



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO

CARRERA DE INGENIERÍA DE SISTEMAS

**HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS APLICADA A DATOS
RECOLECTADOS ACERCA DEL IMPACTO DEL VIRUS SARS-COV-2 (COVID-19)
A NIVEL SOCIAL Y DE SALUD EN EL ECUADOR**

Trabajo de titulación previo a la obtención del
Título de Ingenieros de Sistemas

AUTORES: Jonathan Jair Mejía Toapanta
Carlos Augusto Aguas Ramírez

TUTOR: Walter Fernando Gaibor Naranjo

Quito – Ecuador

2022

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Jonathan Jair Mejía Toapanta con documento de identificación N.º 1721444014 y Carlos Augusto Aguas Ramírez, y N.º 1721041364; manifiesto que:

Somos los autores responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

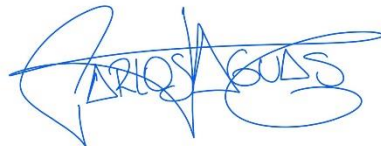
Quito, 8 de marzo del 2022

Atentamente,



Jonathan Jair Mejía Toapanta

1721444014



Carlos Augusto Aguas Ramírez

1721041364

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Nosotros, Jonathan Jair Mejía Toapanta con documento de identificación N.º 1721444014 y Carlos Augusto Aguas Ramírez, y N.º 1721041364, expresamos nuestra voluntad y por medio del presente documentos cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud que somos autores del Proyecto Técnico: “Herramienta De Inteligencia De Negocios Aplicada A Datos Recolectados Acerca Del Impacto Del Virus Sars-Cov-2 (Covid-19) A Nivel Social Y De Salud En El Ecuador”, el cual ha sido desarrollado para optar por el título de: Ingenieros de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega de trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

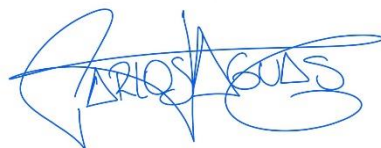
Quito, 8 de marzo del 2022

Atentamente,



Jonathan Jair Mejía Toapanta

1721444014



Carlos Augusto Aguas Ramírez

1721041364

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Walter Fernando Gaibor Naranjo con documento de identificación N.º 1713106647, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS APLICADA A DATOS RECOLECTADOS ACERCA DEL IMPACTO DEL VIRUS SARS-COV-2 (COVID-19) A NIVEL SOCIAL Y DE SALUD EN EL ECUADOR**, realizado por Jonathan Jair Mejía Toapanta con documento de identificación N.º, 1721444014 y Carlos Augusto Aguas Ramírez, y N.º 1721041364, obteniendo como resultado final el trabajo de titulación bajo la opción de Proyecto Técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 8 de marzo del 2022



Ing. Walter Fernando Gaibor Naranjo, MSc.

1713106647

ÍNDICE

| | |
|---|----|
| INTRODUCCIÓN..... | 1 |
| ANTECEDENTES | 1 |
| PROBLEMA..... | 1 |
| JUSTIFICACIÓN | 2 |
| OBJETIVOS GENERALES Y ESPECÍFICOS..... | 2 |
| OBJETIVO GENERAL..... | 2 |
| OBJETIVOS ESPECÍFICOS..... | 2 |
| PROPUESTA DE SOLUCIÓN Y ALCANCE..... | 3 |
| METODOLOGÍA | 3 |
| CAPÍTULO I..... | 5 |
| MARCO TEÓRICO | 5 |
| 1.1. SARS-COV-2..... | 5 |
| 1.2. DATA WAREHOUSE | 5 |
| 1.3. METODOLOGÍAS PARA CREAR UNA DATA WAREHOUSE..... | 6 |
| 1.3.1. METODOLOGÍA KIMBALL..... | 6 |
| 1.3.2. METODOLOGÍA HEFESTO | 6 |
| 1.3.3. <i>Etapas de la metodológica de Hefesto</i> | 7 |
| 1.3.3.1. <i>Etapa de Análisis y Requerimiento</i> | 8 |
| 1.3.3.2. <i>Etapa de Análisis OLTP</i> | 8 |
| 1.3.3.3. <i>Etapa de modelo lógico de la Data Warehouse</i> | 9 |
| 1.3.3.4. <i>Etapa de integración de datos</i> | 9 |
| 1.3.4. METODOLOGÍA SAS RAPID DATA | 9 |
| 1.4. PROCESO ETL EN UNA DATA WAREHOUSE..... | 10 |
| 1.5. MODELAMIENTO DIMENSIONAL DE UNA DATA WAREHOUSE..... | 10 |
| 1.5.1. GRANULARIDAD | 10 |
| 1.5.2. TABLA DE DIMENSIONES..... | 10 |
| 1.5.3. TABLA DE HECHOS..... | 11 |
| 1.5.4. ESQUEMAS DE UNA DATA WAREHOUSE..... | 11 |
| 1.5.4.1. <i>Esquema Estrella</i> | 11 |
| 1.5.4.2. <i>Esquema Copo de nieve</i> | 13 |
| 1.5.4.3. <i>Esquema Constelación</i> | 13 |
| 1.6. TIPOS DE IMPLEMENTACIÓN DE UNA DATA WAREHOUSE | 14 |
| 1.6.1. ROLAP..... | 14 |
| 1.6.2. MOLAP..... | 15 |

| | | |
|--|--|----|
| 1.6.3. | HOLAP..... | 16 |
| 1.7. | POSTGRESQL..... | 16 |
| 1.8. | LENGUAJE DE PROGRAMACIÓN PYTHON | 16 |
| 1.9. | INTELIGENCIA DE NEGOCIO..... | 16 |
| 1.10. | POWER BI..... | 17 |
| CAPÍTULO II | | 18 |
| ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE LA DATA WAREHOUSE | | 18 |
| 2.1. | PLANIFICACIÓN Y DISEÑO..... | 18 |
| 2.2. | PLANIFICACIÓN | 19 |
| 2.3. | FACTIBILIDAD TÉCNICA | 19 |
| 2.3.1. | HARDWARE..... | 19 |
| 2.3.2. | SOFTWARE..... | 19 |
| 2.4. | DISEÑO DE LA DATA WAREHOUSE | 20 |
| 2.4.1. | ANÁLISIS DE LOS REQUERIMIENTOS..... | 20 |
| 2.4.2. | ANÁLISIS DE LOS OLTP | 22 |
| 2.4.3. | MODELO LÓGICO DE LA DATA WAREHOUSE..... | 27 |
| 2.4.3.1. | <i>Dimensión tipo de análisis.</i> | 27 |
| 2.4.3.2. | <i>Dimensión de ubicación geográfica.</i> | 27 |
| 2.4.3.3. | <i>Dimensión del tiempo.</i> | 28 |
| 2.4.3.4. | <i>Dimensión de la población.</i> | 29 |
| 2.4.3.5. | <i>Dimensión de edad.</i> | 29 |
| 2.4.3.6. | <i>Dimensión del género.</i> | 30 |
| 2.4.3.7. | <i>Dimensión tipo de vacuna</i> | 30 |
| 2.4.3.8. | <i>Dimensión dosis</i> | 31 |
| 2.4.3.9. | <i>Tabla de hechos</i> | 31 |
| 2.4.3.9. | <i>Modelo lógico relacional dimensiones y hechos</i> | 33 |
| 2.4.4. | INTEGRACIÓN DE DATOS..... | 34 |
| 2.4.4.1. | <i>Repositorio Ecuacovid</i> | 35 |
| 2.4.4.2. | <i>Ministerio de Salud Pública</i> | 37 |
| 2.4.4.3. | <i>Extraer datos</i> | 38 |
| 2.4.4.4. | <i>Almacenamiento local de datos</i> | 38 |
| 2.4.4.5. | <i>Transformación de datos</i> | 40 |
| 2.4.4.6. | <i>Procedimientos de carga de data a las dimensiones</i> | 43 |
| 2.4.4.5.1. | Carga de datos dimensión Tipo | 43 |
| 2.4.4.5.2. | Carga de datos dimensión Tiempo | 44 |
| 2.4.4.5.3. | Carga de datos dimensión Muestra | 46 |

| | | |
|---|---|-----|
| 2.4.4.5.4. | Carga de datos dimensión Edad | 48 |
| 2.4.4.5.5. | Carga de datos dimensión Género | 49 |
| 2.4.4.5.6. | Carga de datos dimensión tipo de Vacuna | 50 |
| 2.4.4.5.7. | Carga de datos dimensión Dosis | 51 |
| 2.4.4.5.8. | Carga de datos dimensión Geografía | 52 |
| 2.4.4.7. | <i>Diseño base de datos ETL</i> | 54 |
| 2.4.4.8. | <i>Disparadores de carga de datos a la base BI_COVID</i> | 55 |
| 2.4.4.7.1. | Disparador en la tabla positivas_2020 en la base ETL..... | 56 |
| 2.4.4.7.2. | Disparador en la tabla positivas_2021 en la base ETL..... | 57 |
| 2.4.4.7.3. | Disparador en la tabla muertes_2020 en la base ETL..... | 58 |
| 2.4.4.7.4. | Disparador en la tabla muertes_2021 en la base ETL..... | 59 |
| 2.4.4.7.5. | Disparador en la tabla vacunas en la base ETL. | 60 |
| 2.4.4.9. | <i>Funciones de carga de datos a la base ETL</i> | 61 |
| 2.4.4.9.1. | <i>Conexión Python con PostgreSQL</i> | 62 |
| 2.4.4.9.2. | <i>Carga de datos a la tabla positivas_2020 en la base ETL</i> | 63 |
| 2.4.4.9.3. | <i>Carga de datos a la tabla positivas_2021 en la base ETL</i> | 63 |
| 2.4.4.9.4. | <i>Carga de datos a la tabla muertes_2020 en la base ETL</i> | 64 |
| 2.4.4.9.5. | <i>Carga de datos a la tabla muertes_2021 en la base ETL</i> | 65 |
| 2.4.4.9.6. | <i>Carga de data a la tabla vacunas en la base ETL</i> | 66 |
| CAPÍTULO III | | 67 |
| HERRAMIENTA DE INTELIGENCIA DE NEGOCIO | | 67 |
| 3.1. | CONEXIÓN POWER BI Y POSTGRESQL | 67 |
| 3.2. | DISEÑO | 72 |
| 3.2.1. | PANTALLA PRINCIPAL | 72 |
| 3.2.2. | INTERFAZ GRÁFICA DE CASOS POSITIVOS | 73 |
| 3.2.3. | INTERFAZ GRÁFICA DE MUERTES | 74 |
| 3.2.4. | INTERFAZ GRÁFICA DE VACUNAS | 76 |
| CAPÍTULO IV | | 78 |
| ANÁLISIS DE RESULTADOS | | 78 |
| 4.1. | Resultado del eje de análisis de casos positivos | 78 |
| 4.2. | Resultado del eje de análisis de muertes | 96 |
| 4.3. | Resultado del eje de análisis de vacunados | 102 |
| CAPÍTULO V | | 111 |
| CONCLUSIONES Y RECOMENDACIONES | | 111 |
| 5.1. | CONCLUSIONES | 111 |
| 5.2. | RECOMENDACIONES | 112 |

ÍNDICE DE TABLAS

| | |
|--|-----|
| Tabla 1: Región con el mayor índice de contagios en el año 2020..... | 78 |
| Tabla 2: Índice de casos positivos por provincia en el primer trimestre de la pandemia..... | 80 |
| Tabla 3: Provincia con mayor tasa de casos positivos al finalizar el año 2020. | 82 |
| Tabla 4: Provincia de cada región con mayor índice de casos positivos, año 2020 | 84 |
| Tabla 5: Cantón con mayor índice de casos positivos en el primer semestre del año 2020. .. | 88 |
| Tabla 6: Cantón con el mayor índice de casos positivos en el año 2020, región oriente..... | 89 |
| Tabla 7: Cantón con el mayor índice de casos positivos en la pandemia | 91 |
| Tabla 8: Comparativa top 10 provincias de registro de casos positivos, año 2020, 2021 | 93 |
| Tabla 9: Provincia que presenta el mayor índice de muertes en el primer trimestre de la pandemia | 96 |
| Tabla 10: Provincia que presenta el mayor índice de muertes en el año 2020 | 97 |
| Tabla 11: Evolución de muertes en la región sierra del año 2020 al 2021 | 99 |
| Tabla 12: Vacuna con el mayor índice de uso | 102 |
| Tabla 13: Índices totales de cada tipo de vacuna aplicadas como primera dosis..... | 104 |
| Tabla 14: Provincia con la mayor tasa de vacunación | 106 |
| Tabla 15: Análisis de vacuna Sinovac aplicada. | 107 |
| Tabla 16: Análisis de vacuna AstraZeneca aplicada. | 108 |
| Tabla 17: Análisis de la vacuna Pfizer..... | 109 |

ÍNDICE DE FIGURAS

| | |
|--|----|
| Figura 1: Arquitectura de una Data Warehouse..... | 5 |
| Figura 2: Etapas de la metodología Hefesto | 7 |
| Figura 3: Esquema Estrella | 11 |
| Figura 4: Tabla de dimensiones | 12 |
| Figura 5: Tabla de hechos | 12 |
| Figura 6: Esquema Copo de nieve | 13 |
| Figura 7: Esquema Constelación | 14 |
| Figura 8: Flujo de solicitud de información de la herramienta BI con Molap..... | 15 |
| Figura 9: Modelo conceptual etapa 1 de la Data Warehouse..... | 22 |
| Figura 10: Perspectiva del Tipo de análisis | 23 |
| Figura 11: Perspectiva del Tiempo | 23 |
| Figura 12: Perspectiva de Geografía..... | 24 |
| Figura 13: Perspectiva de Género | 24 |
| Figura 14: Perspectiva de la Población..... | 24 |
| Figura 15: Perspectiva de Edad..... | 25 |
| Figura 16: Perspectiva del tipo de Vacuna | 25 |
| Figura 17: Perspectiva de Dosis..... | 25 |
| Figura 18: Modelo conceptual etapa 2 de la Data Warehouse..... | 26 |
| Figura 19: Tabla dimensión tipo de análisis | 27 |
| Figura 20: Tabla dimensión de ubicación geográfica | 28 |

| | |
|--|----|
| Figura 21: Tabla dimensión del tiempo. | 28 |
| Figura 22: Tabla dimensión de la población. | 29 |
| Figura 23: Tabla dimensión de edad. | 29 |
| Figura 24: Tabla dimensión del género. | 30 |
| Figura 25: Tabla dimensión del tipo de vacuna. | 30 |
| Figura 26: Tabla dimensión dosis. | 31 |
| Figura 27: Tabla de hechos de la Data Warehouse. | 32 |
| Figura 28: Datos cargados a la Tabla de hechos. | 32 |
| Figura 29: Modelo lógico de la Data Warehouse. | 34 |
| Figura 30: Flujo del proceso ETL. | 35 |
| Figura 31: Archivo para casos positivos 2020. | 36 |
| Figura 32: Archivo para casos positivos 2021. | 36 |
| Figura 33: Archivo para muertes 2020. | 36 |
| Figura 34: Archivo para muertes 2021. | 36 |
| Figura 35: Archivos del MSP. | 37 |
| Figura 36: Enlace de clonación del repositorio Ecuacovid. | 38 |
| Figura 37: Clonar repositorio Ecuacovid. | 39 |
| Figura 38: Almacenamiento de datos a explotar. | 39 |
| Figura 39: Buscar archivos de texto desde Excel. | 40 |
| Figura 40: Importar archivo Excel seleccionado. | 41 |
| Figura 41: Configuración del asistente de importar texto. | 41 |
| Figura 42: Configuración del formato del archivo generado. | 42 |
| Figura 43: Procedimiento para la dimensión Tipo. | 43 |
| Figura 44: Datos cargados en la dimensión Tipo. | 44 |
| Figura 45: Procedimientos para la dimensión Tiempo. | 45 |
| Figura 46: Datos cargados en la dimensión tiempo. | 46 |
| Figura 47: Procedimiento para la dimensión Población. | 47 |
| Figura 48: Datos cargados en la dimensión Población. | 47 |
| Figura 49: Procedimiento para la dimensión Edad. | 48 |
| Figura 50: Datos cargados en la dimensión edad. | 49 |
| Figura 51: Procedimiento para la dimensión Género. | 49 |
| Figura 52: Datos cargados en la dimensión género. | 50 |
| Figura 53: Procedimiento para la dimensión tipo de vacuna. | 50 |
| Figura 54: Datos cargados en la dimensión vacuna. | 51 |
| Figura 55: Procedimiento para la dimensión dosis. | 51 |
| Figura 56: Datos cargados en la dimensión dosis. | 52 |
| Figura 57: Conexión entre Python y la base de datos BI_COVID. | 53 |
| Figura 58: Carga de datos del archivo a la dimensión de geografía. | 53 |
| Figura 59: Datos cargados en la dimensión geografía. | 54 |
| Figura 60: Tablas de la base de datos ETL. | 55 |
| Figura 61: Disparador carga_positiva_2020. | 57 |
| Figura 62: Disparador carga_positiva_2021. | 58 |
| Figura 63: Disparador carga_muerte_2020. | 59 |
| Figura 64: Disparador carga_muerte_2021. | 60 |
| Figura 65: Disparador carga_vacunas. | 61 |
| Figura 66: Conexión Python con base ETL. | 62 |
| Figura 67: Carga de datos a positivas_2020. | 63 |

| | |
|---|-----|
| Figura 68: Carga de datos a positivas_2021. | 64 |
| Figura 69: Carga de datos a muertes_2020. | 65 |
| Figura 70: Carga de datos a muertes_2021. | 65 |
| Figura 71: Carga de datos a vacunas. | 66 |
| Figura 72: Pantalla de bienvenida de Power BI. | 67 |
| Figura 73: Gestores de base de datos de Power BI. | 68 |
| Figura 74: Conexión a PostgreSQL desde Power BI. | 69 |
| Figura 75: Esquemas para cargar al entorno de Power BI. | 70 |
| Figura 76: Relación de los esquemas cargados a Power BI. | 71 |
| Figura 77: Pantalla principal de la herramienta Power BI. | 72 |
| Figura 78: Interfaz gráfica del eje de análisis de casos positivos. | 74 |
| Figura 79: Interfaz gráfica del eje de análisis muertes. | 75 |
| Figura 80: Interfaz gráfica del eje de análisis vacunados. | 77 |
| Figura 81: Análisis positivos región Costa año 2020. | 78 |
| Figura 82: Análisis positivos región Sierra año 2020. | 79 |
| Figura 83: Análisis positivos región Oriente año 2020. | 79 |
| Figura 84: Análisis positivos región Galápagos año 2020. | 79 |
| Figura 85: Análisis de casos positivos por provincias primer trimestre 2020. | 80 |
| Figura 86: Análisis de casos positivos por provincias al finalizar el año 2020. | 82 |
| Figura 87: Provincia con mayor índice de casos positivos región Costa. | 84 |
| Figura 88: Provincia con mayor índice de casos positivos región Sierra. | 85 |
| Figura 89: Provincia con mayor índice de casos positivos región Oriente. | 86 |
| Figura 90: Provincia con mayor índice de casos positivos región Galápagos. | 87 |
| Figura 91: Cantón con mayor índice de casos positivos hasta el mes de junio 2020. | 88 |
| Figura 92: Cantón con mayor índice de casos positivos año 2020, región Oriente. | 89 |
| Figura 93: Cantón con el mayor índice de contagios en la pandemia. | 91 |
| Figura 94: Top provincias con mayor índice de contagio 2021. | 93 |
| Figura 95: Top provincias con mayor índice de contagio 2020. | 94 |
| Figura 96: Análisis del primer trimestre de pandemia. | 96 |
| Figura 97: Análisis de índice de muertes en el año 2020. | 97 |
| Figura 98: Análisis de muertes región sierra a mes de junio del 2020. | 99 |
| Figura 99: Análisis de muertes región sierra a mes de diciembre del 2020. | 99 |
| Figura 100: Análisis de muertes región sierra a mes de julio del 2021. | 100 |
| Figura 101: Gráfica de áreas del índice de uso de cada vacuna. | 102 |
| Figura 102: Análisis de AstraZeneca, primera dosis. | 104 |
| Figura 103: Análisis de Pfizer, primera dosis. | 104 |
| Figura 104: Análisis de Sinovac, primera dosis. | 105 |
| Figura 105: Análisis de la provincia con mayor índice de vacunados. | 106 |
| Figura 106: Análisis de la vacuna Sinovac. | 107 |
| Figura 107: Análisis de vacuna AstraZeneca. | 108 |
| Figura 108: Análisis de la vacuna Pfizer. | 109 |

RESUMEN

El presente proyecto de titulación se basa en el análisis y desarrollo de una herramienta de inteligencia de negocios, que apoye al procesamiento de datos extraídos de los tres ejes principales de estudio, casos positivos, muertes y vacunados relacionados con el virus SARS-COV-2 dentro del territorio de Ecuador, los datos son extraídos desde reportes generados por el Sistema Nacional de Gestión de Riesgos del Ecuador (SNGRE) y el Ministerio de Salud Pública (MSP). La herramienta de inteligencia de negocios proporciona tres vistas gráficas e intuitivas cada una para cada eje de análisis.

Cada eje de análisis facilita identificar diferentes perspectivas como: provincias con constantes índices elevados de niveles de registro de casos positivos, provincias con el mayor índice de contagios confirmados, muertes y vacunados por región, realizar comparativas de problemáticas sociales y médicas para cada resultado identificado bajo los filtros descritos como dimensiones, cantones con una respuesta baja ante la emergencia sanitaria al verificar su curva de casos verificados, muertes y vacunados entre diferentes periodos de tiempo, facilita el comprobar los resultados con los informes generados por el SNGRE y MSP.

Como resultado del análisis generado para cada eje, se obtuvo a la provincia de Pichincha como la provincia que representa el 35.7% de registro como casos positivos. Además de presentar a la región sierra con el mayor índice de muertes, enfocándose en la capital como la provincia con el mayor número de muertes registradas.

ABSTRACT

This degree project is based on the analysis and development of a business intelligence tool, which supports the analysis of data extracted from the three main axes of analysis, positive cases, deaths and vaccinated people related to the SARS-COV-2 virus within from the territory of Ecuador, the data is extracted from reports generated by the National Risk Management System of Ecuador (SNGRE) and the Ministry of Public Health (MSP). The business intelligence tool provides three graphical and intuitive views each for each axis of analysis.

Each axis of analysis facilitates the identification of different perspectives such as: provinces with constantly high rates of registration levels of positive cases, provinces with the highest rate of positive cases, deaths and vaccinated by region, making comparisons of social and medical problems for each result identified under The filters described as dimensions, cantons with a low response to the health emergency when verifying their curve of positive, dead and vaccinated cases between different periods of time, make it easier to check the results with the reports generated by the SNGRE and MSP.

As a result of the analysis generated for each axis, the province of Pichincha was obtained as the province that represents 35.7% of the registry as positive cases. In addition to presenting the sierra region with the highest death rate, focusing on the capital as the province with the highest number of registered deaths.

INTRODUCCIÓN

ANTECEDENTES

El Servicio Nacional de Gestión de Riesgos y Emergencias del Ecuador (SNGRE) es aquel que garantiza la protección de las personas dentro del territorio del Ecuador, por medio de la creación de políticas, estrategias y normas para controlar y velar por la seguridad de la ciudadanía.

El SNGRE es el organismo que presenta la información consolidada de afectaciones y acciones de respuesta ejecutadas por las instituciones públicas y privadas que conforman el Sistema Nacional Descentralizado de Gestión de Riesgos (SNDGR), referente a la emergencia nacional por la COVID-19 (SNGRE, 29 de febrero del 2020).

Desde el 13 de marzo del 2020 se presentan informes de la situación relacionada a la pandemia e infografías con fuente de datos proporcionados por el Ministerio de Salud Pública (MSP) en los cuales se detallan el número de casos positivos por contagio diario, además del número de fallecidos en la pandemia, y el reporte semestral de vacunados.

PROBLEMA

La principal problemática que se presenta está relacionada al manejo de la información que se publica en el portal de gestión de riesgos proporcionada por el MSP se encuentra en archivos de formato PDF, lo que dificulta analizar en el ámbito de inteligencia de negocios y posteriormente presentar de forma dinámica la evolución que ha tenido el COVID-19, con propuestas a tomar en cuenta o volver a evaluar decisiones tomadas.

Esta información se encuentra almacenada en un repositorio de GitHub, los archivos se encuentran en formato de texto plano los cuales se han generado a partir de la información proporcionada en los PDF de la infografía.

JUSTIFICACIÓN

A raíz de la pandemia declarada por el COVID-19 en 2020, dentro de Ecuador se generan varios aspectos sociales de inconformidad en cuanto a la veracidad y divulgación de la información que se presenta relacionada con la evolución en tema de contagios, recuperaciones, muertes y posterior vacunación.

Bajo el fundamento del poder competitivo que debe tener una empresa al manejar una gran cantidad de información para apoyar la toma de decisiones, o bien para la presentación visual de manera más comprensiva y dinámica de la información; se encuentra la necesidad de implementar una herramienta de Inteligencia de Negocios en la cual se aprovecha los datos alojados en una base de datos de un sistema transaccional que a su vez contendrá los datos de los informes extraídos desde el repositorio de GitHub en el que se encuentran alojados los informes de ECUACOVID.

OBJETIVOS GENERALES Y ESPECÍFICOS

OBJETIVO GENERAL

Efectuar un estudio y posterior análisis mediante la utilización de herramientas de inteligencia de negocios, con base en los datos obtenidos desde ECUACOVID y Ministerio de Salud Pública, para realizar la presentación de los informes de manera dinámica y así facilitar una mejor comprensión visual de esta información.

OBJETIVOS ESPECÍFICOS

- Implementar una herramienta que ayude al modelamiento y análisis de los datos extraídos de las diferentes fuentes de información.
- Ejecutar pruebas rutinarias y comparativas con los informes dados por el MSP para evaluar la precisión de los resultados extraídos.

- Presentar de manera dinámica la evolución que ha tenido el Covid-19 dentro de Ecuador.
- Presentar propuestas a la toma de decisiones, basados en los resultados del análisis de la Data Warehouse

PROPUESTA DE SOLUCIÓN Y ALCANCE

El objetivo de este proyecto es generar una herramienta de inteligencia de negocios, que permita la recolección de datos obtenidos desde ECUACOVID y del Ministerio de Salud Pública, que se apoye sobre el uso de una base de datos para facilitar el análisis de los datos, la limpieza y clasificación por homogeneidad, y de esta manera haciendo uso de sentencias Python completar el proceso de extracción, transformación y carga (ETL), para facilitar la carga de la data desde la base de datos, para que continúe con la carga de datos a la tabla de hechos de la Data Warehouse, los resultados facilitan la predicción de soluciones a tomar en cuenta dentro de los ámbitos sociales y médicos, relacionados con la emergencia sanitaria existente en Ecuador.

METODOLOGÍA

Scrum se ha convertido en un referente principal como marco de trabajo entre las metodologías ágiles, ya que su enfoque se basa en el esfuerzo colaborativo en equipos, esto facilita la elaboración y adaptación continua a los requisitos del usuario. Presenta una estructura de desarrollo por ciclos, definidos como Sprints, tienen un tiempo de duración que va desde una hasta cuatro semanas como máximo, cada sprint se complementa por medio de una reunión de planificación en el cual se expone el alcance y la manera el que se van a ejecutar los avances para alcanzar el resultado final.

Las reuniones diarias dentro de cada sprint mejoran la comunicación continua en la cual se dan a conocer los avances, el plan que se está siguiendo y los obstáculos encontrados en el desarrollo, estas reuniones no deben sobrepasar un máximo de 15 minutos.

Al finalizar cada sprint se realiza una retrospectiva con todo el equipo de scrum y el dueño del producto para analizar y verificar el cumplimiento del avance y de ser el caso también de implementar un plan de mejora para el siguiente, tomando en cuenta las observaciones de cada participante en el proyecto. (Subra y Vannieuwenhuyse, 2018).

CAPÍTULO I

MARCO TEÓRICO

1.1. SARS-COV-2

En la página oficial de la Organización Mundial de la Salud (OMS), se reconoce al virus SARS-CoV-2 (COVID-19) como una enfermedad infecciosa altamente contagiosa, proveniente de la ciudad de Wuhan en China en el mes de diciembre del 2019. (<https://www.who.int/es>)

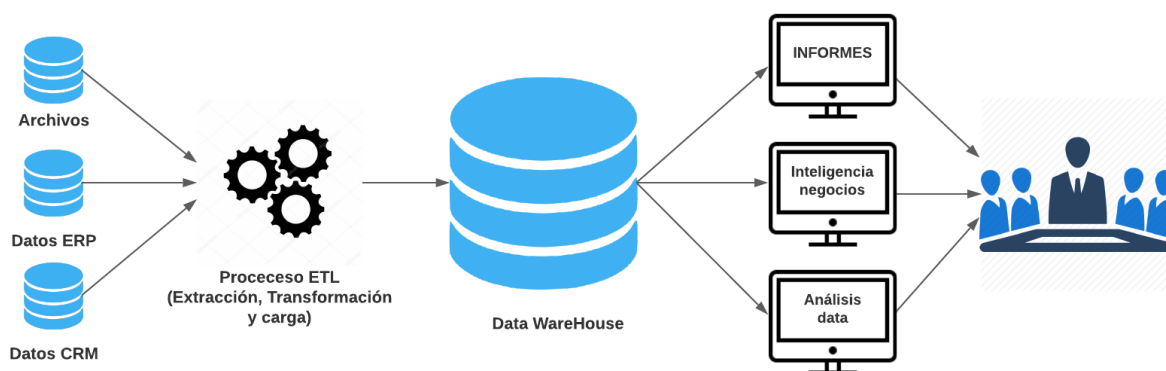
1.2. DATA WAREHOUSE

Es un sistema centralizado que aloja datos obtenidos de las diferentes fuentes de información del negocio, estos sistemas pueden ser transaccionales y convencionales, con el fin de respaldar un futuro análisis a nivel empresarial, con un aporte total a la toma de decisiones.

Según Ortí Rodríguez, J. (2021), “es una colección de datos de distintos orígenes, estructurada y orientada al uso en un determinado ámbito empresarial siendo integrada no volátil y variable en el tiempo” (p. 15).

Figura 1

Arquitectura de una Data Warehouse.



Nota. Modelo del flujo de proceso de un almacén de datos. Fuente: (Mejía J. y Aguas C., 2021)

1.3. METODOLOGÍAS PARA CREAR UNA DATA WAREHOUSE

Para el diseño e implementación de una Data Warehouse y posterior integración con Power BI, se presenta distintas metodologías con su respectiva ventaja, que son:

1.3.1. METODOLOGÍA KIMBALL

También conocida modelo dimensional, parte de la creación de una Data Mart por cada área del negocio que se comunican entre sí a través de elementos comunes, la carga de datos se realiza mediante los procesos de extracción, transformación y carga (ETL), esta metodología facilita que la fuente de datos sea integrada y no volátil, con el objetivo principal de aportar a la toma de decisiones. (Peñañiel S, Geovanny. E, 2018).

1.3.2. METODOLOGÍA HEFESTO

Peñañiel, Yáñez, Guamán, Padilla (2019) plantean que la base de esta metodología son cuatro fases principales: Análisis de requerimientos: define los indicadores y el modelo conceptual, Análisis de los OLTP: identifica el nivel de granularidad y amplía el modelo anterior, Modelo lógico de la Data Warehouse: desarrolla las tablas de dimensiones y la tabla de hechos, e integración de datos: ejecuta la carga y actualización de datos (ETL).

Para Bernabeu, R. D. (2010) las características principales de esta metodología son:

- Los alcances por etapa son fáciles de distinguir y comprender.
- Contiene una estructura que facilita la adaptación de manera ágil y sencilla, así adaptándose a los requerimientos del usuario.
- Involucra a los dueños del producto al finalizar cada etapa, para dar a conocer el comportamiento de la Data Warehouse y evitar cambios a futuro.
- Es independiente del ciclo de vida empleado en la metodología, de las herramientas de implementación, y de las características físicas que tenga el Data Warehouse.

- Al finalizar una etapa, el resultado se convierte en el punto de inicio para continuar con la siguiente etapa.

Figura 2

Etapas de la metodología Hefesto.



Nota: Estructura por etapas de la metodología Hefesto. Fuente: (Bernabeu, R. D. ,2010).

1.3.3. Etapas de la metodológica de Hefesto

La Figura 2 describe cuatro fases para el desarrollo de una Data Warehouse, entre las características destaca que al finalizar una etapa y verificar la ejecución correcta de la misma, se convierte en el punto de partida para la siguiente, Bernabeu, R. D. (2010). describe de la siguiente manera cada etapa:

1.3.3.1. *Etapa de Análisis y Requerimiento.*

El alcance final de esta etapa es: obtener un modelado conceptual, a partir de identificar explícitamente los objetivos que tiene la empresa y definir posibles indicadores de análisis para construir el Data Warehouse.

- Para el proceso de identificar preguntas se utiliza diferentes técnicas de acopio de información, esto permite realizar preguntas complejas, que ayudan a identificar información clave y estratégica de la empresa.
- El reconocimiento de los indicadores de análisis se basa en la información numérica y que se relacione con el alcance final del análisis.
- El modelo conceptual presenta los diferentes alcances identificados y el trabajo sobre ellos.

1.3.3.2. *Etapa de Análisis OLTP.*

El objetivo para esta etapa es: ampliar el modelo conceptual, al identificar la manera para calcular el resultado para los indicadores.

- Un indicador se forma a partir de la identificación de un Hecho y la función para el cálculo del análisis.
- Para identificar las correspondencias, se debe identificar cada OLTP con sus respectivas características, esto facilita el análisis de la relación entre el modelo conceptual y la fuente de datos.
- El nivel de granularidad se obtiene del detalle y valor de los datos encontrados en los OLTP.
- El modelo conceptual ampliado muestra la relación de cada indicador y su fórmula para el análisis.

1.3.3.3. *Etapa de modelo lógico de la Data Warehouse.*

Para realizar el modelamiento lógico, se debe definir qué tipo de modelamiento se va a utilizar.

- Para el modelo lógico de la Data Warehouse, se define el tipo de esquema que mejor se adapte al tipo de relaciones que dispongan la tabla de hechos con las tablas de las dimensiones.
- La tabla de dimensiones representa cada análisis del negocio.
- La tabla de hechos contiene las claves de identificación de cada tabla dimensión que representan los indicadores.
- Las uniones señalan la relación entre la tabla de hechos y dimensiones.

1.3.3.4. *Etapa de integración de datos.*

El alcance total de esta etapa es el de ejecutar el procesamiento ETL.

- La carga inicial es la acción de poblar el modelo lógico construido, se define el software y las funciones que usará para esta carga, después de haber transformado los datos.
- La actualización de datos se realiza después de realizar la carga total a la Data Warehouse, se definen las políticas y estrategias de actualización.

1.3.4. METODOLOGÍA SAS RAPID DATA

Según Peñafiel Silva, G. (2018) manifiesta que “SAS proporciona un almacén de datos que abarca ámbitos como: Gestión de almacenes, organización y explotación, consta de una arquitectura de motor que facilita el desarrollo de sistemas transaccionales, se basa en la experiencia adquirida por cada Hito”.

1.4. PROCESO ETL EN UNA DATA WAREHOUSE

El proceso de extracción, transformación y carga, es parte fundamental dentro de la arquitectura de la Data Warehouse, se encarga de obtener y recoger los datos de las diferentes fuentes de información o de los distintos sistemas transaccionales, esto permite continuar a la etapa de transformación en la cual los datos obtenidos deben ser normalizados, estandarizados, para facilitar la producción de datos procesados, al finalizar el procesamiento de datos se continúa con la etapa de carga de datos para la Data Warehouse. (Trujillo, Mazón, & Pardillo, 2013).

1.5. MODELAMIENTO DIMENSIONAL DE UNA DATA WAREHOUSE

El modelamiento aborda todas las necesidades básicas de un almacén o sistemas de la empresa, garantiza el proceso analítico de datos; el modelamiento necesita de:

1.5.1. GRANULARIDAD

En el escrito realizado por Bernabeu, R. D. (2010) se describe que la granularidad “representa el nivel de detalle al que se desea almacenar la información sobre el negocio. Mientras mayor sea el nivel de detalle se tiene una mejor analítica de datos”, esto indica que mientras mayor sea el detalle del sistema transaccional o área del negocio mayor es la información que se pueda presentar.

1.5.2. TABLA DE DIMENSIONES

La finalidad de las tablas de las dimensiones es de almacenar información descriptiva de las distintas áreas del negocio, estos datos se cargan mediante el proceso ETL, la información alojada en estas tablas, se presenta al generar informes o presentar el resultado de consultas.

Para Curto Díaz, J. (2016), las dimensiones son los puntos de análisis de un hecho, por ejemplo la cantidad de contagios por el virus Covid-19 puede ser analizado por tiempo y ubicación geográfica.

1.5.3. TABLA DE HECHOS

La tabla de hechos simbolizan las métricas para analizar en el proceso de inteligencia de negocios, también representan los campos claves de las tablas de las dimensiones; evitan la redundancia de atributos.

En el escrito realizado por Curto Díaz, J. (2016), reconoce a los hechos como “una representación de un proceso de negocio”; en las cuales se almacena los eventos que se van a representar y las claves foráneas de las dimensiones

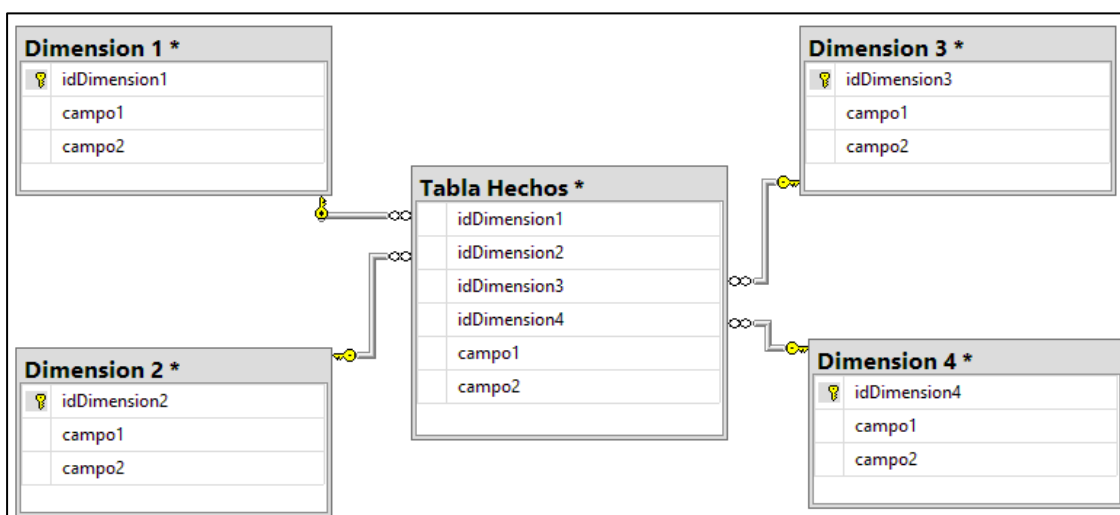
1.5.4. ESQUEMAS DE UNA DATA WAREHOUSE

1.5.4.1. Esquema Estrella.

El esquema estrella centraliza la tabla de hechos, ubicadas alrededor de la tabla de hechos se encuentran las tablas de las dimensiones, presentan una relación por cada dimensión con la tabla de hechos, Manso J. (2021) destaca el esquema estrella por la velocidad y simplicidad al momento de realizar trabajo de consultas, al proporcionar una base de tipo relacional.

Figura 3

Esquema Estrella.



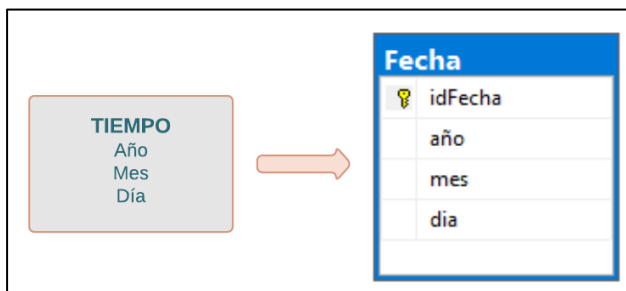
Nota: Estructura de la relación entre la tabla de hechos (central) y las tablas de dimensiones.

Fuente: (Manso J., 2021).

Para el esquema estrella Bernabeu, R. D. (2010), describe que la tabla de dimensiones debe tener las siguientes características, un nombre que identifique el tipo de dimensión, una clave principal y los campos necesarios para obtener resultados de los análisis, en otras palabras, se transforma la perspectiva a una nueva tabla (Figura 4). Para la tabla de hechos el nombre de identificación debe representar el área analizada, se define una clave principal, y los campos necesarios para relacionar con las claves principales de las dimensiones (Figura 5).

Figura 4

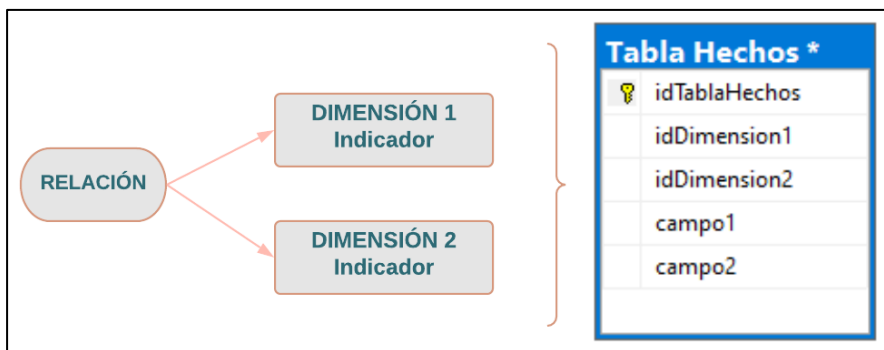
Tabla de dimensiones.



Nota: Cambio conceptual de una perspectiva a una tabla de dimensión. Fuente: (Bernabeu, R. D. 2010).

Figura 5

Tabla de hechos.



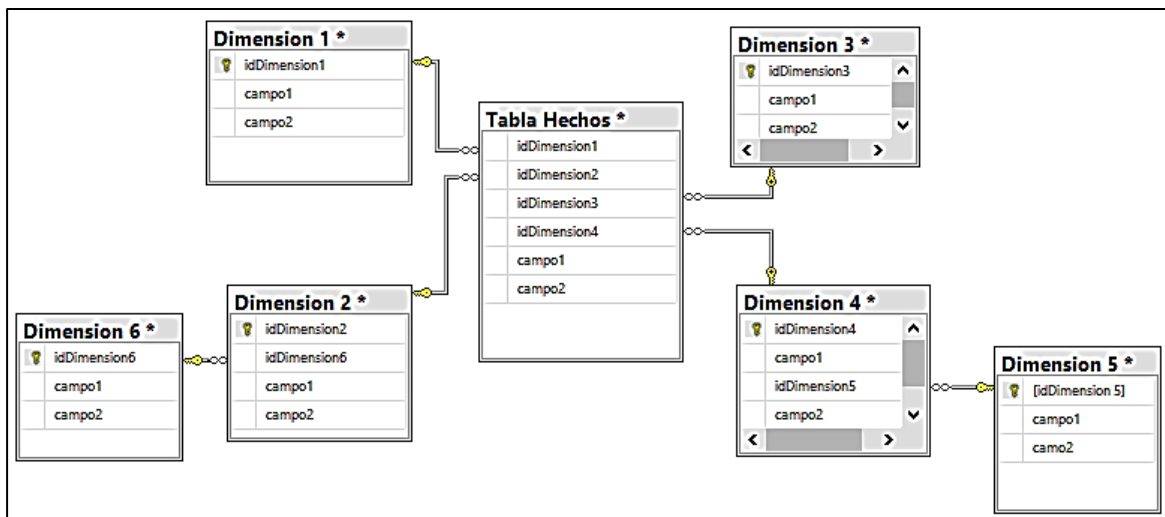
Nota: Paso de un modelo conceptual de dimensiones y la relación a una tabla de hechos. Fuente: (Bernabeu, R. D. 2010).

1.5.4.2. Esquema Copo de nieve

Su estructura es similar al esquema estrella, con la característica que este esquema permite integrar nuevas tablas de dimensiones que están relacionadas directamente a otra tabla de dimensión, estas tablas presentan una organización jerárquica. (Hernández S., 2019).

Figura 6

Esquema Copo de nieve.



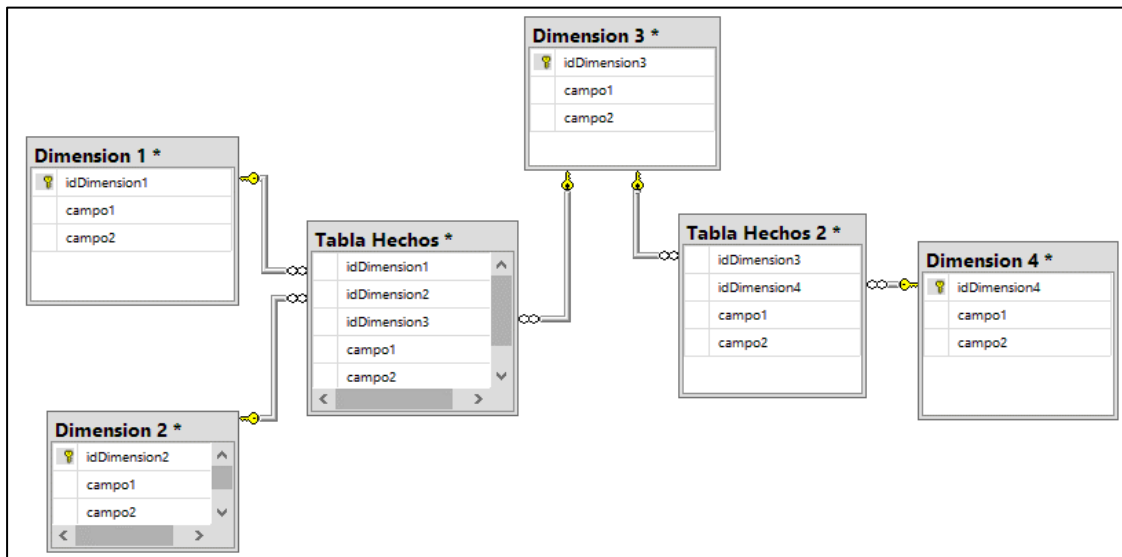
Nota: Ejemplo de la estructura que presenta la relación entre la tabla de hechos y las tablas dimensiones en un esquema copo de nieve para una Data Warehouse. Fuente: (Manso J., 2021).

1.5.4.3. Esquema Constelación

Este esquema destaca la presencia de otra tabla de hechos que comparten la relación entre dimensiones, para este caso las tablas de dimensiones no necesitan estar normalizadas.

Figura 7

Esquema Constelación.



Nota: Ejemplo de la estructura que presenta la relación en un esquema constelación entre las tablas de hechos y las dimensiones para una Data Warehouse. Fuente: (Manso J., 2021).

1.6. TIPOS DE IMPLEMENTACIÓN DE UNA DATA WAREHOUSE

Para la implementación de la Data Warehouse debe estar implícito la ejecución completa del proceso ETL, la data cargada permite que se realice la explotación del Data Warehouse mediante la técnica de Query y Reporting y OLAP para ello se cuenta con:

1.6.1. ROLAP

La estructura de Procesamiento Analítico Online Relacional (ROLAP, siglas en inglés), se constituye por sistemas y herramientas OLAP, construidas sobre base de datos de tipo sistema gestor de base de datos (SGBD), entre sus características está, seleccionar indicadores que facilitan la construcción multidimensional, ejecutar consultas sobre los indicadores desde la tabla de hechos para explotar los datos.

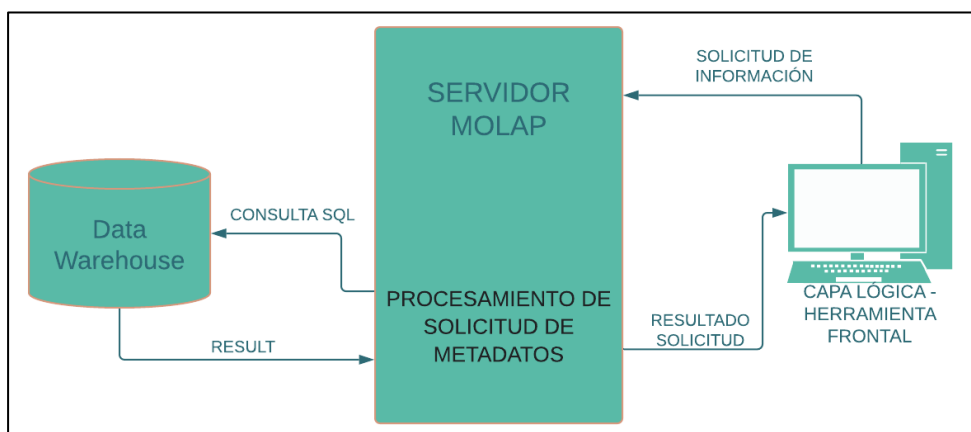
1.6.2. MOLAP

En esta estructura de Procesamiento Analítico Multidimensional en Línea (MOLAP, siglas en inglés), según Benabeu Ricardo (2010), “El objetivo es almacenar físicamente los datos en estructuras multidimensionales de manera que la representación externa y la interna coincidan”, esto permite que las consultas para explotar el Data Warehouse sean con mucha rapidez, esto también representa que para actualizar o modificar un cubo se debe reestructurar desde cero, además representa un mayor espacio de almacenamiento.

Lachev, T. (2005), indica entre las ventajas de Molap que presenta mayor rendimiento al realizar solicitudes de información desde la capa lógica de la herramienta frontal, el servidor Molap procesa las solicitudes y realiza las respectivas consultas a la Data Warehouse que se encuentra alojado del lado del servidor, al obtener el resultado el servidor Molap presenta los resultados en la capa lógica de la herramienta frontal.

Figura 8

Flujo de solicitud de información de la herramienta BI con Molap.



Nota: Flujo del funcionamiento de peticiones desde una herramienta BI en Molap. Fuente: (Mejía y Aguas, 2021).

1.6.3. HOLAP

El Procesamiento Analítico en Línea Híbrido, se basa en la combinación de ROLAP Y MOLAP, estructura el almacenamiento tanto de tipo ROLAP, como de tipo MOLAP en la misma Data Warehouse, para Benabeu Ricardo (2010) al usar ROLAP se puede navegar y explorar los datos, y MOLAP es utilizado para crear tableros.

1.7. POSTGRESQL

Este sistema es una base de tipo relacional con el enfoque orientado a objetos, de código abierto, presenta una comunidad propia de desarrolladores, en términos económicos es de gran beneficio, ya que es una potente herramienta de gestión de base de datos con sus versiones gratuitas, en el escrito de Ordoñez, M. (2017) se menciona que emplea un modelo cliente/servidor que a su vez hace uso de multiprocesos, lo que facilita la continua ejecución del mismo, puesto que si existe una falla en algún proceso este no afecta al resto.

1.8. LENGUAJE DE PROGRAMACIÓN PYTHON

Es un lenguaje de alto nivel, multiplataforma y de código abierto, facilita el trabajo en áreas de IA, Big Data, Machine Learning, ciertas características a destacar son: la sintaxis simple, legibilidad sencilla, mayor nivel de abstracción, presenta una amplia gama de librerías además de permitir el uso de terceros, Multiparadigma. (García Monsálvez, J. C., 2017).

1.9. INTELIGENCIA DE NEGOCIO

La necesidad de transformar la información de una empresa y que el procesamiento de esta facilite la toma de decisiones estratégicas, operativas o tácticas, conlleva a que en la mayoría de ámbitos empresariales se adapte herramientas para la inteligencia de negocios. En el escrito de Garcete, Benítez, Pinto y Vazquez (2017), plantea como objetivos principales, el procesamiento de datos para obtener información de conocimiento, facilitar la disponibilidad

de la información de manera entendible e interactiva y disminuir la incertidumbre en la toma de decisiones estratégicas.

La facilidad que ofrece para la toma de decisiones al realizar el respectivo análisis de datos, ayuda a aportar o guiar las acciones con las que se pueden enfrentar respectivas situaciones del negocio, en este caso el análisis de las diferentes problemáticas ligadas a la alerta de la pandemia por el Covid-19 dentro de Ecuador.

1.10. POWER BI

Microsoft presenta a Power BI como una potente herramienta de análisis de datos necesarios en cada momento, para obtener resultados de apoyo para la toma de decisiones rápida y eficaz, presenta una plataforma intuitiva para facilitar la conectividad con base de datos externas y trabajar mediante consultas SQL y realizar la presentación con una amplia herramienta gráfica. (Microsoft Power BI, 2021)

CAPÍTULO II

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE LA DATA WAREHOUSE

2.1. PLANIFICACIÓN Y DISEÑO

Este capítulo detalla el procedimiento que tuvo el proyecto para la planificación y posterior diseño de la Data Warehouse, presenta el objetivo del apoyo de herramientas y estrategias utilizadas para cumplir con los objetivos planteados y completar el alcance final, describe el proceso de recolección de datos de las diferentes fuentes de información, manteniendo la integridad de la misma, realizar el análisis de los diferentes ejes de estudio relacionado con el virus SARS-COV-2, los enfoques de análisis se presenta como: registro de casos positivos, historial de personas vacunadas y registro de muertes en el Ecuador, posteriormente facilita la construcción de la Data Warehouse y diseño de la herramienta Power BI, la interfaz gráfica para el usuario final pretende presentar de manera clara y precisa los resultados, ofreciendo así una herramienta en la que pueda realizar los respectivos análisis.

Los datos que han sido extraídos de diferentes fuentes confiables, contienen información de los ejes de análisis; El Ministerio de Salud Pública del Ecuador (MSP) proporciona cifras reales de la evolución de vacunados en el Ecuador. El repositorio ECUACOVVID proporciona datos sin procesar extraídos de los informes de la situación nacional frente al COVID-19, los informes se extraen de páginas oficiales del gobierno del Ecuador como son: Servicio nacional de gestión de riesgos y emergencias del Ecuador (SNGRE), Ministerio de Salud Pública del Ecuador (MSP), y Registro civil del Ecuador.

Los datos recolectados por ECUACOVVID representan la fuente principal de información implementada para el desarrollo del proyecto, los datos del repositorio presentan actualizaciones continuas, es decir existen datos desde el 03 de marzo del 2020 hasta la fecha actual (4 de diciembre del 2021) esto permite obtener datos oportunos que reflejan la situación

real del Ecuador frente a la emergencia sanitaria. Los datos extraídos permiten realizar un exhaustivo análisis para determinar la estructura de la base de datos y el posterior desarrollo de la Data Warehouse.

2.2. PLANIFICACIÓN

Esta fase define los puntos clave que se necesita para el desarrollo de la Data Warehouse y el análisis para presentar los resultados, en cada punto se encuentra el detalle de las tareas requeridas para cumplir cada etapa. En total se define cuatro puntos para lograr construir con éxito el DataWarehouse al usar la metodología Hefesto para el diseño del mismo.

2.3. FACTIBILIDAD TÉCNICA

Los requerimientos mínimos de hardware y software para la construcción de la Data Warehouse, el diseño de las interfaces amigables en Power BI y la explotación de datos, requieren de los siguientes recursos.

2.3.1. HARDWARE

| CANTIDAD | DESCRIPCIÓN |
|-----------------|-------------------------------|
| 1 | Memoria RAM 2 GB o superior |
| 1 | 12 GB de disco duro |
| 1 | Procesador de 1 GHZ de 64 bit |

2.3.2. SOFTWARE

| CANTIDAD | DESCRIPCIÓN |
|-----------------|--------------------------------------|
| 1 | Sistema operativo Windows 10 64 bits |
| 1 | Power BI desktop 64 bit |
| 1 | PostgreSQL 13 |
| 1 | Python 3 |

2.4. DISEÑO DE LA DATA WAREHOUSE

El diseño de la Data Warehouse está basado en la metodología HEFESTO que consta de cuatro etapas, para conseguir diseñar de manera correcta el Data Warehouse y así explotar los datos de manera correcta en la herramienta de Power BI.

2.4.1. ANÁLISIS DE LOS REQUERIMIENTOS

Para identificar las perspectivas y variables que apoya para obtener los indicadores de análisis de los diferentes procesos, se realizan interrogantes que respondan a cumplir con los objetivos y el alcance final planteado. Los ejes de análisis son: casos positivos, personas vacunadas y muertes ligadas al virus en el área geográfica de Ecuador, como posibles indicadores se obtienen los siguientes:

- Obtener el total, o parcial, de personas que se registraron como casos positivos dentro de un área geográfica en un período de tiempo determinado.
- Conocer la ubicación geográfica que cumple con el filtro de un rango de población registrada como casos positivos o muertes.
- El total de vacunados y tipo, efectuado por ubicación geográfica con respecto a provincia, por un período de tiempo.
- Adquirir el total de vacunados registrados por rango de edad, género, dentro de una ubicación geográfica en cuanto a provincia, por un período de tiempo.
- El total o parcial de muertes registradas por ubicación geográfica dentro de un período de tiempo.
- El total de vacunados con una primera dosis, segunda dosis.

Al finalizar las preguntas respectivas para identificar las posibles perspectivas y variables de análisis se obtiene que:

Como indicadores se necesita:

- Total, del tipo de análisis (casos positivos, muertes, vacunados).
- Total, de tipo de vacuna registrada.
- Total, de personas vacunadas con primera, segunda dosis.

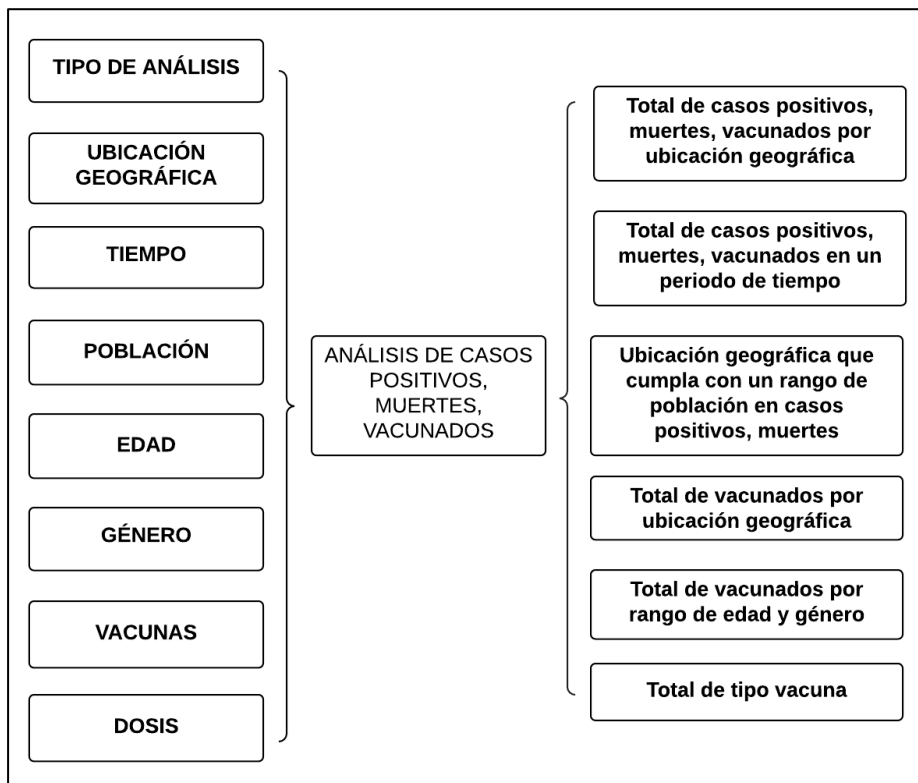
Como perspectivas de análisis se tiene:

- Tipo de análisis (casos positivos, muertes, vacunados)
- Ubicación geográfica
- Tiempo
- Población
- Edad
- Género
- Dosis

El resultado de la identificación de indicadores y perspectiva, permite implementar el siguiente modelo conceptual, que describe la relación que tienen las perspectivas y los indicadores de análisis dentro del proceso.

Figura 9

Modelo conceptual etapa 1 de la Data Warehouse.



Nota: Relación entre las perspectivas e indicadores, etapa 1. Fuente: (Mejía y Aguas, 2021).

2.4.2. ANÁLISIS DE LOS OLTP

Para completar un análisis OLTP se debe indicar la manera para realizar el cálculo de cada indicador identificado en la etapa anterior.

Para el cálculo del total de casos positivos, muertes, y vacunados.

- Hechos: “Total del tipo de análisis”
- Función de suma: SUMA

Aclaración: el indicador “Total del tipo de análisis” representa la sumatoria por separado dependiendo del eje escogido para el análisis (casos positivos, muertes, y vacunados), y de los datos que han sido cargados para filtrar la información solicitada.

Para el cálculo del total y tipo de vacuna.

- Hechos: “Total vacunados”
- Función de suma: SUMA

Aclaración: el indicador “Total vacunados” representa la sumatoria por separado dependiendo de los datos que han sido cargados para filtrar la información solicitada.

Para el cálculo del total de personas vacunadas por dosis

- Hechos: “Total vacunados por dosis”
- Función de suma: SUMA

Aclaración: el indicador “Total vacunados por dosis” representa la sumatoria de la población por separado dependiendo de los datos que han sido cargados para filtrar la información solicitada.

El siguiente procedimiento para esta etapa facilita identificar el nivel de granularidad máximo que puede alcanzar cada perspectiva.

Figura 10

Perspectiva del Tipo de análisis.



Nota: Granularidad que proporciona la perspectiva del Tipo. Fuente: (Mejía y Aguas, 2021).

Figura 11

Perspectiva del Tiempo.



Nota: Granularidad que proporciona la perspectiva del Tiempo. Fuente: (Mejía y Aguas, 2021).

Figura 12

Perspectiva de Geografía.



Nota: Granularidad que brinda la perspectiva de Geografía. Fuente: (Mejía y Aguas, 2021).

Figura 13

Perspectiva de Género.



Nota: Granularidad que tiene la perspectiva de Género. Fuente: (Mejía y Aguas, 2021).

Figura 14

Perspectiva de la Población.



Nota: Granularidad de la perspectiva de Población. Fuente: (Mejía y Aguas, 2021).

Figura 15

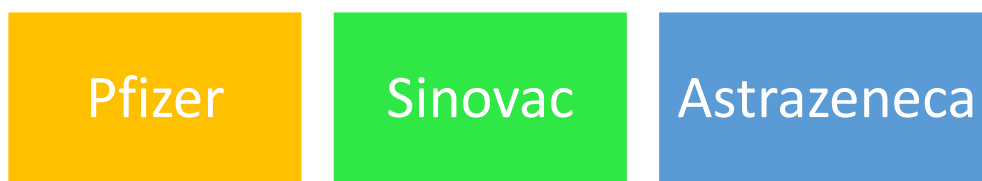
Perspectiva de Edad.



Nota: Granularidad que facilita la perspectiva de Edad. Fuente: (Mejía y Aguas, 2021).

Figura 16

Perspectiva del tipo de Vacuna.



Nota: Granularidad que muestra la perspectiva del tipo de Vacuna. Fuente: (Mejía y Aguas, 2021).

Figura 17

Perspectiva de Dosis.

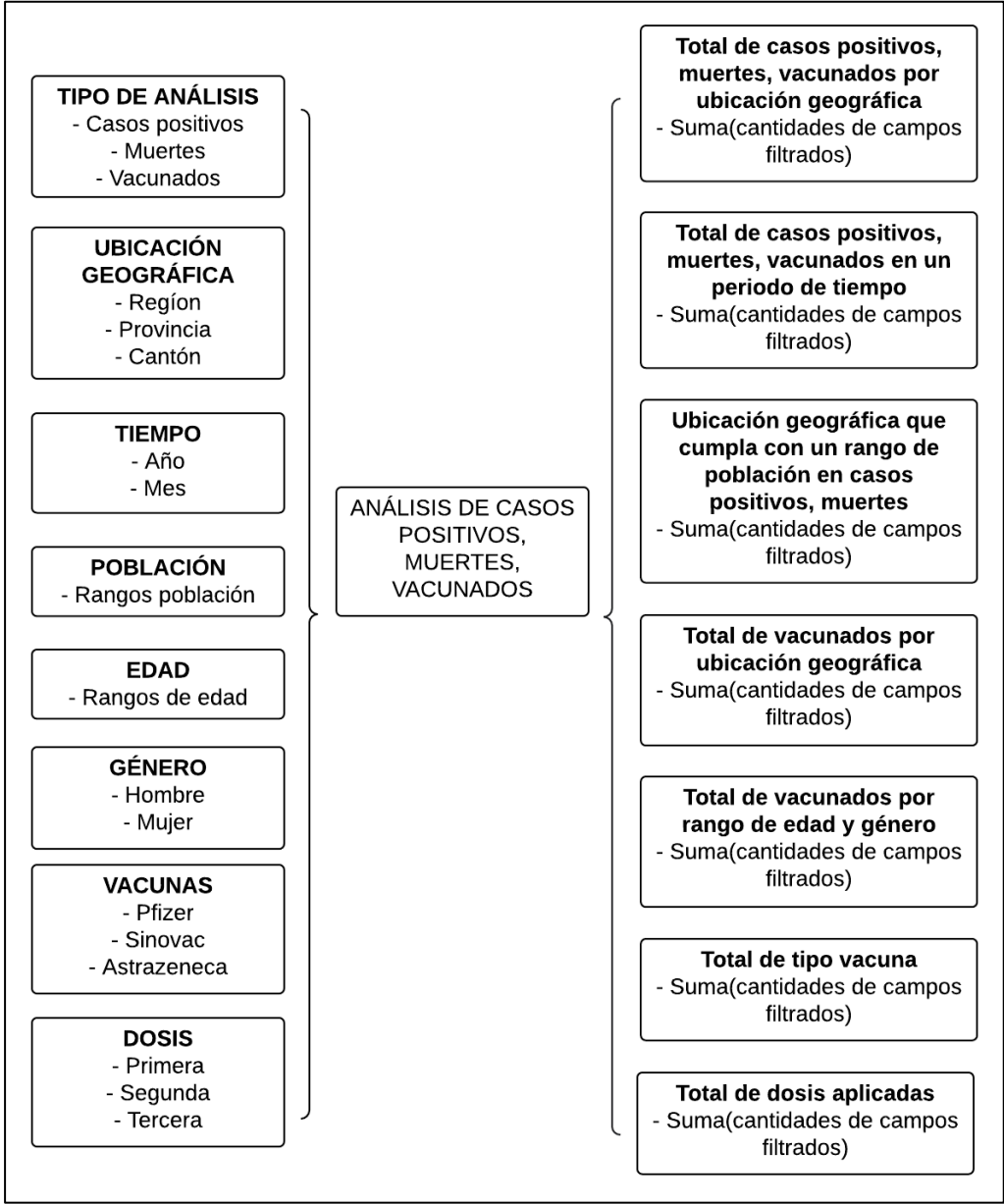


Nota: Granularidad que proporciona la perspectiva de Dosis. Fuente: (Mejía y Aguas, 2021).

El proceso de identificar los indicadores y función, además del nivel de granularidad que presente cada expectativa identificada, proporciona un mejor detalle del modelo conceptual de la Data Warehouse.

Figura 18

Modelo conceptual etapa 2 de la Data Warehouse.



Nota: Relación entre las perspectivas e indicadores, etapa 2. Fuente: (Mejía y Aguas, 2021).

2.4.3. MODELO LÓGICO DE LA DATA WAREHOUSE

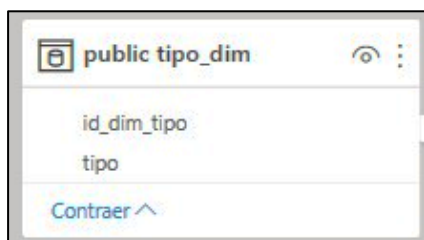
El modelo lógico se basa en el modelo conceptual generado en la etapa 2 de la metodología Hefesto Figura 10, las perspectivas generadas se deben convertir en tablas para almacenar la información y generar las tablas de las dimensiones, estos campos pueden ser modificados, eliminados o se puede aumentar nuevos campos a la perspectiva.

2.4.3.1. *Dimensión tipo de análisis.*

La dimensión del tipo de análisis considera los ejes que se analizan: para este proyecto los ejes son: casos positivos, muertes, y vacunados. El nombre de la tabla es: tipo_dim. Los atributos son: id_dim_tipo: identificador único de cada fila de datos en la tabla, tipo: eje de análisis (casos positivos, muertes, vacunados)

Figura 19

Tabla dimensión tipo de análisis.



| public tipo_dim |
|-----------------|
| id_dim_tipo |
| tipo |
| Contraer ^ |

Nota: Estructura de la dimensión tipo de análisis. Fuente: (Mejía y Aguas, 2021).

2.4.3.2. *Dimensión de ubicación geográfica.*

La dimensión de ubicación geográfica, proporciona información acerca de la región, provincia y cantón de Ecuador. El nombre de la tabla es: geografia_dim. Los atributos son: id_region: identificador único de cada fila de la tabla, inec_canton_id: identificador numérico del cantón, inec_provincia_id: identificador numérico de la provincia, lat y lng: son los valores posicionales respecto a los ejes geográficos, población: total de población en dicha ubicación geográfica,

provincia: nombre de la provincia, región: costa, sierra, amazonia, zona: zona distrital a la que pertenece la provincia.

Figura 20

Tabla dimensión de ubicación geográfica.



Nota: Estructura de la dimensión de ubicación geográfica. Fuente: (Mejía y Aguas, 2021).

2.4.3.3. Dimensión del tiempo.

La dimensión del tiempo permite filtrar por cierto rango de fecha los diferentes análisis de casos positivos, muertes, y vacunados. El nombre de la tabla es: tiempo_dim. Los atributos son: id_dim_tiempo: identificador único asignado a cada fila de la tabla, año: tiene como datos el año 2020 y 2021, mes: es un valor numérico que representa el mes del año, trimestre: es un valor numérico de los trimestres de un año.

Figura 21

Tabla dimensión del tiempo.



Nota: Estructura de la dimensión del tiempo. Fuente: (Mejía y Aguas, 2021).

2.4.3.4. *Dimensión de la población.*

La dimensión de la población brinda información directa para el análisis del total de vacunados por ciertos rangos numéricos del total poblacional. El nombre de la tabla es: muestra_dim. Los datos son: id_dim_muestra: identificador único de cada fila de información de la tabla, muestra: es un campo que almacena ciertos rangos de población total.

Figura 22

Tabla dimensión de la población.



| public muestra_dim | |
|--------------------|--|
| id_dim_muestra | |
| muestra | |

Contraer ^

Nota: Estructura de la dimensión de la población. Fuente: (Mejía y Aguas, 2021).

2.4.3.5. *Dimensión de edad.*

La dimensión de edad brinda información directa para el análisis del total de vacunados por ciertos rangos numéricos de edades. El nombre de la tabla es: edad_dim. Los datos son: id_dim_edad: identificador único de cada fila de información de la tabla, edad: es un campo que almacena ciertos rangos de edad.

Figura 23

Tabla dimensión de edad.



| public edad_dim | |
|-----------------|--|
| edad | |
| id_dim_edad | |

Contraer ^


Nota: Estructura de la dimensión de edad. Fuente: (Mejía y Aguas, 2021).

2.4.3.6. *Dimensión del género.*

La dimensión del género proporciona información concreta de los géneros de la población hombre, y mujer. El nombre de la tabla es: `genero_dim`. Los datos son: `id_dim_genero`: identificador de cada fila de datos de la tabla, `genero`: contiene el nombre del género (hombre, mujer).

Figura 24

Tabla dimensión del género.



| public genero_dim |
|-------------------|
| genero |
| id_dim_genero |
| Contraer ^ |

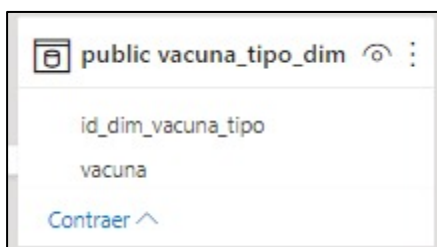
Nota: Estructura de la dimensión del género. Fuente: (Mejía y Aguas, 2021).

2.4.3.7. *Dimensión tipo de vacuna*

En la dimensión tipo de vacuna describe los tipos de vacuna aplicada a la población de Ecuador. El nombre de la tabla es: `vacuna_tipo_dim`. Los campos son: `id_dim_vacuna_tipo`: identificador único de cada fila de datos de la tabla, `vacuna`: almacena el nombre de la vacuna (Pfizer, Sinovac, AstraZeneca).

Figura 25

Tabla dimensión del tipo de vacuna.



| public vacuna_tipo_dim |
|------------------------|
| id_dim_vacuna_tipo |
| vacuna |
| Contraer ^ |

Nota: Estructura de la dimensión de tipo de vacuna. Fuente: (Mejía y Aguas, 2021).

2.4.3.8. *Dimensión dosis*

En la dimensión dosis se describe los tipos de dosis aplicadas de cada vacuna a cada persona de la población de Ecuador. El nombre de la tabla es: dosis_dim. Los campos son: id_dim_dosis: identificador único de cada fila de datos de la tabla, dosis: almacena el nombre de las dosis (primera, segunda).

Figura 26

Tabla dimensión dosis.



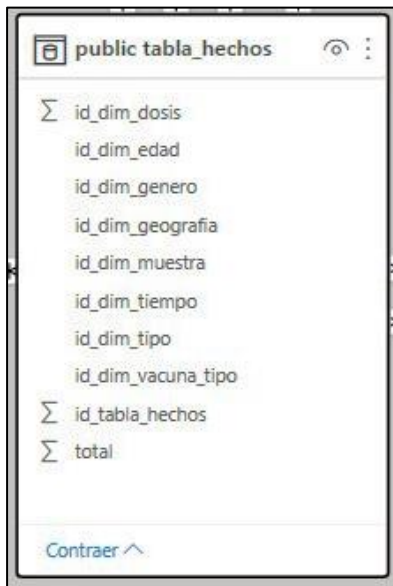
Nota: Estructura de la dimensión dosis. Fuente: (Mejía y Aguas, 2021).

2.4.3.9. *Tabla de hechos*

El nombre de la tabla es: tabla_hechos, se estructura con: id_tabla_hechos: identificador único para cada fila de datos de la tabla, con los identificadores únicos de cada tabla de dimensión, en este caso son siete dimensiones y debe registrarse siete campos para relacionar las dimensiones con la tabla de hechos, total: para obtener los resultados del cálculo de los indicadores descritos.

Figura 27

Tabla de hechos de la Data Warehouse.



Nota: Estructura de la tabla de hechos. Fuente: (Mejía y Aguas, 2021).

En la Figura 28, se presenta un extracto de la información almacenada en la tabla hechos al ejecutar el proceso de carga de datos en el punto 2.4.4.9.

Figura 28

Datos cargados a la Tabla de hechos.

| | id_tabla_hechos [PK] integer | id_dim_tiempo bigint | id_dim_edad bigint | id_dim_genero bigint | id_dim_muestra bigint | id_dim_tipo bigint | id_dim_vacuna_tipo bigint | id_dim_geografia bigint | total bigint | id_dim_dosis integer |
|----|------------------------------|----------------------|--------------------|----------------------|-----------------------|--------------------|---------------------------|-------------------------|--------------|----------------------|
| 1 | 438729 | 1908657 | 17 | 11 | 3 | 4 | 7 | 29 | 0 | 7 |
| 2 | 438730 | 1908657 | 17 | 11 | 3 | 4 | 7 | 30 | 0 | 7 |
| 3 | 438731 | 1908657 | 17 | 11 | 6 | 4 | 7 | 31 | 68 | 7 |
| 4 | 438732 | 1908657 | 17 | 11 | 1 | 4 | 7 | 32 | 0 | 7 |
| 5 | 438733 | 1908657 | 17 | 11 | 3 | 4 | 7 | 33 | 0 | 7 |
| 6 | 438734 | 1908657 | 17 | 11 | 1 | 4 | 7 | 34 | 0 | 7 |
| 7 | 438735 | 1908657 | 17 | 11 | 3 | 4 | 7 | 35 | 3 | 7 |
| 8 | 438736 | 1908657 | 17 | 11 | 3 | 4 | 7 | 36 | 0 | 7 |
| 9 | 438737 | 1908657 | 17 | 11 | 1 | 4 | 7 | 37 | 0 | 7 |
| 10 | 438738 | 1908657 | 17 | 11 | 3 | 4 | 7 | 38 | 3 | 7 |
| 11 | 438739 | 1908657 | 17 | 11 | 3 | 4 | 7 | 39 | 0 | 7 |
| 12 | 438740 | 1908657 | 17 | 11 | 1 | 4 | 7 | 40 | 0 | 7 |
| 13 | 438741 | 1908657 | 17 | 11 | 3 | 4 | 7 | 41 | 3 | 7 |
| 14 | 438742 | 1908657 | 17 | 11 | 2 | 4 | 7 | 42 | 0 | 7 |
| 15 | 438743 | 1908657 | 17 | 11 | 3 | 4 | 7 | 43 | 1 | 7 |
| 16 | 438744 | 1908657 | 17 | 11 | 3 | 4 | 7 | 44 | 0 | 7 |
| 17 | 438745 | 1908657 | 17 | 11 | 3 | 4 | 7 | 45 | 2 | 7 |
| 18 | 438746 | 1908657 | 17 | 11 | 3 | 4 | 7 | 46 | 1 | 7 |

Nota: Datos almacenados en la Data Warehouse al realizar la carga automática en la tabla de hechos. Fuente: (Mejía y Aguas, 2021).

2.4.3.9. *Modelo lógico relacional dimensiones y hechos*

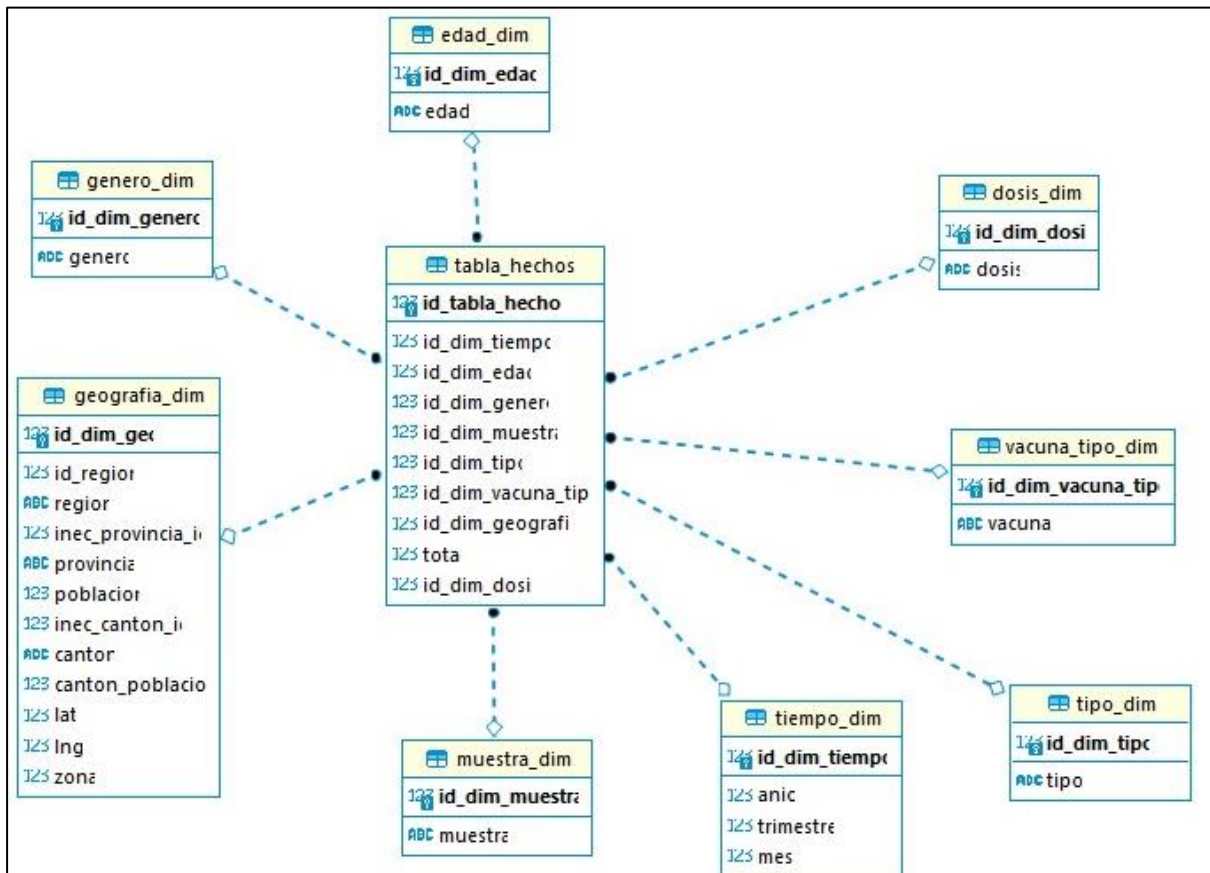
El modelo lógico relacional es el resultado de la relación entre las tablas diseñadas de cada dimensión y la tabla de hechos que debe contener cada clave de las dimensiones. La relación que presentan es de uno a muchos, partiendo de las tablas dimensiones como uno y la tabla de hecho como muchos por las repeticiones del dato que se generen.

Los datos de la tabla hechos presentan la siguiente estructura relacional:

- El campo *id_dim_tiempo* está relacionado con la clave principal *id_dim_tiempo* de la tabla *tiempo_dim*.
- El campo *id_dim_edad* está relacionado con la clave principal *id_dim_edad* de la tabla *edad_dim*.
- El campo *id_dim_genero* está relacionado con la clave principal *id_dim_genero* de la tabla *genero_dim*.
- El campo *id_dim_muestra* está relacionado con la clave principal *id_dim_muestra* de la tabla *muestra_dim*.
- El campo *id_dim_tipo* está relacionado con la clave principal *id_dim_tipo* de la tabla *tipo_dim*.
- El campo *id_dim_vacuna_tipo* está relacionado con la clave principal *id_dim_vacuna_tipo* de la tabla *vacuna_tipo_dim*.
- El campo *id_dim_geografia* está relacionado con la clave principal *id_dim_geo* de la tabla *geografia_dim*.
- El campo *id_dim_dosis* está relacionado con la clave principal *id_dim_dosis* de la tabla *dosis_dim*.

Figura 29

Modelo lógico de la Data Warehouse.



Nota: Estructura tipo estrella del modelo lógico relacional. Fuente: (Mejía y Aguas, 2021).

2.4.4. INTEGRACIÓN DE DATOS

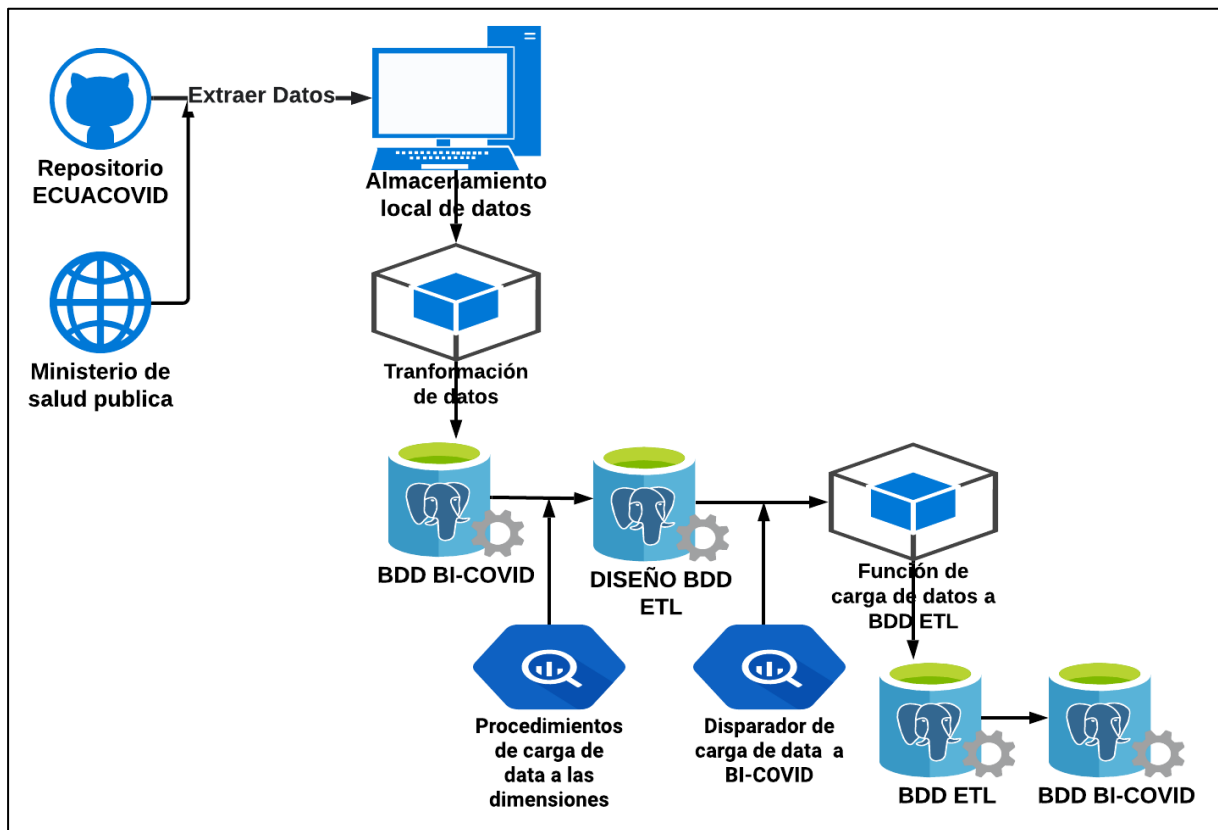
Para la etapa de integración de datos se debe hacer uso de procesos ETL, y del apoyo de las bases, *ETL* y *BI-COVID*. La base de datos *BI-COVID* contiene el modelo lógico de la Data Warehouse diseñado en la etapa tres de la metodología Hefesto, la base de datos *ETL* facilita el procesamiento de datos para la carga de la base *BI-COVID*.

El flujo del proceso ETL parte de la extracción de datos desde el repositorio ECUACOVID, y del MSP, se debe proporcionar un almacenamiento interno para los datos descargados, el uso de funciones Python facilita la carga de datos a la base *ETL*, esta base presenta las tablas de casos positivos (2020, 2021), muertes (2020, 2021), y vacunados, además de que cada tabla

presenta disparadores de carga automáticamente de datos procesados a la tabla de hechos de la base *BI_COVID*, por medio de procedimientos generados en la base *BI_COVID* se realiza la carga de datos a las dimensiones.

Figura 30

Flujo del proceso ETL.



Nota: Estructura del flujo del proceso ETL en el modelado lógico de la etapa tres. Fuente: (Mejía y Aguas, 2021).

2.4.4.1. Repositorio Ecuacovid

La fuente de información Ecuacovid se encuentra alojada en un repositorio GitHub, obtenido de <https://github.com/andrab/ecuacovid.git>, el propietario del repositorio indica que son datos sin procesar extraídos de los diferentes informes generados por el SNGRE, MSP y Registro Civil. Los datos extraídos para ser explotados se encuentran almacenados dentro de la carpeta

datos crudos, esta información se encuentra en archivos de texto plano en formato CSV. Los archivos que van a ser utilizados para los ejes de análisis de casos positivos y muertes son:

Figura 31

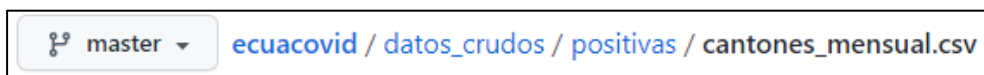
Archivo para casos positivos 2020.



Nota: Ubicación interna dentro del repositorio Ecuacovid del archivo que brinda información para el eje de análisis de casos positivos. Fuente: (Mejía y Aguas, 2021).

Figura 32

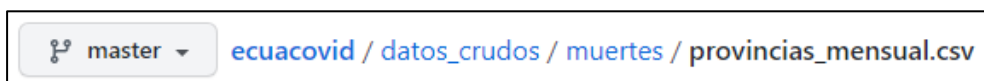
Archivo para casos positivos 2021.



Nota: Ubicación interna dentro del repositorio Ecuacovid del archivo que brinda información para el eje de análisis de casos positivos. Fuente: (Mejía y Aguas, 2021).

Figura 33

Archivo para muertes 2020.



Nota: Ubicación interna dentro del repositorio Ecuacovid del archivo que brinda información para el eje de análisis de muertes. Fuente: (Mejía y Aguas, 2021).

Figura 34

Archivo para muertes 2021.



Nota: Ubicación interna dentro del repositorio Ecuacovid del archivo que brinda información para el eje de análisis de muertes. Fuente: (Mejía y Aguas, 2021).

2.4.4.2. *Ministerio de Salud Pública*

La página oficial del Ministerio de Salud Pública (MSP), brinda entre sus servicios el proceso de la *Salud en Cifras*, este servicio proporciona el eje de información para vacunas es *Datos abiertos* (<https://www.salud.gob.ec/datos-abiertos/>), esta página despliega entre sus archivos de información el de *VACUNAS COVID*, se ingresa al enlace que proporciona: (<https://almacenamiento.msp.gob.ec/index.php/s/LHzKA6BvyNBPWNR>), esta página presenta dos archivos en formato CSV y XLS respectivamente. El archivo que brinda información para el eje de análisis vacunas es *MSP_vacuna_COVID_2021.csv*.

La información que contiene el archivo *MSP_vacuna_COVID_2021.csv* para realizar el eje de análisis de vacunas, proporciona datos limitados hasta la fecha del 30 de junio del 2021.

Figura 35

Archivos del MSP.



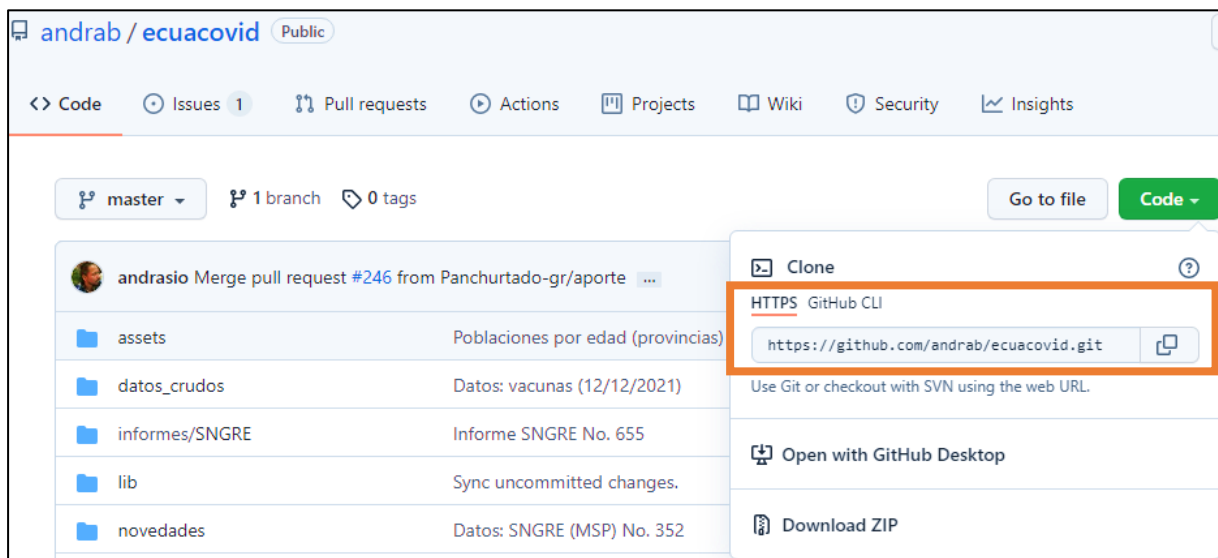
Nota: Archivos que proporciona el MSP con relación a la actividad de vacunas en Ecuador estos archivos brindan información para el eje de análisis de vacunas. Fuente: (Mejía y Aguas, 2021).

2.4.4.3. *Extraer datos*

Para el proceso de extracción de datos desde el repositorio Ecuacovid, se procede a copiar el enlace que brinda el repositorio con la finalidad de clonar el repositorio en la ruta local especificada por el usuario, la ruta de almacenamiento se especifica en el punto 2.4.4.4.

Figura 36

Enlace de clonación del repositorio Ecuacovid.



Nota: Enlace de GitHub que proporciona el repositorio para ser clonado en el almacenamiento interno y así ser utilizada la información del repositorio. Fuente: (Mejía y Aguas, 2021).

Para la extracción de los datos proporcionados por el MSP, debe ingresar a <https://almacenamiento.msp.gob.ec/index.php/s/LHzKA6BvyNBPWNR>, elegir el archivo a descargar (Figura 35), dar clic sobre el archivo elegido y la descarga comienza automáticamente, al archivo *MSP_VACUNA_COVID_2021.csv* se almacena en la carpeta de descargas que el navegador tenga por defecto.

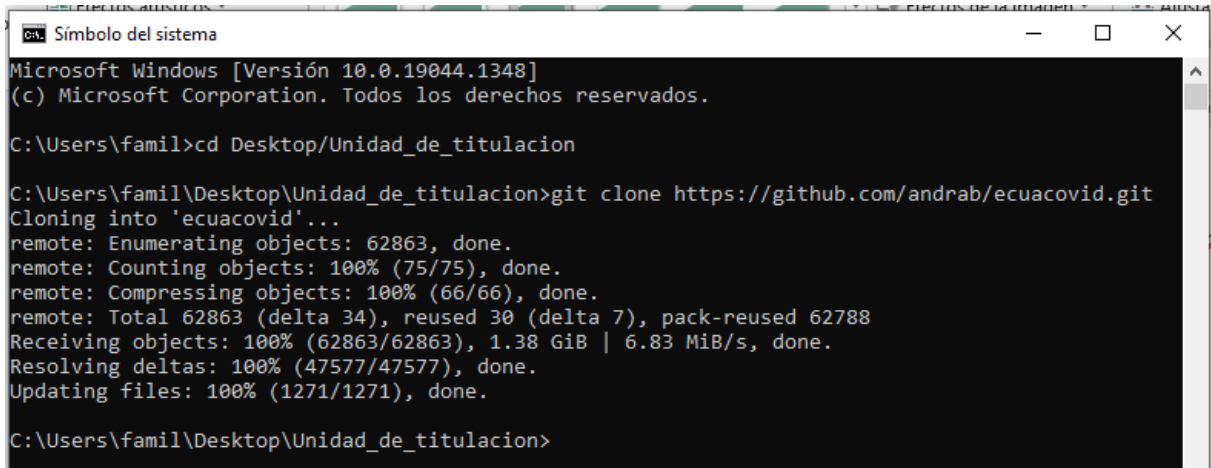
2.4.4.4. *Almacenamiento local de datos*

Los datos descargados para ser explotados en el proceso ETL, se almacenan en la carpeta *Unidad_de_titulacion* alojada en el escritorio del servidor local.

Por medio de la consola de símbolo de sistemas se accede a la carpeta *Unidad_de_titulacion* para almacenar el repositorio clonado Ecuacovid.

Figura 37

Clonar repositorio Ecuacovid.



```
Microsoft Windows [Versión 10.0.19044.1348]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\famil>cd Desktop/Unidad_de_titulacion

C:\Users\famil\Desktop\Unidad_de_titulacion>git clone https://github.com/andrab/ecuacovid.git
Cloning into 'ecuacovid'...
remote: Enumerating objects: 62863, done.
remote: Counting objects: 100% (75/75), done.
remote: Compressing objects: 100% (66/66), done.
remote: Total 62863 (delta 34), reused 30 (delta 7), pack-reused 62788
Receiving objects: 100% (62863/62863), 1.38 GiB | 6.83 MiB/s, done.
Resolving deltas: 100% (47577/47577), done.
Updating files: 100% (1271/1271), done.

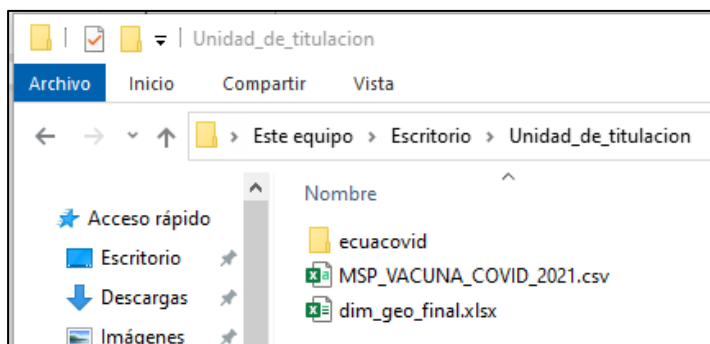
C:\Users\famil\Desktop\Unidad_de_titulacion>
```

Nota: Código git para clonar el repositorio Ecuacovid. Fuente: (Mejía y Aguas, 2021).

El archivo descargado desde los servicios proporcionados por el MSP, se descarga automáticamente en la carpeta *Descargas*, ya que es la carpeta configurada por defecto en el navegador, se debe copiar el archivo y pegar en la carpeta *Unidad_de_titulacion*.

Figura 38

Almacenamiento de datos a explotar.



Nota: Datos almacenados en el servidor local. Fuente: (Mejía y Aguas, 2021).

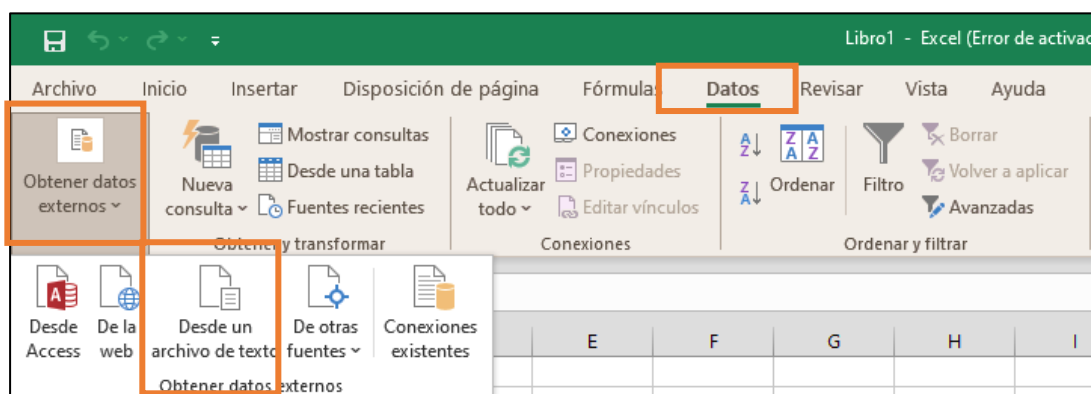
2.4.4.5. *Transformación de datos*

Los archivos que han sido extraídos en el punto 2.4.4.3. del repositorio Ecuacovid, se encuentran en formato CSV, estos contienen caracteres especiales debido al formato con el que se descarga. Al cargar estos archivos en las tablas de la base ETL sin realizar la transformación de datos, las tablas almacenan datos con caracteres especiales que contiene el archivo.

Para la transformación de datos, se debe realizar el cambio de formato del archivo de texto plano, para que pueda reconocer los caracteres especiales (tilde y letra ñ). Para este procedimiento se debe crear un nuevo archivo Excel en blanco, en la vista *Datos*, seleccionar *Obtener datos externos*, y escoger *desde un archivo de texto*.

Figura 39

Buscar archivos de texto desde Excel.

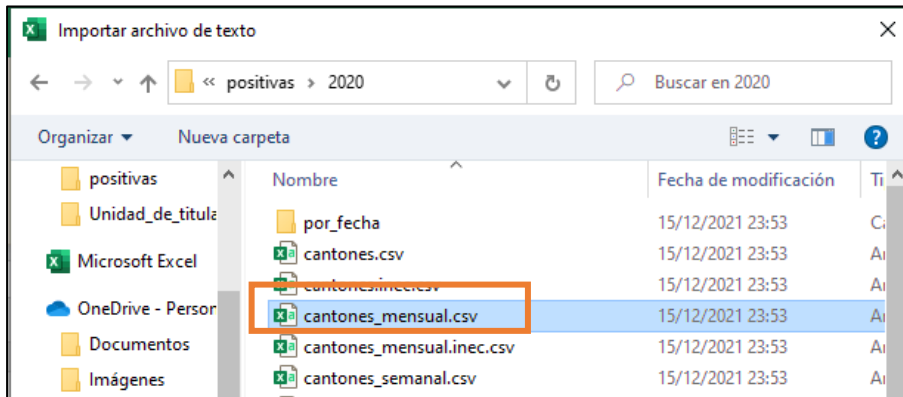


Nota: Procedimiento de carga de un archivo de texto plano. Fuente: (Mejía y Aguas, 2021).

El explorador de archivos, facilita la búsqueda y selección del archivo a usar para la carga de datos e importar.

Figura 40

Importar archivo Excel seleccionado.

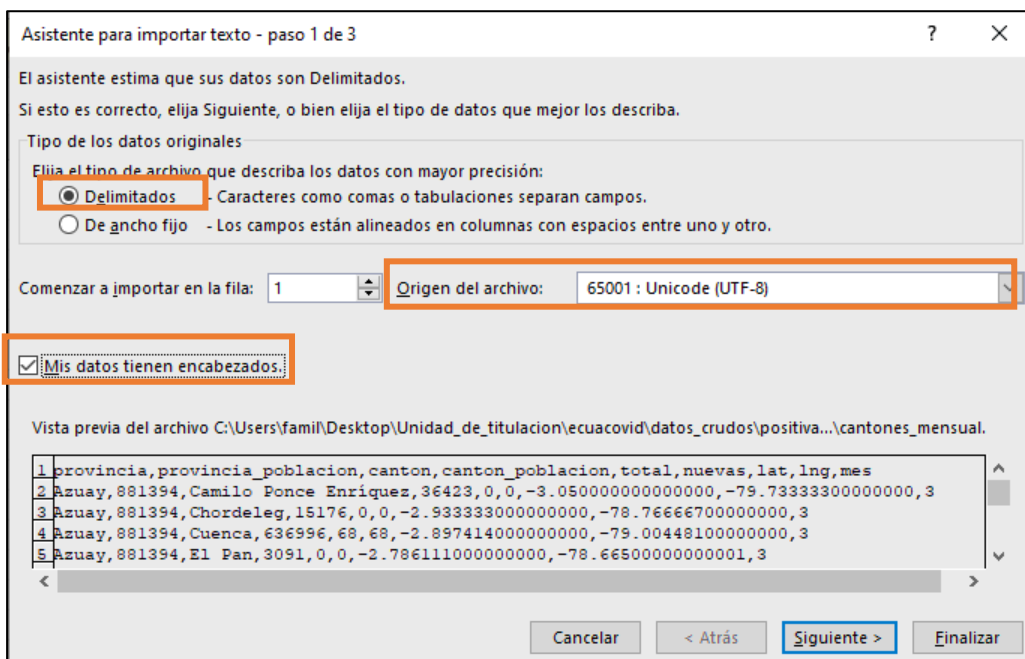


Nota: Seleccionar e importar el archivo CSV a Excel. Fuente: (Mejía y Aguas, 2021).

El asistente para importar texto, proporciona la configuración del tipo de archivo *Delimitado*, el origen del archivo se debe configurar como *65001: Unicode (UTF-8)*, se indica que los datos *tienen encabezados*, y seleccionar finalizar.

Figura 41

Configuración del asistente de importar texto.

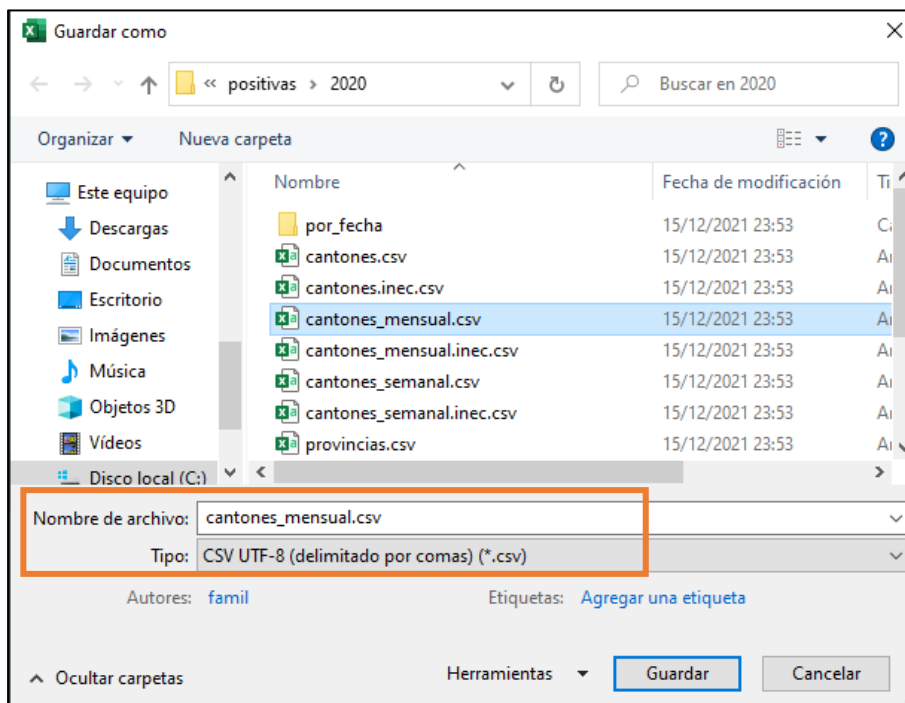


Nota: Configuración del asistente de importar texto a Excel. Fuente: (Mejía y Aguas, 2021).

Aceptar la carga por defecto a la hoja indicada por Excel, guardar el archivo creado en la misma ruta y con el mismo nombre del archivo importado, en tipo seleccionar *CSV UTF-8 (delimitado por comas) (*.csv)*, reemplazar el archivo y seleccionar si al mantener el formato.

Figura 42

Configuración del formato del archivo generado.



Nota: Reemplazar el archivo generado por el existente. Fuente: (Mejía y Aguas, 2021).

Para los otros archivos descargados que se vayan a procesar se debe repetir el proceso 2.4.4.5., de esta manera se garantiza que el formato para la carga de datos reconozca tildes y la letra ñ.

Otro método de transformación de datos se configura en el punto 2.4.4.9., para la carga de datos, la línea de comando `“import sqlalchemyimport Float”`, permite importar el tipo de dato flotante, este tipo importado facilita configurar entre los parámetros de la sentencia `to_sql`, que los datos de latitud `lat` y longitud `lng` son de tipo flotante y que no cambien por datos numéricos enteros al cargar los datos en las tablas.

Para el archivo *MSP_VACUNA_COVID_2021.csv*, en el punto 2.4.4.9.6., para realizar la lectura del archivo mediante el método *read_csv*, se debe especificar mediante el parámetro *encoding* que el tipo de codificación para el archivo es *ISO-8859-1*, esta codificación indica que el texto se encuentra en el alfabeto latino.

2.4.4.6. Procedimientos de carga de data a las dimensiones

Para la carga de data a las diferentes dimensiones se hace uso de procedimientos generados en la base de datos BI_COVID, la tabla de hechos de la Data Warehouse no permite agregar campos no definidos, por este motivo para la carga de datos no relacionables se agrega un campo *no definido* a cada tabla de dimensiones, a continuación, se describen los procesos por dimensiones:

2.4.4.5.1. Carga de datos dimensión Tipo

Para la carga de datos de la dimensión tipo se utiliza el procedimiento *carga_dim_tipo* (), se declara la variable *v_tipo* que almacena el array de los ejes de análisis: *positivos*, *muertes*, y *vacunados*, por medio de la función *foreach* se recorre la variable *v_tipo* y se utiliza la consulta *query insert into* a la tabla *tipo_dim* en el campo *tipo* de la tabla con cada ítem de la variable *v_tipo*.

Figura 43

Procedimiento para la dimensión Tipo.

```
DECLARE
  v_carga int;
  v_tipo text[] := array['positivos', 'muertes', 'vacunados'];
  item text;
BEGIN
  foreach item in array v_tipo loop
    insert into tipo_dim (tipo)
      values (item);
  end loop;
v_carga := 1;
return v_carga;
END
```

Nota: Procedimiento de inserción a la tabla *tipo_dim*. Fuente: (Mejía y Aguas, 2021).

Los datos cargados a la tabla *tipo_dim* al ejecutar el procedimiento *carga_dim_tipo()* son los siguientes.

Figura 44

Datos cargados en la dimensión Tipo.

| id_dim_tipo [PK] bigint | tipo text |
|----------------------------|--------------|
| 1 | positivos |
| 2 | muertes |
| 3 | vacunados |

Nota: Datos almacenados en la tabla de la dimensión Tipo. Fuente: (Mejía y Aguas, 2021).

2.4.4.5.2. Carga de datos dimensión Tiempo

En la dimensión tiempo se hace uso del procedimiento *carga_dim_tiempo()*, en el que se declara la variable *v_anio* que almacena el parámetro que se ingresa al ejecutar el procedimiento sea *2020* o *2021*, la variable *v_trim* contiene el valor del trimestre, mediante la función *while* se recorre la variable *i* hasta que cumpla la condición que sea menor a 12, por medio de la consulta *query insert into* a la dimensión *tiempo_dim* en el campo *anio, mes* se registran los valores de *v_anio, i*.

Figura 45

Procedimientos para la dimensión Tiempo.

```
DECLARE
v_genera      INT;
v_existe      INT;
v_mensaje     character varying(100) := 'Ya se genero el periodo';
i             INT;
j             INT;
v_trim        INT;
v_anio        INT := anio;

BEGIN
IF anio = 2020
  THEN
    i := 3;
ELSIF anio = 2021
  THEN
    i := 1;
END IF;
v_trim = 1;
WHILE i <= 12
  LOOP
    INSERT INTO "Tiempo_dim" (anio, trimestre , mes)
    VALUES (v_anio,v_trim, i);
    i := i+1;
    IF i <= 3
      THEN
        v_trim := 1;
      ELSIF i > 3 AND i<=6
        THEN
          v_trim := 2;
        ELSIF i > 6 AND i<=9
          THEN
            v_trim := 3;
          ELSIF i > 9 AND i<=12
            THEN
              v_trim := 4;
            END IF;
          END IF;
        END LOOP;
v_genera := 1 ;
return v_genera;
END
```

Nota: Código de procedimiento de inserción a la tabla tiempo_dim. Fuente: (Mejía y Aguas, 2021).

Los datos cargados a la tabla *tiempo_dim* al ejecutar el procedimiento *carga_dim_tiempo()* son los siguientes.

Figura 46

Datos cargados en la dimensión tiempo.

| id_dim_tiempo [PK] integer | año smallint | trimestre smallint | mes smallint |
|--------------------------------------|------------------------|------------------------------|------------------------|
| 1908657 | 2020 | 1 | 3 |
| 1908658 | 2020 | 2 | 4 |
| 1908659 | 2020 | 2 | 5 |
| 1908660 | 2020 | 2 | 6 |
| 1908661 | 2020 | 3 | 7 |
| 1908662 | 2020 | 3 | 8 |
| 1908663 | 2020 | 3 | 9 |
| 1908664 | 2020 | 4 | 10 |
| 1908665 | 2020 | 4 | 11 |
| 1908666 | 2020 | 4 | 12 |
| 1908667 | 2021 | 1 | 1 |
| 1908668 | 2021 | 1 | 2 |
| 1908669 | 2021 | 1 | 3 |
| 1908670 | 2021 | 2 | 4 |
| 1908671 | 2021 | 2 | 5 |
| 1908672 | 2021 | 2 | 6 |

Nota: Datos almacenados en la tabla de la dimensión Tiempo. Fuente: (Mejía y Aguas, 2021).

2.4.4.5.3. Carga de datos dimensión Muestra

El procedimiento *carga_dim_muestra()* generado en la dimensión muestra presenta la siguiente estructura: variable *v_muestra* que almacena el array de los parámetros con respecto al total de población: *0 – 5 mil, 5 – 10 mil, 10 – 50 mil, 50 – 100 mil, 100 – 500 mil, 500 mill – 1M y +1M*, por medio de la función *foreach* se recorre la variable *v_muestra*, se ejecuta la consulta query de tipo *insert into* a la dimensión *muestra_dim* en el campo *muestra* con la información de cada ítem de la variable *v_muestra*.

Figura 47

Procedimiento para la dimensión Población.

```
DECLARE
  v_carga int;
  v_muestra text[] := array['0 - 5 mil', '5 mil - 10 mil',
                           '10 mil - 50 mil', '50 mil - 100 mil',
                           '100 mil - 500 mil', '500 mill - 1 M',
                           '+1M', 'No definido'];

  item text;

BEGIN
  foreach item in array v_muestra loop
    insert into muestra_dim (muestra)
      values(item);
  end loop;
v_carga := 1;
return v_carga;
END
```

Nota: Código de procedimiento de inserción a la tabla muestra_dim. Fuente: (Mejía y Aguas, 2021).

Los datos cargados en la tabla muestra_dim al ejecutar el procedimiento carga_dim_muestra() son los siguientes.

Figura 48

Datos cargados en la dimensión Población.

| | id_dim_muestra [PK] bigint | muestra text |
|---|-------------------------------|------------------|
| 1 | 1 | 0 - 5 mil |
| 2 | 2 | 5 mil - 10 mil |
| 3 | 3 | 10 mil - 50 mil |
| 4 | 4 | 50 mil - 100 ... |
| 5 | 5 | 100 mil - 50... |
| 6 | 6 | 500 mill - 1 M |
| 7 | 7 | +1M |
| 8 | 8 | No definido |

Nota: Datos almacenados en la tabla de la dimensión Muestra. Fuente: (Mejía y Aguas, 2021).

2.4.4.5.4. Carga de datos dimensión Edad

Dentro de la tabla de la dimensión edad, se genera el procedimiento *carga_dim_edad* (), la variable *v_edad* almacena el array de los parámetros de edades: 0 – 20 años, 20 – 40 años, 40 – 60 años, 60 – 80 años, por medio de la función *foreach* se recorre la variable *v_edad* y se ejecuta una consulta query de tipo *insert into* a la dimensión *edad_dim* en el campo *edad* con la información de cada ítem de *v_edad*.

Figura 49

Procedimiento para la dimensión Edad.

```
DECLARE
  v_carga int;
  v_edad text[] := array['De 12 a 17 años', 'De 18 a 24 años', 'De 25 a 49 años',
                        'De 50 a 59 años', 'De 60 a 69 años', 'De 70 a 79 años',
                        'De 80 años y más', 'No definido'];
  item text;
BEGIN
  foreach item in array v_edad loop
    insert into edad_dim (edad)
    values (item);
  end loop;
v_carga := 1;
return v_carga;
END
```

Nota: Procedimiento de inserción a la tabla *edad_dim*. Fuente: (Mejía y Aguas, 2021).

Los datos al ejecutar el procedimiento *carga_dim_edad* () son los siguientes. Los valores del campo de clave primaria inician con el valor 6 debido a que para conseguir los valores finales se realizaron cargas de prueba y error, por este motivo los id no inician con el valor 1.

Figura 50

Datos cargados en la dimensión edad.

| | id_dim_edad [PK] bigint | edad text |
|---|----------------------------|------------------|
| 1 | 18 | De 12 a 17 años |
| 2 | 19 | De 18 a 24 años |
| 3 | 20 | De 25 a 49 años |
| 4 | 21 | De 50 a 59 años |
| 5 | 22 | De 60 a 69 años |
| 6 | 23 | De 70 a 79 años |
| 7 | 24 | De 80 años y más |
| 8 | 25 | No definido |

Nota: Datos almacenados en la tabla de la dimensión Edad. Fuente: (Mejía y Aguas, 2021).

2.4.4.5.5. Carga de datos dimensión Género

Para la carga de datos de género, se genera el procedimiento *carga_dim_genero ()*, que declara la variable *v_generos*, esta almacena el array del análisis por género: *hombre, mujer*, la función *foreach* recorre la variable *v_generos* para utilizar la consulta query de tipo *insert into* a la dimensión *genero_dim* en el campo *genero* con la información de cada ítem de *v_generos*.

Figura 51

Procedimiento para la dimensión Género.

```
DECLARE
    v_carga      int;
    v_generos    text[] := array['hombre', 'mujer', 'no definido'];
    item        text;
BEGIN
    foreach item in array v_generos loop
        insert into genero_dim (genero)
            values (item);
    end loop;
v_carga := 1;
return v_carga;
END
```

Nota: procedimiento de inserción a la tabla *genero_dim*. Fuente: (Mejía y Aguas, 2021).

Los datos cargados en la tabla `genero_dim` al ejecutar el procedimiento `carga_dim_genero()` son los siguientes.

Figura 52

Datos cargados en la dimensión género.

| | genero text | id_dim_genero [PK] bigint |
|---|----------------|------------------------------|
| 1 | hombre | 9 |
| 2 | mujer | 10 |
| 3 | no definido | 11 |

Nota: Datos almacenados en la tabla de la dimensión Género. Fuente: (Mejía y Aguas, 2021).

2.4.4.5.6. Carga de datos dimensión tipo de Vacuna

El procedimiento `carga_dim_vacuna_tipo ()`, en el que se declara la variable `v_vacuna`, que almacena el array del análisis por tipo de vacuna: *Pfizer*, *Sinovac*, *AstraZeneca*, la función `foreach` recorre la variable `v_vacuna` que por medio de la consulta query de tipo `insert into` a la tabla `vacuna_tipo_dim`, se registra los datos en el campo `vacuna` de la tabla, con la información de cada ítem de `v_vacuna`.

Figura 53

Procedimiento para la dimensión tipo de vacuna.

```
DECLARE
  v_carga int;
  v_vacuna text[] := array['pfizer', 'sinovac', 'astrazeneca', 'no definido'];
  item text;
BEGIN
  foreach item in array v_vacuna loop
    insert into vacuna_tipo_dim(vacuna)
      values(item);
  end loop;
v_carga := 1;
return v_carga;
END
```

Nota: Procedimiento de inserción a la tabla `vacuna_tipo_dim`. Fuente: (Mejía y Aguas, 2021).

Los datos cargados en la tabla `vacuna_dim` al ejecutar el procedimiento `carga_dim_vacuna_tipo()` son los siguientes.

Figura 54

Datos cargados en la dimensión vacuna.

| | id_dim_vacuna_tipo [PK] bigint | | vacuna text |
|---|-----------------------------------|---|----------------|
| 1 | | 4 | pfizer |
| 2 | | 5 | sinovac |
| 3 | | 6 | astrazeneca |
| 4 | | 7 | no definido |

Nota: Datos almacenados en la tabla de la dimensión Vacuna. Fuente: (Mejía y Aguas, 2021).

2.4.4.5.7. Carga de datos dimensión Dosis

Para la carga de datos en la dimensión de dosis, implementar el procedimiento `carga_dim_dosis()`, este procedimiento declara la variable `v_dosis`, que almacena el array del análisis por tipo de dosis: *primera*, *segunda* la función `foreach` recorre la variable `v_dosis`, mediante la consulta query de tipo `insert into` a la tabla `dosis_dim`, se ingresa los datos en el campo `dosis` de la tabla con la información de cada ítem de `v_dosis`.

Figura 55

Procedimiento para la dimensión dosis.

```
DECLARE
    v_carga int;
    v_dosis text[] := array['primera', 'segunda', 'no definido'];
    item text;
BEGIN
    foreach item in array v_dosis loop
        insert into dosis_dim (dosis)
            values (item);
    end loop;
v_carga := 1;
return v_carga;
END
```

Nota: Código del procedimiento de inserción a la tabla `dosis_dim`. Fuente: (Mejía y Aguas, 2021).

Los datos cargados en la tabla dosis_dim al ejecutar el procedimiento carga_dim_vacuna_tipo() son los siguientes.

Figura 56

Datos cargados en la dimensión dosis.

| | id_dim_dosis [PK] integer | dosis text |
|---|------------------------------|---------------|
| 1 | 4 | primera |
| 2 | 5 | segunda |
| 3 | 7 | no definido |

Nota: Datos almacenados en la tabla de la dimensión Dosis: (Mejía y Aguas, 2021).

2.4.4.5.8. Carga de datos dimensión Geografía

En dimensión de geografía, para la carga de datos se hace uso de código escrito en lenguaje Python, que permite realizar la carga de los datos del archivo *dim_geo_final.xlsx*. La estructura que presenta el archivo es la siguiente:

- id_region: valor numérico del uno al cuatro, que representa cada región del país.
- region: nombre de cada región del país.
- inec_provincia_id: valor numérico del uno al veinticuatro, que representa el código INEC de cada provincia.
- provincia: nombre de cada provincia.
- poblacion: cantidad total de habitantes por provincia.
- inec_canton_id: código generado por el INEC para representar cada cantón de cada provincia.
- canton: nombre del cantón.
- canton_poblacion: total de habitantes por cada cantón.
- lat: valor flotante que representa la latitud geográfica de la provincia o cantón.

- lng: valor flotante que representa la longitud geográfica de la provincia o cantón.

La librería sqlalchemy permite generar la conexión entre Python y PostgreSQL mediante el método `create_engine`, el parámetro que necesita este método es la cadena de conexión a la base externa.

Figura 57

Conexión entre Python y la base de datos BI_COVID.

```
from sqlalchemy import create_engine
import pandas as pd
engine = create_engine('postgresql+psycopg2://postgres:admin@localhost:5432/BI_COVID')
```

Nota: Conexión desde Python a la base de datos BI_COVID. Fuente: (Mejía y Aguas, 2021).

La librería Pandas facilita la lectura de archivos, la variable `df` almacena los datos leídos desde el archivo `dim_geo_final.xlsx`, la sentencia `to_sql` permite la inserción de datos nuevos a la tabla que se indica como parámetro.

Figura 58

Carga de datos del archivo a la dimensión de geografía.

```
#Carga dimension geografía
df = pd.read_excel('/Users/carlo/Desktop/dim_geo/dim_geo_final.xlsx')
from sqlalchemy.types import Float
df.to_sql('geografía_dim', engine, index=True, dtype={"lat": Float(), "lng": Float()}, if_exists="append");
```

Nota: Registro de nuevas filas de datos a la tabla de la dimensión geografía. Fuente: (Mejía y Aguas, 2021).

La data cargada en la tabla `geografía_dim` al ejecutar el código Python desde Jupyter Notebook.

Figura 59

Datos cargados en la dimensión geografía.

| id_dim_geo [FK] bigint | id_region double precision | region text | inec_provincia_id double precision | provincia text | poblacion double precision | inec_canton_id double precision | canton text | canton_poblacion double precision | lat double precision | lng double precision | zona integer |
|---------------------------|-------------------------------|----------------|---------------------------------------|-------------------|-------------------------------|------------------------------------|----------------|--------------------------------------|-------------------------|-------------------------|-----------------|
| 30 | 2 | Sierra | 1 | Azuay | 881394 | 111 | Chordeleg | 15176 | -2.933333 | -78.766667 | 6 |
| 31 | 2 | Sierra | 1 | Azuay | 881394 | 101 | Cuenca | 636996 | -2.897414 | -79.004481 | 6 |
| 32 | 2 | Sierra | 1 | Azuay | 881394 | 112 | El Pan | 3091 | -2.786111 | -78.665 | 6 |
| 33 | 2 | Sierra | 1 | Azuay | 881394 | 102 | Girón | 13037 | -3.159028 | -79.147568 | 6 |
| 34 | 2 | Sierra | 1 | Azuay | 881394 | 114 | Guachapala | 3859 | -2.769 | -78.709 | 6 |
| 35 | 2 | Sierra | 1 | Azuay | 881394 | 103 | Gualaceo | 49104 | -2.9 | -78.783333 | 6 |
| 36 | 2 | Sierra | 1 | Azuay | 881394 | 104 | Nabón | 17292 | -3.33606 | -79.06368 | 6 |
| 37 | 2 | Sierra | 1 | Azuay | 881394 | 110 | Oña | 4128 | -3.46964 | -79.15431 | 6 |
| 38 | 2 | Sierra | 1 | Azuay | 881394 | 105 | Paute | 29214 | -2.7818 | -78.76 | 6 |
| 39 | 2 | Sierra | 1 | Azuay | 881394 | 106 | Pucará | 10584 | -3.23378 | -79.4612 | 6 |
| 40 | 2 | Sierra | 1 | Azuay | 881394 | 107 | San Fernan... | 4156 | -3.1445 | -79.252 | 6 |
| 41 | 2 | Sierra | 1 | Azuay | 881394 | 108 | Santa Isabel | 20935 | -3.27417 | -79.3156 | 6 |
| 42 | 2 | Sierra | 1 | Azuay | 881394 | 113 | Sevilla de ... | 6890 | -2.799833 | -78.655694 | 6 |
| 43 | 2 | Sierra | 1 | Azuay | 881394 | 109 | Sigsig | 30509 | -3.0897 | -78.8032 | 6 |
| 44 | 2 | Sierra | 2 | Bolívar | 209933 | 206 | Caluma | 16429 | -1.633333 | -79.25 | 5 |
| 45 | 2 | Sierra | 2 | Bolívar | 209933 | 202 | Chillanes | 16650 | -1.933333 | -79.066667 | 5 |
| 46 | 2 | Sierra | 2 | Bolívar | 209933 | 203 | Chimbo | 17378 | -1.7 | -79.016667 | 5 |
| 47 | 2 | Sierra | 2 | Bolívar | 209933 | 204 | Echeandía | 14081 | -1.433333 | -79.26667 | 5 |
| 48 | 2 | Sierra | 2 | Bolívar | 209933 | 201 | Guaranda | 108763 | -1.605556 | -79.003056 | 5 |
| 49 | 2 | Sierra | 2 | Bolívar | 209933 | 207 | Las Naves | 7428 | -1.283333 | -79.3 | 5 |
| 50 | 2 | Sierra | 2 | Bolívar | 209933 | 205 | San Miguel | 29004 | -1.7 | -79.033333 | 5 |
| 51 | 2 | Sierra | 4 | Carchi | 186869 | 402 | Bolívar | 15528 | 0.507099 | -77.862454 | 1 |
| 52 | 2 | Sierra | 4 | Carchi | 186869 | 403 | Espejo | 13817 | 0.621 | -77.94 | 1 |
| 53 | 2 | Sierra | 4 | Carchi | 186869 | 404 | Mira | 11969 | 0.5538999999999998 | -78.0493 | 1 |
| 54 | 2 | Sierra | 4 | Carchi | 186869 | 405 | Montúfar | 34229 | 0.61 | -77.84 | 1 |
| 55 | 2 | Sierra | 4 | Carchi | 186869 | 406 | San Pedro ... | 8931 | 0.631111 | -77.725278 | 1 |

Nota: Datos almacenados al realizar la carga desde el archivo `dim_geo_final` a la tabla de la dimensión geografía de la Data Warehouse (Mejía y Aguas, 2021).

2.4.4.7. *Diseño base de datos ETL*

La base de datos *ETL*, facilita el procesamiento y almacenamiento de la información alojada en los ficheros tipo CSV descritos en el proceso de *extracción de datos* y *almacenamiento local*.

El eje de análisis de casos positivos y de muertes divide en dos años que son 2020 y 2021, por esta razón la estructura de la base de datos consta de cinco tablas.

- `Positivas_2020`: almacena la data del eje de análisis de casos positivos del año 2020, manteniendo la estructura del encabezado detallado.
- `Positivas_2021`: almacena la data del eje de análisis de casos positivos del año 2021, manteniendo la estructura del encabezado detallado.
- `Muertes_2020`: almacena la data del eje de análisis de muertes del año 2020, manteniendo la estructura del encabezado detallado.

- Muertes_2021: almacena la data del eje de análisis de muertes del año 2021, manteniendo la estructura del encabezado detallado.
- Vacunas: almacena la data del eje de análisis de vacunas, manteniendo la estructura del encabezado detallado.

Figura 60

Tablas de la base de datos ETL.

| positivas_2020 | positivas_2021 | muertes_2020 | muertes_2021 | vacunas |
|-------------------------|-------------------------|-----------------------|-----------------------|----------------|
| 123 index | 123 index | 123 index | 123 index | 123 index |
| 123 inec_provincia_id | 123 inec_provincia_id | 123 inec_provincia_id | 123 inec_provincia_id | 🕒 fecha_vacuna |
| ABC provincia | ABC provincia | ABC provincia | ABC provincia | ABC zona |
| 123 provincia_poblacion | 123 provincia_poblacion | 123 poblacion | 123 poblacion | ABC provincia |
| 123 inec_canton_id | 123 inec_canton_id | 123 total | 123 total | ABC sexo |
| ABC canton | ABC canton | 123 nuevas | 123 nuevas | ABC edad |
| 123 canton_poblacion | 123 canton_poblacion | 123 lat | 123 lat | ABC nom_vacuna |
| 123 total | 123 total | 123 lng | 123 lng | 123 primera |
| 123 nuevas | 123 nuevas | 123 mes | 123 mes | 123 segunda |
| 123 lat | 123 lat | | | |
| 123 lng | 123 lng | | | |
| 123 mes | 123 mes | | | |

Nota: Tablas de los ejes de análisis en la base de datos ETL. Fuente: (Mejía y Aguas, 2021).

2.4.4.8. Disparadores de carga de datos a la base BI_COVID

Los disparadores facilitan la ejecución automática de consultas query al realizar una inserción en la tabla en la que se ha programado el disparador. Los disparadores para cada tabla de la base *ETL*, facilitan la carga automática de datos a la tabla de hechos de la base *BI_COVID*, la carga depende de los parámetros que se ingresen mediante la consulta query realizada en cada disparador.

La estructura del disparador se basa directamente en el uso de la función *dblink_exec*, esta función facilita ejecutar comandos SQL en una base de datos remota.

Los parámetros que ingresan en la función son: la conexión a la base de datos remota, y la consulta query.

El parámetro de conexión a la base de datos remota se mantiene en todos los disparadores, los datos que se pasa como parámetros para la conexión a la base son:

- dbname = BI_COVID
- user = postgres
- password = admin

2.4.4.7.1. Disparador en la tabla *positivas_2020* en la base ETL.

El disparador *carga_positiva_2020()*, presenta como parámetro de consulta la sentencia SQL *insert into* a la *tabla_hecho*, para identificar los campos que se va a cargar, se dispone de la función *select*: en donde se especifica que se necesita:

El id único de la dimensión geografía, id de tipo de análisis, id de tiempo, id de población y el valor total que se encuentra almacenado en el campo *total* de la tabla *positivas_2020*, los datos se deben extraer desde las tablas de dimensiones en este caso se usan tres dimensiones: geografía, tipo de análisis y tiempo.

Las condiciones para ejecutar la sentencia son: el parámetro *inec_canton_id* de la dimensión geografía debe ser igual al dato del campo *inec_canton_id* de la tabla *positivas_2020*, el mes de la dimensión tiempo debe ser igual al campo *mes* de la tabla *positivas_2020*, el *año* de la dimensión tiempo debe ser igual a 2020 y el *id_dim_tipo* de la dimensión debe ser igual a 1 que es el eje de análisis de positivas.

Figura 61

Disparador carga_positiva_2020.

```
BEGIN
  PERFORM dblink_exec('dbname=BI_COVID user=postgres password=admin',
    'insert into tabla_hechos (id_dim_geografia, id_dim_tipo, id_dim_tiempo, id_dim_muestra,total)
    select DISTINCT(d.id_dim_geo), ti.id_dim_tipo, tie.id_dim_tiempo,
    CASE
      WHEN d.canton_poblacion > 0 and d.canton_poblacion <= 5000 THEN 1
      WHEN d.canton_poblacion > 5000 and d.canton_poblacion <= 10000 THEN 2
      WHEN d.canton_poblacion > 10000 and d.canton_poblacion <= 50000 THEN 3
      WHEN d.canton_poblacion > 50000 and d.canton_poblacion <= 100000 THEN 4
      WHEN d.canton_poblacion > 100000 and d.canton_poblacion <= 500000 THEN 5
      WHEN d.canton_poblacion > 500000 and d.canton_poblacion <= 1000000 THEN 6
      WHEN d.canton_poblacion > 1000000 THEN 7
    END
    ,'|| NEW.total ||'
    from geografia_dim d, tipo_dim ti, tiempo_dim tie
    WHERE d.inec_canton_id = '||NEW.inec_canton_id||'
    and tie.mes = '||NEW.mes||'
    and tie.anio = 2020
    and ti.id_dim_tipo= 1
    GROUP BY d.id_dim_geo, ti.id_dim_tipo, tie.id_dim_tiempo;');
  return NEW ;
END;
```

Nota: Código del disparador carga_positiva_2020 en la tabla positivas_2020. Fuente: (Mejía y Aguas, 2021).

2.4.4.7.2. Disparador en la tabla positivas_2021 en la base ETL.

Los parámetros SQL de consulta *insert into* que toma en cuenta el disparador *carga_positiva_2021()*, a la *tabla_hecho* dispone de la función *select*: en el que se especifica que se necesita:

El id único de la dimensión geografía, id de tipo de análisis, id de tiempo, id de población y un valor total que se encuentra almacenado en el campo *total* de la tabla *positivas_2021*. Los datos se deben extraer desde las tablas dimensiones en este caso se dispone tres dimensiones: geografía, tipo de análisis y tiempo.

Las condiciones que debe cumplir son: el parámetro *inec_canton_id* de la dimensión geografía debe ser igual al dato del campo *inec_canton_id* de la tabla *positivas_2021*, el *mes* de la dimensión tiempo sea igual al campo *mes* de la tabla *positivas_2021*, el *año* de la dimensión

tiempo debe ser igual a 2021 y el *id_dim_tipo* de la dimensión debe ser igual a 1 que es el eje de análisis de positivas.

Figura 62

Disparador carga_positiva_2021.

```
BEGIN
  PERFORM dblink_exec('dbname=BI_COVID user=postgres password=admin',
    'insert into tabla_hechos (id_dim_geografia, id_dim_tipo, id_dim_tiempo, id_dim_muestra,total)
    select DISTINCT(d.id_dim_geo), ti.id_dim_tipo, tie.id_dim_tiempo,
    CASE
      WHEN d.canton_poblacion > 0 and d.canton_poblacion <= 5000 THEN 1
      WHEN d.canton_poblacion > 5000 and d.canton_poblacion <= 10000 THEN 2
      WHEN d.canton_poblacion > 10000 and d.canton_poblacion <= 50000 THEN 3
      WHEN d.canton_poblacion > 50000 and d.canton_poblacion <= 100000 THEN 4
      WHEN d.canton_poblacion > 100000 and d.canton_poblacion <= 500000 THEN 5
      WHEN d.canton_poblacion > 500000 and d.canton_poblacion <= 1000000 THEN 6
      WHEN d.canton_poblacion > 1000000 THEN 7
    END
    ,'|| NEW.total ||'
    from geografia_dim d, tipo_dim ti, tiempo_dim tie
    WHERE d.inec_canton_id = '||NEW.inec_canton_id||'
    and tie.mes = '||NEW.mes||'
    and tie.anio = 2021
    and ti.id_dim_tipo= 1
    GROUP BY d.id_dim_geo, ti.id_dim_tipo, tie.id_dim_tiempo;');
  return NEW ;
END;
```

Nota: Código del disparador *carga_positiva_2021* en la tabla *positivas_2021*. Fuente: (Mejía y Aguas, 2021).

2.4.4.7.3. Disparador en la tabla *muertes_2020* en la base ETL.

El disparador *carga_muertes_2020()*, utiliza la consulta SQL *insert into* a la *tabla_hecho*, para identificar los campos que se va a cargar, la función *select* especifica los parámetros que se necesita:

Un id de la dimensión geografía, un id de tipo de análisis, un id de tiempo, un id de población y un valor total que se encuentra almacenado en el campo *total* de la tabla *muertes_2020*, los datos se deben extraer desde las tablas de dimensiones en este caso se usan tres dimensiones la de geografía, la de tipo y la de tiempo.

Las condiciones que debe cumplir son: el parámetro *inec_canton_id* de la dimensión geografía debe ser igual al dato del campo *inec_canton_id* de la tabla *muertes_2020*, el *mes* de la dimensión tiempo debe ser igual al campo *mes* de la tabla *muertes_2020*, el *año* de la dimensión tiempo debe ser igual a *2020*, y el *id_dim_tipo* de la dimensión debe ser igual a *2* que es el eje de análisis de muertes.

Figura 63

Disparador carga_muerte_2020.

```
BEGIN
PERFORM dbblink_exec('dbname=BI_COVID user=postgres password=admin',
'insert into tabla_hechos (id_dim_geografia, id_dim_tipo, id_dim_tiempo, id_dim_muestra,total)
select d.id_dim_geo, ti.id_dim_tipo, tie.id_dim_tiempo,
CASE
WHEN d.poblacion > 0 and d.poblacion <= 5000 THEN 1
WHEN d.poblacion > 5000 and d.poblacion <= 10000 THEN 2
WHEN d.poblacion > 10000 and d.poblacion <= 50000 THEN 3
WHEN d.poblacion > 50000 and d.poblacion <= 100000 THEN 4
WHEN d.poblacion > 100000 and d.poblacion <= 500000 THEN 5
WHEN d.poblacion > 500000 and d.poblacion <= 1000000 THEN 6
WHEN d.poblacion > 1000000 THEN 7
END
, '||NEW.total||'
from geografia_dim d, tipo_dim ti, tiempo_dim tie
WHERE
d.inec_provincia_id = '||NEW.inec_provincia_id||'
and tie.mes = '||NEW.mes||'
and tie.anio = 2020
and ti.id_dim_tipo= 2
and d.canton is null');
return NEW ;
END;
```

Nota: Código del disparador *carga_muerte_2020* en la tabla *muerte_2020*. Fuente: (Mejía y Aguas, 2021).

2.4.4.7.4. *Disparador en la tabla muertes_2021 en la base ETL.*

El disparador *carga_muertes_2021()*, tiene como parámetro de consulta la sentencia SQL *insert into* a la *tabla_hecho*, que permite identificar los campos que se va a cargar, se hace uso de la función *select*: en el cual se especifica que se necesita:

Un id de la dimensión geografía, un id de tipo de análisis, un id de tiempo, un id de población y un valor total que se encuentra almacenado en el campo *total* de la tabla *muertes_2021*, los datos se deben extraer desde las tablas de dimensiones en este caso se usan tres dimensiones la de geografía, la de tiempo y la de población.

Las condiciones que debe cumplir son: el parámetro *inec_canton_id* de la dimensión geografía debe ser igual al dato del campo *inec_canton_id* de la tabla *muertes_2021*, el *mes* de la dimensión tiempo debe ser igual al campo *mes* de la tabla *muertes_2021*, el *año* de la dimensión tiempo debe ser igual a *2021* y el *id_dim_tipo* de la dimensión debe ser igual a *2* que es el eje de análisis de muertes.

Figura 64

Disparador carga_muerte_2021.

```
BEGIN
PERFORM dblink_exec('dbname=BI_COVID user=postgres password=admin',
'insert into tabla_hechos (id_dim_geografia, id_dim_tipo, id_dim_tiempo, id_dim_muestra,total)
select d.id_dim_geo, ti.id_dim_tipo, tie.id_dim_tiempo,
CASE
WHEN d.poblacion > 0 and d.poblacion <= 5000 THEN 1
WHEN d.poblacion > 5000 and d.poblacion <= 10000 THEN 2
WHEN d.poblacion > 10000 and d.poblacion <= 50000 THEN 3
WHEN d.poblacion > 50000 and d.poblacion <= 100000 THEN 4
WHEN d.poblacion > 100000 and d.poblacion <= 500000 THEN 5
WHEN d.poblacion > 500000 and d.poblacion <= 1000000 THEN 6
WHEN d.poblacion > 1000000 THEN 7
END
, '||NEW.total||'
from geografia_dim d, tipo_dim ti, tiempo_dim tie
WHERE
d.inec_provincia_id = '||NEW.inec_provincia_id||'
and tie.mes = '||NEW.mes||'
and tie.anio = 2021
and ti.id_dim_tipo= 2
and d.canton is null');
return NEW ;
END;
```

Nota: Código del disparador *carga_muerte_2021* en la tabla *muerte_2021*. Fuente: (Mejía y Aguas, 2021).

2.4.4.7.5. *Disparador en la tabla vacunas en la base ETL.*

El disparador *carga_vacunas* (), necesita como parámetros a cumplir para la correcta ejecución: el id de tiempo, la cantidad total de población vacunada con una cierta dosis, id de la edad, id

de geografía, id de género, id de tipo, id tipo de vacuna, id dosis. Para obtener esta información se deben extraer de las dimensiones tiempo, edad, geografía, género, tipo, vacuna, dosis.

Las condiciones que debe cumplir son: el *mes* de la dimensión tiempo debe ser igual al *mes* del campo *fecha_vacuna*, de la tabla *vacunas*, el *año* de la dimensión debe ser igual a 2021, la *edad* de la dimensión edad debe ser igual al campo *gedad* de la tabla *vacunas*, el *id* de la dimensión tipo debe ser igual a 3 que es el eje de análisis de vacunados, se especifica el tipo de vacuna comparando el campo *nom_vacuna* en la tabla *vacunas* con datos quemados, el campo *provincia* de la dimensión geografía debe ser igual al campo *provincia* en la tabla *vacunas*.

Figura 65

Disparador carga_vacunas.

```
BEGIN
PERFORM dblink_exec('dbname=BI_COVID user=postgres password=admin',
                    'insert into tabla_hechos (id_dim_tiempo,total,id_dim_edad, id_dim_genero, id_dim_tipo, id_dim_vacuna_tipo,id_d
                    select ti.id_dim_tiempo as tiempo,sum('||NEW.primer||'), ed.id_dim_edad as edad,
                    ge.id_dim_genero, ty.id_dim_tipo, vt.id_dim_vacuna_tipo, geo.id_dim_geo, dos.id_dim_dosis
                    from tiempo_dim ti, edad_dim ed,
                    genero_dim ge, tipo_dim ty, vacuna_tipo_dim vt, geografia_dim geo,
                    dosis_dim dos
                    where
                    ti.mes = EXTRACT(MONTH from TO_DATE('||NEW.fecha_vacuna||', 'yyyy/mm/dd'))
                    and ti.anio = 2021
                    and ed.edad = '||NEW.gedad||'
                    and ge.genero= lower('||NEW.sexo||')
                    and ty.id_dim_tipo = 3
                    and vt.vacuna =
                    CASE
                    WHEN '||NEW.nom_vacuna||' = 'CoronaVac SINOVAC' THEN 'sinovac'
                    WHEN '||NEW.nom_vacuna||' = 'CHADOX1S RECOMBINANTE ASTRAZENECA' THEN 'astrazeneca'
                    WHEN '||NEW.nom_vacuna||' = 'BNT162b2 PFIZER' THEN 'pfizer'
                    END
                    and unaccent(geo.provincia) = initcap('||NEW.provincia||')
                    and geo.inec_canton_id is null
                    and dos.id_dim_dosis = 1
                    GROUP BY tiempo, ed.id_dim_edad, ge.id_dim_genero,
                    ty.id_dim_tipo, vt.id_dim_vacuna_tipo,geo.id_dim_geo, dos.id_dim_dosis');
return NEW ;
END;
```

Nota: Código del disparador carga_vacunas en la tabla vacunas. Fuente: (Mejía y Aguas, 2021).

2.4.4.9. Funciones de carga de datos a la base ETL

Para la carga de datos a la base ETL, se utiliza código escrito en lenguaje Python apoyándose en las librerías *Pandas* y *sqlalchemy*.

El proceso de carga de datos cumple con la función de transformación de datos del proceso ETL, y a su vez permite realizar nuevas inserciones de datos a las tablas de la base ETL, estas

tablas activan el disparador configurado que realizan la inserción dependiendo del query que tenga a la tabla de hechos en la base BI_COVID.

El procedimiento para realizar la carga de datos es el siguiente.

2.4.4.9.1. *Conexión Python con PostgreSQL.*

La conexión de Python con la base de datos ETL alojada en PostgreSQL se efectúa por medio de la librería *sqlalchemy*, que facilita el acceso a una base de datos relacional y realizar operaciones sobre la misma.

El método *create_engine*, ejecuta la conexión entre Python y la base ETL, el parámetro para el método es la cadena de conexión a la base, el resultado de la conexión se almacena en la variable *engine*.

Figura 66

Conexión Python con base ETL.

```
from sqlalchemy import create_engine
import pandas as pd
engine = create_engine('postgresql+psycopg2://postgres:admin@localhost:5432/ETL')
```

Nota: Código Python para generar la interfaz de conexión a la base de datos ETL. Fuente: (Mejía y Aguas, 2021).

Para la carga de datos desde los diferentes archivos planos alojados en el almacenamiento interno, se utiliza librería *pandas*. La variable *df* almacena la información leída por el método *read_csv* aplicado al archivo plano.

Importar el tipo de dato *Float* desde los *types* de la librería *sqlalchemy*, para controlar que los datos de latitud y longitud se registren con valores decimales.

Se aplica el método *to_sql* a la variable *df*, para que se registren los datos almacenados en la variable *df*.

2.4.4.9.2. Carga de datos a la tabla positivas_2020 en la base ETL.

La carga para la tabla positiva_2020 se realiza desde el fichero cantones_mensual.inec.csv alojado en ‘\Unidad_de_titulacion\ecuacovid\datos_crudos\positivas\2020’.

Los parámetros que se necesitan pasar al método *to_sql* son:

- Nombre de la tabla: positivas_2020.
- Resultado de la conexión entre la base y PostgreSQL: engine.
- Index = true: indica que se conserve la estructura del encabezado para las columnas de la tabla.
- Dtype: se especifica los tipos de datos en caso de ser necesarios para evitar errores.
- If_exists = append: indica que se adjunten los datos en caso de existir data registrada

Figura 67

Carga de datos a positivas_2020.

```
#Positivas 2020
df = pd.read_csv('/Users/carlo/Desktop/Unidad_de_titulacion/ecuacovid/datos_crudos/positivas/2020/'
                +'cantones_mensual.inec.csv')
from sqlalchemy.types import Float
df.to_sql('positivas_2020', engine, index=True, dtype={"lat": Float(), "lng": Float()}, if_exists="append");
```

Nota: Código Python para cargar la data a la tabla positivas_2020 en ETL. Fuente: (Mejía y Aguas, 2021).

2.4.4.9.3. Carga de datos a la tabla positivas_2021 en la base ETL.

Para la tabla positivas_2021 la carga de datos se ejecuta desde el fichero cantones_mensual.inec.csv alojado en la ruta del almacenamiento interno ‘\Unidad_de_titulacion\ecuacovid\datos_crudos\positivas’.

Los parámetros que se necesitan pasar al método *to_sql* son:

- Nombre de la tabla: positivas_2021.
- Resultado de la conexión entre la base y PostgreSQL: engine.

- `Índex = true`: indica que se conserve la estructura del encabezado para las columnas de la tabla.
- `Dtype`: se especifica los tipos de datos en caso de ser necesarios para evitar errores.
- `If_exists = append`: indica que se adjunten los datos en caso de existir data registrada

Figura 68

Carga de datos a positivas_2021.

```
#Positivas 2021
df = pd.read_csv('/Users/carlo/Desktop/Unidad_de_tittulacion/ecuacovid/datos_crudos/positivas/'+
                'cantones_mensual.inec.csv')
from sqlalchemy.types import Float
df.to_sql('positivas_2021', engine, index=True, dtype={"lat": Float(), "lng": Float()}, if_exists="append");
```

Nota: Código Python para cargar la data a la tabla `positivas_2021` en ETL. Fuente: (Mejía y Aguas, 2021).

2.4.4.9.4. Carga de datos a la tabla `muerres_2020` en la base ETL.

Para la tabla `muerres_2020` la carga de datos se realiza desde el fichero `provincias_mensual.inec.csv` alojado en `'\Unidad_de_titulacion\ecuacovid\datos_crudos\muerres\2020\'`.

Los parámetros que se necesitan pasar al método `to_sql` son:

- Nombre de la tabla: `muerres_2020`.
- Resultado de la conexión entre la base y PostgreSQL: `engine`.
- `Índex = true`: indica que se conserve la estructura del encabezado para las columnas de la tabla.
- `Dtype`: se especifica los tipos de datos en caso de ser necesarios para evitar errores.
- `If_exists = append`: indica que se adjunten los datos en caso de existir data registrada

Figura 69

Carga de datos a muertes_2020.

```
#Muertes 2020
df = pd.read_csv('/Users/carlo/Desktop/Unidad_de_tittulacion/ecuacovid/datos_crudos/muertes/2020/'
                +'provincias_mensual.inec.csv')
from sqlalchemy.types import Float
df.to_sql('muertes_2020', engine, index=True, dtype={"lat": Float(), "lng": Float()}, if_exists="append");
```

Nota: Código Python para cargar la data a la tabla muertes_2020 en ETL. Fuente: (Mejía y Aguas, 2021).

2.4.4.9.5. Carga de datos a la tabla muertes_2021 en la base ETL.

Para la tabla muertes_2021 la carga de datos se realiza desde el fichero provincias_mensual.inec.csv alojado en '\Unidad_de_titulacion\ecuacovid\datos_crudos\muertes\'

Los parámetros que se necesitan pasar al método *to_sql* son:

- Nombre de la tabla: muertes_2021.
- Resultado de la conexión entre la base y PostgreSQL: engine.
- Index = true: indica que se conserve la estructura del encabezado para las columnas de la tabla.
- Dtype: se especifica los tipos de datos en caso de ser necesarios para evitar errores.
- If_exists = append: indica que se adjunten los datos en caso de existir data registrada

Figura 70

Carga de datos a muertes_2021.

```
#Muertes 2021
df = pd.read_csv('/Users/carlo/Desktop/Unidad_de_tittulacion/ecuacovid/datos_crudos/muertes/'
                +'provincias_mensual.inec.csv')
from sqlalchemy.types import Float
df.to_sql('muertes_2021', engine, index=True, dtype={"lat": Float(), "lng": Float()}, if_exists="append");
```

Nota: Código Python para cargar la data a la tabla muertes_2021 en ETL. Fuente: (Mejía y Aguas, 2021).

2.4.4.9.6. Carga de data a la tabla vacunas en la base ETL.

Para la carga de la data desde el fichero MSP_VACUNA_COVID_2021.csv, alojada en ‘\Unidad_de_titulacion\’.

Los parámetros que se necesitan pasar al método *to_sql* son:

- Encoding: indica la codificación del escrito del archivo en este caso *ISO-8859-1*.
- Sep: configura el tipo de separación de los datos dentro del archivo plano.
- Nombre de la tabla: vacunas.
- Resultado de la conexión entre la base y PostgreSQL: engine.
- Index = true: indica que se conserve la estructura del encabezado para las columnas de la tabla.
- Dtype: se especifica los tipos de datos en caso de ser necesarios para evitar errores, en este caso se debe configurar la columna *fecha_vacuna* del archivo de tipo *Date*.

Figura 71

Carga de datos a vacunas.

```
#Vacunados 2021
df = pd.read_csv('/Users/carlo/Desktop/Unidad_de_titulacion/MSP_VACUNA_COVID_2021.csv',
                sep = ";", encoding = "ISO-8859-1")
from sqlalchemy.types import Date
df.to_sql('vacunas', engine, index=True, dtype={"fecha_vacuna": Date()}, if_exists="append");
```

Nota: Código Python para cargar la data a la tabla vacunas en ETL. Fuente: (Mejía y Aguas, 2021).

CAPÍTULO III

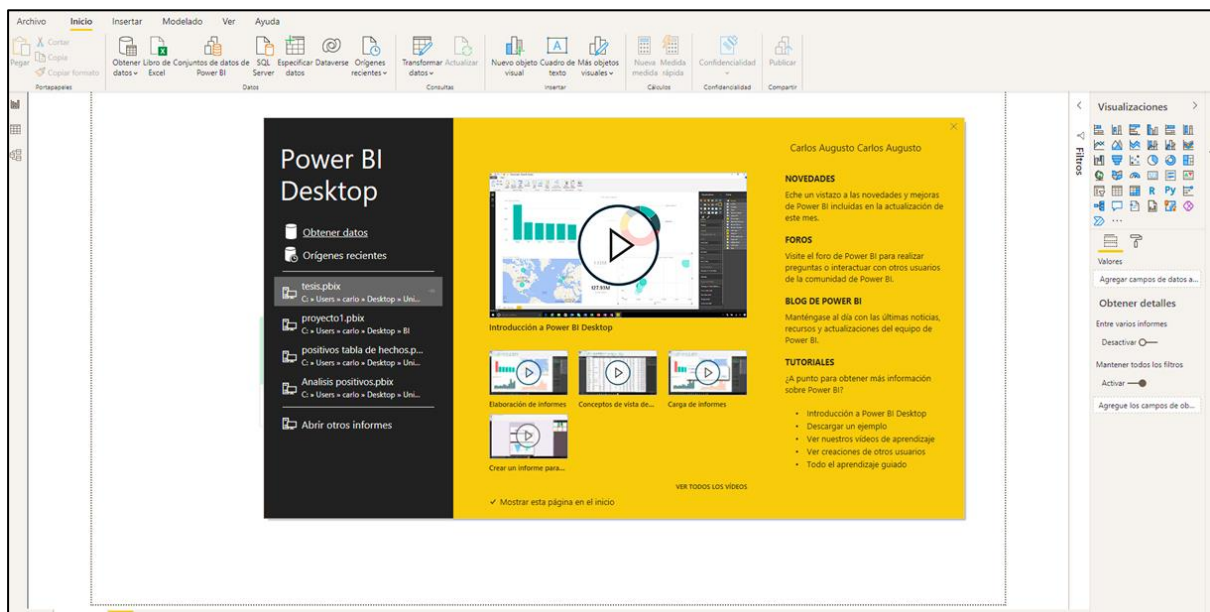
HERRAMIENTA DE INTELIGENCIA DE NEGOCIO

3.1. CONEXIÓN POWER BI Y POSTGRESQL

En el entorno de trabajo de Power BI, seleccionar desde la pestaña home la opción de *Obtener Datos*, o también puede seleccionar esta opción desde la pantalla de bienvenida de Power BI.

Figura 72

Pantalla de bienvenida de Power BI.

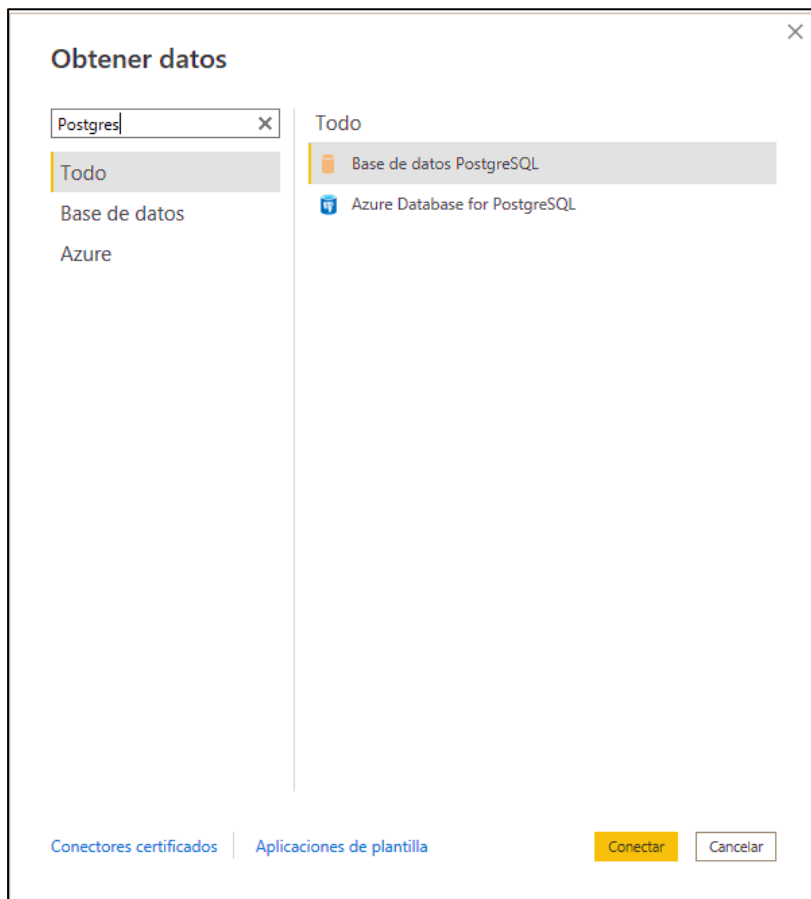


Nota: Pantalla de bienvenida de Power BI. Fuente: (Mejía y Aguas, 2021).

Se despliega un listado de gestores de base de datos, seleccionar *PostgreSQL database*, en caso de que no se encuentre, se permite realizar una búsqueda para el gestor deseado. Una vez seleccionado el gestor de base de datos a utilizar se debe proceder a *Conectar*.

Figura 73

Gestores de base de datos de Power BI.

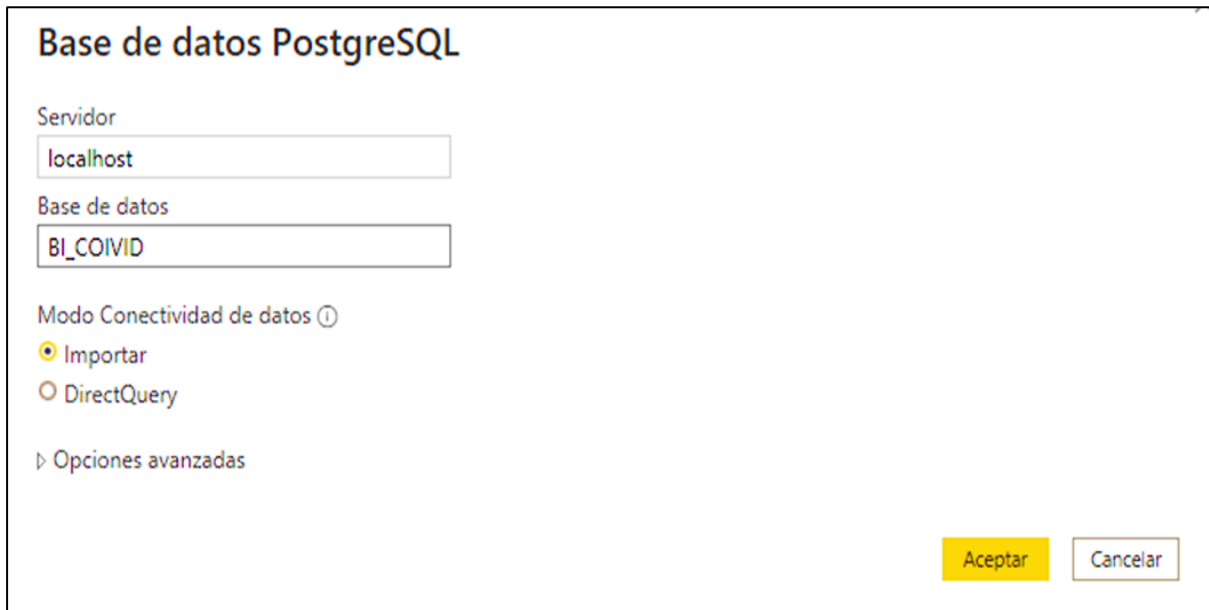


Nota: Gestores de base de datos existentes de Power BI. Fuente: (Mejía y Aguas, 2021).

Indicar el nombre del servidor en el que se encuentra alojada la base de datos *“localhost”*, el nombre de la base de datos a la que se va a conectar *“BI_COVID”*, el modo de conectividad debe ser *“Importar”*, la demás configuración debe ser la indicada por defecto.

Figura 74

Conexión a PostgreSQL desde Power BI.



The image shows a dialog box titled "Base de datos PostgreSQL". It contains the following fields and options:

- Servidor:** A text input field containing "localhost".
- Base de datos:** A text input field containing "BI_COVID".
- Modo Conectividad de datos:** A section with a help icon (i) and two radio button options:
 - Importar
 - DirectQuery
- Opciones avanzadas:** A collapsed section indicated by a right-pointing triangle.
- Buttons:** A yellow "Aceptar" button and a white "Cancelar" button with a grey border.

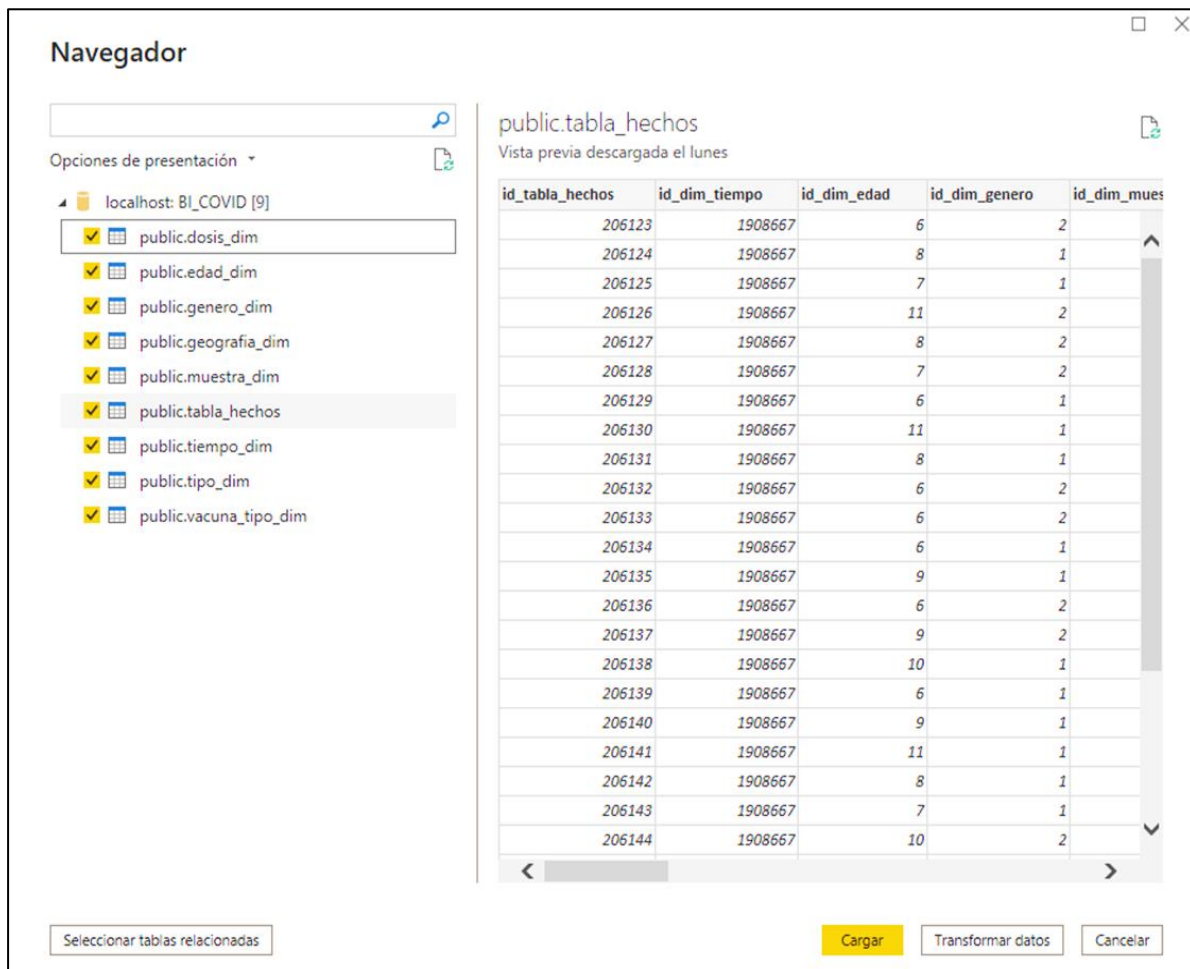
Nota: Configuración de la conexión de la base de datos PostgreSQL desde Power BI. Fuente: (Mejía y Aguas, 2021).

Se despliegan los esquemas generados en la base de datos BI_COVID al cargar al entorno de Power BI, para este caso seleccionar las tablas de las dimensiones: *dosis_dim*, *edad_dim*, *genero_dim*, *geografia_dim*, *muestra_dim*, *tiempo_dim*, *tipo_dim*, *vacuna_tipo_dim*, y la tabla hechos: *tabla_hechos*.

Los esquemas cargados al ser seleccionados, presentan una vista previa de los datos almacenados para confirmar que es correcta la información que se está vinculando,

Figura 75

Esquemas para cargar al entorno de Power BI.

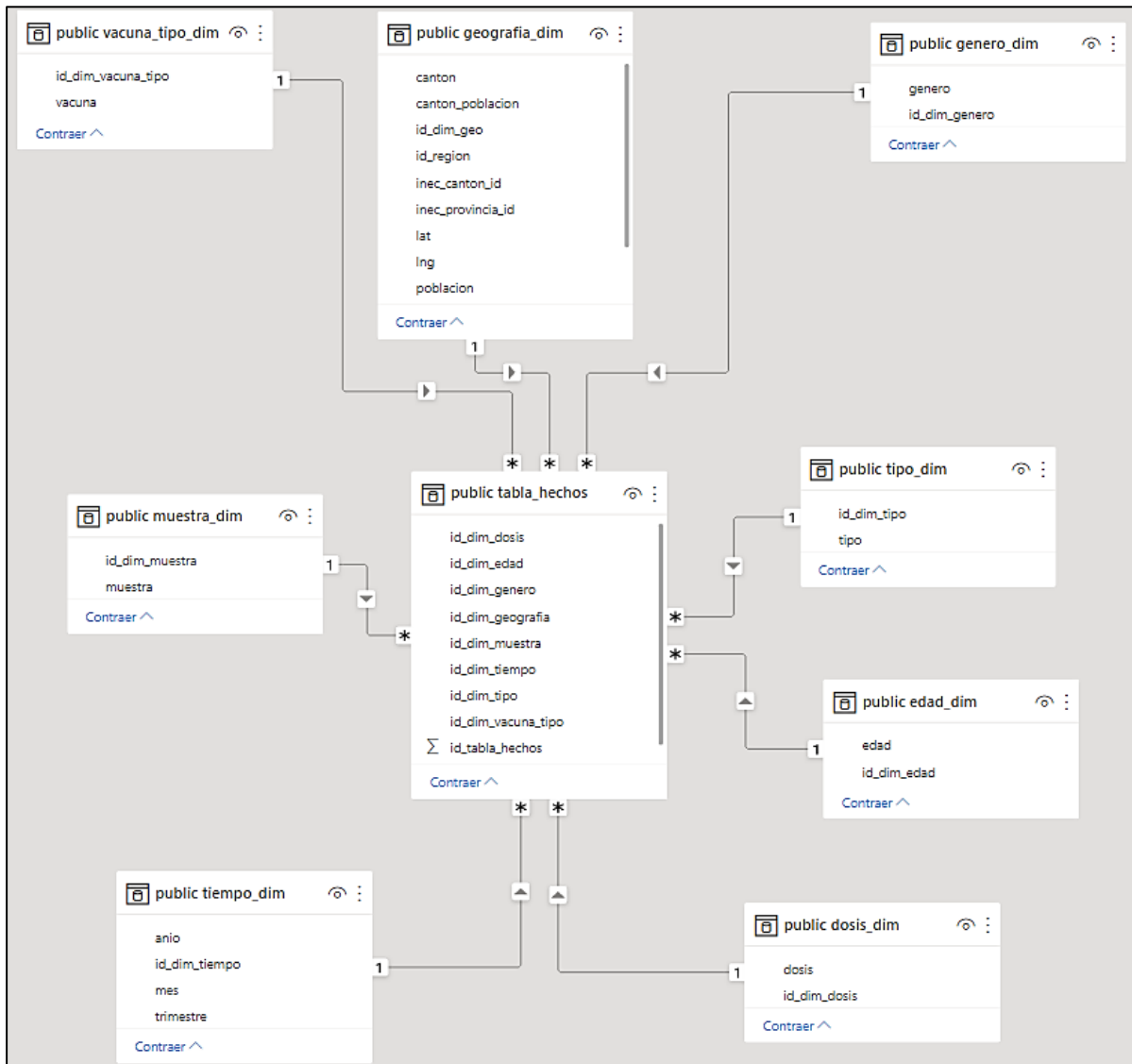


Nota: Lista de los esquemas (tablas de las dimensiones y de hechos) generados en BI_COVID, para seleccionar y cargar al entorno de Power BI. Fuente: (Mejía y Aguas, 2021).

Para verificar que la conexión a la base de datos BI_COVID se ha realizado de manera correcta y que la carga de los esquemas es de manera exitosa, se verifica el esquema generado en el entorno de Power BI.

Figura 76

Relación de los esquemas cargados a Power BI.



Nota: Relación de los esquemas (Tabla de las dimensiones y de hechos) cargados desde BI_COVID al entorno de Power BI. Fuente: (Mejía y Aguas, 2021).

3.2. DISEÑO

Se describe las vistas que presenta la herramienta de negocios, las vistas se utilizan para los tres ejes de análisis. Se presenta el modelado y navegabilidad que tiene cada vista, el funcionamiento de los filtros que presenta y la interpretación de los resultados presentados.

3.2.1. PANTALLA PRINCIPAL.

La pantalla principal presenta:

- Información básica de los involucrados en el desarrollo del presente proyecto.
- Descripción del virus SARS-COV-2 que es el tema principal de análisis.
- La descripción corta de la situación en el país y el actuar de los organismos internos encargados en Ecuador frente a la emergencia sanitaria.
- Ejes de análisis (casos positivos, muertes, vacunas) que se puede escoger, para ello se presenta botones de navegación intuitivos para obtener una fácil navegación entre las diferentes vistas de PowerBI.

Figura 77

Pantalla principal de la herramienta Power BI.



Nota: Pantalla inicial de la herramienta. Fuente: (Mejía y Aguas, 2021).

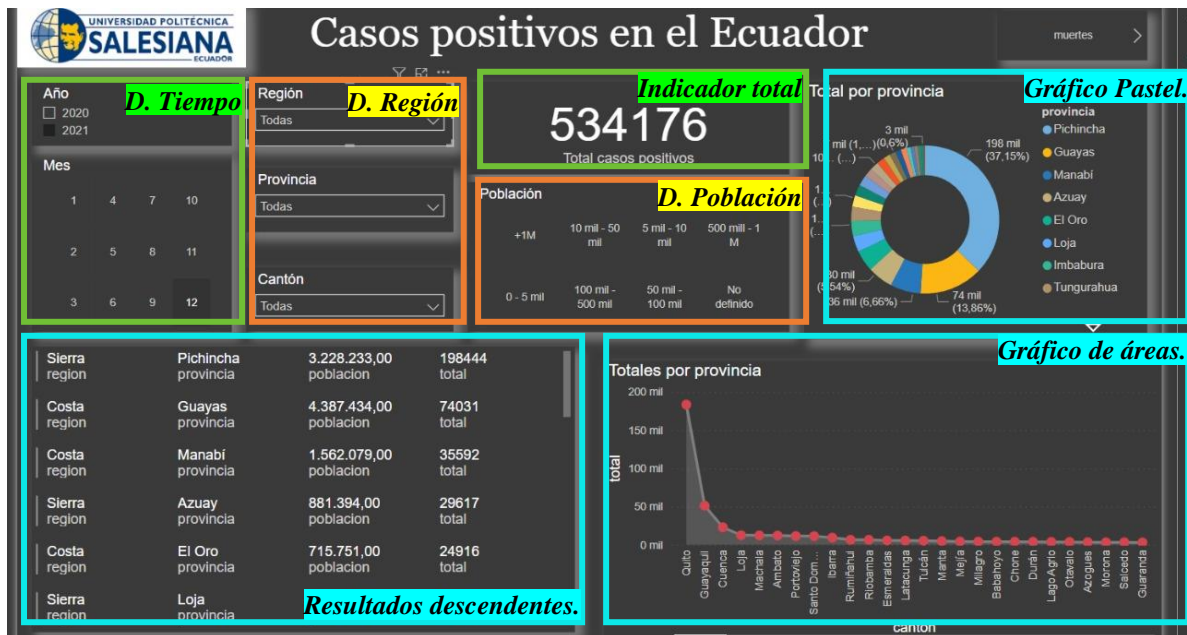
3.2.2. INTERFAZ GRÁFICA DE CASOS POSITIVOS.

La vista presenta el análisis mediante parámetros de filtración, su estructura es:

- Dimensión tiempo: se ubica en la parte superior izquierda de la vista, se divide en dos vistas internas para cada índice de granularidad en este caso: Año 2020 y 2021, y Mes presenta en formato numérico los meses del año.
- Dimensión geografía: se ubica junto a la dimensión tiempo, se divide en tres vistas internas para cada índice de granularidad identificado: Región: presenta cuatro regiones Costa, Sierra, Amazonía y Galápagos, Provincia: presenta a manera de listado el nombre de las 24 provincias del Ecuador, y Cantón: presenta el listado de todos los cantones existentes en el área geográfica de Ecuador.
- Indicador total: se presenta junto a la dimensión tiempo, e indica el resultado total de la búsqueda generada por los filtros escogidos.
- Dimensión población: ubicada bajo el indicador total, se presenta a manera de cuadrícula las opciones que tiene como granularidad para filtrar la información deseada.
- Gráfico pastel: este gráfico muestra el porcentaje obtenido de la sumatoria de población registrada como casos positivos con relación al porcentaje total de la población, el nivel de granularidad que se presenta en esta gráfica es con respecto a cantón.
- Resultados descendentes: presenta la sumatoria total a nivel de provincia de casos registrados como positivos y los ordena en forma descendente.
- Gráfico de áreas: este gráfico indica en provincia o cantón dependiendo de la selección de data filtrada el nivel de región, el total de casos positivos dados en dicha ubicación geográfica dentro de un cierto rango de tiempo.

Figura 78

Interfaz gráfica del eje de análisis de casos positivos.



Nota: Interfaz gráfica eje de análisis de casos positivos. Fuente: (Mejía y Aguas, 2021).

3.2.3. INTERFAZ GRÁFICA DE MUERTES.

La vista presenta el análisis mediante parámetros de filtración, su estructura es:

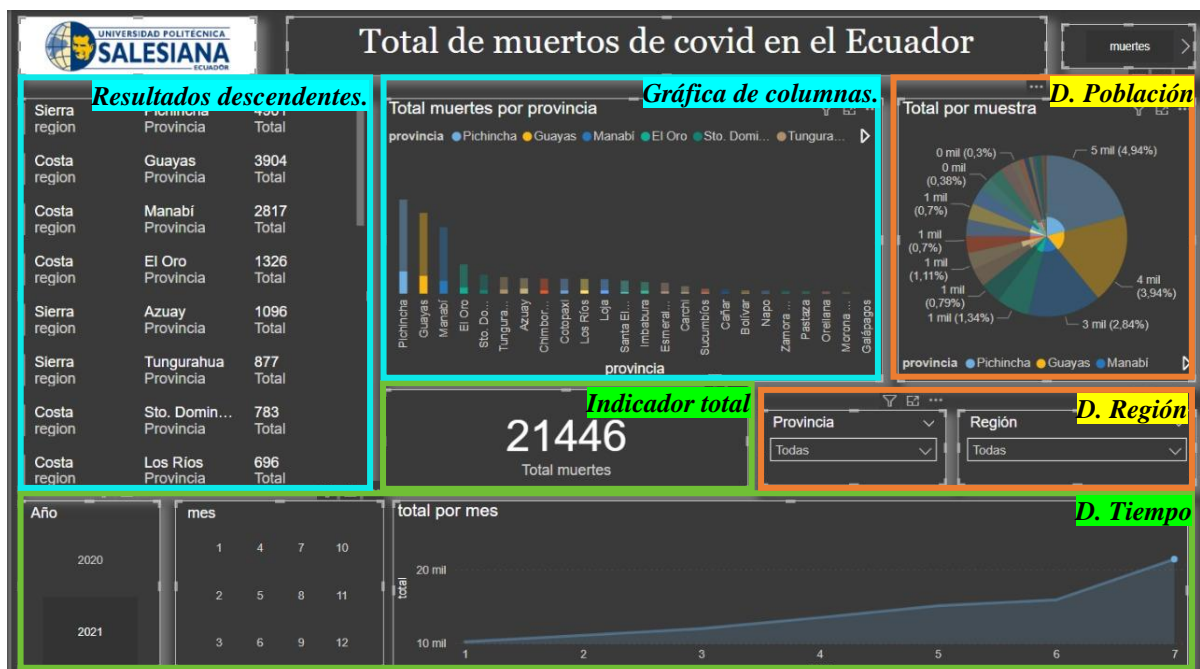
- **Dimensión tiempo:** se ubica en la parte inferior izquierda de la vista, por la granularidad disponible se presenta dos vistas internas para esta dimensión que son: Año: presenta datos de 2020 y 2021, y mes que representa el valor numérico de los 12 meses de cada año.
- **Dimensión región:** se ubica junto a la dimensión de tiempo en la parte inferior central de la vista, el nivel de granularidad para este análisis es de región y provincia: Región: Costa, Sierra, Amazonía y Galápagos, y Provincia: que presenta un listado de tipo select con los nombres de las 24 provincias del Ecuador.
- **Dimensión población:** se presenta a modo de gráfica de pastel ubicada en parte superior derecha, los rangos de filtros se presentan a manera de lista seleccionable.

- Indicador total: se presenta sobre la dimensión región, e indica el resultado total de la búsqueda generada por los filtros escogidos.
- Gráfica de columnas: muestra la totalidad de muertes registradas en cierto período de tiempo por provincia.
- Resultados descendentes: presenta a manera de listado vertical en orden descendente, la totalidad de muertes registradas a nivel de provincia dependiendo del filtro de tiempo que se puede pasar

El uso de dos o más modelos gráficos para representar los resultados obtenidos, facilitan la interpretación y verificación de la veracidad de los resultados.

Figura 79

Interfaz gráfica del eje de análisis muertes.



Nota: Interfaz gráfica eje de análisis de muertes. Fuente: (Mejía y Aguas, 2021).

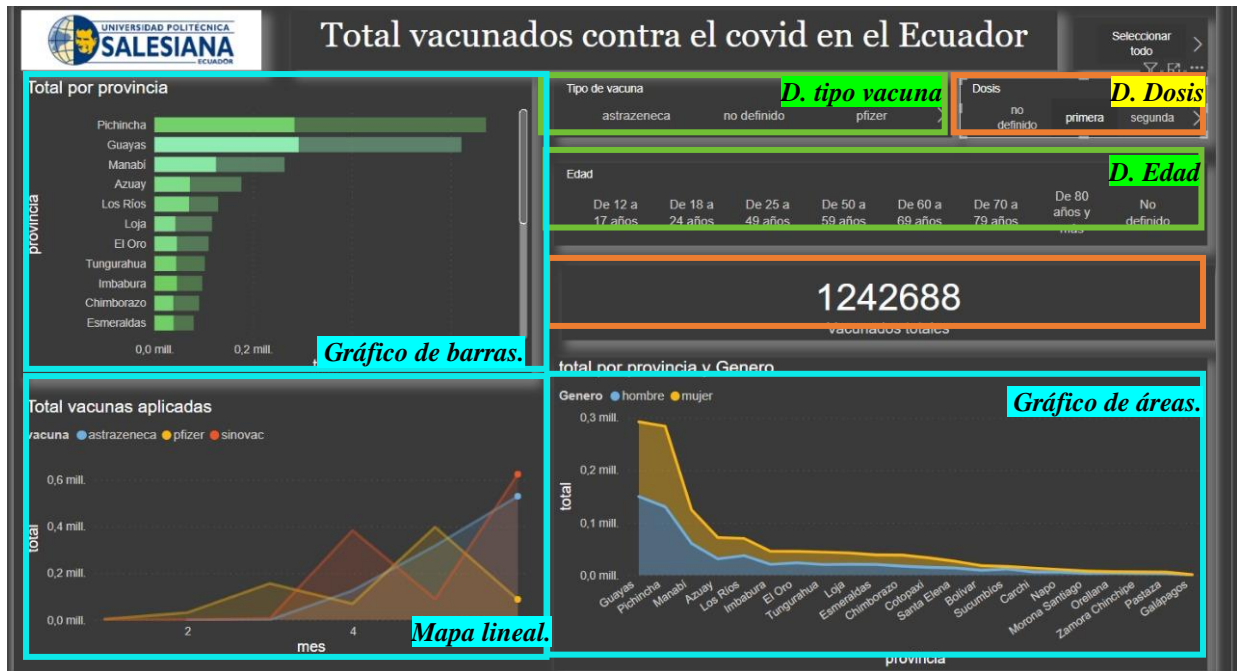
3.2.4. INTERFAZ GRÁFICA DE VACUNAS.

Esta vista presenta el análisis mediante parámetros de filtración y su estructura es:

- Dimensión tipo vacuna: se ubica en la parte superior derecha junto a la dimensión dosis, el nivel de granularidad facilita la filtración de los datos a nivel de AstraZeneca, Pfizer y Sinovac, la información de la dimensión se presenta a manera de lista horizontal formada por botones para cada opción. Facilitando que el usuario final elija una de estas de manera intuitiva.
- Dimensión dosis: ubicada en la parte superior derecha de la vista, esta dimensión facilita a manera de lista horizontal formada por botones las opciones que brinda para filtrar la información. Facilitando que el usuario final elija una de estas de manera intuitiva.
- Dimensión edad: se presenta bajo la dimensión tiempo, esta dimensión despliega sus datos a manera de lista horizontal formada por botones, cada uno toma un valor del campo edad de la dimensión.
- Indicador total: se presenta bajo de la dimensión edad, e indica el resultado total de la búsqueda generada por los filtros escogidos.
- Gráfico de barras: indica el total de vacunados, a mayor índice de personas vacunadas mayor es el tamaño de la barra.
- Mapa lineal: la gráfica indica el total de vacunas por tipo aplicadas por cada mes.
- Gráfico de áreas: esta gráfica indica la totalidad de vacunados en cuanto a provincia y género.

Figura 80

Interfaz gráfica del eje de análisis vacunados.



Nota: Interfaz gráfica eje de análisis de vacunados. Fuente: (Mejía y Aguas, 2021).

CAPÍTULO IV

ANÁLISIS DE RESULTADOS

4.1. Resultado del eje de análisis de casos positivos

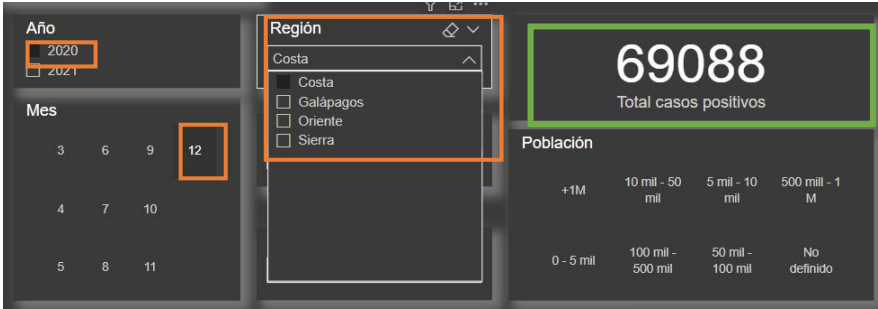
Los resultados que se presentan, pueden ser sostenidos mediante la comparativa entre los resultados y la información presentada por el SNGRE y el MSP en los informes generados, en este caso los informes N.º 67, 094, 124, 186.

Para el análisis de casos positivos, los datos proporcionados por las fuentes de información facilitan realizar el análisis desde el inicio de la pandemia hasta la primera semana de diciembre del 2021.

En el caso de los ejes de análisis de muertes y vacunados los datos han sido actualizados hasta el mes de julio del 2021.

Tabla 1

Región con el mayor índice de contagios en el año 2020.

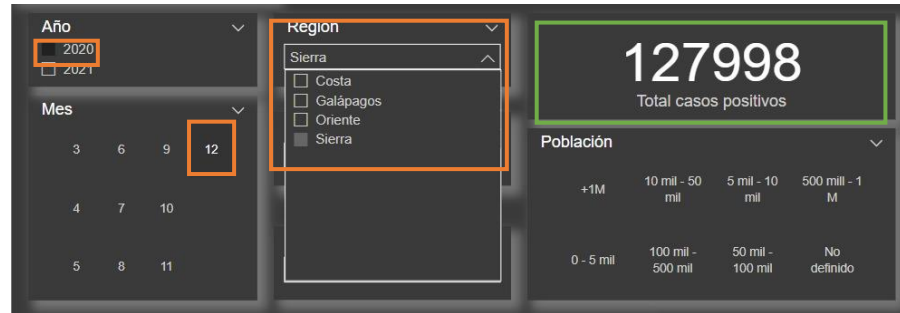
| | |
|---|--|
| Análisis | Región con el mayor índice de contagios en el año 2020. |
| Filtros: <ul style="list-style-type: none">• Año: 2020• Mes: 12• Región: Costa | <p>Figura 81</p> <p><i>Análisis positivos región Costa año 2020.</i></p>  <p>Nota: Análisis región Costa. Fuente: (Mejía y Aguas, 2021).</p> |

Filtros:

- Año: 2020
- Mes: 12
- Región: Sierra

Figura 82

Análisis positivos región Sierra año 2020.



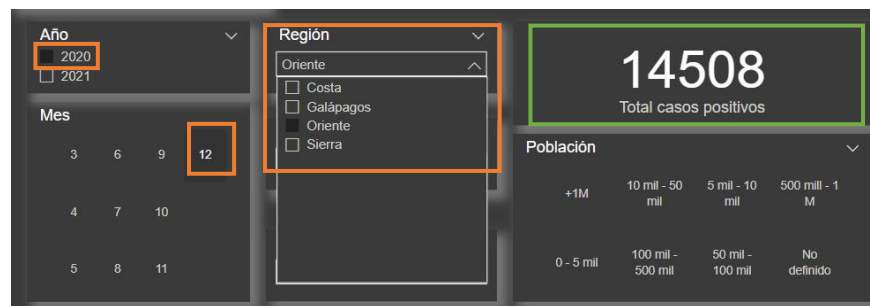
Nota: Análisis región Sierra. Fuente: (Mejía y Aguas, 2021).

Filtros:

- Año: 2020
- Mes: 12
- Región: Oriente

Figura 83

Análisis positivos región Oriente año 2020.



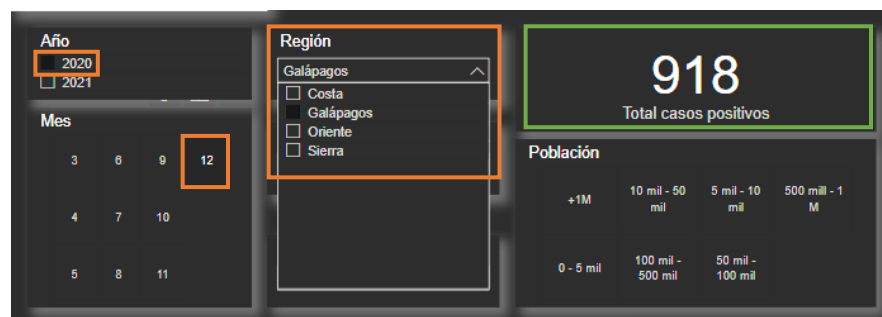
Nota: Análisis región Oriente. Fuente: (Mejía y Aguas, 2021).

Filtros:

- Año: 2020
- Mes: 12
- Región: Galápagos

Figura 84

Análisis positivos región Galápagos año 2020.



Nota: Análisis región Galápagos. Fuente: (Mejía y Aguas, 2021).

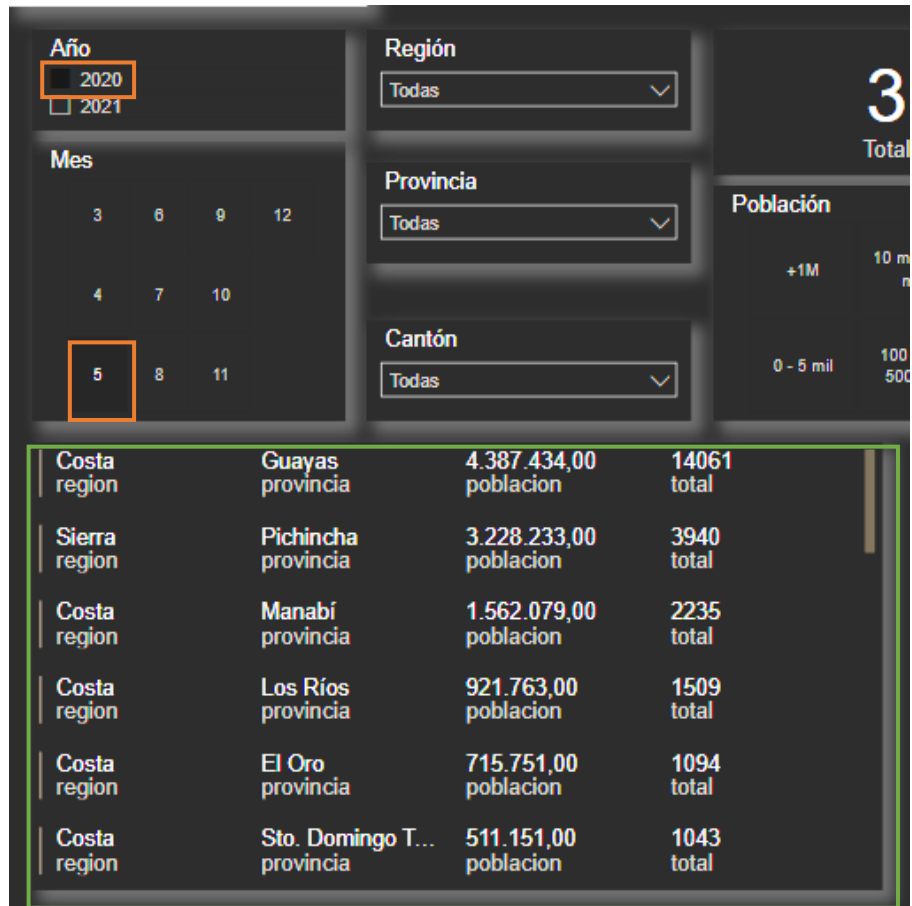
| | |
|-------------------|--|
| Resultados | Los resultados arrojados por el análisis generado, permiten identificar el total de casos positivos por cada región, la región Sierra presenta un mayor nivel de contagio al finalizar el año 2020, el total de la región Sierra duplica el total de registros confirmados de la región Costa que hasta finales del mes de junio del 2020 se muestra como la región que presentaba el mayor índice de registro de casos positivos. |
|-------------------|--|

Nota. Resultados del análisis del mayor índice de contagio por regiones. Elaborado por: Los autores.

Tabla 2

Índice de casos positivos por provincia en el primer trimestre de la pandemia.

| | |
|--|--|
| Análisis | Provincias con el mayor índice de casos en el primer trimestre de pandemia del año 2020. |
| Filtros: <ul style="list-style-type: none"> • Año: 2020 • Mes: 05 | Figura 85 <i>Análisis de casos positivos por provincias primer trimestre 2020.</i> |



Nota: Resultados del análisis de casos positivos por provincia en el primer trimestre de pandemia. Fuente: (Mejía y Aguas, 2021).

Resultados

El análisis de índices de casos positivos en el primer trimestre de la pandemia del año 2020, arroja como resultados, que las provincias que presentan un mayor nivel de contagio hasta el fin de mes de mayo del 2020, son Guayas, Pichincha, y Manabí.

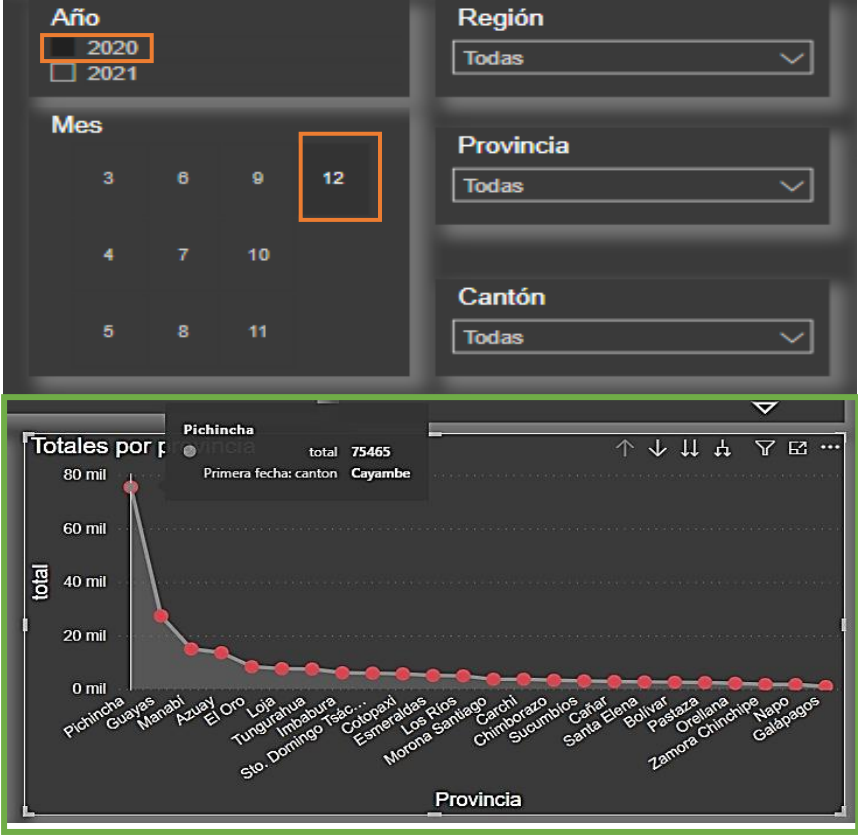
La provincia de Guayas presenta hasta la fecha del mes de mayo un índice de contagio que cuadruplica a la provincia de Pichincha, siendo la de mayor tasa de contagio, para el cierre del mes de mayo del 2020, el sistema de salud de la provincia de Guayas colapso debido a la gran

| | |
|--|--|
| | <p>demanda de enfermos y personas moribundas en las afueras de las casas de salud a raíz del contagio.</p> <p>Existía un desabastecimiento total de medicina en las farmacias, además de que el 40% del personal médico se encontraba contagiado por la falta de insumos de protección adecuados al inicio de la pandemia.</p> |
|--|--|

Nota. Resultados del análisis a nivel de provincias con el mayor índice de contagio en el primer trimestre de la pandemia. Elaborado por: Los autores.

Tabla 3

Provincia con mayor tasa de casos positivos al finalizar el año 2020.

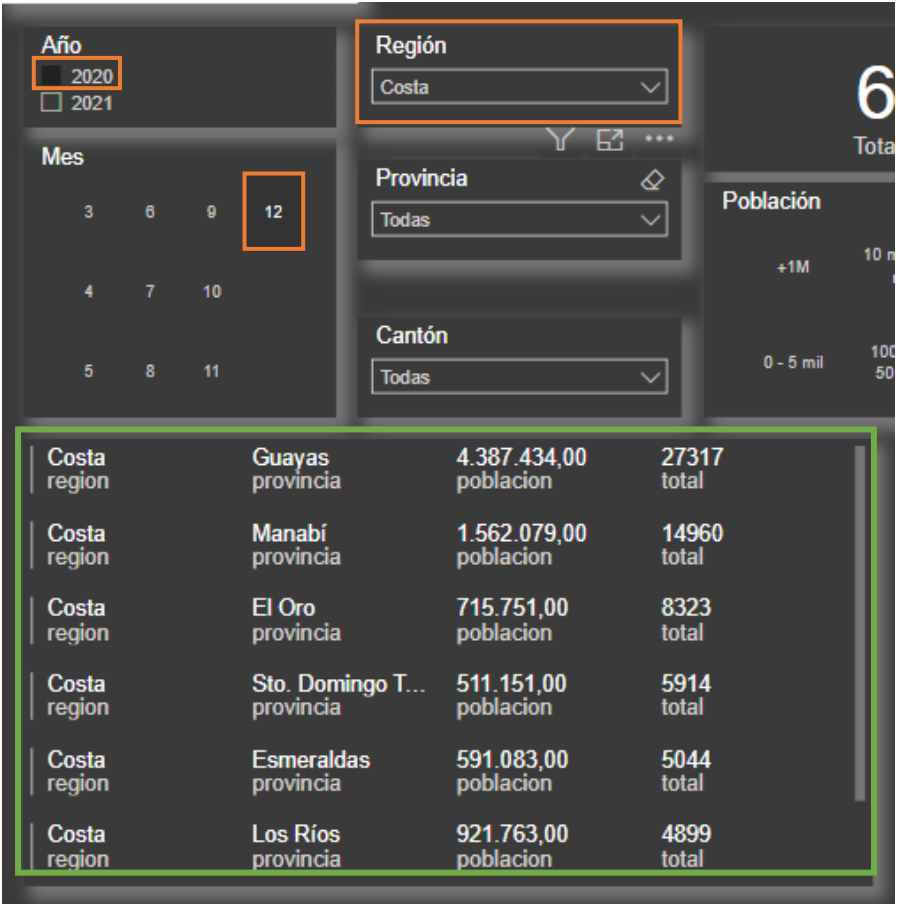
| | |
|---|--|
| <p>Análisis</p> | <p>Provincia con mayor tasa de casos positivos al finalizar el año 2020.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes: 12 | <p>Figura 86 <i>Análisis de casos positivos por provincias al finalizar el año 2020.</i></p>  <p>The dashboard includes filters for 'Año' (Year) set to 2020 and 'Mes' (Month) set to 12. It also has dropdown menus for 'Región', 'Provincia', and 'Cantón', all currently set to 'Todas'. The chart displays 'Totales por Provincia' for Pichincha, with a total of 75,465 cases. The x-axis lists provinces, and the y-axis shows the total number of cases in thousands (mil). Pichincha has the highest number of cases, followed by Guayas, Manabí, Azuay, and El Oro, with other provinces showing significantly lower case counts.</p> |

| | |
|--------------------------|--|
| | <p>Nota: Resultados del análisis de casos positivos con respecto a la provincia al finalizar el año 2020. Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Resultados</p> | <p>En la gráfica de áreas, se observa a Pichincha como la provincia que ocupa el primer lugar en cuanto a la tasa de registro de casos positivos al finalizar el año 2020, seguido de Guayas que para esta fecha reduce notoriamente su índice de contagios. Cabe recalcar que hasta la fecha de julio 2020 Guayas ocupaba el primer sitio en cuanto a registros positivos, a partir del mes de agosto Pichincha pasa a representar un 22.8% de los registros de casos confirmados; a partir de este incremento notorio, Pichincha se mantiene como la provincia con el mayor registro hasta finalizar el año presentando 75.645 casos de contagio como se puede visualizar en la Imagen 84.</p> |

Nota. Resultados del análisis del mayor índice de contagio por regiones. Elaborado por: Los autores.

Tabla 4

Provincia de cada región con mayor índice de casos positivos, año 2020.

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|--|------------------------|------------------|------------------------|-------------|--------------|------------------|------------------------|-------------|--------------|------------------|----------------------|------------|--------------|-----------------------------|----------------------|------------|--------------|----------------------|----------------------|------------|--------------|--------------------|----------------------|------------|
| <p>Análisis</p> | <p>Provincia de cada región con mayor índice de casos positivos año 2020.</p> | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes 12 • Región: Costa | <p>Figura 87</p> <p><i>Provincia con mayor índice de casos positivos región Costa.</i></p>  <p>Nota: Resultados de provincia con el mayor índice de casos positivos en la región Costa, año 2020. Fuente: (Mejía y Aguas, 2021).</p> <table border="1" data-bbox="491 1070 1356 1523"> <tbody> <tr> <td>Costa region</td> <td>Guayas provincia</td> <td>4.387.434,00 poblacion</td> <td>27317 total</td> </tr> <tr> <td>Costa region</td> <td>Manabí provincia</td> <td>1.562.079,00 poblacion</td> <td>14960 total</td> </tr> <tr> <td>Costa region</td> <td>El Oro provincia</td> <td>715.751,00 poblacion</td> <td>8323 total</td> </tr> <tr> <td>Costa region</td> <td>Sto. Domingo T... provincia</td> <td>511.151,00 poblacion</td> <td>5914 total</td> </tr> <tr> <td>Costa region</td> <td>Esmeraldas provincia</td> <td>591.083,00 poblacion</td> <td>5044 total</td> </tr> <tr> <td>Costa region</td> <td>Los Ríos provincia</td> <td>921.763,00 poblacion</td> <td>4899 total</td> </tr> </tbody> </table> | Costa region | Guayas provincia | 4.387.434,00 poblacion | 27317 total | Costa region | Manabí provincia | 1.562.079,00 poblacion | 14960 total | Costa region | El Oro provincia | 715.751,00 poblacion | 8323 total | Costa region | Sto. Domingo T... provincia | 511.151,00 poblacion | 5914 total | Costa region | Esmeraldas provincia | 591.083,00 poblacion | 5044 total | Costa region | Los Ríos provincia | 921.763,00 poblacion | 4899 total |
| Costa region | Guayas provincia | 4.387.434,00 poblacion | 27317 total | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Manabí provincia | 1.562.079,00 poblacion | 14960 total | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | El Oro provincia | 715.751,00 poblacion | 8323 total | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Sto. Domingo T... provincia | 511.151,00 poblacion | 5914 total | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Esmeraldas provincia | 591.083,00 poblacion | 5044 total | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Los Ríos provincia | 921.763,00 poblacion | 4899 total | | | | | | | | | | | | | | | | | | | | | | |
| <p>Resultado región Costa.</p> | <p>Los resultados obtenidos, presenta a la provincia de Guayas como la provincia con el mayor índice de registro de casos positivos en la región</p> | | | | | | | | | | | | | | | | | | | | | | | | |

Costa, el total para la provincia de Guayas es 27.317, cabe aclarar que Guayas presenta el mayor total de población dentro de la región.

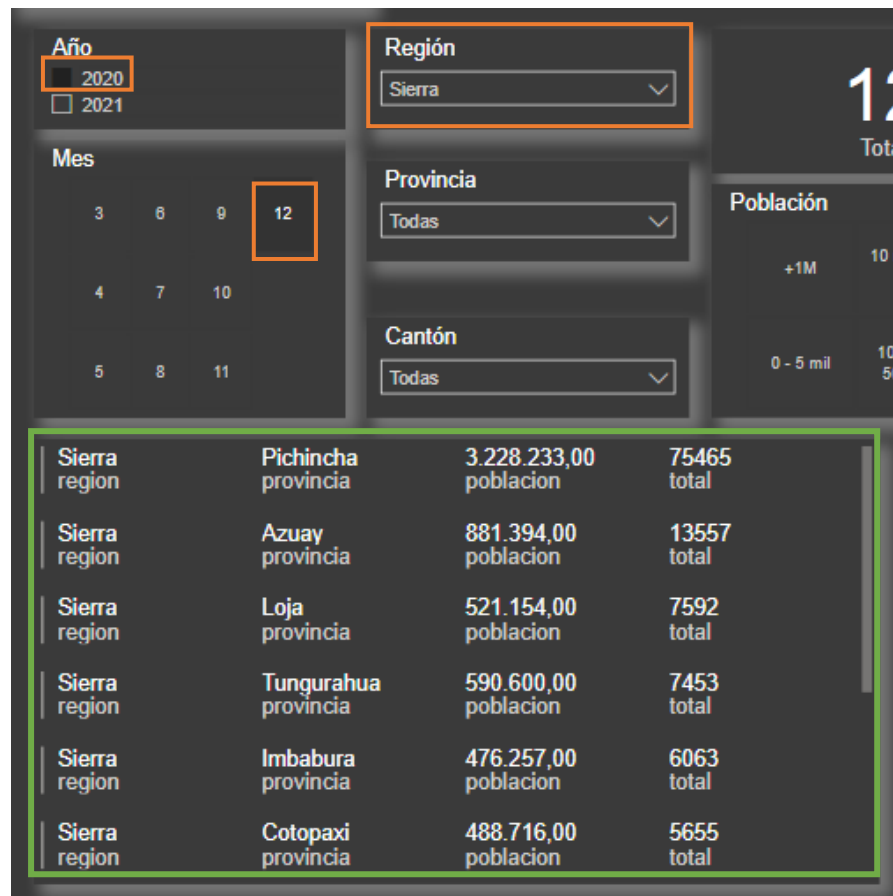
Una estrategia implementada en el ámbito social, fue la distribución de alimentos a domicilio para garantizar que no existan aglomeraciones en los mercados, el comité organizó a un grupo de encuestadores que iban acompañados por médicos o enfermeros que realizaban las respectivas pruebas de COVID-19, y así apoyar a las estadísticas reales del número de contagiados y asintomáticos por cada zona.

Filtros:

- Año: 2020
- Mes 12
- Región: Sierra

Figura 88

Provincia con mayor índice de casos positivos región Sierra.



Nota: Resultados de provincia con el mayor índice de casos positivos en la región Sierra, año 2020. Fuente: (Mejía y Aguas, 2021).

Resultado región Sierra

El análisis efectuado a la región Sierra, demuestra que la provincia de Pichincha presenta un mayor índice de registro de casos confirmados con un total de 75.465.

A partir del mes de julio del 2020 el registro de casos confirmados creció exponencialmente. El sector médico de la provincia fue afectado ya que directamente las áreas de UCI de cada clínica y hospital llegaron a la capacidad máxima de atención, esto debido a la problemática social de aglomeraciones en lugares de abastecimiento de comida, y realización de fiestas clandestinas.

- Filtros:**
- Año: 2020
 - Mes 12
 - Región: Oriente

Figura 89
Provincia con mayor índice de casos positivos región Oriente.

| Región | Provincia | Población | Total |
|----------------|----------------------------|----------------------|------------|
| Oriente region | Morona Santiago provincia | 196.535,00 poblacion | 3641 total |
| Oriente region | Sucumbíos provincia | 230.503,00 poblacion | 3009 total |
| Oriente region | Pastaza provincia | 114.202,00 poblacion | 2411 total |
| Oriente region | Orellana provincia | 161.338,00 poblacion | 2110 total |
| Oriente region | Zamora Chinch... provincia | 120.416,00 poblacion | 1669 total |
| Oriente region | Napo provincia | 133.705,00 poblacion | 1668 total |

Nota: Resultados de provincia con el mayor índice de casos positivos en la región Oriente, año 2020. Fuente: (Mejía y Aguas, 2021).

Resultado región Oriente Los resultados demuestran que la provincia de Morona Santiago presenta el mayor índice de casos positivos en la región Oriente, al finalizar el año 2020, el total obtenido para este análisis Morona Santiago presenta un total de 3.641 casos registrados.

Filtros:

- Año: 2020
- Mes 12
- Región: Galápagos

Figura 90
Provincia con mayor índice de casos positivos región Galápagos.

| | | | |
|------------------|---------------------|---------------------|-----------|
| Galápagos region | Galápagos provincia | 33.042,00 población | 918 total |
|------------------|---------------------|---------------------|-----------|


Nota: Resultados de provincia con el mayor índice de casos positivos en la región Galápagos, año 2020. Fuente: (Mejía y Aguas, 2021).

Resultado región Galápagos Al tener una sola provincia dentro de la región de Galápagos, era evidente que el resultado arroje a la provincia de Galápagos como la provincia que registra el mayor índice de casos positivos con un total de 918 personas contagiadas.

Nota. Resultados del análisis de provincia que presenta el mayor índice de casos positivos por región. Elaborado por: Los autores.

Tabla 5

Cantón con mayor índice de casos positivos en el primer semestre del año 2020.

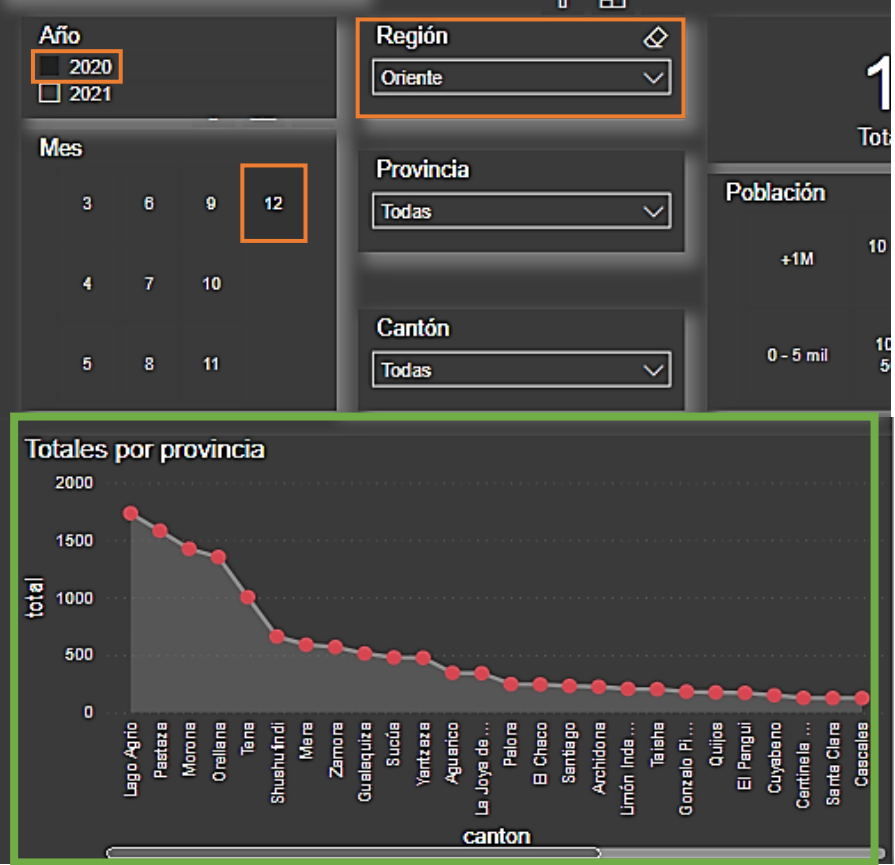
| | |
|--|--|
| <p>Análisis</p> | <p>Cantón con mayor índice de casos positivos en el primer semestre del año 2020.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes 06 | <p>Figura 91</p> <p><i>Cantón con mayor índice de casos positivos hasta el mes de junio 2020.</i></p>  <p>Nota: Resultados de cantón con el mayor índice de casos positivos en el primer semestre del 2020. Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Resultados</p> | <p>La gráfica de áreas, indica al cantón Guayaquil de la provincia de Guayas a la fecha de 30 de junio, presentaba el mayor índice de registro de casos positivos, debido a la falta de atención y control en el comportamiento social de la población, la sobreexplotación de atención en los centros de salud ocasionó la aglomeración de personas a las</p> |

afueras sin ningún control de distanciamiento, otra problemática social era el control y tratamiento de personas fallecidas.

Nota. Resultados del análisis por cantón que presenta el mayor índice de casos positivos. Elaborado por: Los autores.

Tabla 6

Cantón con el mayor índice de casos positivos en el año 2020, región oriente.

| | |
|--|--|
| <p>Análisis</p> | <p>Cantón con mayor índice de casos positivos en el año 2020, región oriente.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes: 12 • Región: Oriente | <p>Figura 92</p> <p><i>Cantón con mayor índice de casos positivos año 2020, región Oriente.</i></p>  <p>Nota: Resultados de cantón con el mayor índice de casos positivos, año 2020, región oriente. Fuente: (Mejía y Aguas, 2021).</p> |

| | |
|--------------------------|---|
| <p>Resultados</p> | <p>En los informes presentados por el MSP, se detalla el listado de cantones con los totales registrados, en él se demuestra que el cantón Lago Agrio presenta el mayor índice de contagio con un total de 1734, la gráfica de áreas posiciona al cantón Lago Agrio al inicio, lo que indica que es el cantón con mayor índice de casos positivos dentro de la región Oriente.</p> <p>Una grave problemática social en la provincia de Sucumbíos es que su ubicación geográfica facilita el tránsito irregular de personas, el Grupo de Trabajo para Refugiadas y Migrantes (GTRM) es el organismo encargado de controlar y apoyar esta situación, este organismo ha apoyado en el pago de recetas médicas, exámenes médicos y entrega de vouchers para el acceso a la atención médica de refugiados y emigrantes.</p> <p>El MSP en cuanto al tema médico, distribuyó insumos médicos, además de equipamiento electrónico para la revisión y la atención de casos positivos en los diferentes centros de salud disponibles.</p> |
|--------------------------|---|

Nota. Resultados del análisis por cantón que presenta el mayor índice de casos positivos. En el año del 2020, región oriente. Elaborado por: Los autores.

Tabla 7

Cantón con el mayor índice de casos positivos en la pandemia.

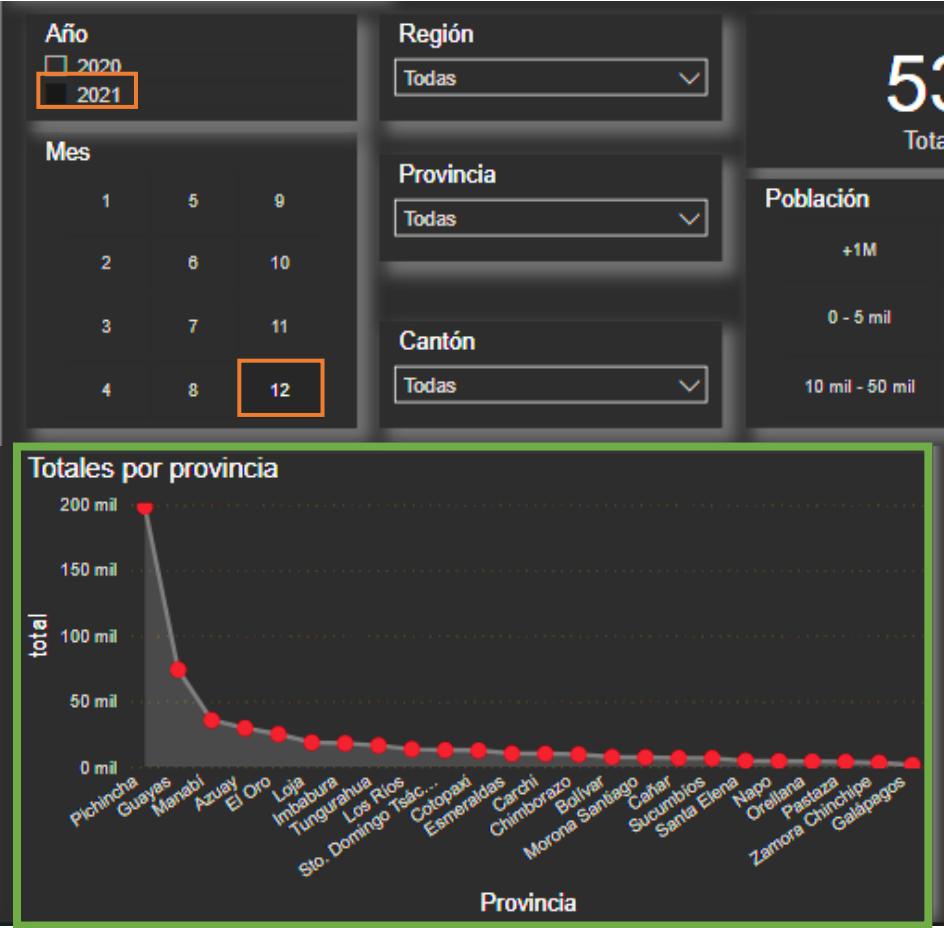
| | |
|---|--|
| <p>Análisis</p> | <p>Cantón con el mayor índice de casos positivos en el tiempo que lleva la pandemia.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2021 • Mes: 12 | <p>Figura 93</p> <p><i>Cantón con el mayor índice de contagios en la pandemia.</i></p> <p>Nota: Resultados de cantón con el mayor índice de casos positivos, durante la pandemia. Fuente: (Mejía y Aguas, 2021).</p> |

| | |
|--------------------------|---|
| <p>Resultados</p> | <p>Al análisis realizado al cálculo de suma de casos positivos desde el inicio de la pandemia hasta la presente fecha con respecto a cantones, presenta al cantón de Quito con el mayor índice de contagios a nivel nacional, representa el 34.36% de registros confirmados, al realizar una comparativa con el cierre de registro del año 2020 se comprueba que el cantón de Quito se mantiene en tendencia al pasar del 32.74% en diciembre del 2020, al 34.36% en diciembre del 2021, pese al plan de vacunación impuesto desde marzo del año 2021.</p> <p>A nivel social las instituciones educativas presentaban luz verde para el retorno a clases presenciales de manera paulatina, se permite el ingreso al país, al mes de junio las fronteras se mantienen cerradas evitando el tránsito regular de personas. En el mes de julio se retorna al trabajo de manera presencial para el sector público una vez que hayan cumplido con el esquema de vacunación.</p> <p>La acción más acertada a nivel social por el COE nacional fue revocar el permiso de hacer eventos deportivos cuando no se cumplían las medidas de seguridad, prevención y gestión de riesgos vigentes.</p> |
|--------------------------|---|

Nota. Resultados del análisis por cantón que presenta el mayor índice de casos positivos. En el año del 2020, región oriente. Elaborado por: Los autores.

Tabla 8

Comparativa top 10 provincias de registro de casos positivos, año 2020, 2021.

| | |
|--|---|
| <p>Análisis</p> | <p>Comparativa de registro de casos positivos top 10 provincias, año 2020, 2021.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2021 • Mes:12 | <p>Figura 94</p> <p><i>Top provincias con mayor índice de contagio 2021.</i></p>  <p>Nota: Resultados del análisis de las provincias con mayor contagio en el año 2021. Fuente: (Mejía y Aguas, 2021).</p> |

Filtros:

- Año: 2020
- Mes: 12

Figura 95

Top provincias con mayor índice de contagio 2020.



Nota: Resultados del análisis de las provincias con mayor contagio en el año 2020. Fuente: (Mejía y Aguas, 2021).

Resultados

De acuerdo a los resultados expuestos, en las seis primeras posiciones se mantienen las provincias en ambos años, en el año 2020 se puede observar a la provincia de Tungurahua en el séptimo lugar con un total de 7.592 casos positivos, que para el año 2021 desciende una posición siendo superada por la provincia de Imbabura en la cual la suma acumulativa hasta la fecha es de 17.981 casos confirmados y de Tungurahua es 16.322 casos confirmados.

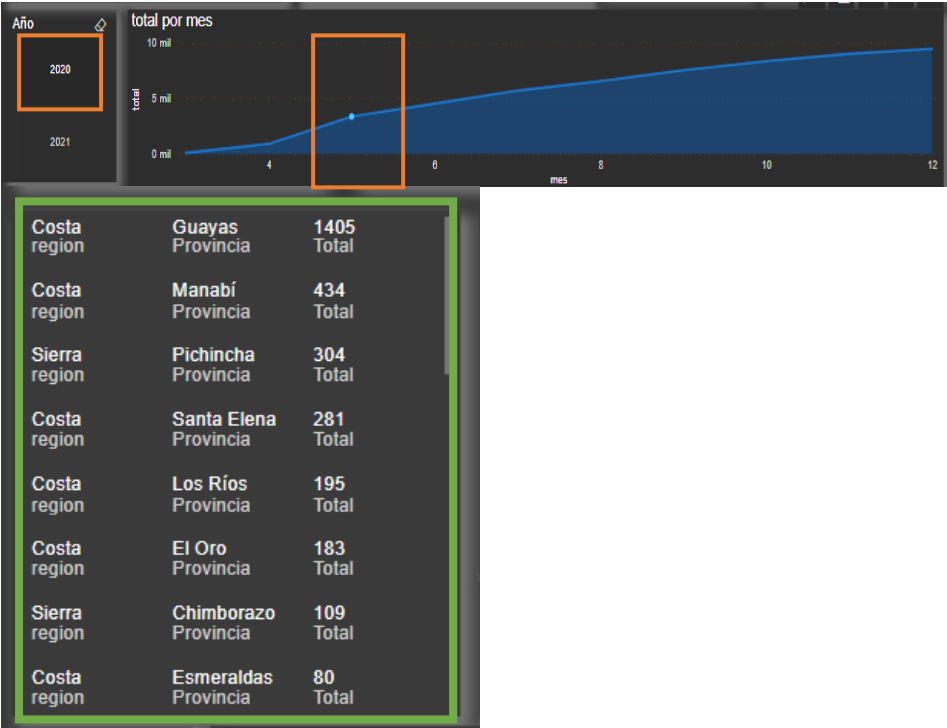
| | |
|--|--|
| | <p>Se puede observar que la provincia de Los Ríos en el año 2020 se muestra por encima de las diez provincias con el mayor índice de contagios, pero para el año 2021, se encuentra ocupando el octavo lugar con un total de 13.499 casos confirmados, desplazando a Cotopaxi de la lista en el año 2021.</p> <p>En temas médicos al mes de septiembre, el hospital general de Latacunga dispuso cerrar la UCI, en caso de presentar un paciente con necesidad de atención de cuidados intensivos las autoridades aseguran que mediante sus mecanismos de derivación y traslados a otros centros de la red pública del servicio la persona sería intervenida con éxito y a tiempo.</p> |
|--|--|

Nota. Resultados del análisis top de provincias con mayor influencia de casos positivos, en el año 2020 y 2021. Elaborado por: Los autores.

4.2. Resultado del eje de análisis de muertes.

Tabla 9

Provincia que presenta el mayor índice de muertes en el primer trimestre de la pandemia.

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|---|--------------|------------------|------------|--------------|------------------|-----------|---------------|---------------------|-----------|--------------|-----------------------|-----------|--------------|--------------------|-----------|--------------|------------------|-----------|---------------|----------------------|-----------|--------------|----------------------|----------|
| <p>Análisis</p> | <p>Provincia que presenta el mayor índice de muertes en el primer trimestre de la pandemia.</p> | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes: 5 | <p>Figura 96</p> <p><i>Análisis del primer trimestre de pandemia.</i></p>  <table border="1" data-bbox="432 1003 911 1541"> <tbody> <tr> <td>Costa region</td> <td>Guayas Provincia</td> <td>1405 Total</td> </tr> <tr> <td>Costa region</td> <td>Manabí Provincia</td> <td>434 Total</td> </tr> <tr> <td>Sierra region</td> <td>Pichincha Provincia</td> <td>304 Total</td> </tr> <tr> <td>Costa region</td> <td>Santa Elena Provincia</td> <td>281 Total</td> </tr> <tr> <td>Costa region</td> <td>Los Ríos Provincia</td> <td>195 Total</td> </tr> <tr> <td>Costa region</td> <td>El Oro Provincia</td> <td>183 Total</td> </tr> <tr> <td>Sierra region</td> <td>Chimborazo Provincia</td> <td>109 Total</td> </tr> <tr> <td>Costa region</td> <td>Esmeraldas Provincia</td> <td>80 Total</td> </tr> </tbody> </table> <p>Nota: Resultados del análisis de las provincias con mayor contagio en el año 2020. Fuente: (Mejía y Aguas, 2021).</p> | Costa region | Guayas Provincia | 1405 Total | Costa region | Manabí Provincia | 434 Total | Sierra region | Pichincha Provincia | 304 Total | Costa region | Santa Elena Provincia | 281 Total | Costa region | Los Ríos Provincia | 195 Total | Costa region | El Oro Provincia | 183 Total | Sierra region | Chimborazo Provincia | 109 Total | Costa region | Esmeraldas Provincia | 80 Total |
| Costa region | Guayas Provincia | 1405 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Manabí Provincia | 434 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Pichincha Provincia | 304 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Santa Elena Provincia | 281 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Los Ríos Provincia | 195 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | El Oro Provincia | 183 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Chimborazo Provincia | 109 Total | | | | | | | | | | | | | | | | | | | | | | | |
| Costa region | Esmeraldas Provincia | 80 Total | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Resultados</p> | <p>Los resultados presentan a Guayas como la provincia con la mayor tasa de mortalidad en la pandemia al mes de mayo, Guayas se presentaba como la provincia con más nivel de contagio a raíz de este alto índice de contagios, las áreas de salud colapsaron por lo que no se permitían recibir en UCI</p> | | | | | | | | | | | | | | | | | | | | | | | | |

nuevos casos confirmados de alto riesgo, esta problemática provocó un descontrol social y pánico provocado a la ciudadanía del resto del país, ya que por medios de difusión de información se indicaba por imágenes a pacientes fallecidos en las casas de residencia propia, personas que fallecían en la espera de ser atendidos.

Nota. Resultados del análisis con el mayor índice de muertes en el primer trimestre de la pandemia. Elaborado por: Los autores.

Tabla 10

Provincia que presenta el mayor índice de muertes en el año 2020.

| | |
|--|---|
| <p>Análisis</p> | <p>La provincia que presenta el mayor índice de muertes en el año 2020.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes:12 | <p>Figura 97</p> <p><i>Análisis de índice de muertes en el año 2020.</i></p> <p>Nota: Resultados del análisis de las provincias con mayor contagio en el año 2020. Fuente: (Mejía y Aguas, 2021).</p> |


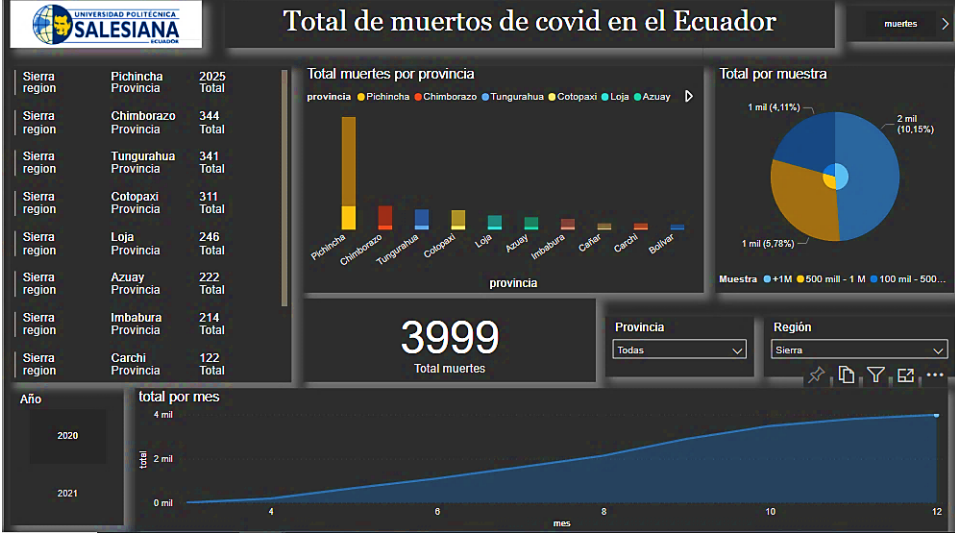
| | |
|--------------------------|--|
| <p>Resultados</p> | <p>El resultado del análisis de la provincia que presente el mayor índice del total registros de muertes en el año 2020, presenta en la gráfica de áreas, a la provincia de Pichincha como la provincia que término el año con el registro más alto de muertes, en el que se indica que el total de muertes para la provincia era de 2.025 registros. Según el informe N.º 067 del MSP el total de fallecidos a nivel nacional era de 9.473, lo que indica que Pichincha representaba el 21.38% de muertes, seguido muy de cerca por la provincia del Guayas que tenía un total de 1.859 lo que representaba el 19.62%, al tener una mayor población es claro que obtenga un mayor índice con relación a muertes o contagios y por ende vacunados respecto al resto de provincias.</p> <p>Al cierre del año 2020, el COE nacional disponía, la prohibición de reuniones por festividades navideñas y de fin de año, esto atrajo un problema social para las personas que laboran por comercio informal, afectando los ingresos que las ventas para estas festividades les podían representar.</p> <p>Otro ámbito social a tomar en cuenta era la campaña política vigente por la presidencia, se solicitó a los líderes de cada movimiento evitar aglomeraciones al realizar campañas en un lugar público, y así evitar un incremento en la sobre demanda de atención hospitalaria, la cual se dificulta atender, así provocando una nueva cifra en las estadísticas de muertes.</p> |
|--------------------------|--|

Nota. Resultados del análisis de la provincia que presenta el mayor índice de muertes en 2020.

Elaborado por: Los autores.

Tabla 11

Evolución de muertes en la región sierra del año 2020 al 2021.

| <p>Análisis</p> | <p>Evolución paulatina de muertes en la región sierra del año 2020 y 2021</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---------------|-----------|-------|---------------|-----------|------|---------------|------------|-----|---------------|------------|-----|---------------|------------|-----|---------------|-------|-----|---------------|-------|-----|---------------|----------|-----|---------------|----------|-----|---------------|--------------|-------------|---------------|---------|----|---------------|--------------|-------------|
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes: 6 | <p>Figura 98 <i>Análisis de muertes región sierra a mes de junio del 2020.</i></p>  <p>The dashboard for June 2020 displays the following data:</p> <table border="1"> <thead> <tr> <th>Sierra region</th> <th>Provincia</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Sierra region</td> <td>Pichincha</td> <td>501</td> </tr> <tr> <td>Sierra region</td> <td>Chimborazo</td> <td>151</td> </tr> <tr> <td>Sierra region</td> <td>Cotopaxi</td> <td>120</td> </tr> <tr> <td>Sierra region</td> <td>Tungurahua</td> <td>107</td> </tr> <tr> <td>Sierra region</td> <td>Azuay</td> <td>50</td> </tr> <tr> <td>Sierra region</td> <td>Cañar</td> <td>45</td> </tr> <tr> <td>Sierra region</td> <td>Loja</td> <td>45</td> </tr> <tr> <td>Sierra region</td> <td>Imbabura</td> <td>39</td> </tr> <tr> <td>Sierra region</td> <td>Carchi</td> <td>12</td> </tr> <tr> <td>Sierra region</td> <td>Bolivar</td> <td>12</td> </tr> <tr> <td>Sierra region</td> <td>Total</td> <td>1109</td> </tr> </tbody> </table> <p>The monthly trend line shows a steady increase in deaths from March to June 2020.</p> | Sierra region | Provincia | Total | Sierra region | Pichincha | 501 | Sierra region | Chimborazo | 151 | Sierra region | Cotopaxi | 120 | Sierra region | Tungurahua | 107 | Sierra region | Azuay | 50 | Sierra region | Cañar | 45 | Sierra region | Loja | 45 | Sierra region | Imbabura | 39 | Sierra region | Carchi | 12 | Sierra region | Bolivar | 12 | Sierra region | Total | 1109 |
| Sierra region | Provincia | Total | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Pichincha | 501 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Chimborazo | 151 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Cotopaxi | 120 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Tungurahua | 107 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Azuay | 50 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Cañar | 45 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Loja | 45 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Imbabura | 39 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Carchi | 12 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Bolivar | 12 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Total | 1109 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Año: 2020 • Mes: 12 | <p>Figura 99 <i>Análisis de muertes región sierra a mes de diciembre del 2020.</i></p>  <p>The dashboard for December 2020 displays the following data:</p> <table border="1"> <thead> <tr> <th>Sierra region</th> <th>Provincia</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Sierra region</td> <td>Pichincha</td> <td>2025</td> </tr> <tr> <td>Sierra region</td> <td>Chimborazo</td> <td>344</td> </tr> <tr> <td>Sierra region</td> <td>Tungurahua</td> <td>341</td> </tr> <tr> <td>Sierra region</td> <td>Cotopaxi</td> <td>311</td> </tr> <tr> <td>Sierra region</td> <td>Loja</td> <td>246</td> </tr> <tr> <td>Sierra region</td> <td>Azuay</td> <td>222</td> </tr> <tr> <td>Sierra region</td> <td>Imbabura</td> <td>214</td> </tr> <tr> <td>Sierra region</td> <td>Carchi</td> <td>122</td> </tr> <tr> <td>Sierra region</td> <td>Total</td> <td>3999</td> </tr> </tbody> </table> <p>The monthly trend line shows a significant increase in deaths throughout 2020, reaching a peak in December.</p> | Sierra region | Provincia | Total | Sierra region | Pichincha | 2025 | Sierra region | Chimborazo | 344 | Sierra region | Tungurahua | 341 | Sierra region | Cotopaxi | 311 | Sierra region | Loja | 246 | Sierra region | Azuay | 222 | Sierra region | Imbabura | 214 | Sierra region | Carchi | 122 | Sierra region | Total | 3999 | | | | | | |
| Sierra region | Provincia | Total | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Pichincha | 2025 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Chimborazo | 344 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Tungurahua | 341 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Cotopaxi | 311 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Loja | 246 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Azuay | 222 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Imbabura | 214 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Carchi | 122 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sierra region | Total | 3999 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Nota: Análisis de la evolución del índice de muertes en la región sierra, período marzo - junio 2020. Fuente: (Mejía y Aguas, 2021).

Nota: Análisis de la evolución del índice de muertes en la región sierra, período marzo - diciembre 2020. Fuente: (Mejía y Aguas, 2021).

Filtros:

- Año: 2021
- Mes: 7

Figura 100

Análisis de muertes región sierra a mes de julio del 2021.



Nota: Análisis de la evolución del índice de muertes en la región sierra, período marzo 2020 - julio 2021. Fuente: (Mejía y Aguas, 2021).

Resultados

Al realizar un análisis por agrupamiento de ciertos meses y año, haciendo hincapié en el índice de total de cada consulta, cabe aclarar que es un total acumulativo de fecha a fecha, por tal motivo haciendo uso de la operación básica de resta se puede observar que en el período de marzo a junio del 2020 el total de muertes es de 1.109 que representa un 6.93% del total de casos positivos en la misma región en el mismo período de tiempo; para el período de julio a diciembre del 2020 se obtiene el total de 2.890, lo que representa un incremento del 60% con relación al período anterior, y en el caso del análisis del período de enero a julio del 2021 se obtiene un total de 10.263 muertes que representa el 3.47% del total de casos confirmados para esta fecha.

| | |
|--|---|
| | <p>A nivel médico se espera un mayor índice de muertes al iniciar la pandemia por el motivo de que las diferentes casas de salud fueron ocupadas al máximo y estas no podían brindar la atención necesaria a todos los pacientes, sin embargo, después de obtener los resultados de estos análisis, se puede concluir que a raíz de la falta de educación de ciertos grupos de población que no respetaron las diferentes normas aplicadas a festividades en el mes de diciembre del 2020 y tomando en cuenta el tiempo que toma el virus en presentar síntomas en un paciente infectado, se indica que las muertes registradas en este año son parte de ese grupo de población, aparte de un mayor nivel de libertinaje social que la población tomo al iniciar el nuevo año, y a raíz de la vacunación la gente tomo tranquilidad total a las medidas de seguridad.</p> |
|--|---|

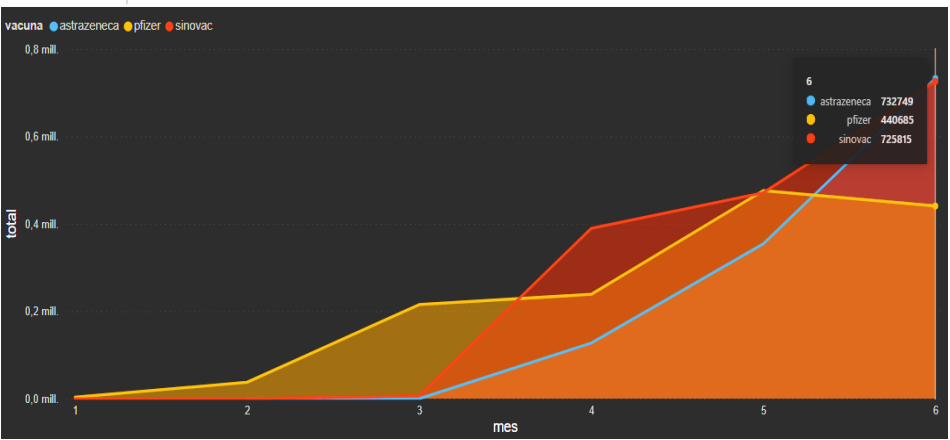
Nota. Resultados del análisis de la provincia que presenta el mayor índice de muertes en 2020.

Elaborado por: Los autores.

4.3. Resultado del eje de análisis de vacunados.

Tabla 12

Vacuna con el mayor índice de uso.

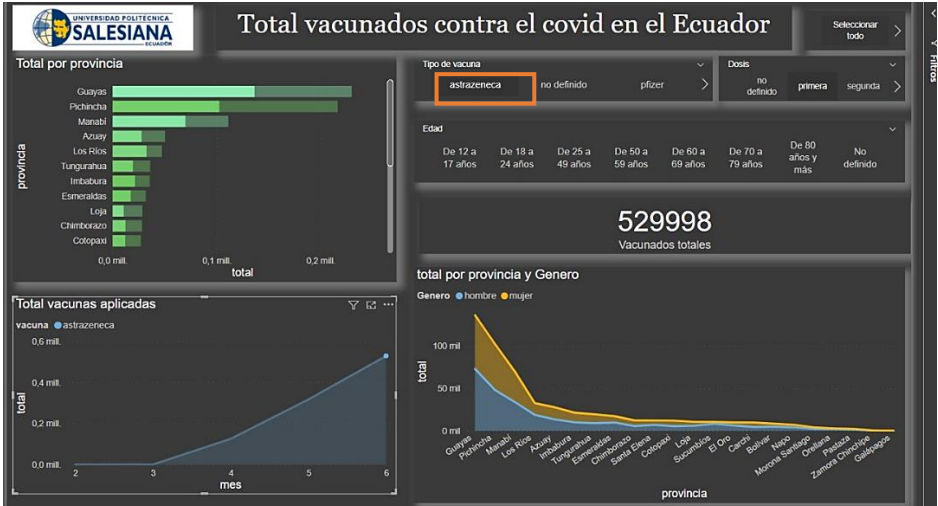
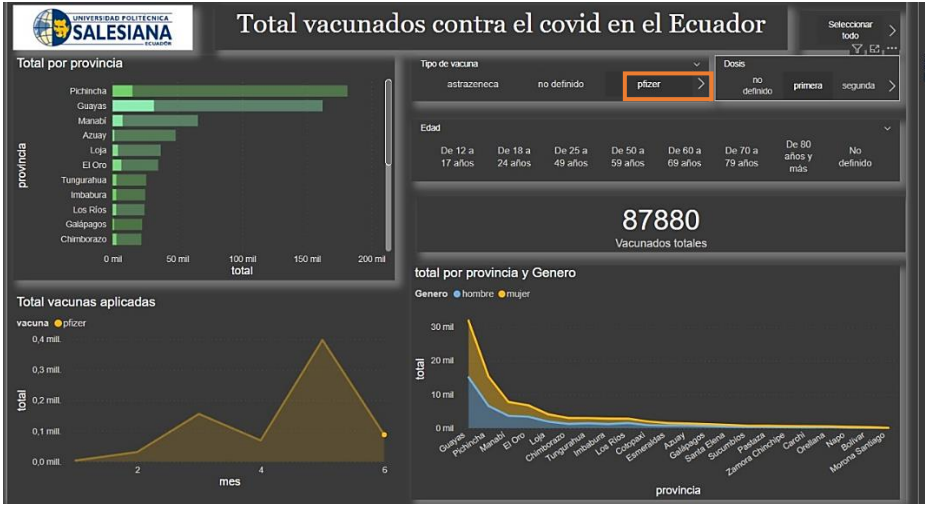
| | |
|--|--|
| <p>Análisis</p> | <p>Vacuna con el mayor índice de utilización.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Ninguno | <p>Figura 101</p> <p>Gráfica de áreas del índice de uso de cada vacuna.</p>  <p>Nota: Gráfica de áreas del progreso de vacunación para cada tipo de vacuna. Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Resultados</p> | <p>La gráfica indica que hasta el mes de marzo Pfizer lidera el proceso de vacunación, debido a que el proceso de vacunación en el Ecuador inicio con las personas de la tercera edad, y la vacuna recomendada y comprobada con mayor éxito para las personas de la tercera edad era Pfizer.</p> <p>A partir del mes de abril la tendencia para Pfizer se mantiene, Sinovac se presenta por encima de Pfizer con un total de 389.619 vacunados, para el mes de abril el proceso de vacunación se extendió para empleados del</p> |

| | |
|--|---|
| | <p>sector público e incremento el rango de edad para que la población se acerque a los distintos centros de vacunación.</p> <p>AstraZeneca era utilizada por un grupo bajo de población, siendo en el mes de mayo que se posicionó cerca del total de vacunados para Pfizer y Sinovac, en mayo se da comienzo a las campañas del sector privado ofertando a empresas privadas la vacuna para sus empleados, esto atrae a la continua vacunación de personas de distintos rangos de edad.</p> <p>Por la efectividad verificada Sinovac se ubica como la vacuna con el mayor índice de vacunados cerrando el mes con un total de 725.815 vacunados.</p> <p>Para el cierre del mes de mayo AstraZeneca es empleada con mayor frecuencia presentando un total de 732.749 vacunados muy por encima de Pfizer que cerraba el mes con un total de 440.685 vacunados.</p> |
|--|---|

Nota. Resultados del análisis realizado al índice total de vacunados por cada tipo de vacuna hasta el mes de junio. Elaborado por: Los autores.

Tabla 13

Índices totales de cada tipo de vacuna aplicada como primera dosis.

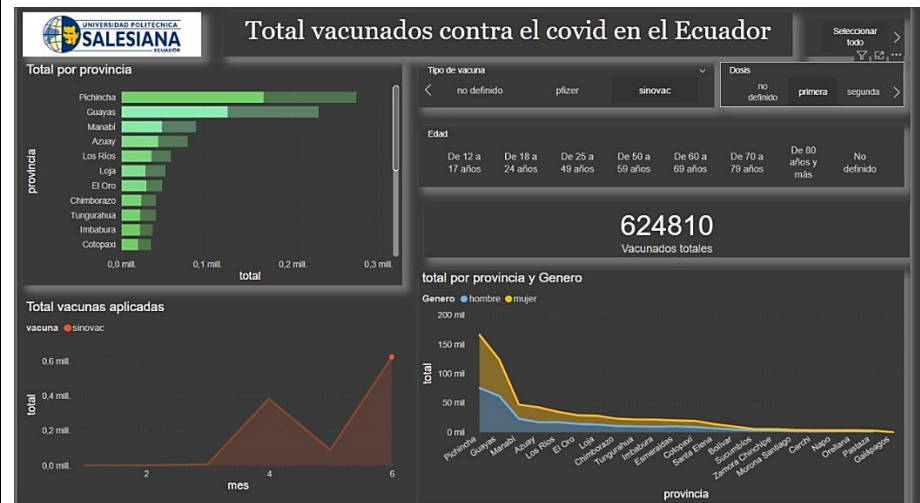
| | |
|--|---|
| <p>Análisis</p> | <p>Índices totales de cada tipo de vacuna aplicada como primera dosis.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Dosis: Primera • Tipo: AstraZeneca | <p>Figura 102 <i>Análisis de AstraZeneca, primera dosis.</i></p>  <p>Nota: Análisis de AstraZeneca del índice total de vacunados con primera dosis. Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Dosis: Primera • Tipo: Pfizer | <p>Figura 103 <i>Análisis de Pfizer, primera dosis.</i></p>  <p>Nota: Análisis de Pfizer del índice total de vacunados con primera dosis. Fuente: (Mejía y Aguas, 2021).</p> |

Filtros:

- Dosis:
Primera
- Tipo:
Sinovac

Figura 104

Análisis de Sinovac, primera dosis.



Nota: Análisis de Pfizer del índice total de vacunados con primera dosis.
Fuente: (Mejía y Aguas, 2021).

Resultados

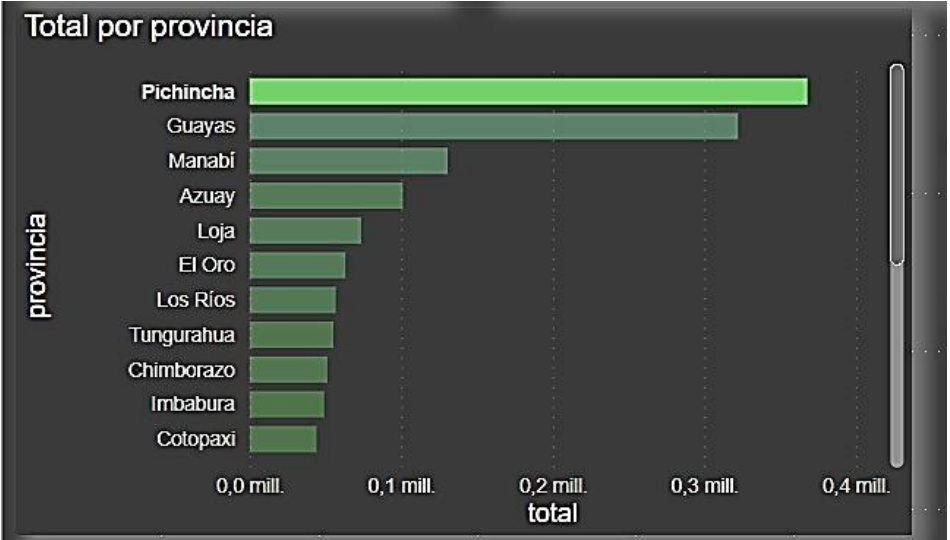
El total de registro de vacunados para el mes de junio es de 1.242.688 personas, el desglose por cada tipo de vacuna, indica a Sinovac como la vacuna que presenta el mayor índice de vacunados con un total de 624.000 personas, y Pfizer hasta la fecha de junio en el tercer lugar con un total de 87.880 vacunados.

En cuanto a provincia se observa que AstraZeneca tuvo una mayor acogida en Guayas con un total de 229.912 personas vacunadas, caso contrario con la acogida AstraZeneca y Pfizer que su más alto índice de vacunados con primera dosis se encuentra en la provincia de Pichincha

*Nota. Resultados del análisis de los índices del tipo de vacuna aplicada en la primera dosis.
Elaborado por: Los autores.*

Tabla 14


Provincia con la mayor tasa de vacunación.

| Análisis | Provincia con la mayor tasa de vacunación. | | | | | | | | | | | | | | | | | | | | | | | | |
|--|---|-----------|-------|-----------|-------------|--------|-------------|--------|-------------|-------|-------------|------|-------------|--------|-------------|----------|-------------|------------|-------------|------------|-------------|----------|-------------|----------|-------------|
| Filtros: <ul style="list-style-type: none">• Año: 2021• Mes: 6 | <p>Figura 105</p> <p><i>Análisis de la provincia con mayor índice de vacunados.</i></p>  <table border="1"><caption>Total por provincia</caption><thead><tr><th>provincia</th><th>total</th></tr></thead><tbody><tr><td>Pichincha</td><td>0,392 mill.</td></tr><tr><td>Guayas</td><td>0,320 mill.</td></tr><tr><td>Manabí</td><td>0,130 mill.</td></tr><tr><td>Azuay</td><td>0,110 mill.</td></tr><tr><td>Loja</td><td>0,080 mill.</td></tr><tr><td>El Oro</td><td>0,070 mill.</td></tr><tr><td>Los Ríos</td><td>0,060 mill.</td></tr><tr><td>Tungurahua</td><td>0,050 mill.</td></tr><tr><td>Chimborazo</td><td>0,050 mill.</td></tr><tr><td>Imbabura</td><td>0,040 mill.</td></tr><tr><td>Cotopaxi</td><td>0,030 mill.</td></tr></tbody></table> <p>Nota: Análisis de la provincia que se encuentra con el mayor índice de vacunados hasta la fecha de mayo 2021. Fuente: (Mejía y Aguas, 2021).</p> | provincia | total | Pichincha | 0,392 mill. | Guayas | 0,320 mill. | Manabí | 0,130 mill. | Azuay | 0,110 mill. | Loja | 0,080 mill. | El Oro | 0,070 mill. | Los Ríos | 0,060 mill. | Tungurahua | 0,050 mill. | Chimborazo | 0,050 mill. | Imbabura | 0,040 mill. | Cotopaxi | 0,030 mill. |
| provincia | total | | | | | | | | | | | | | | | | | | | | | | | | |
| Pichincha | 0,392 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Guayas | 0,320 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Manabí | 0,130 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Azuay | 0,110 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Loja | 0,080 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| El Oro | 0,070 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Los Ríos | 0,060 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Tungurahua | 0,050 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Chimborazo | 0,050 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Imbabura | 0,040 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Cotopaxi | 0,030 mill. | | | | | | | | | | | | | | | | | | | | | | | | |
| Resultados | El resultado arrojado por el análisis realizado a nivel de provincia, indica que la provincia de Pichincha presenta el mayor índice de vacunados hasta la fecha del mes de mayo del 2021, presentando un total de 392.000 vacunados. | | | | | | | | | | | | | | | | | | | | | | | | |

Nota. Resultados del análisis por provincia con mayor índice de vacunación. Elaborado por: Los autores.

Tabla 15

Análisis de vacuna Sinovac aplicada.

| | |
|---|--|
| <p>Análisis</p> | <p>Análisis de vacuna Sinovac aplicada.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Tipo: Sinovac | <p>Figura 106</p> <p><i>Análisis de la vacuna Sinovac.</i></p>  <p>Nota: Análisis de la vacuna Sinovac. Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Resultados</p> | <p>La gráfica de resultados indica un total de 725.815 personas vacunadas, siendo Sinovac la vacuna con la que se ha efectuado el mayor número de inmunizaciones ante el virus, la gráfica de áreas señala que a partir del mes de marzo Sinovac comienza a reportar un incremento de vacunación a nivel nacional completando un total de 389.619 vacunados al mes de abril, para el mes de junio es notable el crecimiento exponencial de inmunización bajo Sinovac realizando 725.815 inmunizaciones. La provincia con el mayor índice de personas vacunadas con Sinovac es Pichincha con un total de 389.619 vacunados.</p> |

Nota. Resultados del análisis de Sinovac. Elaborado por: Los autores.

Tabla 16

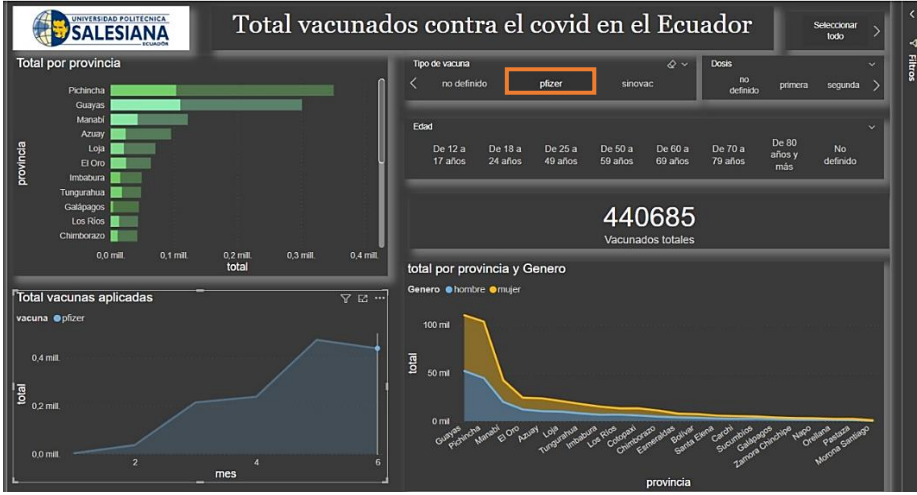
Análisis de vacuna AstraZeneca aplicada.

| | |
|---|---|
| <p>Análisis</p> | <p>Análisis de vacuna AstraZeneca aplicada en el Ecuador.</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Tipo: AstraZeneca • Rango: 25 a 45 años | <p>Figura 107</p> <p><i>Análisis de vacuna AstraZeneca.</i></p>  <p>Nota: Análisis de la vacuna AstraZeneca hasta la fecha de mayo 2021.</p> <p>Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Resultados</p> | <p>Los resultados obtenidos para este tipo de vacuna a la fecha de mayo del 2021, indica que el total de vacunados es de 732.749 personas, según la gráfica de áreas el género “Mujer” una mayor incidencia al momento de vacunar, y la provincia con mayor vacunación con este tipo de vacuna es Pichincha con un total de 73.910 vacunados.</p> |

Nota. Resultados del análisis de AstraZeneca. Elaborado por: Los autores.

Tabla 17

Análisis de la vacuna Pfizer.

| | |
|---|---|
| <p>Análisis</p> | <p>Análisis de la aplicación de vacuna Pfizer</p> |
| <p>Filtros:</p> <ul style="list-style-type: none"> • Tipo: Pfizer | <p>Figura 108 <i>Análisis de la vacuna Pfizer.</i></p>  <p>Nota: Análisis de la vacuna Pfizer hasta la fecha de mayo 2021. Fuente: (Mejía y Aguas, 2021).</p> |
| <p>Resultados</p> | <p>Los resultados que se observan en la Figura 105, del análisis de la vacuna Pfizer, muestra en al cierre del mes de junio un total de 440.685 vacunados, esto debido a que para este mes la vacuna de Sinovac y AstraZeneca fueron elegidas para aplicar a la población que no formaba parte del grupo de tercera edad, aun así, el índice del total de dosis aplicadas indica que se tiene 1.242.688 personas vacunadas. La gráfica de barras posiciona a Pichincha y Guayas en los primeros lugares, indicando que son las provincias con los más altos índices de vacunados con Pfizer, entre las dos provincias representan el 35.46% de población vacunada con</p> |

| | |
|--|---|
| | Pfizer. El 25 de mayo arriban 170.820 dosis lo cual genera una tranquilidad social por parte de la población, ya que genera esperanza de continuar con la fluidez de vacunación a la población. |
|--|---|

Nota. Resultados del análisis de Pfizer. Elaborado por: Los autores

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- El uso de la herramienta de Power Bi facilita el análisis de datos para la obtención y presentación de resultados, ofreciendo una interfaz amigable para el usuario final y que pueda identificar patrones en los resultados que contribuyan la toma de decisiones, además el manejo de datos es sencillo y fácil de actualizar, esto permita obtener datos oportunos y confiables.
- Los resultados presentados en el informe anteriormente, demuestran que la provincia de Pichincha cuenta con los índices más altos de contagios con un total del 35%, hasta el mes de diciembre de 2021 y muertes por COVID-19 con un total del 22 % hasta el mes de julio de 2021. Estos resultados están ligados a diferentes factores externos en la que influye las acciones del gobierno central de Pichincha frente a la emergencia sanitaria, o la responsabilidad de la población para ayudar a contener el virus, puede recaer que el análisis representa estos valores, se debe considerar que la población de Pichincha es una de las más grandes del Ecuador y eso puede influir directamente en los resultados.
- Una vez finalizado el proyecto se concluye que se cumple con los objetivos planteados al desarrollar la herramienta de inteligencia de negocios, aplicando metodologías que cubren las necesidades del proyecto obteniendo como producto final una herramienta intuitiva y fácil de manejar que permite analizar los resultados obtenidos y permitir la colaboración en el análisis de la toma de decisiones.
- El desarrollo de una Data Warehouse con información en constante actualización y que a su vez permita implementar herramientas para el BI, abre un mundo de nuevas posibilidades para las organizaciones de hoy en día, estas dependen de una realidad

donde el análisis de sus datos es un factor clave para su supervivencia y éxito debido a que estos datos se convierten en información sumamente importante para tomar decisiones gerenciales basadas en datos de sus sistemas transaccionales logrando así mejorar su estrategia y poder competir en el mercado.

- Escoger una metodología que se adapte a las necesidades a satisfacer es la base para el desarrollo de un almacén de datos aplicando buenas prácticas. La metodología Hefesto ofrece una guía intuitiva y fácil de comprender en cada una de sus etapas para que el proceso de construcción de la Data Warehouse se realice de una manera correcta.

5.2. RECOMENDACIONES

- Se aconseja que los organismos involucrados en llevar a cabo el recuento de los diferentes ejes de análisis como total de contagios, muertes, vacunados. Frente a la emergencia sanitaria, implementen un sistema web donde se alojen información oportuna, actualizada, veraz. Ya que una vez realizado su análisis permite plantear e implementar medidas eficaces que ofrezcan un accionar correcto a las diferentes autoridades de cada parte del Ecuador dependiendo de su situación.
- La precisión y granularidad de los datos también es un factor muy importante, por el hecho de que permite tomar decisiones con base a datos mucho más precisos y acertados, logrando así poder identificar con mayor facilidad las zonas en donde se deben adoptar medidas de prevención que contribuyan con la disminución de contagiados y afectando directamente y reduciendo el número de muertes por COVID-19.
- La integridad de los datos es altamente relevante, ya que es la base de todo sistema de BI, saber realizar el proceso de limpieza y procesamiento de datos es una de las características más considerables para desarrollar una herramienta de BI logrando de esta manera mostrar resultados reales y sin errores por datos mal procesados.

- Los lenguajes de programación han ido tomando protagonismo en el campo de análisis de datos y Python hoy en día es uno de los pioneros en este tema, ya que ofrece una gran variedad de características y con un gran soporte por parte de la comunidad que ayudan a optimizar el tiempo de trabajo en el desarrollo de la Data Warehouse.

Bibliografía

- Subra, J. P., & Vannieuwenhuysse, A. (2018). Scrum: un método ágil para sus proyectos. Ediciones ENI.
- Garcete Rodríguez, A. D., Benítez, R., Pinto Roa, D. P., & Vazquez, A. (2017). Técnica de pronóstico de la demanda basada en Business Intelligence y Machine Learning. In Simposio Argentino sobre Tecnología y Sociedad (STS)-JAIIO 46 (Córdoba, 2017).
- Power, B. I., Excel, U., Desktop, P. B., & Tiles, P. (2021). Microsoft power bi. Available here: <https://powerbi.microsoft.com/en-us>.
- Bernabeu, R. D. (2010). Hefesto–Data warehousing: Investigación y sistematización de conceptos-Hefesto: Metodología para la construcción de un data warehouse. Córdoba, Argentina. Córdoba, Argentina.
- Bustamante Granda, W. X., Macas Ruiz, E. M., & Cevallos Macas, F. B. (2018). Data Warehouse: Análisis Multidimensional de BAFICI utilizando Power Pivot. *Revista Espacios*, 39(34), 18.
- Ortí Rodríguez, J. (2021). Diseño, implementación y explotación de un almacén de datos (Doctoral dissertation, Universitat Politècnica de València).
- Silva Peñafiel, G. E. (2018). Análisis de metodologías para la implementación de un data warehouse aplicado a la toma de decisiones del Instituto Nacional de Patrimonio Cultural Regional 3 (Master's thesis, Pontificia Universidad Católica del Ecuador).
- Peñafiel, G. E. S., Yáñez, V. M. Z., Guamán, K. P. M., & Padilla, L. M. T. (2019). Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones. *Ciencia Digital*, 3(3.4.), 397-418.
- Manso Millán, J. A. (2021). Retos de la Migración de Datos hacia un Data Warehouse.
- Hernández Mejía, S. (2019). Comparativo de metodologías y herramientas para el desarrollo de un data warehouse.
- Ordóñez, M. P. Z., Ríos, J. R. M., & Castillo, F. F. R. (2017). Administración de Bases de datos con PostgreSQL (Vol. 19). 3Ciencias.
- Lachev, T. (2005). Applied Microsoft analysis services 2005 and Microsoft Business Intelligence platform: a guide to the leading OLAP platform. Prologika Press.
- García Monsálvez, J. C. (2017). Python como primer lenguaje de programación textual en la Enseñanza Secundaria= Python as First Textual Programming Language in Secondary Education. Python como primer lenguaje de programación textual en la Enseñanza Secundaria= Python as First Textual Programming Language in Secondary Education, 147-162.
- Trujillo, J. C., Mazón, J. N., & Pardillo, J. (2013). *DISEÑO Y EXPLOTACIÓN DE ALMACENES DE DATOS*. Alicante: Editorial Club Universitario.
- Curto Díaz, J. (2016). ¿ Cómo crear un data warehouse?. Editorial UOC.