AUTOMATIZACIÓN DE REPORTES Y DASHBOARDS APLICANDO BUSINESS INTELLIGENCE EN EL ÁREA VP DIGITAL DE LA EMPRESA TIGO

Trabajo de Grado en
Modalidad de Práctica Profesional

Santiago Narváez Romero

Director: PhD. José Luis Arciniegas Herrera
Codirector: PhD. Cristhian Nicolás Figueroa Martínez
Asesor: MBA. Nelson Enrique Tolosa Barbosa

Universidad del Cauca

Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telemática
Línea de Investigación: Servicios Avanzados de Telecomunicaciones
Popayán, Enero de 2021

Hoy una vez más como cada día en mi vida, gracias a Dios
por ayudarme a llegar hasta aquí, ahora entiendo con mayor
claridad que su compañía estuvo siempre desde el inicio,
gracias a mis padres, hermano y familia porque son las personas
que motivaron de cerca cada esfuerzo; a ellos,
les dedico no sólo este, sino cada triunfo de mi vida.
Gracias a mis amigos, la rama IEEE, y en general a
todos los ingenieros y administrativos de la facultad, porque son
muy valiosas las experiencias que guardo en mi corazón,
que aparte de la formación profesional me
permitieron crecer humanamente.

Gracias por todo a todos, sin duda su
presencia en mi vida multiplicó los resultados.

TABLA DE CONTENIDO

LISTADO DE TABLAS

LISTADO DE ILUSTRACIONES

Introducción

Las dos primeras décadas del siglo XXI han dado inicio a una nueva era industrial, donde el principal activo de las compañías son los datos. Los sistemas de comunicación han potencializado la producción masiva de datos a nivel mundial, originando nuevas oleadas de crecimiento de la productividad, innovación y formas únicas de interactuar con los clientes.[1] Esta generación intensa y extensa de información, es lo que hoy se conoce con el nombre de Big Data, concepto que abarca todo tipo de datos: video, voz, imágenes, datos de ubicación, redes sociales, datos de IoT, entre muchos más; con la particularidad de que su crecimiento es exponencial a lo largo del tiempo.

No obstante a las oportunidades que Big Data presenta, las organizaciones enfrentan diferentes desafíos para administrarlo, especialmente en las áreas de gestión y gobernanza de datos[2]. Esto, debido a que, cada proceso interno y externo de la compañía asocia un gran volumen de datos a gran velocidad, sin uniformidad en el formato en que se almacenan.

Debido a ello, la actividad que los ingenieros y científicos de datos hacen para extraer, transformar y almacenar los datos útiles, (procesos resumidos con las siglas ETL: extract - transform - load) tiene gran complejidad [3], pues en muchos casos, el procesamiento de la información se hace manualmente, y es susceptible a fallas en el reconocimiento veraz que deben tener los reportes suministrados para la toma de decisiones.

Teniendo en cuenta lo anterior, es preciso afirmar que el talento humano encargado de dichos procesos ETL, desempeña un rol fundamental en una organización, que debe ejecutarse cuidadosamente, pues cumple el objetivo de organizar los flujos de datos de distintos sistemas y canales de comercialización de la compañía, para transferirlos desde múltiples fuentes, a un almacén de datos, que en adelante constituirá los cimientos de la arquitectura de negocios de la empresa.

Sin embargo, tener datos, es insuficiente; a pesar de que representan la base de la estructura empresarial, la toma de decisiones se fundamenta en información, no en datos brutos, ya que no representan nada en sí mismos. Por ello, lo que en la década de los 60s se describió por primera vez como Business Intelligence (BI) [4], se ha potencializado crecientemente en esta última década, ya que, Big Data genera muchos datos de todo tipo, pero es necesario que estos, adquieran valor y significado para influenciar inteligentemente en la toma de decisiones de una organización.

Así, Business Intelligence, se ha establecido en las grandes empresas como una herramienta indispensable para optimizar la toma de decisiones, implementando un análisis descriptivo de los datos históricos y actuales, y un modelamiento predictivo a medida que cambian los escenarios económicos. Todo esto, con el fin de producir un conocimiento integral de los datos de la compañía mediante representaciones visuales (dashboards) basadas en consultas,

que posibiliten el acceso a información de alto valor empresarial en un formato gráfico de fácil interpretación [5].

Lo descrito anteriormente, muestra el panorama empresarial de interés para el presente proyecto, teniendo en cuenta que se desarrolla bajo la modalidad de práctica profesional. Así, este trabajo se proyecta específicamente en Tigo en Colombia [6], la cual, es una de las empresas líder en el sector de Telecomunicaciones que ofrece productos y servicios de telefonía, televisión e internet. Los procesos de comercialización de estos, varían constantemente en el tiempo, ya que se basan en tecnología y tendencias, lo que exige una adaptación inmediata acorde al contexto comercial del momento.

A principios de 2019, Tigo conformó una alianza con Amazon, vinculando el servicio de Amazon Prime Video (APV) en las ofertas móviles y de hogar para sus usuarios [7]. Desde esa fecha se han almacenado continuamente una variedad de datos, registrando las activaciones, facturaciones y cancelaciones diarias del servicio. Su análisis se lleva a cabo utilizando Excel, tablas y fórmulas manualmente, sin embargo, como se evidenciará en este proyecto, era necesario complementar dicho análisis con técnicas de BI acorde a 2 componentes claves: una visualización gráfica automatizada y un modelamiento predictivo de las activaciones del servicio, con el propósito de entender claramente el comportamiento de los datos en el pasado, el presente y un futuro aproximado.

En adelante, el contenido del documento está organizado por capítulos de la siguiente manera: en el capítulo 1 se escribe acerca de la naturaleza del proyecto y el marco referencial que lo soporta, en el capítulo 2 se describe el proceso de planeación e implementación, en el capítulo 3 se redacta sobre los resultados obtenidos, y en el capítulo 4 se describen las conclusiones.

CAPÍTULO I

1.1 Descripción del Problema

Las grandes empresas comercializan múltiples productos/servicios, y poseen diversos canales de distribución para llegar a miles de clientes en diferentes localizaciones. Para ello, su esfuerzo se centra en entender las necesidades de las personas con mayor precisión, con el fin de mejorar los productos o servicios al público que lo requiere.

Acorde a esto, las organizaciones necesitan conocer a sus clientes, aprovechando los datos que estos otorgan en cada interacción a través de los canales de comercialización e información. Por lo que, las empresas del siglo XXI, han dejado de operar basándose en suposiciones e instintos; en su lugar, utilizan los datos como un recurso base para orientar la toma de decisiones basada en hechos.

Así, competir en la industria hoy en día, implica analizar datos para generar información que oriente la toma de decisiones de una compañía. Específicamente en el sector de las Comunicaciones, el informe conjunto de Informatica y Capgemini sobre cómo convertir Big Data en valor empresarial [8], describe que el 60% de las empresas en 2017, ya adoptaban Big Data en sus procesos operacionales, y resalta que la inclusión del análisis de los macro datos en este sector comercial, mejora en un 50% la toma de decisiones estratégica y provee en un 42% mayor agilidad para reaccionar rápidamente a los cambios en el mercado.

Sin embargo, aunque las compañías entienden los beneficios de Big Data para potencializar su competitividad en el mercado, presentan dificultades en el procesamiento de este enorme conjunto de datos para producir información de alto valor empresarial; pues, en muchos casos el procesamiento de los datos y la generación de reportes se realizan manualmente, y por lo tanto, no sólo genera retardos en las actualizaciones de dichos reportes, sino que la información proyectada en estos, es susceptibles a errores y carece de alta fiabilidad, lo cual es clave para las decisiones de negocio.

Tal es el caso de Tigo, una de las empresas líder en el sector de Telecomunicaciones en Colombia, que ofrece productos y servicios de telefonía, televisión e internet; la cual, lleva a cabo múltiples procesos de comercialización, con la particularidad de que varían constantemente en el tiempo, ya que se basan en tecnología y tendencias, exigiendo una adaptación inmediata acorde al contexto comercial del momento.

Para ello, Tigo ha venido adoptando en los últimos años, Business Intelligence como parte de las actividades base para estudiar el comportamiento descriptivo, predictivo y prescriptivo de: los clientes, los productos/servicios y el crecimiento en ventas en el mercado.

No obstante, debido a la cantidad de datos almacenados diariamente en el Data Warehouse de la empresa, muchos de ellos son de fuentes variadas de información, lo que aumenta la complejidad para extraer la información útil que se necesita y proveer reportes; más aún, cuando hay existencia de procesos que se ejecutan manualmente y no hay algoritmos que faciliten el procesamiento de los datos, provocando retrasos en la generación de los informes, duda en la veracidad de los datos procesados y una representación visual poco eficiente para algunos de sus servicios, lo que en últimas resulta haciendo más compleja la toma de decisiones y planeación estratégica de comercialización.

Es por ello que, el área VP Digital de la empresa, requiere implementar mecanismos de automatización de procesos asociados a la generación de reportes y dashboards que puedan optimizar la producción de información veraz, precisa y clara acorde a los insights de KPIs con respecto al servicio Amazon Prime Video, para obtener un rendimiento autónomo de algunas operaciones repetitivas en el análisis de los datos aplicando Business Intelligence en la ejecución integral del proyecto, con el fin de evitar una lectura incorrecta o incompleta de los datos procesados, y aumentar así, la efectividad en la planeación comercial de los directivos del área VP Digital.

1.2 Formulación del problema

En conocimiento del contexto previamente descrito, es propicio fundamentar el desarrollo del proyecto, teniendo en cuenta la siguiente pregunta:

¿Cómo mejorar el procesamiento de datos en el área Vp Digital de la empresa TIGO para tener una interpretación más precisa e intuitiva de la información requerida con respecto a los KPIs estratégicos del negocio?

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar un mecanismo de automatización de los procesos asociados a la generación de reportes y dashboards a través de herramientas de analítica de datos para Business Intelligence en el área VP Digital de Tigo.

1.3.2 Objetivos Específicos

1. Caracterizar los procesos de generación de dashboards susceptibles de automatización teniendo en cuenta un entendimiento amplio del conjunto de datos para los propósitos de visualización del área VP Digital de Tigo.

2. Diseñar un método de procesamiento autónomo de los datos dispuestos en el DataWarehouse de la empresa conforme a los KPIs de interés en el área VP Digital de Tigo.

3. Proponer un esquema de visualización de reportes mediante el software de BI teniendo en cuenta un conjunto de datos ya procesados.

4. Evaluar los procesos de Business Intelligence implementados con relación a la productividad en la interpretación de la información en el área Vp Digital de TIGO.

1.4 Alcances y limitaciones

1.4.1 Alcances

- El proyecto desarrollado detalla la aplicabilidad de Business Intelligence en la empresa TIGO, con relación a un conjunto de datos específico.

- La implementación del trabajo expone la automatización de los procesos asociados a la generación de los reportes visuales en el software de BI disponible en la empresa.

- El proyecto evidencia la construcción de modelos predictivos para brindar un soporte proactivo a los líderes del área de la empresa acerca del comportamiento de los datos en un futuro.

- El trabajo está orientado a brindar una visualización efectiva de un conjunto de datos específico un conjunto de datos específico acorde a los KPIs requeridos.

1.4.2 Limitaciones

- El proyecto no expone el proceso de recolección de los datos de los diferentes canales digitales de la empresa. Simplemente se hace uso de ellos directamente en el Data Warehouse de la empresa, donde se almacenan.

- El presente trabajo no pretende llevar a cabo la toma de decisiones y/o la planificación de estrategias comerciales con base al análisis de los reportes suministrados.

1.5 Marco referencial

1.5.1 Marco teórico

1.5.1.1 Big Data

Oficialmente, Gartner define big data como [9]: "high volume, velocity, and/or variety data assets that demand new, innovative forms of processing for enhanced decision making, business insights or process optimization". Datos, que fluyen incesantemente de sitios web, sistemas de información, dispositivos móviles, redes sociales, sensores, etc; con la particularidad de que no pueden ser almacenados, gestionados y analizados con sistemas de bases de datos tradicionales; debido a su poca flexibilidad a la hora de: almacenar datos sin estructurar (como imágenes, texto y vídeo), alojar datos de "alta velocidad" (en tiempo real) y gestionar volúmenes más grandes de datos (de varios petabytes) [10]

La cantidad de datos que se generan en este orden y sumamente compleja para su análisis. Como se ha descrito, las empresas cada vez exigen que el procesamiento de dichos datos sea lo más cercano posible al tiempo real, con el fin de generar información de alto valor empresarial al instante que lo exige el mercado; pues aprovechar las oportunidades de Big Data, ofrece una ventaja competitiva considerable que impacta tanto en la industria, como en el negocio e incluso en la sociedad.

Así, Big data se refiere al gran volumen de datos complejos, (semi) estructurados y no estructurados que se generan en un gran tamaño, los cuales llegan (en un sistema) a mayor velocidad para que puedan ser analizados y mejorar no sólo el proceso de toma de decisiones, sino también los procesos de organización estratégica y movimientos comerciales. Por lo que, el concepto de Big Data ha ganado alta popularidad en el siglo XXI con las nuevas aplicaciones llegando a definirse con 5 características, conocidas como las 5Vs de Big Data (volumen, velocidad, variedad, veracidad y valor). [11]

1.5.1.1.1 Las 5 características de Big Data (5Vs)

- Volumen: hace referencia a la enorme cantidad de datos que se generan, recopilan y procesan; al respecto, se habla de tamaño en orden de petabytes, exabytes y zettabytes. Por ejemplo, Twitter procesa y recibe millones de tweets de forma regular. De manera similar, Facebook maneja rutinariamente millones de publicaciones e imágenes. Google recibe más de mil millones de consultas de búsqueda. Además, se recopilan millones de registros de datos de tecnologías de sensores asociadas con el transporte, el clima, sistemas ambientales, etc.

- Velocidad: hace referencia a la velocidad a la que se generan, procesan y mueven los datos entre diferentes sistemas y dispositivos. Los ejemplos incluyen la velocidad de las publicaciones en las redes sociales, transacciones en línea y verificación de fraude, datos de transporte en vivo recibidos de autobuses, trenes, aviones, etc.

- Variedad: corresponde a los diferentes tipos de datos que pueden usarse para lograr la información o los resultados deseados. Los tipos y formatos de Big Data incluyen datos estructurados, semiestructurados y no estructurados.

- Veracidad: se refiere a la calidad de los datos, como corrección, coherencia, confianza, seguridad y confiabilidad. Por ejemplo, tener la seguridad de que los datos no están obsoletos y desactualizados para un propósito determinado. De manera similar, los datos deben ser correctos y consistentes, y deben ser generados por un sistema confiable.

- Valor: hace referencia a los diferentes tipos de beneficios que se pueden derivar del procesamiento y análisis de big data. Los ejemplos incluyen, valor monetario, valor social, valor educativo/investigativo, y otros más.

1.5.1.1.2 Importancia de Big Data

Ahora bien, por qué analizar los macro datos y extraer valor de estos en provecho de Big Data, ha pasado de ser un lujo que las grandes compañías adoptan, a ser el combustible con el que potencializan sus negocios la mayoría de organizaciones.

Google, resalta que: "*data analytics returns more value when you have access to more data, so organizations across multiple industries have found big data to be a rich resource for uncovering profound business insights*" [10], expresando que los macro datos, poseen un potencial enorme para explotar, pues entre más datos, más valor se puede revelar mediante técnicas de Machine Learning (ML), apuntando a reconocer que ML y Big Data son aliados altamente complementarios.

1.5.1.1.3 Tipos de Big Data

Big Data puede clasificarse según la estructura de los datos, de la siguiente manera [12]:

- Datos estructurados: son todos los datos que pueden ser almacenados conforme a un esquema o formato determinado, organizándose fácilmente en campos y relaciones entre ellos, según las topologías comunes que comparten las bases de datos relacionales. De esta manera, se pueden almacenar, acceder y procesar acorde al formato fijo sin mayor complejidad. Algunos ejemplos al respecto involucran datos financieros, datos IoT, datos de contacto, registros de ventas.

- Datos no estructurados: son datos heterogéneos y de naturaleza variable, que no presentan un formato definido. Actualmente su crecimiento sobrepasa la generación de datos estructurados, y debido a ese enorme tamaño y su estructura volátil, proponen múltiples desafíos en términos de su procesamiento para obtener valor de ellos. Algunos ejemplos al respecto involucran datos de imágenes, videos, audio, ficheros de texto, mensajería móvil y correos electrónicos.

- Datos semiestructurados: son datos que pueden verse como una forma estructurada, pero en realidad no poseen una estructura definida. Así, este tipo de datos, mantienen etiquetas y marcas internas que identifican elementos de datos separados, posibilitando la agrupación y jerarquías de información. Algunos ejemplos al respecto involucran archivos comprimidos, datos en formato XML y otros lenguajes de marcado, ejecutables binarios, paquetes TCP / IP.

Adicionalmente, Big Data puede identificarse según su fuente de procedencia, de esta manera:

- Datos generados por máquinas: aquellos que son recopilados por instrumentos físicos, tecnología que automáticamente captura y envía los datos para su respectivo procesamiento. Ejemplo de ello son los sensores, registros web, datos financieros, imágenes satelitales, radares, entre muchos más.

- Datos generados por humanos: aquellos que se generan directamente a partir de la interacción de personas. Ejemplo de ello son los datos de entrada suministrados en dispositivos o navegadores, datos de click-stream, posts en redes sociales, datos móviles, mensajes de texto, entre muchos más.

1.5.1.2 Business Intelligence

En provecho de Big Data y las potentes ventajas que trae consigo, las empresas han fijado su atención en los datos, con la intención de extraer valor sustancial para el negocio. Reconociendo que, si algo se puede medir, se lo puede entender, si algo se entiende, se lo puede controlar; si algo se puede controlar, puede mejorarse.

Acorde a ello, Business Intelligence[1], surge como un conjunto de técnicas y conceptos para recopilar, analizar y distribuir información aprovechando herramientas software y servicios para transformar los datos en conocimientos prácticos que informan las decisiones empresariales estratégicas y tácticas de una organización [13].

Así, es posible que las organizaciones obtengan información de alto valor a partir del procesamiento de datos brutos; con el fin de llevar a cabo soluciones inteligentes, que beneficien al negocio, las cuales serán el producto de una combinación de estrategia y tecnología para recopilar, analizar e interpretar datos de fuentes internas y externas, con el resultado final de proporcionar información sobre el estado pasado, presente y futuro del contexto examinado [14].

[1] En el presente documento se utiliza el termino inglés de Business Intelligence (BI), haciendo alusión a inteligencia empresarial, ya que así es comúnmente adoptado tanto en los entornos académicos como en los profesionales.

1.5.1.2.1 Características de BI

Las principales características de Business Intelligence son [15]:

- Observación estratégica: selección de los datos, que deben ser procesados para descubrir su valor respecto a la necesidad de información.

- Comprensión a profundidad: permite el entendimiento integral del comportamiento de los datos.

- Predicción: permite desarrollar pronósticos de los datos en el futuro, con el fin de anteceder los hechos desde una perspectiva analítica de los sucesos que podrían ocurrir.

- Colaboración: permite difundir y compartir los reportes generados entre los usuarios que la necesiten.

- Decisión: brinda reportes intuitivos con la información sustancial del negocio a partir de la cual pueden proponerse planes de acción favorables.

1.5.1.2.2 Características generales de herramientas software de BI

Teniendo en cuenta que Business Intelligence es un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones [16]. A continuación se describen las características comunes más importantes de las herramientas software de BI más utilizadas en el mercado [17].

- Informes en tiempo real: permite realizar reportes con los datos procesados de las operaciones que suceden en el momento, con la posibilidad de incluir alertas sobre los indicadores de rendimiento.

- Agregación y asignación multidimensional: La estructura de datos multidimensional y las numerosas dimensiones de la inteligencia empresarial permiten al usuario navegar rápidamente por el sistema y analizar los datos desde varias perspectivas, posibilitando además, el procesamiento de las consultas casi al instante.

- Amplitud de conexión de datos: permite importar datos ya sea de aplicaciones externas, locales o plataformas en la nube fácilmente, sin necesidad de que el talento humano de la organización invierta tiempo en la conversión a formatos de datos estándares.

- Capacidades de Autoservicio: brinda autonomía a un equipo u organización para generar sus propios reportes, determinar los indicadores de análisis y ejecutar sus propias consultas sin necesidad de soporte externo.

- Interacción con fuentes de datos no estructuradas: permite el procesamiento de la información lingüística, auditiva y visual, a través de métodos automatizados, con el fin de recuperar datos significativos y luego crear datos estructurados sobre la información requerida.

- Análisis reactivo y proactivo: permite el análisis de los datos pasados y presentes, y en base a ellos, desarrollar modelos de predicción, para prever su comportamiento en un futuro.

1.5.1.2.3 Beneficios

Los beneficios más destacados de Business Intelligence son [18]:

- Ayuda a la toma de decisiones: mediante las herramientas disponibles, la representación visual de los datos actualizados, la interacción con cuadros de control y muchas funciones más, permite optimizar la toma de decisiones oportuna de los líderes, asimilando información confiable de alto valor en cualquier momento. Todo esto, para desarrollar estrategias comerciales más sólidas; impulsar mayores ventas y nuevos ingresos; y obtener una ventaja competitiva sobre las empresas rivales.

- Accesibilidad a la información: centraliza la información en un solo repositorio, en lugar de administrar los múltiples canales de datos. Esto, con el fin de que los usuarios que necesitan acceder a dicha información puedan realizar las consultas directamente sin mayor complejidad.

- Ahorro de tiempo y costos: gracias a la posibilidad de procesar los datos automáticamente, no requiere de la manipulación manual por parte del talento humano de la empresa, lo cual evita retrasos en la generación de los informes, aumenta la confiabilidad de la información suministrada y prescinde de invertir recursos para ello.

1.5.1.2.4 Fuentes de Datos para BI

Actualmente, los principales sistemas que sirven de fuente (origen) de datos para operar con Business Intelligence son [19]:

- Data Warehouse: es una base de datos corporativa en la que se integra información depurada de las diversas fuentes que hay en la organización, para luego procesarla permitiendo grandes velocidades de respuesta.

- Data Mart: es una base de almacenamiento especializada en integrar los datos de un área de negocio específica; con la particularidad de que posee una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de esa área determinada.

Un Data Warehouse y un Data Mart se diferencian por el alcance en que operan. Mientras que un Data Warehouse es un sistema centralizado que posee datos globales de la empresa y todos sus procesos operacionales, un Data Mart es un subconjunto de datos, orientado a administrar los datos de un área o proceso específico.

1.5.1.2.5 BI Software

Ahora bien, para implementar BI en una organización, es necesario contar con herramientas software para dar sentido a las enormes cantidades de datos que las organizaciones acumulan a lo largo del tiempo. Las herramientas de BI analizan esta información y la presentan como información procesable que puede guiar la toma de decisiones. Con estas herramientas, los usuarios pueden interactuar con los datos, analizarlos, presentarlos y más. En simples palabras, dar vida a los datos.

Actualmente, las herramientas de BI más populares, por la efectividad que representan para las compañías son [20]:

- Microsoft Power BI
- Tableau
- Google Data Studio
- Looker
- QLIK y QLIKSense

Según Gartner, las herramientas de BI para el 2020, se han posicionado de la siguiente manera [21]:

Ilustración 1. Ranking de las herramientas de BI en 2020. Recuperado de www.gartner.com

Gartner no posiciona en el ranking a Data Studio como software de BI, pues no es una herramienta que en sí misma desarrolla BI, ya que el sistema de inteligencia empresarial de Google es BI Engine, y gracias al servicio que ofrece de análisis de datos en memoria rápidamente, es integrado en Data Studio para procesar los datos y realizar reportes visuales.

Sin embargo, vale aclarar que Gartner reconoce que las capacidades de Data Studio operan como herramienta de BI, aunque no lo posicione en el ranking como se comentó anteriormente [22].

1.5.2 Marco Conceptual

A continuación, se exponen los conceptos fundamentales estudiados y aplicados en la presente propuesta, para mayor amplitud del conocimiento aquí descrito, observar el Anexo (MarcoTeoricoConceptual_complemento.pdf).

1.5.2.1 Conceptos de Inicio

- **Dato:**

Es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades. Sin embargo, cuando se encuentran aislados, no representan nada en sí mismos, es necesario examinar conjuntamente una agrupación de ellos, para obtener información [23].

Así, los datos son considerados como bits de información, que adquieren valor cuando son procesados.

- **Información:**

Es un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre [24].

Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre)

- **Conocimiento:**

Es una mezcla de experiencia, valor, información y saber hacer; representa un marco para integrar nuevas experiencias e información y es útil para la acción. Se origina y se aplica en la mente de las personas que conocen el contexto [24].

En complemento, el diccionario de Oxford lo describe como la facultad del ser humano para comprender por medio de la razón la naturaleza, cualidades y relaciones de las cosas [25].

- **Dashboard:**

Es un panel de datos en el que las empresas visualizan la información más importante, es decir, una representación gráfica de los principales KPIs, permitiendo la optimización de la estrategia de la empresa [26].

Algunas características generales de un dashboard son [27]:

- Diseño agradable y de uso intuitivo
- Acorta el tiempo de reacción, ya que puede supervisarse por varios usuarios al tiempo.
- Posee indicadores para medir el rendimiento, con el fin de evaluar los objetivos del negocio.
- Flexible, ajustable y evolutivo
- Permite filtrar información respecto a dimensiones (ej: fecha, género, localidad)
- Presenta gráficos acordes a la necesidad de información.

- **Toma de decisiones:**

En la literatura, se encuentran múltiples definiciones al respecto, sin embargo el estudio [28], identifica 3 definiciones destacadas en el periodo de tiempo de los últimos 50 años:

- Herbert A Simon (1976): "Systematic process to choose the option that offers the best chances of improving the efficiency and effectiveness of the organization in order to create value in all stakeholders".
- Omar Aktouf: (1998): "Process by which a supposedly clarified, informed and motivated option is reached".
- Koontz, Weihrich y Cannice (2012): "Selection of a course of action among several alternatives"

Basado en las definiciones anteriores, el contexto empresarial y el desarrollo de este proyecto en particular, es propicio proponer la siguiente definición:

La toma de decisiones es un proceso cognitivo que da como resultado la selección de una alternativa entre múltiples opciones posibles, teniendo en cuenta el momento, el contexto espacial y el procesamiento de toda la información requerida, con el fin de mejorar la efectividad, al momento de crear valor en el negocio.

1.5.2.2 Tecnologías y herramientas software

- **BigQuery:**

Oficialmente Google describe Bigquery como un almacén de datos multinube de alta escalabilidad, rentable y sin servidor, diseñado para agilizar un negocio [29].

Ventajas:

Las principales ventajas de usar Bigquery para administrar los datos de una empresa son:

- Obtener información valiosa con analíticas en tiempo real.
- Acceder a los datos y compartir información valiosa de forma sencilla.
- Proteger los datos y trabajar con seguridad.

- **Google Data Studio:**

Es una herramienta gratuita de Google que permite realizar dashboards e informes personalizados mediante datos de las diferentes herramientas de marketing de Google y otras fuentes externas [30]. De esta forma, ayuda a convertir la información, en informes gráficos precisos que ofrezcan respuestas a las preguntas más importantes del negocio.

- **Anaconda:**

Es una suite de código abierto que abarca un conjunto de aplicaciones, librerías y conceptos diseñados para el desarrollo de la ciencia de datos con Python. Desde una perspectiva general, Anaconda funciona como un gestor de entorno, un gestor de paquetes y tiene una colección de más de 720 paquetes de código abierto [31].

- **Jupyter Notebook:**

Es una aplicación web de código abierto que puede utilizarse para crear, compartir y editar documentos que contienen código en vivo, ecuaciones, visualizaciones y texto [32].

Está diseñada generalmente para tener una compatibilidad avanzada con Python; es comúnmente usada para la limpieza y transformación de datos científicos, la simulación numérica, el modelado estadístico y puede abarcar muchas otras áreas [33].

1.5.2.3 Lenguaje de codificación

- **Python:**

Es un lenguaje de programación interpretado, orientado a objetos de alto nivel y con semántica dinámica. Su sintaxis, hace énfasis en la legibilidad del código, lo que ayuda a depurar el código y, por lo tanto, mejora la productividad [34].

Python, es un lenguaje frecuentemente usado para la ciencia de datos, debido a que ofrece múltiples librerías de herramientas científicas, numéricas, herramientas de análisis y estructuras de datos. Lo cual, favorece el desarrollo de proyectos en este contexto, ofreciendo además, la integración de librerías como NumPy, SciPy, Matplotlib, Pandas e incluso la

implementación de algoritmos de Machine Learning como se observará en el capítulo de implementación.

1.5.2.4 Modelos de Predicción

- **Modelos ARIMA**

El acrónimo ARIMA (AutoRegressive Integrated Moving Average) es descriptivo y captura los aspectos clave del modelo en sí. En resumen, son [35]:

- *AR – Auto regresión:* utiliza la relación dependiente entre una observación y cierto número de observaciones retrasadas.
- *I - Integración:* utiliza la diferenciación de observaciones en bruto (por ejemplo, restando una observación de una observación en el paso de tiempo anterior) con el fin de hacer estacionaria a la serie de tiempo.
- *MA - Media móvil:* utiliza la dependencia entre una observación y un error residual de un modelo de promedio móvil aplicado a observaciones retrasadas.

Cada uno de estos componentes se especifica explícitamente en el modelo como parámetro. Se utiliza una notación estándar de ARIMA (p, d, q) donde los parámetros se sustituyen por valores enteros para indicar rápidamente el modelo ARIMA específico que se está implementando.

Los parámetros del modelo ARIMA se definen de la siguiente manera:

- *p:* el número de observaciones de retraso (lag) incluidas en el modelo, también llamado orden de retraso.
- *d:* el número de veces que se diferencian las observaciones sin procesar, también llamado grado de diferenciación.
- *q:* el tamaño de la ventana de media móvil, también llamado orden de media móvil.

Los modelos ARIMA más representativos son:

- ✓ AR (Autoregressive Model)
- ✓ MA (Moving Average)
- ✓ ARMA (p,q) y ARIMA (p,d,q)
- ✓ SARIMA (p,d,q) (P, D, Q) m

- **Deep Learning:**

La inteligencia artificial (IA), es un término genérico que hace referencia a la inteligencia llevada a cabo por máquinas. El aprendizaje automático es un subconjunto de la IA y el aprendizaje profundo (Deep Learning) es un subconjunto del aprendizaje automático.

Específicamente, Deep Learning es una técnica de aprendizaje automático que enseña a las computadoras a hacer lo que es natural para los humanos: aprender con el ejemplo. Su propósito es modelar abstracciones de datos de alto nivel usando arquitecturas computacionales, que admiten transformaciones de datos, expuestos en forma tensorial o matricial [36].

Tiene sus raíces en las redes neuronales, las cuales son conjuntos de algoritmos, modelados libremente a partir del cerebro humano, que están diseñados para reconocer patrones. Los modelos se entrenan utilizando un gran conjunto de datos etiquetados y arquitecturas de redes neuronales que contienen muchas capas [37].

- **Redes Neuronales Recurrentes**

Su aplicación se enfoca en el conjunto de datos secuenciales como por ejemplo: series temporales (ej: registro de ventas), audio, trayectorias de vehículos, música. Pueden procesar tanto a la entrada como a la salida secuencias sin importar su tamaño, teniendo en cuenta la correlación existente entre los elementos de esa secuencia [38].

1.5.3 Antecedentes

El proceso de búsqueda de literatura para conformar los antecedentes del presente proyecto, se realizó teniendo en cuenta la metodología de revisión literaria propuesta por el trabajo**: "La revisión de la literatura científica: Pautas, procedimientos y criterios de calidad"** [39] en complemento con el artículo "**Recuperación de Arquitecturas de Software: Un Mapeo Sistemático de la Literatura"** [40], con el fin de establecer la siguiente estrategia de revisión sistemática: 1. definición de palabras clave y consulta en las bases de datos incluyendo criterios de búsqueda (fecha: 2015-2021, área de investigación, idioma: inglés o español), 2. clasificación de los resultados obtenidos y selección primaria de artículos conforme a la afinidad del presente proyecto, 3. estudio y selección final de los artículos con alto grado de afinidad.

Primero, la búsqueda inicial se realizó mediante el uso de "**Business Intelligence**", "**Business dashboards automation**" y "**Data analytics**" como palabras clave utilizando las bases de datos Science Direct, IEEExplore, ArXIV, Recolecta, 1Findr y BDigital (Universidad Nacional). Los artículos resultantes y los artículos basados en sus referencias se analizaron hasta que surgieron temas comunes más grandes. En algunos casos, la intuición fue lo suficientemente sugerente para proporcionar temas más amplios como: "**Big Data Analysis in Business**", "**Business Report Processing**" y "**Processes in Dashboard Making**" y, por lo tanto, también se analizaron artículos afines para estos temas. En la tabla 1, se exponen la cantidad de resultados obtenidos para cada base de datos consultada.

Tabla 1. Cantidad de artículos encontrados por base de datos bibliográfica. Fuente: el presente trabajo, 2021

Fuente /Palabras Clave	Business Intelligence	Business dashboards automation	Data analytics	Big Data Analysis in Business	Business Report Processing	Processes in Dashboard Making	TOTAL
Science Direct	2395	916	8761	1811	2844	223	16950
IEEExplore	5273	15	7253	1768	767	147	15223
1FINDR	404	3	443	53	759	10	1672
Recolecta	863	23	598	478	4590	103	6655
BDigital	1015	8	724	273	847	54	2921
ArXIV	527	16	2045	301	903	69	3861
TOTAL	10477	981	19824	4684	10710	606	**47282**

De los artículos resultantes en las bases de datos, se filtraron los resultados con respecto al área de aplicación: "**information/data science**" y "**management/business**", obteniendo un total de 221 artículos mediante una revisión superficial, que tenía en cuenta la lectura del título y resumen para verificar la concordancia del conocimiento que se necesitaba para este proyecto. En segundo lugar, se ejecutó una revisión más detallada de los 221 artículos, con el fin de seleccionar los 67 más afines (**ver anexo BúsquedaDeLiteratura.pdf, tabla 1**). Este proceso, conllevó a la tercera etapa de estudio, donde se analizaron a profundidad 30 papers, clasificándolos en una tabla (**ver anexo BúsquedaDeLiteratura.pdf, tabla 2**) con los campos: número, título, base de datos, link de descarga; para poder señalarlos convencionalmente con colores, según el grado de aporte al proyecto, (verde: muy afín, azul: medianamente afín, naranja: poco afín). Como resultado, 7 se descartaron, por brindar un bajo aporte a la investigación, 23 son considerados artículos muy útiles para este trabajo, y 16 de estos, fundamentales para esta investigación, los cuales construyeron el marco de referencia para el estado del arte del proyecto como se observa en la tabla 2.

Tabla 2. Artículos seleccionados que conforman el estado del arte. Fuente: el presente trabajo, 2021

1	Data lakes in business intelligence: reporting from the trenches	**2**	Optimización del mantenimiento industrial mediante técnicas BI. Aplicación de un cuadro de mando integral	**3**	Airport Trends Analytics Engine using the ARIMA Model
4	Using or Not Using Business Intelligence and Big Data for Strategic Management: An Empirical Study Based on Interviews with Executives in Various Sectors	**5**	An artificial intelligence and knowledge-based system to support the decision-making process in sales	**6**	The role of the performance dashboard in the management of modern enterprises
7	Impact of business analytics and enterprise systems on managerial accounting	**8**	Transforming telecom business: Scaling the shift using predictive analytics	**9**	Implementation of Sales Executive Dashboard for A Multistore Company in Yogyakarta
10	Analysis, reporting and forecasting with qlikview	**11**	Measuring the success of changes to Business Intelligence solutions to improve Business Intelligence reporting	**12**	Developing dashboards for SMEs to improve performance of productive equipment and processes

13	Visualization, storyboarding and applications	**14**	Opportunities for the Use of Business Data Analysis Technologies	**15**	Specification and derivation of key performance indicators for business analytics: A semantic approach
16	Analítica de datos con aplicación en un caso práctico, mediante el uso de una herramienta libre				

En las últimas décadas, el sector industrial ha estado enfrentando uno de los retos más complejos, debido al incremento masivo de datos que segundo a segundo se generan en cualquier lugar del mundo [41], exigiendo la atención de los directivos para implementar procesos asociados a la extracción y transformación de datos en información productiva, para la toma de decisiones oportuna. El tiempo es dinero, es un eslogan bien conocido en los negocios; pero actuar a tiempo, tomar decisiones bien fundamentadas basadas en la comprensión adecuada del propio negocio puede ser la diferencia entre el éxito y el fracaso; por lo que, controlar los hechos que genera Big data y el efecto que genera en las empresas es de suma importancia de la mano de las técnicas de Business Intelligence (BI).

En el año 2015, Oracle refirió 7 etapas fundamentales para llevar a cabo los procesos de análisis de datos en el contexto empresarial con BI [42], por lo que, en provecho de ello, la literatura referenciada a continuación, seguirá el ordenamiento descrito por Oracle de esta manera:

1. **Definición de requerimientos:** entendimiento de los objetivos empresariales.
2. **Adquisición de datos:** recolección y preparación de los datos.
3. **Entendimientos de los datos.**
4. **Manipulación de los datos:** filtros, procesamiento y depuración de datos.
5. **Publicación de los datos:** diseño ilustrativo de la información.
6. **Análisis de datos.**
7. **Toma de decisiones.**

Debido a la necesidad latente de la empresa Tigo en el área VP Digital, el objeto de este trabajo y la literatura encontrada está fundamentada en los procesos 3, 4 y 5 principalmente, sin embargo, ha sido necesario entender los procesos 1 y 2 de igual manera, ya que son el lienzo sobre el cual se construirá este proyecto; y los procesos 6 y 7 no se tienen en cuenta para la ejecución de esta propuesta, ya que corresponden a las funciones de operación de los directivos del área.

Teniendo en cuenta lo anterior, fue propicio empezar con **1. definición de requerimientos** a nivel corporativo, con el fin de entender el rol de los KPIs (Key Performance Indicators) y el planteamiento transparente de objetivos en el desarrollo de las actividades empresariales. El artículo "**Specification and Derivation of KPI for business Analytics: A semantic Approach**" [43] propone un enfoque descriptivo que proporciona a los responsables de la toma de decisiones una visión integrada de los objetivos comerciales estratégicos y los KPI del almacén de datos conceptuales. El principal beneficio de la propuesta, es que vincula los modelos comerciales estratégicos con los datos, para monitorearlos y evaluarlos. Lo que en relación con el artículo "**Optimización del mantenimiento industrial mediante técnicas de BI. Aplicación de un cuadro de mandos integral**" [44] fue beneficioso, en la medida que este, resume los detalles conceptuales del proceso de planificación de KPIs y ejemplifica

la manera en que deben redactarse para brindar un entendimiento claro al momento de evaluar mediciones en el proceso de inteligencia empresarial. El documento "**Transforming telecom business- Scaling the shift using predictive analytics**" [45] lleva los conceptos aprendidos previamente, al análisis práctico, resaltando en primera medida, el proceso evolutivo de las empresas de Telecomunicaciones con el surgimiento de los Datos, como un nuevo servicio comercial que demandan los usuarios actualmente; debido a este cambio, el artículo describe cómo las métricas y los KPI heredados que alguna vez se usaron para escalar y comprender el rendimiento de la industria, tuvieron que evolucionar según las necesidades del nuevo entorno, para proporcionar una imagen más precisa de la industria a medida que se sometía a una transformación empresarial.

Así, los artículos permitieron entender cómo las empresas deben ser adaptables a las necesidades latentes del momento, según cómo el mercado y los clientes lo demanden, por lo que, aprender la definición de objetivos y KPIs en el desarrollo de actividades empresariales es de suma importancia, en la medida que orientan los esfuerzos de BI, a un cumplimiento exitoso de las metas planteadas.

En segundo lugar, Oracle describe **2. la adquisición de datos** como un proceso de suma importancia para una compañía, ya que, de la efectividad de esta labor, depende en gran medida la calidad de ejecución de la inteligencia empresarial. El artículo "**Opportunities for the Use of Business Data Analysis Technologies**" [46] empieza describiendo la necesidad de técnicas de minería de datos en el contexto de Big Data, e investiga, cómo hoy en día las empresas buscan y extraen de forma automática, datos provenientes de fuentes internas o externas a la empresa, con el fin de descubrir patrones y tendencias que describan el comportamiento de múltiples variables como: número de ventas, gustos, tipo de productos/servicios, necesidades del mercado actual y futuro, e incluso para llegar a tener una percepción genérica de las personalidades de los clientes. Además, el artículo resalta los algoritmos de minería de datos más usados para la inteligencia empresarial y su aplicabilidad dependiendo de las necesidades de extracción y análisis de los datos.

Con el propósito de proveer un repositorio confiable a partir del cual se ejecutará la inteligencia empresarial, el proceso de extracción de datos debe complementarse mediante una tecnología de almacenamiento que agilice y facilite el acceso para el procesamiento de los datos tal y como lo expresa el documento "**Data Lakes in BI: reporting from the trenches**" [47] cuyo objetivo está en aprender a administrar y almacenar un gran volumen de datos, mediante la aplicabilidad de Data Lakes y Data Warehouses (DW), como tecnologías que facilitan la extracción, almacenamiento y procesamiento de datos brutos, garantizando rápido acceso a un gran volumen de datos sin demandar alta disponibilidad de tiempo para conseguirlo. El artículo, concluye 3 propósitos para la implementación de Data Lakes en una empresa: 1. Como áreas de preparación o fuente para Data Warehouse (almacenes de datos), 2. Como plataforma de experimentación para científicos y analistas de datos, 3. Como fuente directa para autoservicio de inteligencia empresarial.

Estos artículos orientan el estudio de los datos para conformar una fuente de información centralizada a nivel interno de una empresa. Aportan concisamente al proyecto, instruyendo en los métodos y técnicas existentes, para el reconocimiento de fuentes de datos, la minería

de datos y las tecnologías de almacenamiento propicias; lo que construye el repositorio de elementos sobre el cual será soportado el análisis empresarial.

Una vez se han definido los requerimientos y se han capturado y almacenado los datos en los repositorios de la empresa, la inteligencia de negocios, posee los elementos para construir la lógica empresarial mediante los procesos de**: 3. entendimiento de los datos**, **4. manipulación de datos y publicación**. Para aprender su ejecución, la literatura fue lo suficientemente clara en describir y ejemplificar sucesos de interés a nivel corporativo en los que la aplicabilidad de técnicas de BI resultó efectiva. El artículo "**Analysis, reporting and forecasting with QLIKVIEW**" [48] resalta el aporte competitivo que brinda a una empresa, el aplicar inteligencia de negocios a sus actividades de comercialización, y cómo, la industria ha desarrollado herramientas enfocadas al análisis de datos para la generación de reportes. El estudio se fundamenta en el uso del software de BI llamado QLIKVIEW, y argumenta un caso de uso puntual de la empresa respecto a la evolución en ventas. El estudio, se enfoca en el análisis de los datos a partir de los objetivos y KPIs planteados inicialmente, luego procede a explicar la elaboración de un reporte gráfico como resumen de los datos procesados de interés para la toma de decisiones y culmina explicando a groso modo un ejemplo de predicción, ilustrando cómo múltiples variables pueden afectarse si el precio del producto o la cantidad de ventas cambian. El artículo "**Airport Trends Analytics Engine using the ARIMA Model**" [49] expone el desarrollo de un nuevo motor de datos que utiliza R y Tableau para predecir las tendencias del aeropuerto. El motor desarrollado utiliza conjuntos de datos univariados, como: movimiento de pasajeros del aeropuerto de Perth y estadísticas de carga del aeropuerto de Newark; con el fin de analizar y predecir tendencias precisas. El análisis y la predicción de datos se realizó con la implementación del análisis de series temporales y modelos ARIMA para los respectivos módulos. El desarrollo de los módulos se realizó mediante RStudio, mientras que Tableau (Software de BI) fue utilizado para la visualización interactiva y la generación de informes del usuario final.

El artículo "**Using or not using BI and Big Data for Strategic Management: an empirical study base on interviews with Executives in various Sectors**" [50] fue llevado a cabo entrevistando a diez altos ejecutivos de empresas líder en Turquía, que están haciendo negocios en varios sectores industriales, discutiendo cómo o cuánto usan los sistemas de información ejecutiva cuando toman decisiones. El artículo, expresa en detalle los contextos y procesos en los que la inteligencia de negocios potencializa el enriquecimiento de información veraz para la toma de decisiones oportuna. Se enfoca en estudiar el aporte que BI tiene en la generación de reportes gerenciales cuando se hace una manipulación efectiva de los datos, con el fin de producir información acorde a las necesidades de planeación o toma de decisiones en un contexto determinado de la empresa. En complemento a ello, el artículo "**Impact of Business Analytics and Enterprise Systems on Managerial Accountin**g" [51] expone una claridad del concepto de dominio, orientación y técnicas para desarrollar un proyecto de inteligencia empresarial. Este estudio, propone un marco de análisis de datos de contabilidad gerencial (MADA) basado en la teoría del cuadro de mando integral en un contexto de inteligencia empresarial. MADA, brinda a los contadores de gestión la capacidad de utilizar análisis empresariales completos para realizar mediciones de desempeño y proporcionar información relacionada con las decisiones. En este desarrollo, los autores implementan tres tipos de análisis de negocios (descriptivo, predictivo y prescriptivo) en cuatro perspectivas de medición del desempeño corporativo (financiero,

cliente, proceso interno y aprendizaje-crecimiento) en un entorno de sistema empresarial. Este documento contribuye a la prueba de la estructura organizacional, al analizar el impacto de la analítica empresarial en la contabilidad administrativa desde una perspectiva de BI y sistemas empresariales.

Para proveer información veraz de un conjunto de datos, es necesario llevar a cabo el procesamiento de estos, mediante técnicas de analítica de datos. Algunos de los artículos presentados, mencionan técnicas estadísticas, de clasificación y algoritmos sistemáticos para potencializar la producción de información efectiva. El artículo "**An artificial intelligence and Knowledge-based system to support the decision making process in sales**" [52] sustenta el desarrollo de un sistema basado en conocimiento, el cual, fundamentado en las reglas del negocio, permitió apoyar los procesos de toma de decisiones del departamento de ventas de una empresa. El sistema knowledge-based pudo aportar confiabilidad y agilidad al proceso de toma de decisiones, y permitió simular escenarios futuros para la empresa, de acuerdo con el comportamiento combinado de variables clave. La principal contribución de este estudio, es un informe de un caso en el que el sistema Knowledge-based ayudó a encontrar mediante inteligencia artificial una alternativa adecuada a un problema comercial en el sector de ventas de una empresa en el sur de Brasil. El artículo sugiere que, para un proyecto de BI, todas las variables identificadas en el estudio en el proceso de Entendimiento de los datos, tienen alto grado de importancia para el proceso de toma de decisiones; deben analizarse conjuntamente para producir una respuesta confiable, de lo contrario, puede surgir una decisión incorrecta que dañe la ejecución de la estrategia de la empresa.

De igual manera, a nivel nacional, Ingenieros de la Universidad Tecnológica de Pereira, expusieron en 2018 su trabajo de grado titulado "**Analítica de datos con aplicación en un caso práctico, mediante el uso de una herramienta libre**" [53] explicando ampliamente la importancia que la analítica de datos tiene hoy en día en el contexto empresarial. El documento describe el estudio de cuatro herramientas informáticas de analítica de datos con sus características, a partir del cual, se estima la selección de la herramienta WEKA, con el fin de aplicar la analítica de datos a un caso práctico, que tenía como fin, predecir el tipo de fármaco qué se debe administrar a un paciente afectado de rinitis alérgica según distintos parámetros/variables empleando el método de clasificación, de árboles de decisión, con el cual se obtuvo un grado de acierto del 92%-99% para las pruebas realizadas. El caso de estudio llevado a cabo en este artículo es muy descriptivo, permitiendo aprender las herramientas de BI efectivas para un contexto de uso particular y las técnicas de procesamiento de los datos en un ejemplo puntual.

El artículo "**Implementation of Sales Executive Dashboards for a Multistore Company in Yogyakarta**" [54] expresa la necesidad de una empresa en Yogyakarta, de generar informes de ventas en tiempo real de 3 tiendas distribuidas en la ciudad. El documento expone la construcción de un modelo de dimensión de los datos de ventas, los cuales requirieron ser procesados previamente mediante las métricas ETL de Big Data: extracción, transformación y almacenamiento. La investigación resalta la construcción de dicho modelo, usando los datos de 1 año (1 de febrero de 2014 al 31 de enero de 2015), instruyendo en la implementación de dashboards ejecutivos, para monitorear y analizar las condiciones de venta en función de las dimensiones: tiempo, punto de venta/tienda y producto. El aporte de

este trabajo radica en la explicación detallada de los procesos de: filtrado, transformación de los datos, cálculo de medidas y la representación gráfica de informes gerenciales.

A partir del procesamiento de los datos, realizado con base a los objetivos empresariales, la representación de la información producida es crucial para la toma de decisiones, donde, se pretende ilustrar gráficamente, el estado de variables y recursos, teniendo presente, flujos de tiempo para brindar un entendimiento intuitivo de los directivos que visualizan los reportes, con el propósito de extraer información veraz y completa.

Respecto a ello, la búsqueda de la literatura, concluye con el aporte de 3 documentos. Primeramente, el documento "**The role of performance dashboard in the management of modern enterprises**" [55], expone las características de una empresa moderna en el ámbito de gestión y planificación de estrategias comerciales, diseña una serie de etapas para la construcción de Dashboards, teniendo en cuenta que su representación, está ligada directamente a la lectura de los KPIs, como elementos indicativos del progreso para el cumplimiento de objetivos en las actividades empresariales. El artículo, estudia 3 tipos de Dashboards: estratégicos, comerciales y operacionales; los cuales, se pueden implementar, dependiendo del contexto y los usuarios a quien corresponde su lectura.

En complemento a ello, la investigación "**Developing Dashboards for SMEs to improve performance of productivity equipment and processes**" [56] propone un procedimiento para desarrollar Dashboards para pequeñas y medianas empresas, el cual se provisiona de una fuente de datos confiable a nivel interno en la empresa; el objetivo de dicho reporte visual, es mejorar el rendimiento de los equipos y procesos productivos a nivel de planta, proporcionando información de manera eficiente a las áreas productivas y convertir esta información en conocimiento, planes y acciones que promuevan una actividad efectiva en el taller. Una de las fases principales, el desarrollo del diseño del Dashboard, se realizó teniendo en cuenta la gestión visual y los enfoques de mejora continua, como Kaizen y el Mantenimiento Productivo Total.

Finalmente, el capítulo "**Visualization, storyboarding and applications**" [57], extraído del libro **Building Big Data Applications**, publicado el año 2020, aporta singularmente en la implementación de informes visuales para el entendimiento integral de la información procesada a partir de un gran conjunto de datos. Plantea técnicas para aprovechar y desarrollar un guion gráfico (inherente a un Dashboard), utilizando los fundamentos de visualización y storyboard, con una integración de nuevas tecnologías, bases de datos y sistemas analíticos existentes. El aporte del documento radica en relacionar las temáticas de Big Data como: cómputos de datos, procesamiento de datos distribuidos, fórmulas analíticas, ejecuciones de algoritmos de aprendizaje autónomo, entre otras, con el desarrollo de visualizaciones y aplicaciones intuitivas que faciliten la abstracción de conjuntos inmensos de datos, en información fácilmente entendible para la toma de decisiones.

1.5.4 Marco Contextual

El presente trabajo se realiza bajo la modalidad de práctica profesional; se desarrolla en la empresa TIGO en Colombia en la ciudad de Bogotá, en las instalaciones del centro Empresarial Conecta.

TIGO, presta servicios de telecomunicaciones, tecnologías de la información y las comunicaciones y actividades complementarias. Sus accionistas principales son Empresas Públicas de Medellín E.S.P. y Millicom Spain S.L.

A la fecha del 31 de diciembre de 2019, la Compañía estaba organizada así:

- **Negocio de Hogares:** atiende hogares en el territorio nacional, incluyendo regiones donde está presente la filial EDATEL.

  Servicios:

  - Voz
  - Televisión
  - Internet
  - Amazon Prime Video (servicio de valor agregado)
  - HBO Go (servicio de valor agregado)

- **Negocio Móvil:** atiende personas con líneas móviles en el territorio nacional y se atiende desde la filial Colombia Móvil.

  Servicios:

  - Equipos
  - Postpago
  - Prepago
  - Amazon Prime Video (servicio de valor agregado)
  - Más servicios de valor agregado

- **Negocio de Empresas y Gobierno:** atiende entidades gubernamentales, clientes empresariales corporativos y Pymes, incluyendo la gestión de las filiales CTC y OSI.

  Servicios:

  - Conectividad movilidad y seguridad
  - Voz, Cloud y Datacenter
  - Servicios Digitales y Televisión

A inicios del año 2020, TIGO tiene presencia en el 96% de las zonas urbanas del territorio nacional. La tabla 3, especifica su cobertura nacional.

Tabla 3. Cobertura de TIGO en Colombia. Recuperado de www.tigo.com.co/sites/tigounecorp/files/fragmentos/general_listado_archivos/IGS_UNE_2019_Consolidado.PDF

REGIONALES	COBERTURA		
	UNE	COLOMBIA MÓVIL	EDATEL
Centro	Cundinamarca, Meta, Boyacá	Amazonas, Casanare, Cundinamarca, Guainía, Guaviare, Meta, Putumayo, Vaupés, Vichada	
Costa	Atlántico, Bolívar, Cesar, Magdalena, Sucre, Córdoba	Atlántico, Bolívar, Cesar, Córdoba, Guajira, Magdalena, Sucre	Bolívar, Cesar, Sucre
Noroccidente	Antioquia	Antioquia, Archipiélago de San Andrés, Chocó	Antioquia, Caldas, Córdoba
Eje Cafetero	Caldas, Quindío, Risaralda, Tolima	Caldas, Quindío, Risaralda, Tolima, Huila, Caquetá	
Oriente	Boyacá, Norte de Santander, Santander	Arauca, Boyacá, Norte de Santander, Santander	Santander
Sur-Occidente	Cauca, Nariño, Valle del Cauca	Cauca, Nariño, Valle del Cauca	

La información organizacional descrita anteriormente fue consultada en el informe consolidado por Tigo, en el documento Informe de Gestión y Sostenibilidad UNE EPM Telecomunicaciones S.A. (TIGO) 2019 [58].

Teniendo en cuenta los servicios que Tigo ofrece a nivel nacional, sus zonas de cobertura, y su organización comercial, es propicio en este punto, hacer énfasis en el servicio Amazon Prime Video en provecho de los objetivos del presente proyecto. En 2019, la Comisión de Regulación de Comunicaciones (CRC), evidenció el resultado de un estudio acerca del consumo de los servicios en línea audiovisuales en Colombia, afirmando que, el 42% de los colombianos tiene una suscripción a una plataforma de video por streaming, y que el 16% de los hogares locales pagan por usar plataformas con contenidos audiovisuales [59].

Considerando este panorama, en el que los contenidos por streaming se han transformado en un servicio complementario de la televisión lineal; los operadores están buscando alianzas con los gigantes de la industria de video bajo demanda para aventajarse en el mercado, brindando lo que sus usuarios desean.

Así, en Enero de 2019, Tigo anunció su alianza con Amazon para ofrecer el servicio de APV, a sus usuarios HOME que tengan el servicio de internet hogar con velocidad superior o igual a 10 [60] y a sus usuarios MOBILE en los planes de 55 Mil, 60 Mil, 75 Mil y 100 Mil [61]. De esta manera, los usuarios pueden alquilar o comprar programas de televisión, películas, una selección de contenido original de Amazon Studios y adquisiciones con licencias incluidas en la suscripción Prime de Amazon, con el fin de disfrutar de su contenido en cualquier dispositivo electrónico audiovisual.

CAPÍTULO II - IMPLEMENTACIÓN

Teniendo en cuenta lo evidenciado hasta este punto, y de conformidad con el análisis bibliográfico realizado en este documento, fue posible identificar una serie de etapas que promueven la ejecución de BI en una organización. Estas etapas fueron el resultado del estudio de empresas como Oracle [12] e investigaciones como "***A data management and analytic model for Business Intelligence applications***" [62]. El presente proyecto adopta estas etapas con el fin de automatizar los reportes del servicio Amazon Prime Video (APV) en el área Vp Digital de Tigo. Las etapas de ejecución de BI seleccionadas se resumen a continuación:

1. Definición de requerimientos
2. Adquisición de datos.
3. Entendimiento de los datos y preparación.
4. Procesamiento de los datos.
5. Publicación de la información.
6. Toma de decisiones.

Para dar cumplimiento a la implementación de estas etapas en el área Vp Digital, fue necesario realizar un diagnóstico de: los objetivos, las actividades, la tecnología con la que contaba la empresa, y los procesos de BI que se adoptaban, con relación al servicio APV, que como se ha manifestado, es el conjunto de datos base del presente proyecto. Este diagnóstico fue llevado a cabo las dos primeras semanas en la empresa, y permitió proponer actividades claves, para desarrollar este trabajo acorde a un cronograma que dé cumplimiento a los objetivos planteados.

Así, teniendo en cuenta la asesoría del experto en la empresa para conocer el contexto funcional de BI en el área VP Digital, respecto a las etapas previamente descritas y el servicio Amazon Prime Video, fue posible identificar lo siguiente:

1 Etapa - Definición de requerimientos:

El área VP Digital ya había definido los objetivos de análisis sobre el servicio APV y los requerimientos de información que deben suscitarse para analizar su comportamiento en el mercado. En este punto, se establecían indicadores, para validar el estado de cumplimiento de los objetivos, y así garantizar un entendimiento óptimo de los datos del servicio, en concordancia con la necesidad de información definida.

2 Etapa - Adquisición de datos:

En esta etapa, la empresa ya había implementado un sistema automático para el suministro de los datos del servicio APV a su propio repositorio. Dicho sistema, extrae los datos de los diferentes canales digitales que interactúan con el cliente en una transacción, y los almacena

en un formato estructurado en Bigquery, lo cual, no sólo promueve la provisión de datos en tiempo real, sino también, el entendimiento de estos, gracias al orden que adquieren.

3 Etapa - Entendimiento de los datos y preparación:

Involucra el análisis cognitivo del conjunto de datos del servicio. Los expertos del área comprenden e intuyen el valor comercial que hay en ellos y la forma en que podría extraerse. De esta manera, una vez se han entendido las entidades, relaciones, formatos, marcas de tiempo, y otros detalles más del conjunto de datos almacenados; se limpian (pre-procesan), se hacen confiables, y pueden procesarse más adelante para producir el conocimiento que se requiere.

4 Etapa - Procesamiento de los datos:

Con el fin de generar reportes que resuman los detalles sustanciales del negocio, los datos del servicio APV, eran filtrados de acuerdo a la necesidad de información, ordenados en secuencias temporales, analizados acorde al conocimiento descriptivo del momento, clasificados en grupos para discernir información en conjunto, y resumidos conforme a los indicadores estratégicos. Lo anterior, permitía producir resultados concretos con relación al comportamiento presente y futuro del servicio.

5 Etapa - Publicación de la información:

Los expertos del área, construían reportes en hojas de cálculo de Excel, para mostrar los resultados obtenidos a partir del procesamiento de los datos, incluyendo las métricas y dimensiones con relación a los objetivos del negocio. Los reportes del servicio APV, no eran colaborativos en tiempo real, pero se mantenía la actualización mensual entre los miembros del equipo que lo necesitaban.

6 Etapa - Toma de decisiones:

Con base al análisis de la información, esta es la etapa final de Business Intelligence, promueve la acción de decidir con mayor precisión, no en base a intuiciones, sino con información. En el área VP Digital, algunos expertos soportan esta etapa, en colaboración con personas de otras áreas, pues involucra accionar planes estratégicos comerciales en favor del servicio, teniendo en cuenta una amplia visión del negocio.

Lo expuesto aquí, se simplifica en la ilustración 2[2]:

[2] Las ilustraciones presentadas en adelante son de fuente propia del presente proyecto, producto de la experiencia en la empresa, en complemento con la investigación, los resultados obtenidos en las herramientas de analítica de datos y software de BI. Se diseñaron intencionalmente en inglés para dar mayor visibilidad al trabajo y considerando que muchos términos no tienen traducción exacta al español.

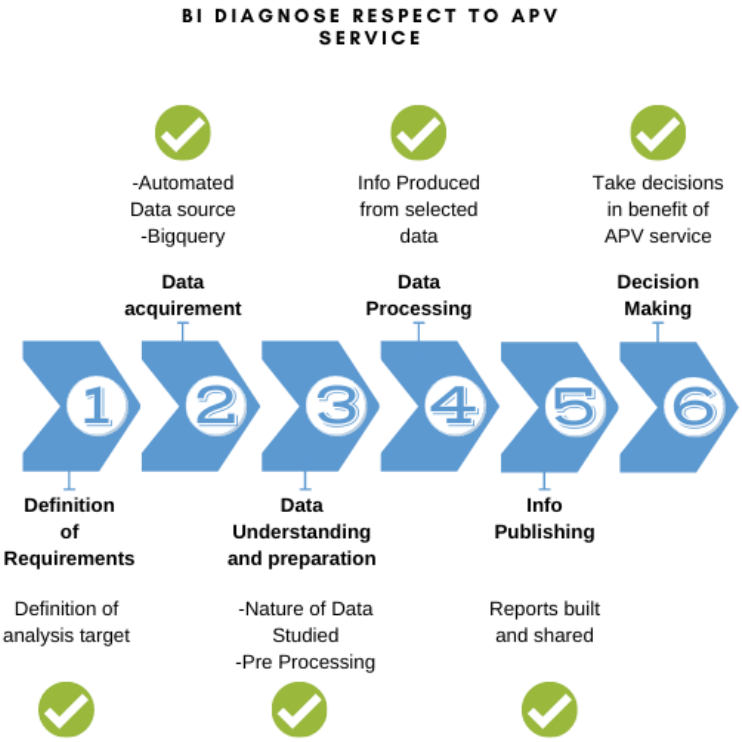


Ilustración 2. Diagnóstico de BI con relación al servicio APV en el área VP Digital. Fuente: el presente trabajo, 2021.

Lo anterior, es muestra de que el área VP Digital ha incluido Business Intelligence en el análisis de los datos para producir información de alto valor en el negocio con respecto al servicio APV. No obstante, es preciso declarar que hay etapas como: el procesamiento de los datos y la publicación de la información, (etapas 4 y 5 respectivamente), que pueden optimizarse a través de la automatización, para lograr el conocimiento sustancial que se necesita con mayor amplitud e incluso en tiempo real.

En consecuencia, se diseña la ilustración 3, que evidencia la integración de Business Intelligence en la implementación del presente proyecto, con el propósito de tener una estructura definida, para planificar su ejecución desde el principio hasta el fin. Además, el esquema pretende aclarar cómo esta propuesta, se vincula a las actividades ya realizadas en el área, y la forma en que complementará la ejecución de las etapas de BI descritas en el diagnóstico.

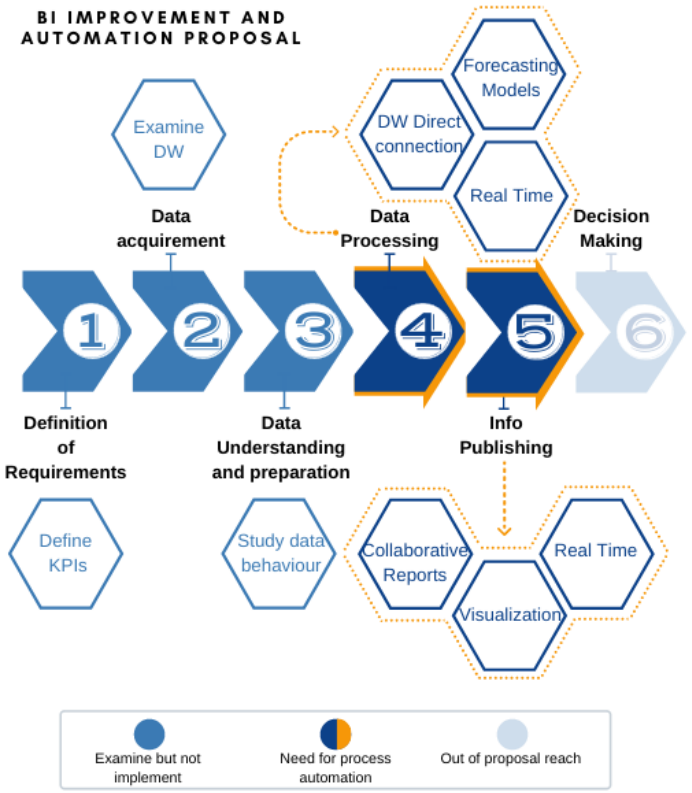


Ilustración 3. Mejoramiento etapas de BI mediante la automatización. Fuente: el presente trabajo, 2021.

Como se visualiza en el esquema, y de acuerdo a los alcances del proyecto, se plantean actividades respecto a las etapas 4 y 5 principalmente, con el fin de optimizar las funcionalidades detalladas a continuación, mediante la automatización de sus procesos.

Etapa 4:

- Conexión directa entre el Data Warehouse y el software de BI para el procesamiento y la visualización de reportes, evitando la carga/descarga manual de los datos.
- Procesamiento en tiempo real.
- Modelamiento predictivo.

Etapa 5:

- Reportes colaborativos.
- Información asociada a gráficos.
- Visualización de la información en tiempo real.

De igual forma, fue necesario proponer actividades para las tres primeras etapas, pues, aunque no requieren de optimización, estudiarlas es indispensable, ya que su entendimiento es el fundamento para las etapas 4 y 5 ya especificadas.

Argumentando esto, es importante resaltar a continuación, las actividades necesarias para la implementación del proyecto, las cuales, no sólo son afines a los objetivos intrínsecos de la

propuesta, sino que se adaptan también, a las fases susceptibles de mejoramiento de Business Intelligence en el área VP Digital con respecto al servicio APV.

2.1 Plan de Trabajo

Tabla 4. Plan de Trabajo. Fuente: el presente trabajo, 2021.

Objetivo	Actividades	Fecha Inicio/ Fin	Indicador de Éxito	Herramientas
1	Establecer las etapas de BI conforme a los procesos del área	Semana 1-2	Evaluación de las técnicas de BI propuestas conforme a las fases teóricas de BI descritas en la revisión de literatura	Asesoramiento del lider del área
	Estudiar los objetivos del área	Semana 1-2	Aprobación por parte del lider del área respecto a los objetivos señalados	
	Identificar los procesos susceptibles de automatización	Semana 3-4	Retroalimentación positiva del asesor del área, conforme a los procesos sugeridos	BigQuery, Excel
	Analizar los datos dispuestos en la empresa	Semana 3-4	Exposición de las fuentes y la naturaleza de los datos útiles para la implementación	
2	Estudiar la tecnología, los recursos y el software de BI dispuesto en la empresa	Semana 5-7	o Ejemplos sencillos acorde al análisis de los datos o Certificaciones	Python, ScikitLearn, Data Studio, Tableau
	Especificar los KPIs propicios acorde a la necesidad de información	Semana 8	Concordancia de los objetivos del negocio con relación a las métricas clave identificadas	Asesoramiento del lider del área
	Construir la lógica de la automatización requerida	Semana 8	Aval del asesor conforme a la explicación de la lógica sugerida	BigQuery, DataStudio, Python, Excel, Asesoramiento del lider del área
	Implementar la lógica diseñada para cada proceso susceptible de automatización	Semana 9-16	o Información producida con relación a los KPIs del negocio o Habilitación de la información que era producida manualmente, de forma autónoma	Bigquery, DataStudio, Python, ScikitLearn, Keras, Pandas, Facebook Prophet
3	Construir Mockups de la visualización de la información requerida	Semana 17	Aval del asesor conforme al diseño sugerido	Google DataStudio, Adobe XD
	Diseñar Dashboards en el software de BI para exponer la información ya procesada	Semana 18-19	oVinculación total de los datos procesados en el Dashboard con relación al diseño previo y la visualización de los KPIs o Visualización de los datos en tiempo real	Google DataStudio
4	Definir criterios para la comparación de los procesos antes y despues de la automatización	Semana 3	Evaluación de los procesos automatizados respecto a los criterios establecidos	Asesoramiento del lider del Área
	Probar la implementación con un experto de la empresa	Semana 20	Retrolimentación del profesional del área	Bigquery, DataStudio, Python, Excel, Asesoramiento del lider del área
	Realizar posibles ajustes y prueba final	Semana 21	Aprobación del proyecto en la empresa	

Siguiendo en este razonamiento, se expone el desarrollo de este capítulo en el orden establecido por las actividades planteadas, con el fin de brindar un entendimiento integral del contenido respecto a la ejecución temporal de los procesos en la construcción, el cumplimiento de los objetivos y la adopción de BI en todo el proyecto.

Expuesto esto, se prosigue con la explicación de la implementación de la propuesta, con respecto a la ejecución de cada actividad.

2.2 Estudiar los objetivos del área VP Digital

Teniendo en cuenta la guía del asesor en la empresa con respecto a las funciones que deben llevarse a cabo en el área VP Digital, y al conocimiento adquirido como aprendiz en las primeras semanas en el equipo. Se identificó que, el trabajo del talento humano en el área, está orientado a cumplir los siguientes objetivos:

2.2.1 Objetivos del Área VP Digital

- Generar la visibilidad de todos los canales digitales dentro de la compañía mediante el desarrollo de Dashboards, Insights de KPIs, Funnels de conversión, y otras herramientas, que garanticen el suministro de información comercial para la toma de decisiones.

- Automatizar fuentes de datos para la generación de informes diarios, que faciliten el rastreo y monitoreo de métricas clave para asegurar que los hitos se cumplan dentro del tiempo y costo estimado.

- Generar bases de audiencias y segmentos de usuarios para las diferentes áreas B2C Mobile, B2C Home y B2B.

- Diseñar e implementar herramientas automáticas de gestión en línea de la información mediante Tableau, Google Data Studio, Mixpanel, Excel, con el fin de incrementar la productividad de la fuerza comercial y la rentabilidad de los canales de venta a nivel de territorio y tipo de producto.

Con base a la comprensión de estos objetivos, fue de provecho entender estas actividades y metas, diseñando el esquema de la ilustración 4, con el fin de ampliar la visibilidad del área VP Digital y poder desarrollar este proyecto en consonancia con las funciones que aquí se realizan.

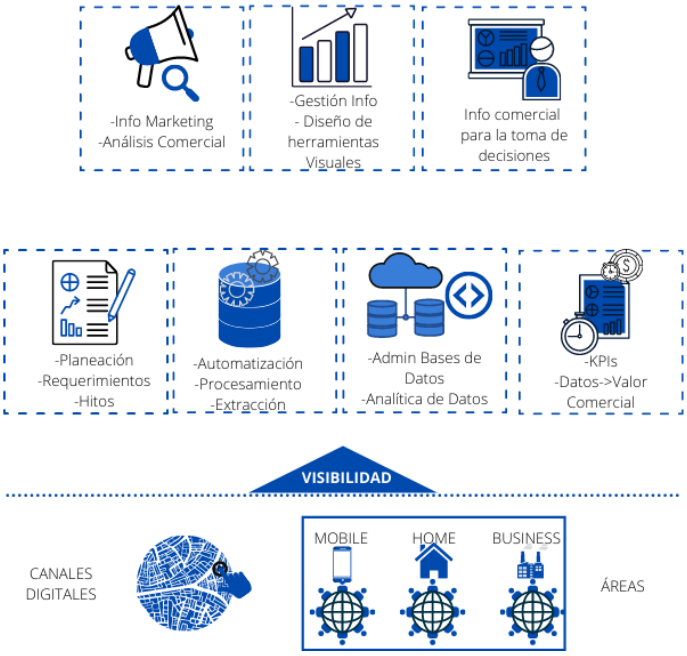


Ilustración 4. Entendimiento de los objetivos y funciones del área VP Digital. Fuente: el presente trabajo, 2021.

2.2.2 Requerimiento

Teniendo presente este contexto, el requerimiento del área al respecto, fue la automatización de procesos asociados a la generación de reportes y dashboards para la producción de información precisa y veraz acorde a uno de los servicios de la empresa, para obtener un rendimiento autónomo de algunas operaciones repetitivas en el procesamiento de los datos, y evitar no sólo una lectura incorrecta o incompleta de estos, sino también el aumento de la efectividad en la planeación comercial del servicio.

2.3 Identificar los procesos susceptibles de automatización

Una vez se han descrito los objetivos y el requerimiento del área Vp digital, que es el entorno, donde se desarrolla y se implementa el presente proyecto, es preciso manifestar que, acorde a las necesidades del momento de ejecución de la práctica profesional, las indicaciones en el área se establecieron respecto a la gestión de los datos del servicio Amazon Prime Video (APV), con el fin de proveer reportes visuales automatizados, que sirvan de base informativa para la toma de decisiones en la planificación comercial del servicio. Por lo que, en consonancia con los objetivos del área, es importante identificar y caracterizar los procesos susceptibles de automatización en este contexto.

Para ello, fue necesario tener en cuenta 4 criterios claves, los cuales fueron analizados con relación al estudio [63] y adecuados al contexto de este trabajo para asegurar su elección con mayor precisión, como se muestra a continuación.

Criterios base:

1. *Tipo de Ejecución*: representa el modo, manual o automático en que se desarrolla el proceso. Incluye el factor de repetición que exige la operación de dicho proceso y la frecuencia de trabajo que demanda del talento humano.

2. *Tiempo:* se relaciona con la duración de ejecución del proceso. De esta manera, entre menos tiempo se requiera, se beneficia la productividad del trabajo en el área.

3. *Veracidad de la información:* se refiere al grado de certeza de la información, libre de incoherencias. Debe asegurarse una correlación entre los datos del servicio APV, y la información que se entrega. Este criterio está ligado a la propensión de errores, que pueden cometerse al ejecutar el proceso.

4. *Amplitud de la información*: es el grado de detalle de la información que puede producirse con base al conjunto de datos. Entre más información clave se procese, más detalles pueden compartirse para orientar la toma de decisiones, y por lo tanto, mayor amplitud en el conocimiento del servicio APV.

5. *Visualización:* hace alusión a la forma en que se presenta la información. Conviene un diseño intuitivo, una selección de gráficos acorde a la información y la posibilidad de filtrar consultas fácilmente.

En consecuencia, fue posible con base a la experiencia guiada en la empresa, identificar 2 procesos susceptibles de automatización:

- Modelamiento predictivo
- Procesamiento de datos

Los cuáles serán descritos brevemente a continuación, acorde al modo en que se llevaban a cabo en ese momento en la empresa. Además, se caracterizan dichos procesos teniendo en cuenta sus funcionalidades y los criterios previamente redactados.

2.3.1 Modelamiento Predictivo

Se refiere a la posibilidad de utilizar datos, algoritmos estadísticos y técnicas de aprendizaje automático para determinar resultados futuros con base en datos históricos. El objetivo es ir más allá de la comprensión de lo que ha sucedido para proporcionar la mejor evaluación de lo que sucederá en el futuro.

El estudio de artículos como "*Desarrollo de un modelo basado en Machine Learning para la predicción de la demanda de habitaciones y ocupación en el sector hotelero*"[64] y "*Opportunities for the Use of Business Data Analysis Technologies"* [12], apuntan a caracterizar el análisis predictivo, de la siguiente forma:

- *Converge a valores lógicos:* la predicción está orientada a la tendencia del comportamiento de los datos. No se afecta por la presencia de valores atípicos, los tiene en cuenta y predice en base a la tendencia de la mayoría de los valores lógicos del conjunto.

- *Ajuste lineal y no lineal:* la predicción se expresa mediante la representación promedio de los valores como una aproximación para determinar futuros valores; en complemento con el ajuste no lineal, puede modelar curvas adecuándose al comportamiento descrito por los datos, para predecir las posibles fluctuaciones más precisamente.

- *Versátil con grandes volúmenes de datos:* la predicción se adapta a conjuntos de datos grandes, y es capaz de escalar el análisis en respuesta a ellos para garantizar la precisión requerida.

- *Asimila los datos:* los valores pronosticados no sólo son el resultado de operaciones matemáticas básicas, sino que los métodos de predicción, pueden analizar estadísticamente y entender el comportamiento de los datos. Se asegura mayor fiabilidad al resultado esperado, cuando se comprende la correlación de los datos entre sí.

Específicamente, el desarrollo de pronósticos de los datos del servicio APV en el área VP Digital, se llevaba a cabo semi-automáticamente. Con la ayuda de Microsoft Excel, se importan los datos históricos, se configuran algunos parámetros (porcentajes), para multiplicar los registros previos, y brindar un aproximado de los valores posibles en un futuro. Algunos valores son ajustados manualmente, con base a los resultados y la experiencia de análisis del comportamiento del servicio en el mercado, ya sea por la temporada o simplemente por la actividad previa de las personas al activarse.

La ejecución del proceso descrito, se resume en la ilustración 5:

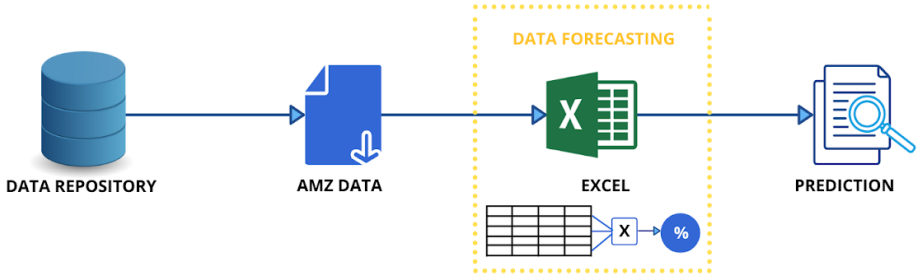


Ilustración 5. Proceso 1 susceptible de automatización. Fuente: el presente trabajo, 2021.

Teniendo presente esto, la caracterización del proceso tal y como se llevaba a cabo en la empresa, de conformidad a sus funcionalidades y los 5 criterios base, puede resumirse en la tabla 5.

Tabla 5. Criterios de comparación para el proceso 1 susceptible. Fuente: el presente trabajo, 2021

Funcionalidad	*Converge a valores lógicos*	✓
	Ajuste lineal y no lineal	Solamente ajuste lineal, representado como el producto de un valor histórico por un parámetro porcentual.
	Versátil con grandes volúmenes de datos	No, demanda más tiempo calcular el pronóstico cuando el volumen de datos incrementa.
	Asimila los datos	No, simplemente se afecta el valor pronosticado con un factor (porcentaje), pero no se comprende automáticamente la correlación del comportamiento de los datos.
Criterios base	*Tipo de Ejecución*	Manual + cálculos matemáticos de Excel, rectificación de valores manualmente.
	Tiempo	1-2 horas.
	Validación de la información	Si, se verifica la fiabilidad de la información, sin embargo, puede ser propenso a errores al ejecutarse manualmente.
	Amplitud de la información	Se focaliza en las métricas clave del negocio y brinda un rango de hasta 10 valores de predicción. Lo cual, es adecuado para la aproximación requerida en el negocio.
	Visualización	Sólo hay resultados numéricos, no hay ilustraciones, ni gráficas, ni la posibilidad de filtrar consultas.

2.3.2 Procesamiento de datos

En las organizaciones, las tareas de analítica de datos generalmente consisten en extraer información oculta y representarla en una forma adecuada para soportar la toma de decisiones. Por lo general, los procesos de este tipo pasan por varias fases en las que la información se extrae de los datos originales, se convierte, se transfiere, se acumula y, en última instancia, se transforma para brindar una fácil interpretación [65]. Simplificadamente,

el procesamiento consiste en recopilar y manipular elementos de datos para producir información significativa.

Así, este proceso, puede caracterizarse funcionalmente de la siguiente manera [65]:

- *Valida:* garantiza que los datos proporcionados sean correctos y relevantes.
- *Agrega:* combina varios datos.
- *Ordena*: organiza los elementos en alguna secuencia y/o en diferentes conjuntos con un sentido lógico.
- *Clasifica:* separación de datos en varias categorías.
- *Resume:* simplificación de los datos a su valor fundamental.
- *Analiza:* involucra la recopilación, la organización, estudio, interpretación y presentación de datos.

Específicamente, el procesamiento de los datos del servicio APV en el área VP Digital, se llevaba a cabo manualmente, en conjunto con algunos cálculos automáticos de Microsoft Excel. El proceso se realizaba de manera local, descargando el conjunto de datos del repositorio de la empresa, e importándolos en la herramienta Excel, donde en conjunto con cálculos proveídos por el entorno y la organización de estos matricialmente, se obtenía el registro contable de la información requerida. Para este proceso, no se tiene un conector vinculado para importar los datos a Excel, por lo que, la integración de los datos demanda su actualización manual cuando se requiera su procesamiento. Por otra parte, el informe producido, es resumido y preciso, pero sólo contiene etiquetas con valores, y colores que resaltan las métricas claves, no presenta gráficos ni la posibilidad de realizar consultas para interactuar con los datos procesados.

La ejecución del proceso descrito, se resume en la ilustración 6:

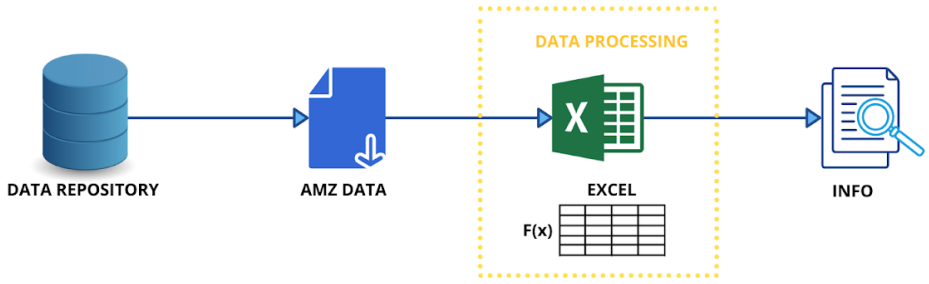


Ilustración 6. Proceso 2 susceptible de automatización. Fuente: el presente trabajo, 2021.

Teniendo presente esto, la caracterización del proceso tal y como se llevaba a cabo en la empresa, de conformidad a sus funcionalidades y los 5 criterios base, puede resumirse en la tabla 6.

Tabla 6. Criterios de comparación para el proceso 1 susceptible. Fuente: el presente trabajo, 2021

Funcionalidad	*Valida*	✓
	Agrega	✓
	Ordena	✓
	Analiza	✓
	Clasifica	✓
	Resume	✓
Criterios base	*Tipo de Ejecución*	Manual + cálculos matemáticos de Excel
	Tiempo	2-4 horas
	Validación de la información	Si, se verifica la fiabilidad de la información, sin embargo, puede ser propenso a errores al ejecutarse manualmente.
	Amplitud de la información	Se focaliza en las métricas clave simplemente, no brinda una visibilidad amplia de la información que podría ser potencial para el negocio.
	Visualización	Sólo hay resultados numéricos, no hay ilustraciones, ni gráficas, ni la posibilidad de filtrar consultas

2.4 Analizar los datos dispuestos en la empresa

Por asuntos de confidencialidad, no se puede documentar a detalle las entidades ni los atributos específicos del conjunto de datos del servicio APV tal y como se almacenan en el repositorio de la compañía. Sin embargo, es posible, explicar la operatividad técnica que se desarrolla con dichos datos.

Inicialmente, cabe exponer que, se cuenta con un conjunto de datos almacenado directamente en el Data Warehouse de Google, BigQuery, con el fin de tener el registro contable de las activaciones y cancelaciones del servicio. Se identificaron algunos elementos relevantes para la automatización del procesamiento y la predicción de los datos; sin embargo, debido al manejo privado de la información que exige la empresa, sólo puede especificarse aquellos atributos que son comunes en toda base de datos de este tipo, es decir, aquellas que se

relacionan con el modo de comercialización por suscripción. Lo cual, no supone ningún problema para el entendimiento de la propuesta, pues estos, que se resaltan a continuación, soportan en gran medida la implementación del trabajo, como se observará más adelante.

Tabla 7. Atributos clave del servicio APV

Nombre Atributo	**Referencia**
Fecha	Activación y cancelación
Estado	Activo o cancelado
Tipo de usuario	Hogar o móvil

Con base a estos atributos, es posible contabilizar el número de activaciones y cancelaciones diarias, mediante la extracción de los datos, ejecutando sentencias SQL en Bigquery. Esto, con el fin de tener un registro temporal de los datos, para facilitar la automatización requerida.

En adición a ello, es importante resaltar que el conjunto de datos del servicio APV ha sido extraído, transformado y almacenado en Bigquery desde los canales digitales, que sirven de interfaz para que los usuarios puedan activarse en el servicio. Estos datos, ya han sido pre procesados, permitiendo disponer de datos limpios y confiables de primera mano. Se caracterizan por tener atributos cualitativos y cuantitativos, en un formato estructurado, lo cual facilita su entendimiento para llevar a cabo su análisis respectivo.

Debido a las continuas transacciones que se hacen, los datos se actualizan en el Data Warehouse diariamente, con el fin de garantizar su disponibilidad en el instante que se requieren, y de esta manera, proveer el fundamento a partir del cual se extraerá valor informativo para la toma de decisiones.

Conviene exponer que, los datos describen una sucesión de registros medidos en determinados instantes y ordenados cronológicamente, debido a que poseen marcas temporales, asociadas al momento en que se procesan las solicitudes del cliente. Esto permite observar que su comportamiento se adapta a la naturaleza de una serie temporal, permitiendo en adelante, analizar las variables descritas en la **tabla 7** con herramientas estadísticas acordes a esta condición.

Observando esto, se procede a especificar algunas características de referencia del conjunto de datos, respecto al momento de ejecución del proyecto:

- # de registros: 1.299.397
- # de variables: 14
- Frecuencia de actualización de registros en el DW: diario
- Promedio cantidad de registros/dia: 1.539,931
- Porcentaje de campos faltantes/registro: 0.00021%
- Tamaño completo del conjunto de datos:473 MB

- Tamaño del conjunto de activaciones resumido diariamente del servicio para pronósticos: 18KB
- Formato de variables: int, double, float, string, date, boolean
- Formato marca temporal: aaaa-mm-dd
- Auto numeración: identificador único del registro (de activación o desactivación).
- Identificador único para cada usuario en el servicio: número de celular de los usuarios
- Mínimo de activaciones diaria: 351
- Máximo de activaciones diaria: 3229

En complemento a lo anterior, con respecto al entendimiento operativo de los datos y la literatura investigada [66], se realiza el siguiente modelo, en el cual, se identifican las múltiples operaciones, anteriormente descritas, en asociación con la arquitectura Multitier. Esto, con el propósito de hacer visible la arquitectura de los procesos involucrados en este contexto en el área, y marcar la relevancia que la capa de datos tiene, para proveer y sustentar los recursos sobre las otras capas.

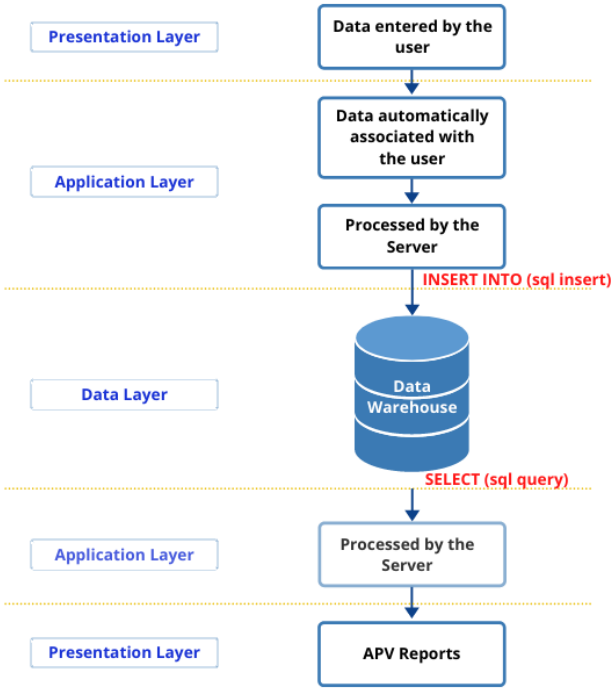


Ilustración 7. Arquitectura Multitier del servicio APV. Fuente: el presente trabajo, 2021.

Así las cosas, se tiene:

- Capa de presentación: es la interfaz que el usuario observa directamente y le permite interactuar con el sistema, le comunica la información del servicio APV y captura la información que el usuario suministra para la transacción. La empresa específicamente, lo hace por medio de sus plataformas digitales, web o móvil, donde se provee de interfaces intuitivas que facilitan su ejecución.

- Capa de negocio: es donde residen los programas que se ejecutan, se reciben las peticiones del usuario y se envían las respuestas correspondientes al proceso. Es aquí donde se establece la lógica del negocio y todas las reglas que deben cumplirse acorde al servicio APV. Esta capa se comunica con la capa de presentación, para recibir las solicitudes y presentar los resultados, y con la capa de datos, para solicitar al gestor de base de datos almacenar o extraer datos de él.

- Capa de datos: es donde residen los datos. La empresa cuenta con más de un gestor de bases de datos que se encargan del almacenamiento, la administración de solicitudes de carga y extracción de información desde la capa de negocio.

Tal como se observa, estos son los niveles que representan los procesos de los datos APV, desde que el usuario interactúa con las plataformas digitales, hasta que la información es procesada, y visualizada para el respectivo análisis ejecutivo.

2.5 Tecnología, recursos y software de BI

Ahora bien, ya en custodia del repositorio de la empresa y la naturaleza del conjunto de datos, se requiere establecer las herramientas y software apropiados para la ejecución de la propuesta de automatización.

En este sentido, las tecnologías, serán descritas no en cuanto a lo que son, puesto que su definición, ya fue documentada en el marco conceptual, sino con relación al aporte funcional que tuvieron para la implementación del trabajo.

2.5.1 Bigquery

Con ello en mente, principalmente se describe la adopción de Bigquery, como el almacén de datos del servicio APV, cabe aclarar que su servicio ya era contratado con Google por la compañía, y no es una decisión particular, sin embargo, dada la amplia funcionalidad que este Data Warehouse proporciona para el tratamiento de datos, es altamente recomendable para organizaciones que requieran administrar grandes volúmenes de datos.

Gracias al lenguaje SQL estándar usado en Bigquery, y la conservación del paradigma de trabajo con tablas, campos y registros, fue posible la construcción de consultas personalizadas para extraer los datos de interés que se enfocaban a los objetivos del negocio.

Para ejemplificar una apertura de su utilidad, se expone 1 consulta personalizada que selecciona registros a través de condiciones, los contabiliza, los agrupa, y los ordena en un formato específico; esto, con el propósito de interactuar con los datos almacenados e interactuar dinámicamente con ellos, de acuerdo a la necesidad de información. Más adelante, en el **punto 2.8** se amplía su aplicación con mayor detalle.

```
SELECT COUNT (distinct(a.[redacted])),Date(a.activation_date )
as date FROM `[redacted]` a
where a.cancelation_date is null
and id home is null
GROUP BY date
ORDER BY date asc
```

Ilustración 8. Consulta SQL personalizada para extraer datos del servicio. Fuente: el presente trabajo, 2021.

2.5.2 Python

Conociendo el propósito de este proyecto, enfocado a modelar, analizar, entender, visualizar y extraer conocimiento a partir de los datos; Python, es un lenguaje óptimo para esta labor, ya que proporciona todas las herramientas necesarias para llevar a cabo estos procesos, en complemento con la distribución de Anaconda, que incluye una variedad de librerías para la ciencia de datos y Jupyter como el entorno de desarrollo interactivo para implementar la programación.

Con base en los datos del servicio APV dispuestos en el Data Warehouse, se procede a visualizar algunos elementos, para comprender de mejor manera cuál es su naturaleza, y cómo se podría automatizar el procesamiento de los datos y la predicción de su comportamiento.

Inicialmente, se identificaron los conjuntos de datos útiles que satisfacen la necesidad de información del área, se exploraron en Bigquery y se los exportó en formato .csv, para trabajar con ellos posteriormente. Así, fue de provecho examinar algunas características de los datos, para comprender mejor, no sólo su estado, sino también, las herramientas y bibliotecas que deberían integrarse para extraer el valor empresarial necesario para la comercialización del servicio.

En este punto se desarrollan los siguientes ejemplos, los cuales, vale aclarar son una introducción para identificar otras herramientas que deberían usarse para la automatización respectiva. La aplicación de estos ejemplos se encaminó a 3 aspectos:

2.5.2.1 Exploración de los datos

Acorde a la necesidad de información, se seleccionó el conjunto de datos relacionado a las activaciones del servicio APV. Como se mencionó previamente, fue exportado directamente desde Bigquery en formato.csv, el cual, representa a los archivos de texto que utilizan comas como delimitadores de campo.

La importación de los datos, se hace mediante la herramienta Pandas, que, trabaja esencialmente con 2 elementos: series y dataframes, siendo el primero, la representación de una matriz etiquetada unidimensional capaz de contener cualquier tipo de datos (enteros,

cadenas, decimales, objetos Python, etc.) y el segundo, la representación de un conjunto de los objetos Series.

De esta manera, es posible trabajar con los datos, como si fuesen tablas con la característica implícita de que los objetos están asociados a etiquetas, las cuales, permiten su indexación con el propósito de identificar elementos y la obtención intuitiva de subconjuntos en el dataset.

Así, habiendo importado los datos, la primera exploración que se hace es, observar su registro, indexando los elementos con la fecha de procesamiento:

fecha	activaciones
2019-02-01	[redacted]
2019-02-02	[redacted]
2019-02-03	[redacted]
2019-02-04	[redacted]
2019-02-05	[redacted]

Ilustración 9. Encabezado. Fuente: el presente trabajo, 2021.

En adelante, se realizan múltiples funciones como se puede observar en el script (**Scripts/ExploraciónDeDatos.ipynb**) anexado para obtener mayor profundidad en la documentación.

Aquí, se incluyó la visualización de los datos y el re muestreo respecto al tiempo, para visualizar los elementos conforme a una frecuencia semanal, anual, o incluso mensual, como lo muestra la ilustración 19[3], por conveniencia para el entendimiento de los reportes realizados en el área.

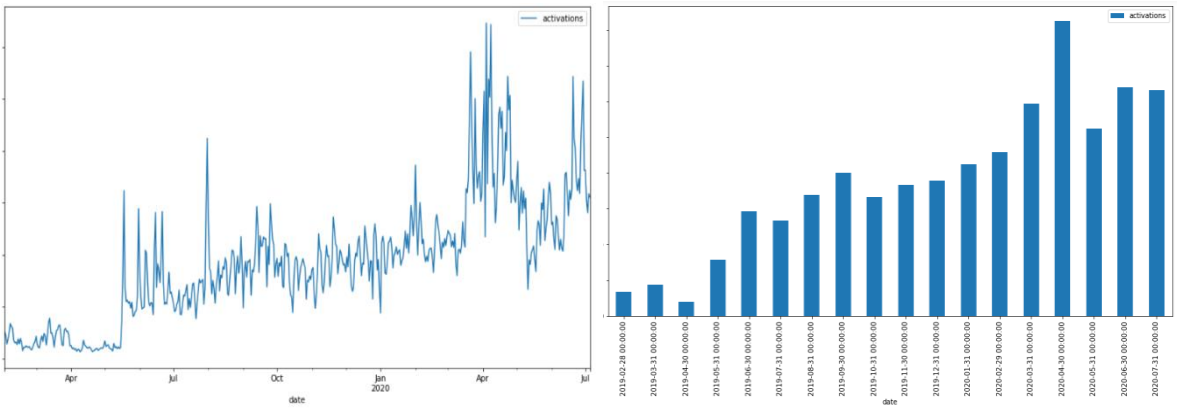


Ilustración 10. Representación gráfica de las activaciones del servicio. Fuente: el presente trabajo, 2021.

[3] Por asuntos de confidencialidad, las ilustraciones de este tipo, relacionadas con los datos privados de la empresa, fueron pixelados o apartados de las gráficas respectivas.

De la misma forma, se exploró el conjunto de datos, a través del análisis de las propiedades estadísticas como la media, el máximo, el mínimo, desviación estándar, percentiles, etc. Propiedades que resultaron relevantes para cuantificar el comportamiento de los datos en determinadas condiciones, con respecto al tiempo.

2.5.2.2 Regresiones

Con el fin de obtener una aproximación para modelar la relación entre el número de activaciones del servicio APV respecto al tiempo, fue propicio tener una introducción a los algoritmos de machine learning supervisados, donde, teniendo en consideración la simplicidad de los modelos como el criterio base de selección, fue de provecho implementar regresiones lineales y logísticas, que permitieran visualizar cómo sería el comportamiento de los datos en un futuro cercano.

La asesoría del profesional en la empresa para este caso, validó la implementación de estos algoritmos, ya que en el momento se requirió una aproximación rápida y general de 2 solicitudes de información: conocer la cantidad de activaciones del servicio APV al finalizar el año, y el número de activaciones mensuales que se tendrían que generar para alcanzar un resultado de activaciones específico.

Por consiguiente, se desarrollaron las siguientes tareas:

- *Usuarios Activos*

Para la implementación de regresiones lineales y logísticas se usó la librería Scikitlearn, debido que, contiene una cantidad de funciones implícitas que permiten tanto el modelamiento como la predicción en sí. De esta forma, conforme a los datos del servicio que contiene el número de usuarios activos en cada mes, se creó una instancia para cada uno de los modelos con las funciones correspondientes de Scikitlearn: LinearRegression() y PolynomialFeatures(degree=n), esta última con un parámetro adicional, que posibilita una configuración manual del grado del polinomio para describir el conjunto más precisamente, donde cabe mencionar, fue necesario tener precaución para no sobre ajustar el modelo.

Posteriormente se transformó la entrada de los datos, con la función fit_transform() para ajustar el modelo, dependiendo del tipo de regresión; posibilitando al final, predecir resultados de las activaciones, mediante la expansión del polinomio generado por la cantidad de espacios en el tiempo que se necesitaban, teniendo en cuenta que entre más lejano estaba el foco de la predicción, más distorsión se presentaba en los resultados.

Así, los modelos implementados dieron como resultado la siguiente aproximación gráfica, y un arreglo de predicciones para diferentes valores en el tiempo, tal y como se evidencia en el script anexado con el nombre '**Scripts/PolynomialRegressionAPV.ipynb**'

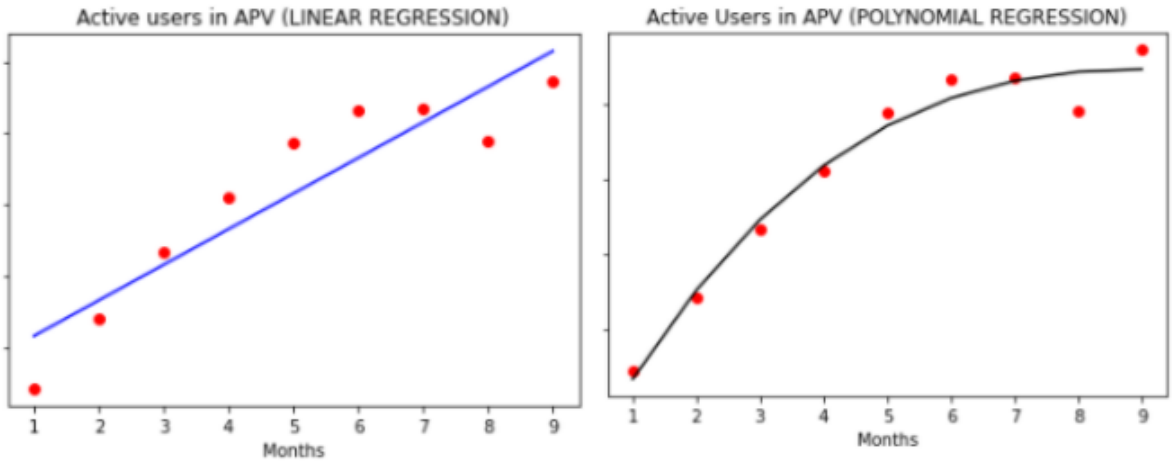


Ilustración 11. Modelamiento lineal y polinomial de la base activa de los usuarios del servicio APV. Fuente: el presente trabajo, 2021.

- *Usuarios activos para el cierre de 2020*

Con la condición de garantizar un número específico de usuarios activos en el servicio para el fin del año 2020, el objetivo era determinar cuántos usuarios debían conformar la base activa de cada mes para cumplir con la meta deseada. De esta manera, se fijó en rojo, el resultado requerido para el mes 20 (fin de año), y se ajustó una regresión polinómica con tendencia incremental.

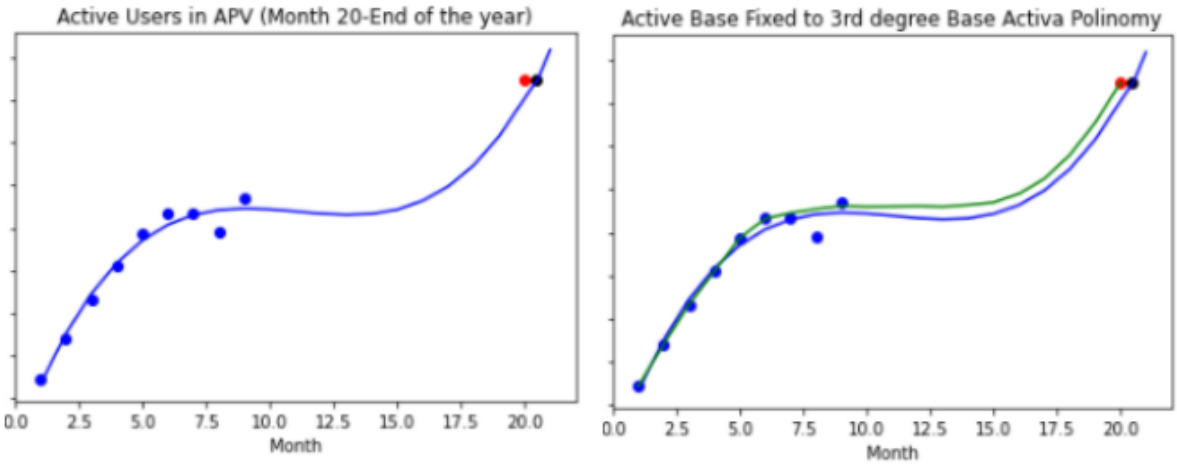


Ilustración 12. Ajuste de activaciones mensuales para un número de cierre específico. Fuente: el presente trabajo, 2021.

La aproximación de esta regresión, generó un conjunto de valores en un arreglo, que representa el número de usuarios activos requeridos, para cada uno los meses posteriores. Adicionalmente, fue posible indicar cuál sería la tasa incremental de los usuarios entre cada mes, para asegurar el cumplimiento del objetivo. El cálculo y desarrollo de este modelo se anexó en el script '**Scripts/APVUsersEnding2020.ipynb**'

2.5.2.3 Modelado de los datos

En complemento con StatsModels, fue posible realizar representaciones de los datos para obtener un acercamiento a la información que se necesitaba, mediante la definición de patrones, pruebas estadísticas y otras herramientas que evaluaban el conjunto de elementos en el tiempo.

Así, los ejemplos incluyen: análisis de tendencias, análisis de estacionalidad, y exploración de elementos residuales, mediante la aplicación de técnicas como la descomposición estacional de StatsModels, utilizando modelos aditivos o multiplicativos, dependiendo de la linealidad en la tendencia denotada por el conjunto. Un resumen de algunas componentes examinadas se puede evidenciar en la ilustración 13.

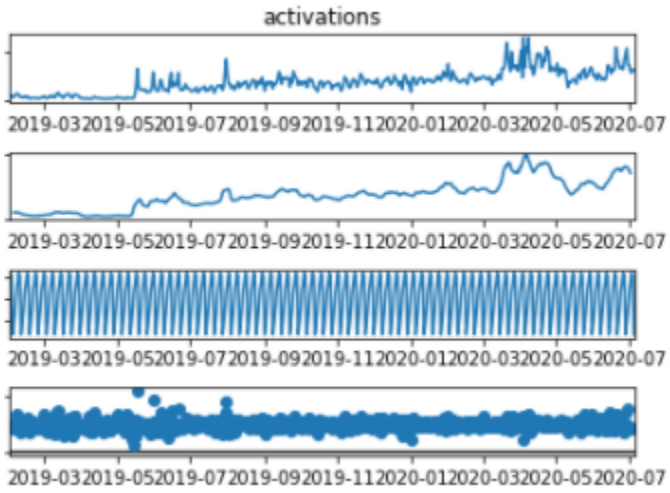


Ilustración 13. Descomposición Estacional Statsmodels. Fuente: el presente trabajo, 2021.

De igual forma, se implementaron modelos de suavizado a los datos existentes. El propósito detrás de esto, fue representar el conjunto de datos actuales, y predecir lo que sucedería próximamente. Para esto, se tomaron en cuenta 3 métodos, desarrollados en el script '**Scripts/HLWMethods.ipynb**':

- *Promedio Móvil Ponderado Exponencialmente (EWMA):* hace una representación de la media ponderada de las n activaciones del servicio APV con un factor de suavizado α, aplicando una función media a una ventana móvil ajustable. Su pronóstico fue simplemente una línea horizontal que se extendía desde el valor más reciente.

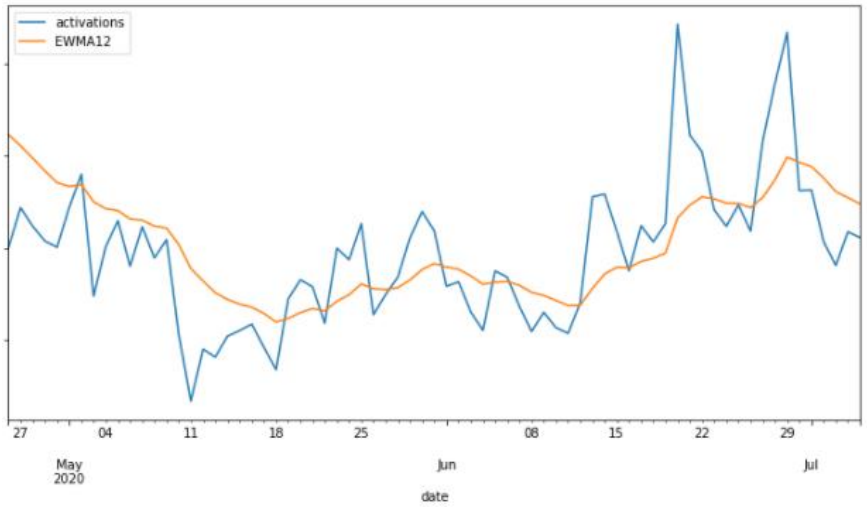


Ilustración 14. Modelo EWMA. Fuente: el presente trabajo, 2021.

De esta manera, el modelo se concentraba en colocar más peso en los valores (activaciones) que ocurrieron más recientemente. La cantidad de ponderación aplicada a estos valores recientes dependía de los parámetros reales utilizados y del número de períodos, dado un tamaño de ventana específico, que para este caso se ajustó a 12.

- *Holt´s Method (DES):* involucra un índice de tendencia en el análisis. Permite realizar un pronóstico con una pendiente asociada a la tendencia del conjunto.

Mientras que, el modelo anterior emplea solo un factor de suavizado α (alfa), el suavizado exponencial doble agrega un segundo factor de suavizado β (beta) que aborda las tendencias en los datos. Al igual que el factor alfa, los valores del factor beta, también están entre cero y uno ($0 < \beta \leq 1$).

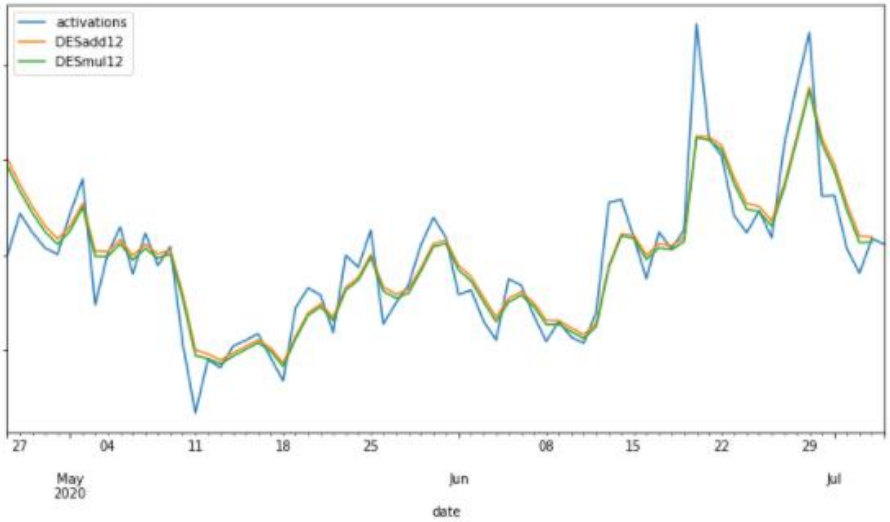


Ilustración 15. Modelo DES con ajuste aditivo y multiplicativo. Fuente: el presente trabajo, 2021.

Como se puede observar, el beneficio aquí, es que el modelo podía anticipar aumentos o disminuciones futuras, donde el modelo EWMA solo funcionaría a partir de cálculos recientes, de ahí la percepción de mayor proximidad al momento de representar el número de activaciones del servicio APV con los datos reales.

- *Holt Winters Method (TES):* este modelo agrega el análisis de la componente estacional. De esta manera, conociendo que datos como los del servicio APV, tienen fluctuaciones regulares en el futuro, se pudieron mapear de mejor forma involucrando el comportamiento de los datos con relación al momento en que se generan las variaciones.

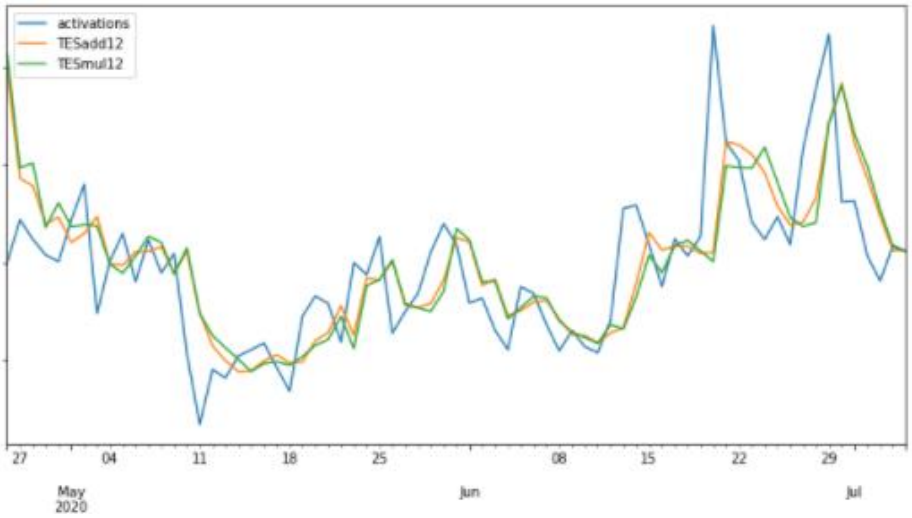


Ilustración 16. Modelo TES con ajuste aditivo y multiplicativo. Fuente: el presente trabajo, 2021.

Ahora bien, para evaluar los resultados de los tres modelos con las activaciones APV en conjunto, se comparó su ajuste en una sola vista, obteniendo el siguiente resultado:

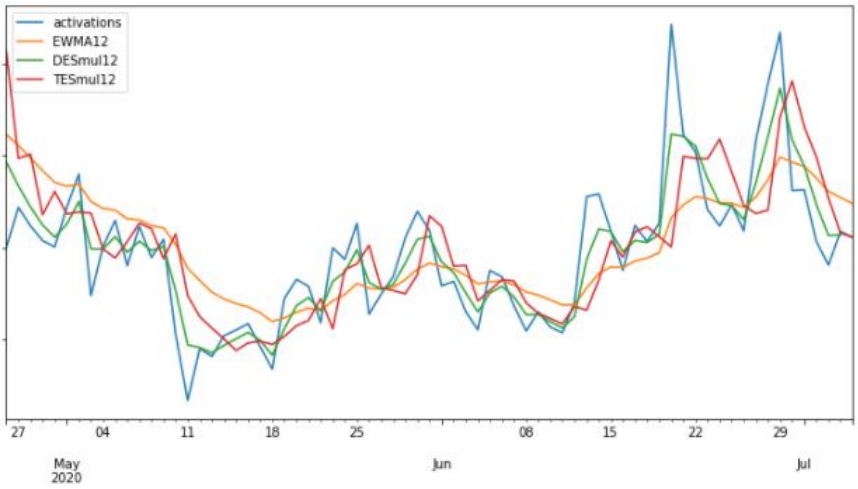


Ilustración 17. Comparación de los resultados al aplicar los modelos EWMA, DES y TES. Fuente: el presente trabajo, 2021.

Así, podría pensarse que el modelo DESmul12, propone un mejor ajuste que TESmul12, teniendo en cuenta la tendencia únicamente y no la estacionalidad, sin embargo, la clave aquí, es la ventaja que da el modelo TESmul12 al tener la capacidad de predecir patrones estacionales fluctuantes, que a pesar de producir una pequeña pérdida de precisión en la representación actual del conjunto, aporta una mejora en las posibles predicciones, abordando diferentes tipos de cambio (crecimiento/disminución) en la tendencia, con observación a los instantes de tiempo en que se generaban.

Esta observación, aportó al estudio de modelos integrales que tengan en cuenta el análisis de estas componentes, para mejorar la calidad de la representación de la información futura, que es un enfoque crucial, para el primer proceso automatizado, como se observará posteriormente.

2.5.3 Data Studio

Google Data Studio es el software de BI que se usa en la implementación de la propuesta para la generación de los reportes y la construcción de dashboards. Su elección se fundamenta en la facilidad de integración con Bigquery, la amplía personalización que permite la herramienta para desarrollar los informes y la posibilidad de actualizar la información en tiempo real.

Su estudio, se llevó a cabo directamente con la documentación y el curso certificado de Google. Inicialmente, se realizaron configuraciones de las fuentes de datos: Excel, hojas de cálculo de Google y Bigquery; se construyeron dashboards personalizados con la variedad de herramientas disponibles para visualizar información en gráficas, y se examinó el entorno en general, para aprender de su funcionamiento al momento de crear reportes asociados a los datos configurados.

Más adelante se describe la aplicación de esta herramienta, y cómo su aplicación permitió la construcción de dashboards con información sustancial del negocio para el servicio APV.

2.6 Definición de KPIs

Los KPIs son indicadores de rendimiento, que permiten visualizar el progreso hacia las metas planteadas. Las empresas, han complementado la definición de objetivos con estos indicadores, para poder medir precisamente en qué grado se están alcanzando.

De acuerdo a esto, la empresa Tigo no es la excepción, y el área Vp Digital específicamente, se especializa en suministrar información de alto valor para el negocio acorde a los KPIs estratégicos para la comercialización de los servicios.

Enfatizando, como se ha descrito previamente, el esfuerzo del presente trabajo se orienta a la automatización de los reportes del servicio APV, por lo que, de igual forma, es indispensable tener en cuenta cuáles son estas métricas clave, que promueven una medición estratégica para el cumplimiento de los objetivos del servicio.

Por asuntos de confidencialidad, no se mencionan los objetivos de la empresa con relación a este servicio, pero, para brindar mayor claridad a la implementación del proyecto, a continuación se describen los KPIs definidos al respecto, teniendo en cuenta la sintaxis en la que deben redactarse, según el estudio "**Un Método para la Definición de Indicadores Clave de Rendimiento con base en Objetivos de Mejoramiento**"[67].

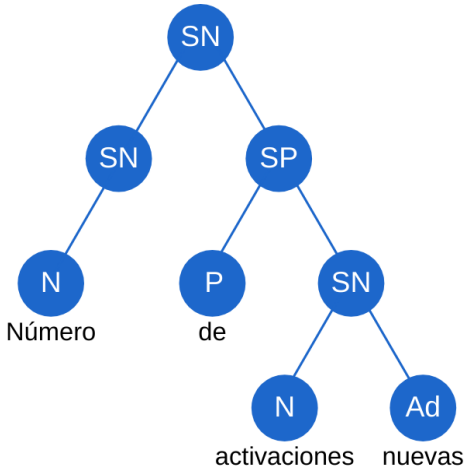
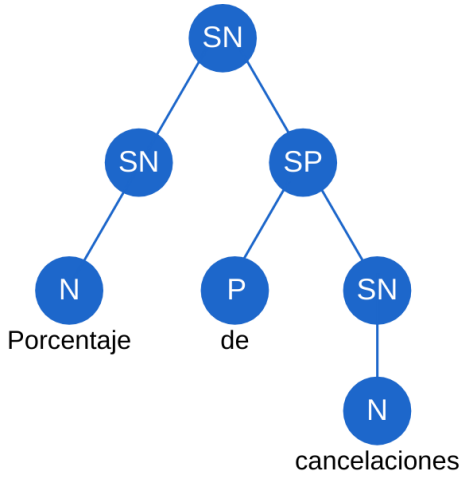
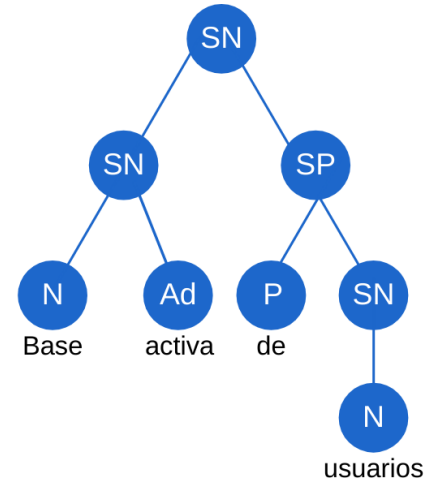
Primeramente, para brindar claridad acerca de la redacción de estos KPIs, se documentan en la tabla 8, cuáles son las reglas gramaticales que deben tenerse en cuenta, al proponer los KPIs.

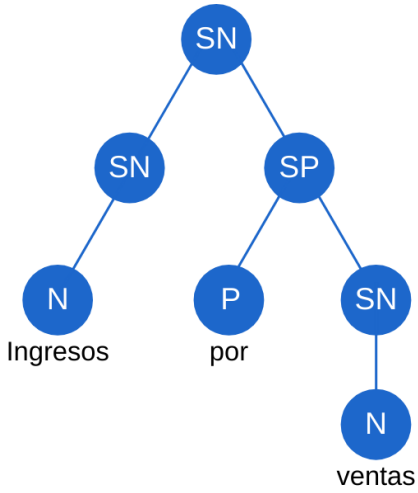
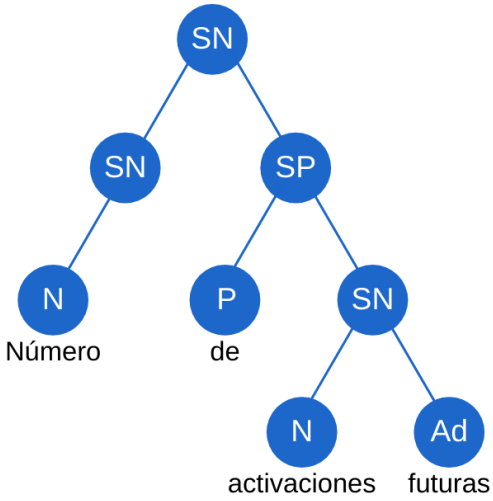
Tabla 8. Reglas gramaticales para redactar un KPI. Recuperado del documento "Un método para la definición de indicadores clave de rendimiento con base en objetivos de mejoramiento" (2019)

Numeral	Regla Gramatical
I.	Todo KPI debe tener la forma de un sintagma nominal (SN).
II.	La forma del sintagma nominal para KPIs siempre es SN = SN + SP, donde SN es sintagma nominal y SP sintagma preposicional.
III.	El SN que hace las veces de primer argumento en el SN enunciado en la regla ii puede tomar cualquiera de las siguientes formas SN = N; SN = N + Ad o SN = D + N, donde N es sustantivo, Ad es adjetivo y D es determinante.
IV.	El SP que hace las veces de segundo argumento en el SN enunciado en la regla ii es de la forma SP = P + SN, donde P es preposición y SN es sintagma nominal.
V.	El último SN del árbol perteneciente al SP enunciado en la regla IV puede tomar nuevamente la forma de SN = SN + SP o cualquiera de las definiciones que se dan en la regla iii para el SN.
VI.	Algunos de los sustantivos cuantificadores que se pueden utilizar para especificar un KPI bien definido son: proporción, porcentaje, número, tasa, probabilidad, cantidad, margen, nivel, aumento, retorno, costo, eficiencia y capacidad, tiempo, relación, monto.

Entendiendo esto, se prosigue a formular los KPIs en la tabla 9, conforme a la necesidad de información empresarial del servicio APV:

Tabla 9. Definición de KPIs. Fuente: el presente trabajo, 2021

KPI	Árbol de Constituyentes
Número de Activaciones Nuevas	
Porcentaje de cancelaciones	
Base activa de usuarios	

Ingresos por ventas	
Número de activaciones futuras	

Así, tal y como lo describe el documento señalado, el árbol de constituyentes señala la composición necesaria de sustantivos, adjetivos y preposiciones para redactar los KPIs propicios para el negocio en la forma de un sintagma nominal de forma resumida, coherente y clara para extraer información de alto valor para la toma de decisiones.

Los 5 KPIs redactados se relacionan con los procesos susceptibles identificados en el **punto 2.3**, los 4 primeros de conformidad al procesamiento de datos y el último con respecto al modelamiento predictivo.

2.7 Lógica de la automatización

La implementación hasta el momento se ha centrado en el estudio de los requerimientos, tecnología y el análisis del conjunto de datos a partir del cual se extraerá valor empresarial. Este conocimiento, que corresponde al desarrollo de procesos en las tres primeras etapas de BI, fue la base para la automatización implementada, es decir, para las etapas 4 y 5 respectivamente.

Con esto en mente, se diseñó un esquema que expone la lógica de la automatización requerida, fijando la atención en el procesamiento y predicción de los datos principalmente. El diseño de esta lógica, se compone de cada uno de los procesos necesarios, desde la extracción de los datos en el Data Warehouse hasta la visualización de la información conforme a los KPIs del negocio.

El esquema realizado a continuación, es la primera versión, construida a partir del conocimiento adquirido en las 8 primeras semanas, y evaluado con la ayuda del experto en la empresa.

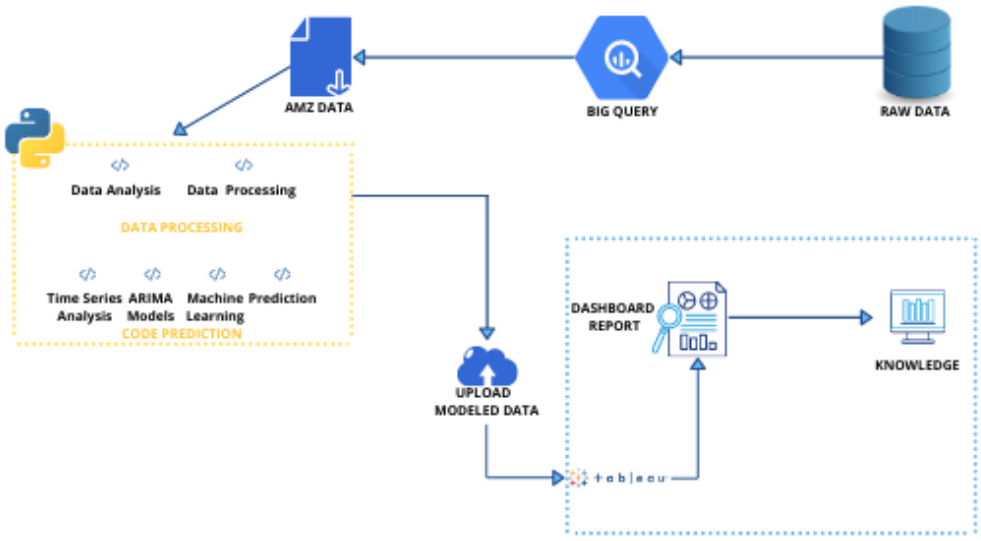


Ilustración 18. Lógica de la automatización requerida (primera versión). Fuente: el presente trabajo, 2021.

Esta primera versión, parte de considerar resueltos los procesos de extracción, transformación y carga de los datos del servicio APV en el repositorio de la compañía, en Bigquery. De ahí en adelante, a través del lenguaje SQL, y la ejecución de consultas personalizadas, se pensaba descargar el conjunto de datos útil directamente del Data Warehouse, para procesarlos y desarrollar modelos predictivos en Python.

A partir de este punto, los datos procesados y pronosticados, podrían ser importados en el software de BI, para construir los reportes y visualizaciones que brinden el conocimiento amplio del servicio. Cabe aclarar que, hasta este instante, se optaba por la utilización de Tableau como herramienta de BI, sin embargo, gracias a la asesoría del experto en el área, fue recomendable usar Google Data Studio, ya que, al permitir la conexión de los datos directamente desde Bigquery, la automatización del procesamiento con la ejecución de consultas personalizadas, podía optimizarse.

En consecuencia, fue necesario diseñar un nuevo método para mejorar la ejecución de los 2 procesos señalados; el resultado expuesto en la ilustración 19, después de ser evaluado por

el experto, fue aprobado, permitiendo establecerlo como la base lógica de la automatización a implementar.

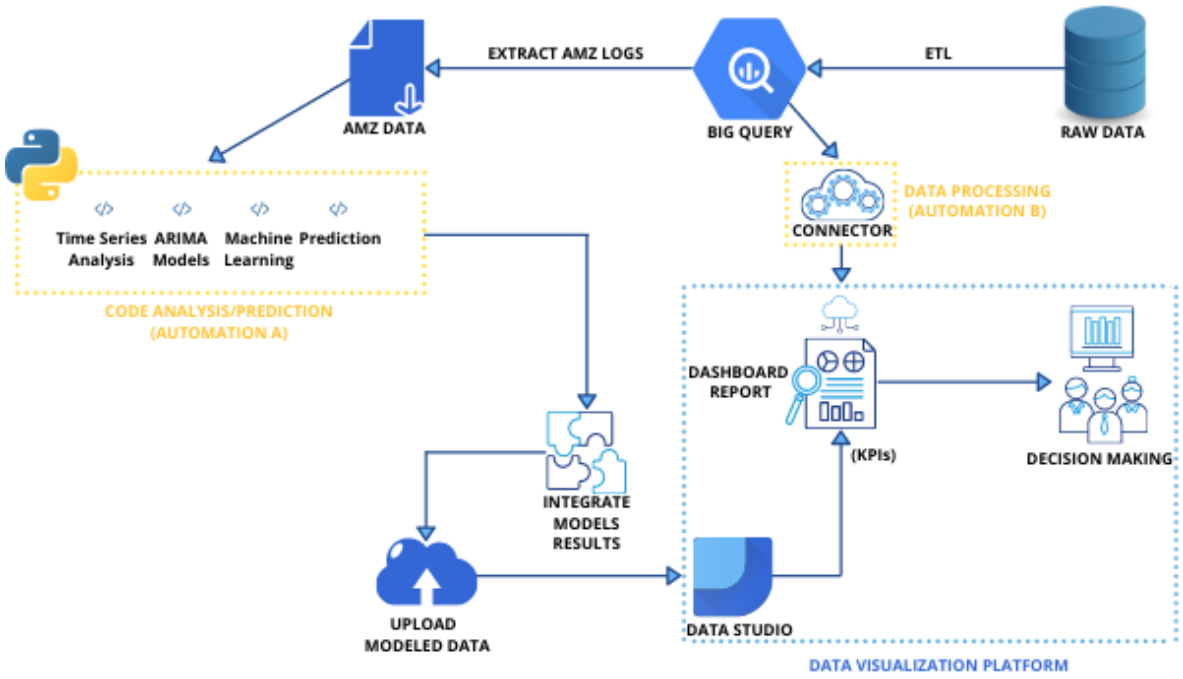


Ilustración 19. Lógica de la automatización requerida (versión final). Fuente: el presente trabajo, 2021.

La lógica actual, coincide en considerar a Bigquery como el repositorio a partir del cual se extraen los datos, con la particularidad de que, ahora sólo un conjunto de estos, es exportado para el análisis y el desarrollo de modelos predictivos en Python, mientras que, el procesamiento de los datos, se realiza a través de SQL con Google Cloud directamente, construyendo información a partir de la manipulación sistematizada de los registros del servicio APV.

Por otro lado, en esta lógica, se utiliza Google Data Studio como herramienta de BI, y por lo tanto, los conjuntos de datos procesados y pronosticados, pueden converger a la herramienta para construir las visualizaciones mediante la configuración de diversos conectores, que vinculan las fuentes de datos correspondientes.

Así, la lógica diseñada resume el aporte central del presente proyecto, estableciendo la guía base para suministrar un rendimiento automático de los procesos desde la concepción de los datos brutos a la generación de información de valor para la toma de decisiones, integrando una solución estable y escalable respecto al volumen de registros mediante la codificación, la aplicación de herramientas estadísticas y la utilización de un software de BI.

2.8 Automatización

En consonancia con lo estudiado, los ejemplos introductorios del **punto 2.5** y la exploración del conjunto de datos del servicio, se ha expuesto que, la naturaleza de los datos es acorde al comportamiento descrito por una serie temporal. De esta manera, es posible desarrollar modelos analíticos, que coincidan con esta característica, con el fin de procesar y predecir los hechos respecto a las marcas de tiempo, que ya sucedieron, que están sucediendo, y que tendrán lugar en un futuro próximo.

Con base a ello, a continuación se explica cómo se implementó la automatización requerida para los dos procesos ya identificados.

2.8.1 Proceso A: Modelamiento Predictivo

Específicamente, este proceso está orientado a satisfacer la necesidad de información, respecto a la cantidad de personas que se activarán en el servicio en los próximos 30 días. La información debe suministrar una predicción diaria y una predicción en conjunto de ese mes en específico.

Al respecto, no se pretende hacer una predicción con base a factores porcentuales de forma manual como se hacía en Excel, según se evidenció anteriormente. Al contrario, se implementaron modelos estadísticos teniendo en cuenta los criterios del punto **2.3.1**, con el fin de que el modelamiento predictivo de las activaciones no sólo tenga convergencia a valores lógicos, sino que llegue a ello, mediante la asimilación de los datos, y el ajuste lineal/no lineal que represente óptimamente la naturaleza de dichos registros de manera automática, garantizando siempre versatilidad para adaptarse al volumen de los datos.

Entendiendo esto, el primer paso fue identificar el conjunto de datos útiles para solventar esta necesidad de información mediante la exploración de las tablas en Bigquery. Teniendo claro ello, se procedió a extraer del DW los datos asociados a las activaciones del servicio y su fecha de procesamiento, desarrollando consultas personalizadas en SQL, para adecuar los elementos en un formato conveniente.

```
SELECT COUNT (distinct(                    )),Date(a.activation_date) as date
FROM `                                                                  ` a
WHERE activation_date >= '2019-05-16'
GROUP BY date
ORDER BY date asc
```

Ilustración 20. Ejemplo de consulta personalizada para extraer el número de activaciones del servicio APV en un rango de fecha determinado. Fuente: el presente trabajo, 2021.

El resultado obtenido, expone en la columna **f0_**, el conteo de las activaciones del servicio, y en la columna **date**, la fecha en que sucedieron esas activaciones. Esto, contribuyó a mantener un formato acorde a una serie temporal, con las marcas en el tiempo requeridas para los modelos predictivos que se explican posteriormente.

Fila	f0_	date
1		2019-05-16
2		2019-05-17
3		2019-05-18
4		2019-05-19
5		2019-05-20
6		2019-05-21

Ilustración 21. Resultados de la consulta personalizada previa. Fuente: el presente trabajo, 2021.

Posteriormente, estos datos fueron exportados en formato .csv, e importados en Python, mediante Pandas, dando como resultado un Dataframe, que, en adelante constituyó, el elemento contenedor de los datos. Cabe mencionar que, para la implementación de los modelos, los elementos de las activaciones **f0_**, fueron indexados a los elementos de tiempo **date**, los cuales, se ajustaron al formato DateTimeIndex, de Pandas, con el parámetro parse_dates=***True***, para reconocerlos como índices de tiempo, y analizar los elementos temporales sin inconvenientes.

Ahora bien, una vez el conjunto de datos (objeto de análisis) fue importado y adecuado en el formato requerido, se procedió a estudiar diferentes modelos, para implementar un pronóstico óptimo de las activaciones del servicio.

Con base a esto, y como se evidenció en los ejemplos de estudio y exploración de los datos, fue necesario tener en cuenta, modelos correspondientes al análisis de series temporales, que asimilen las componentes de tendencia y estacionalidad, las cuales, son características intrínsecas en el conjunto de datos del servicio APV.

Es por ello que, para automatizar este proceso, se recurrió a examinar: la literatura estudiada en el estado del arte, el contenido de cursos como "*Python for Time Series Data Analysis*" y "*Python for Time Series Analysis and Forecasting*" en la plataforma Udemy, y blogs reconocidos en este campo como Towards Data Science [68]. Contenido, que convergía al análisis de series temporales, mediante la aplicación de modelos ARIMA y Redes Neuronales Recurrentes (RNN).

Es necesario tener en cuenta, que el sustento teórico de los modelos, se describió en el Marco Conceptual; este capítulo, solamente está orientado a su implementación, y la utilización de las librerías correspondientes. Además, es preciso manifestar que, los Scripts, que contienen los algoritmos implementados, se adjuntan al presente trabajo, con el nombre específico de cada uno de los modelos como se observará en la descripción respectiva más adelante.

Previo a exponer la implementación de estos modelos, se expone a continuación algunos aspectos generales, que se tuvieron en cuenta en el desarrollo de los pronósticos.

a. Métricas de Error

Con el fin de obtener un margen entre el pronóstico y el valor real, se consideraron 2 tipos de errores, descritos a continuación, que ayudaron a evaluar la precisión de las predicciones obtenidas por los modelos, y orientar un re ajuste de estos, en caso de obtener respuestas poco favorables.

- *MAE:* mide la magnitud promedio de los errores en el conjunto de predicciones, sin considerar su dirección. Fue utilizado para validar la precisión de los modelos implementados, promediando la muestra de prueba de las diferencias absolutas entre la predicción y las observaciones reales, donde todas las diferencias individuales, poseen el mismo peso.

- *RMSE:* es la raíz cuadrada del promedio de las diferencias al cuadrado entre la predicción y la observación real. Se evidenció que este error, daba más importancia a los errores más significativos del conjunto, significando que un gran valor pronosticado de activaciones muy lejos de los conjuntos de pruebas reales, era suficiente para obtener un RMSE muy malo.

Comprendiendo esto, se examinaron los dos tipos de errores para verificar la fiabilidad de las activaciones pronosticadas en cada uno de los modelos, determinando que, RMSE daba mucha más importancia a los errores más significativos, mientras que MAE, al dar la misma importancia a cada error, permitía una percepción más intuitiva para las evaluaciones.

b. Procedimiento:

Por otra parte, los modelos implementados, siguen una secuencia lógica en cuanto a su ejecución, como se manifiesta en seguida, donde cabe aclarar, surgen algunas variaciones para algunos de ellos.

- Selección de un modelo: elección de los modelos ARIMA o RNN, e importar las librerías respectivas.

- División de los datos en un conjunto de entrenamiento y un conjunto de pruebas: el conjunto de pruebas en series temporales está representado por los datos más recientes. El tamaño convencional de este conjunto es alrededor del 20% de la muestra total, aunque realmente, este valor depende de la longitud de la muestra y de la predicción que se desea realizar. Debido a que la cantidad de registros del servicio APV, no sobrepasaba los 2 años de existencia, fue conveniente tener en cuenta un margen circunscrito en esa configuración convencional, garantizando que al menos el conjunto de pruebas, represente el horizonte de pronóstico máximo solicitado de 30 días.

- Ajuste del modelo en el conjunto de entrenamiento: encontrar un patrón en los datos, a través de la aplicación de las funciones específicas de cada modelo para generalizar una respuesta similar al comportamiento de los datos a través del entrenamiento del conjunto destinado para ello.

- Evaluación del modelo en el conjunto de prueba: para este paso se tuvo en cuenta, validar las predicciones estimadas en un rango de 30 días con relación a los datos de prueba, que corresponden a la misma ventana de tiempo. Estos 30 días, varían en cuanto a la fecha de inicio y fin, de conformidad con la división de los datos en entrenamiento/prueba realizada en cada uno de los modelos al inicio. De acuerdo a esta evaluación, se reajustaron algunos parámetros para verificar la precisión de los modelos en la interpretación de los datos, con el fin de establecer la configuración más acertada.

  En este capítulo, simplemente se muestra la aproximación gráfica de la evaluación, en el próximo capítulo se especifican los errores MAE y RMSE, resultantes de las versiones finales de los modelos implementados, que permitieron optar por mayor fiabilidad en los pronósticos entregados.

- Re ajustar el modelo en todo el conjunto de datos: corresponde a la aplicación del modelo (con mejor desempeño en las pruebas) al conjunto de datos completo.

- Pronosticar los resultados: predecir los datos según la configuración del modelo, y adecuarlos en un arreglo ordenado, mediante la librería Numpy, para su exportación en formato .csv con Pandas. Los resultados de este proceso se exponen en el siguiente capítulo.

2.8.1.1 Holt Winters

Holt-Winters es un método que incluye el análisis de cambios y tendencias, así como a patrones estacionales de las series de datos. Habitualmente, es usado por muchas compañías para pronosticar la demanda a corto plazo cuando los datos de venta contienen tendencias y patrones estacionales de un modo subyacente, como en este caso, con el servicio APV. Su implementación se expone en el script (**Scripts/Forecasting/01-HoltWinters-APVforecasting.ipynb**).

Implementación:

1. Importar librerías del Modelo

'***from statsmodels.tsa.holtwinters import ExponentialSmoothing***'

2. División de los datos en conjunto de entrenamiento y pruebas

Se adecuó la configuración convencional, destinando 80% de los datos para el conjunto de entrenamiento y 20% para el conjunto de pruebas.

- Conjunto de entrenamiento: 333
- Conjunto de prueba: 84

3. Ajuste del modelo

fitted_model = ExponentialSmoothing(train_data['activations'],trend='mul',seasonal='mul',seasonal_periods=14).fit()

Configuración de la estacionalidad a 14 registros, teniendo en cuenta la estacionalidad semanal de 7 días, este múltiplo fue la mejor configuración.

4. Evaluación del Modelo

Las pruebas producían resultados muy favorables identificando marcas de tiempo exactas en que se generaba un incremento o decremento en la tendencia de las activaciones.

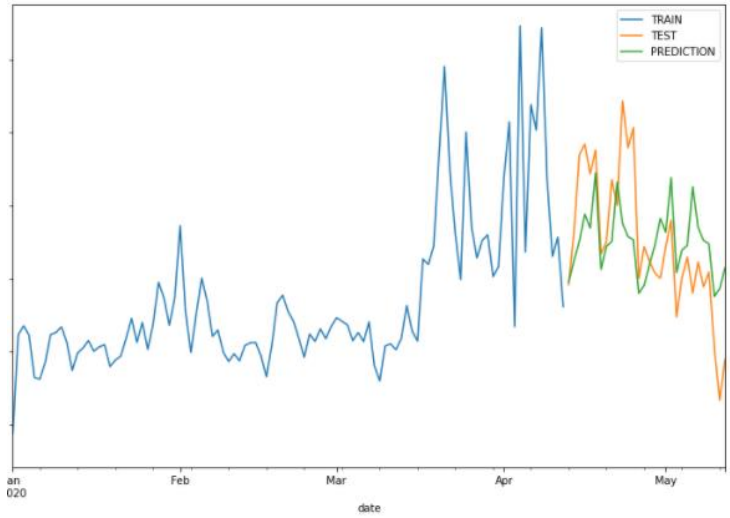


Ilustración 22. Comparación activaciones pronosticadas (Prediction) vs activaciones reales del conjunto de prueba (Test). Fuente: el presente trabajo, 2021.

5. Re ajuste del Modelo en todo el conjunto

Utilizando la función ExponentialSmoothing:

final_model = ExponentialSmoothing(df['activations'],trend='mul',seasonal='mul', seasonal_periods=14).fit()

6. Predicción

Se pronosticaron los datos, una vez el modelo configurado fue ajustado a todo el conjunto.

forecast_predictions = final_model.forecast(30).rename('APV HWinters forecasting')

2.8.1.2 Modelo Autoregresivo (AR)

Fue implementado para pronosticar el número de activaciones del producto APV según un conjunto de predictores de nivel, tendencia y estacionalidad. Su implementación se expone en el script (**Scripts/Forecasting/02-AutoRegressive-APVforecasting.ipynb**).

Implementación:

1. Importar librerías del Modelo

'from statsmodels.tsa.ar_model import AR,ARResults'

2. División de los datos en conjunto de entrenamiento y pruebas

Por conveniencia para garantizar el mejor desempeño del modelo, se requirieron datos de aproximadamente un mes más que el modelo anterior para el conjunto de entrenamiento, ya que se evidenciaba poca afinidad con la representación del conjunto con menores registros.

- Conjunto de entrenamiento: 357
- Conjunto de prueba: 60

3. Ajuste del modelo (Statsmodels determina el valor de p automáticamente)

Instancia del modelo:

newModel = AR(train['activations'])

Entrenamiento:

ARfit = newModel.fit(method='mle')

Obtención del valor de p calculado por Statsmodels:

print(f'Lag: {ARfit.k_ar}')

4. Evaluación del Modelo

La ilustración 23, muestra los resultados pronosticados en comparación con el conjunto de pruebas para evaluar la configuración del modelo. En el script se observa que se implementaron 3 modelos AR, los 2 primeros, con orden p=1 y p=2 respectivamente, y el último con p=16, sugerido directamente por Statsmodels, mediante el reconocimiento automático de los coeficientes requeridos en la combinación lineal de la predicción definida por el modelo.

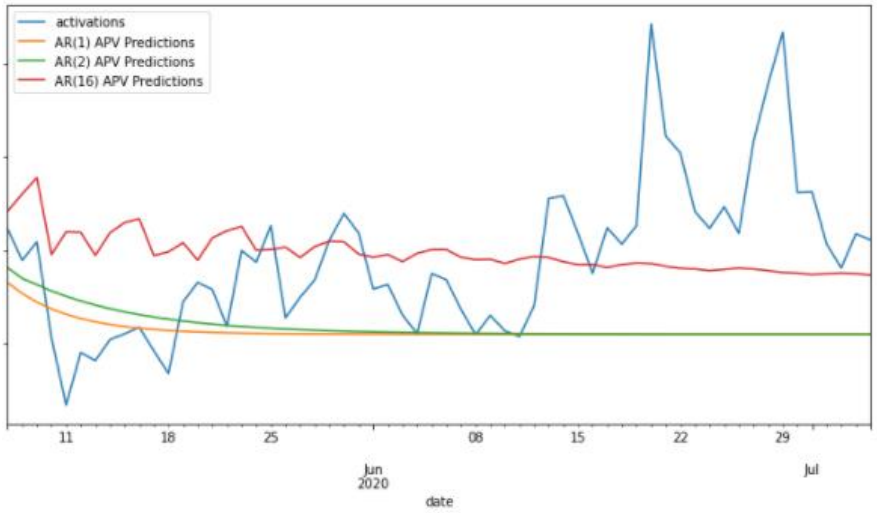


Ilustración 23. Comparación Modelos AR(1), AR(2), AR(16) vs Conjunto de pruebas. Fuente: el presente trabajo, 2021.

Así, siendo el modelo AR(16), el más próximo para la representación de las fluctuaciones estacionales y cambios de tendencia de los datos del servicio, se evaluó nuevamente, esta vez, sesgando el alcance de tiempo de los 30 días requeridos, como se observa a continuación.

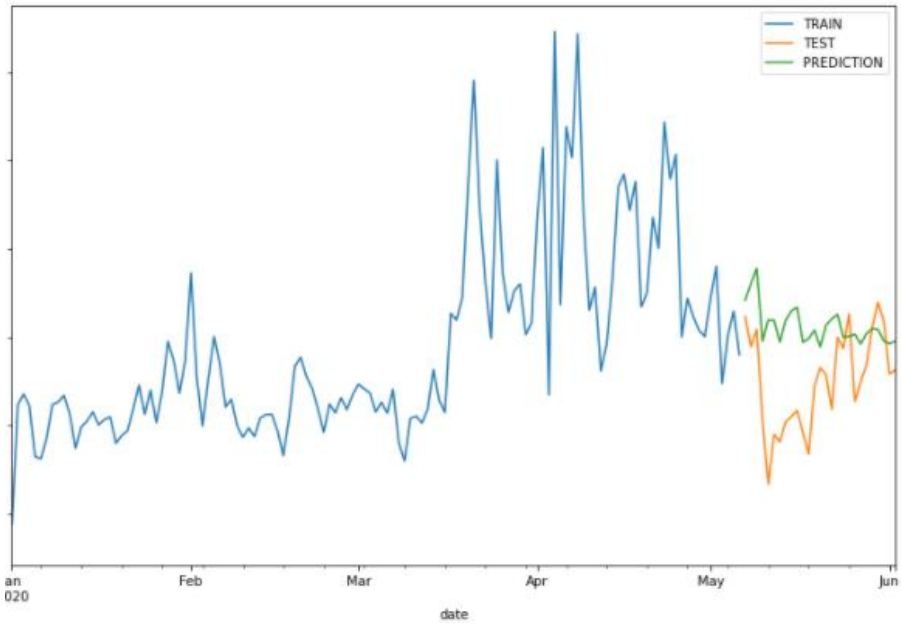


Ilustración 24. Comparación activaciones pronosticadas (Prediction) vs activaciones reales del conjunto de prueba (Test). Fuente: el presente trabajo, 2021.

5. Re ajuste del Modelo en todo el conjunto

Se entrenó el conjunto entero mediante la función AR del modelo:

model = AR(df['activations'])

Se ajustó el modelo, usando la instancia *model* ya realizada:

ARfit = model.fit()

6. Predicción

Se pronosticaron los datos, una vez el modelo configurado fue ajustado a todo el conjunto.

Fcast = Arfit.predict(start=len(df), end=len(df)+30, dynamic=False).rename ('APV AR(16) forecasting')

2.8.1.3 Modelo ARIMA (5,1,2)

Un modelo ARIMA, es usado para descomponer y explicar una serie de tiempo determinada en función de sus propios valores pasados, es decir, sus propios retrasos y los errores de pronóstico retrasados.

De esta manera, es posible predecir resultados con base a la Auto Regresión (AR), que utiliza la relación dependiente entre una observación y cierto número de observaciones pasadas, y Medias Móviles (MA), utilizando la dependencia entre una observación y un error residual del modelo, aplicado a observaciones pasadas.

Lo anterior, es aplicable para series temporales que no presentan tendencia o estacionalidad, es decir, que son estacionarias, de lo contrario, es necesario aplicar en ellas un proceso de diferenciación (I) para convertirlas en estacionarias, tal y como es el caso del conjunto de datos del servicio APV.

La implementación del modelo expuesto en el script (**Scripts/Forecasting/03-ARIMA(1,1,1)pdArima-APVforecasting.ipynb)** consta de los pasos siguientes:

1. Automatización de la prueba de Estacionariedad (Dickey Fuller Aumentada)

Dickey Fuller es una prueba que valida si una serie temporal es o no estacionaria. La función implementada, facilitó su aplicación e incluyó un parámetro configurable para diferenciar la serie, en caso de requerir convertirla en estacionaria como lo exige el modelo ARMA.

2. Importar librerías del Modelo

from statsmodels.tsa.arima_model import ARMA,ARMAResults,ARIMA,ARIMA Results
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
from pmdarima import auto_arima

3. Gráficas ACF y PACF

Primero, es útil comentar que, mediante la función gráfica de retardo para series de tiempo incorporada en Pandas, fue posible trazar los valores $\mathbf{y_t}$ en el eje horizontal contra versiones retrasadas de los valores $\mathbf{y_{t+1}}$ en el eje vertical.

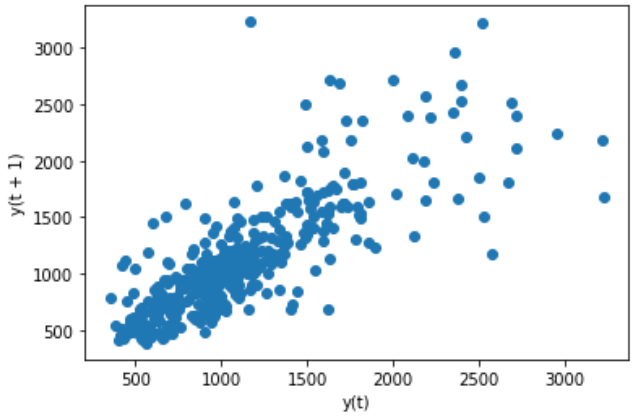


Ilustración 25. Correlograma de las Activaciones y_t vs y_{t+1}. Fuente: el presente trabajo, 2021.

Se observa que, al tener un conjunto de datos no estacionario, con una tendencia ascendente, los valores vecinos asumían la misma tendencia diagonal creciente, en especial, para los primeros valores, evidenciando la existencia de una correlación entre $\mathbf{y_t}$ y $\mathbf{y_{t+1}}$ para las personas que se habían activado en el servicio, la cual, no era constante, pero tendía a una relación lineal positiva.

En consecuencia, se procedió a estudiar los gráficos de **ACF** (Autocorrelation Function) y **PACF** (Partial Autocorrelation Function) para determinar los valores recomendables de **q** y **p** respectivamente.

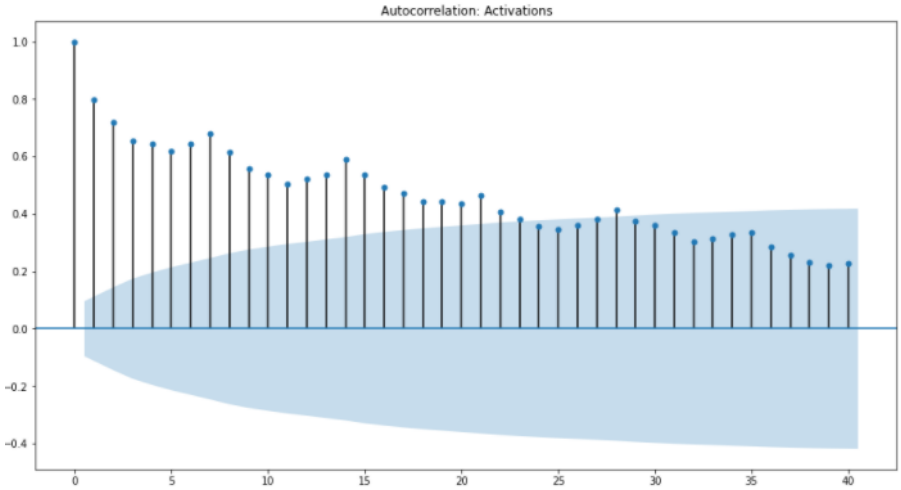


Ilustración 26. Gráfica ACF. Fuente: el presente trabajo, 2021.

De lo anterior, se observó que hay un factor estacional en los datos, debido a la evidencia señalada por los picos (pequeños máximos locales), pues indican que, en cierto punto, la correlación automática empezó a aumentar, con una periodicidad semanal, exponiendo estos picos de la correlación, alrededor de las marcas: 7, 14, 21, 28, etc, tal y como se puede observar en la gráfica.

Ahora bien, la región sombreada de azul, representa un intervalo de confianza del 95%. Y básicamente todo lo que eso significa es que, sugiere que es muy probable que los valores fuera de este intervalo de confianza sean una correlación. Se observa que la región sombreada se hace cada vez más grande a medida que sus retrasos aumentan; lo que tiene sentido, pues se asegura de exponer un índice de correlación con los retrasos más recientes, que aquellos que sucedieron hace más tiempo.

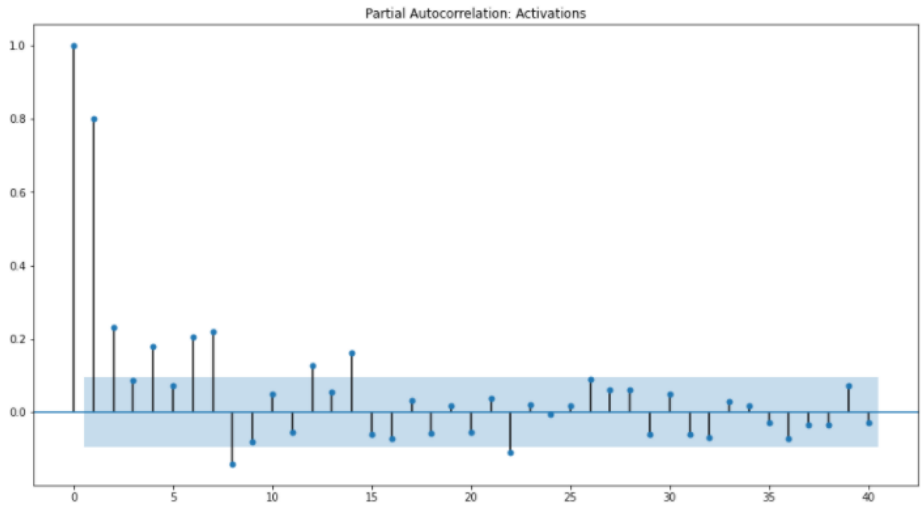


Ilustración 27. Gráfica PACF. Fuente: el presente trabajo, 2021.

De acuerdo con el sitio de pronóstico estadístico de la Universidad de Duke [69] "*si el PACF muestra un corte brusco mientras que el ACF decae más lentamente (es decir, tiene picos significativos en retrasos más altos), se puede afirmar que, la serie (ya estacionaria) muestra una firma AR, lo que significa que el patrón de autocorrelación se puede explicar más fácilmente agregando AR términos que agregando términos MA*".

De lo anterior se pudo concluir que, las gráficas muestran índices de correlación en más de 5 retrasos, por lo que, suponer términos bajos para el modelo ARIMA, a pesar de que garantiza simplicidad, no es suficiente para adecuarse al comportamiento que requieren los datos de las activaciones. Prueba de ello, es la ejecución de la función automática *auto_arim*a, descrita a continuación, que propuso la implementación de un modelo ARIMA (1,1,1), y no se adaptó a las fluctuaciones del conjunto, por lo que, como se evidencia más adelante en el paso 8, se requirió el complemento sugerido por las gráficas ACF y PACF ya vistas, aumentando el orden del factor Autorregresivo (AR).

4. Determinar el valor de los órdenes **p (AR)** y **q (MA)** usando la función **pmdarima.auto_arima**

```
Ajuste ARIMA: orden = (0, 1, 0); AIC = 5964.128
Ajuste ARIMA: orden = (1, 1, 0); AIC = 5925.431
Ajuste ARIMA: orden = (0, 1, 1); AIC = 5902.197
Ajuste ARIMA: orden = (1, 1, 1); AIC = 5873.815
Ajuste ARIMA: orden = (1, 1, 2); AIC = 5875.580
Ajuste ARIMA: orden = (2, 1, 2); AIC = 5874.371
Ajuste ARIMA: orden = (2, 1, 1); AIC = 5875.427
```

Ilustración 28. Selección automática del modelo ARIMA conveniente. Fuente: el presente trabajo, 2021.

La función auto_arima, propuso un modelo ARIMA a partir de la combinación de los diferentes órdenes, evaluando su efectividad con base en la métrica de desempeño AIC (Akaike Information Criterion), la cual, compara un montón de modelos basados en ARIMA con diferentes pesos en sus valores p y q, y estima la calidad de cada configuración con relación a las demás; permitiendo penalizar a los modelos que realmente usan demasiados parámetros e incluso se sobre ajustan al conjunto.

Así, el modelo resultante de esta función automática fue el ARIMA (1,1,1) y se implementó mediante la función del paso 6.

5. División de los datos en conjunto de entrenamiento y pruebas

Teniendo en cuenta el análisis realizado por los modelos ARIMA con la autoregresión y las medias móviles, y examinando el creciente cambio de las activaciones a raíz de la pandemia en el mes de abril, se consideró un conjunto de entrenamiento más amplio, conservando únicamente en el conjunto de prueba, la cantidad de datos requeridos por la ventana de tiempo para la predicción.

Así,

- Conjunto de entrenamiento: 387
- Conjunto de pruebas: 30

6. Ajuste del modelo

Instancia del modelo:

model = ARIMA(train['activations'],order=(1,1,1))

Entrenamiento:

results = model.fit()

7. Evaluar el Modelo

A pesar de obtener errores MAE y RMSE adecuados, cuando se compararon las predicciones con el conjunto de pruebas, se observó, que el modelo no se adaptaba a las fluctuaciones estacionales de las activaciones reales del servicio, pues simplemente definía un promedio, por lo que, fue necesario complementar la aproximación automática de la función auto_arima teniendo en cuenta las gráficas ACF y PACF ya expuestas.

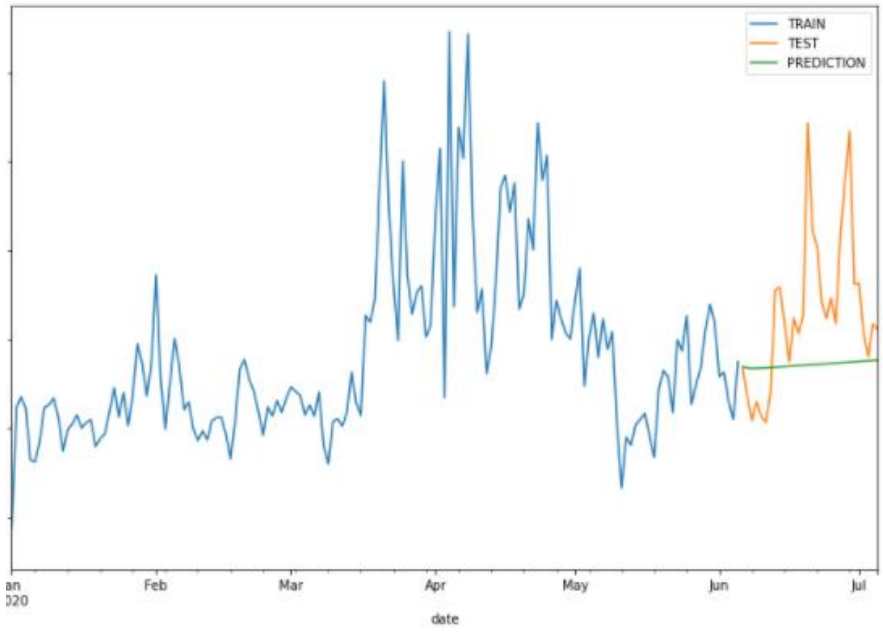


Ilustración 29. Comparación activaciones pronosticadas (Prediction) vs activaciones reales del conjunto de prueba (Test). Fuente: el presente trabajo, 2021.

8. Ajuste de los términos del modelo ARIMA

Con esto en mente, se probaron diferentes configuraciones para ajustar el modelo más precisamente, entre ellas, un ARIMA (16,1,2), teniendo en cuenta el ajuste del modelo autorregresivo anterior con orden 16, sin embargo, pese a que se ajustó muy bien a las fluctuaciones de cambio presente en la serie, su complejidad, fue altamente penalizada por el criterio AIC, ya que puede generar un sobre ajuste al conjunto, por lo que se optó, por una configuración más simple, ARIMA (5,1,2) agregando solamente 4 grados más al término AR con relación al modelo previo ARIMA (1,1,1) y un grado al término de MA, como sugiere el sitio de pronóstico estadístico de la Universidad de Duke.

9. Evaluación del Modelo ARIMA (5,1,2)

Los errores MAE y RMSE, variaron mínimamente con esta configuración, sin embargo, como se puede observar en la gráfica, el modelo se adaptó a la estacionalidad del conjunto, lo cual fue indispensable para determinar una predicción más certera, que reconocía los instantes de tiempo, en que se generaban más y menos activaciones del servicio.

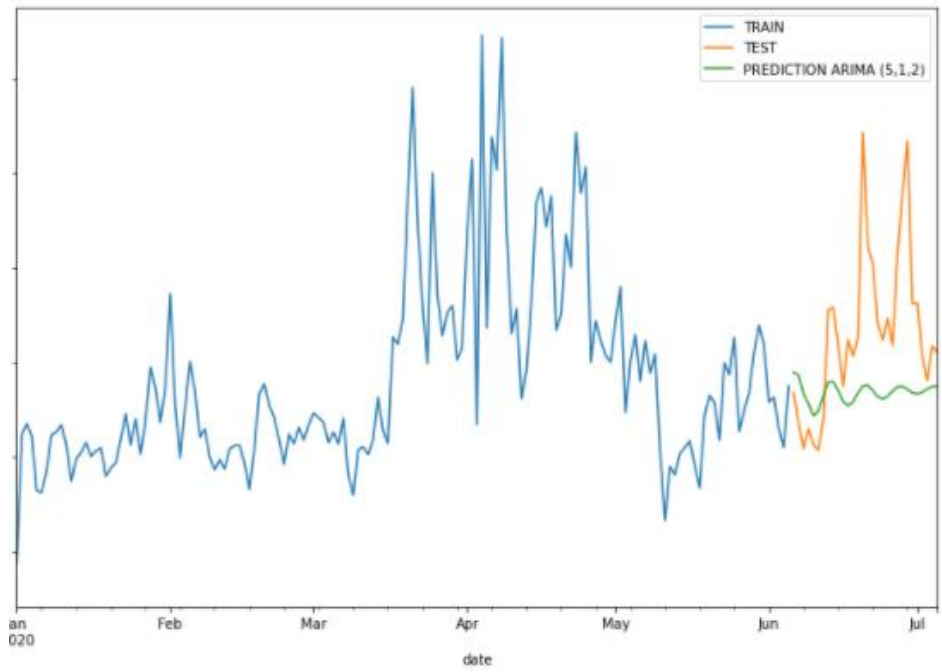


Ilustración 30. Comparación de las activaciones pronosticadas (Prediction) vs activaciones reales del conjunto de prueba (Test) [ARIMA (5,1,2)]. Fuente: el presente trabajo, 2021.

10. Re ajuste del Modelo en todo el conjunto

Se entrenó el conjunto entero mediante la función ARIMA implícita en el modelo.

model1 = ARIMA(df['activations'],order=(5,1,2))

Se ajustó el modelo, usando la instancia *model1*.

results1 = model1.fit()

11. Predicción:

Se pronosticaron los datos, una vez el modelo configurado fue ajustado a todo el conjunto.

fcast = results1.predict(len(df),len(df)+30,typ='levels').rename('ARIMA (5,1,2)Forecast')

2.8.1.4 Modelo SARIMA (2,1,2) (1, 0, 1, 7)

El modelo se desarrolla en el script (**Scripts/Forecasting/ 04-SARIMA-APVforecasting-AUTO.ipynb)**, consta de los pasos siguientes:

1. Importar librerías del Modelo:

from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
from pmdarima import auto_arima

2. Ejecución de la función **pmdarima.auto_arima** para obtener los términos sugeridos del modelo

A diferencia del modelo anterior, se ejecutó la función auto_arima, especificando como parámetro ***seasonal=True***, para considerar el análisis estacional en la serie temporal.

auto_arima(df['activations'],seasonal=True,m=7).summary()

Como resultado, la función sugirió la implementación del modelo: SARIMA (2,1,2) (1,0,1,7)

3. Dividir datos en conjunto de entrenamiento y pruebas

Para este caso fue propicio considerar un 87% de los datos para el conjunto de entrenamiento, y casi 2 meses de los registros para el conjunto de pruebas. Con el fin de asegurar un entrenamiento estable a las variaciones temporales definidas por la pandemia, sin desconocer un 13% para el conjunto de pruebas, que fue suficiente para ratificar la configuración del modelo implementado.

- Conjunto de entrenamiento: 362
- Conjunto de pruebas: 55

4. Ajustar el Modelo SARIMA (2,1,2) (1,0,1,7)

Creación del modelo:

model = SARIMAX(train['activations'],order=(2,1,2),seasonal_order=(1,0,1,7))

Entrenamiento:

results = model.fit()

5. Evaluación del Modelo:

Los errores MAE y RMSE obtenidos fueron adecuados, y la comparación con el conjunto de pruebas, verificó la coherencia interpretada por el modelo para adaptarse a las fluctuaciones de cambio estacional de las activaciones, como se observa en la gráfica.

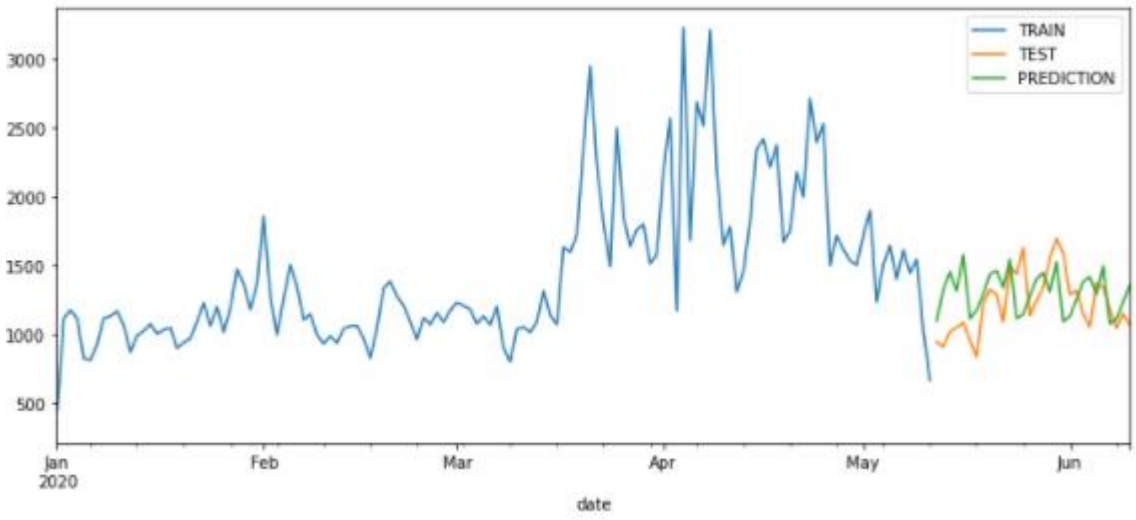


Ilustración 31. Comparación de las activaciones pronosticadas (Prediction) vs activaciones reales del conjunto de prueba (Test). Fuente: el presente trabajo, 2021.

6. Re ajuste del Modelo en todo el conjunto

Se entrenó el conjunto entero mediante la función SARIMAX implícita en el modelo.

model = SARIMAX(df['activations'],order=(2,1,2),seasonal_order=(1,0,1,7))

Se ajustó el modelo, usando la instancia *model* ya realizada.

results = model.fit()

7. Predicción:

Se pronosticaron los datos, una vez el modelo configurado fue ajustado a todo el conjunto.

fcast = results.predict(len(df),len(df)+30,typ='levels').rename('SARIMA(2,1,2)(1,0,1,7) Forecast')

2.8.1.5 Modelo RNN

El modelo se desarrolla en el script (**Scripts/Forecasting/ 05-RNN-APVforecasting.ipynb)**, consta de los pasos siguientes.

1. Importar Librerías del Modelo:

from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM

2. División de los datos en conjunto de entrenamiento y pruebas:

Teniendo en cuenta la forma en que opera el análisis de las series temporales con RNN, y examinando la variación de las activaciones a raíz de la pandemia en el mes de abril, se consideró un conjunto de entrenamiento amplío, conservando únicamente en el conjunto de prueba, la cantidad de datos requerida por la ventana de tiempo para la predicción.

Es decir:

- Conjunto de entrenamiento: 387
- Conjunto de pruebas: 30

3. Escalar los datos:

Mediante la función MinMaxScaler de ScikitLearn, fue posible transformar los conjuntos de entrenamiento y prueba escalando cada registro a un rango determinado, para este caso, entre 0 y 1.

Dicha función, claramente se ajustó sobre el conjunto de entrenamiento, ya que en teoría, el conjunto de pruebas se desconoce; y posteriormente se aplicó la transformación para ambos, reduciendo los datos a este rango (o cerca, pues algunos datos en el conjunto de pruebas pueden ser superiores a los ajustados en el conjunto de entrenamiento).

scaled_train = scaler.transform(train)
scaled_test = scaler.transform(test)

4. Generador de Series de Tiempo:

Debido a que la red neuronal recurrente básicamente necesita obtener un paso de tiempo uno, un paso de tiempo dos, un paso de tiempo tres y luego algún tipo de etiqueta para el paso de tiempo cuatro, como se puede observar a continuación. Keras transforma el arreglo de los datos de prueba, en ‘batches’ (lotes) de la siguiente forma.

Afortunadamente, Keras tiene un objeto generador de series de tiempo de preprocesamiento completo que hace esto automáticamente, proporcionando una secuencia de puntos de datos recopilados a intervalos iguales y devolviéndolos como ‘batches’ de ellos.

Así, el generador de series temporales fue implementado para captar la cantidad de entradas que se deseaban, para producir el siguiente paso. Inicialmente, se suministraron 7 datos de entrada, para predecir las activaciones del siguiente día, sin embargo, debido a la afectación implícita de la estacionalidad semanal en los datos, fue más conveniente suministrar 1 sola

entrada, para predecir con base a ella, el siguiente día. Lo cual, como se ve más adelante, garantizó excelentes resultados.

5. Crear el Modelo:

Una vez implementado esto, se procedió a diseñar y poner en marcha la red LSTM, la cual, es un tipo especial de red neuronal recurrente (RNN), con una amplia aplicabilidad para series temporales. La característica principal de LSTM es que la información puede persistir introduciendo bucles en el diagrama de la red, por lo que, básicamente, se podía recordar estados previos de las activaciones del servicio y utilizar esta información para decidir qué sucederá en los siguientes días.

La red neuronal se definió como una secuencia de dos capas densamente conectadas, es decir, que todas las neuronas de cada capa están conectadas con todas las neuronas de la capa siguiente. La red, tiene una capa visible con 1 entrada, una capa oculta con 150 neuronas y una capa de salida que genera un solo valor para la predicción. Se usó la función de activación relu como recomendación para los bloques de memoria LSTM con el análisis para series temporales.

El modelo fue compilado con el optimizador adam, el cual es un algoritmo de optimización de reemplazo para el descenso de gradiente estocástico para entrenar modelos de aprendizaje profundo. Por otra parte, se tuvo en cuenta la evaluación respecto al error MSE, con el fin de penalizar valores muy diferentes a los esperados, o en otras palabras, valores que producían grandes errores.

Esto, fue implementado de la siguiente manera:

```
# define model
model = Sequential()
model.add(LSTM(150, activation='relu', input_shape=(n_input, n_features)))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mse')
```

Ilustración 32. Definición del modelo. Fuente: el presente trabajo, 2021.

Con la función summary() de Keras fue posible ver un resumen del modelo diseñado.

```
Model: "sequential_7"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_7 (LSTM)                (None, 150)               91200
_________________________________________________________________
dense_8 (Dense)              (None, 1)                 151
=================================================================
Total params: 91,351
Trainable params: 91,351
Non-trainable params: 0
```

Ilustración 33. Resumen del modelo. Fuente: el presente trabajo, 2021.

6. Entrenamiento del Modelo:

En complemento, se procedió a ajustar el modelo al generador, y se establecieron las épocas para el entrenamiento (30 para este caso), las cuales, representan una ejecución completa a través de todos los datos de entrenamiento.

El entrenamiento expuso la efectividad del modelo durante las 30 épocas, evidenciando una pérdida mínima a medida que transcurría el proceso, como se puede observar en la ilustración 34.

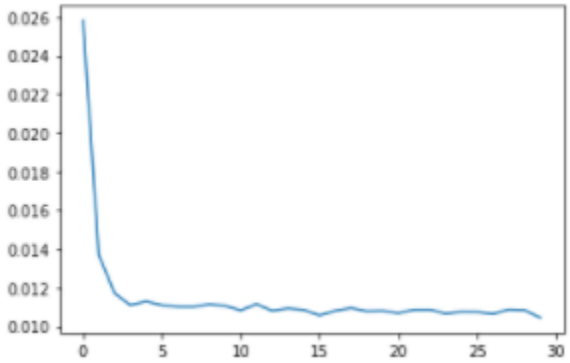


Ilustración 34. Registro de los valores de pérdida para los datos de entrenamiento. Fuente: el presente trabajo, 2021.

7. Evaluación del Modelo

Para evaluar, se compararon los resultados generados por el modelo con el conjunto de pruebas. Para ello, se tuvo en cuenta que los datos habían sido escalados en el rango de 0 a 1, por lo que, fue necesario aplicar la siguiente transformación inversa a los datos para poder compararlos en la medida original.

predictions = scaler.inverse_transform(predictions)

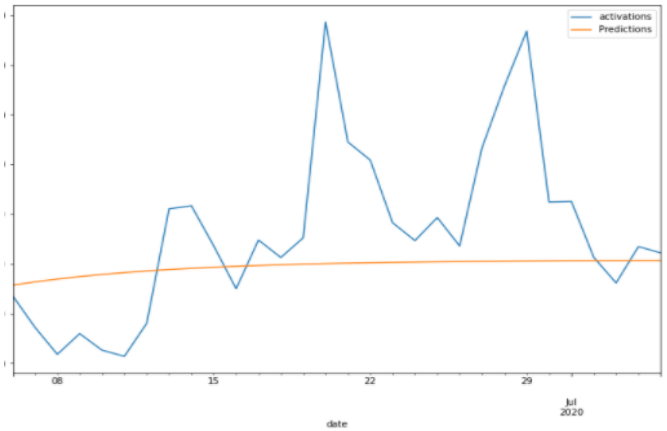


Ilustración 35. Comparación de las activaciones pronosticadas (Prediction) vs activaciones reales del conjunto de prueba (Test). Fuente: el presente trabajo, 2021.

Como se evidencia en la ilustración 35, se compararon los dos conjuntos, y se evidenció una leve curva aproximada a la media descrita por los datos de prueba, sin embargo no se tiene en cuenta fluctuaciones estacionales, pues el modelo RNN, no analiza los elementos con respecto a sus propiedades estadísticas como los anteriores modelos, sino que propone predicciones, con base al comportamiento de los valores previos, por lo que, obtener este resultado, es lógico, ya que RNN predice el día próximo con relación a las activaciones del día previo, según se describió con la función de generador de series temporales y el entrenamiento de la red (1 entrada produce 1 predicción).

Por otra parte, los errores MAE y RMSE fueron muy favorables, pese a que no brinda un ajuste diario óptimo, si lo hace en el contexto mensual, en que se tiene en cuenta el total de activaciones en los 30 días pronosticados.

8. Predicción

La predicción se implementó mediante el desarrollo de una serie de pasos:

- Se creó un arreglo para almacenar los valores pronosticados.
- Se obtuvieron los últimos n_input puntos del conjunto de entrenamiento y se adecuaron al formato requerido por RNN (el mismo formato del generador de series de tiempo).
- Se estableció un ciclo for para iterar el proceso de predicción del siguiente día con relación al día anterior, se almacena el valor pronosticado en el arreglo y se actualiza el batch (lote) para predecir los valores siguientes con este valor generado. Este proceso se repite las veces equivalentes a la cantidad de elementos del conjunto de pruebas más 30 días adicionales, para generar la predicción mensual requerida.

```
predictions = []

#Last n_input points from the training set
first_eval_batch = scaled_train[-n_input:]

#Reshape this to the format RNN wants (same format as Timse Series Generator)
current_batch = first_eval_batch.reshape((1, n_input, n_features))

for i in range(len(test)+30):

    # get prediction 1 time stamp ahead ([0] is for grabbing just the number instead of [array])
    current_pred = model.predict(current_batch)[0]

    # store APV activations prediction
    predictions.append(current_pred)

    # update batch to now include prediction and drop first value
    current_batch = np.append(current_batch[:,1:,:],[[current_pred]],axis=1)
```

Ilustración 36. Algoritmo de predicción. Fuente: el presente trabajo, 2021.

Con esto listo, se aplicó la transformación inversa a los datos escalados tal y como se hizo en la evaluación, y se capturaron los últimos 30 valores pronosticados del arreglo **predictions**.

2.8.2 Proceso B: Procesamiento de los Datos

Para la automatización de este proceso, se consideró la caracterización realizada en el punto **2.3.2**, donde se evidenció las funciones que deben desarrollarse en el procesamiento de los datos para producir información de alto valor empresarial.

En consonancia, su implementación fue posible gracias a la disponibilidad de herramientas como: Bigquery y Google Data Studio, las cuales, permitieron ejecutar consultas SQL personalizadas para extraer y configurar los conjuntos de datos de interés para brindar información acertada del servicio APV acorde a los KPIs planteados.

Al respecto, se requirió primeramente el estudio de los datos como se observó en el **punto 2.4,** para garantizar que aquellos que se han seleccionado sean relevantes, y segundo, la preparación de diversas fuentes de datos, mediante la configuración en Data Studio, con diversos conectores asociados directamente a Bigquery, para ejecutar consultas automáticamente una vez se han vinculado al proyecto que contiene los registros del servicio APV. De esta manera, el procesamiento de los datos queda en manos de las consultas SQL configuradas y su adaptación en Data Studio, con el fin de suministrar los registros necesarios para la construcción de los informes y Dashboards en la herramienta.

Entendiendo esto, se procede a observar las funciones inmersas en este proceso, y cómo se solventaron con la personalización de las consultas SQL en los servicios de Google Cloud.

- Valida

Esta función, está orientada a garantizar que los datos proporcionados sean correctos y relevantes. Se ha comentado que el DW, está dotado de datos confiables, y el hecho de configurar un conector directo con Bigquery, evita realizar una depuración sobre el conjunto de datos, ya que se vinculan automáticamente.

Ahora bien, para conformar las fuentes de datos, esta funcionalidad fue solventada, con relación a evitar que los datos se repitan e incentiven conteos equívocos de los registros, por lo que, una sentencia indispensable al respecto fue DISTINCT, la cual, era aplicada al número móvil de los usuarios, asociando en él, un identificador único para cada una de las personas en las diferentes transacciones almacenadas en el DW.

- Agrega

Las consultas fueron diseñadas para extraer datos de Bigquery, a partir de la combinación de múltiples campos y variables presentes en las tablas del Dataset. Las consultas construidas tienen como fundamento la sentencia SELECT que permite capturar los datos con base a condiciones estructuradas de diferentes locaciones en complemento con la sentencia FROM.

- Ordena

Con el fin de organizar los elementos en alguna secuencia y/o en diferentes conjuntos con un sentido lógico, se procedió a utilizar sentencias como GROUP BY y ORDER BY para

agrupar y ordenar elementos designados bajo la nominación propuesta por la sentencia AS. Fue frecuentemente usado, para organizar los datos en el formato requerido para el análisis como series temporales, y para garantizar la contabilidad de los registros ordenadamente acorde a las fechas de procesamiento de las transacciones del servicio.

- Clasifica

Debido a la necesidad de filtrar información y coordinar los registros del servicio, fue necesario dotar cada consulta de condiciones a través de sentencias WHERE, AND y OR, para extraer los elementos específicos que se necesitaban para cada uno de los informes.

Así, estas sentencias se tuvieron en cuenta para categorizar diversos parámetros como, el estado de activación, los tipos de usuario (Home/ Mobile), el tipo de plan (gratis, 50% o 100%), y muchas otros más, propicios para brindar una amplia visión del servicio en los reportes generados.

- Resume

Esta función, fue resuelta en algunos casos mediante la sentencia COUNT, para hacer un registro contable de los elementos y sintetizar un gran conjunto de estos en resultados generales que los relacionen. Fue complementado en gran medida con Data Studio, ya que fue posible implementar funciones matemáticas para activar cálculos sobre los campos en las fuentes de datos configuradas.

Estos cálculos, fueron configurados con relación al nombre del campo, por lo que, al estar incluido en las consultas programadas de las herramientas, se calculan automáticamente.

Con esta utilidad, fue posible llevar la contabilidad de los ingresos por el número de activaciones con relación al tipo de usuario (Mobile/Home) y obtener porcentajes con respecto a resultados totales de datos evitando la presentación de rellenos de información.

- Analiza

Involucra la recopilación, la organización, la interpretación y la presentación de los datos. Claramente, hasta este momento se ha presentado cada función con relación a una sentencia SQL que aparentemente la solventa, sin embargo, es notable, que no significan nada independientemente, pues su efectividad, sólo es posible en la medida que son parte de una lógica común a la implementación. Es por ello que, para cada reporte fue necesario configurar una fuente de datos, con una consulta específica, que desarrolla una lógica integrada con las funciones previamente mencionadas, acorde a la necesidad de información que se quería sustentar.

Con esto en mente, a continuación se presentan dos ejemplos de las consultas SQL construidas para procesar los datos en beneficio del conocimiento del servicio APV.
(Algunos campos han sido pixelados por cuestiones de confidencialidad)

```
SELECT COUNT(distinct([redacted])),Date(a.activation_date) as date
FROM `[redacted]` a
WHERE activation_date >= '2019-01-01'
GROUP BY date
ORDER BY date asc
```

Ilustración 37. Consulta personalizada 1. Fuente: el presente trabajo, 2021.

La consulta expuesta en la ilustración 37, fue construida simplemente para obtener el número de activaciones del servicio, desde la fecha de inicio hasta el momento actual, con el fin de obtener cada registro asociado al día en que dichos usuarios se activaron.

Para ello, se seleccionó la variable requerida, se aplicó un conteo sobre los registros y se extrajo la fecha de activación, la cual, teniendo un formato establecido, fue utilizada para agrupar y ordenar los elementos.

```
SELECT COUNT (distinct(a.[redacted])),Date(a.activation_date )
as date FROM `[redacted]` a
where a.cancelation_date is null
and id_home is null
and a.status = 'ACTIVE'
and a.activation_date
BETWEEN timestamp_sub(CURRENT_TIMESTAMP(), INTERVAL 30 DAY)
and CURRENT_TIMESTAMP()
and substr(a.[redacted],-10) not in (
    select distinct substr(b.[redacted],-10)
    FROM `[redacted]` b
    where b.cancelation_date is not null
    and b.status = 'CANCELLED'
    and b.[redacted] = a.[redacted]
    and b.cancelation_date >= a.activation_date
)
and substr(a.[redacted],-10) not in (
    select distinct substr(b.[redacted],-10)
    FROM `[redacted]` b
)
GROUP BY date
ORDER BY date asc
```

Ilustración 38. Consulta personalizada 2. Fuente: el presente trabajo, 2021.

La consulta expuesta en la ilustración 38, fue construida para extraer los usuarios Mobile que se han activado en el último mes. De igual forma, se realiza un conteo sobre los usuarios acorde al momento de activación en un rango determinado, entre la fecha actual y 30 días previos, se complementó especificando múltiples condiciones para garantizar que los usuarios seleccionados estén con estado activo, y no estén notificados con el estado cancelado.

La construcción de estas consultas, evidencia la integración de cada una de las funciones inmersas en el procesamiento de los datos, para: garantizar que los datos proporcionados sean correctos y relevantes (**validar**), combinar y suministrar varios datos (**agregar**), organizar los elementos en una secuencia y/o en diferentes conjuntos con un sentido lógico (**ordenar**), filtrar y categorizar los datos (**clasificar**), simplificar los datos a su valor fundamental (**resumir**) y por último, presentar la información interpretada (**analizar**).

Como se ha visto, el procesamiento de los datos fue implementado de esta forma y fue necesario diseñar consultas específicas para cada uno de los reportes expuestos en Data Studio, los cuales, se exponen en el siguiente capítulo.

2.9 Construcción de Mockups para la visualización de la información

Para desarrollar esta actividad, fue necesario estudiar artículos como: "**The role of the performance dashboard in the management of modern enterprises**" y "**Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts**" para aprender la forma en que la información debe ser presentada acorde a los objetivos del negocio en las empresas. Los artículos ayudaron a comprender múltiples conceptos y técnicas relevantes para la construcción de los reportes realizados, los cuáles, fueron bosquejados primeramente a través de Mockups, para servir de base como diseño y esquema funcional de los Dashboards que se implementarían en Data Studio.

Para el desarrollo de este prototipo visual, se utilizó la herramienta Adobe XD, que permite editar gráficos para la creación de interfaces de páginas web y de aplicaciones, enfocándose en la experiencia del usuario al navegar, con un rango mínimo de error y en el menor tiempo posible.

Así, los Mockups realizados se adjuntaron en la carpeta Anexos (**Mockups/dashboardMockups.xd** y **Mockups/dashboardMockups.pdf**), donde se expone cada uno de los elementos configurados para presentar la información procesada. Algunos de estos elementos, se destacan a continuación:

- Contadores: se usan para resumir la información y tener un registro numérico preciso de algunas métricas. En el proyecto, fueron usados para brindar una lectura rápida de valores totales como: recuento de activaciones/cancelaciones, porcentajes, número de usuarios activos, cantidad de ingresos en COP.

Ilustración 39. Ejemplo Contador. Fuente: el presente trabajo, 2021.

- Tablas: usadas para asociar información cuantificable a dimensiones espaciales o temporales. En la visualización del servicio APV, se implementaron para organizar la visibilidad de los datos de activación/cancelación con relación a la fecha en que se procesaron.

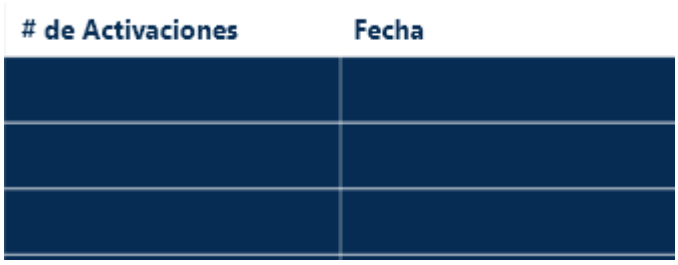


Ilustración 40. Ejemplo Tabla de Registro de activaciones. Fuente: el presente trabajo, 2021.

- Filtros: permiten limitar la visibilidad de la información con respecto a una condición. Lo particular de este elemento, es que las selecciones calculan los datos que aparecen en cualquier informe en el tablero que está asociado con dicho filtro, sin cambiar las definiciones del informe. En el proyecto se implementaron filtros de tiempo, para seleccionar un rango de fechas y poder observar información específica con mayor detalle.

Ilustración 41. Ejemplo Filtro de fecha. Fuente: el presente trabajo, 2021.

- Serie Temporal: describe puntos de datos indexados en orden de tiempo. Se utilizaron este tipo de gráficas para producir un registro contable de la información diaria de métricas como el número de usuarios, el número de activaciones y el número de cancelaciones.

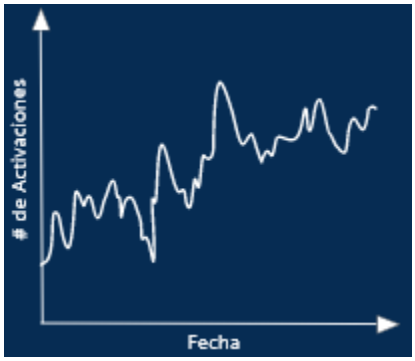


Ilustración 42. Ejemplo Gráfico serie temporal del registro diario de activaciones. Fuente: el presente trabajo, 2021.

- Barras: se usan para comparar elementos clasificados. En el diseño, se construyeron estos gráficos para brindar una visibilidad comparativa de las activaciones y la cantidad de usuarios respecto a su clasificación mensual.

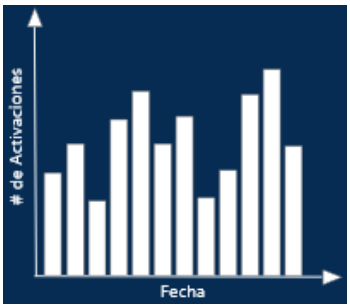


Ilustración 43. Ejemplo Gráfico de barras de activaciones mensuales. Fuente: el presente trabajo, 2021.

- Texto: se utilizó como elemento complementario para orientar la visibilidad de los reportes y etiquetar la representación de los gráficos.

Ilustración 44. Ejemplo Etiqueta texto del registro diario de activaciones. Fuente: el presente trabajo, 2021.

La tabla 10, muestra algunos de estos elementos utilizados para construir los reportes y dashboards del servicio.

Tabla 10. Tabla de componentes utilizados para la visualización de los datos. Fuente: el presente trabajo, 2021.

Tipo de elemento	Número de Categorías	Variables numéricas	Propósito	Nivel de interpretación	Aplicación
Contador	1	1	Resumir métricas (total)	Fácil	Recuento de activaciones/cancelaciones, porcentajes, número de usuarios activos, cantidad de ingresos en COP
Tabla	Múltiple	Múltiple	Resumen +clasificación	Fácil	Organizar datos de activación/cancelación con relación al tiempo
Filtro	-	-	Seleccionar info específica	Medio	Filtrar información en un rango de fecha
Serie Temporal	Múltiple	1 + variable de tiempo	Describir registros con relación al tiempo	Fácil	Describir el número de usuarios, número de activaciones, número de cancelaciones.

Gráfico de barras	Múltiple	1	Comparación de registros	Fácil	Comparar la cantidad de activaciones y usuarios respecto a su clasificación mensual
Texto	-	-	Etiquetar gráficos	Fácil	Etiquetar la representación de los gráficos

2.10 Diseñar Dashboards en Data Studio

En este punto, la actividad se simplifica gracias al desarrollo de las actividades previas. Por esta razón, sólo conviene destacar los procesos desarrollados y su integración en la herramienta de BI a través de la exposición de los pasos implementados.

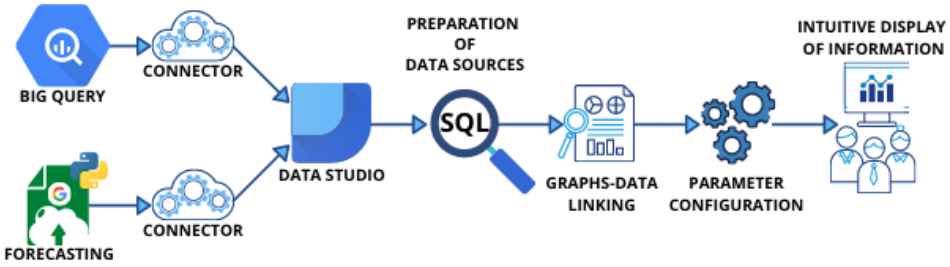


Ilustración 45. Construcción de Dashboards en Data Studio. Fuente: el presente trabajo, 2021.

La ilustración 45 explica los 6 procesos inmersos en la construcción de los reportes en Data Studio:

1. Establecimiento del conector Bigquery (registros reales del servicio APV) para importar los datos a Data Studio.
2. Establecimiento del conector de Hojas de cálculo de Google (registros pronosticados) para importar los datos a Data Studio
3. Preparación de las fuentes con la personalización de consultas SQL para extraer información específica.
4. Vinculación de elementos visuales con las fuentes de datos y filtros.
5. Configuración de parámetros en las gráficas y cálculo de métricas con los campos de datos.
6. Organización de los elementos conforme a los Mockups realizados para entregar información de alto valor de forma intuitiva.

Es relevante manifestar que, la construcción a detalle de los dashboards y reportes del servicio APV no se pueden especificar a detalle, debido a que el manejo de los datos está sujeto a la confidencialidad con la empresa. Sin embargo, en el próximo capítulo se expone la presentación visual de la información procesada del servicio APV, como resultado de la validación constante con el asesor de la empresa, en cumplimiento con las exigencias estudiadas de BI para la optimización de la toma de decisiones.

CAPÍTULO III - RESULTADOS

El mecanismo diseñado, surgió como respuesta al contexto de las actividades desarrolladas en el área VP Digital con respecto a la necesidad de información del servicio APV, muchas de las cuales, al hacerse repetitivamente, consumían tiempo de efectividad en el trabajo del talento humano.

Así, se ha establecido hasta el momento, el mecanismo propuesto para llevar a cabo este proyecto, en adición, se desarrolló una lógica (**punto 2.7**) que evidenciaba la forma en que los dos procesos identificados serían automatizados, mediante la aplicación de Business Intelligence y herramientas de analítica de datos, con el fin de obtener resultados satisfactorios.

Primeramente, es necesario tener en cuenta el requerimiento señalado en el punto **2.2.2**, para verificar su coherencia con los resultados obtenidos. Además, es fundamental incluir para la observación de este capítulo, los KPIs definidos, ya que son las métricas clave, que sustentan la necesidad de información asociada al servicio APV.

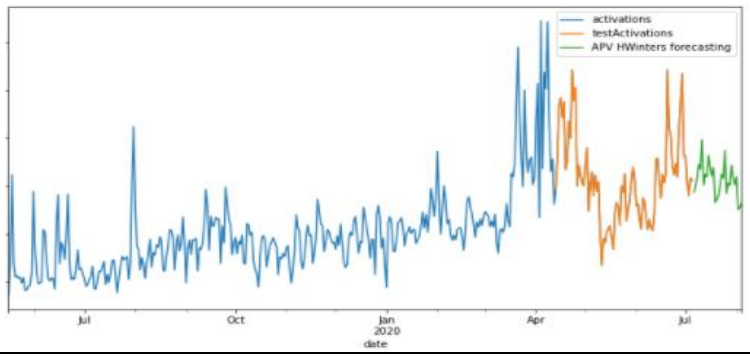
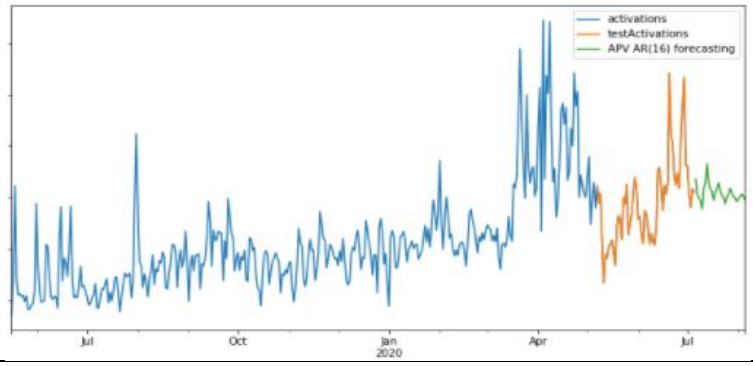
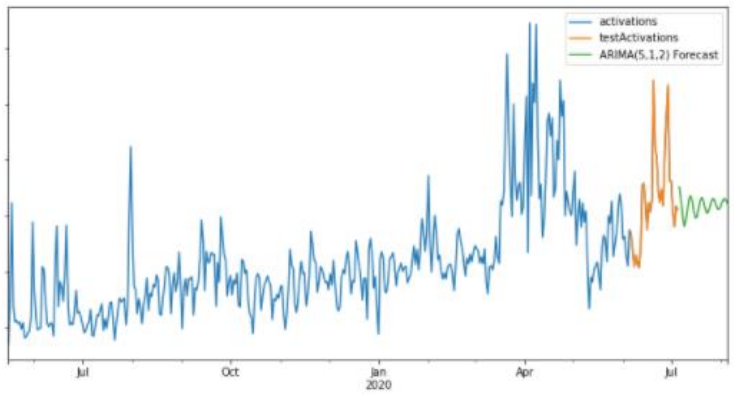
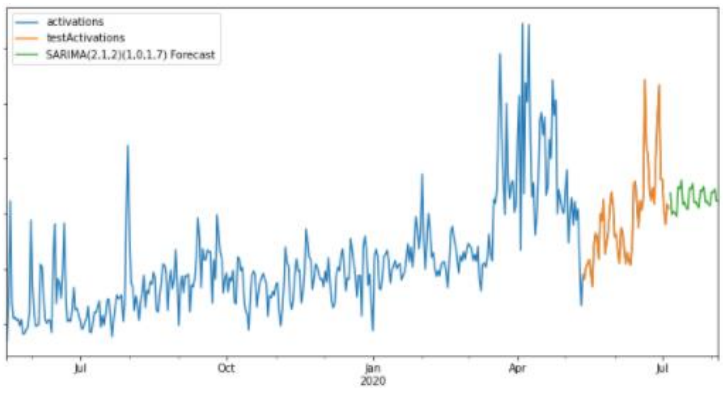
Con esto en mente, y en consonancia con las etapas 4 y 5 de Business Intelligence descritas al inicio del capítulo II, los resultados obtenidos son:

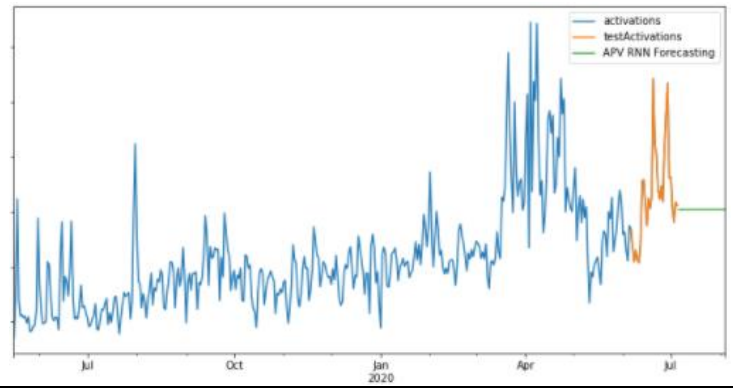
3.1 Automatización Modelamiento predictivo

Acorde a cada uno de los modelos implementados, se obtuvieron los resultados expuestos en la tabla 11, la cual asocia gráficamente cada una de las predicciones generadas por cada modelo y los errores MAE y RMSE evaluados.

Los 5 modelos aportan a la representación de los pronósticos de las activaciones del servicio en los 30 días posteriores, y fueron incluidos en el procesamiento (proceso 2) para el desarrollo del Dashboard.

Tabla 11 Pronóstico de activaciones de los modelos predictivos implementados. Fuente: el presente trabajo, 2021.

Modelo	MAE	RMSE	Predicción
Holt Winters	**330,287**	**408,112**	
AR (16)	**363,297**	**455,017**	
ARIMA (1,1,1)	**398,846**	**516,045**	
SARIMA (2,1,2)(1,0,1,7)	**370,678**	**497,658**	

RNN	**317,856**	**438,832**	

Acorde a los registros producidos, y el análisis implícito de cada modelo, se resalta la efectividad de los modelos Holt Winters, SARIMA y RNN, debido a que, respecto a los dos primeros, se tiene en cuenta el efecto estacional, fundamental para describir las fluctuaciones de los datos del servicio, como se puede observar en las gráficas, y por otra parte, el modelo RNN, es capaz de asimilar los datos de entrada, y evaluarlos simulando memoria a corto plazo, para generar resultados con base al entrenamiento realizado por los valores de las activaciones pasadas en la red neuronal conformada.

Adicionalmente, dichos modelos destacan su supremacía sobre los otros dos, por la efectividad evaluada a través de las métricas de error MAE y RMSE, las cuales, van más acorde a la descripción integral del conjunto de datos completo como se evidencia en los scripts, obteniendo un valor de MAE bajo y un valor de RMSE por debajo o muy cercano a la desviación estándar del conjunto de datos total (std=495,214) como es recomendable para este tipo de aplicaciones, según la sección 8. General Forecasting Models, en el curso de Udemy "*Python for Time Series Analysis and Forecasting"*.

Cabe destacar que, los Scripts anexados de cada modelo, demuestran la implementación del conjunto de funciones y algoritmos para resolver las predicciones de las activaciones del servicio automáticamente, sin necesidad de afectar los resultados manualmente, ya que, los modelos asimilan los datos, y conforme a la interpretación que se genera en el análisis de los registros de entrada, es posible producir los pronósticos respectivos.

3.2 Automatización Procesamiento de datos – Visualización de la información

Los dashboards construidos en Data Studio, son una representación gráfica de los datos del servicio APV, los cuales han sido procesados y configurados para actualizarse automáticamente. Dichos dashboards, fueron organizados en 6 páginas acorde a la necesidad de información, considerando los KPIs del negocio y las categorías implícitas del servicio, como se evidencia a continuación, donde cabe aclarar, se ocultaron los datos, debido a los asuntos de confidencialidad con la empresa.

Ahora bien, es propicio describir con mayor detalle cada una de estas páginas:

3.2.1 Registro diario de Activaciones

Corresponde al procesamiento de las activaciones del servicio APV, se brinda una visualización del registro diario de los datos en una serie temporal, para evidenciar la tendencia y comportamiento de los datos, una tabla que facilita el número de activaciones por día (fijado por defecto para ver los últimos 15 días), dos tarjetas de resultados (total y semanal) y un filtro de fecha para ver la información acorde a un periodo determinado.

En las pruebas, se sugirió fijar 2 límites (línea verde y roja) en la gráfica, para dar una visión intuitiva del rango óptimo de las activaciones, y observar posibles alertas en caso de obtener registros inferiores al umbral rojo.

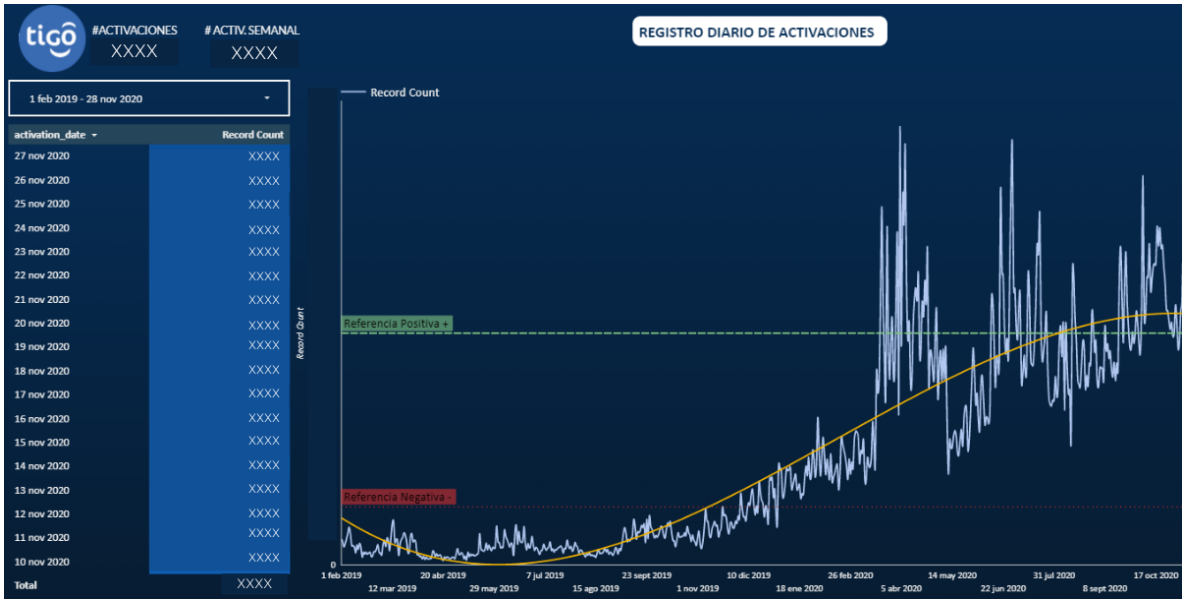


Ilustración 46. Página 1, Registro diario de Activaciones. Fuente: el presente trabajo, 2021.

3.2.2 Registro diario de Cancelaciones

Corresponde al procesamiento de las cancelaciones del servicio APV, se brinda una visualización en una serie temporal de los usuarios que se retiraron, una tabla que asocia el número de cancelaciones por día (fijado por defecto para ver los últimos 15 días), dos tarjetas de resultados (total y semanal) y un filtro de fecha para ver la información acorde a un periodo determinado.
A diferencia del anterior reporte, esta gráfica sólo posee 1 umbral para alertar de la deserción de uso del servicio, en un periodo determinado.

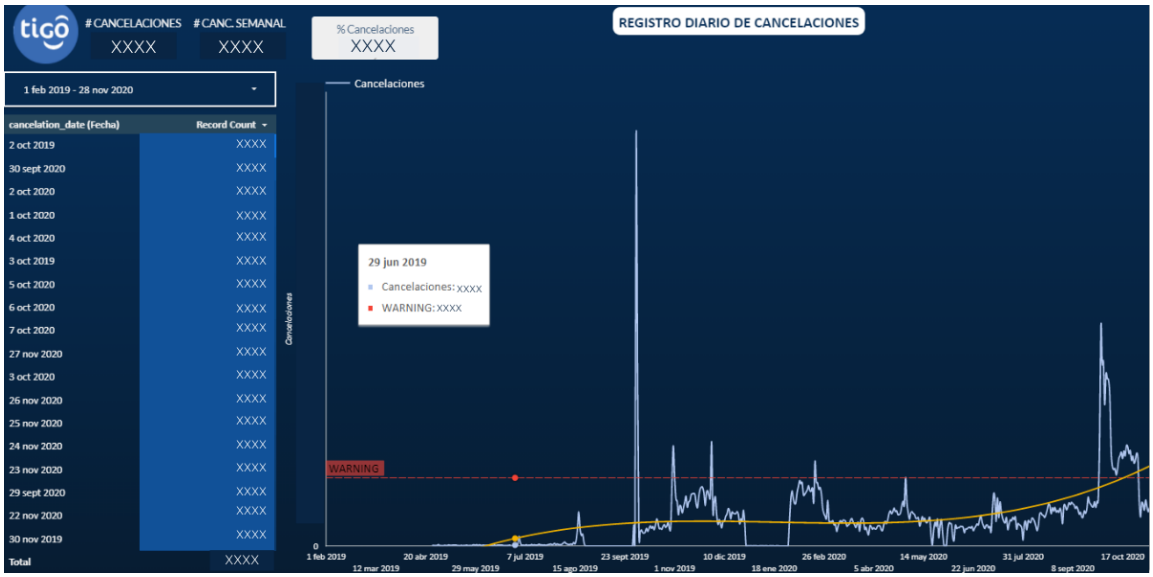


Ilustración 47. Página 2, Registro diario de Cancelaciones. Fuente: el presente trabajo, 2021.

3.2.3 Contabilidad y Registros ADDON/BUNDLE

El presente dashboard representa los ingresos del servicio APV, con relación al tipo de usuario Mobile o Home, se detallan 3 tarjetas de resultados para cada tipo, una que procesa los ingresos totales, y las 2 restantes, asociadas al tipo de plan (pagando 50% o 100%). Por otro lado, se procesan las activaciones mensuales del servicio respecto al modo de suscripción (Addon o Bundle).

En complemento a ello, se calcula el porcentaje de activaciones y cancelaciones, para observar constantemente y a grandes rasgos, el grado de aceptación (usabilidad) del servicio APV.

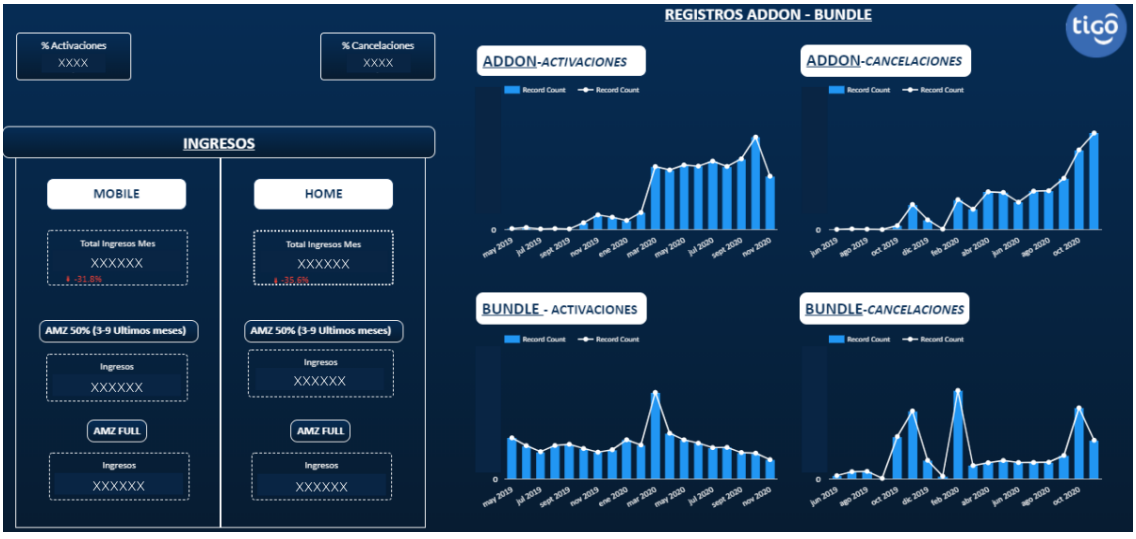


Ilustración 48. Página 3, Contabilidad y Registros ADDON/BUNDLE. Fuente: el presente trabajo, 2021.

3.2.4 Usuarios Activos MOBILE/HOME (Registro diario)

El presente dashboard brinda una visualización de las activaciones respecto al tipo de usuario (Home/Mobile), detallando respectivamente, el registro de los datos con relación al plan (Gratis, pago 50%, pago 100%) y el número de usuarios nuevos que se han vinculado al servicio en el último mes. Los reportes que conforman el dashboard, integran gráficas de series temporales para facilitar una interpretación del comportamiento de los datos, y tarjetas de resultados que exponen las activaciones totales del momento, acorde al filtro de fecha implícito que denota la antigüedad del usuario en el servicio.

Ilustración 49. Página 4, Usuarios activos MOBILE/HOME (Registro diario). Fuente: el presente trabajo, 2021.

3.2.5 Usuarios Activos MOBILE/HOME (Registro mensual)

Para este caso, fue aconsejable por el asesor, re muestrear los datos del dashboard anterior, para brindar información mensual de las activaciones Mobile/Home, debido a que, muchas veces los expertos de la empresa requieren información más simplificada y no tan específica. En adición, este dashboard al igual que el anterior, contiene una modificación añadida en las pruebas, un gráfico circular que compara la cantidad de activaciones anualmente, para observar el crecimiento de los usuarios entre un año y otro.

Ilustración 50. Página 5, Usuarios activos MOBILE/HOME (Registro mensual). Fuente: el presente trabajo, 2021.

3.2.6 Predicción de Activaciones

Como es de esperar, la visualización también incluye el procesamiento de los datos pronosticados, el dashboard presenta el desglose diario de cada uno de las predicciones con relación al modelo respectivo. De igual forma, se publica el registro diario de las activaciones reales, en complemento con las activaciones pronosticadas para comunicar el ajuste que tienen los modelos con relación a las transacciones ya realizadas.

Por otro lado, se expone un resultado obtenido con la librería Facebook Prophet, para analizar la estacionalidad de los datos, evidenciando cuáles son los días en la semana, en que las personas son más propensas a suscribirse al servicio.

Precisa mencionar que, en las pruebas fue aconsejable por el asesor, documentar en la parte inicial, la usabilidad del dashboard y los procesos realizados para gestionar esta visualización, con el fin de que las personas que tengan acceso al informe, puedan interpretar la información aquí descrita con mayor facilidad.

Ilustración 51. Página 6, predicción de activaciones. Fuente: el presente trabajo, 2021.

3.3 Criterios de comparación

Con base en la implementación y los resultados previamente descritos, es relevante evaluar la propuesta de automatización desarrollada, con relación a los criterios de comparación sustentados en el punto **2.3** para cada uno de los procesos (**2.3.1**) y (**2.3.2**). En consecuencia, se describen los resultados en la tabla 12 y la tabla 13 respectivamente.

Tabla 12. Métricas de evaluación para el proceso 1 susceptible. Fuente: el presente trabajo, 2021.

Funcionalidad	*Converge a valores lógicos*	Si, las predicciones satisfacen las expectativas en el rango esperado acorde al conocimiento del negocio. Además, los resultados generados son razonables con relación a las características estadísticas como la media y la desviación estándar (detalladas en los Scripts) intrínsecas del conjunto de datos.
	Ajuste lineal y no lineal	Si, los resultados son el producto de la reacción a las fluctuaciones de cambio de tendencia y estacionalidad de los datos. El análisis de los modelos se adapta a la naturaleza descrita por los registros de entrenamiento.
	Versátil con grandes volúmenes de datos	Si, los modelos se adaptan al volumen creciente de los datos; entre más datos, puede asegurarse mayor precisión en la predicción, ya que más registros pueden vincularse al set de entrenamiento y set de prueba de los modelos.

	Asimila los datos	Si, los modelos y las funciones claves implícitas en ellos, analizan el comportamiento de los datos, comprenden los patrones inmersos en los registros y evalúan el ajuste de las predicciones para adaptar los resultados a valores lógicos en el futuro.
Criterios base	*Tipo de Ejecución*	Automático
	Tiempo	6 min: Importar los datos y ajustar algunos parámetros.
	Validación de la información	Si, se verifica la fiabilidad de los resultados generados mediante la evaluación del modelo, para corroborar que los datos pronosticados sean lógicos.
	Amplitud de la información	Se focaliza en la necesidad de información de las activaciones del servicio, los modelos brindan un rango de hasta 45 valores de predicción confiable. Sin embargo, para la aproximación requerida se satisface con 30 valores, los cuales, claramente garantizan precisión en la ventana de tiempo objetivo.
	Visualización	Los valores pronosticados son presentados en formato numérico y gráfico para facilitar su interpretación. Además, cuando se vinculan en el informe en la herramienta de BI, es posible filtrar su consulta con respecto al tiempo.

Tabla 13. Métricas de evaluación para el proceso 2 susceptible. Fuente: el presente trabajo, 2021.

Funcionalidad	*Valida*	✓
	Agrega	✓
	Ordena	✓
	Analiza	✓
	Clasifica	✓

	Resume	✓
Criterios base	*Tipo de Ejecución*	Automático
	Tiempo	0 min
	Validación de la información	Si, la información es procesada directamente desde el Data Warehouse, no hay descargas manuales de los datos que afecten la fiabilidad de la información, las consultas SQL personalizadas además de extraer automáticamente los datos, garantizan que sean correctos y originales desde la fuente.
	Amplitud de la información	Se focaliza en las métricas clave definidas para el negocio, en complemento, brinda una amplia visibilidad de variables adicionales que pueden soportar información de alto valor para la toma de decisiones.
	Visualización	Los dashboards construidos en Data Studio, contienen el procesamiento de los datos vinculado a múltiples fuentes configuradas; de esta manera, es posible brindar una representación numérica y gráfica de la información, e incluso la posibilidad de filtrar consultas respecto al tiempo.

Los resultados expuestos, fueron evaluados y aprobados en la empresa, como se puede observar en el anexo (CartaDeAprobaciónTIGO.pdf), en adición, la tabla 14 resume algunas de las correcciones/sugerencias solventadas para las etapas clave de la implementación,

Tabla 14. Resumen de las etapas evaluadas en la implementación. Fuente: el presente trabajo, 2021.

Etapas evaluadas en la implementación	**Correcciones - Sugerencias**
Identificación de los procesos susceptibles de automatización	- Seleccionar los procesos con base al estudio de los reportes del momento. - Mejorar la operatividad de los procesos que se realizan manualmente conforme a la lógica de BI llevada a cabo en el área.
Definición de la lógica de automatización	- Integración de recursos y tecnologías existentes en la empresa para mejorar los procesos previamente identificados. - Diseño de un esquema estructurado que defina los pasos y relaciones entre cada proceso de automatización de inicio a fin. - Sustitución de herramienta de BI, Google Data Studio en vez de Tableau.

Modelamiento predictivo	- Ajuste de parámetros y funciones para mayor precisión. - Selección de los modelos predictivos más confiables. - Ajustar la generación de los resultados pronosticados en un formato versátil (array ->.csv) para importar en la herramienta de BI.
Procesamiento de datos	- Corrección en la preparación de conectores de datos (Bigquery, Hojas de cálculo Google). - Necesidad de generar información en tiempo real. - Necesidad de incluir más variables de información. - Rediseño de consultas personalizadas para extraer y procesar datos.
Diseño de reportes y dashboards	- Presentación de la información en el formato adecuado (gráficas, tablas, filtros..) - Proveer un diseño intuitivo para facilitar la lectura de la información. - Ajuste de la temática de colores afín a la empresa.
Visualización integral de la información	- Presentación de la información sustancial del negocio conforme a la necesidad de conocimiento y KPIs del negocio. - Integración de la información (histórica, presente y futura) relevante para el negocio. - Corregir el orden de las páginas. Vista: Información general a específica.

Las retroalimentaciones comunicadas por el asesor de la empresa, fueron tenidas en cuenta en los múltiples procesos de la ejecución como se fue describiendo. De esta manera, el proyecto cumple con los objetivos planteados, en lineamiento con las actividades requeridas por Tigo para automatizar los reportes y dashboards del servicio APV en el área VP Digital.

CAPÍTULO IV - CONCLUSIONES

- ✓ Los datos son la materia prima más importante de las compañías para interpretar el contexto comercial y funcional en el que van operando. En consonancia, Tigo había adoptado Business Intelligence para extraer información de alto valor a partir de ellos, y el presente proyecto fue el complemento a los procesos ya desarrollados en el área VP Digital respecto al servicio APV, aportando a la ejecución del procesamiento de los datos, el modelamiento predictivo y la visualización de la información por medio de la automatización, y no sólo reduciendo los retardos en la generación de los reportes, sino también brindando mayor amplitud en la información suscitada para la toma de decisiones.

- ✓ La lógica de automatización propuesta, logró integrar satisfactoriamente las tecnologías, el conocimiento estudiado y los datos del servicio APV para soportar el rendimiento automático esperado de cada uno de los procesos identificados para la generación de reportes y dashboards conforme a las métricas clave de desempeño (KPIs) del negocio, desempeñando una mejora importante en comparación con el modo de ejecución manual que se tenía al comienzo.

- ✓ A partir de los datos almacenados en el repositorio de la empresa, fue posible transformarlos en información mediante técnicas de procesamiento y analítica de datos; en complemento, el esquema de visualización propuesto para presentar adecuadamente dicha información, fue satisfactorio para facilitar el entendimiento de las variables, dimensiones y registros del servicio APV al talento humano encargado de la toma de decisiones.

- ✓ Los criterios de comparación definidos para evaluar los procesos implementados conforme a la lógica de BI descrita en el proyecto, validaron el mejoramiento de la productividad en la interpretación de la información del servicio APV con relación a la que se tenía antes de la presente propuesta de automatización, en la medida que la información dispuesta, no sólo se genera más rápido, sino también, es más amplia y visualmente más precisa para el entendimiento del negocio.

REFERENCIAS

[1] K. Krishnan, «1 - Big Data introduction», en *Building Big Data Applications*, K. Krishnan, Ed. Academic Press, 2020, pp. 1-16.

[2] «1124-C-05-02-2018-1.pdf». Accedido: nov. 21, 2020. [En línea]. Disponible en: https://www.asobancaria.com/wp-content/uploads/2018/02/1124-C-05-02-2018-1.pdf.

[3] M. Bala, O. Boussaid, y Z. Alimazighi, «Big-ETL: Extracting-Transforming-Loading Approach for Big Data», p. 7.

[4] K. D. Foote, «A Brief History of Business Intelligence», *DATAVERSITY*, sep. 14, 2017. https://www.dataversity.net/brief-history-business-intelligence/ (accedido nov. 21, 2020).

[5] «2018_2_6.pdf». Accedido: nov. 21, 2020. [En línea]. Disponible en: http://economic.upit.ro/repec/pdf/2018_2_6.pdf.

[6] «Tigo | Planes de Internet, Televisión y Móvil para tí y tu hogar. | Conectados siempre». https://www.tigo.com.co/ (accedido nov. 23, 2020).

[7] «Tigo es el primer operador que tiene disponible Amazon Prime Video para usuarios fijos y móviles». http://saladeprensa.une.com.co/index.php/2140-tigo-es-el-primer-operador-que-tiene-disponible-amazon-prime-video-para-usuarios-fijos-y-moviles (accedido nov. 23, 2020).

[8] «the_big_data_payoff-turning_big_data_into_business_value.pdf». Accedido: nov. 23, 2020. [En línea]. Disponible en: https://www.capgemini.com/wp-content/uploads/2017/07/the_big_data_payoff-turning_big_data_into_business_value.pdf.

[9] «Definition of Big Data - Gartner Information Technology Glossary», *Gartner*. https://www.gartner.com/en/information-technology/glossary/big-data (accedido nov. 25, 2020).

[10] «¿Qué es el Big Data? | Cloud Big Data Solutions», *Google Cloud*. https://cloud.google.com/what-is-big-data?hl=es (accedido nov. 25, 2020).

[11] M. Younas, «Research challenges of big data», *Serv. Oriented Comput. Appl.*, 2019, doi: 10.1007/s11761-019-00265-x.

[12] Y. Riahi, «Big Data and Big Data Analytics: Concepts, Types and Technologies», vol. 5, pp. 524-528, 2018, doi: 10.21276/ijre.2018.5.9.5.

[13] Savo Stupar, Mirha Bičo Ćar, y Elvir Šahić, «The Importance of Implementing Big Data Analytics Concepts in Companies», en *Handbook of Research on Integrating Industry 4.0 in Business and Manufacturing*, Isak Karabegović, Ahmed Kovačević, Lejla Banjanović-Mehmedović, y Predrag Dašić, Eds. Hershey, PA, USA: IGI Global, 2020, pp. 53-74.

[14] «What does Business Intelligence mean to you?» https://www.oracle.com/co/what-is-business-intelligence.html (accedido nov. 25, 2020).

[15] N. Dedić y C. Stanier, «Measuring the success of changes to Business Intelligence solutions to improve Business Intelligence reporting», *J. Manag. Anal.*, vol. 4, n.º 2, pp. 130-144, 2017, doi: 10.1080/23270012.2017.1299048.

[16] «First of All, Understand Data Analytics Context and Changes», en *Big Data Analytics for Entrepreneurial Success*, Hershey, PA, USA: IGI Global, 2019, pp. 92-124.

[17] R. Toledo, «7 características esenciales de una solución de Business Intelligence». https://www.grupocibernos.com/blog/business-intelligence/7-caracteristicas-esenciales-de-una-solucion-de-business-intelligence (accedido nov. 25, 2020).

[18] M. E. C. Inocente y J. I. G. Caporal, «IMPLEMENTACIÓN DE BUSINESS INTELLIGENCE PARA MEJORAR LA EFICIENCIA DE LA TOMA DE DECISIONES EN LA GESTIÓN DE PROYECTOS», p. 73.
[19] «¿Qué es Business Intelligence?» https://www.sinnexus.com/business_intelligence/ (accedido nov. 25, 2020).
[20] R. Desai, «Top 10 Business Intelligence Tools of 2020», *Medium*, ago. 26, 2020. https://towardsdatascience.com/top-10-business-intelligence-tools-of-2020-be5bfe22a9b (accedido nov. 25, 2020).
[21] «Gartner Reprint». https://www.gartner.com/doc/reprints?id=1-1YBTIWVR&ct=200211&st=sb (accedido nov. 25, 2020).
[22] G. Inc, «Google Data Studio review in Business Intelligence (BI) Tools», *Gartner*. https://www.gartner.com/market/analytics-business-intelligence-platforms/vendor/google/product/google-data-studio/review/view/1068224 (accedido nov. 25, 2020).
[23] E. D. de la Iglesia, «Los datos del Big Data». https://www.campusbigdata.com/big-data-blog/item/111-datos-big-data (accedido nov. 25, 2020).
[24] «Datos, información, conocimiento». https://www.sinnexus.com/business_intelligence/piramide_negocio.aspx (accedido nov. 25, 2020).
[25] «Oxford Languages and Google - Spanish | Oxford Languages». https://languages.oup.com/google-dictionary-es/ (accedido nov. 25, 2020).
[26] «Para qué sirve un "dashboard"». https://www.expansion.com/economia-digital/protagonistas/2016/11/12/5824c400e5fdea752d8b45d3.html (accedido nov. 25, 2020).
[27] S. E. Dragomirescu y D. C. Solomon, «THE ROLE OF THE PERFORMANCE DASHBOARD IN THE MANAGEMENT OF MODERN ENTERPRISES», *Stud. Sci. Res. - Econ. Ed.*, vol. 0, n.º 18, 2013, doi: 10.29358/sceco.v0i18.221.
[28] A. Pérez, «Teamwork and leadership styles, their relationship with decision making in the organization: a review», p. 11.
[29] «BigQuery: Almacén de datos en la nube», *Google Cloud*. https://cloud.google.com/bigquery?hl=es (accedido nov. 25, 2020).
[30] «Google Data Studio: ¿Qué es y cómo utilizarlo? [Tutorial]», *Bloo Media*, ago. 08, 2019. https://bloo.media/blog/tutorial-google-data-studio/ (accedido nov. 25, 2020).
[31] L. Toro, «Anaconda Distribution: La Suite más completa para la Ciencia de datos con Python», *Desde Linux*, sep. 14, 2017. https://blog.desdelinux.net/ciencia-de-datos-con-python/ (accedido nov. 25, 2020).
[32] R. Python, «Jupyter Notebook: An Introduction – Real Python». https://realpython.com/jupyter-notebook-introduction/ (accedido nov. 25, 2020).
[33] L. Toro, «Jupyter notebook: documenta y ejecuta código desde el navegador», *Desde Linux*, sep. 21, 2017. https://blog.desdelinux.net/jupyter-notebook/ (accedido nov. 25, 2020).
[34] «Welcome to Python.org», *Python.org*. https://www.python.org/about/ (accedido nov. 25, 2020).
[35] «How to Create an ARIMA Model for Time Series Forecasting in Python». https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/ (accedido nov. 25, 2020).
[36] E. Universitat Politècnica de València, «Universitat Politècnica de València», *Ing. Agua*, vol. 18, n.º 1, p. ix, sep. 2014, doi: 10.4995/ia.2014.3293.

[37] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, y P.-A. Muller, «Deep learning for time series classification: a review», *Data Min. Knowl. Discov.*, vol. 33, 2019, doi: 10.1007/s10618-019-00619-1.
[38] B. Wang, «Deep Learning for Time Series Analysis», p. 24.
[39] L. A. Sabatés, «La revisión de la literatura científica»:, p. 22.
[40] M. E. Monroy, J. L. Arciniegas, y J. C. Rodríguez, «Recuperación de Arquitecturas de Software: Un Mapeo Sistemático de la Literatura», *Inf. Tecnológica*, vol. 27, n.º 5, pp. 201-220, 2016, doi: 10.4067/S0718-07642016000500022.
[41] T. Kolajo, O. Daramola, y A. Adebiyi, «Big data stream analysis: a systematic literature review», *J. Big Data*, vol. 6, n.º 1, p. 47, jun. 2019, doi: 10.1186/s40537-019-0210-7.
[42] S. Balina, R. Zuka, y J. Krasts, «Opportunities for the Use of Business Data Analysis Technologies», *Econ. Bus.*, vol. 28, 2016, doi: 10.1515/eb-2016-0003.
[43] A. Maté, J. Trujillo, y J. Mylopoulos, «Specification and Derivation of Key Performance Indicators for Business Analytics: A Semantic Approach», *Data Knowl. Eng.*, vol. 108, 2017, doi: 10.1016/j.datak.2016.12.004.
[44] G. Martínez, «Optimización del mantenimiento industrial mediante técnicas BI, aplicación de un cuadro de mandos integral», p. 53.
[45] S. Joshi, G. S. Jayendran, y R. Dalal, «Transforming Telecom Business: Scaling the Shift Using Predictive Analytics», *Indian J. Sci. Technol.*, vol. 8, n.º S4, p. 34, feb. 2015, doi: 10.17485/ijst/2015/v8iS4/60362.
[46] S. Bāliņa, R. Žuka, y J. Krasts, «Opportunities for the Use of Business Data Analysis Technologies», *Econ. Bus.*, vol. 28, n.º 1, pp. 20–25, 2016, doi: 10.1515/eb-2016-0003.
[47] M. R. Llave, «Data lakes in business intelligence: reporting from the trenches», *CENTERIS 2018 - Int. Conf. Enterp. Inf. Syst. ProjMAN 2018 - Int. Conf. Proj. Manag. HCist 2018 - Int. Conf. Health Soc. Care Inf. Syst. Technol. CENTERISProjMANHCist 2018*, vol. 138, pp. 516-524, ene. 2018, doi: 10.1016/j.procs.2018.10.071.
[48] L. SERBANESCU, «ANALYSIS, REPORTING AND FORECASTING WITH QLIKVIEW», *Sci. Bull. - Econ. Sci.*, vol. 17, n.º 2, pp. 66–71, 2018.
[49] C. Rajesh, Y. Jain, y J. Jayaram, «Airport Trends Analytics Engine using the ARIMA Model», *Int. J. Eng. Technol.*, vol. 7, p. 239, 2018, doi: 10.14419/ijet.v7i3.12.16033.
[50] G. Silahtaroğlu y N. Alayoglu, «Using or Not Using Business Intelligence and Big Data for Strategic Management: An Empirical Study Based on Interviews with Executives in Various Sectors», *12th Int. Strateg. Manag. Conf. ISMC 2016 28-30 Oct. 2016 Antalya Turk.*, vol. 235, pp. 208-215, nov. 2016, doi: 10.1016/j.sbspro.2016.11.016.
[51] D. Appelbaum, A. Kogan, M. Vasarhelyi, y Z. Yan, «Impact of business analytics and enterprise systems on managerial accounting», *Int. J. Account. Inf. Syst.*, vol. 25, pp. 29-44, may 2017, doi: 10.1016/j.accinf.2017.03.003.
[52] I. C. Baierle, M. A. Sellitto, A. F. Habekost, R. Frozza, y J. L. Schaefer, «An artificial intelligence and knowledge-based system to support the decision-making process in sales», *South Afr. J. Ind. Eng.*, vol. 30, n.º 2, pp. 17–25, 2019, doi: 10.7166/30-2-1964.
[53] C. M. Lugo Cabrera y J. López Herrera, «Analítica de datos con aplicación en un caso práctico, mediante el uso de una herramienta libre», 2018, [En línea]. Disponible en: http://hdl.handle.net/11059/9185.
[54] H. B. Santoso, P. A. Sonia Putri, y B. S. D. Oetomo, «Implementation of Sales Executive Dashboard for A Multistore Company in Yogyakarta», *IJNMT Int. J. New Media Technol.*, vol. 4, n.º 1, pp. 59-68, 2017, doi: 10.31937/ijnmt.v4i1.540.

[55] S. E. Dragomirescu y D. C. Solomon, «THE ROLE OF THE PERFORMANCE DASHBOARD IN THE MANAGEMENT OF MODERN ENTERPRISES», *Stud. Sci. Res. - Econ. Ed.*, n.º 18, 2013, doi: 10.29358/sceco.v0i18.221.
[56] S. Vilarinho, I. Lopes, y S. Sousa, «Developing dashboards for SMEs to improve performance of productive equipment and processes», *J. Ind. Inf. Integr.*, vol. 12, pp. 13-22, dic. 2018, doi: 10.1016/j.jii.2018.02.003.
[57] K. Krishnan, «6 - Visualization, storyboarding and applications», en *Building Big Data Applications*, K. Krishnan, Ed. Academic Press, 2020, pp. 113-125.
[58] «IGS_UNE_2019_Consolidado.pdf». Accedido: nov. 25, 2020. [En línea]. Disponible en: https://www.tigo.com.co/sites/tigounecorp/files/fragmentos/general_listado_archivos/IGS_UNE_2019_Consolidado.PDF.
[59] M. de, C. E. L. Silva, y J. M. W. Durán, «EL ROL DE LOS SERVICIOS OTT EN EL SECTOR DE LAS COMUNICACIONES EN COLOMBIA», p. 31.
[60] «Activa Amazon Prime Video para Hogares en Mi Tigo | Hogar – Tigo Colombia». https://ayuda.tigo.com.co/hc/es/articles/360038028013-Activa-Amazon-Prime-Video-para-Hogares-en-Mi-Tigo-Hogar (accedido nov. 25, 2020).
[61] «Activa Amazon Prime Video en tu plan 100 Mil desde Mi Tigo | Móvil», *Tigo Colombia*. https://ayuda.tigo.com.co/hc/es/articles/360015443794-Activa-Amazon-Prime-Video-en-tu-plan-100-Mil-desde-Mi-Tigo-M%C3%B3vil (accedido nov. 25, 2020).
[62] M. Banda y E. K. Ngassam, «A data management and analytic model for business intelligence applications», en *2017 IST-Africa Week Conference (IST-Africa)*, Windhoek, may 2017, pp. 1-10, doi: 10.23919/ISTAFRICA.2017.8102350.
[63] «MariadelPilar_MorenoZuluaga_2020.pdf». Accedido: dic. 01, 2020. [En línea]. Disponible en: https://repository.eafit.edu.co/bitstream/handle/10784/16069/MariadelPilar_MorenoZuluaga_2020.pdf?sequence=2&isAllowed=y.
[64] I. F. P. Cabrera, «Desarrollo de un modelo basado en Machine Learning para la predicción de la demanda de», p. 174.
[65] P. V. Golubtsov, «The Concept of Information in Big Data Processing», *Autom. Doc. Math. Linguist.*, vol. 52, n.º 1, pp. 38-43, ene. 2018, doi: 10.3103/S000510551801003X.
[66] J. C. B. Herrera y N. R. Torres, «Bigdata y analítica para la toma de decisiones en el tratamiento y prevención del cáncer de mama», p. 116.
[67] V. Buitrago y H. Joana, «Un método para la definición de indicadores clave de rendimiento con base en objetivos de mejoramiento», 2019.
[68] «Towards Data Science», *Towards Data Science*. https://towardsdatascience.com (accedido dic. 01, 2020).
[69] «Identifying the orders of AR and MA terms in an ARIMA model». https://people.duke.edu/~rnau/411arim3.htm (accedido dic. 01, 2020).