**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

## **Responsables**

|  |
| --- |
| **Nombre Completo – Documento de Identificación** |
| 1. **Juan Fernando Aldana Bohorquez - cc 1128435626** |
| 1. **Laura Isaza Echeverri - cc 1037600324** |
| 1. **Yuliana Franco Noreña cc 1001359126** |
| **REPO EN GITHUB:** **https://github.com/lisazae/Trabajo-Analitica-Grandes-Datos** |

# *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*](https://www.kaggle.com/datasets) *o* [*https://arxiv.org/*](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***](https://forms.gle/h7ty3yZykaUq5m7y6)

# **1** **Comprensión del negocio**

## **1.1** **Descripción del contexto del negocio.**

El negocio del BPO y Contact center es una tercerización para la atención de clientes, dentro de estas compañías existen gran cantidad de plataformas que permiten a los usuarios una navegación y diferentes opciones para la solución de sus requerimientos.

Regularmente el primer sistema de recepción el usuario es el IVR (*Interactive Voice Response*) el cual es un sistema de audio respuesta que permite al usuario navegar y auto atenderse, en este sistema se busca que el usuario pueda resolver sus inquietudes de manera autónoma, cada log de transacción queda registrado en bases de datos que para este trabajo llamaremos la base de IVR. El usuario no logra solucionar su inquietud es redireccionado con un agente o asesor el cual escuchará la solicitud del usuario y este deberá solucionarlo, todas las interacciones que realice el asesor quedarán registradas en una tabla que para este trabajo llamaremos tabla asesor, en la interacción del asesor este debe dejar registro de los motivos por los cuales el usuario generó la consulta, estas posibilidades de llamada quedan registradas en dos tablas una que se s códigos y otra que es productos, para este trabajo se usarán sólo interacciones de servicio al cliente, no se usarán servicios de salida, por tal razón en el diccionario pueden aparecer variables como no aplica para el problema.

## **1.2** **Identificación del problema:**

Identificar el comportamiento de los clientes que hacen uso del canal telefónico, teniendo en cuenta las interacciones que hayan realizado en un periodo de tiempo estimado, haciendo referencia a la cantidad de veces que se comunica al canal, cantidad de veces que lo hace por el mismo motivo, etc.… para así evaluar cómo mejorar la experiencia del cliente con respecto a la consultas o inconvenientes que tenga que lo hacen comunicarse a una línea específica para resolverlos, además de encontrar usuarios con comportamiento que generen alertas y poder generar estrategias focalizadas para así evitar el deterioro de la experiencia.

## **1.3** **Determinación de objetivos:**

se desea conocer patrones de necesidades que tienen los usuarios con el canal telefónico, entender esa relación clientes o usuarios, para mejorar experiencias en la navegabilidad y en los procesos diseñados por la compañía.

Conocer e identificar aquellos procesos de mayor complejidad que hacen que los usuarios no puedan resolver sus inquietudes en un único contacto.

Mejorar indicadores de satisfacción de usuarios y de resolución de problemas.

## **1.4** **Evaluación de la situación actual:**

Actualmente no se cuentan con modelos o estrategias que satisfagan este requerimiento, por lo que es necesario la implementación de modelos analíticos para dar solución a las preguntas planteadas y poder cumplir los objetivos propuestos inicialmente.

# **2** **Comprensión de los datos**

## **2.1** **recolección de datos**

Se tienen en cuenta 4 fuentes de datos los cuales corresponden a dos tipos de plataforma.

**plataforma 1:** Se compone de una tabla que corresponden al sistema de autoatención o

enrutamiento (IVR), esta tabla contiene los logs de cada interacción en el sistema de audio respuesta.

**plataforma 2:** Se compone de tres tablas, la primera contiene los logs de las interacciones que corresponden a la atención del cliente por medio de un asesor, la segunda es un diccionario que nos permite conocer el motivo por el cual se contacta el usuario y la tercera es una tabla que nos indica el producto por el cual el usuario está haciendo la consulta.

●  **Supuestos:**

* Transacciones con falta de información relevante como fecha, identificación del usuario y teléfono, no será tenida en cuenta.
* El proyecto apunta únicamente a sistemas transaccionales donde el usuario se autentique.

**se adiciona información de las 4 fuentes de datos en el repositorio de github en el archivo Datos.zip.**

## **2.2** **Descripción de datos (diccionario):**

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

**Tabla IVR**

|  |  |  |
| --- | --- | --- |
| Nombre del atributo / variable | Formato o Tipo de Dato | Descripción |
| COD\_INST | STRING | Código de institución a la cual pertenece el cliente |
| DESC\_INST | STRING | Descripción de institución a la cual pertenece el cliente |
| DIA\_LLAM | INTEGER | Dia en que se realizó la llamada |
| MES\_LLAM | INTEGER | Mes en que se realizó la llamada |
| ANO\_LLAM | INTEGER | Año en que se realizó la llamada |
| HORA\_TRAN | INTEGER | Hora en que se realizó la llamada |
| NRO\_TRAN | INTEGER | Consecutivo que indica el orden de las transacciones |
| COD\_TRAN | STRING | Código de la transacción que indica la acción que realiza el usuario |
| ID\_CLTE | STRING | Identificación del usuario |
| COD\_OPC\_MENU | INTEGER | Código que indica la opción que dio el usuario en el IVR |
| DESC\_OPC\_MENU | STRING | Descripción de la opción que dio el usuario en el IVR |
| CONNID | STRING | Identificación única de la llamada |
| DURA\_LLAM | INTEGER | Duración de la llamada |
| NRO\_ORIG\_LLAM | STRING | Numero de donde proviene la llamada |

## **Tabla Asesor**

|  |  |  |
| --- | --- | --- |
| Nombre del atributo / variable | Formato o Tipo de Dato | Descripción |
| CALL\_CONNID | STRING | Identificación única de la llamada |
| INI\_CALL | DATE | Fecha de la llamada |
| INI\_CALL\_2 | TIMESTAMP | Fecha de la llamada |
| REGLANEGOCIO | STRING | La ruta por la cual va a entrar el cliente para ser atendido. |
| CALL\_CODACT | INTEGER | Código que el agente tipifica del motivo de la llamada |
| AG\_USER\_NAME | STRING | Usuario de red del agente que atendió la llamada |
| PL\_EXTENSION | STRING | Número al cual se está comunicando el usuario (Un canal telefónico puede tener múltiples números de contacto. ) |
| CTI\_ID | STRING | Identificación del usuario |
| CTI\_NOMBRE | STRING | Nombre del usuario |
| CTI\_ANI | STRING | Numero de donde proviene la llamada |
| TRANSFER | STRING | Determina si la llamada hay o no transferencia |
| CALL\_CODEMP | STRING | Identificación del cliente en el sistema transaccional |
| CALL\_TYPE | INTEGER | Identifica el tipo de llamada si es salida o entrada |
| DUR\_CALL | INTEGER | Duración de la llamada |
| CALL\_CODPROD | INTEGER | Código que el agente tipifica asignando a un producto el motivo de la llamada |

## **Códigos**

|  |  |  |
| --- | --- | --- |
| Nombre del atributo / variable | Formato o Tipo de Dato | Descripción |
| CODACTIV | STRING | Código de tipificación de motivos de la llamada para ser usado por los agentes |
| DESCCOD | STRING | Descripción de Códigos de tipificación de motivos de la llamada para ser usado por los agentes |
| CODPROD | STRING | Código de tipificación de producto de la llamada para ser usado por los agentes |
| EFECTIVA\_S\_N | STRING | No aplica para el problema. |
| RPC\_S\_N | STRING | No aplica para el problema. |
| ENCONTRADO\_S\_N | STRING | No aplica para el problema. |
| FACTURABLE\_S\_N | STRING | No aplica para el problema. |
| INB2\_OUT3 | STRING | No aplica para el problema. |
| CODEMP | STRING | Identificación del cliente en el sistema transaccional |
| GESTIONADO | STRING | No aplica para el problema. |
| SPC | STRING | No aplica para el problema. |
| REGLA\_NEGOCIO | STRING | La ruta por la cual va a entrar el cliente para ser atendido. |

## 

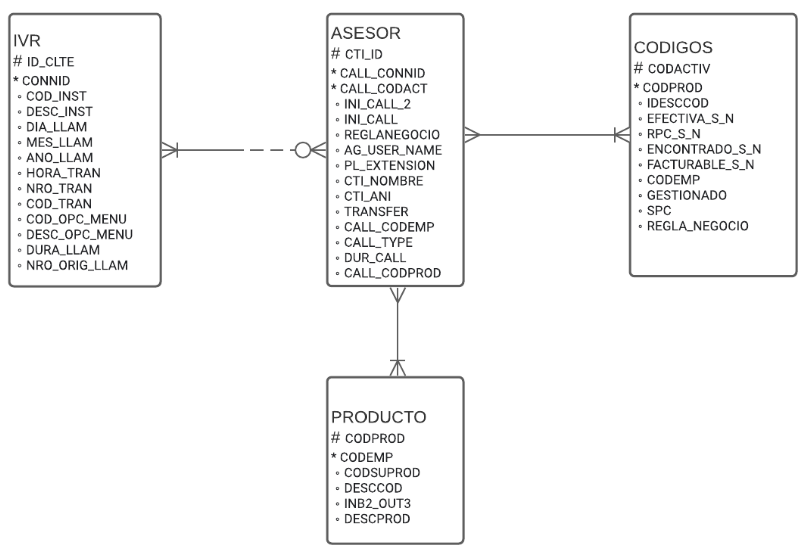
## **Productos**

|  |  |  |
| --- | --- | --- |
| Nombre del atributo / variable | Formato o Tipo de Dato | Descripción |
| CODPROD | STRING | Código del producto que identifica el producto por el cual el cliente generó una consulta en el canal. |
| CODSUPROD | STRING | No aplica para el problema. |
| DESCCOD | STRING | Descripción del producto por el cual el cliente generó una consulta en el canal. |
| CODEMP | STRING | Identificación del cliente en el sistema transaccional |
| INB2\_OUT3 | STRING | No aplica para el problema. |
| DESCPROD | STRING | No aplica para el problema. |

## **2.3** **Modelo de****l dominio**

# **3** **Modelo Entidad-Relación**

## **3.1** **Toma de pantalla del modelo E-R**



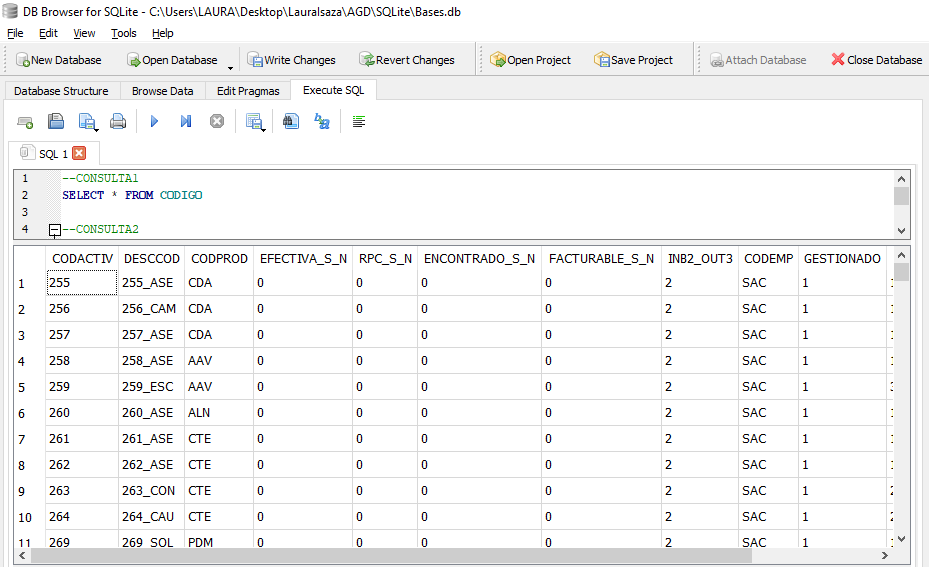
## **3.2** **Sentencia o consulta de creación de las tabla(s)**

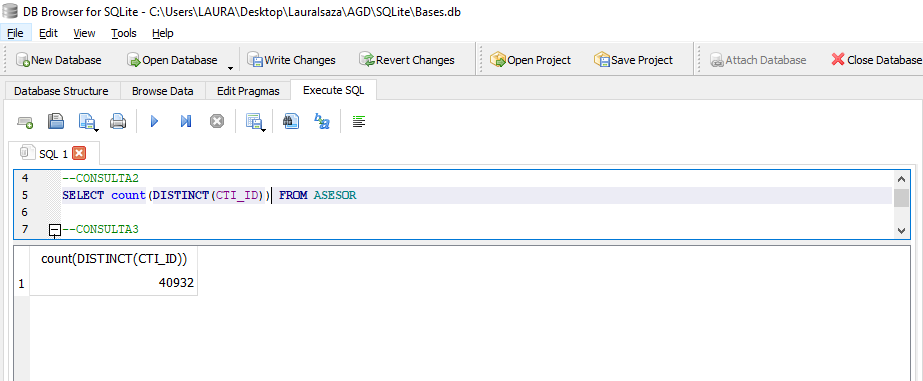
***Se subió el archivo al repo como estaba definido***

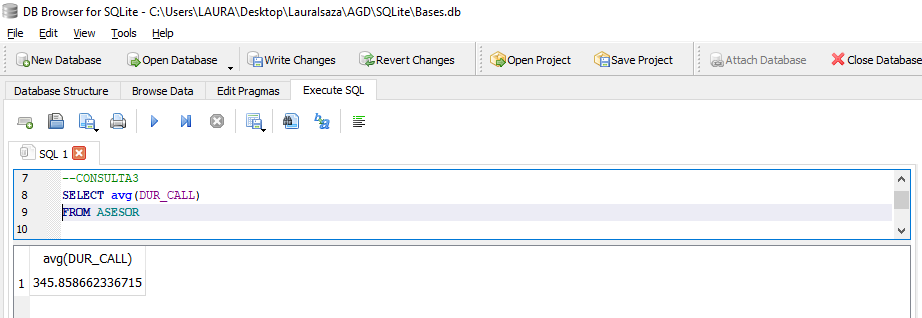
## **3.3** **Sentencias para Insertar datos**

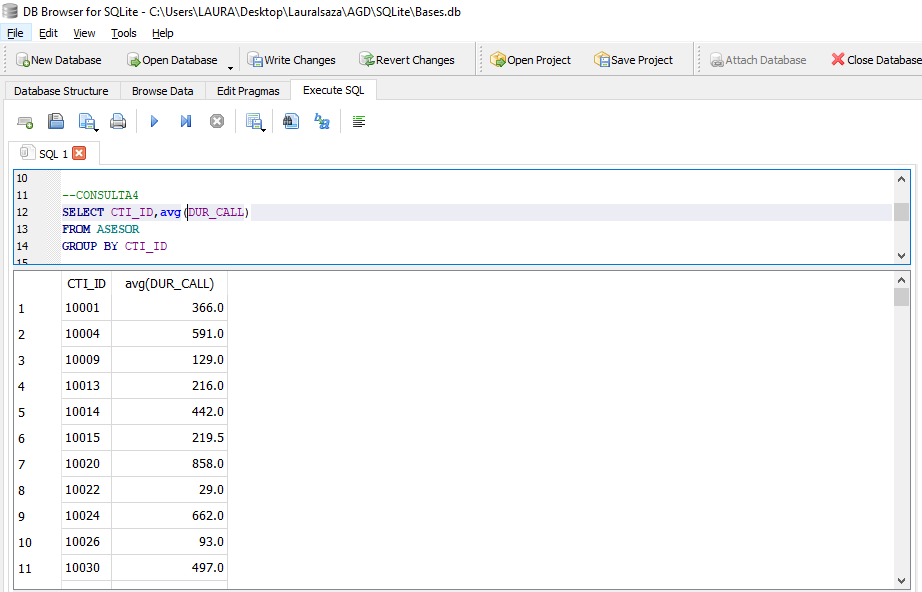
***Se subió el archivo al repo como estaba definido***

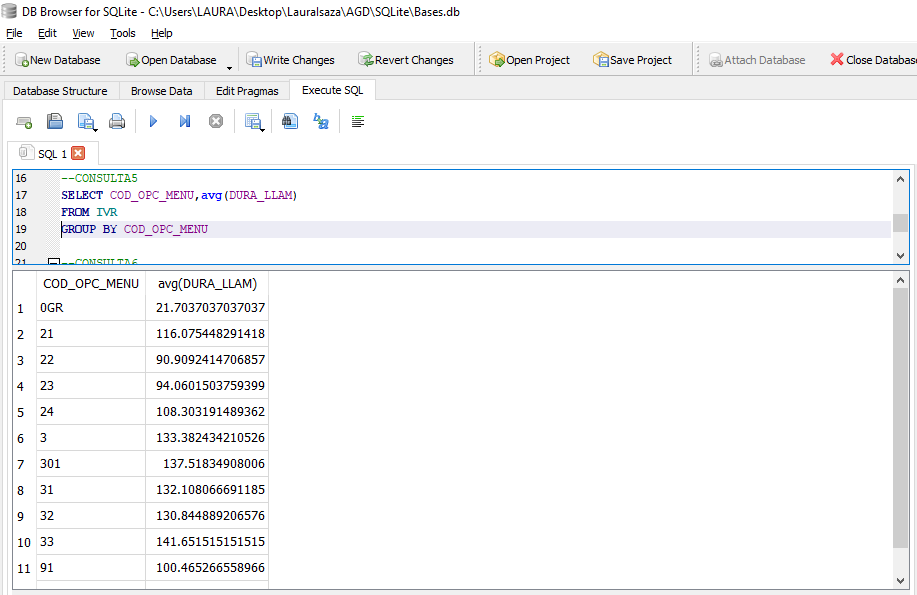
## **3.4** **Sentencia de consulta**

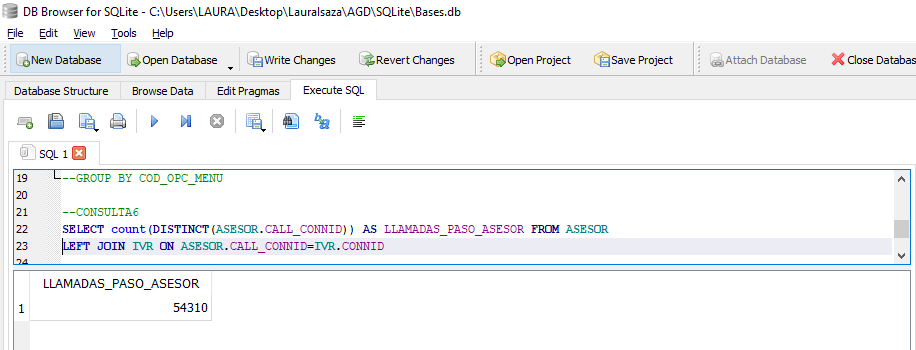
******

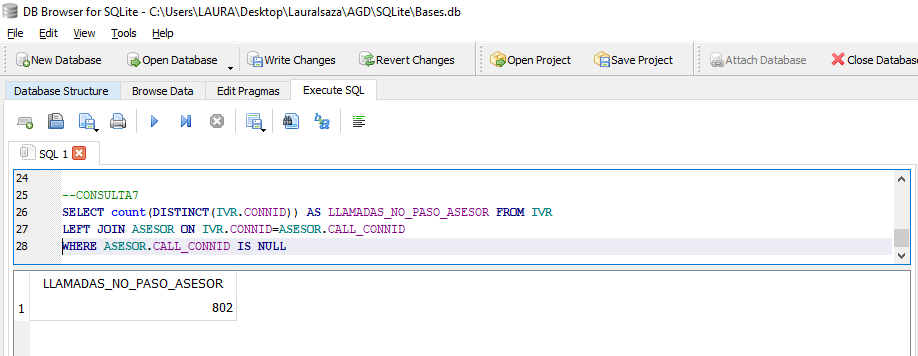










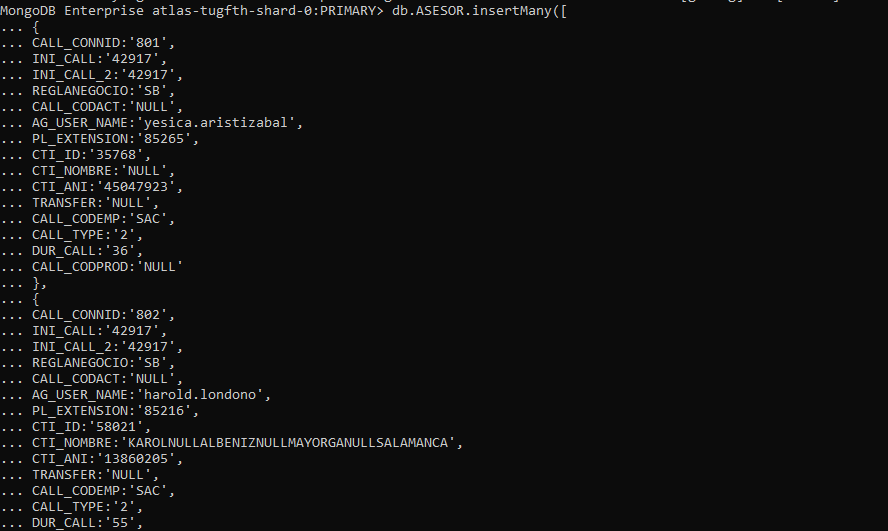


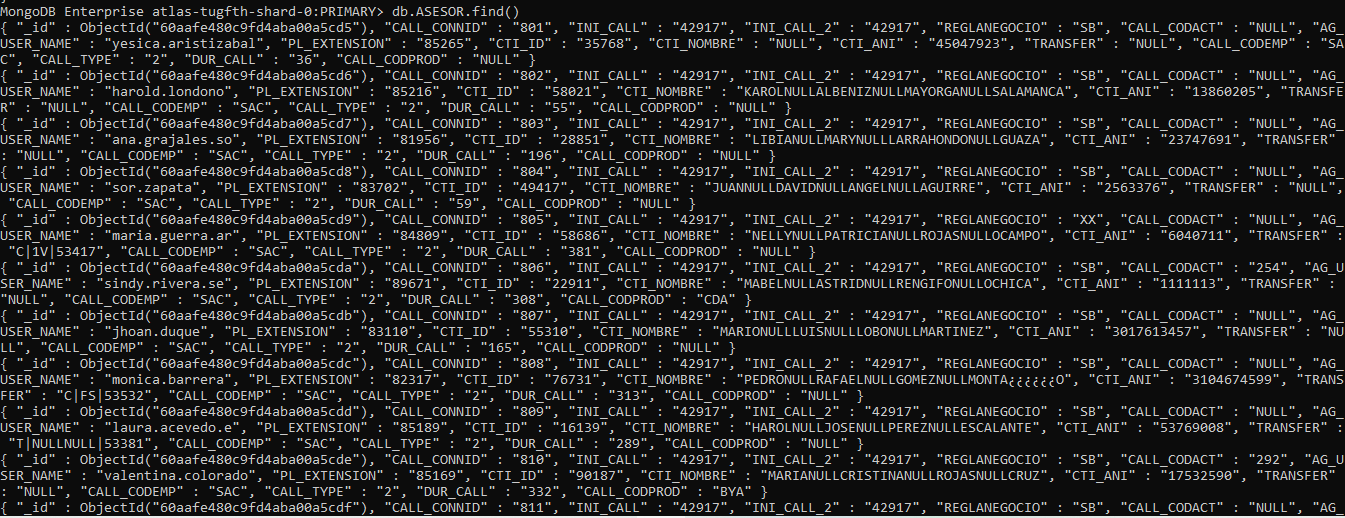
# **4** **MongoDB**

## **4.1** **Sentencia o consulta de creación del documento(s)**

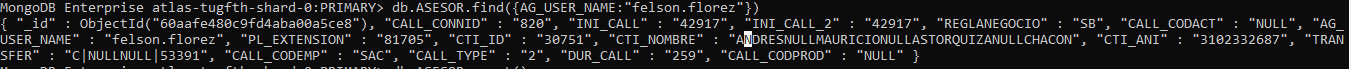
***Se subió el archivo al repo como estaba definido***

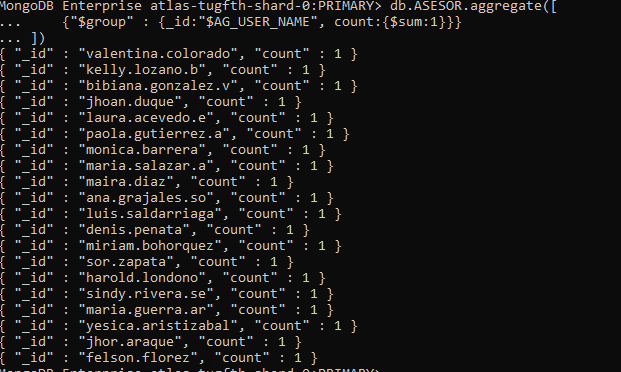
## **4.2** **Sentencia de consulta**











# **5** **Análisis de lectura**

***Observación****: Considerando el artículo: “*The Definitive Guide to Graph Databases for the RDBMS Developer*” de Neo4J. Compartido en la 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?

En la actualidad se quiere que los datos no sean solo un constante flujo de almacenamiento entre filas y columnas, sino que se pueda capturar verdaderas relaciones entre diferentes tipos de datos que se podrían ver como objetos. una de las limitaciones para pasar a un sistema de datos NoSQL que si permita la captura de estas relaciones que la industria demanda es que las bases existentes relacionales no se adaptan al cambio y esto implica un alto mantenimiento y costo, además que se puede volver un proceso lento que en épocas de agilismo en el desarrollo no lo hace viable. Otro aspecto a tener en cuenta es que las bases relacionales ya tienen su uso en particular y funcionan bien para ese uso, llevarlos a un NoSQL implica costos computacionales en consultas simples que pueden afectar el rendimiento al momento de generar información de valor.

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

* **Esquemas dinámicos:** En las bases de datos relacionales es necesario conocer de antemano que tipo de dato se va a almacenar y las bases de datos NoSQL están diseñadas para que se puedan insertar datos sin un esquema predefinido.
* Las bases de datos relacionales solo permiten un escalamiento de forma vertical y las bases de datos NoSQL distribuyen los datos de forma automática. (escalado horizontal)

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

* Se debe pensar en la migración cuando se desean usar estos datos para tomar decisiones, al instante, la creación de documentos en sistemas transaccionales hace que sea más fácil la búsqueda y el uso.
* se debe considerar migración de datos relacional a NoSQL cuando se están generando grandes volúmenes de información y se desea que este sea usado para temas analíticos.