Analítica de Grandes Datos Departamento de Ciencias de la Computación y la Decisión

Facultad de Minas Universidad Nacional de Colombia

Trabajo Nociones de Arquitectura de la Información

Versión:2021.05.17 17:30

Observación: Cada vez que agregue nuevos elementos al documento, o que modifique algún componente del informe, revise la coherencia y consistencia con los otros elementos que hacen parte del mismo.

RESPONSABLES

Nombre Completo – Documento de Identificación				
1. Andres Felipe Ortega Lucero – 1082657705				
2. Daniel Arroyave Lopez – 1128448929				
3. Evelin Calderón Caro – 1152707862				
4. Rafael Eduardo Ropero Layton – 1118566826				
5. Viviana Posada Restrepo – 1033339284				
REPO EN GITHUB: https://github.com/viviposada/AGD2021-1S				

Realiza este trabajo considerando los datos que generan los sistemas transaccionales e información no estructurada de tu dominio (si trabajas por ejemplo para TCC tu dominio es la mensajería; también puedes explorar en la página https://www.kaggle.com/datasets o https://arxiv.org/). Considera tener acceso a esta información, de al menos 10 MB (puede ser uno o varios archivos de texto), y tener al menos cuatro clases conceptuales. Este documento también debe almacenarse en el REPO. Plazo Máximo de Entrega 23 de Mayo, NO SE recibirá por correo electrónico, envío por https://forms.gle/h7ty3yZykaUq5m7y6

1 COMPRENSIÓN DEL NEGOCIO

1.1 DESCRIPCIÓN DEL CONTEXTO DEL NEGOCIO.

Describa en máximo 250 palabras el contexto en el cual se generan los datos y cuál es el proceso que los genera.

El COVID-19 también conocido como SARS-CoV2 es una nueva enfermedad viral inicialmente identificada en la ciudad de Wuhan - China. La enfermedad llegó a Colombia a principios del mes de marzo del año 2020, alertando a las autoridades gubernamentales y sanitarias, lo que conllevo a declarar cuarentena por la vida en todo el país para los primeros 20 días del mes de marzo del año 2020 con el fin de mitigar la propagación del virus. Dicho periodo de tiempo fue usado para preparar a las autoridades de salud sobre el protocolo y las medidas a implementar con el fin de proteger la salud pública, para ello, el gobierno

nacional liberó una aplicación en la cual un miembro o varios de cada familia registraron su estado de salud, permitiendo así realizar seguimiento de los posibles focos del virus. Con el paso del tiempo y en inicio de los contagios, el ministerio de Salud designó un área para realizar seguimiento a las personas que presentaron posibles contactos con contagiados y así construir un cerco epidemiológico, que permitiera mitigar la propagación del virus. Mediante estos seguimientos el ministerio de salud captura información valiosa para los entes gubernamentales, la cual, ayuda a la toma de decisiones frente al comportamiento del contagio del virus, nuestro énfasis será el departamento de Antioquia, el cual al mes de abril del año 2021, presenta la mayor tasa de ocupación de UCI y niveles de contagio.

1.2 | IDENTIFICACIÓN DEL PROBLEMA:

Delimite en máximo 150 palabras la problemática, así como identificar los requisitos, supuestos, restricciones y beneficios de la solución de este.

La tasa de contagios del virus en el departamento de Antioquia ha ido incrementando de manera exponencial. Sin embargo, la obtención de la información ha sido lenta debido al difícil acceso de los datos y limitado contacto con los pacientes, información desactualizada y, en algunas ocasiones, fallos en los procesos de ingreso de la información; por ello es necesario que se implemente como requisito un correcto registro, sistematización y seguimiento de las personas infectadas en el departamento, en pro de establecer las actividades prioritarias para prevenir, responder y mitigar oportunamente la propagación del virus.

Entre los beneficios se espera determinar, específicamente en el departamento de Antioquia, la evolución de los contagios por municipio, género y rango de edad con el fin de establecer medidas preventivas y correctivas dirigidas a la población con mayor susceptibilidad.

El desarrollo de este problema está basado bajo el supuesto que no se presentan falsos positivos, la información obtenida está limpia, es verídica y se encuentra actualizada.

1.3 DETERMINACIÓN DE OBJETIVOS:

Describa en máximo 150 palabras las metas a lograr al proponer una solución basada en un modelo de datos o de analítica (cómo y qué tipo de ventaja competitiva se ganará).

Nuestro objetivo general es identificar la evolución de contagios del virus en la población colombiana, específicamente para el departamento de Antioquia durante el año 2020.

Es vital para nosotros analizar las zonas (municipios) del departamento que presentan una mayor tasa de contagio con el fin de proveer la información a los diferentes entes gubernamentales y de salud, para que se tomen medidas correctivas y preventivas en pro de disminuir la población afectada.

Adicionalmente es necesario establecer los rangos de edades donde el virus afecta de manera crítica, y así determinar una posible tasa de muerte basados en los datos recolectados hasta la fecha.

1.4 EVALUACIÓN DE LA SITUACIÓN ACTUAL:

Describa en máximo 150 palabras el estado actual antes de implementar la solución de analítica, a fin de tener un punto de comparación que permita medir el grado de éxito de la solución.

El COVID-19 ha tenido una variación significativa a nivel mundial con la aparición de nuevas cepas que han alterado el nivel de contagios, el gobierno departamental no cuenta con todos los insumos necesarios para el correcto manejo de la pandemia, además, en los últimos meses se ha evidenciado que la población no ha seguido de manera rigurosa los protocolos de bioseguridad, por esto, el mejor método encontrado hasta ahora para frenar el alto índice de contagios en el departamento de Antioquia ha sido la implementación de toques de queda y cuarentenas sectorizadas que permitan detener la propagación del virus.

Por esta razón, es importante evaluar los datos obtenidos desde la llegada del virus al país, con finalidades de entender su comportamiento, impacto por edad, género y ubicación geográfica, que permita determinar nuevas medidas de cuidado, prever y prepararse para posibles picos y determinar las zonas con alto índice de contagio.

2 COMPRENSIÓN DE LOS DATOS

2.1 RECOLECCIÓN DE DATOS

Describa en máximo 150 palabras los datos a utilizar identificando las fuentes, las técnicas empleadas en su recolección, los problemas encontrados en su obtención y la forma como se resolvieron los mismos. Además, adjunte los datos (archivos de texto, etc.) agréguelos en el github https://github.com/viviposada/AGD2021-15 en un solo archivo, por favor comprímalo(s). Llame el archivo T1.2.1.Datos.zip

La plataforma de Datos Abiertos del Gobierno Nacional de Colombia fue creada con el fin de que todas las entidades del sector público publiquen la información pertinente y de calidad a disposición de los usuarios. Esta información se comparte mediante la plataforma de datos gov mediante formatos digitales estandarizados. Entre las base de datos publicadas, se encuentran los datos de la evolución del COVID-19 en el país, a partir de la cual se obtuvo el seguimiento del comportamiento de este virus para el departamento de Antioquia entre marzo y noviembre del 2020. Entre los problemas identificados se encontró que la cantidad de casos superó el límite máximo disponible en algunos sistemas, por lo que fue necesario realizar ajustes en la información disponible en la plataforma. Otro de los posibles problemas es la actualización de la información (veracidad) por lo que es necesario establecer la fecha de corte del análisis a realizar.

Fuente: https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data

2.2 DESCRIPCIÓN DE DATOS (DICCIONARIO):

Diligencia la siguiente tabla, puede agregar otra columna si lo considera necesario.

Nombre Tabla	Nombre Atributo	Tipo Dato	Descripción
GrupoEtnico	ID_GrupoEtnico	int	Identificador único del grupo étnico al que pertenece la persona. Esta variable depende totalmente del correcto diligenciamiento por los profesionales de salud que notifican en más de 10.000 instituciones de salud en todos los municipios y departamentos, del autorreconocimiento de la persona cuando se le pregunta por esta variable y del listado censal que haga y mantenga actualizado cada departamento.
GrupoEtnico	DescripcionGrupoEtnico	varchar(20)	Nombre de los grupos étnicos, se pueden clasificar en 1-Indígena, 2-ROM, 3-Raizal, 4- Palenquero, 5-Negro y 6-Otro
Personas	ID_Caso	bigint	Identificador único del caso asignado a la persona
Personas	ID_Geografia	bigint	Identificador único de la ubicación de la persona
Personas	ID_RangoEdad	int	Identificador único de la edad de la persona
Personas	ID_GrupoEtnico	int	Identificador unico del grupo etnico de la persona
Personas	Genero	varchar(1)	Género de la persona, se puede clasificar en F-Femenino, M-Masculino,O-Otro
Personas	Edad	int	Número de la edad
RangoEdad	ID_RangoEdad	int	Identificador único del rango de edad

RangoEdad	DescripcionRangoEdad	varchar(20)	La edad de la persona se puede clasificar en los siguientes rangos: 1-De 0 a 2 años, 2-De 2 a 4 años, 3- De 4 a 6 años, 4-De 6 a 8 años, 5-De 8 a 10 años, 6-De 10 a 14 años, 7-De 14 a 18 años, 8-De 18 a 22 años, 9-De 22 a 35 años, 10-De 35 a 45 años, 11-De 45 a 55 años, 12-De 55 a 65 años, 13-De 65 a 75 años, 14-De 75 a 85 años, 15-De 85 a 95 años, 16-De 95 en adelante
Geografia	ID_Geografia	bigint	Identificador único de la ubicación de la persona por municipio
Geografia	CodigoDepartamento	bigint	Código DIVIPOLA que permite identificar Departamento y Municipio
Geografia	DescripcionDepartamento	varchar(20)	Nombre del Departamento
Geografia	DescripcionMunicipio	varchar(30)	Nombre del Municipio
Seguimiento	ID_Caso	bigint	Identificador único del caso asignado al seguimiento
Seguimiento	ID_FechaReporteWeb	date	Fecha de publicación en sitio web
Seguimiento	ID_FechaNotificacion	date	Fecha de notificación a SIVIGILA
Seguimiento	ID_FechalnicioSintomas	date	Fecha de inicio de síntomas
Seguimiento	ID_FechaMuerte	date	Fecha de muerte
Seguimiento	ID_FechaDiagnostico	date	Fecha de confirmación por laboratorio
Seguimiento	ID_FechaRecuperacion	date	Fecha de recuperación
Seguimiento	TipoContagio	varchar(20)	Indica si el caso es Relacionado, Importado, En estudio, Comunitario

Seguimiento	Estado	varchar(20)	Corresponde a muertes no relacionadas con COVID-19, aún si eran casos activos **Hay pacientes recuperados para COVID-19, que pueden permanecer en hospitalización por otras comorbilidades
Seguimiento	Recuperado	varchar(20)	Recuperado Fallecido N/A (Vacío). N/A se refiere a los fallecidos no COVID. Pueden haber casos recuperados con ubicación Hospital u Hospital UCI, ya que permanecen en hospitalización por causas diferentes. Los casos con información en blanco en esta columna corresponde a los casos activos

2.3 MODELO DEL DOMINIO

Observación: Incluya el gráfico del modelo del dominio que representa la estructura de datos de su problema.

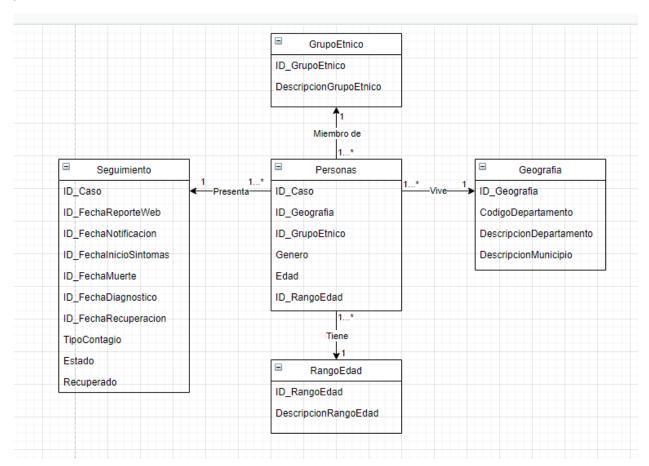


Imagen 1. Modelo del Dominio. Proyecto COVID - Antioquia. Realizado en: draw.io. Fuente: Autores.

3 MODELO ENTIDAD-RELACIÓN

3.1 TOMA DE PANTALLA DEL MODELO E-R

Observación: lo que se pide, puede usar https://draw.io o Microsoft Visio® y modele usando la notación de Barker.

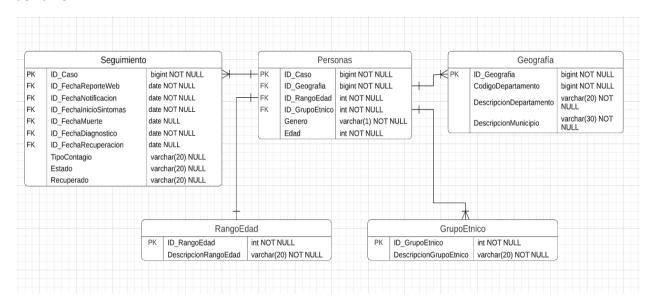


Imagen 2. Modelo E-R. Proyecto COVID - Antioquia. Realizado en: Lucidchart. Fuente: Autores. Disponible en: https://lucid.app/lucidchart/e0f61d2c-978a-478a-8c79-2c1daf64d436/edit?page=0 0#

3.2 SENTENCIA O CONSULTA DE CREACIÓN DEL TABLA(S)

Observación: Escriba el código en el Sistema de Gestión de Bases de Datos Relacionales de su elección (se recomienda SQLite por simplicidad, mediante https://sqlitebrowser.org/) para crear las tablas que corresponda con su conjunto de datos específico. Almacene en el repositorio https://github.com/viviposada/AGD2021-1S el script con el nombre de T1.3.2.Creacion_Tablas.sql

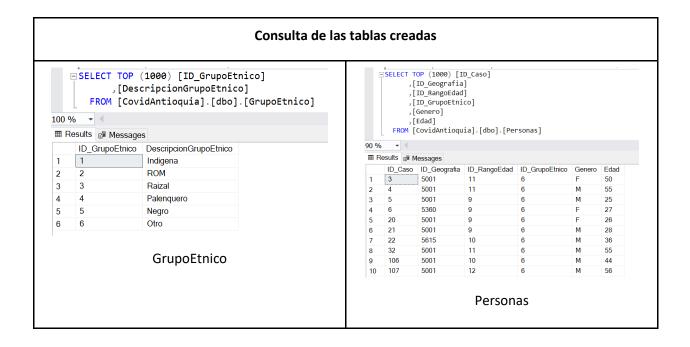
Se recomienda repasar SQL en https://www.w3schools.com/sql/default.asp

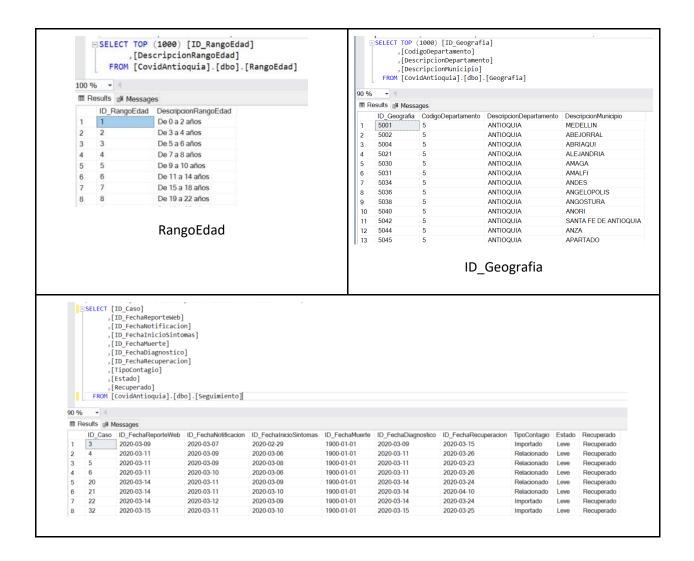


Con el uso del software de Lucidchart, luego de realizar la diagramación del modelo E-R fue posible obtener las líneas en SQL para la creación de tablas que fueron utilizadas en SQL Server. Se trabajó de manera local. Revisar anexo *T1.3.2.Creacion_Tablas.sql*

3.3 SENTENCIAS PARA INSERTAR DATOS

Observación: Escriba el código para insertar los datos en cada una de las tablas creadas. Almacene en el repositorio https://github.com/viviposada/AGD2021-15 el script con el nombre de T1.3.3.Insertar_Datos.sql



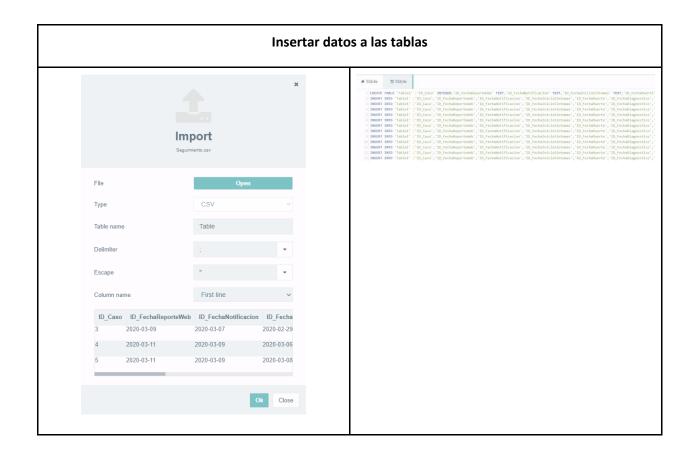


Las sentencias para insertar los datos a las tablas se pudieron obtener de dos maneras, la primera mediante el uso de SQL OnLine IDE desde donde es posible indicar el archivo que contiene la información objetivo, realizar su respectiva configuración de carga, al finalizar este proceso, la interfaz retorna las líneas necesarias para insertar los datos. Revisar anexo *T1.3.3.Insertar_Datos.sql*

Por otro lado también es posible hacer uso del SQLite3 donde con un ciclo for se itera cada una de las líneas que componen un archivo y de esta manera hacer la importación de los datos a las respectivas tablas. Revisar anexo *sqlite3 script.ipynb*

Para el caso particular del presente trabajo, se optó por hacer la carga mediante el primer método ya que era posible ejecutar estas mismas líneas en SQL Server y trabajar de manera local, ya que para el desarrollo de las posteriores consultas SQLite3 poseía ciertas restricciones en ciertas funciones lo que dificulta el desarrollo de queries.

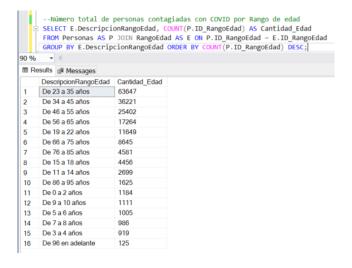
A continuación se muestra la evidencia de la navegación realizada por el primer método para obtener el query para poder insertar los datos a cada una de las tablas que componen al modelo.



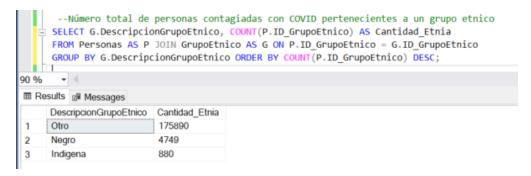
3.4 SENTENCIA DE CONSULTA

Observación: realice la exploración básica de los datos, conteos totales y por categorías, máximos, promedio y mínimos. Es decir, aplique estadística descriptiva con el fin de conocer las propiedades de los datos y entenderlos lo mejor posible. Use solamente sentencias SQL. Anexe las tomas de pantalla donde evidencie la sentencia SQL y su correspondiente ejecución. Además, Almacene en el https://github.com/viviposada/AGD2021-1S el script con el nombre de T1.3.4.Consultar_Datos.sql

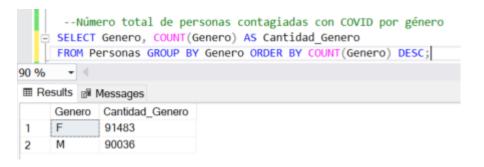
Número total de personas contagiadas con COVID por Rango de edad



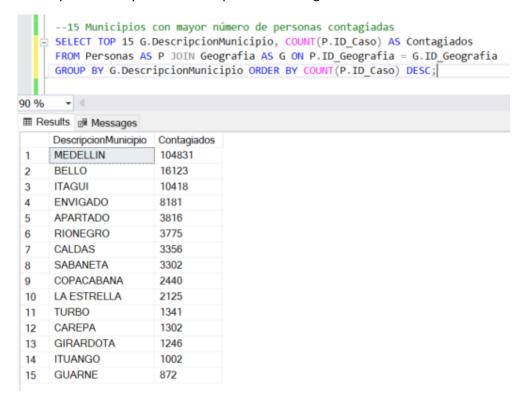
Número total de personas contagiadas con COVID pertenecientes a un grupo etnico



Número total de personas contagiadas con COVID por género



• 15 Municipios con mayor número de personas contagiadas



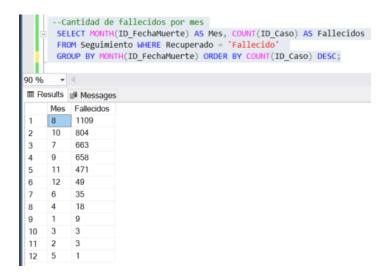
• Cantidad de contagiados por mes



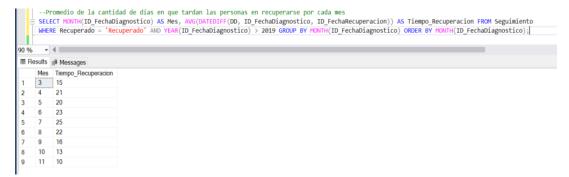
• Cantidad de recuperados por mes



• Cantidad de fallecidos por mes



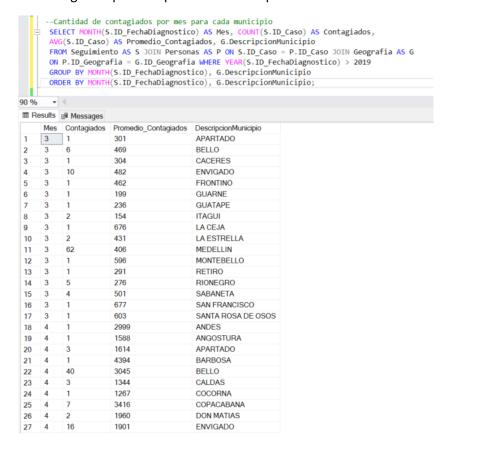
Promedio de la cantidad de días en que tardan las personas en recuperarse por cada mes



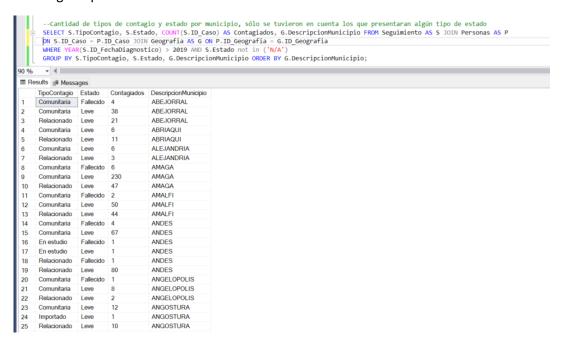
Promedio de la cantidad de días en que tardan las personas en fallecer por cada mes



Cantidad de contagiados por mes para cada municipio



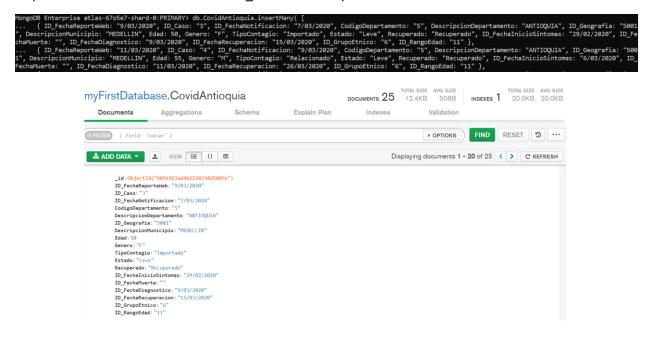
• Cantidad de tipos de contagio y estado por municipio, sólo se tuvieron en cuenta los que presentaran algún tipo de estado



4 MONGODB

4.1 SENTENCIA O CONSULTA DE CREACIÓN DEL DOCUMENTO(S)

Observación: Escriba el código en MongoDB para crear al menos 20 documentos que correspondan a su conjunto de datos específico. Almacene en el repositorio https://github.com/viviposada/AGD2021-15 el script con el nombre de T1.4.1.Creacion_Documentos.sql



4.2 SENTENCIA DE CONSULTA

Observación: Realice la exploración básica de los datos, conteos totales y por categorías, máximos, promedio y mínimos. Es decir, aplique estadística descriptiva con el fin de conocer las propiedades de los datos y entenderlos lo mejor posible. Use solamente sentencias SQL. Anexe las tomas de pantalla donde evidencie la sentencia SQL y su correspondiente ejecución. Además, *Almacene en el repositorio* https://github.com/viviposada/AGD2021-1S el script con el nombre de T1.4.2.Consultar_Datos.sql

Visualización de todos los elementos de la colección

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find().pretty()
          " id" : ObjectId("6091923ad4b2238334d508fe"),
          "ID_FechaReporteWeb" : "9/03/2020",
         "ID_Caso" : "3",
"ID_FechaNotificacion" : "7/03/2020",
"CodigoDepartamento" : "5",
          "DescripcionDepartamento": "ANTIOQUIA",
          "ID_Geografia" : "5001",
"DescripcionMunicipio" : "MEDELLIN",
          "Edad" : 50,
"Genero" : "F"
          Genero" : "F",
"TipoContagio" : "Importado",
          "Estado" : "Leve",
          "Recuperado" : "Recuperado",
"ID_FechaInicioSintomas" : "29/02/2020",
          "ID_FechaMuerte" : ""
         "ID_FechaDiagnostico": "9/03/2020",
          "ID FechaRecuperacion" : "15/03/2020",
          "ID_GrupoEtnico" : "6",
"ID_RangoEdad" : "11"
          "_id" : ObjectId("6091923ad4b2238334d508ff"),
          "ID_FechaReporteWeb" : "11/03/2020",
         "ID_Caso": "4",
"ID_FechaNotificacion": "9/03/2020",
"CodigoDepartamento": "5",
          "DescripcionDepartamento" : "ANTIOQUIA",
         "ID_Geografia" : "5001",
"DescripcionMunicipio" : "MEDELLIN",
"Edad" : 55,
"Genero" : "M",
"TipoContagio" : "Relacionado",
         "Estado" : "Leve",
"Recuperado" : "Recuperado",
"ID_FechaInicioSintomas" : "6/03/2020",
          "ID_FechaMuerte" : ""
          "ID_FechaDiagnostico": "11/03/2020".
          "ID_FechaRecuperacion" : "26/03/2020",
          "ID_GrupoEtnico" : "6",
"ID_RangoEdad" : "11"
```

Conteo del número de casos totales

Selección y cuenta de los contagiados en Medellín

```
WongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find( { DescripcionMunicipio: "MEDELLIN"} ).pretty()

"id": ObjectId("6091923ad4b2238334d508fe"),
"ID_FechaReporteWeb": "9/03/2020",
"ID_Gaso": 3",
"ID_FechaNotificacion": "7/03/2020",
"CodigoDepartamento": "5",
"DescripcionDepartamento": "5",
"DescripcionDepartamento": "ANTIOQUIA",
"ID_Geografia": "5001",
"DescripcionHunicipio": "MEDELLIN",
"Edad": 50,
"Genero": "F",
"ID_Contagio": "Importado",
"Estado": "Leve",
"Recuperado": "Recuperado",
"ID_FechaInicioSintomas": "29/02/2020",
"ID_FechaInicioSintomas": "29/02/2020",
"ID_FechaPecuperacion": "15/03/2020",
"ID_FechaPecuperacion": "15/03/2020",
"ID_FechaPecuperacion": "18/03/2020",
"ID_FechaPecuperacion": "18/03/2020",
"ID_FechaPecuperacion": "9/03/2020",
"ID_FechaPecuperacion": "9/03/2020",
"ID_FechaPecuperacion": "9/03/2020",
"ID_FechaPecuperacion": "9/03/2020",
"ID_FechaPecuperacion": "9/03/2020",
"ID_FechaPecuperacion": "9/03/2020",
"ID_FechaPecuperacion": "Soff,
"DescripcionDepartamento": "S",
"DescripcionDepartamento": "S",
"DescripcionDepartamento": "MEDELLIN",
"Edad": 55,
"Genero": "M",
"ID_FechaPecuperacion": "Relacionado",
"Estado": "Recuperado",
"ID_FechaPecuperacion": "1/03/2020",
"ID_FechaPecuperacion": "26/03/2020",
"ID_FechaPecuperacion": "26/03/2020",
"ID_FechaPecuperacion": "100, "26/03/2020",
"ID_FechaPecuperacion": "26/03/2020",
"ID_FechaPecuperacion": "26/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "26/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion": "11/03/2020",
"ID_FechaPecuperacion":
```

, MongoDB Enterprise atlas-6705e7-shard-0:PRIMARY> db.CovidAntioquia.find({ DescripcionMunicipio: "MEDELLIN"}).count()

Visualización del tipo de contagio, Rango de edad y grupo étnico para el municipio de Medellín

```
MongoDB Enterprise atlas-6705e7-shard-0:PRIMARY> db.CovidAntioquia.find(
... { DescripcionNunicipio: "MEDELLIN" },
... { TipoContagio: 1, ID_RangoEdad: 1, ID_GrupoEtnico: 1}
... }

{ "id": ObjectId("6091923ad4b2238334d508fe"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "11" }

{ "id": ObjectId("6091923ad4b2238334d508ff"), "TipoContagio": "Relacionado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "11" }

{ "id": ObjectId("6091923ad4b2238334d50900"), "TipoContagio": "Relacionado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "9" }

{ "id": ObjectId("6091923ad4b2238334d50900"), "TipoContagio": "Relacionado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "9" }

{ "id": ObjectId("6091923ad4b2238334d50906"), "TipoContagio": "Relacionado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "9" }

{ "id": ObjectId("6091923ad4b2238334d50906"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "11" }

{ "id": ObjectId("6091923ad4b2238334d50906"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "11" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "12" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "12" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "12" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "12" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "12" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "10" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "12" }

{ "id": ObjectId("6091923ad4b2238334d50908"), "TipoContagio": "Importado", "ID_GrupoEtnico": "6", "ID_RangoEdad": "10" }

{ "id": ObjectId("6091923ad4b2
```

Filtro y cuenta de los municipios diferentes de Medellín con casos de contagiados

```
.
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find( { DescripcionMunicipio: {$ne: "MEDELLIN"} } ).count()
8
```

• Visualización del tipo de contagio, Rango de edad y grupo étnico para los municipios diferentes de Medellín

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find(
... { DescripcionMunicipio: {$ne: "MEDELLIN"} },
... { TipoContagio: 1, ID_RangoEdad: 1, ID_GrupoEtnico: 1}
... }

{ "_id" : ObjectId("6091923ad4b2238334d50901"), "TipoContagio" : "Relacionado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "9" }

{ "_id" : ObjectId("6091923ad4b2238334d50904"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "10" }

{ "_id" : ObjectId("6091923ad4b2238334d50904"), "TipoContagio" : "Relacionado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "11" }

{ "_id" : ObjectId("6091923ad4b2238334d5090b"), "TipoContagio" : "Relacionado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "11" }

{ "_id" : ObjectId("6091923ad4b2238334d5090b"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "11" }

{ "_id" : ObjectId("6091923ad4b2238334d50911"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "11" }

{ "_id" : ObjectId("6091923ad4b2238334d50915"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "9" }

{ "_id" : ObjectId("6091923ad4b2238334d50915"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "9" }

{ "_id" : ObjectId("6091923ad4b2238334d50915"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "9" }

{ "_id" : ObjectId("6091923ad4b2238334d50915"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "9" }

{ "_id" : ObjectId("6091923ad4b2238334d50916"), "TipoContagio" : "Importado", "ID_GrupoEtnico" : "6", "ID_RangoEdad" : "9" }
```

Filtro de los contagiados con edades entre 56 y 65 y determinación de la cantidad

Filtro de los contagiados con edades superiores a 50 y determinación de la cantidad

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find(
                                                                                                                                                           { Edad: { $gt: 50 } }, ).pretty()
                "_id" : ObjectId("6091923ad4b2238334d508ff"),
               "_id": ObjectId("6091923ad4b223833dd508'
"ID_FechaReporteWeb": "11/03/2020",
"ID_Caso": "4",
"ID_FechaNotificacion": "9/03/2020",
"CodigoDepartamento": "5",
"DescripcionDepartamento": "ANTIOQUIA",
"ID_Geografia": "5001",
"DescripcionMunicipio": "MEDELLIN",
               "Edad" : 55,
"Genero" : "M",
"TipoContagio" : "Relacionado",
               "Estado" : "Leve",
"Recuperado" : "Recuperado",
"ID_FechaInicioSintomas" : "6/03/2020",
               "ID_FechaMuerte" : "",
"ID_FechaMuerte" : "11/03/2020",
"ID_FechaRecuperacion" : "26/03/2020",
               "ID_GrupoEtnico" : "6",
"ID_RangoEdad" : "11"
               " id" : ObjectId("6091923ad4b2238334d50905"),
                "ID FechaReporteWeb" : "15/03/2020",
               "ID_Caso" : "32",
               "ID_FechaNotificacion" : "11/03/2020",
"CodigoDepartamento" : "5",
                "DescripcionDepartamento" : "ANTIOQUIA",
               "ID_Geografia" : "5001",
"DescripcionMunicipio" : "MEDELLIN",
               "Edad" : 55,
"Genero" : "M",
"TipoContagio" : "Importado",
               "Tipocontagio : importado ;
"Estado" : "Leve",
"Recuperado" : "Recuperado",
"ID_FechaInicioSintomas" : "10/03/2020",
"ID_FechaMuerte" : "",
"ID_FechaDiagnostico" : "15/03/2020",
               "ID_Fechablaghostico : "15/03/2020",
"ID_FechaRecuperacion" : "25/03/2020",
"ID_GrupoEtnico" : "6",
"ID_RangoEdad" : "11"
```

MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find({ Edad: { \$gt: 50 } },).count() 7

Municipios con casos de contagiados con COVID reportados en Antioquia

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.aggregate( [ { $group : { _id : "$DescripcionMunicipio" } } ] )
{ "_id" : "GUARNE" }
{ "_id" : "ITAGUI" }
{ "_id" : "RIONEGRO" }
{ "_id" : "BELLO" }
{ "_id" : "BELLO" }
{ "_id" : "ENVIGADO" }
{ "_id" : "LA ESTRELLA" }
{ "_id" : "LA ESTRELLA" }
{ "_id" : "MEDELLIN" }
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY>
```

Tipos de Contagio reportados para las personas contagiadas en Antioquia

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.aggregate( [ { $group : { _id : "$TipoContagio" } } ] ) { "_id" : "Relacionado" } { "_id" : "Importado" }
```

Selección de los casos recuperados ordenados por edad con su respectiva cantidad

...
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find({ Recuperado: "Recuperado" }).sort({ Edad: -1 }).count()
25

Filtro de los contagiados para cada género (F:Femenino y M:Masculino)

Masculino

```
MongoDB Enterprise atlas-6705e7-shard-0:PRIMARY> db.CovidAntioquia.find( { Genero: "M" } ).pretty() {

"id": ObjectId("6091923ad4b2238334d508ff"),
"ID_FechaReporteWeb": "11/03/2020",
"ID_GeoporteWeb": "11/03/2020",
"OdigODepartamento": "S",
"DescripcionDepartamento": "ANTIOQUIA",
"ID_Geografia": "5001",
"DescripcionMunicipio": "MEDELLIN",
"Edad": 55,
"Genero": "M",
"TipoContagio": "Relacionado",
"Estado": "Leve",
"Recuperado": "Recuperado",
"ID_FechalnicioSintomas": "6/03/2020",
"ID_FechalnicioSintomas": "6/03/2020",
"ID_FechaBlagnostico": "11/03/2020",
"ID_FechaBlagnostico": "26/03/2020",
"ID_Geografia": "500",
"ID_FechaReporteWeb": "11/03/2020",
"ID_FechaReporteWeb": "11/03/2020",
"ID_FechaReporteWeb": "11/03/2020",
"ID_FechaReporteWeb": "11/03/2020",
"ID_Geografia": "500",
"ID_Geografia": "500",
"DoscripcionDepartamento": "S",
"DescripcionDepartamento": "ANTIOQUIA",
"ID_Geografia": "5001",
"DescripcionMunicipio": "MEDELLIN",
"Edad": 25,
"Genero": "M",
"TipoContagio": "Relacionado",
"Estado": "Relacionado",
"Estado": "Relacionado",
"Estado": "Recuperado": "Recuperado",
"Recuperado": "Recuperado",
"Recuperado": "Recuperado",
"Recuperado": "Recuperado",
"Recuperado": "Recuperado",
```

MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find({ Genero: "M" }).count()

Femenino

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find(
                                                                                       { Genero: "F" } ).pretty()
         "_id" : ObjectId("6091923ad4b2238334d508fe"),
        "ID_FechaReporteWeb" : "9/03/2020",
        "ID_Caso" : "3",
        "ID_FechaNotificacion" : "7/03/2020",
         "CodigoDepartamento" : "5"
         "CodigoDepartamento" : "5",
"DescripcionDepartamento" : "ANTIOQUIA",
         "ID_Geografia" : "5001",
"DescripcionMunicipio" : "MEDELLIN",
        "Edad" : 50,
"Genero" : "F"
         "TipoContagio" : "Importado",
         "Estado" : "Leve",
         "Recuperado" : "Recuperado"
         "ID_FechaInicioSintomas" : "29/02/2020",
         "ID_FechaMuerte" : ""
        "ID_FechaDiagnostico": "9/03/2020",
        "ID FechaRecuperacion": "15/03/2020",
        "ID_GrupoEtnico" : "6",
         "ID_RangoEdad" : "11"
         "_id" : ObjectId("6091923ad4b2238334d50901"),
        "ID_FechaReporteWeb" : "11/03/2020",
        "ID_Caso" : "6",
         "ID_FechaNotificacion": "10/03/2020",
        "CodigoDepartamento": "5",
"DescripcionDepartamento": "ANTIOQUIA",
         "ID_Geografia": "5360",
"DescripcionMunicipio": "ITAGUI",
         "Edad" : 27,
"Genero" : "F",
         "TipoContagio" : "Relacionado",
```

```
MongoDB Enterprise atlas-67o5e7-shard-0:PRIMARY> db.CovidAntioquia.find(  { Genero: "F" } ).count()
9
```

5 ANÁLISIS DE LECTURA

Observación: Considerando el artículo: "The Definitive Guide to Graph Databases for the RDBMS Developer" de Neo4J. Compartido en las carpeta de lecturas recomendadas. Analice y responda cada pregunta en máximo 150 palabras:

1. ¿Cuáles son las limitaciones, que se pueden inferir de la lectura, para migrar los conjuntos de datos relacionales a NoSQL?

Para migrar o realizar un cambio estructural de conjuntos de datos relacionales a NoSQL se pueden presentar algunas limitaciones como el tiempo que tarda la refactorización de la base de datos, el cual puede prolongarse hasta meses y generar tiempos de inactividad para el esquema de cambios. Además, cuando los conjuntos de datos relacionales no han sido diseñados u optimizados para exportar cantidades de datos grandes en cortos periodos de tiempo, el proceso de migración puede detenerse significativamente.

2. ¿Cuáles limitaciones adicionales se deben considerar, a parte de las mencionadas en el artículo?

Conocimiento del negocio y conocimientos técnicos: las personas del área de TI deben conocer el funcionamiento del negocio y la manera como hacer las consultas para responder a las necesidades; se debe saber en qué tabla están los datos y cómo hacer las uniones entre estas para obtener la información esperada.

Versión de los motores de base datos: cada vez que salga una nueva versión del motor de base de datos que se use, se debe tener el personal capacitado para el manejo de la herramienta.

Almacenamiento: el crecimiento de los datos puede ser exponencial, lo que implica inversión por parte del negocio para continuar almacenando la data que está en crecimiento constante.

Dependiendo de la infraestructura que se trabaje, si esta es on-premises, se debe tener presente el impacto de las migraciones de software en los servidores, que afectan la disponibilidad para acceder a las bases de datos.

3. ¿Cuáles son las razones (criterios) que se deben considerar para migrar un conjunto de datos relacionados a NoSQL?

Algunas de las razones que se deben considerar para realizar una migración de un conjunto de datos relacionados a NoSQL es cuando el volumen de los datos presenta un rápido crecimiento, que implique cambiar la estructura de los datos, o si se presentan cambios frecuentes de esquemas o se requieren realizar análisis en tiempo real. Asimismo, se debe considerar si los usuarios de la base de datos necesitan frecuentes modificaciones y la estructura de los datos no se adaptan a los nuevos cambios y problemáticas del negocio.

También debe considerarse, que si la base de datos es compleja, las consultas de ejecución no son suficientemente rápidas y generan incrementos en tiempos de procesamiento y en el consumo de recursos.

El negocio continuamente evoluciona y es necesario que el sistema lo haga con él, se requieren bases de datos flexibles que se adecuen a las necesidades cambiantes del negocio.