

## **El rol del Data warehouse y BI en la era digital**

Corporación Universitaria Remington.

Seminario BIG DATA

Estudiantes:

Cristian Silva Gil

Luis Miguel Ocampo Ocampo

Yilmar Camilo Echavarría Quintero

Tutor:

Roberto Carlos Guevara Calume

2024.

**Agradecimientos**

Queremos manifestar nuestro más profundo agradecimiento a todas las personas, amigos, familiares y profesores que, de alguna manera, hicieron posible la realización de este proyecto y nos ayudaron a alcanzar este importante logro académico.

## **Tabla de Contenido**

Resumen.....	4
Esquema teórico.....	4
Introducción .....	4

	3
Big Data .....	4
Casos de uso.....	4
• Análisis de datos clínicos.....	5
• Descubrimiento de medicamentos .....	5
• Medicina de precisión .....	5
• Vigilancia de enfermedades .....	5
• Farmacovigilancia.....	5
Business Intelligence (BI).....	5
Beneficios en el sector salud.....	6
Data mart.....	6
Data warehouse .....	6
Machine Learning .....	7
Casos de uso.....	7
• Identificación de enfermedades y diagnósticos .....	7
• Diagnóstico por imagen médica.....	7
• Medicina personalizada .....	7
• Modificación del comportamiento basada en el aprendizaje automático .....	7
• Ensayos clínicos e investigación.....	7
Roles claves para un equipo de ciencia de datos .....	8
• Científico de Datos .....	8
• Ingeniero de Datos .....	8
• Analista de datos .....	8
• Especialista en dominio .....	8
• Desarrollador.....	8
Desarrollo e implementación del aprendizaje.....	8
Dashboard .....	8
¿Por qué QlikSense? .....	9
Apendice .....	9
• Tendencias de mortalidad por diagnóstico. ....	14
• Comparaciones de mortalidad entre diferentes áreas o servicios. ....	15
• Evolución temporal de la tasa de mortalidad. ....	15
• Factores de riesgo más influyentes en los desenlaces de los pacientes. ....	15
Conclusiones .....	17
Referencias.....	18

## **Resumen**

Este proyecto tiene como objetivo abordar el problema crítico de las tasas de mortalidad en los entornos hospitalarios. Debido a este problema, podemos utilizar diferentes herramientas comerciales para mejorar en muchos aspectos, lo que se reflejará en una mayor calidad de atención médica ofrecida a los pacientes. Para abordar eficazmente el problema, es fundamental obtener una comprensión más profunda de los factores subyacentes que contribuyen al aumento observado. Para lograrlo, contaremos con la colaboración de un equipo formado por un epidemiólogo y un especialista en seguridad del paciente. Intentaremos disminuir las muertes relacionadas con diferentes enfermedades para reducir el daño que causan al hospital, esto nos ayudará a crear un panel para prevenir y actuar sobre estos problemas.

El objetivo es reducir el aumento de las tasas de mortalidad, lo que mejorará la calidad de la atención al paciente.

## **Palabras claves**

Tasa de mortalidad, Hospital, Herramientas BI, Big Data, Datos, Machine learning, Data Mart, Data warehouse, dashboard.

## **Esquema teórico**

### **Introducción**

Este informe analiza la cuestión del aumento de las tasas de mortalidad en los hospitales y analiza ideas básicas sobre cómo los epidemiólogos deben gestionar las crisis sanitarias. El texto examina cómo Big Data y el aprendizaje automático pueden mejorar los métodos de prevención y control en un entorno hospitalario.

### **Big Data**

Según Ballestar, Ribeiro, y Sainz, En 1988, los investigadores de IBM, Barry Devlin y Paul Murphy, fueron los primeros en proponer la idea de un almacén de datos. No obstante, William H. Inmon fue quien se encargó de popularizar el concepto, lo que lo ha llevado a ser reconocido como el pionero en este ámbito del almacenamiento de información.

Este concepto trata sobre el uso de una gran cantidad de datos en bases de datos. Las nuevas tecnologías como la nube y el Internet de las cosas generan más datos sobre las acciones de las personas. Estos conjuntos de datos son muy grandes y tienen diferentes tipos, son difíciles de procesar rápidamente y algunos pueden no ser ciertos. Un buen manejo y estudio de big data necesita herramientas sólidas para conservar, trabajar y encontrar información útil en grandes colecciones de datos.

### **Casos de uso**

Big data puede abarcar una variedad de actividades comerciales, desde interacciones con clientes hasta análisis. Recopilaremos algunos de ellos, según (Tiao, 2024)

- Desarrollo de productos.
- Manteamiento predictivo.
- Experiencia al cliente.
- Fraude y conformidad.
- Eficiencia operativa.

El uso del big data se ha convertido en una herramienta importante en la industria, y su capacidad de analizar datos permite a las grandes empresas mejorar la toma de decisiones, mejorar los servicios, prevenir, predecir y optimizar procesos, ya que el big data seguirá creciendo con el desarrollo de la tecnología.

Reconociendo la importancia del análisis de big data en la atención médica, consideraremos adoptar enfoques analíticos para mejorar los procesos internos y la experiencia del paciente, según (Sydle, 2024)

- **Análisis de datos clínicos**  
Permite identificar patrones que contribuyen a mejorar la atención médica y tomar decisiones más acertadas en el diagnóstico y prevención.
- **Descubrimiento de medicamentos**  
Facilita encontrar posibles objetivos para nuevos tratamientos, prever cómo funcionarán los medicamentos y mejorar la planificación de los ensayos clínicos.
- **Medicina de precisión**  
Adaptando los tratamientos a las características únicas de cada paciente, lo que permite ofrecer una atención más específica y eficaz.
- **Vigilancia de enfermedades**  
Los investigadores pueden identificar patrones y tendencias de enfermedades, facilitar una respuesta temprana y prevenir brotes.
- **Farmacovigilancia**  
Esto permite identificar efectos secundarios y tomar medidas para garantizar un uso seguro de los medicamentos.

Todos estos procesos contribuyen a la investigación clínica y al descubrimiento de fármacos, acelerando el tratamiento y salvando vidas de manera oportuna.

### **Business Intelligence (BI)**

Según SAP, se refiere a los procesos y las herramientas que se usan para analizar los datos de negocio, transformándolos en información útil y ayudar a todos en la organización a tomar mejores decisiones. Un sistema de BI evalúa los datos actuales e

históricos y presenta los resultados en informes fáciles claros, dashboards, gráficos y mapas que pueden compartirse con toda la empresa.

La inteligencia empresarial en el sector sanitario es el uso de herramientas y tecnología para recopilar, analizar y convertir grandes cantidades de datos sanitarios en información. Ayuda a los proveedores de atención médica, administradores de hospitales y otras partes interesadas a recopilar información sobre tendencias, patrones y métodos para mejorar la calidad de la atención al paciente.

### **Beneficios en el sector salud**

- **Mezcla de calidad de la atención:** Identifica áreas de mejora en los procesos clínicos, reduce errores médicos y optimiza el uso de recursos.
- **Toma de decisiones basada en hechos:** Apoya a los altos directivos a tomar decisiones estratégicas, incluida la planificación de recursos, la gestión del riesgo y la inversión en nuevas tecnologías.
- **Predicción de enfermedades:** Se pueden crear modelos que identifiquen a los pacientes con mayor riesgo de desarrollar una enfermedad, permitiendo tomar acciones preventivas de manera anticipada. Esto ayuda a reducir la incidencia de enfermedades y mejorar los resultados de salud.

### **Data mart**

Un data mart es un sistema de almacenamiento de datos que contiene información específica de la organización. Contiene una pequeña porción de datos específicos que la empresa almacena en un sistema de almacenamiento más grande. Las empresas utilizan almacenes de datos para analizar información específica de cada departamento de manera más eficiente y proporcionar datos agregados que las partes interesadas pueden utilizar para tomar decisiones rápidas, según (Amazon aws, 2024)

### **Data warehouse**

Según PowerData, "La idea de un data warehouse surgió en 1988 gracias a los investigadores de IBM, Barry Devlin y Paul Murphy. Sin embargo, fue William H. Inmon quien popularizó el término 'data warehouse', siendo reconocido como el pionero en el campo del almacenamiento de datos."

Un data warehouse es un sistema que se utiliza para almacenar y gestionar datos, pero no debe considerarse un producto en sí mismo. Su objetivo principal es combinar y organizar información de diferentes fuentes, permitiendo a las empresas responder rápidamente a preguntas comerciales y tomar decisiones fundamentadas. Para lograr esto, estos sistemas emplean bases de datos relacionales que están diseñadas para permitir un acceso y análisis eficientes. Además, los almacenes de datos segregan las actividades de

análisis de las operaciones diarias y, por lo general, contienen datos históricos que provienen de diversas transacciones, según (Espinosa)

## **Machine Learning**

El aprendizaje automático es una rama de la inteligencia artificial que permite a las máquinas obtener conocimiento a partir de datos. En lugar de depender de reglas precisas, el sistema utiliza modelos para analizar ejemplos y experiencias pasadas para mejorar su capacidad para resolver problemas y completar tareas. A medida que se recopilan más datos, estos modelos pueden probar de forma independiente su comportamiento, aprendizaje y optimización sin intervención humana, según (Ramírez, 2018)

Hoy en día, tenemos muchas aplicaciones que utilizan regularmente el Machine Learning sin siquiera darnos cuenta o saberlo.

- Las plataformas de streaming más populares utilizan algoritmos para estudiar el comportamiento de los usuarios y proporcionar sugerencias personalizadas.
- Los motores de búsqueda emplean técnicas para perfeccionar los resultados.
- Las plataformas de publicidad utilizan algoritmos para ajustar los anuncios en función del comportamiento de cada usuario. Esto les permite mostrar publicidad más relevante y personalizada.

## **Casos de uso**

La creciente popularidad del aprendizaje automático en la atención sanitaria presagia un futuro en el que los datos, el análisis y la innovación se combinarán para beneficiar a innumerables pacientes. En un futuro próximo, las aplicaciones impulsadas por Machine Learning que se integran con datos de pacientes en tiempo real de múltiples sistemas de salud en múltiples países se volverán comunes, aumentando la efectividad de tratamientos que antes no existían, según (OneClick, 2022)

- **Identificación de enfermedades y diagnósticos**
- **Diagnóstico por imagen médica**
- **Medicina personalizada**
- **Modificación del comportamiento basada en el aprendizaje automático**
- **Ensayos clínicos e investigación**

## **Roles claves para un equipo de ciencia de datos**

Los proyectos de ciencia de datos que impliquen la combinación de big data y aprendizaje automático deben contar con personas que puedan realizar las tareas propuestas. Muchas empresas creen que sólo pueden poner en marcha sus proyectos con la ayuda de científicos de datos, pero no es así. La persona responsable de agregar valor al proyecto.

- **Científico de Datos**  
Según Jiménez, Los científicos de datos son como detectives que utilizan sus habilidades matemáticas e informáticas para resolver misterios ocultos en grandes cantidades de información.
- **Ingeniero de Datos**  
Los ingenieros de datos son responsables de diseñar y construir sistemas de análisis, almacenamiento y recopilación de datos a gran escala con habilidades de SQL, ETL y automatización, según (CloudCamp, 2024)
- **Analista de datos**  
Según soyhenry, Los profesionales analistas de datos son responsables de recopilar, procesar y analizar grandes cantidades de información para descubrir patrones, tendencias y relaciones que ayudan a la toma de decisiones dentro de una organización.
- **Especialista en dominio**  
Le brinda todos los detalles que necesita saber sobre el negocio y el proyecto, para que pueda comprender lo que está pasando.
- **Desarrollador**  
Utiliza la ciencia de datos para crear aplicaciones útiles y prácticas.

Un equipo de ciencia de datos exitoso es aquel que combina las habilidades técnicas y de negocio necesarias para transformar los datos en conocimiento y tomar decisiones informadas.

## **Desarrollo e implementación del aprendizaje**

### **Dashboard**

Es una herramienta adaptable para visualizar datos que recopila información desde diversas fuentes como archivos, servicios, API y documentos adjuntos. Luego presenta



esos datos en diferentes formatos visuales, como tablas o gráficos, para facilitar su interpretación y reducir la necesidad de tareas manuales. Los paneles permiten seguir estos datos en tiempo real, lo que ayuda a las organizaciones a ahorrar tiempo al evitar procesos largos y laboriosos de análisis que solían ser una carga.

### ¿Por qué QlikSense?

Es una herramienta poderosa de Business Intelligence que permite a los usuarios crear visualizaciones de datos interactivas y ajustadas a sus necesidades. Esto facilita la exploración y comprensión de grandes volúmenes de información, ayudando a identificar patrones, tendencias y relaciones clave de manera visual. Además, su interfaz intuitiva permite que personas sin conocimientos técnicos profundos puedan aprovechar al máximo los datos, mejorando la toma de decisiones basadas en información precisa y actualizada.

Según datascientest, ofrece numerosas opciones para el análisis de datos y estas son alguna de las ventajas:

- Intuitivo, gracias a su función de arrastrar y soltar, hace el trabajo más fácil para los usuarios.
- Permite la creación de paneles interactivos que permiten a las partes interesadas y a los tomadores de decisiones presentar datos con mayor claridad.
- Diseño responsivo que permite a los usuarios trabajar en cualquier dispositivo (computadora, celular o tablet) y acceder desde cualquier lugar.

### Apendice

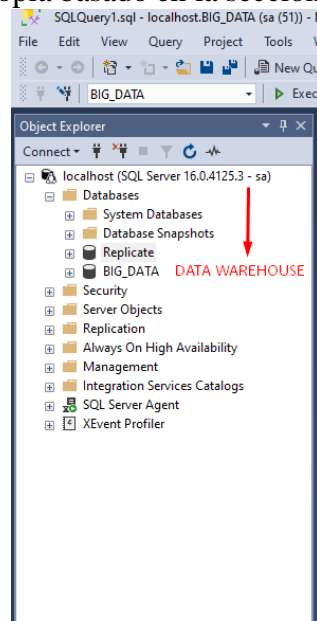
En el taller de Big Data logramos consolidar ideas claras sobre cómo manejar y obtener resultados eficaces a partir de grandes volúmenes de datos. Esto nos impulsó a implementar un almacén de datos que nos permite resolver los desafíos asociados con la integración y organización de grandes cantidades de datos clínicos y administrativos. Estamos creando un almacén de datos centralizado, que será clave para identificar patrones relacionados con la mortalidad y facilitar la toma de decisiones informadas.

El proceso comienza con la integración de diversas fuentes de información hospitalaria, asegurando que toda la información pase primero por un filtro ETL (Extracción, Transformación y Carga). Durante este paso, los datos son normalizados, garantizando que sean consistentes, sin duplicidades y bien estructurados para un análisis más efectivo. Las imágenes muestran claramente la arquitectura del almacén de datos, desde la extracción de la información hasta su transformación y carga final.

### Figura 1

*Data Waterhouse.*

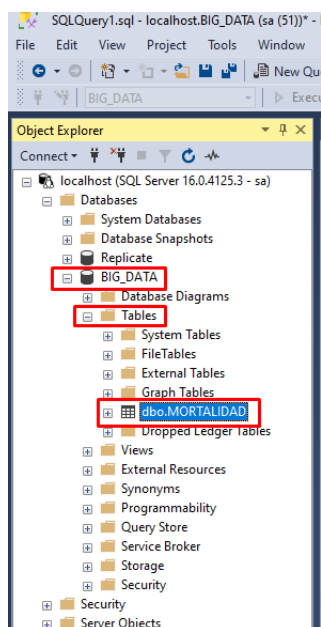
*Nota.* Gráfico de elaboración propia basado en la sección anterior.



Después de haber implementado el almacén de datos, el paso siguiente es desarrollar centros de datos especializados. Estos centros funcionan como partes específicas del almacén de datos principal, permitiendo que diferentes áreas de la organización accedan de manera más directa y personalizada a la información que necesitan.

**Figura 2**

Base de datos *BIG\_DATA*, *Data Mart*.

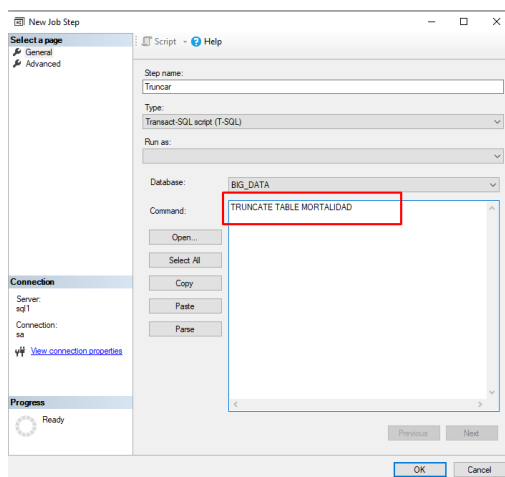


*Nota.* Gráfico de elaboración propia basado en la sección anterior.

El proceso de carga de datos a este Data Mart se realiza a través de Jobs automáticos, que garantizan una actualización constante de la información en los cubos de datos. Estos cubos permiten realizar análisis más profundos los cuales nos sirven para identificar tendencias en los diagnósticos y la atención de los pacientes.

### Figura 3

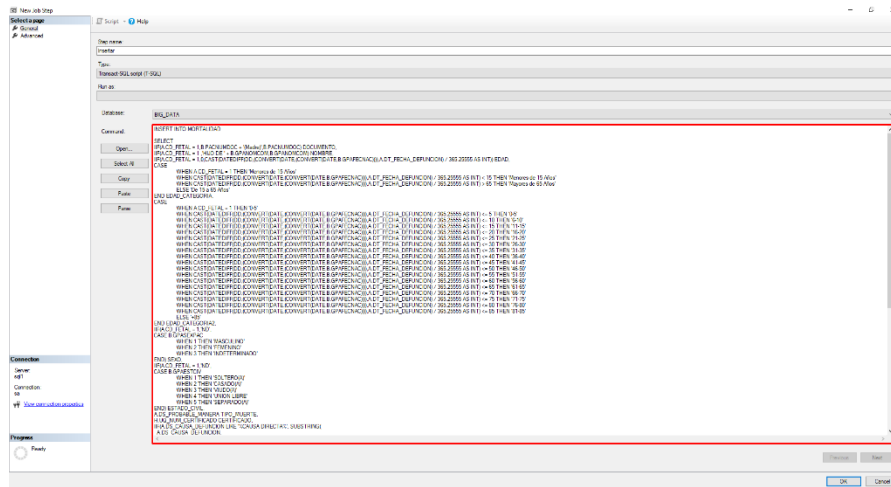
*Eliminación de datos.*



*Nota.* Gráfico de elaboración propia basado en la sección anterior.

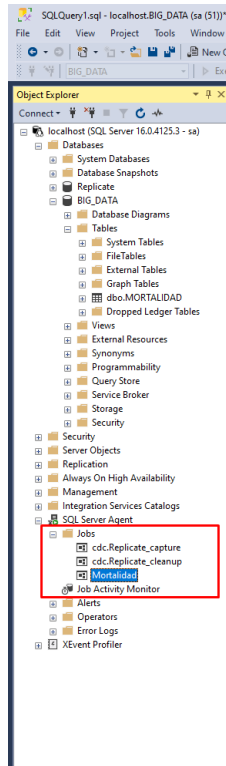
Este proceso se realiza con el fin de agilizar la carga de los datos, ya que si se realizara con un UPDATE el proceso tomaría demasiado tiempo porque este validaría registro a registro para evidenciar que dato debería ser actualizado dentro de la tabla.

**Figura 4**  
*Insertar datos.*



*Nota.* Gráfico de elaboración propia basado en la sección anterior.

**Figura 5**  
*Insertar variables.*

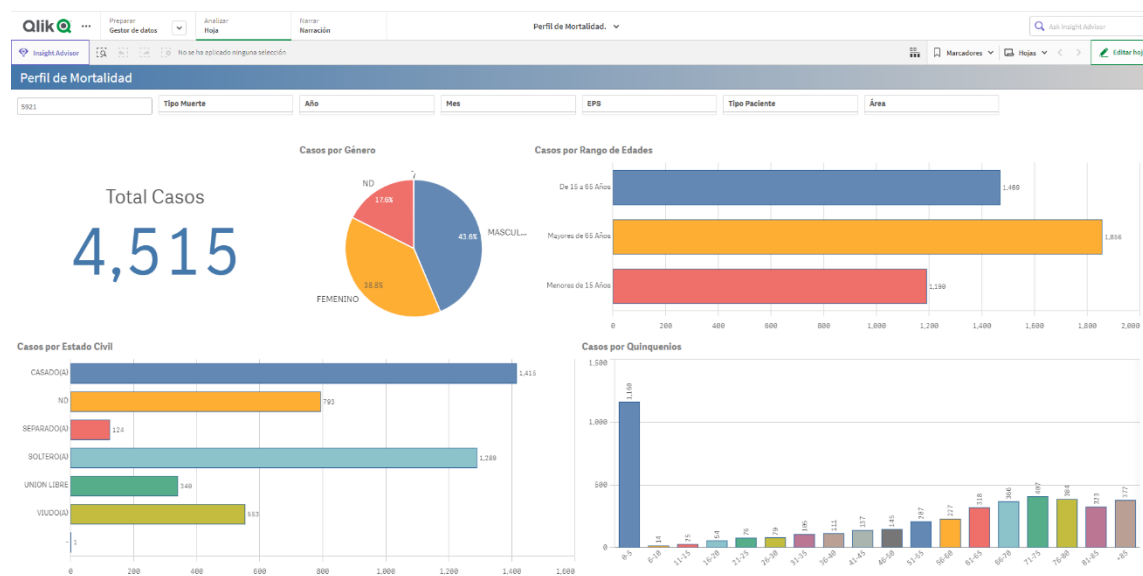


*Nota.* Gráfico de elaboración propia basado en la sección anterior.

Necesitamos recopilar toda la información y detalles sobre la mortalidad para analizarla posteriormente mediante el panel.

El proceso de implementación finalizará creando un panel interactivo llamado "mortalidad" en el panel qlik. Este panel brindará a los profesionales de la salud y a los administradores de hospitales información importante sobre la mortalidad hospitalaria.

**Figura 6**  
*Dashboard interactivo.*

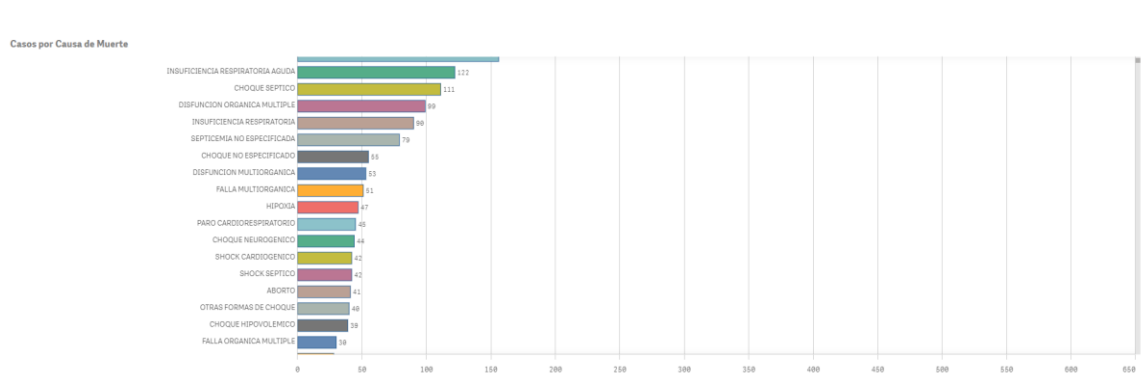


*Nota.* Gráfico de elaboración propia basado en la sección anterior.

Este dashboard ofrece una representación gráfica e intuitiva de los indicadores clave, como:

- **Tendencias de mortalidad por diagnóstico.**

**Figura 7**  
*Tendencias de mortalidad por diagnóstico.*



*Nota.* Gráfico de elaboración propia basado en la sección anterior.

- **Comparaciones de mortalidad entre diferentes áreas o servicios.**

**Figura 8**

*Comparaciones de mortalidad entre diferentes áreas o servicios.*

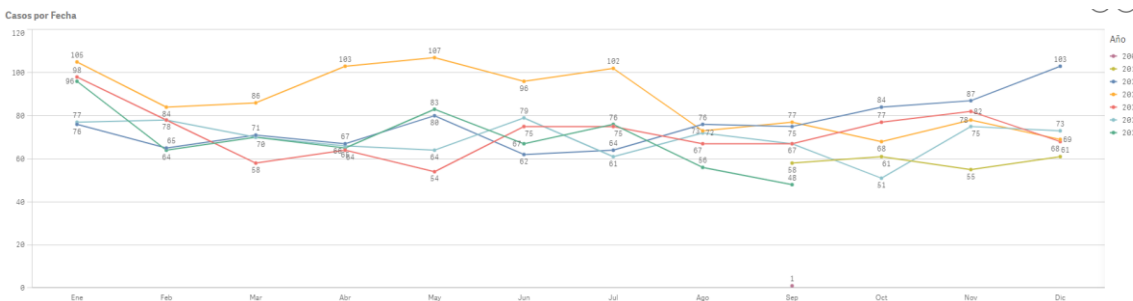


*Nota.* Gráfico de elaboración propia basado en la sección anterior.

- **Evolución temporal de la tasa de mortalidad.**

**Figura 9**

*Evolución temporal de la tasa de mortalidad.*

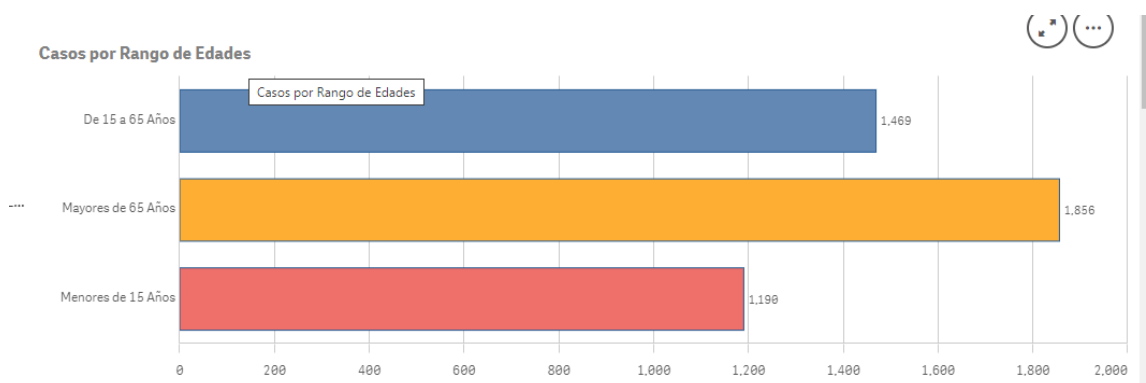


*Nota.* Elaboración propia a partir de escenario de predicción.

- **Factores de riesgo más influyentes en los desenlaces de los pacientes.**

**Figura 10**

*Factores de riesgo más influyentes en los desenlaces de los pacientes.*



*Nota.* Gráfico de elaboración propia basado en la sección anterior.

Además, gracias a las capacidades de análisis visual que ofrece Qlik, los usuarios pueden trabajar con datos en tiempo real, ajustando la información por rangos de fechas, categorías de diagnóstico, estado del paciente y otros criterios importantes. Esta interactividad les permite identificar de manera rápida patrones y áreas críticas que necesitan atención urgente.

El panel de mortalidad está estrechamente vinculado a los cubos de datos generados en el almacén de datos, lo que garantiza que la información esté constantemente actualizada. Las métricas y visualizaciones disponibles no solo son útiles para determinar las principales causas de muerte, sino que también contribuyen a la formulación de estrategias preventivas y a la mejora de la calidad de la atención hospitalaria.



## Conclusiones

Lo que nos deja este estudio es la posibilidad de mejorar los sistemas de salud para reducir el aumento de las muertes hospitalarias. Creemos que podemos mejorar esta situación con la tecnología BIGDATA y estamos enfocados en encontrar mejores formas de comprender las causas de la muerte de los pacientes a través del almacenamiento de datos. Cuando utilizamos estas herramientas para analizar datos, podemos identificar patrones y problemas que son invisibles a simple vista pero que nos brindarán información más rápida y detallada sobre el problema.

Esto es muy importante para la industria porque dará a los trabajadores sanitarios una visión más amplia y clara a la hora de tomar decisiones. Tendrán los datos más relevantes en tiempo real y podrán actuar en consecuencia de la mejor manera posible. Prevenir situaciones más complejas y de difícil solución.

Este enfoque también señala la importancia de los big data en el sector de la salud, la innovación continua y la introducción de nuevas herramientas como el aprendizaje automático ayudarán a prevenir y tratar la mayor cantidad de enfermedades de manera oportuna.

Por último, es importante manejar siempre los datos de forma ética y respetar su confidencialidad. Datos, especialmente de pacientes, para generar confianza en los sistemas implementados, y esto se puede expandir a muchas otras partes del mundo para brindar resultados más positivos y realistas para todos.

## Referencias

- Amazon aws. (2024). *aws.amazon.com*. Fuente: <https://aws.amazon.com/es/what-is/datalake/>
- Ballestar, M. T., Ribeiro, D., & Sainz, J. (s.d.). *¿ES EL BIG DATA EL SIGUIENTE PASO EN LA DIGITALIZACIÓN DE LA EMPRESA?* Fuente: PAPEL DE LA TECNOLOGÍA: <https://www.mintur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/409/BALLESTAR,%20RIBEIRO%20Y%20SAINZ.pdf>
- CloudCamp. (21 de Febrero de 2024). *linkedin*. Fuente: ¿Qué es Ingeniería de Datos? Guía Completa: <https://es.linkedin.com/pulse/qu%C3%A9-es-ingenier%C3%ADa-de-datos-gu%C3%ADa-completa-cloudcamplata-cv8re>
- datascientest. (12 de Marzo de 2024). *Lo que necesitas saber sobre Qlik Sense*. Fuente: <https://datascientest.com/en/what-you-need-to-know-about-qlik-sense>
- Espinosa, I. A. (s.d.). *DATA WAREHOUSE PARA LA GESTIÓN DE LISTA DE ESPERA SANITARIA*. Fuente: [https://oa.upm.es/1095/1/PFC\\_ITZIAR\\_ANGOITIA\\_ESPINOSA.pdf](https://oa.upm.es/1095/1/PFC_ITZIAR_ANGOITIA_ESPINOSA.pdf)
- Jiménez, J. (16 de Octubre de 2020). *xataka*. Fuente: Científico de datos: así es y así se forma uno en esta profesión cada vez más demandada: <https://www.xataka.com/otros/cientifico-datos-asi-profesion-demandada>
- OneClick, H. (23 de Noviembre de 2022). *linkedin*. Fuente: 7 utilidades del Machine Learning en la salud: <https://es.linkedin.com/pulse/7-utilidades-del-machine-learning-en-la-salud-healthoneclick>
- PowerData. (2024). *Data Warehouse*. Fuente: todo lo que necesitas saber sobre almacenamiento de datos: <https://www.powerdata.es/data-warehouse#:~:text=Un%20Data%20Warehouse%20es%20un,recuperar%20y%20f%C3%A1cil%20de%20administrar.>
- Ramírez, D. H. (2018). *EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD*. Fuente: <https://repository.unilibre.edu.co/bitstream/handle/10901/17289/EL%20MACHINE%20LEARNING.pdf?sequence=1&isAllowed=y>
- sap.com. (s.d.). *¿Qué es business intelligence (BI)?* Fuente: <https://www.sap.com/latinamerica/products/technology-platform/cloud-analytics/what-is-business-intelligence.html#:~:text=Definici%C3%B3n%20de%20business%20intelligence&text=Tambi%C3%A9n%20conocido%20como%20sistema%20de,compartirse%20con%20toda%20la%20empr>
- soyhenry. (s.d.). *Analista de Datos*. Fuente: [https://www.soyhenry.com/analista-de-datos?gad\\_source=1&gclid=Cj0KCQjwo8S3BhDeARIsAFRmkONUvMRZ\\_yc8qIHHi3y14ans6e7ZEdeovErGCmtB63r\\_4gBjlfkg3\\_YaArJgEALw\\_wcB](https://www.soyhenry.com/analista-de-datos?gad_source=1&gclid=Cj0KCQjwo8S3BhDeARIsAFRmkONUvMRZ_yc8qIHHi3y14ans6e7ZEdeovErGCmtB63r_4gBjlfkg3_YaArJgEALw_wcB)
- Sydle. (18 de Julio de 2024). *Aplicaciones del Big Data en la salud*. Fuente: Transformación Digital: <https://www.sydle.com/es/blog/aplicaciones-big-data-salud-6511fd80f8955c2bef737329>
- Tiao, S. (11 de Marzo de 2024). *Oracle*. Fuente: What Is Big Data?: <https://www.oracle.com/co/big-data/what-is-big-data/>

