

## Diseño de un Dashboard para el manejo y análisis de los datos de casos positivos del covid-19 en Colombia utilizando herramientas de Inteligencia de Negocios y de Analítica de Datos

### Estudiantes:

Andrés Felipe Barrios Perez  
CC 1.000.126.164  
Código estudiantil: 201812794028  
Correo: [andres.barrios@unisimon.edu.co](mailto:andres.barrios@unisimon.edu.co)

Roberto José De la Cruz Padilla  
CC 1.234.089.830  
Código estudiantil: 201811294220  
Correo: [roberto.delacruz@unisimon.edu.co](mailto:roberto.delacruz@unisimon.edu.co)

Pedro Pablo Pabón Granados  
CC 1.125.230.559  
Código estudiantil: 201711282384  
Correo: [pedro.pabon@unisimon.edu.co](mailto:pedro.pabon@unisimon.edu.co)

Nilson Rafael Ruiz Otero  
CC 1.046.273.901  
Código estudiantil: 201811295068  
Correo: [nilson.ruiz@unisimon.edu.co](mailto:nilson.ruiz@unisimon.edu.co)

Trabajo de Investigación del Programa  
Ingeniería de Sistemas - Ingeniería Multimedia

Tutor:  
Juan Carlos Calabria Sarmiento

## RESUMEN

La pandemia de covid-19 ha tenido un gran impacto en Colombia afectando varios aspectos en nuestro país, de esta se generan una basta cantidad de datos para la producción de reportes y gráficos estadísticos para que informen a las personas, donde cada día se recibe información, a través de diversos medios, de los nuevos casos positivos, muertos, contagiados y demás datos sobre el covid-19. A pesar de que cada día se recibe esta información, no se ha encontrado una plataforma o sitio que evidencie de forma visual y gráfica datos detallados y completos sobre los casos positivos de covid-19.

La pandemia por el covid-19 genera cantidad de datos que son almacenados en archivos CVS, Planos, XML entre otros. Estos son utilizados para generar reportes estadísticos y realizar análisis de: contagios, recuperados, muertos e importados. Estos datos son procesados por distintas ciencias del saber, los cuales son adjuntados a un documento, pero la manera en que se presentan no siempre es clara y fácil de interpretar. Por otro lado, no presentan gráficos que ayuden al lector a interpretar y sacar conclusiones de una forma rápida y eficiente para la toma de decisiones.

En otras palabras debido a la grave situación en la que estamos, al no encontrar una manera de acceder a una información detallada, clara e imparcial de los datos del virus significa no tener un proceso secuencial para generar cambios importantes en la comunidad, en definitiva se requiere mejorar en varios aspectos ya localizados donde los datos puedan llegar de manera clara y concisa a cada persona para su posterior análisis, no obstante se sabe que algunos casos en los reportes no se encuentran los datos actualizados causando en muchas ocasiones desinformación que puede conllevar a la toma de decisiones no acertadas y llegar a niveles críticos en sectores como el económico, social entre otros.

Posibles causas de este problema radican en la extensa cantidad de datos que se generan de esta temática, siendo un poco dificultoso trabajar con estos. Además de la complejidad de los diversos datos, también se puede decir que la falta o carencia de un sistema que no solo tome los datos para registrarlos en un formato o documento, el cual es de difícil comprensión para un usuario sin el conocimiento del tema, si no la construcción de un sistema que ayude a entender a la persona, y que esta pueda interpretar la información dada y sacar sus propias conclusiones.

Por otro lado, para dar solución a esta problemática se construirá un Dashboard, utilizando un DataWareHouse para el almacenamiento de los datos de los casos de covid 19 en Colombia y realizar análisis por medio de herramientas de inteligencia negocio y de Analítica de Datos y por medio de una plataforma con una interfaz gráfica amigable y dinámica, se plasmarán los resultados de los análisis.

### **Antecedentes:**

En los antecedentes de la investigación podemos encontrar a OMS (2020) quien realizó la estimación de la mortalidad de la COVID-19 la investigación se enmarco dentro de un proyecto factible donde la OMS buscaba encontrar un porcentaje de las personas fallecidas a nivel mundial pero enfocados en clasificar la información dependiendo las razones de la muerte uno de sus objetivos era evaluar la proporción de personas infectadas que tienen un desenlace mortal ellos utilizaron dos medidas: la razón de letalidad de la infección (IFR, por su acrónimo en inglés), que estima la proporción de muertes entre todas las personas infectadas, y la razón de letalidad de los casos (CFR, por su acrónimo en inglés), que estima la proporción de muertes entre los casos confirmados.

Y para medir con exactitud la IFR ellos tuvieron en cuenta los datos completos del número de infecciones y de muertes causadas por la enfermedad. Y en consecuencia, en esta fase inicial de la pandemia, la mayoría de las estimaciones de las razones de letalidad se han basado en los casos detectados a través de la vigilancia y se han calculado con métodos brutos, lo que ha dado lugar a estimaciones de la CFR muy variables según el país: de menos del 0,1% a más del 25%.

Ellos toman datos completos de las personas infectadas a nivel mundial para revelar un informe detallado a las partes interesadas que en este caso podría ser cada una de las naciones sus ministerios de salud incluidos sus empresas y otras organizaciones. Por otra parte la nuestra está enfocada en establecer los datos de las personas contagiadas por COVID-19 a nivel nacional.

### **Objetivos:**

#### **General**

Diseñar un DashBoard para el manejo y análisis de los datos de casos positivos del covid-19 en Colombia utilizando herramientas de Inteligencia de Negocios y Herramientas de Analítica de Datos.

#### **Específicos**

- Construir una Base de Datos Transaccional para el almacenamiento de datos obtenidos de las diferentes fuentes oficiales.
- Construir una bodega de datos utilizando el modelo Multidimensional tipo copo de nieve.
- Realizar la extracción, transformación y carga de datos con el fin de obtener información íntegra.
- Desarrollar un tablero de control para el manejo de gráficos dinámicos e informes inteligentes para la toma de decisiones.

### **Materiales y Métodos:**

La metodología planteada para la elaboración del proyecto consta de 3 fases:

**Fase 1. Planeación:** en esta primera fase se identificarán los requisitos del proyecto, se definirá el problema referente al proyecto, así como también se definirán los presupuestos y el cronograma.

**Fase 2. Desarrollo:** Se creará una base de datos, donde se importarán los datos recopilados, para luego implementar el proceso ETL, para la reconstrucción de los datos, y para que estos puedan ser almacenados en un DataWareHouse. Además de construir una plataforma con una interfaz gráfica visualmente agradable, donde se plasmarán los datos de la DataWareHouse.

**Fase 3. Análisis:** a través de una herramienta de inteligencia de negocio se desarrollará un análisis con los datos reunidos, donde se plasmarán resultados y conclusiones sobre el comportamiento de los datos del Covid.

### Resultados:

Como resultado de la presente investigación se obtuvo:

Recopilación de los datos haciendo uso del proceso ETL (Extracción, transformación y carga) para realizar transformación de los datos sobre las fuentes de origen de los datos reunidos. En este caso los datos son extraídos del archivo CSV, originario de datos abiertos de Colombia, del cual se extraen los datos requeridos en la parte de la planeación del proyecto. (Ver figuras 6 y 7)

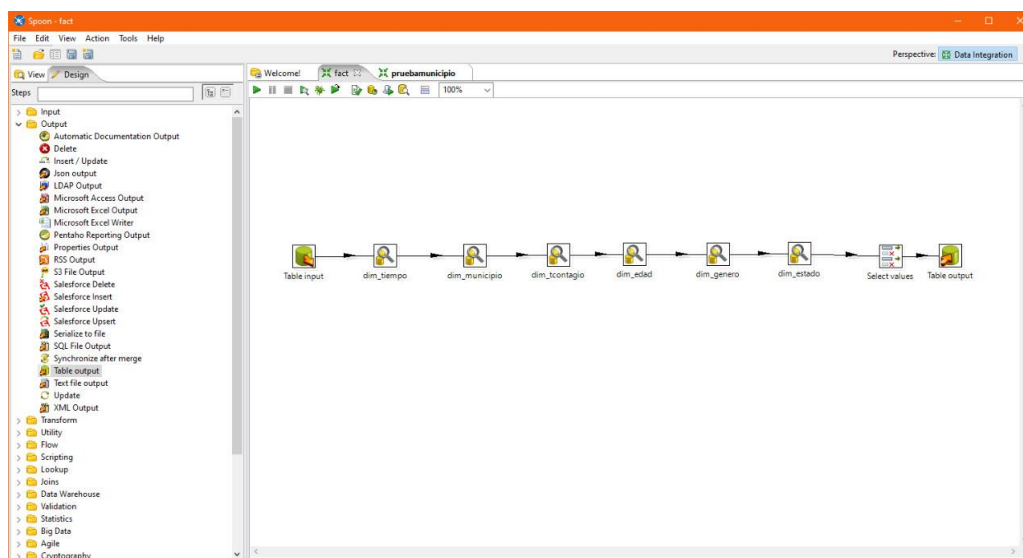


Fig. 6. Aplicación del proceso de ETL en Fact Table con Pentaho

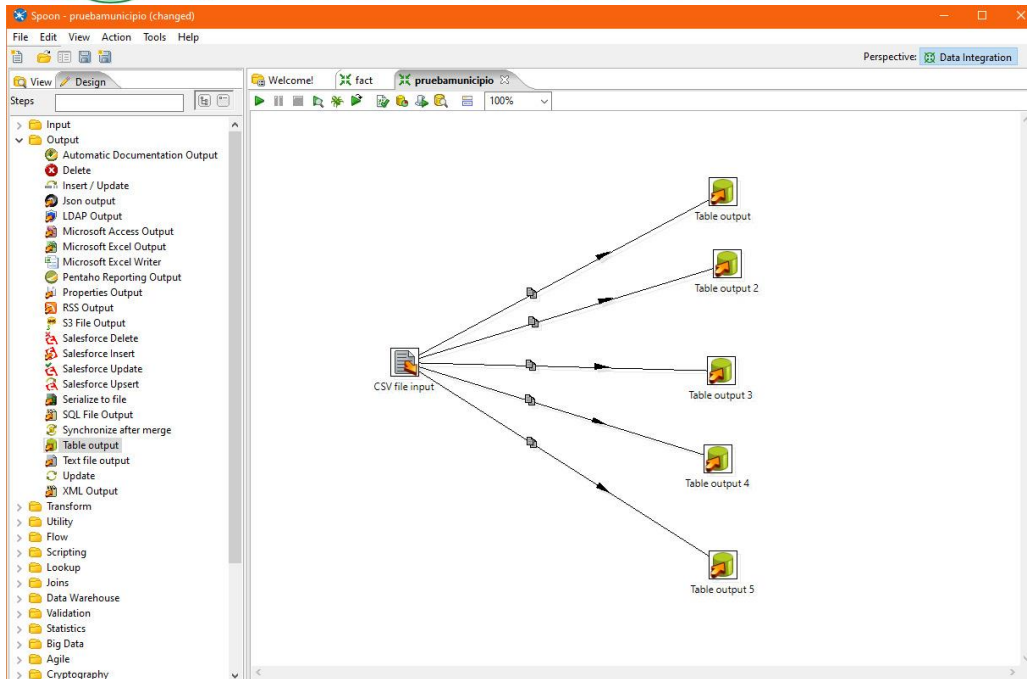


Fig. 7. Aplicación del proceso de ETL en las dimensiones de nuestra fact table con Pentaho

Además del proceso de ETL otro de los resultados, el cual fue el principal objetivo de esta investigación, es el diseño del dashboard de los casos positivos de covid 19 en Colombia, el cual fue desarrollado a través de la herramienta Qlik Sense, donde se importaron los datos recopilados de la datawarehouse, y se plasmaron dichos datos en gráficos estadísticos con las medidas y dimensiones que se establecieron al inicio de la investigación.

(Ver figuras 8 y 9)

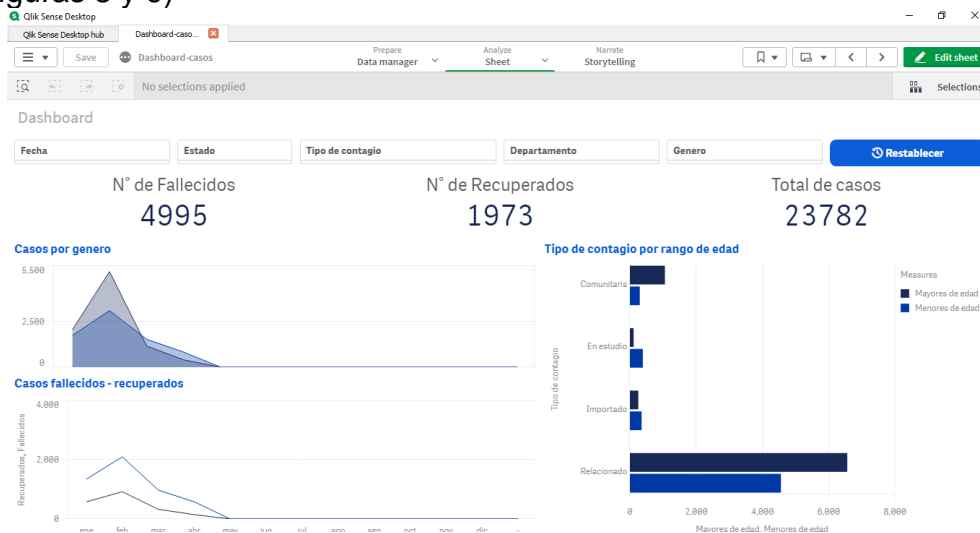


Fig. 8. Dashboard y análisis de los datos con qlik sense

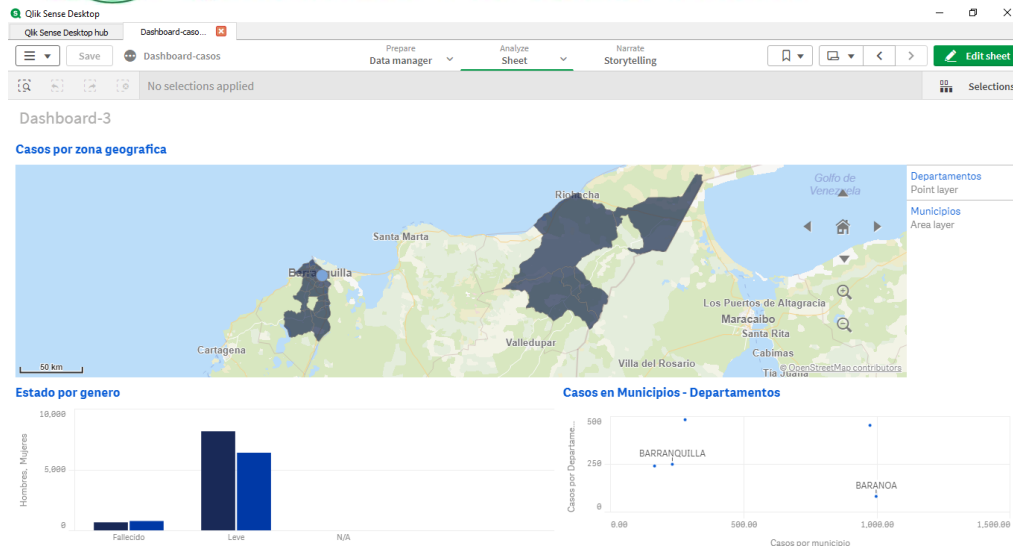


Fig. 9. Dashboard y análisis de los datos geográficos con qlik sense

## Conclusiones:

El objetivo de este proyecto de investigación está en generar un análisis sobre covid 19 en Colombia, ya que la pandemia genera un gran cantidad de datos que son almacenados en archivos CVS, Planos, XML entre otros. Estos son utilizados para generar reportes estadísticos y realizar análisis de: contagios, recuperados, muertos e importados. Estos datos son procesados por distintas ciencias del saber, los cuales son adjuntados a un documento, pero la manera en que se presentan no siempre es clara y fácil de interpretar, por parte de personas sin esos conocimientos. Por lo anterior se toma la determinación de llevar a cabo un proyecto de implementar una data Warehouse, una herramienta rápida y confiable para el análisis y búsqueda de datos, lo que permitirá un manejo eficiente de la información. En conclusión la investigación ayudará en gran medida a las personas a entender y estar informado, para poder tomar decisiones acertadas en cuanto a los datos empleados en la investigación.

**Palabras clave:** Herramientas de inteligencia de Negocios, DataWareHouse, Bases de datos, ETL, Covid 19, Colombia, datos

## ABSTRACT

The covid-19 pandemic has had a great impact in Colombia affecting several aspects in our country, from this a large amount of data is generated for the production of statistical reports and graphs to inform people, where every day information is received through various media, new positive cases, deaths, infected and other data on covid-19. Although this information is received every day, no



platform or site has been found that visually and graphically shows detailed and complete data on positive cases of covid-19.

The covid-19 pandemic generates a large amount of data that is stored in CVS, Plans, XML and other files. These are used to generate statistical reports and perform analysis of: infections, recovered, dead and imported. These data are processed by different knowledge sciences, which are attached to a document, but the way they are presented is not always clear and easy to interpret. On the other hand, they do not present graphs that help the reader to interpret and draw conclusions in a quick and efficient way for decision making.

In other words, due to the serious situation we are in, not finding a way to access detailed, clear and unbiased information on the virus data means not having a sequential process to generate important changes in the community, in short, it is necessary to improve in several aspects already located where the data can reach each person in a clear and concise manner for further analysis, However, it is known that in some cases the reports do not contain updated data, causing in many cases misinformation that can lead to unsound decision making and reach critical levels in sectors such as economic, social and others.

Possible causes of this problem lie in the extensive amount of data generated on this subject, making it difficult to work with them. In addition to the complexity of the various data, it can also be said that the lack or lack of a system that not only takes the data to record them in a format or document, which is difficult to understand for a user without knowledge of the subject, but the construction of a system that helps to understand the person, and that this can interpret the information given and draw their own conclusions.

On the other hand, to solve this problem, a Dashboard will be built, using a DataWareHouse to store the data of the covid 19 cases in Colombia and perform analysis through business intelligence tools and Data Analytics and through a platform with a friendly and dynamic graphical interface, the results of the analysis will be shown.

### **Background:**

In the background of the research we can find WHO (2020) who carried out the estimation of the mortality of the COVID-19 the research was framed within a feasible project where WHO sought to find a percentage of the deceased persons worldwide but focused on classifying the information depending on the reasons for death one of its objectives was to evaluate the proportion of infected persons who have a fatal outcome they used two measures: the infection case fatality ratio (IFR), which estimates the proportion of deaths among all infected persons, and the case

fatality ratio (CFR), which estimates the proportion of deaths among confirmed cases.

And to accurately measure the IFR they took into account the complete data on the number of infections and deaths caused by the disease. And consequently, in this early phase of the pandemic, most of the estimates of case fatality ratios have been based on cases detected through surveillance and calculated using crude methods, resulting in estimates of CFR that vary widely by country: from less than 0.1% to more than 25%.

They take complete data of infected people worldwide to disclose a detailed report to interested parties which in this case could be each of the nations, their health ministries including their companies and other organizations. On the other hand, ours is focused on establishing the data of people infected by COVID-19 at the national level.

## **Objectives:**

### **General**

To design a DashBoard for the management and analysis of covid-19 positive case data in Colombia using Business Intelligence tools and Data Analytics tools.

### **Specific**

- Build a Transactional Database for the storage of data obtained from different official sources.
- Build a data warehouse using the Multidimensional snowflake model.
- Perform data extraction, transformation and loading in order to obtain complete information.
- Develop a dashboard for the management of dynamic graphics and intelligent reports for decision making.

## **Materials and Methods:**

The methodology proposed for the elaboration of the project consists of 3 phases:

**Phase 1. Planning:** in this first phase the project requirements will be identified, the problem related to the project will be defined, as well as the budgets and schedule.

**Phase 2. Development:** A database will be created, where the collected data will be imported, to then implement the ETL process, for the reconstruction of the data, and so that they can be stored in a DataWareHouse. In addition to building a platform





with a visually pleasing graphical interface, where the DataWareHouse data will be displayed.

**Phase 3. Analysis:** Through a business intelligence tool, an analysis will be developed with the collected data, where results and conclusions about the behavior of Covid's data will be shown.

**Results:**

As a result of the present investigation, it was obtained:

Data collection using the ETL process (Extraction, transformation and loading) to carry out transformation of the data on the sources of origin of the collected data. In this case, the data is extracted from the CSV file, originating from open data in Colombia, from which the data required in the planning part of the project are extracted. (See figures 6 and 7)

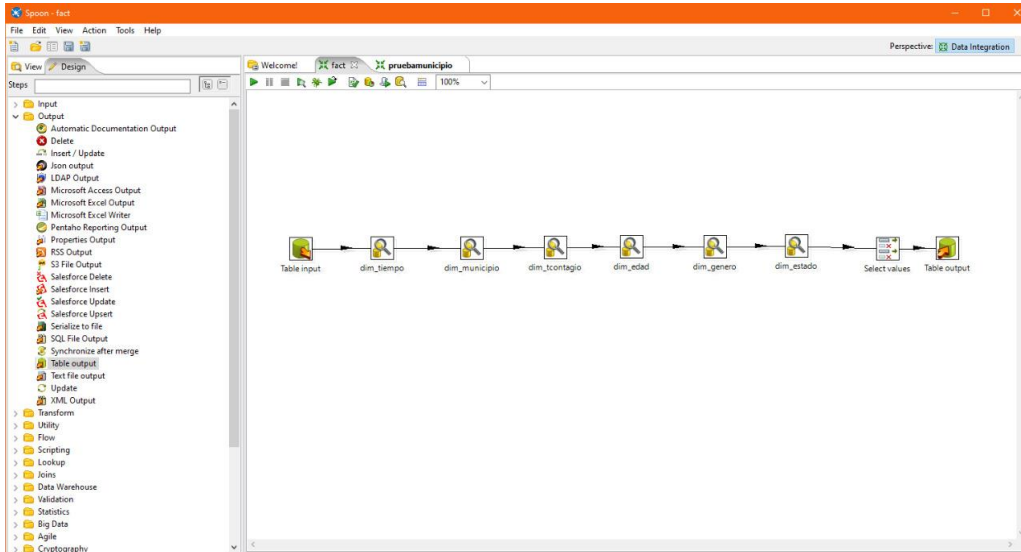


Fig. 6. Application of the ETL process in Fact Table with Pentaho

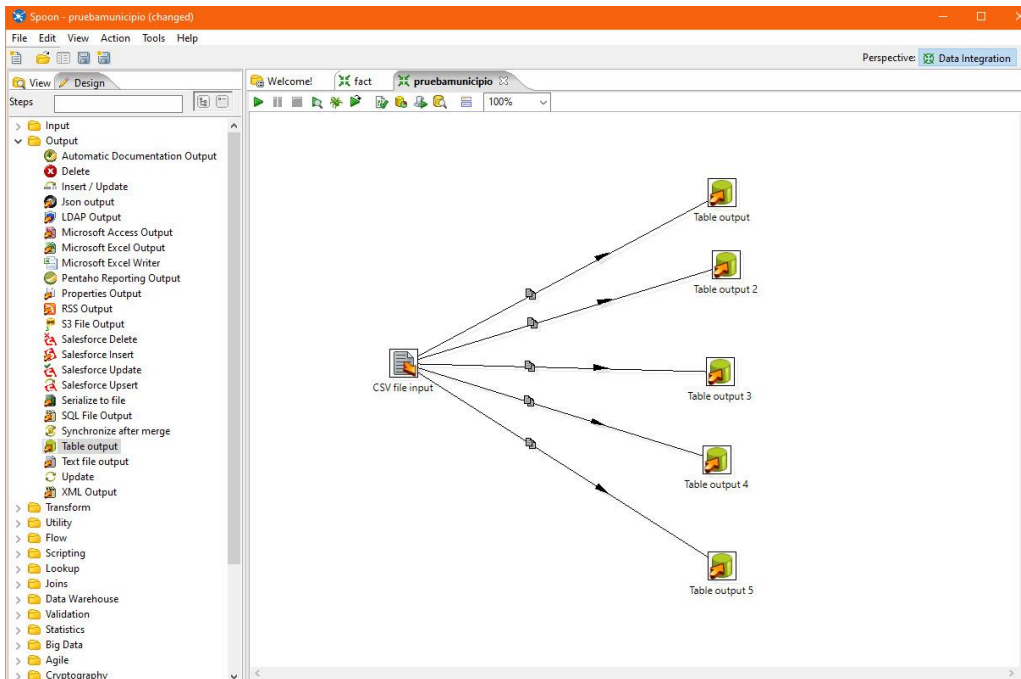


Fig. 7. Application of the ETL process in the dimensions of our fact table with Pentaho

In addition to the ETL process, another of the results, which was the main objective of this research, is the design of the dashboard of the positive cases of covid 19 in Colombia, which was developed through the Qlik Sense tool, where they were imported the data collected from the data warehouse, and these data were reflected in statistical graphs with the measures and dimensions that were established at the beginning of the investigation.

(See figures 8 and 9)

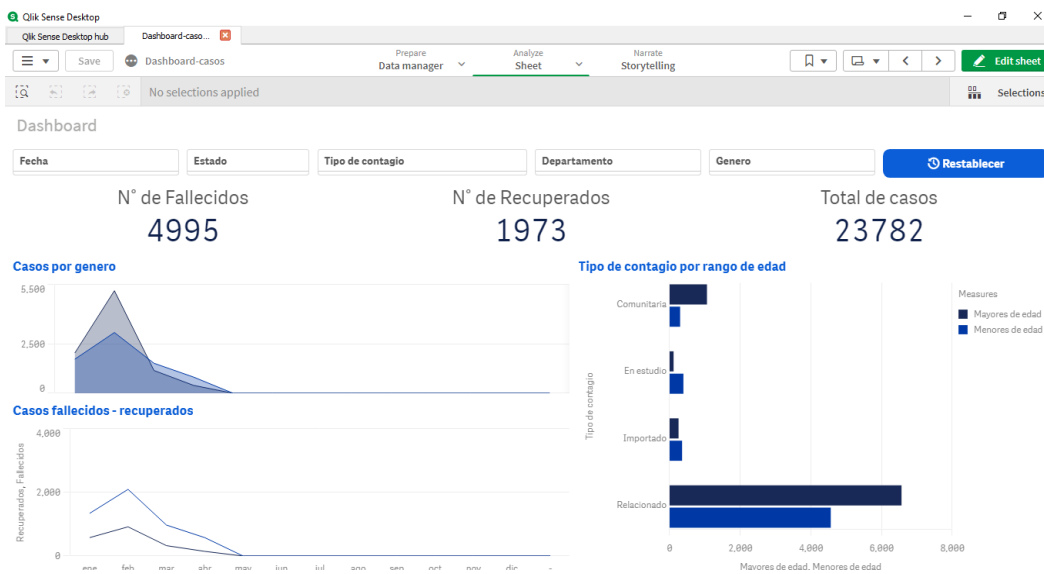


Fig. 8. Dashboard and data analysis with qlik sense

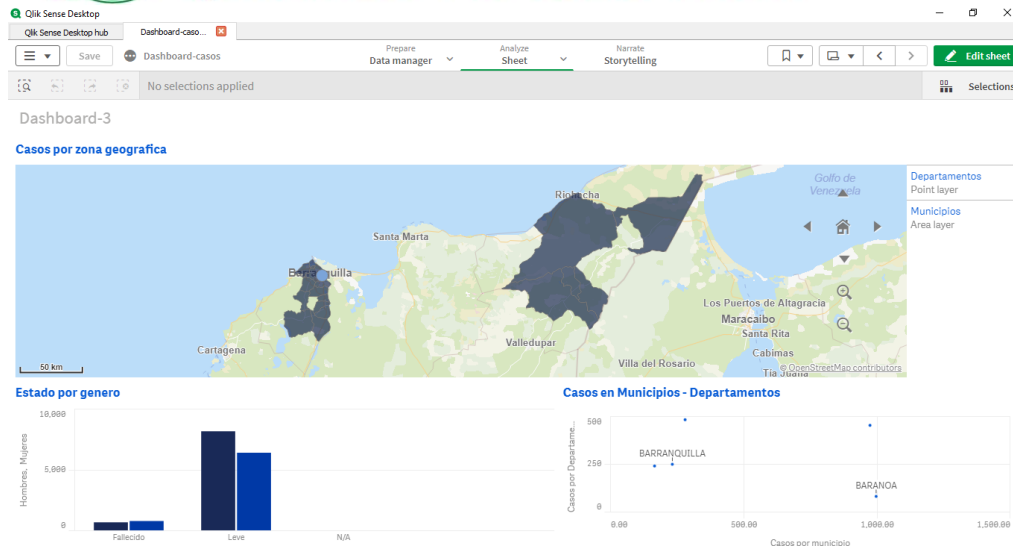


Fig. 9. Dashboard and analysis of geographic data with qlik sense

### Conclusions:

The objective of this research project is to generate an analysis of covid 19 in Colombia, since the pandemic generates a large amount of data that are stored in CVS, Plans, XML files, among others. These are used to generate statistical reports and perform analysis of: infections, recovered, dead and imported. These data are processed by different sciences of knowledge, which are attached to a document, but the way they are presented is not always clear and easy to interpret by people without such knowledge. Therefore, it was decided to carry out a project to implement a data warehouse, a fast and reliable tool for the analysis and search of data, which will allow an efficient management of the information. In conclusion, the research will greatly help people to understand and be informed, in order to be able to make the right decisions regarding the data used in the research.

**KeyWords:** Business Intelligence Tools, DataWareHouse, Databases, ETL, Covid 19, Colombia, data

## REFERENCIAS

- [1] Nauca Torres, Enrique Santos, Gómez Martínez, Jimy Yohan, *IMPLEMENTACIÓN DE UNA PLATAFORMA DE BUSINESS INTELLIGENCE BASADO EN ANÁLISIS MULTIDIMENSIONAL PARA MONITOREAR EL COMPORTAMIENTO DE CASOS COVID 19 EN EL PERÚ*, 2020.
- [2] Agapito G, Zucco C, Cannataro M. COVID-WAREHOUSE: A Data Warehouse of Italian COVID-19, Pollution, and Climate Data. *Revista Internacional de Investigación Ambiental y Salud Pública* . 2020; 17 (15): 5596. <https://doi.org/10.3390/ijerph17155596>
- [3] D. Quintero, K. Mejia, F. Carpio, K. Ariza, Y Sanchez, « Diseño e Implementación de un Datawarehouse para la presentación de informes inteligentes sobre medicamentos entregados basados en herramientas OLAP con el fin de seleccionar múltiples criterios y entrega de información personalizada en la distribuidora SEMEDICAL,» Universidad Simón Bolívar, 2018.
- [4] .R. A. Abad and H. K. Hameed, Construction Of Drug Company Data System Using Multidimensional Database Insertion, 2013. [http://www.jeasd.org/images/2013edition/issue\\_4/15Construction.of.Drug.Compan.y.Data.System.using.Multidimensional.Database.Insertion.pdf](http://www.jeasd.org/images/2013edition/issue_4/15Construction.of.Drug.Compan.y.Data.System.using.Multidimensional.Database.Insertion.pdf)
- [5] Shaker H, Abdeltawab M, AhmedHendawi, A proposed model for data warehouse ETL processes,2011. <https://www.sciencedirect.com/science/article/pii/S131915781100019X/?imgSel=Y>
- [6] Díaz Pinzón, J. E. (2020). Descripción estadística del COVID- 19 según el grupo etario en Colombia. *Revista Repertorio De Medicina Y Cirugía*, 79-85. <https://doi.org/10.31260/RepertMedCir.01217372.1098>
- [7] Otoyá Tono, A. M., García, M., Jaramillo Moncayo, C., Wills, C., & Campos Mahecha, Ángela M. (2020). COVID-19: generalidades, comportamiento epidemiológico y medidas adoptadas en medio de la pandemia en Colombia. *ACTA DE OTORRINOLARINGOLOGÍA & CIRUGÍA DE CABEZA Y CUELLO*, 4 13. Recuperado a partir de <https://www.revista.acorl.org.co/index.php/acorl/article/view/475>
- [8] Díaz Pinzón, J. E. (2020). Uso de modelo predictivo para la dinámica de transmisión del COVID-19 en Colombia. *Revista Repertorio De Medicina Y Cirugía*, 34-44. <https://doi.org/10.31260/RepertMedCir.01217372.1056>
- [9] Pontificia Universidad Javeriana. Facultad de Medicina. Departamento de Epidemiología Clínica y Bioestadística, 2020. <https://repository.javeriana.edu.co/handle/10554/53012>

[10] Amirhoshang Hoseinpour Dehkordi, Majid Alizadeh, Pegah Derakhshan, Peyman Babazadeh, Arash Jahandideh,(2020). Understanding epidemic data and statistics: A case study of COVID-19 <https://doi.org/10.1002/jmv.25885>

[11] Neil Pearce PhD, Jan P. Vandenbroucke MD, PhD, Tyler J. VanderWeele PhD, and Sander Greenland DrPH, MS, 2020. Accurate Statistics on COVID-19 Are Essential for Policy Guidance and Decisions.

[12] Department of Laboratory Medicine, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China. Laboratory data analysis of novel coronavirus (COVID-19) screening in 2510. patients

[13] L. J. Muhammad, Md. Milon Islam, Sani Sharif Usman & Safial Islam Ayon, 2020. Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery.

[14] Department of Engineering Sciences, University of Oxford, United Kingdom, 2020. Data mining and analysis of scientific research data records on Covid-19 mortality, immunity, and vaccine development - In the first wave of the Covid-19 pandemic.

[15] J.C Calabria Sarmiento Construcción y poblamiento de un datawarehouse basado en el paradigma de bases de datos objeto relacional PROSPECTIVA, vol. 9, núm. 1, enero-junio, 2011, pp. 69-77

[16] OMS, «Organización mundial de la salud,» 12 octubre 2020. [En línea]. Available: <https://www.who.int/es/news-room/q-a-detail/coronavirus-disease-covid-19>. [Último acceso: 1 noviembre 2021]

[17] M. Ligia, «Business Intelligence,» 2005.

[18] L. Yachambay, DESARROLLO DE UN BUSINESS INTELLIGENCE EN SOFTWARE LIBRE, 2015.

[19] . J. Y. Gómez Martínez y E. S. Nauca Torres, «IMPLEMENTACIÓN DE UNA PLATAFORMA DE BUSINESS INTELLIGENCE BASADO EN ANÁLISIS MULTIDIMENSIONAL PARA MONITOREAR EL COMPORTAMIENTO DE CASOS COVID 19 EN EL PERÚ, PERIODO MARZO-JULIO 2020,» PERÚ, Universidad de Lambayeque, 2020.

[20] Arrasco y Chaname, «Desarrollo de un modelo de inteligencia de negocios para apoyar la toma de decisiones en nla Clínica del Pacífico S.A,» Chiclayo - Perú, 2018.





[21]

C. Hevia y A. Neumeyer, «PNUD América Latina y el Caribe,» 20 Marzo 2020. [En línea]. Available: [www.latinamerica.undp.org](http://www.latinamerica.undp.org)