

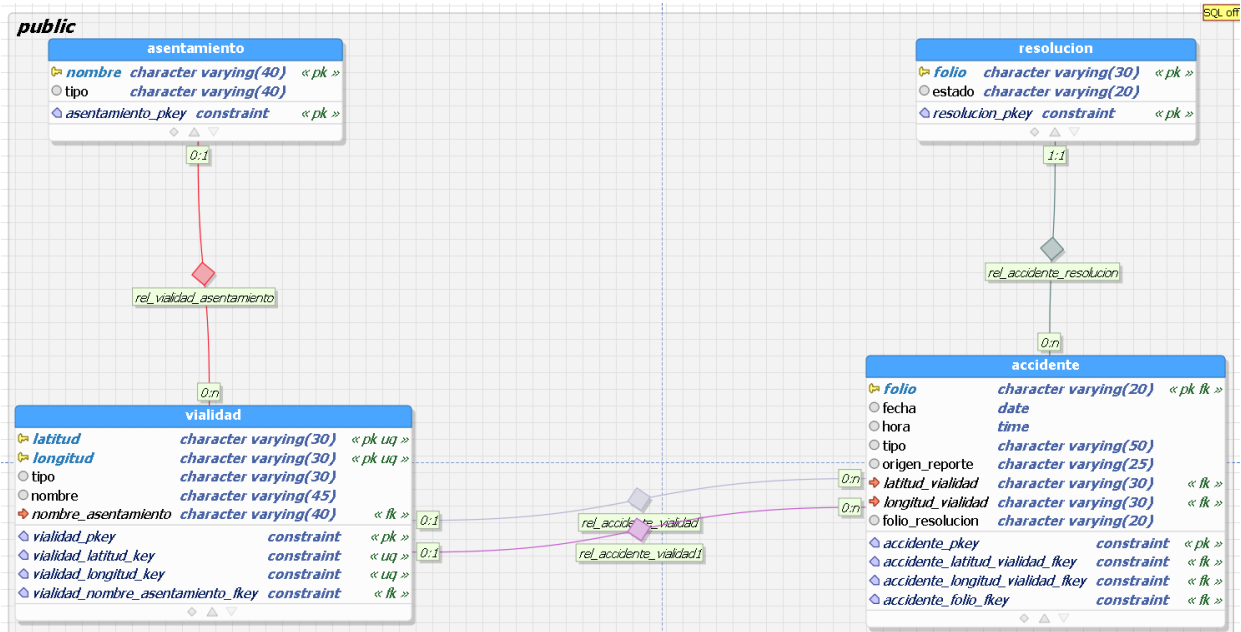
Universidad del Rosario

Proyecto final - Ingeniería de datos

Juan Camilo Pugliese y Juan Andrés Castro
semestre 2023-1

Entrega 3

1. Modelo relacional normalizado en tercera forma normal.



2. Descripción del proceso de carga de información en la base de datos.

- Para la carga masiva de datos, primero se normalizó la base de datos, eliminando valores nulos, errores en los registros que no correspondían con el tipo de dato, y columnas repetidas. Posterior a esto se hizo un documento de Excel para cada tabla correspondiente. Para llevar a cabo la carga masiva, se debe descargar esta carpeta que en el repositorio tiene el nombre de Carga_Masiva_F.zip, después, se ejecuta en PostgreSQL el script de las tablas y se cargan los archivos csv con el comando copy desde PostgreSQL.

3. Descripción de análisis identificados.

Encontramos 8 formas de ver la base de datos encontrando 8 posibles escenarios de análisis.

- Podemos observar la cantidad y los detalles de los accidentes automovilísticos que ocurren durante un período de tiempo para ver si hay algún patrón. Usando las tablas accidente, específicamente las columnas de hora y tipo podemos determinar esto.
- Podemos observar los accidentes por tipo y ver en que tipo de vialidad ocurrieron estos accidentes en búsqueda de ver en que vialidades ocurren que accidente, y en que vialidad ocurre el tipo de accidente mas frecuente. Esto lo podemos conseguir usando la tabla accidente, la columna tipo_accidente, y, usando en la tabla vialidad, la columna tipo_vialidad. Estas se pueden relacionar debido a las llaves foráneas de latitud y longitud en la tabla accidente.
- Podemos averiguar la cantidad de accidentes por asentamiento, definiendo en que asentamiento ocurren mas accidentes. Esto se puede determinar por

medio de la tabla asentamiento, mas específicamente la columna nombre que se relaciona posteriormente en la tabla vialidad en la columna nombre_asentamiento como llave foránea. Esta tabla también se relaciona con accidente logrando así el resultado final del escenario.

- d. Se puede determinar la cantidad de accidentes según su tipo y que cantidad de estos se resuelven. Esto usando la columna tipo y folio_resolución de la tabla accidente.

e.

4. URL del repositorio GIT del proyecto.

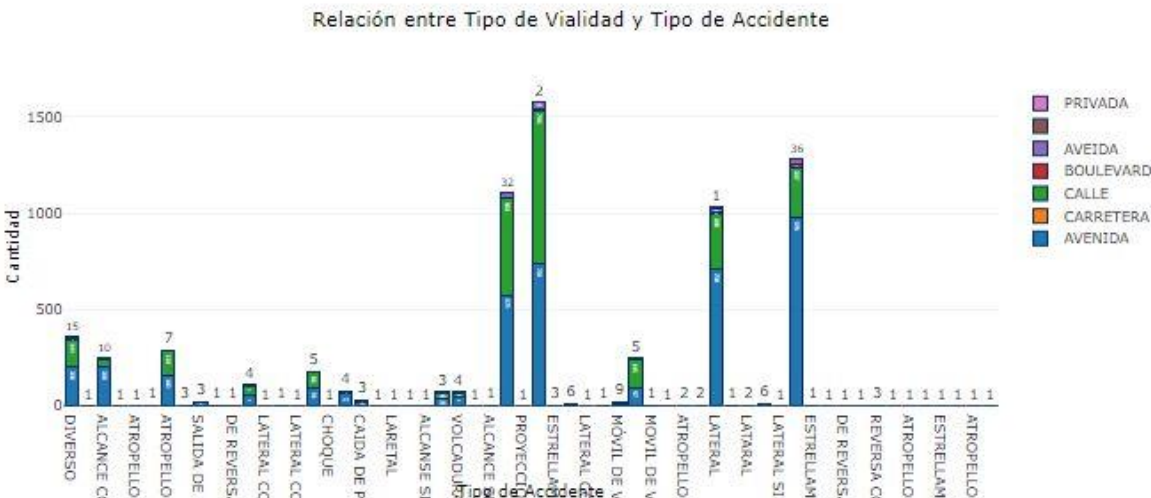
- a. https://github.com/juanandrescastro/Base_accidentes_viales

5. Escenarios en DASH.

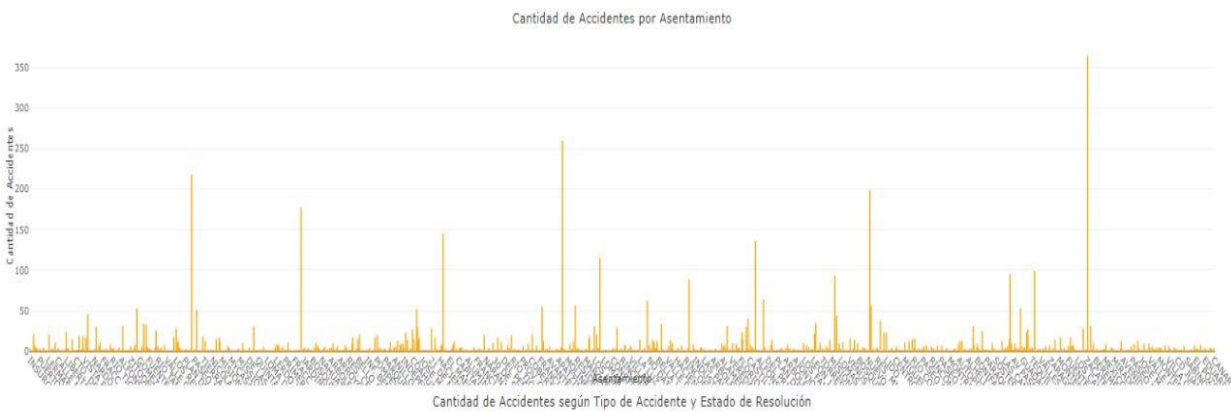
1.



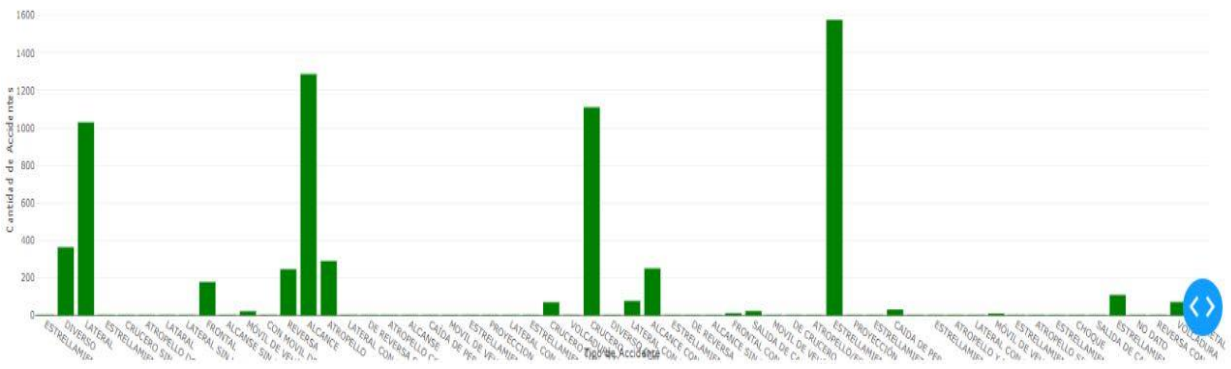
2.



3.



4.



6. Sección de discusión.

- a. Viendo los análisis desarrollados vemos que es complicado para una cantidad bastante grande de datos tener distintos tipos de representación gráfica. En nuestro caso, la grafica de barras fue la mas efectiva ya que era

la única en la que se podía determinar una diferencia de algunos datos en específico que resaltaban entre el resto, esto nos hizo bastante fácil ver que dato se lleva la primera posición en el estudio correspondiente y cuales si tenían una cantidad considerable. Sin embargo, también notamos que algunos datos llegaban a un nivel tan ínfimo que casi ni se puede determinar bien o considerar como un dato relevante en la tabla.

7. Sección de Conclusiones.

- a. Durante el proyecto, tuvimos la oportunidad de trabajar con fuentes de datos reales, lo cual nos brindó una experiencia práctica y significativa.
- b. Identificamos y comprendimos las reglas de negocio asociadas a las fuentes de datos seleccionadas.
- c. La construcción del diagrama Entidad-Relación y del diagrama relacional nos permitió tener una representación clara de la estructura de la base de datos.
- d. La normalización de la base de datos en tercera forma normal garantiza la eficiencia y la integridad de los datos almacenados.
- e. La carga de los datos en la base de datos nos permitió trabajar con una cantidad muy grande de datos, enfrentándonos a la dificultad de manejar tal magnitud.
- f. Desarrollamos escenarios de análisis utilizando Dash, lo cual nos permitió visualizar los resultados de manera interactiva.
- g. Las visualizaciones desarrolladas en Dash revelaron patrones interesantes y hallazgos inesperados en los datos.
- h. Experimentamos cómo la correcta estructuración y visualización de los datos pueden proporcionar información valiosa y facilitar la toma de decisiones.
- i. Encontramos desafíos en el proceso de búsqueda y selección de fuentes de datos reales debido a la disponibilidad y accesibilidad limitadas de los datos.
- j. Identificar las reglas de negocio fue un proceso complejo que requirió un análisis exhaustivo de los conjuntos de datos seleccionados.