

**DATA MART PARA EL ALMACENAMIENTO Y MINERÍA DE DATOS DE  
CASOS CONFIRMADOS DE COVID-19 EN LA CIUDAD DE BOGOTÁ**

**MILTON FERNANDO QUINTERO ABRIL  
PEDRO AGUIRRE ARIAS**

**UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
ESPECIALIZACIÓN EN BASES DE DATOS  
TUNJA  
2021**

**DATA MART PARA EL ALMACENAMIENTO Y MINERÍA DE DATOS DE  
CASOS CONFIRMADOS DE COVID-19 EN LA CIUDAD DE BOGOTÁ**

**MILTON FERNANDO QUINTERO ABRIL  
PEDRO AGUIRRE ARIAS**

**TRABAJO DIRIGIDO  
INFORME FINAL**

**INGENIERO:  
JAVIER ANTONIO BALLESTEROS RICAURTE**

**UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
ESPECIALIZACIÓN EN BASES DE DATOS  
TUNJA  
2021**

## TABLA DE CONTENIDO

<b>1. TABLA DE ILUSTRACIONES</b>	<b>6</b>
<b>2. INTRODUCCIÓN</b>	<b>8</b>
<b>3. PROBLEMÁTICA</b>	<b>9</b>
<b>4. OBJETIVOS</b>	<b>10</b>
1.1 OBJETIVO GENERAL	10
2.1 OBJETIVOS ESPECÍFICOS	10
<b>5. ALCANCE</b>	<b>11</b>
<b>6. MARCO CONCEPTUAL</b>	<b>12</b>
3.1 ETL	12
4.1 OLAP	12
5.1 VISUAL STUDIO MICROSOFT	12
6.1 SQL SERVER DEVELOPER	12
7.1 SERVER MANAGEMENT STUDIO	12
8.1 INTEGRATION SERVICES	13
9.1 ANALYSIS SERVICE	13
10.1 Power BI	13
<b>7. METODOLOGÍA</b>	<b>14</b>
11.1 PLAN DE TRABAJO SEGÚN OBJETIVOS PLANTEADOS	14
<b>8. Metadatos Número de casos confirmados por el laboratorio de COVID- 19</b>	<b>16</b>
<b>9. RECOLECCIÓN</b>	<b>17</b>
1.1.1 Diagrama de bus	17
<b>10. PRODUCTOS</b>	<b>18</b>
12.1 MODELO DE DATOS.	18
13.1 DIMENSIONES:	18
14.1 HECHOS:	18
15.1 ORIGEN DE DATOS	19
16.1 FLUJO DE TAREAS ETL	21
17.1 TAREA DE CARGA DE DISENSIÓN TIEMPO	22
18.1 TAREA CARGA DE TABLA DE HECHOS	23

<b>19.1</b>	<b>DESTINO DE CARGUE DE DATOS</b>	<b>23</b>
<b>20.1</b>	<b>Tablas de hechos:</b>	<b>23</b>
<b>21.1</b>	<b>Tablas de dimensiones:</b>	<b>23</b>
<b>22.1</b>	<b>DESARROLLO DEL CUBO</b>	<b>24</b>
2.1.1	Implementación del cubo	25
3.1.1	Resultados del análisis.	26
<b>11.</b>	<b>ANÁLISIS DE NECESIDADES</b>	<b>28</b>
<b>23.1</b>	<b>Cantidad de contagiados en: día de la semana, semana, mes, trimestre y año</b>	<b>28</b>
4.1.1	Día de la semana	28
<b>24.1</b>	<b>Semana del año 2020</b>	<b>28</b>
5.1.1	Semana del año 2021	29
6.1.1	Mes del año 2020	30
7.1.1	Mes del año 2021	31
8.1.1	Trimestre del año	32
9.1.1	Año	33
<b>25.1</b>	<b>Cantidad de hombres y mujeres contagiados en: día de la semana, semana, mes, trimestre y año</b>	<b>35</b>
10.1.1	Día de la semana año 2020	35
11.1.1	Día de la semana año 2021	35
12.1.1	Semana del año 2020	36
13.1.1	Semana del año 2021	36
14.1.1	Mes del año 2020	37
15.1.1	Mes del año 2021	37
16.1.1	Trimestre año 2020	38
17.1.1	Trimestre año 2021	39
18.1.1	Año 2020	39
19.1.1	Año 2021	41
<b>26.1</b>	<b>Cantidad de casos confirmados por localidad teniendo en cuenta: mes, trimestre y año.</b>	<b>42</b>
20.1.1	Día de la semana año 2020	43
21.1.1	Día de la semana año 2021	43
22.1.1	Semana del año 2020	44
23.1.1	Semana del año 2021	45
24.1.1	Mes del año 2020	45
25.1.1	Mes del año 2021	46
26.1.1	Trimestre año 2020	47
27.1.1	Trimestre año 2021	47
28.1.1	Año 2020	48
29.1.1	Año 2021	49
<b>27.1</b>	<b>cantidad de personas fallecidas: día de la semana, semana, mes, trimestre, año.</b>	<b>49</b>
30.1.1	Día de la semana año 2020	49
31.1.1	Día de la semana año 2021	50
32.1.1	Semana del año 2020	50
33.1.1	Semana del año 2021	51
34.1.1	Mes del año 2020	52
35.1.1	Mes del año 2021	52
36.1.1	Trimestre año 2020	53

37.1.1	Trimestre año 2021	53
38.1.1	Año	54
<b>28.1</b>	<b>Cantidad de casos recuperados: día de la semana, semana, mes, trimestre, año.</b>	<b>55</b>
39.1.1	Día de la semana año 2020	55
40.1.1	Día de la semana año 2021	56
41.1.1	Semana del año 2020	57
42.1.1	Semana del año 2021	57
43.1.1	Mes del año 2020	58
44.1.1	Mes del año 2021	59
45.1.1	Trimestre año 2020	59
46.1.1	Trimestre año 2021	60
47.1.1	Año	61
<b>29.1</b>	<b>Contagios por edad.</b>	<b>62</b>
<b>12.</b>	<b>CONCLUSIONES</b>	<b>64</b>
<b>13.</b>	<b>10.PROBLEMAS ENCONTRADOS DURANTE EL ESTUDIO</b>	<b>65</b>
<b>14.</b>	<b>BIBLIOGRAFÍA</b>	<b>66</b>

## 1. TABLA DE ILUSTRACIONES

Imagen 1. Matriz de bus	13
Imagen 2. Modelo en estrella	15
Imagen 3. Origen de datos	16
Imagen 4. Origen de datos rangos edad	17
Imagen 5. Flujo de tareas ETL	18
Imagen 6. Tarea carga de datos dimensión tiempo	18
Imagen 7. Carga de datos tabla de hechos	19
Imagen 8. Destino carga de datos	20
Imagen 9. Modelo cubo OLAP	20
Imagen 10. Jerarquía dimensión fechas A	21
Imagen 11. Jerarquía dimensión de fechas B	21
Imagen 12. Implementación del cubo OLAP	22
Imagen 13. Cubo en Analysis Services	23
Imagen 14. Cantidad de contagios por día de la semana año 2020 y 2021	24
Imagen 15. Cantidad de contagios por semana año 2020	25
Imagen 16. Cantidad de contagios por semana año 2021	26
Imagen 17. Cantidad de contagios por mes del año 2020	27
Imagen 18. Cantidad de contagios por mes del año 2021	28
Imagen 19. Cantidad de contagios por trimestre del año 2020 y 2021	29
Imagen 20. Cantidad de contagios por año 2020 y 2021	30
Imagen 21. Imagen 20. Cantidad de contagios por año 2020 y 2021 barras	30
Imagen 22. Cantidad de hombres y mujeres contagiados día de la semana año 2020	31
Imagen 23. Cantidad de hombres y mujeres contagiados día de la semana año 2021	32
Imagen 24. Cantidad de hombres y mujeres contagiados POR SEMANA año 2020	32
Imagen 25. Cantidad de hombres y mujeres contagiados por semana año 2021	33
Imagen 26. Cantidad de hombres y mujeres contagiados POR MES año 2020	33
Imagen 27. Cantidad de hombres y mujeres contagiados POR MES año 2021	34
Imagen 28. Cantidad de hombres y mujeres contagiados por trimestre año 2020	34
Imagen 29. Cantidad de hombres y mujeres contagiados por trimestre año 2021	35
Imagen 30. Cantidad de hombres y mujeres contagiados por año 2020 A	36
Imagen 31. Cantidad de hombres y mujeres contagiados por año 2020 B	36
Imagen 32. Cantidad de hombres y mujeres contagiados por año 2020 C	37
Imagen 33. Cantidad de hombres y mujeres contagiados por año 2020 D	37
Imagen 34. Cantidad de hombres y mujeres contagiados por año 2020 F	38
Imagen 35. Cantidad de hombres y mujeres contagiados por año 2021 A	38
Imagen 36. Cantidad de hombres y mujeres contagiados por año 2021 B	39
Imagen 37. Cantidad de casos confirmados por localidad y día de la semana año 2020	40
Imagen 38. Cantidad de casos confirmados por localidad y día de la semana año 2021	40
Imagen 39. Cantidad de casos confirmados por LOCALIDAD Y semana año 2020	41

Imagen 40. Cantidad de casos confirmados por LOCALIDAD Y semana año 2021	42
Imagen 41. Cantidad de casos confirmados por LOCALIDAD Y mes año 2020	43
Imagen 42. Cantidad de casos confirmados por localidad y mes año 2021	44
Imagen 43. Cantidad de casos confirmados por LOCALIDAD Y trimestre año 2020	45
Imagen 44. Cantidad de casos confirmados por localidad y trimestre año 2021	46
Imagen 45. Cantidad de casos confirmados por localidad año 2020	47
Imagen 46. Cantidad de casos confirmados por localidad año 2021	48
Imagen 47. Cantidad de personas fallecidas por día de la semana año 2020	49
Imagen 48. Cantidad de personas fallecidas por día de la semana año 2021	50
Imagen 49. Cantidad de personas fallecidas por semana año 2020	50
Imagen 50. Cantidad de personas fallecidas por semana año 2021	51
Imagen 51. Cantidad de personas fallecidas por mes año 2020	52
Imagen 52. Cantidad de personas fallecidas por mes año 2021	53
Imagen 53. Cantidad de personas fallecidas por trimestre año 2020	53
Imagen 54. Cantidad de personas fallecidas por trimestre año 2021	54
Imagen 55. Cantidad de personas fallecidas por año 2020 y 2021 A	55
Imagen 56. Cantidad de personas fallecidas por año 2020 y 2021 B	55
Imagen 57. Cantidad de casos recuperados por día de la semana año 2020	56
Imagen 58. Cantidad de casos recuperados por día de la semana año 2021	57
Imagen 59. Cantidad de casos recuperados por semana año 2020	57
Imagen 60. Cantidad de casos recuperados por semana año 2021	58
Imagen 61. Cantidad de casos recuperados por mes año 2020	59
Imagen 62. Cantidad de casos recuperados por mes año 2021	60
Imagen 63. Cantidad de casos recuperados por trimestre año 2020	61
Imagen 64. Cantidad de casos recuperados por trimestre año 2021	62
Imagen 65. Cantidad de casos recuperados por año 2020 y 2021 A	63
Imagen 66. Cantidad de casos recuperados por año 2020 y 2021 B	64
Imagen 67. Cantidad de contagios por edad año 2020 y 2021	65

## 2. INTRODUCCIÓN

El observatorio de salud de Bogotá (Salud Data) diariamente recopila, depura, sigue los casos confirmados de Covid19 y publica dicha información al público. Maneja información relacionada con fecha de síntomas, fecha de diagnóstico, ciudad, localidad, edad, sexo, fuente o tipo de contagio, lugar donde se encuentra la persona actualmente y el estado de la misma, para que la persona ingrese en esta clasificación según el instituto nacional de salud debe presentar la siguiente sintomatología: fiebre cuantificada mayor o igual a 38 °C, dificultad respiratoria, odinofagia, fatiga, anosmia o hipogeusia, además se tiene en cuenta las siguientes condiciones, historial de viaje dentro o fuera de Colombia, antecedentes de contacto estrecho en los últimos 14 días con un caso confirmado o probable de contagio.

Salud Data contribuye con el sector de la salud en el manejo, análisis e interpretación de datos de covid 19 para la ciudad de Bogotá, con esto dispone al público mediante su página web, un análisis detallado del comportamiento del virus en la capital del país, lo que ayuda a la toma de decisiones relacionadas con las medidas de prevención.

La implementación de este tipo de soluciones y otras como: Inteligencia artificial (IA), Big Data y relacionadas con el análisis masivo de datos, ayuda a complementar el sistema de salud, contribuyendo a su fortalecimiento.



### 3. PROBLEMÁTICA

La pandemia COVID-19 es una emergencia sanitaria que amenaza de manera particular a todos los países y Colombia ha venido con una curva bastante aplanada y está probablemente en el número más alto del primer brote (Social n.d.), y de manera particular Bogotá es una de la ciudades que ha presentado mayor número de contagios, por las complejas dinámicas sociales, económicas y sanitarias de la capital de Colombia. tal complejidad ha exigido una gran capacidad del sistema de salud e identificar aspectos relevantes para su intervención mediante métodos de recolección de datos donde la Secretaria de salud de Bogotá ha contado con entidades cooperantes que generan procesos y proyectos de investigación y gestión del conocimiento relacionadas con la vigencia en salud pública.

La Organización Mundial de la Salud (OMS) especifica la vigilancia epidemiológica como “la recolección sistemática, análisis e interpretación de datos de salud necesarios para la planificación, implementación y evaluación de políticas de salud pública. Combinado con la difusión oportuna de los datos que los actores necesitan saber”(Covida 2020),

Según los autores (Díaz 2020) el problema surge en el procedimiento de recolección de las pruebas que se realiza manualmente, a lo que se suma la ingente cantidad de información que se requiere procesar, tales como datos relacionados con la rápida transmisión, dinámica molecular y celular del virus, trazabilidad acerca de la susceptibilidad poblacional y étnica asociadas con la pandemia, incluso el monitoreo del nivel de riesgo de empleados en una empresa, entre otros aspectos.

Por esta razón, se propone la creación de un Data Mart para el almacenamiento de datos de casos confirmados de covid-19, y lograr un análisis de baja granularidad que permita evidenciar los efectos del covid en la ciudad de Bogotá, además validar la veracidad de los datos respecto a la información expuesta en Salud Data y poder contribuir con el sistema de Salud.

## **4. OBJETIVOS**

A continuación, se enlistan los objetivos que se proponen alcanzar en el desarrollo del proyecto minería de datos covid-19 Bogotá

### **1.1 OBJETIVO GENERAL**

- Analizar el comportamiento del covid-19 en la ciudad de Bogotá, durante el año 2020 hasta la fecha, mediante la implementación de un Data mart en un periodo de dos meses.

### **2.1 OBJETIVOS ESPECÍFICOS**

- Caracterizar las preguntas frecuentes de covid 19 para orientar el análisis de los datos.
- Recolectar los datos de casos confirmados de Covid-19 en Bogotá D.C para poblar la bodega de datos.
- Modelar el conjunto de dimensiones y hechos de la bodega de datos para la construcción del ETL.
- Establecer el ETL para la implementación del cubo para análisis OLAP.
- Diseñar e Implementar el cubo para análisis OLAP.
- Graficar análisis sobre cubo OLAP para visualizar los resultados del análisis y dar respuesta a las preguntas planteadas.

## 5. ALCANCE

El resultado de este proyecto es la creación de un Data Mart de casos confirmados de COVID-19 en la ciudad de Bogotá, para dar respuesta a las siguientes métricas que se plantearon en base a las siguientes necesidades:

- Cantidad de hombres y mujeres contagiados en: día de la semana, semana, mes, trimestre y año
- Cantidad de casos confirmados por localidad teniendo en cuenta: mes, trimestre y año.
- cantidad de personas fallecidas: día de la semana, semana, mes, trimestre, año.
- Cantidad de casos recuperados: día de la semana, semana, mes, trimestre, año.
- Promedio de edad con más contagios.
- Promedio de días transcurridos desde la fecha de los síntomas hasta la fecha del diagnóstico.
- Contagios por rango de edades.

## **6. MARCO CONCEPTUAL**

### **3.1 ETL**

Extraer, transformar y cargar (ETL) es una canalización de datos que se utiliza para recopilar datos de varias fuentes, transformar los datos de acuerdo con las reglas comerciales y cargarlos en un almacén de datos de destino. El trabajo de transformación en ETL se lleva a cabo en un motor especializado y, a menudo, implica el uso de tablas de preparación para almacenar temporalmente los datos a medida que se transforman y, en última instancia, se cargan en su destino.

La transformación de datos que tiene lugar generalmente implica varias operaciones, como filtrar, clasificar, agregar, unir datos, limpiar datos, quitar duplicados y validar datos. (Microsoft, 2019a)

### **4.1 OLAP**

Los cubos de procesamiento analítico en línea (OLAP) son una función de Service Manager que utilizan la infraestructura de almacenamiento de datos existente para proporcionar capacidades de inteligencia empresarial de autoservicio a los usuarios finales.

Un cubo OLAP es una estructura de datos que supera las limitaciones de las bases de datos relacionales al proporcionar un análisis rápido de los datos. Los cubos pueden mostrar y sumar grandes cantidades de datos al mismo tiempo que brindan a los usuarios acceso de búsqueda a cualquier punto de datos. De esta manera, los datos se pueden acumular, dividir y dividir según sea necesario para manejar la más amplia variedad de preguntas que son relevantes para el área de interés de un usuario. (Microsoft, 2019b)

### **5.1 VISUAL STUDIO MICROSOFT**

El entorno de desarrollo integrado de Visual Studio es una plataforma de lanzamiento creativa que se puede usar para editar, depurar y compilar código, y luego publicar una aplicación. Un entorno de desarrollo integrado (IDE) es un programa rico en funciones que se puede utilizar para muchos aspectos del desarrollo de software. Además del editor y depurador estándar que ofrecen la mayoría de los IDE, Visual Studio incluye compiladores, herramientas de finalización de código, diseñadores gráficos y muchas más funciones para facilitar el proceso de desarrollo de software. (*Overview of Visual Studio | Microsoft Docs*, n.d.)

### **6.1 SQL SERVER DEVELOPER**

Es una edición gratuita con todas las características que se puede usar como base de datos de desarrollo y pruebas en un entorno que no sea de producción. (Microsoft-SqlServer, 2020)

### **7.1 SERVER MANAGEMENT STUDIO**

SQL Server Management Studio (SSMS) es un entorno integrado para administrar cualquier infraestructura SQL. Utilice SSMS para acceder, configurar, gestionar, administrar y desarrollar todos los componentes de SQL Server, Azure SQL Database y Azure Synapse Analytics. SSMS proporciona una única utilidad integral

que combina un amplio grupo de herramientas gráficas con varios editores de scripts enriquecidos para brindar acceso a SQL Server para desarrolladores y administradores de bases de datos de todos los niveles. (Microsoft, 2019c)

## **8.1 INTEGRATION SERVICES**

SQL Server Integration Services es una plataforma para crear soluciones de transformación de datos e integración de datos a nivel empresarial. Utilice Integration Services para resolver problemas comerciales complejos copiando o descargando archivos, cargando almacenes de datos, limpiando y extrayendo datos y administrando objetos y datos de SQL Server. (Microsoft, 2018)

## **9.1 ANALYSIS SERVICE**

Es un motor de datos analíticos (VertiPaq) que se utiliza en soporte de decisiones y análisis de negocios. Proporciona capacidades de modelo de datos semánticos de nivel empresarial para inteligencia empresarial (BI), análisis de datos y aplicaciones de informes como Power BI, Excel, Reporting Services y otras herramientas de visualización de datos. (Microsoft, 2020)

## **10.1 POWER BI**

Es una colección de servicios de software, aplicaciones y conectores que funcionan en conjunto para convertir sus fuentes de datos no relacionadas en información coherente, visualmente inmersiva e interactiva. Sus datos pueden ser una hoja de cálculo de Excel o una colección de almacenes de datos híbridos locales y basados en la nube. Power BI le permite conectarse fácilmente a sus fuentes de datos, visualizar y descubrir lo que es importante y compartirlo con quien quiera. (Microsoft, 2019d)

## 7. METODOLOGÍA

A continuación, se describe la metodología que permite el desarrollo de los objetivos planteados

### 11.1 PLAN DE TRABAJO SEGÚN OBJETIVOS PLANTEADOS

A continuación, se presentan los objetivos con sus respectivas actividades a desarrollar, además se presenta el resultado obtenido del desarrollo de las mismas.

Objetivo	Actividad	Tarea	Producto
Caracterizar las preguntas frecuentes de Covid 19 para orientar el análisis de los datos.	Recolección de preguntas acerca del Covid 19 en diferentes páginas web.	Extraer preguntas de las diferentes páginas web.	Preguntas acerca de los contagios de Covid 19.
	Analizar preguntas acerca del covid 19.	Extraer métricas de las preguntas analizadas.	Métricas que responden a las preguntas.
Recolectar los datos de casos confirmados de Covid-19 en Bogotá D.C.	Buscar páginas web que contengan datos relacionados con covid 19.	Extraer datos abiertos de covid-19 en la ciudad de Bogotá.	Datos relacionados con Covid 19 en la ciudad de Bogotá.
Modelar el conjunto de dimensiones y hechos de la bodega de datos.	Establecer el modelo que mejor se ajuste a la necesidad.	Definir la tabla de hechos.	Tabla de hechos.
		Definir la granularidad de los datos.	
		Definir las dimensiones.	Tablas de dimensiones.
Establecer el ETL para el conjunto de datos de covid-19.	Definir recursos.	Definir fuente de datos.	Fuente de datos.
		Definir destino de carga de datos	Destino de almacenamiento de datos.
		Definir	Herramienta de

		herramienta para transformación y cargue de datos.	transformación y cargue de datos.
	Transformación y carga de datos.	Instalación de herramientas.	Herramientas de integración configuradas en el ambiente de desarrollo.
		Procesar fuente de datos (archivos, bases de datos, etc).	Datos en formato consistente.
		Carga de datos.	Datos cargados en el destino de almacenamiento.
Diseñar e Implementar el cubo para análisis OLAP.	Desarrollar el cubo.	Instalación de herramientas.	Configuración de ambiente de trabajo.
		procesar el cubo.	Cubo procesado.
		Realizar análisis OLAP.	Métricas del covid-19.

Tabla 1. Plan de trabajo según objetivos

## 8. METADATOS NÚMERO DE CASOS CONFIRMADOS POR EL LABORATORIO DE COVID- 19

En el siguiente cuadro se presentan las etiquetas y su respectiva descripción de los datos de casos confirmados de Covid-19 en Bogotá D.C.

Nombre dato	Descripción
Serie de tiempo desde	Marzo de 2020
Definición del indicador	Número de casos confirmados por el laboratorio de COVID-19
Utilidad	Permite conocer la evolución del número de casos confirmados con nuevo coronavirus COVID-19 en el contexto de la pandemia.
Unidad de medición	Número absoluto
Periodicidad	Diario
Fuente de Información	SIVIGILA D.C.
Fecha inicio de los síntomas	Fecha en la que el caso refiere el inicio de los síntomas
Fecha de diagnóstico	Fecha en la que se confirmó el diagnóstico del laboratorio
Ciudad de residencia	Ciudad donde reside el caso
Localidad	Localidad de residencia del caso
Edad	Edad de la persona diagnosticada dada en años meses o días.
Sexo	Sexo al que pertenece la persona diagnosticada
Tipo de caso	Forma como adquirió la infección de COVID-19.
Ubicación	Lugar en donde se encuentra el caso la cual tiene 3 salidas: importado, hospitalización y UCI
Estado	Estado de la persona diagnosticada. tiene 5 estados: leve, moderado, fallecido, grave y recuperado.

*Tabla 2. Metadatos casos covid-19*



## 9. RECOLECCIÓN

Los datos recolectados para el desarrollo de este proyecto se obtuvieron en su totalidad del portal de datos de Covid 19 en la ciudad de Bogotá, estos se consolidaron gracias a los filtros de búsqueda que la misma página facilita, resultando así el conjunto de datos con las características específicas planteadas en los objetivos.

El total de datos arrojados por el portal se puede consultar en: <https://datosabiertos.bogota.gov.co/dataset/numero-de-casos-confirmados-por-el-laboratorio-de-covid-19-bogota-d-c/resource/b64ba3c4-9e41-41b8-b3fd-2da21d627558>

con un total 1079120 registros.

El proceso de obtención de los datos se hizo ingresando a Salud data y dar clic en casos confirmados de covid 19 y descargar el formato CSV (Archivo delimitado por comas).

### 1.1.1 Diagrama de bus

Como parte de la representación del Data mart se utilizó la siguiente matriz de bus en donde se representan las dimensiones que hacen parte de los hechos: Dimensión de tiempo, ubicación, rango de edades, estados, fuentes de contagio, casos y dimensión de localidades que describen las métricas de contagios.

Grupos de medida	
Dimensiones	Contagios Fact
Dim Fechas (Fecha Sintomas)	Fecha
Dim Fechas (Fecha Diagnost...)	Fecha
Dim Ubicaciones	NombreUbicacion
Dim Rango Edades	Rango
Dim Estados Salud	NombreEstadoSalud
Dim Fuentes Contagio	NombreFuenteContagio
Dim Casos	Id
Dim Localidades	NombreLocalidad

Imagen 1. Matriz de bus

## **10. PRODUCTOS**

### **12.1 MODELO DE DATOS.**

Se utilizó un modelo de datos basado en estrella el cual se compone de siete tablas de dimensiones y una tabla de hechos como se muestra a continuación:

#### **13.1 DIMENSIONES:**

DimFechas: Representa las fechas de inicio de síntomas y fecha de diagnóstico de Covid 19.

DimCaso: Representa la edad y sexo de las personas contagiadas.

DimRangoEdad: Representa el rango de edad al que pertenece la persona contagiada.

DimLocalidad: Representa la localidad de Bogotá a donde pertenece el caso de Covid 19.

DimEstadoSalud: Representa el estado de salud de la persona (Fallecido, Fallecido no aplica causa directa, Leve, Recuperado).;

DimUbicacion: Representa en qué parte se encuentra la persona en el momento (Casa, Fallecido, Fallecido no aplica causa directa, Hospital, Hospital UCI).

DimFuenteContagio: Representa el origen del contagio (Desconocido, En estudio, Importado, Relacionado).

#### **14.1 HECHOS:**

ContagiosFact: Representa las métricas de contagio (número días demora en el diagnóstico de contagio).

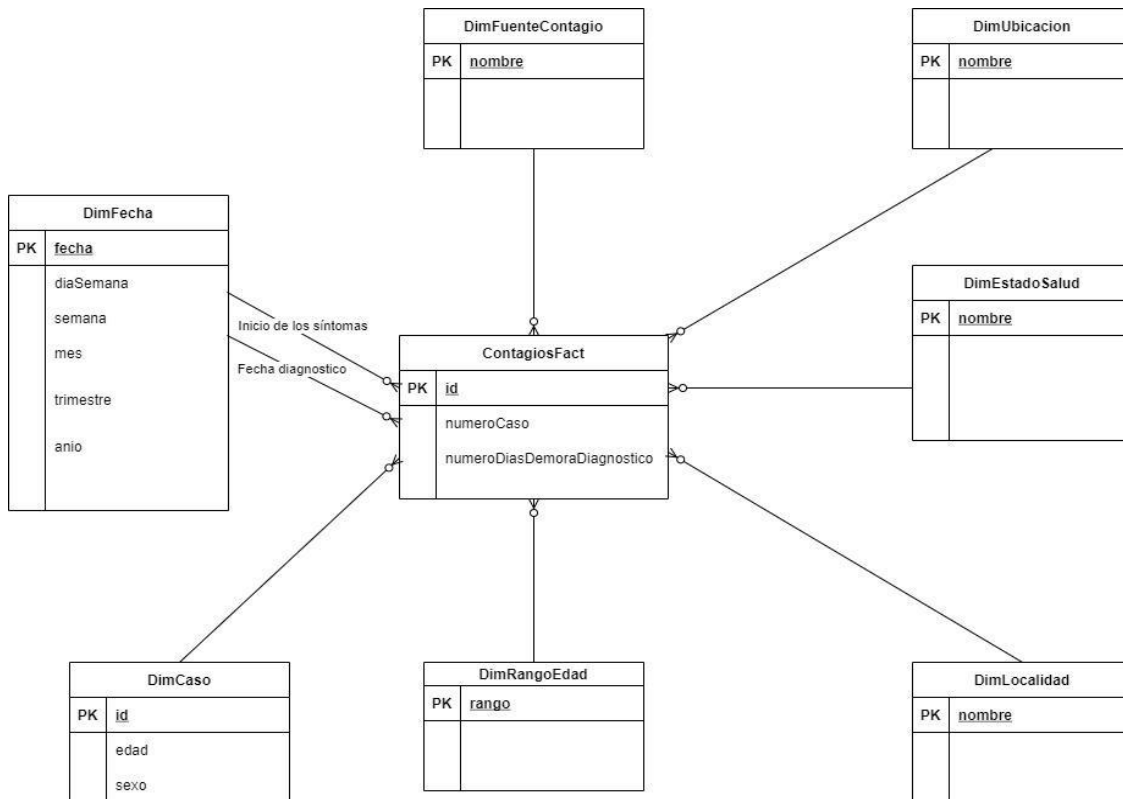


Imagen 2. Modelo en estrella

## 15.1 ORIGEN DE DATOS

En este apartado se selecciona nuestro origen de datos, los cuales corresponden al número de casos activos de covid 19 en formato CSV (Archivo delimitado por comas), en el contexto de la pandemia, además se realiza una asignación del tipo de dato correspondiente a cada campo para que permita la carga a la base de datos y no se presenten errores en el ETL.

Editor del administrador de conexiones de archivos planos

Nombre del administrador de conexiones:

Descripción:

General

Columnas

Avanzadas

Vista previa

La vista previa muestra el archivo de origen dividido en las columnas especificadas. No se muestran las filas de datos iniciales que se omiten cuando el archivo se analiza en tiempo de ejecución.

Filas de datos que se omitirán:

Vista previa de las filas 2-101:

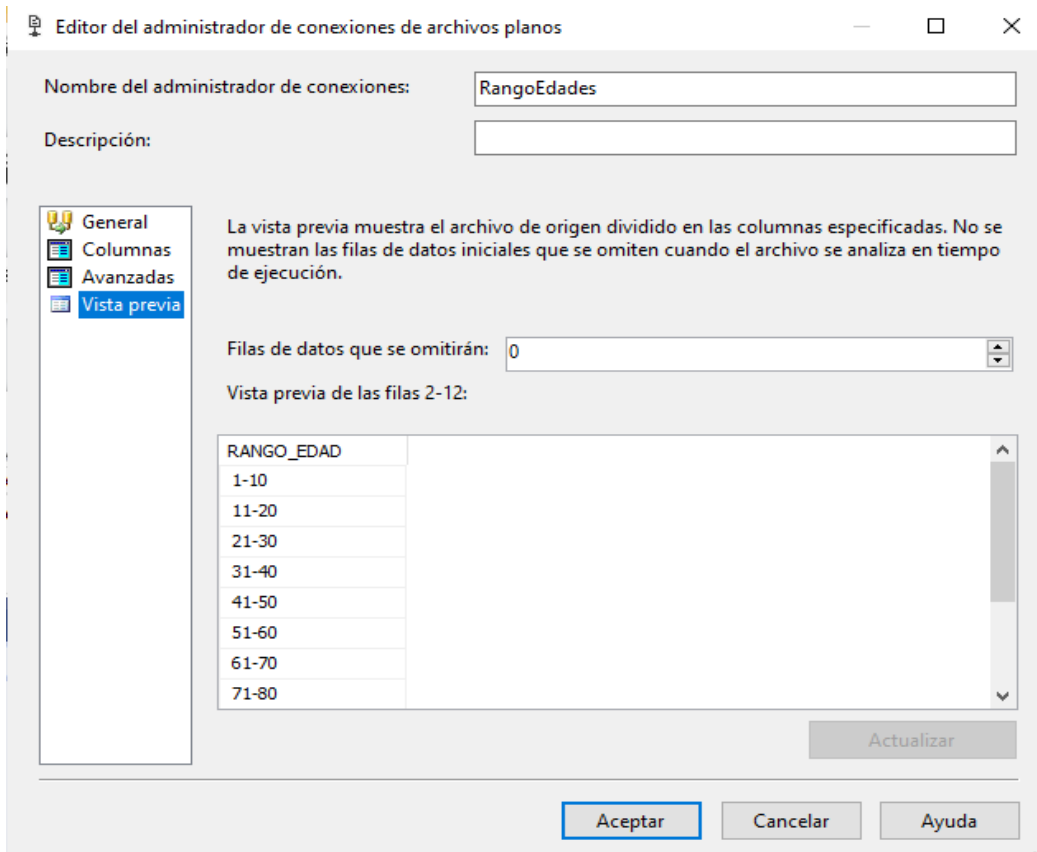
CASO	FECHA_DE_INICI...	FECHA_DIAGNOS...	CIUDAD	LC
1	26/02/2020	6/03/2020	Bogota	U
2	4/03/2020	10/03/2020	Bogota	Ei
3	7/03/2020	10/03/2020	Bogota	Ei
4	6/03/2020	12/03/2020	Bogota	Fi
5	6/03/2020	12/03/2020	Bogota	Ki
6	8/03/2020	13/03/2020	Bogota	Si
7	28/02/2020	13/03/2020	Bogota	Ti

Actualizar

Aceptar Cancelar Ayuda

Imagen 3. Origen de datos

En la siguiente imagen se muestra la carga del archivo rango de edades, los cuales serán parte fundamental para dar respuesta a las métricas que se plantearon en base a las necesidades del cliente.



*Imagen 4. Origen de datos rangos edad*

## 16.1 FLUJO DE TAREAS ETL

Se definieron las siguientes tareas de transformación y cargue de datos las cuales empiezan desde la dimensión de tiempo hasta la tabla de hechos llamada contagios, en cada una de las tareas se extraen los datos de un archivo .csv y para el caso específico de los rangos se definió un archivo .txt, se realizan ajustes a los datos y se les da un formato apropiado para ser cargados posteriormente en una Base de datos Sql Server.

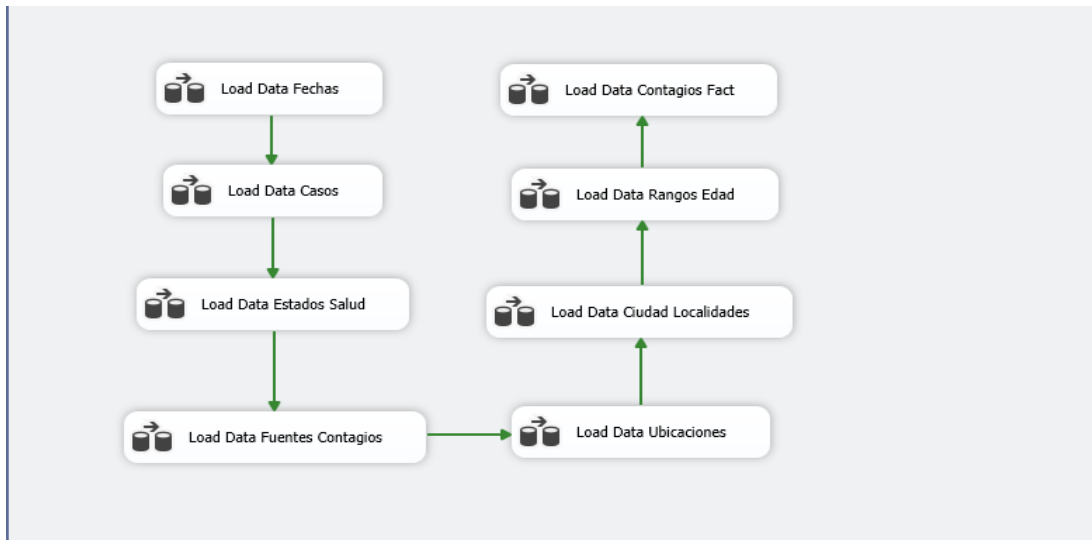


Imagen 5. Flujo de tareas ETL

### 17.1 TAREA DE CARGA DE DISENSIÓN TIEMPO

En la siguiente imagen se muestra un ejemplo de la estructura interna de la dimensión Tiempo. Esta tarea empieza su flujo con la extracción de los datos necesarios, luego hace unas transformaciones y formatos intermedios antes de hacer el cargue a la su tabla correspondiente en SQL Server.

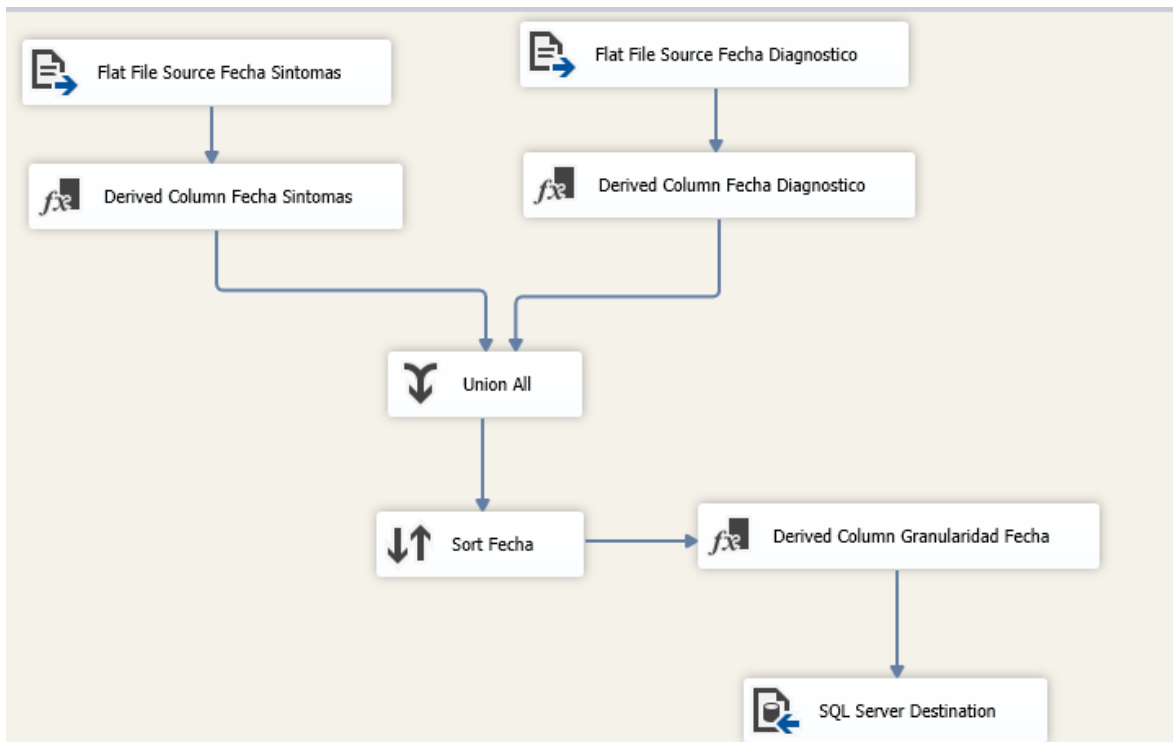
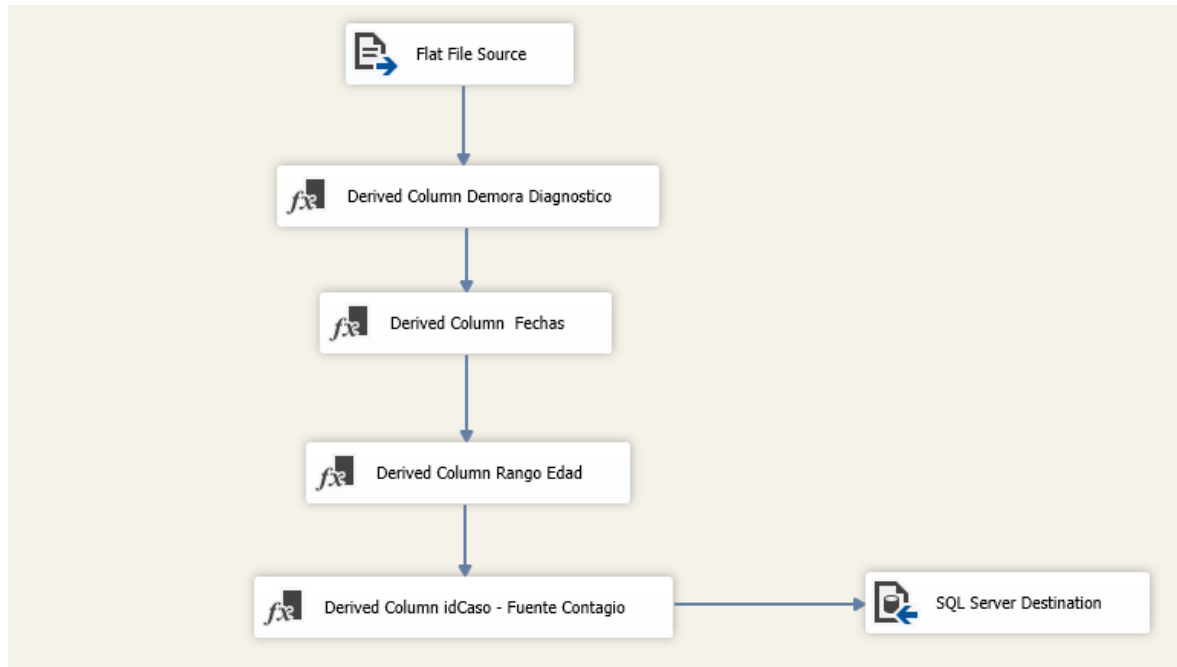


Imagen 6. Tarea carga de datos dimensión tiempo

## 18.1 TAREA CARGA DE TABLA DE HECHOS

A continuación, se presenta el ETL para la tabla de hechos, esta es una de las partes fundamentales para la creación del Data mart. iniciando con la selección del origen de datos, luego se realizan unas transformaciones y formatos intermedios para que los datos queden en limpios y finalmente hacer el cargue a la su tabla correspondiente en la base de datos de Covid-19 en SQL Server.



*Imagen 7. Carga de datos tabla de hechos*

## 19.1 DESTINO DE CARGUE DE DATOS

El destino de cargue de datos como se muestra en la siguiente imagen es una base de datos SQL Server, compuesta de ocho tablas de hechos y dimensiones:

### 20.1 TABLAS DE HECHOS:

- ContagiosFact.

### 21.1 TABLAS DE DIMENSIONES:

- DimFechas.
- DimCasos.
- DimEstadosSalud.
- DimFuentesContagio.
- DimLocalidades.
- DimRangoEdades.
- DimUbicaciones.

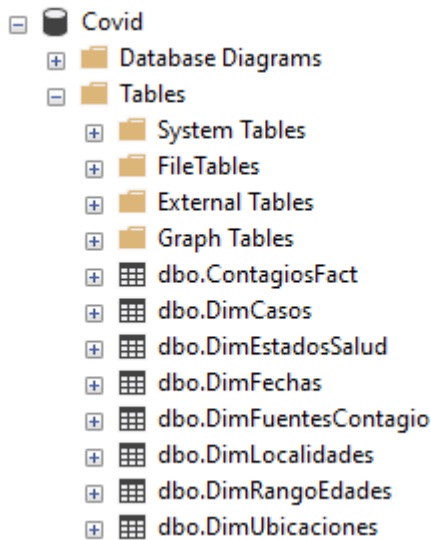


Imagen 8. Destino carga de datos

## 22.1 DESARROLLO DEL CUBO

En este apartado se describe el desarrollo del cubo OLAP, como se presenta en la siguiente imagen se selecciona el origen de datos el cual es la base de datos Covid 19 creada en SQL Server, después de esto crea el cubo indicando la base de datos existente, una vez ahí se selecciona la tabla de hechos y nos carga las dimensiones como se presenta en la imagen.

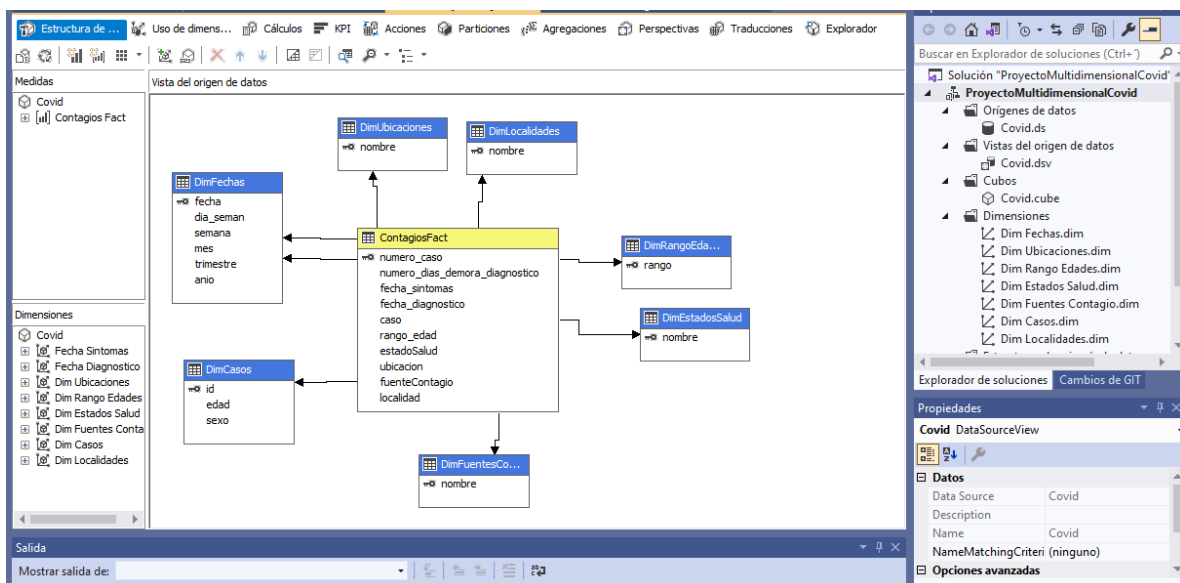


Imagen 9. Modelo cubo OLAP

En la siguiente imagen se muestra la creación de las jerarquías para la dimensión de fecha donde se establece desde el de mayor nivel hasta el nivel más bajo.



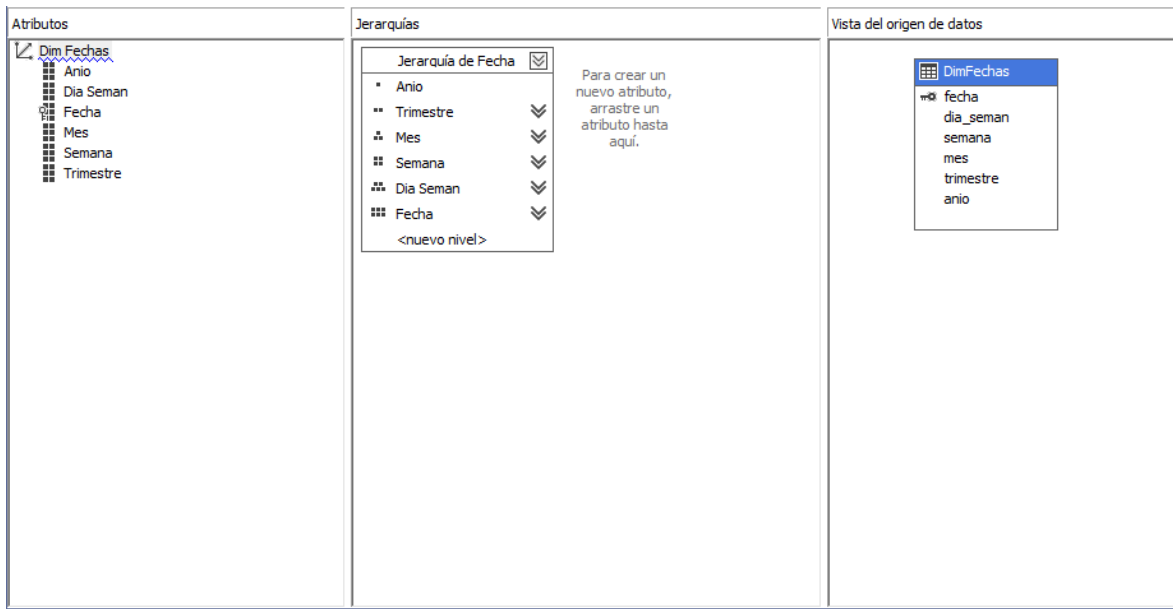


Imagen 10. Jerarquía dimensión fechas A

Una vez establecidas las jerarquías se procede a realizar unión iniciando por el nivel más bajo hasta el mayor nivel, luego se establece el tipo de relación rígida, como se observa en la imagen.



Imagen 11. Jerarquía dimensión de fechas B

### 2.1.1 Implementación del cubo

En la siguiente imagen se muestra el procesamiento realizado para la generación del cubo, en donde se procesan cada una de las dimensiones y hechos que intervienen en nuestro modelo de datos.

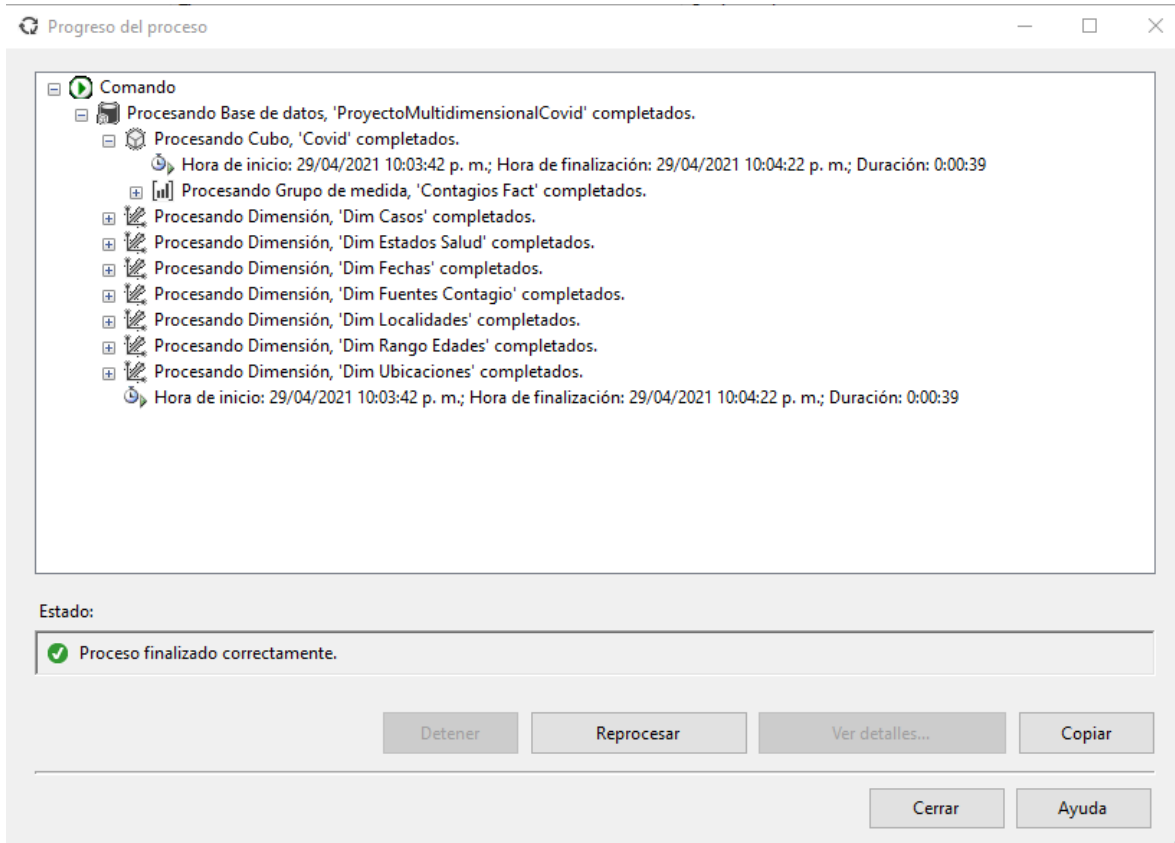


Imagen 12. Implementación del cubo OLAP

En la siguiente imagen se el proceso terminado de la implementación y queda el cubo OLAP para iniciar el análisis correspondiente a este proceso.

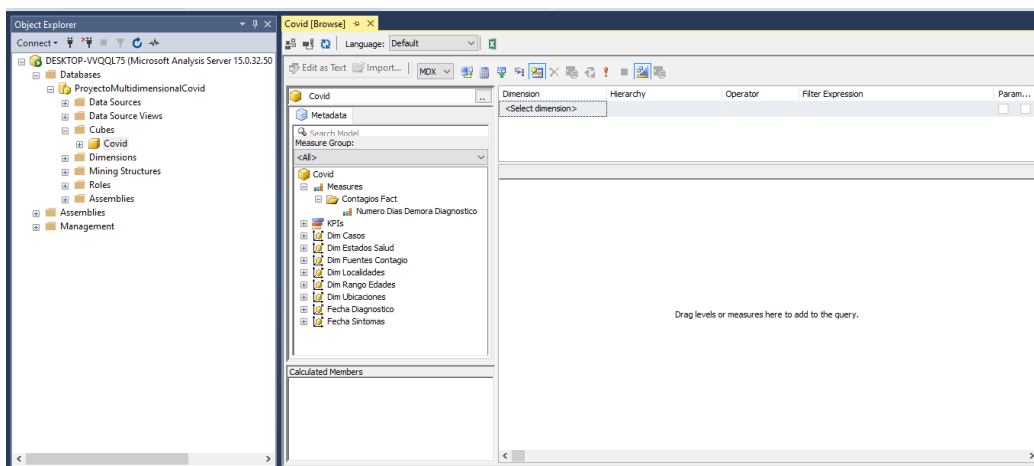
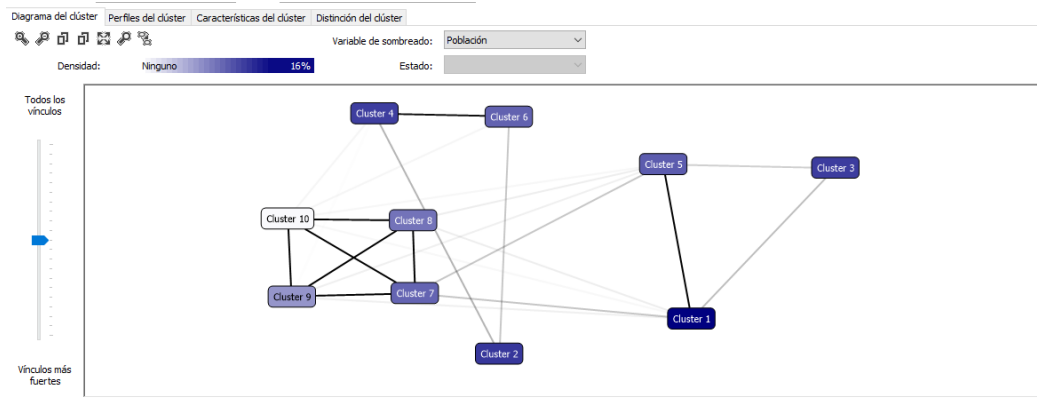


Imagen 13. Cubo en Analysis Services

### 3.1.1 Resultados del análisis.

A continuación, se presenta el modelo de minería de datos, en el cual se establece un modelo en clustering donde podemos observar las diferentes características y probabilidades de contagios por género y edad.



Atributos		Perfiles del clúster										
Variables	Estados	Población ... Tamaño: ...	Cluster 1 Tamaño: ...	Cluster 2 Tamaño: ...	Cluster 3 Tamaño: ...	Cluster 4 Tamaño: ...	Cluster 5 Tamaño: ...	Cluster 6 Tamaño: ...	Cluster 7 Tamaño: ...	Cluster 8 Tamaño: ...	Cluster 9 Tamaño: ...	Cluster 10 Tamaño: ...
Cantidad Contagio	1.20											
	1.00											
Edad	93.00											
	39.00											
Sexo	1.00											
	F M ausente NA											

## 11. ANÁLISIS DE NECESIDADES

A continuación, se realiza el análisis de las necesidades propuestas y se agregan gráficas de visualización de los resultados esperados de dicho análisis.

### 23.1 CANTIDAD DE CONTAGIADOS EN: DÍA DE LA SEMANA, SEMANA, MES, TRIMESTRE Y AÑO

#### 4.1.1 Día de la semana

En la siguiente gráfica se observa el comportamiento del Covid -19 en los días de la semana, lo cual nos indica que la tendencia de máximos contagios son los días martes.

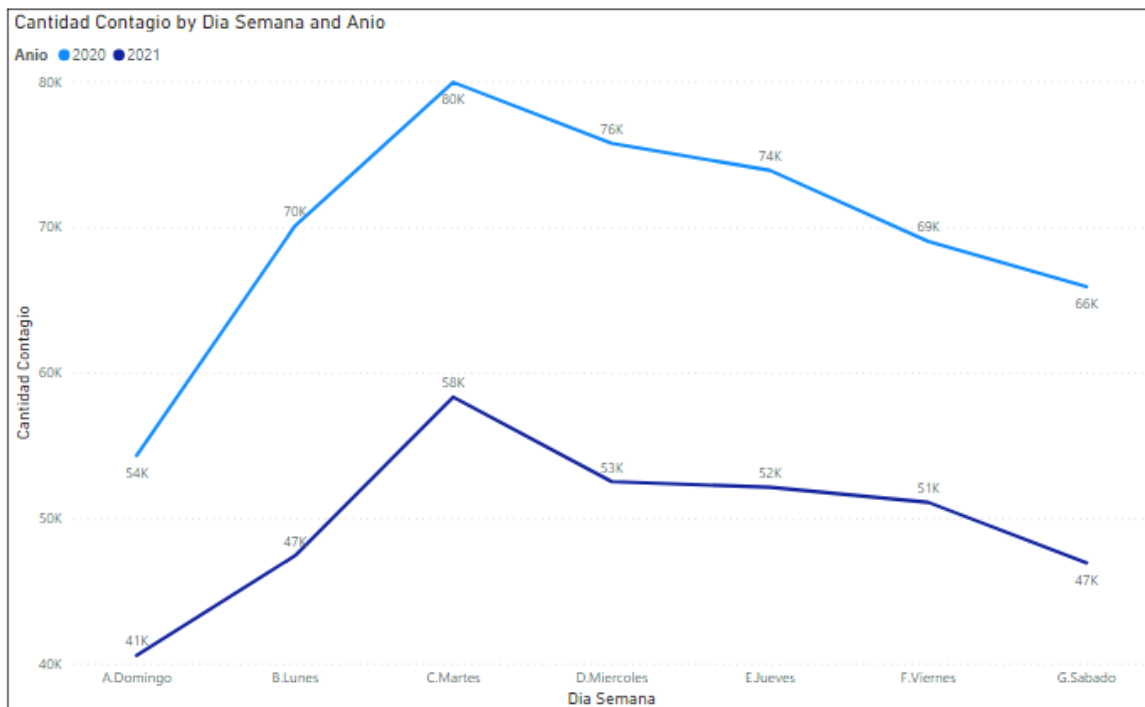


Imagen 14. Cantidad de contagios por día de la semana año 2020 y 2021

#### 24.1 SEMANA DEL AÑO 2020

A continuación, se presenta la gráfica del comportamiento del Covid por semana en el año 2020, en la cual se observa que los picos de contagios ocurren en la semana 32.

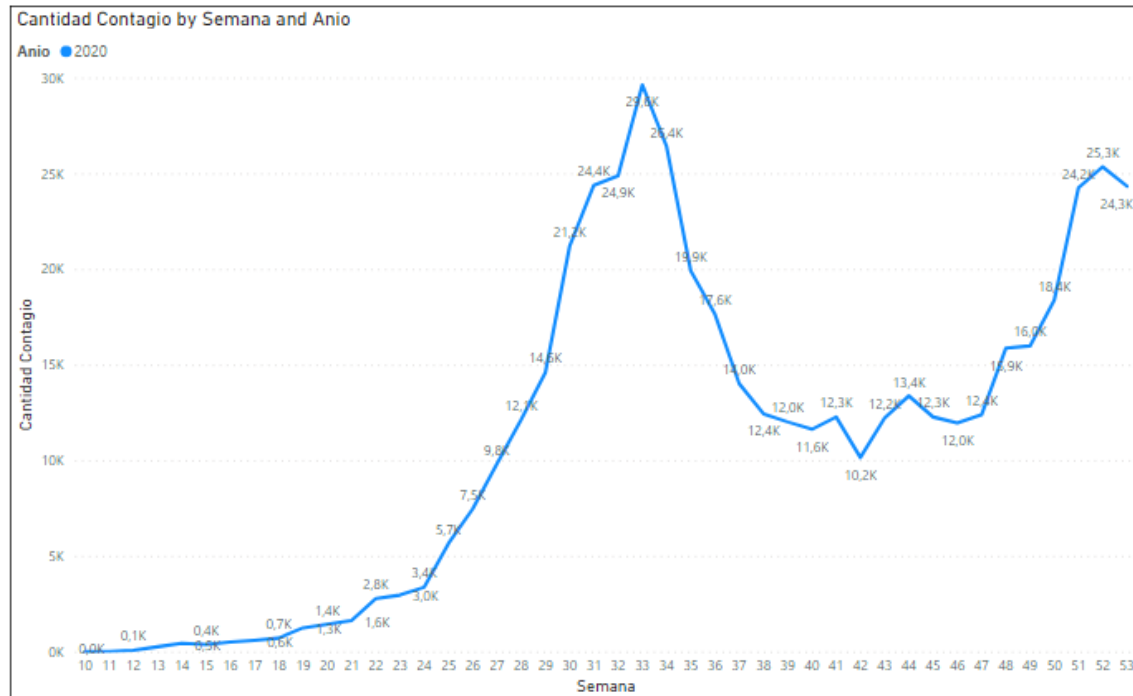
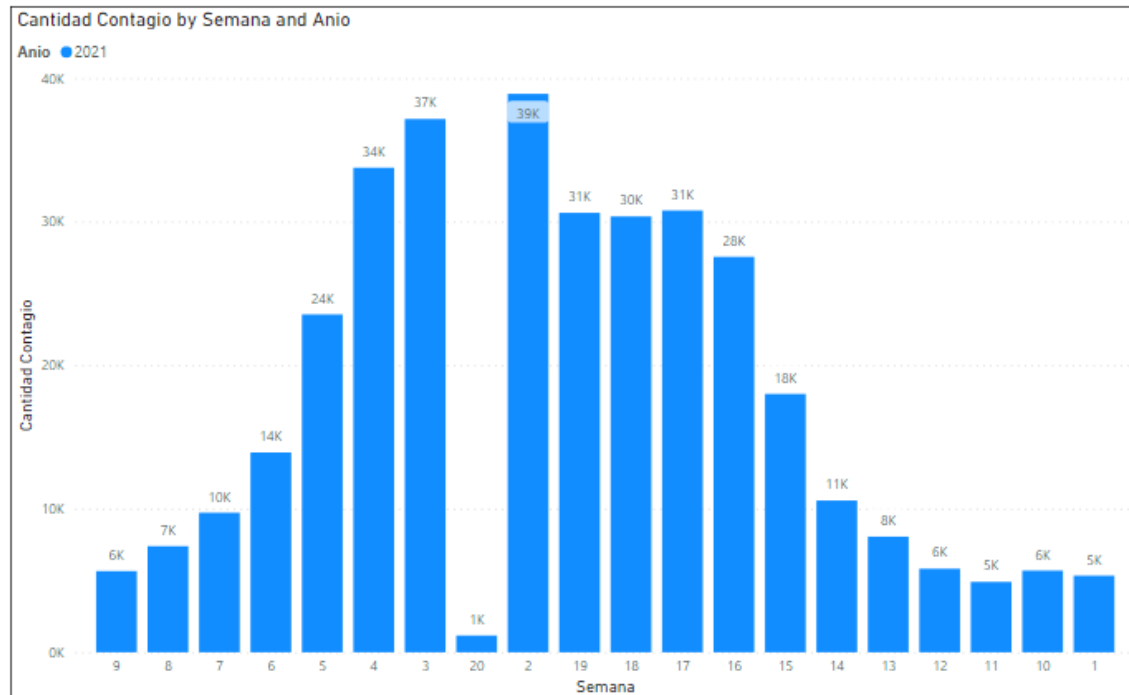


Imagen 15. Cantidad de contagios por semana año 2020

### 5.1.1 Semana del año 2021

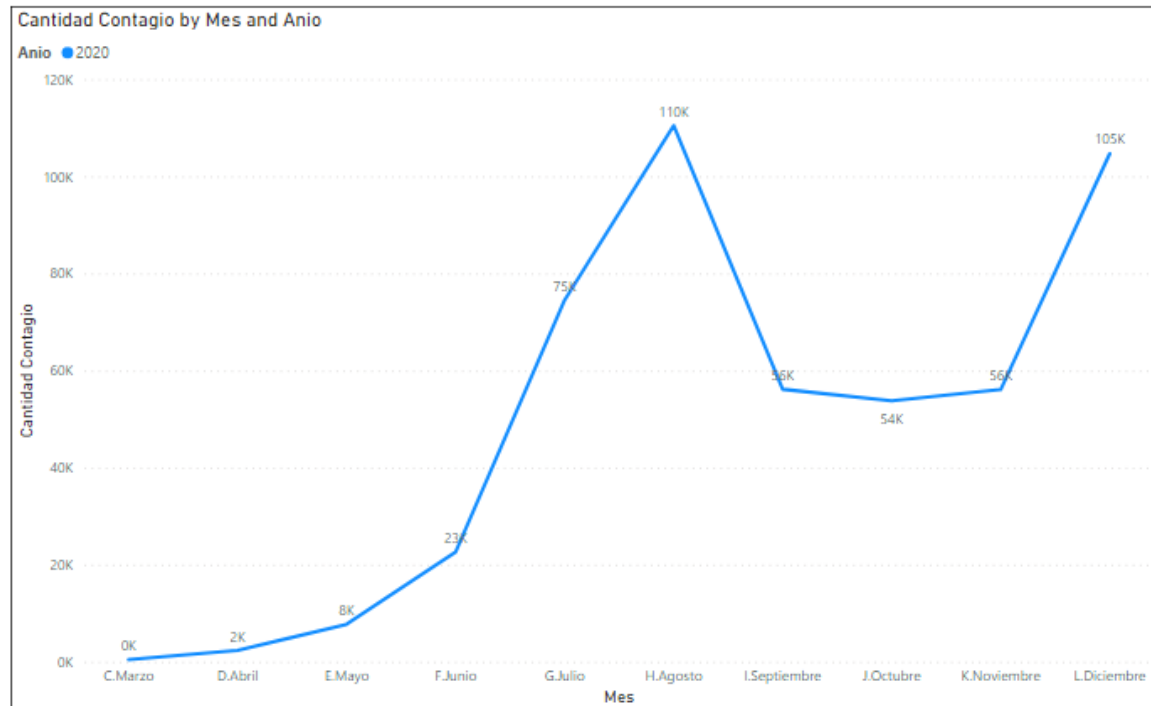
En la siguiente gráfica se presenta el comportamiento del Covid por semana en el año 2021, en la cual se observa que el pico de contagios máximo ocurrió en la semana 2, se observa que hasta la fecha los casos han descendido.



*Imagen 16. Cantidad de contagios por semana año 2021*

### 6.1.1 Mes del año 2020

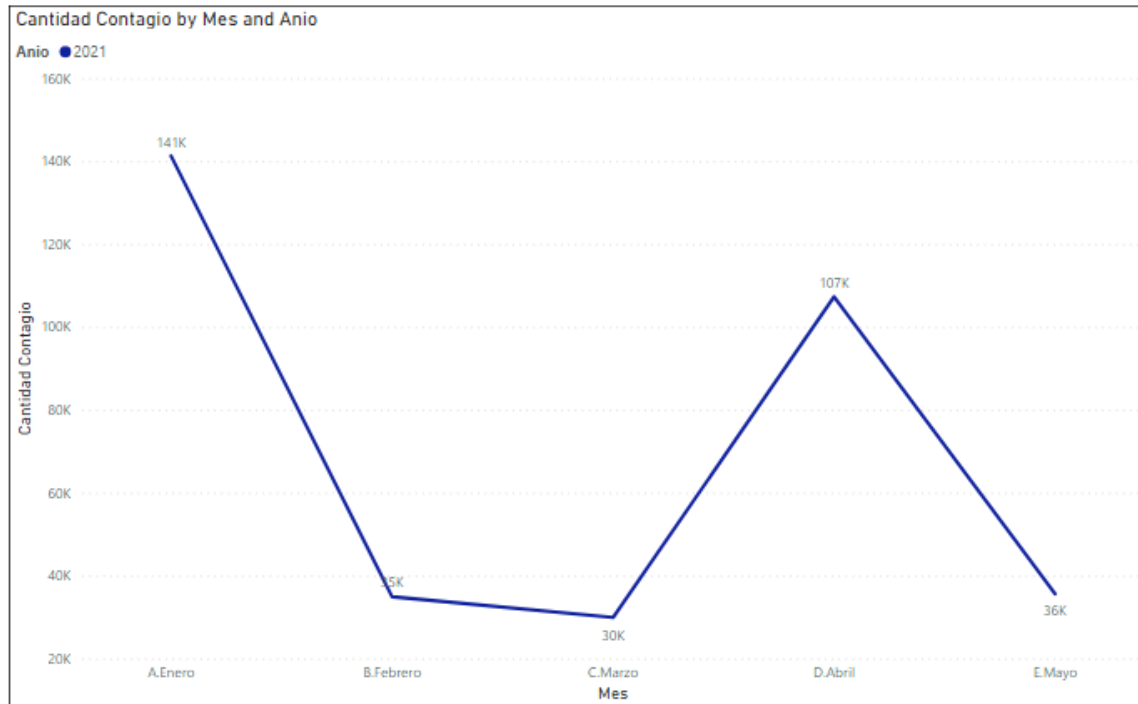
En la siguiente gráfica se presenta el comportamiento del covid por mes en el año 2020, en la cual se observa que el pico de contagios máximo ocurrió en el mes de agosto, además se observa que la cantidad de contagios decayó hasta el mes de noviembre, luego se evidencia un crecimiento exponencial de casos confirmados de covid 19.



*Imagen 17. Cantidad de contagios por mes del año 2020*

### 7.1.1 Mes del año 2021

En la siguiente gráfica se presenta el comportamiento del covid por mes en el año 2021, en la cual se observa que el pico de contagios máximo ocurrió en el mes de enero, además se ve que la cantidad de contagios decayó hasta el mes de marzo, luego se evidencia un crecimiento exponencial de casos confirmados hasta el mes de abril y disminuyen significativamente hasta el mes de mayo.

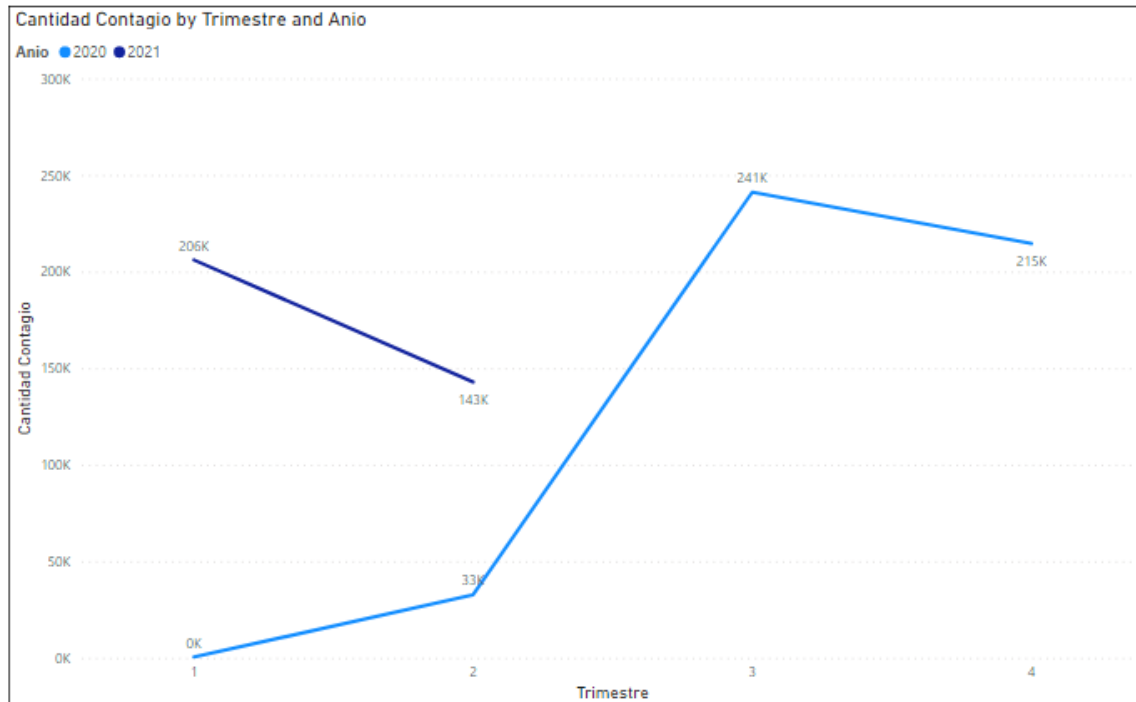


*Imagen 18. Cantidad de contagios por mes del año 2021*

### 8.1.1 Trimestre del año

A continuación, se presenta la gráfica del comportamiento del covid por trimestre en el año 2020 y 2021, en la cual se observa que para el trimestre 1 del año 2020 inician los contagios teniendo un crecimiento exponencial hasta el trimestre 3, seguido de un descenso en el trimestre 4, para el año 2021 en el primer trimestre se observa el pico de contagios máximo y posteriormente se evidencia un descenso exponencial.





*Imagen 19. Cantidad de contagios por trimestre del año 2020 y 2021*

### 9.1.1 Año

En la siguiente gráfica se evidencia el comportamiento del covid por año 2020 y 2021, en la cual se observa que desde el mes de julio hasta el mes de septiembre el número de contagios aumentan significativamente, también se observa que en el mes de diciembre del 2020 hasta finales de enero del 2021 el número de contagios alcanzó su punto máximo.

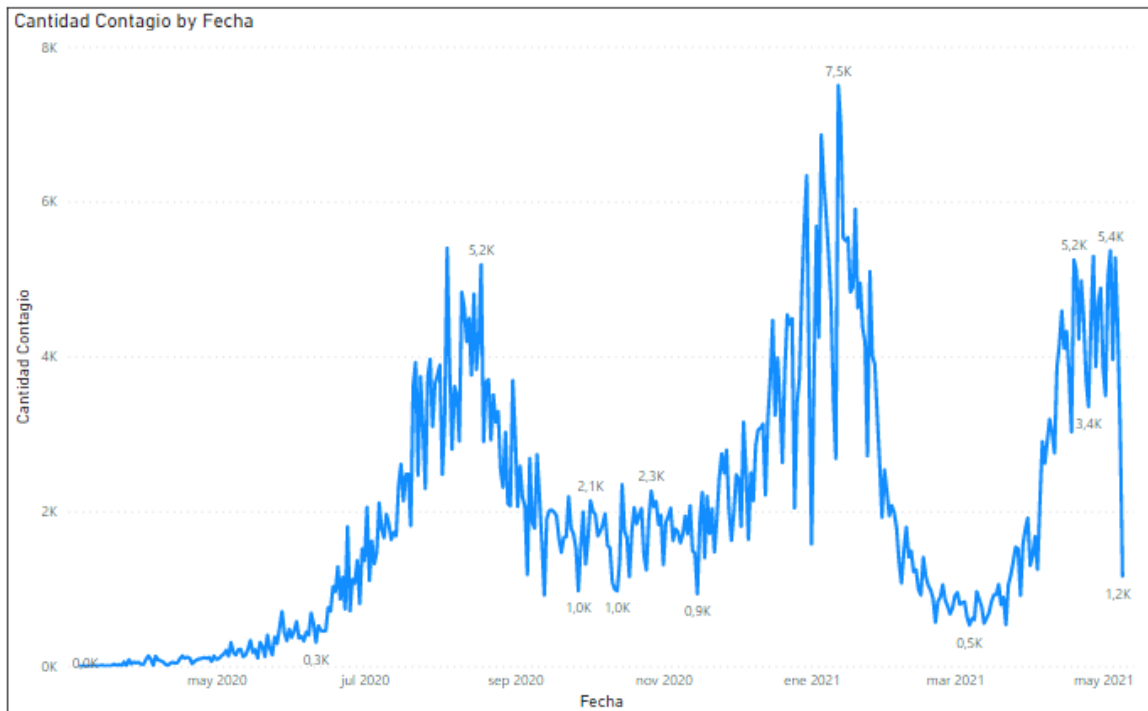


Imagen 20. Cantidad de contagios por año 2020 y 2021

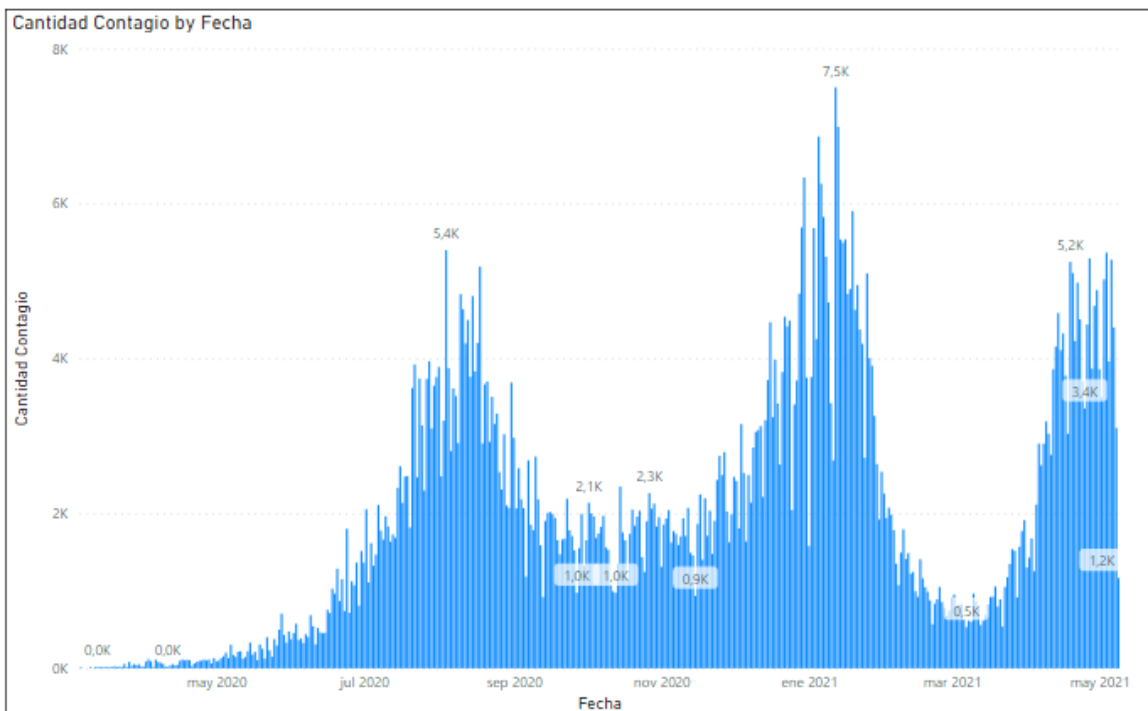


Imagen 21. Imagen 20. Cantidad de contagios por año 2020 y 2021 barras

## 25.1 CANTIDAD DE HOMBRES Y MUJERES CONTAGIADOS EN: DÍA DE LA SEMANA, SEMANA, MES, TRIMESTRE Y AÑO

### 10.1.1 Día de la semana año 2020

Se observa un mayor número de contagiados en el día martes como se muestra en la siguiente imagen, evidenciando así mismo un mayor número de mujeres que hombres contagiados.

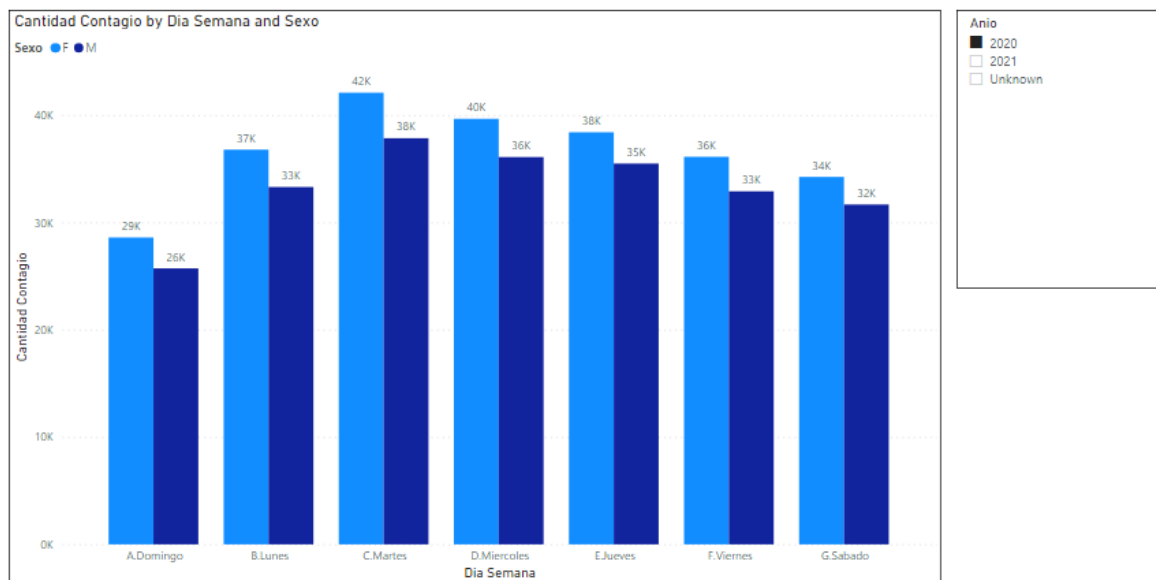


Imagen 22. Cantidad de hombres y mujeres contagiados día de la semana año 2020

### 11.1.1 Día de la semana año 2021

Se observa un mayor número de contagiados en el día martes como se muestra en la siguiente imagen, evidenciando así mismo un mayor número de mujeres que hombres contagiados. Además, se evidencia que tiene un comportamiento similar al del año anterior.

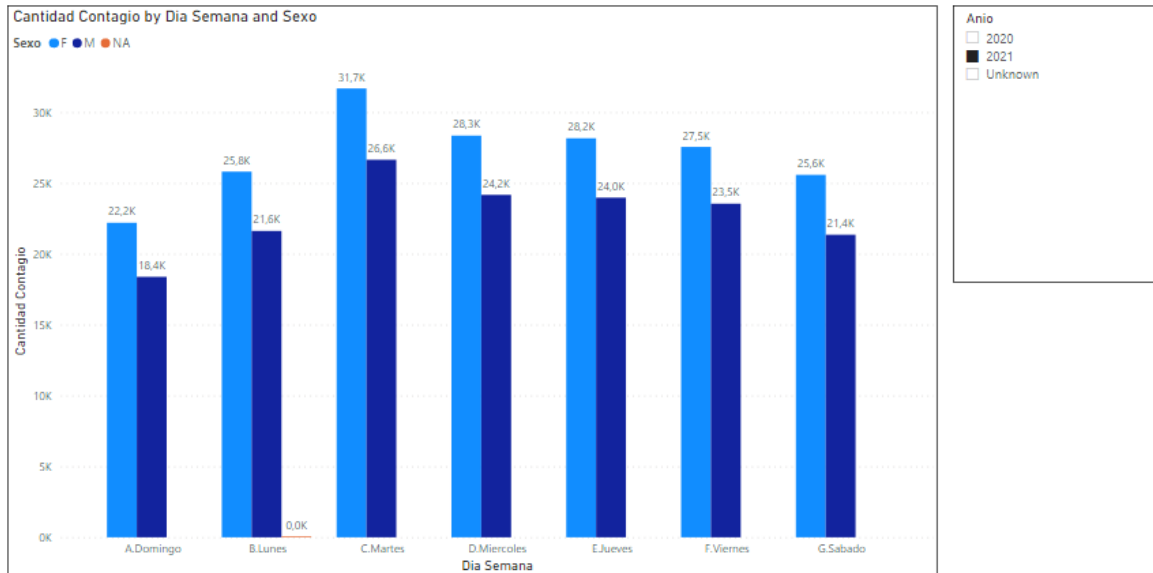


Imagen 23. Cantidad de hombres y mujeres contagiados día de la semana año 2021

### 12.1.1 Semana del año 2020

En la siguiente imagen se observa que el pico máximo de contagio ocurrió en la semana 33 en donde el número de mujeres contagiadas fue superior al de hombres.

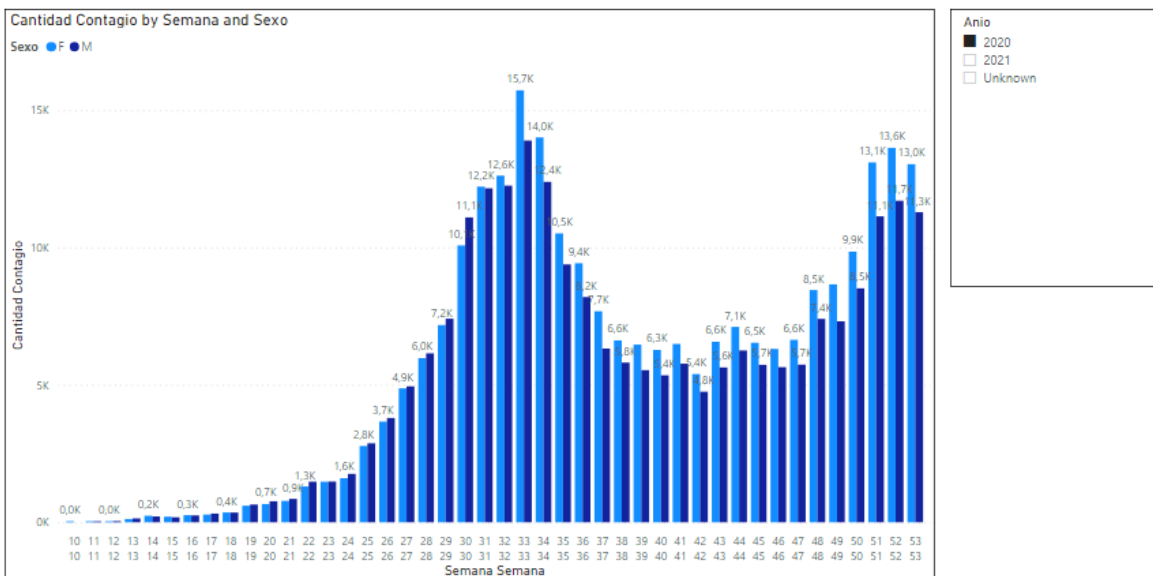


Imagen 24. Cantidad de hombres y mujeres contagiados POR SEMANA año 2020

### 13.1.1 Semana del año 2021

Para lo corrido del año 2021 la semana 22 es la de más número de contagios, siguiendo el comportamiento más alto para las mujeres que hombres contagiados.

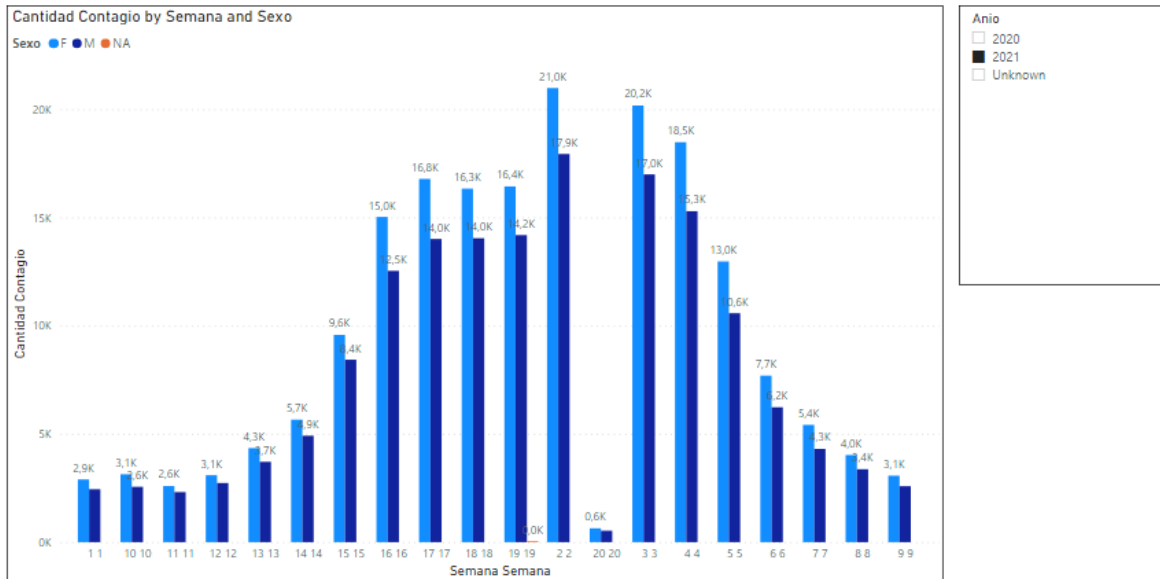


Imagen 25. Cantidad de hombres y mujeres contagiados por semana año 2021

#### 14.1.1 Mes del año 2020

En la siguiente imagen se observa como la mayor cantidad de contagios se presentó en los meses de agosto y diciembre para el año 2022, siguiendo la tendencia de más mujeres que hombres contagiados.

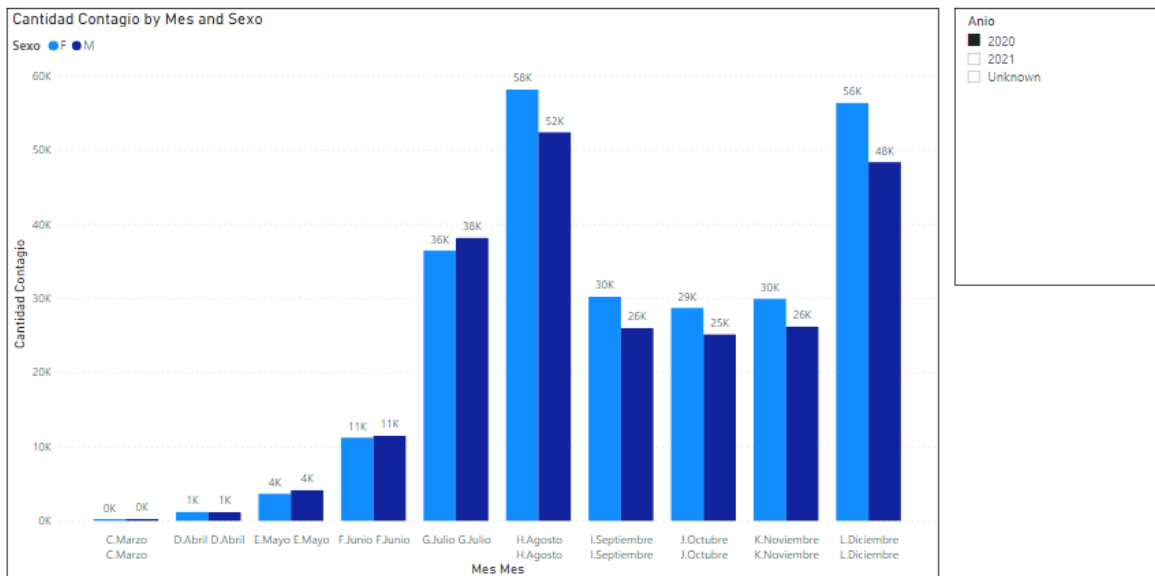


Imagen 26. Cantidad de hombres y mujeres contagiados POR MES año 2020

#### 15.1.1 Mes del año 2021

Para lo corrido del año 2021 el mes de enero fue en donde se presentó el mayor número de contagios, seguido del mes de abril, y continúa la tendencia de más mujeres que hombres contagiados.

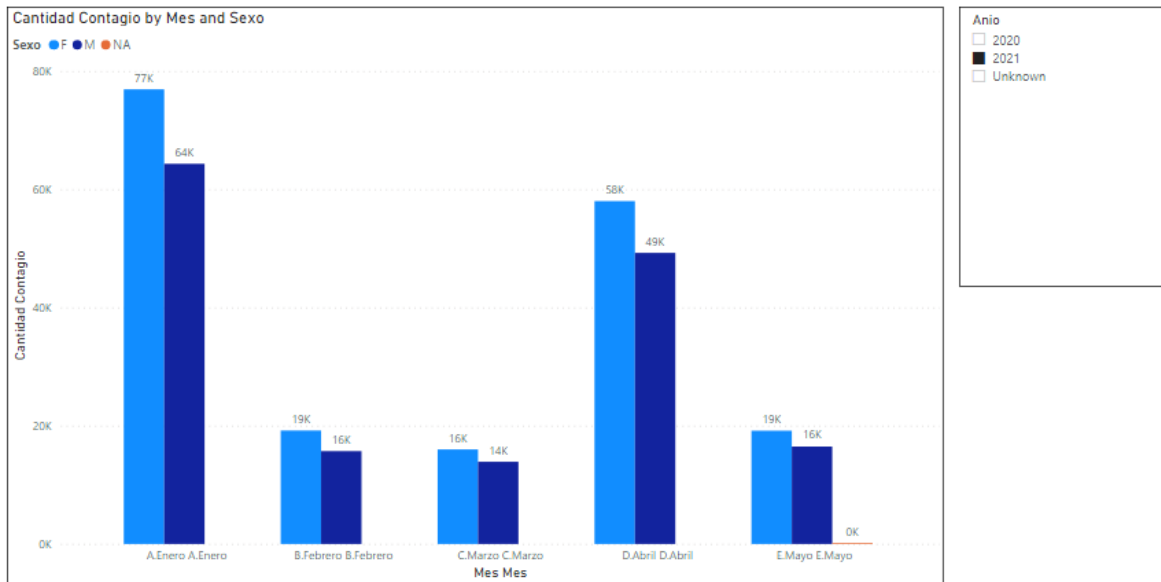


Imagen 27. Cantidad de hombres y mujeres contagiados POR MES año 2021

### 16.1.1 Trimestre año 2020

Para los trimestres del año 2020, el tercer trimestre fue donde se presentó el mayor número de contagios, seguido del cuarto semestre y evidenciando el comportamiento de las mujeres contagiadas que hombres.

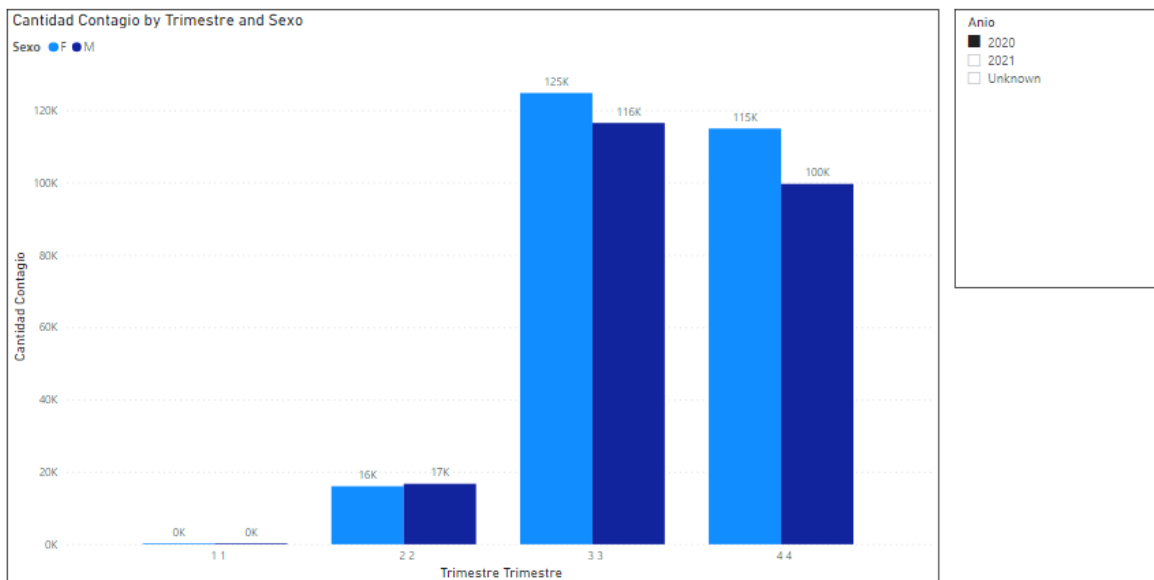


Imagen 28. Cantidad de hombres y mujeres contagiados por trimestre año 2020

### 17.1.1 Trimestre año 2021

Para lo corrido del año 2021 el primer trimestre fue en donde hubo el mayor número de contagios, siguiendo la tendencia de más mujeres contagiadas que hombres.

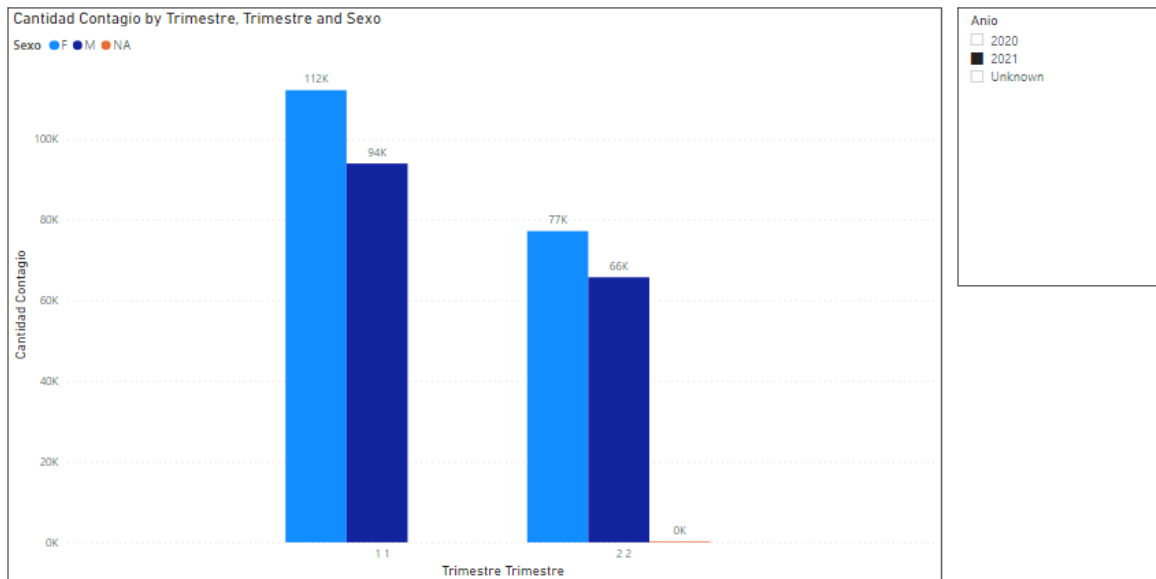


Imagen 29. Cantidad de hombres y mujeres contagiados por trimestre año 2021

### 18.1.1 Año 2020

A través de la línea de tiempo del año 2020 se observa que desde el mes de julio hasta el mes de septiembre el número de contagios aumentan significativamente y durante el mes de diciembre, en donde hay mayor número de mujeres contagiadas que hombres.

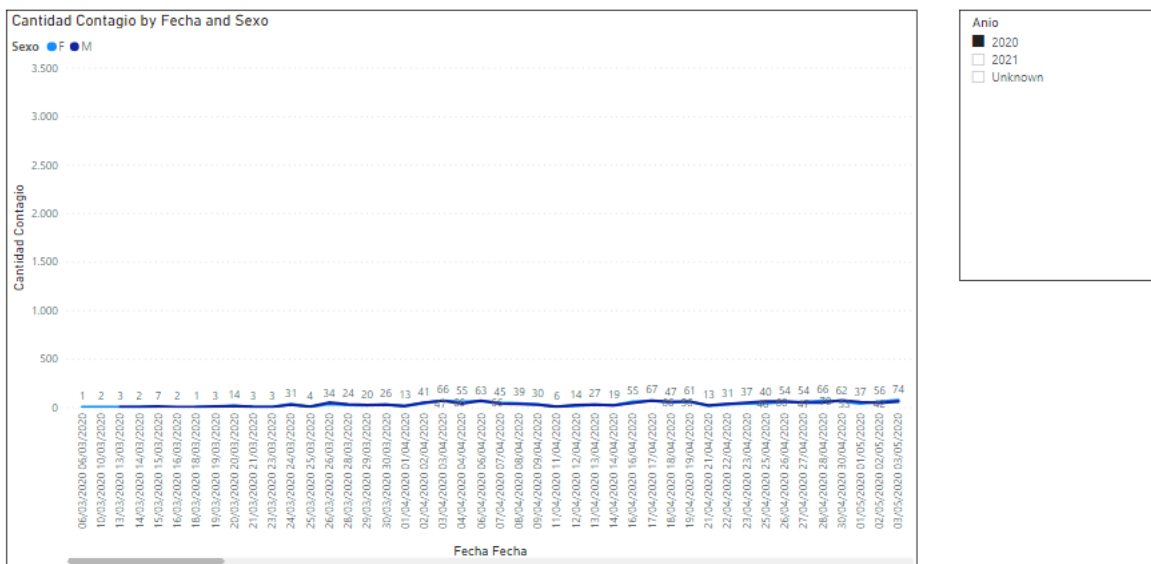


Imagen 30. Cantidad de hombres y mujeres contagiados por año 2020 A

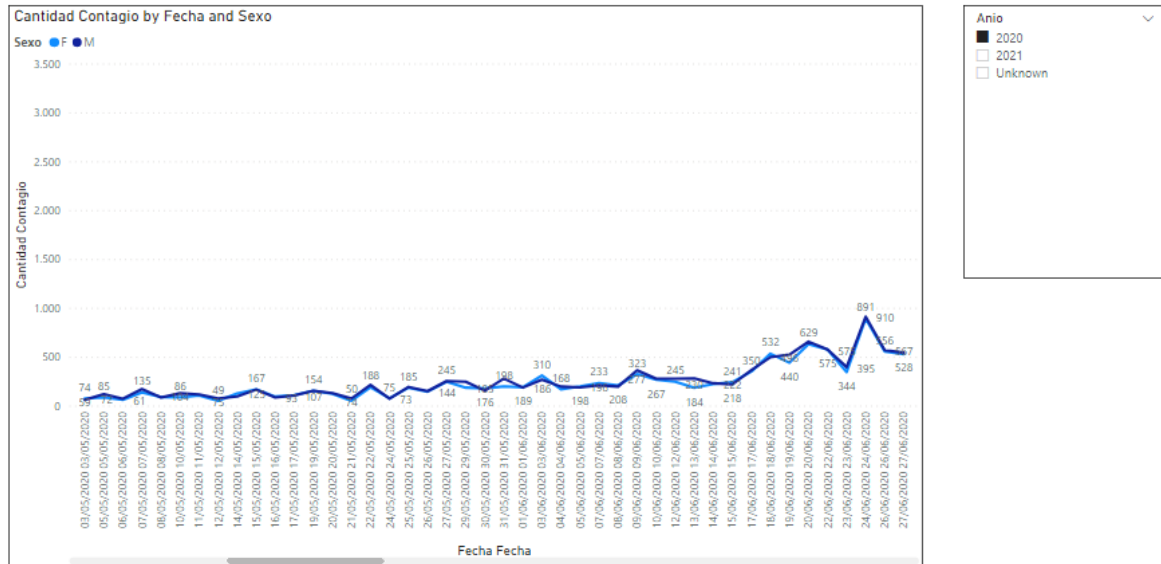


Imagen 31. Cantidad de hombres y mujeres contagiados por año 2020 B

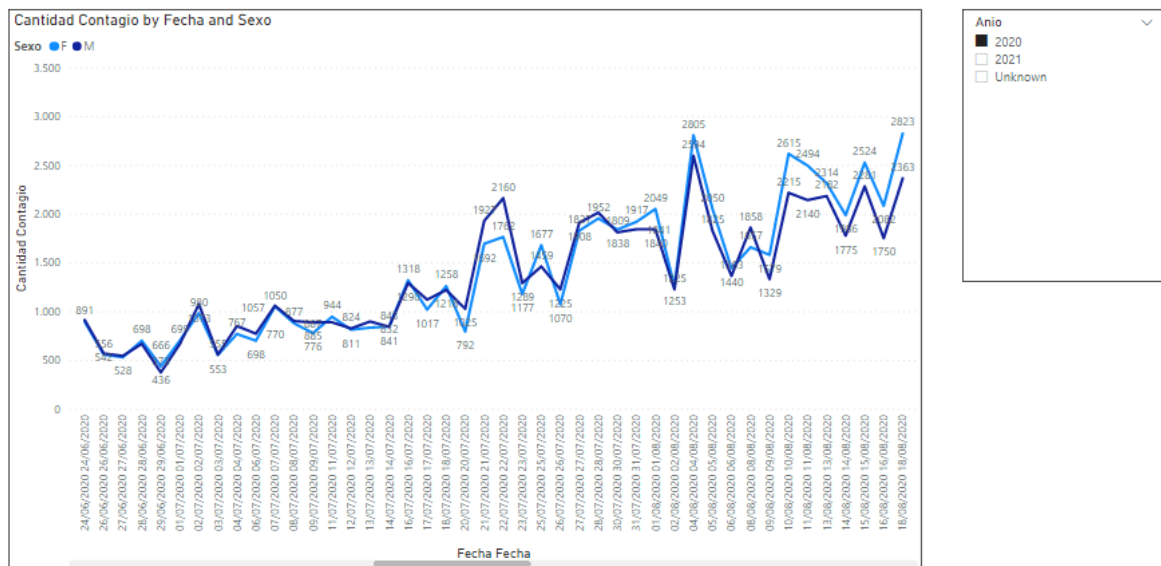


Imagen 32. Cantidad de hombres y mujeres contagiados por año 2020 C



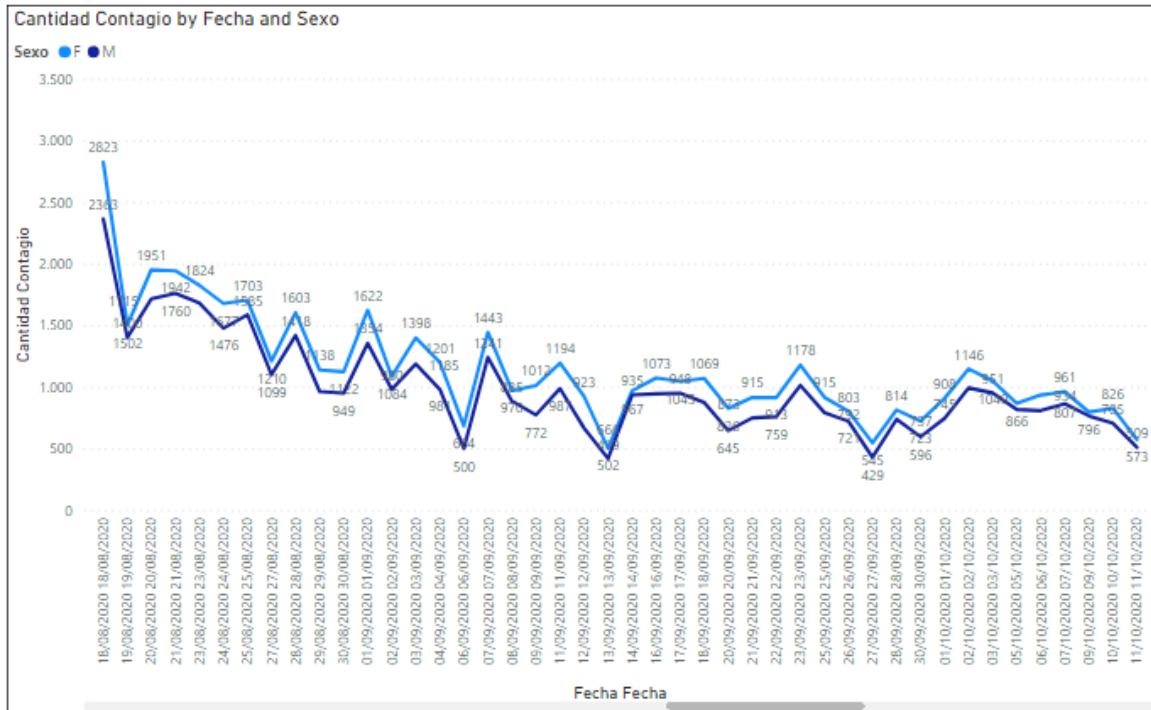


Imagen 33. Cantidad de hombres y mujeres contagiados por año 2020 D

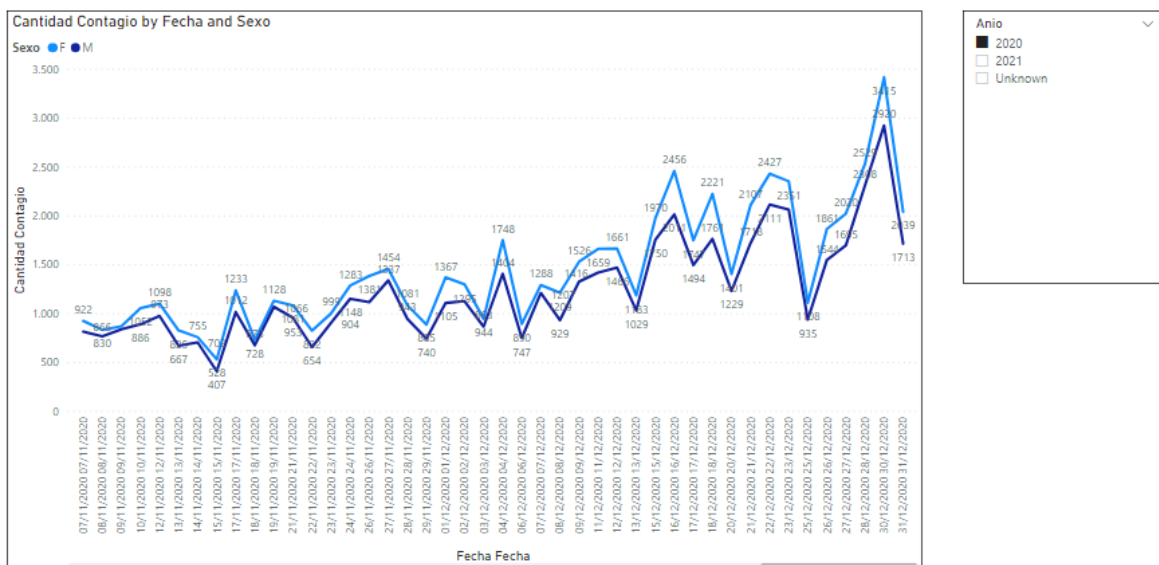


Imagen 34. Cantidad de hombres y mujeres contagiados por año 2020 F

### 19.1.1 Año 2021

Para lo corrido del año 2021 se observa que el mes Enero el número de contagios aumentan significativamente teniendo en cuenta la misma tendencia de contagios de mujeres y hombres observado en el año 2020.

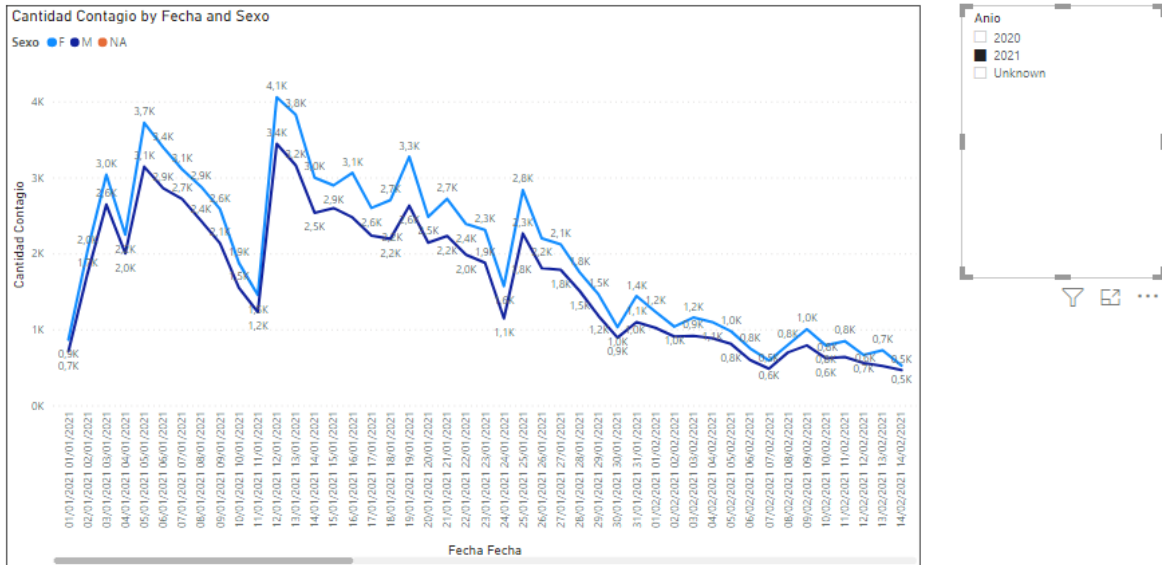


Imagen 35. Cantidad de hombres y mujeres contagiados por año 2021 A

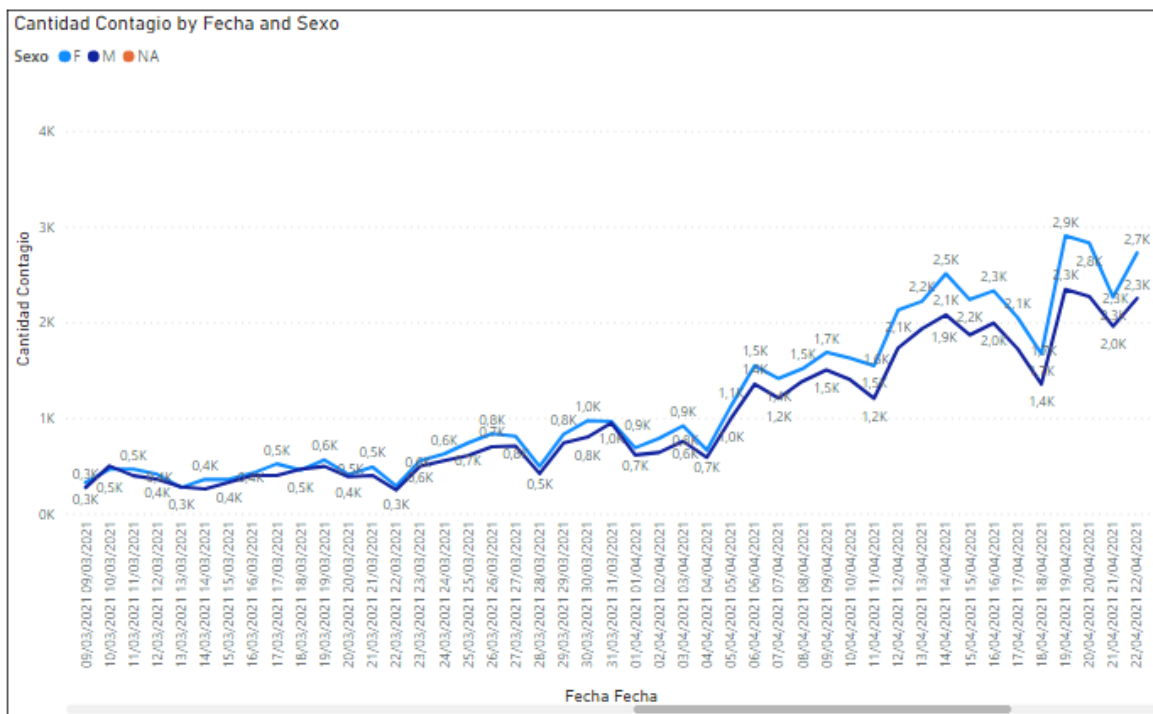


Imagen 36. Cantidad de hombres y mujeres contagiados por año 2021 B

## 26.1 CANTIDAD DE CASOS CONFIRMADOS POR LOCALIDAD TENIENDO EN CUENTA: MES, TRIMESTRE Y AÑO.

### 20.1.1 Día de la semana año 2020

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2020 por día de la semana, en la cual se observa que la localidad de Suba presenta el mayor número de contagios teniendo el 14,77% de estos en el día martes, seguida por la localidad de Kennedy con 13,44% el día lunes, y la tendencia es que dichos días es donde ocurren el mayor número de contagios.

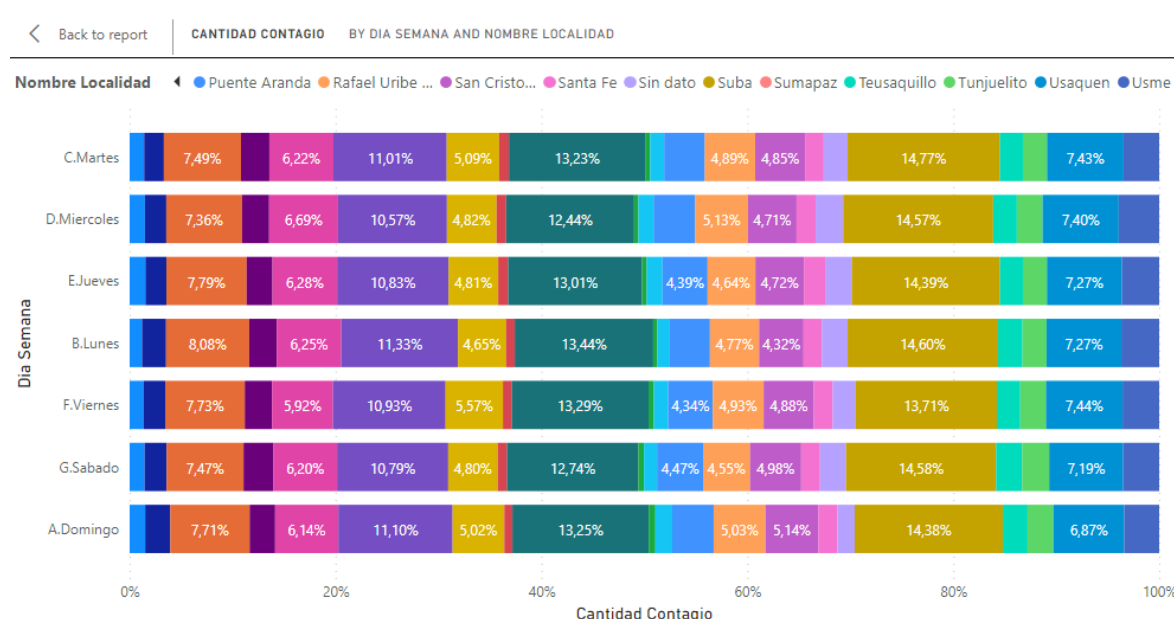


Imagen 37. Cantidad de casos confirmados por localidad y día de la semana año 2020

### 21.1.1 Día de la semana año 2021

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2021 por día de la semana, en la cual se observa que la localidad de Suba presenta el mayor número de contagios teniendo el 15,28% de estos en el día Domingo, seguida por la localidad de Kennedy con 12,94 el día Domingo.



Imagen 38. Cantidad de casos confirmados por localidad y día de la semana año 2021

### 22.1.1 Semana del año 2020

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2020 por semana, en la cual se observa que las localidades presentan su mayor número de contagios en la semana 33.

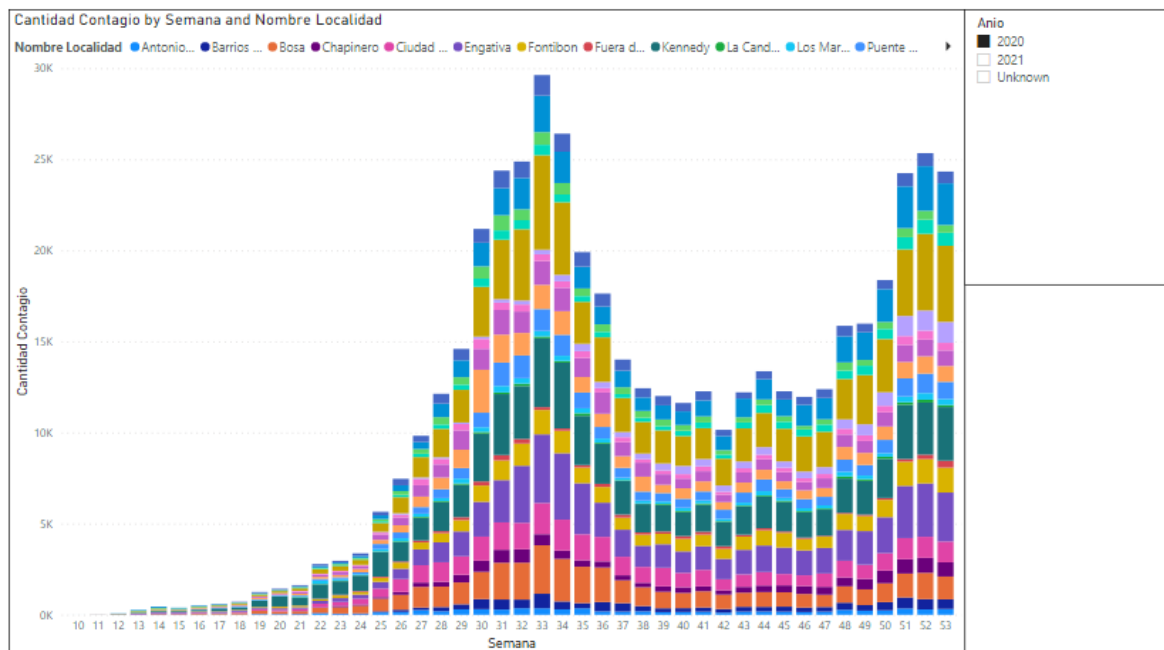


Imagen 39. Cantidad de casos confirmados por LOCALIDAD Y semana año 2020

### 23.1.1 Semana del año 2021

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2020 por semana, en la cual se observa que las localidades presentan su mayor número de contagios en la semana 33.

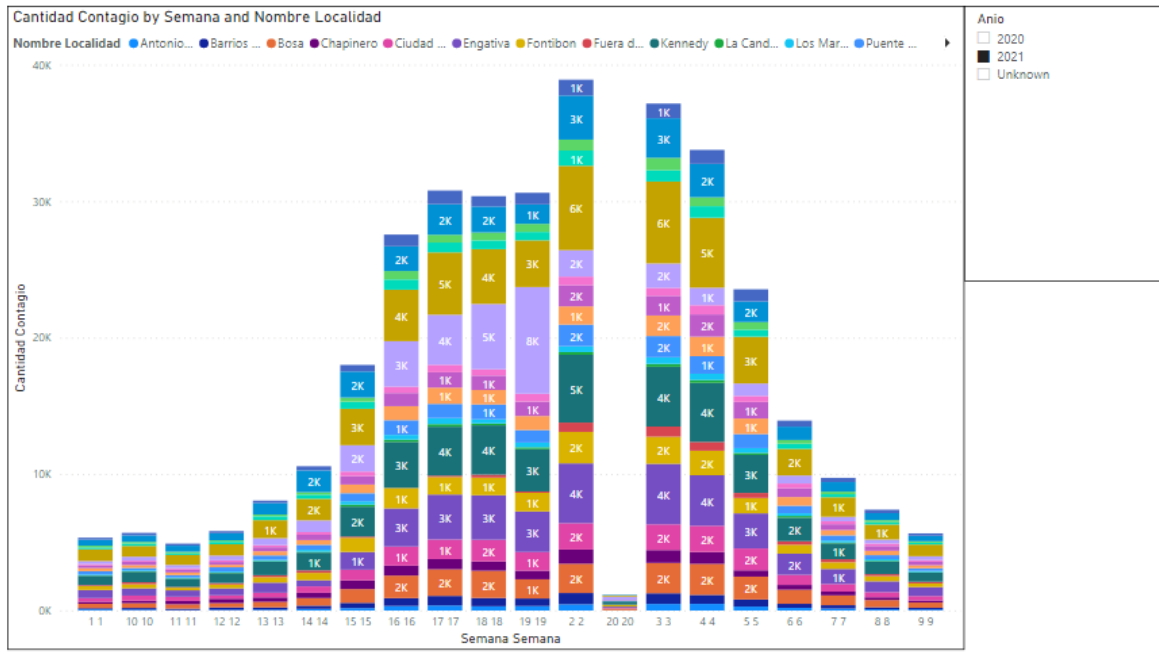


Imagen 40. Cantidad de casos confirmados por LOCALIDAD Y semana año 2021

### 24.1.1 Mes del año 2020

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2020 por mes, en la cual se observa que las localidades presentan su mayor número de contagios en el mes de agosto, de las cuales la localidad de Suba es la que más casos presenta, seguida por la localidad de Kennedy.

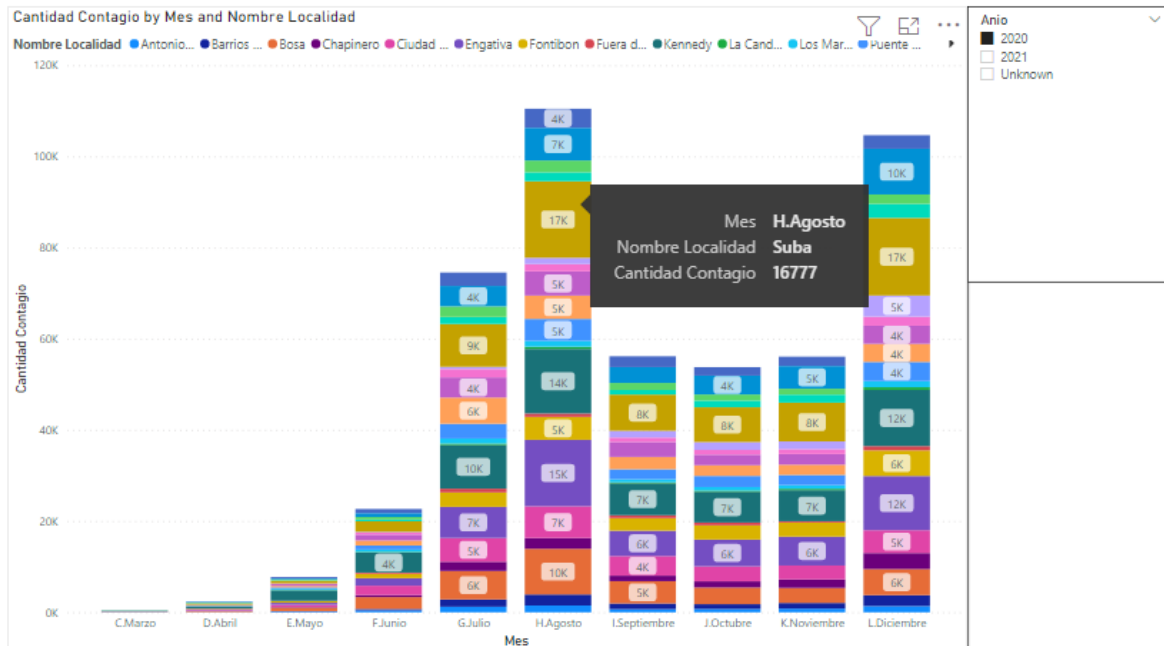


Imagen 41. Cantidad de casos confirmados por LOCALIDAD Y mes año 2020

### 25.1.1 Mes del año 2021

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2021 por mes, en la cual se observa que las localidades presentan su mayor número de contagios en el mes de enero, de las cuales la localidad de Suba es la que más casos presenta, seguida por la localidad de Kennedy.

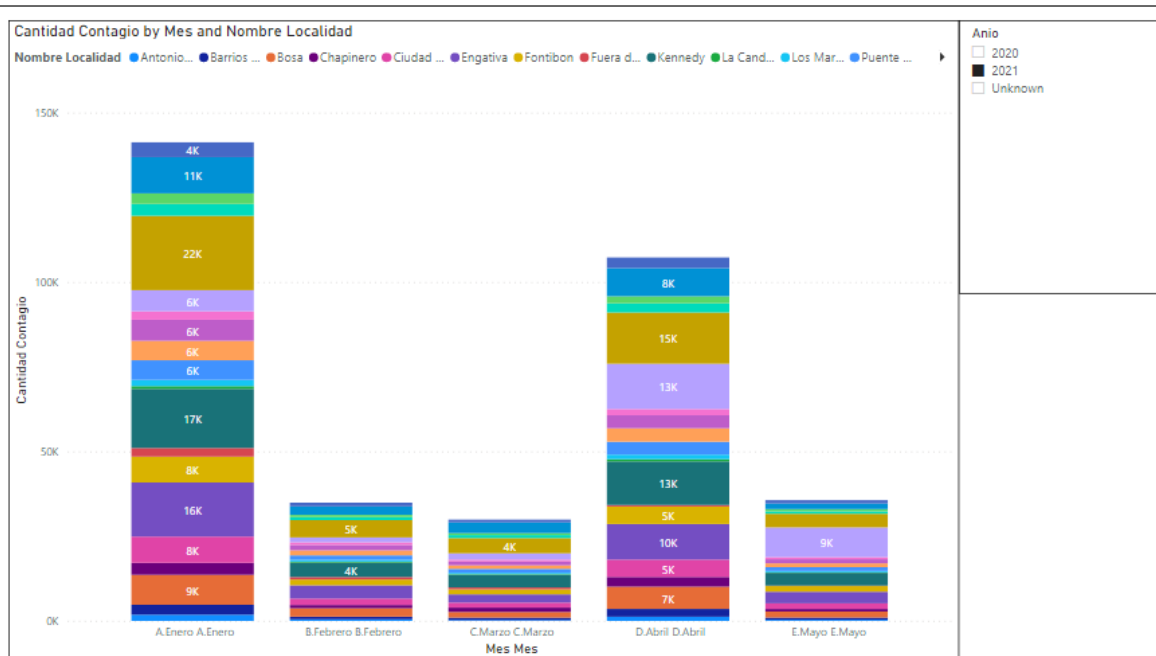


Imagen 42. Cantidad de casos confirmados por localidad y mes año 2021

### 26.1.1 Trimestre año 2020

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2020 por trimestre, en la cual se observa que las localidades presentan su mayor número de contagios en el trimestre 3, de las cuales la localidad de Suba es la que más casos presenta, seguida por la localidad de Kennedy.



Imagen 43. Cantidad de casos confirmados por LOCALIDAD Y trimestre año 2020

### 27.1.1 Trimestre año 2021

En la siguiente gráfica se presenta el comportamiento del Covid por localidades en el año 2021 por trimestre, en la cual se observa que las localidades presentan su mayor número de contagios en el trimestre 1, de las cuales la localidad de Suba es la que más casos presenta, seguida por la localidad de Kennedy y Engativá, para el trimestre 2 datos sin localidad presenta el mayor número de contagios seguido por Suba.



Imagen 44. Cantidad de casos confirmados por localidad y trimestre año 2021

### 28.1.1 Año 2020

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2020, en la cual se observa que la localidad que presentó mayor número de contagios fue Suba con un total de 70601 contagios, seguido por Kennedy con 63810 contagios y Engativá con 53446 casos contagiados.

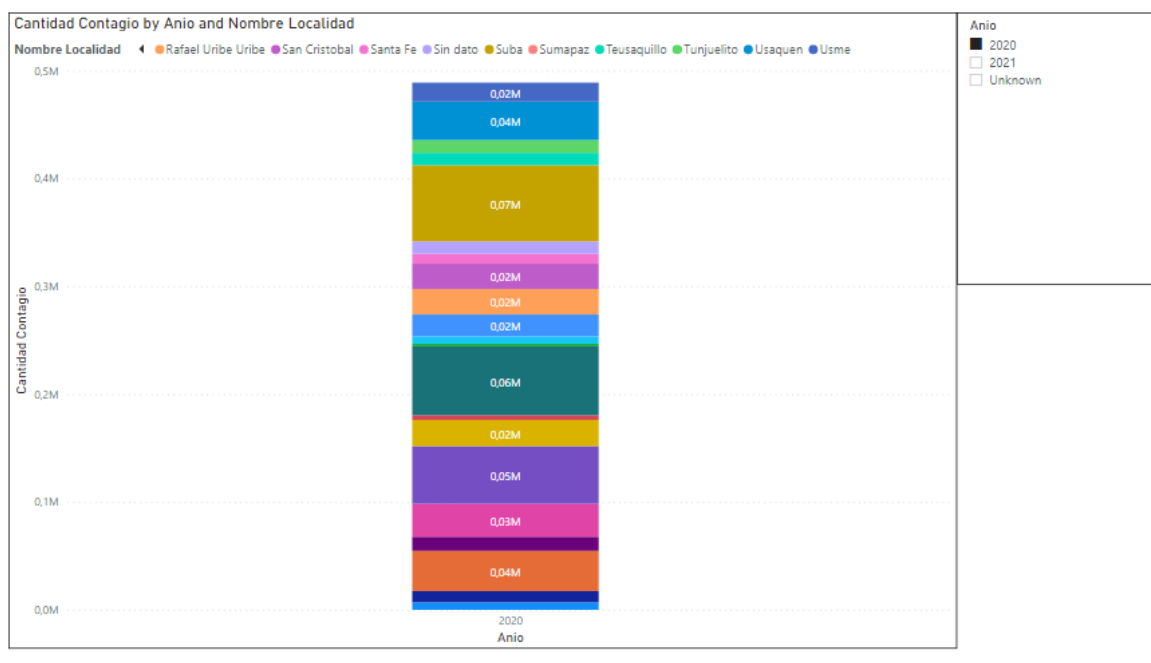


Imagen 45. Cantidad de casos confirmados por localidad año 2020



### 29.1.1 Año 2021

En la siguiente gráfica se presenta el comportamiento del covid por localidades en el año 2021, en la cual se observa que la localidad que presentó mayor número de contagios fue suba con un total de 50664 contagios, seguido por Kennedy con 42100 contagios y Engativá con 36252 casos contagiados.

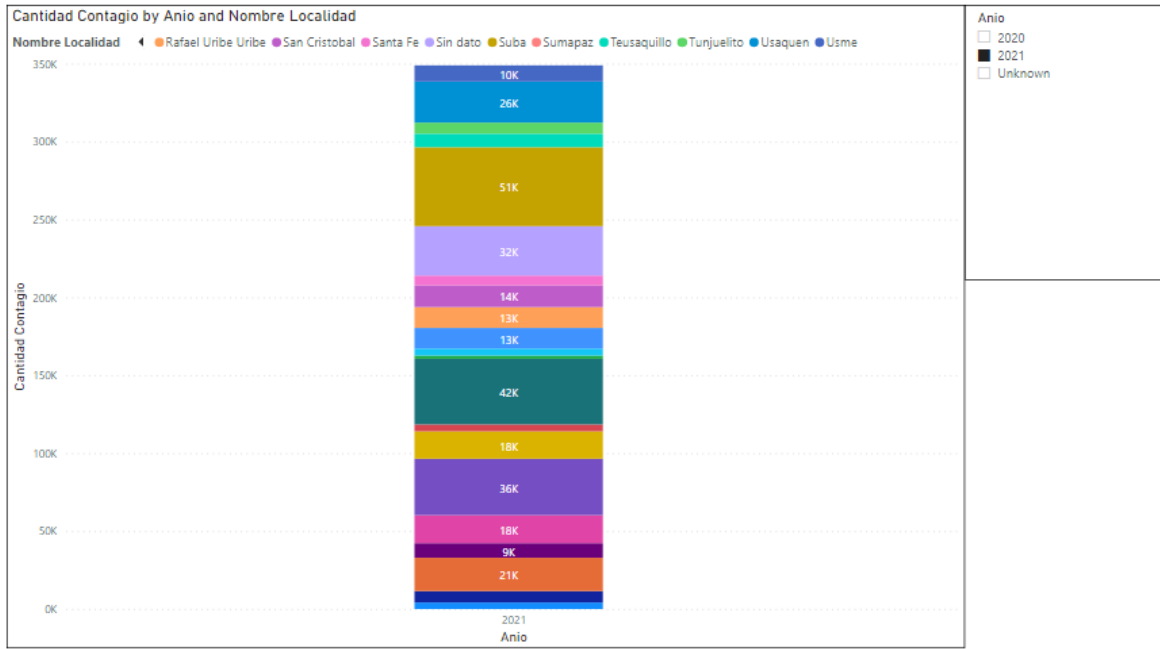


Imagen 46. Cantidad de casos confirmados por localidad año 2021

## 27.1 CANTIDAD DE PERSONAS FALLECIDAS: DÍA DE LA SEMANA, SEMANA, MES, TRIMESTRE, AÑO.

### 30.1.1 Día de la semana año 2020

Para el año 2020 el día martes fue el día en donde hubo mayor número de fallecidos y el día domingo como el día con menos fallecidos como se observa en la siguiente imagen.

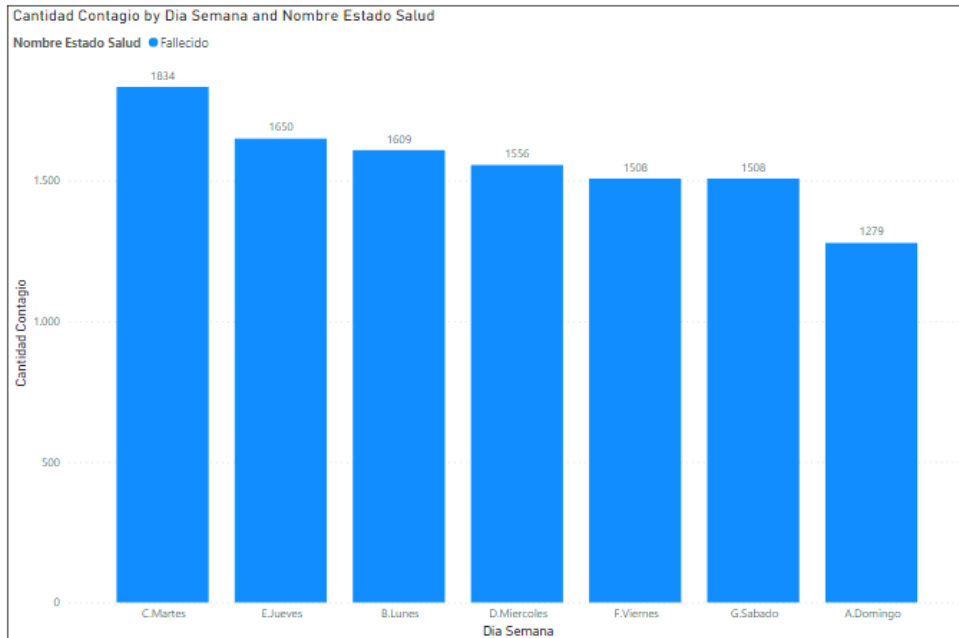


Imagen 47. Cantidad de personas fallecidas por día de la semana año 2020

### 31.1.1 Día de la semana año 2021

Para lo corrido del año 2021 continúa la tendencia del 2020; El día martes con más fallecidos y el día domingo con menos como se observa en la siguiente imagen.

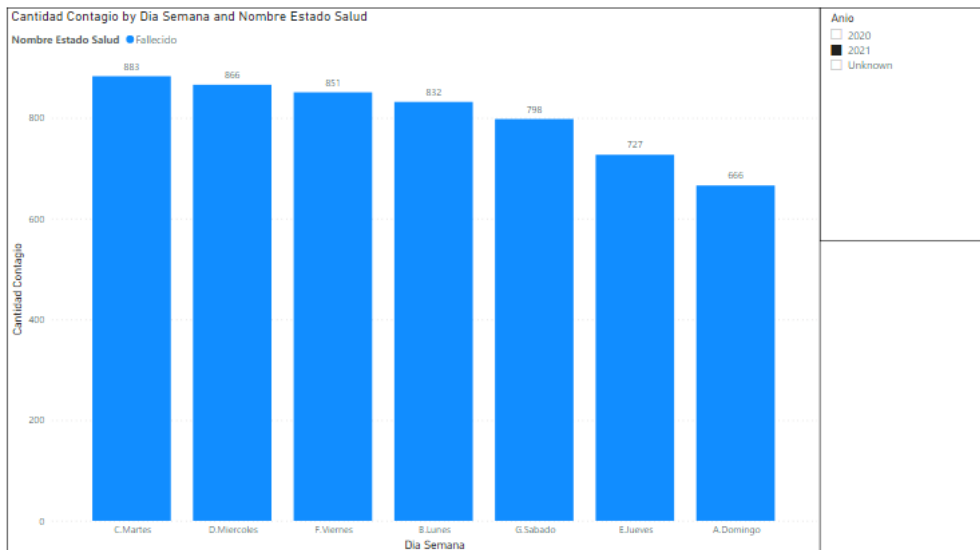


Imagen 48. Cantidad de personas fallecidas por día de la semana año 2021

### 32.1.1 Semana del año 2020

En las semanas 30 y 31 hubo más fallecidos y las semanas 12 y 13 del año 2020 hubo menos fallecidos como se observa en la siguiente imagen.

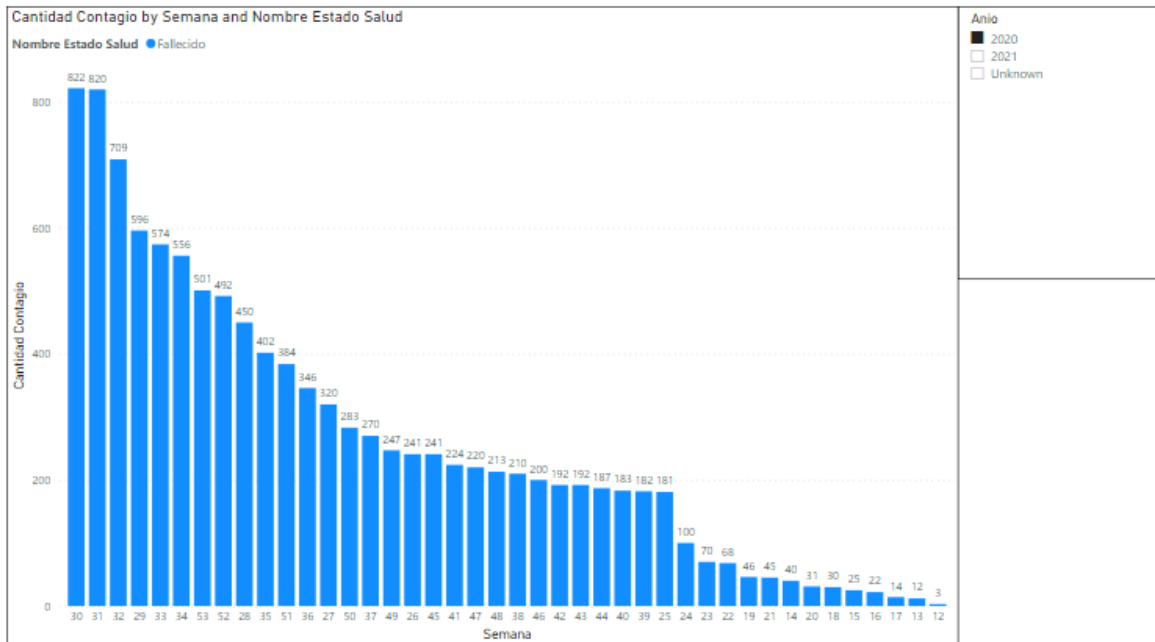


Imagen 49. Cantidad de personas fallecidas por semana año 2020

### 33.1.1 Semana del año 2021

Para lo corrido del año 2021 las semanas 2 y 3 hubo mayor número de fallecidos y en las semanas 10 y 11 tuvieron menos fallecidos como se observa en la siguiente imagen.

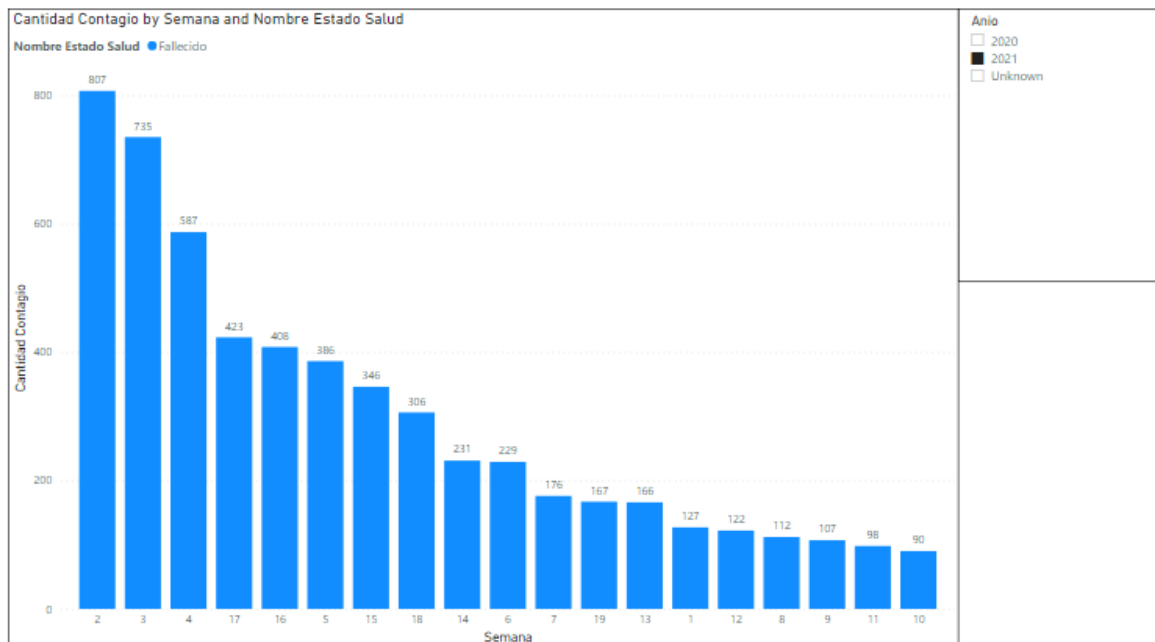


Imagen 50. Cantidad de personas fallecidas por semana año 2021

### 34.1.1 Mes del año 2020

En los meses de agosto, septiembre y diciembre del año 2020 hubo el mayor número de fallecidos con se observa en la siguiente gráfica, siendo agosto el mes en que más personas fallecieron.

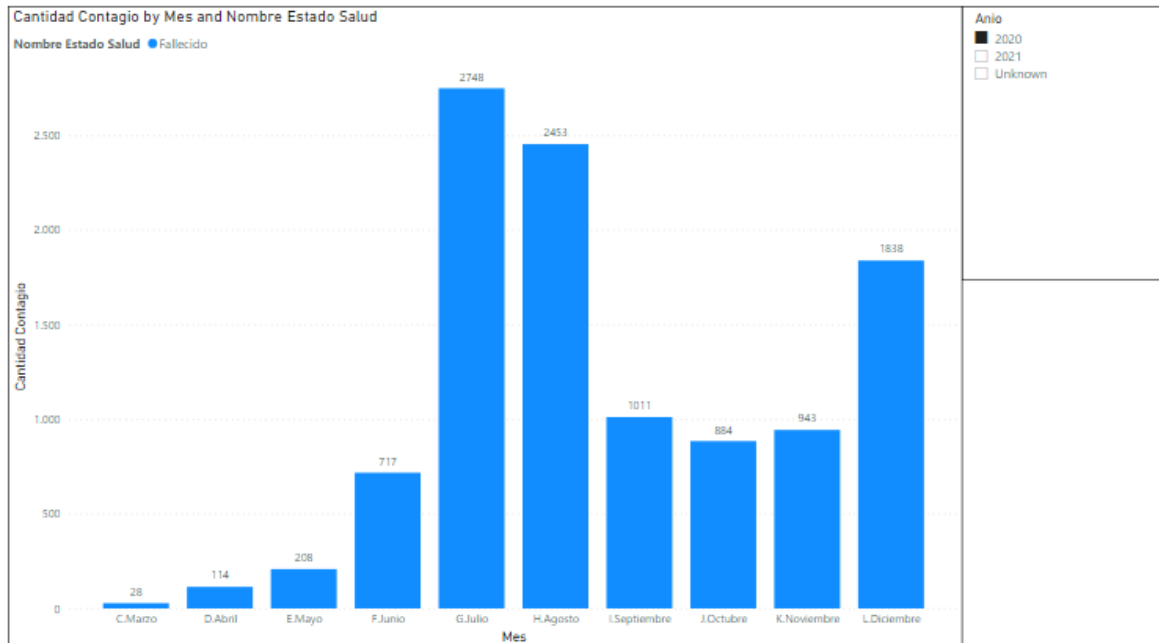


Imagen 51. Cantidad de personas fallecidas por mes año 2020

### 35.1.1 Mes del año 2021

Para lo corrido del año 2021 el mes de enero y abril fue en donde más fallecidos hubieron, como se aprecia en la siguiente imagen, siendo el mes de enero por el momento en donde más personas fallecieron.

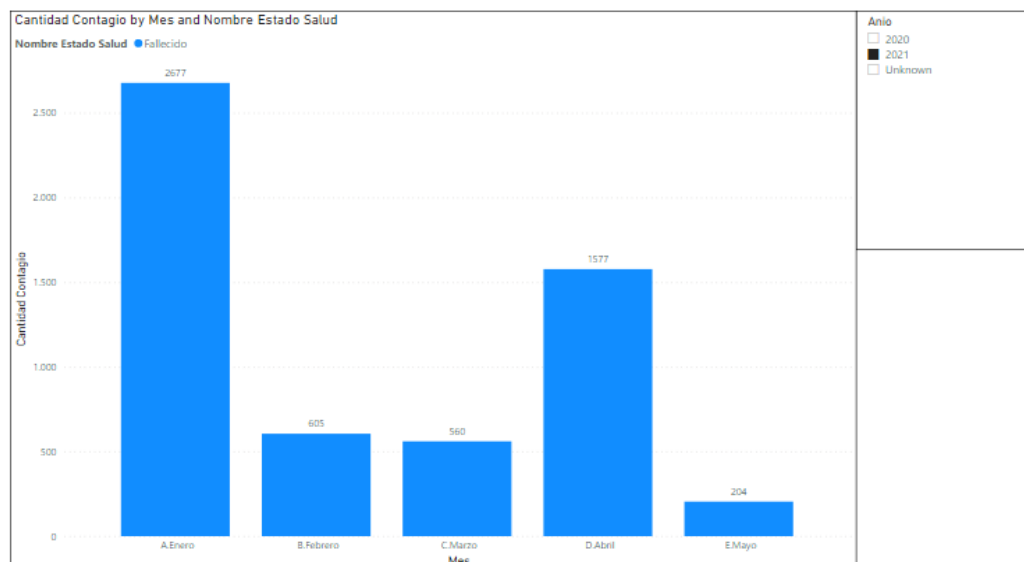


Imagen 52. Cantidad de personas fallecidas por mes año 2021

### 36.1.1 Trimestre año 2020

Para el tercer trimestre del año 2020 fallecieron más personas como se evidencia en la siguiente imagen, seguido del cuarto trimestre.

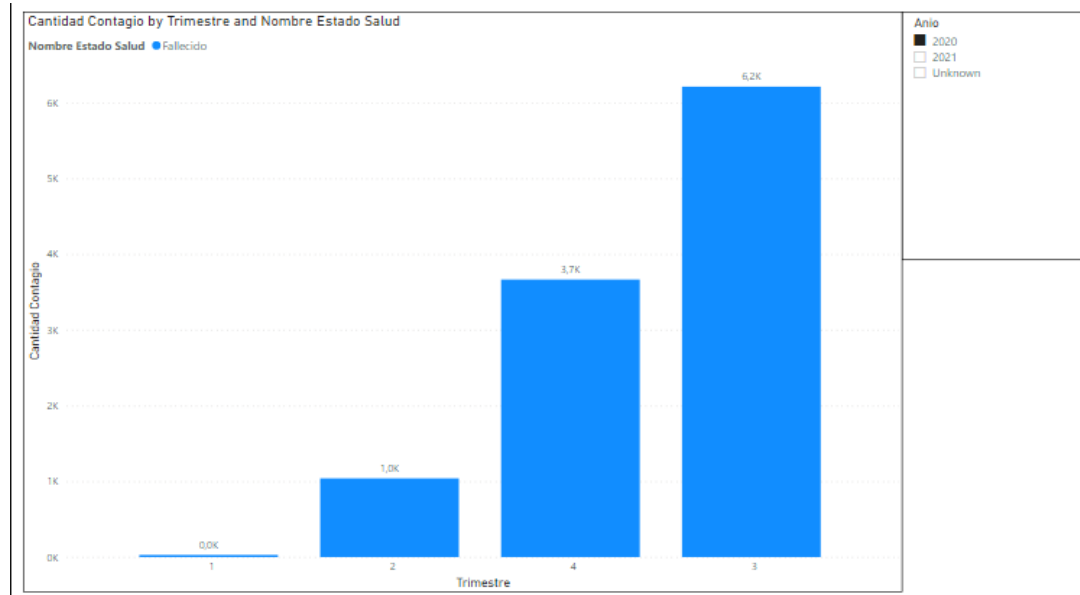


Imagen 53. Cantidad de personas fallecidas por trimestre año 2020

### 37.1.1 Trimestre año 2021

Para lo corrido del año 2021 el primer trimestre por el momento es en el cual ha habido más fallecidos como se observa en la siguiente imagen.

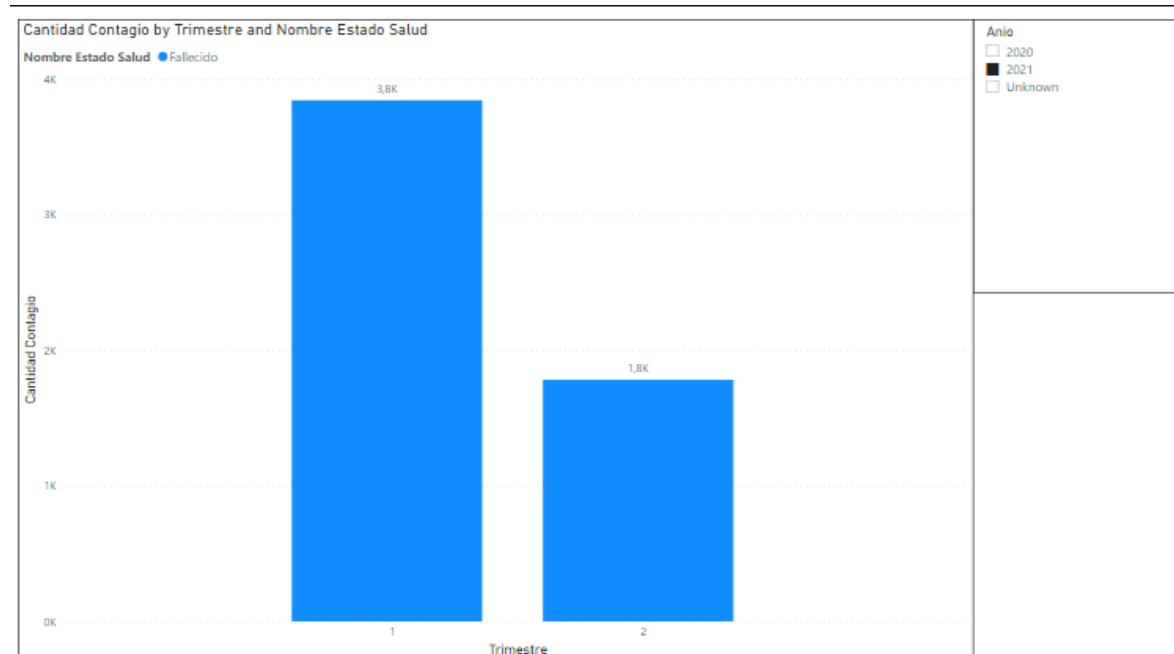


Imagen 54. Cantidad de personas fallecidas por trimestre año 2021

### 38.1.1 Año

Para el año 2020 hubo aproximadamente 10900 fallecidos y para lo corrido del año 2021 ha habido aproximadamente 5600 personas fallecidas, lo que equivale aproximadamente al 2% de los contagios, y de lo corrido del año 2021 va el 51,4% aproximadamente de los fallecidos respecto del año anterior como se evidencia en las siguientes imágenes.

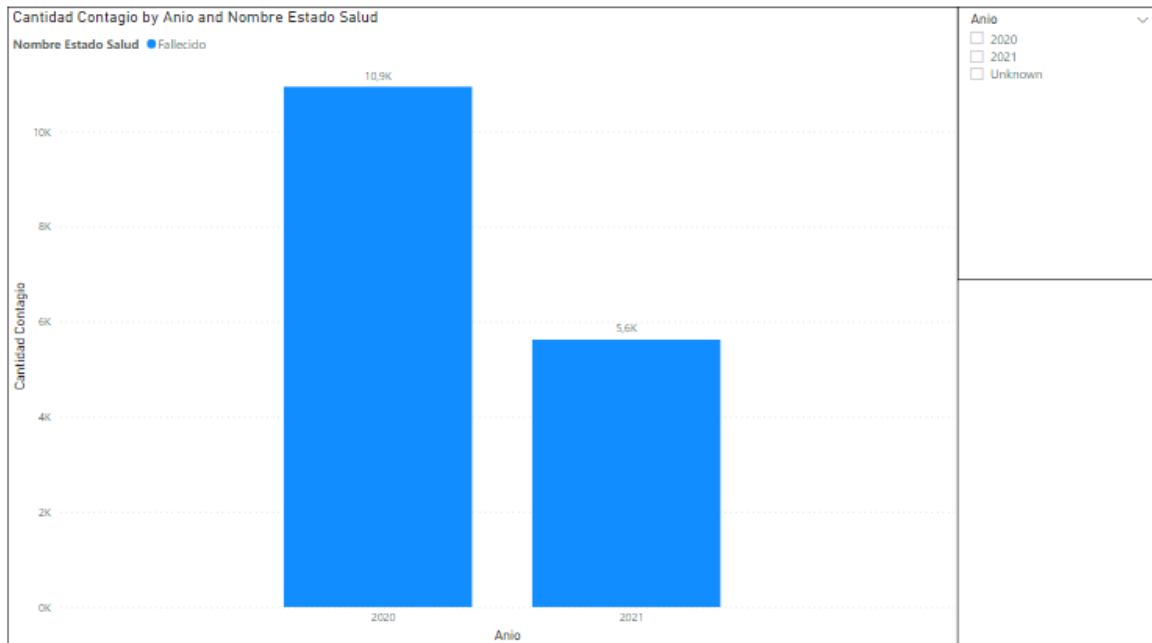
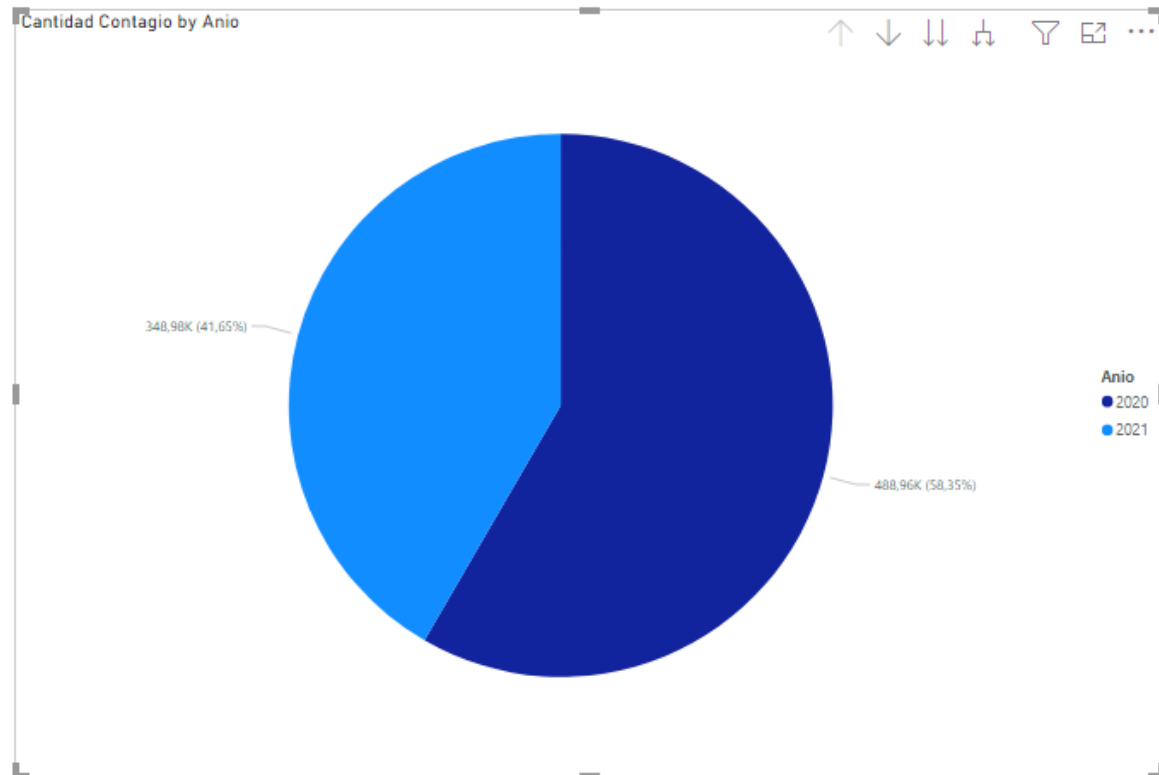


Imagen 55. Cantidad de personas fallecidas por año 2020 y 2021 A



*Imagen 56. Cantidad de personas fallecidas por año 2020 y 2021 B*

## 28.1 CANTIDAD DE CASOS RECUPERADOS: DÍA DE LA SEMANA, SEMANA, MES, TRIMESTRE, AÑO.

### 39.1.1 Día de la semana año 2020

Para el año 2020 el día martes fue el día en donde hubo mayor número de recuperados y el día domingo como el día con menos recuperados como se observa en la siguiente imagen.

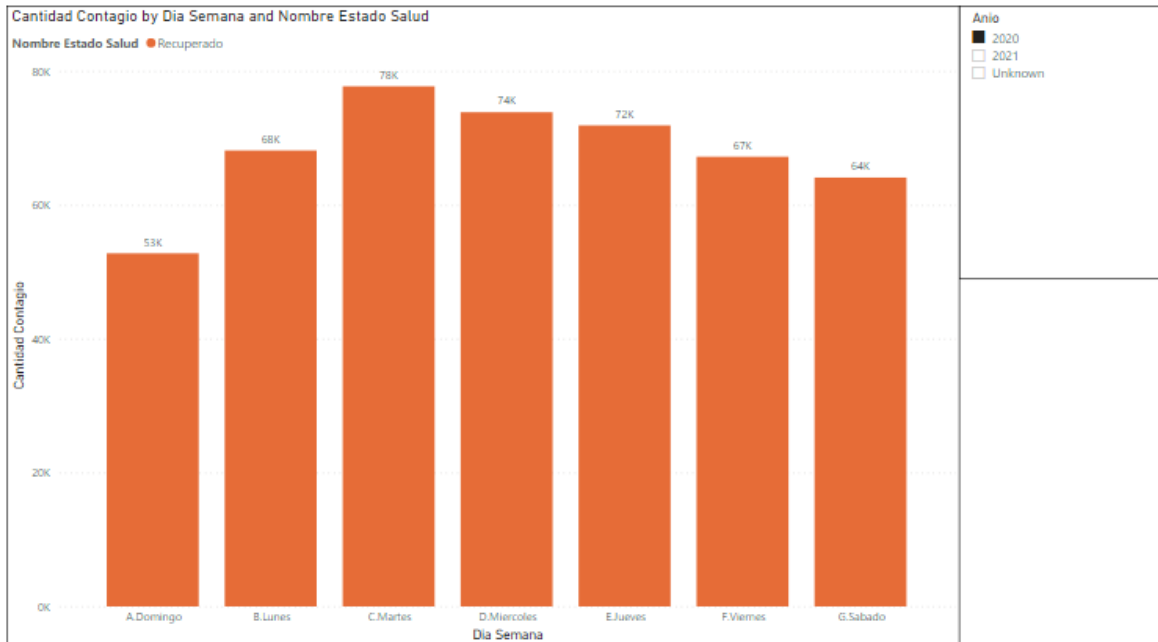


Imagen 57. Cantidad de casos recuperados por día de la semana año 2020

#### 40.1.1 Día de la semana año 2021

Para lo corrido del año 2021 continúa la tendencia del 2020; El día martes con más recuperados y el día domingo con menos como se observa en la siguiente imagen.

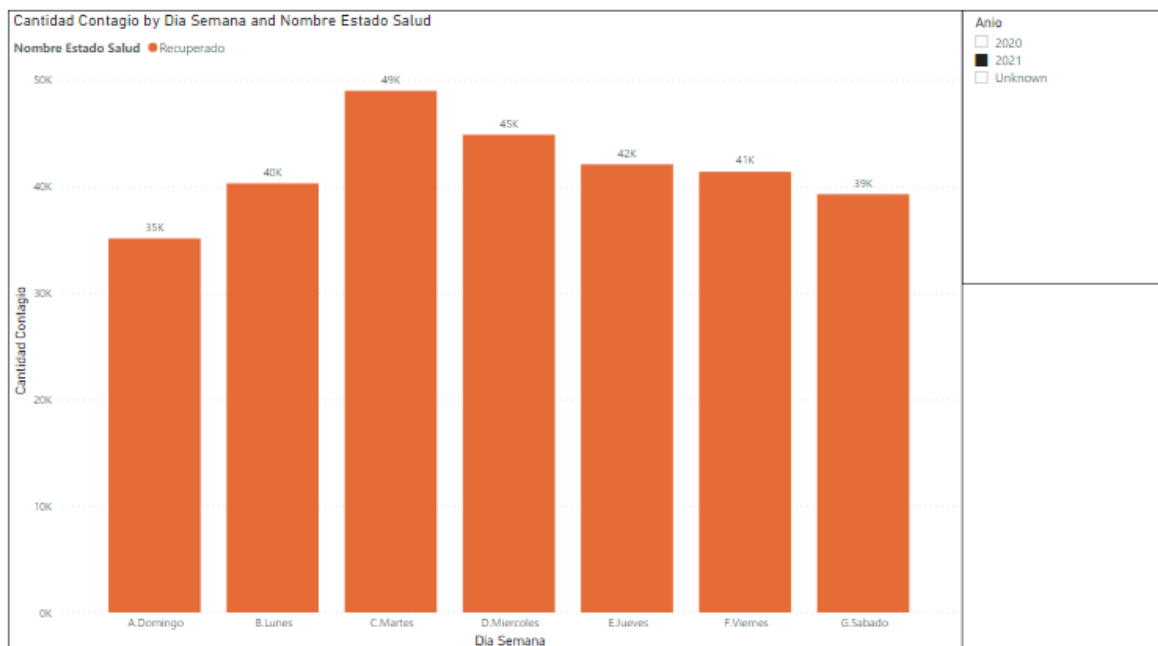


Imagen 58. Cantidad de casos recuperados por día de la semana año 2021



#### 41.1.1 Semana del año 2020

En las semanas 33 y 34 hubo más recuperados y las semanas 10 y 11 del año 2020 hubo menos recuperados como se observa en la siguiente imagen.

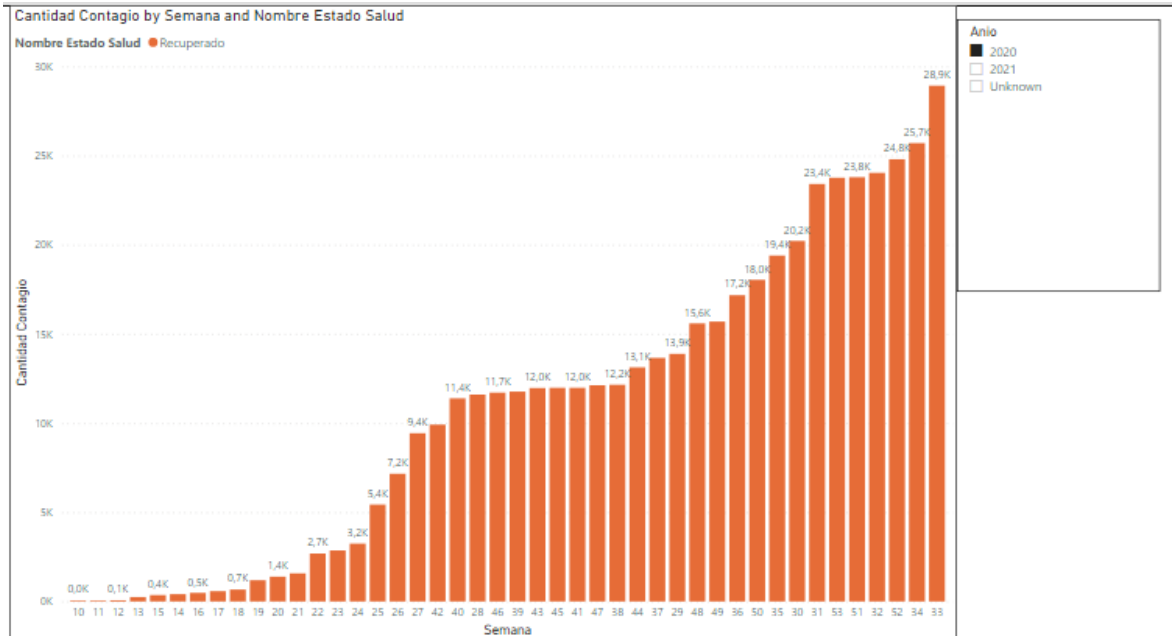


Imagen 59. Cantidad de casos recuperados por semana año 2020

#### 42.1.1 Semana del año 2021

Para lo corrido del año 2021 las semanas 2 y 3 hubo mayor número de recuperados y en las semanas 19 y 20 tuvieron menos recuperados como se observa en la siguiente imagen.

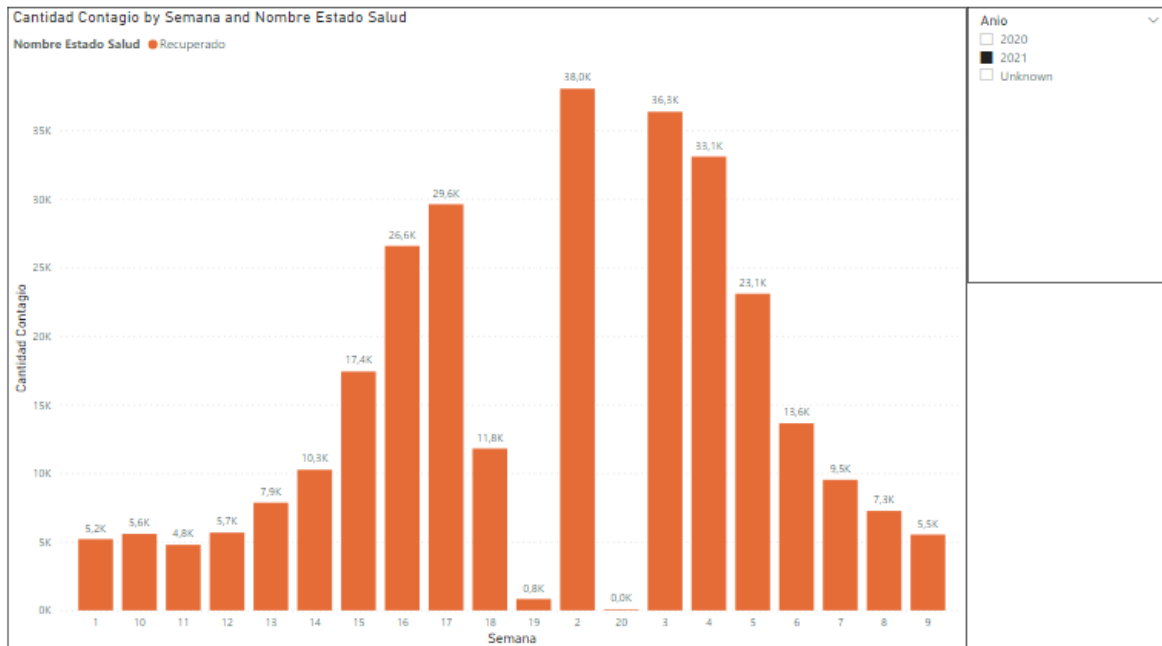


Imagen 60. Cantidad de casos recuperados por semana año 2021

#### 43.1.1 Mes del año 2020

En los meses de agosto, septiembre y diciembre del año 2020 hubo el mayor número de recuperados con se observa en la siguiente gráfica, siendo agosto el mes en que más personas se recuperaron.

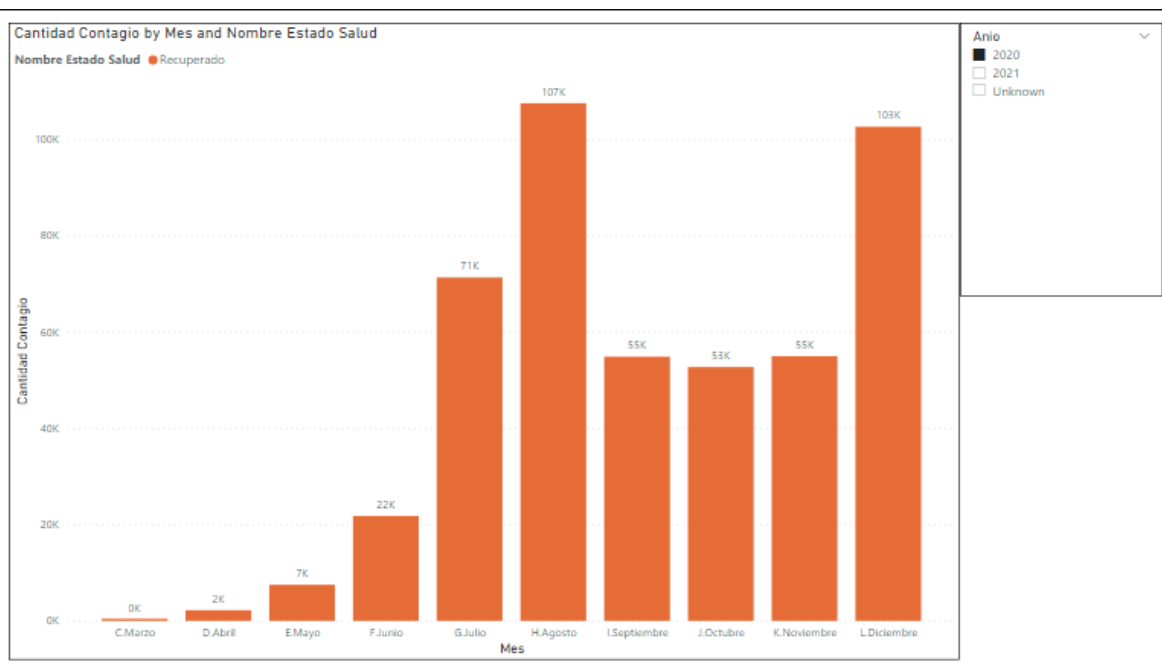
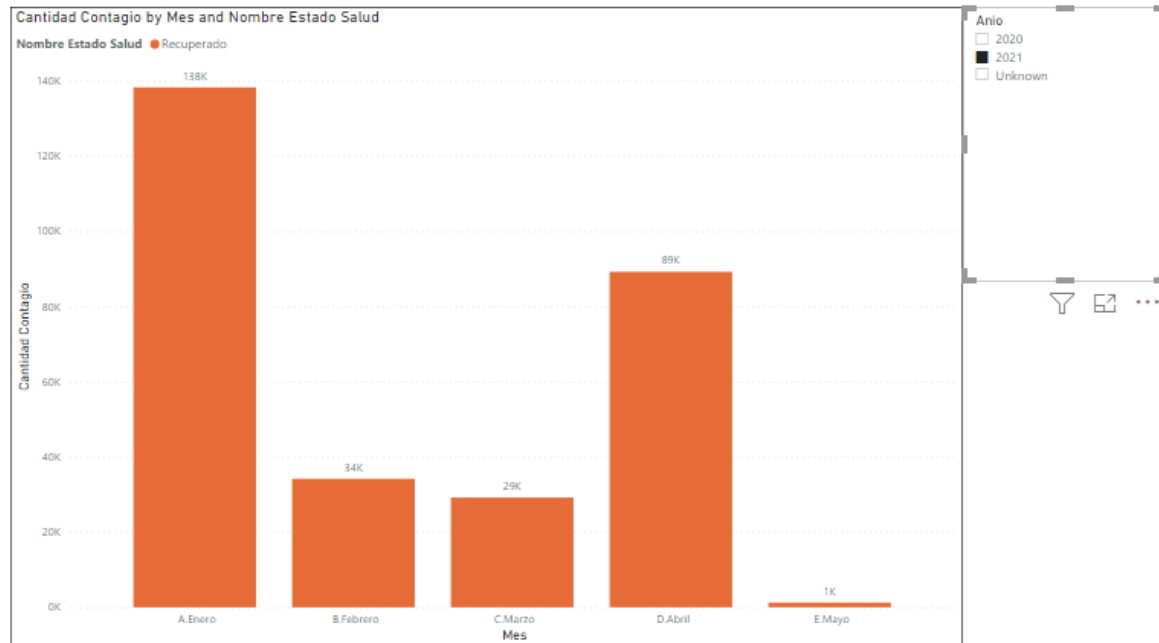


Imagen 61. Cantidad de casos recuperados por mes año 2020

#### 44.1.1 Mes del año 2021

Para lo corrido del año 2021 el mes de enero y abril fue en donde más recuperados hubieron, como se aprecia en la siguiente imagen, siendo el mes de enero por el momento en donde más personas se recuperaron.



*Imagen 62. Cantidad de casos recuperados por mes año 2021*

#### 45.1.1 Trimestre año 2020

Para el tercer trimestre del año 2020 se recuperaron más personas como se evidencia en la siguiente imagen, seguido del cuarto trimestre.

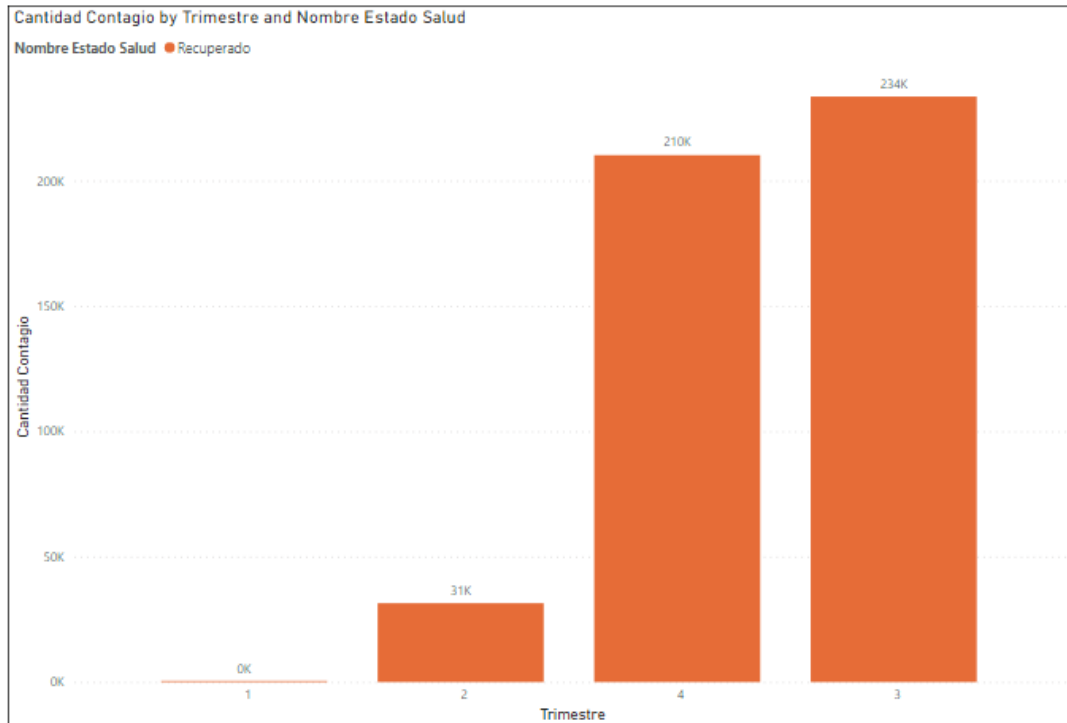


Imagen 63. Cantidad de casos recuperados por trimestre año 2020

#### 46.1.1 Trimestre año 2021

Para lo corrido del año 2021 el primer trimestre por el momento es en el cual ha habido más recuperados como se observa en la siguiente imagen.

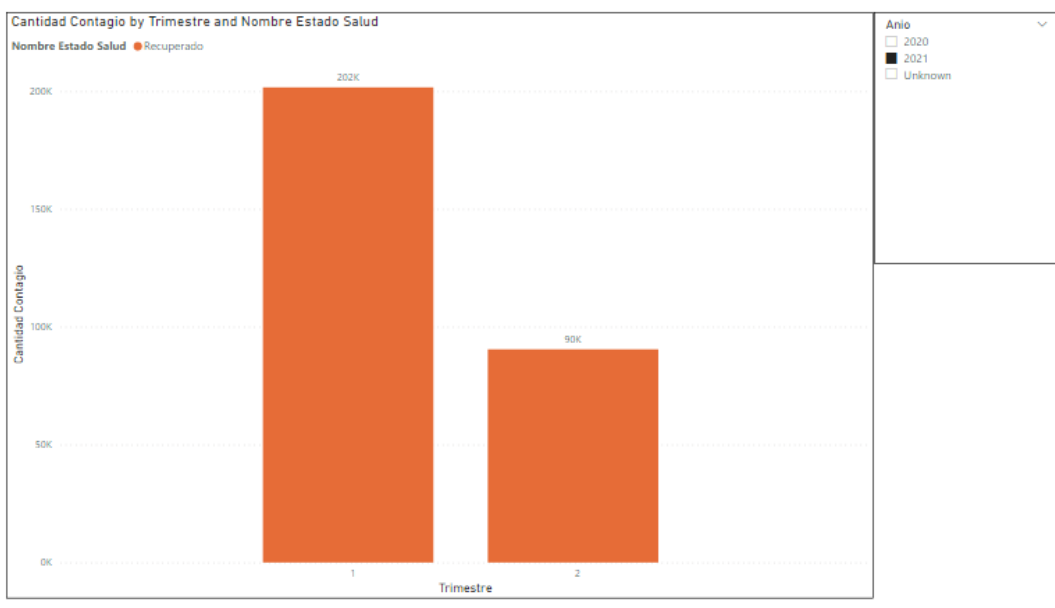


Imagen 64. Cantidad de casos recuperados por trimestre año 2021

### 47.1.1 Año

Para el año 2020 hubo aproximadamente 475686 recuperados y para lo corrido del año 2021 ha habido aproximadamente 292043 personas recuperadas, lo que equivale aproximadamente al 92% de los contagios, y de lo corrido del año 2021 va el 62% aproximadamente de los recuperados respecto del año anterior como se evidencia en las siguientes imágenes.

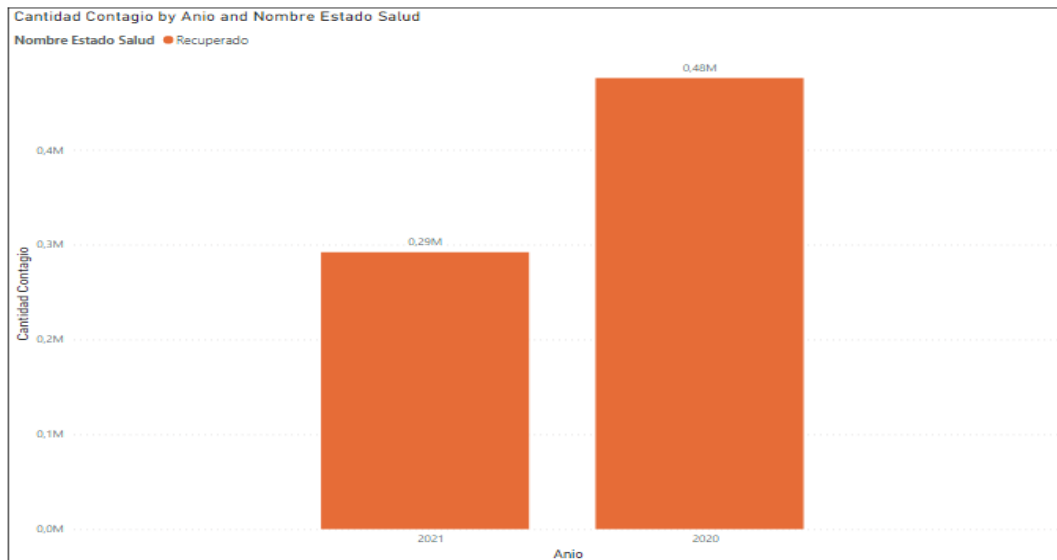
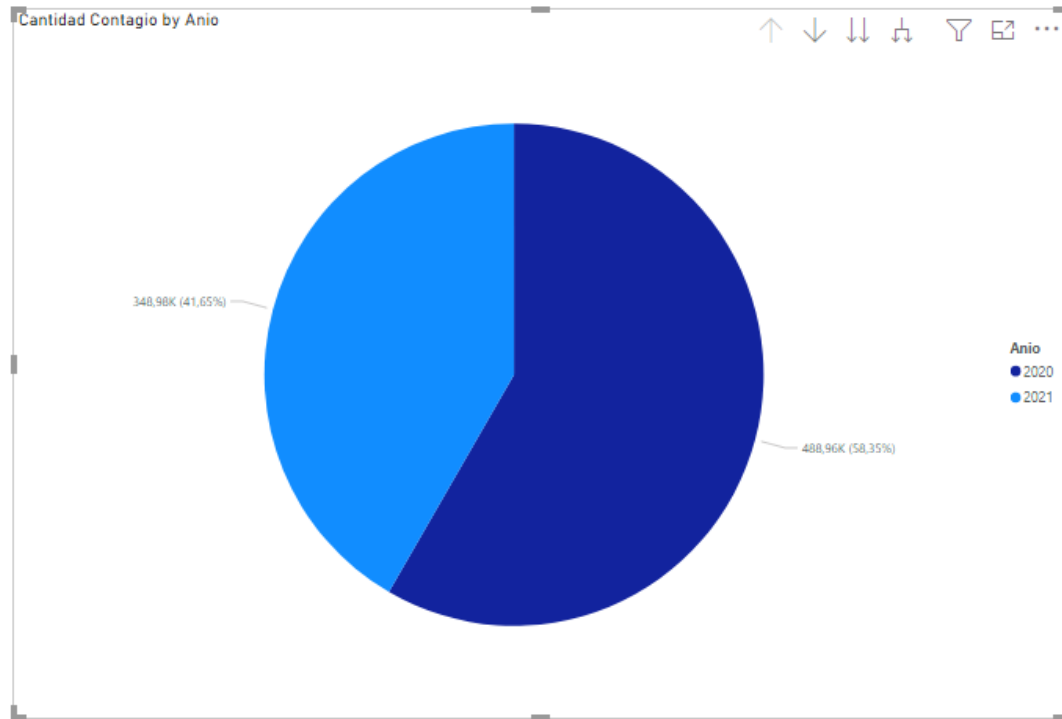


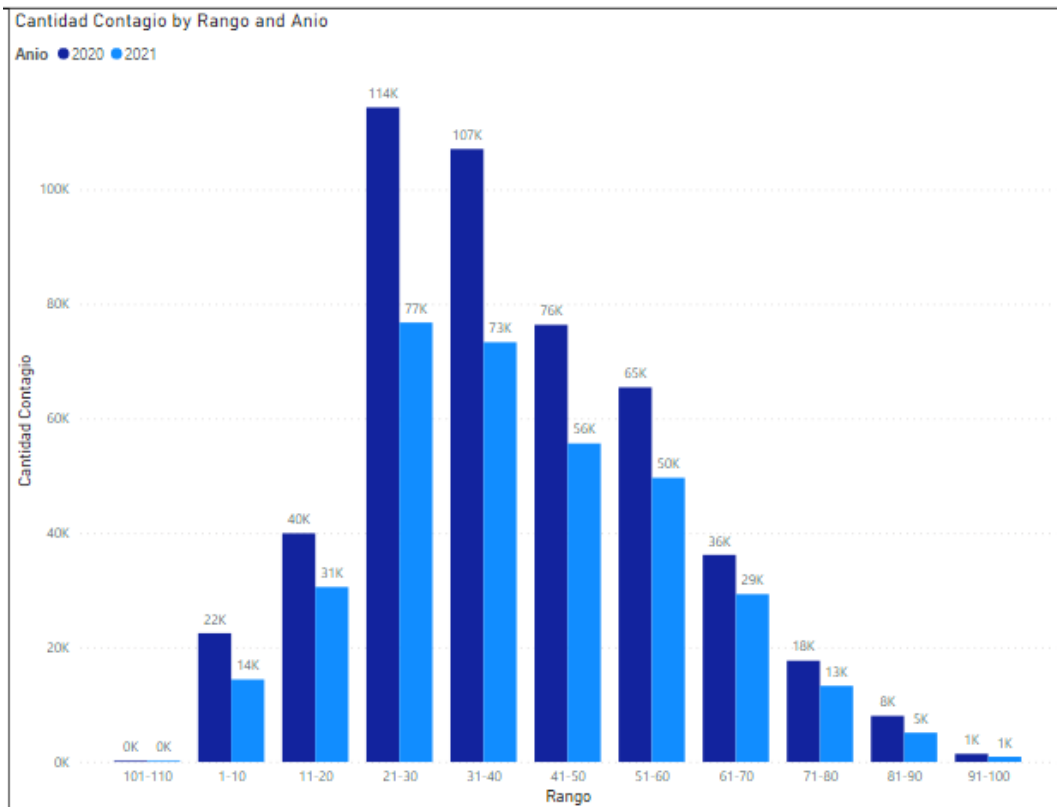
Imagen 65. Cantidad de casos recuperados por año 2020 y 2021 A



*Imagen 66. Cantidad de casos recuperados por año 2020 y 2021 B*

## 29.1 CONTAGIOS POR EDAD.

A continuación, se presenta la gráfica de rango de edades con más contagios entre el año 2020 y 2021, en la cual se puede observar que de 21 a 40 años fueron las edades donde más se presentaron contagios en lo corrido de los dos años.



*Imagen 67. Cantidad de contagios por edad año 2020 y 2021*

## 12. CONCLUSIONES

Obtener información de las necesidades del cliente directamente permite definir claramente el panorama de lo que se debe de realizar, es el punto de partida para empezar con la solución de minería de datos.

Los datos abiertos son una opción al momento de no tener un dataset por parte de un cliente, para el caso puntual de la recolección de datos de casos confirmados de covid-19 en la ciudad de Bogotá D.C se obtuvieron de datos abiertos de salud data.

Una estrategia para la implementación de un Data mart es realizar un buen diseño del modelo de datos a implementar.

Tras realizar el ETL a los datos de casos confirmados de covid-19 se obtienen los datos limpios y con un formato para que puedan ser utilizados para la analítica y no presentar problemas en este proceso.

La implementación del Cubo OLAP facilita la analítica sobre un dataset.

La graficación de resultados permite una mejor comprensión de los datos.



### **13. 10.PROBLEMAS ENCONTRADOS DURANTE EL ESTUDIO**

- Problemas en la instalación y configuración de las herramientas debido a incompatibilidades de versiones.
- Mala calidad de los datos, mezclas entre mayúsculas y minúsculas, fecha inicio de los síntomas en null y fechas inconsistentes, lo cual dificultó el ETL de los datos.

## 14. BIBLIOGRAFÍA

Covida, Proyecto. 2020. "Boletín."

Díaz, Jairo Márquez. 2020. "Obra Está Bajo Una Licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional DOSSIER CUESTIONES BIOÉTICAS DE LA PANDEMIA COVID-19 Inteligencia Artificial y Big Data Como Soluciones Frente a La COVID-19 Artificial Intelligen."

Social, Ministerio de Salud y Protección. n.d. Situación de Colombia Frente a Otros Países Durante La Pandemia.

Microsoft-SqlServer. (2020). Descargas de SQL Server | Microsoft. <https://www.microsoft.com/es-es/sql-server/sql-server-downloads>

Microsoft. (2018). SQL Server Integration Services - SQL Server Integration Services (SSIS) | Microsoft Docs. Microsoft Docs. <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver15>

Microsoft. (2019a). Extract, transform, and load (ETL) - Azure Architecture Center | Microsoft Docs. <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>

Microsoft. (2019b). Overview of OLAP cubes for advanced analytics | Microsoft Docs. <https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2019>

Microsoft. (2019c). SQL Server Management Studio (SSMS) - SQL Server Management Studio (SSMS) | Microsoft Docs. 11/09/2019. <https://docs.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver15>

Microsoft. (2019d). What is Power BI? - Power BI | Microsoft Docs. Website. <https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>

Microsoft. (2020). What is Analysis Services? <https://docs.microsoft.com/en-us/analysis-services/analysis-services-overview?view=asallproducts-allversions>

Overview of Visual Studio | Microsoft Docs. (n.d.). Retrieved May 12, 2021, from <https://docs.microsoft.com/en-us/visualstudio/get-started/visual-studio-ide?view=vs-2019>