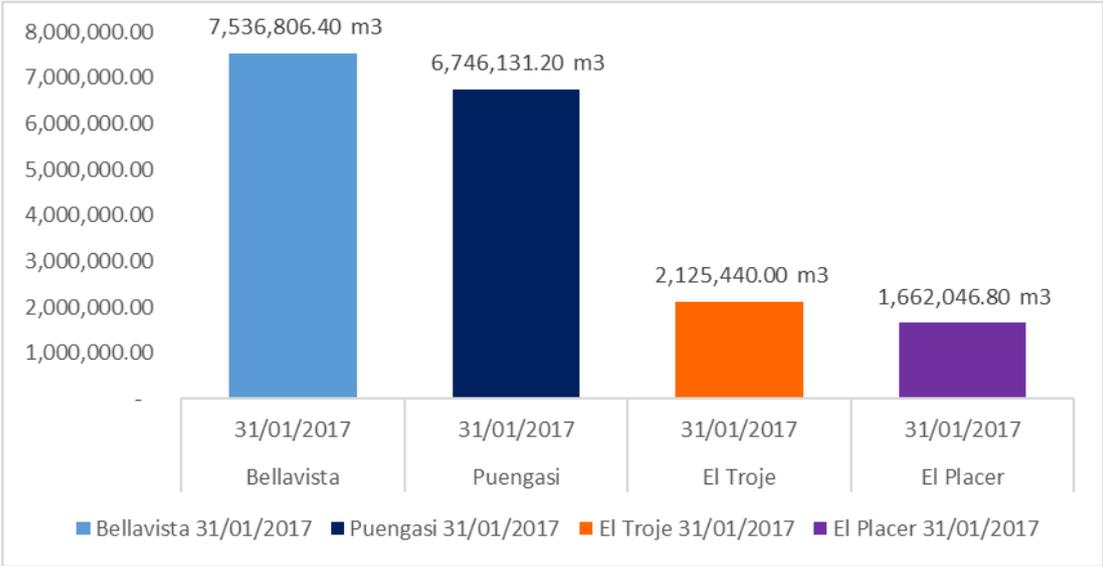


Figura 2.16 Volumen de captación por planta en la fecha 31/01/2017

La figura 2.17 muestra la obtención del volumen de captación de una planta por fecha, en esta figura se aprecia el volumen de captación de la planta El Troje tomando como referencia la fecha 31 de enero del 2017.

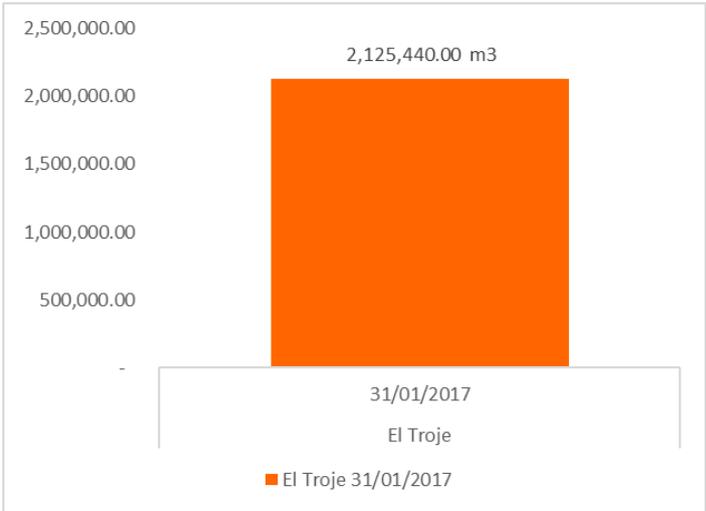


Figura 2.17 Volumen de captación de la planta El Troje en la fecha 31/01/2017

La figura 2.18 muestra la obtención de la suma del volumen de captación de una planta de todos los años, en esta figura se aprecia el volumen de captación de la planta El Troje en los años 2017, 2018, 2019, y los dos primeros meses del 2020.

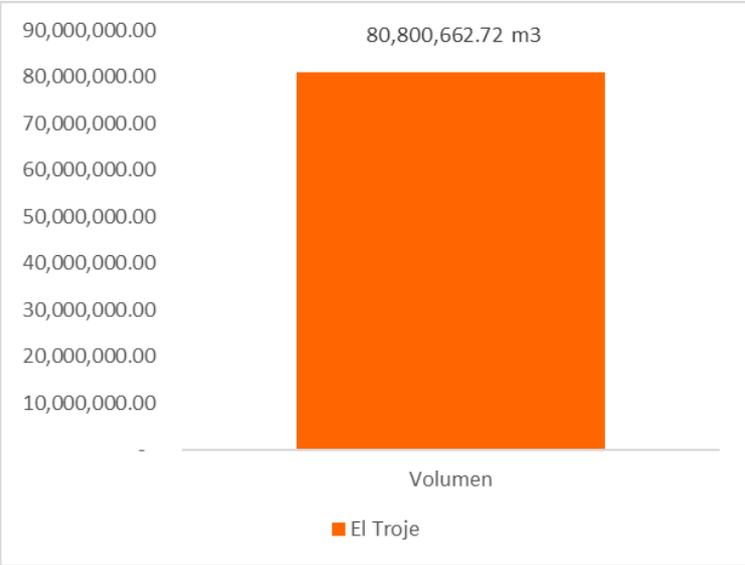


Figura 2.18 Volumen de captación de la planta El Troje de todos los años

La figura 2.19 muestra la obtención de la suma de volúmenes de captación de todas las plantas por fecha, en esta figura se aprecia el volumen total de captación de todas las plantas tomando como referencia la fecha 31 de enero del 2017.

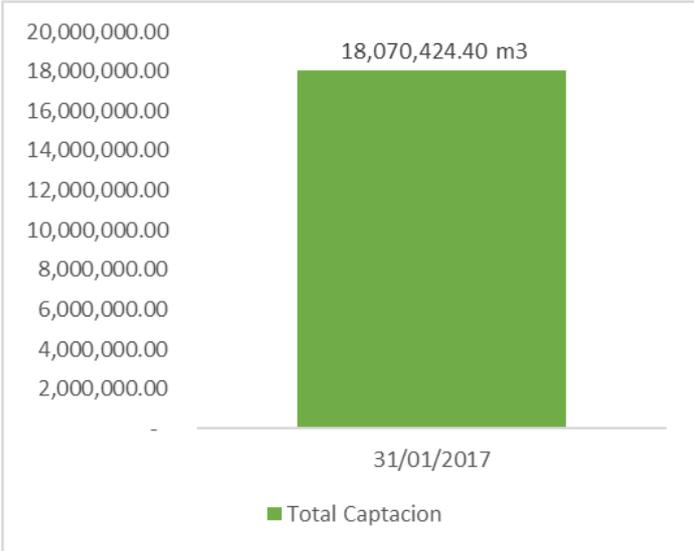


Figura 2.19 Volumen de captación total de todas las plantas en la fecha 31/01/2017

La figura 2.20 muestra la obtención del porcentaje del volumen de captación de todas las plantas en un rango de fechas, en esta figura se aprecia el porcentaje del volumen de captación de todas las plantas entre las fechas 31 de enero del 2017 y 31 de marzo del 2017.

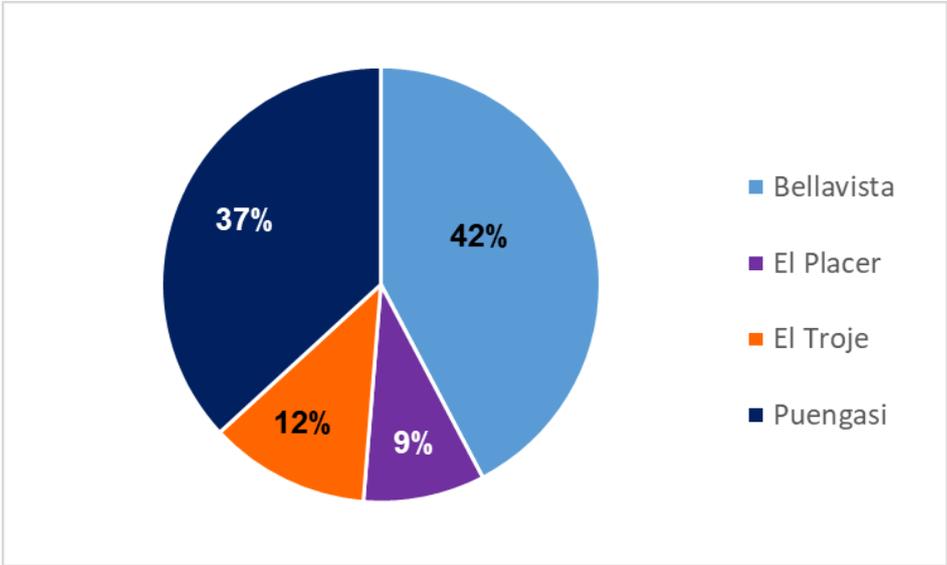


Figura 2.20 Porcentaje del volumen de captación de todas las plantas en un rango de fechas

La figura 2.21 muestra la obtención del volumen de producción de todas las plantas por fecha, en esta figura se aprecia el volumen de producción de cada una de las plantas tomando como referencia la fecha 30 de abril del 2018.

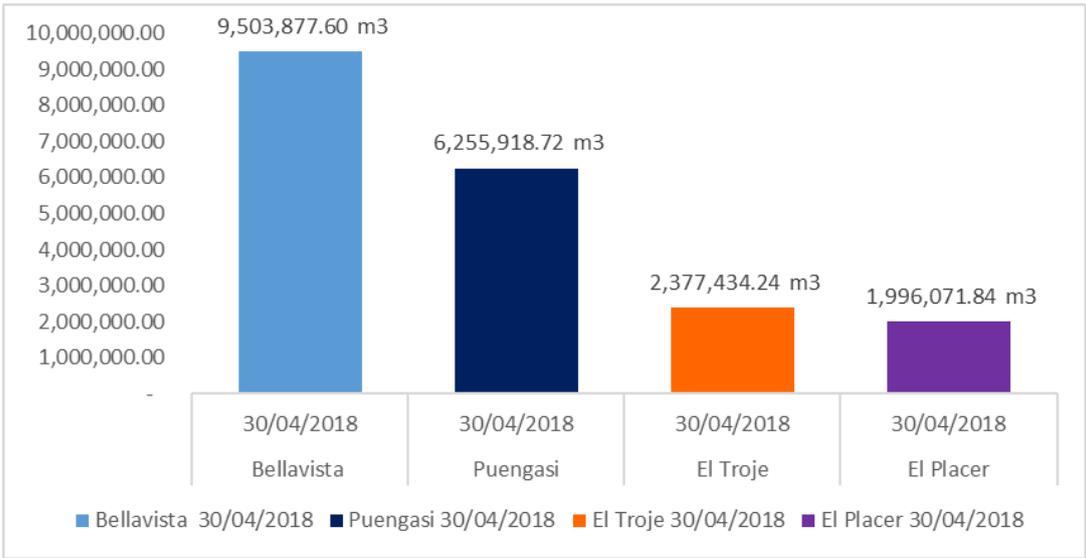


Figura 2.21 Volumen de producción por planta en la fecha 30/04/2018

La figura 2.22 muestra la obtención del volumen de producción de una planta por fecha, en esta figura se aprecia el volumen de producción de la planta Bellavista tomando como referencia la fecha 30 de abril del 2018.

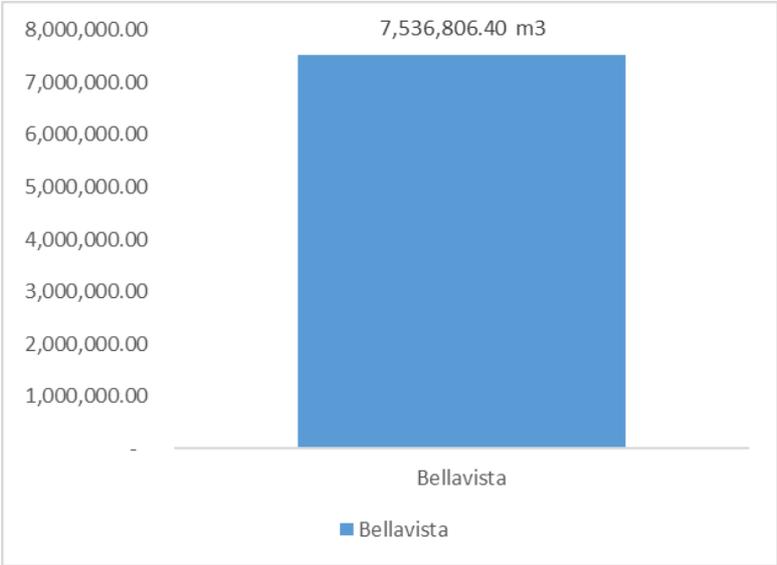


Figura 2.22 Volumen de producción de la planta Bellavista en la fecha 31/04/2018

La figura 2.23 muestra la obtención de la suma del volumen de producción de una planta de todos los años, en esta figura se aprecia el volumen de producción de la planta Bellavista en los años 2017, 2018, 2019, y los dos primeros meses del 2020.

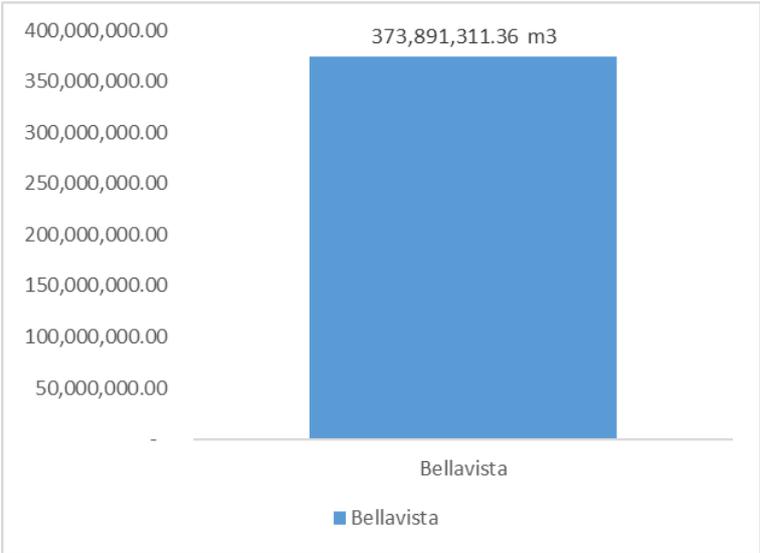


Figura 2.23 Volumen de producción de la planta Bellavista de todos los años

La figura 2.24 muestra la obtención de la suma de volúmenes de producción de todas las plantas por fecha, en esta figura se aprecia el volumen total de producción de todas las plantas tomando como referencia la fecha 30 de abril del 2018.

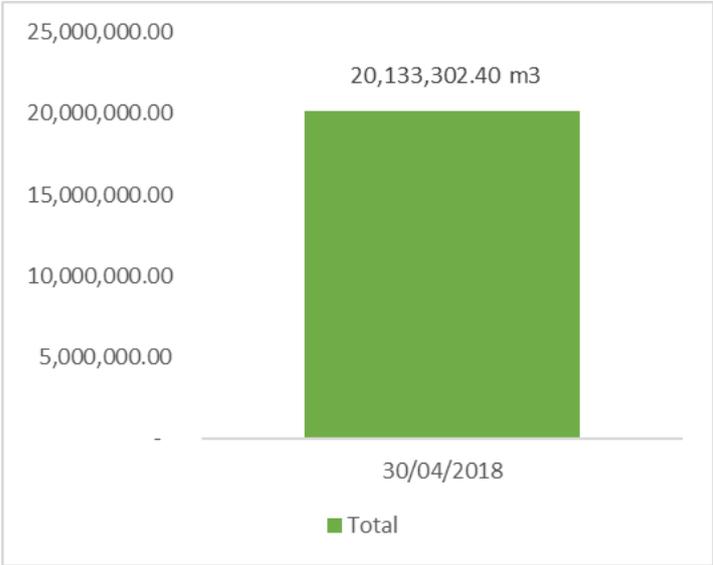


Figura 2.24 Volumen de producción total de todas las plantas en la fecha 30/04/2018

La figura 2.25 muestra la obtención del porcentaje de volumen de producción de todas las plantas en un rango de fechas, en esta figura se aprecia el porcentaje del volumen de producción de todas las plantas entre las fechas 30 de abril del 2018 y 30 de junio del 2018.

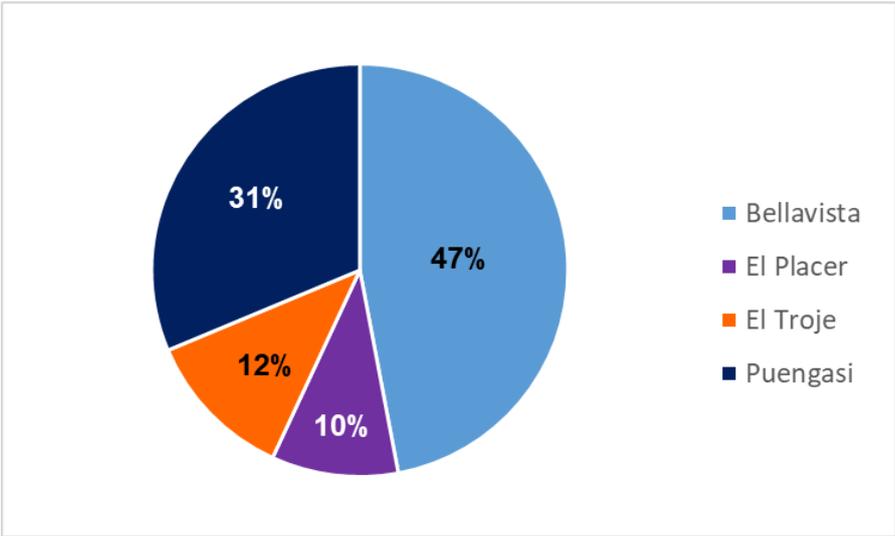


Figura 2.25 Porcentaje del volumen de producción de todas las plantas en un rango de fechas

La figura 2.26 muestra la obtención de volumen de captación y producción de una planta en un rango de fechas, en esta figura se aprecia la comparativa entre los volúmenes de producción y captación de la planta El Placer entre las fechas 30 de septiembre del 2019 y 31 de diciembre del 2019.

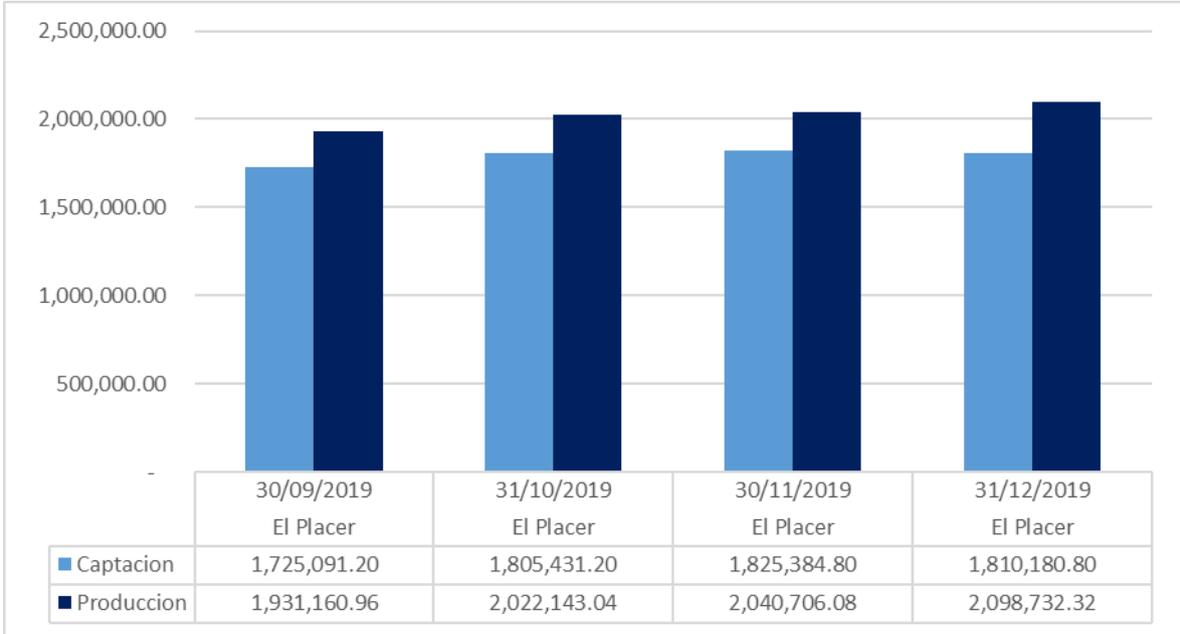


Figura 2.26 Comparativa entre los volúmenes de producción y captación de la planta El Placer en un rango de fechas

2.7.2.4 Verificación de la calidad de los datos

Al finalizar la exploración inicial de los datos se procede a realizar una revisión junto con el respectivo personal de la EPMAPS para asegurar la calidad de los datos y se llega a la conclusión que estos son los correctos para el desarrollo del proyecto, debido que cumplen con los requerimientos de cada caso porque al obtener los resultados esperados se tomarán decisiones para beneficio de la empresa. Los datos no contienen errores, ya que estos fueron generados automáticamente por el sistema SCADA. En cuanto a los valores nulos, no se encontró ninguno por cuanto la base fue creada localmente con la información necesaria.

2.7.3 Preparación de los datos

2.7.3.1 Selección de datos

Se van a utilizar todos los registros dentro de cada tabla que contiene la base de datos, ya que la misma fue creada para propósitos específicos de la empresa, la cantidad de registros

que se han insertado en la base de datos fueron esenciales para los análisis estadísticos que realiza la empresa. Pero existen algunos campos dentro de estos registros que no son de relevancia para la toma de decisiones, por lo que se puede prescindir de estos.

Los campos seleccionados para el análisis son los siguientes:

Tabla Plantas

- Cod_Planta
- Nombre

Tabla Captación

- Cod_Captacion
- Volumen
- Fecha

Tabla Producción

- Cod_Produccion
- Volumen
- Fecha

2.7.3.2 Limpieza de datos

La base de datos con la que se trabaja para el desarrollo del proyecto consta de la información necesaria para poder cumplir con una correcta toma de decisiones, son datos limpios y por tanto no existe la necesidad de hacer una limpieza más profunda sobre ellos, tampoco se presentan campos en los que falten valores.

2.7.3.3 Estructuración e integración de los datos

En esta etapa se ve la necesidad de crear una tabla Calendario con el campo Fecha y Fecha (ubicaciones) generada por el Power BI como se visualiza en la Figura 2.27. Esta tabla sirve para relacionar las fechas de las tablas de captación y producción con el fin de que la comparativa entre las dos tablas contenga el mismo rango de fechas.

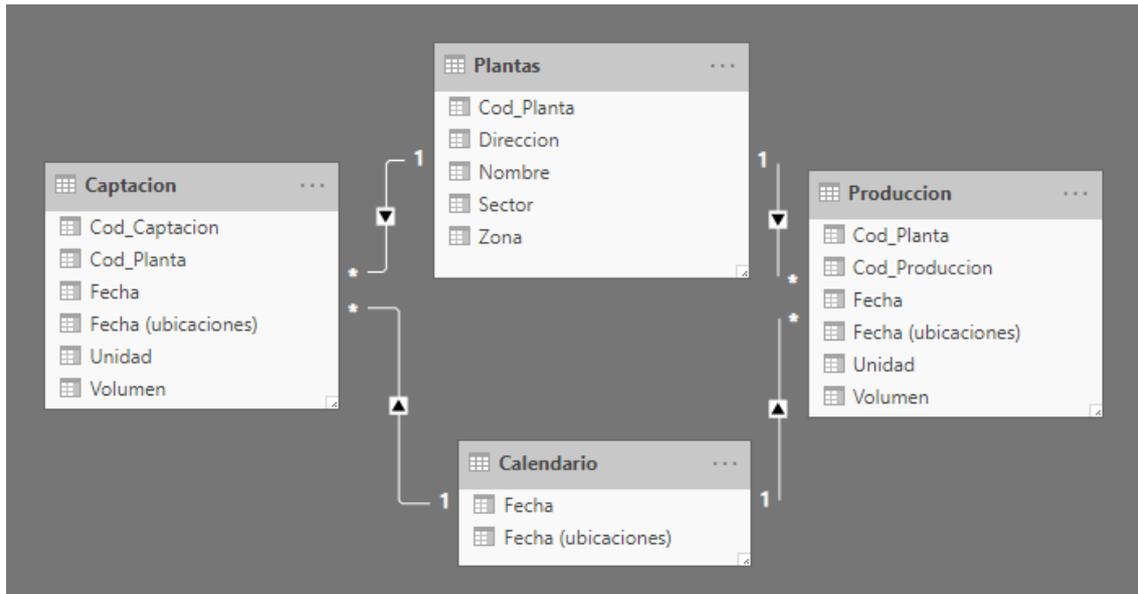


Figura 2.27 Integración de nuevos datos

Además, en las tablas Captación y Producción se crea el campo Fecha (ubicaciones) con el objetivo que la gráfica estadística de cada proceso muestre con mayor detalle la fecha para un mejor entendimiento de la misma.

2.7.3.4 Formateo de los datos

En esta etapa no fue necesario formatear los datos de ningún campo dentro de los registros, ya que los mismos fueron normalizados en el momento de insertarlos en la base de datos. Se utilizó la normalización proporcionada por la empresa, como se muestra en el Anexo2.

2.7.4 Modelado

En esta fase se adaptará la metodología a las necesidades del proyecto, ya que en el mismo no se realiza predicciones, más bien es una herramienta de BI para toma de decisiones, por lo que no se seguirá las etapas descritas en esta fase en el capítulo 2.

2.7.4.1 Especificación de características de los reportes

El miércoles 29 de enero del 2020 se lleva a cabo una reunión con el Departamento de Investigación Desarrollo y el Departamento de Sistemas de Información Geográfica y SCADA de la EPMAPS para la especificación de los reportes y la información detallada que se presenta dentro de los mismos. Se establece las siguientes características específicas de cada uno de ellos como se presenta continuación:

- **Reporte “Captación EPMAPS”**: primer Dashboard que se muestra al usuario, contiene la siguiente información:
 - a) Filtro por planta.
 - b) Filtro por fecha.
 - c) Volumen de captación total de acuerdo a los filtros definidos.
 - d) Gráfico representando el volumen de captación de acuerdo a los filtros definidos.

- **Reporte “Comparativa Captación EPMAPS”**: segundo Dashboard que se muestra al usuario, contiene la siguiente información:
 - a) Filtro por planta.
 - b) Filtro por rango de fechas.
 - c) Gráfico representando el porcentaje de volumen de captación de acuerdo con los filtros definidos.
 - d) Gráfico representando el volumen de captación de acuerdo con los filtros definidos.

- **Reporte “Producción EPMAPS”**: tercer Dashboard que se muestra al usuario, contiene la siguiente información:
 - a) Filtro por planta.
 - b) Filtro por fecha.
 - c) Volumen de producción total de acuerdo con los filtros definidos.
 - d) Gráfico representando el volumen de producción de acuerdo con los filtros definidos.

- **Reporte “Comparativa Producción EPMAPS”**: cuarto Dashboard que se muestra al usuario, contiene la siguiente información:
 - a) Filtro por planta.
 - b) Filtro por rango de fechas.
 - c) Gráfico representando el porcentaje de volumen de producción de acuerdo con los filtros definidos.
 - d) Gráfico representando el volumen de producción de acuerdo con los filtros definidos.

- **Reporte “Comparativa Captación vs Producción”**: quinto Dashboard que se muestra al usuario, contiene la siguiente información:
 - a) Filtro por planta.
 - b) Filtro por rango de fechas.
 - c) Tabla representando el volumen de producción vs captación de acuerdo con los filtros definidos.
 - d) Gráfico representando el volumen de producción vs captación de acuerdo con los filtros definidos.

2.7.4.2 Creación de los reportes y dashboards

Los gráficos y reportes mencionados anteriormente se generaron con la ayuda de la herramienta de Microsoft – Power BI. Ya que la empresa se encuentra actualmente trabajando con esta herramienta. A continuación, se procede a explicar la creación de los dashboards que responderán a las necesidades de la empresa.

1. **Verificación y acceso a la herramienta Power BI.** Como este es un servicio en la nube no se requiere infraestructura especial. En la Figura 2.28 se muestra la creación de un proyecto nuevo en la herramienta Microsoft Power BI.

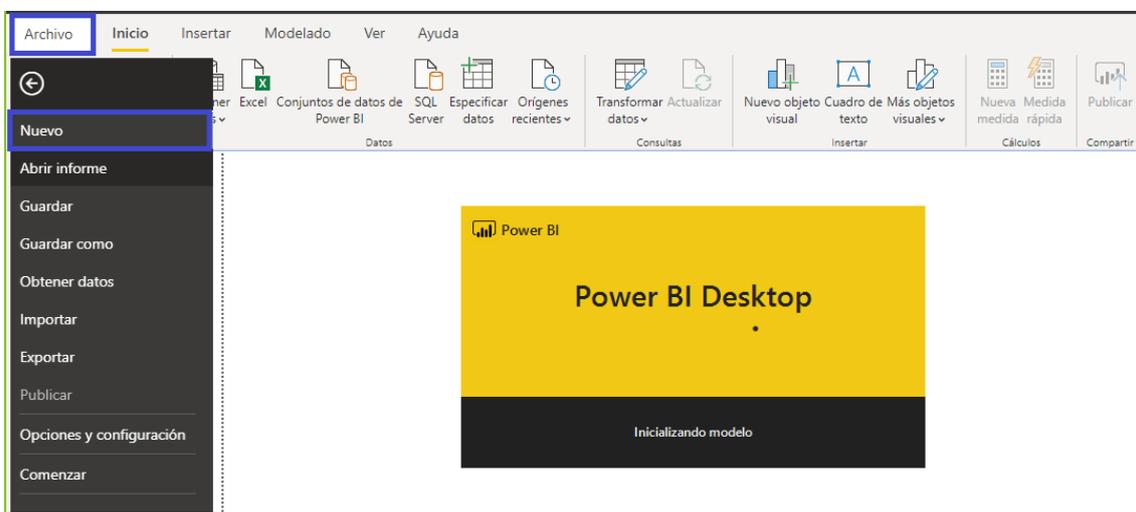


Figura 2.28 Creación de nuevo proyecto en Power BI

2. Obtención de los datos desde SLQ.

En la Figura 2.29 se muestra como obtener los datos desde la base de datos SQL Server.

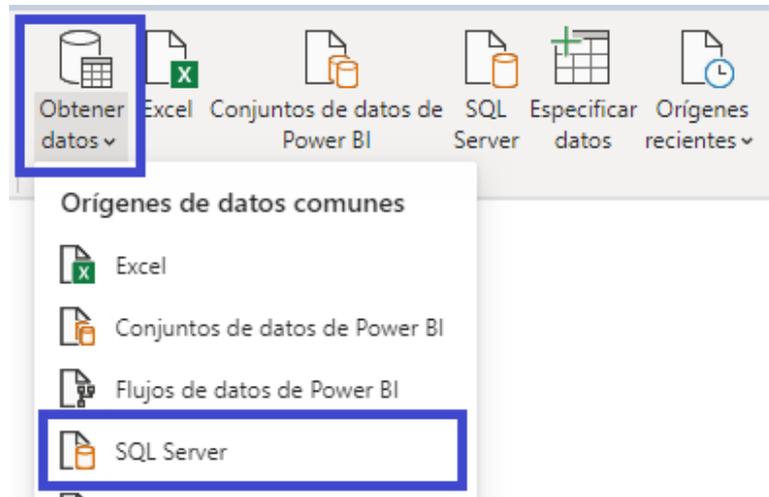


Figura 2.29 Obtención de datos en Power BI

3. **Generación de la estructura del reporte para la visualización de datos.** Dentro de Power BI se presentan tres interfaces las cuales se visualizan en la Figura 2.30, estas permiten generar los componentes necesarios para el diseño de los dashboards.

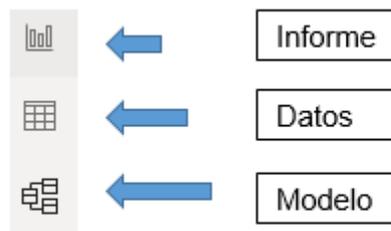


Figura 2.30 Interfaces de Power BI

4. **Generación de filtros.** Dentro de la interfaz Informe se encuentra el componente Filtros como se visualiza en la figura 2.31, éste muestra los datos que serán filtrados al momento de generar el informe.

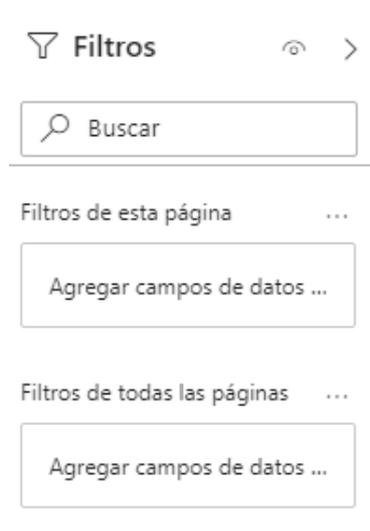


Figura 2.31 Componente Filtros

- 5. Generación del informe.** Dentro de la interfaz Informe se encuentra también el componente Visualizaciones como se visualiza en la figura 2.32, éste permitirá generar gráficos, tablas o selectores que contienen los resultados para el análisis de datos.

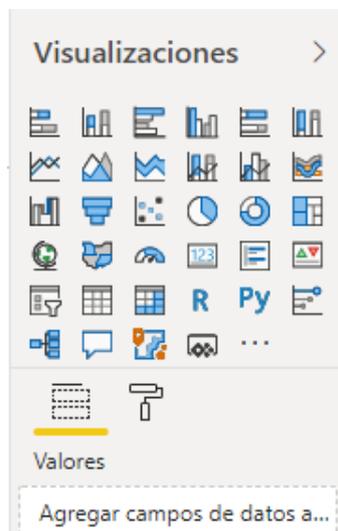


Figura 2.32 Componente Visualizaciones

- 6. Visualización de los atributos de la base de datos.** Dentro de la interfaz Informe se encuentra el componente Campos como se visualiza en la figura 2.33, éste presenta los atributos de la base de datos obtenida del SQL Server, los cuales nos servirán para visualizar de mejor manera los datos en el componente Visualizaciones.

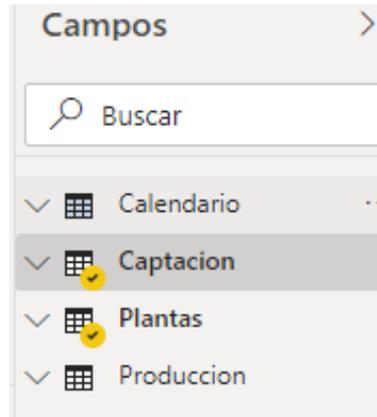


Figura 2.33 Componente Campos

7. Consulta de información en la base de datos. Dentro de la interfaz Datos, se encuentran todos los atributos disponibles para la consulta de datos en la base de datos obtenida del SQL Server para la presentación de datos concretos. En la Figura 3.34 se muestra los atributos de las diferentes tablas utilizadas en el proyecto.

Cod_Produccion	Cod_Planta	Volumen	Unidad	Fecha	Fecha (ubicaciones)
PC001PRO001	PC001		m3	<i>martes, 31 de enero de 2017</i>	31/01/2017
PC001PRO002	PC001		m3	<i>martes, 28 de febrero de 2017</i>	28/02/2017
PC001PRO003	PC001		m3	<i>viernes, 31 de marzo de 2017</i>	31/03/2017
PC001PRO004	PC001		m3	<i>domingo, 30 de abril de 2017</i>	30/04/2017
PC001PRO005	PC001		m3	<i>miércoles, 31 de mayo de 2017</i>	31/05/2017
PC001PRO006	PC001		m3	<i>viernes, 30 de junio de 2017</i>	30/06/2017
PC001PRO007	PC001		m3	<i>lunes, 31 de julio de 2017</i>	31/07/2017
PC001PRO008	PC001		m3	<i>jueves, 31 de agosto de 2017</i>	31/08/2017
PC001PRO009	PC001		m3	<i>sábado, 30 de septiembre de 2017</i>	30/09/2017
PC001PRO010	PC001		m3	<i>martes, 31 de octubre de 2017</i>	31/10/2017
PC001PRO011	PC001		m3	<i>jueves, 30 de noviembre de 2017</i>	30/11/2017
PC001PRO012	PC001		m3	<i>domingo, 31 de diciembre de 2017</i>	31/12/2017
PC001PRO013	PC001		m3	<i>miércoles, 31 de enero de 2018</i>	31/01/2018
PC001PRO014	PC001		m3	<i>miércoles, 28 de febrero de 2018</i>	28/02/2018
PC001PRO015	PC001		m3	<i>sábado, 31 de marzo de 2018</i>	31/03/2018
PC001PRO016	PC001		m3	<i>lunes, 30 de abril de 2018</i>	30/04/2018
PC001PRO017	PC001		m3	<i>jueves, 31 de mayo de 2018</i>	31/05/2018
PC001PRO018	PC001		m3	<i>sábado, 30 de junio de 2018</i>	30/06/2018

Figura 2.34 Atributos presentados en la interfaz Datos

8. Visualización del modelo la base de datos. En la interfaz Modelo, se presenta el modelo de la base de datos extraída del SQL Server para la presentación de datos concretos. Además, se muestra la tabla Calendario generado por la herramienta Microsoft Power BI. En la Figura 2.35 se muestra el modelo de la base de datos.

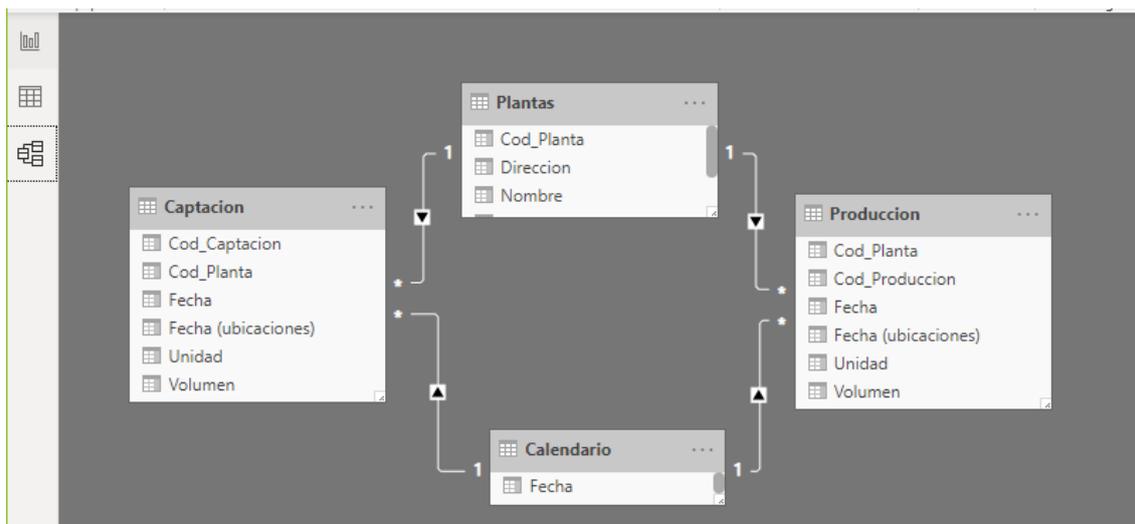


Figura 2.35 Visualización del Modelo de la base de datos

2.7.4.3 Reporte de usuario final

El martes 19 de mayo del 2020 se envió por correo electrónico el reporte final al Departamento de Investigación Desarrollo e Innovación de la EPMAPS quien socializó con el Departamento de Sistemas de Información Geográfica y SCADA para su respectiva aprobación. A continuación, se presenta las características específicas de cada uno de los dashboards:

En la parte inferior de la aplicación final, se presenta un menú Hover Carousel donde se despliega toda la información a través de los reportes y estadísticas relacionadas con los procesos de captación y producción de agua. En el dashboard Captación EPMAPS (Figura 2.36) se muestra una tarjeta indicativa donde se presenta el volumen (m³) total de captación de agua. Además, se agrega los diferentes filtros requeridos por la empresa para visualizar el volumen (m³) de captación de agua por planta y fecha. De acuerdo a estos filtros también se genera un gráfico estadístico que representa el volumen (m³) de agua captada por cada planta.

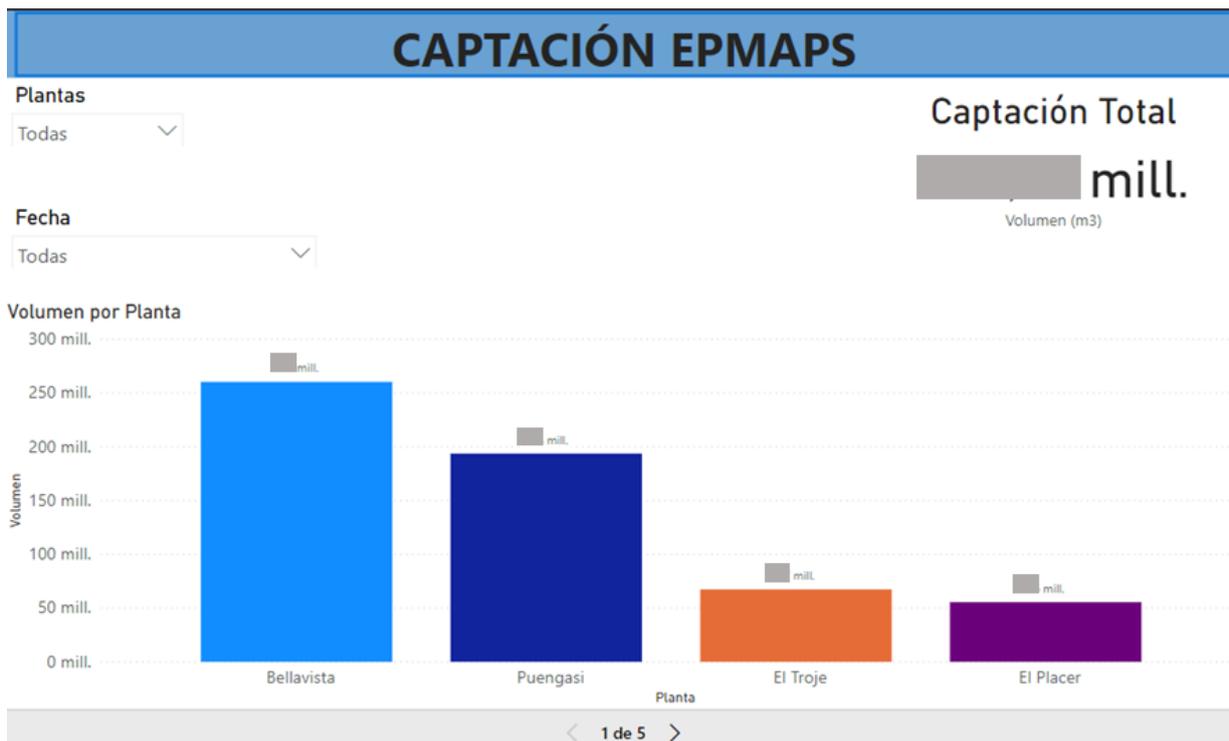


Figura 2.36 Dashboard "Captación EPMAPS"

En el dashboard Comparativa Producción EPMAPS (Figura 2.37) se muestra los diferentes filtros requeridos por la empresa para comparar el volumen (m³) de captación de agua por planta y un rango de fechas. De acuerdo a estos filtros también se genera dos gráficos estadísticos que representan el volumen (m³) y el porcentaje de agua captada, y la comparación entre el volumen (m³) de agua captada por cada planta.

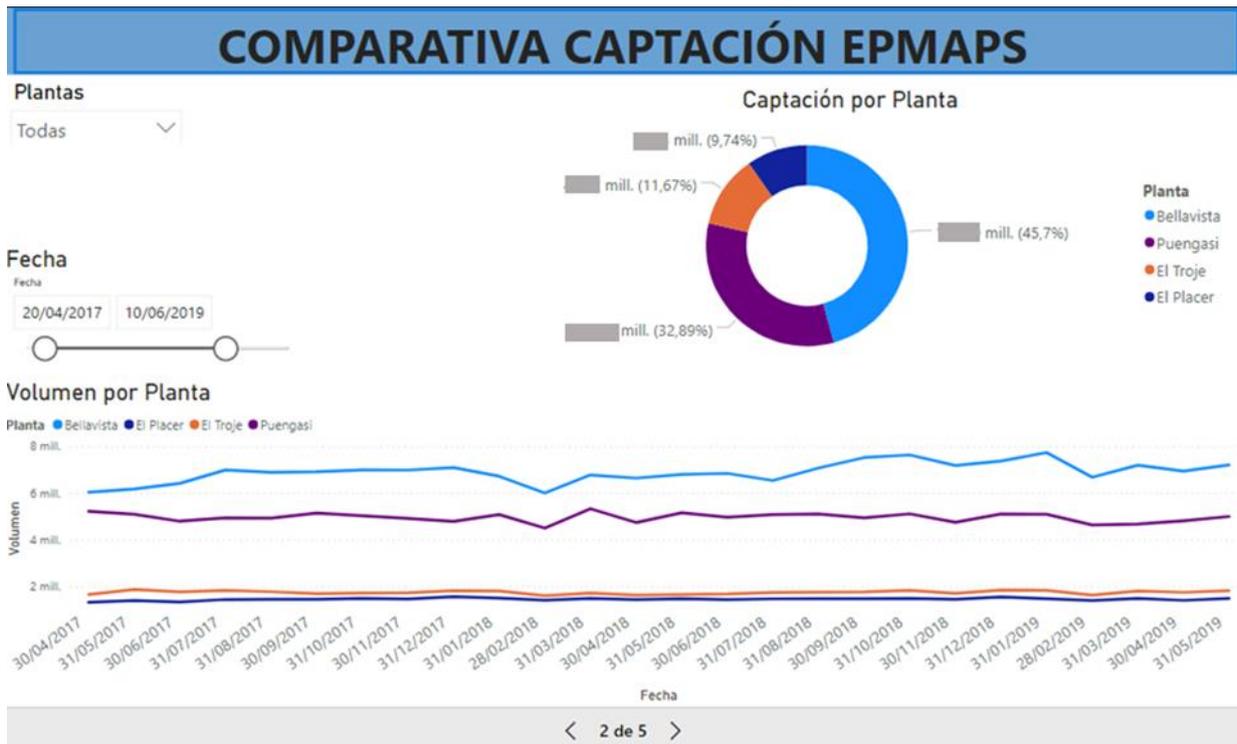


Figura 2.37 Dashboard "Comparativa Captación EPMAPS"

En el dashboard Producción EPMAPS (Figura 2.38) se muestra una tarjeta indicativa donde se presenta el volumen (m³) total de producción de agua. Además, se agrega los diferentes filtros requeridos por la empresa para visualizar el volumen (m³) de producción de agua por planta y fecha. De acuerdo a estos filtros también se genera un gráfico estadístico que representa el volumen (m³) de agua producida por cada planta.

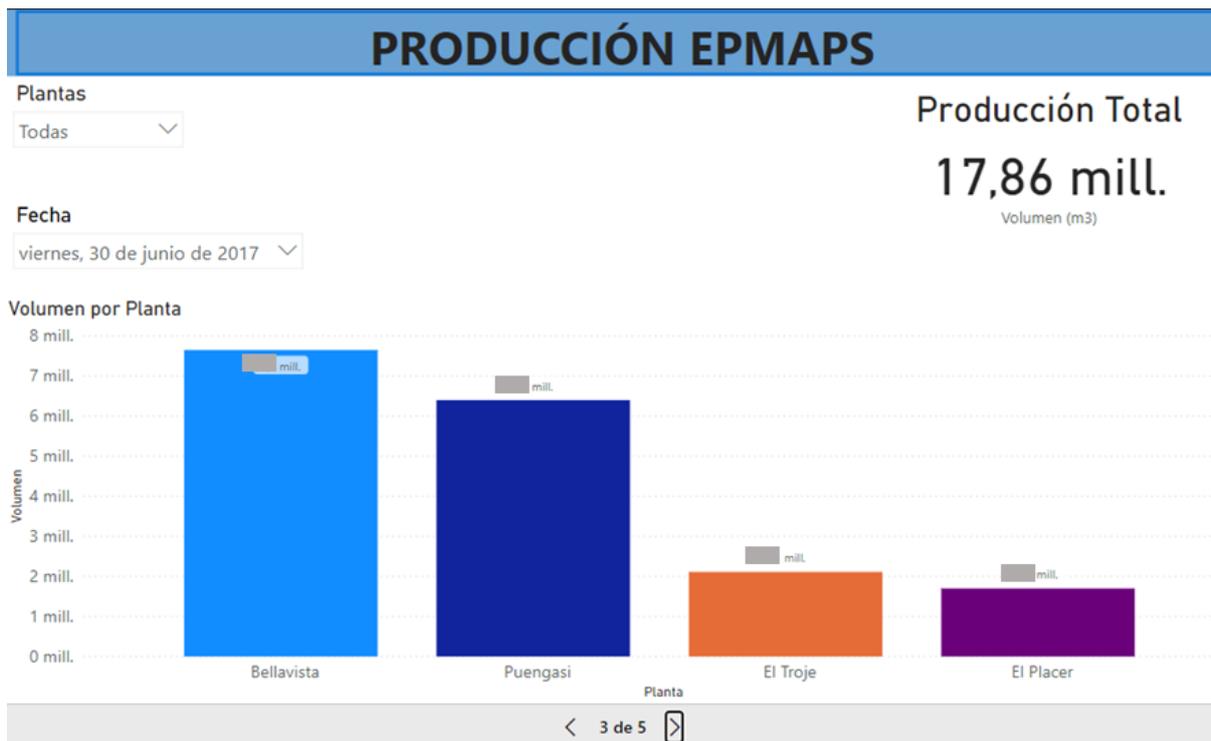


Figura 2.38 Dashboard "Producción EPMAPS"

En el dashboard Comparativa Producción EPMAPS (Figura 2.39) se muestra los diferentes filtros requeridos por la empresa para comparar el volumen (m³) de producción de agua por planta y un rango de fechas. De acuerdo a estos filtros también se genera dos gráficos estadísticos que representan el volumen (m³) y el porcentaje de agua producida, y la comparación entre el volumen (m³) de agua producida por cada planta.

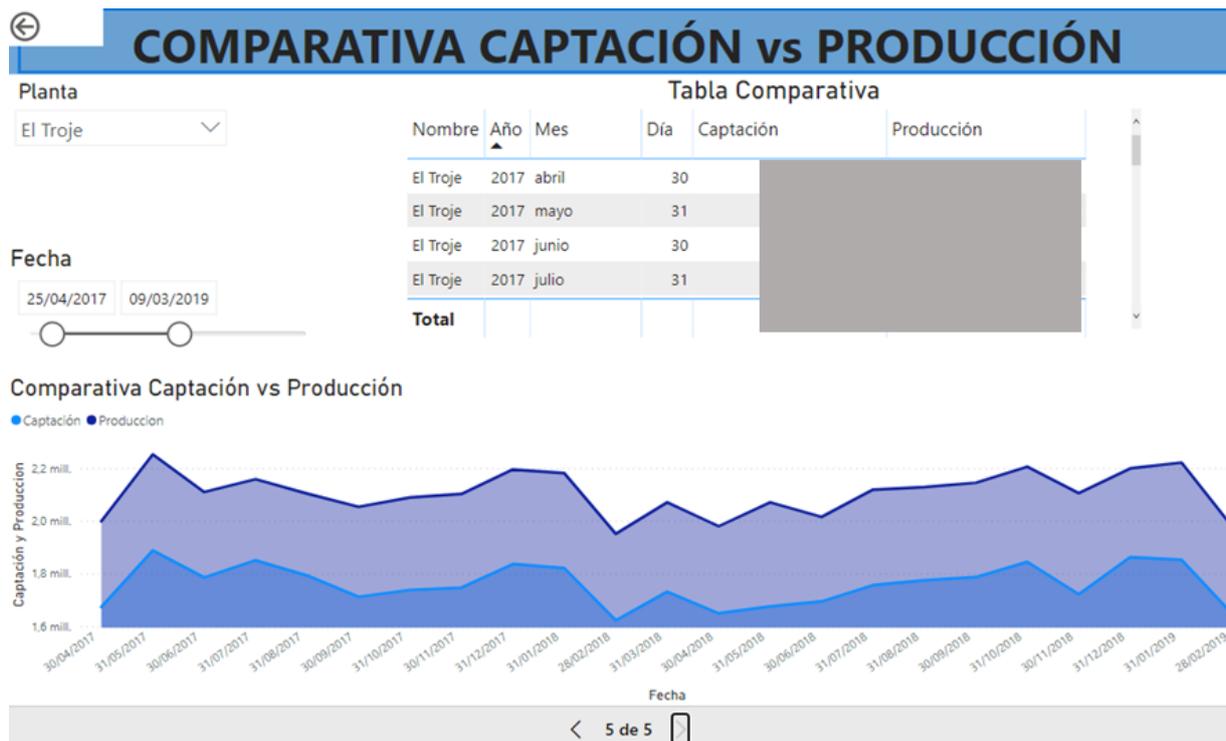


Figura 2.40 Dashboard "Comparativa Captación vs Producción"

2.7.5 Evaluación

Para llegar al reporte de usuario final se realizaron las versiones descritas a continuación.

2.7.5.1 Versionamiento de la herramienta de BI

Para solventar las necesidades a nivel estratégico (Requerimiento: tener un reporte del proceso de producción con periodicidad mensual con el fin de asegurar la cobertura del servicio de agua potable), y a nivel táctico (Requerimiento: tener un reporte del proceso de captación llevando un monitoreo y control mensual con el fin que el departamento correspondiente tome las decisiones necesarias (repotenciación o cierre de la planta)), se presenta las siguientes versiones:

- **Mockups**
 - Un reporte con dos dashboards para el proceso de captación:
 - ✓ Los mockups descritos en las figuras 2.41 y 2.42 fueron realizados con datos ficticios, los mismos cumplen con la necesidad requerida a nivel táctico; ya que cuentan con las gráficas estadísticas y los filtros respectivos para el análisis del proceso de captación.



Figura 2.41 Mockup Captación 1

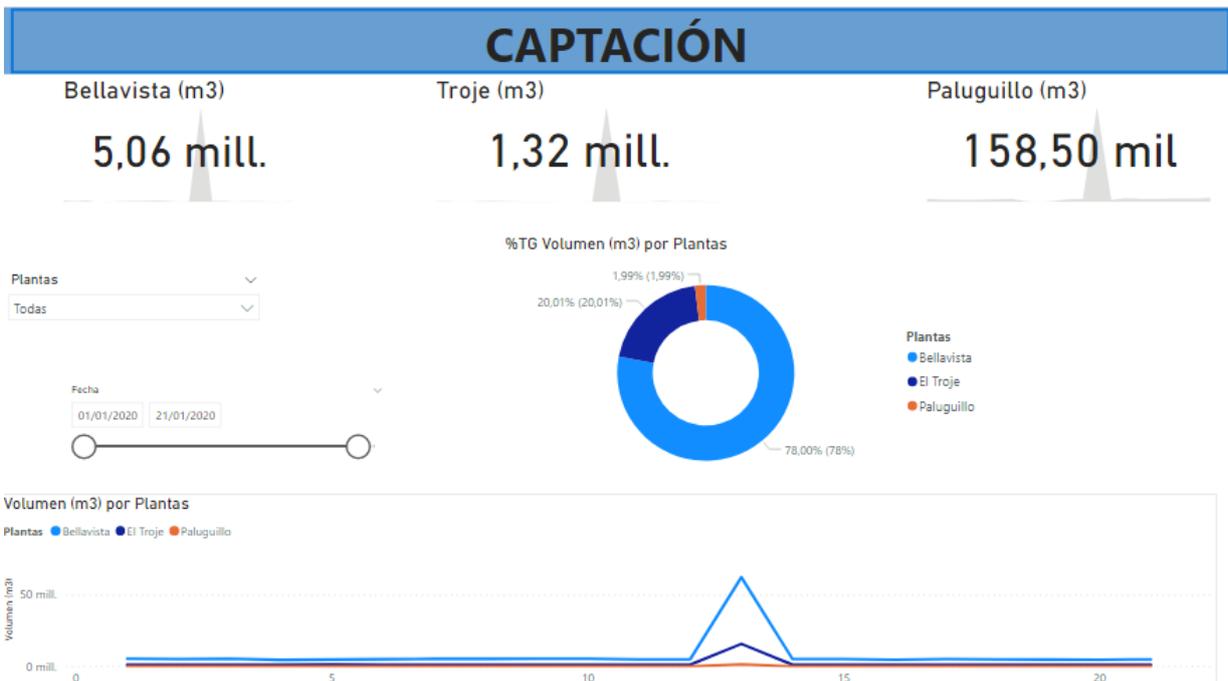


Figura 2.42 Mockup Captación 2

- Un reporte con dos dashboards para el proceso de producción:
 - ✓ Los mockups descritos en las figuras 2.43 y 2.44 fueron realizados con datos ficticios, los mismos cumplen con la necesidad requerida a nivel estratégico; ya que cuentan con las gráficas estadísticas y los filtros respectivos para el aseguramiento y cobertura del agua potable.



Figura 2.43 Mockup Producción 1

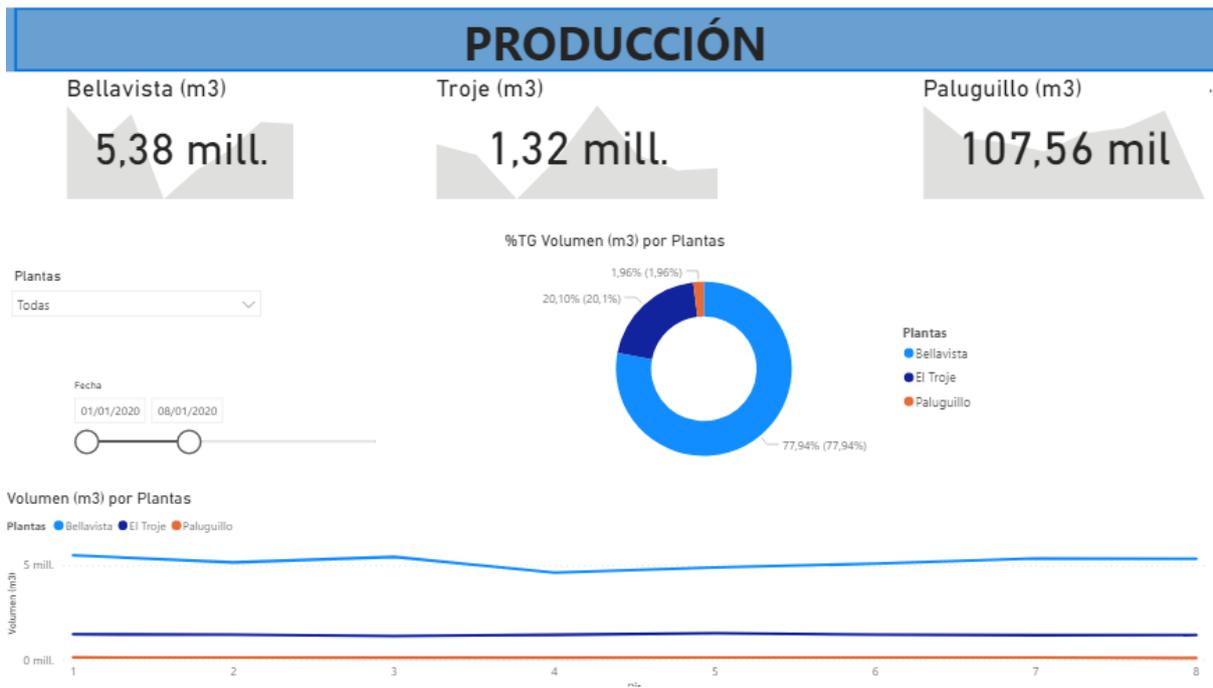


Figura 2.44 Mockup Producción 2

- **Versión 1**

El jueves 07 de mayo del 2020 se envió por correo electrónico la primera versión del reporte final al Departamento de Investigación Desarrollo e Innovación de la EPMAPS quien socializó con el Departamento de Sistemas de Información Geográfica y SCADA para su respectiva revisión.

- Reporte Captación descrito en las figuras 2.45 y 2.46.

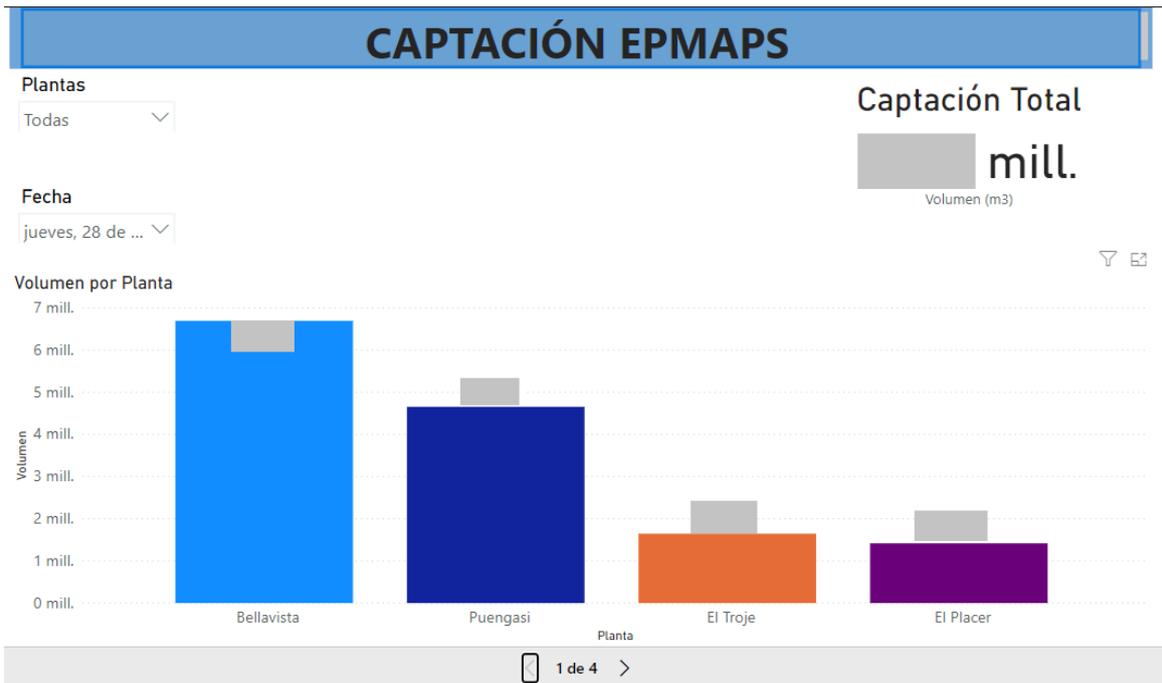


Figura 2.45 Primera versión Captación 1

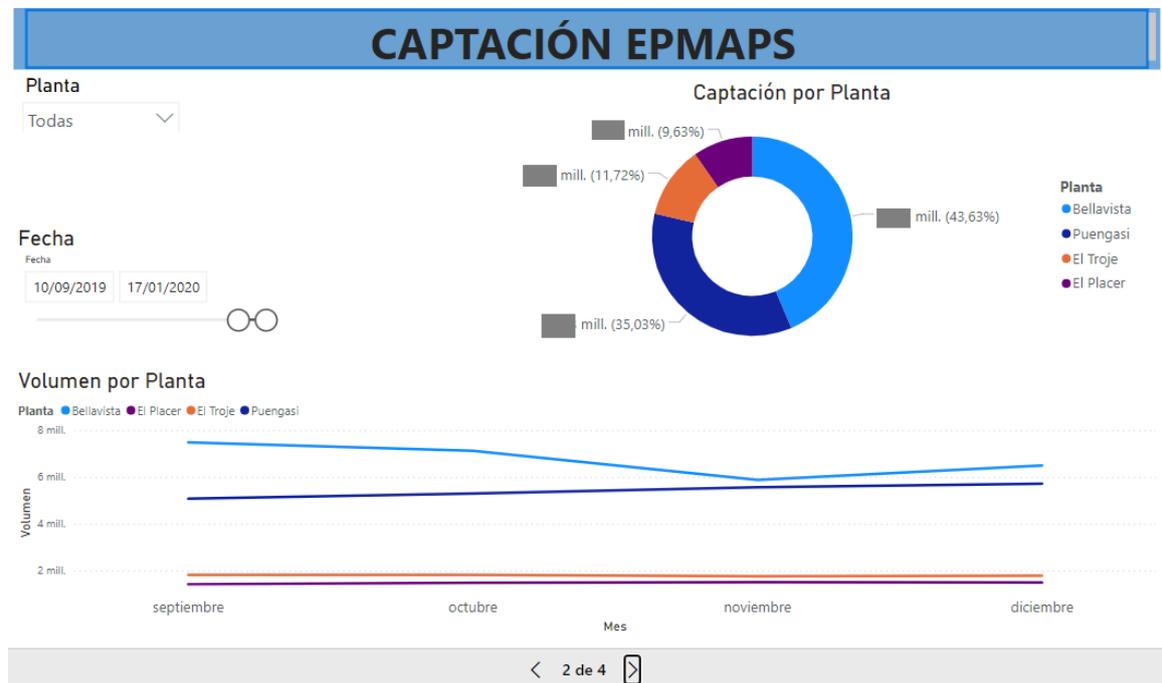


Figura 2.46 Primera versión Captación 2

La tabla 2.3 presenta los cambios que se realizaron del mockup a la versión 1 del proceso de captación.

	Mockup	Versión 1	Observación
Filtro por planta	X	X	N/A
Filtro por fecha	X	X	N/A
Tarjeta indicativa (Total Captación)	X	X	N/A
Tarjeta indicativa de varias filas (Por cada planta)	X		La gráfica estadística de barras presenta la misma información que las tarjetas indicativas.
Graficas estadísticas	X	X	N/A
KPI (Por cada planta)	X		No se requiere la clave de desempeño de la empresa.

Tabla 2.3 Mackup vs Versión 1 (Captación)

- Reporte Producción descrito en las figuras 2.47 y 2.48.

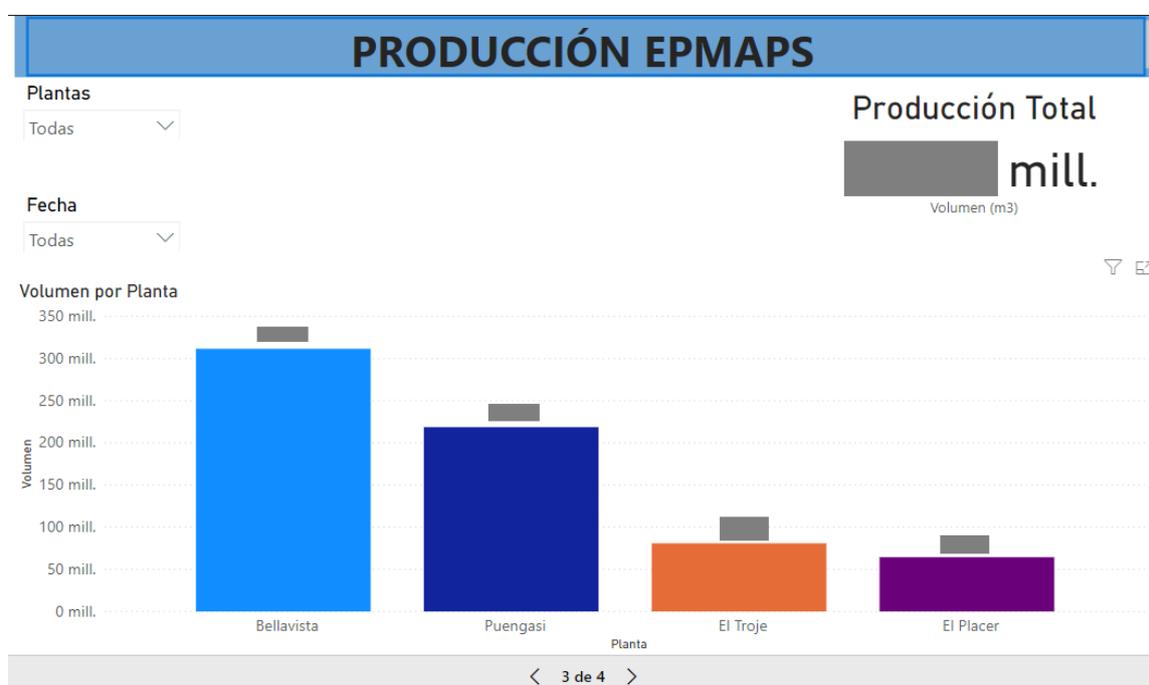


Figura 2.47 Primera versión Producción 1

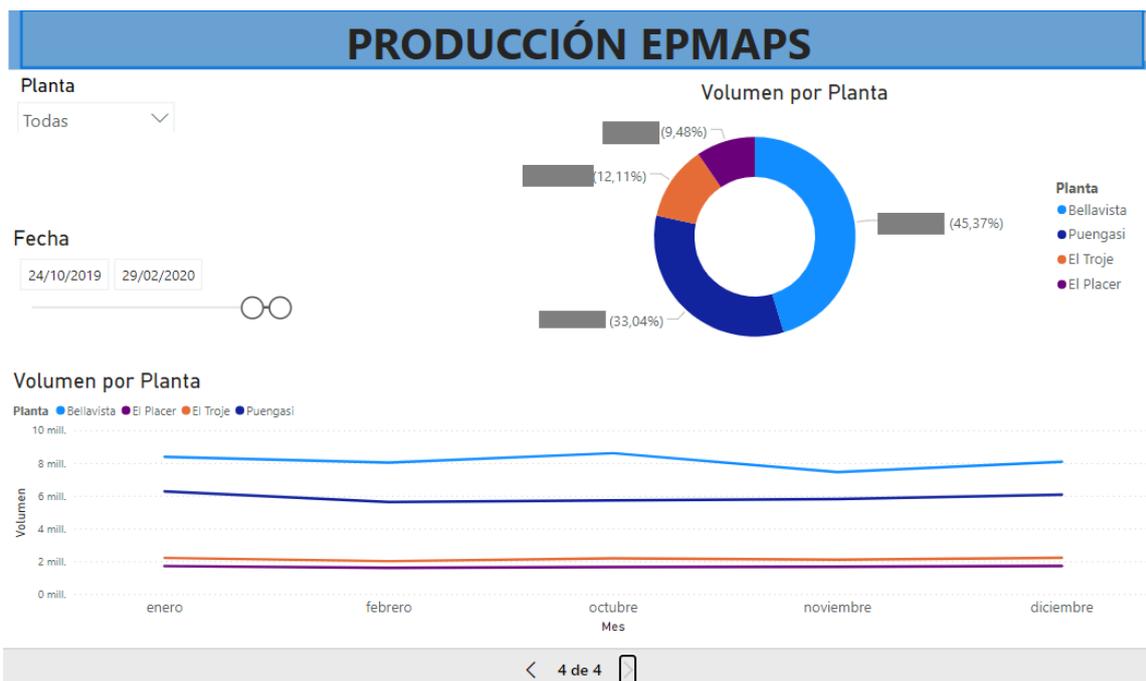


Figura 2.48 Primera versión Producción 2

La tabla 2.4 presenta los cambios que se realizaron del mockup a la versión 1 del proceso de producción.

	Mockup	Versión 1	Observación
Filtro por planta	X	X	N/A
Filtro por fecha	X	X	N/A
Tarjeta indicativa (Total Producción)	X	X	N/A
Tarjeta indicativa de varias filas (Por cada planta)	X		La gráfica estadística de barras presenta la misma información que las tarjetas indicativas.
Graficas estadísticas	X	X	N/A
KPI (Por cada planta)	X		No se requiere la clave de desempeño de la empresa.

Tabla 2.4 Mackup vs Versión 1 (Producción)

El viernes 08 de mayo del 2020 se recibe por correo electrónico las observaciones de la primera versión del reporte final de usuario por parte del Departamento de

Investigación y Desarrollo e Innovación de la EPMAPS a continuación se detalla las mismas:

- Implementar un reporte donde se compare el volumen de captación y producción por planta.
- Mejorar la visualización de las fechas en los gráficos estadísticos.
- Modificación de títulos de los dashboard.

- **Versión 2**

El lunes 18 de mayo del 2020 se envió por correo electrónico la segunda versión del reporte final de usuario al Departamento de Investigación Desarrollo e Innovación de la EPMAPS quien socializó con el Departamento de Sistemas de Información Geográfica y SCADA para su respectiva revisión y aprobación.

- Reporte Captación descrito en las figuras 2.49 y 2.50.

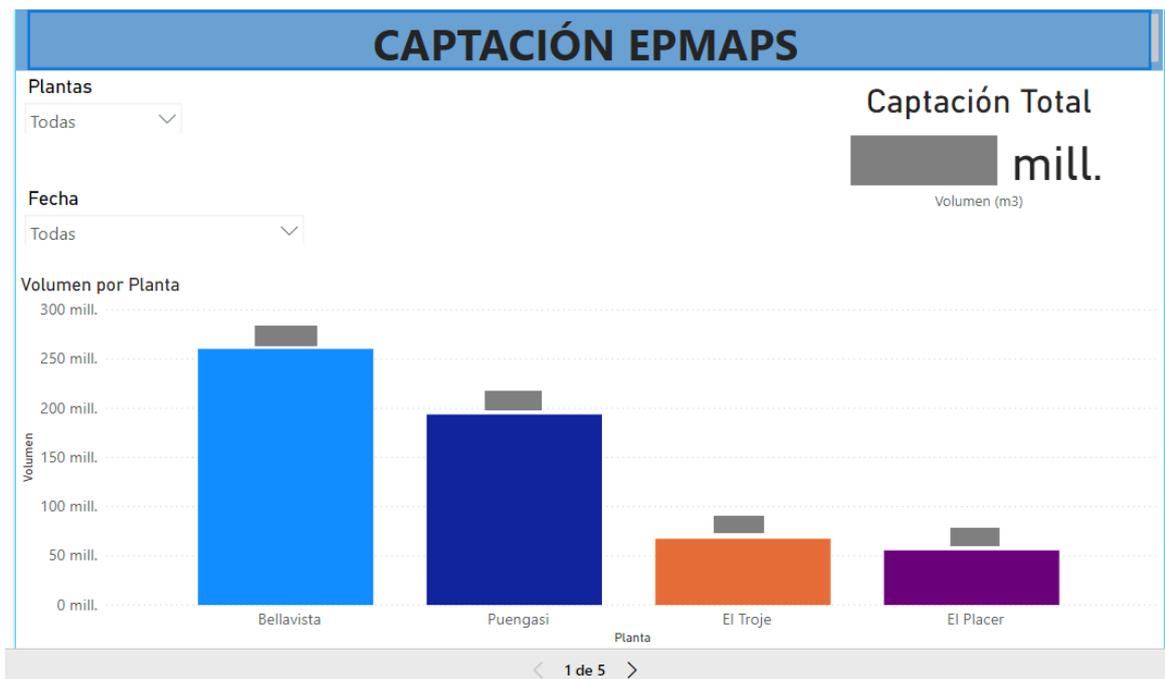


Figura 2.49 Segunda versión Captación 1

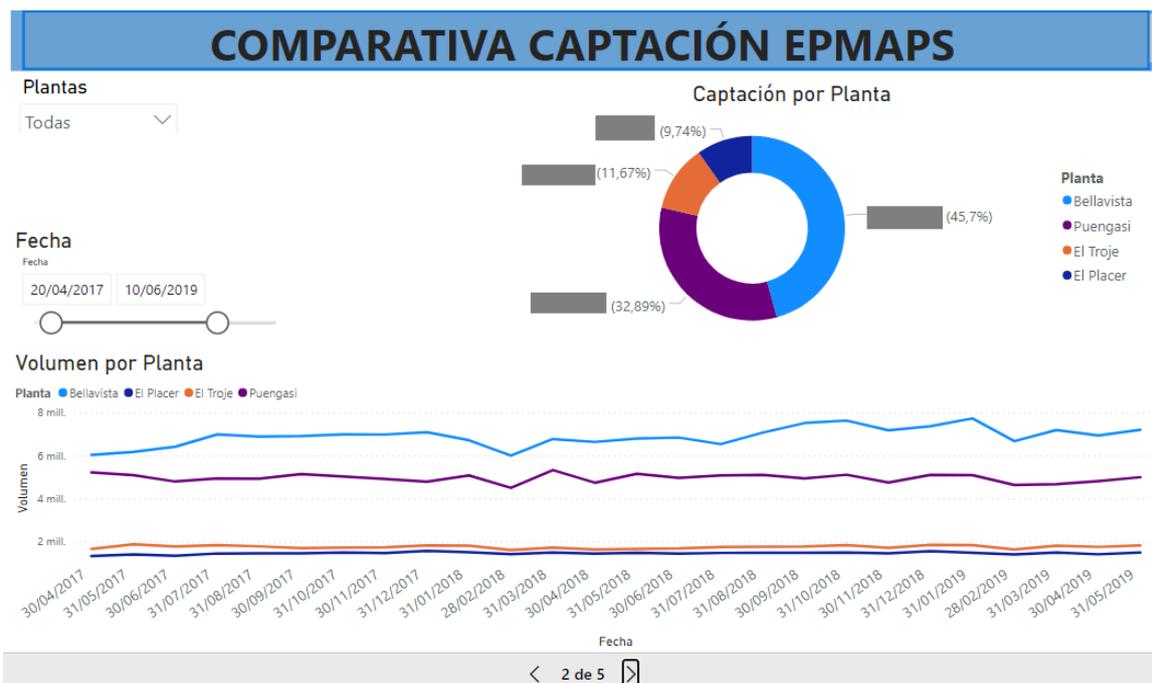


Figura 2.50 Segunda versión Captación 2

La tabla 2.5 presenta los cambios que se realizaron de la versión 1 a la versión 2 del proceso de captación.

	Versión 1	Versión 2	Observación
Filtro por planta	X	X	N/A
Filtro por fecha	X	X	N/A
Tarjeta indicativa (Total Captación)	X	X	N/A
Gráficas estadísticas	X	X	Se mejoró la visualización de las fechas en las gráficas estadísticas.
Títulos	X	X	Se realizó la modificación en los títulos de los dashboards de captación.

Tabla 2.5 Versión 1 vs Versión 2 (Captación)

- Reporte Producción descrito en las figuras 2.51 y 2.52.

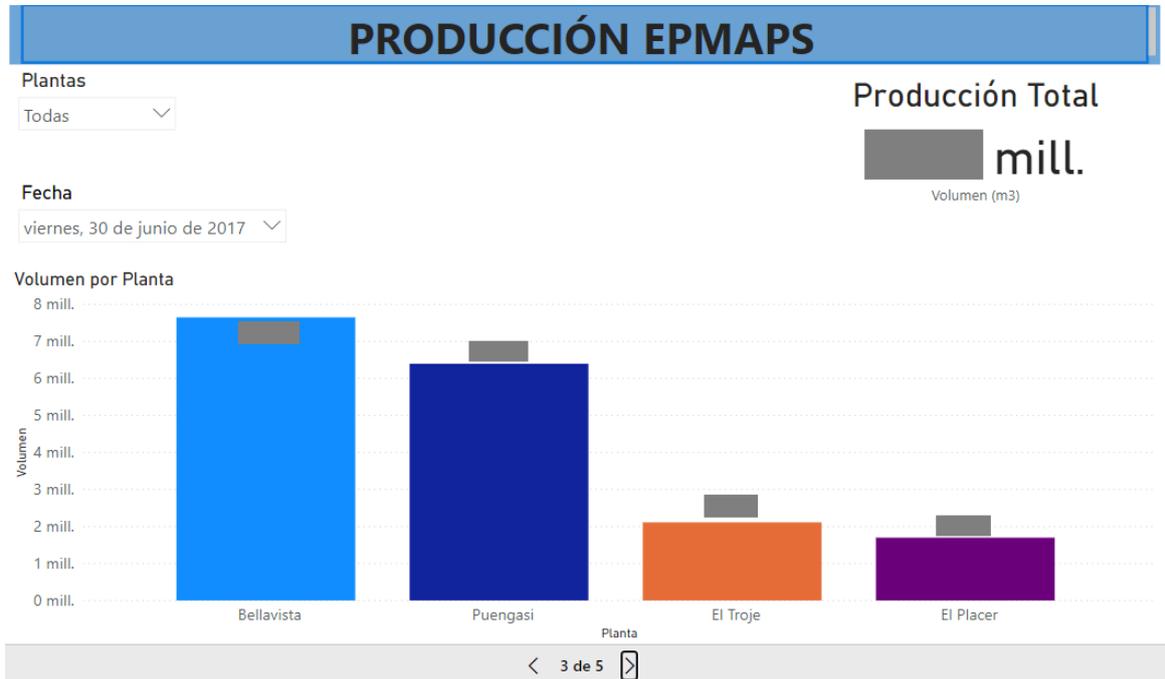


Figura 2.51 Segunda versión Producción 1

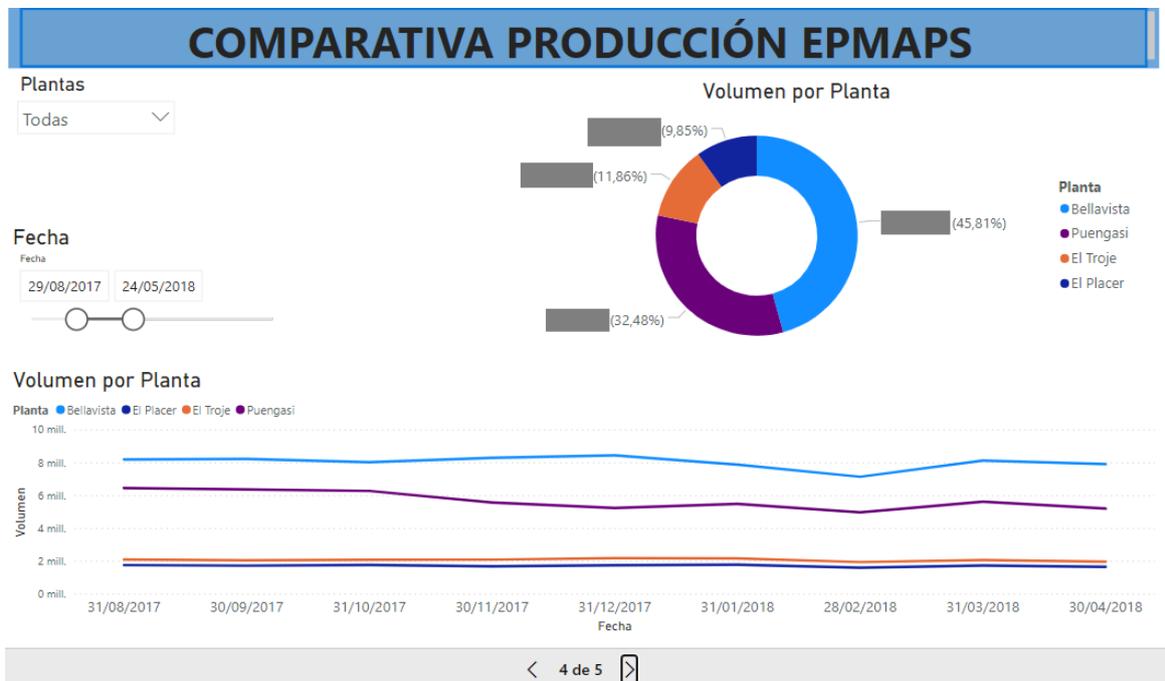


Figura 2.52 Segunda versión Producción 2

La tabla 2.6 presenta los cambios que se realizaron de la versión 1 a la versión 2 del proceso de producción.

	Versión 1	Versión 2	Observación
Filtro por planta	X	X	N/A
Filtro por fecha	X	X	N/A
Tarjeta indicativa (Total Producción)	X	X	N/A
Graficas estadísticas	X	X	Se mejoró la visualización de las fechas en las gráficas estadísticas.
Títulos	X	X	Se realizó la modificación en los títulos de los dashboards de producción.

Tabla 2.6 Versión 1 vs Versión 2 (Producción)

- Reporte Comparativo descrito en la figura 2.53.

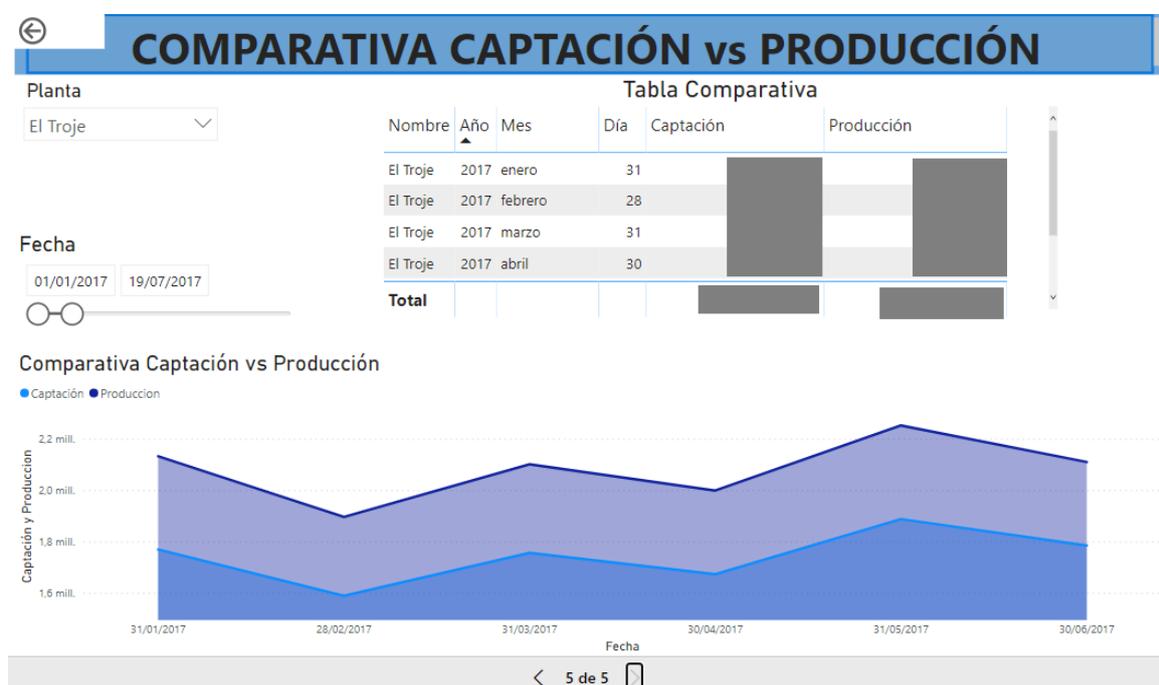


Figura 2.53 Segunda versión Comparativa Captación vs Producción

De acuerdo con el requerimiento establecido en la revisión de la versión 1 se lleva a cabo la implementación del dashboard Comparativa Captación vs Producción, el cual presenta una

comparación entre los volúmenes de captación y producción. Y esta comparación se muestra en una tabla y en una gráfica estadística.

2.7.6 Implementación

Para esta última fase se elabora un manual de usuario donde se explica el uso de la herramienta de BI enfocándose en 4 puntos importantes:

- Descripción del Sistema
- Mapa del Sistema
- Diseño del Sistema
- FAQ

El manual de usuario detallado se lo presenta en el Anexo 4.

La implantación de la herramienta de BI se realizó directamente por el personal de la EPMAPS por el tema de la pandemia, luego de haber realizado las respectivas pruebas nos notifican por medio de correo electrónico que la implementación fue satisfactoria. Concluyendo así la fase de implementación de este proyecto.

El correo de la notificación de satisfacción se presenta en el Anexo 5.

3 DISCUSIÓN DE RESULTADOS

- Generar los datos para el desarrollo de este proyecto no fue una tarea sencilla, debido que las plantas no tenían conectividad con la matriz por lo que la base de datos no tenía la información en tiempo real, además la información histórica presentaba datos repetitivos y nulos. Por esta razón se optó por elaborar una base de datos local con la información específica para solventar las necesidades del negocio brindando un apoyo para que la toma de decisiones sea más segura y confiable.
- Fue necesaria la implementación de un esquema dimensional en estrella para la creación de la base de datos debido a que los requerimientos y las necesidades de los diferentes niveles de la empresa no eran los mismos. Este esquema permitió el manejo de múltiples tablas en la base de datos y la reutilización de las mismas en los diferentes dashboards. De esta forma se logró obtener una mejor abstracción de las necesidades de la empresa.
- La metodología facilitó el desarrollo de la herramienta de BI, debido que la misma es bidireccional, lo cual permitió regresar a fases anteriores para realizar las respectivas correcciones en dichas fases.
- Los dashboards generados para el usuario final fueron enfocados en tres aspectos que están relacionados con los procesos de captación y producción de volumen de agua en la EPMAPS: Captación EPMAPS y su comparativa, Producción EPMAPS y su comparativa, Comparativa Captación vs Producción.
- Los dashboards generados para el usuario final presenta en sus graficas las respuestas a las siguientes preguntas:

A nivel táctico

¿Cuál es el volumen de agua captado por cada planta al mes, durante los tres últimos años?

¿Cuál es la planta que capta mayor volumen de agua al mes, durante los tres últimos años?

¿Cuál es el volumen de captación total de todas las plantas?

A nivel estratégico

¿Cuál es el volumen de agua producido por cada planta al mes, durante los tres últimos años?

¿Cuál es la planta que produce mayor volumen de agua al mes, durante los tres últimos años?

¿Cuál es el volumen de producción total de todas las plantas?

- En el dashboard Comparativa Captación vs Producción se puede visualizar que el volumen de agua producido por las plantas es mayor que el volumen de agua captado por las mismas, esto se debe a que las plantas solo cuentan con un sensor de captación en la entrada principal, y en la mayoría de las plantas también existen entradas secundarias las cuales no cuentan con este sensor, lo que produce el desfase mencionado anteriormente.

4 CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

- Este proyecto cumplió con el diseño y construcción de un módulo de Inteligencia de Negocios para la Empresa Publica Metropolitana de Agua Potable y Saneamiento Quito el cual fue enfocado en la data de caudales de los procesos de captación y producción de la misma. (Objetivo General)
- Previo a la preparación de datos, se procedió a la exploración del sistema SCADA, en el cual se realizaron consultas Sql generando archivos xlsx, los cuales fueron de gran ayuda para identificar y documentar los datos necesarios para generar la base de datos tanto del proceso de captación como del proceso de producción de cada planta. (Objetivo Especifico 1)
- Para asegurar la disponibilidad y calidad de los datos se procedió a crear una base de datos local con la información necesaria para solventar las necesidades del negocio, siguiendo el formato establecido por la codificación UT de la empresa para crear los respectivos IDs, con esto se logró identificar de una forma más sencilla los atributos utilizados en el desarrollo de este proyecto. (Objetivo Especifico 2)
- Debido que la metodología se adaptó a las necesidades del proyecto no fue necesario realizar un análisis de modelos ya que los indicadores asociados a la data de caudales fueron descritos en el apartado 3.1.2 (Evaluación de la situación), esta descripción permitió que los indicadores y objetivos del negocio se relacionen cumpliendo así con el propósito de la herramienta de BI desarrollada en este proyecto. (Objetivo Especifico 3)
- Debido que la carga de los datos se realizó de forma manual en la base de datos y aplicando la codificación de la empresa antes mencionada no hubo la necesidad de realizar sobre estos una limpieza exhaustiva, ni de transfórmalos, lo que permitió obtener un ahorro de tiempo el cual fue aprovechado de mejor manera en las siguientes fases. (Objetivo Especifico 4)
- La información que se genera en la base datos se presenta a los usuarios finales mediante dashboards los cuales contienen gráficos estadísticos, tarjetas informativas y tablas permitiendo a estos usuarios visualizar de mejor manera la información antes mencionada. Para el desarrollo de los dashboards se trabajó en conjunto con los departamentos correspondientes, con el fin que los involucrados realicen las respectivas revisiones y aprobaciones de estos. (Objetivo Especifico 5)

- La última fase de este proyecto (Implementación) se realizará directamente por los usuarios finales de la empresa, ya que los mismos deben integrar su base de datos interna con la herramienta de BI realizada en este proyecto, lo que permitirá a la empresa en un determinado tiempo la toma de decisiones no solo de los sistemas de agua más grandes sino de todos los existentes en la ciudad de Quito.

4.2 Recomendaciones

- Se recomienda analizar los requerimientos e identificar las diferentes actividades dentro del negocio, con el fin que se pueda obtener ventajas que ayuden a solventar las necesidades del negocio, sirviendo como apoyo en el proceso de toma de decisiones.
- Para futuros proyectos se recomienda trabajar con una base de datos local con los datos necesarios para solventar el problema, para después aplicarlo a la base de datos central de la empresa.
- Para evitar hacer limpiezas exhaustivas de datos en las cuales se invierte demasiado tiempo del trabajo del proyecto se recomienda a la empresa contar con una base de datos integra y de calidad.
- Se recomienda a la empresa contar con una base de datos integra y de calidad o poner en marcha un proceso de calidad de datos de forma que se pueda evitar datos duplicados.
- Para un mejor análisis en la comparativa de Captación vs Producción se recomienda a la empresa implementar sensores en las entradas secundarias de cada planta.

5 BIBLIOGRAFIA

- [1] EPMAPS-AGUA DE QUITO, "Sistema de Distribución - Bienvenido a EPMAPS." [Online]. Available: <https://www.aguaquito.gob.ec/sistema-de-distribucion/>. [Accessed: 08-Jun-2020].
- [2] P. Rodríguez *et al.*, "Apoyando la formulación de políticas públicas y toma de decisiones en educación utilizando técnicas de análisis de datos masivos: el caso de Chile," *Minist. Educ.*, 2016.
- [3] EPMAPS-AGUA DE QUITO, "Memoria de sostenibilidad," 2012. [Online]. Available: <https://www.aguaquito.gob.ec/Alojamientos/Memoria de Sostenibilidad 2017.pdf>. [Accessed: 22-Oct-2019].
- [4] EPMAPS-Agua de Quito, "Organigrama - Bienvenido a EPMAPS." [Online]. Available: <https://www.aguaquito.gob.ec/organigrama/>. [Accessed: 07-Nov-2019].
- [5] Wikipedia, "Melecitosa - Wikipedia, la enciclopedia libre." [Online]. Available: [https://es.wikipedia.org/wiki/Caudal_\(hidrografía\)](https://es.wikipedia.org/wiki/Caudal_(hidrografía)). [Accessed: 07-Nov-2019].
- [6] EPMAPS-AGUA DE QUITO, "MEMORIA DE SOSTENIBILIDAD E P M A P S 2 0 1 5 2 3 Memoria de Sostenibilidad, EPMAPS 2015 Empresa Pública Metropolitana de Agua Potable y Saneamiento." [Online]. Available: https://www.aguaquito.gob.ec/downloads/InformesGestion/memoria_de_sostenibilidad_2015_web.pdf. [Accessed: 07-Nov-2019].
- [7] Andrea Encalada, "Funciones ecosistémicas y diversidad de los ríos." [Online]. Available: https://www.usfq.edu.ec/publicaciones/polemika/Documents/polemika005/polemika005_007_articulo004.pdf. [Accessed: 07-Nov-2019].
- [8] EPMAPS-AGUA DE QUITO, "DIRECCIONAMIENTO Y PLANIFICACIÓN ESTRATÉGICA PLAN ESTRATÉGICO (RESUMEN EJECUTIVO)," 2018. [Online]. Available: <https://www.aguaquito.gob.ec/wp-content/uploads/2018/05/REVISION-ANUAL-RESUMEN-EJECUTIVO-PLAN-ESTRATÉGICO-2018-2021-DIRECTORIO-FINAL.pdf>. [Accessed: 07-Nov-2019].
- [9] Oracle, "¿Qué es Inteligencia de Negocios?" [Online]. Available: https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf. [Accessed: 07-Nov-2019].
- [10] Julio Castro, "¿Qué es la inteligencia de negocios y cómo beneficia a tu empresa?," 2015. [Online]. Available: <https://blog.corponet.com.mx/que-es-la-inteligencia-de-negocios>. [Accessed: 11-Nov-2019].
- [11] ISOTools, "Business Intelligence: 8 etapas para el proceso de toma de decisiones,"

2019. [Online]. Available: <https://www.isotools.org/2019/01/04/business-intelligence-8-etapas-proceso-toma-decisiones/>. [Accessed: 11-Nov-2019].
- [12] PowerData, "Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos." [Online]. Available: <https://www.powerdata.es/data-warehouse>. [Accessed: 11-Nov-2019].
- [13] Sinnexus, "Datawarehouse." [Online]. Available: https://www.sinnexus.com/business_intelligence/datawarehouse.aspx. [Accessed: 11-Nov-2019].
- [14] Sinnexus, "Datamining (Minería de datos)." [Online]. Available: https://www.sinnexus.com/business_intelligence/datamining.aspx. [Accessed: 30-Jan-2020].
- [15] MICROTECH, "Business Intelligence: 5 aspectos importantes para la toma de decisiones." [Online]. Available: <https://www.microtech.es/blog/business-intelligence-5-aspectos-importantes-para-la-toma-de-decisiones>. [Accessed: 11-Nov-2019].
- [16] I. Bernabeu and R. Dario, "HEFESTO DATA WAREHOUSING: Investigación y Sistematización de Conceptos HEFESTO: Metodología para la Construcción de un Data Warehouse." [Online]. Available: <https://www.businessintelligence.info/resources/assets/hefesto-v2.pdf>. [Accessed: 11-Nov-2019].
- [17] O. De Postgrado, G. Euclides, and S. Peñafiel, "Ambato-Ecuador," 2018. [Online]. Available: <http://repositorio.pucesa.edu.ec/bitstream/123456789/2367/1/76540.pdf>. [Accessed: 11-Nov-2019].
- [18] R. Espinosa, "15.2.Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. « El Rincon del BI," 2010. [Online]. Available: <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>. [Accessed: 11-Nov-2019].
- [19] Espinoza Carlos, "PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR FACULTAD DE INGENIERÍA ESCUELA DE SISTEMAS DISERTACIÓN DE TESIS PREVIO A LA OBTENCIÓN DEL TÍTULO DE." [Online]. Available: <http://repositorio.puce.edu.ec/bitstream/handle/22000/6216/T-PUCE-6392.pdf?sequence=1&isAllowed=y>. [Accessed: 11-Nov-2019].
- [20] "CRISP-DM, Una metodología para proyectos de Minería de Datos | anibal goicochea," 2019. [Online]. Available: <https://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/>. [Accessed: 11-Nov-2019].
- [21] EPMAPS-AGUA DE QUITO, "Objetivos Estratégicos - Bienvenido a EPMAPS."

[Online]. Available: <https://www.aguaquito.gob.ec/objetivos/>. [Accessed: 05-Jan-2020].

6 ANEXOS

6.1 Anexo 1 Actas de Reunión

Se Anexa el siguiente archivo "Actas de Reunión .pdf"

6.2 Anexo 2 Scripts de Creación de Base de Datos y Consultas SQL

6.2.1 Creación de la Base de Datos

```
/* creación de la base de datos */
CREATE DATABASE CaudalesEPMAPS
go

/* acceso a la base */
use CaudalesEPMAPS
go

/* creación tablas */
CREATE TABLE [dbo].[Plantas](
    [Cod_Planta] [char] (5) NOT NULL,
    [Nombre] [char](30) NOT NULL,
    [Direccion] [char](50) NOT NULL,
    [Sector] [char](30) NOT NULL,
    [Zona] [char](20) NOT NULL,

CONSTRAINT [PK_Plantas] PRIMARY KEY NONCLUSTERED (Cod_Planta))

CREATE TABLE [dbo].[Captacion](
    [Cod_Captacion] [char] (11) NOT NULL,
    [Cod_Planta] [char] (5) NOT NULL,
    [Volumen] [numeric] (10,2)NOT NULL,
    [Unidad] [char] (2) NOT NULL,
    [Fecha] [date] NOT NULL,

CONSTRAINT [PK_Captacion] PRIMARY KEY NONCLUSTERED ([Cod_Captacion]) )

CREATE TABLE [dbo].[Produccion](
    [Cod_Produccion] [char] (11) NOT NULL,
    [Cod_Planta] [char] (5) NOT NULL,
    [Volumen] [numeric] (10,2) NOT NULL,
    [Unidad] [char] (2) NOT NULL,
    [Fecha] [date] NOT NULL,

CONSTRAINT [PK_Produccion] PRIMARY KEY NONCLUSTERED ([Cod_Produccion]) )
```

```
ALTER TABLE [dbo].[Captacion] WITH CHECK ADD CONSTRAINT  
[FK_Planta_Captacion] FOREIGN KEY([Cod_Planta])  
REFERENCES [dbo].[Plantas] ([Cod_Planta])
```

```
ALTER TABLE [dbo].[Produccion] WITH CHECK ADD CONSTRAINT  
[FK_Planta_Produccion] FOREIGN KEY([Cod_Planta])  
REFERENCES [dbo].[Plantas] ([Cod_Planta])
```

6.2.2 Consultas empleadas durante la realización del proyecto:

- Obtención de volumen de captación de todas las plantas por fecha.

```
select Nombre,Volumen from  
Plantas p, Captacion c  
where p.Cod_Planta = c.Cod_Planta  
and Fecha = '2017-01-31'  
order by Volumen desc
```

- Obtención de volumen de captación de una planta por fecha.

```
select Nombre,Volumen from  
Plantas p, Captacion c  
where p.Cod_Planta = c.Cod_Planta  
and Nombre = 'El Troje'  
and Fecha = '2017-01-31'
```

- Obtención de la suma del volumen de captación de una planta de todos los años.

```
select nombre, sum (Volumen) from  
Plantas p, Captacion c  
where p.Cod_Planta = c.Cod_Planta  
and Nombre = 'El Troje'  
group by p.Nombre
```

- Obtención de la suma de volúmenes de captación de todas las plantas por fecha.

```
select sum (Volumen) as 'Captacion Total' from  
Plantas p, Captacion c  
where p.Cod_Planta = c.Cod_Planta  
and Fecha = '2017-01-31'
```

- Obtención del porcentaje de volumen de captación de todas las plantas en un rango de fechas.

```
select Nombre,(sum (Volumen)/(select sum (Volumen) from  
Plantas p, Captacion c  
where p.Cod_Planta = c.Cod_Planta  
and Fecha between '2017-01-31' and '2017-03-31'  
Plantas p, Captacion c  
where p.Cod_Planta = c.Cod_Planta  
and Fecha between '2017-01-31' and '2017-03-31'  
group by nombre
```

- Selección de volumen de producción de todas las plantas por fecha.

```
select Nombre,Volumen from
Plantas p, Produccion pr
where p.Cod_Planta = pr.Cod_Planta
and Fecha = '2018-04-30'
order by Volumen desc
```

- Selección de volumen de producción de una planta por fecha.

```
select Nombre,Volumen from
Plantas p, Produccion pr
where p.Cod_Planta = pr.Cod_Planta
and Nombre = 'Bellavista'
and Fecha = '2018-04-30'
```

- Selección de la suma del volumen de producción de una planta de todos los años.

```
select nombre, sum (Volumen) from
Plantas p, Produccion pr
where p.Cod_Planta = pr.Cod_Planta
and Nombre = 'Bellavista'
group by p.Nombre
```

- Obtención de la suma de volumen de producción de todas las plantas por fecha.

```
select sum (Volumen) as 'Produccion Total'from
Plantas p, Produccion pr
where p.Cod_Planta = pr.Cod_Planta
and Fecha = '2018-04-30'
```

- Obtención del porcentaje de volúmenes de producción de todas las plantas en un rango de fechas.

```
select Nombre,(sum (Volumen)/(select sum (Volumen) from
Plantas p, Produccion pr
where p.Cod_Planta = pr.Cod_Planta
and Fecha between '2018-04-30' and '2018-06-30'
))*100 as 'Porcentaje' from
Plantas p, Produccion pr
where p.Cod_Planta = pr.Cod_Planta
and Fecha between '2018-04-30' and '2018-06-30'
group by nombre
```

- Obtención de volumen de captación y producción de una planta en un rango de fechas.

```
select Nombre, c.Volumen, pr.Volumen from
Plantas p inner join Captacion c
on (p.Cod_Planta = c.Cod_Planta )
inner join Produccion pr
on (p.Cod_Planta = pr.Cod_Planta)
and c.Cod_Captacion <> pr.Cod_Produccion
and Nombre = 'El Placer'
and c.Fecha = pr.Fecha
and c.Fecha between '2017-01-31' and '2017-03-31'
and pr.Fecha between '2017-01-31' and '2017-03-31'
order by pr.Fecha asc
```

6.3 Anexo 3 Codificación UT EPMAPS

Código	Descripción
CAP	Captación de Agua Cruda
PRO	Tratamiento de Agua Cruda

Tabla 6.1 Codificación UT Nivel 1

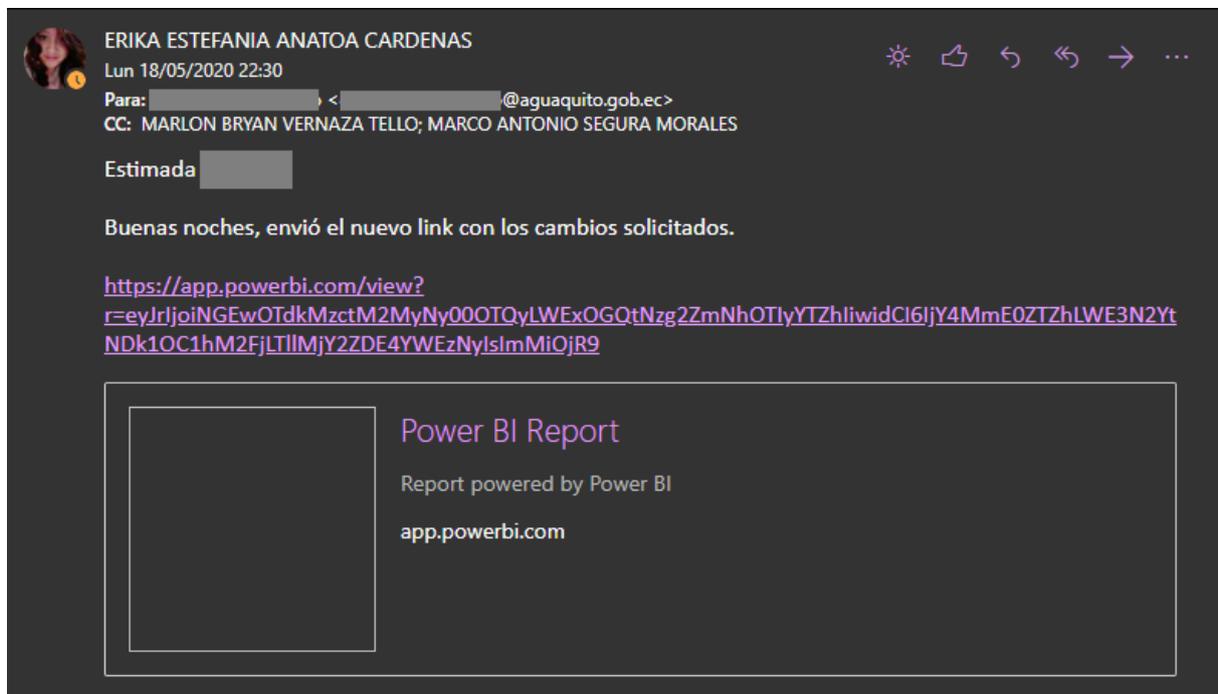
Código	Descripción	Código
PC001	Planta Bellavista	PC001
PC005	Planta Puengasí	PC005
PC007	Planta El Placer	PC007
PC013	Planta El Troje	PC013

Tabla 6.2 Codificación UT Nivel 2

6.4 Anexo 4 Manual de Usuario

Se adjunta el Manual de Usuario

6.5 Anexo 5 Correo de Satisfacción EPMAPS



  <@aguaquito.g

ob.ec>      

Mar 19/05/2020 11:25
Para: ERIKA ESTEFANIA ANATOA CARDENAS
CC: MARLON BRYAN VERNAZA TELLO; MARCO ANTONIO SEGURA MORALES

Muchas gracias por nuestra parte estaria ok.

Saludos cordiales,



¿Las sugerencias anteriores son útiles? Sí No

[Responder](#) [Responder a todos](#) [Reenviar](#)